Research Reports from the Communications Research Laboratory
at Ilmenau University of Technology

Advanced Array Signal Processing Algorithms for Multi-Dimensional Parameter Estimation

Jens Steinwandt

ILMENAU UNIVERSITY OF TECHNOLOGY

Fakultät für Elektrotechnik und Informationstechnik
der Technischen Universität Ilmenau

Advanced Array Signal Processing Algorithms for Multi-Dimensional Parameter Estimation

Jens Steinwandt

Dissertation zur Erlangung des
akademischen Grades Doktor-Ingenieur (Dr.-Ing.)

Anfertigung im:	Fachgebiet Nachrichtentechnik Institut für Informationstechnik Fakultät für Elektrotechnik und Informationstechnik
Gutachter:	Univ.-Prof. Dr.-Ing. Martin Haardt
	Univ.-Prof. Dr.-Ing. Marius Pesavento
	Prof. Dr. Sergiy A. Vorobyov
Vorgelegt am:	09.04.2018
Verteidigt am:	30.11.2018

urn:nbn:de:gbv:ilm1-2018000614

Acknowledgements

There were many people who have contributed to the completion of the work in this thesis. I would like to take this opportunity to thank them.

First and foremost, I would like to express my appreciation and deepest gratitude to my supervisor and mentor Prof. Martin Haardt for giving me the opportunity to work under his supervision. His constant support, encouragement, strive for perfection, and numerous ideas have inspired me throughout the years of pursuing my doctoral degree. The many fruitful discussions with him, his great attention to detail, and his immense knowledge have continuously motivated me to go beyond what I thought was possible and made me become the researcher I am today. This thesis would not have been possible without his guidance and supervision.

I would also like to sincerely thank Prof. Marius Pesavento and Prof. Sergiy Vorobyov for the fruitful cooperation, their brilliant ideas in research, their valuable advice, and their friendship during the recent years. All the mutual discussions were very helpful in continuously improving my research work. I am also very grateful to both of them for volunteering to review the thesis.

Furthermore, I want to give special thanks to Dr. Florian Roemer, who has introduced me to the research topic of my thesis. His invaluable guidance, his genius as well as his knowledge and skills have very much inspired me and motivated me to discover my potential. I also thank him for helping me out when help was needed and for being very good friends. Special thanks also goes to Dr. Christian Steffens, and Dr. Mike Wolf for their support, encouragement, and the many inspiring conversations on research and the less profound things in life.

I also owe a debt of gratitude to my former and current colleagues Kristina Naskovska, Sher Ali Cheema, Bilal Zafar, Mikus Grasis, Jianshu Zhang, Yao Cheng, and Marko Hennhöfer for countless technical or personal discussions, for exchanging ideas, sharing their knowledge and skills, and supporting me throughout the years. I was very lucky to have such helpful and friendly colleagues and such a wonderful working environment. Moreover, I want to thank Christina Patotschka and Wolfgang Erdtmann for providing optimism and administrative help whenever needed.

Last but not least, I would like to thank my mother and my family for their tremendous support in many ways. Their love, patience, and sacrifices will always be an inspiration for me. I would also like to thank my mother for supporting my education, discovering my abilities, and for encouraging me to be curious, open-minded, and inquiring. None of all this would have been possible if you had not believed in me and motivated me to discover my full potential.

Abstract

Multi-dimensional high-resolution parameter estimation is a fundamental problem in a variety of array signal processing applications, including radar, mobile communications, multiple-input multiple-output (MIMO) channel estimation, and biomedical imaging. The objective is to estimate the frequency parameters of noise-corrupted multi-dimensional harmonics that are sampled on a multi-dimensional grid. Among the proposed parameter estimation algorithms to solve this problem, multi-dimensional (R-D) ESPRIT-type algorithms have been widely used due to their computational efficiency and their simplicity. Their performance in various scenarios has been objectively evaluated by means of an analytical performance assessment framework. Recently, a relatively new class of parameter estimators based on sparse signal reconstruction has gained popularity due to their robustness under challenging conditions such as a small sample size or strong signal correlation. A common approach towards further improving the performance of parameter estimation algorithms is to exploit prior knowledge on the structure of the signals. In this thesis, we develop enhanced versions of R-D ESPRIT-type algorithms and the relatively new class of sparsity-based parameter estimation algorithms by exploiting the multi-dimensional structure of the signals and the statistical properties of strictly non-circular (NC) signals.

First, we derive analytical expressions for the gain from forward-backward averaging (FBA) and tensor-based processing in R-D ESPRIT-type and R-D Tensor-ESPRIT-type algorithms for the special case of two sources. This is accomplished by simplifying the generic analytical MSE expressions from the performance analysis of R-D ESPRIT-type algorithms. The analytical expressions are analyzed in terms of the physical parameters, e.g., the number of sensors, the signal correlation, and the source separation, to identify the scenarios for which the maximum gain or no gain is achieved.

Second, we propose the generalized least squares (GLS) algorithm to solve the overdetermined shift invariance equation in R-D ESPRIT-type algorithms. GLS directly incorporates the statistics of the subspace estimation error into the shift invariance solution through its covariance matrix, which is found via a first-order perturbation expansion. To objectively assess the estimation accuracy, we derive performance analysis expressions for the mean square error (MSE) of GLS-based ESPRIT-type algorithms, which are asymptotic in the effective SNR, i.e., the results become exact for a high SNR or a small sample size. Based on the performance analysis, we show that the simplified MSE expressions of GLS-based 1-D ESPRIT-type algorithms for a single source and two sources can be transformed into the corresponding Cramér-Rao bound (CRB) expressions, which provide a lower limit on the estimation error. Thereby, ESPRIT-type algorithms can become asymptotically efficient, i.e., they asymptotically achieve the CRB. Numerical simulations show that this can also be the case for more than two sources.

In the third contribution, we derive matrix-based and tensor-based R-D NC ESPRIT-type algorithms for multi-dimensional strictly non-circular signals, where R-D NC Tensor-ESPRIT-type algorithms exploit both the multi-dimensional structure and the strictly non-circular structure of the signals. Exploiting the NC signal structure by means of a preprocessing step leads to a virtual doubling of the original sensor array, which provides an improved estimation accuracy and doubles the number of resolvable signals. We derive an analytical performance analysis and compute simplified MSE expressions for a single source and two sources. These expressions are used to analytically compute the NC gain for these cases, which has so far only been studied via Monte-Carlo simulations. Moreover, the NC gain is analyzed with respect to the rotation phase separation and the correlation of the two signals to find the scenarios that provide the maximum gain. We additionally consider spatial smoothing preprocessing for R-D ESPRIT-type algorithms, which has been widely used to improve the estimation performance for highly correlated signals or a small sample size. Once more, we derive performance analysis expressions for R-D ESPRIT-type algorithms and their corresponding NC versions with spatial smoothing and derive the optimal number of subarrays for spatial smoothing that minimizes the MSE for a single source.

In the next part, we focus on the relatively new concept of parameter estimation via sparse signal reconstruction (SSR), in which the sparsity of the received signal power spectrum in the spatio-temporal domain is exploited. We develop three NC SSR-based parameter estimation algorithms for strictly non-circular sources and show that the benefits of exploiting the signals' NC structure can also be achieved via sparse reconstruction. We develop two grid-based NC SSR algorithms with a low-complexity off-grid estimation procedure, and a gridless NC SSR algorithm based on atomic norm minimization.

As the final contribution of this thesis, we derive the deterministic R-D NC CRB for strictly non-circular sources, which serves as a benchmark for the presented R-D NC ESPRIT-type algorithms and the NC SSR-based parameter estimation algorithms. We show for the special cases of, e.g., full coherence, a single snapshot, or a single strictly non-circular source, that the deterministic R-D NC CRB reduces to the existing deterministic R-D CRB for arbitrary signals. Therefore, no NC gain can be achieved in these cases. For the special case of two closely-spaced NC sources, we simplify the NC CRB expression and compute the NC gain for two closely-spaced NC signals. Finally, its behavior in terms of the physical parameters is studied to determine the parameter settings that provide the largest NC gain.

Zusammenfassung

Die hochauflösende Parameterschätzung für mehrdimensionale Signale findet Anwendung in vielen Bereichen der Signalverarbeitung in Mehrantennensystemen. Zu den Anwendungsgebieten zählen beispielsweise Radar, die Mobilkommunikation, die Kanalschätzung in *multiple-input multiple-output* (MIMO)-Systemen und bildgebende Verfahren in der Biosignalverarbeitung. In letzter Zeit sind eine Vielzahl von Algorithmen zur Parameterschätzung entwickelt worden, deren Schätzgüte durch eine analytische Beschreibung der Leistungsfähigkeit objektiv bewertet werden kann. Eine verbreitete Methode zur Verbesserung der Schätzgenauigkeit von Parameterschätzverfahren ist das Ausnutzen von Vorwissen bezüglich der Signalstruktur. In dieser Arbeit werden mehrdimensionale ESPRIT-Verfahren als Beispiel für Unterraum-basierte Verfahren entwickelt und analysiert, die explizit die mehrdimensionale Signalstruktur mittels Tensorsignalverarbeitung ausnutzen und die statistischen Eigenschaften von nicht-zirkulären Signalen einbeziehen. Weiterhin werden neuartige auf Signalrekonstruktion basierende Algorithmen vorgestellt, die die nicht-zirkuläre Signalstruktur bei der Rekonstruktion ausnutzen. Die vorgestellten Verfahren ermöglichen eine deutlich verbesserte Schätzgüte und erhöhen die Anzahl der auflösbaren Signale. Die Vielzahl der Forschungsbeiträge in dieser Arbeit setzt sich aus verschiedenen Teilen zusammen.

Im ersten Teil wird die analytische Beschreibung der Leistungsfähigkeit von Matrix-basierten und Tensor-basierten ESPRIT-Algorithmen betrachtet. Die Tensor-basierten Verfahren nutzen explizit die mehrdimensionale Struktur der Daten aus. Es werden für beide Algorithmenarten vereinfachte analytische Ausdrücke für den mittleren quadratischen Schätzfehler für zwei Signalquellen hergeleitet, die lediglich von den physikalischen Parametern, wie zum Beispiel die Anzahl der Antennenelemente, das Signal-zu-Rausch-Verhältnis, oder die Anzahl der Messungen, abhängen. Ein Vergleich dieser Ausdrücke ermöglicht die Berechnung einfacher Ausdrücke für den Schätzgenauigkeitsgewinn durch den *forward-backward averaging* (FBA)-Vorverarbeitungsschritt und die Tensorsignalverarbeitung, die die analytische Abhängigkeit von den physikalischen Parametern enthalten.

Im zweiten Teil entwickeln wir einen neuartigen *general least squares* (GLS)-Ansatz zur Lösung der Verschiebungs-Invarianz-Gleichung, die die Grundlage der ESPRIT-Algorithmen darstellt. Der neue Lösungsansatz berücksichtigt die statistische Beschreibung des Schätzfehlers bei der Unterraumschätzung durch dessen Kovarianzmatrix und ermöglicht unter bestimmten Annahmen eine optimale Lösung der Invarianz-Gleichung. Mittels einer Performanzanalyse der GLS-basierten ESPRIT-Verfahren und der Vereinfachung der analytischen Ausdrücke für den Schätzfehler für eine Signalquelle und zwei zeitlich unkorrelierte Signalquellen wird gezeigt, dass die Cramér-Rao-Schranke, eine untere Schranke für die Varianz eines Schätzers, erreicht werden kann.

Im nächsten Teil werden Matrix-basierte und Tensor-basierte ESPRIT-Algorithmen für nicht-zirkuläre Signale vorgestellt. Unter Ausnutzung der Signalstruktur gelingt es, die Schätzgenauigkeit

zu erhöhen und die doppelte Anzahl an Quellen aufzulösen. Dabei ermöglichen die vorgeschlagenen Tensor-ESPRIT-Verfahren sogar die gleichzeitige Ausnutzung der mehrdimensionalen Signalstruktur und der nicht-zirkuläre Signalstruktur. Die Leistungsfähigkeit dieser Verfahren wird erneut durch eine analytische Beschreibung objektiv bewertet und Spezialfälle für eine und zwei Quellen betrachtet. Es zeigt sich, dass für eine Quelle keinerlei Gewinn durch die nicht-zirkuläre Struktur erzielt werden kann. Für zwei nicht-zirkuläre Quellen werden vereinfachte Ausdrücke für den Gewinn sowohl im Matrixfall also auch im Tensorfall hergeleitet und die Abhängigkeit von den physikalischen Parametern analysiert. Sind die Signale stark korreliert oder ist die Anzahl der Messdaten sehr gering, kann der *spatial smoothing*-Vorverarbeitungsschritt mit den verbesserten ESPRIT-Verfahren kombiniert werden. Anhand der Performanzanalyse wird die Anzahl der Mittelungen für das *spatial smoothing*-Verfahren analytisch für eine Quelle bestimmt, die den Schätzfehler minimiert.

Der nächste Teil befasst sich mit einer vergleichsweise neuen Klasse von Parameterschätzern, die auf der Rekonstruktion überlagerter dünnbesetzter Signale basiert. Als Vorteil gegenüber den Algorithmen, die eine Signalunterraumschätzung voraussetzen, sind die Rekonstruktions-Verfahren verhältnismäßig robust im Falle einer geringen Anzahl zeitlicher Messungen oder einer starken Korrelation der Signale. In diesem Teil der vorliegenden Arbeit werden drei solcher Verfahren entwickelt, die bei der Rekonstruktion zusätzlich die nicht-zirkuläre Signalstruktur ausnutzen. Dadurch kann auch für diese Art von Verfahren eine höhere Schätzgenauigkeit erreicht werden und eine höhere Anzahl an Signalen rekonstruiert werden.

Im letzten Kapitel der Arbeit wird schließlich die Cramér-Rao-Schranke für mehrdimensionale nicht-zirkuläre Signale hergeleitet. Sie stellt eine untere Schranke für den Schätzfehler aller Algorithmen dar, die speziell zur Ausnutzung dieser Signalstruktur entwickelt wurden. Im Vergleich zur bekannten Cramér-Rao-Schranke für beliebige Signale, zeigt sich, dass im Fall von zeitlich kohärenten Signalen, für einen Messvektor oder für nur eine Quelle, beide Schranken äquivalent sind. In diesen Fällen kann daher keine Verbesserung der Schätzgüte erzielt werden. Zusätzlich wird die Cramér-Rao-Schranke für zwei benachbarte nicht-zirkuläre Signalquellen vereinfacht und der maximal mögliche Gewinn in Abhängigkeit der physikalischen Parameter ermittelt. Dieser Ausdruck gilt als Maßstab für den erzielbaren Gewinn durch die Anwendung der ESPRIT-Verfahren für zwei nicht-zirkuläre Signalquellen.

Contents

List of Figures

List of Tables

List of Algorithms

1. Introduction and scope of the thesis

In the first chapter of this thesis, we start with a general introduction to multi-dimensional parameter estimation in Section 1.1. Then, we describe the goals of this thesis in Section 1.2 and summarize the various contributions in Section 1.3. Finally, in Section 1.4, we introduce the mathematical notation used throughout this thesis.

1.1. Introduction to multi-dimensional parameter estimation

The problem of multi-dimensional parameter estimation is of fundamental importance in a wide range of signal processing related applications. Many real-world signals can be modeled as a linear superposition (or linear mixture) of multi-dimensional harmonics (also sinusoids or exponentials), which are separable across R dimensions. The crucial task of estimating the unknown frequency parameters of a noise-corrupted mixture of harmonics sampled on a separable multi-dimensional lattice is referred to as the *harmonic retrieval* problem [KAB83, AR88, SM05]. The underlying mathematical model is described in Chapter 2. The multi-dimensional harmonic retrieval problem arises in a broad variety of application fields. A prominent example is MIMO channel sounding [ZHM+00, HTR04, GS05], where the multi-dimensional parameters are extracted from the sounding measurements to construct a MIMO channel model [ZFDW00, SMB01]. In particular, it has been shown that sampling the channels in multiple dimensions such as time, frequency, space, and polarization, the multi-dimensional channel transfer function can be modeled by R-D harmonics [HZN95, HN98, RHST01, HTR04]. Another important application is bistatic MIMO radar imaging [JLL09, NS10], where the R-D harmonic retrieval problem arises due to the transmission of multi-pulse signals in a bistatic antenna configuration. A related special case of R-D harmonic retrieval is direction of arrival (DOA) estimation in the context of antenna array signal processing [Van02]. For instance, if the array manifold is separable in two or more spatial dimensions, e.g., a uniform rectangular array or a uniform cubic array, and the antenna elements posses the same complex beam pattern, the azimuth and elevation angles of incident signals from sources in the far-field can be estimated [SK93]. Further applications in wireless communications include the initial synchronization in orthogonal frequency division multiple access (OFDMA) systems [LM15], where the timing and frequency offsets of multiple users are estimated, and the localization of multiple frequency-hopped spread-spectrum signals in code division multiple access (CDMA)-based systems [LSS02]. The R-D harmonic retrieval problem also occurs in biomedical signal processing applications. For instance, in multi-dimensional nuclear magnetic resonance (NMR) spectroscopy [BL86, LRL98], where the protein structures are determined by estimating the frequency and damping factors of damped harmonics for $R = 2$. Another biomedical application is the iden-

tification and reconstruction of dipoles in biomagnetic inverse source problems [NSA06, NOH+07] that occur in magnetoencephalography (MEG) and electroencephalography (EEG). Interestingly, R-D harmonic retrieval is also found in image motion estimation, which plays an important role in computer vision and in video communications [CGN98]. In all these various applications, the signals can be modeled as a superposition of R-D complex exponentials or R-D damped (exponentially decaying) complex exponentials in the case of NMR spectroscopy and therefore, the goal is to estimate the frequency parameters (and damping factors), which are nonlinear functions of the observed data.

The R-D harmonic retrieval problem can be addressed by either *nonparametric* or *parametric* estimation techniques [SM05]. The nonparametric harmonic retrieval algorithms are based on the discrete Fourier transform (DFT) and include the periodogram or beamforming as its spatial equivalent. The benefit of nonparametric methods is that no model assumptions are made, i.e., they can estimate the entire spectrum of arbitrary signals. However, as a major disadvantage, their resolution, i.e., the ability to resolve closely-spaced frequencies, is fundamentally limited by the "Rayleigh" resolution limit. If the harmonics are sampled by M uniformly spaced sample points, the resolution limit states that two frequencies that are closer than $2\pi/M$ cannot be resolved [Sch91, Van02]. A higher spectral resolution can be achieved by parametric algorithms, which are therefore often preferred to the nonparametric techniques. The parametric methods assume that the signals satisfy a specific model with known basic functions, e.g., a linear superposition of harmonics or a sparse linear mixture of very few components. Hence, instead of estimating the entire spectrum, only the model parameters, i.e., the frequencies and amplitudes, are estimated. By exploiting this prior knowledge, a significantly higher resolution can be attained.

The performance of high-resolution parameter estimation algorithms is often evaluated by comparing them to the Cramér-Rao bound (CRB), which provides a lower limit on the estimation error of any unbiased estimator [SM05]. Thus, it is desirable to develop parametric harmonic retrieval algorithms, which achieve the CRB or perform as close to the CRB as possible. The literature distinguishes between the deterministic (conditional) and stochastic (unconditional) CRBs derived in [SN89] and [SLG01], respectively. Whereas the stochastic data assumption requires both the signals and the noise to be complex Gaussian-distributed, the deterministic model assumes that the signals are arbitrary non-random sequences while only the noise follows a complex Gaussian distribution. Therefore, the deterministic CRB is more general and at the same time easier to derive by means of the Slepian-Bangs formula [SM05].

The parametric R-D harmonic retrieval algorithms can be classified into *maximum likelihood* methods [SM05, KV96], *subspace-based* algorithms [KV96, Haa97a], and the relatively new class of *sparsity-based* algorithms [MCW05]. Maximum likelihood algorithms are known to be asymptotically efficient [SN90]. This means that they attain the Cramér-Rao bound [SN89] if the signal-to-noise ratio (SNR) is sufficiently high and the sample size is large. However, as maximum likelihood

methods require the minimization of a multi-dimensional objective function, which generally has a complicated multi-modal shape, the achieved asymptotic efficiency comes at a high computational cost. A very attractive alternative is provided by subspace-based and sparsity-based high-resolution parameter estimation algorithms, which admit a lower computational complexity. In this thesis, we are particularly interested in low-complexity algorithms. Therefore, we focus on subspace-based and sparsity-based high-resolution parameter estimation algorithms, which are further discussed in Section 1.1.1 and Section 1.1.2, respectively.

1.1.1. Subspace-based parameter estimation

1-D and R-D parameter estimation

Subspace-based high-resolution parameter estimation algorithms [KV96, Haa97a] are known for their simplicity and computational efficiency. Therefore, their development and advancement has been a very active research area over the last few decades and a broad variety of algorithms has been proposed. The existing 1-D subspace-based parameter estimation algorithms can be classified into the following three categories [Haa97a, HPRE14]:

- *extrema-searching techniques*, e.g., the spectral MUSIC algorithm [Sch79], the rank-reduction estimator (RARE) algorithm [PGW02], the weighted subspace fitting algorithm [VOK91]
- *polynomial rooting techniques*, e.g., Pisarenko's harmonic decomposition [Pis73], the Min-Norm algorithm [KT83], the Root-MUSIC algorithm [Bar83], the Unitary Root-MUSIC algorithm [PGH00], the Fourier domain root-MUSIC algorithm [RG09], the Root-RARE algorithm [PGW02], the method of direction of arrival estimation (MODE) algorithm [Van02]
- *matrix-shifting techniques*: State-Space methods [KAB83, RA92], matrix pencil methods [HS91], the Standard ESPRIT algorithm [RPK86], the multiple invariance ESPRIT algorithm [SORK92], the optimally weighted ESPRIT algorithm [ES94], the Unitary ESPRIT algorithm [HN95].

Among these subspace-based high-resolution parameter estimation algorithms, ESPRIT-type algorithms are often preferred due to their simplicity and their low computational complexity as they provide closed-form parameter estimates. Moreover, they perform very close to the CRB. After the subspace estimation, the key task in ESPRIT-type algorithms is solving a highly structured overdetermined linear system of equations, termed shift invariance equation. The shift invariance equation is usually solved by means of least squares (LS) algorithms, such as the simple LS method [RK89], total least squares (TLS) [OVK91], structured least squares (SLS) [Haa97b], or weighted least squares (WLS) [SN91]. A prominent extension of the ESPRIT algorithm is the Unitary ESPRIT algorithm [HN95], which includes the forward-backward averaging (FBA)

preprocessing step [PK89a] and enables an entirely real-valued implementation with a reduced computational complexity. The advantage of FBA is that it can decorrelate two coherent, i.e., fully correlated, sources. If more than two coherent sources are present or only a single snapshot $N = 1$ for more than two sources is available, spatial smoothing can be applied to decorrelate multiple coherent signals [EJS82, SWK85].

These 1-D subspace-based parameter estimation algorithms can be extended to the multi-dimensional case by applying them to each dimension separately and jointly processing the estimates across all dimensions to obtain the correct pairing [RZZ93]. This approach has led to a number of R-D versions of the aforementioned algorithms. For example, [HN98] proposes an R-D version of Unitary ESPRIT, the work in [PMB04, MSPM04] presents R-D versions of RARE, and [HF94] introduces R-D polynomial rooting. A general literature review of search-free R-D parameter estimation methods is provided in [GRP10, HPRE14].

A common approach towards improving the estimation accuracy of parameter estimation algorithms is to exploit inherent signal structure. Among many types of signal structure, in this thesis, we will focus on the *multi-dimensional* signal structure of the R-D harmonics and the *strictly second-order (SO) non-circular (NC)* signal structure of signals with real-valued modulation schemes.

Exploiting the multi-dimensional signal structure

The aforementioned R-D subspace-based parameter estimation algorithms have in common that the multi-dimensional measurement data is stacked into a measurement matrix. Therefore, the inherent R-D nature of the signals is lost, which leads to a potential performance degradation in the estimation accuracy. The multi-dimensional signal structure can be fully exploited by means of tensor-based signal processing. For instance, the canonical polyadic (CP) tensor decomposition, also known as the parallel factor analysis (PARAFAC) decomposition, has been applied to multi-dimensional harmonic retrieval in [SBG00], which, however, requires computationally complex iterations. A computationally more tractable tensor decomposition is the higher-order singular value decomposition (HOSVD), also known as the multi-linear singular value decomposition (MLSVD) [dLdMV00a], which is a multi-linear extension of the matrix-based singular value decomposition (SVD) and easy to compute. Tensor-based harmonic retrieval algorithms that apply the HOSVD to estimate the subspaces in the individual modes separately include, for instance, a tensor-based spectral MUSIC algorithm [MLM05] and a tensor-based Root-MUSIC algorithm [Boy08]. At the same time, the R-D Standard Tensor-ESPRIT and R-D Unitary Tensor-ESPRIT algorithms [HRD08], which rely on an enhanced HOSVD-based signal subspace estimate across all dimensions instead of the matrix SVD-based estimate, have been developed in [RHD06, HRD08] as an example for subspace-based parameter estimation algorithms. Thereby, the signals' R-D structure is fully exploited. Based on this concept, the R-D Tensor-MUSIC algorithm [BGP+13] and the R-D Tensor

MODE algorithm [WS15] have been proposed. For the Tensor-ESPRIT-type algorithms, the R-D structure can be exploited even further by solving the overdetermined shift invariance equation via the tensor-structure SLS (TS-SLS) algorithm [RH07b], which outperforms the matrix-based R-D SLS algorithm [Haa97b]. In all these references, it has been shown that by exploiting the R-D structure of the signals via tensor-based algorithms, the estimation accuracy can be improved significantly if the number of signals is less than the number of sensors in at least one dimension [HRD08]. This improvement is referred to as the tensor gain [HRD08].

Exploiting the strictly non-circular signal structure

In some harmonic retrieval applications, the multi-dimensional signals additionally exhibit specific statistical properties such as a second-order (SO) strictly non-circular (NC) signal structure. This structure occurs if the received complex signals carry symbols from real-valued modulation schemes such as binary phase shift keying (BPSK), amplitude phase shift keying (ASK), offset quadrature phase shift keying (OQPSK), and minimum phase shift keying (MSK), which undergo a phase rotation due to the channel. These modulation schemes are used, for example, in wireless communications, global navigation satellite systems (GNSS), cognitive radio, radar, tracking, channel sounding, etc. The non-circularity of complex signals implies that the amplitudes are no longer circularly symmetric [Pic94, SS10]. Hence, the SO statistics of NC signals are not fully described by the covariance matrix anymore and the pseudo covariance matrix [NM93], also known as the complementary covariance matrix [SS03] has to be considered. Exploiting this additional information in the pseudo covariance matrix by means of widely-linear processing has been an active field of research in many applications. The research work includes, for instance, widely-linear minimum mean square error (MMSE) estimation [PC95], Wiener filtering [SS03], detection and estimation [SS05], widely-linear adaptive beamforming [CB07, SSW$^+$11, SdW$^+$11], interference-suppression [SdHW12], widely-linear precoding [Ste07], and widely-linear distributed beamforming [SH13a, SH13c]. Therein, it is demonstrated that exploiting the non-circular signal structure provides substantial performance benefits. In the context of harmonic retrieval, it has been shown that the statistical properties of strictly non-circular signals can also be exploited in subspace-based parameter estimation algorithms. Several algorithms including NC MUSIC [AD06], NC Root-MUSIC [CWS01], 1-D NC Standard ESPRIT [ZCW03], and 2-D NC Unitary ESPRIT [HR04] have been proposed. Both the R-D structure and the strict non-circularity of the signals have been exploited in R-D NC Tensor-ESPRIT [RH09]. Assuming only NC signals, these subspace-based NC parameter estimation algorithms provide a significant improvement in the estimation accuracy and can resolve twice as many sources compared to their conventional versions. The more general case of coexisting circular and strictly non-circular signals for MUSIC-like algorithms has been considered in [GNW08, LLXZ12].

Performance analysis of parameter estimation algorithms

With the constant development of subspace-based parameter estimation algorithms, the analytical performance of these algorithms has been of equal importance in the literature. The two most popular performance analysis frameworks for subspace-based algorithms have been developed in [Bri75] and [LLV93]. The framework in [Bri75] derives the analytical performance based on the eigenvector distribution of the sample covariance matrix and has been applied to the MUSIC algorithm [KB86, PF88, PK89a, Fri90], the Root-MUSIC algorithm [RH89b], and the ESPRIT algorithm in [RH89a, MHZ96]. However, the approach in [Bri75] is associated with two main disadvantages. First, it has a limited applicability as the impinging signals as well as the noise contribution are required to be Gaussian distributed. Second, the analytical expressions are only asymptotic in the sample size N, i.e, the results become exact only if the sample size N is very large. In contrast, the performance analysis framework from [LLV93], which has been applied in [LLV93] to MUSIC, Root-MUSIC, and ESPRIT, is based on a first-order perturbation expansion of the SVD, which models the estimation error of the signal subspace as a linear function of the noise. Thus, [LLV93] only assumes that the noise is small compared to the received signal power but no assumptions on the statistics of the signals or the noise are required. Moreover, the analytical expressions from [LLV93] are asymptotic in the effective SNR, i.e., the results become accurate for either high SNRs or a large sample size N. Thus, they are even valid in the case of a single snapshot $N = 1$ if the SNR is sufficiently high. Due to these advantages, the framework in [LLV93] is preferable over that in [Bri75].

For ESPRIT-type parameter estimation based on the simple LS method to solve the shift invariance equation, the work in [RBHW09, RBH10, RHD14] extends the performance analysis framework of [LLV93] to the case of matrix-based R-D ESPRIT-type algorithms, i.e., R-D Standard ESPRIT and R-D Unitary ESPRIT as well as the tensor-based R-D ESPRIT-type algorithms, i.e., R-D Standard Tensor-ESPRIT and R-D Unitary Tensor-ESPRIT. Moreover, [RHD14] derives analytical mean square error (MSE) expressions that only require the noise to be zero-mean with finite SO moments regardless of its statistics. Using this framework, a performance analysis of 1-D Standard ESPRIT using SLS to solve the shift invariance equation is proposed in [RH11].

Performance analysis for special cases

The above-mentioned analytical MSE expressions are formulated in terms of the subspaces of the measurement matrix. However, in many applications, it is desirable to know how the MSE scales with the explicit physical system parameters, e.g., the SNR, the number of sensors, the sample size, etc. To this end, for special cases of a small number of sources, i.e., a single source or two sources, the MSE expressions can be simplified into very compact formulas. Such expressions are very valuable in practice as they can facilitate array design decisions on the number of the required

sensors to achieve a certain performance for a specific SNR. Moreover, they allow to objectively compare different parameter estimation algorithms to find the best estimator for particular scenarios. In the literature, the special case of the performance analysis of ESPRIT for a single source is considered in [RH89a] and the asymptotic efficiency, i.e., the ratio of the CRB and the MSE, of MUSIC, Root-MUSIC, and TLS-ESPRIT for a single source was presented in [PF88, RH89b] and [OVK91], respectively. However, these results are asymptotic in the sample size N or even in the number of sensors M. Based on the performance analysis framework in [LLV93], analytical expressions for the single source case of R-D ESPRIT-type and R-D Tensor-ESPRIT-type algorithms, which are also accurate for small values of M and asymptotic in the effective SNR have been derived in [RH12].

1.1.2. Sparsity-based parameter estimation

A relatively new concept to solve the multi-dimensional harmonic retrieval problem is sparse signal reconstruction (SSR), which is also known as sparse signal recovery or compressed sensing. SSR has been a rapidly emerging research area and has already found many applications in signal processing related fields [CW08, EK12] such as spectral analysis, image processing, and parameter estimation. In its abstract formulation, SSR aims at reconstructing a high-dimensional signal vector, which is sparse in an overcomplete basis from a low-dimensional measurement vector by solving an underdetermined system of linear equations [CW08]. Exploiting the prior knowledge on the sparse structure of the signals in the overcomplete basis, which is assumed known, a unique solution to the underdetermined system of equations can be found [CW08, EK12]. This result has, for instance, great implications in the field of sampling theory, where perfect signal construction can be achieved with fewer samples than required by the Nyquist sampling theorem [CW08, ME11]. However, finding the sparsest solution requires solving an ℓ_0-norm minimization problem, which is NP-hard for large problem instances. Therefore, a number of approximate solutions to the SSR problem based on the convex ℓ_1-norm relaxation [CDS98] or greedy algorithms [MZ93, PRK93] have been proposed. These methods provide a good reconstruction performance while being computationally tractable.

Sparse signal reconstruction concepts for parameter estimation

The concept of SSR has recently been applied to multi-dimensional harmonic retrieval for a single snapshot, which is also referred to as the single measurement vector (SMV) case [Don92, Don06, CW08]. Therein, the array response is modeled as the superposition of the received signal power of very few wavefronts, i.e., the power spectrum is sparse in an overcomplete basis corresponding to the spatial domain. Such a finite basis is obtained by discretizing the continuous spatial dimension with a predefined grid. The corresponding ℓ_0-norm minimization problem is then addressed by

the above-mentioned convex ℓ_1-norm relaxation or greedy algorithms. The obtained support, i.e., the indices of the non-zero elements, of the sparsest solution vector corresponds to the spatial directions.

In the case of multiple snapshots, also termed multiple measurement vectors (MMV) case, which is encountered more often in harmonic retrieval, many sparsity-exploiting parameter estimation algorithms [MCW05, HM10, SBL11, MZ06, SPP17, SPP18b] have been developed. These algorithms additionally take advantage of the joint sparsity, i.e., the common row support of the solution matrix, by addressing the corresponding $\ell_{2,0}$-norm minimization problem (see Appendix C for the norm definitions). For instance, the recent work in [SPP17, SPP18b] proposes a compact formulation of the relaxed $\ell_{2,1}$-norm minimization problem for the MMV case. An efficient approach for the SSR in the case of partly calibrated arrays has been proposed in [SPP14, SPP18a]. It has been observed that all these SSR methods exhibit super-resolution capabilities without facing the drawbacks of the conventional DOA methods [KV96], i.e., a performance degradation for a high source correlation, unknown model order, a low sample size, etc. These challenging conditions often occur in multipath environments, fast-changing tracking applications, and in co-array processing [PV10, VP11]. However, the discretization of the continuous spatial domain often results in an off-grid problem, where the true parameters lie off the discretized grid, which results in a model mismatch and a severe performance degradation. Solutions to the off-grid problem include an adaptive grid refinement [MCW05], statistical fitting of the offset error [YXZ13], and a low-complexity grid offset estimation procedure based on local interpolation [IRA+14].

Another approach that completely avoids the off-grid problem is provided by the gridless sparse recovery framework developed in [CFG14, TBSR13]. This framework poses an atomic ℓ_0-norm minimization (ANM) problem as defined in Appendix C.4 for the SMV case, whose convex ℓ_1-norm relaxation can be equivalently formulated as a semi-definite programming (SDP) problem. The Hermitian-Toeplitz structured solution matrix then admits a unique Vandermonde decomposition, which allows to uniquely recover the frequency parameters via conventional parameter estimation algorithms [KV96]. It has been shown that in the noiseless case, the sparse spatial line spectrum can be recovered in the continuous parameter domain with infinite precision [CFG14, TBSR13]. Note that this is also the case for the subspace-based parameter estimation algorithms mentioned in Section 1.1.1. An extension of the ANM framework to the MMV case is given in [YX16b, SPP18b], while an extension to multi-dimensional parameter estimation is provided in [CC15, YXS16]. However, a major drawback of the ANM approach is that it suffers from the Rayleigh resolution limit, i.e., the spatial frequencies can only be recovered if they are sufficiently separated [CFG14, TBSR13].

Sparse signal reconstruction for strictly non-circular signals

As motivated for subspace-based parameter estimation algorithms, the received signals can possess a strictly second-order (SO) non-circular (NC) signal structure [SS10]. NC signals result from real-valued modulation schemes such as BPSK, PAM, ASK, or Offset-QPSK (after a de-rotation) and are introduced in Section 2.2. It has been shown that the performance of parameter estimation algorithms can be significantly improved if the NC structure of the received signals is exploited. The concept of exploiting the non-circularity property has recently been introduced for sparsity-based parameter estimation [LHZ12], [YLZ15]. While [LHZ12] proposes a sparse covariance matrix representation of the SO statistics of the non-circular data, in [YLZ15], the authors adopt a strategy, which relies on a sparsity-based fitting of the NC subspaces. However, both algorithms require a rather complex setting of the sparsity-inducing parameters depending on the scenario, are limited to the case of uncorrelated sources, and do not deal with the critical off-grid problem.

1.2. Goals of the thesis

Multi-dimensional parameter estimation has been an active area of research for several decades, which is emphasized by the extensive literature overview given in the previous section. However, despite the many publications in this research area, there are still a number of fundamental questions, which have so far remained unanswered. Several open problems will be identified in this section.

The first open problem is related to the tensor-based R-D parameter estimation algorithms derived in [HRD08]. It has been shown that the tensor-based R-D parameter estimation algorithms achieve a tensor gain over the matrix-based R-D parameter estimation algorithms in several scenarios. The tensor gain has so far only been analyzed by simulative studies. However, by means of the tensor-based R-D performance analysis framework for R-D Tensor-ESPRIT-type algorithms in [RHD14], an analytical study of the tensor gain can be performed to investigate its explicit dependence on the physical parameters. For the special case of a single source, it has been shown that no tensor gain can be achieved [RH12]. Therefore, to derive an analytical expression of the tensor gain of R-D Tensor-ESPRIT-type algorithms, the special case of two sources needs to be considered.

Another open problem is the development of matrix-based and tensor-based R-D NC ESPRIT-type algorithms to exploit the signal structure of strictly non-circular (NC) sources. In [HR04], only 2-D NC ESPRIT-type algorithms have been proposed. Moreover, the R-D performance analysis framework from [RHD14] can be extended to the NC case to obtain analytical performance expressions for the matrix-based and tensor-based R-D NC ESPRIT-type algorithms. Based on the analytical expressions for the MSE, the NC gain can be analytically computed for a single source and two sources. Thereby, the behavior of the NC gain as a function of the physical parameters

can be studied.

So far, the analytical performance of R-D ESPRIT-type algorithms and R-D NC ESPRIT-type algorithms with spatial smoothing has not been considered in the literature. In practice, the choice of the number of subarrays for spatial smoothing is usually unclear. Based on the performance analysis expressions for the spatially smoothed R-D ESPRIT-type algorithms, the optimal number of subarrays for a single source can be computed analytically.

ESPRIT-type parameter estimation algorithms are search-free and have a low computational complexity. After the subspace estimation, the performance mainly depends on the accuracy of the solution to the shift invariance equation via least squares methods. The existing least squares solutions [RK89, OVK91, Haa97b] usually perform close to the CRB in the asymptotic case but do not attain it. It is still an open problem to develop a least squares algorithm for ESPRIT-type algorithms to achieve the CRB asymptotically.

Recently, the concept of parameter estimation based on sparse signal reconstruction (SSR) has been introduced. The proposed SSR algorithms for parameter estimation can be extended to exploit the structure of NC signals. From the literature, it is unclear whether the same benefits from processing NC sources can be achieved for SSR algorithms.

The R-D extension of the deterministic NC CRB for NC signals with a rigorous proof is still an open problem. The deterministic R-D NC CRB serves as a benchmark for the R-D parameter estimation algorithms for NC signals. In order to find a closed-form expression for the maximum achievable NC gain, the deterministic R-D NC CRB can be simplified for a single source and two sources.

The aforementioned open problems will be addressed in this thesis.

1.3. Summary of the contributions

In this section, we summarize the contributions to the field of multi-dimensional parameter estimation. We distinguish between the contributions covered in this thesis in Section 1.3.1 and other contributions that are related to the topic but not discussed in this thesis in Section 1.3.2.

1.3.1. Contributions in this thesis

This section provides a detailed summary of the contributions in this thesis and how they are organized into the chapters.

After presenting the data model in Chapter 2, we first review the matrix-based and tensor-based R-D ESPRIT-type algorithms, i.e., R-D Standard ESPRIT and R-D Unitary ESPRIT, in Chapter 3.

- *Analytical FBA gain and tensor gain of R-D ESPRIT-type algorithms for two sources*

Chapter 4 first reviews the performance analysis framework for matrix-based and tensor-based R-D ESPRIT-type algorithms using LS to solve the shift invariance equation, which was first proposed in [RHD14]. The resulting analytical expressions for the parameter estimation error and the MSE only require the noise to be zero-mean with finite SO moments and are asymptotic in the effective SNR. Our contribution starts in Section 4.5 with the simplification of these MSE expressions of the matrix-based and tensor-based R-D ESPRIT-type algorithms using LS for two sources. The resulting expressions only depend on the physical parameters, e.g., the number of antennas, the SNR, etc. For a single source, the matrix-based and tensor-based R-D ESPRIT-type algorithms yield the same MSE and perform identical. For the two source case of the matrix-based R-D ESPRIT-type algorithms, we analytically compute the gain from forward-backward averaging and analyze its dependence on the physical parameters. For the simplified MSE of the tensor-based R-D ESPRIT-type algorithms for two sources, we compute and analyze the tensor gain as well as the forward-backward averaging gain in the tensor case.

- *Generalized least squares (GLS) for R-D ESPRIT-type algorithms*

In Chapter 5, we propose a novel least squares algorithm, termed general least squares (GLS), to solve the shift invariance equation of the matrix-based R-D Standard ESPRIT and R-D Unitary ESPRIT algorithms for multiple sources. The results have been published in [SRH17a]. Assuming a uniform R-D array with maximally overlapping subarrays and zero-mean circularly symmetric white noise, the GLS algorithm takes the statistics of the subspace estimation error on both sides of the shift invariance equation into account through its covariance matrix. The covariance matrix of the subspace estimation error is found analytically via the first-order perturbation expansion from [LLV93]. As the error covariance matrix requires an initial estimate of the unknown shift invariance solution, an iterative procedure by repeatedly performing GLS updates is possible. However, we show that if GLS is initialized by the simple LS solution, only one GLS iteration is required to achieve a significantly improved estimation accuracy in the asymptotic case, i.e., at either high SNRs or a large sample size. However, at low SNRs and for a small sample size, performing additional GLS iterations further improves the estimation accuracy. Note that R-D Unitary ESPRIT with GLS involves only real-valued operations, thus reducing the computational complexity. In the second contribution, we develop a performance analysis for R-D ESPRIT-type algorithms using a single GLS iteration. The derived analytical MSE expressions are based on the framework in Chapter 4 and are asymptotic in the effective SNR. We simplify the MSE expressions for a single source and two orthogonal sources and show that they coincide with the deterministic Cramér-Rao bound (CRB). This implies that in these cases, R-D ESPRIT-type algorithms in combination with one GLS iteration are asymptotically efficient, i.e., the ratio of the CRB and the MSE is equal to 1.

- *Matrix-based and tensor-based R-D ESPRIT-type algorithms for strictly non-circular sources*

and their performance analysis

In Chapter 6, we first present the matrix-based and tensor-based versions of R-D NC Standard ESPRIT and the R-D NC Unitary ESPRIT algorithms, which exploit the strict SO non-circularity of stationary sources by means of an NC preprocessing step. The results have been published in [SRH13a, SRHD14]. We show that in the matrix case as well as in the tensor case, the NC preprocessing step always results in a centro-symmetric virtual array with a doubled number of sensors. This property makes the R-D NC Unitary ESPRIT algorithms, which can be formulated in terms of entirely real-valued operations, also applicable to physical non-centro-symmetric arrays. Both matrix-based and tensor-based R-D NC ESPRIT-type algorithms achieve a significantly lower estimation error than their conventional non-NC counterparts reviewed in Chapter 3. In our second contribution, we derive a first-order performance analysis of the matrix-based and tensor-based R-D NC ESPRIT-type algorithms using least squares (LS). Due to its advantages, we adopt the R-D performance analysis framework in [RHD14] introduced in Section 4.4. We derive MSE expressions, where apart from a zero mean and finite SO moments no further assumptions on the noise statistics are needed. Moreover, the expressions are asymptotic in the effective SNR, i.e., they become accurate for either high SNRs or large sample sizes. Additionally, we analytically prove that both R-D NC Standard ESPRIT and R-D NC Unitary ESPRIT as well as R-D NC Standard Tensor-ESPRIT and R-D NC Unitary Tensor-ESPRIT perform asymptotically identical. Finally, we present simplified R-D MSE expressions for both matrix-based and tensor-based NC ESPRIT-type algorithms in the special case of a single NC source and two NC sources, where a uniform sampling grid and circularly symmetric white noise are assumed. The resulting expressions only depend on the physical parameters and admit the derivation of closed-form expressions of the NC gain in the matrix as well as in the tensor case and their asymptotic efficiency.

- *R-D ESPRIT-type algorithms and their performance analysis with spatial smoothing for general and strictly non-circular sources*

Chapter 7 presents a first-order performance analysis for the matrix-based R-D ESPRIT-type as well as the R-D NC ESPRIT-type algorithms with spatial smoothing preprocessing, which has been published in [SRH14b, SRHD17]. Spatial smoothing is widely used to decorrelate more than two coherent sources and to virtually increase the sample size, which yields an improved estimation accuracy in these cases but sacrifices array aperture. For the performance analysis, we adopt the framework from [RHD14, SRHD14] for a uniform R-D array geometry and use least squares (LS) to solve the shift invariance equation. The derived analytical MSE expressions are explicit in the noise realization such that apart from a zero mean and finite SO moments, no further assumptions on the noise statistics are required. We show that due to the NC preprocessing both R-D NC ESPRIT-type algorithms with spatial smoothing perform identical in the high effective SNR. Moreover, we perform a case study of a single source to obtain further insights into the

dependence of the MSE expressions on the physical parameters. In particular, we first show that R-D spatial smoothing improves the MSE for a single source. Moreover, the considered spatial smoothing based R-D ESPRIT-type algorithms provide the same MSE result, i.e., asymptotically, no additional gain is obtained from FBA and NC preprocessing in case of a single source. Based on these results, we analytically find the optimal number of subarrays L that minimizes the MSE in each of the R dimensions, which extends the 1-D results in [PK89a, RH90, RH93, HS90, HR99, LvdVD03, WF93]. This enables us to compute the maximum asymptotic R-D spatial smoothing gain and the asymptotic efficiency for a single source in closed-form.

- *Sparsity-based parameter estimation algorithms for strictly non-circular sources*

In Chapter 8, we address the problem of multi-dimensional harmonic retrieval via sparse signal reconstruction (SSR). In particular, we develop three different SSR algorithms published in [SRH16c, SSPH16, SRS$^+$16] for exploiting the statistical properties of strictly non-circular signals. The concept of strictly non-circular signals is introduced in Section 2.2 and assumes that the received complex symbols result from real-valued constellations rotated by an arbitrary phase ϕ. This property can be exploited via an NC preprocessing step that virtually doubles the number of sensors at the receiver and thus, yields a higher estimation accuracy and doubles the number of identifiable signals. In the first contribution [SRH16c], we show that since ϕ is usually unknown, the harmonic retrieval problem becomes a two-dimensional (2-D) sparse recovery problem (in the spatial domain and in the rotation phase domain), which can be solved by $\ell_{2,1}$-mixed norm relaxation using MMV. For the resulting 2-D off-grid problem, we propose an off-grid estimation procedure based on [IRA$^+$14] by means of local interpolation. The second contribution [SSPH16] reduces the high computational complexity of [SRH16c] by proposing a solution based on nuclear norm minimization after lifting the original problem to a semidefinite programming (SDP) problem in a higher-dimensional space. This procedure effectively reduces the 2-D estimation problem to a 1-D estimation problem only in the sampled spatial domain, which requires a significantly lower computational complexity while providing the same performance. Additionally, we present a 1-D grid-offset estimator for the spatial domain using the concept of [IRA$^+$14]. In the third contribution [SRS$^+$16], we extend the grid-less atomic norm minimization (ANM) framework to the NC case. After the NC preprocessing step, the ANM-equivalent SDP problem provides a solution matrix with a two-level Hermitian-Toeplitz structure, which is used to uniquely extract the frequency estimates via the NC ESPRIT-type algorithms presented in Chapter 6. All these three proposed NC SSR algorithms provide a superior estimation accuracy over the original methods for arbitrary signals and can resolve more sources than the number of sensors in the array.

- *Deterministic R-D Cramér-Rao bound for strictly non-circular sources and its analysis*

Finally, in Chapter 9, we derive a closed-form expression of the deterministic R-D NC Cramér-Rao bound (CRB) for strictly non-circular signals, which is based on the Slepian-Bangs formula.

This result is published in [SRHD16]. We identify the special cases, e.g., full coherence, a single snapshot, and a single strictly non-circular source, where the deterministic R-D NC CRB reduces to the existing deterministic R-D CRB for arbitrary signals [SN89]. This suggests that under the deterministic data assumption no NC gain can be achieved in these cases. In the second contribution, we simplify both the 1-D NC CRB and the 1-D CRB for the special case of two closely-spaced NC signals and a uniform linear array (ULA) to obtain compact formulas in terms of the physical parameters. These simplified expressions are then used to analytically compute the maximum achievable NC gain, which is then studied in terms of the above-mentioned physical parameters.

- *Summary and future research directions*

In Chapter 10, we provide the conclusions and identify directions for future work related to subspace-based and sparsity-based multi-dimensional parameter estimation. A detailed literature overview of the various ESPRIT-type algorithms along with their performance analysis is presented in Table 10.1. Furthermore, Table 10.2 provides a survey of the various least squares algorithms to solve the overdetermined shift invariance equation of ESPRIT-type algorithms and the corresponding performance analysis expressions. Here, we want to put emphasis on the novel GLS algorithm introduced in Chapter 5, which achieves the asymptotic efficiency of ESPRIT-type algorithms. The tables also outline the cases, where the performance analysis is still an open topic.

1.3.2. Other related contributions

In this section, we provide a list of contributions that are related[1] to the topic of multi-dimensional parameter estimation but for the sake of brevity not further discussed in this thesis.

- *Performance analysis of 1-D ESPRIT-type algorithms using structured least squares*

In [SRH13b], we apply the first-order analytical performance framework for R-D NC ESPRIT-type algorithms from [SRHD14], which is described in Section 6.4 to 1-D NC ESPRIT-type algorithms using structured least squares (SLS) to solve the shift invariance equation. For overlapping subarrays, structured least squares takes into account that the subspace estimation error on both sides of the shift invariance equation is common and therefore, provides a more accurate solution. We derive a first-order approximation of the estimation error, which is explicit in the noise, and analytical MSE expressions, which only require the assumptions of a zero mean and finite SO moments of the noise. The expressions are asymptotic in the effective signal-to-noise ratio (SNR). In accordance with Section 6.4, 1-D NC Standard ESPRIT and 1-D NC Unitary ESPRIT perform asymptotically identical.

[1] Non-related contributions include [SdH11a, SdH11b, SdH13, SH13b, SVH14c, SVH14a, SVH14b, BCS+16].

- *1-D C-NC ESPRIT-type algorithms for a mixture of circular and strictly non-circular signals, their performance analysis, and the deterministic Cramér-Rao bound*

The presented R-D NC ESPRIT-type algorithms assume that all the received signals are strictly non-circular. In the references [SRH15c, SRH16a, SRH15b], we consider the more realistic case of a mixture of circular and strictly non-circular signals. In [SRH15c], we propose the 1-D C-NC Standard ESPRIT and 1-D C-NC Unitary ESPRIT algorithms for this scenario, which yield closed-form estimates while C-NC Unitary ESPRIT also enables an entirely real-valued implementation. It is shown that the estimation accuracy of both algorithms improves with an increasing number of strictly non-circular signals among a fixed number of sources. Thereby, not only the estimation accuracy of the strictly non-circular signals themselves is improved, but also the estimation accuracy of the circular signals. The corresponding performance analysis for both 1-D C-NC ESPRIT-type algorithms is derived in [SRH16a]. As we have again applied the framework in [SRHD14], the derived analytical expressions for the estimation error and the MSE have the same above-mentioned properties. Moreover, [SRH15b] presents the deterministic Cramér-Rao bound (CRB), termed deterministic C-NC CRB, for coexisting circular and strictly non-circular signals. This bound serves as a benchmark for the proposed C-NC ESPRIT-type algorithms and the MUSIC-like algorithms from [GNW08, LLXZ12].

- *Performance analysis of 1-D ESPRIT-type algorithms for co-array structures*

In [SRH17b], we have proposed a first-order performance analysis of 1-D Standard ESPRIT and 1-D Unitary ESPRIT algorithms for co-array structures [PV10, VP11]. In the recent field of co-array signal processing, sparse linear arrays are processed to form a virtual uniform linear array (ULA), termed co-array, that allows to resolve more sources than physical sensors. The extra degrees of freedom (DOFs) are leveraged by the assumption that the signals are uncorrelated, which requires a large sample size. We again derive analytical first-order expressions for the parameter estimation error and the MSE using the framework in [RHD14]. Based on the obtained expressions, we study the effects of a small sample size such as the residual sample signal correlation and the sample noise contribution on the estimation accuracy of the proposed algorithms.

- *Second-order performance analysis of 1-D Standard ESPRIT*

The first-order performance analysis framework in [LLV93] and its R-D extension [RHD14] is asymptotic in the effective SNR, i.e., the expressions become exact for either high SNRs or a large sample size. Moreover, this framework does not require any statistical assumptions on the noise, except that it should be small compared to the signal component. However, this poses the question of how small exactly the noise needs to be in order for the first-order performance framework to be applicable. Furthermore, these first-order expressions do not capture the algorithmic behavior in the threshold region at low SNRs or for a small sample size. Yet, such conditions are often

encountered in practice. Therefore, in [SRH17c], we propose a second-order (SO) performance analysis framework of the 1-D Standard ESPRIT algorithm. We present a closed-form expression for the parameter estimation error of 1-D Standard ESPRIT up to the SO that is valid in a wider effective SNR range. In addition, we derive an analytical MSE expression, which assumes a zero-mean circularly symmetric complex Gaussian noise distribution. Finally, the existing first-order MSE expression and the derived second-order MSE expression are used to analytically compute the SNR breakdown threshold of the MSE threshold region. The threshold region refers to a rapid deterioration of the MSE of parameter estimators, i.e., a performance breakdown, when the SNR falls below a threshold SNR [FLB04].

1.4. Algebraic concepts and notation

This section introduces the notation and reviews several matrix-based and tensor-based (multi-linear) algebraic concepts that are used throughout this thesis. These useful properties, operators, and definitions are organized in a compact and systematic form to facilitate the readability and the comprehensibility of the presented proofs and derivations without the need to consult other references. In Section 1.4.1, we first introduce the general notation and review a number of matrix properties. In Section 1.4.2, we review the fundamentals of multi-linear (tensor) algebra, which are used to store and process the multi-dimensional data in its natural form. These multi-linear algebraic concepts are used for the Tensor-ESPRIT-type parameter estimation algorithms discussed in the Chapters 3-6. Throughout the thesis, we will often refer to these properties whenever they are needed.

1.4.1. Matrix-based algebraic concepts and properties

We denote scalars by lowercase italic letters (a, b), vectors by lowercase boldface letters $(\boldsymbol{a}, \boldsymbol{b})$, matrices by uppercase boldface letters $(\boldsymbol{A}, \boldsymbol{B})$, and tensors by uppercase calligraphic letters $(\boldsymbol{\mathcal{A}}, \boldsymbol{\mathcal{B}})$. The transpose, conjugate, conjugate-transpose, inverse, and Moore-Penrose pseudo-inverse of a matrix are represented by $^{\mathrm{T}}$,* ,$^{\mathrm{H}}$,$^{-1}$, and $^{+}$, respectively. As the (pseudo)-inverse and the (hermitian)-transpose operations commute, we also introduce the short-hand notations $^{-\mathrm{T}}$,$^{-*}$, and $^{-\mathrm{H}}$. The operator $\mathrm{vec}\{\boldsymbol{A}\}$ stacks the columns of the matrix $\boldsymbol{A} \in \mathbb{C}^{M \times N}$ into a column vector of length $MN \times 1$, The trace-operator $\mathrm{Tr}\{\boldsymbol{A}\}$ returns the trace of the matrix $\boldsymbol{A}$, $\mathrm{diag}\{\boldsymbol{a}\}$ returns a diagonal matrix with the elements of $\boldsymbol{a}$ placed on its diagonal, and $\mathrm{blkdiag}\{\boldsymbol{A}, \boldsymbol{B}\}$ is the block-diagonal extension with the matrices $\boldsymbol{A}, \boldsymbol{B}$ aligned on the diagonal. The rank of a matrix $\boldsymbol{A}$ is denoted by $\mathrm{rank}\{\boldsymbol{A}\}$. The symbol $\geq$ indicates element-wise inequality for vectors and positive semi-definiteness for matrices, correspondingly. Furthermore, $\boldsymbol{I}_M$ represents the $M \times M$ identity matrix, the matrix $\boldsymbol{\Pi}_M$ is the $M \times M$ exchange matrix with ones on its anti-diagonal and zeros elsewhere, and the matrices $\mathbf{1}_{M \times M}$ and $\mathbf{0}_{M \times M}$ denote the $M \times M$ matrices of ones and zeros, respectively. Moreover,

$\mathrm{Re}\{\cdot\}$ and $\mathrm{Im}\{\cdot\}$ extract the respective real and imaginary parts of a complex number or a matrix, $|\cdot|$ represents the absolute value of a complex number, and $\mathbb{E}\{\cdot\}$ stands for the statistical expectation. Finally, the Euclidean norm of a vector, the Frobenius norm of a matrix, and the higher-order norm of a tensor are denoted by $\|\cdot\|_2$, $\|\cdot\|_{\mathrm{F}}$, and $\|\cdot\|_{\mathrm{H}}$. A complete list of all the symbols and operators used in this thesis is provided in Appendix A.2.

The Kronecker product of two matrices $\boldsymbol{A} \in \mathbb{C}^{M\times N}$ and $\boldsymbol{B} \in \mathbb{C}^{P\times Q}$ is represented by $\boldsymbol{A} \otimes \boldsymbol{B} \in \mathbb{C}^{MP\times NQ}$. Its definition is given in Appendix A.2. The Khatri-Rao product (column-wise Kronecker product) of two matrices $\boldsymbol{A} \in \mathbb{C}^{M\times d}$ and $\boldsymbol{B} \in \mathbb{C}^{P\times d}$ is denoted by $\boldsymbol{A} \diamond \boldsymbol{B} \in \mathbb{C}^{MP\times d}$. Moreover, the Hadamard product (element-wise multiplication) of two matrices $\boldsymbol{A} \in \mathbb{C}^{M\times N}$ and $\boldsymbol{B} \in \mathbb{C}^{M\times N}$ of the same size is represented by $\boldsymbol{A} \odot \boldsymbol{B} \in \mathbb{C}^{M\times N}$.

For the sake of notational convenience, we define the following short-hand notation for multiple Kronecker, Khatri-Rao, and Hadarmard products as

$$\bigotimes_{r=1}^{R} \boldsymbol{X}^{(r)} = \boldsymbol{X}^{(1)} \otimes \boldsymbol{X}^{(2)} \otimes \ldots \otimes \boldsymbol{X}^{(R)} \tag{1.1}$$

$$\bigdiamond_{r=1}^{R} \boldsymbol{X}^{(r)} = \boldsymbol{X}^{(1)} \diamond \boldsymbol{X}^{(2)} \diamond \cdots \diamond \boldsymbol{X}^{(R)} \tag{1.2}$$

$$\bigodot_{r=1}^{R} \boldsymbol{X}^{(r)} = \boldsymbol{X}^{(1)} \odot \boldsymbol{X}^{(2)} \odot \cdots \odot \boldsymbol{X}^{(R)}. \tag{1.3}$$

We next present a list of algebraic properties for the Kronecker, the Khatri-Rao, and the Hadamard product that are used frequently within this thesis. For the matrices $\boldsymbol{A} \in \mathbb{C}^{M\times N}, \boldsymbol{B} \in \mathbb{C}^{P\times Q}, \boldsymbol{C} \in \mathbb{C}^{N\times R}, \boldsymbol{D} \in \mathbb{C}^{Q\times S}, \boldsymbol{E} \in \mathbb{C}^{Q\times R}$ and the vector $\boldsymbol{x} \in \mathbb{C}^{N}$, we have [Bre78, PP08]

$$\alpha \cdot (\boldsymbol{A} \otimes \boldsymbol{B}) = (\alpha \cdot \boldsymbol{A}) \otimes \boldsymbol{B} = \boldsymbol{A} \otimes (\boldsymbol{B} \cdot \alpha) \tag{1.4}$$

$$(\boldsymbol{A} \otimes \boldsymbol{B})^{\mathrm{H}} = \boldsymbol{A}^{\mathrm{H}} \otimes \boldsymbol{B}^{\mathrm{H}} \tag{1.5}$$

$$(\boldsymbol{A} \otimes \boldsymbol{B})^{+} = \boldsymbol{A}^{+} \otimes \boldsymbol{B}^{+} \tag{1.6}$$

$$\|\boldsymbol{A} \otimes \boldsymbol{B}\|_{\mathrm{F}} = \|\boldsymbol{A}\|_{\mathrm{F}} \cdot \|\boldsymbol{B}\|_{\mathrm{F}} \tag{1.7}$$

$$(\boldsymbol{A} \otimes \boldsymbol{B}) \cdot (\boldsymbol{C} \otimes \boldsymbol{D}) = (\boldsymbol{A} \cdot \boldsymbol{C}) \otimes (\boldsymbol{B} \cdot \boldsymbol{D}) \tag{1.8}$$

$$(\boldsymbol{A} \otimes \boldsymbol{B}) \cdot (\boldsymbol{C} \diamond \boldsymbol{E}) = (\boldsymbol{A} \cdot \boldsymbol{C}) \diamond (\boldsymbol{B} \cdot \boldsymbol{E}) \tag{1.9}$$

$$(\boldsymbol{A} \diamond \boldsymbol{B})^{\mathrm{H}} \cdot (\boldsymbol{C} \otimes \boldsymbol{D}) = \left[\boldsymbol{A}^{\mathrm{H}} \cdot \boldsymbol{C} \diamond \boldsymbol{B}^{\mathrm{H}} \cdot \boldsymbol{D}\right]^{\mathrm{H}} \tag{1.10}$$

$$\boldsymbol{A} \diamond \boldsymbol{x}^{\mathrm{T}} = \boldsymbol{x}^{\mathrm{T}} \diamond \boldsymbol{A} = \boldsymbol{A} \cdot \mathrm{diag}\{\boldsymbol{x}\} \tag{1.11}$$

$$(\boldsymbol{C} \diamond \boldsymbol{E})^{\mathrm{H}} \cdot (\boldsymbol{C} \diamond \boldsymbol{E}) = \left(\boldsymbol{C}^{\mathrm{H}} \cdot \boldsymbol{C}\right) \odot \left(\boldsymbol{E}^{\mathrm{H}} \cdot \boldsymbol{E}\right) \tag{1.12}$$

Note that all these properties generalize to multiple Kronecker, Khatri-Rao and Schur products by applying them repeatedly.

The operator $\mathrm{vec}\{\boldsymbol{X}\}$ stacks the columns of a matrix $\boldsymbol{X} = [\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N] \in \mathbb{C}^{M\times N}$ into a column

vector according to

$$\operatorname{vec}\left\{\begin{bmatrix}\boldsymbol{x}_1 & \dots & \boldsymbol{x}_N\end{bmatrix}\right\} = \begin{bmatrix}\boldsymbol{x}_1^{\mathrm{T}} & \dots & \boldsymbol{x}_N^{\mathrm{T}}\end{bmatrix}^{\mathrm{T}} \in \mathbb{C}^{MN\times 1} \tag{1.13}$$

It satisfies the following properties

$$\operatorname{vec}\{\boldsymbol{A}\cdot\boldsymbol{X}\cdot\boldsymbol{B}\} = \left(\boldsymbol{B}^{\mathrm{T}}\otimes\boldsymbol{A}\right)\cdot\operatorname{vec}\{\boldsymbol{X}\} \tag{1.14}$$

$$\operatorname{vec}\{\alpha\cdot\boldsymbol{A}\} = \alpha\cdot\operatorname{vec}\{\boldsymbol{A}\} \tag{1.15}$$

for arbitrary matrices $\boldsymbol{A}\in\mathbb{C}^{M\times N}$, $\boldsymbol{X}\in\mathbb{C}^{N\times P}$, and $\boldsymbol{B}\in\mathbb{C}^{P\times Q}$. A special case of (1.14) is given by

$$\operatorname{vec}\{\boldsymbol{A}\cdot\boldsymbol{B}\} = (\boldsymbol{I}_Q\otimes\boldsymbol{A})\cdot\operatorname{vec}\{\boldsymbol{B}\} = \left(\boldsymbol{B}^{\mathrm{T}}\otimes\boldsymbol{I}_M\right)\cdot\operatorname{vec}\{\boldsymbol{A}\}\,. \tag{1.16}$$

The vectorization of a Kronecker product in terms of one of its arguments is accomplished via the following identities [Roe13]

$$\operatorname{vec}\{\boldsymbol{A}\otimes\boldsymbol{X}\} = \left(\begin{bmatrix}\boldsymbol{I}_Q\otimes\boldsymbol{a}_1\\ \vdots\\ \boldsymbol{I}_Q\otimes\boldsymbol{a}_N\end{bmatrix}\otimes\boldsymbol{I}_P\right)\cdot\operatorname{vec}\{\boldsymbol{X}\} \tag{1.17}$$

$$\operatorname{vec}\{\boldsymbol{X}\otimes\boldsymbol{A}\} = \left(\boldsymbol{I}_Q\otimes\begin{bmatrix}\boldsymbol{I}_P\otimes\boldsymbol{a}_1\\ \vdots\\ \boldsymbol{I}_P\otimes\boldsymbol{a}_N\end{bmatrix}\right)\cdot\operatorname{vec}\{\boldsymbol{X}\}\,, \tag{1.18}$$

where $\boldsymbol{A} = [\boldsymbol{a}_1,\dots,\boldsymbol{a}_N]\in\mathbb{C}^{M\times N}$ and $\boldsymbol{X}\in\mathbb{C}^{P\times Q}$. The vectorization of a Kronecker product of three matrices, i.e., expanding $\operatorname{vec}\{\boldsymbol{A}\otimes\boldsymbol{X}\otimes\boldsymbol{B}\}$ in terms of $\operatorname{vec}\{\boldsymbol{X}\}$, can be obtained by applying (1.17) and (1.18) as

$$\operatorname{vec}(\boldsymbol{A}\otimes\boldsymbol{X}\otimes\boldsymbol{B}) = \operatorname{vec}(\boldsymbol{A}\otimes(\boldsymbol{X}\otimes\boldsymbol{B})) \tag{1.19}$$

$$= \left(\begin{bmatrix}\boldsymbol{I}_{Q\cdot L}\otimes\boldsymbol{a}_1\\ \vdots\\ \boldsymbol{I}_{Q\cdot L}\otimes\boldsymbol{a}_N\end{bmatrix}\otimes\boldsymbol{I}_{P\cdot K}\right)\cdot\left(\boldsymbol{I}_Q\otimes\begin{bmatrix}\boldsymbol{I}_P\otimes\boldsymbol{b}_1\\ \vdots\\ \boldsymbol{I}_P\otimes\boldsymbol{b}_L\end{bmatrix}\right)\cdot\operatorname{vec}\{\boldsymbol{X}\} \tag{1.20}$$

for $\boldsymbol{A} = [\boldsymbol{a}_1,\dots,\boldsymbol{a}_N]\in\mathbb{C}^{M\times N}, \boldsymbol{X}\in\mathbb{C}^{P\times Q}, \boldsymbol{B} = [\boldsymbol{b}_1,\dots,\boldsymbol{b}_L]\in\mathbb{C}^{K\times L}$.

In some cases, it is required to exchange the order of the matrices in a Kronecker product. This can be achieved by commutation matrices [MN95]. The commutation matrix $\boldsymbol{K}_{M,N}\in\mathbb{R}^{MN\times MN}$ is a matrix that satisfies

$$\boldsymbol{K}_{M,N}\cdot\operatorname{vec}\left\{\boldsymbol{A}^{\mathrm{T}}\right\} = \operatorname{vec}\{\boldsymbol{A}\} \tag{1.21}$$

for any matrix $\boldsymbol{A} \in \mathbb{C}^{M\times N}$. As a consequence of (1.21), commutation matrices can be used to permute Kronecker products according to

$$\boldsymbol{K}_{M,N}^{\mathrm{T}} \cdot (\boldsymbol{A} \otimes \boldsymbol{B}) \cdot \boldsymbol{K}_{P,Q} = \boldsymbol{B} \otimes \boldsymbol{A} \tag{1.22}$$

for $\boldsymbol{A} \in \mathbb{C}^{M\times P}$ and $\boldsymbol{B} \in \mathbb{C}^{N\times Q}$.

Furthermore, we will use the matrix inversion lemma [GvL96], which is given by

$$(\boldsymbol{A} + \boldsymbol{U} \cdot \boldsymbol{C} \cdot \boldsymbol{V})^{-1} = \boldsymbol{A}^{-1} - \boldsymbol{A}^{-1} \cdot \boldsymbol{U} \cdot \left(\boldsymbol{C}^{-1} + \boldsymbol{V} \cdot \boldsymbol{A}^{-1} \cdot \boldsymbol{U}\right)^{-1} \cdot \boldsymbol{V} \cdot \boldsymbol{A}^{-1}, \tag{1.23}$$

where $\boldsymbol{A} \in \mathbb{C}^{M\times N}, \boldsymbol{U} \in \mathbb{C}^{M\times K}, \boldsymbol{C} \in \mathbb{C}^{K\times K}$, and $\boldsymbol{V} \in \mathbb{C}^{K\times M}$.

Finally, we define the short-hand notation

$$\boldsymbol{T}_{a:b}^{\otimes} = \begin{cases} \boldsymbol{T}_a \otimes \ldots \otimes \boldsymbol{T}_b & a \leq b \\ 1 & a > b \end{cases} \tag{1.24}$$

and

$$M_{a:b} = \begin{cases} \prod_{r=a}^{b} M_r & a \leq b \\ 1 & a > b. \end{cases} \tag{1.25}$$

1.4.2. Tensor algebra

In many signal processing applications, the multi-dimensional data is represented in a multi-dimensional (multi-way) array of numbers, which we refer to as a tensor. Thus, an R-dimensional (R-D) tensor is defined as a collection of numbers that are referenced by R indices. A 1-D tensor represents a vector and a 2-D tensor is a matrix. In the sequel, we will focus on higher-order tensors with $R \geq 2$ and introduce new notation and operations to manipulate the multi-linear data in its native R-dimensional form. These algebraic rules are often referred to as multi-linear algebra or tensor algebra.

Let $\boldsymbol{\mathcal{A}} \in \mathbb{C}^{M_1\times\ldots\times M_R}$ denote an R-D tensor or R-way with size M_r along the r-th dimension (or mode) for $r = 1, 2, \ldots, R$. The total number of elements contained in the R-way array is given by $M = \prod_{r=1}^{R} M_r$. The r-mode vectors of $\boldsymbol{\mathcal{A}}$ are obtained by varying the r-th index while keeping all the other indices fixed. Thus, r-mode vectors are a generalization of row vectors and columns vectors of matrices. An example for the r-mode vectors of a 3-D tensor is illustrated in Figure 1.1. The vector space, which is spanned by the r-mode vectors of $\boldsymbol{\mathcal{A}}$ is termed r-mode subspace or r-mode space of $\boldsymbol{\mathcal{A}}$. The collection of all r-mode vectors in a matrix of size $M_r \times \frac{M}{M_r}$ is called r-mode unfolding of $\boldsymbol{\mathcal{A}}$ and is denoted by $[\boldsymbol{\mathcal{A}}]_{(r)}$. The ordering of the columns is arbitrary when used consistently. Two popular choices are

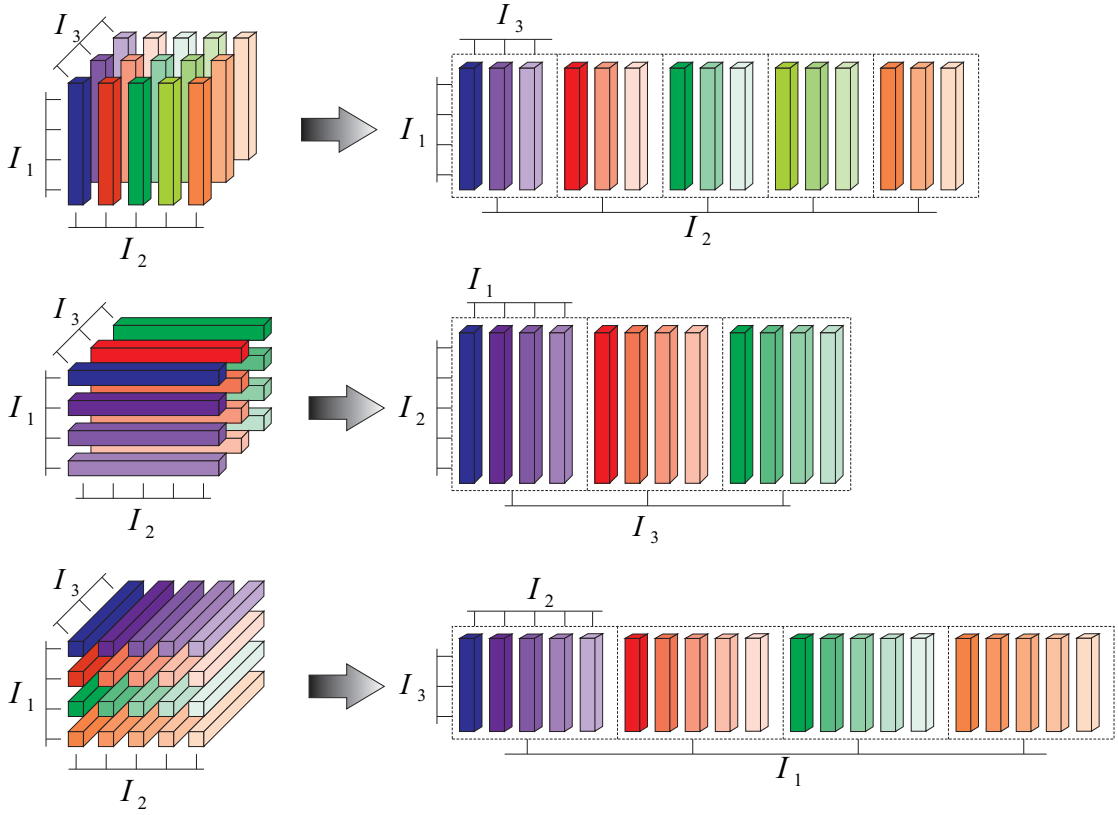


Figure 1.1.: Illustration of the unfoldings of a $4 \times 5 \times 3$ tensor in reverse cyclical column ordering.

- Forward column ordering: This ordering starts with the first index and increases up to the R-th index, while leaving out the r-th index. This ordering complies with the `reshape` command in MATLAB.

- Reverse cyclical [dLdMV00b]: This ordering starts with the $(r-1)$-th index and decreases up to the $(r+1)$-th index. As this is most commonly used ordering in signal processing, it is also chosen in this thesis. A 3-D example is shown in Figure 1.1.

The r-mode product of a tensor $\boldsymbol{\mathcal{A}} \in \mathbb{C}^{M_1 \times \ldots \times M_R}$ and a matrix $\boldsymbol{U}_r \in \mathbb{C}^{P_r \times M_r}$ is denoted as $\boldsymbol{\mathcal{B}} = \boldsymbol{\mathcal{A}} \times_r \boldsymbol{U}_r$. It can be interpreted as a linear transformation in the r-th mode, which is computed by multiplying all r-mode vectors by the matrix $\boldsymbol{U}_r$ from the left, i.e.,

$$[\boldsymbol{\mathcal{B}}]_{(r)} = \boldsymbol{U}_r \cdot [\boldsymbol{\mathcal{A}}]_{(r)} \,. \tag{1.26}$$

Repeated r-mode products are represented by the short-hand notation

$$\boldsymbol{\mathcal{A}} \bigtimes_{r=1}^{R}{}_r \boldsymbol{U}^{(r)} = \boldsymbol{\mathcal{A}} \times_1 \boldsymbol{U}_1 \times_2 \boldsymbol{A}_2 \ldots \times_R \boldsymbol{A}_R. \tag{1.27}$$

The rank of the r-mode unfolding, i.e., the dimension of the r-space of a tensor $\boldsymbol{\mathcal{A}}$, is called r-rank of $\boldsymbol{\mathcal{A}}$. Note that in contrast to the matrix case, where the column rank is equal to the row rank, the r-ranks of a tensor can be different. As a matrix of rank r can be constructed as a sum of r rank-one matrices, a tensor of rank r can be constructed as a sum of r rank-one tensors. The smallest possible integer r for this case is defined as the tensor rank. An R-D tensor of rank-one can be written as the outer product of R non-zero vectors. Note that the tensor rank is not directly related to the r-ranks as it only provides an upper bound [KB09], i.e., $\operatorname{rank}(\boldsymbol{\mathcal{X}}) \geq r\text{-rank}(\boldsymbol{\mathcal{X}})$ for $r = 1, 2, \ldots, R$.

The outer product of the tensors $\boldsymbol{\mathcal{A}} \in \mathbb{C}^{I_1 \times I_2 \times \ldots \times I_N}$ and $\boldsymbol{\mathcal{B}} \in \mathbb{C}^{J_1 \times J_2 \times \ldots \times J_M}$ is given by

$$\begin{aligned} &\boldsymbol{\mathcal{C}} = \boldsymbol{\mathcal{A}} \circ \boldsymbol{\mathcal{B}} \in \mathbb{C}^{I_1 \times \ldots \times I_N \times J_1 \times \ldots \times J_M}, \quad \text{where} \\ &c_{i_1, i_2, \ldots, i_N, j_1, j_2, \ldots, j_M} = a_{i_1, i_2, \ldots, i_N} \cdot b_{j_1, j_2, \ldots, j_M}. \end{aligned} \tag{1.28}$$

In other words, the tensor $\boldsymbol{\mathcal{C}}$ contains all possible combinations of pairwise products between all the elements of $\boldsymbol{\mathcal{A}}$ and all the elements of $\boldsymbol{\mathcal{B}}$. It is the generalization of the outer product of two vectors $\boldsymbol{a}$ and $\boldsymbol{b}$ for $R_1 = R_2 = 1$, which represents a matrix, i.e., $\boldsymbol{a} \circ \boldsymbol{b} = \boldsymbol{a} \cdot \boldsymbol{b}^{\mathrm{T}}$.

The concatenation of two tensors $\boldsymbol{\mathcal{A}}$ and $\boldsymbol{\mathcal{B}}$ along the r-th mode is defined in [HRD08] via $[\boldsymbol{\mathcal{A}} \llcorner_r \boldsymbol{\mathcal{B}}]$. Note that the two tensors can only be concatenated along the r-th mode if they have the same size in all modes $q \neq r$ for $q = 1, 2, \ldots, R$. The r-mode vectors of $[\boldsymbol{\mathcal{A}} \llcorner_r \boldsymbol{\mathcal{B}}]$ are given by the r-mode vectors of $\boldsymbol{\mathcal{A}}$ stacked on top of the r-mode vectors of $\boldsymbol{\mathcal{B}}$ according to

$$[\boldsymbol{\mathcal{A}} \llcorner_r \boldsymbol{\mathcal{B}}]_{(r)} = \begin{bmatrix} [\boldsymbol{\mathcal{A}}]_{(r)} \\ [\boldsymbol{\mathcal{B}}]_{(r)} \end{bmatrix}. \tag{1.29}$$

The r-mode products, r-mode unfoldings, and r-mode concatenations satisfy the following properties [Roe13]:

$$[\boldsymbol{\mathcal{A}} \times_1 \boldsymbol{U}_1 \ldots \times_R \boldsymbol{U}_R]_{(r)} = \boldsymbol{U}_r \cdot [\boldsymbol{\mathcal{A}}]_{(r)} \cdot (\boldsymbol{U}_{r+1} \otimes \ldots \otimes \boldsymbol{U}_R \otimes \boldsymbol{U}_1 \otimes \ldots \otimes \boldsymbol{U}_{r-1})^{\mathrm{T}} \tag{1.30}$$

$$\boldsymbol{\mathcal{A}} \times_r \boldsymbol{U}_r \times_p \boldsymbol{U}_p = \boldsymbol{\mathcal{A}} \times_p \boldsymbol{U}_p \times_r \boldsymbol{U}_r \quad \text{with } r \neq p \tag{1.31}$$

$$\boldsymbol{\mathcal{A}} \times_r \boldsymbol{U}_r \times_r \boldsymbol{V}_r = \boldsymbol{\mathcal{A}} \times_r (\boldsymbol{V}_r \cdot \boldsymbol{U}_r) \tag{1.32}$$

$$[\boldsymbol{\mathcal{A}} \llcorner_r \boldsymbol{\mathcal{B}}] + [\boldsymbol{\mathcal{C}} \llcorner_r \boldsymbol{\mathcal{D}}] = [(\boldsymbol{\mathcal{A}} + \boldsymbol{\mathcal{C}}) \llcorner_r (\boldsymbol{\mathcal{B}} + \boldsymbol{\mathcal{D}})] \tag{1.33}$$

$$[\boldsymbol{\mathcal{A}} \llcorner_r \boldsymbol{\mathcal{B}}] \times_p \boldsymbol{U}_p = [\boldsymbol{\mathcal{A}} \times_p \boldsymbol{U}_p \llcorner_r \boldsymbol{\mathcal{B}} \times_p \boldsymbol{U}_p] \quad \text{where } r \neq p \tag{1.34}$$

$$[\boldsymbol{\mathcal{A}} \llcorner_r \boldsymbol{\mathcal{B}}] \times_r \begin{bmatrix} \boldsymbol{U}_r & \boldsymbol{W}_r \end{bmatrix} = \boldsymbol{\mathcal{A}} \times_r \boldsymbol{U}_r + \boldsymbol{\mathcal{B}} \times_r \boldsymbol{W}_r, \tag{1.35}$$

$$\boldsymbol{\mathcal{A}} \times_r \begin{bmatrix} \boldsymbol{X}_r \\ \boldsymbol{Y}_r \end{bmatrix} = \left[(\boldsymbol{\mathcal{A}} \times_r \boldsymbol{X}_r) \sqcup_r (\boldsymbol{\mathcal{A}} \times_r \boldsymbol{Y}_r) \right], \tag{1.36}$$

where $r, p \in \{1, 2, \ldots, R\}$ and the tensors and matrices have the dimensions $\boldsymbol{\mathcal{A}}, \boldsymbol{\mathcal{B}} \in \mathbb{C}^{M_1 \times \ldots \times M_R}$, $\boldsymbol{U}_r, \boldsymbol{W}_r \in \mathbb{C}^{N_r \times M_r}$, $\boldsymbol{V}_r \in \mathbb{C}^{P_r \times N_r}$, $\boldsymbol{X}_r \in \mathbb{C}^{N_r \times M_r}$, and $\boldsymbol{Y}_r \in \mathbb{C}^{Q_r \times M_r}$.

The tensor $\boldsymbol{\mathcal{I}}_{R,N}$ denotes the R-D identity tensor of size $N \times N \ldots \times N$, whose elements are equal to one if all R indices are equal and zero elsewhere [KB09]. The identity tensor represents an extension of the identity matrix $\boldsymbol{I}_N$ of size $N \times N$.

The unfoldings of repeated r-mode products of the identity tensor $\boldsymbol{\mathcal{I}}_{R,N}$ in (1.30) simplify to

$$\left[\boldsymbol{\mathcal{I}}_{R,N} \bigtimes_{r=1}^{R} \boldsymbol{F}_r \right]_{(p)} = \boldsymbol{F}_p \cdot (\boldsymbol{F}_{p+1} \diamond \ldots \diamond \boldsymbol{F}_R \diamond \boldsymbol{F}_1 \diamond \ldots \diamond \boldsymbol{F}_{p-1})^{\mathrm{T}}. \tag{1.37}$$

As an extension of the vec-operator applied to matrices (cf. Equation (1.13)), we define the vec-operator applied to a tensor $\mathrm{vec}\{\boldsymbol{\mathcal{A}}\}$, where all elements are stacked into a column vector of size $M \times 1$ by aligning the indices in ascending order starting with the first up to the R-th index. This ordering complies with the command `A(:)` in `MATLAB`.

According to (1.30), the r-mode multiplication can be transformed into a matrix multiplication by means of the r-mode unfolding. However, the vectorized versions of different unfoldings contain the same elements in different order. To restore the correct ordering, permutation matrices can be defined. For every tensor $\boldsymbol{\mathcal{A}} \in \mathbb{C}^{M_1 \times M_2 \ldots \times M_R}$, there exists a unique set of permutation matrices $\boldsymbol{P}^{(r)}_{M_1, M_2, \ldots, M_R}$ such that [Roe13]

$$\mathrm{vec}\{\boldsymbol{\mathcal{A}}\} = \boldsymbol{P}^{(r)}_{M_1, M_2, \ldots, M_R} \cdot \mathrm{vec}\left\{ [\boldsymbol{\mathcal{A}}]_{(r)} \right\}, \quad r = 1, \ldots, R. \tag{1.38}$$

The permutation matrices satisfy $\boldsymbol{P}^{(r)^{-1}}_{M_1, M_2, \ldots, M_R} = \boldsymbol{P}^{(r)^{\mathrm{T}}}_{M_1, M_2, \ldots, M_R}$ for $r = 1, \ldots, R$ and can be interpreted as an extension of the commutation matrices $\boldsymbol{K}_{M,N}$ defined in (1.21). Note that we have $\boldsymbol{K}_{M,N} = \boldsymbol{P}^{(2)}_{N,M}$ since $[\boldsymbol{A}]_{(2)} = \boldsymbol{A}^{\mathrm{T}}$ for a matrix $\boldsymbol{A}$. As a special property, the permutation matrices in (1.38) can perform cyclic shifts on R-fold repeated Kronecker products. For instance, the matrices $\boldsymbol{P}^{(R)}_{J_1, J_2, \ldots, J_R}$ reverse the order of a Kronecker product, i.e., for a set of matrices $\boldsymbol{X}_r \in \mathbb{C}^{J_r \times I_r}$, $r = 1, \ldots, R$, we have

$$\bigotimes_{r=R}^{1} \boldsymbol{X}_r \cdot \boldsymbol{P}^{(R)}_{I_1, I_2, \ldots, I_R} = \boldsymbol{P}^{(R)}_{J_1, J_2, \ldots, J_R} \cdot \bigotimes_{r=1}^{R} \boldsymbol{X}_r. \tag{1.39}$$

The matrices $\boldsymbol{P}^{(r)}_{J_1, J_2, \ldots, J_R}$ for $r < R$ first reverse the order and then accomplish a cyclic shift of

Kronecker products by q terms as

$$\bigotimes_{r=R}^{1} \boldsymbol{X}_r \cdot \boldsymbol{P}^{(q)}_{I_1,I_2,\ldots,I_R} = \boldsymbol{P}^{(q)}_{J_1,J_2,\ldots,J_R} \cdot \bigotimes_{r=q+1}^{R} \boldsymbol{X}_r \otimes \bigotimes_{r=1}^{q} \boldsymbol{X}_r. \tag{1.40}$$

In the case of a repeated Kronecker product of vectors $\boldsymbol{x}_r \in \mathbb{C}^{1\times I_r}, r = 1,\ldots,R$, the matrix $\boldsymbol{P}^{(q)}_{J_1,\ldots,J_R}$ simplifies to $\boldsymbol{P}^{(q)}_{1,\ldots,1} = 1$ and consequently, the property in (1.40) results in

$$\bigotimes_{r=R}^{1} \boldsymbol{x}_r \cdot \boldsymbol{P}^{(q)}_{I_1,I_2,\ldots,I_R} = \bigotimes_{r=q+1}^{R} \boldsymbol{x}_r \otimes \bigotimes_{r=1}^{q} \boldsymbol{x}_r. \tag{1.41}$$

Moreover, for $r = 1,2,\ldots,R-1$, we have the general rule [Roe13]

$$\operatorname{vec}\left\{[\boldsymbol{\mathcal{A}}]_{(r)}\right\} = \operatorname{vec}\left\{[\boldsymbol{\mathcal{A}}]^{\mathrm{T}}_{(r+1)}\right\} = \boldsymbol{K}_{M_{r+1},\frac{M}{M_{r+1}}} \cdot \operatorname{vec}\left\{[\boldsymbol{\mathcal{A}}]_{(r+1)}\right\}. \tag{1.42}$$

The higher-order norm of a tensor $\boldsymbol{\mathcal{A}}$, which is the extension of the Frobenius norm of a matrix, satisfies the property [dLdMV00b]

$$\|\boldsymbol{\mathcal{A}}\|_{\mathrm{H}} = \left\|[\boldsymbol{\mathcal{A}}]_{(r)}\right\|_{\mathrm{F}} = \left\|\operatorname{vec}\left\{[\boldsymbol{\mathcal{A}}]_{(r)}\right\}\right\|_2 \tag{1.43}$$

for $r = 1,2,\ldots,R$. Thus, it is equal to the Frobenius norm of an arbitrary unfolding of $\boldsymbol{\mathcal{A}}$.
A generalization of the singular value decomposition (SVD) of matrices to R-D tensors is given by the higher-order singular value decomposition (HOSVD), which is also called multi-linear SVD (MLSVD) [dLdMV00b]. The HOSVD of a tensor $\boldsymbol{\mathcal{X}}_0 \in \mathbb{C}^{M_1\times\ldots\times M_R}$ is given by

$$\boldsymbol{\mathcal{X}}_0 = \boldsymbol{\mathcal{S}} \times_1 \boldsymbol{U}_1 \ldots \times_R \boldsymbol{U}_R = \boldsymbol{\mathcal{S}} \bigtimes_{r=1}^{R}{}_r \boldsymbol{U}_r, \tag{1.44}$$

where $\boldsymbol{\mathcal{S}} \in \mathbb{C}^{p_1\times\ldots\times p_R}$ is the core tensor and $p_r = \operatorname{rank}\left\{[\boldsymbol{\mathcal{X}}]_{(r)}\right\}$ represents the r-rank of $\boldsymbol{\mathcal{X}}$ for $r = 1,\ldots,R$. Moreover, $\boldsymbol{U}_r \in \mathbb{C}^{M_r\times p_r}$ are the unitary matrices that span the r-spaces of $\boldsymbol{\mathcal{X}}$ and can be computed from the SVDs of the r-mode unfoldings of $\boldsymbol{\mathcal{X}}$. An example of the HOSVD of a 3-D tensor is visualized in Figure 1.2.

In many multi-dimensional signal processing applications, we often observe low-rank tensors $\boldsymbol{\mathcal{X}}_0$ that are superimposed by a noise tensor $\boldsymbol{\mathcal{N}}$, i.e., we have

$$\boldsymbol{\mathcal{X}} = \boldsymbol{\mathcal{X}}_0 + \boldsymbol{\mathcal{N}}. \tag{1.45}$$

A low-rank approximation $\hat{\boldsymbol{\mathcal{X}}}_0$ of $\boldsymbol{\mathcal{X}}$ can be obtained from the truncated HOSVD, which is achieved by truncating the core tensor and the r-spaces provided that p_r of $\boldsymbol{\mathcal{X}}_0$ is known or has been

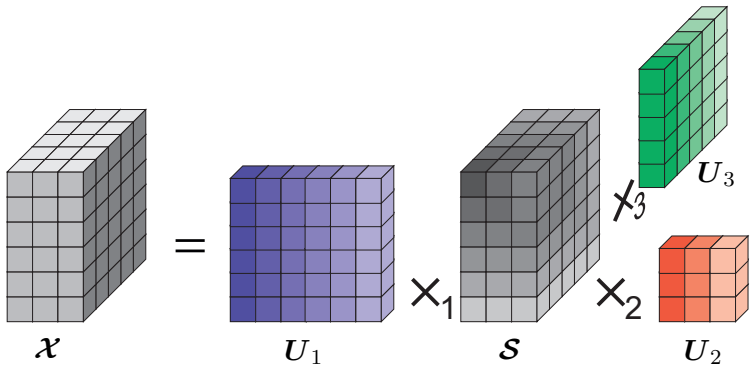


Figure 1.2.: Illustration of the HOSVD of a 3-D tensor $\boldsymbol{\mathcal{X}}$ of size $6 \times 3 \times 5$.

estimated. Thus, the truncated HOSVD of a tensor $\boldsymbol{\mathcal{X}}$ is given by

$$\hat{\boldsymbol{\mathcal{X}}}_0 \approx \hat{\boldsymbol{\mathcal{S}}}^{[\mathrm{s}]} \times_1 \hat{\boldsymbol{U}}_1^{[\mathrm{s}]} \ldots \times_R \hat{\boldsymbol{U}}_R^{[\mathrm{s}]} = \hat{\boldsymbol{\mathcal{S}}}^{[\mathrm{s}]} \bigtimes_{r=1}^{R}{}_r \hat{\boldsymbol{U}}_r^{[\mathrm{s}]}. \tag{1.46}$$

where $\hat{\boldsymbol{\mathcal{S}}}^{[\mathrm{s}]}$ and $\hat{\boldsymbol{U}}_r^{[\mathrm{s}]}$ are the estimated versions of $\boldsymbol{\mathcal{S}}$ and $\boldsymbol{U}_r$. It has been shown in [dLdMV00b] that the truncated HOSVD is asymptotically optimal in the high signal-to-noise ratio (SNR) regime.

2. Data model

Multi-dimensional harmonic retrieval is a fundamental problem in signal processing and arises in a broad variety of applications. The objective of this problem is to estimate the frequency parameters of multi-dimensional harmonics that are sampled on a multi-dimensional lattice. A number of signal processing applications, where the harmonic retrieval problem occurs, as well as a literature review are given in the introduction in Section 1.1. In this chapter, we introduce the underlying multi-dimensional data model, which will be the basis for the multi-dimensional parameter estimation algorithms developed and analyzed in this thesis. A common approach towards improving the estimation accuracy of parameter estimation algorithms is to exploit inherent signal structure. In this thesis, we develop and analytically evaluate the performance of algorithms that take two specific types of signal structure into account. On the one hand, we will show in Section 3.5 that the multi-dimensional structure of the measurement data can be exploited by using tensor-based signal processing, which preserves the multilinear structure. On the other hand, in Section 6.2, we will exploit the statistical properties of the received signal waveforms if the signals exhibit a strictly second-order (SO) non-circular (NC) structure by applying an algebraic preprocessing step. Furthermore, we show in Section 6.3 that exploiting both types of structure significantly reduces the parameter estimation error compared to the conventional parameter estimation algorithms.

In this chapter, we first present the multi-dimensional data model in Section 2.1 and then present the model for strictly non-circular signals in Section 2.2.

2.1. Multi-dimensional data model

In order to formulate the data model in the most generic way, we start by mathematically defining R-dimensional signals in Section 2.1.1. A matrix-based and a tensor-based formulation of the model is then shown in Section 2.1.2 and Section 2.1.3.

2.1.1. Scalar multi-dimensional data model

In this section, we first introduce the abstract definition of a separable multi-dimensional complex exponential with respect to a dimension p. As shown later using examples, this abstract dimension can be associated with several physical dimensions, such as, time, frequency, or space.

The formal definition is given as follows:

Definition 2.1.1. A noise-free R-dimensional ($R > 1$) undamped[1] exponential (complex sinusoid)

[1] Note that damped exponentials with exponentially decaying amplitudes can be considered by allowing μ to be complex, i.e., $\mu = \eta + \mathrm{j}\zeta$. Here, however, we assume undamped or non-decaying exponentials ($\zeta = 0$), where the amplitudes $s(t)$ are constant.

$x^{(0)}(t) \in \mathbb{C}$, which is separable in R continuous parameters $p^{(1)}, p^{(2)}, \ldots, p^{(R)} \in \mathbb{R}$ can be expressed as

$$x^{(0)}\left(p^{(1)}, p^{(2)}, \ldots, p^{(R)}, t\right) = s(t) \cdot \prod_{r=1}^{R} \mathrm{e}^{\mathrm{j}p^{(r)}\mu^{(r)}}, \tag{2.1}$$

where $s(t) \in \mathbb{C}$ is the constant complex amplitude of the exponential and $\mu^{(r)} \in \mathbb{R}$ denotes its frequency in the r-th dimension.

The multi-dimensional signal model defined in (2.1) arises in many signal processing applications. In practice, it is often the task to estimate the frequency μ of the complex exponential. Note that since μ is a generic parameter, estimating μ can refer to estimating a number of physical parameters. For example, if p is the time dimension, μ becomes the radial frequency $\omega = 2\pi f_\mathrm{c}$, where f_c is the frequency of the electromagnetic wave. On the other hand, if p is a spatial dimension, μ becomes the spatial frequency $\mu = k_\mathrm{c} \cdot f(\theta)$, where $k_\mathrm{c} = 2\pi/\lambda_\mathrm{c}$ is the wave number with wavelength λ_c of the exponential and $f(\theta)$ is the projection of k_c onto the spatial dimension p. Since $f(\theta)$ contains information on the angle of arrival θ of the wave relative to the receiver, the direction of arrival (DOA) of an impinging planar wavefront can be estimated. Other examples for estimating μ include estimating the delay, the Doppler, the polarization, etc. of the wave.

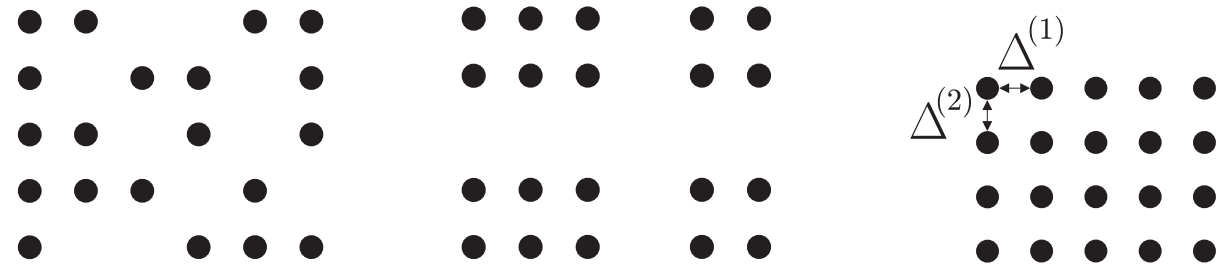
In order to estimate the frequency parameters $\mu_i^{(r)}$ for $i = 1, 2, \ldots, d$ and $r = 1, 2, \ldots, R$ of a linear superposition of d R-dimensional exponentials defined by (2.2), the exponentials need to be sampled in all R dimensions. To simplify the corresponding sampling model, we make the following assumptions:

(A1) The sources transmitting the complex exponentials are in the far-field region of the antenna array such that the electromagnetic waves impinging on the array approximately have a planar wavefront. In other words, each sensor element receives the electronic waves from the same angle.

(A2) The sources are point sources, i.e., the physical dimensions of the emitting sources are negligibly small such that each source is associated with a distinct angular direction.

(A3) The source and sensor positions are stationary during the observation time such that the (spatial) frequencies remain constant.

(A4) The received signals are narrow-band signals, i.e., the radio frequency (RF) bandwidth of the signals is negligible, such that the delay of the signals when traveling across the array translates into a phase shift. Moreover, the amplitudes of the narrow-band signals vary slowly



(a) Non-separable 2-D grid (b) Separable 2-D grid (c) Uniform 2-D grid

Figure 2.1.: Examples of 2-D sampling grids: (a) non-separable 2-D sampling grid; (b) separable 2-D sampling grid composed of the outer product of two (non-uniform) linear arrays; (c) uniform separable 2-D grid.

with respect to the propagation time across the array τ, i.e., $s(t-\tau) \approx s(t)$.

(A5) The source signals have an identical and known radio center frequency $f_c = c/\lambda_c$ such that the RF measurement data can be properly converted to the baseband. Here, λ_c and c denote the signals' wavelength and the speed of light, respectively.

(A6) To avoid spatial aliasing, the dimensions p are assumed to be orthogonal and the spacing $\Delta^{(r)} = p^{(r)}_{m_r} - p^{(r)}_{m_r-1}$ for $m_r = 2, \ldots, M_r$ and $r = 1, \ldots, R$ between the sampling points should be less than or equal to $\lambda_c/2$.

(A7) The sensor elements are omni-directional[2], there is no coupling between them and the array is perfectly calibrated.

Since the multi-dimensional exponentials in (2.1) are defined to be separable in the R variables, a separable R-dimensional sampling grid is required.

An R-dimensional sampling grid is called separable if it can be constructed from an outer product of R one-dimensional sampling grids. In other words, for each dimension, we can design the sampling arbitrarily but then all combinations of sampling points must be present in the sampled data. Examples of non-separable and separable 2-D sampling grids are shown in Figure 2.1a and Figure 2.1b, respectively. Figure 2.1c shows the special case of a 2-D uniform sampling grid introduced at the end of this section.

Let $p^{(r)}_1 < p^{(r)}_2 < \ldots < p^{(r)}_{M_r}$ be the sampling grid in the r-th dimension for $r = 1, 2, \ldots, R$. Then, an observation of the linear superposition of d R-dimensional exponentials on the sampling grid

[2]Note that the results are also applicable to arrays with common beam pattern as discussed in Section 2.1.4.1

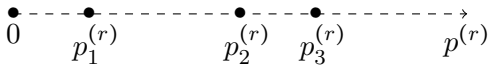


Figure 2.2.: Sampling grid along the dimension $p^{(r)}$.

under additive noise is given by

$$x_{m_1,m_2,\ldots,m_R}[n] = \sum_{i=1}^{d} \left(s_i[n] \cdot \prod_{r=1}^{R} \mathrm{e}^{\mathrm{j}p_{m_r}^{(r)}\mu_i^{(r)}} \right) + n_{m_1,m_2,\ldots,m_R}[n], \tag{2.2}$$

where $i = 1, 2, \ldots, d$ enumerates the signals, $m_r = 1, 2, \ldots, M_r$, specifies the M_r sampling points in the r-th dimension for $r = 1, 2, \ldots, R$, n is an integer index enumerating the N snapshots, i.e., $n = 1, 2, \ldots, N$, and $s[n] = s(t_n) \in \mathbb{C}$ is the amplitude of the signal x at the snapshot n. Moreover, $n_{m_1,m_2,\ldots,m_R}[n]$ represents the additive noise.

The sampling grid points along the r-th dimension are illustrated in Figure 2.2. Note that according to (A.1), the spatial sampling theorem needs to be satisfied to avoid spatial aliasing, i.e., the frequencies can be uniquely extracted. This aspect is further addressed in Section 2.1.4.1, where special sampling structures such as uniform sampling are discussed.

2.1.2. Matrix formulation

In order to obtain a more compact formulation of (2.2), we introduce a matrix-based model to describe the measurement data [HN98]. To this end, we collect the observed samples into an R-D measurement matrix $\boldsymbol{X} \in \mathbb{C}^{M \times N}$ with $M = \prod_{r=1}^{R} M_r$ by stacking the spatial dimensions $1, 2, \ldots, R$ along the rows and aligning the snapshots $n = 1, 2, \ldots, N$ along the columns. Subsequently, we obtain the matrix-based data model

$$\boldsymbol{X} = \boldsymbol{A} \cdot \boldsymbol{S} + \boldsymbol{N}, \tag{2.3}$$

where $\boldsymbol{S} \in \mathbb{C}^{d \times N}$ contains the amplitudes $s_i[n]$, $i = 1, 2, \ldots, d$, $n = 1, 2, \ldots, N$ and $\boldsymbol{N} \in \mathbb{C}^{M \times N}$ collects all the noise samples $n_{m_1,m_2,\ldots,m_R}[n]$.

Note that subspace-based parameter estimation schemes require the rank of the matrix $\boldsymbol{S}$ to be equal to d. If the $\operatorname{rank}\{\boldsymbol{S}\} < d$, preprocessing has to be applied, cf. Section 3.2.3.

Furthermore, $\boldsymbol{A} = [\boldsymbol{a}(\boldsymbol{\mu}_1), \ldots, \boldsymbol{a}(\boldsymbol{\mu}_d)] \in \mathbb{C}^{M \times d}$ is referred to as the array steering matrix, which consists of the array steering vectors $\boldsymbol{a}(\boldsymbol{\mu}_i)$ with $\boldsymbol{\mu}_i = [\mu_i^{(1)}, \ldots, \mu_i^{(R)}]^{\mathrm{T}} \in \mathbb{R}^{R \times 1}$ defined by

$$\boldsymbol{a}(\boldsymbol{\mu}_i) = \boldsymbol{a}^{(1)}\left(\mu_i^{(1)}\right) \otimes \cdots \otimes \boldsymbol{a}^{(R)}\left(\mu_i^{(R)}\right) \in \mathbb{C}^{M \times 1}, \tag{2.4}$$

where $\boldsymbol{a}^{(r)}(\mu_i^{(r)})$ is the array steering vector of the i-th frequency parameter in the r-th mode given

by

$$\boldsymbol{a}^{(r)}\left(\mu_i^{(r)}\right) = \left[\mathrm{e}^{\mathrm{j}p_1^{(r)}\mu_i^{(r)}} \quad \mathrm{e}^{\mathrm{j}p_2^{(r)}\mu_i^{(r)}} \quad \dots \quad \mathrm{e}^{\mathrm{j}p_{M_r}^{(r)}\mu_i^{(r)}}\right]^{\mathrm{T}} \in \mathbb{C}^{M_r \times 1}. \tag{2.5}$$

Note that $\boldsymbol{a}^{(r)}(\mu_i^{(r)})$ describes the response of the array in the r-th dimension with respect to the i-th exponential with the frequency parameter $\mu_i^{(r)}$.

An alternative formulation of $\boldsymbol{A}$ is given by

$$\boldsymbol{A} = \boldsymbol{A}^{(1)} \diamond \boldsymbol{A}^{(2)} \diamond \cdots \diamond \boldsymbol{A}^{(R)}, \tag{2.6}$$

where $\boldsymbol{A}^{(r)} = [\boldsymbol{a}^{(r)}(\mu_1^{(r)}), \dots, \boldsymbol{a}^{(r)}(\mu_d^{(r)})] \in \mathbb{C}^{M_r \times d}$ represents the array steering matrix in the r-th mode.

2.1.3. Tensor formulation

The drawback of the matrix-based model in (2.3) is that it does not represent the signal in its natural R-dimensional structure. This is due to the stacking of the multi-dimensional measurements into a matrix.

In order to arrive at a more structured formulation of the generic data model in (2.2), we collect the samples of the R-D signal $x_{m_1,m_2,\dots,m_R}[n]$ at N subsequent snapshots into an array. As our signal is referenced by $R+1$ indices, the most natural way of formulating the model is to employ an $(R+1)$-way array $\boldsymbol{\mathcal{X}} \in \mathbb{C}^{M_1 \times M_2 \dots \times M_R \times N}$ which contains $x_{m_1,m_2,\dots,m_R}[n]$ for $m_r = 1, 2, \dots, M_r$, $r = 1, 2, \dots, R$, and $n = 1, 2, \dots, N$. We can then conveniently express the model from (2.2) via the n-mode product introduced in Section 1.4.2 and obtain

$$\boldsymbol{\mathcal{X}} = \boldsymbol{\mathcal{A}} \times_{R+1} \boldsymbol{S}^{\mathrm{T}} + \boldsymbol{\mathcal{N}}, \tag{2.7}$$

where $\boldsymbol{S} \in \mathbb{C}^{d \times N}$ contains the amplitudes $s_i[n]$, $i = 1, 2, \dots, d$, $n = 1, 2, \dots, N$ as in (2.3).

Moreover, $\boldsymbol{\mathcal{N}} \in \mathbb{C}^{M_1 \times M_2 \dots \times M_R \times N}$ collects all the noise samples $n_{m_1,m_2,\dots,m_R}[n]$ in the same manner as $\boldsymbol{\mathcal{X}}$. Finally, $\boldsymbol{\mathcal{A}} \in \mathbb{C}^{M_1 \times M_2 \dots \times M_R \times d}$ is referred to as the "array steering tensor" [HRD08]. It can be expressed by virtue of the concatenation operator defined in (1.29) via the concatenation of rank-1 tensors as

$$\boldsymbol{\mathcal{A}} = [\boldsymbol{\mathcal{A}}_1 \sqcup_{R+1} \boldsymbol{\mathcal{A}}_2 \sqcup_{R+1} \dots \sqcup_{R+1} \boldsymbol{\mathcal{A}}_d] \tag{2.8}$$

$$\boldsymbol{\mathcal{A}}_i = \boldsymbol{a}^{(1)}\left(\mu_i^{(1)}\right) \circ \boldsymbol{a}^{(2)}\left(\mu_i^{(2)}\right) \circ \dots \circ \boldsymbol{a}_i^{(R)}\left(\mu_i^{(R)}\right) \in \mathbb{C}^{M_1 \times M_2 \times \dots \times M_R}, \tag{2.9}$$

where $\boldsymbol{a}^{(r)}(\mu_i^{(r)})$ is given in (2.5) and $\circ$ represents the outer product operator introduced in Section 1.4.2. Equation (2.9) shows how the necessary assumption of separability in defining R-D

signals translates to the outer product structure across dimensions in the array steering tensors.

An alternative expression for the array steering tensor is given by

$$\boldsymbol{\mathcal{A}} = \boldsymbol{\mathcal{I}}_{R+1,d} \times_1 \boldsymbol{A}^{(1)} \times_2 \boldsymbol{A}^{(2)} \ldots \times_R \boldsymbol{A}^{(R)} = \boldsymbol{\mathcal{I}}_{R+1,d} \bigtimes_{r=1}^{R}{}_r \boldsymbol{A}^{(r)}, \tag{2.10}$$

where $\boldsymbol{A}^{(r)} = \left[\boldsymbol{a}_1^{(r)}, \ldots, \boldsymbol{a}_d^{(r)}\right] \in \mathbb{C}^{M_r \times d}$ is referred to as the array steering matrix in the r-th mode.

The strength of the data model in (2.7) is that it represents the signal in its natural R-dimensional structure by virtue of the measurement tensor $\boldsymbol{\mathcal{X}}$. Applying the stacking operation introduced in Section 2.1.2 to (2.7), which is equivalent to writing $\boldsymbol{X} = [\boldsymbol{\mathcal{X}}]_{(R+1)}^{\mathrm{T}} \in \mathbb{C}^{M \times N}$, we again arrive at the matrix-based data model (2.3) [HN98]. Here, $\boldsymbol{A} = [\boldsymbol{\mathcal{A}}]_{(R+1)}^{\mathrm{T}} \in \mathbb{C}^{M \times d}$ and $\boldsymbol{N} = [\boldsymbol{\mathcal{N}}]_{(R+1)}^{\mathrm{T}} \in \mathbb{C}^{M \times N}$. Note that since $\boldsymbol{\mathcal{A}}$ obeys (2.10) we immediately find [HN98]

$$\boldsymbol{A} = [\boldsymbol{\mathcal{A}}]_{(R+1)}^{\mathrm{T}} = \boldsymbol{A}^{(1)} \diamond \boldsymbol{A}^{(2)} \diamond \ldots \diamond \boldsymbol{A}^{(R)} \tag{2.11}$$

by applying identity (1.37).

Therefore, whenever we encounter a linear mixture model like (2.3) where the "mixing matrix" $\boldsymbol{A}$ can be decomposed into a Khatri-Rao product of smaller mixing matrices according to (2.11), an alternative representation of the matrix-valued data model is given by the tensor-valued data model (2.7). We can typically benefit from the more explicit representation of the R-dimensional structure in (2.7) via a tensor, e.g., filter out the unstructured noise more efficiently, as shown in Section 3.2.2.

2.1.4. Special sampling structures

In this section, we discuss specific array structures that are often used in practice. In particular, we introduce uniform array geometries in Section 2.1.4.1 and centro-symmetric array geometries in Section 2.1.4.2.

2.1.4.1. Uniform sampling grid

So far, the data model is applicable to any separable R-D sampling grid. An important special case is given by uniform sampling, which means that the sampling points are uniformly spaced. Mathematically, we have $p_{m_r}^{(r)} - p_{m_r-1}^{(r)} = \Delta^{(r)}$ for $m_r = 2, \ldots, M_r$ and $r = 1, \ldots, R$, where $\Delta^{(r)}$ denotes the spacing between the sampling points. This implies that the elements of the antenna array are uniformly spaced, i.e., we have a uniform linear array (ULA) in the one-dimensional case (shown in Figure 2.3) or a uniform rectangular array (URA) in the two-dimensional case (see Figure 2.1c). In order to avoid aliasing, where the frequencies cannot be uniquely recovered anymore, the sample interval $\Delta^{(r)}$ is required to fulfill the sampling theorem [Van02].

Therefore, if the frequencies are known to be within the range $-\mu_{\max}^{(r)} \leq \mu^{(r)} \leq \mu_{\max}^{(r)}$, we require $\Delta^{(r)} \leq \frac{\pi}{\mu_{\max}^{(r)}}$. For example, if $p_{m_r}^{(r)}$ denotes the sampling of a spatial dimension, the corresponding spatial frequencies are within the range $-2\pi/\lambda_c \leq \mu^{(r)} \leq 2\pi/\lambda_c$, i.e., we have $\mu_{\max}^{(r)} = 2\pi/\lambda_c$. Thus, the sampling theorem reduces to $\Delta^{(r)} \leq \lambda_c/2$.

Often, the grid spacing $\Delta^{(r)}$, $\forall r = 1, 2, \ldots, R$, is included in the frequency parameter, which can then be defined as $\mu_i = \Delta^{(r)} \cdot \bar{\mu}_i$. In this case, the sampling grid reduces to integer numbers such that $p_{m_r}^{(r)} = p_1^{(r)} + m_r - 1$, where $p_1^{(r)}$ is an arbitrary phase reference point. Choosing, for instance, the first sensor element as the phase reference, i.e., $p_1^{(r)} = 0$, we obtain $p_{m_r}^{(r)} = m_r - 1$. Therefore, in the case of a uniform R-D array, i.e., uniform sampling is applied in each of the R dimensions, the model in 2.2 at time snapshot n is given by

$$x_{m_1,m_2,\ldots,m_R}[n] = \sum_{i=1}^{d} \left(s_i[n] \cdot \prod_{r=1}^{R} \mathrm{e}^{\mathrm{j}(m_r-1)\mu_i^{(r)}} \right) + n_{m_1,m_2,\ldots,m_R}[n] \tag{2.12}$$

for $m_r = 1, 2, \ldots, M_r$, $r = 1, 2, \ldots, R$ and $n = 1, 2, \ldots, N$. The generalization to arbitrary phase reference points is considered in Section 2.1.4.2. Then, for the case of uniform sampling, we obtain a simplified version of the array steering vectors $\boldsymbol{a}^{(r)}(\mu_i^{(r)})$ in (2.5) that are used in the matrix-based model (2.4) and the tensor-based model (2.9) as

$$\boldsymbol{a}^{(r)}\left(\mu_i^{(r)}\right) = \begin{bmatrix} 1 & \mathrm{e}^{\mathrm{j}\mu_i^{(r)}} & \mathrm{e}^{\mathrm{j}2\mu_i^{(r)}} & \ldots & \mathrm{e}^{\mathrm{j}(M_r-1)\mu_i^{(r)}} \end{bmatrix}^{\mathrm{T}} \in \mathbb{C}^{M_r \times 1}. \tag{2.13}$$

It is apparent that (2.13) exhibits a Vandermonde structure due to the fact that we assumed omnidirectional sensor elements. Note that if the antenna elements have a complex beam pattern represented by $g_{m_r}(\mu_i^{(r)}) \in \mathbb{C}$, where m_r denotes the m_r-th antenna element, and all the elements have the same beam pattern, i.e., $g_1(\mu_i^{(r)}) = \ldots = g_{M_r}(\mu_i^{(r)})$, we can write $\boldsymbol{a}^{(r)}(\mu) = g(\mu_i^{(r)}) \cdot \bar{\boldsymbol{a}}^{(r)}(\mu_i^{(r)})$, where the array steering vector $\bar{\boldsymbol{a}}^{(r)}(\mu_i^{(r)})$ still has a Vandermonde structure. For simplicity, we here assume $g(\mu) = 1$.

The special cases of a one-dimensional (1-D) uniform linear array (ULA) and a two-dimensional (2-D) uniform rectangular array are introduced in Example 2.1.1 and Example 2.1.2, where we discuss the problem of direction of arrival (DOA) estimation as an example of parameter estimation.

Example 2.1.1. 1-D DOA estimation: Assume that d narrow-band signals impinging from distinct directions in the far-field are sampled by a uniform linear array (ULA) composed of M isotropic sensor elements as illustrated in Figure 2.3. The task is to estimate the azimuth angles θ_i, i.e., the directions of arrival of the d signals. In the 1-D case, the model (2.12) for uniform sampling grids simplifies significantly, such that the n-th time observation at the m-th antenna

element is given by

$$x_m[n] = \sum_{i=1}^{d} s_i[n] \cdot \mathrm{e}^{\mathrm{j}(m-1)\mu_i} + n_m[n] \tag{2.14}$$

for $m = 1, 2, \ldots, M$ and $n = 1, 2, \ldots, N$. Using matrix notation, the observed samples in (2.14) can be compactly collected into the measurement matrix $\boldsymbol{X} \in \mathbb{C}^{M \times N}$, which is the 1-D version of (2.3), according to

$$\boldsymbol{X} = \boldsymbol{A} \cdot \boldsymbol{S} + \boldsymbol{N}, \tag{2.15}$$

where $\boldsymbol{S} \in \mathbb{C}^{d \times N}$ denotes the symbol matrix and $\boldsymbol{N} \in \mathbb{C}^{M \times N}$ represents the additive sensor noise. Moreover, $\boldsymbol{A} = [\boldsymbol{a}(\mu_1), \ldots, \boldsymbol{a}(\mu_d)] \in \mathbb{C}^{M \times d}$ is the array steering matrix, which contains the array steering vectors $\boldsymbol{a}(\mu_i)$ corresponding to the spatial frequency μ_i given by

$$\boldsymbol{a}(\mu_i) = \begin{bmatrix} 1 & \mathrm{e}^{\mathrm{j}\mu_i} & \mathrm{e}^{\mathrm{j}2\mu_i} & \ldots & \mathrm{e}^{\mathrm{j}(M-1)\mu_i} \end{bmatrix}^{\mathrm{T}} \in \mathbb{C}^{M \times 1}. \tag{2.16}$$

The spatial frequencies μ_i are given by

$$\mu_i = \frac{2\pi}{\lambda_{\mathrm{c}}} \cdot \Delta \cdot \sin(\theta_i), \tag{2.17}$$

where λ_{c} is the signal wavelength and Δ is the interelement spacing as shown in Figure 2.3. If we assume $\Delta = \lambda_{\mathrm{c}}/2$, we have $\mu_i = \pi \cdot \sin(\theta_i)$. Thus, the spatial frequencies μ_i are restricted to the range $-\pi \leq \mu_i < \pi$. Note that the azimuth angle θ_i cannot be uniquely recovered since $\mu_i(\theta_i) = \mu_i(\pi - \theta_i)$ due to $\sin(\theta_i) = \sin(\pi - \theta_i), \forall \theta_i$, i.e., the azimuth angle θ_i is indistinguishable from the azimuth angle $\pi - \theta_i$. This implies that a ULA is unable to distinguish signals from its front or back side, which is known as "front-back ambiguity". To avoid these ambiguities, we assume a one-to-one mapping from the spatial frequencies to the azimuth angles θ_i, which restricts θ_i to the front side of the ULA with the angle range $-90° \leq \theta_i < 90°$.

Example 2.1.2. 2-D direction of arrival (DOA) estimation: Suppose that an $M_1 \times M_2$ uniform rectangular array (URA) with isotropic elements captures d narrow-band signals from far-field sources. It is our goal to estimate the azimuth and co-elevation angles θ_i and ϕ_i, i.e., the 2-D directions of arrival as defined in Figure 2.4.

Following the model (2.12) for uniform sampling grids, the observations at antenna m_1, m_2 and time snapshot n can be expressed as [MHZ96]

$$x_{m_1,m_2}[n] = \sum_{i=1}^{d} s_i[n] \cdot \prod_{r=1}^{2} \mathrm{e}^{\mathrm{j}(m_r-1)\mu_i^{(r)}} + n_{m_1,m_2}[n], \tag{2.18}$$

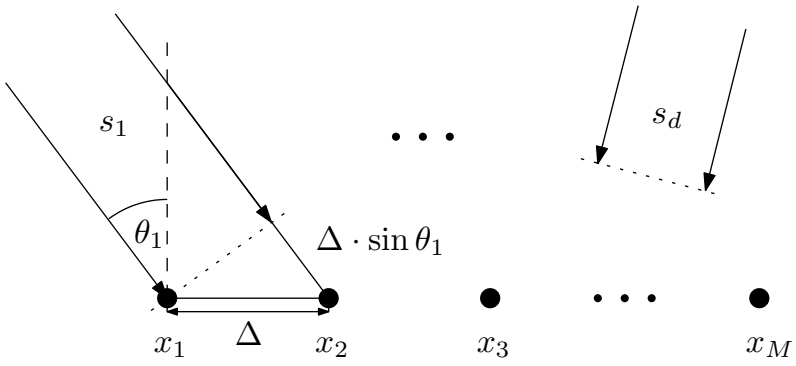


Figure 2.3.: Definition of the azimuth (θ) angles of an impinging planar wavefront for the 1-D DOA estimation in Example 2.1.1.

for $m_r = 1, 2, \ldots, M_r$, $r = 1, 2$ and $n = 1, 2, \ldots, N$. In this example, the parameters $\mu_i^{(1)}$ and $\mu_i^{(2)}$ are the spatial frequencies of the i-th signal, which are associated with the azimuth angle θ_i ($-180° < \theta_i \leq 180°$) and the co-elevation angle ϕ_i ($0° \leq \phi_i \leq 90°$) relative to the URA (illustrated in Figure 2.4) as [MHZ96]

$$\mu_i^{(1)} = \frac{2\pi}{\lambda_c} \cdot \Delta^{(1)} \cdot \underbrace{\cos(\theta_i) \cdot \sin(\phi_i)}_{u_i} \tag{2.19}$$

$$\mu_i^{(2)} = \frac{2\pi}{\lambda_c} \cdot \Delta^{(2)} \cdot \underbrace{\sin(\theta_i) \cdot \sin(\phi_i)}_{v_i}, \tag{2.20}$$

where u_i and v_i are the direction cosines relative to the x- and y-axes as defined in Figure 2.4. Assuming $\Delta^{(1)} = \Delta^{(2)} = \lambda_c/2$ and since $\xi_i = u_i + \mathrm{j}v_i = \sin(\phi) \cdot \mathrm{e}^{\mathrm{j}\theta_i}$, we can extract θ_i and ϕ_i via $\theta_i = \arg\{\xi_i\}$ and $\phi_i = \arcsin(|\xi_i|)$. Note that ϕ_i cannot be uniquely extracted from $\mu_i^{(1)}$ and $\mu_i^{(2)}$ as $\sin(\phi) = \sin(\pi - \phi), \forall\phi$. Geometrically, this ambiguity refers to the half space above and below the antenna array, which, however, can often be resolved via plausible reasoning.

2.1.4.2. Centro-symmetric sampling grid

Another important special type of array geometries are antenna arrays that are invariant under mirroring around their array centroids. Such antenna arrays are referred to as centro-symmetric arrays [XRK94]. A sensor configuration that is centro-symmetric in each of its R modes is said to be R-D centro-symmetric. For instance, ULAs and URAs introduced in the previous subsection are 1-D and 2-D centro-symmetric arrays. However, the 2-D antenna configuration illustrated in Figure 2.1b is only centro-symmetric in the vertical direction but not centro-symmetric in the horizontal direction. Therefore, this array is not 2-D centro-symmetric.

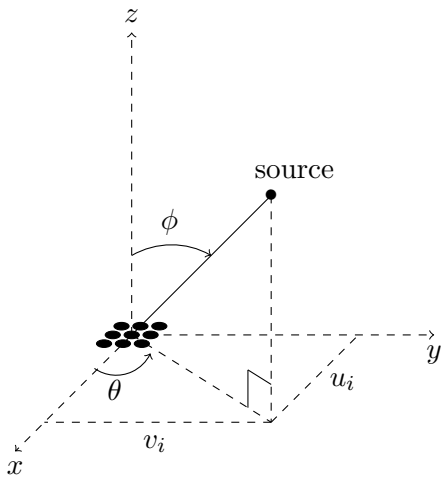


Figure 2.4.: Definition of azimuth (θ) and co-elevation (ϕ) angles of an impinging planar wavefront for the 2-D DOA estimation Example 2.1.2.

Mathematically, an R-D array is centro-symmetric in the r-th mode, if the corresponding steering matrix $\boldsymbol{A}^{(r)} \in \mathbb{C}^{M_r \times d}$ satisfies [HN95]

$$\boldsymbol{\Pi}_{M_r} \cdot \boldsymbol{A}^{(r)^*} = \boldsymbol{A}^{(r)} \cdot \boldsymbol{\Delta}_{\mathrm{c}}^{(r)}, \tag{2.21}$$

where $\boldsymbol{\Delta}_{\mathrm{c}}^{(r)} \in \mathbb{C}^{d \times d}$ is a unitary diagonal matrix which depends on the phase reference of the array in the r-th mode. Accordingly, an R-D array is R-D centro-symmetric if (2.21) is satisfied for each of the R modes. In this case, the overall array steering matrix $\boldsymbol{A} \in \mathbb{C}^{M \times d}$ satisfies

$$\boldsymbol{\Pi}_M \cdot \boldsymbol{A}^* = \boldsymbol{A} \cdot \boldsymbol{\Delta}_{\mathrm{c}}, \tag{2.22}$$

where $\boldsymbol{\Delta}_{\mathrm{c}} \in \mathbb{C}^{d \times d}$ now depends on the overall R-D phase reference of the R-D array.

In order to investigate the impact of the phase reference of the R-D array on the performance of R-D ESPRIT-type parameter estimation algorithms and the R-D Cramér-Rao bound, we extract the phase reference as a separate parameter by refining the matrix-based model in (2.3) using the array steering matrix formulation in (2.6). To this end, let the array centroid of the array along the r-th mode be defined by

$$\delta^{(r)} = \frac{1}{M_r} \sum_{m_r=1}^{M_r} p_{m_r}^{(r)}. \tag{2.23}$$

Note that $\delta^{(r)}$ is a property of the array and independent of $\mu_i^{(r)}$. Then, we can decompose the

r-mode array steering matrix $\boldsymbol{A}^{(r)}$ from (2.6) with an arbitrary array centroid as

$$\boldsymbol{A}^{(r)} = \boldsymbol{A}_{\mathrm{c}}^{(r)} \cdot \boldsymbol{\Delta}^{(r)}, \tag{2.24}$$

where $\boldsymbol{A}_{\mathrm{c}}^{(r)} \in \mathbb{C}^{M_r \times d}$ is the array steering matrix whose array centroid $\delta^{(r)}$ serves as the phase reference in the r-th mode such that $\boldsymbol{\Pi}_M \cdot \boldsymbol{A}_{\mathrm{c}}^{(r)^*} = \boldsymbol{A}_{\mathrm{c}}^{(r)}$. Furthermore, the diagonal matrix $\boldsymbol{\Delta}^{(r)} = \mathrm{diag}\{\mathrm{e}^{\mathrm{j}\delta^{(r)}\mu_i^{(r)}}\}_{i=1}^{d}$ contains the phase shifts depending on $\mu_i^{(r)}$ due to a non-centered phase reference $\delta^{(r)}$ on its diagonal.

Further, assuming the R-D array to be R-D centro-symmetric, i.e., (2.21) holds in the r-th mode, by inserting (2.24) into (2.21), we can easily establish the relation $\boldsymbol{\Delta}_{\mathrm{c}}^{(r)} = \boldsymbol{\Delta}^{(r)^*} \cdot \boldsymbol{\Delta}^{(r)^*}$. Alternatively, we can write $\boldsymbol{\Delta}^{(r)} = \boldsymbol{\Delta}_{\mathrm{c}}^{(r)^{-1/2}}$ since $\boldsymbol{\Delta}^{(r)^*} = \boldsymbol{\Delta}^{(r)^{-1}}$. Thus, if the actual phase reference of $\boldsymbol{A}^{(r)}$ is at the centroid of the r-th mode, we have $\boldsymbol{\Delta}^{(r)} = \boldsymbol{\Delta}_{\mathrm{c}}^{(r)} = \boldsymbol{I}_d$, and consequently $\boldsymbol{A}^{(r)} = \boldsymbol{A}_{\mathrm{c}}^{(r)}$. Based on (2.24), we can rewrite $\boldsymbol{A}$ in (2.6) as

$$\begin{aligned}
\boldsymbol{A} &= \left(\boldsymbol{A}_{\mathrm{c}}^{(1)} \cdot \boldsymbol{\Delta}^{(1)}\right) \diamond \left(\boldsymbol{A}_{\mathrm{c}}^{(2)} \cdot \boldsymbol{\Delta}^{(2)}\right) \diamond \cdots \diamond \left(\boldsymbol{A}_{\mathrm{c}}^{(R)} \cdot \boldsymbol{\Delta}^{(R)}\right) \\
&= \left(\boldsymbol{A}_{\mathrm{c}}^{(1)} \diamond \boldsymbol{A}_{\mathrm{c}}^{(2)} \diamond \cdots \diamond \boldsymbol{A}_{\mathrm{c}}^{(R)}\right) \cdot \left(\boldsymbol{\Delta}^{(1)} \cdot \boldsymbol{\Delta}^{(2)} \cdot \ldots \cdot \boldsymbol{\Delta}^{(R)}\right) \\
&= \boldsymbol{A}_{\mathrm{c}} \cdot \boldsymbol{\Delta},
\end{aligned} \tag{2.25}$$

where $\boldsymbol{A}_{\mathrm{c}} = \boldsymbol{A}_{\mathrm{c}}^{(1)} \diamond \boldsymbol{A}_{\mathrm{c}}^{(2)} \diamond \cdots \diamond \boldsymbol{A}_{\mathrm{c}}^{(R)} \in \mathbb{C}^{M \times d}$ and $\boldsymbol{\Delta} = \boldsymbol{\Delta}^{(1)} \cdot \boldsymbol{\Delta}^{(2)} \cdot \ldots \cdot \boldsymbol{\Delta}^{(R)} \in \mathbb{C}^{d \times d}$.

As an example, we consider a uniform sampling grid with half-wavelength sampling and choose the array centroid as the phase reference along the r-th mode. In this case, the corresponding steering vector $\boldsymbol{a}_{\mathrm{c}}^{(r)}(\mu_i^{(r)})$ is given by

$$\boldsymbol{a}_{\mathrm{c}}^{(r)}\left(\mu_i^{(r)}\right) = \left[\mathrm{e}^{-\mathrm{j}\frac{M_r-1}{2}\mu_i^{(r)}} \quad \mathrm{e}^{-\mathrm{j}\frac{M_r-3}{2}\mu_i^{(r)}} \quad \ldots \quad \mathrm{e}^{\mathrm{j}\frac{M_r-3}{2}\mu_i^{(r)}} \quad \mathrm{e}^{\mathrm{j}\frac{M_r-1}{2}\mu_i^{(r)}}\right]^{\mathrm{T}}. \tag{2.26}$$

Then, by applying (2.24) and setting $\delta^{(r)} = \frac{M_r-1}{2}$, we can move the array phase reference to the first sensor element as follows:

$$\begin{aligned}
\boldsymbol{a}^{(r)}\left(\mu_i^{(r)}\right) &= \boldsymbol{a}_{\mathrm{c}}^{(r)}\left(\mu_i^{(r)}\right) \cdot \mathrm{e}^{\mathrm{j}\delta^{(r)}\mu_i^{(r)}} \\
&= \left[\mathrm{e}^{-\mathrm{j}\frac{M_r-1}{2}\mu_i^{(r)}} \quad \mathrm{e}^{-\mathrm{j}\frac{M_r-3}{2}\mu_i^{(r)}} \quad \ldots \quad \mathrm{e}^{\mathrm{j}\frac{M_r-3}{2}\mu_i^{(r)}} \quad \mathrm{e}^{\mathrm{j}\frac{M_r-1}{2}\mu_i^{(r)}}\right]^{\mathrm{T}} \cdot \mathrm{e}^{\mathrm{j}\delta^{(r)}\mu_i^{(r)}} \\
&= \left[1 \quad \mathrm{e}^{\mathrm{j}\mu_i^{(r)}} \quad \mathrm{e}^{\mathrm{j}2\mu_i^{(r)}} \quad \ldots \quad \mathrm{e}^{\mathrm{j}(M_r-1)\mu_i^{(r)}}\right]^{\mathrm{T}}.
\end{aligned} \tag{2.27}$$

Using the relation in (2.25), we can express the data model in (2.3) in terms of $\boldsymbol{A}_{\mathrm{c}}$ as

$$\boldsymbol{X} = \boldsymbol{A}_{\mathrm{c}} \cdot \boldsymbol{\Delta} \cdot \boldsymbol{S} + \boldsymbol{N} = \boldsymbol{A}_{\mathrm{c}} \cdot \bar{\boldsymbol{S}} + \boldsymbol{N}, \tag{2.28}$$

where we have defined $\bar{\boldsymbol{S}} = \boldsymbol{\Delta} \cdot \boldsymbol{S}$.

In the tensor case, the tensor-based equivalent of the centry-symmetry property in (2.22) is given by

$$\boldsymbol{\mathcal{A}}^{*} \bigtimes_{r=1}^{R}{}_{r} \boldsymbol{\Pi}_{M_r} = \boldsymbol{\mathcal{A}} \times_{R+1} \boldsymbol{\Delta}_{\mathrm{c}}. \tag{2.29}$$

where $\boldsymbol{\Delta}_{\mathrm{c}}$ is the unitary diagonal matrix from (2.22).

2.2. Signal model for second-order non-circular signals

In the multi-dimensional data model (2.2), we have so far made no further assumptions on the amplitudes of the multi-dimensional signals, which represent the complex baseband data symbols. In the matrix-based data model (2.3) and in the tensor-based data model 2.7, the symbols are collected in the matrix $\boldsymbol{S}$. The symbols in $\boldsymbol{S}$ are usually assumed to be arbitrary complex random variables and the only restriction so far is that the rank of $\boldsymbol{S}$ should be equal to d. However, often times the complex baseband symbols of the received multi-dimensional signals possess specific statistical properties such as a strictly second-order (SO) non-circular (NC) structure. This particular type of signal structure often occurs in communication-type scenarios, where the transmitters employ real-valued digital modulation schemes, e.g., binary phase shift keying (BPSK), amplitude phase shift keying (ASK), minimum shift keying (MSK), offset-quadrature phase shift keying (OQPSK) (after a de-rotation), etc. As mentioned in Section 1.1.1, previous work [CWS01, ZCW03, HR04, RH09, AD06, LLXZ12] has shown that taking advantage of the strict non-circularity of the signals helps to improve the estimation accuracy and doubles the number of identifiable sources of the conventional parameter estimation algorithms. Specific applications for strictly non-circular signals are wireless communications, GNSS, cognitive radio, etc., where strictly non- circular sources are known to be present, and radar, tracking, channel sounding, etc., where the signals can be designed as strictly non-circular signals.

In Section 3.4, we will discuss the matrix-based R-D NC Standard ESPRIT and R-D NC Unitary ESPRIT algorithms and analyze their analytical performance in Section 4.3. Furthermore, in Section 3.5, we will discuss the tensor-based ESPRIT-type algorithms R-D NC Standard Tensor-ESPRIT and R-D NC Unitary Tensor-ESPRIT that exploit both the R-D structure and the NC structure of the signals. Their analytical performance expressions are derived and analyzed in Section 4.4.

Before we introduce the concept and the statistical properties of second-order non-circular signals, we first consider non-circular random variables in the next section.

2.2.1. Non-circular random variables

We first start with the statistical properties of circular random variables before moving on to non-circular random variables.

2.2.1.1. Circularly symmetric random variables

The full description of the statistical properties of complex random variables is not only given by the individual distribution of their real and imaginary parts but also by their joint distribution as there may be a correlation between the real and imaginary parts.

As a simplifying assumption, it is often presumed that complex random variables are second-order *circularly symmetric*. This property is defined as follows:

Definition 2.2.1. **[SS10]** A zero-mean complex random variable $Z = X + \mathrm{j} \cdot Y \in \mathbb{C}$, i.e., $\mathbb{E}\{Z\} = 0$, is termed second-order *circularly symmetric* iff it satisfies

$$\mathbb{E}\left\{Z^2\right\} = 0. \tag{2.30}$$

Expanding (2.30) yields

$$\mathbb{E}\left\{Z^2\right\} = \mathbb{E}\left\{(X + \mathrm{j} \cdot Y)^2\right\} = \mathbb{E}\left\{X^2\right\} - \mathbb{E}\left\{Y^2\right\} + 2 \cdot \mathrm{j} \cdot \mathbb{E}\{X \cdot Y\}. \tag{2.31}$$

Hence, we conclude from (2.31) that the condition $\mathbb{E}\left\{Z^2\right\} = 0$ implies that the real part X and the imaginary part Y are uncorrelated and have the same variance.

Geometrically, for example, if the random variables X and Y are Gaussian distributed, the contour lines of equal probability in the joint probability density function of X and Y represent circles as seen in Figure 2.5(a), which provides the reason why this property is termed circular symmetry.

2.2.1.2. Non-circular random variables

Consequently, if $\mathbb{E}\left\{Z^2\right\} \neq 0$, the complex random variable Z is second-order non-circular. The degree of non-circularity is usually measured via the "non-circularity rate" [DA04], which also referred to as "circularity coefficient" [EK06], or "circularity quotient" [Oll08] in the literature. The non-circularity rate is defined as the a complex scalar parameter

$$\zeta = \frac{\mathbb{E}\left\{Z^2\right\}}{\mathbb{E}\left\{|Z|^2\right\}} = |\zeta| \cdot \mathrm{e}^{\mathrm{j}\psi}, \tag{2.32}$$

where ψ is the non-circularity phase. It is apparent that $\zeta = |\zeta| = 0$ only if Z is circularly symmetric (see Section 2.2.1.1). Hence, the non-circularity rate can be interpreted as a measure for the

deviation from the circular symmetry. The maximum value of $|\zeta|$ is given in the following theorem:

Theorem 2.2.1. *For any complex random variable $Z = X + j \cdot Y$, the maximum non-circularity rate is given by*

$$|\zeta| \leq 1, \tag{2.33}$$

which holds with equality iff $c \cdot X = Y$ for some $c \in \mathbb{R}$, i.e., the real part and the imaginary part of Z are linearly dependent.

A proof of this theorem is given in Appendix B.1. A complex random variable with $0 < |\zeta| < 1$ is termed *weak-sense* second-order non-circular, while the case $|\zeta| = 1$ defines a *strictly* (or *strict-sense*) second-order non-circular random variable. Strictly non-circular random variables are also referred to as rectilinear [CP06] in the literature. From Theorem 2.2.1, we know that strict non-circularity implies a linear dependence between the real and the imaginary part of Z. Therefore, Z can, for instance, be represented as a real-valued random variable which is rotated by a complex phase term, i.e.,

$$Z = W \cdot \mathrm{e}^{\mathrm{j}\varphi} = W \cdot \cos(\varphi) + \mathrm{j} \cdot W \cdot \sin(\varphi), \tag{2.34}$$

where $W \in \mathbb{R}$ is a real-valued random variable and φ with $-\pi/2 \leq \varphi \leq \pi/2$ is the deterministic (fixed) rotation phase parameter. It is obvious that (2.34) fulfills the condition $c \cdot X = Y$ for some $c \in \mathbb{R}$. Alternatively, it is easily verified that

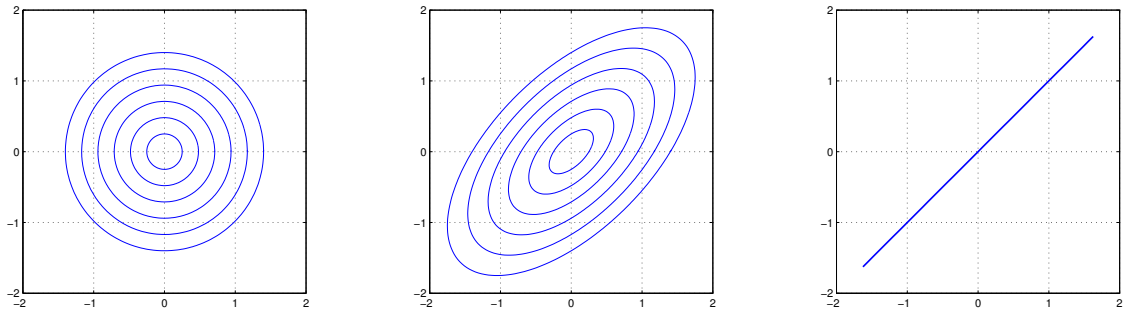
$$\zeta = \frac{\mathbb{E}\left\{Z^2\right\}}{\mathbb{E}\left\{|Z|^2\right\}} = \frac{\mathbb{E}\left\{W^2\right\} \cdot \mathrm{e}^{\mathrm{j}2\varphi}}{\mathbb{E}\left\{W^2\right\}} = \mathrm{e}^{\mathrm{j}2\varphi}, \tag{2.35}$$

which has $|\zeta| = 1$ and where the non-circularity phase is given by $\psi = 2 \cdot \varphi$.

The geometrical interpretation of non-circular Gaussian random variables is visualized in Figure 2.5(b) and Figure 2.5(c). Figure 2.5(b) shows the contour lines of constant probability density for a weak-sense non-circular Gaussian random variable with non-circularity rate $\zeta = 0.6 \cdot \mathrm{e}^{\mathrm{j}\pi/2}$ and Figure 2.5(c) illustrates the strictly non-circular case with $\zeta = \mathrm{e}^{\mathrm{j}\pi/2}$.

2.2.2. Strictly non-circular signals

In a communication system, the complex baseband symbols $s_i[n]$ of the received signals represent non-circular random variables if the symbols are drawn from constellation diagrams in the complex plane that are not circularly symmetric (sometimes also termed rotation invariant). The case of strictly non-circular (rectilinear) symbols presumes that the sources transmit signals with symbols from real-valued constellations (such as BPSK or M-ASK). Due to the distinct transmission delays



(a) $\zeta = 0$ (b) $\zeta = 0.6 \cdot e^{j\pi/2}$ (c) $\zeta = e^{j\pi/2}$

Figure 2.5.: Examples for the contour lines of constant probability density for complex Gaussian random variables with different ζ: (a) circularly symmetric case, (b) weak-sense non-circular case, and (c) strictly non-circular case.

of the signals received from different sources, the symbols at the receiver appear on straight lines with different phase rotations in the complex plane. Note that OQPSK and MSK symbols can be transformed into real-valued amplitudes by applying an appropriate de-rotation at the receiver [CP06]. Figure 2.6 shows an example of the Inphase and Quadrature (I/Q) components of the received symbols of two sources that transmit a real-valued constellation. As the transmission of each source is subject to a different phase rotation φ_i, the receiver observes rotated real-valued random variables, which satisfy the strict non-circularity property.

In analogy to the model for strictly non-circular random variables in (2.34), the symbol matrix $\boldsymbol{S} \in \mathbb{C}^{d \times N}$ for strictly non-circular signals is modeled as [ZCW03]

$$\boldsymbol{S} = \boldsymbol{\Psi} \cdot \boldsymbol{S}_0, \tag{2.36}$$

where $\boldsymbol{S}_0 \in \mathbb{R}^{d \times N}$ is the real-valued symbol matrix and $\boldsymbol{\Psi} = \operatorname{diag}\left\{[e^{j\varphi_1}, \ldots, e^{j\varphi_d}]\right\}$ contains the arbitrary complex phase shifts on its diagonal. Note that in this model, the complex phases are assumed to be stationary, i.e., they do not change over time.

It is clear from (2.36) that the factorization of $\boldsymbol{S}$ reveals a particular signal structure, i.e., only real-valued random variables as well as different stationary complex phase shift for each source. This specific structure can be exploited in signal processing applications if prior knowledge about the presence of strictly non-circular sources is available.

If $\boldsymbol{s}[n]$ is a non-circular random variable, its full second-order statistics are not only described by the covariance matrix $\boldsymbol{R}_{ss} = \mathbb{E}\left\{\boldsymbol{s}[n] \cdot \boldsymbol{s}^H[n]\right\}$, but also by the pseudo-covariance matrix $\boldsymbol{C}_{ss} = \mathbb{E}\left\{\boldsymbol{s}[n] \cdot \boldsymbol{s}^T[n]\right\}$. For circular random variables, the pseudo-covariance matrix $\boldsymbol{C}_{ss}$ is equal to the zero matrix. However, for non-circular sources the pseudo-covariance matrix contains additional statistical information about $\boldsymbol{s}[n]$ that can be taken advantage of.

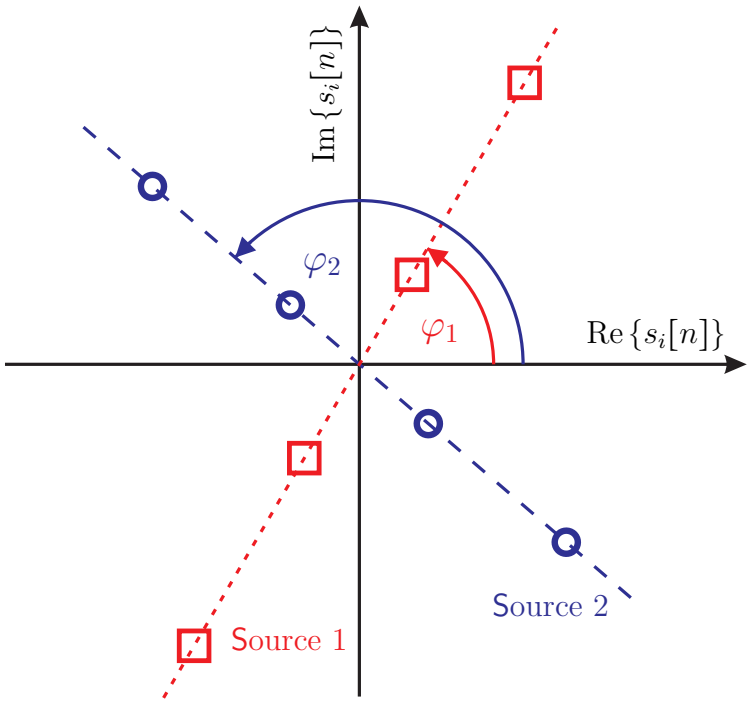


Figure 2.6.: Example for strictly non-circular amplitudes: Two sources (red, blue) transmit symbols drawn from real-valued constellations (4-ASK). As they undergo different transmission delays, the I/Q diagram at the receiver consists of differently rotated real-valued random variables, i.e., the complex symbols $s_i[n]$ represent strictly non-circular random variables.

Inserting the factorization (2.36) into the matrix-based data model (2.3), we obtain the following data model for strictly non-circular sources:

$$\boldsymbol{X} = \boldsymbol{A} \cdot \boldsymbol{\Psi} \cdot \boldsymbol{S}_0 + \boldsymbol{N}. \tag{2.37}$$

Analogously, the corresponding tensor model for strictly non-circular sources is given by

$$\boldsymbol{\mathcal{X}} = \boldsymbol{\mathcal{A}} \times_{R+1} \left(\boldsymbol{\Psi} \cdot \boldsymbol{S}_0\right)^{\mathrm{T}} + \boldsymbol{\mathcal{N}}. \tag{2.38}$$

3. ESPRIT-type parameter estimation algorithms

This chapter provides a state of the art of ESPRIT-type parameter estimation algorithms and will serve as a basis for the contributions presented in the other chapters of this thesis.

3.1. Overview

In this chapter, we review multi-dimensional ESPRIT-type algorithms to extract the parameters of R-dimensional signals sampled on a separable R-D grid as defined in Section 2. The advantage of ESPRIT-type parameter estimation algorithms compared to other subspace-based algorithms is that they provide closed-form parameter estimates and therefore only require a low computational complexity as no peak search is needed. Specifically, we will discuss the matrix-based R-D Standard ESPRIT and R-D Unitary ESPRIT algorithms [HN98] as well as their tensor-extensions R-D Standard Tensor-ESPRIT and R-D Unitary Tensor-ESPRIT [HRD08].

In Section 3.2, we first recall the matrix-based subspace estimation via the singular value decomposition (SVD) and then show how the estimation accuracy can be improved by incorporating the multi-dimensional structure of the signals via the higher-order singular value decomposition (HOSVD), which is also referred to as the multi-linear singular value decomposition (MLSVD). As the subspace estimation is required for any multi-dimensional subspace-based parameter estimation algorithm, the HOSVD-based subspace estimate [RHD06, HRD08] can be applied to R-D versions of, e.g., MODE [Van02], RARE [PGW02], or MUSIC [Sch79] as well (cf. Section 1.1.1 for a literature overview) to improve the estimation accuracy. The preprocessing steps forward-backward averaging (FBA) and spatial smoothing are discussed as well.

After the subspace estimation, the second step of ESPRIT-type algorithms is finding the solution to the shift invariance equation, an overdetermined, linear system of equations. Section 3.3 first reviews the matrix-based version of the shift invariance equation and then formulates its tensor-based extension in order to preserve the multi-linear structure.

We then summarize the matrix-based R-D ESPRIT-type algorithms along with the R-D Tensor-ESPRIT-type algorithms in Section 3.4 and highlight that both are algebraically equivalent apart from replacing the SVD-based subspace estimate by the HOSVD-based subspace estimate [HRD08]. Here, we apply the simple least squares (LS) method to solve the shift invariance equation.

Finally, conclusions are drawn in Section 3.6.

3.2. Subspace estimation

The fundamental step of any subspace-based parameter estimation algorithm is the estimation of the signal subspace. The signal subspace is defined as the vector space spanned by the columns of the array steering matrix $\boldsymbol{A} \in \mathbb{C}^{M \times d}$ (cf. Section 2.1.2), which represent the principal components of the model and serve as a basis for the signal subspace. The orthogonal complement of the signal subspace in the vector space $\mathbb{C}^M$ is referred to as the noise subspace. The signal subspace estimation can be viewed as a dimensionality reduction and a noise reduction.

In this section, we first review the matrix-based subspace estimation in Section 3.2.1. In Section 3.2.2, we review the enhanced tensor-based subspace estimation, which improves the accuracy of the signal subspace estimate by imposing the multi-dimensional structure of R-dimensional signals onto the unstructured subspace estimate. The tensor-based subspace estimate can be combined with any multi-dimensional subspace-based parameter estimation schemes, e.g., R-D Tensor MODE [WS15], R-D Tensor MUSIC [BGP+13], or R-D Tensor ESPRIT [HRD08]. We also provide the link of the tensor-based estimate subspace to the matrix-based subspace. Finally, the forward-backward averaging preprocessing and real-valued subspace estimation in the matrix and tensor case as well as the spatial smoothing preprocessing step are discussed in Section 3.2.3 and Section 3.2.4.

3.2.1. Matrix-based subspace estimation

We have seen in Section 2.1.2 that the N observed samples of the R-D signals can be collected into an $M \times N$ measurement matrix $\boldsymbol{X}$, where the R dimensions are stacked along the rows and the columns represent the N snapshots. The resulting data model (2.3) is given by

$$\boldsymbol{X} = \boldsymbol{A} \cdot \boldsymbol{S} + \boldsymbol{N} = \boldsymbol{X}_0 + \boldsymbol{N}, \tag{3.1}$$

where $\boldsymbol{X} = [\boldsymbol{\mathcal{X}}]^{\mathrm{T}}_{(R+1)} \in \mathbb{C}^{M \times N}$, $\boldsymbol{A} = [\boldsymbol{\mathcal{A}}]^{\mathrm{T}}_{(R+1)} \in \mathbb{C}^{M \times d}$, and $\boldsymbol{N} = [\boldsymbol{\mathcal{N}}]^{\mathrm{T}}_{(R+1)} \in \mathbb{C}^{M \times N}$. Moreover, $\boldsymbol{X}_0 = \boldsymbol{A} \cdot \boldsymbol{S} \in \mathbb{C}^{M \times d}$ is the noise-free part of the measurement matrix $\boldsymbol{X}$, which is at most rank-d. Then, the column space of $\boldsymbol{X}_0$ is spanned by the d dominant left singular vectors of $\boldsymbol{X}_0$. The SVD of $\boldsymbol{X}_0$ is given by

$$\boldsymbol{X}_0 = \begin{bmatrix} \boldsymbol{U}_{\mathrm{s}}, & \boldsymbol{U}_{\mathrm{n}} \end{bmatrix} \cdot \begin{bmatrix} \boldsymbol{\Sigma}_{\mathrm{s}} & \boldsymbol{0}_{d \times (N-d)} \\ \boldsymbol{0}_{(M-d) \times d} & \boldsymbol{0}_{(M-d) \times (N-d)} \end{bmatrix} \cdot \begin{bmatrix} \boldsymbol{V}_{\mathrm{s}}, & \boldsymbol{V}_{\mathrm{n}} \end{bmatrix}^{\mathrm{H}}, \tag{3.2}$$

where $\boldsymbol{U}_{\mathrm{s}} \in \mathbb{C}^{M \times d}$ and $\boldsymbol{U}_{\mathrm{n}} \in \mathbb{C}^{M \times (M-d)}$ are the orthonormal bases for the signal subspace and the noise subspace. Furthermore, $\boldsymbol{\Sigma}_{\mathrm{s}} = \operatorname{diag}\left\{\begin{bmatrix}\sigma_1, \ldots, \sigma_d\end{bmatrix}\right\} \in \mathbb{R}^{d \times d}$ contains the d non-zero singular values on its main diagonal.

Note that due to $\boldsymbol{X}_0 = \boldsymbol{A} \cdot \boldsymbol{S}$, the column space of $\boldsymbol{A}$ is identical to the subspace spanned by the

columns of $\boldsymbol{U}_\mathrm{s}$, i.e., we have

$$\operatorname{span}\{\boldsymbol{A}\} = \operatorname{span}\{\boldsymbol{U}_\mathrm{s}\}. \tag{3.3}$$

This property is exploited by ESPRIT-type parameter estimation algorithms described in Section 3.4.1.

For the signal subspace estimation of the noisy measurement matrix $\boldsymbol{X}$, we can perform a rank-d-approximation of $\boldsymbol{X}$ via the truncated SVD, which is optimal in the Frobenius norm sense [EY36]. The SVD of $\boldsymbol{X}$ is expressed as

$$\boldsymbol{X} = \begin{bmatrix}\hat{\boldsymbol{U}}_\mathrm{s}, & \hat{\boldsymbol{U}}_\mathrm{n}\end{bmatrix} \cdot \begin{bmatrix}\hat{\boldsymbol{\Sigma}}_\mathrm{s} & \boldsymbol{0}_{d\times(N-d)} \\ \boldsymbol{0}_{(M-d)\times d} & \hat{\boldsymbol{\Sigma}}_\mathrm{n}\end{bmatrix} \cdot \begin{bmatrix}\hat{\boldsymbol{V}}_\mathrm{s}, & \hat{\boldsymbol{V}}_\mathrm{n}\end{bmatrix}^\mathrm{H}, \tag{3.4}$$

where $\hat{\boldsymbol{U}}_\mathrm{s} \in \mathbb{C}^{M\times d}$, $\hat{\boldsymbol{U}}_\mathrm{n} \in \mathbb{C}^{M\times(M-d)}$, and $\hat{\boldsymbol{\Sigma}}_\mathrm{s}$ are the respective estimates of $\boldsymbol{U}_\mathrm{s}$, $\boldsymbol{U}_\mathrm{n}$, and $\boldsymbol{\Sigma}_\mathrm{s}$.

Then, the signal subspace spanned by the columns of $\hat{\boldsymbol{U}}_\mathrm{s}$ is an estimate of the column space spanned by the columns of $\boldsymbol{A}$, i.e., we have

$$\operatorname{span}\{\boldsymbol{A}\} \approx \operatorname{span}\{\hat{\boldsymbol{U}}_\mathrm{s}\}. \tag{3.5}$$

An analytical first-order approximation of the estimation error contained in the matrix $\hat{\boldsymbol{U}}_\mathrm{s}$ from the SVD is provided in Section 4.2.1 as part of the performance analysis of matrix-based R-D ESPRIT-type algorithms.

3.2.2. Tensor-based subspace estimation

The natural multi-dimensional structure of the signals can be taken into account for the signal subspace estimation by employing a multi-dimensional extension of the SVD in form of a suitable tensor decomposition. Often, the Higher-Order SVD (HOSVD) is chosen for this purpose as it is easily computed via SVDs of the unfoldings of the tensor. Moreover, the truncated HOSVD allows for a multilinear low-rank approximation similar to the truncated SVD [HRD08].

Recall the tensor model from (2.7)

$$\boldsymbol{\mathcal{X}} = \boldsymbol{\mathcal{A}} \times_{R+1} \boldsymbol{S}^\mathrm{T} + \boldsymbol{\mathcal{N}} = \boldsymbol{\mathcal{X}}_0 + \boldsymbol{\mathcal{N}} \in \mathbb{C}^{M_1\times\ldots\times M_R\times N}, \tag{3.6}$$

where $\boldsymbol{\mathcal{X}}_0$ is the noise-free measurement tensor. Then, the truncated HOSVDs of $\boldsymbol{\mathcal{X}}_0$ and $\boldsymbol{\mathcal{X}}$ are given by

$$\boldsymbol{\mathcal{X}}_0 = \boldsymbol{\mathcal{S}}^{[\mathrm{s}]} \times_1 \boldsymbol{U}_1^{[\mathrm{s}]} \ldots \times_R \boldsymbol{U}_R^{[\mathrm{s}]} \times_{R+1} \boldsymbol{U}_{R+1}^{[\mathrm{s}]} \tag{3.7}$$

$$\boldsymbol{\mathcal{X}} \approx \hat{\boldsymbol{\mathcal{S}}}^{[\mathrm{s}]} \times_1 \hat{\boldsymbol{U}}_1^{[\mathrm{s}]} \ldots \times_R \hat{\boldsymbol{U}}_R^{[\mathrm{s}]} \times_{R+1} \hat{\boldsymbol{U}}_{R+1}^{[\mathrm{s}]}, \tag{3.8}$$

where $\boldsymbol{\mathcal{S}}^{[\mathrm{s}]} \in \mathbb{C}^{p_1 \times \ldots \times p_R \times p_{R+1}}$ and its estimate $\hat{\boldsymbol{\mathcal{S}}}^{[\mathrm{s}]}$ denote the truncated core tensors

$$\boldsymbol{\mathcal{S}}^{[\mathrm{s}]} = \boldsymbol{\mathcal{X}}_0 \times_1 \boldsymbol{U}_1^{[\mathrm{s}]\mathrm{H}} \ldots \times_R \boldsymbol{U}_R^{[\mathrm{s}]\mathrm{H}} \times_{R+1} \boldsymbol{U}_{R+1}^{[\mathrm{s}]\mathrm{H}} \tag{3.9}$$

$$\hat{\boldsymbol{\mathcal{S}}}^{[\mathrm{s}]} = \boldsymbol{\mathcal{X}} \times_1 \hat{\boldsymbol{U}}_1^{[\mathrm{s}]\mathrm{H}} \ldots \times_R \hat{\boldsymbol{U}}_R^{[\mathrm{s}]\mathrm{H}} \times_{R+1} \hat{\boldsymbol{U}}_{R+1}^{[\mathrm{s}]\mathrm{H}} \tag{3.10}$$

and $p_r,\ r = 1, \ldots, R$ represents the r-rank of $\boldsymbol{\mathcal{X}}_0$ and is given by $p_r = \min(M_r, d)$. Moreover, $\boldsymbol{U}_r^{[\mathrm{s}]} \in \mathbb{C}^{M_r \times p_r}$ and $\hat{\boldsymbol{U}}_r^{[\mathrm{s}]} \in \mathbb{C}^{M_r \times p_r}$ are the r-spaces, i.e., the matrices of the dominant r-mode left singular vectors, which can be computed by performing an SVD of the r-mode unfoldings of $\boldsymbol{\mathcal{X}}_0$ and $\boldsymbol{\mathcal{X}}$ according to

$$[\boldsymbol{\mathcal{X}}_0]_{(r)} = \begin{bmatrix} \boldsymbol{U}_r^{[\mathrm{s}]}, & \boldsymbol{U}_r^{[\mathrm{n}]} \end{bmatrix} \cdot \begin{bmatrix} \boldsymbol{\Sigma}_r^{[\mathrm{s}]} & \boldsymbol{0}_{d \times (N-d)} \\ \boldsymbol{0}_{(M-d) \times d} & \boldsymbol{0}_{(M-d) \times (N-d)} \end{bmatrix} \cdot \begin{bmatrix} \boldsymbol{V}_r^{[\mathrm{s}]}, & \boldsymbol{V}_r^{[\mathrm{n}]} \end{bmatrix}^{\mathrm{H}} \tag{3.11}$$

$$[\boldsymbol{\mathcal{X}}]_{(r)} = \begin{bmatrix} \hat{\boldsymbol{U}}_r^{[\mathrm{s}]}, & \hat{\boldsymbol{U}}_r^{[\mathrm{n}]} \end{bmatrix} \cdot \begin{bmatrix} \hat{\boldsymbol{\Sigma}}_r^{[\mathrm{s}]} & \boldsymbol{0}_{d \times (N-d)} \\ \boldsymbol{0}_{(M-d) \times d} & \hat{\boldsymbol{\Sigma}}_r^{[\mathrm{n}]} \end{bmatrix} \cdot \begin{bmatrix} \hat{\boldsymbol{V}}_r^{[\mathrm{s}]}, & \hat{\boldsymbol{V}}_r^{[\mathrm{n}]} \end{bmatrix}^{\mathrm{H}}. \tag{3.12}$$

Comparing the truncated HOSVD of $\boldsymbol{\mathcal{X}}$ in (3.8) to the truncated SVD in (3.4), we observe that the HOSVD performs low-rank approximations in all $R+1$ modes and hence, takes the multilinear structure into account for more efficient denoising.

Recall that the measurement matrix $\boldsymbol{X}$ and the measurement tensor $\boldsymbol{\mathcal{X}}$ are related via $\boldsymbol{X} = [\boldsymbol{\mathcal{X}}]_{(R+1)}^{\mathrm{T}}$. By considering the SVD of $[\boldsymbol{\mathcal{X}}]_{(R+1)}^{\mathrm{T}}$ using (3.12) and the SVD of $\boldsymbol{X}$ in (3.4), we find that the subspaces are linked through the following identities:

$$\hat{\boldsymbol{U}}_{\mathrm{s}} = \hat{\boldsymbol{V}}_{R+1}^{[\mathrm{s}]*},\ \hat{\boldsymbol{U}}_{\mathrm{n}} = \hat{\boldsymbol{V}}_{R+1}^{[\mathrm{n}]*},\ \hat{\boldsymbol{V}}_{\mathrm{s}} = \hat{\boldsymbol{U}}_{R+1}^{[\mathrm{s}]*},\ \hat{\boldsymbol{V}}_{\mathrm{n}} = \hat{\boldsymbol{U}}_{R+1}^{[\mathrm{n}]*},\ \hat{\boldsymbol{\Sigma}}_{\mathrm{s}} = \hat{\boldsymbol{\Sigma}}_{R+1}^{[\mathrm{s}]}. \tag{3.13}$$

Based on (3.8), the tensor-based subspace tensor $\hat{\boldsymbol{\mathcal{U}}}^{[\mathrm{s}]} \in \mathbb{C}^{M_1 \times \ldots \times M_R \times d}$, which is the multilinear extension of the matrix-based subspace estimate $\hat{\boldsymbol{U}}_{\mathrm{s}}$, is given by [HRD08, RHD14]

$$\hat{\boldsymbol{\mathcal{U}}}^{[\mathrm{s}]} = \hat{\boldsymbol{\mathcal{S}}}^{[\mathrm{s}]} \times_1 \hat{\boldsymbol{U}}_1^{[\mathrm{s}]} \ldots \times_R \hat{\boldsymbol{U}}_R^{[\mathrm{s}]} \times_{R+1} \hat{\boldsymbol{\Sigma}}_{R+1}^{[\mathrm{s}]-1} \in \mathbb{C}^{M_1 \times \ldots \times M_R \times d}. \tag{3.14}$$

Note that in its original definition in [HRD08], $\hat{\boldsymbol{\mathcal{U}}}^{[\mathrm{s}]}$ was defined without the multiplication by $\hat{\boldsymbol{\Sigma}}_{R+1}^{[\mathrm{s}]-1}$ in the $(R+1)$-th mode. This normalization was included in [RHD14] to simplify the notation in the derivation of the performance analysis of R-D Tensor-ESPRIT-type algorithms presented in [RHD14].

Consequently, the enhanced tensor-based signal subspace estimate, which can be used to replace the matrix-based subspace estimate $\hat{\boldsymbol{U}}_{\mathrm{s}}$ is given by $\left[\hat{\boldsymbol{\mathcal{U}}}^{[\mathrm{s}]}\right]_{(R+1)}^{\mathrm{T}}$.

According to [RHD14], the link between the HOSVD-based signal subspace estimate $\left[\hat{\boldsymbol{\mathcal{U}}}^{[\mathrm{s}]}\right]_{(R+1)}^{\mathrm{T}}$

and the SVD-based subspace estimate $\hat{\boldsymbol{U}}_{\mathrm{s}}$ is given by

$$\left[\hat{\boldsymbol{\mathcal{U}}}^{[\mathrm{s}]}\right]^{\mathrm{T}}_{(R+1)} = \left(\hat{\boldsymbol{T}}_1 \otimes \hat{\boldsymbol{T}}_2 \otimes \ldots \hat{\boldsymbol{T}}_R\right) \cdot \hat{\boldsymbol{U}}_{\mathrm{s}}, \tag{3.15}$$

where $\hat{\boldsymbol{T}}_r = \hat{\boldsymbol{U}}_r^{[\mathrm{s}]} \cdot \hat{\boldsymbol{U}}_r^{[\mathrm{s}]^{\mathrm{H}}} \in \mathbb{C}^{M_r \times M_r}$ are estimates of the projection matrices onto the r-spaces of $\boldsymbol{\mathcal{X}}_0$. Hence, the HOSVD-based subspace estimate can be seen as the projection of the unstructured matrix-based subspace estimate onto the Kronecker structure represented by the Kronecker product of the r-space projection matrices $\hat{\boldsymbol{T}}_r$, which is inherent in the data. Since this projection does not affect the true signal subspace, the improvement of the HOSVD-based subspace estimate comes from the fact that the noise components that do not obey the required Kronecker structure are removed. In other words, the multi-dimensional structure is imprinted onto the matrix-based subspace estimate. It is shown in [HRD08] that if $d \geq \max\limits_{r=1,2,\ldots,R} M_r$, we have $\left[\hat{\boldsymbol{\mathcal{U}}}^{[\mathrm{s}]}\right]^{\mathrm{T}}_{(R+1)} = \hat{\boldsymbol{U}}_{\mathrm{s}}$, i.e., there is no improvement in terms of the subspace estimation accuracy from the HOSVD-based subspace estimate if the number of signals d is greater than or equal to the number of sensors in all R modes. Note that in [CRKH14], the relation (3.15) has been used to derive a tensor-based subspace tracking algorithm for time-varying harmonic retrieval.

Similar to the matrix case, the error of the HOSVD-based subspace estimate $\left[\hat{\boldsymbol{\mathcal{U}}}^{[\mathrm{s}]}\right]^{\mathrm{T}}_{(R+1)}$ can be approximated by an analytical first-order expansion. Details are provided in Section 4.2.2.

3.2.3. Forward-backward averaging and real-valued subspace estimation

In the case of a centro-symmetric array, i.e., the array steering matrix $\boldsymbol{A}$ satisfies $\boldsymbol{\Pi}_M \cdot \boldsymbol{A}^* = \boldsymbol{A} \cdot \boldsymbol{\Delta}$ (cf. Section 2.1.4.2), the forward-backward averaging (FBA) preprocessing step [EJS82, PK89a] can be applied to the data to improve the estimation accuracy of the signal subspace. This is achieved by exploiting the symmetry in the data due to the array symmetry. The centro-symmetry property of the array implies that its array structure is invariant under mirroring around the centroid. Hence, we can exploit the fact that $\boldsymbol{A}$ and $\boldsymbol{\Pi}_M \cdot \boldsymbol{A}^*$ span the same column space. Therefore, the measurements $\boldsymbol{X} \in \mathbb{C}^{M \times N}$ can be augmented by $\boldsymbol{\Pi}_M \cdot \boldsymbol{X}^* \cdot \boldsymbol{\Pi}_N$, which is conjugated and row- as well as column-flipped version of the original measurement matrix, along the columns without changing the column space [HN95]. Mathematically speaking, FBA preprocessing is performed by constructing an augmented measurement matrix $\boldsymbol{X}^{(\mathrm{fba})}$ according to

$$\boldsymbol{X}^{(\mathrm{fba})} = \begin{bmatrix} \boldsymbol{X} & \boldsymbol{\Pi}_M \cdot \boldsymbol{X}^* \cdot \boldsymbol{\Pi}_N \end{bmatrix} \in \mathbb{C}^{M \times 2N}. \tag{3.16}$$

Notice that $\boldsymbol{X}^{(\mathrm{fba})}$ has $2N$ columns, i.e., the number of snapshots has been virtually doubled. Thus, processing $\boldsymbol{X}^{(\mathrm{fba})}$ instead of $\boldsymbol{X}$ improves the accuracy of the signal subspace estimate especially if only a small number of snapshots is available or highly correlated signals are present. In fact, two

coherent sources can be perfectly decorrelated.

Another advantage of the FBA preprocessing step in (3.16) is that the resulting augmented measurement matrix $\boldsymbol{X}^{(\mathrm{fba})}$ is a centro-Hermitian matrix[1] and hence it can be transformed into the real-valued domain [HN95]. This is achieved by a bijective mapping of the set of centro-Hermitian matrices onto the set of real-valued matrices [Lee80]. To this end, let us define left-$\boldsymbol{\Pi}$-real matrices, i.e., matrices $\boldsymbol{Q} \in \mathbb{C}^{p\times q}$ that satisfy $\boldsymbol{\Pi}_p \cdot \boldsymbol{Q}^* = \boldsymbol{Q}$. Sparse and square unitary left-$\boldsymbol{\Pi}$-real matrices [HN95] denoted by $\boldsymbol{Q}_p^{(\mathrm{s})} \in \mathbb{C}^{d\times d}$, which allow for an efficient implementation of the real-valued transformation are given by (cf. Appendix A.2)

$$\boldsymbol{Q}_{2n}^{(\mathrm{s})} = \frac{1}{\sqrt{2}} \cdot \begin{bmatrix} \boldsymbol{I}_n & \mathrm{j}\boldsymbol{I}_n \\ \boldsymbol{\Pi}_n & -\mathrm{j}\boldsymbol{\Pi}_n \end{bmatrix} \quad \text{and} \quad \boldsymbol{Q}_{2n+1}^{(\mathrm{s})} = \frac{1}{\sqrt{2}} \cdot \begin{bmatrix} \boldsymbol{I}_n & \boldsymbol{0}_{n\times 1} & \mathrm{j}\boldsymbol{I}_n \\ \boldsymbol{0}_{n\times 1}^{\mathrm{T}} & \sqrt{2} & \boldsymbol{0}_{n\times 1}^{\mathrm{T}} \\ \boldsymbol{\Pi}_n & \boldsymbol{0}_{n\times 1} & -\mathrm{j}\boldsymbol{\Pi}_n \end{bmatrix}. \tag{3.17}$$

Then, the real-valued transformed augmented measurement matrix is computed via

$$\varphi\left(\boldsymbol{X}^{(\mathrm{fba})}\right) = \boldsymbol{Q}_M^{\mathrm{H}} \cdot \boldsymbol{X}^{(\mathrm{fba})} \cdot \boldsymbol{Q}_{2N}. \tag{3.18}$$

The benefit of (3.18) is that since the matrix is real-valued, its signal subspace estimate can be obtained by applying a real-valued SVD, which has a lower computational complexity compared to its complex-valued counterpart.

Thus, in analogy to (3.4), the real-valued SVD of $\varphi\left(\boldsymbol{X}^{(\mathrm{fba})}\right)$ is given by

$$\varphi\left(\boldsymbol{X}^{(\mathrm{fba})}\right) = \left[\hat{\boldsymbol{E}}_{\mathrm{s}}, \quad \hat{\boldsymbol{E}}_{\mathrm{n}}\right] \cdot \begin{bmatrix} \hat{\boldsymbol{\Sigma}}_{\mathrm{s}}^{(\varphi)} & \boldsymbol{0}_{d\times(2N-d)} \\ \boldsymbol{0}_{(M-d)\times d} & \hat{\boldsymbol{\Sigma}}_{\mathrm{n}}^{(\varphi)} \end{bmatrix} \cdot \left[\hat{\boldsymbol{W}}_{\mathrm{s}}, \quad \hat{\boldsymbol{W}}_{\mathrm{n}}\right]^{\mathrm{H}} \tag{3.19}$$

and hence, the real-valued signal subspace estimate $\hat{\boldsymbol{E}}_{\mathrm{s}} \in \mathbb{R}^{M\times d}$ contains the d dominant left singular vectors of $\varphi\left(\boldsymbol{X}^{(\mathrm{fba})}\right)$.

Based on the real-valued signal subspace estimate $\hat{\boldsymbol{E}}_{\mathrm{s}}$, several "unitary" versions of the conventional DOA estimation algorithms can be defined, e.g., the Unitary Root-MUSIC algorithm [PGH00] or the Unitary ESPRIT algorithm [HN95] reviewed in Section 3.4.2.

In [HRD08], it is shown that forward-backward averaging and the real-valued transformation can also be defined for the measurement tensor $\boldsymbol{\mathcal{X}}$ according to

$$\boldsymbol{\mathcal{X}}^{(\mathrm{fba})} = \left[\boldsymbol{\mathcal{X}} \sqcup_{R+1} \boldsymbol{\mathcal{X}}^* \times_1 \boldsymbol{\Pi}_{M_1} \ldots \times_R \boldsymbol{\Pi}_{M_R} \times_{R+1} \boldsymbol{\Pi}_N\right] \in \mathbb{C}^{M_1\times M_2 \ldots \times M_R \times N} \tag{3.20}$$

[1] A matrix $\boldsymbol{X} \in \mathbb{C}^{p\times q}$ is called centro-Hermitian if it satisfies $\boldsymbol{\Pi}_p \cdot \boldsymbol{X}^* \cdot \boldsymbol{\Pi}_q = \boldsymbol{X}$.

and

$$\varphi\left(\boldsymbol{\mathcal{X}}^{(\mathrm{fba})}\right) = \boldsymbol{\mathcal{X}}^{(\mathrm{fba})} \times_1 \boldsymbol{Q}_{M_1}^{\mathrm{H}} \ldots \times_R \boldsymbol{Q}_{M_R}^{\mathrm{H}} \times_{R+1} \boldsymbol{Q}_{2N}^{\mathrm{H}} \in \mathbb{R}^{M_1 \times M_2 \ldots \times M_R \times N}, \tag{3.21}$$

which are the tensor extensions of (3.16) and (3.18). We can then estimate the real-valued signal subspace tensor $\hat{\boldsymbol{\mathcal{E}}}^{[\mathrm{s}]} \in \mathbb{R}^{M_1 \times \ldots \times M_R \times d}$ via a real-valued truncated HOSVD of the transformed tensor $\varphi\left(\boldsymbol{\mathcal{X}}^{(\mathrm{fba})}\right) \in \mathbb{R}^{M_1 \times \ldots \times M_R \times 2N}$ from (3.21) given by

$$\varphi\left(\boldsymbol{\mathcal{X}}^{(\mathrm{fba})}\right) \approx \hat{\boldsymbol{\mathcal{S}}}_{\mathcal{T}}^{[\mathrm{s}]} \times_1 \hat{\boldsymbol{E}}_1^{[\mathrm{s}]} \ldots \times_R \hat{\boldsymbol{E}}_R^{[\mathrm{s}]} \times_{R+1} \hat{\boldsymbol{E}}_{R+1}^{[\mathrm{s}]}, \tag{3.22}$$

where the real-valued r-spaces $\hat{\boldsymbol{E}}_r^{[\mathrm{s}]} \in \mathbb{R}^{M_r \times p_r}$ can be computed via the SVD of the r-mode unfoldings of $\varphi\left(\boldsymbol{\mathcal{X}}^{(\mathrm{fba})}\right)$ for $r = 1, \ldots, R+1$ as

$$\left[\varphi\left(\boldsymbol{\mathcal{X}}^{(\mathrm{fba})}\right)\right]_{(r)} = \left[\hat{\boldsymbol{E}}_r^{[\mathrm{s}]}, \quad \hat{\boldsymbol{E}}_r^{[\mathrm{n}]}\right] \cdot \begin{bmatrix} \hat{\boldsymbol{\Sigma}}_r^{[\mathrm{s}]} & \boldsymbol{0}_{d \times (N-d)} \\ \boldsymbol{0}_{(M-d) \times d} & \hat{\boldsymbol{\Sigma}}_r^{[\mathrm{n}]} \end{bmatrix} \cdot \left[\hat{\boldsymbol{W}}_r^{[\mathrm{s}]}, \quad \hat{\boldsymbol{W}}_r^{[\mathrm{n}]}\right]^{\mathrm{H}}. \tag{3.23}$$

Then, the real-valued subspace tensor $\hat{\boldsymbol{\mathcal{E}}}^{[\mathrm{s}]}$ is defined as

$$\hat{\boldsymbol{\mathcal{E}}}^{[\mathrm{s}]} = \hat{\boldsymbol{\mathcal{S}}}_{\mathcal{T}}^{[\mathrm{s}]} \times_1 \hat{\boldsymbol{E}}_1^{[\mathrm{s}]} \ldots \times_R \hat{\boldsymbol{E}}_R^{[\mathrm{s}]} \times_{R+1} \hat{\boldsymbol{\Sigma}}_{R+1}^{[\mathrm{s}]^{-1}} \tag{3.24}$$

following (3.14). Note that we have again used the normalization by $\boldsymbol{\Sigma}_{R+1}^{[\mathrm{s}]^{-1}}$, which is obtained from (3.23) for $r = R+1$.

3.2.4. Spatial smoothing

Forward-backward averaging can only decorrelate two coherent, i.e., fully correlated, sources. If more than two coherent sources are present or only a single snapshot $N = 1$ is available, spatial smoothing can be applied to decorrelate multiple coherent wavefronts [SWK85] by dividing the array into L identical displaced subarrays and averaging their spatial covariance matrices. As a result, L coherent wavefronts are decorrelated. However, the number of antenna elements is also reduced to $M - L + 1$. Spatial smoothing can be combined with forward-backward averaging [PK89b], in which case $2L$ coherent wavefronts are decorrelated when using L subarrays. Spatial smoothing is readily formulated in terms of tensors via the concatenation operator, as shown in [HRD08].

However, as shown in [THRG10, THG09b, THG09a], the tensor structure can be exploited further to derive tensor-based spatial smoothing techniques that outperform matrix-based approaches.

In Chapter 7, we will combine spatial smoothing preprocessing with R-D ESPRIT-type algorithms as well as R-D NC ESPRIT-type algorithms for strictly non-circular sources. We will also develop analytical performance evaluation expressions for these algorithms with spatial smoothing.

3.3. R-D shift invariance

A requirement for the applicability of ESPRIT-type parameter estimation algorithms is the "shift invariance" property of the array, which implies that the array can be divided into two subarrays that are identical except for a displacement. An example is shown in Figure 3.1. Due to the narrowband assumption of the impinging exponentials, a spatial displacement results in a phase offset, which is proportional to the spatial frequency. Hence, the spatial frequency estimates can be obtained by estimating the phase offsets of all signals. Note that to avoid ambiguities, the spatial displacement should not exceed $\lambda_c/2$ (cf. Section 2.1.4.1).

It was shown in [HN98] that the spatial frequencies of R-dimensional signals in all dimensions can be estimated efficiently if the sampling grid exhibits the shift invariance structure in all R dimensions. The matrix-based R-D shift invariance equation in terms of the array steering matrix $\boldsymbol{A} = [\boldsymbol{\mathcal{A}}]^{\mathrm{T}}_{(R+1)}$ is given in [HN98] by

$$\tilde{\boldsymbol{J}}_1^{(r)} \cdot \boldsymbol{A} \cdot \boldsymbol{\Phi}^{(r)} = \tilde{\boldsymbol{J}}_2^{(r)} \cdot \boldsymbol{A}, \qquad r = 1, \ldots, R, \tag{3.25}$$

where $\tilde{\boldsymbol{J}}_1^{(r)}$ and $\tilde{\boldsymbol{J}}_2^{(r)} \in \mathbb{R}^{\frac{M}{M_r} M_r^{(\mathrm{sel})} \times M}$ are the effective R-D selection matrices, which select $M_r^{(\mathrm{sel})}$ elements for the first and the second subarray in the r-th mode, respectively. They are compactly defined as

$$\tilde{\boldsymbol{J}}_n^{(r)} = (\boldsymbol{I}_{M_1} \otimes \ldots \otimes \boldsymbol{I}_{M_{r-1}}) \otimes \boldsymbol{J}_n^{(r)} \otimes (\boldsymbol{I}_{M_{r+1}} \otimes \ldots \otimes \boldsymbol{I}_{M_R}), \quad n = 1, 2, \tag{3.26}$$

where $\boldsymbol{J}_n^{(r)} \in \mathbb{R}^{M_r^{(\mathrm{sel})} \times M_r}$ are the r-mode selection matrices for the first and second subarray. The diagonal matrix $\boldsymbol{\Phi}^{(r)} = \mathrm{diag}\{[\mathrm{e}^{\mathrm{j}\mu_1^{(r)}}, \ldots, \mathrm{e}^{\mathrm{j}\mu_d^{(r)}}]^{\mathrm{T}}\} \in \mathbb{C}^{d \times d}$ contains the spatial frequencies in the r-th mode to be estimated. For the special case of a uniform sampling grid introduced in Section 2.1.4.1, the selection matrices $\boldsymbol{J}_1^{(r)}$ and $\boldsymbol{J}_2^{(r)}$ are chosen as

$$\boldsymbol{J}_1^{(r)} = \begin{bmatrix} \boldsymbol{I}_{M_r-1} & \boldsymbol{0}_{(M_r-1)\times 1} \end{bmatrix}, \quad \boldsymbol{J}_2^{(r)} = \begin{bmatrix} \boldsymbol{0}_{(M_r-1)\times 1} & \boldsymbol{I}_{M_r-1} \end{bmatrix} \tag{3.27}$$

such that $M_r^{(\mathrm{sel})} = M_r - 1$, which corresponds to maximally overlapping subarrays.

In the 1-D case, the shift invariance equation simplifies into

$$\boldsymbol{J}_1 \cdot \boldsymbol{A} \cdot \boldsymbol{\Phi} = \boldsymbol{J}_2 \cdot \boldsymbol{A} \tag{3.28}$$

with the selection matrices

$$\boldsymbol{J}_1 = \begin{bmatrix} \boldsymbol{I}_{M-1} & \boldsymbol{0}_{(M-1)\times 1} \end{bmatrix}, \quad \boldsymbol{J}_2 = \begin{bmatrix} \boldsymbol{0}_{(M-1)\times 1} & \boldsymbol{I}_{M-1} \end{bmatrix}. \tag{3.29}$$

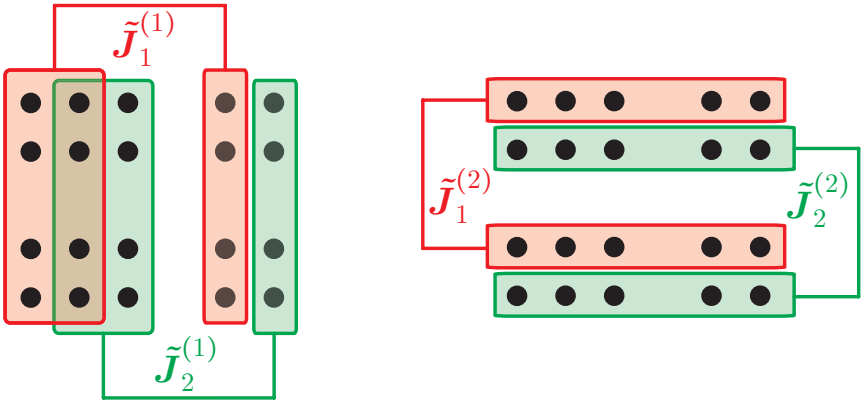


Figure 3.1.: 2-D shift invariance for the 5×4 separable 2-D sampling grid from Figure 2.1b. Left: subarrays for the horizontal dimension, right: subarrays for the vertical dimension.

An example for the 2-D shift invariance of the 5×4 separable 2-D sampling grid introduced in Figure 2.1b is provided in Figure 3.1. On the left-hand side, the selected subarrays for the horizontal dimension are shown and the right-hand side illustrates the selected subarrays for the vertical dimension. The corresponding selection matrices are given by

$$\boldsymbol{J}_1^{(1)} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \end{bmatrix}, \quad \boldsymbol{J}_2^{(1)} = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}, \quad \boldsymbol{J}_1^{(2)} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}, \quad \boldsymbol{J}_2^{(2)} = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

such that $\tilde{\boldsymbol{J}}_n^{(1)} = \boldsymbol{J}_n^{(1)} \otimes \boldsymbol{I}_4$ and $\tilde{\boldsymbol{J}}_n^{(2)} = \boldsymbol{I}_5 \otimes \boldsymbol{J}_n^{(2)}$ for $n = 1, 2$.

Using tensor calculus, the R-D shift invariance can be expressed in a more natural way in terms of the array steering tensor $\boldsymbol{\mathcal{A}} \in \mathbb{C}^{M_1 \ldots \times M_R \times d}$ defined in (2.10). Then, the shift invariance of $\boldsymbol{\mathcal{A}}$ in the r-th mode can be expressed as [HRD08]

$$\boldsymbol{\mathcal{A}} \times_r \boldsymbol{J}_1^{(r)} \times_{R+1} \boldsymbol{\Phi}^{(r)} = \boldsymbol{\mathcal{A}} \times_r \boldsymbol{J}_2^{(r)}. \tag{3.30}$$

Note that we arrive again at the matrix-based R-D shift invariance equation in (3.25) if we compute the transpose of the $(R+1)$-mode unfolding of (3.30) and apply the property (1.30).

3.4. R-D matrix-based ESPRIT-type algorithms

In this section, we first review the R-D Standard ESPRIT algorithm in Section 3.4.1 before considering the R-D Unitary ESPRIT algorithm in Section 3.4.2.

3.4.1. R-D Standard ESPRIT

After the subspace estimation, the key task in ESPRIT-type parameter estimation algorithms is to solve the shift invariance equations (3.25) for the matrices $\boldsymbol{\Phi}^{(r)}$. As the array steering matrix $\boldsymbol{A}$ is unknown, we make use of the property that

$$\operatorname{span}\{\boldsymbol{A}\} \approx \operatorname{span}\{\hat{\boldsymbol{U}}_{\mathrm{s}}\}, \tag{3.31}$$

where the column space spanned by the columns of $\hat{\boldsymbol{U}}_{\mathrm{s}}$ obtained from the SVD of $\boldsymbol{X}$ (cf. Section 3.2.1) is an estimate of the true signal subspace spanned by the columns of $\boldsymbol{A}$. As $\hat{\boldsymbol{U}}_{\mathrm{s}}$ and $\boldsymbol{A}$ approximately span the same subspace, we can write

$$\boldsymbol{A} \approx \hat{\boldsymbol{U}}_{\mathrm{s}} \cdot \boldsymbol{T}, \tag{3.32}$$

where $\boldsymbol{T} \in \mathbb{C}^{d \times d}$ is a full-rank matrix that contains the coefficients for the linear combination of the columns of $\boldsymbol{A}$ to form the basis $\hat{\boldsymbol{U}}_{\mathrm{s}}$.

Inserting this relation into (3.25), we have

$$\begin{aligned} \tilde{\boldsymbol{J}}_1^{(r)} \cdot \hat{\boldsymbol{U}}_{\mathrm{s}} \cdot \boldsymbol{T} \cdot \boldsymbol{\Phi}^{(r)} &\approx \tilde{\boldsymbol{J}}_2^{(r)} \cdot \hat{\boldsymbol{U}}_{\mathrm{s}} \cdot \boldsymbol{T} \\ \tilde{\boldsymbol{J}}_1^{(r)} \cdot \hat{\boldsymbol{U}}_{\mathrm{s}} \cdot \boldsymbol{T} \cdot \boldsymbol{\Phi}^{(r)} \cdot \boldsymbol{T}^{-1} &\approx \tilde{\boldsymbol{J}}_2^{(r)} \cdot \hat{\boldsymbol{U}}_{\mathrm{s}} \\ \tilde{\boldsymbol{J}}_1^{(r)} \cdot \hat{\boldsymbol{U}}_{\mathrm{s}} \cdot \boldsymbol{\Psi}^{(r)} &\approx \tilde{\boldsymbol{J}}_2^{(r)} \cdot \hat{\boldsymbol{U}}_{\mathrm{s}}, \end{aligned} \tag{3.33}$$

where $\boldsymbol{\Psi}^{(r)} \approx \boldsymbol{T} \cdot \boldsymbol{\Phi}^{(r)} \cdot \boldsymbol{T}^{-1}$. Note that (3.33) represents an overdetermined set of linear equations in $\boldsymbol{\Psi}^{(r)}$, which can be solved by least squares (LS) methods. The simple LS solution [RK89] of (3.33) for $\boldsymbol{\Psi}^{(r)}$ is obtained by

$$\hat{\boldsymbol{\Psi}}_{\mathrm{LS}}^{(r)} = \arg\min_{\boldsymbol{\Psi}} \left\| \tilde{\boldsymbol{J}}_1^{(r)} \cdot \hat{\boldsymbol{U}}_{\mathrm{s}} \cdot \boldsymbol{\Psi} - \tilde{\boldsymbol{J}}_2^{(r)} \cdot \hat{\boldsymbol{U}}_{\mathrm{s}} \right\|_{\mathrm{F}}^2 = \left(\tilde{\boldsymbol{J}}_1^{(r)} \cdot \hat{\boldsymbol{U}}_{\mathrm{s}} \right)^{+} \cdot \tilde{\boldsymbol{J}}_2^{(r)} \cdot \hat{\boldsymbol{U}}_{\mathrm{s}}. \tag{3.34}$$

Lastly, an estimate of $\boldsymbol{\Phi}^{(r)}$ is obtained via an eigenvalue decomposition (EVD) of $\hat{\boldsymbol{\Psi}}_{\mathrm{LS}}^{(r)}$. In order to ensure the correct pairing across the R dimensions, the matrices $\hat{\boldsymbol{\Phi}}^{(r)}$ should be estimated via a joint EVD of $\hat{\boldsymbol{\Psi}}_{\mathrm{LS}}^{(r)}$ (e.g., via [FG06]).

More accurate and structured solutions of (3.33) can be obtained by applying TLS [OVK91], SLS [Haa97b], R-D SLS [Haa97b], or the novel generalized least squares (GLS) algorithm [SRH17a], which is derived in Chapter 5.

The R-D Standard ESPRIT algorithm is summarized in Algorithm 1.

Algorithm 1 Summary of R-D Standard ESPRIT

1. Estimate the signal subspace $\hat{\boldsymbol{U}}_{\mathrm{s}}$ via the truncated SVD of the observation matrix $\boldsymbol{X} \in \mathbb{C}^{M \times N}$.
2. Solve the overdetermined shift invariance equations
$$\tilde{\boldsymbol{J}}_1^{(r)} \cdot \hat{\boldsymbol{U}}_{\mathrm{s}} \cdot \boldsymbol{\Psi}^{(r)} \approx \tilde{\boldsymbol{J}}_2^{(r)} \cdot \hat{\boldsymbol{U}}_{\mathrm{s}}$$
for the matrices $\boldsymbol{\Psi}^{(r)}$ for $r = 1, 2, \ldots, R$ via least squares (LS) methods, e.g., LS, TLS, SLS, GLS.
3. Compute the eigenvalues $\hat{\lambda}_i^{(r)}$ for $i = 1, \ldots, d$ of $\hat{\boldsymbol{\Psi}}^{(r)}$ jointly for all $r = 1, 2, \ldots, R$, e.g., via the joint diagonalization scheme proposed in [FG06]. Recover the correctly paired frequencies $\hat{\mu}_i^{(r)}$ via $\hat{\mu}_i^{(r)} = \arg\left\{\hat{\lambda}_i^{(r)}\right\}$.

3.4.2. R-D Unitary ESPRIT

We have outlined in Section 3.2.3 that FBA preprocessing can be applied to create another set of N virtual snapshots if the array geometry is centro-symmetric. As a result, the augmented measurement matrix $\boldsymbol{X}^{(\mathrm{fba})} \in \mathbb{C}^{M \times 2N}$ defined in (3.16) is further processed instead of $\boldsymbol{X} \in \mathbb{C}^{M \times N}$. Moreover, FBA allows to decorrelate two coherent sources if their correlation phases and the diagonal elements of $\boldsymbol{\Delta}$ in (2.28) are distinct. As an additional feature, the augmented complex measurement matrix $\boldsymbol{X}^{(\mathrm{fba})}$ can be transformed into the real-valued domain by means of a one-to-one mapping. Thus, the entire subsequent processing can be performed using real-valued additions and multiplications only.

The FBA preprocessing step and the real-valued transformation are described in (3.16) and (3.18), respectively. Then, we define the transformed steering matrix as $\boldsymbol{D} = \boldsymbol{Q}_M^{\mathrm{H}} \cdot \boldsymbol{A}$. Based on the R-D shift invariance property of $\boldsymbol{A}$, it is easy to see that $\boldsymbol{D}$ also obeys the shift invariance

$$\tilde{\boldsymbol{K}}_1^{(r)} \cdot \boldsymbol{D} \cdot \boldsymbol{\Omega}^{(r)} = \tilde{\boldsymbol{K}}_2^{(r)} \cdot \boldsymbol{D} \tag{3.35}$$

for $r = 1, \ldots, R$, where the R pairs of transformed selection matrices are given as [HN98]

$$\tilde{\boldsymbol{K}}_1^{(r)} = 2 \cdot \mathrm{Re}\left\{\boldsymbol{Q}_{M_r^{(\mathrm{sel})} \cdot M/M_r}^{\mathrm{H}} \cdot \tilde{\boldsymbol{J}}_2^{(r)} \cdot \boldsymbol{Q}_M\right\} \tag{3.36}$$

$$\tilde{\boldsymbol{K}}_2^{(r)} = 2 \cdot \mathrm{Im}\left\{\boldsymbol{Q}_{M_r^{(\mathrm{sel})} \cdot M/M_r}^{\mathrm{H}} \cdot \tilde{\boldsymbol{J}}_2^{(r)} \cdot \boldsymbol{Q}_M\right\}. \tag{3.37}$$

Moreover, the real-valued set of diagonal matrices $\boldsymbol{\Omega}^{(r)} = \mathrm{diag}\left\{[\omega_1^{(r)}, \ldots, \omega_d^{(r)}]^{\mathrm{T}}\right\} \in \mathbb{R}^{d \times d}$ with $\omega_i^{(r)} = \tan(\mu_i^{(r)}/2)$ contain the spatial frequencies in the r-th mode.

Based on the preprocessed measurement matrix $\varphi\left(\boldsymbol{X}^{(\mathrm{fba})}\right)$ from (3.18), the real-valued aug-

Algorithm 2 [HN98] Summary of R-D Unitary ESPRIT

1. Estimate the real-valued signal subspace $\hat{\boldsymbol{E}}_{\mathrm{s}}$ via the truncated SVD of the transformed real-valued observation matrix $\mathcal{T}(\boldsymbol{X}) = \boldsymbol{Q}_M^{\mathrm{H}} \cdot \left[\boldsymbol{X} \quad \boldsymbol{\Pi}_M \boldsymbol{X}^* \boldsymbol{\Pi}_N\right] \cdot \boldsymbol{Q}_{2N} \in \mathbb{R}^{M \times 2N}$, where $\boldsymbol{Q}_p$ is a unitary $p \times p$ left-$\boldsymbol{\Pi}$-real matrix (cf. Section 3.2.3).

2. Solve the overdetermined shift invariance equations

$$\tilde{\boldsymbol{K}}_1^{(r)} \cdot \hat{\boldsymbol{E}}_{\mathrm{s}} \cdot \boldsymbol{\Upsilon}^{(r)} \approx \tilde{\boldsymbol{K}}_2^{(r)} \cdot \hat{\boldsymbol{E}}_{\mathrm{s}}$$

   for the matrices $\boldsymbol{\Upsilon}^{(r)}$ for $r = 1, 2, \ldots, R$ via LS methods, e.g., LS, TLS, SLS, GLS. The transformed selection matrices $\tilde{\boldsymbol{K}}_1^{(r)}$ and $\tilde{\boldsymbol{K}}_2^{(r)}$ are given in (3.36) and (3.37).

3. Compute the eigenvalues $\hat{\omega}_i^{(r)}$ for $i = 1, 2, \ldots, d$ of $\hat{\boldsymbol{\Upsilon}}^{(r)}$ jointly for all $r = 1, \ldots, R$, e.g., via the joint diagonalization scheme proposed in [FG06] or via the Simultaneous Schur Decomposition proposed in [HN98]. Recover the correctly paired frequencies $\hat{\mu}_i^{(r)}$ via $\hat{\mu}_i^{(r)} = 2 \cdot \arctan(\hat{\omega}_i^{(r)})$.

mented signal subspace $\hat{\boldsymbol{E}}_{\mathrm{s}} \in \mathbb{R}^{M \times d}$ is estimated via the SVD of $\varphi\left(\boldsymbol{X}^{(\mathrm{fba})}\right)$ as shown in (3.19). Since $\boldsymbol{D}$ and $\hat{\boldsymbol{E}}_{\mathrm{s}}$ span approximately the same column space, we can find a non-singular matrix $\boldsymbol{T} \in \mathbb{C}^{d \times d}$ such that $\boldsymbol{D} \approx \hat{\boldsymbol{E}}_{\mathrm{s}} \cdot \boldsymbol{T}$ holds. Substituting this relation into (3.35), the overdetermined set of R real-valued shift invariance equations is given by [HN98]

$$\tilde{\boldsymbol{K}}_1^{(r)} \cdot \hat{\boldsymbol{E}}_{\mathrm{s}} \cdot \boldsymbol{\Upsilon}^{(r)} \approx \tilde{\boldsymbol{K}}_2^{(r)} \cdot \hat{\boldsymbol{E}}_{\mathrm{s}}, \quad r = 1, \ldots, R \tag{3.38}$$

with $\boldsymbol{\Upsilon}^{(r)} \approx \boldsymbol{T} \cdot \boldsymbol{\Omega}^{(r)} \cdot \boldsymbol{T}^{-1}$. The R unknown real-valued diagonal matrices $\boldsymbol{\Upsilon}^{(r)}$ can be estimated via, for instance, the simple LS method, i.e.,

$$\hat{\boldsymbol{\Upsilon}}_{\mathrm{LS}}^{(r)} = \left(\tilde{\boldsymbol{K}}_1^{(r)} \cdot \hat{\boldsymbol{E}}_{\mathrm{s}}\right)^{+} \cdot \tilde{\boldsymbol{K}}_2^{(r)} \cdot \hat{\boldsymbol{E}}_{\mathrm{s}} \in \mathbb{R}^{d \times d}. \tag{3.39}$$

Finally, the correctly paired spatial frequency estimates are obtained by $\hat{\mu}_i^{(r)} = 2 \cdot \arctan(\hat{\omega}_i^{(r)})$, $i = 1, \ldots, d$. The eigenvalues $\hat{\omega}_i^{(r)}$ of $\hat{\boldsymbol{\Upsilon}}_{\mathrm{LS}}^{(r)}$ are computed by performing a joint eigendecomposition across all R dimensions [FG06] or via the simultaneous Schur decomposition [HN98]. If all the eigenvalues are real, they provide reliable estimates [HN95].

Alternatively, more accurate solutions of (3.38) can be obtained by applying TLS [OVK91], SLS [Haa97b], R-D SLS [Haa97b], or the novel generalized least squares (GLS) algorithm [SRH17a], which is derived in Chapter 5.

A summary of the R-D Unitary ESPRIT algorithm is provided in Algorithm 2.

3.5. R-D tensor-based ESPRIT-type algorithms

In this section, we review the tensor-based R-D ESPRIT-type algorithms. We first discuss the R-D Standard Tensor-ESPRIT algorithm in Section 3.5.1 and then resort to the R-D Unitary Tensor-ESPRIT algorithm in Section 3.5.2

3.5.1. R-D Standard Tensor-ESPRIT

As seen in Section 3.3, a more natural multilinear formulation of the R-D shift invariance equation can be found by means of tensor algebra. Thereby, the stacking operation, which does not fully preserve the multi-dimensional structure, can be avoided. Based on the enhanced tensor-based signal subspace estimate introduced in Section 3.2.2, a tensor-based version of the R-D Standard ESPRIT algorithm with an improved estimation accuracy has been developed in [HRD08].

Starting from the R-D shift invariance equation in terms of the array steering tensor $\boldsymbol{\mathcal{A}}$ in (3.30), we first eliminate the unknown $\boldsymbol{\mathcal{A}}$ by applying the tensor-based extension of the relation in (3.32), which is given as [HRD08]

$$\boldsymbol{\mathcal{A}} = \boldsymbol{\mathcal{U}}^{[\mathrm{s}]} \times_{R+1} \bar{\boldsymbol{T}}_{\mathrm{T}}, \tag{3.40}$$

where $\bar{\boldsymbol{T}}_{\mathrm{T}} \in \mathbb{C}^{d \times d}$ is a non-singular transform matrix. Note that (3.40) shows that the r-spaces for $r = 1, \ldots, R$ of the r-mode unfoldings as well as the row spaces of the $(R+1)$-mode unfoldings of $\boldsymbol{\mathcal{A}}$ and $\boldsymbol{\mathcal{U}}^{[\mathrm{s}]}$ are identical.

Motivated by (3.40), the unknown array steering tensor in the shift invariance equation can be replaced by the estimated signal subspace tensor $\hat{\boldsymbol{\mathcal{U}}}^{[\mathrm{s}]}$ using $\boldsymbol{\mathcal{A}} \approx \hat{\boldsymbol{\mathcal{U}}}^{[\mathrm{s}]} \times_{R+1} \bar{\boldsymbol{T}}_{\mathrm{T}}$ to obtain the overdetermined sets of R shift invariance equations

$$\hat{\boldsymbol{\mathcal{U}}}^{[\mathrm{s}]} \times_r \boldsymbol{J}_1^{(r)} \times_{R+1} \boldsymbol{\Psi}^{(r)} \approx \hat{\boldsymbol{\mathcal{U}}}^{[\mathrm{s}]} \times_r \boldsymbol{J}_2^{(r)}, \tag{3.41}$$

where $\boldsymbol{\Psi}^{(r)} = \bar{\boldsymbol{T}}_{\mathrm{T}}^{-1} \cdot \boldsymbol{\Phi}^{(r)} \cdot \bar{\boldsymbol{T}}_{\mathrm{T}}$, $r = 1, \ldots, R$ follows due to property (1.32) for repeated n-mode products, which also reverses the order of the matrices in $\boldsymbol{\Psi}^{(r)}$ compared to the matrix case in (3.33). As shown in [HRD08], a solution of (3.41) is given by the simple LS estimate

$$\hat{\boldsymbol{\Psi}}_{\mathrm{LS}}^{(r)} = \arg\min_{\boldsymbol{\Psi}} \left\| \hat{\boldsymbol{\mathcal{U}}}^{[\mathrm{s}]} \times_r \boldsymbol{J}_1^{(r)} \times_{R+1} \boldsymbol{\Psi} - \hat{\boldsymbol{\mathcal{U}}}^{[\mathrm{s}]} \times_r \boldsymbol{J}_2^{(r)} \right\|_{\mathrm{H}}^2 \tag{3.42}$$

$$\Rightarrow \hat{\boldsymbol{\Psi}}_{\mathrm{LS}}^{(r)^{\mathrm{T}}} = \left(\tilde{\boldsymbol{J}}_1^{(r)} \cdot \left[\hat{\boldsymbol{\mathcal{U}}}^{[\mathrm{s}]} \right]_{(R+1)}^{\mathrm{T}} \right)^{+} \cdot \tilde{\boldsymbol{J}}_2^{(r)} \cdot \left[\hat{\boldsymbol{\mathcal{U}}}^{[\mathrm{s}]} \right]_{(R+1)}^{\mathrm{T}}. \tag{3.43}$$

Note that $\tilde{\boldsymbol{J}}_n^{(r)}$ and $\boldsymbol{J}_n^{(r)}$ for $n = 1, 2$ are related via (3.26). The remaining steps including the joint eigendecomposition of $\hat{\boldsymbol{\Psi}}_{\mathrm{LS}}^{(r)}$ to extract the frequencies $\mu_i^{(r)}$ are the same as for the matrix-based

Algorithm 3 [HRD08] Summary of R-D Standard Tensor-ESPRIT

1. Estimate the signal subspace tensor $\hat{\boldsymbol{\mathcal{U}}}^{[\mathrm{s}]} \in \mathbb{C}^{M_1 \times \ldots \times M_R \times d}$ via the truncated HOSVD of the observation tensor $\boldsymbol{\mathcal{X}} \in \mathbb{C}^{M_1 \times \ldots \times M_R \times N}$ following (3.14).

2. Solve the overdetermined shift invariance equations

$$\hat{\boldsymbol{\mathcal{U}}}^{[\mathrm{s}]} \times_r \boldsymbol{J}_1^{(r)} \times_{R+1} \hat{\boldsymbol{\Psi}}^{(r)} \approx \hat{\boldsymbol{\mathcal{U}}}^{[\mathrm{s}]} \times_r \boldsymbol{J}_2^{(r)}$$

   for the matrices $\hat{\boldsymbol{\Psi}}^{(r)}$ for $r = 1, \ldots, R$ via LS methods, e.g., LS, TLS, SLS, TS-SLS, GLS.

3. Compute the eigenvalues $\hat{\lambda}_i^{(r)}$ for $i = 1, 2, \ldots, d$ of $\hat{\boldsymbol{\Psi}}^{(r)}$ jointly for all $r = 1, \ldots, R$, e.g., via the joint diagonalization scheme proposed in [FG06]. Recover the correctly paired frequencies $\hat{\mu}_i^{(r)}$ via $\hat{\mu}_i^{(r)} = \arg\left\{\hat{\lambda}_i^{(r)}\right\}$.

R-D Standard ESPRIT algorithm described in (3.4.1).

Notice that if we compare the solution (3.43) to the least squares solution (3.34) of the matrix-based shift invariance equations for R-D Standard ESPRIT, we find that they only differ in the signal subspace estimate. Hence, R-D Standard Tensor-ESPRIT is algebraically equivalent to R-D Standard ESPRIT and the improvement in the estimation accuracy of the spatial frequencies is due to replacing the SVD-based subspace estimate $\hat{\boldsymbol{U}}_{\mathrm{s}}$ by the HOSVD-based subspace estimate $\left[\hat{\boldsymbol{\mathcal{U}}}^{[\mathrm{s}]}\right]_{(R+1)}^{\mathrm{T}}$. Note that this is not the case when the tensor-structure SLS (TS-SLS) algorithm [RH07b] is used to solve the R-D shift invariance equation as it exploits additional structure.

The resulting R-D Standard ESPRIT algorithm is summarized in Algorithm 3.

3.5.2. R-D Unitary Tensor-ESPRIT

We have seen in the previous section that a tensor version of R-D Standard ESPRIT, termed R-D Standard Tensor-ESPRIT can be derived by simply replacing the SVD-based signal subspace estimate in the matrix-based R-D Standard ESPRIT algorithm by the improved HOSVD-based signal subspace estimate.

Therefore, the R-D Unitary Tensor-ESPRIT algorithm can be derived in a similar way. As discussed in Section 3.2.3 and shown in [HRD08], for centro-symmetric arrays, forward-backward averaging can be applied to the measurement tensor $\boldsymbol{\mathcal{X}}$, which enables the transformation into the real-valued domain to reduce the computational complexity. Applying the real-valued transformation in (3.21) to the shift invariance equation in (3.41) and using the real-valued signal subspace

tensor $\hat{\boldsymbol{\mathcal{E}}}^{[\mathrm{s}]}$ defined in (3.24), we obtain the transformed shift invariance equation

$$\hat{\boldsymbol{\mathcal{E}}}^{[\mathrm{s}]} \times_r \boldsymbol{K}_1^{(r)} \times_{R+1} \boldsymbol{\Upsilon}^{(r)} \approx \hat{\boldsymbol{\mathcal{E}}}^{[\mathrm{s}]} \times_r \boldsymbol{K}_2^{(r)} \tag{3.44}$$

for $r = 1, \ldots, R$, where the transformed selection matrices $\boldsymbol{K}_1^{(r)}$ and $\boldsymbol{K}_2^{(r)}$ are given by

$$\boldsymbol{K}_1^{(r)} = 2 \cdot \mathrm{Re}\left\{ \boldsymbol{Q}_{M_r^{(\mathrm{sel})}}^{\mathrm{H}} \cdot \boldsymbol{J}_2^{(r)} \cdot \boldsymbol{Q}_{M_r} \right\} \tag{3.45}$$

$$\boldsymbol{K}_2^{(r)} = 2 \cdot \mathrm{Im}\left\{ \boldsymbol{Q}_{M_r^{(\mathrm{sel})}}^{\mathrm{H}} \cdot \boldsymbol{J}_2^{(r)} \cdot \boldsymbol{Q}_{M_r} \right\}. \tag{3.46}$$

Notice that the real-valued shift invariance equation in (3.44) has the same form as the shift invariance equation for R-D Standard Tensor-ESPRIT in (3.41). Hence, in analogy to (3.43), the simple LS solution to (3.44) is found via

$$\hat{\boldsymbol{\Upsilon}}_{\mathrm{LS}}^{(r)^{\mathrm{T}}} = \left(\tilde{\boldsymbol{K}}_1^{(r)} \cdot \left[\hat{\boldsymbol{\mathcal{E}}}^{[\mathrm{s}]} \right]_{(R+1)}^{\mathrm{T}} \right)^{+} \cdot \tilde{\boldsymbol{K}}_2^{(r)} \cdot \left[\hat{\boldsymbol{\mathcal{E}}}^{[\mathrm{s}]} \right]_{(R+1)}^{\mathrm{T}}. \tag{3.47}$$

Note that the matrix-based R-D Unitary ESPRIT algorithm from Section 3.4.2 and the R-D Unitary Tensor-ESPRIT algorithm are algebraically equivalent except for the fact that the real-valued SVD-based subspace estimate $\boldsymbol{E}_{\mathrm{s}}$ is replaced by the real-valued HOSVD-based subspace estimate $\hat{\boldsymbol{\mathcal{E}}}^{[\mathrm{s}]}$.

Note also that the transformed selection matrices $\tilde{\boldsymbol{K}}_1^{(r)}$ and $\tilde{\boldsymbol{K}}_2^{(r)}$ in (3.47) are obtained from (3.45) and (3.46) via [HRD08]

$$\tilde{\boldsymbol{K}}_n^{(r)} = \left(\boldsymbol{I}_{M_1} \otimes \ldots \otimes \boldsymbol{I}_{M_{r-1}} \right) \otimes \boldsymbol{K}_n^{(r)} \otimes \left(\boldsymbol{I}_{M_{r+1}} \otimes \ldots \otimes \boldsymbol{I}_{M_R} \right), \quad n = 1, 2. \tag{3.48}$$

They only coincide with the selection matrices $\tilde{\boldsymbol{K}}_n^{(r)}$ for R-D Unitary ESPRIT that are defined in (3.36) and (3.37) if the unitary left-$\boldsymbol{\Pi}$-real matrices $\boldsymbol{Q}_M$ and $\boldsymbol{Q}_{M_r^{(\mathrm{sel})} \cdot M/M_r}$ for R-D Unitary ESPRIT are chosen as

$$\boldsymbol{Q}_M = \boldsymbol{Q}_{M_1} \otimes \boldsymbol{Q}_{M_2} \otimes \ldots \otimes \boldsymbol{Q}_{M_r} \otimes \ldots \otimes \boldsymbol{Q}_{M_R} \tag{3.49}$$

$$\boldsymbol{Q}_{M_r^{(\mathrm{sel})} \cdot M/M_r} = \boldsymbol{Q}_{M_1} \otimes \boldsymbol{Q}_{M_2} \otimes \ldots \otimes \boldsymbol{Q}_{M_r^{(\mathrm{sel})}} \otimes \ldots \otimes \boldsymbol{Q}_{M_R}, \tag{3.50}$$

where the matrices $\boldsymbol{Q}_{M_r}$ are arbitrary unitary left-$\boldsymbol{\Pi}$-real matrices.

The R-D Unitary Tensor-ESPRIT algorithm is summarized in Algorithm 4.

Algorithm 4 [HRD08] Summary of R-D Unitary Tensor-ESPRIT

1. Estimate the real-valued signal subspace tensor $\hat{\boldsymbol{\mathcal{E}}}^{[\mathrm{s}]} \in \mathbb{R}^{M_1 \times \ldots \times M_R \times d}$ via the truncated HOSVD of the transformed observation tensor $\varphi\left(\boldsymbol{\mathcal{X}}^{(\mathrm{fba})}\right) \in \mathbb{R}^{M_1 \times \ldots \times M_R \times 2N}$ shown in (3.24).
2. Solve the overdetermined shift invariance equations
$$\hat{\boldsymbol{\mathcal{E}}}^{[\mathrm{s}]} \times_r \boldsymbol{K}_1^{(r)} \times_{R+1} \hat{\boldsymbol{\Upsilon}}^{(r)} \approx \hat{\boldsymbol{\mathcal{E}}}^{[\mathrm{s}]} \times_r \boldsymbol{K}_2^{(r)}$$
for the matrices $\hat{\boldsymbol{\Upsilon}}^{(r)}$ for $r = 1, 2, \ldots, R$ via LS methods, e.g., LS, TLS, SLS, TS-SLS, GLS.
3. Compute the eigenvalues $\hat{\omega}_i^{(r)}$ for $i = 1, 2, \ldots, d$ of $\hat{\boldsymbol{\Upsilon}}^{(r)}$ jointly for all $r = 1, \ldots, R$, e.g., via the joint diagonalization scheme proposed in [FG06] or via the Simultaneous Schur Decomposition proposed in [HN98]. Recover the correctly paired frequencies $\hat{\mu}_i^{(r)}$ via $\hat{\mu}_i^{(r)} = 2 \cdot \arctan(\hat{\omega}_i^{(r)})$.

3.6. Summary

In this chapter, we have reviewed R-D ESPRIT-type algorithms for the parameter estimation of multi-dimensional signals sampled on a separable R-D grid. We have discussed the matrix-based R-D Standard ESPRIT and R-D Unitary ESPRIT algorithms as well as the R-D Standard Tensor-ESPRIT and R-D Unitary Tensor-ESPRIT algorithms.

It has been shown that the matrix-based subspace estimation step using the SVD can be improved by enforcing the multi-dimensional structure of the signals onto the signal subspace estimate by means of the truncated HOSVD. Incorporating this structure, the denoising can be performed more efficiently in each dimension. Hence, it serves as a generic approach to improve subspace-based parameter estimation methods, e.g., R-D ESPRIT, R-D MODE, R-D RARE, or R-D MUSIC.

For R-D ESPRIT, the tensor-based signal subspace estimate then leads to a tensor extension of the matrix-based shift invariance equation, which is solved by ESPRIT-type parameter estimation algorithms. As a result, the tensor versions R-D Standard Tensor-ESPRIT and R-D Unitary Tensor-ESPRIT, which efficiently exploit the multi-dimensional structure of the signals have been reviewed. It has been demonstrated that the Tensor-ESPRIT-type algorithms using LS to solve the shift invariance equation are identical to their matrix-based version except for the signal subspace estimate which is replaced by the enhanced truncated HOSVD-based signal subspace estimate. This is not the case when TS-SLS, which exploits additional structure, is used to solve the R-D shift invariance equation. A summary of the different ESPRIT-type algorithms including their performance analysis as well as the different solutions to the overdetermined shift invariance

equation is provided in Chapter 10.

Simulation results to illustrate the empirical performance of the considered R-D ESPRIT-type algorithms along with their analytical performance expressions are provided in the next chapter.

4. Performance analysis of ESPRIT-type parameter estimation algorithms

In this chapter, we first review the performance analysis framework from [RHD14] for the matrix-based and tensor-based R-D ESPRIT-type algorithms using LS discussed in Chapter 3. The resulting analytical expressions for the parameter estimation error and the MSE only require the noise to be zero-mean with finite SO moments and are asymptotic in the effective SNR. In the contribution of this chapter, we simplify the MSE expressions of the matrix-based and tensor-based R-D ESPRIT-type algorithms using LS for a single source and two sources. Based on the resulting expressions, which only depend on the physical parameters, we analytically compute the gain from forward-backward averaging as well as the tensor gain for two sources. A detailed motivation including the state of the art and the specific contributions is given in Section 4.1. In Section 4.2, we review the performance analysis for the matrix-based and tensor-based subspace estimation. The analytical MSE expressions for the matrix-based and tensor-based R-D ESPRIT-type algorithms are provided in Section 4.3 and in Section 4.4, respectively. The simplified MSE expressions for the special cases of a single source and two sources are derived in Section 4.5 followed by numerical simulations in Section 4.6 and a summary in Section 4.7.

4.1. Overview

The performance of parameter estimation algorithms is usually compared by means of Monte-Carlo simulations. Thereby, for a specific parameter setting, the input to the algorithm in each realization is generated randomly and the performance with respect to some statistical measure is evaluated by averaging over several random realizations. However, despite their simplicity, Monte-Carlo simulations entail a number of disadvantages. For instance, they are only valid for a specific setting of the input parameters and they are non-objective as random fluctuations such as outliers can obscure the results. These aspects serve as a motivation for the development of analytical performance evaluation frameworks for the parameter estimation algorithms of interest. These frameworks allow for a performance prediction of the corresponding algorithms for different parameter settings and allow for an objective comparison of different algorithms without performing Monte-Carlo simulations.

In this chapter, we consider an asymptotic performance analysis framework for subspace-based parameter estimation algorithms. This framework provides an explicit first-order approximation of the error in the signal subspace estimate obtained from the singular value decomposition due to the noise. Therefore, it can be used to derive analytical expressions for the parameter estimation

error and the mean square error (MSE) of any subspace-based parameter estimation algorithm. As the resulting analytical expressions are deterministic in the noise term, no assumptions on the statistics of the noise are required. The noise only needs to be small[1] compared to the signal component. Moreover, the analytical MSE expressions can be simplified for special cases on the number of sources in order to provide insights into the dependence of the MSE on the physical parameters, e.g., the number of sensors, the SNR, and the correlation.

With the focus of ESPRIT-type parameter estimation algorithms, we provide a literature overview of the performance analysis frameworks for matrix-based and tensor-based multi-dimensional ESPRIT algorithms in Section 4.1.1 and list the contributions in Section 4.1.2.

4.1.1. State of the art

Subspace-based parameter estimation has long been a fundamental research area in the field of array signal processing for many decades [KV96]. This is mainly due to the development of the MUSIC algorithm [Sch86] and the ESPRIT algorithm [RPK86], which provide high-resolution capabilities in estimating the parameters of impinging signals.

With the growing popularity of subspace-based parameter estimation algorithms, their analytical performance assessment has attracted considerable attention. The two most prominent performance analysis frameworks have been proposed in [Bri75] and [LLV93]. The framework in [Bri75], which has, for instance, been applied to the MUSIC algorithm in [KB86, PF88, PK89a, Fri90], Root-MUSIC [RH89b], and to the ESPRIT algorithm in [RH89a, MHZ96], analyzes the analytical performance based on the eigenvector distribution of the sample covariance matrix. A major drawback of [Bri75] is that the impinging signals as well as the noise contribution are required to be Gaussian distributed. Hence, the concept in [Bri75] and its above mentioned follow-up papers have a limited applicability. Moreover, the resulting analytical expressions for the perturbation of the eigenvectors of the sample covariance matrix are only asymptotic in the sample size N, i.e, the results become exact only if the sample size N is very large. In addition, the expressions are rather long and difficult to simplify.

In contrast, the performance analysis framework in [LLV93] is based on a first-order perturbation expansion of the singular value decomposition (SVD), which models the estimation error of the signal subspace as an explicit function of the additive noise component. Therefore, it directly describes the leakage of the noise subspace into the signal subspace. The authors of [LLV93] have shown that the explicit subspace estimation error can be applied to derive analytical expressions for the parameter estimation error of any subspace-based parameter algorithm, e.g, MUSIC, Root-MUSIC, or ESPRIT. A major advantage of the framework in [LLV93] is that no assumptions on the statistics of the signals or the noise are required. It only assumes that the noise is small compared

[1] Note that by deriving a second-order performance analysis for 1-D Standard ESPRIT in [SRH17c], we have addressed the question of how small the noise should be for the first-order performance analysis to be valid.

to the received signal power. As a result, [LLV93] is applicable to a wider range of applications and is, for instance, even valid for non-Gaussian and non-circular perturbations often caused by clutter environments in radar applications [Gue03].

Moreover, the analytical expressions obtained from [LLV93] are asymptotic in the effective signal-to-noise ratio (SNR), i.e., the results become accurate for either high SNRs or a large sample size N. Thus, it is even valid in the case of a single snapshot $N = 1$ if the SNR is sufficiently high. Due to these advantages, the framework in [LLV93] is preferable over that in [Bri75]. There are several extensions of [LLV93] available in the literature. The work in [Xu02], for example, generalizes the first-order perturbation expansion of the SVD to the second-order (SO) performance expansion, while [LLM08] also derives a first-order performance expression for the perturbation within the basis of the signal subspace, which was ignored in [LLV93] and [Xu02]. However, it was shown in [LLM08] that the perturbation within the basis is irrelevant for subspace-based algorithms.

Considering ESPRIT-type algorithms using least squares (LS) to solve the shift invariance equation, the asymptotic performance of 1-D Standard ESPRIT for harmonic retrieval from time series has been studied in [SS91, ESS93]. For ESPRIT-based parameter estimation, the authors of [LLV93] also provide a closed-form mean square error (MSE) expression for 1-D Standard ESPRIT, which, however, assumes a circularly symmetric noise distribution and does not generalize to multi-dimensional parameter estimation. As an extension, the contributions in [RH12] and [RHD14] derive analytical MSE expressions that only require the noise to be zero-mean with finite SO moments regardless of its statistics. Moreover, the work in [RBHW09, RBH10, RHD14] extends the performance analysis framework of [LLV93] to the case of matrix-based R-D ESPRIT-type algorithms, i.e., R-D Standard ESPRIT and R-D Unitary ESPRIT as well as tensor-based R-D ESPRIT-type algorithms, i.e., R-D Standard Tensor-ESPRIT and R-D Unitary Tensor-ESPRIT. These R-D ESPRIT-type algorithms are discussed in Chapter 3. A performance analysis of 1-D Standard ESPRIT using SLS to solve the shift invariance equation is proposed in [RH11].

The analytical MSE expressions are formulated in terms of the subspaces of the measurement matrix. However, in many applications, it is desirable to know how the MSE scales with the explicit physical system parameters, e.g., the number of sensors, the SNR, the number of snapshots, etc. This can be achieved by simplifying the analytical MSE expressions for special cases on the number of sources, i.e., a single source or two sources. The performance analysis of ESPRIT for the special case of a single source is considered in [RH89a] and the asymptotic efficiency of MUSIC, Root-MUSIC, and TLS-ESPRIT was presented in [PF88, RH89b] and [OVK91], respectively. However, these results are asymptotic in the sample size N or even in the number of sensors M. The results presented in this chapter are also accurate for small values of M and asymptotic in the effective SNR.

4.1.2. Contribution

In this chapter, we first review the R-D performance analysis framework in [RHD14] for the matrix-based R-D ESPRIT-type algorithms, i.e., R-D Standard ESPRIT and R-D Unitary ESPRIT as well as for the tensor-based R-D ESPRIT-type algorithms, i.e., R-D Standard Tensor-ESPRIT and R-D Unitary Tensor-ESPRIT. The analytical expressions for the parameter estimation error and the MSE only require the noise to be zero-mean with finite SO moments and are asymptotic in the effective SNR. In the contribution of this chapter described in Section 4.5, we derive simplified MSE expressions of the matrix-based and tensor-based R-D ESPRIT-type algorithms using LS for a single source and two sources. The resulting expressions only depend on the physical parameters, e.g., the number of antennas M, the SNR, etc. For the single source case considered in Section 4.5.1, the matrix-based and tensor-based R-D ESPRIT-type algorithms yield the same MSE and perform identical. For the two source case discussed in Section 4.5.2, we start with the matrix-based R-D ESPRIT-type algorithms and derive closed-form MSE expressions for two sources. Some parts of these results have been published in [SRH14a]. Then, we use these MSE expressions to analytically compute the gain from forward-backward averaging and analyze its dependence on the physical parameters. For the simplified MSE of the tensor-based R-D ESPRIT-type algorithms for two sources, we compute and analyze the tensor gain as well as the forward-backward averaging gain in the tensor case and analyze its behavior.

Note that the performance analysis expressions for the matrix-based and tensor-based R-D ESPRIT-type algorithms using LS as well as the considered special source cases in this chapter form the basis of the derivation of the performance analysis for R-D ESPRIT-type algorithms using GLS to solve the shift invariance equation as well as the analytical performance of R-D NC ESPRIT-type algorithms, which are provided in Chapter 5 and Chapter 6, respectively.

4.2. Performance analysis of subspace estimation

In this section, we review the performance analysis result of the subspace estimation, which is the basis of subspace-based parameter estimation algorithms. In Section 4.2.1, we first review the results from [LLV93] for the matrix-based subspace estimation based on the SVD and in Section 4.2.2, we discuss the tensor-based subspace estimation using the truncated HOSVD presented in [RHD14].

4.2.1. Performance analysis of matrix-based subspace estimation

In this section, we review the performance analysis framework for the SVD-based subspace estimation from [LLV93]. Recall the matrix-based model

$$\boldsymbol{X} = \boldsymbol{X}_0 + \boldsymbol{N} \in \mathbb{C}^{M \times N} \tag{4.1}$$

introduced in (3.1) of Section 3.2, where $\boldsymbol{X}_0 \in \mathbb{C}^{M \times N}$ denotes the noise-free measurement matrix comprising N subsequent snapshots and $\boldsymbol{N} \in \mathbb{C}^{M \times N}$ represents the additive noise component. Following (3.2) and (3.4), the SVDs of $\boldsymbol{X}_0$ and $\boldsymbol{X}$ can be written as

$$\boldsymbol{X}_0 = \begin{bmatrix} \boldsymbol{U}_\mathrm{s} & \boldsymbol{U}_\mathrm{n} \end{bmatrix} \cdot \begin{bmatrix} \boldsymbol{\Sigma}_\mathrm{s} & \boldsymbol{0}_{d \times (N-d)} \\ \boldsymbol{0}_{(M-d) \times d} & \boldsymbol{0}_{(M-d) \times (N-d)} \end{bmatrix} \cdot \begin{bmatrix} \boldsymbol{V}_\mathrm{s} & \boldsymbol{V}_\mathrm{n} \end{bmatrix}^\mathrm{H} \tag{4.2}$$

$$\boldsymbol{X} = \begin{bmatrix} \hat{\boldsymbol{U}}_\mathrm{s} & \hat{\boldsymbol{U}}_\mathrm{n} \end{bmatrix} \cdot \begin{bmatrix} \hat{\boldsymbol{\Sigma}}_\mathrm{s} & \boldsymbol{0}_{d \times (N-d)} \\ \boldsymbol{0}_{(M-d) \times d} & \hat{\boldsymbol{\Sigma}}_\mathrm{n} \end{bmatrix} \cdot \begin{bmatrix} \hat{\boldsymbol{V}}_\mathrm{s} & \hat{\boldsymbol{V}}_\mathrm{n} \end{bmatrix}^\mathrm{H}, \tag{4.3}$$

where $\boldsymbol{U}_\mathrm{s} \in \mathbb{C}^{M \times d}$, $\boldsymbol{U}_\mathrm{n} \in \mathbb{C}^{M \times (M-d)}$, and $\boldsymbol{V}_\mathrm{s} \in \mathbb{C}^{N \times d}$ span the signal subspace, the noise subspace, and the row space, respectively, and $\boldsymbol{\Sigma}_\mathrm{s} = \mathrm{diag}\left\{ \begin{bmatrix} \sigma_1, \sigma_2, \ldots, \sigma_d \end{bmatrix}^\mathrm{T} \right\} \in \mathbb{R}^{d \times d}$ contains the d non-zero singular values on its diagonal. Note that their estimated equivalents obtained from the SVD of $\boldsymbol{X}$ are denoted by a "hat".

As we are mainly interested in estimating the signal subspace, we can express the perturbed signal subspace estimate as $\hat{\boldsymbol{U}}_\mathrm{s} = \boldsymbol{U}_\mathrm{s} + \Delta\boldsymbol{U}_\mathrm{s}$, where $\Delta\boldsymbol{U}_\mathrm{s}$ represents the estimation error. Then, according to [LLV93], the first-order approximation of $\Delta\boldsymbol{U}_\mathrm{s}$ is given by

$$\Delta\boldsymbol{U}_\mathrm{s} = \boldsymbol{U}_\mathrm{n} \cdot \boldsymbol{U}_\mathrm{n}^\mathrm{H} \cdot \boldsymbol{N} \cdot \boldsymbol{V}_\mathrm{s} \cdot \boldsymbol{\Sigma}_\mathrm{s}^{-1} + \mathcal{O}\left\{\nu^2\right\} \in \mathbb{C}^{M \times d}, \tag{4.4}$$

where $\nu = \|\boldsymbol{N}\|$ and $\|.\|$ stands for an arbitrary sub-multiplicative[2] norm such as the Frobenius norm. Note that the term $\mathcal{O}\left\{\nu^2\right\}$ in (4.4) represents the second-order term, which is neglected due to the first-order approximation. In what follows, we drop the term $\mathcal{O}\left\{\nu^2\right\}$ and use the notation "$\approx$" instead of "=" to denote "up to the first-order". Since the first-order expansion of the signal subspace estimation error $\Delta\boldsymbol{U}_\mathrm{s}$ in (4.4) is formulated in terms of the noise subspace $\boldsymbol{U}_\mathrm{n}$, it models the leakage of the noise subspace into the signal subspace due to the effect of the perturbation $\boldsymbol{N}$. Moreover, (4.4) is deterministic in $\boldsymbol{N}$ such that no assumptions about the statistics of $\boldsymbol{N}$ are required.

We should highlight that the expansion in (4.4) ignores the perturbation of $\Delta\boldsymbol{U}_\mathrm{s}$ in the particular basis for the signal subspace, i.e., the perturbation within the columns of $\boldsymbol{U}_\mathrm{s}$. This term is, for instance, taken into account in [LLM08], which is an extension of (4.4). However, for subspace-

[2] A matrix norm is called submultiplicative if $\|\boldsymbol{A} \cdot \boldsymbol{B}\| \leq \|\boldsymbol{A}\| \cdot \|\boldsymbol{B}\|$ for arbitrary matrices $\boldsymbol{A}$ and $\boldsymbol{B}$.

based parameter estimation algorithms, the particular choice of the basis is irrelevant and does not affect the estimation performance. In [LLM08], this fact was proven for the example of ESPRIT-type algorithms. Consequently, we only consider (4.4) in the following sections.

4.2.2. Performance analysis of tensor-based subspace estimation

First recall the tensor model from (3.6) as

$$\boldsymbol{\mathcal{X}} = \boldsymbol{\mathcal{X}}_0 + \boldsymbol{\mathcal{N}}, \tag{4.5}$$

where $\boldsymbol{\mathcal{X}}_0$ is the noise-free measurement tensor and $\boldsymbol{\mathcal{N}}$ represents the additive noise tensor. It was shown in Section 3.2.2 that an enhanced signal subspace estimate, which takes the multi-dimensional signal structure into account can be computed via the HOSVD of the measurement tensor $\boldsymbol{\mathcal{X}}$. As the HOSVD is constructed from SVDs of the r-mode unfoldings (cf. Section 3.2.2), the same performance analysis framework from the previous section can be applied to obtain a first-order perturbation expansion of the HOSVD-based subspace estimate. Following (3.11) and (3.12), the SVDs of the r-mode unfoldings of the noise-free tensor $\boldsymbol{\mathcal{X}}_0$ and the noisy tensor $\boldsymbol{\mathcal{X}}$ and are given by

$$[\boldsymbol{\mathcal{X}}_0]_{(r)} = \begin{bmatrix} \boldsymbol{U}_r^{[\mathrm{s}]} & \boldsymbol{U}_r^{[\mathrm{n}]} \end{bmatrix} \cdot \begin{bmatrix} \boldsymbol{\Sigma}_r^{[\mathrm{s}]} & \boldsymbol{0}_{d\times(M\cdot N/M_r-d)} \\ \boldsymbol{0}_{(M_r-d)\times d} & \boldsymbol{0}_{(M_r-d)\times(M\cdot N/M_r-d)} \end{bmatrix} \cdot \begin{bmatrix} \boldsymbol{V}_r^{[\mathrm{s}]} & \boldsymbol{V}_r^{[\mathrm{n}]} \end{bmatrix}^{\mathrm{H}} \tag{4.6}$$

$$[\boldsymbol{\mathcal{X}}]_{(r)} = \begin{bmatrix} \hat{\boldsymbol{U}}_r^{[\mathrm{s}]} & \hat{\boldsymbol{U}}_r^{[\mathrm{n}]} \end{bmatrix} \cdot \begin{bmatrix} \hat{\boldsymbol{\Sigma}}_r^{[\mathrm{s}]} & \boldsymbol{0}_{d\times(M\cdot N/M_r-d)} \\ \boldsymbol{0}_{(M_r-d)\times d} & \hat{\boldsymbol{\Sigma}}_r^{[\mathrm{n}]} \end{bmatrix} \cdot \begin{bmatrix} \hat{\boldsymbol{V}}_r^{[\mathrm{s}]} & \hat{\boldsymbol{V}}_r^{[\mathrm{n}]} \end{bmatrix}^{\mathrm{H}} \tag{4.7}$$

for $r = 1, 2, \ldots, R$. Writing $\hat{\boldsymbol{U}}_r^{[\mathrm{s}]} = \boldsymbol{U}_r^{[\mathrm{s}]} + \Delta\boldsymbol{U}_r^{[\mathrm{s}]}$ and applying (4.4), the first-order expansion of $\Delta\boldsymbol{U}_r^{[\mathrm{s}]}$ can be expressed as

$$\Delta\boldsymbol{U}_r^{[\mathrm{s}]} \approx \boldsymbol{U}_r^{[\mathrm{n}]} \cdot \boldsymbol{U}_r^{[\mathrm{n}]^{\mathrm{H}}} \cdot [\boldsymbol{\mathcal{N}}]_{(r)} \cdot \boldsymbol{V}_r^{[\mathrm{s}]} \cdot \boldsymbol{\Sigma}_r^{[\mathrm{s}]^{-1}}. \tag{4.8}$$

Then, from Section 3.2.2, we recall that the HOSVD-based subspace estimate is given by $\left[\hat{\boldsymbol{\mathcal{U}}}^{[\mathrm{s}]}\right]_{(R+1)}^{\mathrm{T}}$, where the signal subspace tensor $\hat{\boldsymbol{\mathcal{U}}}^{[\mathrm{s}]}$ is defined in (3.14). In order to find the first-order expansion of $\left[\hat{\boldsymbol{\mathcal{U}}}^{[\mathrm{s}]}\right]_{(R+1)}^{\mathrm{T}}$, the algebraic relation between the HOSVD-based subspace estimate $\left[\hat{\boldsymbol{\mathcal{U}}}^{[\mathrm{s}]}\right]_{(R+1)}^{\mathrm{T}}$ and the SVD-based estimate $\hat{\boldsymbol{U}}_{\mathrm{s}}$ stated in (3.15) can be considered. It was derived

in [RBHW09, RHD14] that a first-order expansion of $\left[\hat{\boldsymbol{U}}^{[\mathrm{s}]}\right]^{\mathrm{T}}_{(R+1)}$ can be expressed as

$$\left[\hat{\boldsymbol{U}}^{[\mathrm{s}]}\right]^{\mathrm{T}}_{(R+1)} = \boldsymbol{U}_{\mathrm{s}} + \left[\Delta\hat{\boldsymbol{U}}^{[\mathrm{s}]}\right]^{\mathrm{T}}_{(R+1)}, \tag{4.9}$$

where the HOSVD-based signal subspace error is given by

$$\left[\Delta\hat{\boldsymbol{U}}^{[\mathrm{s}]}\right]^{\mathrm{T}}_{(R+1)} \approx \boldsymbol{T}^{\otimes}_{1:R} \cdot \Delta\boldsymbol{U}_{\mathrm{s}} + \sum_{r=1}^{R} \left(\boldsymbol{T}^{\otimes}_{1:r-1} \otimes \left[\Delta\boldsymbol{U}^{[\mathrm{s}]}_{r} \cdot \boldsymbol{U}^{[\mathrm{s}]^{\mathrm{H}}}_{r}\right] \otimes \boldsymbol{T}^{\otimes}_{r+1:R}\right) \cdot \boldsymbol{U}_{\mathrm{s}}. \tag{4.10}$$

Here, $\boldsymbol{T}^{\otimes}_{a:b}$ is defined in (1.24), the SVD-based signal subspace perturbation $\Delta\boldsymbol{U}_{\mathrm{s}}$ is given by (4.4), and the perturbation of the r-spaces is computed via (4.8). Note that (4.10) is a more compact formulation of the expression given in [RHD14] and that the special case for $R = 2$ was first derived in [RBHW09].

4.3. Performance of R-D matrix-based ESPRIT-type algorithms

This section reviews the first-order performance analysis framework for R-D matrix-based ESPRIT-type algorithms using least squares (LS) from [RHD14]. In particular, the multi-dimensional extensions of the analytical performance results for 1-D Standard ESPRIT from [LLV93] to R-D Standard ESPRIT and R-D Unitary ESPRIT are summarized in Section 4.3.1 and Section 4.3.2, respectively.

4.3.1. Performance of R-D Standard ESPRIT

The authors of [LLV93] show that the first-order expansion of the subspace estimation error (cf. Section 4.2.1) can be used to find a corresponding first-order expansion of the estimation error of 1-D Standard ESPRIT algorithm based on LS. In [RHD14], it is shown that this result is easily generalized to the R-D case. This is due to the fact that for R-D LS-based ESPRIT, the R shift invariance equations are solved independently. To ensure the correct pairing of the parameters across dimensions, a joint eigendecomposition of all R dimensions is performed. Note that this step is not included in the performance analysis as it has no impact on the asymptotic estimation error of the spatial frequencies for high SNRs since the eigenvectors become asymptotically equal [LT78, RHD14, BCW+17].

Hence, the results from [LLV93] can be applied to each of the modes individually. Thus, the first-order approximation for the estimation error of the i-th spatial frequency in the r-th mode is

given by [RHD14]

$$\Delta\mu_i^{(r)} \approx \operatorname{Im}\left\{\boldsymbol{p}_i^{\mathrm{T}} \cdot \left(\tilde{\boldsymbol{J}}_1^{(r)} \cdot \boldsymbol{U}_{\mathrm{s}}\right)^{+} \cdot \left[\tilde{\boldsymbol{J}}_2^{(r)}/\lambda_i^{(r)} - \tilde{\boldsymbol{J}}_1^{(r)}\right] \cdot \Delta\boldsymbol{U}_{\mathrm{s}} \cdot \boldsymbol{q}_i\right\}, \tag{4.11}$$

where $\lambda_i^{(r)} = \mathrm{e}^{\mathrm{j}\mu_i^{(r)}}$ and $\boldsymbol{q}_i$ and $\boldsymbol{p}_i^{\mathrm{T}}$ are the i-th column of $\boldsymbol{Q}$ and the i-th row vector of the matrix $\boldsymbol{P} = \boldsymbol{Q}^{-1}$, respectively. The matrix $\boldsymbol{Q}$ is obtained from the eigendecomposition $\boldsymbol{\Psi}^{(r)} = \boldsymbol{Q} \cdot \boldsymbol{\Lambda}^{(r)} \cdot \boldsymbol{Q}^{-1}$ in the r-th mode. Moreover, the R-D selection matrices $\tilde{\boldsymbol{J}}_1^{(r)}, \tilde{\boldsymbol{J}}_2^{(r)} \in \mathbb{R}^{\frac{M}{M_r} \cdot M_r^{(\mathrm{sel})} \times M}$ for the first and the second subarray in the r-th mode are computed according to (3.26) by [HN98]

$$\tilde{\boldsymbol{J}}_\ell^{(r)} = \boldsymbol{I}_{\prod_{n=1}^{r-1} M_n} \otimes \boldsymbol{J}_\ell^{(r)} \otimes \boldsymbol{I}_{\prod_{n=r+1}^{R} M_n} \tag{4.12}$$

for $\ell = 1, 2$ and $r = 1, 2, \ldots, R$ and where $\boldsymbol{J}_\ell^{(r)} \in \mathbb{R}^{M_r^{(\mathrm{sel})} \times M_r}$ are the r-mode selection matrices that select the $M_r^{(\mathrm{sel})}$ elements for the first and the second subarray in the r-th mode. The derivation of (4.11) is shown in Appendix B.2.

Notice that upon applying (1.14) to (4.11) and inserting (4.4), a compact formulation of (4.11) is given by

$$\Delta\mu_i^{(r)} \approx \operatorname{Im}\left\{\boldsymbol{r}_i^{(r)^{\mathrm{T}}} \cdot \operatorname{vec}\{\Delta\boldsymbol{U}_{\mathrm{s}}\}\right\} = \operatorname{Im}\left\{\boldsymbol{r}_i^{(r)^{\mathrm{T}}} \cdot \boldsymbol{W}_{\mathrm{mat}} \cdot \boldsymbol{n}\right\}, \tag{4.13}$$

where the vector $\boldsymbol{r}_i^{(r)} \in \mathbb{C}^{Md \times 1}$ and the matrix $\boldsymbol{W}_{\mathrm{mat}} \in \mathbb{C}^{Md \times MN}$ are defined as

$$\boldsymbol{r}_i^{(r)} = \boldsymbol{q}_i \otimes \left(\left[\left(\tilde{\boldsymbol{J}}_1^{(r)}\boldsymbol{U}_{\mathrm{s}}\right)^{+} \cdot \left(\tilde{\boldsymbol{J}}_2^{(r)}/\mathrm{e}^{\mathrm{j}\cdot\mu_i^{(r)}} - \tilde{\boldsymbol{J}}_1^{(r)}\right)\right]^{\mathrm{T}} \cdot \boldsymbol{p}_i\right) \tag{4.14}$$

$$\boldsymbol{W}_{\mathrm{mat}} = \left(\boldsymbol{\Sigma}_{\mathrm{s}}^{-1} \cdot \boldsymbol{V}_{\mathrm{s}}^{\mathrm{T}}\right) \otimes \left(\boldsymbol{U}_{\mathrm{n}} \cdot \boldsymbol{U}_{\mathrm{n}}^{\mathrm{H}}\right) \tag{4.15}$$

and $\boldsymbol{n} = \operatorname{vec}\{\boldsymbol{N}\} \in \mathbb{C}^{MN \times 1}$ is the vectorized noise contribution.

In the special case of 1-D parameter estimation presented in [LLV93], the estimation error of the i-th spatial frequency in (4.11) simplifies to

$$\Delta\mu_i \approx \operatorname{Im}\left\{\boldsymbol{p}_i^{\mathrm{T}} \cdot (\boldsymbol{J}_1 \cdot \boldsymbol{U}_{\mathrm{s}})^{+} \cdot [\boldsymbol{J}_2/\lambda_i - \boldsymbol{J}_1] \cdot \Delta\boldsymbol{U}_{\mathrm{s}} \cdot \boldsymbol{q}_i\right\}, \tag{4.16}$$

where $\boldsymbol{J}_1$ and $\boldsymbol{J}_2$ are the 1-D selection matrices.

Note that since $\Delta\boldsymbol{U}_{\mathrm{s}}$ is deterministic in the perturbation $\boldsymbol{N}$, the same holds true for $\Delta\mu_i^{(r)}$. Thus, no statistical assumptions on the noise are required for (4.11), which is a major advantage of the performance analysis framework. However, the estimation accuracy of parameter estimation algorithms is usually compared with respect to the mean square error (MSE), where the ensemble average over all possible noise realizations is computed.

In [RHD14], the following analytical expression for the MSE of the R-D Standard ESPRIT algo-

rithm has been derived. Consider (4.13) and suppose the noise samples in $\boldsymbol{n}$ are zero-mean random variables with finite second-order (SO) moments, which are described by the covariance matrix $\boldsymbol{R}_{\mathrm{nn}} = \mathbb{E}\left\{\boldsymbol{n} \cdot \boldsymbol{n}^{\mathrm{H}}\right\} \in \mathbb{C}^{MN \times MN}$ and the pseudo-covariance matrix $\boldsymbol{C}_{\mathrm{nn}} = \mathbb{E}\left\{\boldsymbol{n} \cdot \boldsymbol{n}^{\mathrm{T}}\right\} \in \mathbb{C}^{MN \times MN}$. Then, the MSE for the i-th spatial frequency in the r-th mode can be expressed as

$$\mathbb{E}\left\{\left(\Delta\mu_i^{(r)}\right)^2\right\} \approx \frac{1}{2} \cdot \left(\boldsymbol{z}_i^{(r)^{\mathrm{H}}} \cdot \boldsymbol{R}_{\mathrm{nn}}^{\mathrm{T}} \cdot \boldsymbol{z}_i^{(r)} - \mathrm{Re}\left\{\boldsymbol{z}_i^{(r)^{\mathrm{T}}} \cdot \boldsymbol{C}_{\mathrm{nn}}^{\mathrm{T}} \cdot \boldsymbol{z}_i^{(r)}\right\}\right), \tag{4.17}$$

where $\boldsymbol{z}_i^{(r)} = \boldsymbol{W}_{\mathrm{mat}}^{\mathrm{T}} \cdot \boldsymbol{r}_i^{(r)} \in \mathbb{C}^{MN \times 1}$ for $i = 1, \ldots, d$ and $r = 1, \ldots, R$, and $\boldsymbol{r}_i^{(r)}$ and $\boldsymbol{W}_{\mathrm{mat}}$ are given in (4.14) and (4.15), respectively. The proof of (4.17) is shown in [RHD14]. We stress again that the MSE in (4.17) does not require any assumptions on the noise statistics except for a zero mean and finite SO statistics.

Note that in the special case of circularly symmetric white noise, which corresponds to $\boldsymbol{R}_{\mathrm{nn}} = \sigma_{\mathrm{n}}^2 \cdot \boldsymbol{I}_{MN}$ and $\boldsymbol{C}_{\mathrm{nn}} = \boldsymbol{0}_{MN \times MN}$, the MSE in (4.17) for the i-th spatial frequency in the r-th mode simplifies to

$$\mathbb{E}\left\{\left(\Delta\mu_i^{(r)}\right)^2\right\} \approx \frac{\sigma_{\mathrm{n}}^2}{2} \cdot \left\|\boldsymbol{z}_i^{(r)}\right\|_2^2. \tag{4.18}$$

To emphasize the practical significance of the presented performance analysis framework, we return to the 2-D DOA estimation problem in Example 2.1.2 and apply the analytical expressions for 2-D Standard ESPRIT.

Example 2.1.2 (continued): In Example 2.1.2, we have considered the task of estimating the azimuth and co-elevation angles of d signals impinging on a $M_1 \times M_2$ URA. We assume that the spacing between the sensor elements is equal in both dimensions, i.e., $\Delta^{(1)} = \Delta^{(2)} = \Delta$. In this case, the relation between the two spatial frequencies $\mu_i^{(1)}$ and $\mu_i^{(2)}$ and the azimuth angle θ_i ($-180° < \theta_i \leq 180°$) and co-elevation angle ϕ_i ($0° \leq \phi_i \leq 90°$) of the i-th signal is given by

$$\mu_i^{(1)} = \frac{2\pi}{\lambda_{\mathrm{c}}} \cdot \Delta \cdot \cos(\theta_i) \cdot \sin(\phi_i) \tag{4.19}$$

$$\mu_i^{(2)} = \frac{2\pi}{\lambda_{\mathrm{c}}} \cdot \Delta \cdot \sin(\theta_i) \cdot \sin(\phi_i). \tag{4.20}$$

Writing the parameter estimates as $\hat{\theta}_i = \theta_i + \Delta\theta_i$ and $\hat{\phi}_i = \phi_i + \Delta\phi_i$, where $\Delta\theta_i$ and $\Delta\phi_i$ are the parameter errors, the MSE expressions of 2-D Standard ESPRIT for this example are given in the following theorem:

Theorem 4.3.1. *Assuming a URA, the first-order MSE expressions of 2-D Standard ESPRIT for*

the angles θ_i and ϕ_i of the i-th signal are given by

$$\mathbb{E}\left\{(\Delta\theta_i)^2\right\} \approx \left(\frac{1}{\frac{2\pi}{\lambda_c}\cdot\Delta\cdot\sin\phi_i}\right)^2 \cdot \Big((\cos\theta_i)^2\cdot\mathbb{E}\left\{(\Delta\mu_i^{(2)})^2\right\} + (\sin\theta_i)^2\cdot\mathbb{E}\left\{(\Delta\mu_i^{(1)})^2\right\} - 2\cdot\sin\theta_i\cdot\cos\theta_i\cdot\mathbb{E}\left\{\Delta\mu_i^{(1)}\cdot\Delta\mu_i^{(2)}\right\}\Big) \tag{4.21}$$

and

$$\mathbb{E}\left\{(\Delta\phi_i)^2\right\} \approx \left(\frac{1}{\frac{2\pi}{\lambda_c}\cdot\Delta\cdot\cos\phi_i}\right)^2 \cdot \Big((\sin\theta_i)^2\cdot\mathbb{E}\left\{(\Delta\mu_i^{(2)})^2\right\} + (\cos\theta_i)^2\cdot\mathbb{E}\left\{(\Delta\mu_i^{(1)})^2\right\} + 2\cdot\sin\theta_i\cdot\cos\theta_i\cdot\mathbb{E}\left\{\Delta\mu_i^{(1)}\cdot\Delta\mu_i^{(2)}\right\}\Big), \tag{4.22}$$

where $\mathbb{E}\left\{(\Delta\mu_i^{(r)})^2\right\}$ for $r = 1,2$ is given in (4.17). Moreover, the expression for $\mathbb{E}\left\{\Delta\mu_i^{(1)}\cdot\Delta\mu_i^{(2)}\right\}$ has been derived by us in [SRH16a] as

$$\mathbb{E}\left\{\Delta\mu_i^{(1)}\cdot\Delta\mu_i^{(2)}\right\} \approx \frac{1}{2}\cdot\left(\boldsymbol{z}_i^{(1)^{\mathrm{H}}}\cdot\boldsymbol{R}_{\mathrm{nn}}^{\mathrm{T}}\cdot\boldsymbol{z}_i^{(2)} - \mathrm{Re}\left\{\boldsymbol{z}_i^{(1)^{\mathrm{T}}}\cdot\boldsymbol{C}_{\mathrm{nn}}^{\mathrm{T}}\cdot\boldsymbol{z}_i^{(2)}\right\}\right),$$

where $\boldsymbol{z}_i^{(r)} = \boldsymbol{W}_{\mathrm{mat}}^{\mathrm{T}}\cdot\boldsymbol{r}_i^{(r)}$ for $r = 1,2$ and $\boldsymbol{r}_i^{(r)}$ and $\boldsymbol{W}_{\mathrm{mat}}$ are given by (4.14) and (4.15), respectively.

The proof is provided in Appendix B.3.

4.3.2. Performance of R-D Unitary ESPRIT

In this section, we review the extension of the performance analysis expressions for R-D Standard ESPRIT from the previous section to R-D Unitary ESPRIT. The additional features of R-D Unitary ESPRIT are the incorporation of forward-backward averaging (FBA) and the transformation into the real-valued domain to reduce the computational complexity (cf. Section 3.4.2).

As discussed in Section 3.2.3, FBA virtually doubles the number of snapshots N of the original data $\boldsymbol{X} \in \mathbb{C}^{M\times N}$ by defining a column augmented measurement matrix $\boldsymbol{X}^{(\mathrm{fba})} \in \mathbb{C}^{M\times 2N}$ according to (3.16), which is restated here for convenience as

$$\boldsymbol{X}^{(\mathrm{fba})} = \begin{bmatrix}\boldsymbol{X} & \boldsymbol{\Pi}_M\cdot\boldsymbol{X}^*\cdot\boldsymbol{\Pi}_N\end{bmatrix} \in \mathbb{C}^{M\times 2N}. \tag{4.23}$$

Considering the model $\boldsymbol{X} = \boldsymbol{X}_0 + \boldsymbol{N}$ from (4.1), we can express $\boldsymbol{X}^{(\mathrm{fba})}$ as

$$\boldsymbol{X}^{(\mathrm{fba})} = \begin{bmatrix}\boldsymbol{X}_0 & \boldsymbol{\Pi}_M\cdot\boldsymbol{X}_0^*\cdot\boldsymbol{\Pi}_N\end{bmatrix} + \begin{bmatrix}\boldsymbol{N} & \boldsymbol{\Pi}_M\cdot\boldsymbol{N}^*\cdot\boldsymbol{\Pi}_N\end{bmatrix} = \boldsymbol{X}_0^{(\mathrm{fba})} + \boldsymbol{N}^{(\mathrm{fba})}. \tag{4.24}$$

The latter expression shows that $\boldsymbol{X}^{(\mathrm{fba})}$ can be written as the superposition of the noise-free FBA-processed measurement matrix $\boldsymbol{X}_0^{(\mathrm{fba})} \in \mathbb{C}^{M\times 2N}$ and a small additive noise component $\boldsymbol{N}^{(\mathrm{fba})} \in$

$\mathbb{C}^{M\times 2N}$. Hence, $\boldsymbol{X}^{(\mathrm{fba})}$ is modeled in the same way as (4.1) and does not violate the assumption of a small noise perturbation. Consequently, the performance analysis framework from [LLV93] and [RHD14] discussed in Section 4.3.1 is directly applicable to (4.24) and we only need to replace the noise-free subspaces of $\boldsymbol{X}_0$ in (4.11) by the corresponding subspaces of $\boldsymbol{X}_0^{(\mathrm{fba})}$. Thus, for the parameter estimation error of R-D Standard ESPRIT with FBA, we immediately obtain the first-order approximation [RHD14]

$$\Delta\mu_i^{(r)} \approx \mathrm{Im}\left\{\boldsymbol{p}_i^{(\mathrm{fba})^{\mathrm{T}}} \cdot \left(\tilde{\boldsymbol{J}}_1^{(r)} \cdot \boldsymbol{U}_{\mathrm{s}}^{(\mathrm{fba})}\right)^{+} \cdot \left[\tilde{\boldsymbol{J}}_2^{(r)}/\lambda_i^{(r)} - \tilde{\boldsymbol{J}}_1^{(r)}\right] \cdot \Delta\boldsymbol{U}_{\mathrm{s}}^{(\mathrm{fba})} \cdot \boldsymbol{q}_i^{(\mathrm{fba})}\right\} \tag{4.25}$$

where the signal subspace estimation error $\Delta\boldsymbol{U}_{\mathrm{s}}^{(\mathrm{fba})} \in \mathbb{C}^{M\times d}$ is given by

$$\Delta\boldsymbol{U}_{\mathrm{s}}^{(\mathrm{fba})} \approx \boldsymbol{U}_{\mathrm{n}}^{(\mathrm{fba})} \cdot \boldsymbol{U}_{\mathrm{n}}^{(\mathrm{fba})^{\mathrm{H}}} \cdot \boldsymbol{N}^{(\mathrm{fba})} \cdot \boldsymbol{V}_{\mathrm{s}}^{(\mathrm{fba})} \cdot \boldsymbol{\Sigma}_{\mathrm{s}}^{(\mathrm{fba})^{-1}} \tag{4.26}$$

and the matrices $\boldsymbol{U}_{\mathrm{s}}^{(\mathrm{fba})} \in \mathbb{C}^{M\times d}$, $\boldsymbol{U}_{\mathrm{n}}^{(\mathrm{fba})} \in \mathbb{C}^{M\times(M-d)}$, $\boldsymbol{V}_{\mathrm{s}}^{(\mathrm{fba})} \in \mathbb{C}^{2N\times d}$, and $\boldsymbol{\Sigma}_{\mathrm{s}}^{(\mathrm{fba})} \in \mathbb{R}^{d\times d}$ denote the signal subspace, the noise subspace, the row space, and the diagonal matrix of the singular values of $\boldsymbol{X}_0^{(\mathrm{fba})}$, respectively. Moreover, $\boldsymbol{q}_i$ and $\boldsymbol{p}_i$ are replaced by the corresponding versions $\boldsymbol{q}_i^{(\mathrm{fba})}$ and $\boldsymbol{p}_i^{(\mathrm{fba})}$ obtained from the eigendecomposition $\boldsymbol{\Psi}^{(r)} = \boldsymbol{Q}^{(\mathrm{fba})} \cdot \boldsymbol{\Lambda}^{(r)} \cdot \boldsymbol{Q}^{(\mathrm{fba})^{-1}}$, which results from the shift invariance equation $\tilde{\boldsymbol{J}}_1^{(r)} \cdot \boldsymbol{U}_{\mathrm{s}}^{(\mathrm{fba})} \cdot \boldsymbol{\Phi}^{(r)} = \tilde{\boldsymbol{J}}_2^{(r)} \cdot \boldsymbol{U}_{\mathrm{s}}^{(\mathrm{fba})}$.

In the second step of R-D Unitary ESPRIT, the complex-valued augmented measurement matrix $\boldsymbol{X}_0^{(\mathrm{fba})}$ is transformed into the real-valued domain. However, it was argued in [RHD14] and proven in Appendix D.13 of [Roe13] that asymptotically in the effective SNR, this step has no impact on the performance. Thus, the real-valued transformation can be ignored and only the FBA preprocessing step needs to be taken into account for the asymptotic performance of Unitary-ESPRIT-type algorithms. Therefore, the analytical expression for the parameter estimation error for R-D Standard ESPRIT with FBA in (4.25) is also valid for R-D Unitary ESPRIT.

In order to derive the analytical MSE expression for R-D Unitary ESPRIT, it is apparent that FBA does not the violate the assumptions of a zero-mean noise component with finite SO moments. FBA does, however, affect the SO moments of the physical noise $\boldsymbol{N}$. Let $\boldsymbol{n}^{(\mathrm{fba})} = \mathrm{vec}\left\{\boldsymbol{N}^{(\mathrm{fba})}\right\}$ be the vectorized FBA-processed noise component with the covariance matrix $\boldsymbol{R}_{\mathrm{nn}}^{(\mathrm{fba})} = \mathbb{E}\left\{\boldsymbol{n}^{(\mathrm{fba})} \cdot \boldsymbol{n}^{(\mathrm{fba})^{\mathrm{H}}}\right\} \in \mathbb{C}^{2MN\times 2MN}$ and the pseudo-covariance matrix $\boldsymbol{C}_{\mathrm{nn}}^{(\mathrm{fba})} = \mathbb{E}\left\{\boldsymbol{n}^{(\mathrm{fba})} \cdot \boldsymbol{n}^{(\mathrm{fba})^{\mathrm{T}}}\right\} \in \mathbb{C}^{2MN\times 2MN}$. Then, it was shown in [RHD14] that for R-D Unitary ESPRIT, the MSE of the i-th spatial frequency in the r-th mode is given by

$$\mathbb{E}\left\{\left(\Delta\mu_i^{(r)}\right)^2\right\} \approx \frac{1}{2} \cdot \left(\boldsymbol{z}_i^{(r)^{(\mathrm{fba})^{\mathrm{H}}}} \cdot \boldsymbol{R}_{\mathrm{nn}}^{(\mathrm{fba})^{\mathrm{T}}} \cdot \boldsymbol{z}_i^{(r)^{(\mathrm{fba})}} - \mathrm{Re}\left\{\boldsymbol{z}_i^{(r)^{(\mathrm{fba})^{\mathrm{T}}}} \cdot \boldsymbol{C}_{\mathrm{nn}}^{(\mathrm{fba})^{\mathrm{T}}} \cdot \boldsymbol{z}_i^{(r)^{(\mathrm{fba})}}\right\}\right), \tag{4.27}$$

where $\boldsymbol{z}_i^{(r)(\mathrm{fba})} = \boldsymbol{W}_{\mathrm{mat}}^{(\mathrm{fba})^{\mathrm{T}}} \cdot \boldsymbol{r}_i^{(r)(\mathrm{fba})} \in \mathbb{C}^{2MN\times 1}$, and $\boldsymbol{r}_i^{(r)(\mathrm{fba})}$ and $\boldsymbol{W}_{\mathrm{mat}}^{(\mathrm{fba})}$ are computed as

$$\boldsymbol{r}_i^{(r)(\mathrm{fba})} = \boldsymbol{q}_i^{(\mathrm{fba})} \otimes \left(\left[\left(\tilde{\boldsymbol{J}}_1^{(r)} \boldsymbol{U}_{\mathrm{s}}^{(\mathrm{fba})} \right)^{+} \cdot \left(\tilde{\boldsymbol{J}}_2^{(r)} / \mathrm{e}^{\mathrm{j}\cdot\mu_i^{(r)}} - \tilde{\boldsymbol{J}}_1^{(r)} \right) \right]^{\mathrm{T}} \cdot \boldsymbol{p}_i^{(\mathrm{fba})} \right) \in \mathbb{C}^{Md\times 1} \tag{4.28}$$

$$\boldsymbol{W}_{\mathrm{mat}}^{(\mathrm{fba})} = \left(\boldsymbol{\Sigma}_{\mathrm{s}}^{(\mathrm{fba})^{-1}} \cdot \boldsymbol{V}_{\mathrm{s}}^{(\mathrm{fba})^{\mathrm{T}}} \right) \otimes \left(\boldsymbol{U}_{\mathrm{n}}^{(\mathrm{fba})} \cdot \boldsymbol{U}_{\mathrm{n}}^{(\mathrm{fba})^{\mathrm{H}}} \right) \in \mathbb{C}^{Md\times 2MN} \tag{4.29}$$

by consistently replacing all quantities in (4.14) and (4.15) by their FBA-processed equivalents.

The covariance matrix $\boldsymbol{R}_{\mathrm{nn}}^{(\mathrm{fba})}$ and the pseudo-covariance matrix $\boldsymbol{C}_{\mathrm{nn}}^{(\mathrm{fba})}$ of $\boldsymbol{n}^{(\mathrm{fba})}$ have been shown in [RHD14] to be equal to

$$\boldsymbol{R}_{\mathrm{nn}}^{(\mathrm{fba})} = \begin{bmatrix} \boldsymbol{R}_{\mathrm{nn}} & \boldsymbol{C}_{\mathrm{nn}} \cdot \boldsymbol{\Pi}_{MN} \\ \boldsymbol{\Pi}_{MN} \cdot \boldsymbol{C}_{\mathrm{nn}}^{*} & \boldsymbol{\Pi}_{MN} \cdot \boldsymbol{R}_{\mathrm{nn}}^{*} \cdot \boldsymbol{\Pi}_{MN} \end{bmatrix} \in \mathbb{C}^{2MN\times 2MN} \tag{4.30}$$

$$\boldsymbol{C}_{\mathrm{nn}}^{(\mathrm{fba})} = \begin{bmatrix} \boldsymbol{C}_{\mathrm{nn}} & \boldsymbol{R}_{\mathrm{nn}} \cdot \boldsymbol{\Pi}_{MN} \\ \boldsymbol{\Pi}_{MN} \cdot \boldsymbol{R}_{\mathrm{nn}}^{*} & \boldsymbol{\Pi}_{MN} \cdot \boldsymbol{C}_{\mathrm{nn}}^{*} \cdot \boldsymbol{\Pi}_{MN} \end{bmatrix} \in \mathbb{C}^{2MN\times 2MN}, \tag{4.31}$$

where $\boldsymbol{R}_{\mathrm{nn}} = \mathbb{E}\left\{ \boldsymbol{n} \cdot \boldsymbol{n}^{\mathrm{H}} \right\} \in \mathbb{C}^{MN\times MN}$ and $\boldsymbol{C}_{\mathrm{nn}} = \mathbb{E}\left\{ \boldsymbol{n} \cdot \boldsymbol{n}^{\mathrm{T}} \right\} \in \mathbb{C}^{MN\times MN}$ are the covariance matrix and the pseudo-covariance matrix of the physical noise contribution $\boldsymbol{n}$. Therefore, the SO moments of the FBA-processed noise $\boldsymbol{n}^{(\mathrm{fba})}$ can be written in terms of the physical noise $\boldsymbol{n}$.

In the special case of circularly symmetric white noise with $\boldsymbol{R}_{\mathrm{nn}} = \sigma_{\mathrm{n}}^2 \cdot \boldsymbol{I}_{MN}$ and $\boldsymbol{C}_{\mathrm{nn}} = \boldsymbol{0}_{MN\times MN}$, (4.30) and (4.31) reduce to $\boldsymbol{R}_{\mathrm{nn}}^{(\mathrm{fba})} = \sigma_{\mathrm{n}}^2 \cdot \boldsymbol{I}_{2MN}$ and $\boldsymbol{C}_{\mathrm{nn}}^{(\mathrm{fba})} = \sigma_{\mathrm{n}}^2 \cdot \boldsymbol{\Pi}_{2MN}$. Consequently, the MSE of R-D Unitary ESPRIT in (4.27) for the i-th spatial frequency in the r-th mode simplifies to

$$\mathbb{E}\left\{ \left(\Delta\mu_i^{(r)} \right)^2 \right\} \approx \frac{\sigma_{\mathrm{n}}^2}{2} \cdot \left(\left\| \boldsymbol{z}_i^{(r)(\mathrm{fba})} \right\|_2^2 - \mathrm{Re}\left\{ \boldsymbol{z}_i^{(r)(\mathrm{fba})^{\mathrm{T}}} \cdot \boldsymbol{\Pi}_{2MN} \cdot \boldsymbol{z}_i^{(r)(\mathrm{fba})} \right\} \right). \tag{4.32}$$

4.4. Performance of R-D tensor-based ESPRIT-type algorithms

In this section, we briefly discuss the first-order performance analysis framework for R-D tensor-based ESPRIT-type algorithms using LS from [RHD14]. We first consider the R-D Standard Tensor-ESPRIT algorithm in Section 4.4.1 and then resort to the R-D Unitary Tensor-ESPRIT algorithm in Section 4.4.2.

4.4.1. Performance of R-D Standard Tensor-ESPRIT

We have seen in Section 3.5 that R-D Standard Tensor-ESPRIT only differs from R-D Standard ESPRIT in the improved HOSVD-based subspace estimate. Since the analytical expression for the parameter estimation error of R-D Standard ESPRIT in (4.17) is explicit in the subspace estimation error $\Delta\boldsymbol{U}_{\mathrm{s}}$, it is concluded in [RHD14] that for the analytical parameter estimation error of R-D Standard Tensor-ESPRIT algorithm, the SVD-based subspace error $\Delta\boldsymbol{U}_{\mathrm{s}}$ in (4.17)

can be replaced by the HOSVD-based subspace error $\left[\Delta\hat{\boldsymbol{\mathcal{U}}}^{[\mathrm{s}]}\right]^{\mathrm{T}}_{(R+1)}$. Therefore, we immediately arrive at the following first-order approximation of the parameter estimation for R-D Standard Tensor-ESPRIT:

$$\Delta\mu_i^{(r)} \approx \operatorname{Im}\left\{\boldsymbol{p}_i^{\mathrm{T}} \cdot \left(\tilde{\boldsymbol{J}}_1^{(r)} \cdot \boldsymbol{U}_{\mathrm{s}}\right)^{+} \cdot \left[\tilde{\boldsymbol{J}}_2^{(r)}/\lambda_i^{(r)} - \tilde{\boldsymbol{J}}_1^{(r)}\right] \cdot \left[\Delta\hat{\boldsymbol{\mathcal{U}}}^{[\mathrm{s}]}\right]^{\mathrm{T}}_{(R+1)} \cdot \boldsymbol{q}_i\right\}. \tag{4.33}$$

Note that by inserting the analytical expression for $\left[\Delta\hat{\boldsymbol{\mathcal{U}}}^{[\mathrm{s}]}\right]^{\mathrm{T}}_{(R+1)}$ from (4.10), $\Delta\mu_i^{(r)}$ in (4.33) can be expressed explicitly in terms of the noise tensor $\boldsymbol{\mathcal{N}}$.

Then, as derived in [RHD14], the analytical MSE expression of R-D Standard Tensor-ESPRIT for the i-th spatial frequency in the r-th mode is given by

$$\mathbb{E}\left\{\left(\Delta\mu_i^{(r)}\right)^2\right\} \approx \frac{1}{2} \cdot \left(\boldsymbol{z}_i^{(r)\mathrm{H}} \cdot \boldsymbol{R}_{\mathrm{nn}}^{\mathrm{T}} \cdot \boldsymbol{z}_i^{(r)} - \operatorname{Re}\left\{\boldsymbol{z}_i^{(r)^{\mathrm{T}}} \cdot \boldsymbol{C}_{\mathrm{nn}}^{\mathrm{T}} \cdot \boldsymbol{z}_i^{(r)}\right\}\right), \tag{4.34}$$

where $\boldsymbol{z}_i^{(r)} = \boldsymbol{W}_{\mathrm{ten}}^{\mathrm{T}} \cdot \boldsymbol{r}_i^{(r)} \in \mathbb{C}^{MN\times 1}$ for $i = 1, 2, \ldots, d$ and $r = 1, 2, \ldots, R$. The vector $\boldsymbol{r}_i^{(r)} \in \mathbb{C}^{Md\times 1}$ is already given in (4.14) and the matrix $\boldsymbol{W}_{\mathrm{ten}}$ can be expressed as

$$\boldsymbol{W}_{\mathrm{ten}} = \boldsymbol{W}_0 + \sum_{r=1}^{R} \boldsymbol{W}_r \cdot \boldsymbol{P}^{(r)^{\mathrm{T}}}_{M_1,\ldots,M_R,N} \cdot \boldsymbol{P}^{(r)}_{M_1,\ldots,M_R,N} \in \mathbb{C}^{Md\times MN},$$

with the definitions

$$\begin{aligned}
\boldsymbol{W}_0 &= \left(\boldsymbol{\Sigma}_{\mathrm{s}}^{-1} \cdot \boldsymbol{V}_{\mathrm{s}}^{\mathrm{T}}\right) \otimes \left(\boldsymbol{T}^{\otimes}_{1:R} \cdot \boldsymbol{U}_{\mathrm{n}} \cdot \boldsymbol{U}_{\mathrm{n}}^{\mathrm{H}}\right) \\
\boldsymbol{W}_r &= \left(\boldsymbol{U}_{\mathrm{s}}^{\mathrm{T}} \otimes \boldsymbol{I}_M\right) \cdot \left(\bar{\boldsymbol{T}}_{1:r-1} \otimes \boldsymbol{I}_{M_{r:R}}\right) \cdot \left(\boldsymbol{I}_{M_r} \otimes \bar{\boldsymbol{T}}_{r+1:R}\right) && (4.35) \\
&\quad \cdot \left[\left(\boldsymbol{V}_r^{[\mathrm{s}]} \cdot \boldsymbol{\Sigma}_r^{[\mathrm{s}]^{-1}} \cdot \boldsymbol{U}_r^{[\mathrm{s}]^{\mathrm{H}}}\right)^{\mathrm{T}} \otimes \left(\boldsymbol{U}_r^{[\mathrm{n}]} \cdot \boldsymbol{U}_r^{[\mathrm{n}]^{\mathrm{H}}}\right)\right]. && (4.36)
\end{aligned}$$

Moreover, the matrices $\bar{\boldsymbol{T}}_{1:r-1}$ and $\bar{\boldsymbol{T}}_{r+1:R}$ are given by

$$\bar{\boldsymbol{T}}_{1:r-1} = \begin{bmatrix} \boldsymbol{I}_{M_{r:R}} \otimes \boldsymbol{t}_{1:r-1,1} \\ \vdots \\ \boldsymbol{I}_{M_{r:R}} \otimes \boldsymbol{t}_{1:r-1,M_{1:r-1}} \end{bmatrix}, \quad \bar{\boldsymbol{T}}_{r+1:R} = \begin{bmatrix} \boldsymbol{I}_{M_r} \otimes \boldsymbol{t}_{r+1:R,1} \\ \vdots \\ \boldsymbol{I}_{M_r} \otimes \boldsymbol{t}_{r+1:R,M_{r+1:R}} \end{bmatrix}, \tag{4.37}$$

where we have introduced the short-hand notation $\boldsymbol{t}_{a:b,n}$ as the n-th column of $\boldsymbol{T}^{\otimes}_{a:b}$ defined in (1.24) and $M_{a:b}$ is defined in (1.25). Note that (4.35) was only stated in [RHD14] and not proven. The proof of (4.35) is given in Appendix B.4.

It can be seen from (4.34) that compared to the analytical MSE expression for R-D Standard ESPRIT, the matrix $\boldsymbol{W}_{\mathrm{mat}}$ only needed to be replaced by $\boldsymbol{W}_{\mathrm{ten}}$ to take into account the enhanced HOSVD-based signal subspace estimate.

Similarly to (4.18), for the special case of circularly symmetric white noise with $\boldsymbol{R}_{\mathrm{nn}} = \sigma_{\mathrm{n}}^2 \cdot \boldsymbol{I}_{MN}$ and $\boldsymbol{C}_{\mathrm{nn}} = \boldsymbol{0}_{MN \times MN}$, the MSE in (4.34) for the i-th spatial frequency in the r-th mode simplifies to

$$\mathbb{E}\left\{\left(\Delta\mu_i^{(r)}\right)^2\right\} \approx \frac{\sigma_{\mathrm{n}}^2}{2} \cdot \left\|\boldsymbol{z}_i^{(r)}\right\|_2^2, \tag{4.38}$$

where $\boldsymbol{z}_i^{(r)} = \boldsymbol{W}_{\mathrm{ten}}^{\mathrm{T}} \cdot \boldsymbol{r}_i^{(r)}$ and $\boldsymbol{r}_i^{(r)} \in \mathbb{C}^{MN \times 1}$ and the matrix $\boldsymbol{W}_{\mathrm{ten}}$ are given by (4.14) and (4.35).

4.4.2. Performance of R-D Unitary Tensor-ESPRIT

In Section 4.3.2, we have seen that for the asymptotic performance of Unitary-ESPRIT-type algorithms, only forward-backward averaging needs to be taken into account and the real-valued transformation can be neglected. Thus, parameter estimation error expression for R-D Unitary Tensor-ESPRIT is directly obtained by consistently replacing $\left[\Delta\hat{\boldsymbol{\mathcal{U}}}^{[\mathrm{s}]}\right]_{(R+1)}^{\mathrm{T}}$ in (4.33) by its corresponding FBA-processed version $\left[\Delta\hat{\boldsymbol{\mathcal{U}}}^{[\mathrm{s}](\mathrm{fba})}\right]_{(R+1)}^{\mathrm{T}}$, i.e., we get

$$\Delta\mu_i^{(r)} \approx \mathrm{Im}\left\{\boldsymbol{p}_i^{(\mathrm{fba})^{\mathrm{T}}} \cdot \left(\tilde{\boldsymbol{J}}_1^{(r)} \cdot \boldsymbol{U}_{\mathrm{s}}^{(\mathrm{fba})}\right)^{+} \cdot \left[\tilde{\boldsymbol{J}}_2^{(r)}/\lambda_i^{(r)} - \tilde{\boldsymbol{J}}_1^{(r)}\right] \cdot \left[\Delta\hat{\boldsymbol{\mathcal{U}}}^{[\mathrm{s}](\mathrm{fba})}\right]_{(R+1)}^{\mathrm{T}} \cdot \boldsymbol{q}_i^{(\mathrm{fba})}\right\}, \tag{4.39}$$

where the HOSVD-based signal subspace estimation error $\left[\Delta\hat{\boldsymbol{\mathcal{U}}}^{[\mathrm{s}](\mathrm{fba})}\right]_{(R+1)}^{\mathrm{T}}$ is computed via (4.10) from the augmented measurement tensor $\boldsymbol{\mathcal{X}}_0^{(\mathrm{fba})}$ given in (3.20) in Section 3.2.3.

Following the reasoning in the previous section, the MSE of R-D Unitary Tensor-ESPRIT for the i-th spatial frequency in the r-th mode is given by

$$\mathbb{E}\left\{\left(\Delta\mu_i^{(r)}\right)^2\right\} \approx \frac{1}{2} \cdot \left(\boldsymbol{z}_i^{(r)^{(\mathrm{fba})^{\mathrm{H}}}} \cdot \boldsymbol{R}_{\mathrm{nn}}^{(\mathrm{fba})^{\mathrm{T}}} \cdot \boldsymbol{z}_i^{(r)^{(\mathrm{fba})}} - \mathrm{Re}\left\{\boldsymbol{z}_i^{(r)^{(\mathrm{fba})^{\mathrm{T}}}} \cdot \boldsymbol{C}_{\mathrm{nn}}^{(\mathrm{fba})^{\mathrm{T}}} \cdot \boldsymbol{z}_i^{(r)^{(\mathrm{fba})}}\right\}\right), \tag{4.40}$$

where $\boldsymbol{z}_i^{(r)^{(\mathrm{fba})}} = \boldsymbol{W}_{\mathrm{ten}}^{(\mathrm{fba})^{\mathrm{T}}} \cdot \boldsymbol{r}_i^{(r)^{(\mathrm{fba})}}$. The vector $\boldsymbol{r}_i^{(r)^{(\mathrm{fba})}}$ is given in (4.28) and $\boldsymbol{W}_{\mathrm{ten}}^{(\mathrm{fba})}$ is obtained by replacing the quantities in (4.35) by their FBA-processed equivalents.

Again, in the special case of circularly symmetric white noise considered in (4.32), we obtain $\boldsymbol{R}_{\mathrm{nn}}^{(\mathrm{fba})} = \sigma_{\mathrm{n}}^2 \cdot \boldsymbol{I}_{2MN}$ and $\boldsymbol{C}_{\mathrm{nn}}^{(\mathrm{fba})} = \sigma_{\mathrm{n}}^2 \cdot \boldsymbol{\Pi}_{2MN}$ such that the MSE of R-D Unitary Tensor-ESPRIT in (4.40) for the i-th spatial frequency in the r-th mode simplifies to

$$\mathbb{E}\left\{\left(\Delta\mu_i^{(r)}\right)^2\right\} \approx \frac{\sigma_{\mathrm{n}}^2}{2} \cdot \left(\left\|\boldsymbol{z}_i^{(r)^{(\mathrm{fba})}}\right\|_2^2 - \mathrm{Re}\left\{\boldsymbol{z}_i^{(r)^{(\mathrm{fba})^{\mathrm{T}}}} \cdot \boldsymbol{\Pi}_{2MN} \cdot \boldsymbol{z}_i^{(r)^{(\mathrm{fba})}}\right\}\right), \tag{4.41}$$

where $\boldsymbol{z}_i^{(r)^{(\mathrm{fba})}} = \boldsymbol{W}_{\mathrm{mat}}^{(\mathrm{fba})^{\mathrm{T}}} \cdot \boldsymbol{r}_i^{(r)^{(\mathrm{fba})}} \in \mathbb{C}^{2MN \times 1}$, and $\boldsymbol{r}_i^{(r)^{(\mathrm{fba})}}$ is given in (4.40).

4.5. Special source cases

The MSE expressions for both matrix-based R-D ESPRIT-type algorithms and tensor-based R-D ESPRIT-type algorithms reviewed in Section 4.3 and Section 4.4 are deterministic and do no require Monte-Carlo simulations. Therefore, the MSE can be plotted as a function of the varying system parameters to analyze their effect on the estimation accuracy via simulations. However, the disadvantage of the MSE expressions is that they are formulated in terms of the subspaces of the unperturbed measurement matrix and hence, provide no explicit insights into the influence of the physical parameters, e.g., the SNR, the number of sensors M, the sample size N, the correlation of the signals, and their source separation. Knowing how the performance scales with these system parameters can facilitate array design decisions on the number of required sensors to achieve a certain performance for a specific SNR. Moreover, different parameter estimators can be objectively compared to find the best estimator for particular scenarios. Establishing a general formulation for an arbitrary number of sources is an intricate task given the complex dependence of the subspaces on the physical parameters. However, special cases for the number of sources can be considered to gain more insights by such an analytical performance assessment.

In [RH12], simplified MSE expressions of the matrix-based and tensor-based R-D ESPRIT-type algorithms for a single source ($d = 1$) have been derived under the assumption of a R-D uniform sampling grid, e.g., a ULA for $R = 1$ or a URA for $R = 2$. It has been shown that for a single source, no performance improvement from forward-backward averaging (FBA) and tensor-based processing can be achieved. This motivates us to simplify the analytical MSE expressions of the R-D ESPRIT-type algorithms for two sources ($d = 2$) in order to obtain analytical expressions for the gain from FBA and tensor processing in this case. These closed-form expressions can then be analyzed in terms of the correlation and the source separation of the two signals to identify the scenarios for which the maximum gain or no gain is achieved.

In Section 4.5.1, we first review the simplified MSE expressions of the matrix-based and tensor-based R-D ESPRIT-type algorithms for $d = 1$ from [RH12] in Section 4.5.1. Then, in Section 4.5.2, assuming an R-D uniform sampling grid, we extend these results to the special case of $d = 2$.

4.5.1. Single source case

The special case of a single source for the analytical MSE expressions of the matrix-based and tensor-based R-D ESPRIT-type algorithms using LS has already been considered in [RBH10, RH12, Roe13]. It has been found that assuming an M-element uniform R-D array with an M_r-element ULA in the r-th mode, circularly symmetric white noise, and $d = 1$, the MSE of the spatial frequency for R-D Standard ESPRIT, R-D Unitary ESPRIT, R-D Standard Tensor-ESPRIT, and

R-D Unitary Tensor-ESPRIT is given by

$$\mathbb{E}\left\{\left(\Delta\mu^{(r)}\right)^2\right\} \approx \frac{1}{\hat{\varrho}} \cdot \frac{M_r}{M \cdot (M_r - 1)^2}, \tag{4.42}$$

where $\hat{\varrho}$ represents the effective SNR $\hat{\varrho} = N\hat{P}/\sigma_{\mathrm{n}}^2$ with $\hat{P}$ being the empirical source power given by $\hat{P} = \|\boldsymbol{s}\|_2^2/N$ and $\boldsymbol{s} \in \mathbb{C}^{N\times 1}$. This result shows that for a single source, there is neither an improvement in terms of the estimation accuracy from forward-backward averaging nor from the HOSVD-based subspace estimate. The latter is indeed surprising as the HOSVD-based subspace estimate itself is more accurate than the SVD-based subspace estimate for a single source.

In order to analytically compute the asymptotic efficiency, i.e., the ratio of the CRB and MSE, of the considered R-D ESPRIT-type algorithms for a single source, we require the deterministic R-D Cramér-Rao bound for a single source [RH12], which is computed in (9.39). The expression is given by

$$\boldsymbol{C} = \operatorname{diag}\left\{\left[C^{(1)}, \ldots, C^{(R)}\right]\right\} \in \mathbb{R}^{R\times R} \tag{4.43}$$

where

$$C^{(r)} = \frac{1}{\hat{\varrho}} \cdot \frac{6}{M \cdot (M_r^2 - 1)}. \tag{4.44}$$

Using (4.42) and (4.44), we can determine the asymptotic efficiency of the R-D ESPRIT-type algorithms for an arbitrary number of dimensions R. The result for $R = 1$ is given in [RH12] by

$$\eta = \lim_{\hat{\varrho}\to\infty} \frac{\mathrm{CRB}}{\mathbb{E}\{(\Delta\mu)^2\}} = \frac{6(M-1)}{M(M+1)}. \tag{4.45}$$

It should be noted that η is only a function of the array geometry, i.e., the number of sensors M. From (4.45), it is apparent that 1-D ESPRIT-type algorithms using LS are asymptotically efficient, i.e., $\eta = 1$, for $M = 2$ and $M = 3$ for a single source. However, they become less efficient when the number of sensors grows, in fact, for $M \to \infty$ we have $\eta \to 0$. Note that the same behavior can be observed for $R = 2$, where $\eta = 1$ for $M_1 \in [2, 3]$ and $M_2 \in [2, 3]$ as shown in [RH12]. In Chapter 5, we derive a new LS solution, termed generalized least squares (GLS), for matrix-based R-D ESPRIT-type algorithms, which is asymptotically efficient, i.e., $\eta = 1$, for an arbitrary number of sensors M in the single source case. In fact, we show in Section 5.5.1 that the analytical MSE expression of GLS-based 1-D ESPRIT-type algorithms for a single source can be reformulated into the corresponding deterministic CRB expression.

4.5.2. Two source case

In the previous section, we have seen that no gain from FBA and tensor-based processing can be achieved using matrix-based and tensor-based R-D ESPRIT-type algorithms for single source. Therefore, in this section, we simplify the analytical MSE expressions of the reviewed R-D ESPRIT-type algorithms for the special case of two sources. Based on these expressions, we compute the FBA gain as well as the tensor gain for two sources and investigate their behavior with respect to the physical parameters of interest. The derived MSE expressions will depend on the following parameters:

1. The number of sensors M_r in the r-th mode for $r = 1, \ldots, R$.

2. The empirical signal powers $\hat{P}_i = \frac{1}{N} \cdot \|s_i\|_2^2$ for $i = 1, 2$.

3. The spatial correlation of the two sources, which is reflected by various versions of the inner product of the array steering vectors $\boldsymbol{a}_i$ contained in $\boldsymbol{A}_\mathrm{c} = [\boldsymbol{a}_1, \boldsymbol{a}_2] \in \mathbb{C}^{M \times 2}$, where $M = \prod_{r=1}^{R} M_r$, $\boldsymbol{A}_\mathrm{c}$ has a centered phase reference and $\|\boldsymbol{a}_i\|_2^2 = M$, as

$$\alpha = \boldsymbol{a}_1^\mathrm{H} \cdot \boldsymbol{a}_2 \tag{4.46}$$

$$\alpha^{(r)} = \boldsymbol{a}_1^{(r)^\mathrm{H}} \cdot \boldsymbol{a}_2^{(r)} \tag{4.47}$$

$$\alpha_\mathrm{sel}^{(r)} = \boldsymbol{a}_1^{(r)^\mathrm{H}} \cdot \boldsymbol{J}_1^{(r)^\mathrm{H}} \cdot \boldsymbol{J}_1^{(r)} \cdot \boldsymbol{a}_2^{(r)} \tag{4.48}$$

$$\tilde{\alpha}^{(r)} = \boldsymbol{a}_1^\mathrm{H} \cdot \tilde{\boldsymbol{J}}_1^{(r)^\mathrm{H}} \cdot \tilde{\boldsymbol{J}}_1^{(r)} \cdot \boldsymbol{a}_2 = \frac{\alpha}{\alpha^{(r)}} \cdot \alpha_\mathrm{sel}^{(r)} \tag{4.49}$$

$$\alpha_{\mathrm{sel},0}^{(r)} = \boldsymbol{a}_{\mathrm{1st},1}^{(r)^\mathrm{H}} \cdot \boldsymbol{J}_1^{(r)^\mathrm{H}} \cdot \boldsymbol{J}_1^{(r)} \cdot \boldsymbol{a}_{\mathrm{1st},2}^{(r)} = \sum_{m_r=0}^{M_r-2} \mathrm{e}^{\mathrm{j} m_r \Delta\mu^{(r)}}, \tag{4.50}$$

where $\Delta\mu^{(r)} = \left|\mu_2^{(r)} - \mu_1^{(r)}\right|$ represents the spatial separation of the sources in the r-th mode, $\boldsymbol{a}_{\mathrm{1st},i}^{(r)} = [1, \mathrm{e}^{\mathrm{j}\mu_i^{(r)}}, \ldots, \mathrm{e}^{\mathrm{j}(M-1)\mu_i^{(r)}}]^\mathrm{T}$ with the phase reference being at the first element, and the selection matrices $\boldsymbol{J}_1^{(r)}$ and $\tilde{\boldsymbol{J}}_1^{(r)}$ are given in (3.27) and (3.26), respectively.

4. The complex-valued empirical correlation $\hat{\rho}$ of the two sources, which is defined by

$$\hat{\rho} = \frac{1}{N} \cdot \frac{\boldsymbol{s}_1^\mathrm{H} \cdot \boldsymbol{s}_2}{\sqrt{\hat{P}_1 \cdot \hat{P}_2}} = |\hat{\rho}| \cdot \mathrm{e}^{\mathrm{j}\Delta\varphi_\mathrm{corr}}, \tag{4.51}$$

where $\Delta\varphi_\mathrm{corr}$ is the correlation phase between the two sources.

5. The parameter for the array phase reference $\delta^{(r)}$ as introduced in Section 2.1.4.2. In the simplified expressions, the array phase reference occurs in conjunction with the correlation

phase of the source symbols (cf. Equation 2.28). Therefore, we obtain

$$\Delta\varphi = \Delta\varphi_{\text{ref}} + \Delta\varphi_{\text{corr}}. \tag{4.52}$$

where $\Delta\varphi_{\text{ref}} = \sum_{r=1}^{R} \delta^{(r)} \cdot \Delta\mu^{(r)}$.

Note that for the special case of two closely spaced sources, i.e., for $\Delta\mu^{(r)} \to 0$, we can further simplify the obtained MSE expressions by applying a first-order Taylor approximation of the spatial correlation terms. Thereby, we can reduce the dependence on the spatial source correlation to a function of the number of sensors M_r.

We first consider the matrix-based versions in Section 4.5.3 before we proceed to the tensor-based R-D ESPRIT-type algorithms in Section 4.5.4. An analysis of the results is provided in Section 4.5.5.

4.5.3. Two source case for matrix-based R-D ESPRIT-type algorithms

The resulting MSE expressions of R-D Standard ESPRIT and R-D Unitary ESPRIT for two sources are given in Theorem 4.5.1 and in Theorem 4.5.2, respectively.

For the MSE of R-D Standard ESPRIT for two sources, we obtain the following result:

Theorem 4.5.1. *For the case of an M-element R-D uniform sampling grid with an M_r-element ULA in the r-th mode, circularly symmetric white noise, and two sources ($d = 2$), the MSE of R-D Standard ESPRIT is given by*

$$\text{MSE}_{\text{mat}} = \frac{\sigma_{\text{n}}^2}{2} \cdot \frac{\hat{P}_1 + \hat{P}_2}{N \cdot \hat{P}_1 \cdot \hat{P}_2} \cdot \sum_{r=1}^{R} a_{\text{mat}}^{(r)}, \tag{4.53}$$

where the scalar $a_{\text{mat}}^{(r)}$ is given by

$$a_{\text{mat}}^{(r)} = \frac{M}{M_r} \cdot \Bigg[\left(\frac{M}{M_r}\right)^2 \cdot (M_r - 1)^2 \cdot 2 + \left|\tilde{\alpha}^{(r)}\right|^2 \cdot \left(2 + (M_r - 2) \cdot \left|\mathrm{e}^{\mathrm{j}\Delta\mu^{(r)}} - 1\right|^2\right) - 4 \cdot (M_r - 1) \cdot \frac{|\alpha|^2}{\left|\alpha^{(r)}\right|^2} \cdot \text{Re}\left\{\alpha_{\text{sel},0}^{(r)}\right\} \Bigg] \cdot \frac{1}{D_{\text{sel}}^{(r)^2} \cdot \left(1 - |\hat{\rho}|^2\right)} \tag{4.54}$$

and $D_{\text{sel}}^{(r)}$ is given by

$$D_{\text{sel}}^{(r)} = \left(\frac{M}{M_r}\right)^2 \cdot (M_r - 1)^2 - \frac{|\alpha|^2}{\left|\alpha^{(r)}\right|^2} \cdot \left|\alpha_{\text{sel}}^{(r)}\right|^2. \tag{4.55}$$

The proof is given in Appendix B.5. Note that preliminary results for $R = 1$ are provided by us in [Gra15]. It is apparent from (4.53) that the MSE of the matrix-based R-D Standard ESPRIT algorithm is directly proportional to the term $\frac{1}{1-|\hat{\rho}|^2}$, i.e., the MSE increases as the correlation $|\hat{\rho}|$ increases. Moreover, (4.53) does not depend on the array phase reference $\Delta\varphi_{\text{ref}}$ and the correlation phase $\Delta\varphi_{\text{corr}}$. Hence, we conclude that the performance of R-D Standard ESPRIT does not depend on these parameters.

For small values of the spatial source separation $\Delta\mu^{(r)}$, the term $a_{\text{mat}}^{(r)}$ can be approximated via a first-order Taylor series expansion as [Gra15]

$$\tilde{a}_{\text{mat}}^{(r)} = \frac{1}{1-|\hat{\rho}|^2} \cdot \frac{24 \cdot \left(S_r + 3 \cdot \Delta\mu^{(r)^2} \cdot M_r \cdot (M_r - 2)\right)}{\frac{M}{M_r} \cdot (M_r - 1)^2 \cdot A_r^2}, \tag{4.56}$$

where $A_r = \Delta\mu^{(r)^2} \cdot M_r \cdot (M_r - 2) + S_r$ and $S_r = \sum\limits_{\substack{q=1 \\ q \neq r}}^{R} \Delta\mu^{(q)^2} \cdot (M_q^2 - 1)$.

For the MSE of R-D Unitary ESPRIT for two sources, we obtain the result:

Theorem 4.5.2. *For the case of an M-element R-D uniform sampling grid with an M_r-element ULA in the r-th mode, circularly symmetric white noise, and two sources ($d = 2$), the MSE of R-D Unitary ESPRIT is given by*

$$\text{MSE}_{\text{mat}}^{(\text{fba})} = \frac{\sigma_{\text{n}}^2}{2} \cdot \frac{\hat{P}_1 + \hat{P}_2}{N \cdot \hat{P}_1 \cdot \hat{P}_2} \cdot \sum_{r=1}^{R} a_{\text{mat}}^{(\text{fba})(r)}, \tag{4.57}$$

where the scalar $a_{\text{mat}}^{(\text{fba})(r)}$ is given by

$$a_{\text{mat}}^{(\text{fba})(r)} = \frac{M}{M_r} \cdot \left[\left(\frac{M}{M_r}\right)^2 \cdot (M_r - 1)^2 \cdot 2 + \left|\tilde{\alpha}^{(r)}\right|^2 \cdot \left(2 + (M_r - 2) \cdot \left|\mathrm{e}^{\mathrm{j}\Delta\mu^{(r)}} - 1\right|^2\right) - 4 \cdot (M_r - 1) \cdot \frac{|\alpha|^2}{\left|\alpha^{(r)}\right|^2} \cdot \text{Re}\left\{\alpha_{\text{sel},0}^{(r)}\right\} \right] \cdot \frac{1}{\left(1 - \cos^2(\Delta\varphi) \cdot |\hat{\rho}|^2\right) \cdot D_{\text{sel}}^{(r)^2}}.$$

The proof is given in Appendix B.6. Again, the results for $R = 1$ have been derived by us in [Gra15]. The MSE of R-D Unitary ESPRIT for two sources in (4.57) has the same form as the MSE of R-D Standard ESPRIT. They only differ in the term $\cos^2(\Delta\varphi)$, where $\Delta\varphi = \Delta\varphi_{\text{ref}} + \Delta\varphi_{\text{corr}}$, which represents the algorithms' ability to decorrelate the sources depending on how the correlation phase $\Delta\varphi_{\text{corr}}$ is related to the array phase $\Delta\varphi_{\text{ref}}$. This behavior is discussed in detail in Section 4.5.5.

If the sources are uncorrelated, i.e., $|\hat{\rho}| = 0$, we find from (4.57) that $\mathrm{MSE}_{\mathrm{mat}}^{(\mathrm{fba})}$ reduces to

$$\mathrm{MSE}_{\mathrm{mat}}^{(\mathrm{fba})}\Big|_{|\hat{\rho}|\to 0} = \mathrm{MSE}_{\mathrm{mat}}\Big|_{|\hat{\rho}|\to 0}. \tag{4.58}$$

Thus, no improvement from forward-backward averaging (FBA) can be achieved in this case.

Using a first-order Taylor series expansion for small $\Delta\mu^{(r)}$, an approximation of $a_{\mathrm{mat}}^{(\mathrm{fba})(r)}$ for small $\Delta\mu^{(r)}$ can be computed similarly to (4.56) as

$$\tilde{a}_{\mathrm{mat}}^{(\mathrm{fba})(r)} = \frac{1}{1-\cos^2(\Delta\varphi)\cdot|\hat{\rho}|^2}\cdot\frac{24\cdot\left(S_r+3\cdot\Delta\mu^{(r)^2}\cdot M_r\cdot(M_r-2)\right)}{\frac{M}{M_r}\cdot(M_r-1)^2\cdot A_r^2}, \tag{4.59}$$

where A_r and S_r are defined as in (4.56).

4.5.4. Two source case for tensor-based R-D ESPRIT-type algorithms

In this section, we simplify the analytical MSE expressions of the R-D Tensor-ESPRIT-type algorithms for two sources. The results for R-D Standard Tensor-ESPRIT and R-D Unitary Tensor-ESPRIT are provided in Theorem 4.5.3 and in Theorem 4.5.4, respectively.

For R-D Standard Tensor-ESPRIT, we can state the result:

Theorem 4.5.3. *For the case of an M-element R-D uniform sampling grid with an M_r-element ULA in the r-th mode, circularly symmetric white noise, and two sources ($d = 2$), the MSE of R-D Standard Tensor-ESPRIT is given by*

$$\mathrm{MSE}_{\mathrm{ten}} = \frac{\sigma_{\mathrm{n}}^2}{2}\cdot\frac{\hat{P}_1+\hat{P}_2}{N\cdot\hat{P}_1\cdot\hat{P}_2}\cdot\sum_{r=1}^{R}a_{\mathrm{ten}}^{(r)}, \tag{4.60}$$

where the scalar $a_{\mathrm{ten}}^{(r)}$ is given by

$$a_{\mathrm{ten}}^{(r)} = \frac{1}{D_{\mathrm{sel}}^{(r)^2}}\cdot\left(\frac{b^{(r)}}{D_r}\cdot\left[\frac{1}{1-|\hat{\rho}|^2}-\frac{\left(\frac{M}{M_r}\right)^2-\frac{|\alpha|^2}{|\alpha^{(r)}|^2}}{\left(\frac{M}{M_r}\right)^2-|\hat{\rho}|^2\cdot\left|\frac{\alpha}{\alpha^{(r)}}\right|^2}\right]+c^{(r)}\cdot\frac{\frac{M}{M_r}}{\left(\frac{M}{M_r}\right)^2-|\hat{\rho}|^2\cdot\frac{|\alpha|^2}{|\alpha^{(r)}|^2}}\right) \tag{4.61}$$

and

$$b^{(r)} = M\cdot(M_r-1)^2\cdot\left|\mathrm{e}^{\mathrm{j}\Delta\mu^{(r)}}-1\right|^2\cdot\left|\alpha_{\mathrm{sel}}^{(r)}\right|^2\cdot\left[\left(\frac{M}{M_r}\right)^2-\frac{|\alpha|^2}{|\alpha^{(r)}|^2}\right] \tag{4.62}$$

$$c^{(r)} = \left(\frac{M}{M_r}\right)^4\cdot(M_r-1)^2\cdot 2+\left|\frac{\alpha}{\alpha^{(r)}}\right|^4\cdot\left|\alpha_{\mathrm{sel}}^{(r)}\right|^2\cdot\left(2+(M_r-2)\cdot\left|\mathrm{e}^{\mathrm{j}\Delta\mu^{(r)}}-1\right|^2\right) \tag{4.63}$$

$$- 4 \cdot \left(\frac{M}{M_r}\right)^2 \cdot (M_r - 1) \cdot \left|\frac{\alpha}{\alpha^{(r)}}\right|^2 \cdot \mathrm{Re}\left\{\alpha_{\mathrm{sel},0}^{(r)}\right\}. \tag{4.64}$$

Moreover, $D_{\mathrm{sel}}^{(r)}$ *is defined in* (4.55) *and* $D_r = M_r^2 - |\alpha^{(r)}|^2$.

A sketch of the proof is provided in Appendix B.7. For the sake of brevity, some of the lengthy derivations have been omitted, however, the full proof is provided by us in [Gra16]. Note that as in the matrix-case, the MSE of R-D Standard Tensor-ESPRIT is independent of $\cos(\Delta\varphi)$, which contains the array phase reference and the correlation phase.

By applying a first-order Taylor approximation, which is valid for small $\Delta\mu^{(r)}$, $r = 1, \ldots, R$, the MSE expression given in (4.60) simplifies by replacing a_{ten} with $\tilde{a}_{\mathrm{ten}}$ given as [Gra16]

$$\tilde{a}_{\mathrm{ten}} = \frac{144}{A_r^2} \cdot \left(\frac{S_r}{M \cdot (M_r^2 - 1) \cdot \left(1 - |\hat{\rho}|^2\right)} + \frac{6 \cdot \Delta\mu^{(r)^2} \cdot M_r \cdot (M_r - 2)}{\frac{M}{M_r} \cdot (M_r - 1)^2 \cdot \left(12 + |\hat{\rho}|^2 \cdot [S_r - 12]\right)} \right). \tag{4.65}$$

For R-D Unitary Tensor-ESPRIT, we obtain the yresult:

Theorem 4.5.4. *For the case of an* M*-element* R*-D uniform sampling grid with an* M_r*-element ULA in the* r*-th mode, circularly symmetric white noise, and two sources (*$d = 2$*), the MSE of* R*-D Unitary Tensor-ESPRIT is given by*

$$\mathrm{MSE}_{\mathrm{ten}}^{(\mathrm{fba})(r)} = \frac{\sigma_{\mathrm{n}}^2}{2} \cdot \frac{\hat{P}_1 + \hat{P}_2}{N \cdot \hat{P}_1 \cdot \hat{P}_2} \cdot \sum_{r=1}^{R} a_{\mathrm{ten}}^{(\mathrm{fba})(r)} \tag{4.66}$$

where

$$a_{\mathrm{ten}}^{(\mathrm{fba})(r)} = \frac{1}{D_{\mathrm{sel}}^{(r)^2}} \cdot \left(\frac{b^{(r)}}{D_r} \cdot \left[\frac{1}{1 - \cos^2(\Delta\varphi) \cdot |\hat{\rho}|^2} - \frac{\left(\frac{M}{M_r}\right)^2 - \frac{|\alpha|^2}{|\alpha^{(r)}|^2}}{\left(\frac{M}{M_r}\right)^2 - \cos^2(\Delta\varphi) \cdot |\hat{\rho}|^2 \cdot \left|\frac{\alpha}{\alpha^{(r)}}\right|^2} \right] + c^{(r)} \cdot \frac{\frac{M}{M_r}}{\left(\frac{M}{M_r}\right)^2 - \cos^2(\Delta\varphi) \cdot |\hat{\rho}|^2 \cdot \frac{|\alpha|^2}{|\alpha^{(r)}|^2}} \right). \tag{4.67}$$

The terms $b^{(r)}$, $c^{(r)}$, *and* $D_{\mathrm{sel}}^{(r)^2}$ *are defined in* (4.62), (4.63), *and* (4.55), *respectively, and* $D_r = M_r^2 - |\alpha^{(r)}|^2$.

For a sketch of the proof, the reader is referred to Appendix B.8. Again, the full proof is provided by us in [Gra16]. As the MSE of R-D Unitary Tensor-ESPRIT depends on $\cos(\Delta\varphi)$, its performance is affected by the phase reference and the correlation phase. Similarly to the matrix case, the MSEs of R-D Standard Tensor-ESPRIT and R-D Unitary Tensor-ESPRIT have the same form and only differ in the term $\cos(\Delta\varphi)$.

By applying a first-order Taylor approximation for small $\Delta\mu^{(r)}$, $r = 1, \ldots, R$ the MSE expression given in (4.66) simplifies by replacing $a_{\text{ten}}^{(\text{fba})(r)}$ by $\tilde{a}_{\text{ten}}^{(\text{fba})(r)}$, which is given by [Gra16]

$$\tilde{a}_{\text{ten}}^{(\text{fba})(r)} = \frac{144}{A_r^2} \cdot \left(\frac{S_r}{M \cdot (M_r^2 - 1) \cdot \left(1 - \cos^2(\Delta\varphi) \cdot |\hat{\rho}|^2\right)} + \frac{6 \cdot \Delta\mu^{(r)^2} \cdot M_r \cdot (M_r - 2)}{\frac{M}{M_r} \cdot (M_r - 1)^2 \cdot \left(12 + \cos^2(\Delta\varphi) \cdot |\hat{\rho}|^2 \cdot [S_r - 12]\right)} \right).$$

4.5.5. Analysis of the results for two sources

In this section, we use the simplified MSE expressions of the matrix-based and tensor-based R-D ESPRIT-type algorithms for two sources to analytically analyze the FBA gain and the tensor gain. We first consider the FBA gain in the matrix case and in the tensor case in Section 4.5.5.1 and in Section 4.5.5.2, respectively. Then, in Section 4.5.5.3, we resort to the tensor gain.

4.5.5.1. FBA gain in the matrix case

We start with the FBA gain in the matrix case. Using MSE_{mat} for R-D Standard ESPRIT from Equation (4.53) and $\text{MSE}_{\text{mat}}^{(\text{fba})}$ for R-D Unitary ESPRIT from (4.57), we define the matrix-based FBA gain $\eta_{\text{mat}}^{(\text{fba})}$ as

$$\eta_{\text{mat}}^{(\text{fba})} = \frac{\text{MSE}_{\text{mat}}}{\text{MSE}_{\text{mat}}^{(\text{fba})}} = \frac{1 - |\hat{\rho}|^2 \cdot \cos^2(\Delta\varphi)}{1 - |\hat{\rho}|^2}. \tag{4.68}$$

Note that $\eta_{\text{mat}}^{(\text{fba})}$ is independent of the dimensions r. From (4.68), we find the following limiting behavior:

$$\eta_{\text{mat}}^{(\text{fba})}\Big|_{|\hat{\rho}|\to 0} = 1, \quad \text{and} \quad \eta_{\text{mat}}^{(\text{fba})}\Big|_{|\hat{\rho}|\to 1} = \infty, \text{ for } \cos^2(\Delta\varphi) < 1. \tag{4.69}$$

Thus, the FBA gain can be infinitely large for $|\hat{\rho}| \to 1$. The term $\cos^2(\Delta\varphi)$ determines how much of the achievable FBA gain is actually extracted. We find the following special cases

$$\eta_{\text{mat}}^{(\text{fba})}\Big|_{\cos^2(\Delta\varphi)=1} = 1 \quad \text{and} \quad \eta_{\text{mat}}^{(\text{fba})}\Big|_{\cos^2(\Delta\varphi)=0} = \frac{1}{1 - |\hat{\rho}|^2} = \eta_{\text{mat,max}}^{(\text{fba})}. \tag{4.70}$$

It is apparent that if $\cos^2(\Delta\varphi) = 1$, there is no FBA gain for arbitrary $|\hat{\rho}|$. In contrast, the maximum FBA gain $\eta_{\text{mat,max}}^{(\text{fba})}$ is achieved for $\cos^2(\Delta\varphi) = 0$ when

$$\Delta\varphi = \Delta\varphi_{\text{ref}} + \Delta\varphi_{\text{corr}} = \pm\frac{\pi}{2}, \tag{4.71}$$

i.e., the displacement of the array phase $\Delta\varphi_{\text{ref}}$ and the correlation phase sum up to $\pi/2$, which can be interpreted as the array viewing two orthogonal signals on which a full decorrelation can be performed. Note that for $\cos(\Delta\varphi) = 0$, the analytical performance of R-D Unitary ESPRIT is equal to the performance of R-D Standard ESPRIT as if no correlation were present at all, i.e.,

$$a_{\text{mat}}^{(\text{fba})(r)}\Big|_{\cos(\Delta\varphi)=0} = a_{\text{mat}}^{(r)}\Big|_{|\hat{\rho}|=0}. \tag{4.72}$$

4.5.5.2. FBA gain in the tensor case

In order to compute the FBA gain in the tensor case, we use the MSEs of R-D Standard Tensor-ESPRIT in (4.60) and R-D Unitary Tensor-ESPRIT in (4.66), respectively. Let us define the tensor-based FBA gain in the r-th mode as

$$\eta_{\text{ten}}^{(\text{fba})(r)} = \frac{\text{MSE}_{\text{ten}}^{(r)}}{\text{MSE}_{\text{ten}}^{(\text{fba})(r)}} = \frac{a_{\text{ten}}^{(r)}}{a_{\text{ten}}^{(\text{fba})(r)}}. \tag{4.73}$$

Note that similarly to the observation for the FBA gain in the matrix case, the decorrelation of the sources by using R-D Unitary Tensor-ESPRIT instead of R-D Standard Tensor ESPRIT is reflected by the term $\cos^2(\Delta\varphi)$. Apparently, in the optimum case $\cos(\Delta\varphi) = 0$, the analytical performance of R-D Unitary Tensor ESPRIT is equal to the performance of R-D Standard Tensor ESPRIT as if no correlation were present. We have

$$\text{MSE}_{\text{ten}}^{(\text{fba})(r)}\Big|_{\cos(\Delta\varphi)=0} = \text{MSE}_{\text{ten}}^{(r)}\Big|_{|\hat{\rho}|=0}. \tag{4.74}$$

However, when comparing the FBA gain in the tensor case to the FBA gain in the matrix case, the novelty lies in the fact that the decorrelation term $\cos(\Delta\varphi)$ also occurs together with the terms $\frac{|\alpha|^2}{|\alpha^{(r)}|^2}$ that are determined by the spatial correlation of the array steering vectors. Thus, the FBA gain in the tensor case also depends on the spatial correlation of the sources.

4.5.5.3. Analysis of the tensor gain

Based on the MSE expressions for R-D Standard ESPRIT and R-D Standard Tensor-ESPRIT in (4.53) and (4.60), respectively, we can formulate an analytical expression for the tensor gain. Let us compute the tensor gain $\eta_{\text{ten}}^{(r)}$ in the r-th dimension as

$$\eta_{\text{ten}}^{(r)} = \frac{\text{MSE}_{\text{mat}}^{(r)}}{\text{MSE}_{\text{ten}}^{(r)}} = \frac{a_{\text{mat}}^{(r)}}{a_{\text{ten}}^{(r)}}, \tag{4.75}$$

where we restate the definitions of $a_{\mathrm{mat}}^{(r)}$ and $a_{\mathrm{ten}}^{(r)}$ from (4.54) and (4.61) again as

$$a_{\mathrm{mat}}^{(r)} = \frac{M}{M_r} \cdot \left[\left(\frac{M}{M_r}\right)^2 \cdot (M_r-1)^2 \cdot 2 + \left|\frac{\alpha}{\alpha^{(r)}}\right|^2 \cdot \left|\alpha_{\mathrm{sel}}^{(r)}\right|^2 \cdot \left(2 + (M_r-2) \cdot \left|\mathrm{e}^{\mathrm{j}\Delta\mu^{(r)}} - 1\right|^2\right) \right. \\ \left. - 4 \cdot (M_r-1) \cdot \frac{|\alpha|^2}{\left|\alpha^{(r)}\right|^2} \cdot \mathrm{Re}\left\{\alpha_{\mathrm{sel},0}^{(r)}\right\} \right] \cdot \frac{1}{1-|\hat{\rho}|^2} \cdot \frac{1}{D_{\mathrm{sel}}^{(r)^2}} \tag{4.76}$$

$$a_{\mathrm{ten}}^{(r)} = \frac{1}{D_{\mathrm{sel}}^{(r)^2}} \cdot \left(\frac{b^{(r)}}{D_r} \cdot \left[\frac{1}{1-|\hat{\rho}|^2} - \frac{\left(\frac{M}{M_r}\right)^2 - \frac{|\alpha|^2}{|\alpha^{(r)}|^2}}{\left(\frac{M}{M_r}\right)^2 - |\hat{\rho}|^2 \cdot \left|\frac{\alpha}{\alpha^{(r)}}\right|^2} \right] + c^{(r)} \cdot \frac{\frac{M}{M_r}}{\left(\frac{M}{M_r}\right)^2 - |\hat{\rho}|^2 \cdot \frac{|\alpha|^2}{|\alpha^{(r)}|^2}} \right) \tag{4.77}$$

with

$$b^{(r)} = M \cdot (M_r-1)^2 \cdot \left|\mathrm{e}^{\mathrm{j}\Delta\mu^{(r)}} - 1\right|^2 \cdot \left|\alpha_{\mathrm{sel}}^{(r)}\right|^2 \cdot \left[\left(\frac{M}{M_r}\right)^2 - \frac{|\alpha|^2}{\left|\alpha^{(r)}\right|^2} \right]$$

$$c^{(r)} = \left(\frac{M}{M_r}\right)^4 \cdot (M_r-1)^2 \cdot 2 + \left|\frac{\alpha}{\alpha^{(r)}}\right|^4 \cdot \left|\alpha_{\mathrm{sel}}^{(r)}\right|^2 \cdot \left(2 + (M_r-2) \cdot \left|\mathrm{e}^{\mathrm{j}\Delta\mu^{(r)}} - 1\right|^2\right) \\ - 4 \cdot \left(\frac{M}{M_r}\right)^2 \cdot (M_r-1) \cdot \left|\frac{\alpha}{\alpha^{(r)}}\right|^2 \cdot \mathrm{Re}\left\{\alpha_{\mathrm{sel},0}^{(r)}\right\}.$$

In order to simplify these expressions, we define $\phi^{(r)}$ as the normalized complex-valued correlation of the array steering vectors, i.e., $\phi^{(r)} = \frac{1}{M_r} \cdot \boldsymbol{a}_1^{(r)^{\mathrm{H}}} \cdot \boldsymbol{a}_2^{(r)}$. Moreover, let $\phi = \prod_{r=1}^{R} \phi^{(r)}$. Then, we can write

$$\frac{|\alpha|^2}{\left|\alpha^{(r)}\right|^2} = \left(\frac{M}{M_r}\right)^2 \cdot \frac{|\phi|^2}{\left|\phi^{(r)}\right|^2}, \quad \text{where} \quad \frac{|\phi|^2}{\left|\phi^{(r)}\right|^2} \in [0,1]. \tag{4.78}$$

This allows us to formulate $a_{\mathrm{mat}}^{(r)}$ in (4.76) in compact form as

$$a_{\mathrm{mat}}^{(r)} = \frac{d^{(r)}}{D_{\mathrm{sel}}^{(r)^2}} \cdot \frac{1}{1-|\hat{\rho}|^2}, \tag{4.79}$$

where

$$d^{(r)} = \left(\frac{M}{M_r}\right)^3 \cdot (M_r-1)^2 \cdot \left[2 + \left|\frac{\phi}{\phi^{(r)}}\right|^2 \cdot \left|\phi_{\mathrm{sel}}^{(r)}\right|^2 \cdot \left(2 + (M_r-2) \cdot \left|\mathrm{e}^{\mathrm{j}\Delta\mu^{(r)}} - 1\right|^2\right) \right. \\ \left. - 4 \cdot \left|\frac{\phi}{\phi^{(r)}}\right|^2 \cdot \mathrm{Re}\left\{\phi_{\mathrm{sel},0}^{(r)}\right\} \right], \tag{4.80}$$

and the determinants $D_{\text{sel}}^{(r)}$ and D_r are given by

$$D_{\text{sel}}^{(r)} = \left(\frac{M}{M_r}\right)^2 \cdot (M_r - 1)^2 \cdot \left(1 - \left|\frac{\phi}{\phi^{(r)}} \cdot \phi_{\text{sel}}^{(r)}\right|^2\right) \tag{4.81}$$

$$D_r = M_r \cdot \left(1 - \left|\phi^{(r)}\right|^2\right). \tag{4.82}$$

Similarly, we can write $a_{\text{ten}}^{(r)}$ in (4.77) as

$$a_{\text{ten}}^{(r)} = \frac{1}{D_{\text{sel}}^{(r)^2}} \cdot \left(\frac{b^{(r)}}{D_r} \cdot \left[\frac{1}{1 - |\hat{\rho}|^2} - \frac{1 - \left|\frac{\phi}{\phi^{(r)}}\right|^2}{1 - |\hat{\rho}|^2 \cdot \left|\frac{\phi}{\phi^{(r)}}\right|^2}\right] + \frac{\frac{M_r}{M} \cdot c^{(r)}}{1 - |\hat{\rho}|^2 \cdot \left|\frac{\phi}{\phi^{(r)}}\right|^2}\right), \tag{4.83}$$

where

$$b^{(r)} = M \cdot \left(\frac{M}{M_r}\right)^2 \cdot (M_r - 1)^3 \cdot \left|\mathrm{e}^{\mathrm{j}\Delta\mu^{(r)}} - 1\right|^2 \cdot \left|\phi_{\text{sel}}^{(r)}\right|^2 \cdot \left[1 - \frac{|\phi|^2}{\left|\phi^{(r)}\right|^2}\right] \tag{4.84}$$

$$\begin{aligned}\frac{M_r}{M} \cdot c^{(r)} = \left(\frac{M}{M_r}\right)^3 \cdot (M_r - 1)^2 \cdot \Bigg[2 + \left|\frac{\phi}{\phi^{(r)}}\right|^4 \cdot \left|\phi_{\text{sel}}^{(r)}\right|^2 \cdot \left(2 + (M_r - 2) \cdot \left|\mathrm{e}^{\mathrm{j}\Delta\mu^{(r)}} - 1\right|^2\right) \\ - 4 \cdot \left|\frac{\phi}{\phi^{(r)}}\right|^2 \cdot \mathrm{Re}\left\{\phi_{\text{sel},0}^{(r)}\right\}\Bigg].\end{aligned} \tag{4.85}$$

It is apparent, that the term $a_{\text{ten}}^{(r)}$ consists of two terms. The first term is related to $b^{(r)}$, which originates from the projections onto the r-mode column spaces (cf. (B.110) and (B.119) in Appendix B.7), and the second term is related to $c^{(r)}$, which is in fact very similar to $a_{\text{mat}}^{(r)}$ from the MSE expression of the matrix-based R-D Standard ESPRIT algorithm.

We proceed to investigate the limiting behavior of $a_{\text{ten}}^{(r)}$ in terms of the temporal correlation of the source signals. In the case of uncorrelated sources, we have $|\hat{\rho}| \to 0$ such that $a_{\text{ten}}^{(r)}$ simplifies to

$$a_{\text{ten}}^{(r)}\Big|_{|\hat{\rho}|\to 0} = \frac{1}{D_{\text{sel}}^{(r)^2}} \cdot \left(\frac{b^{(r)}}{D_r} \cdot \frac{|\phi|^2}{\left|\phi^{(r)}\right|^2} + c^{(r)} \cdot \frac{M_r}{M}\right). \tag{4.86}$$

Equation (4.86) highlights, that for temporally uncorrelated sources the MSE for R-D Standard Tensor ESPRIT only depends on the spatial correlation. Moreover, for fully correlated sources we have

$$a_{\text{ten}}^{(r)}\Big|_{|\hat{\rho}|\to 1} = \frac{1}{D_{\text{sel}}^{(r)^2}} \cdot \left(\underbrace{\frac{b^{(r)}}{D_r} \cdot \left[\frac{1}{0^+} - 1\right]}_{\infty} + \frac{\frac{M_r}{M} \cdot c^{(r)}}{1 - \left|\frac{\phi}{\phi^{(r)}}\right|^2}\right). \tag{4.87}$$

This result indicates that for temporally correlated sources and $\left|\frac{\phi}{\phi^{(r)}}\right| < 1$ the first term of (4.87) becomes dominant compared to the second term.

4.6. Simulation results

In this section, we first provide simulation results for the reviewed R-D performance analysis framework of the matrix-based and tensor-based R-D ESPRIT-type algorithms in Section 4.6.1. Then, in Section 4.6.2, we verify the analytical results for the FBA gain and the tensor gain for two sources and analyze its behavior in terms of the physical parameters.

4.6.1. R-D Tensor-ESPRIT-type algorithms

In this subsection, we show numerical results to demonstrate the asymptotic behavior of the analytical performance assessment for R-D ESPRIT-type algorithms and R-D Tensor-ESPRIT-type algorithms, which is reviewed in this chapter. To this end, we compare the square root of the analytical MSE expressions "ana" from (4.17) and (4.27) for R-D Standard ESPRIT (SE) and R-D Unitary ESPRIT (UE) as well as those in (4.34) and (4.40) for R-D Standard Tensor-ESPRIT (STE) and R-D Unitary Tensor-ESPRIT (UTE) to the root mean squared error (RMSE) of the empirical estimation error "emp" obtained by averaging over Monte-Carlo trials. For all ESPRIT-type algorithms, LS is used to solve the shift invariance equations. The RMSE is defined as

$$\text{RMSE} = \sqrt{\mathbb{E}\left\{\sum_{r=1}^{R}\sum_{i=1}^{d}\left(\mu_i^{(r)} - \hat{\mu}_i^{(r)}\right)^2\right\}}, \tag{4.88}$$

where $\hat{\mu_i}^{(r)}$ is the estimate of i-th spatial frequency in the r-th mode. For the comparison, we also include the deterministic Cramér-Rao bound (CRB) [SN89], which also applies to the tensor-base R-D ESPRIT-type algorithms, which only use a different representation of the same data model.

For all the simulations, we employ a 2-D uniform rectangular array (URA) with uniform spacing in all dimensions and isotropic sensor elements. The phase reference is chosen to be at the centroid of the array. It is assumed for all algorithms that a known number of d narrow-band signals with the symbols $s_i[n]$ drawn from a complex Gaussian distribution impinge on the array and that N subsequent snapshots $n = 1,,2,\ldots,N$ are observed. The signals are assumed to have unit power, i.e., $\mathbb{E}\left\{|s_i(t)|^2\right\} = 1$. To consider source correlation, we generate the symbols such that $\mathbb{E}\left\{|s_i^*(t)\cdot s_j(t)|^2\right\} = \rho\cdot \mathrm{e}^{\mathrm{j}\varphi_{c_{i,j}}}$ for $i \neq j = 1,2,\ldots,d$, where $\rho \in [0,1]$ is the magnitude of the correlation coefficient between each pair of sources and $\varphi_{c_{i,j}}$ is the uniform distributed correlation phase in $[0,2\pi]$. The additive noise is generated according to a circularly symmetric complex Gaussian distribution with zero mean and variance σ_{n}^2.

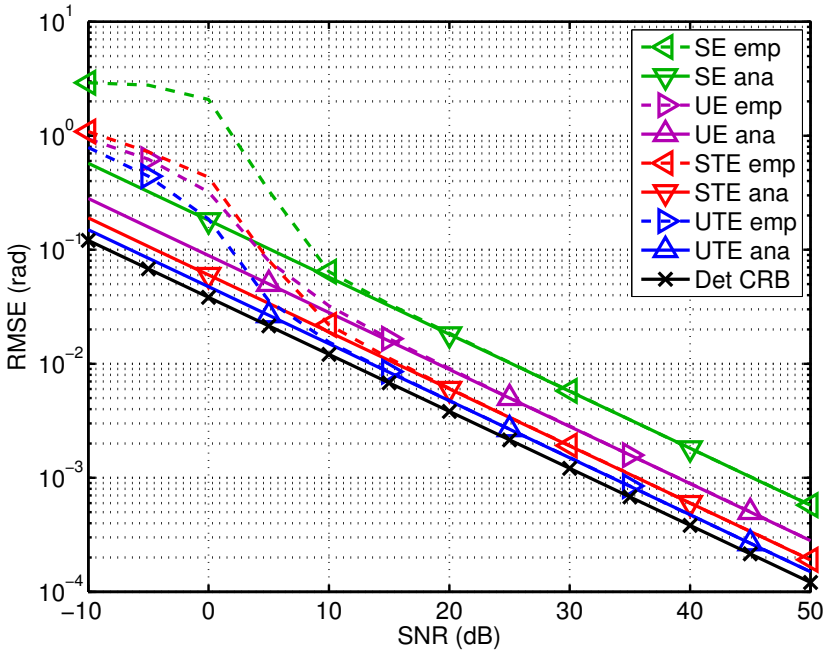


Figure 4.1.: RMSE versus the SNR for a 5×6 URA, $N = 20$, and $d = 2$ correlated signals ($\rho = 0.97$) at $\mu_1^{(1)} = 1$, $\mu_2^{(1)} = -0.5$, $\mu_1^{(2)} = -0.5$, and $\mu_2^{(2)} = 1$.

In Figure 4.1, we illustrate the RMSE versus the SNR for a 5×6 URA and $N = 20$ snapshots. We have $d = 2$ widely-spaced sources located at $\mu_1^{(1)} = 1$, $\mu_2^{(1)} = -0.5$, $\mu_1^{(2)} = -0.5$, and $\mu_2^{(2)} = 1$. The sources are highly correlated with a correlation of $\rho = 0.97$. Figure 4.2 shows the RMSE versus the SNR for $d = 3$ closely-spaced sources with the spatial frequencies $\mu_1^{(1)} = 0.7, \mu_2^{(1)} = 0.9, \mu_3^{(1)} = 1.1, \mu_1^{(2)} = -0.1, \mu_2^{(2)} = -0.3, \mu_3^{(2)} = -0.5$. The correlation coefficient is $\rho = 0.97$ and we employ a 5×5 URA with $N = 20$ snapshots. In Figure 4.3, we depict the RMSE versus the number of snapshots N for a 5×6 URA. The SNR is fixed at 10 dB and we have $d = 3$ uncorrelated sources with the spatial frequencies $\mu_1^{(1)} = 0.25$, $\mu_2^{(1)} = 0.5$, $\mu_3^{(1)} = 0.75$, $\mu_1^{(2)} = 0.25$, $\mu_2^{(2)} = 0.5$, and $\mu_3^{(2)} = 0.75$.

These simulation results demonstrate that the empirical estimation errors agree well with the analytical ones for high effective SNRs, i.e., when either the SNR is large or the number of snapshots is large. This is also expected as the performance analysis framework is asymptotically accurate for high effective SNRs.

4.6.2. Analytical results for two sources

In this section, we provide numerical results to verify the derived analytical MSE expressions for $d = 2$ sources. We consider a URA ($R = 2$) with isotropic sensor elements and the source positions are fixed at $\mu_1^{(1)} = 1$, $\mu_1^{(2)} = 1$, $\mu_2^{(1)} = 1.1$, $\mu_2^{(2)} = 1.1$. Note that the source positions are chosen to be symmetric across the dimensions as this simplifies the analysis of the achievable gains, which are equal in each dimension. Since the MSE expressions are asymptotic in the effective SNR, the

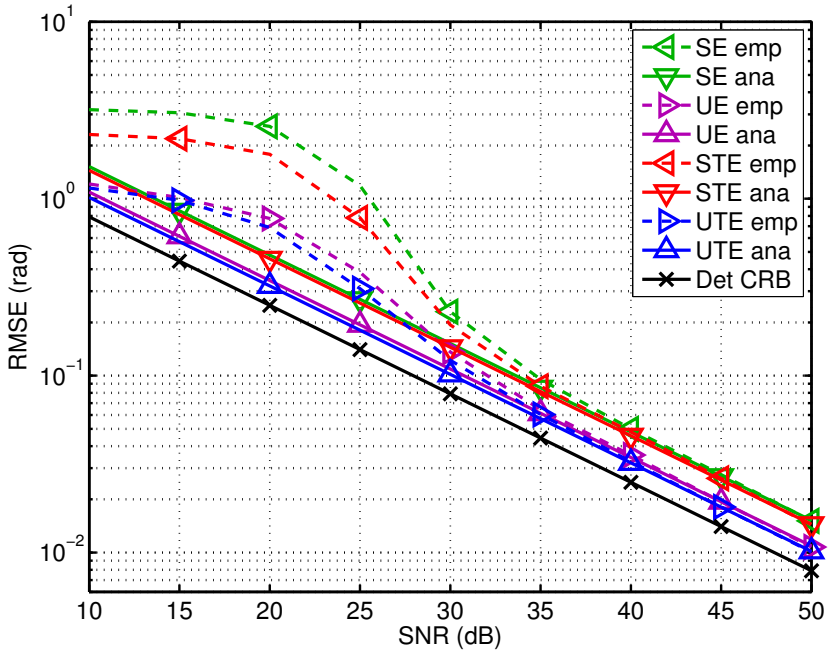


Figure 4.2.: RMSE versus the SNR for a 5×6 URA, $N = 20$, and $d = 3$ correlated signals ($\rho = 0.97$) at $\mu_1^{(1)} = 0.7, \mu_2^{(1)} = 0.9, \mu_3^{(1)} = 1.1, \mu_1^{(2)} = -0.1, \mu_2^{(2)} = -0.3, \mu_3^{(2)} = -0.5$.

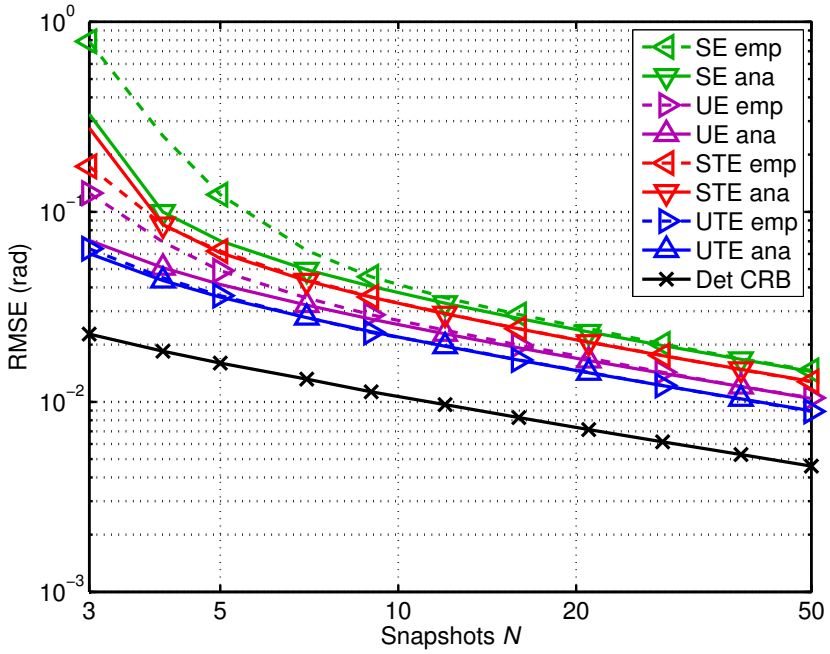


Figure 4.3.: RMSE versus the number of snapshots N for for a 5×6 URA, SNR $= 20$ dB, and $d = 3$ correlated signals ($\rho = 0.97$) at $\mu_1^{(1)} = 0.7, \mu_2^{(1)} = 0.9, \mu_3^{(1)} = 1.1, \mu_1^{(2)} = -0.1, \mu_2^{(2)} = -0.3, \mu_3^{(2)} = -0.5$.

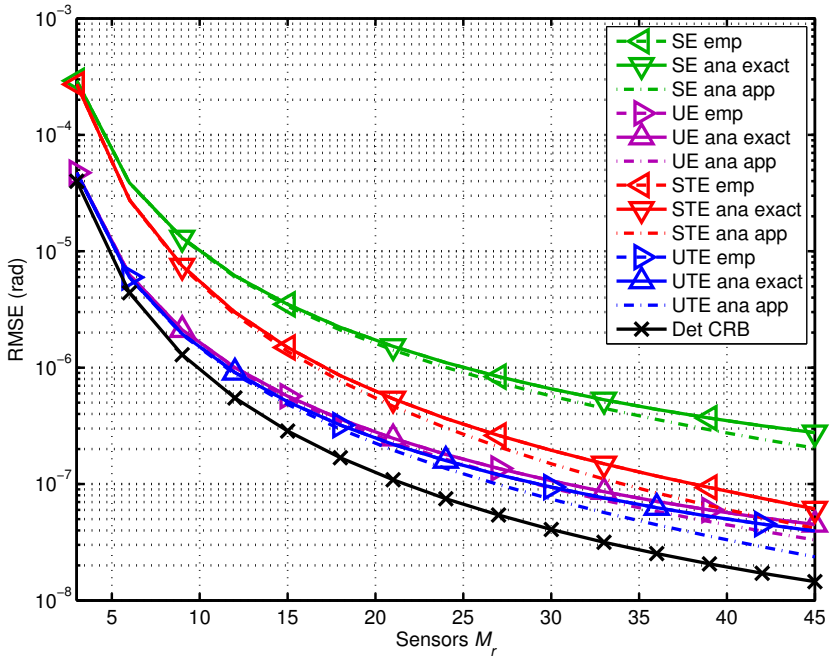


Figure 4.4.: RMSE versus the number of sensors M_r, $r = 1, 2$ with $\delta^{(1)} = \delta^{(2)} = 0$ for $d = 2$ correlated sources with $|\rho| = 0.99$ and $\varphi_{\text{corr}} = \pi/3$ placed at $\mu_1^{(1)} = 1.0$, $\mu_1^{(2)} = 1.0$, $\mu_2^{(1)} = 1.1$, $\mu_2^{(2)} = 1.1$.

effective SNR is set to $\varrho = \frac{NP_1}{\sigma_n^2} = 100\,\text{dB}$ with ($P_1 = 40\,\text{dBm}, P_2 = 33.01\,\text{dBm}$, $N = 100$, $\sigma_n^2 = 10^{-9}$).

In the first part of our analysis, we investigate the effect of the array phase reference on the decorrelation capabilities of 2-D Unitary ESPRIT (UE) and 2-D Unitary Tensor-ESPRIT (UTE) compared to 2-D Standard ESPRIT (SE) and 2-D Standard Tensor-ESPRIT (STE). Moreover, we include the approximations for small $\Delta\mu^{(r)}$ for each of the algorithms. It is clear from (4.57) and (4.66) that the expressions for UE and UTE only differ from the ones for SE and STE by the factor $\cos^2(\Delta\varphi)$, where $\Delta\varphi = \Delta\varphi_{\text{ref}} + \Delta\varphi_{\text{corr}}$.

Figure 4.4 and Figure 4.5 illustrate the RMSE versus the number of sensors M_r for $d = 2$ highly correlated sources with $|\hat{\rho}| = 0.99$ and $\Delta\varphi_{\text{corr}} = \pi/3$ for different phase reference parameters $\delta^{(r)}$. Due to the symmetry of the sources, it is sufficient to consider only the r-th mode. In Figure 4.4, the array phase reference is located at the array centroid in each mode by setting $\delta^{(r)} = 0, \forall r$. In this case $\Delta\varphi_{\text{ref}} = \sum_{r=1}^{R} \delta^{(r)} \cdot \Delta\mu^{(r)} = 0$ and hence $\Delta\varphi$ is determined by the source correlation phase $\Delta\varphi_{\text{corr}}$ only as $\Delta\varphi = \Delta\varphi_{\text{corr}}$. In Figure 4.5, the curves for UE and UTE become dependent on the number of sensors M_r, which is explained by the contribution of the parameter for the array phase reference if $\delta^{(r)} \neq 0$. It is also apparent from Figure 4.4 and Figure 4.5 that the curves associated with the approximated terms only agree well with the exact curves for small $\Delta\Delta\mu(r)$, which is where the Taylor approximation is valid.

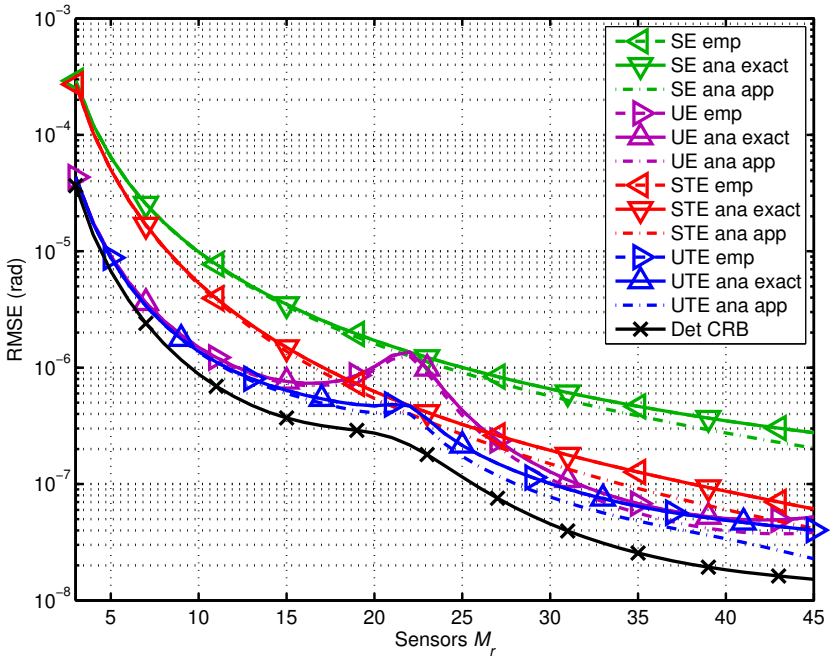


Figure 4.5.: RMSE versus the number of sensors M_r, $r = 1, 2$ with $\delta^{(r)} = \frac{M_r - 1}{2}$ for $d = 2$ correlated sources with $|\rho| = 0.99$ and $\varphi_{\text{corr}} = \pi/3$ placed at $\mu_1^{(1)} = 1.0$, $\mu_1^{(2)} = 1.0$, $\mu_2^{(1)} = 1.1$, $\mu_2^{(2)} = 1.1$.

4.7. Summary

In this chapter, we have first reviewed the R-D performance analysis framework in [RHD14] for matrix-based and tensor-based R-D ESPRIT-type algorithms using LS to solve the shift invariance equation. The analytical expressions for the parameter estimation error and the MSE only require the noise to be zero-mean with finite SO moments and are asymptotic in the effective SNR. Hence, no further assumptions on the noise statistics are required.

The analytical MSE expressions are formulated in terms of the subspaces of the measurement matrix and provide no explicit insights into the influence of the physical parameters, e.g., the SNR, the number of sensors, the sample size, etc. Therefore, we have simplified the MSE expressions of the matrix-based and tensor-based R-D ESPRIT-type algorithms for the special cases of a single and two sources. The resulting expressions are compact and only depend on the physical parameters of significance.

For a single source, we have seen that the matrix-based and the tensor-based R-D ESPRIT-type algorithms result in the same MSE expression. This implies that they perform identical and no gain from forward-backward averaging or tensor processing can be achieved in this case.

In the next step, we have simplified the matrix-based R-D ESPRIT-type algorithms for two sources. Based on the obtained MSE expressions for R-D Standard ESPRIT and R-D Unitary ESPRIT, we have analytically computed the gain from forward-backward averaging and analyzed

its dependence on the physical parameters. Moreover, we compute the asymptotic efficiency for $R = 1$.

Finally, we have simplified MSE of the tensor-based R-D ESPRIT-type algorithms for two sources. We have used the MSE expressions for R-D Standard ESPRIT and R-D Standard Tensor-ESPRIT to analytically compute and analyze the tensor gain. Comparing the MSE of R-D Standard Tensor-ESPRIT and R-D Unitary Tensor-ESPRIT, we also compute the forward-backward averaging gain in the tensor case.

5. ESPRIT-type parameter estimation algorithms using generalized least squares

In Chapter 3, we have reviewed R-D ESPRIT-type parameter estimation algorithms, which provide closed-form estimates at a low computational complexity. R-D ESPRIT-type algorithms exploit the shift invariance structure of the array by solving the shift invariance equation (cf. Equation (3.33)) in terms of the estimated signal subspace via least squares methods (cf. Chapter 3). In this chapter, we propose a novel least squares algorithm, termed general least squares (GLS), for matrix-based R-D ESPRIT-type algorithms, which takes the statistics of the subspace estimation error into account for the solution. Moreover, we present a performance analysis for R-D ESPRIT-type algorithms with GLS based on the R-D performance analysis framework [RHD14] reviewed in Chapter 4. For 1-D parameter estimation and the special cases of a single source and two temporally orthogonal sources, we show that the analytical MSE expressions coincide with the deterministic Cramér-Rao bound (CRB), which implies that 1-D ESPRIT-type algorithms in combination with GLS are asymptotically efficient in these cases, i.e., the ratio of the CRB and the MSE is equal to 1.

We first provide a brief literature review and list the specific contributions in Section 5.1. After discussing the existing least squares algorithms for R-D ESPRIT-type algorithms in Section 5.2, we derive the novel GLS algorithm for matrix-based R-D ESPRIT-type algorithms in Section 5.3. In Section 5.4, we present a first-order performance analysis of the GLS-based R-D ESPRIT-type algorithms and in Section 5.5, we prove that the GLS-based 1-D ESPRIT-type algorithms are asymptotically efficient for the special cases of a single source and two temporally orthogonal sources. Finally, simulation results in Section 5.6 verify the analytical results and a summary is provided in Section 5.7.

5.1. Overview

In Section 5.1.1, we provide a literature overview of least squares algorithms to solve the shift invariance equation (cf. Equation (3.33)) for the R-D ESPRIT-type algorithms from Section 3.4 and summarize the contributions in Section 5.1.2.

5.1.1. State of the art

Estimating the directions of arrival (DOAs) of noise-corrupted signals has long been a fundamental problem in signal processing and is required in many applications such as radar, sonar, and wireless communications. Due to their low complexity and high-resolution capabilities, ESPRIT-type algorithms [RK89, HN95] are among the most popular subspace-based parameter estimation

schemes. After the subspace estimation, ESPRIT-based algorithms solve a highly structured linear system of equations, termed shift invariance equation. The structure is imposed by applying two selection matrices with potential overlap to an estimate of the signal subspace in which case the perturbations on both sides of the shift invariance equation are highly correlated. The shift invariance equation is usually solved by means of least squares (LS) [RK89], total least squares (TLS) [OVK91], or structured least squares (SLS) [Haa97b].

The simple LS solution to a linear system of equations is optimal and identical to the maximum-likelihood (ML) solution if the perturbation is zero-mean Gaussian and only on one side of the equations. As both assumptions are not satisfied by the shift invariance equation, LS is an inconsistent estimator. The TLS solution [OVK91] allows for perturbations on both sides of the shift invariance equation but only achieves the ML solution if both perturbations are zero-mean Gaussian. More importantly, TLS treats the perturbations on either side as independent irrespective of the fact that they can be highly correlated if the subarrays overlap. Hence, TLS is only appropriate for non-overlapping subarrays. In the case of overlapping subarrays, SLS [Haa97b] explicitly accounts for the dependence of the subspace perturbations on both sides. SLS iteratively solves the corresponding optimization problem by successive local linearization. Nevertheless, computing only one iteration is sufficient in many cases according to [Haa97b].

However, as a common drawback, the above mentioned shift invariance equation solutions do not take into account the statistics of the subspace perturbation. For the parameter estimation from general linear models with arbitrary noise distribution, the generalized least squares (GLS) algorithm [Ame85, Kay93] can be applied. GLS exploits the noise statistics by using the noise covariance matrix as a weighting matrix. In the literature, GLS is also known as Gauss-Markov estimator [Kay93] or as weighted least squares (WLS). Nevertheless, to be precise, WLS is considered a special case of GLS, where the weighting matrix is diagonal [Ame85]. By the Gauss-Markov theorem [Kay93], it has been shown that GLS is in fact the best linear unbiased estimator (BLUE) of the parameter vector in general linear models with arbitrary noise covariance. Moreover, in the special case of Gaussian noise, it is equivalent to the maximum likelihood estimator [Kay93], which asymptotically achieves the Cramér-Rao bound (CRB).

In [SN91, ES94, QHS+15], a number of ESPRIT-based algorithms based on the GLS solution of the shift invariance equation has been derived, where a weighting matrix in form of the covariance matrix of the subspace estimation error is introduced to minimize the parameter estimation error. In the work of [SN91, ES94], the weighting matrix is optimized via a first-order performance analysis [Bri75, RH89a] in order to minimize the analytical parameter estimation error. While [SN91] optimizes the weighting matrix only for one of the DOAs, [ES94] presents a GLS-based ESPRIT algorithm similar to the algorithms proposed in this chapter. However, [ES94] provides no proof of the resulting expression. A real-valued GLS-based 2-D parameter estimation algorithm using the principal-singular-vector utilization for modal analysis (PUMA) criterion has been derived in

[QHS+15] for a single source. However, this algorithm is indeed equivalent to Unitary ESPRIT for a single source and is a special case of our proposed algorithms. Several variations of the actual PUMA algorithm with GLS have been presented in [SCLC10] for a single source, in [CSS12] for multiple damped 2-D exponentials, and a tensor-based approach is provided in [SS12]. More recently in [QHSS16], an enhanced 1-D PUMA algorithm with GLS for multiple sources has been proposed. However, [QHSS16] requires several iterations followed by a root selection procedure, which was originally proposed in [SV15] for the Root-MUSIC algorithm [Bar83]. The equivalence of [QHSS16] to the MODE algorithm [Van02] was very recently shown in [ZSJ17].

For the GLS-based algorithms in [SCLC10, QHS+15, QHSS16], the authors also provide a performance analysis based on the framework in [Bri75, RH89a]. Additionally, they consider the special case of a single source and show that these algorithms are asymptotically efficient in this case, i.e., the ratio of the CRB and the MSE is equal to 1.

5.1.2. Contributions

In this chapter, we propose a GLS solution to the shift invariance equation of R-D ESPRIT-type algorithms, i.e., R-D Standard ESPRIT and R-D Unitary ESPRIT for multiple sources, where a uniform R-D sampling grid as defined in Section 2.1.4.1 and maximally overlapping subarrays in the R dimensions are assumed. The GLS algorithm for 1-D ESPRIT-type algorithms has been published in [SRH17a]. First, the shift invariance equation in matrix form is reorganized into a linear system of vector equations, which is solved using the GLS criterion that directly accounts for the statistics of the subspace perturbation through its covariance matrix. The covariance matrix of the subspace estimation error assuming white Gaussian sensor noise is found analytically via a first-order perturbation expansion from [LLV93], which is more general than [Bri75] as discussed in Section 4.1.1. As the error covariance matrix requires an initial estimate of the unknown shift invariance equation solution, an iterative procedure by repeatedly performing GLS updates is possible. We show that if GLS is initialized by the simple LS solution, only one GLS iteration is required to achieve a significantly improved estimation accuracy in the asymptotic case, i.e., at either high SNRs or a large sample size. However, at low SNRs and for a small sample size N, performing additional GLS iterations further improves the estimation accuracy. It should be mentioned that Unitary ESPRIT with GLS involves only real-valued operations, thus reducing the computational complexity.

In our second contribution in Section 5.3.2, we develop a performance analysis for R-D ESPRIT-type algorithms using a single GLS iteration. The derived analytical MSE expressions are based on the framework in Chapter 4 and are asymptotic in the effective SNR. We simplify the MSE expressions for a single source and two temporally orthogonal sources and show that they coincide with the deterministic Cramér-Rao bound (CRB), which provides a lower limit for the MSE. This implies that in these cases, R-D ESPRIT-type algorithms in combination with one GLS iteration

are asymptotically efficient, i.e., the ratio of the CRB and the MSE is equal to 1. However, we have observed via a simulative study that R-D Standard ESPRIT with GLS is indeed asymptotically efficient for an arbitrary number of uncorrelated sources while R-D Unitary ESPRIT with GLS is asymptotically efficient even for an arbitrary number of correlated but incoherent sources.

5.2. Shift invariance equation solutions

In this section, we review the commonly-used least squares algorithms, which are applied to solve the R-D shift invariance equation in R-D ESPRIT-type algorithms.

Consider the matrix-based R-D data model in (2.3) and assume a uniform R-D sampling grid as defined in Section 2.1.4.1. The corresponding measurement matrix $\boldsymbol{X}$ can be modeled as

$$\boldsymbol{X} = \boldsymbol{A} \cdot \boldsymbol{S} + \boldsymbol{N} = \boldsymbol{X}_0 + \boldsymbol{N} \in \mathbb{C}^{M \times N}, \tag{5.1}$$

where $\boldsymbol{S} \in \mathbb{C}^{d \times N}$ represents the source symbol matrix, $\boldsymbol{N} \in \mathbb{C}^{M \times N}$ contains the additive sensor noise samples, and $\boldsymbol{X}_0$ is the noise-free measurement matrix. Moreover, $\boldsymbol{A} = [\boldsymbol{a}(\boldsymbol{\mu}_1), \ldots, \boldsymbol{a}(\boldsymbol{\mu}_d)] \in \mathbb{C}^{M \times d}$ is the array steering matrix, which consists of the array steering vectors $\boldsymbol{a}(\boldsymbol{\mu}_i)$ corresponding to the i-th spatial frequency, which are given by

$$\boldsymbol{a}(\boldsymbol{\mu}_i) = \boldsymbol{a}^{(1)}\left(\mu_i^{(1)}\right) \otimes \cdots \otimes \boldsymbol{a}^{(R)}\left(\mu_i^{(R)}\right) \in \mathbb{C}^{M \times 1}, \tag{5.2}$$

where $\boldsymbol{a}^{(r)}\left(\mu_i^{(r)}\right) \in \mathbb{C}^{M_r \times 1}$ is the r-mode array steering vector defined in (2.13).

As explained in detail in Section 3.2.1, the true signal subspace and its estimate can be obtained by computing the SVD of the noise-free measurement matrix $\boldsymbol{X}_0$ and its noise-corrupted version $\boldsymbol{X}$ as

$$\boldsymbol{X}_0 = \begin{bmatrix} \boldsymbol{U}_{\mathrm{s}}, & \boldsymbol{U}_{\mathrm{n}} \end{bmatrix} \cdot \begin{bmatrix} \boldsymbol{\Sigma}_{\mathrm{s}} & \boldsymbol{0}_{d \times (N-d)} \\ \boldsymbol{0}_{(M-d) \times d} & \boldsymbol{0}_{(M-d) \times (N-d)} \end{bmatrix} \cdot \begin{bmatrix} \boldsymbol{V}_{\mathrm{s}}, & \boldsymbol{V}_{\mathrm{n}} \end{bmatrix}^{\mathrm{H}} \tag{5.3}$$

$$\boldsymbol{X} = \begin{bmatrix} \hat{\boldsymbol{U}}_{\mathrm{s}}, & \hat{\boldsymbol{U}}_{\mathrm{n}} \end{bmatrix} \cdot \begin{bmatrix} \hat{\boldsymbol{\Sigma}}_{\mathrm{s}} & \boldsymbol{0}_{d \times (N-d)} \\ \boldsymbol{0}_{(M-d) \times d} & \hat{\boldsymbol{\Sigma}}_{\mathrm{n}} \end{bmatrix} \cdot \begin{bmatrix} \hat{\boldsymbol{V}}_{\mathrm{s}}, & \hat{\boldsymbol{V}}_{\mathrm{n}} \end{bmatrix}^{\mathrm{H}}, \tag{5.4}$$

where $\boldsymbol{U}_{\mathrm{s}} \in \mathbb{C}^{M \times d}$, $\boldsymbol{U}_{\mathrm{n}} \in \mathbb{C}^{M \times (M-d)}$, and $\boldsymbol{V}_{\mathrm{s}} \in \mathbb{C}^{N \times d}$ span the signal subspace, the noise subspace, and the row subspace, and $\boldsymbol{\Sigma}_{\mathrm{s}} \in \mathbb{R}^{d \times d}$ contains the d non-zero singular values on its diagonal. Moreover, $\hat{\boldsymbol{U}}_{\mathrm{s}}$, $\hat{\boldsymbol{U}}_{\mathrm{n}}$, $\hat{\boldsymbol{V}}_{\mathrm{s}}$, and $\hat{\boldsymbol{\Sigma}}_{\mathrm{s}}$ are their respective estimates.

Then, following (3.25) and (3.33), the set of R shift invariance equations with respect to the estimated signal subspace is given by [HN98]

$$\tilde{\boldsymbol{J}}_1^{(r)} \cdot \hat{\boldsymbol{U}}_\mathrm{s} \cdot \boldsymbol{\Psi}^{(r)} \approx \tilde{\boldsymbol{J}}_2^{(r)} \cdot \hat{\boldsymbol{U}}_\mathrm{s}, \quad r = 1, \ldots, R \tag{5.5}$$

where $\boldsymbol{\Psi}^{(r)} = \boldsymbol{T} \cdot \boldsymbol{\Phi}^{(r)} \cdot \boldsymbol{T}^{-1}$, where $\boldsymbol{\Phi}^{(r)} = \mathrm{diag}\{[\mathrm{e}^{\mathrm{j}\mu_1^{(r)}}, \ldots, \mathrm{e}^{\mathrm{j}\mu_d^{(r)}}]^\mathrm{T}\} \in \mathbb{C}^{d \times d}$ contains the spatial frequencies in the r-th mode. The selection matrices $\tilde{\boldsymbol{J}}_n^{(r)} \in \mathbb{R}^{\frac{M}{M_r}(M_r-1) \times M}$ for $n = 1, 2$ are given in (3.26) based on $\boldsymbol{J}_n^{(r)} \in \mathbb{R}^{(M_r-1) \times M_r}$, $n = 1, 2$ in (3.27) for maximally overlapping subarrays. Suppose we have obtained an estimate $\hat{\boldsymbol{\Psi}}^{(r)}$ of the unknown matrix $\boldsymbol{\Psi}^{(r)}$ in the r-th mode from (5.5). Then, after performing a joint EVD of the R estimates $\hat{\boldsymbol{\Psi}}^{(r)}$ (e.g., via [FG06]), the spatial frequency estimates $\hat{\mu}_i^{(r)}$, $i = 1, \ldots, d$, are extracted from the eigenvalues $\hat{\lambda}_i^{(r)}$ of the solution $\hat{\boldsymbol{\Psi}}^{(r)}$ via $\hat{\mu}_i^{(r)} = \arg\left\{\hat{\lambda}_i^{(r)}\right\}$.

The shift invariance equation in (5.5) is an overdetermined linear system of equations. As it is expressed in terms of the estimate $\hat{\boldsymbol{U}}_\mathrm{s}$ of the signal subspace $\boldsymbol{U}_\mathrm{s}$, (5.5) does not have an exact solution in general. Additionally, the shift invariance equation in (5.5) can be highly structured, where the structure is imposed by applying two selection matrices with potential overlap to the signal subspace estimate $\hat{\boldsymbol{U}}_\mathrm{s}$, in which case the perturbations on both sides of the shift invariance equation are highly correlated. Often, least squares algorithms are used to approximately solve (5.5) for $\boldsymbol{\Psi}^{(r)}$. These include, for instance, the simple least squares (LS) algorithm [RK89], total least squares (TLS) [OVK91], or structured least squares (SLS) [Haa97b]. These three LS algorithms will be briefly discussed in what follows.

First, we write the signal subspace estimate $\hat{\boldsymbol{U}}_\mathrm{s}$ as $\hat{\boldsymbol{U}}_\mathrm{s} = \boldsymbol{U}_\mathrm{s} + \Delta\boldsymbol{U}_\mathrm{s}$, where $\Delta\boldsymbol{U}_\mathrm{s}$ denotes the signal subspace estimation error. Then, the simple LS solution of (5.5) in [RK89] assumes that

$$\tilde{\boldsymbol{J}}_1^{(r)} \cdot \boldsymbol{U}_\mathrm{s} \cdot \boldsymbol{\Psi}_\mathrm{LS}^{(r)} = \tilde{\boldsymbol{J}}_2^{(r)} \cdot (\boldsymbol{U}_\mathrm{s} + \Delta\boldsymbol{U}_\mathrm{s}), \tag{5.6}$$

is satisfied such that the perturbation term $\left\|\tilde{\boldsymbol{J}}_2^{(r)} \cdot \Delta\boldsymbol{U}_\mathrm{s}\right\|_\mathrm{F}$ is minimized. The LS solution is only optimal and identical to the maximum-likelihood (ML) solution if there is no perturbation on the left hand side and the perturbation in $\Delta\boldsymbol{U}_\mathrm{s}$ is zero-mean Gaussian distributed. As both assumptions are clearly not satisfied by the shift invariance equation in (5.5), the LS algorithm provides a suboptimal solution and constitutes an inconsistent estimator that produces a bias.

The TLS algorithm in [OVK91] allows for perturbation terms on both sides of (5.5) and assumes that

$$\tilde{\boldsymbol{J}}_1^{(r)} \cdot (\boldsymbol{U}_\mathrm{s} + \Delta\boldsymbol{U}_\mathrm{s}) \cdot \boldsymbol{\Psi}_\mathrm{TLS}^{(r)} = \tilde{\boldsymbol{J}}_2^{(r)} \cdot (\boldsymbol{U}_\mathrm{s} + \Delta\boldsymbol{U}_\mathrm{s}) \tag{5.7}$$

is satisfied, while minimizing the perturbation $\left\|\left[\tilde{\boldsymbol{J}}_1^{(r)} \cdot \Delta\boldsymbol{U}_\mathrm{s},\ \tilde{\boldsymbol{J}}_2^{(r)} \cdot \Delta\boldsymbol{U}_\mathrm{s}\right]\right\|_\mathrm{F}$. The TLS algorithm provides a consistent solution to (5.5) but only achieves the ML solution if the perturbation terms on the left and the right hand side of (5.5) are zero-mean Gaussian distributed. More impor-

tantly, TLS treats the perturbation terms on either side as independent and does not account for their strong correlation in the case of overlapping subarrays. Hence, the TLS algorithm is only appropriate for non-overlapping subarray configurations [Haa97b].

The SLS algorithm in [Haa97b] explicitly accounts for the dependence of the subspace perturbation terms on both sides of (5.5) in the case of overlapping subarrays. The SLS algorithm solves the optimization problem

$$\min_{\Delta\boldsymbol{\Psi}^{(r)},\Delta\boldsymbol{U}_{\mathrm{s}}} \left\| \tilde{\boldsymbol{J}}_1^{(r)} \cdot \left(\hat{\boldsymbol{U}}_{\mathrm{s}} + \Delta\boldsymbol{U}_{\mathrm{s}}\right) \cdot \left(\boldsymbol{\Psi}_{\mathrm{LS}}^{(r)} + \Delta\boldsymbol{\Psi}^{(r)}\right) - \tilde{\boldsymbol{J}}_2^{(r)} \cdot \left(\hat{\boldsymbol{U}}_{\mathrm{s}} + \Delta\boldsymbol{U}_{\mathrm{s}}\right) \right\|_{\mathrm{F}}^2 + \kappa^2 \cdot \|\Delta\boldsymbol{U}_{\mathrm{s}}\|_{\mathrm{F}}^2 , \tag{5.8}$$

where κ is a regularization parameter. The solution to (5.8) is obtained iteratively by successive local linearization. However, according to [Haa97b], computing one iteration is sufficient in many cases, so that the resulting algorithm is not iterative in nature. Note that an R-D version of SLS, termed R-D SLS, which solves (5.8) across all R modes jointly, is also proposed in [Haa97b]. An advantage of the SLS algorithm is that it can also incorporate more complicated selection matrices used in, e.g., the DFT beamspace [ZHM96].

The common drawback of the above mentioned LS solutions to the shift invariance equation in (5.5) is that in the case of overlapping subarrays, they are suboptimal as they do not take into account the explicit statistics of the subspace perturbation term $\Delta\boldsymbol{U}_{\mathrm{s}}$.

Therefore, assuming a uniform R-D array according to (5.2) and maximally overlapping subarrays, we present a new least squares algorithm, termed generalized least squares (GLS) to solve (5.5) in the next section.

5.3. R-D matrix-based ESPRIT-type algorithms using GLS

In this section, we derive R-D matrix-based ESPRIT-type algorithms using GLS to solve the shift invariance equation. For simplicity, we start with the derivation of the 1-D ESPRIT-type algorithms using GLS in Section 5.3.1 and then extend these algorithms to the case of R-D parameter estimation in Section 5.3.2.

5.3.1. 1-D ESPRIT-type algorithms using GLS

The derivation of the GLS solution for 1-D Standard ESPRIT and for 1-D Unitary ESPRIT is presented in Section 5.3.1.1 and in Section 5.3.1.2, respectively.

5.3.1.1. 1-D Standard ESPRIT using GLS

For the derivation of the GLS solution for 1-D Standard ESPRIT, we consider the 1-D version (cf. Equation (3.28)) of the shift invariance equation in (5.5) given by

$$\boldsymbol{J}_1 \cdot \hat{\boldsymbol{U}}_{\mathrm{s}} \cdot \boldsymbol{\Psi} \approx \boldsymbol{J}_2 \cdot \hat{\boldsymbol{U}}_{\mathrm{s}} \tag{5.9}$$

with $\boldsymbol{\Psi} \approx \boldsymbol{T} \cdot \boldsymbol{\Phi} \cdot \boldsymbol{T}^{-1}$ and the selection matrices $\boldsymbol{J}_n,\ n = 1, 2$ are given in (3.29).

In the absence of noise, the shift invariance equation in (5.9) can be written in terms of the true signal subspace $\boldsymbol{U}_{\mathrm{s}} \in \mathbb{C}^{M \times d}$ as

$$\boldsymbol{J}_1 \cdot \boldsymbol{U}_{\mathrm{s}} \cdot \boldsymbol{\Psi} = \boldsymbol{J}_2 \cdot \boldsymbol{U}_{\mathrm{s}}. \tag{5.10}$$

Expressing $\boldsymbol{U}_{\mathrm{s}}$ in terms of its estimate $\hat{\boldsymbol{U}}_{\mathrm{s}}$ via $\boldsymbol{U}_{\mathrm{s}} = \hat{\boldsymbol{U}}_{\mathrm{s}} + \Delta\boldsymbol{U}_{\mathrm{s}}$, where $\Delta\boldsymbol{U}_{\mathrm{s}}$ denotes the subspace estimation error, (5.10) becomes

$$\boldsymbol{J}_1 \cdot \left(\hat{\boldsymbol{U}}_{\mathrm{s}} + \Delta\boldsymbol{U}_{\mathrm{s}}\right) \cdot \boldsymbol{\Psi} = \boldsymbol{J}_2 \cdot \left(\hat{\boldsymbol{U}}_{\mathrm{s}} + \Delta\boldsymbol{U}_{\mathrm{s}}\right). \tag{5.11}$$

The expansion of (5.11) yields

$$\boldsymbol{J}_1 \cdot \hat{\boldsymbol{U}}_{\mathrm{s}} \cdot \boldsymbol{\Psi} + \boldsymbol{J}_1 \cdot \Delta\boldsymbol{U}_{\mathrm{s}} \cdot \boldsymbol{\Psi} = \boldsymbol{J}_2 \cdot \hat{\boldsymbol{U}}_{\mathrm{s}} + \boldsymbol{J}_2 \cdot \Delta\boldsymbol{U}_{\mathrm{s}}, \tag{5.12}$$

which can be expressed as the following linear model in $\boldsymbol{\Psi}$:

$$\begin{aligned} \boldsymbol{J}_2 \cdot \hat{\boldsymbol{U}}_{\mathrm{s}} &= \boldsymbol{J}_1 \cdot \hat{\boldsymbol{U}}_{\mathrm{s}} \cdot \boldsymbol{\Psi} + \boldsymbol{J}_1 \cdot \Delta\boldsymbol{U}_{\mathrm{s}} \cdot \boldsymbol{\Psi} - \boldsymbol{J}_2 \cdot \Delta\boldsymbol{U}_{\mathrm{s}} \\ &= \boldsymbol{J}_1 \cdot \hat{\boldsymbol{U}}_{\mathrm{s}} \cdot \boldsymbol{\Psi} + \boldsymbol{N}_{\Delta\boldsymbol{U}_{\mathrm{s}}}, \end{aligned} \tag{5.13}$$

where

$$\boldsymbol{N}_{\Delta\boldsymbol{U}_{\mathrm{s}}} = \boldsymbol{J}_1 \cdot \Delta\boldsymbol{U}_{\mathrm{s}} \cdot \boldsymbol{\Psi} - \boldsymbol{J}_2 \cdot \Delta\boldsymbol{U}_{\mathrm{s}} \in \mathbb{C}^{(M-1) \times d} \tag{5.14}$$

is the effective perturbation as a function of the signal subspace error $\Delta\boldsymbol{U}_{\mathrm{s}}$, which is caused by the additive sensor noise $\boldsymbol{N}$ in (5.1). Next, we reorganize the matrix form in (5.13) into a linear system of vector equations by vectorizing (5.13). Using the property (1.14), the vectorization of (5.13) leads to the linear vector model

$$\begin{aligned} \hat{\boldsymbol{b}} \overset{\Delta}{=} \mathrm{vec}\left\{\boldsymbol{J}_2 \cdot \hat{\boldsymbol{U}}_{\mathrm{s}}\right\} &= \left(\boldsymbol{I}_d \otimes \boldsymbol{J}_1 \cdot \hat{\boldsymbol{U}}_{\mathrm{s}}\right) \cdot \mathrm{vec}\left\{\boldsymbol{\Psi}\right\} + \boldsymbol{n}_{\Delta\boldsymbol{u}_{\mathrm{s}}} \\ &= \hat{\boldsymbol{F}}_1 \cdot \boldsymbol{\psi} + \boldsymbol{n}_{\Delta\boldsymbol{u}_{\mathrm{s}}}, \end{aligned} \tag{5.15}$$

where we have defined $\hat{\boldsymbol{b}} \overset{\Delta}{=} \mathrm{vec}\left\{\boldsymbol{J}_2 \cdot \hat{\boldsymbol{U}}_{\mathrm{s}}\right\} \in \mathbb{C}^{(M-1)d \times 1}$ and $\hat{\boldsymbol{F}}_1 = \boldsymbol{I}_d \otimes \boldsymbol{J}_1 \cdot \hat{\boldsymbol{U}}_{\mathrm{s}} \in \mathbb{C}^{(M-1)d \times d^2}$. Moreover,

the vector $\boldsymbol{\psi} = \mathrm{vec}\{\boldsymbol{\Psi}\} \in \mathbb{C}^{d^2 \times 1}$ is the parameter vector to be estimated and $\boldsymbol{n}_{\Delta \boldsymbol{u}_\mathrm{s}} = \mathrm{vec}\left\{\boldsymbol{N}_{\Delta \boldsymbol{u}_\mathrm{s}}\right\} \in \mathbb{C}^{(M-1)d \times 1}$ is the residual error vector obtained by vectorizing (5.14) as

$$\begin{aligned} \boldsymbol{n}_{\Delta \boldsymbol{u}_\mathrm{s}} &= \left(\boldsymbol{\Psi}^\mathrm{T} \otimes \boldsymbol{J}_1\right) \cdot \mathrm{vec}\{\Delta \boldsymbol{U}_\mathrm{s}\} - (\boldsymbol{I}_d \otimes \boldsymbol{J}_2) \cdot \mathrm{vec}\{\Delta \boldsymbol{U}_\mathrm{s}\} \\ &= \left(\left(\boldsymbol{\Psi}^\mathrm{T} \otimes \boldsymbol{J}_1\right) - (\boldsymbol{I}_d \otimes \boldsymbol{J}_2)\right) \cdot \mathrm{vec}\{\Delta \boldsymbol{U}_\mathrm{s}\} \\ &= \boldsymbol{F}_2 \cdot \Delta \boldsymbol{u}_\mathrm{s}, \end{aligned} \tag{5.16}$$

where we have defined $\boldsymbol{F}_2 = \left(\boldsymbol{\Psi}^\mathrm{T} \otimes \boldsymbol{J}_1\right) - (\boldsymbol{I}_d \otimes \boldsymbol{J}_2) \in \mathbb{C}^{(M-1)d \times Md}$ and $\Delta \boldsymbol{u}_\mathrm{s} = \mathrm{vec}\{\Delta \boldsymbol{U}_\mathrm{s}\} \in \mathbb{C}^{Md \times 1}$. It is important to emphasize again that (5.15) represents a linear vector model in the parameter vector $\boldsymbol{\psi}$. In order to take the statistics of the error vector $\boldsymbol{n}_{\Delta \boldsymbol{u}_\mathrm{s}}$ into account for the estimation of $\boldsymbol{\psi}$ from (5.15), the generalized least squares (GLS) algorithm [Ame85] can be applied. Note that GLS has been shown to be the best linear unbiased estimator (BLUE) of the parameter vector [Ame85] in linear models with arbitrary error covariance matrix. In the literature, GLS is also known as Gauss-Markov estimator [Kay93] or as weighted least squares (WLS) [QHS+15].

Applying the GLS concept [Ame85, Kay93] to the linear model in (5.15), we can formulate the optimization problem

$$\hat{\boldsymbol{\psi}}_\mathrm{GLS} = \arg\min_{\boldsymbol{\psi}} \left(\hat{\boldsymbol{b}} - \hat{\boldsymbol{F}}_1 \cdot \boldsymbol{\psi}\right)^\mathrm{H} \cdot \boldsymbol{R}^{-1} \cdot \left(\hat{\boldsymbol{b}} - \hat{\boldsymbol{F}}_1 \cdot \boldsymbol{\psi}\right), \tag{5.17}$$

where $\boldsymbol{R} = \mathbb{E}\left\{\boldsymbol{n}_{\Delta \boldsymbol{u}_\mathrm{s}} \cdot \boldsymbol{n}_{\Delta \boldsymbol{u}_\mathrm{s}}^\mathrm{H}\right\} \in \mathbb{C}^{(M-1)d \times (M-1)d}$ is the covariance matrix of the effective noise vector $\boldsymbol{n}_{\Delta \boldsymbol{u}_\mathrm{s}}$ in (5.16) caused by the signal subspace estimation error $\Delta \boldsymbol{u}_\mathrm{s}$. Thus, the statistics of the signal subspace perturbation $\Delta \boldsymbol{u}_\mathrm{s}$ are explicitly taken into account for the ESPRIT-based parameter estimation. The closed-form solution to the unconstrained minimization problem in (5.17) is given by [Ame85, Kay93]

$$\hat{\boldsymbol{\psi}}_\mathrm{GLS} = \left(\hat{\boldsymbol{F}}_1^\mathrm{H} \cdot \boldsymbol{R}^{-1} \cdot \hat{\boldsymbol{F}}_1\right)^{-1} \cdot \hat{\boldsymbol{F}}_1^\mathrm{H} \cdot \boldsymbol{R}^{-1} \cdot \hat{\boldsymbol{b}} \tag{5.18}$$

assuming that $\boldsymbol{R}$ is invertible. The covariance matrix $\boldsymbol{R}$ of the perturbation $\boldsymbol{n}_{\Delta \boldsymbol{u}_\mathrm{s}}$ can be expressed as

$$\begin{aligned} \boldsymbol{R} &= \mathbb{E}\left\{\boldsymbol{n}_{\Delta \boldsymbol{u}_\mathrm{s}} \cdot \boldsymbol{n}_{\Delta \boldsymbol{u}_\mathrm{s}}^\mathrm{H}\right\} = \mathbb{E}\left\{\boldsymbol{F}_2 \cdot \Delta \boldsymbol{u}_\mathrm{s} \cdot \Delta \boldsymbol{u}_\mathrm{s}^\mathrm{H} \cdot \boldsymbol{F}_2^\mathrm{H}\right\} \\ &= \boldsymbol{F}_2 \cdot \mathbb{E}\left\{\Delta \boldsymbol{u}_\mathrm{s} \cdot \Delta \boldsymbol{u}_\mathrm{s}^\mathrm{H}\right\} \cdot \boldsymbol{F}_2^\mathrm{H} = \boldsymbol{F}_2 \cdot \boldsymbol{Q} \cdot \boldsymbol{F}_2^\mathrm{H}, \end{aligned} \tag{5.19}$$

where $\boldsymbol{Q} = \mathbb{E}\left\{\Delta \boldsymbol{u}_\mathrm{s} \cdot \Delta \boldsymbol{u}_\mathrm{s}^\mathrm{H}\right\} \in \mathbb{C}^{Md \times Md}$ denotes the covariance matrix of the vectorized signal subspace perturbation $\Delta \boldsymbol{u}_\mathrm{s} = \mathrm{vec}\{\Delta \boldsymbol{U}_\mathrm{s}\} \in \mathbb{C}^{Md \times 1}$. In order to compute $\boldsymbol{Q}$, we first need to find an approximation for $\Delta \boldsymbol{U}_\mathrm{s}$. The signal subspace estimate $\hat{\boldsymbol{U}}_\mathrm{s}$ is usually computed by applying the SVD of the measurement matrix $\boldsymbol{X}$ in (5.1) as shown in (5.4). Assuming that the noise component $\boldsymbol{N}$ is small compared to the signal component $\boldsymbol{X}_0$ in the model (5.1), we can approximate $\Delta \boldsymbol{U}_\mathrm{s}$

by applying the explicit first-order perturbation expansion of the SVD[1] from [LLV93]. Thus, $\Delta \boldsymbol{U}_\mathrm{s}$ can be approximated as

$$\Delta \boldsymbol{U}_\mathrm{s} \approx \boldsymbol{U}_\mathrm{n} \cdot \boldsymbol{U}_\mathrm{n}^\mathrm{H} \cdot \boldsymbol{N} \cdot \boldsymbol{V}_\mathrm{s} \cdot \boldsymbol{\Sigma}_\mathrm{s}^{-1}, \tag{5.20}$$

where the matrices $\boldsymbol{U}_\mathrm{n} \in \mathbb{C}^{M \times (M-d)}$, $\boldsymbol{V}_\mathrm{s} \in \mathbb{C}^{N \times d}$, and $\boldsymbol{\Sigma}_\mathrm{s} \in \mathbb{R}^{d \times d}$ are obtained from the SVD of the noise-free measurement matrix $\boldsymbol{X}_0$ according to the 1-D case of (5.3).

The main feature of (5.20) is that it is explicit in the noise realization $\boldsymbol{N}$ such that no assumptions on the noise statistics are required. Moreover, (5.20) is asymptotic in the effective SNR, i.e., the expression becomes exact for either a high SNR or a large sample size. Applying the property (1.14), the vectorized version of (5.20) is given by

$$\Delta \boldsymbol{u}_\mathrm{s} \approx \left(\boldsymbol{\Sigma}_\mathrm{s}^{-1} \cdot \boldsymbol{V}_\mathrm{s}^\mathrm{T} \otimes \boldsymbol{U}_\mathrm{n} \cdot \boldsymbol{U}_\mathrm{n}^\mathrm{H}\right) \cdot \boldsymbol{n}, \tag{5.21}$$

where $\boldsymbol{n} = \mathrm{vec}\{\boldsymbol{N}\} \in \mathbb{C}^{MN \times 1}$ is the vectorized version of $\boldsymbol{N}$. Inserting (5.21) into the expression for $\boldsymbol{Q}$ in (5.19), we obtain

$$\boldsymbol{Q} = \left(\boldsymbol{\Sigma}_\mathrm{s}^{-1} \cdot \boldsymbol{V}_\mathrm{s}^\mathrm{T} \otimes \boldsymbol{U}_\mathrm{n} \cdot \boldsymbol{U}_\mathrm{n}^\mathrm{H}\right) \cdot \boldsymbol{R}_\mathrm{nn} \cdot \left(\boldsymbol{\Sigma}_\mathrm{s}^{-1} \cdot \boldsymbol{V}_\mathrm{s}^\mathrm{T} \otimes \boldsymbol{U}_\mathrm{n} \cdot \boldsymbol{U}_\mathrm{n}^\mathrm{H}\right)^\mathrm{H}, \tag{5.22}$$

where $\boldsymbol{R}_\mathrm{nn} = \mathbb{E}\left\{\boldsymbol{n} \cdot \boldsymbol{n}^\mathrm{H}\right\} \in \mathbb{C}^{MN \times MN}$ is the covariance matrix of $\boldsymbol{n}$. Thus, the sensor noise $\boldsymbol{n}$ is only required to be zero-mean with finite second-order (SO) moments such that $\boldsymbol{R}_\mathrm{nn}$ can be computed. It is important to emphasize that $\boldsymbol{Q}$ in (5.22) can be explicitly computed for any noise distribution that satisfies these assumptions by inserting the respective covariance matrix $\boldsymbol{R}_\mathrm{nn}$.

For simplicity, we here assume the sensor noise to be zero-mean and white[2], i.e., we have $\boldsymbol{R}_\mathrm{nn} = \sigma_\mathrm{n}^2 \cdot \boldsymbol{I}_{MN}$. Then, the matrix $\boldsymbol{Q}$ reduces to

$$\begin{aligned}
\boldsymbol{Q} &= \sigma_\mathrm{n}^2 \cdot \left(\boldsymbol{\Sigma}_\mathrm{s}^{-1} \cdot \boldsymbol{V}_\mathrm{s}^\mathrm{T} \otimes \boldsymbol{U}_\mathrm{n} \cdot \boldsymbol{U}_\mathrm{n}^\mathrm{H}\right) \cdot \left(\boldsymbol{\Sigma}_\mathrm{s}^{-1} \cdot \boldsymbol{V}_\mathrm{s}^\mathrm{T} \otimes \boldsymbol{U}_\mathrm{n} \cdot \boldsymbol{U}_\mathrm{n}^\mathrm{H}\right)^\mathrm{H} \\
&= \sigma_\mathrm{n}^2 \cdot \left(\boldsymbol{\Sigma}_\mathrm{s}^{-1} \cdot \boldsymbol{V}_\mathrm{s}^\mathrm{T} \cdot \boldsymbol{V}_\mathrm{s}^* \cdot \boldsymbol{\Sigma}_\mathrm{s}^{-1}\right) \otimes \left(\boldsymbol{U}_\mathrm{n} \cdot \boldsymbol{U}_\mathrm{n}^\mathrm{H} \cdot \boldsymbol{U}_\mathrm{n} \cdot \boldsymbol{U}_\mathrm{n}^\mathrm{H}\right) \\
&= \sigma_\mathrm{n}^2 \cdot \left(\boldsymbol{\Sigma}_\mathrm{s}^{-2} \otimes \boldsymbol{U}_\mathrm{n} \cdot \boldsymbol{U}_\mathrm{n}^\mathrm{H}\right).
\end{aligned} \tag{5.23}$$

Upon inserting $\boldsymbol{Q} \in \mathbb{C}^{Md \times Md}$ from (5.23) into (5.19), we obtain

$$\boldsymbol{R} = \sigma_\mathrm{n}^2 \cdot \boldsymbol{F}_2 \cdot \left(\boldsymbol{\Sigma}_\mathrm{s}^{-2} \otimes \boldsymbol{U}_\mathrm{n} \cdot \boldsymbol{U}_\mathrm{n}^\mathrm{H}\right) \cdot \boldsymbol{F}_2^\mathrm{H}. \tag{5.24}$$

where $\boldsymbol{F}_2 \in \mathbb{C}^{(M-1)d \times Md}$ is given in (5.16). Notice that $\boldsymbol{R} \in \mathbb{C}^{(M-1)d \times (M-1)d}$ only has full rank and

[1] If the signal subspace $\boldsymbol{U}_\mathrm{s}$ is not computed via the SVD, e.g., via Krylov subspaces [SW01], or in subspace tracking [Yan95], the signal subspace error $\Delta \boldsymbol{U}_\mathrm{s}$ in (5.20) can be replaced by its corresponding first-order expansion.

[2] Note that we only require the covariance matrix $\boldsymbol{R}_\mathrm{nn}$ of $\boldsymbol{n}$ and the pseudo-covariance matrix $\boldsymbol{C}_\mathrm{nn} = \mathbb{E}\left\{\boldsymbol{n} \cdot \boldsymbol{n}^\mathrm{T}\right\} \in \mathbb{C}^{MN \times MN}$ introduced in Section 4.3.1 is not needed. Therefore, $\boldsymbol{n}$ can even be non-circular, i.e., $\boldsymbol{C}_\mathrm{nn} \neq \boldsymbol{0}_{MN}$.

is invertible[3] for $d = 1$. In this case, we have $\operatorname{rank}\left\{\boldsymbol{U}_{\mathrm{n}} \cdot \boldsymbol{U}_{\mathrm{n}}^{\mathrm{H}}\right\} = M - d = M - 1$ since $\boldsymbol{U}_{\mathrm{n}} \in \mathbb{C}^{M \times (M-d)}$ becomes $\boldsymbol{U}_{\mathrm{n}} \in \mathbb{C}^{M \times (M-1)}$. Then, we have $\operatorname{rank}\{\boldsymbol{Q}\} = (M-d)d = M-1$ and thus $\operatorname{rank}\{\boldsymbol{R}\} = (M-1)d = M-1$, where $\boldsymbol{R}$ has full rank. However, for $d > 1$, we have $\operatorname{rank}\left\{\boldsymbol{U}_{\mathrm{n}} \cdot \boldsymbol{U}_{\mathrm{n}}^{\mathrm{H}}\right\} = M - d$ such that $\operatorname{rank}\{\boldsymbol{Q}\} = (M-d)d$ and therefore $\boldsymbol{R}$ is rank-deficient with $\operatorname{rank}\{\boldsymbol{R}\} = (M-d)d < (M-1)d$ and not invertible. To circumvent the rank-deficiency of $\boldsymbol{R}$ for $d > 1$, we introduce a regularized version $\bar{\boldsymbol{Q}}$ of $\boldsymbol{Q}$, which is given by

$$\bar{\boldsymbol{Q}} = \sigma_{\mathrm{n}}^2 \cdot \left(\boldsymbol{\Sigma}_{\mathrm{s}}^{-2} \otimes \left(\boldsymbol{U}_{\mathrm{n}} \cdot \boldsymbol{U}_{\mathrm{n}}^{\mathrm{H}} + \lambda \cdot \boldsymbol{I}_M\right)\right), \tag{5.25}$$

where $\lambda > 0$ is the regularization parameter. Note that inserting $\bar{\boldsymbol{Q}}$ from (5.25) into (5.19) renders

$$\bar{\boldsymbol{R}} = \boldsymbol{F}_2 \cdot \bar{\boldsymbol{Q}} \cdot \boldsymbol{F}_2^{\mathrm{H}} = \sigma_{\mathrm{n}}^2 \cdot \boldsymbol{F}_2 \cdot \left(\boldsymbol{\Sigma}_{\mathrm{s}}^{-2} \otimes \left(\boldsymbol{U}_{\mathrm{n}} \cdot \boldsymbol{U}_{\mathrm{n}}^{\mathrm{H}} + \lambda \cdot \boldsymbol{I}_M\right)\right) \cdot \boldsymbol{F}_2^{\mathrm{H}} \tag{5.26}$$

invertible. After replacing $\boldsymbol{R}$ in (5.18) by its regularized version $\bar{\boldsymbol{R}}$ in (5.26), the choice of the parameter λ is unclear. To address this part, we can state the following result:

Theorem 5.3.1. *Assuming a ULA and maximum subarray overlap, the solution $\hat{\boldsymbol{\psi}}_{\mathrm{GLS}}$ based on $\bar{\boldsymbol{R}}$ in (5.26) is independent of the choice of the regularization parameter $\lambda > 0$.*

The proof is given in Appendix B.9.

As a consequence of Theorem 5.3.1, we can choose any value of $\lambda > 0$. A particularly convenient form is found by considering the case $\lim_{\lambda \to \infty} \hat{\boldsymbol{\psi}}_{\mathrm{GLS}}$ in (5.18) using $\bar{\boldsymbol{R}}$ from (5.26), which yields

$$\begin{aligned}
\lim_{\lambda \to \infty} \hat{\boldsymbol{\psi}}_{\mathrm{GLS}} &= \lim_{\lambda \to \infty} \left(\hat{\boldsymbol{F}}_1^{\mathrm{H}} \cdot \left(\boldsymbol{F}_2 \cdot \left(\boldsymbol{\Sigma}_{\mathrm{s}}^{-2} \otimes \left(\boldsymbol{U}_{\mathrm{n}} \cdot \boldsymbol{U}_{\mathrm{n}}^{\mathrm{H}} + \lambda \cdot \boldsymbol{I}_M\right)\right) \cdot \boldsymbol{F}_2^{\mathrm{H}}\right)^{-1} \cdot \hat{\boldsymbol{F}}_1\right)^{-1} \\
&\qquad \cdot \hat{\boldsymbol{F}}_1^{\mathrm{H}} \cdot \left(\boldsymbol{F}_2 \cdot \left(\boldsymbol{\Sigma}_{\mathrm{s}}^{-2} \otimes \left(\boldsymbol{U}_{\mathrm{n}} \cdot \boldsymbol{U}_{\mathrm{n}}^{\mathrm{H}} + \lambda \cdot \boldsymbol{I}_M\right)\right) \cdot \boldsymbol{F}_2^{\mathrm{H}}\right)^{-1} \cdot \boldsymbol{b} \\
&= \left(\hat{\boldsymbol{F}}_1^{\mathrm{H}} \cdot \boldsymbol{R}_0^{-1} \cdot \hat{\boldsymbol{F}}_1\right)^{-1} \cdot \hat{\boldsymbol{F}}_1^{\mathrm{H}} \cdot \boldsymbol{R}_0^{-1} \cdot \boldsymbol{b},
\end{aligned} \tag{5.27}$$

where we have defined $\boldsymbol{R}_0 = \boldsymbol{F}_2 \cdot \left(\boldsymbol{\Sigma}_{\mathrm{s}}^{-2} \otimes \boldsymbol{I}_M\right) \cdot \boldsymbol{F}_2^{\mathrm{H}}$ and the regularization parameter λ as well as the noise power σ_{n}^2 have canceled. Hence, the covariance matrix $\boldsymbol{R}$ in (5.18) can be replaced by $\boldsymbol{R}_0$. Note, however, that the matrix $\boldsymbol{\Psi}$, which is required to compute $\boldsymbol{F}_2$ in (5.16), and the matrix $\boldsymbol{\Sigma}_{\mathrm{s}}$ required to compute $\boldsymbol{R}_0$ are unknown. Therefore, we replace $\boldsymbol{\Sigma}_{\mathrm{s}}$ by its estimate $\hat{\boldsymbol{\Sigma}}_{\mathrm{s}}$ from (5.4) and initialize $\boldsymbol{\Psi}$ by its LS solution

$$\hat{\boldsymbol{\Psi}}_{\mathrm{LS}} = \left(\boldsymbol{J}_1 \cdot \hat{\boldsymbol{U}}_{\mathrm{s}}\right)^{+} \cdot \boldsymbol{J}_2 \cdot \hat{\boldsymbol{U}}_{\mathrm{s}}. \tag{5.28}$$

Consequently, the estimate $\hat{\boldsymbol{R}}_0$ of the matrix $\boldsymbol{R}_0$ is given by

$$\hat{\boldsymbol{R}}_0 = \hat{\boldsymbol{F}}_2 \cdot \left(\hat{\boldsymbol{\Sigma}}_{\mathrm{s}}^{-2} \otimes \boldsymbol{I}_M\right) \cdot \hat{\boldsymbol{F}}_2^{\mathrm{H}}, \tag{5.29}$$

[3]Note that only the special case $d = 1$, where $\boldsymbol{R}$ has full rank was considered in [QHS+15].

where $\hat{\boldsymbol{F}}_2 = \left(\hat{\boldsymbol{\Psi}}_{\mathrm{LS}}^{\mathrm{T}} \otimes \boldsymbol{J}_1\right) - \left(\boldsymbol{I}_d \otimes \boldsymbol{J}_2\right) \in \mathbb{C}^{(M-1)d \times Md}$. Finally, the GLS solution is obtained as

$$\hat{\boldsymbol{\psi}}_{\mathrm{GLS}} = \left(\hat{\boldsymbol{F}}_1^{\mathrm{H}} \cdot \hat{\boldsymbol{R}}_0^{-1} \cdot \hat{\boldsymbol{F}}_1\right)^{-1} \cdot \hat{\boldsymbol{F}}_1^{\mathrm{H}} \cdot \hat{\boldsymbol{R}}_0^{-1} \cdot \hat{\boldsymbol{b}} \tag{5.30}$$

and the matrix-based GLS estimate $\hat{\boldsymbol{\Psi}}_{\mathrm{GLS}}$ of the shift invariance equation in (5.9) is constructed by $\hat{\boldsymbol{\Psi}}_{\mathrm{GLS}} = \mathrm{unvec}_{d \times d}\{\hat{\boldsymbol{\psi}}_{\mathrm{GLS}}\}$. The main steps of the GLS-based 1-D Standard ESPRIT algorithm are summarized in Table 5.

Remark 1: Note that the above derivation naturally gives rise to an iterative estimation procedure, where after the initialization of $\hat{\boldsymbol{R}}_0$ in (5.29) with $\hat{\boldsymbol{\Psi}}_{\mathrm{LS}}$, alternating updates of $\hat{\boldsymbol{\Psi}}_{\mathrm{GLS}}$ and $\hat{\boldsymbol{R}}_0$ can be performed. However, the simulation results in Section 5.6 show that additional GLS iterations only improve the estimation accuracy at low SNRs or for a small sample size, where convergence is usually reached after 3-5 iterations. In the asymptotic case, i.e., at a high SNR or with a large sample size, only a single iteration is required to achieve a significant improvement in the estimation accuracy. Thus, in the asymptotic case, the presented GLS algorithm is not iterative in nature.

Remark 2: To analyze the computational complexity of 1-D Standard ESPRIT with GLS, we first calculate the complexity of 1-D Standard ESPRIT with LS and then compute the additional cost of one GLS iteration. The computational cost to obtain the LS solution $\hat{\boldsymbol{\Psi}}_{\mathrm{LS}}$ in (5.28) is dominated by the SVD of $\boldsymbol{X}$ and by the pseudo-inverse $\left(\boldsymbol{J}_1 \cdot \hat{\boldsymbol{U}}_{\mathrm{s}}\right)^{+}$, which require $\mathcal{O}(M^2N)$ for $M < N$ and $\mathcal{O}(Md^2)$ arithmetic operations, respectively. The cost of the EVD of $\hat{\boldsymbol{\Psi}}_{\mathrm{LS}}$ to extract the DOAs is $\mathcal{O}(d^3)$. Thus, 1-D Standard ESPRIT with LS requires $\mathcal{O}(M^2N + Md^2 + d^3)$ operations. The additional GLS update requires the computation of $\hat{\boldsymbol{R}}_0^{-1}$ in (5.29) and $\hat{\boldsymbol{F}}_1^{\mathrm{H}} \cdot \hat{\boldsymbol{R}}_0^{-1} \cdot \hat{\boldsymbol{F}}_1$ for (5.30), which need $\mathcal{O}(M^3d^3)$ and $\mathcal{O}(M^2d^4)$ arithmetic operations, respectively. Moreover, $\left(\hat{\boldsymbol{F}}_1^{\mathrm{H}} \cdot \hat{\boldsymbol{R}}_0^{-1} \cdot \hat{\boldsymbol{F}}_1\right)^{-1}$ and $\hat{\boldsymbol{F}}_1^{\mathrm{H}} \cdot \hat{\boldsymbol{R}}_0^{-1} \cdot \hat{\boldsymbol{b}}$ also required for (5.30) take $\mathcal{O}(d^6)$ and $\mathcal{O}(M^2d^4)$ operations. Hence, the additional cost of a single GLS update requires $\mathcal{O}(M^3d^3 + M^2d^4 + d^6)$ arithmetic operations.

Remark 3: The presented GLS algorithm has been derived for the assumption of zero-mean white noise with $\boldsymbol{R}_{\mathrm{nn}} = \sigma_{\mathrm{n}}^2 \cdot \boldsymbol{I}_{MN}$, which was used to simplify $\boldsymbol{Q}$ in (5.22). However, it is important to emphasize that a GLS solution can be derived for any other noise distribution with zero-mean and finite SO moments by inserting the respective covariance matrix $\boldsymbol{R}_{\mathrm{nn}}$ in (5.22) and computing $\boldsymbol{Q}$ and $\boldsymbol{R}$ correspondingly.

5.3.1.2. 1-D Unitary ESPRIT using GLS

In this section, we derive the GLS solution presented in the previous section for the 1-D Unitary ESPRIT algorithm [HN95], which is introduced in Section 3.4.2. The additional features of Unitary ESPRIT compared to Standard ESPRIT are the incorporation of forward-backward averaging (FBA) and the transformation into the real-valued domain to reduce the computational complexity. Both preprocessing steps are discussed in detail for the R-D case in Section 3.2.3. However, we

Algorithm 5 [SRH17a] Summary of 1-D Standard ESPRIT with GLS

1. Estimate the signal subspace $\hat{\boldsymbol{U}}_{\mathrm{s}}$ and $\hat{\boldsymbol{\Sigma}}_{\mathrm{s}}$ from the truncated SVD of $\boldsymbol{X}$.
2. Solve the shift invariance equation by means of LS via (5.28) to obtain the initialization $\hat{\boldsymbol{\Psi}}_{\mathrm{LS}}$. Form the matrix $\hat{\boldsymbol{R}}_0$ via

$$\hat{\boldsymbol{R}}_0 = \hat{\boldsymbol{F}}_2 \cdot \left(\hat{\boldsymbol{\Sigma}}_{\mathrm{s}}^{-2} \otimes \boldsymbol{I}_M\right) \cdot \hat{\boldsymbol{F}}_2^{\mathrm{H}} \tag{5.31}$$

   and obtain the GLS solution $\hat{\boldsymbol{\Psi}}_{\mathrm{GLS}}$ of the shift invariance equation by solving

$$\hat{\boldsymbol{\psi}}_{\mathrm{GLS}} = \left(\hat{\boldsymbol{F}}_1^{\mathrm{H}} \cdot \hat{\boldsymbol{R}}_0^{-1} \cdot \hat{\boldsymbol{F}}_1\right)^{-1} \cdot \hat{\boldsymbol{F}}_1^{\mathrm{H}} \cdot \hat{\boldsymbol{R}}_0^{-1} \cdot \hat{\boldsymbol{b}} \tag{5.32}$$

   In the non-asymptotic case, perform GLS iterations by repeatedly computing (5.31) and (5.32).
3. Calculate the d eigenvalues $\hat{\lambda}_i$ of $\hat{\boldsymbol{\Psi}}_{\mathrm{GLS}}$ and extract the spatial frequencies $\hat{\mu}_i,\ \forall i$ via $\hat{\mu}_i = \arg\left\{\hat{\lambda}_i\right\}$.

restate both steps for the 1-D case here again for convenience.

The augmented measurement matrix $\boldsymbol{X}^{(\mathrm{fba})}$ after applying FBA is given by

$$\boldsymbol{X}^{(\mathrm{fba})} = \begin{bmatrix} \boldsymbol{X} & \boldsymbol{\Pi}_M \cdot \boldsymbol{X}^* \cdot \boldsymbol{\Pi}_N \end{bmatrix} \in \mathbb{C}^{M \times 2N}. \tag{5.33}$$

Considering the model $\boldsymbol{X} = \boldsymbol{X}_0 + \boldsymbol{N}$ in (5.1), we can expand $\boldsymbol{X}^{(\mathrm{fba})}$ as

$$\boldsymbol{X}^{(\mathrm{fba})} = \begin{bmatrix} \boldsymbol{X}_0 & \boldsymbol{\Pi}_M \cdot \boldsymbol{X}_0^* \cdot \boldsymbol{\Pi}_N \end{bmatrix} + \begin{bmatrix} \boldsymbol{N} & \boldsymbol{\Pi}_M \cdot \boldsymbol{N}^* \cdot \boldsymbol{\Pi}_N \end{bmatrix} = \boldsymbol{X}_0^{(\mathrm{fba})} + \boldsymbol{N}^{(\mathrm{fba})}. \tag{5.34}$$

where $\boldsymbol{X}_0^{(\mathrm{fba})} \in \mathbb{C}^{M \times 2N}$ and $\boldsymbol{N}_0^{(\mathrm{fba})} \in \mathbb{C}^{M \times 2N}$ are the noise-free augmented measurement matrix and the augmented noise matrix, respectively. As $\boldsymbol{X}^{(\mathrm{fba})}$ is centro-Hermitian, i.e., it satisfies $\boldsymbol{X}^{(\mathrm{fba})} = \boldsymbol{\Pi}_M \cdot \boldsymbol{X}^{(\mathrm{fba})^*} \cdot \boldsymbol{\Pi}_{2N}$, it can be transformed into the real-valued domain by means of the transformation

$$\begin{aligned} \varphi\left(\boldsymbol{X}^{(\mathrm{fba})}\right) &= \boldsymbol{Q}_M^{\mathrm{H}} \cdot \boldsymbol{X}^{(\mathrm{fba})} \cdot \boldsymbol{Q}_{2N} \in \mathbb{R}^{M \times 2N} && (5.35)\\ &= \boldsymbol{Q}_M^{\mathrm{H}} \cdot \boldsymbol{X}_0^{(\mathrm{fba})} \cdot \boldsymbol{Q}_{2N} + \boldsymbol{Q}_M^{\mathrm{H}} \cdot \boldsymbol{N}^{(\mathrm{fba})} \cdot \boldsymbol{Q}_{2N} && (5.36)\\ &= \varphi\left(\boldsymbol{X}_0^{(\mathrm{fba})}\right) + \varphi\left(\boldsymbol{N}^{(\mathrm{fba})}\right), && (5.37) \end{aligned}$$

where $\boldsymbol{Q}_M$ and $\boldsymbol{Q}_{2N}$ are left $\boldsymbol{\Pi}$-real matrices that satisfy $\boldsymbol{\Pi}_p \cdot \boldsymbol{Q}^* = \boldsymbol{Q}$, and $\varphi\left(\boldsymbol{X}_0^{(\mathrm{fba})}\right)$ and $\varphi\left(\boldsymbol{N}^{(\mathrm{fba})}\right)$ are the transformed FBA-processed noise-free measurement matrix and the transformed FBA-processed noise, respectively.

Following Section 3.2.3, the real-valued SVD of $\varphi\left(\boldsymbol{X}_0^{(\mathrm{fba})}\right)$ and the real-valued SVD of $\varphi\left(\boldsymbol{X}^{(\mathrm{fba})}\right)$ corresponding to the 1-D versions of (5.3) and (5.4) are given by

$$\varphi\left(\boldsymbol{X}_0^{(\mathrm{fba})}\right) = \begin{bmatrix}\boldsymbol{E}_\mathrm{s}, & \boldsymbol{E}_\mathrm{n}\end{bmatrix} \cdot \begin{bmatrix}\boldsymbol{\Sigma}_\mathrm{s}^{(\varphi)} & \boldsymbol{0}_{d\times(2N-d)} \\ \boldsymbol{0}_{(M-d)\times d} & \boldsymbol{0}_{(M-d)\times(2N-d)}\end{bmatrix} \cdot \begin{bmatrix}\boldsymbol{W}_\mathrm{s}, & \boldsymbol{W}_\mathrm{n}\end{bmatrix}^\mathrm{H} \tag{5.38}$$

$$\varphi\left(\boldsymbol{X}^{(\mathrm{fba})}\right) = \begin{bmatrix}\hat{\boldsymbol{E}}_\mathrm{s}, & \hat{\boldsymbol{E}}_\mathrm{n}\end{bmatrix} \cdot \begin{bmatrix}\hat{\boldsymbol{\Sigma}}_\mathrm{s}^{(\varphi)} & \boldsymbol{0}_{d\times(2N-d)} \\ \boldsymbol{0}_{(M-d)\times d} & \hat{\boldsymbol{\Sigma}}_\mathrm{n}^{(\varphi)}\end{bmatrix} \cdot \begin{bmatrix}\hat{\boldsymbol{W}}_\mathrm{s}, & \hat{\boldsymbol{W}}_\mathrm{n}\end{bmatrix}^\mathrm{H} \tag{5.39}$$

where $\boldsymbol{E}_\mathrm{s} \in \mathbb{R}^{M\times d}$, $\boldsymbol{E}_\mathrm{n} \in \mathbb{R}^{M\times(M-d)}$, and $\boldsymbol{W}_\mathrm{s} \in \mathbb{R}^{2N\times d}$ span the real-valued signal subspace, the real-valued noise subspace, and the real-valued row space of $\varphi\left(\boldsymbol{X}_0^{(\mathrm{fba})}\right)$, and $\boldsymbol{\Sigma}_\mathrm{s}^{(\varphi)} \in \mathbb{R}^{d\times d}$ contains the d non-zero singular values on its diagonal. Moreover, $\hat{\boldsymbol{E}}_\mathrm{s}$, $\hat{\boldsymbol{E}}_\mathrm{n}$, $\hat{\boldsymbol{W}}_\mathrm{s}$, and $\hat{\boldsymbol{\Sigma}}_\mathrm{s}^{(\varphi)}$ are their estimates.

Then, the equivalent real-valued shift invariance equation to (5.9) is given by

$$\boldsymbol{K}_1 \cdot \hat{\boldsymbol{E}}_\mathrm{s} \cdot \boldsymbol{\Upsilon} \approx \boldsymbol{K}_2 \cdot \hat{\boldsymbol{E}}_\mathrm{s}, \tag{5.40}$$

where $\boldsymbol{K}_1 = 2 \cdot \mathrm{Re}\left\{\boldsymbol{Q}_{M-1}^\mathrm{H} \cdot \boldsymbol{J}_2 \cdot \boldsymbol{Q}_M\right\} \in \mathbb{R}^{(M-1)\times M}$ and $\boldsymbol{K}_2 = 2 \cdot \mathrm{Im}\left\{\boldsymbol{Q}_{M-1}^\mathrm{H} \cdot \boldsymbol{J}_2 \cdot \boldsymbol{Q}_M\right\} \in \mathbb{R}^{(M-1)\times M}$ are the transformed selection matrices. The spatial frequency estimates $\hat{\mu}_i$, $i = 1, \ldots, d$, are recovered from the eigenvalues $\hat{\omega}_i$ of the solution $\hat{\boldsymbol{\Upsilon}}$ via $\hat{\mu}_i = 2 \cdot \arctan(\hat{\omega}_i)$. Moreover, an additional reliability test that examines if the estimates $\hat{\mu}_i$ are real-valued is obtained [HN95].

For the derivation of the GLS algorithm for 1-D Unitary ESPRIT, we follow the same steps as in the previous section. As in the derivation of 1-D Standard ESPRIT with GLS, we assume a ULA and maximum subarray overlap. Thus, the corresponding real-valued version of the linear vector model in (5.15) is given by[4]

$$\hat{\boldsymbol{b}}^{(\varphi)} = \hat{\boldsymbol{F}}_1^{(\varphi)} \cdot \boldsymbol{\upsilon} + \boldsymbol{n}_{\Delta\boldsymbol{e}_\mathrm{s}}^{(\varphi)}, \tag{5.41}$$

where $\hat{\boldsymbol{b}}^{(\varphi)} = \mathrm{vec}\left\{\boldsymbol{K}_2 \cdot \hat{\boldsymbol{E}}_\mathrm{s}\right\} \in \mathbb{R}^{(M-1)d\times 1}$, $\hat{\boldsymbol{F}}_1^{(\varphi)} = \boldsymbol{I}_d \otimes \boldsymbol{K}_1 \cdot \hat{\boldsymbol{E}}_\mathrm{s} \in \mathbb{R}^{(M-1)d\times d^2}$, and $\boldsymbol{\upsilon} = \mathrm{vec}\left\{\boldsymbol{\Upsilon}\right\} \in \mathbb{R}^{d\times d}$. Moreover, $\boldsymbol{n}_{\Delta\boldsymbol{e}_\mathrm{s}}^{(\varphi)} = \boldsymbol{F}_2^{(\varphi)} \cdot \Delta\boldsymbol{e}_\mathrm{s} \in \mathbb{R}^{(M-1)d\times 1}$ represents the real-valued effective noise vector with $\boldsymbol{F}_2^{(\varphi)} = \left(\boldsymbol{\Upsilon}^\mathrm{T} \otimes \boldsymbol{K}_1\right) - \left(\boldsymbol{I}_d \otimes \boldsymbol{K}_2\right) \in \mathbb{R}^{(M-1)d\times Md}$ and $\Delta\boldsymbol{e}_\mathrm{s} = \mathrm{vec}\left\{\Delta\boldsymbol{E}_\mathrm{s}\right\} \in \mathbb{R}^{Md\times 1}$, where $\Delta\boldsymbol{E}_\mathrm{s} = \hat{\boldsymbol{E}}_\mathrm{s} - \boldsymbol{E}_\mathrm{s}$ is the real-valued signal subspace estimation error.

Then, in analogy to (5.17), the associated GLS optimization problem for the linear model in (5.41) and its solution are given by

$$\hat{\boldsymbol{\upsilon}}_\mathrm{GLS} = \arg\min_{\boldsymbol{\upsilon}} \left(\hat{\boldsymbol{b}}^{(\varphi)} - \hat{\boldsymbol{F}}_1^{(\varphi)} \cdot \boldsymbol{\upsilon}\right)^\mathrm{H} \cdot \boldsymbol{R}^{(\varphi)^{-1}} \cdot \left(\hat{\boldsymbol{b}}^{(\varphi)} - \hat{\boldsymbol{F}}_1^{(\varphi)} \cdot \boldsymbol{\upsilon}\right) \tag{5.42}$$

$$= \left(\hat{\boldsymbol{F}}_1^{(\varphi)^\mathrm{H}} \cdot \boldsymbol{R}^{(\varphi)^{-1}} \cdot \hat{\boldsymbol{F}}_1^{(\varphi)}\right)^{-1} \cdot \hat{\boldsymbol{F}}_1^{(\varphi)^\mathrm{H}} \cdot \boldsymbol{R}^{(\varphi)^{-1}} \cdot \hat{\boldsymbol{b}}^{(\varphi)}, \tag{5.43}$$

[4] Note that we have introduced the superscript (φ) to denote real-valued quantities.

where $\boldsymbol{R}^{(\varphi)} = \mathbb{E}\left\{\boldsymbol{n}_{\Delta\boldsymbol{e}_\mathrm{s}}^{(\varphi)} \cdot \boldsymbol{n}_{\Delta\boldsymbol{e}_\mathrm{s}}^{(\varphi)^\mathrm{H}}\right\} = \boldsymbol{F}_2^{(\varphi)} \cdot \boldsymbol{Q}^{(\varphi)} \cdot \boldsymbol{F}_2^{(\varphi)^\mathrm{H}} \in \mathbb{R}^{(M-1)d\times(M-1)d}$ is the covariance matrix of $\boldsymbol{n}_{\Delta\boldsymbol{e}_\mathrm{s}}^{(\varphi)}$ and $\boldsymbol{Q}^{(\varphi)} = \mathbb{E}\left\{\Delta\boldsymbol{e}_\mathrm{s} \cdot \Delta\boldsymbol{e}_\mathrm{s}^\mathrm{H}\right\} \in \mathbb{R}^{Md\times Md}$. In order to approximate $\Delta\boldsymbol{E}_\mathrm{s}$, we again apply the first-order perturbation expansion from [LLV93] to obtain

$$\begin{aligned}\Delta\boldsymbol{E}_\mathrm{s} &\approx \boldsymbol{E}_\mathrm{n} \cdot \boldsymbol{E}_\mathrm{n}^\mathrm{H} \cdot \varphi\left(\boldsymbol{N}^{(\mathrm{fba})}\right) \cdot \boldsymbol{W}_\mathrm{s} \cdot \boldsymbol{\Sigma}_\mathrm{s}^{(\varphi)^{-1}} \\ &\approx \boldsymbol{E}_\mathrm{n} \cdot \boldsymbol{E}_\mathrm{n}^\mathrm{H} \cdot \boldsymbol{Q}_M^\mathrm{H} \cdot \boldsymbol{N}^{(\mathrm{fba})} \cdot \boldsymbol{Q}_{2N} \cdot \boldsymbol{W}_\mathrm{s} \cdot \boldsymbol{\Sigma}_\mathrm{s}^{(\varphi)^{-1}},\end{aligned} \tag{5.44}$$

where we have inserted $\varphi\left(\boldsymbol{N}^{(\mathrm{fba})}\right)$ from (5.37). The vectorized version $\Delta\boldsymbol{e}_\mathrm{s}$ of (5.44) is obtained by applying property (1.14) as

$$\Delta\boldsymbol{e}_\mathrm{s} \approx \left(\boldsymbol{\Sigma}_\mathrm{s}^{(\varphi)^{-1}} \cdot \boldsymbol{W}_\mathrm{s}^\mathrm{T} \cdot \boldsymbol{Q}_{2N}^\mathrm{T} \otimes \boldsymbol{E}_\mathrm{n} \cdot \boldsymbol{E}_\mathrm{n}^\mathrm{H} \cdot \boldsymbol{Q}_M^\mathrm{H}\right) \cdot \boldsymbol{n}^{(\mathrm{fba})}, \tag{5.45}$$

where $\boldsymbol{n}^{(\mathrm{fba})} = \mathrm{vec}\left\{\boldsymbol{N}^{(\mathrm{fba})}\right\} \in \mathbb{C}^{2MN\times 1}$ is the vectorized version of the FBA-processed noise $\boldsymbol{N}^{(\mathrm{fba})}$ in (5.34). Thus, the covariance matrix $\boldsymbol{Q}^{(\varphi)} = \mathbb{E}\left\{\Delta\boldsymbol{e}_\mathrm{s} \cdot \Delta\boldsymbol{e}_\mathrm{s}^\mathrm{H}\right\}$ of $\Delta\boldsymbol{e}_\mathrm{s}$ can be expressed as

$$\boldsymbol{Q}^{(\varphi)} = \left(\boldsymbol{\Sigma}_\mathrm{s}^{(\varphi)^{-1}} \cdot \boldsymbol{W}_\mathrm{s}^\mathrm{T} \cdot \boldsymbol{Q}_{2N}^\mathrm{T} \otimes \boldsymbol{E}_\mathrm{n} \cdot \boldsymbol{E}_\mathrm{n}^\mathrm{H} \cdot \boldsymbol{Q}_M^\mathrm{H}\right) \cdot \boldsymbol{R}_{\mathrm{nn}}^{(\mathrm{fba})} \cdot \left(\boldsymbol{\Sigma}_\mathrm{s}^{(\varphi)^{-1}} \cdot \boldsymbol{W}_\mathrm{s}^\mathrm{T} \cdot \boldsymbol{Q}_{2N}^\mathrm{T} \otimes \boldsymbol{E}_\mathrm{n} \cdot \boldsymbol{E}_\mathrm{n}^\mathrm{H} \cdot \boldsymbol{Q}_M^\mathrm{H}\right)^\mathrm{H}, \tag{5.46}$$

where $\boldsymbol{R}_{\mathrm{nn}}^{(\mathrm{fba})} = \mathbb{E}\left\{\boldsymbol{n}^{(\mathrm{fba})} \cdot \boldsymbol{n}^{(\mathrm{fba})^\mathrm{H}}\right\} \in \mathbb{C}^{2MN\times 2MN}$ is the covariance matrix of $\boldsymbol{n}^{(\mathrm{fba})}$. Note that in (4.30) of Section 4.3.2, we have already computed $\boldsymbol{R}_{\mathrm{nn}}^{(\mathrm{fba})}$ as a function of $\boldsymbol{n} = \mathrm{vec}\{\boldsymbol{N}\} \in \mathbb{C}^{MN\times 1}$, where $\boldsymbol{N}$ is the original noise contribution in (5.1). Therefore, $\boldsymbol{R}_{\mathrm{nn}}^{(\mathrm{fba})}$ is given by

$$\boldsymbol{R}_{\mathrm{nn}}^{(\mathrm{fba})} = \begin{bmatrix} \boldsymbol{R}_{\mathrm{nn}} & \boldsymbol{C}_{\mathrm{nn}} \cdot \boldsymbol{\Pi}_{MN} \\ \boldsymbol{\Pi}_{MN} \cdot \boldsymbol{C}_{\mathrm{nn}}^* & \boldsymbol{\Pi}_{MN} \cdot \boldsymbol{R}_{\mathrm{nn}}^* \cdot \boldsymbol{\Pi}_{MN} \end{bmatrix}, \tag{5.47}$$

where $\boldsymbol{R}_{\mathrm{nn}} = \mathbb{E}\left\{\boldsymbol{n} \cdot \boldsymbol{n}^\mathrm{H}\right\} \in \mathbb{C}^{MN\times MN}$ and $\boldsymbol{C}_{\mathrm{nn}} = \mathbb{E}\left\{\boldsymbol{n} \cdot \boldsymbol{n}^\mathrm{T}\right\} \in \mathbb{C}^{MN\times MN}$ are the covariance matrix and the pseudo-covariance matrix of $\boldsymbol{n}$.

For simplicity, we assume zero-mean circularly symmetric white sensor noise[5] with $\boldsymbol{R}_{\mathrm{nn}} = \sigma_\mathrm{n}^2 \cdot \boldsymbol{I}_{MN}$ and $\boldsymbol{C}_{\mathrm{nn}} = \boldsymbol{0}_{MN}$. In this case, $\boldsymbol{R}_{\mathrm{nn}}^{(\mathrm{fba})}$ in (5.47) simplifies to $\boldsymbol{R}_{\mathrm{nn}}^{(\mathrm{fba})} = \sigma_\mathrm{n}^2 \cdot \boldsymbol{I}_{2MN}$. With this assumption, the matrix $\boldsymbol{Q}^{(\varphi)}$ simplifies to

$$\begin{aligned}\boldsymbol{Q}^{(\varphi)} &= \sigma_\mathrm{n}^2 \cdot \left(\boldsymbol{\Sigma}_\mathrm{s}^{(\varphi)^{-1}} \cdot \boldsymbol{W}_\mathrm{s}^\mathrm{T} \cdot \boldsymbol{Q}_{2N}^\mathrm{T} \otimes \boldsymbol{E}_\mathrm{n} \cdot \boldsymbol{E}_\mathrm{n}^\mathrm{H} \cdot \boldsymbol{Q}_M^\mathrm{H}\right) \cdot \left(\boldsymbol{\Sigma}_\mathrm{s}^{(\varphi)^{-1}} \cdot \boldsymbol{W}_\mathrm{s}^\mathrm{T} \cdot \boldsymbol{Q}_{2N}^\mathrm{T} \otimes \boldsymbol{E}_\mathrm{n} \cdot \boldsymbol{E}_\mathrm{n}^\mathrm{H} \cdot \boldsymbol{Q}_M^\mathrm{H}\right)^\mathrm{H} \\ &= \sigma_\mathrm{n}^2 \cdot \left(\boldsymbol{\Sigma}_\mathrm{s}^{(\varphi)^{-1}} \cdot \boldsymbol{W}_\mathrm{s}^\mathrm{T} \cdot \boldsymbol{Q}_{2N}^\mathrm{T} \cdot \boldsymbol{Q}_{2N}^* \cdot \boldsymbol{W}_\mathrm{s}^* \cdot \boldsymbol{\Sigma}_\mathrm{s}^{(\varphi)^{-1}}\right) \otimes \left(\boldsymbol{E}_\mathrm{n} \cdot \boldsymbol{E}_\mathrm{n}^\mathrm{H} \cdot \boldsymbol{Q}_M^\mathrm{H} \cdot \boldsymbol{Q}_M \cdot \boldsymbol{E}_\mathrm{n} \cdot \boldsymbol{E}_\mathrm{n}^\mathrm{H}\right) \\ &= \sigma_\mathrm{n}^2 \cdot \left(\boldsymbol{\Sigma}_\mathrm{s}^{(\varphi)^{-2}} \otimes \boldsymbol{E}_\mathrm{n} \cdot \boldsymbol{E}_\mathrm{n}^\mathrm{H}\right).\end{aligned} \tag{5.48}$$

[5] In contrast to (5.23), we additionally require the noise to be circularly symmetric, i.e., $\boldsymbol{C}_{\mathrm{nn}} = \boldsymbol{0}_{MN}$, such that $\boldsymbol{R}_{\mathrm{nn}}^{(\mathrm{fba})}$ corresponds to white noise with $\boldsymbol{R}_{\mathrm{nn}}^{(\mathrm{fba})} = \sigma_\mathrm{n}^2 \cdot \boldsymbol{I}_{2MN}$. This is due to the FBA preprocessing step.

Algorithm 6 [SRH17a] Summary of 1-D Unitary ESPRIT with GLS

1. Estimate the real-valued signal subspace $\hat{\boldsymbol{E}}_{\mathrm{s}}$ and $\hat{\boldsymbol{\Sigma}}_{\mathrm{s}}^{(\varphi)}$ from the truncated SVD of $\varphi\big(\boldsymbol{X}^{(\mathrm{fba})}\big)$.

2. Solve the shift invariance equation by means of LS via (5.51) to obtain the initialization $\hat{\boldsymbol{\Upsilon}}_{\mathrm{LS}}$. Form the matrix $\hat{\boldsymbol{R}}_0^{(\varphi)}$ via

$$\hat{\boldsymbol{R}}_0^{(\varphi)} = \hat{\boldsymbol{F}}_2^{(\varphi)} \cdot \Big(\hat{\boldsymbol{\Sigma}}_{\mathrm{s}}^{(\varphi)^{-2}} \otimes \boldsymbol{I}_M\Big) \cdot \hat{\boldsymbol{F}}_2^{(\varphi)^{\mathrm{H}}} \tag{5.52}$$

   and obtain the GLS solution $\hat{\boldsymbol{\Upsilon}}_{\mathrm{GLS}}$ of the real-valued shift invariance equation by solving

$$\hat{\boldsymbol{v}}_{\mathrm{GLS}} = \Big(\hat{\boldsymbol{F}}_1^{(\varphi)^{\mathrm{H}}} \cdot \hat{\boldsymbol{R}}_0^{(\varphi)^{-1}} \cdot \hat{\boldsymbol{F}}_1^{(\varphi)}\Big)^{-1} \cdot \hat{\boldsymbol{F}}_1^{(\varphi)^{\mathrm{H}}} \cdot \hat{\boldsymbol{R}}_0^{(\varphi)^{-1}} \cdot \hat{\boldsymbol{b}}^{(\varphi)}. \tag{5.53}$$

   In the non-asymptotic case, perform GLS iterations by repeatedly computing (5.52) and (5.53).

3. Calculate the d eigenvalues $\hat{\omega}_i$ of $\hat{\boldsymbol{\Upsilon}}_{\mathrm{GLS}}$ and extract the spatial frequencies $\hat{\mu}_i,\ \forall i$ via $\hat{\mu}_i = 2 \cdot \arctan(\hat{\omega}_i)$.

Note that (5.48) resembles (5.23) and we can follow the same steps as in the derivation of GLS for 1-D Standard ESPRIT including the regularization of $\boldsymbol{Q}^{(\varphi)}$ and $\boldsymbol{R}^{(\varphi)}$. Note that a corresponding real-valued version of Theorem 5.3.1 assuming a ULA and maximum subarray overlap can be proven straightforwardly.

Finally, the real-valued GLS solution to the shift invariance equation in (5.40) is given by

$$\hat{\boldsymbol{v}}_{\mathrm{GLS}} = \Big(\hat{\boldsymbol{F}}_1^{(\varphi)^{\mathrm{H}}} \cdot \hat{\boldsymbol{R}}_0^{(\varphi)^{-1}} \cdot \hat{\boldsymbol{F}}_1^{(\varphi)}\Big)^{-1} \cdot \hat{\boldsymbol{F}}_1^{(\varphi)^{\mathrm{H}}} \cdot \hat{\boldsymbol{R}}_0^{(\varphi)^{-1}} \cdot \hat{\boldsymbol{b}}^{(\varphi)}. \tag{5.49}$$

Note that $\hat{\boldsymbol{R}}_0^{(\varphi)}$ in (5.49) can be computed similarly to (5.29) as

$$\hat{\boldsymbol{R}}_0^{(\varphi)} = \hat{\boldsymbol{F}}_2^{(\varphi)} \cdot \Big(\hat{\boldsymbol{\Sigma}}_{\mathrm{s}}^{(\varphi)^{-2}} \otimes \boldsymbol{I}_M\Big) \cdot \hat{\boldsymbol{F}}_2^{(\varphi)^{\mathrm{H}}}, \tag{5.50}$$

where we have replaced $\boldsymbol{\Sigma}_{\mathrm{s}}^{(\varphi)^{-2}}$ by its estimate $\hat{\boldsymbol{\Sigma}}_{\mathrm{s}}^{(\varphi)^{-2}}$ from (5.39). Furthermore, we have initialized the matrix $\hat{\boldsymbol{F}}_2^{(\varphi)} = \big(\hat{\boldsymbol{\Upsilon}}_{\mathrm{LS}}^{\mathrm{T}} \otimes \boldsymbol{K}_1\big) - \big(\boldsymbol{I}_d \otimes \boldsymbol{K}_2\big)$ with the real-valued LS solution

$$\hat{\boldsymbol{\Upsilon}}_{\mathrm{LS}} = \big(\boldsymbol{K}_1 \cdot \hat{\boldsymbol{E}}_{\mathrm{s}}\big)^{+} \cdot \boldsymbol{K}_2 \cdot \hat{\boldsymbol{E}}_{\mathrm{s}}. \tag{5.51}$$

Eventually, from (5.49), the GLS estimate $\hat{\boldsymbol{\Upsilon}}_{\mathrm{GLS}}$ to (5.40) is obtained by $\hat{\boldsymbol{\Upsilon}}_{\mathrm{GLS}} = \mathrm{unvec}_{d\times d}\{\hat{\boldsymbol{v}}_{\mathrm{GLS}}\}$. Note again that all the operations required for the GLS-based Unitary ESPRIT algorithm are entirely real-valued. The 1-D Unitary ESPRIT algorithm with GLS is summarized in Table 6.

5.3.2. R-D ESPRIT-type algorithms using GLS

In this section, we extend the GLS solutions for 1-D Standard ESPRIT from Section 5.3.1.1 and 1-D Unitary ESPRIT from Section 5.3.1.2 to the case of R-D parameter estimation. As in the R-D matrix case, the R shift invariance equations in (5.5) are solved independently, we can apply the GLS solution from the 1-D case in Section 5.3.1 to all R modes individually. We start with the R-D Standard ESPRIT algorithm using GLS in Section 5.3.3 and then present the R-D Unitary ESPRIT algorithm using GLS in Section 5.3.4.

5.3.3. R-D Standard ESPRIT using GLS

In the case of GLS for R-D Standard ESPRIT, we consider the R shift invariance equations in (5.5). The corresponding R-D generalization of the GLS solution is given by

$$\hat{\boldsymbol{\psi}}_{\mathrm{GLS}}^{(r)} = \left(\hat{\boldsymbol{F}}_1^{(r)\mathrm{H}} \cdot \hat{\boldsymbol{R}}_0^{(r)^{-1}} \cdot \hat{\boldsymbol{F}}_1^{(r)}\right)^{-1} \cdot \hat{\boldsymbol{F}}_1^{(r)\mathrm{H}} \cdot \hat{\boldsymbol{R}}_0^{(r)^{-1}} \cdot \hat{\boldsymbol{b}}^{(r)}, \tag{5.54}$$

where $\hat{\boldsymbol{F}}_1^{(r)} = \boldsymbol{I}_d \otimes \tilde{\boldsymbol{J}}_1^{(r)} \cdot \hat{\boldsymbol{U}}_{\mathrm{s}} \in \mathbb{C}^{\frac{M}{M_r}(M_r-1)d \times d^2}$, $\hat{\boldsymbol{b}}^{(r)} = \mathrm{vec}\left\{\tilde{\boldsymbol{J}}_2^{(r)} \cdot \hat{\boldsymbol{U}}_{\mathrm{s}}\right\} \in \mathbb{C}^{\frac{M}{M_r}(M_r-1)d \times 1}$ and the selection matrices $\tilde{\boldsymbol{J}}_n^{(r)} \in \mathbb{R}^{\frac{M}{M_r}(M_r-1) \times M}$ for $n = 1, 2$ are given as in (5.5). Moreover, the matrix $\hat{\boldsymbol{R}}_0^{(r)}$ is computed as

$$\hat{\boldsymbol{R}}_0^{(r)} = \hat{\boldsymbol{F}}_2^{(r)} \cdot \left(\hat{\boldsymbol{\Sigma}}_{\mathrm{s}}^{-2} \otimes \boldsymbol{I}_M\right) \cdot \hat{\boldsymbol{F}}_2^{(r)\mathrm{H}} \in \mathbb{C}^{\frac{M}{M_r}(M_r-1)d \times \frac{M}{M_r}(M_r-1)d}, \tag{5.55}$$

where we have initialized the matrix $\hat{\boldsymbol{F}}_2^{(r)} = \left(\hat{\boldsymbol{\Psi}}_{\mathrm{LS}}^{(r)^{\mathrm{T}}} \otimes \tilde{\boldsymbol{J}}_1^{(r)}\right) - \left(\boldsymbol{I}_d \otimes \tilde{\boldsymbol{J}}_2^{(r)}\right) \in \mathbb{C}^{\frac{M}{M_r}(M_r-1)d \times M_r d}$ with the LS solution

$$\hat{\boldsymbol{\Psi}}_{\mathrm{LS}}^{(r)} = \left(\tilde{\boldsymbol{J}}_1^{(r)} \cdot \hat{\boldsymbol{U}}_{\mathrm{s}}\right)^{+} \cdot \tilde{\boldsymbol{J}}_2^{(r)} \cdot \hat{\boldsymbol{U}}_{\mathrm{s}}. \tag{5.56}$$

The matrix-based GLS estimate $\hat{\boldsymbol{\Psi}}_{\mathrm{GLS}}^{(r)}$ of the shift invariance equation in (5.5) in the r-th mode is obtained via $\hat{\boldsymbol{\Psi}}_{\mathrm{GLS}}^{(r)} = \mathrm{unvec}_{d \times d}\{\hat{\boldsymbol{\psi}}_{\mathrm{GLS}}^{(r)}\}$. Then, after performing a joint EVD of the R estimates $\hat{\boldsymbol{\Psi}}_{\mathrm{GLS}}^{(r)}$ (e.g., via [FG06]), the spatial frequency estimates $\hat{\mu}_i^{(r)}$, $i = 1, \ldots, d$, are extracted from the eigenvalues $\hat{\lambda}_i^{(r)}$ of $\hat{\boldsymbol{\Psi}}_{\mathrm{GLS}}^{(r)}$ via $\hat{\mu}_i^{(r)} = \arg\left\{\hat{\lambda}_i^{(r)}\right\}$. The GLS-based R-D Standard ESPRIT algorithm is summarized in Table 7.

Algorithm 7 Summary of R-D Standard ESPRIT with GLS

1. Estimate the signal subspace $\hat{\boldsymbol{U}}_{\mathrm{s}}$ and $\hat{\boldsymbol{\Sigma}}_{\mathrm{s}}$ from the truncated SVD of $\boldsymbol{X}$.
2. Solve the shift invariance equation by means of LS via (5.56) to obtain the initialization $\hat{\boldsymbol{\Psi}}_{\mathrm{LS}}^{(r)}$ in the r-th mode for $r = 1, 2, \ldots, R$. Form the matrix $\hat{\boldsymbol{R}}_0^{(r)}$ via

$$\hat{\boldsymbol{R}}_0^{(r)} = \hat{\boldsymbol{F}}_2^{(r)} \cdot \left(\hat{\boldsymbol{\Sigma}}_{\mathrm{s}}^{-2} \otimes \boldsymbol{I}_M\right) \cdot \hat{\boldsymbol{F}}_2^{(r)\mathrm{H}} \in \mathbb{C}^{\frac{M}{M_r}(M_r-1)d \times \frac{M}{M_r}(M_r-1)d} \tag{5.57}$$

   and obtain the GLS solution $\hat{\boldsymbol{\Psi}}_{\mathrm{GLS}}^{(r)}$ of the shift invariance equation by solving

$$\hat{\boldsymbol{\psi}}_{\mathrm{GLS}}^{(r)} = \left(\hat{\boldsymbol{F}}_1^{(r)\mathrm{H}} \cdot \hat{\boldsymbol{R}}_0^{(r)^{-1}} \cdot \hat{\boldsymbol{F}}_1^{(r)}\right)^{-1} \cdot \hat{\boldsymbol{F}}_1^{(r)\mathrm{H}} \cdot \hat{\boldsymbol{R}}_0^{(r)^{-1}} \cdot \hat{\boldsymbol{b}}^{(r)}. \tag{5.58}$$

   In the non-asymptotic case, perform GLS iterations by repeatedly computing (5.57) and (5.58).
3. Compute the eigenvalues $\hat{\lambda}_i^{(r)}$ for $i = 1, \ldots, d$ of $\hat{\boldsymbol{\Psi}}_{\mathrm{GLS}}^{(r)}$ jointly for all $r = 1, 2, \ldots, R$, e.g., via the joint diagonalization scheme proposed in [FG06]. Recover the correctly paired frequencies $\hat{\mu}_i^{(r)}$ via $\hat{\mu}_i^{(r)} = \arg\left\{\hat{\lambda}_i^{(r)}\right\}$.

5.3.4. R-D Unitary ESPRIT using GLS

In the case of GLS for R-D Unitary ESPRIT, the generalization of the real-valued GLS solution in Section 5.3.1.2 to the R-D case is given by

$$\hat{\boldsymbol{\upsilon}}_{\mathrm{GLS}}^{(r)} = \left(\hat{\boldsymbol{F}}_1^{(r)^{(\varphi)\mathrm{H}}} \cdot \hat{\boldsymbol{R}}_0^{(r)^{(\varphi)^{-1}}} \cdot \hat{\boldsymbol{F}}_1^{(r)^{(\varphi)}}\right)^{-1} \cdot \hat{\boldsymbol{F}}_1^{(r)^{(\varphi)\mathrm{H}}} \cdot \hat{\boldsymbol{R}}_0^{(r)^{(\varphi)^{-1}}} \cdot \hat{\boldsymbol{b}}^{(r)^{(\varphi)}}, \tag{5.59}$$

where $\hat{\boldsymbol{F}}_1^{(r)^{(\varphi)}} = \boldsymbol{I}_d \otimes \tilde{\boldsymbol{K}}_1^{(r)} \cdot \hat{\boldsymbol{E}}_{\mathrm{s}} \in \mathbb{R}^{\frac{M}{M_r}(M_r-1)d \times d^2}$, $\hat{\boldsymbol{b}}^{(r)^{(\varphi)}} = \mathrm{vec}\left\{\tilde{\boldsymbol{K}}_2^{(r)} \cdot \hat{\boldsymbol{E}}_{\mathrm{s}}\right\} \in \mathbb{R}^{\frac{M}{M_r}(M_r-1)d \times 1}$, and the selection matrices $\tilde{\boldsymbol{K}}_n^{(r)} \in \mathbb{R}^{\frac{M}{M_r}(M_r-1) \times M}$ for $n = 1, 2$ are given in (3.36) and (3.37). The matrix $\hat{\boldsymbol{R}}_0^{(r)^{(\mathrm{fba})}}$ can be computed as

$$\hat{\boldsymbol{R}}_0^{(r)^{(\varphi)}} = \hat{\boldsymbol{F}}_2^{(r)^{(\varphi)}} \cdot \left(\hat{\boldsymbol{\Sigma}}_{\mathrm{s}}^{(\varphi)^{-2}} \otimes \boldsymbol{I}_M\right) \cdot \hat{\boldsymbol{F}}_2^{(r)^{(\varphi)\mathrm{H}}} \in \mathbb{R}^{\frac{M}{M_r}(M_r-1)d \times \frac{M}{M_r}(M_r-1)d}, \tag{5.60}$$

where we have initialized the matrix $\hat{\boldsymbol{F}}_2^{(r)^{(\varphi)}} = \left(\hat{\boldsymbol{\Upsilon}}_{\mathrm{LS}}^{(r)\mathrm{T}} \otimes \tilde{\boldsymbol{K}}_1^{(r)}\right) - \left(\boldsymbol{I}_d \otimes \tilde{\boldsymbol{K}}_2^{(r)}\right) \in \mathbb{R}^{\frac{M}{M_r}(M_r-1)d \times M_r d}$ with the real-valued LS solution

$$\hat{\boldsymbol{\Upsilon}}_{\mathrm{LS}}^{(r)} = \left(\tilde{\boldsymbol{K}}_1^{(r)} \cdot \hat{\boldsymbol{E}}_{\mathrm{s}}\right)^{+} \cdot \tilde{\boldsymbol{K}}_2^{(r)} \cdot \hat{\boldsymbol{E}}_{\mathrm{s}}. \tag{5.61}$$

Algorithm 8 Summary of R-D Unitary ESPRIT with GLS

1. Estimate the real-valued signal subspace $\hat{\boldsymbol{E}}_{\mathrm{s}}$ and $\hat{\boldsymbol{\Sigma}}_{\mathrm{s}}^{(\varphi)}$ from the truncated SVD of $\varphi\big(\boldsymbol{X}^{(\mathrm{fba})}\big)$.

2. Solve the shift invariance equation by means of LS via (5.61) to obtain the initialization $\hat{\boldsymbol{\Upsilon}}_{\mathrm{LS}}^{(r)}$ in the r-th mode for $r = 1, 2, \ldots, R$. Form the matrix $\hat{\boldsymbol{R}}_0^{(r)^{(\varphi)}}$ via

$$\hat{\boldsymbol{R}}_0^{(r)^{(\varphi)}} = \hat{\boldsymbol{F}}_2^{(r)^{(\varphi)}} \cdot \left(\hat{\boldsymbol{\Sigma}}_{\mathrm{s}}^{(\varphi)^{-2}} \otimes \boldsymbol{I}_M\right) \cdot \hat{\boldsymbol{F}}_2^{(r)^{(\varphi)^{\mathrm{H}}}} \in \mathbb{R}^{\frac{M}{M_r}(M_r-1)d \times \frac{M}{M_r}(M_r-1)d} \tag{5.62}$$

   and obtain the GLS solution $\hat{\boldsymbol{\Upsilon}}_{\mathrm{GLS}}^{(r)}$ of the shift invariance equation by solving

$$\hat{\boldsymbol{v}}_{\mathrm{GLS}}^{(r)} = \left(\hat{\boldsymbol{F}}_1^{(r)^{(\varphi)^{\mathrm{H}}}} \cdot \hat{\boldsymbol{R}}_0^{(r)^{(\varphi)^{-1}}} \cdot \hat{\boldsymbol{F}}_1^{(r)^{(\varphi)}}\right)^{-1} \cdot \hat{\boldsymbol{F}}_1^{(r)^{(\varphi)^{\mathrm{H}}}} \cdot \hat{\boldsymbol{R}}_0^{(r)^{(\varphi)^{-1}}} \cdot \hat{\boldsymbol{b}}^{(r)^{(\varphi)}}. \tag{5.63}$$

   In the non-asymptotic case, perform GLS iterations by repeatedly computing (5.62) and (5.63).

3. Compute the eigenvalues $\hat{\omega}_i^{(r)}$ for $i = 1, 2, \ldots, d$ of $\hat{\boldsymbol{\Upsilon}}_{\mathrm{GLS}}^{(r)}$ jointly for all $r = 1, \ldots, R$, e.g., via the joint diagonalization scheme proposed in [FG06] or via the Simultaneous Schur Decomposition proposed in [HN98]. Recover the correctly paired frequencies $\hat{\mu}_i^{(r)}$ via $\hat{\mu}_i^{(r)} = 2 \cdot \arctan(\hat{\omega}_i^{(r)})$.

Eventually, from the solution of (5.59), the GLS estimate $\hat{\boldsymbol{\Upsilon}}_{\mathrm{GLS}}^{(r)}$ in the r-th mode is obtained by $\hat{\boldsymbol{\Upsilon}}_{\mathrm{GLS}}^{(r)} = \mathrm{unvec}_{d\times d}\left\{\hat{\boldsymbol{v}}_{\mathrm{GLS}}^{(r)}\right\}$. Then, after performing a joint EVD of the R estimates $\hat{\boldsymbol{\Upsilon}}_{\mathrm{GLS}}^{(r)}$ (e.g., via [FG06] or [HN98]), the spatial frequency estimates $\hat{\mu}_i^{(r)}$, $i = 1, \ldots, d$, are extracted from the eigenvalues $\hat{\omega}_i^{(r)}$ of $\hat{\boldsymbol{\Upsilon}}_{\mathrm{GLS}}^{(r)}$ via $\hat{\mu}_i^{(r)} = 2 \cdot \arctan(\hat{\omega}_i^{(r)})$. The R-D Unitary ESPRIT algorithm with GLS is summarized in Table 8.

5.4. Performance of R-D matrix-based ESPRIT-type algorithms using GLS

In this section, we present a first-order performance analysis based on the R-D performance framework in [RHD14] for the matrix-based R-D ESPRIT-type algorithms using GLS derived in Section 5.3. A detailed motivation and a literature overview of performance analysis expressions is provided in Section 4.1 of Chapter 4. As discussed in Section 4.1.1, the R-D performance analysis framework in [RHD14], which is based on the 1-D results in [LLV93], is more general than the performance analysis framework in [Bri75, RH89a]. Therefore, we apply the first-order R-D performance framework for R-D ESPRIT-type algorithms using LS from [RHD14] and extend the

performance analysis results to GLS-based R-D ESPRIT-type algorithms. Note that since the analytical expressions from [RHD14] are asymptotic in the high effective SNR, i.e., the results become exact for either high SNRs or a large sample size N, we only need to consider a single iteration of GLS, which is sufficient in the asymptotic case. In Section 5.4.1, we derive the analytical expressions for the parameter estimation error and the MSE of R-D Standard ESPRIT using GLS and in Section 5.4.2, we derive the corresponding analytical expressions for R-D Unitary ESPRIT with GLS.

5.4.1. Performance of R-D Standard ESPRIT using GLS

The results for the parameter estimation error of R-D Standard ESPRIT using GLS are stated in Theorem 5.4.1 as follows:

Theorem 5.4.1. *A first-order approximation of the estimation error of R-D Standard ESPRIT with GLS for the i-th spatial frequency in the r-th mode is given by*

$$\Delta\mu_i^{(r)} \approx \operatorname{Im}\left\{\boldsymbol{r}_{i,\mathrm{GLS}}^{(r)^{\mathrm{T}}} \cdot \operatorname{vec}\left\{\Delta\boldsymbol{U}_{\mathrm{s}}\right\}\right\} = \operatorname{Im}\left\{\boldsymbol{r}_{i,\mathrm{GLS}}^{(r)^{\mathrm{T}}} \cdot \boldsymbol{W}_{\mathrm{mat}} \cdot \boldsymbol{n}\right\}, \tag{5.64}$$

where $\boldsymbol{W}_{\mathrm{mat}}$ *is given in* (4.15) *in Section 4.3.1 and*

$$\boldsymbol{r}_{i,\mathrm{GLS}}^{(r)^{\mathrm{T}}} = -\frac{1}{\lambda_i^{(r)}} \cdot \left(\boldsymbol{q}_i^{\mathrm{T}} \otimes \boldsymbol{p}_i^{\mathrm{T}}\right) \cdot \left(\boldsymbol{F}_1^{(r)^{\mathrm{H}}} \cdot \boldsymbol{R}_0^{(r)^{-1}} \cdot \boldsymbol{F}_1^{(r)}\right)^{-1} \cdot \boldsymbol{F}_1^{(r)^{\mathrm{H}}} \cdot \boldsymbol{R}_0^{(r)^{-1}} \cdot \boldsymbol{F}_2^{(r)}, \tag{5.65}$$

where $\lambda_i^{(r)} = \mathrm{e}^{\mathrm{j}\mu_i^{(r)}}$*, and* $\boldsymbol{q}_i$ *and* $\boldsymbol{p}_i^{\mathrm{T}}$ *are the i-th column and the i-th row vector of the matrices* $\boldsymbol{Q}$ *and* $\boldsymbol{P} = \boldsymbol{Q}^{-1}$ *obtained from the eigendecomposition* $\boldsymbol{\Psi}^{(r)} = \boldsymbol{Q} \cdot \boldsymbol{\Lambda}^{(r)} \cdot \boldsymbol{Q}^{-1}$ *in the r-th mode.*

The proof of Theorem (5.4.1) is shown in Appendix B.12. Note that the GLS-based error expression for R-D Standard ESPRIT in (5.64) has the same form as the LS-based error expression for R-D Standard ESPRIT in (4.13). Therefore, the MSE expression for (5.64) can be expressed analogously to (4.17). However, the GLS algorithm for R-D Standard ESPRIT is only derived for zero-mean white noise $\boldsymbol{n} = \operatorname{vec}\{\boldsymbol{N}\} \in \mathbb{C}^{MN\times 1}$ with $\boldsymbol{R}_{\mathrm{nn}} = \mathbb{E}\left\{\boldsymbol{n} \cdot \boldsymbol{n}^{\mathrm{H}}\right\} = \sigma_{\mathrm{n}}^2 \cdot \boldsymbol{I}_{MN}$. The pseudo-covariance matrix $\boldsymbol{C}_{\mathrm{nn}} = \mathbb{E}\left\{\boldsymbol{n} \cdot \boldsymbol{n}^{\mathrm{T}}\right\}$ can be arbitrary. Thus, the MSE of R-D Standard ESPRIT with GLS for the i-th spatial frequency in the r-th mode is given by

$$\mathbb{E}\left\{\left(\Delta\mu_i^{(r)}\right)^2\right\} = \frac{1}{2} \cdot \left(\sigma_{\mathrm{n}}^2 \cdot \left\|\boldsymbol{z}_{i,\mathrm{GLS}}^{(r)}\right\|_2^2 - \operatorname{Re}\left\{\boldsymbol{z}_{i,\mathrm{GLS}}^{(r)^{\mathrm{T}}} \cdot \boldsymbol{C}_{\mathrm{nn}}^{\mathrm{T}} \cdot \boldsymbol{z}_{i,\mathrm{GLS}}^{(r)}\right\}\right), \tag{5.66}$$

where $\boldsymbol{z}_{i,\mathrm{GLS}}^{(r)} = \boldsymbol{W}_{\mathrm{mat}}^{\mathrm{T}} \cdot \boldsymbol{r}_{i,\mathrm{GLS}}^{(r)} \in \mathbb{C}^{MN\times 1}$ and $\boldsymbol{C}_{\mathrm{nn}} = \mathbb{E}\left\{\boldsymbol{n} \cdot \boldsymbol{n}^{\mathrm{T}}\right\} \in \mathbb{C}^{MN\times MN}$ is the pseudo-covariance matrix of $\boldsymbol{n}$.

In the special case of zero-mean circularly symmetric white noise with $\boldsymbol{R}_{\mathrm{nn}} = \sigma_{\mathrm{n}}^2 \cdot \boldsymbol{I}_{MN}$ and

$\boldsymbol{C}_{\mathrm{nn}} = \boldsymbol{0}_{MN}$, the MSE of R-D Standard ESPRIT with GLS in (5.66) simplifies to

$$\mathbb{E}\left\{\left(\Delta\mu_i^{(r)}\right)^2\right\} = \frac{\sigma_{\mathrm{n}}^2}{2} \cdot \left\|\boldsymbol{z}_{i,\mathrm{GLS}}^{(r)}\right\|_2^2. \tag{5.67}$$

5.4.2. Performance of R-D Unitary ESPRIT using GLS

This section presents the analytical expressions for the parameter estimation error and the MSE of R-D Unitary ESPRIT with GLS. Recall that Unitary ESPRIT includes forward-backward averaging (FBA) and the real-valued transformation as preprocessing steps (cf. Section 3.4.2). In Section 4.3.2, we have reviewed the performance analysis of R-D Unitary ESPRIT with LS. We have seen that FBA is easily incorporated into the performance analysis by replacing the noise-free subspaces of $\boldsymbol{X}_0$ by the corresponding subspaces of $\boldsymbol{X}_0^{(\mathrm{fba})}$. Moreover, it was shown in [RHD14] that asymptotically in the effective SNR, the real-valued transformation has no impact on the performance and hence, can be ignored. As a result, the performance analysis of the LS-based R-D Standard ESPRIT algorithm with FBA is also valid for the LS-based R-D Unitary ESPRIT algorithm. In this section, we show that the same findings also apply to the performance analysis of R-D Unitary ESPRIT with GLS.

We first consider the FBA preprocessing step for the performance of R-D Unitary ESPRIT with GLS. Following Section 4.3.2, we replace the noise-free subspaces by their corresponding FBA-processed versions and immediately obtain the first-order approximation for the parameter estimation error of the GLS-based R-D Standard ESPRIT algorithm with FBA as

$$\Delta\mu_i^{(r)} \approx \mathrm{Im}\left\{\boldsymbol{r}_{i,\mathrm{GLS}}^{(r)^{(\mathrm{fba})^{\mathrm{T}}}} \cdot \boldsymbol{W}_{\mathrm{mat}}^{(\mathrm{fba})} \cdot \boldsymbol{n}^{(\mathrm{fba})}\right\}, \tag{5.68}$$

where $\boldsymbol{W}_{\mathrm{mat}}^{(\mathrm{fba})}$ is given in (4.29) in Section 4.3.2 and

$$\begin{aligned}\boldsymbol{r}_{i,\mathrm{GLS}}^{(r)^{(\mathrm{fba})^{\mathrm{T}}}} = &-\frac{1}{\lambda_i^{(r)}} \cdot \left(\boldsymbol{q}_i^{(\mathrm{fba})^{\mathrm{T}}} \otimes \boldsymbol{p}_i^{(\mathrm{fba})^{\mathrm{T}}}\right) \\ &\cdot \left(\boldsymbol{F}_1^{(r)^{(\mathrm{fba})^{\mathrm{H}}}} \cdot \boldsymbol{R}_0^{(r)^{(\mathrm{fba})^{-1}}} \cdot \boldsymbol{F}_1^{(r)^{(\mathrm{fba})}}\right)^{-1} \cdot \boldsymbol{F}_1^{(r)^{(\mathrm{fba})^{\mathrm{H}}}} \cdot \boldsymbol{R}_0^{(r)^{(\mathrm{fba})^{-1}}} \cdot \boldsymbol{F}_2^{(r)^{(\mathrm{fba})}}.\end{aligned} \tag{5.69}$$

Note that $\boldsymbol{q}_i$ and $\boldsymbol{p}_i$ are replaced by the corresponding versions $\boldsymbol{q}_i^{(\mathrm{fba})}$ and $\boldsymbol{p}_i^{(\mathrm{fba})}$ obtained from the eigendecomposition $\boldsymbol{\Psi}^{(r)} = \boldsymbol{Q}^{(\mathrm{fba})} \cdot \boldsymbol{\Lambda}^{(r)} \cdot \boldsymbol{Q}^{(\mathrm{fba})^{-1}}$, which results from the shift invariance equation $\tilde{\boldsymbol{J}}_1^{(r)} \cdot \boldsymbol{U}_{\mathrm{s}}^{(\mathrm{fba})} \cdot \boldsymbol{\Phi}^{(r)} = \tilde{\boldsymbol{J}}_2^{(r)} \cdot \boldsymbol{U}_{\mathrm{s}}^{(\mathrm{fba})}$.

Next, we analyze the real-valued transformation as the second preprocessing step of R-D Unitary ESPRIT with GLS. Similarly to the result for R-D Unitary ESPRIT with LS in Section 4.3.2, we can formulate the following theorem:

Theorem 5.4.2. *The GLS-based R-D Unitary ESPRIT algorithm and the GLS-based R-D Stan-*

dard ESPRIT algorithm with FBA perform asymptotically identical in the high effective SNR.

The proof is shown in Appendix B.13.

From Theorem 5.4.2, we can conclude that as in the LS case in Section 4.3.2, the performance analysis of the GLS-based R-D Standard ESPRIT algorithm with FBA is also valid for the GLS-based R-D Unitary ESPRIT algorithm[6]. Recall that the GLS algorithm for R-D Unitary ESPRIT is derived for zero-mean circularly symmetric white noise $\boldsymbol{n} = \mathrm{vec}\{\boldsymbol{N}\} \in \mathbb{C}^{MN\times 1}$ with $\boldsymbol{R}_{\mathrm{nn}} = \sigma_{\mathrm{n}}^2 \cdot \boldsymbol{I}_{MN}$ and $\boldsymbol{C}_{\mathrm{nn}} = \boldsymbol{0}_{MN}$ such that the FBA-processed noise $\boldsymbol{n}^{(\mathrm{fba})} = \mathrm{vec}\left\{\boldsymbol{N}^{(\mathrm{fba})}\right\} \in \mathbb{C}^{2MN\times 1}$ has the SO moments $\boldsymbol{R}_{\mathrm{nn}}^{(\mathrm{fba})} = \mathbb{E}\left\{\boldsymbol{n}^{(\mathrm{fba})} \cdot \boldsymbol{n}^{(\mathrm{fba})^{\mathrm{H}}}\right\} = \sigma_{\mathrm{n}}^2 \cdot \boldsymbol{I}_{2MN}$ and $\boldsymbol{C}_{\mathrm{nn}}^{(\mathrm{fba})} = \mathbb{E}\left\{\boldsymbol{n}^{(\mathrm{fba})} \cdot \boldsymbol{n}^{(\mathrm{fba})^{\mathrm{T}}}\right\} = \sigma_{\mathrm{n}}^2 \cdot \boldsymbol{\Pi}_{2MN}$ as shown in Section 4.3.2. Thus, the analytical MSE of R-D Unitary ESPRIT using GLS for the i-th spatial frequency in the r-th mode is given by

$$\mathbb{E}\left\{\left(\Delta\mu_i^{(r)}\right)^2\right\} = \frac{\sigma_{\mathrm{n}}^2}{2} \cdot \left(\left\|\boldsymbol{z}_{i,\mathrm{GLS}}^{(r)(\mathrm{fba})}\right\|_2^2 - \mathrm{Re}\left\{\boldsymbol{z}_{i,\mathrm{GLS}}^{(r)(\mathrm{fba})^{\mathrm{T}}} \cdot \boldsymbol{\Pi}_{2MN} \cdot \boldsymbol{z}_{i,\mathrm{GLS}}^{(r)(\mathrm{fba})}\right\}\right), \tag{5.70}$$

where $\boldsymbol{z}_{i,\mathrm{GLS}}^{(r)(\mathrm{fba})} = \boldsymbol{W}_{\mathrm{mat}}^{(\mathrm{fba})^{\mathrm{T}}} \cdot \boldsymbol{r}_{i,\mathrm{GLS}}^{(r)(\mathrm{fba})} \in \mathbb{C}^{2MN\times 1}$.

5.5. Special source cases

The analytical MSE expressions for R-D ESPRIT-type algorithms with GLS, which are derived in Section 5.4, are deterministic, asymptotic in the effective SNR, and valid for an arbitrary number of sources. Thus, the analytical MSE expressions can be analyzed as a function of the physical parameters such as the SNR, the number of sensors M, and the sample size N via a simulative study without the need of Monte-Carlo simulations. However, the analytical MSE expressions depend on the subspaces and are not explicit in the physical parameters. As motivated in Section 4.5, it is therefore desirable to simplify the MSE expressions for special cases, e.g., a single source ($d = 1$) or two sources ($d = 2$), to obtain simple formulas that directly depend on the physical parameters of significance. Therefore, in this section, we simplify the analytical MSE expressions of both GLS-based R-D Standard ESPRIT and GLS-based R-D Unitary ESPRIT for a single source and the MSE expression of R-D Standard ESPRIT with GLS for two temporally orthogonal sources, i.e., the sample signal covariance matrix $\hat{\boldsymbol{R}}_{\mathrm{ss}}$ is given by $\hat{\boldsymbol{R}}_{\mathrm{ss}} = \mathrm{diag}\left\{\hat{P}_i\right\}_{i=1}^{d}$, where $\hat{P}_i = \|\boldsymbol{s}_i\|_2^2/N$ is the empirical power of the i-th signal $\boldsymbol{s}_i$. Moreover, we show that the simplified MSE expressions for both special cases can be directly reformulated into the corresponding simplified expressions for the deterministic Cramér-Rao bound (CRB) from [SN89]. As discussed in Section 4.5 and in Chapter 9, the deterministic CRB provides a lower limit on the MSE of any unbiased parameter estimation algorithm. The fact that the simplified analytical MSE expressions coincide with the deterministic

[6]Note that this property is not true for R-D Unitary ESPRIT with SLS, where according to a simulative study, the real-valued transformation slightly improves the MSE in the high effective SNR.

CRB for a single source and two temporally orthogonal sources implies that R-D Standard ESPRIT and R-D Unitary ESPRIT with GLS are asymptotically efficient for a single source and R-D Standard ESPRIT with GLS is asymptotically efficient for two temporally orthogonal sources, i.e., the ratio of the CRB and the MSE is equal to 1. For the GLS-based 1-D Unitary PUMA algorithm for a single source from [QHS+15], which corresponds to our GLS-based 1-D Unitary ESPRIT algorithm, it is shown that the CRB can be achieved asymptotically. However, our result that this is also the case for R-D Standard ESPRIT with GLS for two temporally orthogonal sources is very significant as it has so far not been shown that ESPRIT-type algorithms can be asymptotically efficient for $d > 1$.

In Section 5.5.1, we present the results for a single source and in Section 5.5.2, we present the results for two temporally orthogonal sources.

5.5.1. Single source case

In this section, we consider the case of 1-D parameter estimation and simplify the analytical MSE expressions for 1-D Standard ESPRIT and 1-D Unitary ESPRIT both with GLS from (5.66) and (5.70) for a single source. The result is stated in the following theorem:

Theorem 5.5.1. *For an M-element ULA with omnidirectional sensors, a single source ($d = 1$), and circularly symmetric white complex Gaussian noise[7], the analytical MSEs of the GLS-based 1-D Standard ESPRIT and 1-D Unitary ESPRIT algorithms coincide with the deterministic CRB. The simplified CRB expression for a single source is provided in (4.44) and given by*

$$C = \frac{1}{\hat{\varrho}} \cdot \frac{6}{M \cdot (M^2 - 1)}, \tag{5.71}$$

where $\hat{\varrho}$ represents the effective SNR $\hat{\varrho} = N\hat{P}/\sigma_{\mathrm{n}}^2$ with $\hat{P}$ being the empirical source power given by $\hat{P} = \|\boldsymbol{s}\|_2^2/N$ and $\boldsymbol{s} \in \mathbb{C}^{N\times 1}$.

The proof is shown in Appendix B.14.

From Theorem 5.5.1, we can conclude that 1-D Standard ESPRIT and 1-D Unitary ESPRIT with GLS are asymptotically efficient for a single source, i.e., the ratio of the CRB and the MSE is equal to one. This result agrees with the findings in [QHS+15] for the GLS-based 1-D Unitary PUMA algorithm developed for a single source, which is a special case of our 1-D Unitary ESPRIT algorithm with GLS for multiple sources.

Note that the R-D ESPRIT-type algorithms with LS discussed in Chapter 3 and the R-D ESPRIT-type algorithms with LS and spatial smoothing preprocessing discussed in Chapter 7 are

[7] The assumption of the noise to be Gaussian distributed is only required for the CRB and not for the GLS-based ESPRIT-type algorithms.

not asymptotically efficient. We have computed their asymptotic efficiency, i.e., the ratio of the CRB and the MSE in (4.45) and (7.59), respectively.

Furthermore, similarly to the R-D ESPRIT-type algorithms with LS in Section 4.5.1, no gain from FBA can be achieved for a single source.

5.5.2. Two source case

In this section, we simplify the analytical MSE expressions for 1-D Standard ESPRIT with GLS for two temporally orthogonal sources. We obtain the following result:

Theorem 5.5.2. *For an M-element ULA, two temporally orthogonal sources ($d = 2$), and circularly symmetric white complex Gaussian noise[8], the analytical MSEs of 1-D Standard ESPRIT and 1-D Unitary ESPRIT algorithms coincide with the deterministic CRB. The simplified CRB expression for $d = 2$ is given in* (9.45).

The proof is shown in Appendix B.15.

Note that this result is quite significant as so far, ESPRIT-type algorithms have not been shown to become asymptotically efficient for $d > 1$.

5.6. Simulation results

In this section, we show numerical results to demonstrate the performance of the presented R-D ESPRIT-type algorithms with GLS. In Section 5.6.1, we first consider the 1-D parameter estimation of the GLS-based Standard ESPRIT and Unitary ESPRIT algorithms. In Section 5.6.2, we evaluate the performance of the R-D Standard ESPRIT and R-D Unitary ESPRIT algorithms with GLS along with the asymptotic behavior of the presented performance analysis.

5.6.1. 1-D ESPRIT-type algorithms with GLS

This subsection provides numerical simulations, which illustrate the performance of the proposed Standard ESPRIT using GLS (SE GLS) and Unitary ESPRIT using GLS (UE GLS) algorithms both with a single GLS iteration and 5 GLS iterations (where convergence is reached) compared to the existing algorithms. For the comparison, we include SE LS and SE SLS as well as their Unitary versions UE LS and UE SLS, where we use one iteration of SLS. We also consider the optimally weighted ESPRIT (OW ESPRIT) algorithm [SN91]. The algorithms are benchmarked by the deterministic Cramér-Rao bound (Det CRB) [SN89]. In the simulations, we assume that a ULA with $M = 20$ isotropic sensor elements with half-wavelength spacing receives unit-power

[8]The assumption of the noise to be Gaussian distributed is only required for the CRB and not for the GLS-based ESPRIT-type algorithms.

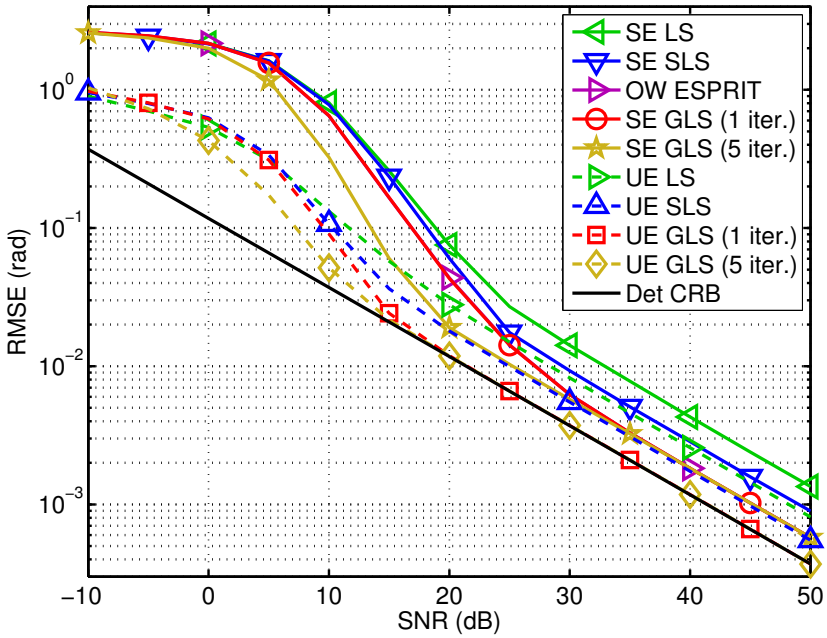


Figure 5.1.: RMSE versus the SNR for $d = 3$ correlated signals ($\rho = 0.95$) at $\boldsymbol{\mu} = [0.25, 0.5, 0.75]^{\mathrm{T}}$ with $M = 20$ and $N = 5$.

signals carrying symbols drawn from a complex Gaussian distribution. The sensor noise is zero-mean circularly symmetric white Gaussian distributed and the curves are averaged over 5000 Monte Carlo trials.

Figure 5.1 shows the total root mean square error (RMSE) versus the SNR. We assume $d = 3$ highly correlated signals with the spatial frequencies $\boldsymbol{\mu} = [0.25, 0.5, 0.75]^{\mathrm{T}}$ and a real-valued pairwise correlation of $\rho = 0.95$. The number of snapshots is $N = 5$. It is apparent that for low SNRs, SE GLS (5 iter.) clearly outperforms SE GLS (1 iter.) and OW ESPRIT, which perform identically. However, both SE GLS versions converge in the high SNR regime, such that a single GLS iteration is sufficient in the asymptotic case. Moreover, the UE algorithms can resolve the high signal correlation due to the implicit FBA processing, whereby UE GLS (5 iter.) is superior over UE GLS (1 iter.) at low SNRs. Yet, both UE GLS versions achieve the Det CRB.

In Figure 5.2, we depict the total RMSE versus the number of snapshots N. The SNR is fixed at 20 dB and we have $d = 3$ signals from the positions $\boldsymbol{\mu} = [0.2, 0.4, 0.6]^{\mathrm{T}}$ with a correlation of $\rho = 0.95$. In Figure 5.2, the same behavior of the algorithms is observed as the GLS (5 iter.) versions outperform the GLS (1 iter.) versions for a small sample size but both converge for increasing N. Again, both UE GLS versions achieve the Det CRB. From the Figures 5.1 and 5.2, we have seen that in the non-asymptotic case, i.e., at a low SNR or for a small sample size, additional GLS iterations can improve the estimation accuracy. However, in the asymptotic case, i.e., a high SNR or a large sample size, only one GLS iteration is sufficient. In Figure 5.3, we

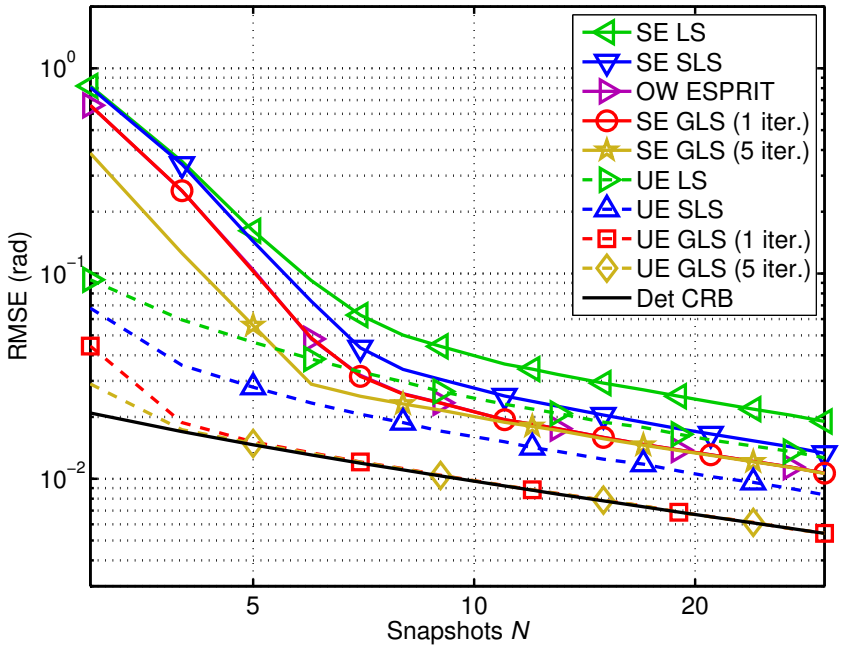


Figure 5.2.: RMSE versus the sample size N for $d = 3$ correlated signals ($\rho = 0.95$) at $\boldsymbol{\mu} = [0.2, 0.4, 0.6]^{\mathrm{T}}$ with $M = 20$ and SNR = 20 dB.

further investigate this observation and show the RMSE of the presented SE GLS and UE GLS algorithms as a function of the number of iterations for two different fixed SNRs. The parameters for the simulation setup are given by $M = 20$, $N = 5$, and $d = 3$ uncorrelated signals impinge from $\boldsymbol{\mu} = [0.25, 0.5, 0.75]^{\mathrm{T}}$. We show the results for SNR = 10 dB and SNR = 30 dB. It can be seen that in the case of SNR = 10 dB, both SE GLS and UE GLS have converged after about 3-5 GLS iterations. However, at SNR = 30 dB, the RMSEs of both algorithms remain constant over the number of iterations. Hence, one GLS iteration is sufficient in the asymptotic case.

5.6.2. Performance analysis of R-D ESPRIT-type algorithms with GLS

In this subsection, we demonstrate the asymptotic behavior of the performance analysis of the GLS-based R-D ESPRIT-type algorithms presented in Section 5.4. We compare the square root of the analytical MSE expressions ("ana") in (5.66) and (5.70) to the empirical ("emp") root mean square errors (RMSE)s of R-D Standard ESPRIT with GLS (SE GLS) and R-D Unitary ESPRIT with GLS (UE GLS). For both GLS-based algorithms, we only consider a single GLS iteration, which is sufficient in the asymptotic case. We again include the corresponding R-D versions using LS and SLS with one iteration of SLS into comparison. We also consider the analytical performance expressions for the R-D ESPRIT-type algorithms with LS discussed in Chapter 4 and an R-D extension of the analytical expressions for SLS-based 1-D ESPRIT-type algorithms in [RH11]. The algorithms are benchmarked by the deterministic R-D Cramér-Rao bound (Det

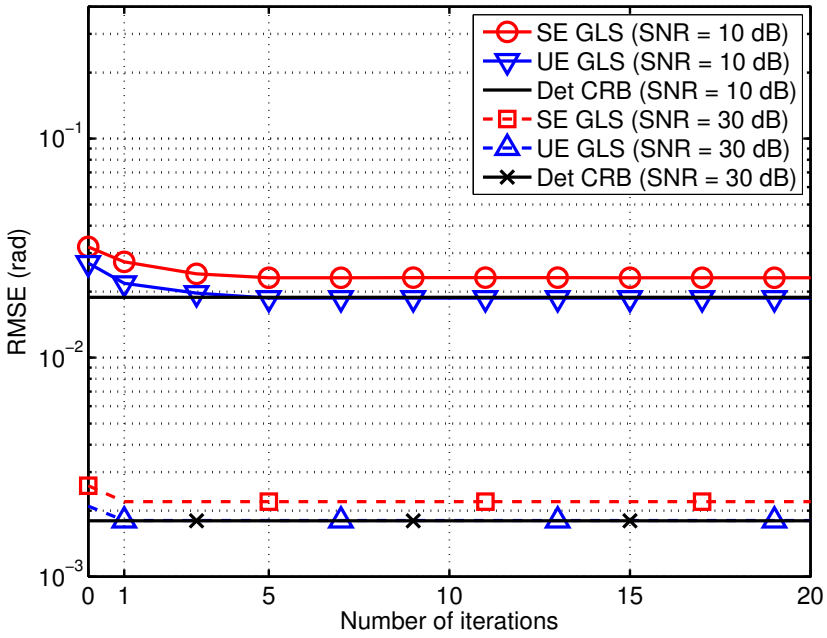


Figure 5.3.: RMSE versus the number of GLS iterations $d = 3$ uncorrelated signals at $\boldsymbol{\mu} = [0.25, 0.5, 0.75]^{\mathrm{T}}$ with $M = 20$ and $N = 5$ for SNR = 10 dB and SNR = 30 dB.

CRB) [SN89] discussed in Chapter 9. The total RMSE is defined as

$$\mathrm{RMSE} = \sqrt{\mathbb{E}\left\{\sum_{r=1}^{R}\sum_{i=1}^{d}\left(\mu_i^{(r)} - \hat{\mu}_i^{(r)}\right)^2\right\}}, \tag{5.72}$$

where $\hat{\mu}_i^{(r)}$ is the estimate of i-th spatial frequency in the r-th mode. The In the simulations, we assume that a $M_1 \times M_2$ uniform rectangular array (URA) with isotropic sensor elements and half-wavelength spacing in both dimensions. The remaining assumptions used for the simulation setup from the previous section are still valid.

In Figure 5.4, we depict the total RMSE versus the SNR of $d = 3$ sources impinging on a 5×65 URA with $N = 20$. The sources are located at $\mu_1^{(1)} = 1$, $\mu_2^{(1)} = 1.4$, $\mu_3^{(1)} = 1.8$, $\mu_1^{(2)} = -0.5$, $\mu_2^{(2)} = -0.1$, and $\mu_3^{(2)} = 0.8$. The magnitude of the correlation coefficient is $\rho = 0.99$. Figure 5.5 investigates the total RMSE versus the number of snapshots N for a 7×7 URA, where the SNR is 30 dB. We have $d = 2$ correlated sources with $\rho = 0.95$ located at $\mu_1^{(1)} = 1$, $\mu_2^{(1)} = 0.8$, $\mu_1^{(2)} = 1$, $\mu_2^{(2)} = 0.8$.

It is apparent from Figure 5.4 and Figure 5.5 that the analytical results agree well with the empirical results for high effective SNRs, i.e., either high SNRs or a large sample size. Furthermore, R-D SE GLS and R-D UE GLS provide the lowest estimation errors and perform very close to the CRB.

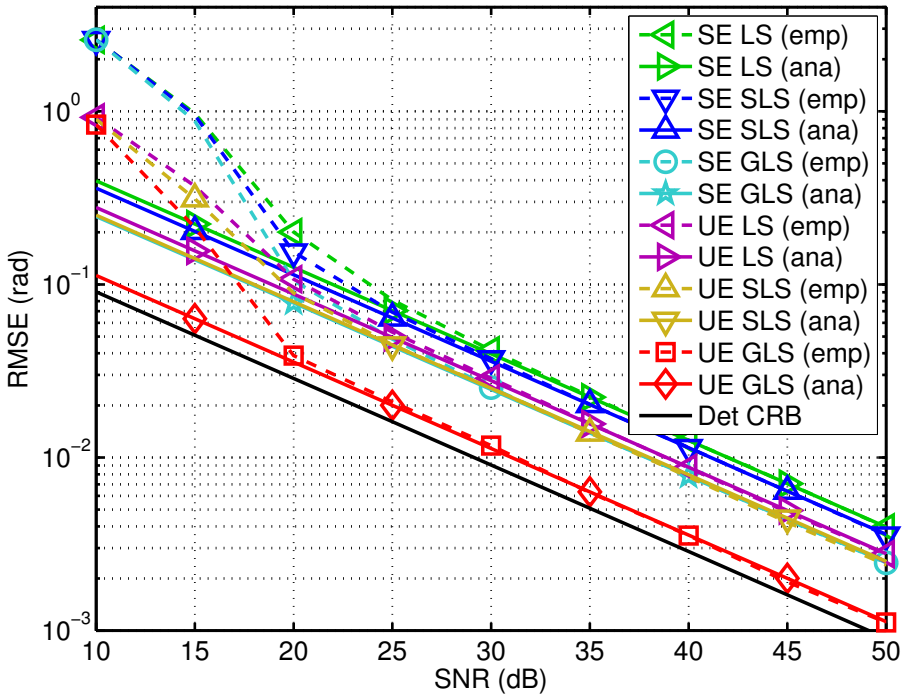


Figure 5.4.: RMSE versus the SNR for a 5×5 URA and $d = 3$ highly correlated signals ($\rho = 0.99$) at $\mu_1^{(1)} = 1$, $\mu_2^{(1)} = 1.4$, $\mu_3^{(1)} = 1.8$, $\mu_1^{(2)} = -0.5$, $\mu_2^{(2)} = -0.1$, and $\mu_3^{(2)} = 0.8$ with $N = 20$.

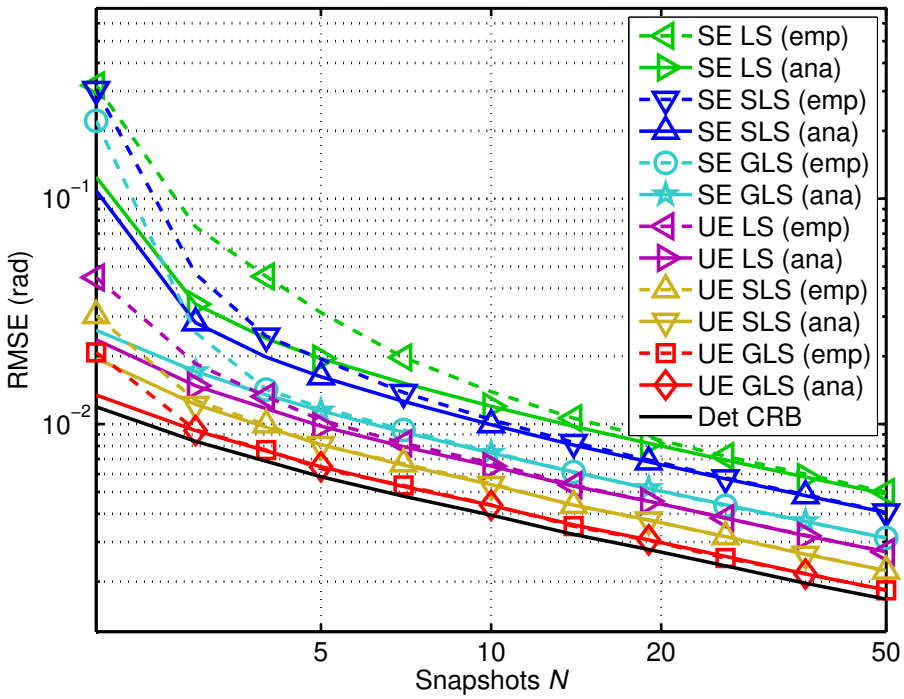


Figure 5.5.: RMSE versus the sample size N for a 7×7 URA and $d = 2$ highly correlated signals ($\rho = 0.95$) at $\mu_1^{(1)} = 1$, $\mu_2^{(1)} = 0.8$, $\mu_1^{(2)} = 1$, $\mu_2^{(2)} = 0.8$ with SNR = 30 dB.

Moreover, it becomes clear that the presented ESPRIT-type algorithms with GLS only achieve the CRB for $R = 1$. This is not surprising as we have only considered the matrix-based R-D ESPRIT-type algorithms and thus, the multi-dimensional structure of the signals is not exploited. This motivates the development of tensor-based R-D ESPRIT-type algorithms with GLS, which takes this structure into account. However, this extension is beyond the scope of this thesis and left for future work.

5.7. Summary

In this chapter, we have proposed a GLS solution to the shift invariance equation of R-D Standard ESPRIT and R-D Unitary ESPRIT for multiple sources, assuming a uniform R-D array and maximum subarray overlap in the R dimensions. The GLS algorithm directly incorporates the statistics of the subspace estimation error into the shift invariance equation solution via its covariance matrix. A closed-form expression for the subspace estimation error and its corresponding covariance matrix is found through a first-order perturbation expansion. We have seen that if the error covariance matrix is initialized by the simple LS solution, only one GLS iteration is required to achieve a significantly improved estimation accuracy in the asymptotic case, i.e., at either high SNRs or a large sample size. However, at low SNRs and for a small sample size N, performing additional GLS iterations further improves the estimation accuracy.

Furthermore, we have developed a performance analysis for R-D ESPRIT-type algorithms using a single GLS iteration. The derived analytical MSE expressions are based on the framework in Chapter 4 and are asymptotic in the effective SNR. For the special cases of a single source and two temporally orthogonal sources, we have simplified the general MSE expressions and prove that they coincide with the deterministic Cramér-Rao bound (CRB), which implies that in these cases, R-D ESPRIT-type algorithms in combination with one GLS iteration are asymptotically efficient, i.e., the ratio of the CRB and the MSE is equal to 1. However, simulation results have shown that R-D Standard ESPRIT with GLS also achieves the CRB for an arbitrary number of uncorrelated sources while R-D Unitary ESPRIT with GLS achieves the CRB for an arbitrary number of correlated but incoherent sources. These results are very significant as ESPRIT-type algorithms have so far not been shown to become asymptotically efficient for $d > 1$.

6. NC ESPRIT-type parameter estimation algorithms for strictly non-circular sources

In this chapter, we develop multi-dimensional ESPRIT-type algorithms to estimate the parameters of R-dimensional signals with strictly non-circular (NC) structure as introduced in Section 2.2. By exploiting the statistical properties of the strictly non-circular signal structure, an improved estimation accuracy can be achieved and a larger number of signals can be resolved. An overview on parameter estimation for NC signals including our contributions is given in Section 6.1. In Section 6.2 and Section 6.3, we present the matrix-based and tensor-based R-D NC ESPRIT-type algorithms for strictly non-circular signals. Their corresponding performance analysis to analytically assess their achievable improvements over conventional parameter estimation algorithms is provided in Section 6.4 and Section 6.5. Moreover, simplified expressions for the special cases of a single and two NC signals are derived in Section 6.6 while simulation results are presented in Section 6.7 and conclusions are drawn in Section 6.8.

6.1. Overview

In multi-dimensional parameter estimation problems, the received multi-dimensional signals are often assumed to carry symbols from complex-valued modulation schemes including quadrature phase shift keying (QPSK) or quadrature amplitude modulation (QAM), which represent second-order (SO) circular complex random variables [Pic94] as discussed in Section 2.2.1.1. The SO statistics of circularly symmetric random variables are fully described by the conventional covariance matrix. However, this assumption is not always justified as in some applications including wireless communications, cognitive radio, GNSS satellite systems [HC08], radar, tracking, and channel sounding, symbols from real-valued modulations schemes such as binary phase shift keying (BPSK), amplitude shift keying (ASK), offset-quadrature phase shift keying (OQPSK) or minimum shift keying (MSK) are received. In this case, the symbols represent SO strictly non-circular (NC) random variables (cf. Section 2.2.1.2) [PB97], which implies that the correlation between the random variables and their complex conjugates is equal to one [Pic94]. Consequently, the conventional covariance matrix no longer fully describes the SO statistics and the pseudo covariance matrix [NM93, Pic96], also known as complementary covariance matrix [SS03] needs to be taken into account [PB97]. Fully exploiting the statistical properties of NC signals can provide significant algorithmic performance improvements in many signal processing applications including parameter estimation. A detailed state of the art and a summary of our contributions in this field are given in the following two subsections.

6.1.1. State of the art

Considering the full SO statistics of non-circular signals by exploiting the additional information contained in the pseudo covariance matrix, which is termed widely-linear signal processing, has a long standing research history. A literature survey on non-circular signals and their many signal processing applications is provided in [SS10]. Examples include, for instance, widely-linear minimum mean squared error (MMSE) estimation [PC95], Wiener filtering [SS03], detection and estimation [SS05], widely-linear adaptive beamforming [CB07, SSW$^+$11, SdW$^+$11], interference-suppression [SdHW12], widely-linear precoding [Ste07], and more recently, widely-linear distributed beamforming [SH13a, SH13c, SRH15a]. In these applications, it has been shown that a considerable gain from processing weak-sense non-circular signals (with the non-circularity coefficient of $0 \leq |\zeta| < 1$) and strictly non-circular signals ($|\zeta| = 1$) can be achieved. However, the gain is more pronounced for strictly non-circular signals. Furthermore, it has been observed that apart from the non-circularity coefficient $|\zeta|$, the gain also depends on the rotation phase. A detailed discussion on non-circular signals and their properties is provided in Section 2.2.1.2.

In the last two decades, the concept of exploiting the signal structure of non-circular signals has also been applied to parameter estimation. It has been shown that the performance of subspace-based parameter estimation algorithms can be significantly improved if the signals' non-circularity is taken into account. For instance, NC versions of Root-MUSIC and Standard ESPRIT for strictly non-circular sources have been proposed in [CWS01] and [ZCW03], respectively. A spectral MUSIC algorithm for weak-sense non-circular signals is introduced in [AD06], which shows that the maximum gain is achieved for strictly non-circular signals. In [HR04], a 2-D NC Unitary ESPRIT algorithm has been proposed as an extension of [ZCW03]. It only requires real-valued operations at a reduced computational complexity and admits the 2-D parameter estimation. As a preliminary result of the contributions in this thesis, tensor-based R-D NC ESPRIT-type algorithms have been proposed in [RH09]. These algorithms are presented in Section 6.3. All these NC algorithms for strictly non-circular signals efficiently exploit the prior knowledge of the NC signal structure via a preprocessing step, which virtually doubles the antenna array. Therefore, for specific scenarios, they provide a substantial improvement in the estimation accuracy and can resolve twice as many sources compared to their conventional versions. It has been observed through simulations that the NC gain mostly depends on the rotation phase and the correlation of the signals. However, the exact dependence of the NC gain on these parameters has not been investigated in the literature.

The more general case of coexisting circular and strictly non-circular signals for MUSIC-like algorithms has been considered in [GNW08, LLXZ12]. In [SRH15c, SRH16a], we have presented the C-NC Standard ESPRIT and the C-NC Unitary ESPRIT algorithms developed for the mixture of circular and strictly non-circular signals. We show that the estimation accuracy of both algorithms improves with an increasing number of NC signals among a fixed number of sources. Thereby, not

only the estimation accuracy of the strictly non-circular signals themselves is improved, but also the estimation accuracy of the circular signals. Moreover, the corresponding deterministic Cramér-Rao bound has been derived by us in [SRH15b]. Note, however, that the C-NC ESPRIT-type algorithms will not be discussed in this thesis.

The observed benefits associated with NC signals have raised a considerable research interest in the analytical performance evaluation of the subspace-based NC parameter estimation algorithms in order to quantify the achievable improvements objectively. A detailed motivation and a literature review on performance analysis frameworks for subspace-based parameter estimation methods is provided in Section 4.1. Therein, we have introduced the two most prominent performance analysis frameworks, i.e., [RH89a] and [LLV93], and highlighted that [LLV93] is more general and should therefore be preferred. The performance of the spectral NC MUSIC algorithm has been derived in [AD06] based on the framework in [RH89a] and its source resolvability has been analytically investigated in [AD08]. A performance analysis of NC MUSIC in combination with mutual coupling has been considered in [HLLZ10], which shows how the coupling affects the performance.

6.1.2. Contributions

In Section 6.2, we first present the matrix-based R-D NC Standard ESPRIT and the R-D NC Unitary ESPRIT algorithms, which have been published in [SRHD14], as an extension of 1-D NC Standard ESPRIT in [ZCW03] and 2-D NC Unitary ESPRIT in [HR04]. Both algorithms exploit the strict SO non-circularity of stationary sources. We show that the preprocessing step for NC sources always results in a centro-symmetric virtual array with a doubled number of sensors, which makes R-D NC Unitary ESPRIT also applicable to physical non-centro-symmetric arrays. Moreover, R-D NC Unitary ESPRIT can also be efficiently implemented in terms of only real-valued computations, which reduces the computational complexity.

In Section 6.3, we develop the tensor-based R-D NC Standard Tensor-ESPRIT and the R-D NC Unitary Tensor-ESPRIT algorithms that exploit both the R-D structure and the NC structure of the signals. Preliminary results have been published in [RH09] and [Roe13]. Note that in comparison to [Roe13], the derivations and proofs in this thesis are simpler and more details are provided. In the tensor case, the NC preprocessing step is performed in each dimension independently and the corresponding solutions of the shift invariance equations are processed jointly. As in the matrix case, the virtual array steering tensor is always centro-symmetric and R-D NC Unitary Tensor-ESPRIT can be formulated in terms of real-valued operations. Both matrix-based and tensor-based R-D NC ESPRIT-type algorithms achieve a significantly lower estimation error than their conventional non-NC counterparts reviewed in Chapter 3.

In our second contribution, we derive a first-order performance analysis of the matrix-based and tensor-based R-D NC ESPRIT-type algorithms using least squares (LS). The results have been published in [SRHD14] and [SRH16b], respectively. Due to its advantages, we adopt the

performance analysis framework in [LLV93] in combination with its R-D tensor extension [RHD14] reviewed in Section 4.4 and incorporate the NC preprocessing step. We derive MSE expressions, where apart from a zero mean and finite SO moments no further assumptions on the noise statistics are needed. Moreover, the expressions are asymptotic in the effective SNR, i.e., they become accurate for either high SNRs or large sample sizes. Furthermore, we analytically prove that both R-D NC Standard ESPRIT and R-D NC Unitary ESPRIT as well as R-D NC Standard Tensor-ESPRIT and R-D NC Unitary Tensor-ESPRIT perform asymptotically identical. However, the unitary versions should be preferred due to the better performance in the non-asymptotic case and their lower computational complexity.

Finally, we simplify the analytical R-D MSE expressions for both matrix-based and tensor-based NC ESPRIT-type algorithms in the special case of a single NC source and two NC sources, where a uniform sampling grid and circularly symmetric white noise are assumed. The resulting expressions only depend on the physical parameters, e.g., the source correlation and the rotation phase of the signals. Based on these expressions, we derive analytical expressions for the NC gain of the matrix-based and tensor-based NC ESPRIT-type algorithms for two sources. Thereby, the exact dependence of the NC gain on the physical parameters is revealed and the parameter settings for the largest NC gain are identified.

Note that NC ESPRIT-type algorithms and their analytical performance combined with spatial smoothing is considered in Chapter 7 and the deterministic R-D NC Cramér-Rao bound (CRB) as a benchmark in the NC case is derived in Chapter 9.

6.2. R-D matrix-based NC ESPRIT-type algorithms

In this section, we present the matrix-based R-D NC ESPRIT-type algorithms. We start with the preprocessing step for R-D NC signals, termed R-D NC preprocessing, in Section 6.2.1 and show that the shift invariance property after the NC preprocessing is still satisfied. In Section 6.2.2, we introduce the R-D NC Standard ESPRIT algorithm before we resort to the R-D NC Unitary ESPRIT algorithm in Section 6.2.3.

6.2.1. R-D NC preprocessing

In this section, we derive the NC model resulting from the preprocessing for strictly non-circular sources. We show that the shift invariance equations also hold in the NC case and that the virtual array always possesses a centro-symmetric structure, even if the physical array is not centro-symmetric.

In Section 2.2, we have introduced the specific structure of strictly non-circular signals. This property implies that the complex baseband symbols can be expressed as $s_i[n] = \mathrm{e}^{\mathrm{j}\varphi_i} \cdot s_{0,i}[n]$, where $s_{0,i}[n] \in \mathbb{R}$ are real-valued symbols and the rotation phases φ_i do not vary in time $[n]$. Therefore,

the symbol matrix $\boldsymbol{S} \in \mathbb{C}^{d\times N}$ can be decomposed as [HR04, SRHD14]

$$\boldsymbol{S} = \boldsymbol{\Psi} \cdot \boldsymbol{S}_0, \tag{6.1}$$

where $\boldsymbol{S}_0 \in \mathbb{R}^{d\times N}$ is a real-valued symbol matrix and $\boldsymbol{\Psi} = \operatorname{diag}\left\{[\mathrm{e}^{\mathrm{j}\varphi_1}, \ldots, \mathrm{e}^{\mathrm{j}\varphi_d}]^{\mathrm{T}}\right\} \in \mathbb{C}^{d\times d}$ contains stationary complex phase shifts on its diagonal that can be different for each source (cf. (2.36)). Geometrically, the complex symbols of each source form a rotated line in the complex plane.

In order to exploit the strictly non-circular signal structure, we apply a preprocessing procedure and define the augmented measurement matrix $\boldsymbol{X}^{(\mathrm{nc})}$ as [CWS01, HR04][1]

$$\boldsymbol{X}^{(\mathrm{nc})} = \begin{bmatrix} \boldsymbol{X} \\ \boldsymbol{\Pi}_M \cdot \boldsymbol{X}^* \end{bmatrix} \in \mathbb{C}^{2M\times N}. \tag{6.2}$$

Upon inserting the matrix model $\boldsymbol{X} = \boldsymbol{A}\cdot\boldsymbol{S} + \boldsymbol{N}$ along with (6.1), the augmented matrix $\boldsymbol{X}^{(\mathrm{nc})}$ in (6.2) can be rewritten as

$$\begin{aligned} \boldsymbol{X}^{(\mathrm{nc})} &= \begin{bmatrix} \boldsymbol{A}\cdot\boldsymbol{S} \\ \boldsymbol{\Pi}_M \cdot \boldsymbol{A}^* \cdot \boldsymbol{S}^* \end{bmatrix} + \begin{bmatrix} \boldsymbol{N} \\ \boldsymbol{\Pi}_M \cdot \boldsymbol{N}^* \end{bmatrix} \\ &= \begin{bmatrix} \boldsymbol{A} \\ \boldsymbol{\Pi}_M \cdot \boldsymbol{A}^* \cdot \boldsymbol{\Psi}^* \cdot \boldsymbol{\Psi}^* \end{bmatrix} \cdot \boldsymbol{S} + \begin{bmatrix} \boldsymbol{N} \\ \boldsymbol{\Pi}_M \cdot \boldsymbol{N}^* \end{bmatrix} \qquad (6.3) \\ &= \boldsymbol{A}^{(\mathrm{nc})} \cdot \boldsymbol{S} + \boldsymbol{N}^{(\mathrm{nc})} = \boldsymbol{X}_0^{(\mathrm{nc})} + \boldsymbol{N}^{(\mathrm{nc})}, \qquad (6.4) \end{aligned}$$

where $\boldsymbol{A}^{(\mathrm{nc})} \in \mathbb{C}^{2M\times d}$ and $\boldsymbol{N}^{(\mathrm{nc})} \in \mathbb{C}^{2M\times N}$ are the augmented array steering matrix and the augmented noise matrix, respectively, and we have used the fact that $\boldsymbol{S}_0 = \boldsymbol{\Psi}^* \cdot \boldsymbol{S}$ in (6.3). Moreover, $\boldsymbol{X}_0^{(\mathrm{nc})} \in \mathbb{C}^{2M\times N}$ denotes the noise-free augmented measurement matrix. The extended dimensions of $\boldsymbol{A}^{(\mathrm{nc})}$ can be interpreted as a virtual doubling of the number of sensor elements, which also doubles the number of detectable sources and provides a lower estimation error.

Based on the assumption that the array steering matrix $\boldsymbol{A}$ is shift-invariant, we next analyze the properties of the augmented array steering matrix $\boldsymbol{A}^{(\mathrm{nc})}$. The shift invariance properties for the physical array described by $\boldsymbol{A}$ are given by (cf. Equation 3.25)

$$\tilde{\boldsymbol{J}}_1^{(r)} \cdot \boldsymbol{A} \cdot \boldsymbol{\Phi}^{(r)} = \tilde{\boldsymbol{J}}_2^{(r)} \cdot \boldsymbol{A}, \quad r = 1, \ldots, R, \tag{6.5}$$

where $\tilde{\boldsymbol{J}}_1^{(r)}$ and $\tilde{\boldsymbol{J}}_2^{(r)} \in \mathbb{R}^{\frac{M}{M_r}M_r^{(\mathrm{sel})}\times M}$ are the effective R-D selection matrices, which select $M_r^{(\mathrm{sel})}$ elements for the first and the second subarray in the r-th mode, respectively. They are compactly defined as $\tilde{\boldsymbol{J}}_k^{(r)} = \boldsymbol{I}_{\prod_{l=1}^{r-1} M_l} \otimes \boldsymbol{J}_k^{(r)} \otimes \boldsymbol{I}_{\prod_{l=r+1}^{R} M_l}$ for $k = 1, 2$, where $\boldsymbol{J}_k^{(r)} \in \mathbb{R}^{M_r^{(\mathrm{sel})}\times M_r}$ are the r-

[1] Note that [CWS01] defines $\boldsymbol{X}^{(\mathrm{nc})}$ for NC Root-MUSIC without the matrix $\boldsymbol{\Pi}_M$. The preprocessing step in (6.2) including $\boldsymbol{\Pi}_M$ was first proposed in [HR04] to facilitate the real-valued implementation for Unitary ESPRIT.

mode selection matrices for the first and second subarray [HN98]. The diagonal matrix $\boldsymbol{\Phi}^{(r)} = \mathrm{diag}\{[\mathrm{e}^{\mathrm{j}\mu_1^{(r)}}, \ldots, \mathrm{e}^{\mathrm{j}\mu_d^{(r)}}]^{\mathrm{T}}\} \in \mathbb{C}^{d\times d}$ contains the spatial frequencies in the r-th mode to be estimated.

The first important property of the augmented steering matrix $\boldsymbol{A}^{(\mathrm{nc})}$ is formulated in the following theorem:

Theorem 6.2.1. *If the array steering matrix $\boldsymbol{A}$ is shift-invariant* (6.5)*, then $\boldsymbol{A}^{(\mathrm{nc})}$ is also shift-invariant and satisfies*

$$\tilde{\boldsymbol{J}}_1^{(\mathrm{nc})(r)} \cdot \boldsymbol{A}^{(\mathrm{nc})} \cdot \boldsymbol{\Phi}^{(r)} = \tilde{\boldsymbol{J}}_2^{(\mathrm{nc})(r)} \cdot \boldsymbol{A}^{(\mathrm{nc})}, \quad r = 1, \ldots, R, \tag{6.6}$$

where the selection matrices $\tilde{\boldsymbol{J}}_k^{(\mathrm{nc})(r)}$ for $k = 1, 2$ are given by

$$\tilde{\boldsymbol{J}}_k^{(\mathrm{nc})(r)} = \boldsymbol{I}_{\prod_{l=1}^{r-1} M_l} \otimes \boldsymbol{J}_k^{(\mathrm{nc})(r)} \otimes \boldsymbol{I}_{\prod_{l=r+1}^{R} M_l} \tag{6.7}$$

with

$$\boldsymbol{J}_1^{(\mathrm{nc})(r)} = \begin{bmatrix} \boldsymbol{J}_1^{(r)} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{\Pi}_{M_r^{(\mathrm{sel})}} \cdot \boldsymbol{J}_2^{(r)} \cdot \boldsymbol{\Pi}_{M_r} \end{bmatrix} \in \mathbb{R}^{2M_r^{(\mathrm{sel})} \times 2M_r}, \tag{6.8}$$

$$\boldsymbol{J}_2^{(\mathrm{nc})(r)} = \begin{bmatrix} \boldsymbol{J}_2^{(r)} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{\Pi}_{M_r^{(\mathrm{sel})}} \cdot \boldsymbol{J}_1^{(r)} \cdot \boldsymbol{\Pi}_{M_r} \end{bmatrix} \in \mathbb{R}^{2M_r^{(\mathrm{sel})} \times 2M_r}. \tag{6.9}$$

The proof is given in Appendix B.16. If the physical array is R-D centro-symmetric (cf. Section 2.1.4.2), i.e., it is symmetric with respect to its centroid, the array steering matrix $\boldsymbol{A}$ satisfies the property (2.22), which we restate here as

$$\boldsymbol{\Pi}_M \cdot \boldsymbol{A}^* = \boldsymbol{A} \cdot \boldsymbol{\Delta}_{\mathrm{c}}, \tag{6.10}$$

where $\boldsymbol{\Delta}_{\mathrm{c}} \in \mathbb{C}^{d\times d}$ is a unitary diagonal matrix[2]. If the centro-symmetry in (6.10) holds, we have $\boldsymbol{J}_1^{(r)} = \boldsymbol{\Pi}_{M_r^{(\mathrm{sel})}} \cdot \boldsymbol{J}_2^{(r)} \cdot \boldsymbol{\Pi}_{M_r}$ and $\boldsymbol{J}_2^{(r)} = \boldsymbol{\Pi}_{M_r^{(\mathrm{sel})}} \cdot \boldsymbol{J}_1^{(r)} \cdot \boldsymbol{\Pi}_{M_r}$ such that the augmented selection matrices $\boldsymbol{J}_1^{(\mathrm{nc})(r)}$ and $\boldsymbol{J}_2^{(\mathrm{nc})(r)}$ simplify to

$$\boldsymbol{J}_k^{(\mathrm{nc})(r)} = \boldsymbol{I}_2 \otimes \boldsymbol{J}_k^{(r)}, \ k = 1, 2. \tag{6.11}$$

Note that this special case was assumed in [ZCW03] and [HR04].

The second important property of $\boldsymbol{A}^{(\mathrm{nc})}$ is stated in the following theorem:

Theorem 6.2.2. *The augmented steering matrix $\boldsymbol{A}^{(\mathrm{nc})}$ always exhibits centro-symmetry even if*

[2] In case of a physical centro-symmetric array, we have established in Section 2.1.4.2 how $\boldsymbol{\Delta}_{\mathrm{c}}$ depends on the phase reference of the array. Recall that if the phase reference coincides with the array's centroid, we have $\boldsymbol{\Delta}_{\mathrm{c}} = \boldsymbol{I}_d$.

$\boldsymbol{A}$ *is not centro-symmetric.*

For the proof, the reader is referred to Appendix B.17. Hence, the NC preprocessing step always provides a centro-symmetric virtually augmented array. As a result, we can draw the important conclusion that R-D NC Unitary ESPRIT derived in the Section 6.2.3 can be applied to a broader variety of array geometries than R-D Unitary ESPRIT, which requires a centro-symmetric array. A numerical example is provided in the simulation results of Figure 6.3 in Section 6.7. In the next section, we first derive the R-D NC Standard ESPRIT algorithm before turning our attention to its unitary version.

6.2.2. R-D NC Standard ESPRIT

In the first step of subspace-based parameter estimation algorithm, we estimate the signal subspace. Based on the noisy augmented data model (6.4), we obtain the signal subspace $\hat{\boldsymbol{U}}_{\mathrm{s}}^{(\mathrm{nc})} \in \mathbb{C}^{2M \times d}$ by computing the d dominant left singular vectors of $\boldsymbol{X}^{(\mathrm{nc})}$. Using the property that $\boldsymbol{A}^{(\mathrm{nc})}$ and $\hat{\boldsymbol{U}}_{\mathrm{s}}^{(\mathrm{nc})}$ span approximately the same column space, we can find a non-singular matrix $\boldsymbol{T} \in \mathbb{C}^{d \times d}$ such that $\boldsymbol{A}^{(\mathrm{nc})} \approx \hat{\boldsymbol{U}}_{\mathrm{s}}^{(\mathrm{nc})} \cdot \boldsymbol{T}$. Using this relation, the overdetermined set of R augmented shift invariance equations from (6.6) can be expressed in terms of the estimated augmented signal subspace, yielding

$$\tilde{\boldsymbol{J}}_1^{(\mathrm{nc})(r)} \cdot \hat{\boldsymbol{U}}_{\mathrm{s}}^{(\mathrm{nc})} \cdot \boldsymbol{\Psi}^{(r)} \approx \tilde{\boldsymbol{J}}_2^{(\mathrm{nc})(r)} \cdot \hat{\boldsymbol{U}}_{\mathrm{s}}^{(\mathrm{nc})} \tag{6.12}$$

for $r = 1, \ldots, R$ with $\boldsymbol{\Psi}^{(r)} \approx \boldsymbol{T} \cdot \boldsymbol{\Phi}^{(r)} \cdot \boldsymbol{T}^{-1}$. Often, the R unknown matrices $\boldsymbol{\Psi}^{(r)} \in \mathbb{C}^{d \times d}$ are estimated using least squares (LS), i.e.,

$$\hat{\boldsymbol{\Psi}}_{\mathrm{LS}}^{(r)} = \left(\tilde{\boldsymbol{J}}_1^{(\mathrm{nc})(r)} \cdot \hat{\boldsymbol{U}}_{\mathrm{s}}^{(\mathrm{nc})} \right)^{+} \cdot \tilde{\boldsymbol{J}}_2^{(\mathrm{nc})(r)} \cdot \hat{\boldsymbol{U}}_{\mathrm{s}}^{(\mathrm{nc})} \in \mathbb{C}^{d \times d}. \tag{6.13}$$

Finally, after solving (6.13) for $\hat{\boldsymbol{\Psi}}^{(r)}$ in each mode independently, the correctly paired spatial frequency estimates are given by $\hat{\mu}_i^{(r)} = \arg\{\hat{\lambda}_i^{(r)}\}$, $i = 1, \ldots, d$. The eigenvalues $\hat{\lambda}_i^{(r)}$ of $\hat{\boldsymbol{\Psi}}^{(r)}$ are obtained by performing a joint eigendecomposition across all R dimensions [FG06] or via the simultaneous Schur decomposition [HN98]. The R-D NC Standard ESPRIT algorithm is summarized in Table 9.

6.2.3. R-D NC Unitary ESPRIT

In this subsection, we extend the concept of Unitary ESPRIT introduced in Section 3.4.2 to the augmented measurement matrix $\boldsymbol{X}^{(\mathrm{nc})}$ in (6.4) and derive the R-D NC Unitary ESPRIT algorithm. To this end, we first apply forward-backward averaging (FBA) as well as the real-valued transformation (cf. Section 3.2.3) to $\boldsymbol{X}^{(\mathrm{nc})}$. Following (3.16), FBA is performed by replacing the NC measurement matrix $\boldsymbol{X}^{(\mathrm{nc})} \in \mathbb{C}^{2M \times N}$ by the column-wise augmented measurement matrix

Algorithm 9 [SRHD14] Summary of R-D NC Standard ESPRIT

1. Estimate the augmented signal subspace $\hat{\boldsymbol{U}}_s^{(\mathrm{nc})} \in \mathbb{C}^{2M \times d}$ via the truncated SVD of the augmented measurement matrix $\boldsymbol{X}^{(\mathrm{nc})} \in \mathbb{C}^{2M \times N}$.

2. Solve the set of overdetermined augmented shift invariance equations

$$\tilde{\boldsymbol{J}}_1^{(\mathrm{nc})(r)} \cdot \hat{\boldsymbol{U}}_{\mathrm{s}}^{(\mathrm{nc})} \cdot \boldsymbol{\Psi}^{(r)} \approx \tilde{\boldsymbol{J}}_2^{(\mathrm{nc})(r)} \cdot \hat{\boldsymbol{U}}_{\mathrm{s}}^{(\mathrm{nc})}$$

   for the R matrices $\boldsymbol{\Psi}^{(r)} \in \mathbb{C}^{d \times d}$, $r = 1, \ldots, R$, via LS methods, e.g., LS, TLS, SLS, GLS, and $\tilde{\boldsymbol{J}}_k^{(\mathrm{nc})(r)} \in \mathbb{R}^{\frac{M}{M_r} M_r^{(\mathrm{sel})} \times 2M}$, $k = 1, 2$, is defined in (6.7).

3. Compute the eigenvalues $\hat{\lambda}_i^{(r)}$, $i = 1, \ldots, d$ of $\boldsymbol{\Psi}_{\mathrm{LS}}^{(r)}$ jointly for all $r = 1, \ldots, R$, e.g., via the joint diagonalization scheme proposed in [FG06]. Recover the correctly paired spatial frequencies $\hat{\mu}_i^{(r)}$ via $\hat{\mu}_i^{(r)} = \arg\left\{\hat{\lambda}_i^{(r)}\right\}$.

$\boldsymbol{X}^{(\mathrm{nc})(\mathrm{fba})}$ defined as

$$\boldsymbol{X}^{(\mathrm{nc})(\mathrm{fba})} = \begin{bmatrix} \boldsymbol{X}^{(\mathrm{nc})} & \boldsymbol{\Pi}_{2M} \cdot \boldsymbol{X}^{(\mathrm{nc})^*} \cdot \boldsymbol{\Pi}_N \end{bmatrix} \in \mathbb{C}^{2M \times 2N}. \tag{6.14}$$

Inserting the definition of $\boldsymbol{X}^{(\mathrm{nc})}$ in (6.2) yields

$$\begin{aligned} \boldsymbol{X}^{(\mathrm{nc})(\mathrm{fba})} &= \begin{bmatrix} \boldsymbol{X} & \boldsymbol{X} \cdot \boldsymbol{\Pi}_N \\ \boldsymbol{\Pi}_M \cdot \boldsymbol{X}^* & \boldsymbol{\Pi}_M \cdot \boldsymbol{X}^* \cdot \boldsymbol{\Pi}_N \end{bmatrix} \\ &= \begin{bmatrix} \boldsymbol{X}^{(\mathrm{nc})} & \boldsymbol{X}^{(\mathrm{nc})} \cdot \boldsymbol{\Pi}_N \end{bmatrix}. \end{aligned} \tag{6.15}$$

From (6.15), it is straightforward to see that

$$\boldsymbol{X}^{(\mathrm{nc})(\mathrm{fba})} \cdot \boldsymbol{X}^{(\mathrm{nc})(\mathrm{fba})^{\mathrm{H}}} = 2 \cdot \boldsymbol{X}^{(\mathrm{nc})} \cdot \boldsymbol{X}^{(\mathrm{nc})^{\mathrm{H}}}. \tag{6.16}$$

Consequently, FBA does not alter the column space of $\boldsymbol{X}^{(\mathrm{nc})}$. This observation has two implications. Firstly, FBA does not improve the performance. In fact, it will be proven in Section 6.4.2 that the performance of R-D NC Standard ESPRIT and R-D NC Unitary ESPRIT is asymptotically identical. Secondly, after the NC preprocessing step, two coherent sources cannot be decorrelated anymore by additionally applying FBA. Thus, unlike R-D Unitary ESPRIT, R-D NC Unitary ESPRIT cannot resolve two coherent sources. This effect will be demonstrated using numerical simulations in Section 6.7.

Since the FBA-processed augmented measurement matrix $\boldsymbol{X}^{(\mathrm{nc})(\mathrm{fba})}$ is a centro-Hermitian matrix, we can apply the real-valued transformation analogously to (3.18). If the sparse left-$\boldsymbol{\Pi}$-real

matrices $\boldsymbol{Q}_p^{(\mathrm{s})}$ from (3.17) (also Appendix A.2) are used, we obtain the simple form

$$\varphi\left(\boldsymbol{X}^{(\mathrm{nc})^{(\mathrm{fba})}}\right) = \boldsymbol{Q}_{2M}^{(\mathrm{s})^{\mathrm{H}}} \cdot \boldsymbol{X}^{(\mathrm{nc})^{(\mathrm{fba})}} \cdot \boldsymbol{Q}_{2N}^{(\mathrm{s})} \tag{6.17}$$

$$= 2 \cdot \begin{bmatrix} \mathrm{Re}\{\boldsymbol{X}\} & \boldsymbol{0}_{M\times N} \\ \mathrm{Im}\{\boldsymbol{X}\} & \boldsymbol{0}_{M\times N} \end{bmatrix}. \tag{6.18}$$

The proof is given in Appendix B.18. Note that the factor 2 and the zero entries can be skipped as they do not affect the signal subspace estimate. Therefore, it is sufficient to process the matrix $\left[\mathrm{Re}\{\boldsymbol{X}\}^{\mathrm{T}},\ \ \mathrm{Im}\{\boldsymbol{X}\}^{\mathrm{T}}\right]^{\mathrm{T}} \in \mathbb{R}^{2M\times N}$, which is real-valued. Thereby, the computational complexity is reduced significantly.

In the next step, we define the transformed augmented steering matrix as $\boldsymbol{D}^{(\mathrm{nc})} = \boldsymbol{Q}_{2M}^{\mathrm{H}} \cdot \boldsymbol{A}^{(\mathrm{nc})}$. Based on the R-D shift invariance property of $\boldsymbol{A}^{(\mathrm{nc})}$ proven in Theorem 6.2.1, it can easily be verified that $\boldsymbol{D}^{(\mathrm{nc})}$ obeys

$$\tilde{\boldsymbol{K}}_1^{(\mathrm{nc})(r)} \cdot \boldsymbol{D}^{(\mathrm{nc})} \cdot \boldsymbol{\Omega}^{(r)} = \tilde{\boldsymbol{K}}_2^{(\mathrm{nc})(r)} \cdot \boldsymbol{D}^{(\mathrm{nc})} \tag{6.19}$$

for $r = 1, \ldots, R$, where the R pairs of augmented selection matrices in (6.7) are transformed similarly to [HN98] as

$$\tilde{\boldsymbol{K}}_1^{(\mathrm{nc})(r)} = 2 \cdot \mathrm{Re}\left\{\boldsymbol{Q}^{\mathrm{H}}_{\frac{M}{M_r}M_r^{(\mathrm{sel})}} \cdot \tilde{\boldsymbol{J}}_2^{(\mathrm{nc})(r)} \cdot \boldsymbol{Q}_{2M}\right\} \tag{6.20}$$

$$\tilde{\boldsymbol{K}}_2^{(\mathrm{nc})(r)} = 2 \cdot \mathrm{Im}\left\{\boldsymbol{Q}^{\mathrm{H}}_{\frac{M}{M_r}M_r^{(\mathrm{sel})}} \cdot \tilde{\boldsymbol{J}}_2^{(\mathrm{nc})(r)} \cdot \boldsymbol{Q}_{2M}\right\}. \tag{6.21}$$

Moreover, the real-valued set of diagonal matrices $\boldsymbol{\Omega}^{(r)} = \mathrm{diag}\left\{[\omega_1^{(r)}, \ldots, \omega_d^{(r)}]^{\mathrm{T}}\right\} \in \mathbb{R}^{d\times d}$ with $\omega_i^{(r)} = \tan(\mu_i^{(r)}/2)$ contain the spatial frequencies in the r-th mode.

Using the preprocessed noisy data in (6.18), we then estimate the real-valued augmented signal subspace $\hat{\boldsymbol{E}}_{\mathrm{s}}^{(\mathrm{nc})} \in \mathbb{R}^{2M\times d}$ by computing the d dominant left singular vectors of $\varphi\left(\boldsymbol{X}^{(\mathrm{nc})^{(\mathrm{fba})}}\right)$. As $\boldsymbol{D}^{(\mathrm{nc})}$ and $\hat{\boldsymbol{E}}_{\mathrm{s}}^{(\mathrm{nc})}$ span approximately the same column space, we can find a non-singular matrix $\boldsymbol{T} \in \mathbb{C}^{d\times d}$ such that $\boldsymbol{D}^{(\mathrm{nc})} \approx \hat{\boldsymbol{E}}_{\mathrm{s}}^{(\mathrm{nc})} \cdot \boldsymbol{T}$. Substituting this relation into (6.19), the overdetermined set of R real-valued shift invariance equations in terms of the estimated augmented signal subspace is given by

$$\tilde{\boldsymbol{K}}_1^{(\mathrm{nc})(r)} \cdot \hat{\boldsymbol{E}}_{\mathrm{s}}^{(\mathrm{nc})} \cdot \boldsymbol{\Upsilon}^{(r)} \approx \tilde{\boldsymbol{K}}_2^{(\mathrm{nc})(r)} \cdot \hat{\boldsymbol{E}}_{\mathrm{s}}^{(\mathrm{nc})}, \quad r = 1, \ldots, R \tag{6.22}$$

with $\boldsymbol{\Upsilon}^{(r)} \approx \boldsymbol{T} \cdot \boldsymbol{\Omega}^{(r)} \cdot \boldsymbol{T}^{-1}$. Often, the R unknown real-valued diagonal matrices $\boldsymbol{\Upsilon}^{(r)}$ are estimated

Algorithm 10 [SRHD14] Summary of R-D NC Unitary ESPRIT

1. Estimate the augmented real-valued signal subspace $\hat{\boldsymbol{E}}_s^{(\mathrm{nc})} \in \mathbb{R}^{2M \times d}$ via the truncated SVD of the stacked measurement matrix
$$\left[\mathrm{Re}\{\boldsymbol{X}\}^{\mathrm{T}}, \quad \mathrm{Im}\{\boldsymbol{X}\}^{\mathrm{T}}\right]^{\mathrm{T}} \in \mathbb{R}^{2M \times N}.$$
2. Solve the overdetermined set of augmented shift invariance equations
$$\tilde{\boldsymbol{K}}_1^{(\mathrm{nc})(r)} \cdot \hat{\boldsymbol{E}}_{\mathrm{s}}^{(\mathrm{nc})} \cdot \boldsymbol{\Upsilon}^{(r)} \approx \tilde{\boldsymbol{K}}_2^{(\mathrm{nc})(r)} \cdot \hat{\boldsymbol{E}}_{\mathrm{s}}^{(\mathrm{nc})}$$
for $\boldsymbol{\Upsilon}^{(r)} \in \mathbb{R}^{d \times d}$, $r = 1, \ldots, R$, via LS methods, e.g., LS, TLS, SLS, GLS. The selection matrices $\tilde{\boldsymbol{K}}_k^{(\mathrm{nc})(r)} \in \mathbb{R}^{\frac{M}{M_r} M_r^{(\mathrm{sel})} \times 2M}$, $k = 1, 2$ and $\tilde{\boldsymbol{J}}_2^{(\mathrm{nc})(r)}$ are defined in (6.20), (6.21), and (6.7), respectively.
3. Compute the eigenvalues $\hat{\omega}_i^{(r)}$, $i = 1, \ldots, d$ of $\boldsymbol{\Upsilon}_{\mathrm{LS}}^{(r)}$ jointly for all $r = 1, \ldots, R$, e.g., via the joint diagonalization scheme proposed in [FG06] or via the Simultaneous Schur Decomposition proposed in [HN98]. Recover the correctly paired spatial frequencies $\hat{\mu}_i^{(r)}$ via $\hat{\mu}_i^{(r)} = 2 \cdot \arctan(\hat{\omega}_i^{(r)})$.

using least squares (LS), i.e.,

$$\hat{\boldsymbol{\Upsilon}}_{\mathrm{LS}}^{(r)} = \left(\tilde{\boldsymbol{K}}_1^{(\mathrm{nc})(r)} \cdot \hat{\boldsymbol{E}}_{\mathrm{s}}^{(\mathrm{nc})}\right)^{+} \cdot \tilde{\boldsymbol{K}}_2^{(\mathrm{nc})(r)} \cdot \hat{\boldsymbol{E}}_{\mathrm{s}}^{(\mathrm{nc})} \in \mathbb{R}^{d \times d}. \tag{6.23}$$

Finally, the correctly paired spatial frequency estimates are obtained by $\hat{\mu}_i^{(r)} = 2 \cdot \arctan(\hat{\omega}_i^{(r)})$, $i = 1, \ldots, d$. The eigenvalues $\hat{\omega}_i^{(r)}$ of $\hat{\boldsymbol{\Upsilon}}_{\mathrm{LS}}^{(r)}$ are computed by performing a joint eigendecomposition across all R dimensions [FG06] or via the simultaneous Schur decomposition [HN98]. If all the eigenvalues are real, they provide reliable estimates [HN95]. A summary of R-D NC Unitary ESPRIT is given in Table 10.

6.3. R-D tensor-based NC ESPRIT-type algorithms

In this section, we combine the NC preprocessing step with tensor algebra and present the tensor-based R-D NC ESPRIT-type algorithms. We will see that many results from the NC matrix case carry over to the tensor case. The R-D tensor-based NC preprocessing step is introduced in Section 6.3.1 and the R-D NC Standard Tensor-ESPRIT and R-D NC Unitary Tensor-ESPRIT algorithms are described in Section 6.3.2 and in Section 6.3.3, respectively.

6.3.1. R-D tensor-based NC preprocessing

We have seen in Section 3.5 that the multi-dimensional structure of the signals in the R-D harmonic retrieval problem can be exploited by processing a measurement tensor instead of stacking the multi-dimensional measurements into a matrix. As an example of tensor-based parameter estimation algorithms, we have discussed R-D Tensor-ESPRIT-type algorithms, which provide an improved parameter estimation accuracy by taking the R-D structure into account.

In some applications, the multi-dimensional signals additionally exhibit specific statistical properties, i.e., a strictly second-order (SO) non-circular (NC) structure. In Section 6.2 and in [SRHD14], we have shown for the matrix-based R-D NC ESPRIT-type algorithms that exploiting the strict non-circularity of the signals helps to enhance the estimation accuracy and doubles the number of identifiable sources of the conventional ESPRIT-type algorithms. This is also demonstrated in the numerical simulations in Section 6.7.1.

In this section, we demonstrate how both the multi-dimensional structure of the signals and their SO strictly non-circular structure can be exploited simultaneously by means of R-D NC Tensor-ESPRIT-type algorithms. Note that exploiting both types of structure is not straightforward. In the matrix case (cf. Section 6.2), a preprocessing step in terms of an augmentation along the rows is applied to exploit the signals' NC structure. However, in the tensor case, this operation would destroy the separability property of the R-D sampling grid, which is required for R-D Tensor-ESPRIT-type algorithms. An example is shown in Figure 6.1, where a 3×3 URA is augmented according to the matrix-based NC preprocessing step and flipped in both dimensions. It as apparent that the resulting virtual array composed of 18 sensors cannot be expressed as the outer product of 1-D sampling grids and hence, is not a separable 2-D sampling grid as per the definition in Section 2.1.1.

Thus, in order to take advantage of the R-D structure and the strict non-circularity simultaneously, a tensor-based equivalent of exploiting the non-circularity is proposed in [RH09]. Thereby, the preprocessing is performed by augmenting the measurement tensor along the individual modes separately and exploiting all these augmentations jointly. In the 2-D example in Figure 6.1, the augmentation is separately applied along the rows and along the columns and then both augmented tensors are processed jointly. Mathematically, the r-mode augmented measurement tensor $\boldsymbol{\mathcal{X}}^{(\mathrm{nc},r)} \in \mathbb{C}^{M_1 \times \ldots \times M_{r-1} \times 2M_r \times M_{r+1} \times \ldots \times M_R \times N}$ is defined as [RH09]

$$\boldsymbol{\mathcal{X}}^{(\mathrm{nc},r)} = \left[\boldsymbol{\mathcal{X}} \amalg_r \left(\boldsymbol{\mathcal{X}}^* \bigtimes_{q=1}^{R}{}_q \boldsymbol{\Pi}_{M_q} \right) \right], \tag{6.24}$$

where $\boldsymbol{\mathcal{X}}^{(\mathrm{nc},r)}$ is of size $2M_r$ along the r-th mode. Note that (6.24) can be factorized similarly

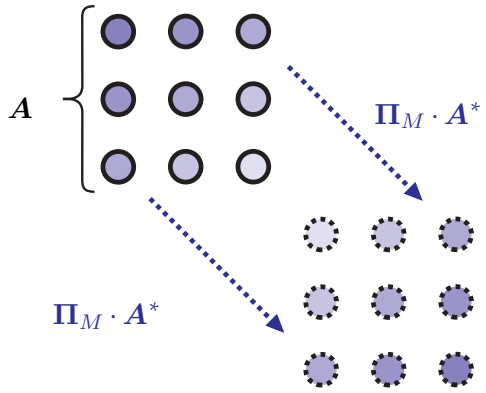


Figure 6.1.: Virtually doubled 2-D array after matrix-based augmentation of the measurements. The virtually doubled 3×3 URA is augmented by a second URA flipped in both dimensions. The resulting array is not a separable 2-D sampling grid.

to (6.4) as

$$\boldsymbol{\mathcal{X}}^{(\mathrm{nc},r)} = \boldsymbol{\mathcal{A}}^{(\mathrm{nc},r)} \times_{R+1} \boldsymbol{S}^{\mathrm{T}} + \boldsymbol{\mathcal{N}}^{(\mathrm{nc},r)} = \boldsymbol{\mathcal{X}}_0^{(\mathrm{nc},r)} + \boldsymbol{\mathcal{N}}^{(\mathrm{nc},r)}, \tag{6.25}$$

where the augmented array steering tensor $\boldsymbol{\mathcal{A}}^{(\mathrm{nc},r)} \in \mathbb{C}^{M_1 \times \ldots \times M_{r-1} \times 2M_r \times M_{r+1} \times \ldots \times M_R \times d}$ augmented in the r-th mode is given by

$$\boldsymbol{\mathcal{A}}^{(\mathrm{nc},r)} = \left[\boldsymbol{\mathcal{A}} \hookrightarrow_r \left(\boldsymbol{\mathcal{A}}^* \bigtimes_{q=1}^{R}{}_q \mathbf{\Pi}_{M_q} \times_{R+1} \left(\mathbf{\Psi}^* \cdot \mathbf{\Psi}^* \right) \right) \right]. \tag{6.26}$$

Moreover, $\boldsymbol{\mathcal{X}}_0^{(\mathrm{nc},r)}$ and $\boldsymbol{\mathcal{N}}^{(\mathrm{nc},r)}$ denote the noise-free augmented measurement tensor and the augmented noise tensor (cf. Equation (6.4) for the matrix case). From the augmented array steering tensor in (6.26), it is evident that the number of antennas in the r-th mode has been virtually doubled to $2M_r$.

Since ESPRIT-type algorithms require the shift invariance property, we show that if the array steering tensor $\boldsymbol{\mathcal{A}}$ is shift invariant, the augmented array steering tensor $\boldsymbol{\mathcal{A}}^{(\mathrm{nc},r)}$ is shift invariant as well. This property is stated in the next theorem:

Theorem 6.3.1. *If the array steering tensor $\boldsymbol{\mathcal{A}}$ satisfies the shift invariance property $\boldsymbol{\mathcal{A}} \times_r \boldsymbol{J}_1^{(r)} \times_{R+1} \mathbf{\Psi}^{(r)} = \boldsymbol{\mathcal{A}} \times_r \boldsymbol{J}_2^{(r)}$, then the r-mode augmented array steering tensor $\boldsymbol{\mathcal{A}}^{(\mathrm{nc},r)}$ in (6.26) satisfies the shift invariance equation*

$$\boldsymbol{\mathcal{A}}^{(\mathrm{nc},r)} \times_r \boldsymbol{J}_1^{(\mathrm{nc})(r)} \times_{R+1} \mathbf{\Phi}^{(r)} = \boldsymbol{\mathcal{A}}^{(\mathrm{nc},r)} \times_r \boldsymbol{J}_2^{(\mathrm{nc})(r)} \tag{6.27}$$

for $r = 1, 2, \ldots R$, where $\boldsymbol{J}_1^{(\mathrm{nc})(r)}$ and $\boldsymbol{J}_2^{(\mathrm{nc})(r)}$ are given in (6.8) and (6.9), respectively.

The proof of this theorem is given in Appendix B.19.

As the augmented array steering tensor is shift invariant in the r-th mode[3], we can exploit both the R-D tensor structure and the signals' strict non-circularity jointly by obtaining estimates of $\boldsymbol{\Phi}^{(r)}$ from (6.27) after performing the corresponding r-mode augmentations $\boldsymbol{\mathcal{X}}^{(\mathrm{nc},r)}$ for $r = 1, 2, \ldots, R$.

In analogy to the centro-symmetry property of the augmented array steering matrix in Theorem 6.2.2, the same property also holds for $\boldsymbol{\mathcal{A}}^{(\mathrm{nc},r)}$ in the tensor case, which is stated in the following theorem:

Theorem 6.3.2. *The augmented steering tensor $\boldsymbol{\mathcal{A}}^{(\mathrm{nc},r)}$ always exhibits centro-symmetry in the r-th mode even if $\boldsymbol{\mathcal{A}}$ is not centro-symmetric in the r-th mode.*

The proof can be found in Appendix B.20. Thus, similarly to the matrix case, the r-mode augmentation always results in a virtual array that is centro-symmetric in the r-th mode. Consequently, unlike R-D Unitary Tensor-ESPRIT, we can always apply its NC extension R-D NC Unitary Tensor-ESPRIT (cf. Section 6.2.3), even if the original R-D array is not centro-symmetric. Numerical examples to demonstrate this effect are provided in the simulations in Section 6.7. In the next two sections, we derive the R-D NC Standard Tensor-ESPRIT algorithm and the R-D NC Unitary Tensor-ESPRIT algorithm, respectively.

6.3.2. R-D NC Standard Tensor-ESPRIT

The shift invariance equation in (6.27) cannot be solved directly as the augmented array steering tensors are unknown. Thus, they are usually replaced by the estimates of the respective signal subspace tensors, which can be found in a similar manner as for the non-structured signal case described in Section 3.2.2.

Let the truncated HOSVD of the r-mode augmented measurement tensors $\boldsymbol{\mathcal{X}}_0^{(\mathrm{nc},r)}$ and $\boldsymbol{\mathcal{X}}^{(\mathrm{nc},r)}$ from (6.25) be given by

$$\boldsymbol{\mathcal{X}}_0^{(\mathrm{nc},r)} = \boldsymbol{\mathcal{S}}^{[\mathrm{s}](r)} \times_1 \boldsymbol{U}_1^{[\mathrm{s}](r)} \ldots \times_R \boldsymbol{U}_R^{[\mathrm{s}](r)} \times_{R+1} \boldsymbol{U}_{R+1}^{[\mathrm{s}](r)} \tag{6.28}$$

$$\boldsymbol{\mathcal{X}}^{(\mathrm{nc},r)} = \hat{\boldsymbol{\mathcal{S}}}^{[\mathrm{s}](r)} \times_1 \hat{\boldsymbol{U}}_1^{[\mathrm{s}](r)} \ldots \times_R \hat{\boldsymbol{U}}_R^{[\mathrm{s}](r)} \times_{R+1} \hat{\boldsymbol{U}}_{R+1}^{[\mathrm{s}](r)}, \tag{6.29}$$

where $\boldsymbol{\mathcal{S}}^{[\mathrm{s}](r)} \in \mathbb{C}^{p_1 \times \ldots \times p_R \times d}$ and its estimate $\hat{\boldsymbol{\mathcal{S}}}^{[\mathrm{s}](r)}$ are the truncated core tensors with $p_r = \min\{M_r, d\}$ if the array is centro-symmetric in all R modes. The case where the array is not centro-symmetric in all R modes is discussed in Theorem 6.3.3. The matrices $\boldsymbol{U}_q^{[\mathrm{s}](r)} \in \mathbb{C}^{M_q \times p_q}$ for $q = 1, \ldots, r-1, r+1, \ldots, R$ denote the q-spaces and $\boldsymbol{U}_r^{[\mathrm{s}](r)}$ is the augmented r-space, which is of size $2M_r \times p_r$ due to the r-mode augmentation in (6.24). Moreover, $\hat{\boldsymbol{U}}_q^{[\mathrm{s}](r)}$ and $\hat{\boldsymbol{U}}_r^{[\mathrm{s}](r)}$ are their

[3]Note that $\boldsymbol{\mathcal{A}}^{(\mathrm{nc},r)}$ can also be shift invariant in the other modes $q = 1, 2, \ldots, R$, $q \neq r$ if the array is centro-symmetric in the q-th mode. However, these additional shift invariances are not required for R-D NC Tensor-ESPRIT-type algorithms.

estimates, respectively. These matrices can be obtained from the SVD of the p-mode unfoldings for $p = 1, \ldots, R$ of $\boldsymbol{\mathcal{X}}_0^{(\mathrm{nc},r)}$ and $\boldsymbol{\mathcal{X}}^{(\mathrm{nc},r)}$ according to

$$\left[\boldsymbol{\mathcal{X}}_0^{(\mathrm{nc},r)}\right]_{(p)} = \left[\boldsymbol{U}_p^{[\mathrm{s}](r)}, \quad \boldsymbol{U}_p^{[\mathrm{n}](r)}\right] \cdot \begin{bmatrix} \boldsymbol{\Sigma}_p^{[\mathrm{s}](r)} & \boldsymbol{0}_{d\times(N-d)} \\ \boldsymbol{0}_{(M-d)\times d} & \boldsymbol{0}_{(M-d)\times(N-d)} \end{bmatrix} \cdot \left[\boldsymbol{V}_p^{[\mathrm{s}](r)}, \quad \boldsymbol{V}_p^{[\mathrm{n}](r)}\right]^{\mathrm{H}} \tag{6.30}$$

$$\left[\boldsymbol{\mathcal{X}}^{(\mathrm{nc},r)}\right]_{(p)} = \left[\hat{\boldsymbol{U}}_p^{[\mathrm{s}](r)}, \quad \hat{\boldsymbol{U}}_p^{[\mathrm{n}](r)}\right] \cdot \begin{bmatrix} \hat{\boldsymbol{\Sigma}}_p^{[\mathrm{s}](r)} & \boldsymbol{0}_{d\times(N-d)} \\ \boldsymbol{0}_{(M-d)\times d} & \hat{\boldsymbol{\Sigma}}_p^{[\mathrm{n}](r)} \end{bmatrix} \cdot \left[\hat{\boldsymbol{V}}_p^{[\mathrm{s}](r)}, \quad \hat{\boldsymbol{V}}_p^{[\mathrm{n}](r)}\right]^{\mathrm{H}}. \tag{6.31}$$

The r-mode augmented signal subspace tensor $\boldsymbol{\mathcal{U}}^{[\mathrm{s}](r)}$ and its estimate $\hat{\boldsymbol{\mathcal{U}}}^{[\mathrm{s}](r)}$ are then defined as

$$\boldsymbol{\mathcal{U}}^{[\mathrm{s}](r)} = \boldsymbol{\mathcal{S}}^{[\mathrm{s}](r)} \times_1 \boldsymbol{U}_1^{[\mathrm{s}](r)} \ldots \times_R \boldsymbol{U}_R^{[\mathrm{s}](r)} \times_{R+1} \boldsymbol{\Sigma}_{R+1}^{[\mathrm{s}](r)^{-1}} \tag{6.32}$$

$$\hat{\boldsymbol{\mathcal{U}}}^{[\mathrm{s}](r)} = \hat{\boldsymbol{\mathcal{S}}}^{[\mathrm{s}](r)} \times_1 \hat{\boldsymbol{U}}_1^{[\mathrm{s}](r)} \ldots \times_R \hat{\boldsymbol{U}}_R^{[\mathrm{s}](r)} \times_{R+1} \hat{\boldsymbol{\Sigma}}_{R+1}^{[\mathrm{s}](r)^{-1}}, \tag{6.33}$$

where the $(R+1)$-mode normalizations by $\boldsymbol{\Sigma}_{R+1}^{[\mathrm{s}](r)^{-1}}$ and $\hat{\boldsymbol{\Sigma}}_{R+1}^{[\mathrm{s}](r)^{-1}}$ are used for notational convenience. The enhanced tensor-based NC signal subspace estimate, which replaces the matrix-based NC subspace estimate $\hat{\boldsymbol{U}}_{\mathrm{s}}^{(\mathrm{nc})}$ is given by $\left[\hat{\boldsymbol{\mathcal{U}}}^{[\mathrm{s}](r)}\right]_{(R+1)}^{\mathrm{T}}$.

The augmented measurement matrix $\boldsymbol{X}^{(\mathrm{nc})(r)} \in \mathbb{C}^{2M\times N}$ corresponding to the augmented measurement tensor $\boldsymbol{\mathcal{X}}^{(\mathrm{nc},r)}$ is obtained via $\boldsymbol{X}^{(\mathrm{nc})(r)} = \left[\boldsymbol{\mathcal{X}}^{(\mathrm{nc},r)}\right]_{(R+1)}^{\mathrm{T}}$. The SVD of $\boldsymbol{X}^{(\mathrm{nc})(r)}$ is given as

$$\boldsymbol{X}^{(\mathrm{nc})(r)} = \left[\hat{\boldsymbol{U}}_{\mathrm{s}}^{(\mathrm{nc})(r)}, \quad \hat{\boldsymbol{U}}_{\mathrm{n}}^{(\mathrm{nc})(r)}\right] \cdot \begin{bmatrix} \hat{\boldsymbol{\Sigma}}_{\mathrm{s}}^{(\mathrm{nc})(r)} & \boldsymbol{0}_{d\times(N-d)} \\ \boldsymbol{0}_{(M-d)\times d} & \hat{\boldsymbol{\Sigma}}_{\mathrm{n}}^{(\mathrm{nc})(r)} \end{bmatrix} \cdot \left[\hat{\boldsymbol{V}}_{\mathrm{s}}^{(\mathrm{nc})(r)}, \quad \hat{\boldsymbol{V}}_{\mathrm{n}}^{(\mathrm{nc})(r)}\right]^{\mathrm{H}}, \tag{6.34}$$

where $\hat{\boldsymbol{U}}_{\mathrm{s}}^{(\mathrm{nc})(r)} \in \mathbb{C}^{2M\times d}$, $\hat{\boldsymbol{U}}_{\mathrm{n}}^{(\mathrm{nc})(r)} \in \mathbb{C}^{2M\times(2M-d)}$, and $\hat{\boldsymbol{V}}_{\mathrm{s}}^{(\mathrm{nc})(r)} \in \mathbb{C}^{N\times d}$ span the estimated signal subspace, the estimated noise subspace, and the estimated row space, respectively, and $\hat{\boldsymbol{\Sigma}}_{\mathrm{s}}^{(\mathrm{nc})(r)} \in \mathbb{R}^{d\times d}$ contains the d largest singular values on its diagonal. Considering (6.34) and the SVD of $[\boldsymbol{\mathcal{X}}]_{(R+1)}^{\mathrm{T}}$ using (6.31), the corresponding subspaces are related through

$$\begin{aligned} &\hat{\boldsymbol{U}}_{\mathrm{s}}^{(\mathrm{nc})(r)} = \hat{\boldsymbol{V}}_{R+1}^{[\mathrm{s}](r)^*}, \ \hat{\boldsymbol{U}}_{\mathrm{n}}^{(\mathrm{nc})(r)} = \hat{\boldsymbol{V}}_{R+1}^{[\mathrm{n}](r)^*}, \ \hat{\boldsymbol{V}}_{\mathrm{s}}^{(\mathrm{nc})(r)} = \hat{\boldsymbol{U}}_{R+1}^{[\mathrm{s}](r)^*}, \\ &\hat{\boldsymbol{V}}_{\mathrm{n}}^{(\mathrm{nc})(r)} = \hat{\boldsymbol{U}}_{R+1}^{[\mathrm{n}](r)^*}, \ \hat{\boldsymbol{\Sigma}}_{\mathrm{s}}^{(\mathrm{nc})(r)} = \hat{\boldsymbol{\Sigma}}_{R+1}^{[\mathrm{s}](r)}. \end{aligned} \tag{6.35}$$

In order to ensure the correct truncation of the estimated q-spaces $\hat{\boldsymbol{U}}_q^{[\mathrm{s}](r)}$ for $q = 1, \ldots, R+1$ required in (6.33), we formulate the following theorem:

Theorem 6.3.3. *If the R-D array is not R-D centro-symmetric, i.e., it is not centro-symmetric in each of the R modes, then the n-ranks of $\boldsymbol{\mathcal{A}}^{(\mathrm{nc},r)}$ can exceed d up to $2d$.*

The proof is given in Appendix B.21. A consequence of Theorem 6.3.3 is that the n-ranks must be taken into account when computing the truncated HOSVD in order to obtain the signal subspace tensor $\boldsymbol{\mathcal{U}}^{[\mathrm{s}](r)}$. If the the array is not centro-symmetric in any mode, the n-ranks are equal to $2d$ and the HOSVD should be truncated to $2d$ in the first R modes, while truncating to d in mode $R+1$.

In order to solve the shift invariance equation in (6.27), the unknown array steering tensor $\boldsymbol{\mathcal{A}}^{(\mathrm{nc},r)}$ in (6.27) can be eliminated using the following relation between $\boldsymbol{\mathcal{A}}^{(\mathrm{nc},r)}$ and $\boldsymbol{\mathcal{U}}^{[\mathrm{s}](r)}$:

Theorem 6.3.4. *The augmented array steering tensor $\boldsymbol{\mathcal{A}}^{(\mathrm{nc},r)}$ and the signal subspace tensor $\boldsymbol{\mathcal{U}}^{[\mathrm{s}](r)}$ are related via*

$$\boldsymbol{\mathcal{A}}^{(\mathrm{nc},r)} = \boldsymbol{\mathcal{U}}^{[\mathrm{s}](r)} \times_{R+1} \boldsymbol{T}, \tag{6.36}$$

where $\boldsymbol{T} \in \mathbb{C}^{d\times d}$ is a non-singular matrix, which is independent of r.

The proof is given in Appendix B.22. The crucial part of this theorem is the fact that $\boldsymbol{T}$ is independent of r, which implies that the matrices $\boldsymbol{\Psi}^{(r)}$ have a common set eigenvectors and the automatic pairing in R-D ESPRIT-type algorithms is ensured.

In the noisy case, the relation (6.36) only holds approximately and after its insertion into the shift invariance equation (6.27), we can write the shift invariance equation in terms of the estimated signal subspace tensor as

$$\hat{\boldsymbol{\mathcal{U}}}^{[\mathrm{s}](r)} \times_r \boldsymbol{J}_1^{(\mathrm{nc})(r)} \times_{R+1} \boldsymbol{\Psi}^{(r)} \approx \hat{\boldsymbol{\mathcal{U}}}^{[\mathrm{s}](r)} \times_r \boldsymbol{J}_2^{(\mathrm{nc})(r)}, \tag{6.37}$$

where $\boldsymbol{\Psi}^{(r)} \approx \boldsymbol{T} \cdot \boldsymbol{\Phi}^{(r)} \cdot \boldsymbol{T}^{-1}$ and $\boldsymbol{J}_1^{(\mathrm{nc})(r)}$ and $\boldsymbol{J}_2^{(\mathrm{nc})(r)}$ are given in (6.8) and (6.9), respectively. From (6.37), the R-D NC Standard Tensor-ESPRIT follows naturally and is summarized in Algorithm 11. Note that in analogy to R-D Standard Tensor-ESPRIT in Section 3.5.1, R-D NC Standard Tensor-ESPRIT provides a tensor gain if $d \leq \min_r\{2M_r\}$.

6.3.3. R-D NC Unitary Tensor-ESPRIT

In this section, we derive the R-D NC Unitary Tensor-ESPRIT algorithm as an extension of the R-D NC Standard Tensor-ESPRIT algorithm presented in the previous subsection. As in the matrix-based version R-D NC Unitary ESPRIT, we first apply forward-backward averaging (FBA) to the augmented measurement tensor $\boldsymbol{\mathcal{X}}^{(\mathrm{nc},r)}$ in (6.24) and then use the real-valued transformation introduced in Section 3.2.3 to obtain a real-valued measurement tensor. In the same way as in (3.20), FBA can be applied to $\boldsymbol{\mathcal{X}}^{(\mathrm{nc},r)}$ by defining the further augmented measurement tensor

$$\boldsymbol{\mathcal{X}}^{(\mathrm{nc},r)(\mathrm{fba})} = \left[\boldsymbol{\mathcal{X}}^{(\mathrm{nc},r)} \,\llcorner_{R+1}\, \left(\boldsymbol{\mathcal{X}}^{(\mathrm{nc},r)*} \bigtimes_{p=1}^{R}{}_p \boldsymbol{\Pi}_{M_p} \times_{R+1} \boldsymbol{\Pi}_N\right)\right], \tag{6.38}$$

Algorithm 11 Summary of R-D NC Standard Tensor-ESPRIT

1. Estimate the augmented signal subspace tensors $\hat{\boldsymbol{\mathcal{U}}}^{[\mathrm{s}](r)} \in \mathbb{C}^{M_1 \times \ldots \times 2M_r \times \ldots \times M_R \times d}$ via the truncated HOSVD of the r-mode augmented observation tensors $\boldsymbol{\mathcal{X}}^{(\mathrm{nc},r)} \in \mathbb{C}^{M_1 \times \ldots \times M_{r-1} \times 2M_r \times M_{r+1} \times \ldots \times M_R \times N}$ in (6.24) according to (6.32) for $r = 1, 2, \ldots, R$.

2. Solve the overdetermined shift invariance equations

$$\hat{\boldsymbol{\mathcal{U}}}^{[\mathrm{s}](r)} \times_r \boldsymbol{J}_1^{(\mathrm{nc})(r)} \times_{R+1} \hat{\boldsymbol{\Psi}}^{(r)} \approx \hat{\boldsymbol{\mathcal{U}}}^{[\mathrm{s}](r)} \times_r \boldsymbol{J}_2^{(\mathrm{nc})(r)}$$

   for the matrices $\hat{\boldsymbol{\Psi}}^{(r)}$ for $r = 1, 2, \ldots, R$ via LS methods, e.g., LS, TLS, SLS, TS-SLS.

3. Compute the eigenvalues $\hat{\lambda}_i^{(r)}$ for $i = 1, 2, \ldots, d$ of $\hat{\boldsymbol{\Psi}}^{(r)}$ jointly for all $r = 1, 2, \ldots, R$, e.g., via the joint diagonalization scheme proposed in [FG06]. Recover the correctly paired frequencies $\hat{\mu}_i^{(r)}$ via $\hat{\mu}_i^{(r)} = \arg\left\{\hat{\lambda}_i^{(r)}\right\}$.

which is processed instead of $\boldsymbol{\mathcal{X}}^{(\mathrm{nc},r)}$. Considering (6.38), many of the results established for the matrix-based version R-D NC Unitary ESPRIT in Section 6.2.3 also carry over to the tensor case. For instance, the augmented array steering tensor $\boldsymbol{\mathcal{A}}^{(\mathrm{nc},r)}$ is always centro-symmetric in the r-th mode even if the original array steering tensor $\boldsymbol{\mathcal{A}}$ is not centro-symmetric the r-th mode as shown in Theorem (6.3.2), R-D NC Unitary Tensor-ESPRIT algorithm is also applicable to non-centro-symmetric arrays. Furthermore, it is easily shown that

$$\left[\boldsymbol{\mathcal{X}}^{(\mathrm{nc},r)(\mathrm{fba})}\right]_{(R+1)}^{\mathrm{T}} \cdot \left(\left[\boldsymbol{\mathcal{X}}^{(\mathrm{nc},r)(\mathrm{fba})}\right]_{(R+1)}^{\mathrm{T}}\right)^{\mathrm{H}} = 2 \cdot \left[\boldsymbol{\mathcal{X}}^{(\mathrm{nc},r)}\right]_{(R+1)}^{\mathrm{T}} \cdot \left(\left[\boldsymbol{\mathcal{X}}^{(\mathrm{nc},r)}\right]_{(R+1)}^{\mathrm{T}}\right)^{\mathrm{H}}. \tag{6.39}$$

This implies that as in the matrix case FBA has no effect on the r-mode augmented tensor $\boldsymbol{\mathcal{X}}^{(\mathrm{nc},r)}$. Recall that in the matrix case in (6.17), we have shown that by using sparse left-$\boldsymbol{\Pi}$-real matrices $\boldsymbol{Q}_p^{(\mathrm{s})}$ (cf. (3.17) or Appendix A.2), the transformed real-valued measurement matrix takes the very simple form in (6.18) with reduced dimensionality. An equivalent simple form of the transformed measurement tensor $\varphi\left(\boldsymbol{\mathcal{X}}^{(\mathrm{nc},r)(\mathrm{fba})}\right)$ can be found by applying (3.21) from Section 3.2.3 and using the sparse matrices $\boldsymbol{Q}_{2M_r}^{(\mathrm{s})}$ in the r-th mode and $\boldsymbol{Q}_{2N}^{(\mathrm{s})}$ to obtain

$$\begin{aligned}
\varphi\left(\boldsymbol{\mathcal{X}}^{(\mathrm{nc},r)(\mathrm{fba})}\right) &= \boldsymbol{\mathcal{X}}^{(\mathrm{nc},r)(\mathrm{fba})} \bigtimes_{\substack{q=1 \\ q \neq r}}^{R}{}_q \, \boldsymbol{Q}_{M_q}^{\mathrm{H}} \times_r \boldsymbol{Q}_{2M_r}^{(\mathrm{s})\mathrm{H}} \times_{R+1} \boldsymbol{Q}_{2N}^{(\mathrm{s})\mathrm{H}} \\
&= \left[\left[2 \cdot \mathrm{Re}\left\{\bar{\boldsymbol{\mathcal{X}}}^{(r)}\right\} \hookrightarrow_r 2 \cdot \mathrm{Im}\left\{\bar{\boldsymbol{\mathcal{X}}}^{(r)}\right\}\right] \hookrightarrow_{R+1} \left[\boldsymbol{\mathcal{O}}_{M_1 \times \ldots \times M_R \times N} \hookrightarrow_r \boldsymbol{\mathcal{O}}_{M_1 \times \ldots \times M_R \times N}\right]\right] \\
&\quad \in \mathbb{R}^{M_1 \times \ldots \times M_{r-1} \times 2M_r \times M_{r+1} \times \ldots \times M_R \times 2N},
\end{aligned} \tag{6.40}$$

Algorithm 12 [RH09] Summary of R-D NC Unitary Tensor-ESPRIT

1. Estimate the real-valued augmented signal subspace tensors $\hat{\boldsymbol{\mathcal{E}}}^{[\mathrm{s}](r)} \in \mathbb{R}^{M_1 \times \ldots \times 2M_r \times \ldots \times M_R \times d}$ via the truncated HOSVD of the real-valued concatenated r-mode observation tensors
$$\left[\mathrm{Re}\left\{\bar{\boldsymbol{\mathcal{X}}}^{(r)}\right\} \hookrightarrow_r \mathrm{Im}\left\{\bar{\boldsymbol{\mathcal{X}}}^{(r)}\right\}\right] \in \mathbb{R}^{M_1 \times \ldots \times M_{r-1} \times 2M_r \times M_{r+1} \times \ldots \times M_R \times N} \tag{6.42}$$
for $r = 1, 2, \ldots, R$.

2. Solve the overdetermined shift invariance equations
$$\hat{\boldsymbol{\mathcal{E}}}^{[\mathrm{s}](r)} \times_r \boldsymbol{K}_1^{(\mathrm{nc})(r)} \times_{R+1} \hat{\boldsymbol{\Upsilon}}^{(r)} \approx \hat{\boldsymbol{\mathcal{E}}}^{[\mathrm{s}](r)} \times_r \boldsymbol{K}_2^{(\mathrm{nc})(r)}$$
for the matrices $\hat{\boldsymbol{\Upsilon}}^{(r)}$ for $r = 1, 2, \ldots, R$ via LS methods, e.g., LS, TLS, SLS, TS-SLS, where
$$\boldsymbol{K}_1^{(\mathrm{nc})(r)} = 2 \cdot \mathrm{Re}\left\{\boldsymbol{Q}_{M_r^{(\mathrm{sel})}}^{\mathrm{H}} \cdot \boldsymbol{J}_2^{(\mathrm{nc})(r)} \cdot \boldsymbol{Q}_{M_r}\right\} \tag{6.43}$$
$$\boldsymbol{K}_2^{(\mathrm{nc})(r)} = 2 \cdot \mathrm{Im}\left\{\boldsymbol{Q}_{M_r^{(\mathrm{sel})}}^{\mathrm{H}} \cdot \boldsymbol{J}_2^{(\mathrm{nc})(r)} \cdot \boldsymbol{Q}_{M_r}\right\} \tag{6.44}$$
and $\boldsymbol{J}_n^{(\mathrm{nc})(r)}$ for $n = 1, 2$ are defined in (6.8) and (6.9).

3. Compute the eigenvalues $\hat{\omega}_i^{(r)}$ for $i = 1, 2, \ldots, d$ of $\hat{\boldsymbol{\Upsilon}}^{(r)}$ jointly for all $r = 1, 2, \ldots, R$, e.g., via the joint diagonalization scheme proposed in [FG06] or via the Simultaneous Schur Decomposition proposed in [HN98]. Recover the correctly paired frequencies $\hat{\mu}_i^{(r)}$ via $\hat{\mu}_i^{(r)} = 2 \cdot \arctan(\hat{\omega}_i^{(r)})$.

where $\bar{\boldsymbol{\mathcal{X}}}^{(r)} = \boldsymbol{\mathcal{X}} \overset{R}{\underset{\substack{q=1 \\ q \neq r}}{\times_q}} \boldsymbol{Q}_{M_q}^{\mathrm{H}}$ and $\boldsymbol{Q}_{M_q}$ are arbitrary unitary left-$\boldsymbol{\Pi}$-real matrices. For the proof, see Appendix B.23. Similarly to the matrix case, the factor 2 and the zero entries in $\varphi\left(\boldsymbol{\mathcal{X}}^{(\mathrm{nc},r)(\mathrm{fba})}\right)$ can be omitted. As a result, it is sufficient to process
$$\left[\mathrm{Re}\left\{\bar{\boldsymbol{\mathcal{X}}}^{(r)}\right\} \hookrightarrow_r \mathrm{Im}\left\{\bar{\boldsymbol{\mathcal{X}}}^{(r)}\right\}\right] \in \mathbb{R}^{M_1 \times \ldots \times M_{r-1} \times 2M_r \times M_{r+1} \times \ldots \times M_R \times N} \tag{6.41}$$
with a reduced column dimension from $2N$ to N.

The remaining steps of the R-D NC Unitary Tensor-ESPRIT algorithm are equivalent to those for R-D Unitary Tensor-ESPRIT in Section 3.5.2. The R-D NC Unitary Tensor-ESPRIT algorithm is summarized in Algorithm 12.

6.4. Performance of R-D matrix-based NC ESPRIT-type algorithms

Along with the development of subspace-based parameter estimation algorithms, their analytical performance evaluation has received considerable attention. A motivation and a detailed overview of the most important performance analysis frameworks is given in Section 4.1. In this section, we present a first-order analytical performance assessment of the matrix-based R-D NC Standard ESPRIT and R-D NC Unitary ESPRIT algorithms using least squares. The derivations are based on the R-D performance analysis framework in [RHD14], which is described in detail in Chapter 4. Note that [RHD14] is an R-D extension of the analytical results for 1-D Standard ESPRIT in [LLV93]. First, we will show that the R-D performance analysis framework is still applicable after the NC preprocessing step in the case of NC signals. Then, we will derive analytical expressions for the parameter estimation error and the MSE. Moreover, we analytically prove in Section 6.4.2 that the performance of R-D NC Standard ESPRIT and R-D NC Unitary ESPRIT is asymptotically identical. Therefore, we start with the simpler derivation of the analytical expressions for R-D NC Standard ESPRIT in Section 6.4.1 and then show their equivalence in Section 6.4.2. These results have been published in [SRHD14].

6.4.1. Performance of R-D NC Standard ESPRIT

For the first-order perturbation analysis of R-D NC Standard ESPRIT, we apply the analytical R-D performance framework proposed in [RHD14] and reviewed in Chapter 4. As [RHD14] adopts a two-step procedure, we first develop a first-order approximation of the signals subspace estimation error in terms of the noise perturbation and then find a corresponding first-order expansion for the parameter estimation error $\Delta\mu_i$.

Consider the model after the NC preprocessing step in (6.4), which is restated as

$$\begin{aligned}\boldsymbol{X}^{(\mathrm{nc})} &= \begin{bmatrix} \boldsymbol{A} \\ \boldsymbol{\Pi}_M \cdot \boldsymbol{A}^* \cdot \boldsymbol{\Psi}^* \cdot \boldsymbol{\Psi}^* \end{bmatrix} \cdot \boldsymbol{S} + \begin{bmatrix} \boldsymbol{N} \\ \boldsymbol{\Pi}_M \cdot \boldsymbol{N}^* \end{bmatrix} \\ &= \boldsymbol{A}^{(\mathrm{nc})} \cdot \boldsymbol{S} + \boldsymbol{N}^{(\mathrm{nc})} = \boldsymbol{X}_0^{(\mathrm{nc})} + \boldsymbol{N}^{(\mathrm{nc})} \in \mathbb{C}^{2M \times N}. \end{aligned} \tag{6.45}$$

It is evident from (6.45) that the NC preprocessing does not violate the assumption from [LLV93] that the noise perturbation $\boldsymbol{N}^{(\mathrm{nc})}$ is small compared to the signal component. Hence, we can directly apply the framework of [LLV93, RHD14] to the augmented measurement matrix $\boldsymbol{X}^{(\mathrm{nc})}$ in (6.45). The resulting expressions are asymptotic in the high effective SNR and explicit in the noise term $\boldsymbol{N}^{(\mathrm{nc})}$.

Starting with the subspace error expression based on (6.45), we can express the SVD of the

noise-free augmented observations $\boldsymbol{X}_0^{(\mathrm{nc})}$ as

$$\boldsymbol{X}_0^{(\mathrm{nc})} = \begin{bmatrix} \boldsymbol{U}_\mathrm{s}^{(\mathrm{nc})} & \boldsymbol{U}_\mathrm{n}^{(\mathrm{nc})} \end{bmatrix} \cdot \begin{bmatrix} \boldsymbol{\Sigma}_\mathrm{s}^{(\mathrm{nc})} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} \end{bmatrix} \cdot \begin{bmatrix} \boldsymbol{V}_\mathrm{s}^{(\mathrm{nc})} & \boldsymbol{V}_\mathrm{n}^{(\mathrm{nc})} \end{bmatrix}^\mathrm{H},$$

where $\boldsymbol{U}_\mathrm{s}^{(\mathrm{nc})} \in \mathbb{C}^{2M \times d}$, $\boldsymbol{U}_\mathrm{n}^{(\mathrm{nc})} \in \mathbb{C}^{2M \times (2M-d)}$, and $\boldsymbol{V}_\mathrm{s}^{(\mathrm{nc})} \in \mathbb{C}^{N \times d}$ span the signal subspace, the noise subspace, and the row space respectively, and $\boldsymbol{\Sigma}_\mathrm{s}^{(\mathrm{nc})} \in \mathbb{R}^{d \times d}$ contains the non-zero singular values on its diagonal. Next, we write the perturbed signal subspace estimate $\hat{\boldsymbol{U}}_\mathrm{s}^{(\mathrm{nc})}$ from (6.12) in Section 6.2.2 as $\hat{\boldsymbol{U}}_\mathrm{s}^{(\mathrm{nc})} = \boldsymbol{U}_\mathrm{s}^{(\mathrm{nc})} + \Delta\boldsymbol{U}_\mathrm{s}^{(\mathrm{nc})}$, where $\Delta\boldsymbol{U}_\mathrm{s}^{(\mathrm{nc})}$ denotes the signal subspace estimation error. Then, following (4.4), we directly obtain the first-order subspace error approximation

$$\Delta\boldsymbol{U}_\mathrm{s}^{(\mathrm{nc})} = \boldsymbol{U}_\mathrm{n}^{(\mathrm{nc})} \cdot \boldsymbol{U}_\mathrm{n}^{(\mathrm{nc})^\mathrm{H}} \cdot \boldsymbol{N}^{(\mathrm{nc})} \cdot \boldsymbol{V}_\mathrm{s}^{(\mathrm{nc})} \cdot \boldsymbol{\Sigma}_\mathrm{s}^{(\mathrm{nc})^{-1}} + \mathcal{O}\{\nu^2\}, \tag{6.46}$$

where $\nu = \|\boldsymbol{N}^{(\mathrm{nc})}\|$, and $\|\cdot\|$ represents an arbitrary sub-multiplicative norm. As in Section 4.2.1, we drop the second-order (SO) term $\mathcal{O}\{\nu^2\}$ in what follows and write "≈" to refer to "up to the first order". Equation (6.46) models the leakage of the noise subspace into the signal subspace due to the augmented noise $\boldsymbol{N}^{(\mathrm{nc})}$. The perturbation of the particular basis for the signal subspace $\boldsymbol{U}_\mathrm{s}^{(\mathrm{nc})}$, which is taken into account in [LLM08] can be ignored as the choice of this basis is irrelevant for R-D NC Standard ESPRIT (cf. Section 4.2.1).

For the parameter estimation error of R-D NC Standard ESPRIT for the i-th spatial frequency in the r-th mode obtained by the LS solution in (6.13), we follow the derivation in Section 4.3 to obtain

$$\Delta\mu_i^{(r)} \approx \mathrm{Im}\left\{\boldsymbol{p}_i^\mathrm{T} \cdot \left(\tilde{\boldsymbol{J}}_1^{(\mathrm{nc})(r)} \cdot \boldsymbol{U}_\mathrm{s}^{(\mathrm{nc})}\right)^+ \cdot \left[\tilde{\boldsymbol{J}}_2^{(\mathrm{nc})(r)}/\lambda_i^{(r)} - \tilde{\boldsymbol{J}}_1^{(\mathrm{nc})(r)}\right] \cdot \Delta\boldsymbol{U}_\mathrm{s}^{(\mathrm{nc})} \cdot \boldsymbol{q}_i\right\}, \tag{6.47}$$

where $\lambda_i^{(r)} = \mathrm{e}^{\mathrm{j}\mu_i^{(r)}}$ is the i-th eigenvalue of $\boldsymbol{\Psi}^{(r)}$ in the r-th mode, $\boldsymbol{q}_i$ represents the i-th eigenvector of $\boldsymbol{\Psi}^{(r)}$, i.e., the i-th column vector of the eigenvector matrix $\boldsymbol{Q}$, and $\boldsymbol{p}_i^\mathrm{T}$ is the i-th row vector of $\boldsymbol{P} = \boldsymbol{Q}^{-1}$. Hence, the eigendecomposition of $\boldsymbol{\Psi}^{(r)}$ in the r-th mode is given by

$$\boldsymbol{\Psi}^{(r)} = \boldsymbol{Q} \cdot \boldsymbol{\Lambda}^{(r)} \cdot \boldsymbol{Q}^{-1}, \tag{6.48}$$

where $\boldsymbol{\Lambda}^{(r)}$ contains the eigenvalues $\lambda_i^{(r)}$ on its diagonal. Then, by inserting (6.46) into (6.47) similarly to (4.13) in Section 4.3.1, we can write the first-order approximation of the estimation error $\Delta\mu_i^{(r)}$ explicitly in terms of the vectorized noise perturbation $\boldsymbol{n}^{(\mathrm{nc})} = \mathrm{vec}\left\{\boldsymbol{N}^{(\mathrm{nc})}\right\} \in \mathbb{C}^{2MN \times 1}$ as

$$\Delta\mu_i^{(r)} \approx \mathrm{Im}\left\{\boldsymbol{r}_i^{(\mathrm{nc})(r)^\mathrm{T}} \cdot \mathrm{vec}\left\{\Delta\boldsymbol{U}_\mathrm{s}^{(\mathrm{nc})}\right\}\right\} = \mathrm{Im}\left\{\boldsymbol{r}_i^{(\mathrm{nc})(r)^\mathrm{T}} \cdot \boldsymbol{W}_\mathrm{mat}^{(\mathrm{nc})} \cdot \boldsymbol{n}^{(\mathrm{nc})}\right\}, \tag{6.49}$$

where $\boldsymbol{r}_i^{(\mathrm{nc})(r)} \in \mathbb{C}^{2Md\times 1}$ and $\boldsymbol{W}_{\mathrm{mat}}^{(\mathrm{nc})} \in \mathbb{C}^{2Md\times 2MN}$ are given by

$$\boldsymbol{r}_i^{(\mathrm{nc})(r)} = \boldsymbol{q}_i \otimes \left(\left[\left(\tilde{\boldsymbol{J}}_1^{(\mathrm{nc})(r)} \cdot \boldsymbol{U}_{\mathrm{s}}^{(\mathrm{nc})} \right)^{+} \cdot \left(\tilde{\boldsymbol{J}}_2^{(\mathrm{nc})(r)} / \lambda_i^{(r)} - \tilde{\boldsymbol{J}}_1^{(\mathrm{nc})(r)} \right) \right]^{\mathrm{T}} \cdot \boldsymbol{p}_i \right) \tag{6.50}$$

$$\boldsymbol{W}_{\mathrm{mat}}^{(\mathrm{nc})} = \left(\boldsymbol{\Sigma}_{\mathrm{s}}^{(\mathrm{nc})^{-1}} \cdot \boldsymbol{V}_{\mathrm{s}}^{(\mathrm{nc})^{\mathrm{T}}} \right) \otimes \left(\boldsymbol{U}_{\mathrm{n}}^{(\mathrm{nc})} \cdot \boldsymbol{U}_{\mathrm{n}}^{(\mathrm{nc})^{\mathrm{H}}} \right). \tag{6.51}$$

In order to derive an analytical expression for the MSE of R-D NC Standard ESPRIT, we apply the concept of [RHD14] in (4.17), which presents an MSE expression that only depends on the SO statistics of the noise, i.e., the covariance matrix and the pseudo-covariance matrix, assuming the noise to be zero-mean. As the preprocessing in (6.45) does not violate the zero-mean assumption, [RHD14] is applicable once the corresponding SO statistics are found. Therefore, defining the covariance matrix $\boldsymbol{R}_{\mathrm{nn}}^{(\mathrm{nc})} = \mathbb{E}\left\{ \boldsymbol{n}^{(\mathrm{nc})} \cdot \boldsymbol{n}^{(\mathrm{nc})^{\mathrm{H}}} \right\} \in \mathbb{C}^{2MN\times 2MN}$ and the pseudo-covariance matrix $\boldsymbol{C}_{\mathrm{nn}}^{(\mathrm{nc})} = \mathbb{E}\left\{ \boldsymbol{n}^{(\mathrm{nc})} \cdot \boldsymbol{n}^{(\mathrm{nc})^{\mathrm{T}}} \right\} \in \mathbb{C}^{2MN\times 2MN}$ of $\boldsymbol{n}^{(\mathrm{nc})}$, the MSE of R-D NC Standard ESPRIT for the i-th spatial frequency in the r-th mode is given by

$$\mathbb{E}\left\{ \left(\Delta\mu_i^{(r)} \right)^2 \right\} \approx \frac{1}{2} \cdot \left(\boldsymbol{z}_i^{(\mathrm{nc})(r)^{\mathrm{H}}} \cdot \boldsymbol{R}_{\mathrm{nn}}^{(\mathrm{nc})^{\mathrm{T}}} \cdot \boldsymbol{z}_i^{(\mathrm{nc})(r)} - \mathrm{Re}\left\{ \boldsymbol{z}_i^{(\mathrm{nc})(r)^{\mathrm{T}}} \cdot \boldsymbol{C}_{\mathrm{nn}}^{(\mathrm{nc})^{\mathrm{T}}} \cdot \boldsymbol{z}_i^{(\mathrm{nc})(r)} \right\} \right), \tag{6.52}$$

where $\boldsymbol{z}_i^{(\mathrm{nc})(r)} = \boldsymbol{W}_{\mathrm{mat}}^{(\mathrm{nc})^{\mathrm{T}}} \cdot \boldsymbol{r}_i^{(\mathrm{nc})(r)} \in \mathbb{C}^{2MN\times 1}$ for $i = 1, \ldots, d$ and $r = 1, \ldots, R$.

In the next step, we derive the covariance matrix and the pseudo-covariance matrix of the augmented noise contribution $\boldsymbol{n}^{(\mathrm{nc})}$ required in (6.52). The result is stated in the following theorem:

Theorem 6.4.1. *In the case of zero-mean noise $\boldsymbol{n} = \mathrm{vec}\{\boldsymbol{N}\} \in \mathbb{C}^{MN\times 1}$ with finite SO moments, the covariance and pseudo covariance matrices of $\boldsymbol{n}^{(\mathrm{nc})} \in \mathbb{C}^{2MN\times 1}$ are given by*

$$\boldsymbol{R}_{\mathrm{nn}}^{(\mathrm{nc})} = \mathbb{E}\left\{ \boldsymbol{n}^{(\mathrm{nc})} \cdot \boldsymbol{n}^{(\mathrm{nc})^{\mathrm{H}}} \right\} = \tilde{\boldsymbol{K}} \cdot \begin{bmatrix} \boldsymbol{R}_{\mathrm{nn}} & \boldsymbol{C}_{\mathrm{nn}} \\ \boldsymbol{C}_{\mathrm{nn}}^{*} & \boldsymbol{R}_{\mathrm{nn}}^{*} \end{bmatrix} \cdot \tilde{\boldsymbol{K}}^{\mathrm{H}} \tag{6.53}$$

$$\boldsymbol{C}_{\mathrm{nn}}^{(\mathrm{nc})} = \mathbb{E}\left\{ \boldsymbol{n}^{(\mathrm{nc})} \cdot \boldsymbol{n}^{(\mathrm{nc})^{\mathrm{T}}} \right\} = \tilde{\boldsymbol{K}} \cdot \begin{bmatrix} \boldsymbol{C}_{\mathrm{nn}} & \boldsymbol{R}_{\mathrm{nn}} \\ \boldsymbol{R}_{\mathrm{nn}}^{*} & \boldsymbol{C}_{\mathrm{nn}}^{*} \end{bmatrix} \cdot \tilde{\boldsymbol{K}}^{\mathrm{T}}, \tag{6.54}$$

where $\boldsymbol{R}_{\mathrm{nn}} = \mathbb{E}\left\{ \boldsymbol{n} \cdot \boldsymbol{n}^{\mathrm{H}} \right\} \in \mathbb{C}^{MN\times MN}$ and $\boldsymbol{C}_{\mathrm{nn}} = \mathbb{E}\left\{ \boldsymbol{n} \cdot \boldsymbol{n}^{\mathrm{T}} \right\} \in \mathbb{C}^{MN\times MN}$ are the covariance matrix and the pseudo covariance matrix of the physical noise $\boldsymbol{n}$, respectively.

The proof is given in Appendix B.24. Thus, the SO moments of the augmented noise $\boldsymbol{n}^{(\mathrm{nc})}$ can be expressed in terms of the SO statistics of $\boldsymbol{n}$. In the special case of circularly symmetric white noise with $\boldsymbol{R}_{\mathrm{nn}} = \sigma_{\mathrm{n}}^2 \cdot \boldsymbol{I}_{MN}$ and $\boldsymbol{C}_{\mathrm{nn}} = \boldsymbol{0}_{MN}$, the SO moments in (6.53) and (6.54) simplify to

$$\boldsymbol{R}_{\mathrm{nn}}^{(\mathrm{nc})} = \sigma_{\mathrm{n}}^2 \cdot \boldsymbol{I}_{2MN} \quad \text{and} \quad \boldsymbol{C}_{\mathrm{nn}}^{(\mathrm{nc})} = \sigma_{\mathrm{n}}^2 \cdot (\boldsymbol{I}_N \otimes \boldsymbol{\Pi}_{2M}). \tag{6.55}$$

Note that the pseudo-covariance matrix $\boldsymbol{C}_{\mathrm{nn}}^{(\mathrm{nc})}$ is always non-zero, even in the case of circularly

symmetric white noise. This is due to the NC preprocessing step in (6.45). Furthermore, it is worth mentioning that the step of solving the R augmented shift invariance equations for $\boldsymbol{\Psi}^{(r)}$ independently and then performing a joint eigendecomposition across all R dimensions to obtain $\boldsymbol{\Lambda}^{(r)}$ has no impact on the asymptotic estimation error for high SNRs since the eigenvectors become asymptotically equal [RHD14, BCW+17].

Inserting (6.55) into (6.52), the MSE of R-D NC Standard ESPRIT assuming circularly symmetric white noise for the i-th spatial frequency in the r-th mode becomes

$$\mathbb{E}\left\{\left(\Delta\mu_i^{(r)}\right)^2\right\} \approx \frac{\sigma_{\mathrm{n}}^2}{2} \cdot \left(\left\|\boldsymbol{z}_i^{(\mathrm{nc})(r)}\right\|_2^2 - \mathrm{Re}\left\{\boldsymbol{z}_i^{(\mathrm{nc})(r)^{\mathrm{T}}} \cdot (\boldsymbol{I}_N \otimes \boldsymbol{\Pi}_{2M}) \cdot \boldsymbol{z}_i^{(\mathrm{nc})(r)}\right\}\right), \tag{6.56}$$

where $\boldsymbol{z}_i^{(\mathrm{nc})(r)}$ is given below (6.52).

6.4.2. Performance of R-D NC Unitary ESPRIT

In the previous section, we have derived the analytical first-order approximation of the parameter estimation error and the corresponding analytical MSE expression of R-D NC Standard ESPRIT. In this section, we show that the analytical performance of R-D NC Unitary ESPRIT and R-D NC Standard ESPRIT is identical in the high effective SNR regime. To this end, we recall from Section 6.2.3 that R-D NC Unitary ESPRIT includes forward-backward averaging (FBA) in (6.14) as well as the transformation into the real-valued domain in (6.17) as preprocessing steps.

We have already established in (6.16) that the FBA-processed augmented measurement matrix $\boldsymbol{X}^{(\mathrm{nc})^{(\mathrm{fba})}}$ in (6.15) satisfies

$$\boldsymbol{X}^{(\mathrm{nc})^{(\mathrm{fba})}} \cdot \boldsymbol{X}^{(\mathrm{nc})^{(\mathrm{fba})^{\mathrm{H}}}} = 2 \cdot \boldsymbol{X}^{(\mathrm{nc})} \cdot \boldsymbol{X}^{(\mathrm{nc})^{\mathrm{H}}}. \tag{6.57}$$

Thus, the column space of $\boldsymbol{X}^{(\mathrm{nc})^{(\mathrm{fba})}}$ is the same as the column space of $\boldsymbol{X}^{(\mathrm{nc})}$, which implies that FBA does not improve the signal subspace.

Next, we analyze the real-valued transformation as the second preprocessing step of R-D NC Unitary ESPRIT and formulate the following theorem:

Theorem 6.4.2. *R-D NC Unitary ESPRIT and R-D NC Standard ESPRIT with FBA preprocessing perform asymptotically identical in the high effective SNR.*

The proof is shown in Appendix B.25.

As a result of (6.57) and Theorem 6.4.2, we can conclude that the performance of R-D NC Standard ESPRIT and R-D NC Unitary ESPRIT is asymptotically identical in the high effective SNR. However, in practice, R-D NC Unitary ESPRIT is preferable due to its lower computational complexity and its better performance at a low SNR and for a small number of snapshots.

6.5. Performance of R-D tensor-based NC ESPRIT-type algorithms

In this section, we derive a first-order analytical performance assessment of the tensor-based R-D NC Standard ESPRIT and R-D NC Unitary ESPRIT algorithms using least squares. The resulting analytical expressions for the parameter estimation error and the MSE are derived using the tensor-based R-D performance analysis framework from [RHD14], which is still applicable after the tensor-based NC preprocessing step. Similarly to the matrix case in Section 6.4, we prove that both R-D NC Standard Tensor-ESPRIT and R-D NC Unitary Tensor-ESPRIT perform asymptotically identical in the high effective SNR. To this end, we first derive the analytical expressions for R-D NC Standard Tensor-ESPRIT and then prove the equivalence of the expressions for R-D NC Unitary Tensor-ESPRIT. Note that the 2-D case of both algorithms has already been presented by us in [SRH16b].

6.5.1. Performance of R-D NC Standard Tensor-ESPRIT

For the derivation of the first-order performance analysis expressions of R-D NC Standard Tensor-ESPRIT using LS, we adopt the R-D performance framework from [RHD14], which is derived for the tensor case and described in Section 4.4. Note that this framework is still applicable in the NC case as the augmented signal model in (6.25)

$$\boldsymbol{\mathcal{X}}^{(\mathrm{nc},r)} = \boldsymbol{\mathcal{X}}_0^{(\mathrm{nc},r)} + \boldsymbol{\mathcal{N}}^{(\mathrm{nc},r)} \tag{6.58}$$

still has the same properties and is written in the same form as (4.5) in Section 4.2.2. However, the extension of the tensor-based performance analysis framework to the NC case is not straightforward due to the joint processing of the R augmented shift invariance equations that contain the NC augmentations in each of the R dimensions individually.

We start by deriving the relation between the improved HOSVD-based signal subspace estimate $\left[\hat{\boldsymbol{\mathcal{U}}}^{[\mathrm{s}](r)}\right]_{(R+1)}^{\mathrm{T}}$ in (6.33) and the corresponding SVD-based subspace estimate $\hat{\boldsymbol{U}}_{\mathrm{s}}^{(\mathrm{nc})(r)}$ obtained from the SVD of $\boldsymbol{X}^{(\mathrm{nc})(r)} = \left[\boldsymbol{\mathcal{X}}^{(\mathrm{nc},r)}\right]_{(R+1)}^{\mathrm{T}}$ in (6.34). The result is shown in the following theorem:

Theorem 6.5.1. *The HOSVD-based augmented signal subspace estimate* $\left[\hat{\boldsymbol{\mathcal{U}}}^{[\mathrm{s}](r)}\right]_{(R+1)}^{\mathrm{T}}$ *can be computed from the SVD-based subspace estimate* $\hat{\boldsymbol{U}}_{\mathrm{s}}^{(\mathrm{nc})(r)}$ *via the following relation*

$$\left[\hat{\boldsymbol{\mathcal{U}}}^{[\mathrm{s}](r)}\right]_{(R+1)}^{\mathrm{T}} = \left(\hat{\boldsymbol{T}}_{1:r-1}^{(r)\otimes} \otimes \hat{\boldsymbol{T}}_{r}^{(\mathrm{nc},r)} \otimes \hat{\boldsymbol{T}}_{r+1:R}^{(r)\otimes}\right) \cdot \hat{\boldsymbol{U}}_{\mathrm{s}}^{(\mathrm{nc})(r)}, \tag{6.59}$$

where $\hat{\boldsymbol{T}}_{a:b}^{(r)\otimes}$ *is defined as*

$$\hat{\boldsymbol{T}}_{a:b}^{(r)\otimes} = \begin{cases} \hat{\boldsymbol{T}}_{a}^{(r)} \otimes \ldots \otimes \hat{\boldsymbol{T}}_{b}^{(r)} & a \leq b \\ 1 & a > b \end{cases} \tag{6.60}$$

and $\hat{\boldsymbol{T}}_{q}^{(r)} = \hat{\boldsymbol{U}}_{q}^{[\mathrm{s}](r)} \cdot \hat{\boldsymbol{U}}_{q}^{[\mathrm{s}](r)\mathrm{H}} \in \mathbb{C}^{M_q \times M_q}$ *for* $q = 1, \ldots, r-q, r+1, \ldots, R$ *and* $\hat{\boldsymbol{T}}_{r}^{(\mathrm{nc},r)} = \hat{\boldsymbol{U}}_{r}^{[\mathrm{s}](r)} \cdot \hat{\boldsymbol{U}}_{r}^{[\mathrm{s}](r)\mathrm{H}} \in \mathbb{C}^{2M_r \times 2M_r}$ *are the projection matrices onto the estimated* q*-spaces and the* r*-space of* $\boldsymbol{\mathcal{X}}^{(\mathrm{nc},r)}$.

For the proof, the reader is referred to Appendix B.26. Note that this result is very similar to that in the non-NC counterpart in (3.15) as the multi-dimensional structure is enforced onto the matrix-based subspace estimate $\hat{\boldsymbol{U}}_{\mathrm{s}}^{(\mathrm{nc})(r)}$ via the Kronecker product of the projection matrices. It should also be highlighted that in contrast to the non-NC tensor case in (3.15), where an overall improved HOSVD-based subspace estimate is obtained, we now have R improved HOSVD-based subspace estimates in each of the individual modes. Therefore, a separate perturbation analysis in each of the R modes is required to evaluate the analytical performance in the NC tensor case.

As a result of (6.59), the first-order performance evaluation expression of $\left[\hat{\boldsymbol{\mathcal{U}}}^{[\mathrm{s}](r)}\right]_{(R+1)}^{\mathrm{T}}$ can be derived in the same way as in the non-NC case in (4.10). The result is stated in the theorem:

Theorem 6.5.2. *The first-order performance expansion for* $\left[\hat{\boldsymbol{\mathcal{U}}}^{[\mathrm{s}](r)}\right]_{(R+1)}^{\mathrm{T}}$ *is given by*

$$\left[\hat{\boldsymbol{\mathcal{U}}}^{[\mathrm{s}](r)}\right]_{(R+1)}^{\mathrm{T}} = \boldsymbol{U}_{\mathrm{s}}^{(\mathrm{nc},r)} + \left[\Delta\boldsymbol{\mathcal{U}}^{[\mathrm{s}](r)}\right]_{(R+1)}^{\mathrm{T}}, \tag{6.61}$$

where

$$\begin{aligned} \left[\Delta\boldsymbol{\mathcal{U}}^{[\mathrm{s}](r)}\right]_{(R+1)}^{\mathrm{T}} &= \left(\boldsymbol{T}_{1:r-1}^{(r)\otimes} \otimes \boldsymbol{T}_{r}^{(\mathrm{nc},r)} \otimes \boldsymbol{T}_{r+1:R}^{(r)\otimes}\right) \cdot \Delta\boldsymbol{U}_{\mathrm{s}}^{(\mathrm{nc},r)} \\ &+ \Bigg[\sum_{q=1}^{r-1} \left(\boldsymbol{T}_{1:q-1}^{(r)\otimes} \otimes \left[\Delta\boldsymbol{U}_{q}^{[\mathrm{s}](r)} \cdot \boldsymbol{U}_{q}^{[\mathrm{s}](r)\mathrm{H}}\right] \otimes \boldsymbol{T}_{q+1:R}^{(\mathrm{nc},r)\otimes}\right) + \boldsymbol{T}_{1:r-1}^{(r)\otimes} \otimes \left[\Delta\boldsymbol{U}_{r}^{[\mathrm{s}](r)} \cdot \boldsymbol{U}_{r}^{[\mathrm{s}](r)\mathrm{H}}\right] \otimes \boldsymbol{T}_{r+1:R}^{(r)\otimes} \\ &+ \sum_{q=r+1}^{R} \boldsymbol{T}_{1:q-1}^{(\mathrm{nc},r)\otimes} \otimes \left[\Delta\boldsymbol{U}_{q}^{[\mathrm{s}](r)} \cdot \boldsymbol{U}_{q}^{[\mathrm{s}](r)\mathrm{H}}\right] \otimes \boldsymbol{T}_{q+1:R}^{(r)\otimes} \Bigg] \cdot \boldsymbol{U}_{\mathrm{s}}^{(\mathrm{nc},r)}. \end{aligned} \tag{6.62}$$

Note that $\boldsymbol{T}_{a:b}^{(r)\otimes}$ *is the noise-free version of* (6.60) *and* $\boldsymbol{T}_{a:b}^{(\mathrm{nc},r)\otimes}$ *denotes that the NC augmented projection matrix* $\boldsymbol{T}_{r}^{(\mathrm{nc},r)}$ *is contained in* (6.60). *Moreover, the SVD-based signal subspace perturbation* $\Delta\boldsymbol{U}_{\mathrm{s}}^{(\mathrm{nc},r)}$ *is obtained as*

$$\Delta\boldsymbol{U}_{\mathrm{s}}^{(\mathrm{nc},r)} = \boldsymbol{U}_{\mathrm{n}}^{(\mathrm{nc},r)} \cdot \boldsymbol{U}_{\mathrm{n}}^{(\mathrm{nc},r)\mathrm{H}} \cdot \left[\boldsymbol{\mathcal{N}}^{(\mathrm{nc},r)}\right]_{(R+1)}^{\mathrm{T}} \cdot \boldsymbol{V}_{\mathrm{s}}^{(\mathrm{nc},r)} \cdot \boldsymbol{\Sigma}_{\mathrm{s}}^{(\mathrm{nc},r)^{-1}} \tag{6.63}$$

and the perturbation of the p-space $\Delta\boldsymbol{U}_p^{[\mathrm{s}](r)}$ *for* $p = 1, \ldots, R$ *can be computed via*

$$\Delta\boldsymbol{U}_p^{[\mathrm{s}](r)} = \boldsymbol{U}_p^{[\mathrm{n}](r)} \cdot \boldsymbol{U}_p^{[\mathrm{n}](r)^{\mathrm{H}}} \cdot \left[\boldsymbol{\mathcal{N}}^{(\mathrm{nc},r)}\right]_{(p)} \cdot \boldsymbol{V}_p^{[\mathrm{s}](r)} \cdot \boldsymbol{\Sigma}_p^{[\mathrm{s}](r)^{-1}}. \tag{6.64}$$

The proof is provided in Appendix B.27.

Then, the first-order parameter estimation error $\Delta\mu_i^{(r)}$ of the i-th spatial frequency in the r-th mode is given by

$$\Delta\mu_i^{(r)} \approx \mathrm{Im}\left\{\boldsymbol{p}_i^{\mathrm{T}} \cdot \left(\tilde{\boldsymbol{J}}_1^{(\mathrm{nc})(r)} \cdot \boldsymbol{U}_{\mathrm{s}}^{(\mathrm{nc},r)}\right)^{+} \cdot \left[\tilde{\boldsymbol{J}}_2^{(\mathrm{nc})(r)}/\lambda_i^{(r)} - \tilde{\boldsymbol{J}}_1^{(\mathrm{nc})(r)}\right] \cdot \left[\Delta\boldsymbol{\mathcal{U}}^{[\mathrm{s}](r)}\right]_{(R+1)}^{\mathrm{T}} \cdot \boldsymbol{q}_i\right\}, \tag{6.65}$$

where $\lambda_i^{(r)} = \mathrm{e}^{\mathrm{j}\mu_i^{(r)}}$ and the vectors $\boldsymbol{q}_i$ and $\boldsymbol{p}_i$ denote the respective i-th columns of the matrices $\boldsymbol{Q}$ and $\boldsymbol{P} = \boldsymbol{Q}^{-1}$, which result from the eigendecomposition $\boldsymbol{\Psi}^{(r)} = \boldsymbol{Q} \cdot \boldsymbol{\Lambda}^{(r)} \cdot \boldsymbol{Q}^{-1}$. It is important to emphasize that $\boldsymbol{Q}$ is independent of r, which was already established in Theorem 6.3.4. As a result, the R augmented shift invariance equations are connected via $\boldsymbol{Q}$, i.e., they share the same set of eigenvectors. This enables the correct pairing of the estimates.

Finally, the first-order approximation for the MSE of the R-D NC Standard Tensor-ESPRIT algorithm for the i-th spatial frequency in the r-th mode is given by

$$\mathbb{E}\left\{\left(\Delta\mu_i^{(r)}\right)^2\right\} \approx \frac{1}{2} \cdot \left(\boldsymbol{z}_i^{(\mathrm{nc},r)^{\mathrm{H}}} \cdot \boldsymbol{R}_{\mathrm{nn}}^{(\mathrm{nc},r)^{\mathrm{T}}} \cdot \boldsymbol{z}_i^{(\mathrm{nc},r)} - \mathrm{Re}\left\{\boldsymbol{z}_i^{(\mathrm{nc},r)^{\mathrm{T}}} \cdot \boldsymbol{C}_{\mathrm{nn}}^{(\mathrm{nc},r)^{\mathrm{T}}} \cdot \boldsymbol{z}_i^{(\mathrm{nc},r)}\right\}\right), \tag{6.66}$$

where $\boldsymbol{z}_i^{(\mathrm{nc},r)} = \boldsymbol{W}_{\mathrm{ten}}^{(\mathrm{nc},r)^{\mathrm{T}}} \cdot \boldsymbol{r}_i^{(\mathrm{nc},r)}$ with

$$\boldsymbol{r}_i^{(\mathrm{nc},r)} = \boldsymbol{q}_i \otimes \left(\left[\left(\tilde{\boldsymbol{J}}_1^{(\mathrm{nc})(r)} \cdot \boldsymbol{U}_{\mathrm{s}}^{(\mathrm{nc},r)}\right)^{+} \cdot \left(\tilde{\boldsymbol{J}}_2^{(\mathrm{nc})(r)}/\lambda_i^{(r)} - \tilde{\boldsymbol{J}}_1^{(\mathrm{nc})(r)}\right)\right]^{\mathrm{T}} \cdot \boldsymbol{p}_i\right) \tag{6.67}$$

and

$$\boldsymbol{W}_{\mathrm{ten}}^{(\mathrm{nc},r)} = \boldsymbol{W}_0^{(\mathrm{nc},r)} + \sum_{q=1}^{R} \boldsymbol{W}_q^{(\mathrm{nc},r)} \cdot \boldsymbol{P}_{M_1,\ldots,M_{r-1},2M_r,M_{r+1},\ldots,M_R,N}^{(q)^{\mathrm{T}}} \cdot \boldsymbol{P}_{M_1,\ldots,M_{r-1},2M_r,M_{r+1},\ldots,M_R,N}^{(R)}. \tag{6.68}$$

The matrix $\boldsymbol{W}_0^{(\mathrm{nc},r)}$ is defined as

$$\boldsymbol{W}_0^{(\mathrm{nc},r)} = \left(\boldsymbol{\Sigma}_{\mathrm{s}}^{(\mathrm{nc},r)^{-1}} \cdot \boldsymbol{V}_{\mathrm{s}}^{(\mathrm{nc},r)^{\mathrm{T}}}\right) \otimes \left(\left(\hat{\boldsymbol{T}}_{1:r-1}^{(r)\otimes} \otimes \hat{\boldsymbol{T}}_r^{(\mathrm{nc},r)} \otimes \hat{\boldsymbol{T}}_{r+1:R}^{(r)\otimes}\right) \cdot \boldsymbol{U}_{\mathrm{n}}^{(\mathrm{nc},r)} \cdot \boldsymbol{U}_{\mathrm{n}}^{(\mathrm{nc},r)^{\mathrm{H}}}\right) \tag{6.69}$$

and the matrices $\boldsymbol{W}_q^{(\mathrm{nc},r)}$ are computed according to the following three cases, which depend on the index q. For $q < r$, we have

$$\boldsymbol{W}_q^{(\mathrm{nc},r)} = \left(\boldsymbol{U}_{\mathrm{s}}^{(\mathrm{nc},r)^{\mathrm{T}}} \otimes \boldsymbol{I}_{2\cdot M}\right) \cdot \left(\bar{\boldsymbol{T}}_{1:q-1}^{(\mathrm{nc},r)} \otimes \boldsymbol{I}_{2\cdot M_{q:R}}\right) \cdot \left(\boldsymbol{I}_{M_q} \otimes \bar{\boldsymbol{T}}_{q+1:R}^{(\mathrm{nc},r)}\right) \tag{6.70}$$

$$\cdot \left[\left(\boldsymbol{V}_q^{[\mathrm{s}](r)} \cdot \boldsymbol{\Sigma}_q^{[\mathrm{s}](r)^{-1}} \cdot \boldsymbol{U}_q^{[\mathrm{s}](r)^{\mathrm{H}}} \right)^{\mathrm{T}} \otimes \left(\boldsymbol{U}_q^{[\mathrm{n}](r)} \cdot \boldsymbol{U}_q^{[\mathrm{n}](r)^{\mathrm{H}}} \right) \right] \tag{6.71}$$

with

$$\bar{\boldsymbol{T}}_{1:q-1}^{(\mathrm{nc},r)} = \begin{bmatrix} \boldsymbol{I}_{2 \cdot M_{q:R}} \otimes \boldsymbol{t}_{1:q-1,1} \\ \vdots \\ \boldsymbol{I}_{2 \cdot M_{q:R}} \otimes \boldsymbol{t}_{1:q-1,M_{1:q-1}} \end{bmatrix}, \quad \bar{\boldsymbol{T}}_{q+1:R}^{(\mathrm{nc},r)} = \begin{bmatrix} \boldsymbol{I}_{M_q} \otimes \boldsymbol{t}_{q+1:R,1}^{(\mathrm{nc},r)} \\ \vdots \\ \boldsymbol{I}_{M_q} \otimes \boldsymbol{t}_{q+1:R,2 \cdot M_{q+1:R}}^{(\mathrm{nc},r)} \end{bmatrix}, \tag{6.72}$$

for $q = r$, we have

$$\begin{aligned} \boldsymbol{W}_r^{(\mathrm{nc},r)} = & \left(\boldsymbol{U}_{\mathrm{s}}^{(\mathrm{nc},r)^{\mathrm{T}}} \otimes \boldsymbol{I}_{2 \cdot M} \right) \cdot \left(\bar{\boldsymbol{T}}_{1:r-1}^{(\mathrm{nc},r)} \otimes \boldsymbol{I}_{2 \cdot M_{r:R}} \right) \cdot \left(\boldsymbol{I}_{2 \cdot M_r} \otimes \bar{\boldsymbol{T}}_{r+1:R}^{(\mathrm{nc},r)} \right) \\ & \cdot \left[\left(\boldsymbol{V}_r^{[\mathrm{s}](r)} \cdot \boldsymbol{\Sigma}_r^{[\mathrm{s}](r)^{-1}} \cdot \boldsymbol{U}_r^{[\mathrm{s}](r)^{\mathrm{H}}} \right)^{\mathrm{T}} \otimes \left(\boldsymbol{U}_r^{[\mathrm{n}](r)} \cdot \boldsymbol{U}_r^{[\mathrm{n}](r)^{\mathrm{H}}} \right) \right] \end{aligned} \tag{6.73}$$

with

$$\bar{\boldsymbol{T}}_{1:r-1}^{(\mathrm{nc},r)} = \begin{bmatrix} \boldsymbol{I}_{2 \cdot M_{r:R}} \otimes \boldsymbol{t}_{1:r-1,1} \\ \vdots \\ \boldsymbol{I}_{2 \cdot M_{r:R}} \otimes \boldsymbol{t}_{1:r-1,M_{1:r-1}} \end{bmatrix}, \quad \bar{\boldsymbol{T}}_{r+1:R}^{(\mathrm{nc},r)} = \begin{bmatrix} \boldsymbol{I}_{2 \cdot M_r} \otimes \boldsymbol{t}_{r+1:R,1} \\ \vdots \\ \boldsymbol{I}_{2 \cdot M_r} \otimes \boldsymbol{t}_{r+1:R,M_{r+1:R}} \end{bmatrix}, \tag{6.74}$$

and, finally, for $q > r$

$$\boldsymbol{W}_q^{(\mathrm{nc},r)} = \left(\boldsymbol{U}_{\mathrm{s}}^{(\mathrm{nc},r)^{\mathrm{T}}} \otimes \boldsymbol{I}_{2 \cdot M} \right) \cdot \left(\bar{\boldsymbol{T}}_{1:q-1}^{(\mathrm{nc},r)} \otimes \boldsymbol{I}_{M_{q:R}} \right) \cdot \left(\boldsymbol{I}_{M_q} \otimes \bar{\boldsymbol{T}}_{q+1:R}^{(\mathrm{nc},r)} \right) \tag{6.75}$$

$$\cdot \left[\left(\boldsymbol{V}_q^{[\mathrm{s}](r)} \cdot \boldsymbol{\Sigma}_q^{[\mathrm{s}](r)^{-1}} \cdot \boldsymbol{U}_q^{[\mathrm{s}](r)^{\mathrm{H}}} \right)^{\mathrm{T}} \otimes \left(\boldsymbol{U}_q^{[\mathrm{n}](r)} \cdot \boldsymbol{U}_q^{[\mathrm{n}](r)^{\mathrm{H}}} \right) \right] \tag{6.76}$$

with

$$\bar{\boldsymbol{T}}_{1:q-1}^{(\mathrm{nc},r)} = \begin{bmatrix} \boldsymbol{I}_{M_{q:R}} \otimes \boldsymbol{t}_{1:q-1,1}^{(\mathrm{nc},r)} \\ \vdots \\ \boldsymbol{I}_{M_{q:R}} \otimes \boldsymbol{t}_{1:q-1,2 \cdot M_{1:q-1}}^{(\mathrm{nc},r)} \end{bmatrix}, \quad \bar{\boldsymbol{T}}_{q+1:R}^{(\mathrm{nc},r)} = \begin{bmatrix} \boldsymbol{I}_{M_q} \otimes \boldsymbol{t}_{q+1:R,1} \\ \vdots \\ \boldsymbol{I}_{M_q} \otimes \boldsymbol{t}_{q+1:R,M_{q+1:R}} \end{bmatrix}. \tag{6.77}$$

The proof of Equation (6.68) is given in Appendix B.28. We would like to emphasize that the cases $q < r$, $q = r$, and $q > r$ need to be introduced as a result of the different dimensions of the matrices $\bar{\boldsymbol{T}}_{1:q-1}^{(\mathrm{nc},r)}$ and $\bar{\boldsymbol{T}}_{q+1:R}^{(\mathrm{nc},r)}$ due to the NC augmentation in the r-th mode. Note that the special case of 2-D NC Standard Tensor-ESPRIT and 2-D NC Unitary Tensor-ESPRIT was considered by us in [SRH16b].

The covariance matrix $\boldsymbol{R}_{\mathrm{nn}}^{(\mathrm{nc},r)} = \mathbb{E}\left\{ \boldsymbol{n}^{(\mathrm{nc},r)} \cdot \boldsymbol{n}^{(\mathrm{nc},r)^{\mathrm{H}}} \right\} \in \mathbb{C}^{2MN \times 2MN}$ and the pseudo covariance matrix $\boldsymbol{C}_{\mathrm{nn}}^{(\mathrm{nc},r)} = \mathbb{E}\left\{ \boldsymbol{n}^{(\mathrm{nc},r)} \cdot \boldsymbol{n}^{(\mathrm{nc},r)^{\mathrm{T}}} \right\} \in \mathbb{C}^{2MN \times 2MN}$ for $\boldsymbol{n}^{(\mathrm{nc},r)} = \mathrm{vec}\left\{ \boldsymbol{N}^{(\mathrm{nc},r)} \right\} \in \mathbb{C}^{2MN \times 1}$ with

$\boldsymbol{N}^{(\mathrm{nc},r)} = \left[\boldsymbol{\mathcal{N}}^{(\mathrm{nc},r)}\right]^{\mathrm{T}}_{(R+1)}$ are given by

$$\boldsymbol{R}_{\mathrm{nn}}^{(\mathrm{nc},r)} = \tilde{\boldsymbol{K}} \cdot \begin{bmatrix} \boldsymbol{R}_{\mathrm{nn}}^{(r)} & \boldsymbol{C}_{\mathrm{nn}}^{(r)} \\ \boldsymbol{C}_{\mathrm{nn}}^{(r)^*} & \boldsymbol{R}_{\mathrm{nn}}^{(r)^*} \end{bmatrix} \cdot \tilde{\boldsymbol{K}}^{\mathrm{T}}, \; \boldsymbol{C}_{\mathrm{nn}}^{(\mathrm{nc},r)} = \tilde{\boldsymbol{K}} \cdot \begin{bmatrix} \boldsymbol{C}_{\mathrm{nn}}^{(r)} & \boldsymbol{R}_{\mathrm{nn}}^{(r)} \\ \boldsymbol{R}_{\mathrm{nn}}^{(r)^*} & \boldsymbol{C}_{\mathrm{nn}}^{(r)^*} \end{bmatrix} \cdot \tilde{\boldsymbol{K}}^{\mathrm{T}}, \tag{6.78}$$

where $\tilde{\boldsymbol{K}} = \boldsymbol{K}_{2M,N}^{\mathrm{T}} \cdot \mathrm{blkdiag}\{\boldsymbol{K}_{M,N}, \boldsymbol{K}_{M,N}(\boldsymbol{I}_N \otimes \boldsymbol{\Pi}_M)\}$. Moreover, $\boldsymbol{R}_{\mathrm{nn}}^{(r)} = \mathbb{E}\{\boldsymbol{n}^{(r)}\boldsymbol{n}^{(r)^{\mathrm{H}}}\} \in \mathbb{C}^{MN \times MN}$ and $\boldsymbol{C}_{\mathrm{nn}}^{(r)} = \mathbb{E}\{\boldsymbol{n}^{(r)}\boldsymbol{n}^{(r)^{\mathrm{T}}}\} \in \mathbb{C}^{MN \times MN}$ are the covariance matrix and the pseudo-covariance matrix of the physical noise component $\boldsymbol{n}^{(r)} = \mathrm{vec}\{[\boldsymbol{\mathcal{N}}]_{(r)}\} \in \mathbb{C}^{MN \times 1}$ in the r-th mode.

In the special case of circularly symmetric white noise with $\boldsymbol{R}_{\mathrm{nn}}^{(r)} = \sigma_{\mathrm{n}}^2 \cdot \boldsymbol{I}_{MN}$ and $\boldsymbol{C}_{\mathrm{nn}}^{(r)} = \boldsymbol{0}_{MN \times MN}$, the covariance matrix $\boldsymbol{R}_{\mathrm{nn}}^{(\mathrm{nc},r)}$ and the pseudo covariance matrix $\boldsymbol{C}_{\mathrm{nn}}^{(\mathrm{nc},r)}$ in (6.78) simplify to $\boldsymbol{R}_{\mathrm{nn}}^{(\mathrm{nc},r)} = \sigma_{\mathrm{n}}^2 \cdot \boldsymbol{I}_{2MN}$ and $\boldsymbol{C}_{\mathrm{nn}}^{(\mathrm{nc},r)} = \sigma_{\mathrm{n}}^2 \cdot (\boldsymbol{I}_N \otimes \boldsymbol{\Pi}_{2M})$ in analogy to (6.55) in the matrix case. Therefore, the MSE of R-D NC Standard Tensor-ESPRIT in (6.66) for the i-th spatial frequency in the r-th mode reduces to

$$\mathbb{E}\left\{\left(\Delta\mu_i^{(r)}\right)^2\right\} \approx \frac{\sigma_{\mathrm{n}}^2}{2} \cdot \left(\left\|\boldsymbol{z}_i^{(\mathrm{nc},r)}\right\|_2^2 - \mathrm{Re}\left\{\boldsymbol{z}_i^{(\mathrm{nc},r)^{\mathrm{T}}} \cdot (\boldsymbol{I}_N \otimes \boldsymbol{\Pi}_{2M}) \cdot \boldsymbol{z}_i^{(\mathrm{nc},r)}\right\}\right), \tag{6.79}$$

where $\boldsymbol{z}_i^{(\mathrm{nc},r)} = \boldsymbol{W}_{\mathrm{ten}}^{(\mathrm{nc},r)^{\mathrm{T}}} \cdot \boldsymbol{r}_i^{(\mathrm{nc},r)}$ is given below (6.66).

6.5.2. Performance of R-D NC Unitary Tensor-ESPRIT

In Section 6.4.2, we have proven that in the matrix case, the R-D NC Standard ESPRIT and the R-D NC Unitary ESPRIT algorithm have the same asymptotic performance in the high effective SNR. In this section, we show that this result also carries over to the tensor case, i.e., R-D NC Standard Tensor-ESPRIT and R-D NC Unitary Tensor-ESPRIT perform asymptotically identical in the high effective SNR. As described in Section 6.3.3, the additional features of R-D NC Unitary Tensor-ESPRIT compared to R-D NC Standard Tensor-ESPRIT are the incorporation of the forward-backward averaging (FBA) preprocessing step in (6.38) and the transformation into the real-valued domain in (6.40) to reduce the computational complexity.

We have already shown in (6.39) that the augmented measurement tensor $\boldsymbol{\mathcal{X}}^{(\mathrm{nc},r)^{(\mathrm{fba})}}$ with FBA from (6.38) fulfills

$$\left[\boldsymbol{\mathcal{X}}^{(\mathrm{nc},r)^{(\mathrm{fba})}}\right]^{\mathrm{T}}_{(R+1)} \cdot \left(\left[\boldsymbol{\mathcal{X}}^{(\mathrm{nc},r)^{(\mathrm{fba})}}\right]^{\mathrm{T}}_{(R+1)}\right)^{\mathrm{H}} = 2 \cdot \left[\boldsymbol{\mathcal{X}}^{(\mathrm{nc},r)}\right]^{\mathrm{T}}_{(R+1)} \cdot \left(\left[\boldsymbol{\mathcal{X}}^{(\mathrm{nc},r)}\right]^{\mathrm{T}}_{(R+1)}\right)^{\mathrm{H}}. \tag{6.80}$$

Therefore, as in the matrix case, FBA has no effect on the augmented measurement tensor $\boldsymbol{\mathcal{X}}^{(\mathrm{nc},r)}$.

In order to account for the real-valued transformation, we can formulate the following theorem:

Theorem 6.5.3. *R-D NC Unitary Tensor-ESPRIT and R-D NC Standard Tensor-ESPRIT with FBA preprocessing perform asymptotically identical in the high effective SNR.*

The proof is shown in Appendix B.29.

Thus, in analogy to the matrix case, we can conclude that the performance of R-D NC Standard Tensor-ESPRIT and R-D NC Unitary Tensor-ESPRIT is asymptotically identical. However, the latter is preferable due to its lower computational complexity and better performance at low SNRs.

6.6. Special NC source cases

So far, we have derived analytical MSE expressions for the matrix-based and the tensor-based R-D NC ESPRIT-type algorithms for strictly non-circular sources. In our previous work [HR04, SRHD14], we have seen that via simulations, which are also shown in Section 6.7.1, that the gain from NC sources depends on the signal correlation, the source separation, and the phase reference of the array. In order to establish the exact dependence of the NC gain on these parameters, we adopt the derived analytical MSE expressions for the R-D NC ESPRIT-type algorithms. In this section, we simplify these expressions for the special cases of a single NC source in Section 6.6.1 and for two NC sources in Section 6.6.2.

6.6.1. Single NC source case

In this section, we consider the special case of a single NC source for the analytical MSE expressions of the matrix-based and tensor-based R-D NC ESPRIT-type algorithms using LS. Note that we have already shown in Section 6.4.2 that the asymptotic performance of R-D NC Standard ESPRIT and R-D NC Unitary ESPRIT is identical. In Section 6.5.2, we have found the same property for R-D NC Standard Tensor-ESPRIT and R-D NC Unitary Tensor-ESPRIT as well. Therefore, for the special case of a single NC source, it is sufficient to simplify the MSE expression in (6.52) for R-D NC Standard ESPRIT and the MSE expression for R-D NC Standard Tensor-ESPRIT in (6.66). We state the following result:

Theorem 6.6.1. *For the case of an M-element R-D uniform sampling grid with an M_r-element ULA in the r-th mode, a single strictly non-circular source ($d = 1$), and circularly symmetric white noise, the MSE of R-D NC Standard ESPRIT, R-D NC Unitary ESPRIT, R-D NC Standard Tensor-ESPRIT, and R-D NC Unitary Tensor-ESPRIT in the r-th mode is given by*

$$\mathbb{E}\left\{\left(\Delta\mu^{(r)}\right)^2\right\} \approx \frac{1}{\hat{\varrho}} \cdot \frac{M_r}{M(M_r-1)^2}, \tag{6.81}$$

where $\hat{\varrho}$ represents the effective SNR $\hat{\varrho} = N\hat{P}/\sigma_{\mathrm{n}}^2$ with $\hat{P}$ being the empirical source power given by $\hat{P} = \|\boldsymbol{s}\|_2^2/N$ and $\boldsymbol{s} \in \mathbb{C}^{N\times 1}$.

The proof is given in Appendix B.30. Note that the expression in (6.81) is equivalent to the single source result obtained in (4.42) for the conventional R-D ESPRIT-type algorithms for arbitrary

sources. Hence, we can conclude that no improvement in terms of the estimation accuracy can be achieved by applying the matrix-based or the tensor-based R-D NC ESPRIT-type algorithms if only a single NC source is present. In Section 9.4.1, we will confirm this property by deriving a single source expression of the deterministic R-D NC Cramér-Rao bound (CRB), which is derived in Chapter 9. Therein, we find that the single source expression of the NC CRB is identical to that of the conventional CRB. As a consequence of this result, the asymptotic efficiency, i.e., the ratio of the CRB and the MSE, of the matrix-based or the tensor-based R-D NC ESPRIT-type algorithms for a single NC source is also given by the expression in (4.45).

6.6.2. Two NC source case

In the previous section, we have seen that similarly to FBA and tensor processing (cf. Section 4.5.1), no gain can be achieved by applying the matrix-based or the tensor-based R-D NC ESPRIT-type algorithms for a single NC source. Thus, in this section, we simplify the analytical MSE expressions of the presented matrix-based and tensor-based R-D NC ESPRIT-type algorithms for two NC sources. Based on these expressions, we compute the NC gain for two sources and investigate its dependence on the physical parameters of interest. The derived MSE expressions will depend on the same parameters as in Section 4.5.1. The only differences for strictly non-circular sources are:

1. The empirical correlation $\hat{\rho}$ in (4.51) reduces to

$$\hat{\rho} = |\hat{\rho}| \cdot \mathrm{e}^{\mathrm{j}\Delta\varphi_{\mathrm{rot}}}, \tag{6.82}$$

   where $|\hat{\rho}| = \frac{1}{N} \cdot \frac{\boldsymbol{s}_{0_1}^{\mathrm{H}} \cdot \boldsymbol{s}_{0_2}}{\sqrt{\hat{P}_1 \cdot \hat{P}_2}} \in \mathbb{R}$, and $\Delta\varphi_{\mathrm{rot}} = |\varphi_2 - \varphi_1|$ is the rotation phase separation of the two sources, which is introduced in (6.1).

2. The dependence on the array phase reference parameter in (4.52) becomes

$$\Delta\varphi = \Delta\varphi_{\mathrm{ref}} + \Delta\varphi_{\mathrm{rot}}, \tag{6.83}$$

   where $\Delta\varphi_{\mathrm{ref}} = \sum_{r=1}^{R} \delta^{(r)} \cdot \Delta\mu^{(r)}$.

As in Section 4.5.1, we also provide approximations of the MSE expressions for two closely spaced sources, i.e., for $\Delta\mu^{(r)} \to 0$, by using a first-order Taylor approximation of the spatial correlation terms for this case.

We first consider the matrix-based versions in Section 6.6.2.1 before we proceed to the tensor-based R-D ESPRIT-type algorithms in Section 6.6.2.2. An analysis of the NC gain for two NC sources is provided in Section 6.6.2.3.

6.6.2.1. Two source case for matrix-based R-D NC ESPRIT-type algorithms

In this section, we simplify the analytical MSE expressions of the R-D NC ESPRIT-type algorithms for two sources. As we have shown that the performance of R-D NC Standard ESPRIT and R-D NC Unitary ESPRIT is asymptotically identical, we only need to simplify the MSE of R-D NC Standard ESPRIT for two sources. The result is provided in Theorem 6.6.2.

Theorem 6.6.2. *For the case of an M-element R-D uniform sampling grid with an M_r-element ULA in the r-th mode, circularly symmetric white noise, and two NC sources ($d = 2$), the MSE of R-D NC Standard ESPRIT and R-D NC Unitary ESPRIT is given by*

$$\mathrm{MSE}_{\mathrm{mat}}^{(\mathrm{nc})} = \frac{\sigma_{\mathrm{n}}^2}{2} \cdot \frac{\hat{P}_1 + \hat{P}_2}{N \cdot \hat{P}_1 \cdot \hat{P}_2} \cdot \sum_{r=1}^{R} a_{\mathrm{mat}}^{(\mathrm{nc})(r)} \tag{6.84}$$

where the scalar $a_{\mathrm{mat}}^{(\mathrm{nc})(r)}$ is derived as

$$a_{\mathrm{mat}}^{(\mathrm{nc})(r)} = \frac{M}{M_r} \cdot \Bigg[\left(\frac{M}{M_r}\right)^2 \cdot (M_r - 1)^2 \cdot 2 + \cos^2(\Delta\varphi) \cdot \Bigg\{ \left|\tilde{\alpha}^{(r)}\right|^2 \cdot \left(2 + (M_r - 2) \cdot \left|\mathrm{e}^{\mathrm{j}\Delta\mu^{(r)}} - 1\right|^2\right) - 4 \cdot (M_r - 1) \cdot \frac{|\alpha|^2}{\left|\alpha^{(r)}\right|^2} \cdot \mathrm{Re}\left\{\alpha_{\mathrm{sel},0}^{(r)}\right\} \Bigg\} \Bigg] \cdot \frac{1}{D_{\mathrm{sel}}^{(\mathrm{nc})(r)^2} \cdot \left(1 - |\hat{\rho}|^2\right)}$$

and the determinant $D_{\mathrm{sel}}^{(\mathrm{nc})(r)}$ is given by

$$D_{\mathrm{sel}}^{(\mathrm{nc})(r)} = \left(\frac{M}{M_r}\right)^2 \cdot (M_r - 1)^2 - \cos^2(\Delta\varphi) \cdot \frac{|\alpha|^2}{\left|\alpha^{(r)}\right|^2} \cdot \left|\alpha_{\mathrm{sel}}^{(r)}\right|^2. \tag{6.85}$$

The proof is given in Appendix B.31. Preliminary results for $R = 1$ are provided by us in [Gra15]. Note that similar to the MSE of R-D Unitary ESPRIT for two sources, the MSE in (6.84) depends on $\cos^2(\Delta\varphi)$, which contains the rotation phase and the phase reference.

An approximation for $a_{\mathrm{mat}}^{(\mathrm{nc})(r)}$ can be found by means of a first-order Taylor expansion as [Gra15]

$$\tilde{a}_{\mathrm{mat}}^{(\mathrm{nc})(r)} = \frac{1}{1 - |\hat{\rho}|^2} \cdot \frac{24 \cdot \left(12 + \cos^2(\Delta\varphi) \cdot \left[\left(S_r + 3 \cdot \Delta\mu^{(r)^2} \cdot M_r \cdot (M_r - 2)\right) - 12\right]\right)}{\frac{M}{M_r} \cdot (M_r - 1)^2 \cdot \left(12 + \cos^2(\Delta\varphi) \cdot [A_r - 12]\right)^2}. \tag{6.86}$$

where $A_r = \Delta\mu^{(r)^2} \cdot M_r \cdot (M_r - 2) + S_r$ and $S_r = \sum\limits_{\substack{q=1 \\ q \neq r}}^{R} \Delta\mu^{(q)^2} \cdot (M_q^2 - 1)$.

6.6.2.2. Two source case for tensor-based R-D NC ESPRIT-type algorithms

In this section, we simplify the analytical MSE expressions of the R-D NC Tensor-ESPRIT-type algorithms for two sources. As we have shown that the performance of R-D NC Standard Tensor-ESPRIT and R-D NC Unitary Tensor-ESPRIT is asymptotically identical, we only need to simplify the MSE of R-D NC Standard Tensor-ESPRIT for two sources. The result is provided in Theorem 6.6.3.

Theorem 6.6.3. *For the case of an M-element R-D uniform sampling grid with an M_r-element ULA in the r-th mode, circularly symmetric white noise, and two NC sources ($d = 2$), the MSE of R-D NC Standard Tensor-ESPRIT and R-D NC Unitary Tensor-ESPRIT is given by*

$$\mathrm{MSE}_{\mathrm{ten}}^{(\mathrm{nc})} = \frac{\sigma_{\mathrm{n}}^2}{2} \cdot \frac{\hat{P}_1 + \hat{P}_2}{N \cdot \hat{P}_1 \cdot \hat{P}_2} \cdot \sum_{r=1}^{R} a_{\mathrm{ten}}^{(\mathrm{nc},r)} \tag{6.87}$$

where $a_{\mathrm{ten}}^{(\mathrm{nc},r)}$ is given by

$$a_{\mathrm{ten}}^{(\mathrm{nc},r)} = \frac{1}{{D_{\mathrm{sel}}^{(\mathrm{nc},r)}}^2} \cdot \left(\frac{b^{(r)} \cdot \cos^2(\Delta\varphi)}{D_r^{(\mathrm{nc},r)}} \cdot \left[\frac{1}{1-|\hat{\rho}|^2} - \frac{\left(\frac{M}{M_r}\right)^2 - \frac{|\alpha|^2}{|\alpha^{(r)}|^2}}{\left(\frac{M}{M_r}\right)^2 - |\hat{\rho}|^2 \cdot \left|\frac{\alpha}{\alpha^{(r)}}\right|^2} \right] + c^{(\mathrm{nc},r)} \cdot \frac{\frac{M}{M_r}}{\left(\frac{M}{M_r}\right)^2 - |\hat{\rho}|^2 \cdot \frac{|\alpha|^2}{|\alpha^{(r)}|^2}} \right) \tag{6.88}$$

with

$$b^{(r)} = M \cdot (M_r - 1)^2 \cdot \left| \mathrm{e}^{\mathrm{j}\Delta\mu^{(r)}} - 1 \right|^2 \cdot \left|\alpha_{\mathrm{sel}}^{(r)}\right|^2 \cdot \left[\left(\frac{M}{M_r}\right)^2 - \frac{|\alpha|^2}{|\alpha^{(r)}|^2} \right] \tag{6.89}$$

$$\begin{aligned} c^{(\mathrm{nc},r)} = \left(\frac{M}{M_r}\right)^4 \cdot (M_r-1)^2 \cdot 2 + \cos^2(\Delta\varphi) \cdot \Bigg\{ & \left|\frac{\alpha}{\alpha^{(r)}}\right|^4 \cdot \left|\alpha_{\mathrm{sel}}^{(r)}\right|^2 \cdot \left(2 + (M_r-2) \cdot \left| \mathrm{e}^{\mathrm{j}\Delta\mu^{(r)}} - 1 \right|^2\right) \\ & - 4 \cdot \left(\frac{M}{M_r}\right)^2 \cdot (M_r - 1) \cdot \left|\frac{\alpha}{\alpha^{(r)}}\right|^2 \cdot \mathrm{Re}\left\{\alpha_{\mathrm{sel},0}^{(r)}\right\} \Bigg\}. \end{aligned} \tag{6.90}$$

Moreover, $D_{\mathrm{sel}}^{(\mathrm{nc},r)}$ is defined in (6.85) *and $D^{(\mathrm{nc},r)} = M_r^2 - \cos^2(\Delta\varphi)\,|\alpha^{(r)}|^2$.*

A sketch of the proof is given in Appendix B.32. The detailed proof is provided by us in [Gra16].

By applying a first-order Taylor approximation for small $\Delta\mu^{(r)}$, $r = 1, \ldots, R$, the MSE expression in (6.87) simplifies by replacing $a_{\mathrm{ten}}^{(\mathrm{nc},r)}$ with $\tilde{a}_{\mathrm{ten}}^{(\mathrm{nc},r)}$, which is given by [Gra16]

$$\tilde{a}_{\mathrm{ten}}^{(\mathrm{nc},r)} = \frac{144}{\left(12 + \cos^2(\Delta\varphi) \cdot (A_r - 12)\right)^2}$$

$$\cdot \left[\frac{\cos^2(\Delta\varphi) \cdot S_r \cdot \Delta\mu^{(r)^2}}{M \cdot \left(12 + \cos^2(\Delta\varphi) \cdot \left(\Delta\mu^{(r)^2} \cdot (M_r^2 - 1) - 12\right)\right) \cdot \left(1 - |\hat{\rho}|^2\right)} + \frac{6 \cdot \left(4 + \cos^2(\Delta\varphi) \cdot \left(\Delta\mu^{(r)^2} \cdot M_r \cdot (M_r - 2) - 4\right)\right)}{\frac{M}{M_r} \cdot (M_r - 1)^2 \cdot \left(12 + |\hat{\rho}|^2 \cdot (S_r - 12)\right)} \right].$$

6.6.2.3. Analysis of the NC gain

In this section, we analyze the NC gain of R-D NC ESPRIT-type algorithms in the matrix case in Section 6.6.2.4 and the NC gain of R-D NC Tensor-ESPRIT-type algorithms in the tensor case in Section 6.6.2.5 in terms of the rotation phase separation and the temporal correlation of the signals. In Section 9.4.2, we additionally investigate the NC gain based on the deterministic NC Cramér-Rao bound (CRB), which serves as a benchmark for the NC gain of R-D NC ESPRIT-type algorithms.

6.6.2.4. NC gain in the matrix case

In order to compute the NC gain, we require the MSE for R-D Standard ESPRIT in (4.53) and the MSE for R-D Standard ESPRIT in (6.84). Let us define the NC gain in the r-th mode as

$$\eta_{\text{mat}}^{(\text{nc})} = \frac{\text{MSE}_{\text{mat}}^{(r)}}{\text{MSE}_{\text{mat}}^{(\text{nc})(r)}} = \frac{a_{\text{mat}}^{(r)}}{a_{\text{mat}}^{(\text{nc})(r)}}, \tag{6.91}$$

where we restate $a_{\text{mat}}^{(r)}$ and $a_{\text{mat}}^{(\text{nc})(r)}$ again as

$$\begin{aligned} a_{\text{mat}}^{(r)} &= \frac{M}{M_r} \cdot \left[\left(\frac{M}{M_r}\right)^2 \cdot (M_r - 1)^2 \cdot 2 + d^{(\text{nc})(r)} \right] \cdot \frac{1}{D_{\text{sel}}^{(r)^2} \cdot \left(1 - |\hat{\rho}|^2\right)} \\ a_{\text{mat}}^{(\text{nc})(r)} &= \frac{M}{M_r} \cdot \left[\left(\frac{M}{M_r}\right)^2 \cdot (M_r - 1)^2 \cdot 2 + \cos^2(\Delta\varphi) \cdot d^{(\text{nc})(r)} \right] \cdot \frac{1}{D_{\text{sel}}^{(\text{nc})(r)^2} \cdot \left(1 - |\hat{\rho}|^2\right)}, \end{aligned} \tag{6.92}$$

where $D_{\text{sel}}^{(r)}$ and $D_{\text{sel}}^{(\text{nc})(r)}$ are given in (4.55) and in (6.85), and we have defined

$$d^{(\text{nc})(r)} = \left|\tilde{\alpha}^{(r)}\right|^2 \cdot \left(2 + (M_r - 2) \cdot \left|\mathrm{e}^{\mathrm{j}\Delta\mu^{(r)}} - 1\right|^2\right) - 4 \cdot (M_r - 1) \cdot \frac{|\alpha|^2}{\left|\alpha^{(r)}\right|^2} \cdot \text{Re}\left\{\alpha_{\text{sel},0}^{(r)}\right\}. \tag{6.93}$$

Then, we can express (6.91) as

$$\eta_{\text{mat}}^{(\text{nc})(r)} = \frac{\left(\left(\frac{M}{M_r}\right)^2 \cdot (M_r - 1)^2 - \cos^2(\Delta\varphi) \cdot \frac{|\alpha|^2}{\left|\alpha^{(r)}\right|^2} \cdot \left|\alpha_{\text{sel}}^{(r)}\right|^2\right)^2}{\left(\left(\frac{M}{M_r}\right)^2 \cdot (M_r - 1)^2 - \frac{|\alpha|^2}{\left|\alpha^{(r)}\right|^2} \cdot \left|\alpha_{\text{sel}}^{(r)}\right|^2\right)^2}$$

$$\cdot \frac{\left(\frac{M}{M_r}\right)^2 \cdot (M_r - 1)^2 \cdot 2 + d^{(\mathrm{nc})(r)}}{\left(\frac{M}{M_r}\right)^2 \cdot (M_r - 1)^2 \cdot 2 + \cos^2(\Delta\varphi) \cdot d^{(\mathrm{nc})(r)}}. \tag{6.94}$$

First, we note that $\eta_{\mathrm{mat}}^{(\mathrm{nc})(r)}$ is independent of $|\hat{\rho}|$. Moreover, we can easily identify the special cases

$$\eta_{\mathrm{mat}}^{(\mathrm{nc})(r)}\Big|_{\cos^2(\Delta\varphi)=1} = 1 \quad \text{and} \quad \eta_{\mathrm{mat}}^{(\mathrm{nc})(r)}\Big|_{\cos^2(\Delta\varphi)=0,\ \Delta\mu\to 0} = \infty. \tag{6.95}$$

Thus, theoretically, the NC gain can become inifitely large for $\Delta\mu \to 0$ and $\cos^2(\Delta\varphi) = 0$, which implies that

$$\Delta\varphi = \Delta\varphi_{\mathrm{ref}} + \Delta\varphi_{\mathrm{rot}} = \pm\frac{\pi}{2}. \tag{6.96}$$

This result can be interpreted as two orthogonal signals with a full spatial decorrelation. Note that we have a similar behavior for the FBA gain in Section 4.5.5.1, where a full temporal decorrelation can be achieved.

6.6.2.5. NC gain in the tensor case

For the computation of the NC gain in the tensor case, we use the MSEs of R-D Standard Tensor-ESPRIT in (4.60) and R-D NC Standard Tensor-ESPRIT in (6.87), respectively. Let us define the tensor-based NC gain in the r-th mode as

$$\eta_{\mathrm{ten}}^{(\mathrm{nc})(r)} = \frac{\mathrm{MSE}_{\mathrm{ten}}^{(r)}}{\mathrm{MSE}_{\mathrm{ten}}^{(\mathrm{nc})(r)}} = \frac{a_{\mathrm{ten}}^{(r)}}{a_{\mathrm{ten}}^{(\mathrm{nc})(r)}}. \tag{6.97}$$

Similarly to the NC gain in the matrix case, we can observe for the NC gain in the tensor case that

$$\eta_{\mathrm{ten}}^{(\mathrm{nc})(r)}\Big|_{\cos^2(\Delta\varphi)=1} = 1 \quad \text{and} \quad \eta_{\mathrm{ten}}^{(\mathrm{nc})(r)}\Big|_{\cos^2(\Delta\varphi)=0,\ \Delta\mu\to 0} = \infty. \tag{6.98}$$

However, the difference to the matrix case is that the NC gain in the tensor case is dependent on the temporal correlation $|\hat{\rho}|$ for $\cos^2(\Delta\varphi) < 1$.

6.7. Numerical results

In this section, we provide numerical simulations to illustrate the performance of the presented matrix-based and tensor-based R-D NC ESPRIT-type algorithms and assess their corresponding

analytical performance analysis expressions. We start with the matrix-based R-D NC ESPRIT-type algorithms in Section 6.7.1 before we consider their tensor-based versions in Section 6.7.2. Finally, we verify the analytical results for the special cases of a single NC source and two NC sources in Section 6.7.3.

6.7.1. Performance of R-D NC ESPRIT-type algorithms

In this set of simulation results, we evaluate the performance of the proposed R-D NC Standard ESPRIT and R-D NC Unitary ESPRIT algorithms along with the asymptotic behavior of the presented performance analysis. We compare the square root of the analytical MSE expression ("ana") in (6.52) to the root mean squared error (RMSE) of the empirical estimation error ("emp") of R-D NC Standard ESPRIT (NC SE) and R-D NC Unitary ESPRIT (NC UE) obtained by averaging over 5000 Monte Carlo trials. The RMSE is defined as

$$\text{RMSE} = \sqrt{\mathbb{E}\left\{\sum_{r=1}^{R}\sum_{i=1}^{d}\left(\mu_i^{(r)} - \hat{\mu_i}^{(r)}\right)^2\right\}}, \tag{6.99}$$

where $\hat{\mu_i}^{(r)}$ is the estimate of i-th spatial frequency in the r-th mode. Furthermore, we compare our results to R-D Standard ESPRIT (SE), R-D Unitary ESPRIT (UE) as well as the deterministic Cramér-Rao bounds for circular (Det CRB) and strictly SO non-circular sources (Det NC CRB) [RH07a]. In the simulations, we employ different array configurations consisting of isotropic sensor elements with interelement spacing $\delta = \lambda/2$ in all dimensions. The phase reference is chosen to be at the centroid of the array. It is assumed for all algorithms that a known number of signals with unit power and symbols $\boldsymbol{S}_0$ (cf. Equation (6.1)) drawn from a real-valued Gaussian distribution impinge on the array. Moreover, we assume zero-mean circularly symmetric white Gaussian sensor noise according to (6.55).

Figure 6.2 illustrates the RMSE versus the SNR, where we consider a $4\times4\times4$ uniform cubic array with $N = 5$ available observations of $d = 2$ sources with the spatial frequencies $\mu_1^{(1)} = 0$, $\mu_2^{(1)} = 0.1$, $\mu_1^{(2)} = 0$, $\mu_2^{(2)} = 0.1$, $\mu_1^{(3)} = 0$, and $\mu_2^{(3)} = 0.1$, and a real-valued pair-wise correlation of $\rho = 0.9$. The rotation phases contained in $\boldsymbol{\Psi}$ are given by $\varphi_1 = 0$ and $\varphi_2 = \pi/2$. In Figure 6.3, we depict the RMSE versus the number of snapshots N for the non-centro-symmetric 2-D array with $M = 20$ given in Figure 6.4, where we also provide the subarrays in both dimensions. The SNR is fixed at 10 dB and we have $d = 3$ uncorrelated sources with the spatial frequencies $\mu_1^{(1)} = 0.25$, $\mu_2^{(1)} = 0.5$, $\mu_3^{(1)} = 0.75$, $\mu_1^{(2)} = 0.25$, $\mu_2^{(2)} = 0.5$, and $\mu_3^{(2)} = 0.75$. The rotation phases are given by $\varphi_1 = 0$, $\varphi_2 = \pi/4$, and $\varphi_3 = \pi/2$. Note that 2-D Unitary ESPRIT cannot be applied as the array is not centro-symmetric. It is apparent from Figure 6.2 and Figure 6.3 that in general, the NC schemes perform better than their non-NC counterparts. Specifically, R-D NC Unitary ESPRIT provides a lower estimation error than R-D NC Standard ESPRIT for low SNRs and a low sample size.

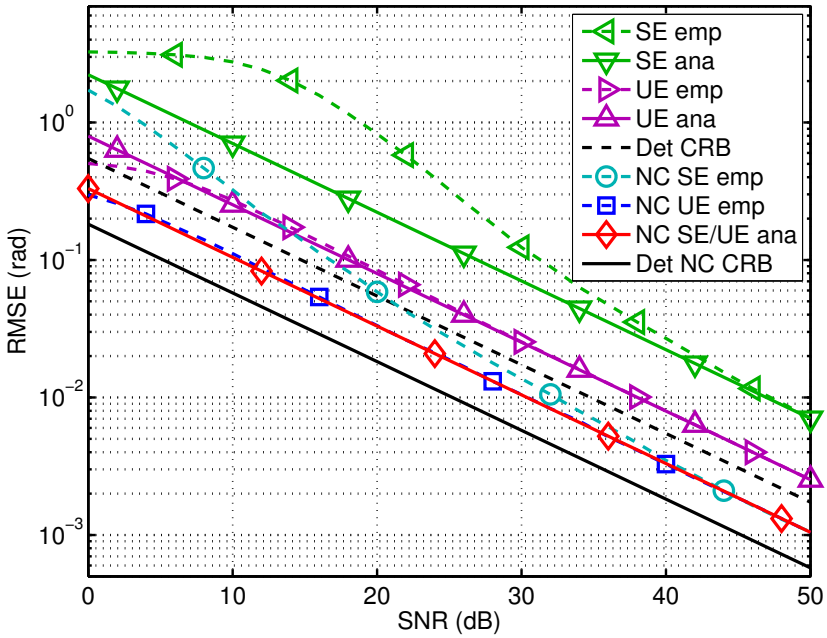


Figure 6.2.: Analytical and empirical RMSEs versus SNR for a $4\times4\times4$ cubic uniform array ($R = 3$), and $N = 5$, $d = 2$ correlated sources ($\rho = 0.9$) at $\mu_1^{(1)} = 0$, $\mu_2^{(1)} = 0.1$, $\mu_1^{(2)} = 0$, $\mu_2^{(2)} = 0.1$, $\mu_1^{(3)} = 0$, $\mu_2^{(3)} = 0.1$ with rotation phases $\varphi_1 = 0$, $\varphi_2 = \pi/2$.

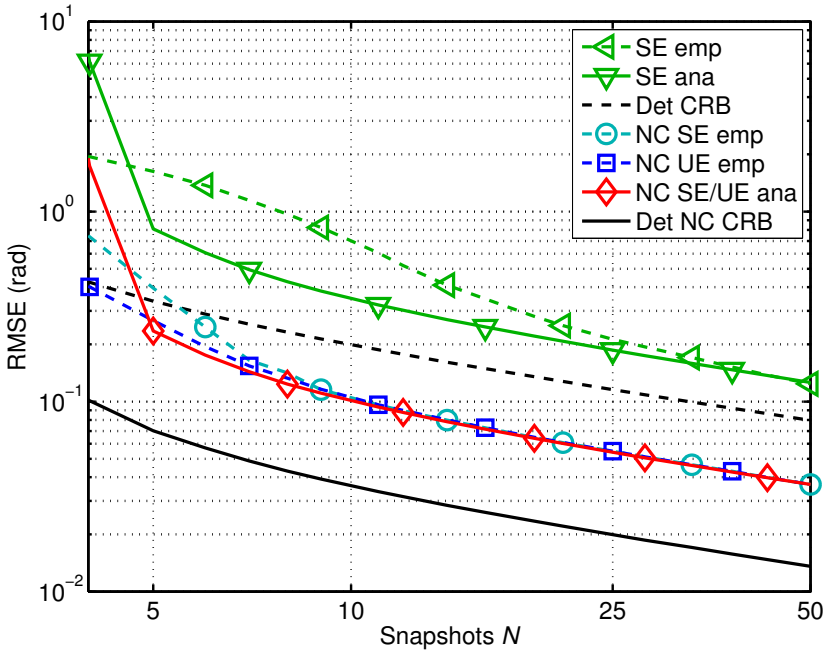


Figure 6.3.: Analytical and empirical RMSEs versus the snapshots N for the 20-element 2-D array ($R = 2$) from Figure 6.4 and SNR = 10 dB, $d = 3$ uncorrelated sources at $\mu_1^{(1)} = 0.25$, $\mu_2^{(1)} = 0.5$, $\mu_3^{(1)} = 0.75$, $\mu_1^{(2)} = 0.25$, $\mu_2^{(2)} = 0.5$, $\mu_3^{(2)} = 0.75$ with rotation phases $\varphi_1 = 0$, $\varphi_2 = \pi/4$, $\varphi_3 = \pi/2$.

Moreover, the analytical results agree well with the empirical estimation errors for high effective SNRs, i.e., when either the SNR or the number of samples becomes large. This also validates that the asymptotic performance of R-D NC Standard ESPRIT and R-D NC Unitary ESPRIT is identical as both coincide with the analytical curve. Note that the performance of the proposed algorithms can degrade if the signals' non-circularity is not perfectly strict.

In Figure 6.5, we show the RMSE as a function of the separation ("sep") between $d = 2$ uncorrelated sources located at $\mu_1^{(1)} = -\text{sep}/2$, $\mu_2^{(1)} = 0$, $\mu_1^{(2)} = \text{sep}/2$, $\mu_2^{(2)} = \text{sep}$ with the rotation phases $\varphi_1 = 0$, $\varphi_2 = \pi/2$. We employ a 5×6 uniform rectangular array (URA), $N = 5$ snapshots, and the SNR is fixed at 30 dB. Figure 6.6 demonstrates the RMSE as a function of the non-circularity phase separation $\Delta\varphi = |\varphi_2 - \varphi_1|$ of the $d = 2$ uncorrelated sources with the spatial frequencies $\mu_1^{(1)} = 1$, $\mu_2^{(1)} = 0.8$, $\mu_1^{(2)} = 1$, and $\mu_2^{(2)} = 0.8$. The remaining parameters are kept the same. Again, it can be seen from Figure 6.5 and Figure 6.6 that the analytical results match the empirical ones. But more importantly, the gain of the NC ESPRIT-type methods increases if the sources approach each other. Furthermore, as a substantial feature of strictly non-circular sources, it is observed that for two uncorrelated sources with a phase separation of $\Delta\varphi = \pi/2$, the sources entirely decouple as if each of them was present alone. In this case, the achievable gain from strictly non-circular sources is largest, which is verified by Figure 6.6. This decoupling effect was also shown analytically for the Det NC CRB in [RH07a] and recently for NC Standard ESPRIT in [SRH14a].

6.7.2. Performance of R-D NC Tensor-ESPRIT-type algorithms

This section provides simulation results for the proposed performance evaluation of R-D NC Standard Tensor-ESPRIT and R-D NC Unitary Tensor-ESPRIT for strictly non-circular sources using

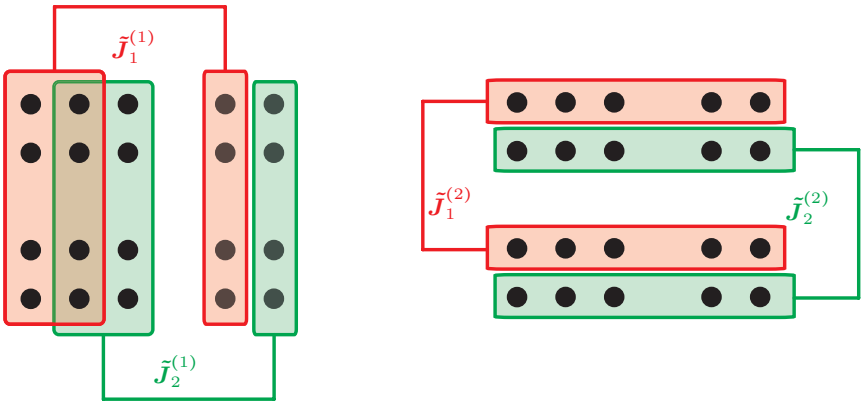


Figure 6.4.: 2-D shift invariance for the depicted non-centro-symmetric 5×4 sampling grid, left: subarrays for the first (horizontal) dimension, right: subarrays for the second (vertical) dimension.

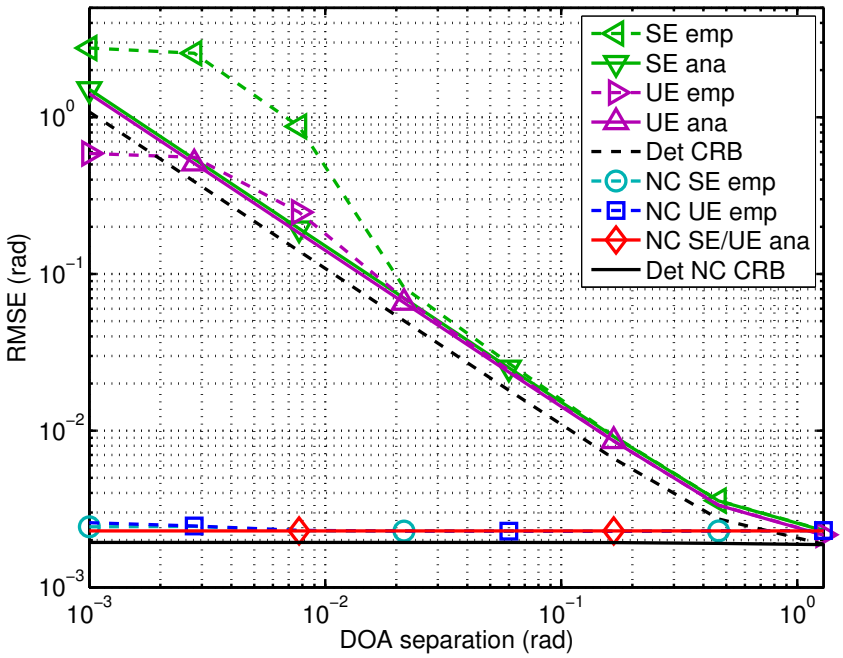


Figure 6.5.: Analytical and empirical RMSEs versus the separation ("sep") of $d = 2$ uncorrelated sources at $\mu_1^{(1)} = -\text{sep}/2$, $\mu_2^{(1)} = 0$, $\mu_1^{(2)} = \text{sep}/2$, $\mu_2^{(2)} = \text{sep}$ for a 5×6 URA ($R = 2$), $N = 5$, SNR = 30 dB, with rotation phases $\varphi_1 = 0$, $\varphi_2 = \pi/2$.

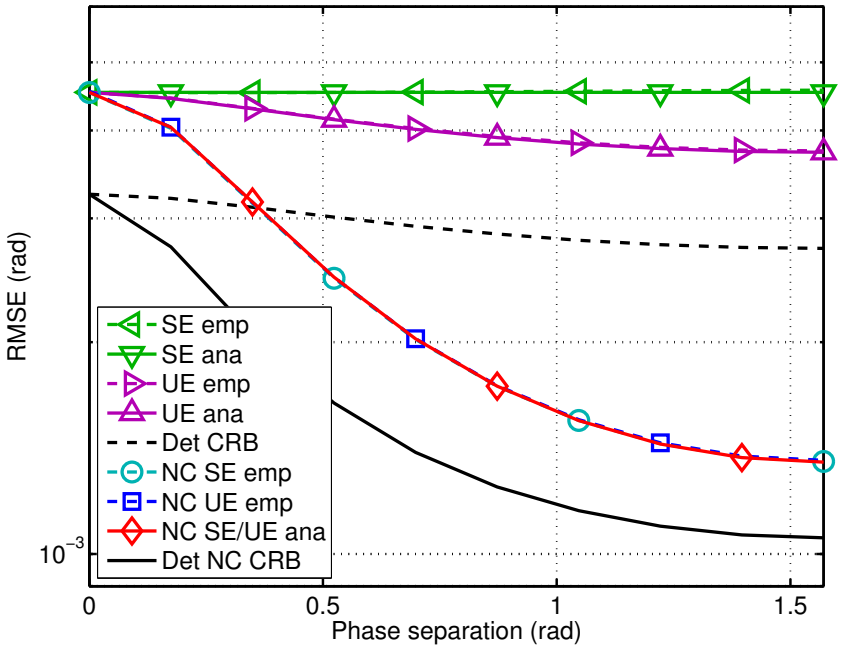


Figure 6.6.: Analytical and empirical RMSEs versus the phase separation for a 5×6 URA ($R = 2$), $N = 5$, SNR = 30 dB, $d = 2$ uncorrelated sources at $\mu_1^{(1)} = 1$, $\mu_2^{(1)} = 0.8$, $\mu_1^{(2)} = 1$, $\mu_2^{(2)} = 0.8$.

LS. Specifically, we compare the results found analytically (ana) to the empirical estimation errors (emp) obtained by averaging over Monte-Carlo trials. For the comparison, we consider 2-D versions of NC Standard Tensor-ESPRIT (NC STE) and NC Unitary Tensor-ESPRIT (NC UTE), their counterparts Standard Tensor-ESPRIT (STE) and Unitary Tensor-ESPRIT (UTE) [HRD08] from Chapter 4, and the corresponding deterministic Cramér-Rao bounds (Det CRB) [SN89] and its NC version (Det NC CRB) [SRHD16] presented in Chapter 9. We assume that the sources have unit power and that the transmit symbols are drawn from a real-valued Gaussian distribution. The noise is white Gaussian circularly symmetric. We have used 5000 Monte-Carlo trials to simulate the curves.

Figure 6.7 shows the total root mean square error (RMSE) versus the SNR for a 5×7 uniform rectangular array (URA) with $N = 10$ and $d = 3$ sources at $\mu_1^{(1)} = \mu_1^{(2)} = 1,\ \mu_2^{(1)} = \mu_2^{(2)} = 0.85,\ \mu_3^{(1)} = \mu_3^{(2)} = 1.15$. The sources have the pair-wise correlation $\rho = 0.99$ and the rotation phases $\varphi_1 = 0,\ \varphi_2 = \pi/2,\ \varphi_3 = \pi/4$. It is apparent from Figure 6.7 that the analytical curves match the empirical curves at high SNRs. Moreover, NC STE and NC UTE both outperform the non-NC algorithms and perform asymptotically identical.

In Figure 6.8, we illustrate the RMSE versus the number of snapshots N for a 6×6 URA with $d = 3$ correlated sources ($\rho = 0.9$) positioned at $\mu_1^{(1)} = \mu_1^{(2)} = 1,\ \mu_2^{(1)} = \mu_2^{(2)} = 0.9,\ \mu_3^{(1)} = \mu_3^{(2)} = 0.7$ and with the rotation phases $\varphi_1 = 0,\ \varphi_2 = \pi/6,\ \varphi_3 = \pi/3$. We set SNR = 40 dB. From Figure 6.8, we observe that the analytical results agree well with empirical results for a large sample size. Again, NC STE and NC UTE perform identically in this case.

In summary, the numerical results verify that the presented analytical performance evaluation is indeed asymptotic in the high effective SNR, i.e., the expressions become exact for either high SNRs or a large sample size. Moreover, as reasoned before, NC STE and NC UTE perform identically in the high effective SNR.

6.7.3. Analysis of special NC source cases

In the final set of simulations, we consider verify and study the analytical results for a single NC source and two NC sources.

Single NC source

In this simulation, we consider the single source case, which was used in Section 6.6.1 to express the analytical MSE equations of R-D NC Standard ESPRIT and R-D NC Unitary ESPRIT only in terms of the physical parameters, i.e., the array size M and the effective SNR. Figure 6.9 shows the asymptotic efficiency for the case $R = 1$ versus the number of sensors M of a ULA. The effective SNR is set to 46 dB, where $P_s = 0$ dB, $N = 4$, and $\sigma_n^2 = 10^{-4}$. This plot validates the fact that 1-D NC Standard ESPRIT and 1-D NC Unitary ESPRIT using LS become increasingly inefficient for

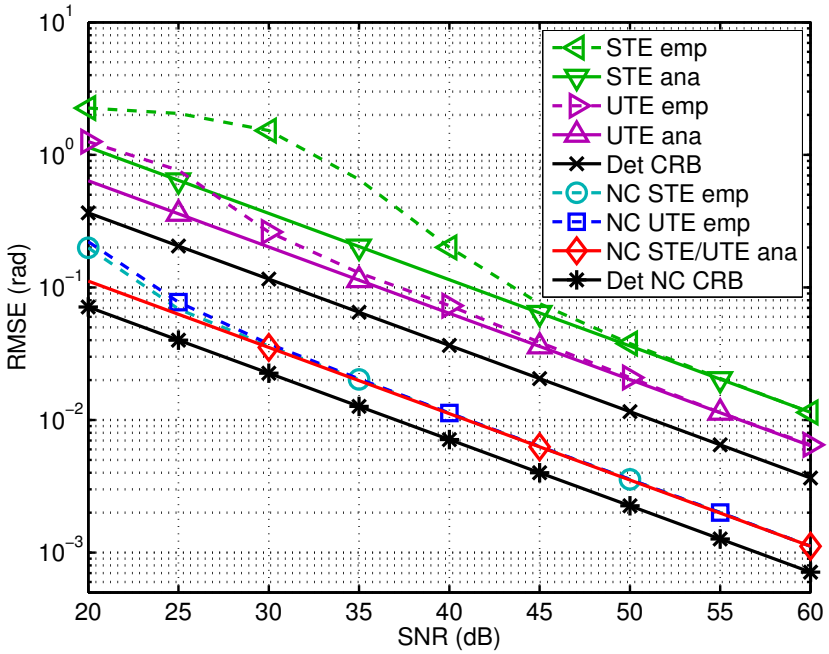


Figure 6.7.: RMSE versus SNR for a 5×7 URA with $N = 10$, $d = 3$ correlated ($\rho = 0.99$) sources at $\mu_1^{(1)} = \mu_1^{(2)} = 1$, $\mu_2^{(1)} = \mu_2^{(2)} = 0.85$, $\mu_3^{(1)} = \mu_3^{(2)} = 1.15$ and rotation phases $\varphi_1 = 0$, $\varphi_2 = \pi/2$, $\varphi_3 = \pi/4$.

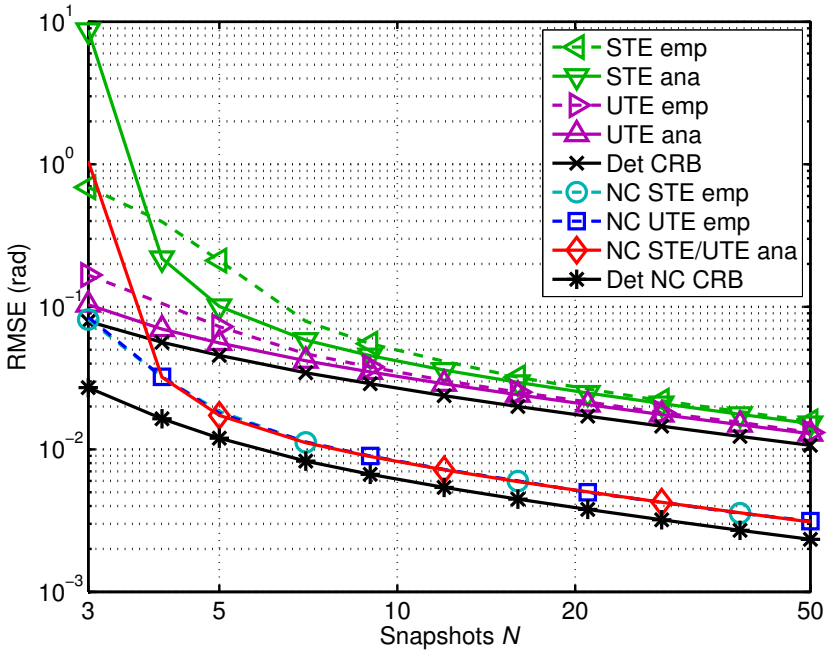


Figure 6.8.: RMSE versus N for a 6×6 URA with SNR $= 40$ dB, $d = 3$ correlated ($\rho = 0.9$) sources at $\mu_1^{(1)} = \mu_1^{(2)} = 1$, $\mu_2^{(1)} = \mu_2^{(2)} = 0.9$, $\mu_3^{(1)} = \mu_3^{(2)} = 0.7$ with rotation phases $\varphi_1 = 0$, $\varphi_2 = \pi/6$, $\varphi_3 = \pi/3$.

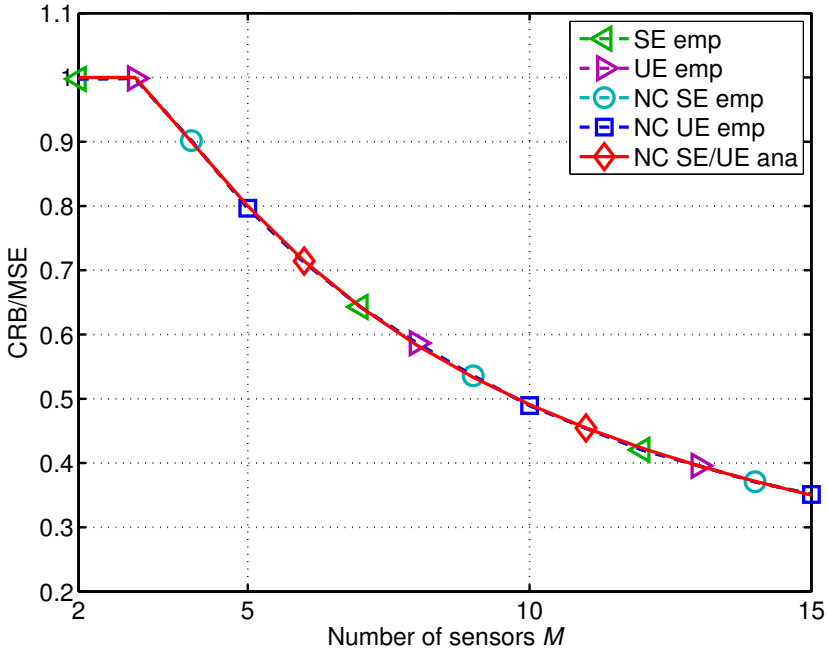


Figure 6.9.: Asymptotic efficiency versus M of a ULA ($R = 1$) for a single strictly non-circular source with an effective SNR of 46 dB ($P_s = 0$ dB, $N = 4$, $\sigma_n^2 = 10^{-4}$).

$M > 3$. It should be stressed that the same curves are obtained for 1-D Standard ESPRIT and 1-D Unitary ESPRIT from (4.45). Hence, no gain is achieved from a single strictly non-circular source.

Two NC sources

In this set of simulation results, we verify the analytical results for two NC sources derived in Section 6.6.2 and investigate the effect of the rotation phase separation and the temporal correlation of the two sources on the NC gain To this end, we compare the empirical estimation error (emp) to the analytical MSEs (ana exact) of R-D NC Standard ESPRIT (NC SE) in (6.84) and R-D NC Tensor-ESPRIT (NC STE) in (6.87). Moreover, we also consider the corresponding Taylor approximations (ana app). For comparison, we include the empirical and analytical MSEs of R-D Standard ESPRIT (SE) and R-D Standard Tensor-ESPRIT (STE) in (4.53) and (4.60), respectively. A benchmark for the algorithms is provided by the curves for the deterministic CRB (Det CRB) for arbitrary signals and the deterministic NC CRB (Det NC CRB) for strictly non-circular sources, which is derived in Chapter 9.

For the simulations, we consider a URA ($R = 2$) with isotropic elements and the positions of the $d = 2$ sources are fixed at $\mu_1^{(1)} = 1$, $\mu_1^{(2)} = 1$, $\mu_2^{(1)} = 1.1$, $\mu_2^{(2)} = 1.1$. The effective SNR is set to $\varrho = \frac{NP_1}{\sigma_n^2} = 100\,\text{dB}$ with ($P_1 = 40\,\text{dBm}, P_2 = 33.01\,\text{dBm}$, $N = 100$, $\sigma_n^2 = 10^{-9}$). Moreover, we assume that the array phase reference is located at the array centroid by choosing $\delta^{(r)} = 0, r = 1, 2$.

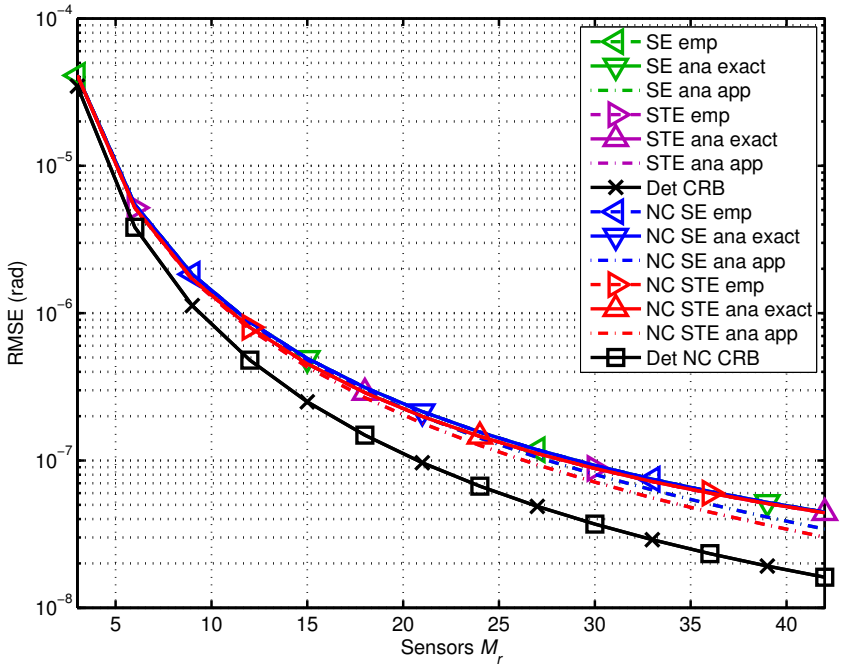


Figure 6.10.: RMSE versus the number of sensors M_r for $d = 2$ sources with $|\hat{\rho}| = 0$, $\varphi_{\text{rot}} = 0$.

Hence, the term $\cos(\Delta\varphi)$ that determines the NC gain as seen from (6.84) and (6.87) is entirely determined by the non-circularity phase $\Delta\varphi_{\text{rot}}$, i.e., $\Delta\varphi = \Delta\varphi_{\text{rot}}$.

Figure 6.10 to Figure 6.13 illustrate the MSE of the R-D NC ESPRIT-type algorithms as a function of M_r, $r = 1, 2$ in the r-th mode. We start our analysis by first considering uncorrelated sources, i.e., $|\hat{\rho}| = 0$. In Figure 6.10, we additionally set $\Delta\varphi_{\text{rot}} = 0$ and it is apparent that no NC gain can be achieved in this case. In Figure 6.11, we use $\Delta\varphi_{\text{rot}} = \pi/2$, where we observe that the NC gain is most pronounced for small M_r and decreases with increasing M_r. Then, we consider highly correlated sources with $|\hat{\rho}| = 0.99$. Again, we choose $\Delta\varphi_{\text{rot}} = 0$ and $\Delta\varphi_{\text{rot}} = \pi/2$ in Figure 6.12 and in Figure 6.13, respectively. From Figure 6.12, we only observe the tensor gain but no NC gain can be obtained due to $\Delta\varphi_{\text{rot}} = 0$. However, as expected in Figure 6.12, we obtain the combination of the tensor gain and the NC gain as $\Delta\varphi_{\text{rot}} = \pi/2$. Moreover, the analytical agree well with the empirical ones and the approximated MSEs are only valid for a small source separation.

6.8. Summary

In this chapter, we have presented the R-D NC ESPRIT-type algorithms and the R-D NC Tensor-ESPRIT-type algorithms that exploit the signal structure of strictly SO non-circular sources. We have shown that applying the NC preprocessing step always results in a centro-symmetric virtual array. Therefore, R-D NC Unitary ESPRIT and R-D NC Unitary Tensor-ESPRIT can even be applied if the original array is not centro-symmetric.

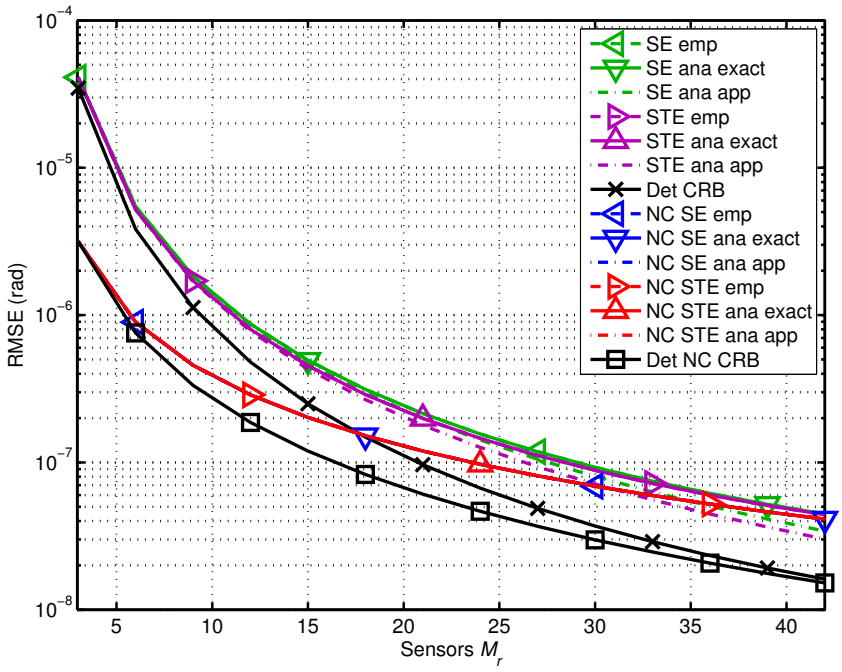


Figure 6.11.: RMSE versus the number of sensors M_r for $d = 2$ sources with $|\hat{\rho}| = 0$, $\varphi_{\text{rot}} = \pi/2$.

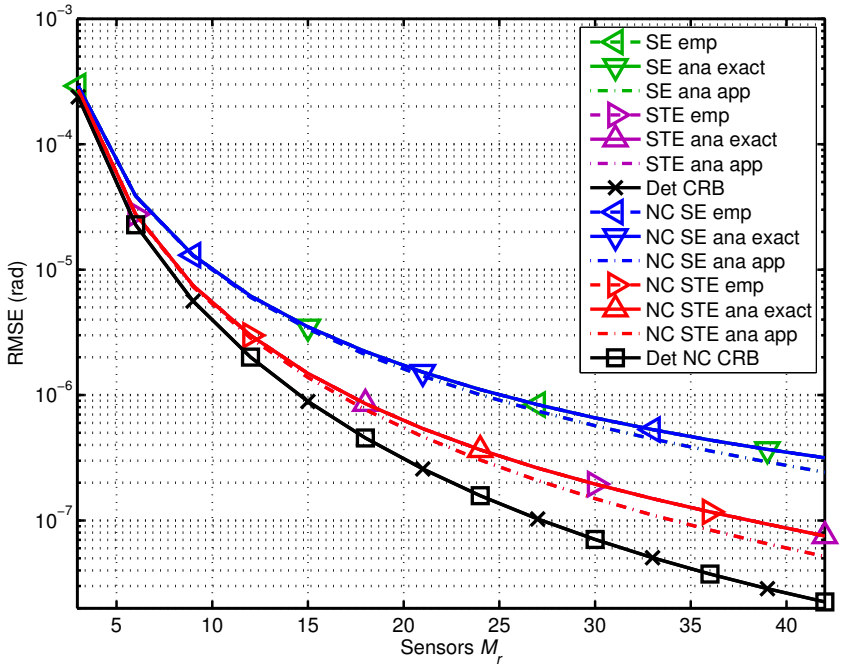


Figure 6.12.: RMSE versus the number of sensors M_r for $d = 2$ sources with $|\hat{\rho}| = 0.99$, $\varphi_{\text{rot}} = 0$.

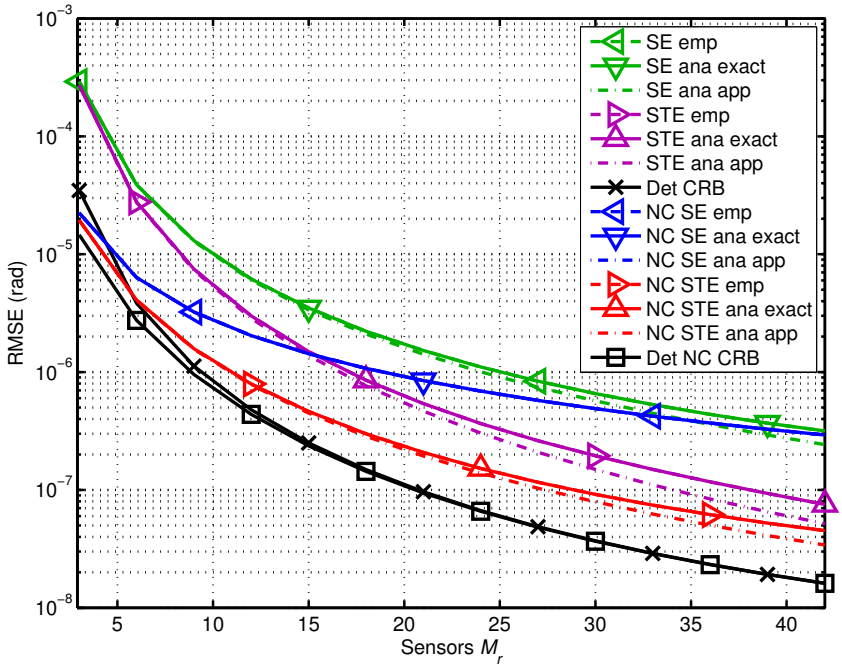


Figure 6.13.: RMSE versus the number of sensors M_r for $d = 2$ sources with $|\hat{\rho}| = 0.99$, $\varphi_{\text{rot}} = \pi/2$.

Moreover, we have derived a first-order analytical performance assessment for the matrix-based and the tensor-based R-D NC ESPRIT-type algorithms. The analytical MSE expressions only assumes the noise to be zero-mean with finite second-order moments such that no statistics of the noise are required. Moreover, all the results are asymptotic in the effective SNR, i.e., they become accurate for either high SNRs or a large sample size. Furthermore, we have analytically proven that the respective matrix-based R-D ESPRIT-type algorithms and the respective tensor-based R-D ESPRIT-type algorithms perform identical in the high effective SNR regime. However, R-D NC Unitary ESPRIT and R-D NC Unitary Tensor-ESPRIT should be preferred due to their real-valued operations and the better performance at low effective SNRs.

Finally, we have simplified the analytical MSE expressions for both matrix-based and tensor-based NC ESPRIT-type algorithms for the special case of a single NC source and two NC sources, assuming a uniform sampling grid and circularly symmetric white noise. The resulting expressions only depend on the physical parameters, e.g., the source correlation and the rotation phase of the signals. Based on these expressions, we have derived analytical expressions for the NC gain of the matrix-based and tensor-based NC ESPRIT-type algorithms for two sources. We have seen that the NC gain is largest for closely-spaced sources and a rotation phase separation of $\pi/2$.

7. NC ESPRIT-type parameter estimation algorithms with spatial smoothing

In Chapter 3 and in Chapter 6, we have presented matrix-based and tensor-based R-D ESPRIT-type algorithms as well as matrix-based and tensor-based R-D NC ESPRIT-type algorithms for strictly non-circular (NC) sources. However, these algorithms fail when more than two sources are coherent or only a single snapshot for more than two sources is present. In these cases, spatial smoothing can be applied as a preprocessing step to the aforementioned R-D ESPRIT-type and R-D NC ESPRIT-type algorithms to estimate the parameters of coherent signals. In this chapter, we present a first-order performance analysis for the spatially smoothed versions of the matrix-based R-D ESPRIT-type algorithms and R-D NC ESPRIT-type algorithms. In Section 7.1, we provide a literature overview and summarize the contributions. The R-D spatial smoothing preprocessing step is introduced in Section 7.2. In Section 7.3 and Section 7.4, we present the performance analysis results for the R-D ESPRIT-type algorithms and the R-D NC ESPRIT-type algorithms both with spatial smoothing, while the special case of a single source to compute the optimum number of subarrays for spatial smoothing is considered in Section 7.5. Finally, simulation results are provided in Section 7.6, and Section 7.7 summarizes the results.

7.1. Overview

In this section, we review the state of the art of the existing performance analysis expressions of parameter estimation algorithms in combination with spatial smoothing in Section 7.1.1 and summarize the contributions in Section 7.1.2.

7.1.1. State of the art

The problem of high-resolution parameter estimation from R-D signals with $R \geq 1$ has long been a fundamental research area in the field of array signal processing. As motivated in Chapter 3, R-D ESPRIT-type parameter estimation algorithms [HN98] are particularly appealing due to their fully algebraic estimates and their low complexity. As a result of their growing popularity, their analytical performance assessment has also been of great research interest. In Chapter 4, we have reviewed the R-D performance analysis framework in [RHD14], which is based on the concept in [LLV93]. The two fundamental performance analysis concepts for 1-D parameter estimation are derived in [Bri75, RH89a] and [LLV93]. While [RH89a] relies on the eigenvector distribution of the sample covariance matrix and is only asymptotic in the sample size N, the framework in [LLV93] provides an explicit first-order approximation of the parameter estimation error based on

the superposition of the signal component by a small noise perturbation. The latter is asymptotic in the effective signal-to-noise ratio (SNR), i.e., the results become accurate for either high SNRs or a large sample size. Therefore, [LLV93] is more general than [RH89a] as it is even valid for $N = 1$ if the SNR is sufficiently high. In [RHD14], this performance analysis framework has been extended to R-D parameter estimation, where no assumptions on the noise statistics apart from a zero mean and finite second-order (SO) moments are required for the analytical mean square error (MSE) expressions. A more detailed literature review on performance analysis concepts is given in Section 4.1.

In Chapter 6, we have shown how exploiting the signal structure of strictly non-circular (NC) signals can improve the performance of the conventional R-D ESPRIT-type parameter estimation algorithms as an example for subspace-based algorithms. A detailed introduction into the statistical properties of strictly non-circular signals is provided in Section 2.2. NC signals appear at the receiver when transmit signals using real-valued modulation schemes including BPSK, PAM, ASK, and Offset-QPSK and MSK after a de-rotation, undergo a phase shift due to the complex nature of the channel. Recently, a number of improved subspace-based parameter estimation schemes, e.g., NC MUSIC [AD06, FC10, FC14], NC Root-MUSIC [CWS01], NC Standard ESPRIT [ZCW03], NC Unitary ESPRIT [HR04] and the matrix-based and tensor-based R-D NC ESPRIT-type algorithms [SRHD14, RH09] presented in Chapter 6 have been developed. We have seen in Chapter 6 that exploiting the prior knowledge on the signals' strict non-circularity via the matrix-based and tensor-based R-D NC ESPRIT-type algorithms significantly improves the estimation accuracy and doubles the number of identifiable sources [SRHD14, RH09]. In Chapter 6, we have also extended the R-D performance analysis framework in [RHD14] to the NC case to obtain analytical MSE expressions for the presented matrix-based and tensor-based R-D NC ESPRIT-type algorithms for NC signals. For the special case of a single source, it is shown that neither forward-backward averaging (FBA) nor NC preprocessing in combination with ESPRIT-type algorithms improve the asymptotic MSE.

The aforementioned R-D ESPRIT-type algorithms and R-D NC ESPRIT-type algorithms are known to yield a high resolution even in the case of correlated sources. However, they fail when more than two signals[1] are coherent (fully correlated) or if $N = 1$ for more than two sources, as both render the signal covariance matrix rank-deficient. In practice, coherent signals often occur in a multipath environment [TV05] and the single snapshot case is often encountered in, e.g., MIMO channel sounding [LSJ05], co-array processing for nested and co-prime arrays [PV10, VP11, SRH17b], and tracking applications [CG90]. Assuming a uniform array geometry, preprocessing via spatial smoothing [EJS82, SWK85, PK89a] can be applied to estimate the parameters of multiple coherent signals. Spatial smoothing decorrelates multiple coherent signals by dividing the array into L identical displaced subarrays with $M_{\text{sub}} = M - L + 1$ sensor elements each and stacking

[1] Two coherent signals can be separated by FBA if the array phase reference is not located at the array centroid [Haa97b].

the corresponding measurement data into a column-wise augmented measurement matrix with an extended number of snapshots from N to NL but a reduced effective aperture from M to M_{sub}. As a result, L coherent signals are decorrelated. Alternatively, in terms of the spatial covariance matrix of the augmented measurement matrix, spatial smoothing can be understood as averaging the spatial covariance matrices corresponding to the L subarrays. If spatial smoothing is combined with forward-backward averaging (FBA) [PK89b] discussed in Section 3.2.3, the number of effective snapshots is extended to $2NL$, in which case $2L$ coherent signals are decorrelated with L subarrays.

As the resulting parameter estimation error after applying spatial smoothing depends on the number of subarrays L, it is a design parameter that can be optimized to achieve the best estimation accuracy. Several performance analyses of parameter estimation schemes using spatial smoothing based on the framework in [Bri75, RH89a], which is, however, only asymptotic in N, have been presented in [PK89a, RH90, RH93, HS90, HR99, LvdVD03, WF93]. While the authors of [PK89a, RH90, RH93] consider spatially smoothed MUSIC-type algorithms, the references [HS90, HR99, LvdVD03] study ESPRIT-type algorithms. In [WF93], a performance analysis for an interpolated spatial smoothing algorithm for non-uniform linear arrays was proposed. The special case of spatial smoothing for a single source was considered in [RH90, RH93], and in [HS90] for harmonic retrieval in time series analysis. It was observed that in this case a gain from spatial smoothing can be achieved. However, these existing performance analysis results only concern the 1-D parameter estimation. Analytical expressions for R-D parameter estimation algorithms such as R-D Standard ESPRIT and R-D Unitary ESPRIT with spatial smoothing as well as their recently proposed NC-versions R-D NC Standard ESPRIT and R-D NC Unitary ESPRIT with spatial smoothing have first been reported by us in [SRHD17].

7.1.2. Contributions

In this chapter, we present a first-order performance analysis for the spatially smoothed versions of the matrix-based R-D ESPRIT-type algorithms, i.e., R-D Standard ESPRIT and R-D Unitary ESPRIT, as well as the matrix-based R-D NC ESPRIT-type algorithms, i.e., R-D NC Standard ESPRIT and R-D NC Unitary ESPRIT, based on the R-D performance analysis framework in [RHD14], which is asymptotic in the high effective SNR. The results have been published in [SRH14b, SRHD17]. We assume a uniform R-D array geometry and use least squares (LS) to solve the shift invariance equation. However, as LS and total least squares (TLS) have been shown to perform asymptotically identical [RH89a], the results obtained for LS are also valid for TLS. The derived closed-form MSE expressions are explicit in the noise realizations such that apart from a zero mean and finite SO moments, no further assumptions on the noise statistics are required. We show that due to the NC preprocessing both R-D NC ESPRIT-type algorithms with spatial smoothing perform identical in the high effective SNR.

Further insights into the dependence of the MSE expressions on the physical parameters are

provided by the case study of a single source ($d = 1$). In particular, we first show that R-D spatial smoothing improves the estimation accuracy and that all the considered spatial smoothing based R-D ESPRIT-type algorithms and R-D NC ESPRIT-type algorithms provide the same MSE result for a single source, i.e., asymptotically, no additional gain is obtained from FBA and NC preprocessing. Based on these results, we analytically compute the optimal number of subarrays L that minimizes the MSE in each of the R dimensions, which extends the 1-D results in [PK89a, RH90, RH93, HS90, HR99, LvdVD03, WF93]. This enables us to compute the maximum asymptotic R-D spatial smoothing gain for a single source in closed-form. Additionally, we analytically compute the asymptotic efficiency, i.e., the ratio of the MSE and the corresponding Cramér-Rao bound expression, of the spatial smoothing based algorithms for $R = 1$.

7.2. R-D spatial smoothing preprocessing

In Section 7.2.1, we first restate the multi-dimensional matrix-based data model for arbitrary signals from (2.3) before reviewing the preprocessing step for the strictly non-circular data model from Section 6.2.1 in Section 7.2.2. Then, we apply R-D spatial smoothing to the data model for arbitrary signals in Section 7.2.3 and introduce R-D spatial smoothing for strictly non-circular signals in Section 7.2.4.

7.2.1. Matrix-based data model

Consider the R-D matrix-based data model in (2.3), where the measurement matrix $\boldsymbol{X}$ can be modeled as

$$\boldsymbol{X} = \boldsymbol{A} \cdot \boldsymbol{S} + \boldsymbol{N} \in \mathbb{C}^{M \times N}, \tag{7.1}$$

where $\boldsymbol{S} \in \mathbb{C}^{d \times N}$ represents the source symbol matrix, $\boldsymbol{N} \in \mathbb{C}^{M \times N}$ contains the noise samples, and $\boldsymbol{A} = [\boldsymbol{a}(\boldsymbol{\mu}_1), \ldots, \boldsymbol{a}(\boldsymbol{\mu}_d)] \in \mathbb{C}^{M \times d}$ is the array steering matrix. The latter consists of the array steering vectors $\boldsymbol{a}(\boldsymbol{\mu}_i)$ corresponding to the i-th spatial frequency, which are given by

$$\boldsymbol{a}(\boldsymbol{\mu}_i) = \boldsymbol{a}^{(1)}\left(\mu_i^{(1)}\right) \otimes \cdots \otimes \boldsymbol{a}^{(R)}\left(\mu_i^{(R)}\right) \in \mathbb{C}^{M \times 1}, \tag{7.2}$$

where $\boldsymbol{a}^{(r)}(\mu_i^{(r)}) \in \mathbb{C}^{M_r \times 1}$ is the array steering vector in the r-th mode. Alternatively, $\boldsymbol{A}$ can be expressed as

$$\boldsymbol{A} = \boldsymbol{A}^{(1)} \diamond \boldsymbol{A}^{(2)} \diamond \cdots \diamond \boldsymbol{A}^{(R)}, \tag{7.3}$$

where $\boldsymbol{A}^{(r)} = [\boldsymbol{a}^{(r)}(\mu_1^{(r)}), \ldots, \boldsymbol{a}^{(r)}(\mu_d^{(r)})] \in \mathbb{C}^{M_r \times d}$ represents the array steering matrix in the r-th mode. For an arbitrary phase reference along the r-th mode, $\boldsymbol{A}^{(r)}$ can be decomposed according

to (2.24) as $\boldsymbol{A}^{(r)} = \boldsymbol{A}_{\mathrm{c}}^{(r)} \cdot \boldsymbol{\Delta}^{(r)}$, where $\boldsymbol{A}_{\mathrm{c}}^{(r)} = [\boldsymbol{a}_{\mathrm{c}}^{(r)}(\mu_1^{(r)}), \cdots, \boldsymbol{a}_{\mathrm{c}}^{(r)}(\mu_d^{(r)})] \in \mathbb{C}^{M_r \times d}$ satisfies $\boldsymbol{A}_{\mathrm{c}}^{(r)} = \boldsymbol{\Pi}_{M_r} \cdot \boldsymbol{A}_{\mathrm{c}}^{(r)^*}$ and contains the steering vectors $\boldsymbol{a}_{\mathrm{c}}^{(r)}(\mu_i^{(r)})$, $i = 1, \ldots, d$, whose phase reference is located at the centroid of the r-th mode, i.e.,

$$\boldsymbol{a}_{\mathrm{c}}^{(r)}\left(\mu_i^{(r)}\right) = \left[\mathrm{e}^{-\mathrm{j}\frac{(M_r-1)}{2}\mu_i^{(r)}} \quad \ldots \quad \mathrm{e}^{\mathrm{j}\frac{(M_r-1)}{2}\mu_i^{(r)}}\right]^{\mathrm{T}} \in \mathbb{C}^{M_r \times 1}. \tag{7.4}$$

If the phase reference of $\boldsymbol{A}^{(r)}$ is not at the centroid of the r-th mode, the diagonal matrix $\boldsymbol{\Delta}^{(r)} = \operatorname{diag}\left\{\mathrm{e}^{\mathrm{j}\delta^{(r)}\mu_i^{(r)}}\right\}_{i=1}^{d}$ defines the shifts of the phase reference $\delta^{(r)} \in [\frac{-(M_r-1)}{2}, \frac{(M_r-1)}{2}]$ for each $\mu_i^{(r)}$. An example for this case is shown in (2.26). If the phase reference is at the array centroid of the r-th mode, we have $\delta^{(r)} = 0$, $\boldsymbol{\Delta}^{(r)} = \boldsymbol{I}_d$, and consequently $\boldsymbol{A}^{(r)} = \boldsymbol{A}_{\mathrm{c}}^{(r)}$. Thus, we can rewrite $\boldsymbol{A}$ in (7.3) according to (2.25) as $\boldsymbol{A} = \boldsymbol{A}_{\mathrm{c}} \cdot \boldsymbol{\Delta}$, where $\boldsymbol{A}_{\mathrm{c}} = \boldsymbol{A}_{\mathrm{c}}^{(1)} \diamond \boldsymbol{A}_{\mathrm{c}}^{(2)} \diamond \cdots \diamond \boldsymbol{A}_{\mathrm{c}}^{(R)} \in \mathbb{C}^{M \times d}$ and $\boldsymbol{\Delta} = \boldsymbol{\Delta}^{(1)} \cdot \boldsymbol{\Delta}^{(2)} \cdot \ldots \cdot \boldsymbol{\Delta}^{(R)} = \operatorname{diag}\left\{\mathrm{e}^{\mathrm{j}\delta_i}\right\}_{i=1}^{d} \in \mathbb{C}^{d \times d}$ with $\delta_i = \sum_{r=1}^{R} \delta^{(r)}\mu_i^{(r)}$. Again, if $\delta^{(r)} = 0 \ \forall\, r$, we have $\boldsymbol{A} = \boldsymbol{A}_{\mathrm{c}}$. Using these relations, we obtain the model

$$\boldsymbol{X} = \boldsymbol{A}_{\mathrm{c}} \cdot \boldsymbol{\Delta} \cdot \boldsymbol{S} + \boldsymbol{N} = \boldsymbol{A}_{\mathrm{c}} \cdot \bar{\boldsymbol{S}} + \boldsymbol{N} \ \in \mathbb{C}^{M \times N}, \tag{7.5}$$

where we have defined[2] $\bar{\boldsymbol{S}} = \boldsymbol{\Delta} \cdot \boldsymbol{S}$.

Due to the assumption that the R-D sampling grid is uniform, the array steering matrix $\boldsymbol{A}$ satisfies the shift invariance equations given by

$$\tilde{\boldsymbol{J}}_1^{(r)} \cdot \boldsymbol{A}_{\mathrm{c}} \cdot \boldsymbol{\Phi}^{(r)} = \tilde{\boldsymbol{J}}_2^{(r)} \cdot \boldsymbol{A}_{\mathrm{c}}, \quad r = 1, \ldots, R, \tag{7.6}$$

where $\tilde{\boldsymbol{J}}_1^{(r)}$ and $\tilde{\boldsymbol{J}}_2^{(r)} \in \mathbb{R}^{\frac{M}{M_r}(M_r-1) \times M}$ are the effective R-D selection matrices, which select $M_r - 1$ elements (maximum overlap) for the first and the second subarray in the r-th mode, respectively. They are compactly defined as $\tilde{\boldsymbol{J}}_k^{(r)} = \boldsymbol{I}_{\prod_{l=1}^{r-1} M_l} \otimes \boldsymbol{J}_k^{(r)} \otimes \boldsymbol{I}_{\prod_{l=r+1}^{R} M_l}$ for $k = 1, 2$, where $\boldsymbol{J}_k^{(r)} \in \mathbb{R}^{(M_r-1) \times M_r}$ are the r-mode selection matrices for the first and second subarray [HN98]. The diagonal matrix $\boldsymbol{\Phi}^{(r)} = \operatorname{diag}\left\{\mathrm{e}^{\mathrm{j}\mu_i^{(r)}}\right\}_{i=1}^{d} \in \mathbb{C}^{d \times d}$ contains the spatial frequencies in the r-th mode to be estimated.

From (7.6), it is evident that up to

$$d \leq \min\left\{\min_r \left((M_r - 1)M/M_r\right), N\right\} \tag{7.7}$$

incoherent sources can be resolved.

7.2.2. R-D Preprocessing for strictly non-circular signals

The signal model of strictly second-order non-circular signals is described in Section 2.2. In a communication system, the case of strictly non-circular signals presumes that the sources transmit

[2]The definition of $\bar{\boldsymbol{S}}$ facilitates the simplification of the analytical MSE expression for a single source in Section 7.5.

signals with real-valued modulation schemes including BPSK, ASK, and Offset-QPSK (after a derotation). Due to the distinct transmission delays of the signals received from different sources, the symbol amplitudes at the receiver lie on lines with different phase rotations in the complex plane. Therefore, the symbol matrix $\boldsymbol{S}$ in (7.1) can be decomposed as shown in (2.36) as

$$\boldsymbol{S} = \boldsymbol{\Psi} \cdot \boldsymbol{S}_0, \tag{7.8}$$

where $\boldsymbol{S}_0 \in \mathbb{R}^{d \times N}$ is a real-valued symbol matrix and $\boldsymbol{\Psi} = \operatorname{diag}\left\{\mathrm{e}^{\mathrm{j}\varphi_i}\right\}_{i=1}^{d}$ contains the stationary complex phase shifts on its diagonal that are usually different for each source. Then, $\bar{\boldsymbol{S}}$ in (7.5) is given by

$$\bar{\boldsymbol{S}} = \boldsymbol{\Delta} \cdot \boldsymbol{\Psi} \cdot \boldsymbol{S}_0 = \boldsymbol{\Xi} \cdot \boldsymbol{S}_0, \tag{7.9}$$

where we have defined $\boldsymbol{\Xi} = \boldsymbol{\Delta} \cdot \boldsymbol{\Psi} = \operatorname{diag}\left\{\mathrm{e}^{\mathrm{j}(\varphi_i+\delta_i)}\right\}_{i=1}^{d}$.

In order to take advantage of the strict non-circularity of the signals, we apply the preprocessing scheme from (6.4) to the data model in (7.1) and define the augmented measurement matrix $\boldsymbol{X}^{(\mathrm{nc})} \in \mathbb{C}^{2M \times N}$ as

$$\begin{aligned}\boldsymbol{X}^{(\mathrm{nc})} &= \begin{bmatrix} \boldsymbol{X} \\ \boldsymbol{\Pi}_M \cdot \boldsymbol{X}^* \end{bmatrix} = \begin{bmatrix} \boldsymbol{A}_{\mathrm{c}} \\ \boldsymbol{A}_{\mathrm{c}} \cdot \boldsymbol{\Xi}^* \cdot \boldsymbol{\Xi}^* \end{bmatrix} \cdot \bar{\boldsymbol{S}} + \begin{bmatrix} \boldsymbol{N} \\ \boldsymbol{\Pi}_M \cdot \boldsymbol{N}^* \end{bmatrix} \\ &= \boldsymbol{A}_{\mathrm{c}}^{(\mathrm{nc})} \cdot \bar{\boldsymbol{S}} + \boldsymbol{N}^{(\mathrm{nc})}, \end{aligned} \tag{7.10}$$

where we have used the property $\boldsymbol{\Pi}_M \cdot \boldsymbol{A}_{\mathrm{c}}^* = \boldsymbol{A}\mathrm{c}$. Moreover, $\boldsymbol{A}_{\mathrm{c}}^{(\mathrm{nc})} \in \mathbb{C}^{2M \times d}$ and $\boldsymbol{N}^{(\mathrm{nc})} \in \mathbb{C}^{2M \times N}$ are the augmented array steering matrix and the augmented noise matrix, respectively.

It was shown in Theorem 6.2.1 that if the array steering matrix $\boldsymbol{A}_{\mathrm{c}}$ is shift-invariant (7.6), then $\boldsymbol{A}_{\mathrm{c}}^{(\mathrm{nc})}$ is also shift-invariant and satisfies

$$\tilde{\boldsymbol{J}}_1^{(\mathrm{nc})(r)} \cdot \boldsymbol{A}_{\mathrm{c}}^{(\mathrm{nc})} \cdot \boldsymbol{\Phi}^{(r)} = \tilde{\boldsymbol{J}}_2^{(\mathrm{nc})(r)} \cdot \boldsymbol{A}_{\mathrm{c}}^{(\mathrm{nc})}, \quad r = 1, \ldots, R, \tag{7.11}$$

where $\tilde{\boldsymbol{J}}_k^{(\mathrm{nc})(r)} = \boldsymbol{I}_{\prod_{l=1}^{r-1} M_l} \otimes \boldsymbol{J}_k^{(\mathrm{nc})(r)} \otimes \boldsymbol{I}_{\prod_{l=r+1}^{R} M_l}$ and $\boldsymbol{J}_k^{(\mathrm{nc})(r)} = \boldsymbol{I}_2 \otimes \boldsymbol{J}_k^{(r)}$, $k = 1, 2$. Note that the extended dimensions of $\boldsymbol{A}_{\mathrm{c}}^{(\mathrm{nc})}$ can be interpreted as a virtual doubling of the number of sensors, which leads to a lower estimation error and doubles the number of resolvable sources as shown in Chapter 6.

Thus, after applying the NC preprocessing in (7.10), we can estimate up to

$$d \leq \min\left\{ \min_r \left(2 \cdot (M_r - 1) M / M_r\right), N \right\} \tag{7.12}$$

incoherent sources as compared to (7.7).

7.2.3. R-D Spatial smoothing for arbitrary signals

In the case of coherent signals (fully correlated), or for a single snapshot $N = 1$, the symbol matrix $\bar{\boldsymbol{S}}$ becomes row rank deficient, i.e., $\operatorname{rank}\{\bar{\boldsymbol{S}}\} < d$. If only two signals are coherent, forward-backward averaging (FBA) [PK89a] can separate these signals if the corresponding diagonal elements of $\boldsymbol{\Delta}$ are distinct [Haa97a], i.e., the phase reference is not at the array centroid. For more than two coherent signals, however, the conventional subspace-based parameter estimators fail to estimate the directions of the coherent signals. In case of a uniform array geometry, spatial smoothing preprocessing can be applied to restore the full row rank d of $\bar{\boldsymbol{S}}$ albeit reducing the effective array aperture.

In order to perform R-D spatial smoothing, we apply 1-D spatial smoothing to each of the R dimensions independently [HN98]. To this end, the M_r uniform sampling grid points in the r-th dimension are divided into L_r maximally overlapping subarrays, each containing $M_{\mathrm{sub}_r} = M_r - L_r + 1$ elements. The corresponding $M_{\mathrm{sub}_r} \times M_r$ selection matrix for the ℓ_r-th subarray, $1 \leq \ell_r \leq L_r$ for $1 \leq r \leq R$, is defined as

$$\boldsymbol{J}_{\ell_r}^{(M_r)} = \left[\boldsymbol{0}_{M_{\mathrm{sub}_r} \times (\ell_r - 1)} \quad \boldsymbol{I}_{M_{\mathrm{sub}_r}} \quad \boldsymbol{0}_{M_{\mathrm{sub}_r} \times (L_r - \ell_r)}\right]. \tag{7.13}$$

Next, we define the $L = \prod_{r=1}^{R} L_r$ multi-dimensional selection matrices

$$\boldsymbol{J}_{\underline{\boldsymbol{\ell}}} = \boldsymbol{J}_{\ell_1,\ldots,\ell_{R-1},\ell_R} = \boldsymbol{J}_{\ell_1}^{(M_1)} \otimes \cdots \otimes \boldsymbol{J}_{\ell_{R-1}}^{(M_{R-1})} \otimes \boldsymbol{J}_{\ell_R}^{(M_R)} \in \mathbb{R}^{M_{\mathrm{sub}} \times M} \tag{7.14}$$

for $1 \leq \ell_r \leq L_r$ with $M_{\mathrm{sub}} = \prod_{r=1}^{R} M_{\mathrm{sub}_r}$. Then, the spatially smoothed data matrix $\boldsymbol{X}_{\mathrm{SS}} \in \mathbb{C}^{M_{\mathrm{sub}} \times NL}$, which is subsequently processed instead of $\boldsymbol{X}$, is given by

$$\begin{aligned}
\boldsymbol{X}_{\mathrm{SS}} = &\Big[\boldsymbol{J}_{1,\cdots,1,1} \cdot \boldsymbol{X} \quad \boldsymbol{J}_{1,\cdots,1,2} \cdot \boldsymbol{X} \quad \cdots \quad \boldsymbol{J}_{1,\cdots,1,L_R} \cdot \boldsymbol{X} \\
&\quad \boldsymbol{J}_{1,\cdots,2,1} \cdot \boldsymbol{X} \quad \boldsymbol{J}_{1,\cdots,2,2} \cdot \boldsymbol{X} \quad \cdots \quad \boldsymbol{J}_{L_1,\cdots,L_{R-1},L_R} \cdot \boldsymbol{X}\Big] \\
= &\Big[\boldsymbol{J}_{1,\cdots,1,1} \cdot \boldsymbol{A}_{\mathrm{c}} \cdot \bar{\boldsymbol{S}} \quad \boldsymbol{J}_{1,\cdots,1,2} \cdot \boldsymbol{A}_{\mathrm{c}} \cdot \bar{\boldsymbol{S}} \quad \cdots \quad \boldsymbol{J}_{1,\cdots,1,L_R} \cdot \boldsymbol{A}_{\mathrm{c}} \cdot \bar{\boldsymbol{S}} \\
&\quad \boldsymbol{J}_{1,\cdots,2,1} \cdot \boldsymbol{A}_{\mathrm{c}} \cdot \bar{\boldsymbol{S}} \quad \boldsymbol{J}_{1,\cdots,2,2} \cdot \boldsymbol{A}_{\mathrm{c}} \cdot \bar{\boldsymbol{S}} \quad \cdots \quad \boldsymbol{J}_{L_1,\cdots,L_{R-1},L_R} \cdot \boldsymbol{A}_{\mathrm{c}} \cdot \bar{\boldsymbol{S}}\Big] \\
&+ \Big[\boldsymbol{J}_{1,\cdots,1,1} \cdot \boldsymbol{N} \quad \boldsymbol{J}_{1,\cdots,1,2} \cdot \boldsymbol{N} \quad \cdots \quad \boldsymbol{J}_{1,\cdots,1,L_R} \cdot \boldsymbol{N} \\
&\quad \boldsymbol{J}_{1,\cdots,2,1} \cdot \boldsymbol{N} \quad \boldsymbol{J}_{1,\cdots,2,2} \cdot \boldsymbol{N} \quad \cdots \quad \boldsymbol{J}_{L_1,\cdots,L_{R-1},L_R} \cdot \boldsymbol{N}\Big].
\end{aligned} \tag{7.15}$$

Note that by using (7.3) and (7.14), the array steering matrix of the ℓ-th subarray in all R modes can be expressed as

$$\begin{aligned}
\boldsymbol{J}_{\ell_1,\ldots,\ell_{R-1},\ell_R} \cdot \boldsymbol{A}_{\mathrm{c}} &= \left(\boldsymbol{J}_{\ell_1}^{(M_1)} \cdot \boldsymbol{A}_{\mathrm{c}}^{(1)}\right) \diamond \cdots \diamond \left(\boldsymbol{J}_{\ell_R}^{(M_R)} \cdot \boldsymbol{A}_{\mathrm{c}}^{(R)}\right) \\
&= \left(\boldsymbol{A}_{\mathrm{c}_1}^{(1)} \cdot (\boldsymbol{\Phi}^{(1)})^{\ell_1 - 1}\right) \diamond \cdots \diamond \left(\boldsymbol{A}_{\mathrm{c}_1}^{(R)} \cdot (\boldsymbol{\Phi}^{(R)})^{\ell_R - 1}\right)
\end{aligned}$$

$$= \boldsymbol{A}_{\mathrm{SS_c}} \cdot \boldsymbol{\Phi}_{\ell_1,\ldots,\ell_{R-1},\ell_R}, \tag{7.16}$$

where we have defined

$$\begin{aligned} \boldsymbol{A}_{\mathrm{c}_1}^{(r)} &= \boldsymbol{J}_{1_r}^{(M_r)} \cdot \boldsymbol{A}_{\mathrm{c}}^{(r)} \in \mathbb{C}^{M_{\mathrm{sub}_r} \times d}, \\ \boldsymbol{A}_{\mathrm{SS_c}} &= \boldsymbol{A}_{\mathrm{c}_1}^{(1)} \diamond \cdots \diamond \boldsymbol{A}_{\mathrm{c}_1}^{(R)} = \boldsymbol{J}_{1,\cdots,1,1} \cdot \boldsymbol{A}_{\mathrm{c}} \in \mathbb{C}^{M_{\mathrm{sub}} \times d}, \end{aligned} \tag{7.17}$$

and

$$\boldsymbol{\Phi}_{\ell_1,\ldots,\ell_{R-1},\ell_R} = \boldsymbol{\Phi}^{\ell_1 - 1} \cdot \ldots \cdot \boldsymbol{\Phi}^{\ell_{R-1}-1} \cdot \boldsymbol{\Phi}^{\ell_R - 1}$$

Consequently, we can rewrite (7.15) by applying (7.16) as

$$\boldsymbol{X}_{\mathrm{SS}} = \boldsymbol{A}_{\mathrm{SS_c}} \cdot \boldsymbol{\Phi} \cdot \left(\boldsymbol{I}_L \otimes \bar{\boldsymbol{S}} \right) + \boldsymbol{N}_{\mathrm{SS}} = \boldsymbol{X}_{\mathrm{SS}_0} + \boldsymbol{N}_{\mathrm{SS}}, \tag{7.18}$$

where $\boldsymbol{\Phi} = \left[\boldsymbol{\Phi}_{1,\ldots,1,1}, \cdots, \boldsymbol{\Phi}_{1,\ldots,1,L_R}, \boldsymbol{\Phi}_{1,\ldots,2,1}, \cdots, \boldsymbol{\Phi}_{L_1,\ldots,L_{R-1},L_R} \right] \in \mathbb{C}^{d \times Ld}$, $\boldsymbol{X}_{\mathrm{SS}_0} \in \mathbb{C}^{M_{\mathrm{sub}} \times NL}$ is the noise-free spatially smoothed data matrix, and $\boldsymbol{N}_{\mathrm{SS}} \in \mathbb{C}^{M_{\mathrm{sub}} \times NL}$ is the spatially smoothed noise. Thus, spatial smoothing preprocessing reduces the array aperture to M_{sub} sensors and increases the number of snapshots by the factor L to NL.

It is apparent that $\boldsymbol{A}_{\mathrm{SS_c}}$ still satisfies the shift-invariance equation and we can write

$$\tilde{\boldsymbol{J}}_{\mathrm{SS}_1}^{(r)} \cdot \boldsymbol{A}_{\mathrm{SS_c}} \cdot \boldsymbol{\Phi}^{(r)} = \tilde{\boldsymbol{J}}_{\mathrm{SS}_2}^{(r)} \cdot \boldsymbol{A}_{\mathrm{SS_c}}, \quad r = 1, \ldots, R, \tag{7.19}$$

where $\tilde{\boldsymbol{J}}_{\mathrm{SS}_1}^{(r)}$ and $\tilde{\boldsymbol{J}}_{\mathrm{SS}_2}^{(r)} \in \mathbb{R}^{\frac{M_{\mathrm{sub}}}{M_{\mathrm{sub}_r}}(M_{\mathrm{sub}_r}-1) \times M_{\mathrm{sub}}}$ are the R-D selection matrices that select $M_{\mathrm{sub}_r} - 1$ elements for the first and the second subarray in the r-th mode, respectively. They are compactly defined as $\tilde{\boldsymbol{J}}_{\mathrm{SS}_k}^{(r)} = \boldsymbol{I}_{\prod_{l=1}^{r-1} M_{\mathrm{sub}_l}} \otimes \boldsymbol{J}_{\mathrm{SS}_k}^{(r)} \otimes \boldsymbol{I}_{\prod_{l=r+1}^{R} M_{\mathrm{sub}_l}}$ for $k = 1, 2$, where $\boldsymbol{J}_{\mathrm{SS}_k}^{(r)} \in \mathbb{R}^{(M_{\mathrm{sub}_r}-1) \times M_{\mathrm{sub}_r}}$ are the r-mode selection matrices for the first and second subarray.

We can conclude that up to

$$d \leq \min \left\{ \min_r \left((M_{\mathrm{sub}_r} - 1) M_{\mathrm{sub}} / M_{\mathrm{sub}_r} \right), NL \right\} \tag{7.20}$$

sources can be resolved after applying R-D spatial smoothing.

As (7.19) holds, the $R{\cdot}d$ spatial frequencies can be estimated by applying the R-D ESPRIT-type algorithms presented in Section 3.4 to $\boldsymbol{X}_{\mathrm{SS}}$. In R-D Standard ESPRIT, the signal subspace $\hat{\boldsymbol{U}}_{\mathrm{SS_s}} \in \mathbb{C}^{M_{\mathrm{sub}} \times d}$ is estimated by computing the d dominant left singular vectors of $\boldsymbol{X}_{\mathrm{SS}}$. As $\boldsymbol{A}_{\mathrm{SS_c}}$ and $\hat{\boldsymbol{U}}_{\mathrm{SS_s}}$ span approximately the same column space, a non-singular matrix $\boldsymbol{T} \in \mathbb{C}^{d \times d}$ can be found such that $\boldsymbol{A}_{\mathrm{SS_c}} \approx \hat{\boldsymbol{U}}_{\mathrm{SS_s}} \cdot \boldsymbol{T}$. Using this relation, the overdetermined set of R shift invariance

equations (7.19) can be expressed in terms of the estimated signal subspace, yielding

$$\tilde{\boldsymbol{J}}_{\mathrm{SS}_1}^{(r)} \cdot \hat{\boldsymbol{U}}_{\mathrm{SS_s}} \cdot \boldsymbol{\Psi}^{(r)} \approx \tilde{\boldsymbol{J}}_{\mathrm{SS}_2}^{(r)} \cdot \hat{\boldsymbol{U}}_{\mathrm{SS_s}}, \quad r = 1, \ldots, R \tag{7.21}$$

with $\boldsymbol{\Psi}^{(r)} \approx \boldsymbol{T} \cdot \boldsymbol{\Phi}^{(r)} \cdot \boldsymbol{T}^{-1}$. The R unknown matrices $\boldsymbol{\Psi}^{(r)} \in \mathbb{C}^{d \times d}$ can be estimated, e.g., via least squares (LS), i.e.,

$$\hat{\boldsymbol{\Psi}}_{\mathrm{LS}}^{(r)} = \left(\tilde{\boldsymbol{J}}_{\mathrm{SS}_1}^{(r)} \cdot \hat{\boldsymbol{U}}_{\mathrm{SS_s}} \right)^{+} \cdot \tilde{\boldsymbol{J}}_{\mathrm{SS}_2}^{(r)} \cdot \hat{\boldsymbol{U}}_{\mathrm{SS_s}} \in \mathbb{C}^{d \times d}. \tag{7.22}$$

Finally, after solving (7.22) for $\hat{\boldsymbol{\Psi}}^{(r)}$ in each mode independently, the correctly paired spatial frequency estimates are given by $\hat{\mu}_i^{(r)} = \arg\left\{ \hat{\lambda}_i^{(r)} \right\}$, $i = 1, \ldots, d$. The eigenvalues $\hat{\lambda}_i^{(r)}$ of $\hat{\boldsymbol{\Psi}}_{\mathrm{LS}}^{(r)}$ are obtained by performing a joint eigendecomposition across all R dimensions [FG06] or via the simultaneous Schur decomposition [HN98].

Alternatively, the R-D Unitary ESPRIT algorithm [HN98] from Section 3.4.2 can be applied to $\boldsymbol{X}_{\mathrm{SS}}$ to estimate the $R{\cdot}d$ parameters, which is preferable due to its better performance at low SNRs and its real-valued implementation.

7.2.4. R-D spatial smoothing for strictly non-circular sources

If only NC sources are present, a modified spatial smoothing concept can be applied to the NC model in (7.10), where we select $2M_{\mathrm{sub}}$ out of $2M$ virtual sensors. Thus, the L selection matrices in (7.14) are extended to

$$\boldsymbol{J}_{\ell_1,\ldots,\ell_{R-1},\ell_R}^{(\mathrm{nc})} = \boldsymbol{I}_2 \otimes \boldsymbol{J}_{\ell_1,\ldots,\ell_{R-1},\ell_R} \in \mathbb{R}^{2M_{\mathrm{sub}} \times 2M}. \tag{7.23}$$

The resulting spatially smoothed data matrix $\boldsymbol{X}_{\mathrm{SS}}^{(\mathrm{nc})}$ of size $2M_{\mathrm{sub}} \times NL$ is then given by

$$\begin{aligned} \boldsymbol{X}_{\mathrm{SS}}^{(\mathrm{nc})} = \Big[& \boldsymbol{J}_{1,\cdots,1,1}^{(\mathrm{nc})} \cdot \boldsymbol{X}^{(\mathrm{nc})} \quad \boldsymbol{J}_{1,\cdots,1,2}^{(\mathrm{nc})} \cdot \boldsymbol{X}^{(\mathrm{nc})} \quad \cdots \quad \boldsymbol{J}_{1,\cdots,1,L_r}^{(\mathrm{nc})} \cdot \boldsymbol{X}^{(\mathrm{nc})} \\ & \boldsymbol{J}_{1,\cdots,2,1}^{(\mathrm{nc})} \cdot \boldsymbol{X}^{(\mathrm{nc})} \quad \boldsymbol{J}_{1,\cdots,2,2}^{(\mathrm{nc})} \cdot \boldsymbol{X}^{(\mathrm{nc})} \quad \cdots \quad \boldsymbol{J}_{L_1,\cdots,L_{R-1},L_R}^{(\mathrm{nc})} \cdot \boldsymbol{X}^{(\mathrm{nc})} \Big]. \end{aligned} \tag{7.24}$$

Following the lines of the previous subsection, we can compactly express (7.24) as

$$\begin{aligned} \boldsymbol{X}_{\mathrm{SS}}^{(\mathrm{nc})} &= \boldsymbol{A}_{\mathrm{SS_c}}^{(\mathrm{nc})} \cdot \boldsymbol{\Phi} \cdot \left(\boldsymbol{I}_L \otimes \bar{\boldsymbol{S}} \right) + \boldsymbol{N}_{\mathrm{SS}}^{(\mathrm{nc})} \\ &= \boldsymbol{X}_{\mathrm{SS}_0}^{(\mathrm{nc})} + \boldsymbol{N}_{\mathrm{SS}}^{(\mathrm{nc})} \in \mathbb{C}^{2M_{\mathrm{sub}} \times NL}, \end{aligned} \tag{7.25}$$

where

$$\boldsymbol{A}_{\mathrm{SS_c}}^{(\mathrm{nc})} = \boldsymbol{J}_{1,\cdots,1,1}^{(\mathrm{nc})} \cdot \boldsymbol{A}_{\mathrm{c}}^{(\mathrm{nc})} \in \mathbb{C}^{2M_{\mathrm{sub}} \times d} \tag{7.26}$$

and $\boldsymbol{X}_{\mathrm{SS}_0}^{(\mathrm{nc})}$ is the unperturbed spatially smoothed NC data matrix. Note that spatial smoothing cannot be applied before $\boldsymbol{X}^{(\mathrm{nc})}$ is formed (7.10) as this would destroy the NC structure of the source signals.

As in the previous cases, $\boldsymbol{A}_{\mathrm{SS_c}}^{(\mathrm{nc})}$ is shift-invariant and satisfies

$$\tilde{\boldsymbol{J}}_{\mathrm{SS}_1}^{(\mathrm{nc})(r)} \cdot \boldsymbol{A}_{\mathrm{SS_c}}^{(\mathrm{nc})} \cdot \boldsymbol{\Phi}^{(r)} = \tilde{\boldsymbol{J}}_{\mathrm{SS}_2}^{(\mathrm{nc})(r)} \cdot \boldsymbol{A}_{\mathrm{SS_c}}^{(\mathrm{nc})}, \quad r = 1, \ldots, R, \tag{7.27}$$

where $\tilde{\boldsymbol{J}}_{\mathrm{SS}_k}^{(\mathrm{nc})(r)} \in \mathbb{R}^{2\frac{M_{\mathrm{sub}}}{M_{\mathrm{sub}_r}} M_{\mathrm{sub}_r}^{(\mathrm{sel})} \times 2M_{\mathrm{sub}}}$, $k = 1, 2$ are the corresponding selection matrices that select $2M_{\mathrm{sub}_r}^{(\mathrm{sel})}$ elements for the first and the second subarray in the r-th mode. They are defined as $\tilde{\boldsymbol{J}}_{\mathrm{SS}_k}^{(\mathrm{nc})(r)} = \boldsymbol{I}_{\prod_{l=1}^{r-1} M_l} \otimes \boldsymbol{J}_{\mathrm{SS}_k}^{(\mathrm{nc})(r)} \otimes \boldsymbol{I}_{\prod_{l=r+1}^{R} M_l}$, where $\boldsymbol{J}_{\mathrm{SS}_k}^{(\mathrm{nc})(r)} = \boldsymbol{I}_2 \otimes \boldsymbol{J}_{\mathrm{SS}_k}^{(r)} \in \mathbb{R}^{2M_{\mathrm{sub}_r}^{(\mathrm{sel})} \times 2M_{\mathrm{sub}_r}}$ are the r-mode selection matrices for the first and second subarray.

Thus, after applying spatial smoothing to the NC model in (7.10), we can estimate up to

$$d \leq \min\left\{ \min_r \left(2 \cdot (M_{\mathrm{sub}_r} - 1) M_{\mathrm{sub}} / M_{\mathrm{sub}_r}\right), NL \right\} \tag{7.28}$$

incoherent sources, i.e., twice as many sources as compared to (7.20).

In order to estimate the R·d parameters, the R-D NC ESPRIT-type algorithms, i.e., R-D NC Standard ESPRIT and R-D NC Unitary ESPRIT, from Section 6.2 can be applied.

7.3. Performance of R-D ESPRIT-type algorithms with spatial smoothing

In this section, we present first-order performance analysis expressions of R-D Standard ESPRIT and R-D Unitary ESPRIT both with spatial smoothing. The derived expressions rely on the data model (7.18) in Section 7.2.3.

7.3.1. Performance of R-D Standard ESPRIT with spatial smoothing

For the perturbation analysis of the estimation error, we adopt the R-D performance analysis framework in [RHD14], which is described in Chapter 4 and based on [LLV93]. The authors of [LLV93] assume a small additive noise perturbation and derive an explicit first-order error expansion of the subspace estimation error in terms of the noise $\boldsymbol{N}$, which is followed by a corresponding expression for the parameter estimation error $\Delta\mu_i$. As a follow-up, analytical expressions for the MSE that only require a zero mean and finite SO moments of the noise have been derived in [RHD14]. From (7.18), it is clear that these assumptions are not violated by spatial smoothing such that the R-D performance analysis framework in [RHD14] is still applicable for the performance analysis of the spatially smoothed R-D ESPRIT-type algorithms.

To derive the signal subspace estimation error for (7.18), we express the SVD of the noise-free spatially smoothed observations $\boldsymbol{X}_{\mathrm{SS}_0}$ as

$$\boldsymbol{X}_{\mathrm{SS}_0} = \begin{bmatrix} \boldsymbol{U}_{\mathrm{SS_s}} & \boldsymbol{U}_{\mathrm{SS_n}} \end{bmatrix} \cdot \begin{bmatrix} \boldsymbol{\Sigma}_{\mathrm{SS_s}} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} \end{bmatrix} \cdot \begin{bmatrix} \boldsymbol{V}_{\mathrm{SS_s}} & \boldsymbol{V}_{\mathrm{SS_n}} \end{bmatrix}^{\mathrm{H}}, \tag{7.29}$$

where $\boldsymbol{U}_{\mathrm{SS_s}} \in \mathbb{C}^{M_{\mathrm{sub}} \times d}$, $\boldsymbol{U}_{\mathrm{SS_n}} \in \mathbb{C}^{M_{\mathrm{sub}} \times (NL-d)}$, and $\boldsymbol{V}_{\mathrm{SS_s}} \in \mathbb{C}^{NL \times d}$ span the signal subspace, the noise subspace, and the row space, respectively, and $\boldsymbol{\Sigma}_{\mathrm{SS_s}} \in \mathbb{R}^{d \times d}$ contains the non-zero singular values on its diagonal. Writing the perturbed signal subspace estimate $\hat{\boldsymbol{U}}_{\mathrm{SS_s}}$ computed from the SVD of $\boldsymbol{X}_{\mathrm{SS}}$ as $\hat{\boldsymbol{U}}_{\mathrm{SS_s}} = \boldsymbol{U}_{\mathrm{SS_s}} + \Delta\boldsymbol{U}_{\mathrm{SS_s}}$, where $\Delta\boldsymbol{U}_{\mathrm{SS_s}}$ denotes the signal subspace error, the first-order approximation using is given according to (4.4) by

$$\Delta\boldsymbol{U}_{\mathrm{SS_s}} \approx \boldsymbol{U}_{\mathrm{SS_n}} \cdot \boldsymbol{U}_{\mathrm{SS_n}}^{\mathrm{H}} \cdot \boldsymbol{N}_{\mathrm{SS}} \cdot \boldsymbol{V}_{\mathrm{SS_s}} \cdot \boldsymbol{\Sigma}_{\mathrm{SS_s}}^{-1}. \tag{7.30}$$

For the estimation error $\Delta\mu_i^{(r)}$ of the i-th spatial frequency in the r-th mode obtained by the LS solution, we have

$$\Delta\mu_i^{(r)} \approx \mathrm{Im}\left\{ \boldsymbol{p}_i^{\mathrm{T}} \cdot \left(\tilde{\boldsymbol{J}}_{\mathrm{SS}_1}^{(r)} \cdot \boldsymbol{U}_{\mathrm{SS_s}} \right)^{+} \cdot \left[\tilde{\boldsymbol{J}}_{\mathrm{SS}_2}^{(r)} / \lambda_i^{(r)} - \tilde{\boldsymbol{J}}_{\mathrm{SS}_1}^{(r)} \right] \cdot \Delta\boldsymbol{U}_{\mathrm{SS_s}} \cdot \boldsymbol{q}_i \right\}, \tag{7.31}$$

where $\lambda_i^{(r)} = \mathrm{e}^{\mathrm{j}\mu_i^{(r)}}$ is the i-th eigenvalue of $\boldsymbol{\Psi}^{(r)}$, $\boldsymbol{q}_i$ represents the i-th eigenvector of $\boldsymbol{\Psi}^{(r)}$ and the i-th column vector of the eigenvector matrix $\boldsymbol{Q}$, and $\boldsymbol{p}_i^{\mathrm{T}}$ is the i-th row vector of $\boldsymbol{P} = \boldsymbol{Q}^{-1}$. Hence, the eigendecomposition of $\boldsymbol{\Psi}^{(r)}$ is given by $\boldsymbol{\Psi}^{(r)} = \boldsymbol{Q} \cdot \boldsymbol{\Lambda}^{(r)} \cdot \boldsymbol{Q}^{-1}$, where $\boldsymbol{\Lambda}^{(r)}$ contains the eigenvalues $\lambda_i^{(r)}$ on its diagonal.

Finally, to compute the first-order MSE expression for R-D Standard ESPRIT with spatial smoothing, we extend the results in [RHD14] reviewed in Chapter 4. The MSE for the i-th spatial frequency in the r-th mode is given by

$$\mathbb{E}\left\{ (\Delta\mu_i^{(r)})^2 \right\} \approx \frac{1}{2} \cdot \left(\boldsymbol{z}_{\mathrm{SS}_i}^{(r)^{\mathrm{H}}} \cdot \boldsymbol{R}_{\mathrm{SS}}^{\mathrm{T}} \cdot \boldsymbol{z}_{\mathrm{SS}_i}^{(r)} - \mathrm{Re}\left\{ \boldsymbol{z}_{\mathrm{SS}_i}^{(r)^{\mathrm{T}}} \cdot \boldsymbol{C}_{\mathrm{SS}}^{\mathrm{T}} \cdot \boldsymbol{z}_{\mathrm{SS}_i}^{(r)} \right\} \right), \tag{7.32}$$

where $\boldsymbol{z}_{\mathrm{SS}_i}^{(r)} = \boldsymbol{W}_{\mathrm{SS}}^{\mathrm{T}} \cdot \boldsymbol{r}_{\mathrm{SS}_i}^{(r)} \in \mathbb{C}^{M_{\mathrm{sub}}NL \times 1}$ with

$$\boldsymbol{r}_{\mathrm{SS}_i}^{(r)} = \boldsymbol{q}_i \otimes \left(\left[\left(\tilde{\boldsymbol{J}}_{\mathrm{SS}_1}^{(r)} \cdot \boldsymbol{U}_{\mathrm{SS_s}} \right)^{+} \cdot \left(\tilde{\boldsymbol{J}}_{\mathrm{SS}_2}^{(r)} / \lambda_i^{(r)} - \tilde{\boldsymbol{J}}_{\mathrm{SS}_1}^{(r)} \right) \right]^{\mathrm{T}} \cdot \boldsymbol{p}_i \right) \in \mathbb{C}^{M_{\mathrm{sub}}d \times 1} \tag{7.33}$$

$$\boldsymbol{W}_{\mathrm{SS}} = \left(\boldsymbol{\Sigma}_{\mathrm{SS_s}}^{-1} \cdot \boldsymbol{V}_{\mathrm{SS_s}}^{\mathrm{T}} \right) \otimes \left(\boldsymbol{U}_{\mathrm{SS_n}} \cdot \boldsymbol{U}_{\mathrm{SS_n}}^{\mathrm{H}} \right) \in \mathbb{C}^{M_{\mathrm{sub}}d \times M_{\mathrm{sub}}NL}. \tag{7.34}$$

In order to compute (7.32), we require the covariance matrix $\boldsymbol{R}_{\mathrm{SS}} = \mathbb{E}\{\boldsymbol{n}_{\mathrm{SS}} \cdot \boldsymbol{n}_{\mathrm{SS}}^{\mathrm{H}}\} \in \mathbb{C}^{M_{\mathrm{sub}}NL \times M_{\mathrm{sub}}NL}$ and the pseudo-covariance matrix $\boldsymbol{C}_{\mathrm{SS}} = \mathbb{E}\{\boldsymbol{n}_{\mathrm{SS}} \cdot \boldsymbol{n}_{\mathrm{SS}}^{\mathrm{T}}\} \in \mathbb{C}^{M_{\mathrm{sub}}NL \times M_{\mathrm{sub}}NL}$ of the spatially smoothed noise $\boldsymbol{n}_{\mathrm{SS}} = \mathrm{vec}\{\boldsymbol{N}_{\mathrm{SS}}\} \in \mathbb{C}^{M_{\mathrm{sub}}NL \times 1}$. It is clear that the spatial smoothing preprocessing step

modifies the prior noise statistics, resulting in colored noise. In what follows, we analytically derive the SO noise statistics of the spatially smoothed noise. We first expand $\boldsymbol{n}_{\mathrm{SS}}$ as

$$\begin{aligned}\boldsymbol{n}_{\mathrm{SS}} &= \mathrm{vec}\left\{\begin{bmatrix}\boldsymbol{J}_{1,\cdots,1,1}\cdot\boldsymbol{N} & \cdots & \boldsymbol{J}_{L_1,\cdots,L_{R-1},L_R}\cdot\boldsymbol{N}\end{bmatrix}\right\}\\ &= \begin{bmatrix}(\boldsymbol{I}_N\otimes\boldsymbol{J}_{1,\cdots,1,1})\\ \vdots\\ (\boldsymbol{I}_N\otimes\boldsymbol{J}_{L_1,\cdots,L_{R-1},L_R})\end{bmatrix}\cdot\boldsymbol{n} = \boldsymbol{M}\cdot\boldsymbol{n},\end{aligned} \tag{7.35}$$

where $\boldsymbol{M}\in\mathbb{R}^{M_{\mathrm{sub}}NL\times MN}$, $\boldsymbol{n}=\mathrm{vec}\{\boldsymbol{N}\}\in\mathbb{C}^{MN\times 1}$ is the unsmoothed noise contribution, and we have used the property (1.14). Thus, the SO statistics of $\boldsymbol{n}_{\mathrm{SS}}$ can be expressed in terms of the covariance matrix $\boldsymbol{R}_{\mathrm{nn}}=\mathbb{E}\{\boldsymbol{n}\cdot\boldsymbol{n}^{\mathrm{H}}\}\in\mathbb{C}^{MN\times MN}$ and the pseudo-covariance matrix $\boldsymbol{C}_{\mathrm{nn}}=\mathbb{E}\{\boldsymbol{n}\cdot\boldsymbol{n}^{\mathrm{T}}\}\in\mathbb{C}^{MN\times MN}$ of $\boldsymbol{n}$. We obtain

$$\boldsymbol{R}_{\mathrm{SS}}=\boldsymbol{M}\cdot\boldsymbol{R}_{\mathrm{nn}}\cdot\boldsymbol{M}^{\mathrm{T}},\quad \boldsymbol{C}_{\mathrm{SS}}=\boldsymbol{M}\cdot\boldsymbol{C}_{\mathrm{nn}}\cdot\boldsymbol{M}^{\mathrm{T}}. \tag{7.36}$$

7.3.2. Performance of R-D Unitary ESPRIT with spatial smoothing

It was proven in [Roe13] and discussed in Section 4.3.2 that the asymptotic performance of R-D Unitary-ESPRIT is found once forward-backward averaging (FBA) (cf. Section 3.2.3) is taken into account. FBA is performed by replacing the spatially smoothed data matrix $\boldsymbol{X}_{\mathrm{SS}}\in\mathbb{C}^{M_{\mathrm{sub}}\times NL}$ by the column-augmented data matrix $\tilde{\boldsymbol{X}}_{\mathrm{SS}}\in\mathbb{C}^{M_{\mathrm{sub}}\times 2NL}$ defined by

$$\tilde{\boldsymbol{X}}_{\mathrm{SS}}=\begin{bmatrix}\boldsymbol{X}_{\mathrm{SS}} & \boldsymbol{\Pi}_{M_{\mathrm{sub}}}\cdot\boldsymbol{X}_{\mathrm{SS}}^{*}\cdot\boldsymbol{\Pi}_{NL}\end{bmatrix}=\tilde{\boldsymbol{X}}_{\mathrm{SS}_0}+\tilde{\boldsymbol{N}}_{\mathrm{SS}}, \tag{7.37}$$

where $\tilde{\boldsymbol{X}}_{\mathrm{SS}_0}$ is the noiseless FBA-processed spatially smoothed data matrix. Following the steps of the previous subsection, the first-order MSE expression for R-D Unitary ESPRIT with spatial smoothing for the i-th spatial frequency in the r-th mode is given by

$$\mathbb{E}\left\{(\Delta\mu_i^{(r)})^2\right\}\approx\frac{1}{2}\cdot\left(\tilde{\boldsymbol{z}}_{\mathrm{SS}_i}^{(r)^{\mathrm{H}}}\cdot\tilde{\boldsymbol{R}}_{\mathrm{SS}}^{\mathrm{T}}\cdot\tilde{\boldsymbol{z}}_{\mathrm{SS}_i}^{(r)}-\mathrm{Re}\left\{\tilde{\boldsymbol{z}}_{\mathrm{SS}_i}^{(r)^{\mathrm{T}}}\cdot\tilde{\boldsymbol{C}}_{\mathrm{SS}}^{\mathrm{T}}\cdot\tilde{\boldsymbol{z}}_{\mathrm{SS}_i}^{(r)}\right\}\right), \tag{7.38}$$

where $\tilde{\boldsymbol{z}}_{\mathrm{SS}_i}^{(r)}=\tilde{\boldsymbol{W}}_{\mathrm{SS}}^{\mathrm{T}}\cdot\tilde{\boldsymbol{r}}_{\mathrm{SS}_i}^{(r)}\in\mathbb{C}^{2M_{\mathrm{sub}}NL\times 1}$ with

$$\begin{aligned}\tilde{\boldsymbol{r}}_{\mathrm{SS}_i}^{(r)}&=\tilde{\boldsymbol{q}}_i\otimes\left(\left[\left(\tilde{\boldsymbol{J}}_{\mathrm{SS}_1}^{(r)}\cdot\tilde{\boldsymbol{U}}_{\mathrm{SS}_{\mathrm{s}}}\right)^{+}\cdot\left(\tilde{\boldsymbol{J}}_{\mathrm{SS}_2}^{(r)}/\lambda_i^{(r)}-\tilde{\boldsymbol{J}}_{\mathrm{SS}_1}^{(r)}\right)\right]^{\mathrm{T}}\cdot\tilde{\boldsymbol{p}}_i\right)\in\mathbb{C}^{M_{\mathrm{sub}}d\times 1},\\ \tilde{\boldsymbol{W}}_{\mathrm{SS}}&=\left(\tilde{\boldsymbol{\Sigma}}_{\mathrm{SS}_{\mathrm{s}}}^{-1}\cdot\tilde{\boldsymbol{V}}_{\mathrm{SS}_{\mathrm{s}}}^{\mathrm{T}}\right)\otimes\left(\tilde{\boldsymbol{U}}_{\mathrm{SS}_{\mathrm{n}}}\cdot\tilde{\boldsymbol{U}}_{\mathrm{SS}_{\mathrm{n}}}^{\mathrm{H}}\right)\in\mathbb{C}^{M_{\mathrm{sub}}d\times 2M_{\mathrm{sub}}NL},\end{aligned} \tag{7.39}$$

where we have replaced the noise-free subspaces of $\boldsymbol{X}_{\mathrm{SS}_0}$ in (7.32) by the corresponding subspaces of $\tilde{\boldsymbol{X}}_{\mathrm{SS}_0}$, and $\boldsymbol{p}_i$ and $\boldsymbol{q}_i$ by $\tilde{\boldsymbol{p}}_i$ and $\tilde{\boldsymbol{q}}_i$, respectively. It can be shown that $\tilde{\boldsymbol{n}}_{\mathrm{SS}}=\mathrm{vec}\{\tilde{\boldsymbol{N}}_{\mathrm{SS}}\}\in\mathbb{C}^{2M_{\mathrm{sub}}NL\times 1}$

is given by

$$\begin{aligned}\tilde{\boldsymbol{n}}_{\mathrm{SS}} &= \mathrm{vec}\left\{\begin{bmatrix}\boldsymbol{N}_{\mathrm{SS}} & \boldsymbol{\Pi}_{M_{\mathrm{sub}}} \cdot \boldsymbol{N}_{\mathrm{SS}}^{*} \cdot \boldsymbol{\Pi}_{NL}\end{bmatrix}\right\} \\ &= \begin{bmatrix}\mathrm{vec}\left\{\boldsymbol{N}_{\mathrm{SS}}\right\} \\ \mathrm{vec}\left\{\boldsymbol{\Pi}_{M_{\mathrm{sub}}} \cdot \boldsymbol{N}_{\mathrm{SS}}^{*} \cdot \boldsymbol{\Pi}_{NL}\right\}\end{bmatrix} = \begin{bmatrix}\boldsymbol{n}_{\mathrm{SS}} \\ \boldsymbol{\Pi}_{M_{\mathrm{sub}}NL} \cdot \boldsymbol{n}_{\mathrm{SS}}^{*}\end{bmatrix}. \end{aligned} \tag{7.40}$$

Therefore, the expressions for $\tilde{\boldsymbol{R}}_{\mathrm{SS}} = \mathbb{E}\{\tilde{\boldsymbol{n}}_{\mathrm{SS}} \cdot \tilde{\boldsymbol{n}}_{\mathrm{SS}}^{\mathrm{H}}\} \in \mathbb{C}^{2M_{\mathrm{sub}}NL \times 2M_{\mathrm{sub}}NL}$ and $\tilde{\boldsymbol{C}}_{\mathrm{SS}} = \mathbb{E}\{\tilde{\boldsymbol{n}}_{\mathrm{SS}} \cdot \tilde{\boldsymbol{n}}_{\mathrm{SS}}^{\mathrm{T}}\} \in \mathbb{C}^{2M_{\mathrm{sub}}NL \times 2M_{\mathrm{sub}}NL}$ can be derived in terms of (7.36) as

$$\tilde{\boldsymbol{R}}_{\mathrm{SS}} = \boldsymbol{P} \cdot \begin{bmatrix}\boldsymbol{R}_{\mathrm{SS}} & \boldsymbol{C}_{\mathrm{SS}} \\ \boldsymbol{C}_{\mathrm{SS}}^{*} & \boldsymbol{R}_{\mathrm{SS}}^{*}\end{bmatrix} \cdot \boldsymbol{P}^{\mathrm{T}}, \quad \tilde{\boldsymbol{C}}_{\mathrm{SS}} = \boldsymbol{P} \cdot \begin{bmatrix}\boldsymbol{C}_{\mathrm{SS}} & \boldsymbol{R}_{\mathrm{SS}} \\ \boldsymbol{R}_{\mathrm{SS}}^{*} & \boldsymbol{C}_{\mathrm{SS}}^{*}\end{bmatrix} \cdot \boldsymbol{P}^{\mathrm{T}},$$

where $\boldsymbol{P} = \mathrm{blkdiag}\{\boldsymbol{I}_{M_{\mathrm{sub}}NL},\ \boldsymbol{\Pi}_{M_{\mathrm{sub}}NL}\}$.

7.4. Performance of R-D NC ESPRIT-type algorithms with spatial smoothing

In this section, we derive first-order analytical error expressions of R-D NC Standard ESPRIT and R-D NC Unitary ESPRIT for strictly non-circular sources both with spatial smoothing. As will be shown in Subsection 7.4.2, the performance of both algorithms is asymptotically identical in the high effective SNR. Therefore, we first resort to the simpler derivation for the spatially smoothed R-D NC Standard ESPRIT algorithm and then show its equivalence to the spatially smoothed R-D NC Unitary ESPRIT algorithm. Our results are based on the data model (7.25) in Section 7.2.4.

7.4.1. Performance of R-D NC Standard ESPRIT with spatial smoothing

In Chapter 6, we have shown that the R-D performance analysis framework in [RHD14] is still applicable to the augmented measurement matrix $\boldsymbol{X}^{(\mathrm{nc})}$ (7.10) obtained by the preprocessing scheme for non-circular sources. From (7.25), it is apparent that adding spatial smoothing as a second preprocessing step does not violate the assumptions, such that the steps from Section 7.3.1 can be applied to the spatially smoothed augmented data matrix $\boldsymbol{X}_{\mathrm{SS}}^{(\mathrm{nc})}$.

As a result, equivalently to (7.32), the first-order MSE expression for R-D NC Standard ESPRIT with spatial smoothing for the i-th spatial frequency in the r-th mode is given by

$$\mathbb{E}\left\{(\Delta\mu_i^{(r)})^2\right\} \approx \frac{1}{2} \cdot \left(\boldsymbol{z}_{\mathrm{SS}_i}^{(\mathrm{nc})(r)^{\mathrm{H}}} \cdot \boldsymbol{R}_{\mathrm{SS}}^{(\mathrm{nc})^{\mathrm{T}}} \cdot \boldsymbol{z}_{\mathrm{SS}_i}^{(\mathrm{nc})(r)} - \mathrm{Re}\left\{\boldsymbol{z}_{\mathrm{SS}_i}^{(\mathrm{nc})(r)^{\mathrm{T}}} \cdot \boldsymbol{C}_{\mathrm{SS}}^{(\mathrm{nc})^{\mathrm{T}}} \cdot \boldsymbol{z}_{\mathrm{SS}_i}^{(\mathrm{nc})(r)}\right\}\right), \tag{7.41}$$

where $\boldsymbol{z}_{\mathrm{SS}_i}^{(\mathrm{nc})(r)} = \boldsymbol{W}_{\mathrm{SS}}^{(\mathrm{nc})^{\mathrm{T}}} \cdot \boldsymbol{r}_{\mathrm{SS}_i}^{(\mathrm{nc})(r)} \in \mathbb{C}^{2M_{\mathrm{sub}}NL \times 1}$ with

$$\tilde{\boldsymbol{r}}_{\mathrm{SS}_i}^{(\mathrm{nc})(r)} = \boldsymbol{q}_i^{(\mathrm{nc})} \otimes \left(\left[\left(\tilde{\boldsymbol{J}}_{\mathrm{SS}_1}^{(\mathrm{nc})(r)} \cdot \boldsymbol{U}_{\mathrm{SS_s}}^{(\mathrm{nc})} \right)^{+} \cdot \left(\tilde{\boldsymbol{J}}_{\mathrm{SS}_2}^{(\mathrm{nc})(r)} / \lambda_i^{(r)} - \tilde{\boldsymbol{J}}_{\mathrm{SS}_1}^{(\mathrm{nc})(r)} \right) \right]^{\mathrm{T}} \cdot \boldsymbol{p}_i^{(\mathrm{nc})} \right) \in \mathbb{C}^{2M_{\mathrm{sub}}d \times 1}, \tag{7.42}$$

$$\boldsymbol{W}_{\mathrm{SS}}^{(\mathrm{nc})} = \left(\boldsymbol{\Sigma}_{\mathrm{SS_s}}^{(\mathrm{nc})^{-1}} \cdot \boldsymbol{V}_{\mathrm{SS_s}}^{(\mathrm{nc})^{\mathrm{T}}} \right) \otimes \left(\boldsymbol{U}_{\mathrm{SS_n}}^{(\mathrm{nc})} \cdot \boldsymbol{U}_{\mathrm{SS_n}}^{(\mathrm{nc})^{\mathrm{H}}} \right) \in \mathbb{C}^{2M_{\mathrm{sub}}d \times 2M_{\mathrm{sub}}NL}, \tag{7.43}$$

where $\boldsymbol{p}_i^{(\mathrm{nc})}$ and $\boldsymbol{q}_i^{(\mathrm{nc})}$ replace $\boldsymbol{p}_i$ and $\boldsymbol{q}_i$, respectively, we have used the corresponding subspaces of $\boldsymbol{X}_{\mathrm{SS}_0}^{(\mathrm{nc})}$ defined in (7.25), and the selection matrices $\tilde{\boldsymbol{J}}_{\mathrm{SS}_k}^{(\mathrm{nc})(r)}, k = 1, 2$, are given in (7.27).

The spatially smoothed augmented noise contribution $\boldsymbol{n}_{\mathrm{SS}}^{(\mathrm{nc})} = \mathrm{vec}\{\boldsymbol{N}_{\mathrm{SS}}^{(\mathrm{nc})}\} \in \mathbb{C}^{2M_{\mathrm{sub}}NL \times 1}$ can be expressed similarly to (7.35) as

$$\begin{aligned} \boldsymbol{n}_{\mathrm{SS}}^{(\mathrm{nc})} &= \mathrm{vec}\left\{ \left[\boldsymbol{J}_{1,\cdots,1,1}^{(\mathrm{nc})} \cdot \boldsymbol{N}^{(\mathrm{nc})} \quad \cdots \quad \boldsymbol{J}_{L_1,\cdots,L_{R-1},L_R}^{(\mathrm{nc})} \cdot \boldsymbol{N}^{(\mathrm{nc})} \right] \right\} \\ &= \begin{bmatrix} (\boldsymbol{I}_N \otimes \boldsymbol{J}_{1,\cdots,1,1}^{(\mathrm{nc})}) \\ \vdots \\ (\boldsymbol{I}_N \otimes \boldsymbol{J}_{L_1,\cdots,L_{R-1},L_R}^{(\mathrm{nc})}) \end{bmatrix} \cdot \boldsymbol{n}^{(\mathrm{nc})} = \boldsymbol{M}^{(\mathrm{nc})} \cdot \cdot \boldsymbol{n}^{(\mathrm{nc})}, \end{aligned} \tag{7.44}$$

where $\boldsymbol{M}^{(\mathrm{nc})} \in \mathbb{R}^{2M_{\mathrm{sub}}NL \times 2MN}$ and $\boldsymbol{n}^{(\mathrm{nc})} = \mathrm{vec}\left\{\boldsymbol{N}^{(\mathrm{nc})}\right\} \in \mathbb{C}^{2MN \times 1}$. We have shown in Theorem 6.4.1 in Chapter 6 that $\boldsymbol{n}^{(\mathrm{nc})}$ can be represented as

$$\boldsymbol{n}^{(\mathrm{nc})} = \tilde{\boldsymbol{K}} \cdot \begin{bmatrix} \boldsymbol{n} \\ \boldsymbol{n}^* \end{bmatrix}, \tag{7.45}$$

where $\tilde{\boldsymbol{K}} = \boldsymbol{K}_{2M,N}^{\mathrm{T}} \cdot \mathrm{blkdiag}\{\boldsymbol{K}_{M,N}, \boldsymbol{K}_{M,N} \cdot (\boldsymbol{I}_N \otimes \boldsymbol{\Pi}_M)\}$ and $\boldsymbol{K}_{M,N} \in \mathbb{R}^{MN \times MN}$ is the commutation matrix defined in (1.21). Then, $\boldsymbol{R}_{\mathrm{SS}}^{(\mathrm{nc})} = \mathbb{E}\{\boldsymbol{n}_{\mathrm{SS}}^{(\mathrm{nc})} \cdot \boldsymbol{n}_{\mathrm{SS}}^{(\mathrm{nc})^{\mathrm{H}}}\} \in \mathbb{C}^{2M_{\mathrm{sub}}NL \times 2M_{\mathrm{sub}}NL}$ and $\boldsymbol{C}_{\mathrm{SS}}^{(\mathrm{nc})} = \mathbb{E}\{\boldsymbol{n}_{\mathrm{SS}}^{(\mathrm{nc})} \cdot \boldsymbol{n}_{\mathrm{SS}}^{(\mathrm{nc})^{\mathrm{T}}}\} \in \mathbb{C}^{2M_{\mathrm{sub}}NL \times 2M_{\mathrm{sub}}NL}$ can be computed as

$$\boldsymbol{R}_{\mathrm{SS}}^{(\mathrm{nc})} = \boldsymbol{M}^{(\mathrm{nc})} \cdot \boldsymbol{R}_{\mathrm{nn}}^{(\mathrm{nc})} \cdot \boldsymbol{M}^{(\mathrm{nc})^{\mathrm{T}}}, \quad \boldsymbol{C}_{\mathrm{SS}}^{(\mathrm{nc})} = \boldsymbol{M}^{(\mathrm{nc})} \cdot \boldsymbol{C}_{\mathrm{nn}}^{(\mathrm{nc})} \cdot \boldsymbol{M}^{(\mathrm{nc})^{\mathrm{T}}}, \tag{7.46}$$

where $\boldsymbol{R}_{\mathrm{nn}}^{(\mathrm{nc})} \in \mathbb{C}^{2MN \times 2MN}$ and $\boldsymbol{C}_{\mathrm{nn}}^{(\mathrm{nc})} \in \mathbb{C}^{2MN \times 2MN}$ are given according to Theorem 6.4.1 as

$$\boldsymbol{R}_{\mathrm{nn}}^{(\mathrm{nc})} = \mathbb{E}\left\{ \boldsymbol{n}^{(\mathrm{nc})} \cdot \boldsymbol{n}^{(\mathrm{nc})^{\mathrm{H}}} \right\} = \tilde{\boldsymbol{K}} \cdot \begin{bmatrix} \boldsymbol{R}_{\mathrm{nn}} & \boldsymbol{C}_{\mathrm{nn}} \\ \boldsymbol{C}_{\mathrm{nn}}^* & \boldsymbol{R}_{\mathrm{nn}}^* \end{bmatrix} \cdot \tilde{\boldsymbol{K}}^{\mathrm{T}}, \tag{7.47}$$

$$\boldsymbol{C}_{\mathrm{nn}}^{(\mathrm{nc})} = \mathbb{E}\left\{ \boldsymbol{n}^{(\mathrm{nc})} \boldsymbol{n}^{(\mathrm{nc})^{\mathrm{T}}} \right\} = \tilde{\boldsymbol{K}} \cdot \begin{bmatrix} \boldsymbol{C}_{\mathrm{nn}} & \boldsymbol{R}_{\mathrm{nn}} \\ \boldsymbol{R}_{\mathrm{nn}}^* & \boldsymbol{C}_{\mathrm{nn}}^* \end{bmatrix} \cdot \tilde{\boldsymbol{K}}^{\mathrm{T}}. \tag{7.48}$$

7.4.2. Performance of R-D NC Unitary ESPRIT with spatial smoothing

We have shown in Section 6.4.2 that R-D NC Standard ESPRIT and R-D NC Unitary ESPRIT

both have the same asymptotic performance in the high effective SNR regime. It was established that applying FBA to the augmented matrix $\boldsymbol{X}^{(\mathrm{nc})}$ does not improve the signal subspace estimate and that the real-valued transformation has no effect on the asymptotic performance in the high effective SNR. In this subsection, we prove that these properties still hold when spatial smoothing is applied to both algorithms. To this end, we first investigate the effect of FBA and state the following theorem:

Theorem 7.4.1. *Applying FBA to* $\boldsymbol{X}_{\mathrm{SS}}^{(\mathrm{nc})}$ *does not improve the signal subspace estimate.*

The proof is given in Appendix B.33.

Next, we analyze the real-valued transformation as the second preprocessing step of R-D NC Unitary ESPRIT with spatial smoothing and formulate the theorem:

Theorem 7.4.2. *The spatially smoothed R-D NC Unitary ESPRIT algorithm and the spatially smoothed R-D NC Standard ESPRIT algorithm with FBA preprocessing perform asymptotically identical in the high effective SNR.*

The proof of this theorem follows the same steps as the one for the case without spatial smoothing considered in Appendix B.25 for the proof of Theorem 6.4.2. This is due to the fact that spatial smoothing modifies the NC signal subspace of R-D NC Standard ESPRIT and R-D NC Unitary ESPRIT in the same way.

As a result of Theorem 7.4.1 and Theorem 7.4.2, we can conclude that the asymptotic performance of R-D NC Standard ESPRIT and R-D NC Unitary ESPRIT both with spatial smoothing is identical in the high effective SNR.

7.5. Single source case

In this section, we simplify the derive MSE expressions for the matrix-based R-D ESPRIT-type and R-D NC ESPRIT-type algorithms for the special case of a single source.

The derived analytical MSE expressions for the R-D ESPRIT-type methods with spatial smoothing are deterministic and formulated in terms of the subspaces of the noise-free observations. In Section 4.5 and in Section 6.6, we have already considered the special case of a single source for R-D ESPRIT-type algorithms and R-D NC ESPRIT-type algorithms without spatial smoothing to gain explicit insights into how the MSE expressions depend on the physical parameters, e.g., the number of sensors M, the sample size N, and the SNR. The knowledge of how the MSE expressions depend on these parameters can be of practical significance. For instance, this enables an objective comparison of different estimators or facilitates array design decisions on the value of M required to achieve a target MSE for a specific SNR. Note that establishing general MSE expressions for an arbitrary number of sources is challenging given the complex dependence of the subspaces on

the physical parameters. For the single source case, it was shown in Section 4.5 and in Section 6.6 that neither FBA nor NC preprocessing can improve the MSE. However, in this section, we show that a significant gain can be achieved for the MSE of R-D ESPRIT-type methods for a single source when spatial smoothing is applied. Assuming an R-D uniform sampling grid, i.e., a ULA in each mode, and circularly symmetric white noise, we simplify the derived MSE expressions in (7.32), (7.38), and (7.41) for this special case. The result depends on the number of subarrays L_r in the r-th mode as a design parameter, which we analytically compute in the R-D case by minimizing the MSE. It should be emphasized that these results for the special case $R = 1$ are in line with those derived in [RH90, RH93, HS90] for harmonic retrieval. Here, the R-D extension is provided. Based on our R-D results, we explicitly compute the asymptotic spatial smoothing gain for arbitrary R and the asymptotic efficiency for $R = 1$ in closed-form.

7.5.1. R-D ESPRIT-type algorithms with spatial smoothing

The final result for the simplified MSE expressions is summarized in the following theorem:

Theorem 7.5.1. *For the case of an M-element R-D uniform sampling grid with an M_r-element ULA in the r-th mode, a single source ($d = 1$), and circularly symmetric white noise, the MSE in the r-th mode of R-D Standard ESPRIT and R-D Unitary ESPRIT with spatial smoothing as well as the MSE in the r-th mode of R-D NC Standard ESPRIT and R-D NC Unitary ESPRIT with spatial smoothing for a single source are given by $\mathrm{MSE}_{\mathrm{SS}}^{(r)} = \mathbb{E}\left\{(\Delta\mu^{(r)})^2\right\}$, yielding*

$$\mathrm{MSE}_{\mathrm{SS}}^{(r)} \approx \begin{cases} \frac{1}{\hat{\rho}} \cdot \frac{1}{(M_r - L_r)^2 L_r} \cdot \prod_{\substack{p=1 \\ p \neq r}}^{R} \frac{c_p}{M_{\mathrm{sub}_p}^2 L_p^2} & \text{if } L_r \leq \frac{M_r}{2} \\ \frac{1}{\hat{\rho}} \cdot \frac{1}{(M_r - L_r) L_r^2} \cdot \prod_{\substack{p=1 \\ p \neq r}}^{R} \frac{c_p}{M_{\mathrm{sub}_p}^2 L_p^2} & \text{if } L_r > \frac{M_r}{2}, \end{cases} \tag{7.49}$$

where c_p is given as

$$\begin{aligned} c_p =& \frac{1}{3} \cdot \Big(\min\{L_p, M_p - L_p\} + 1 \Big) \\ & \cdot \Big(\min\{L_p, M_p - L_p\} \big(2 \cdot \min\{L_p, M_p - L_p\} - 3 \cdot M_p - 2 \big) + 6 \cdot M_{\mathrm{sub}_p} L_p \Big) - M_{\mathrm{sub}_p} L_p \end{aligned} \tag{7.50}$$

and $\hat{\rho}$ represents the effective SNR $\hat{\rho} = N\hat{P}_{\mathrm{s}}/\sigma_{\mathrm{n}}^2$ with $\hat{P}_{\mathrm{s}}$ being the empirical source power given by $\hat{P}_{\mathrm{s}} = \|\boldsymbol{s}\|_2^2/N$ and $\boldsymbol{s} \in \mathbb{C}^{N \times 1}$.

For the proof, the reader is referred to Appendix B.34.

Note that (7.49) as a function of L_r is symmetric with respect to $L_r = M_r/2$. Moreover, (7.49) does not depend on the location of the phase reference $\delta^{(r)}$ of the array. In the special case of

$R = 1$, where $M_r = M$ and $L_r = L$, the MSE in (7.49) simplifies to

$$\mathrm{MSE}_{\mathrm{SS}} \approx \begin{cases} \frac{1}{\hat{\rho}} \cdot \frac{1}{(M-L)^2 L} & \text{if } L \leq \frac{M}{2} \\ \frac{1}{\hat{\rho}} \cdot \frac{1}{(M-L)L^2} & \text{if } L > \frac{M}{2}. \end{cases} \tag{7.51}$$

Interestingly, we arrive at the same result for the MSE of all the considered spatially smoothed R-D ESPRIT-type algorithms for a single source, i.e., no additional gain from FBA or NC preprocessing can be achieved.

7.5.2. Optimal number of subarrays for spatial smoothing

In the MSE expression in (7.49), the number of subarrays L_r in each mode is a design parameter that can be optimized. Therefore, minimizing the MSE expression (7.49) with respect to L_r, yields[3]

$$L_r^{\mathrm{opt}} = \begin{cases} \frac{1}{3} \cdot M_r & \text{if } L_r \leq \frac{M_r}{2} \\ \frac{2}{3} \cdot M_r & \text{if } L_r > \frac{M_r}{2}, \end{cases} \tag{7.52}$$

where it is assumed that M_r is a multiple of 3. A short proof is provided in Appendix B.35. If M_r is not a multiple of 3, we round to the nearest integer. Then, L_r^{opt} for the case $L_r \leq \frac{M_r}{2}$, for instance, is given by

$$L_r^{\mathrm{opt}} = \begin{cases} \frac{1}{3} \cdot (M_r - 1) & \text{if } M_r \bmod 3 = 1 \\ \frac{1}{3} \cdot (M_r + 1) & \text{if } M_r \bmod 3 = 2. \end{cases} \tag{7.53}$$

It is worth highlighting that L_r^{opt} is independent of L_p and M_p for $p \neq r$, which is due to the separability of the array. Inserting L_r^{opt} from (7.52) and (7.53) into expression (7.49), we obtain $\mathrm{MSE}_{\mathrm{SS}_{\mathrm{opt}}}^{(r)} = \mathrm{MSE}_{\mathrm{SS}}^{(r)}(L_r^{\mathrm{opt}})$ as

$$\mathrm{MSE}_{\mathrm{SS}_{\mathrm{opt}}}^{(r)} \approx \begin{cases} \frac{1}{\hat{\rho}} \cdot \frac{27}{4} \cdot \frac{a}{M_r^3} & \text{if } M_r \bmod 3 = 0 \\ \frac{1}{\hat{\rho}} \cdot \frac{27}{4} \cdot \frac{a}{\left(M_r + \frac{1}{2}\right)^2 (M_r - 1)} & \text{if } M_r \bmod 3 = 1 \\ \frac{1}{\hat{\rho}} \cdot \frac{27}{4} \cdot \frac{a}{\left(M_r - \frac{1}{2}\right)^2 (M_r + 1)} & \text{if } M_r \bmod 3 = 2, \end{cases} \tag{7.54}$$

where $a = \prod_{\substack{p=1 \\ p \neq r}}^{R} \frac{c_p}{M_{\mathrm{sub}_p}^2 L_p^2}$. It is clear that the MSE for a fixed $\hat{\rho}$ is lowest when M_r is a multiple of 3. Again, for $R = 1$, these results are in line with those derived in [RH90, RH93, HS90] for harmonic retrieval.

[3] As (7.49) is symmetric with respect to $L_r = M_r/2$, we obtain two values for L_r^{opt} that both minimize the MSE and are equally valid.

7.5.3. Asymptotic spatial smoothing gain

Based on the result for L_r^{opt}, the maximum asymptotic gain obtained from spatial smoothing can be explicitly quantified. To this end, we contrast $\text{MSE}_{\text{SS}}^{(r)}(L_r^{\text{opt}})$ from above with the result $\text{MSE}^{(r)} = \frac{1}{\hat{\rho}} \cdot \frac{M_r}{M(M_r-1)^2}$ from Section 4.5 and Section 6.6 without spatial smoothing. The maximum asymptotic spatial smoothing gain in the r-th mode defined as $\eta_{\text{SS}}^{(r)}(L_r^{\text{opt}}) = \text{MSE}^{(r)}/\text{MSE}_{\text{SS}}^{(r)}(L_r^{\text{opt}})$ can be computed as

$$\eta_{\text{SS}}^{(r)}(L_r^{\text{opt}}) \approx \begin{cases} \frac{4}{27} \cdot \frac{M_r^4}{(M_r-1)^2} \cdot \frac{1}{Ma} & \text{if } M_r \bmod 3 = 0 \\ \frac{4}{27} \cdot \frac{M_r(M_r+\frac{1}{2})^2}{(M_r-1)} \cdot \frac{1}{Ma} & \text{if } M_r \bmod 3 = 1 \\ \frac{4}{27} \cdot \frac{M_r(M_r-\frac{1}{2})^2(M_r+1)}{(M_r-1)^2 Ma} & \text{if } M_r \bmod 3 = 2. \end{cases} \tag{7.55}$$

7.5.4. Asymptotic efficiency of 1-D ESPRIT-type algorithms with spatial smoothing

Furthermore, the optimal value for L_r^{opt} from Subsection 7.5.2 allows to analytically compute the asymptotic efficiency of the considered R-D ESPRIT-type and R-D NC ESPRIT-type algorithms with spatial smoothing for a single source. To this end, we utilize the simplified single source expressions of the deterministic R-D Cramér-Rao bound (CRB) in [RH12] and the R-D NC CRB in Section 9.4.1 of Chapter 9, respectively. As both expressions are the same, we here only state the conventional case from [RH12].

For the case of an M-element R-D uniform sampling grid with an M_r-element ULA in the r-th mode and a single source ($d = 1$), the deterministic R-D CRB can be simplified to [RH12]

$$\boldsymbol{C} = \text{diag}\left\{\left[C^{(1)}, \ldots, C^{(R)}\right]^{\text{T}}\right\}, \tag{7.56}$$

where

$$C^{(r)} = \frac{1}{\hat{\rho}} \cdot \frac{6}{M(M_r^2 - 1)}. \tag{7.57}$$

This result is also shown in (9.39). Using (7.49) and (7.56), the asymptotic efficiency

$$\eta^{(r)}(L_r^{\text{opt}}) = \lim_{\hat{\rho}\to\infty} \frac{C^{(r)}}{\text{MSE}_{\text{SS}}^{(r)}(L_r^{\text{opt}})} \tag{7.58}$$

of the spatially smoothed versions of R-D Standard and R-D Unitary ESPRIT as well as R-D NC Standard and R-D NC Unitary ESPRIT can be computed in closed-form for arbitrary dimensions

R. As an example, the asymptotic efficiency $\eta(L^{\text{opt}})$ for $R = 1$ is given by

$$\eta(L^{\text{opt}}) \approx \begin{cases} \frac{8}{9} \cdot \frac{M^2}{M^2-1} & \text{if } M \bmod 3 = 0 \\ \frac{8}{9} \cdot \frac{(M+\frac{1}{2})^2}{M(M+1)} & \text{if } M \bmod 3 = 1 \\ \frac{8}{9} \cdot \frac{(M-\frac{1}{2})^2}{M(M-1)} & \text{if } M \bmod 3 = 2. \end{cases} \tag{7.59}$$

It should be noted that η is only a function of the array geometry, i.e., the number of sensors M. Moreover, it is straightforward to see that the asymptotic efficiency is larger when M is a multiple of 3. As one of the main results from (7.59), we observe that $\lim_{M\to\infty} \eta(L^{\text{opt}}) = 8/9$ for 1-D ESPRIT-type and 1-D NC ESPRIT-type algorithms with spatial smoothing. In contrast, it was shown in [RH12, SRHD14] and in Section 4.5.1 that their counterparts without spatial smoothing become less efficient for increasing M, i.e., for $M \to \infty$, we have $\eta \to 0$. Consequently, spatial smoothing provides a significant gain for large M.

7.6. Simulation results

In this section, we present two sets of simulation results to assess the behavior of the derived performance analysis expressions of the R-D ESPRIT-type and R-D NC ESPRIT-type algorithms based on spatial smoothing and to illustrate the analytical expressions for the single source case.

7.6.1. Performance analysis

We first compare the square root of the analytical MSE expressions ("ana") in (7.32), (7.38), and (7.41) to the empirical ("emp") root mean square errors (RMSE)s of the spatially smoothed (SpSm) versions of R-D Standard ESPRIT (SE SpSm), R-D Unitary ESPRIT (UE SpSm) as well R-D NC Standard ESPRIT (NC SE SpSm) and R-D NC Unitary ESPRIT (NC UE SpSm). For all ESPRIT-type algorithms, LS is used to solve the shift invariance equations. We also include the deterministic Cramér-Rao bounds for arbitrary signals (Det CRB) and strictly SO non-circular sources (Det NC CRB) derived in Section 9.2. The total RMSE is defined as

$$\text{RMSE} = \sqrt{\mathbb{E}\left\{\sum_{r=1}^{R}\sum_{i=1}^{d}\left(\mu_i^{(r)} - \hat{\mu}_i^{(r)}\right)^2\right\}}, \tag{7.60}$$

where $\hat{\mu}_i^{(r)}$ is the estimate of i-th spatial frequency in the r-th mode. It is assumed that a known number of signals with unit power impinge on uniform R-D array structures consisting of isotropic sensor elements with $\lambda/2$-interelement spacing in all dimensions. The phase reference is located at the array centroid. The symbols $\boldsymbol{S}_0$ are drawn from a real-valued Gaussian distribution and we assume zero-mean circularly symmetric white Gaussian noise. The curves are averaged over 5000

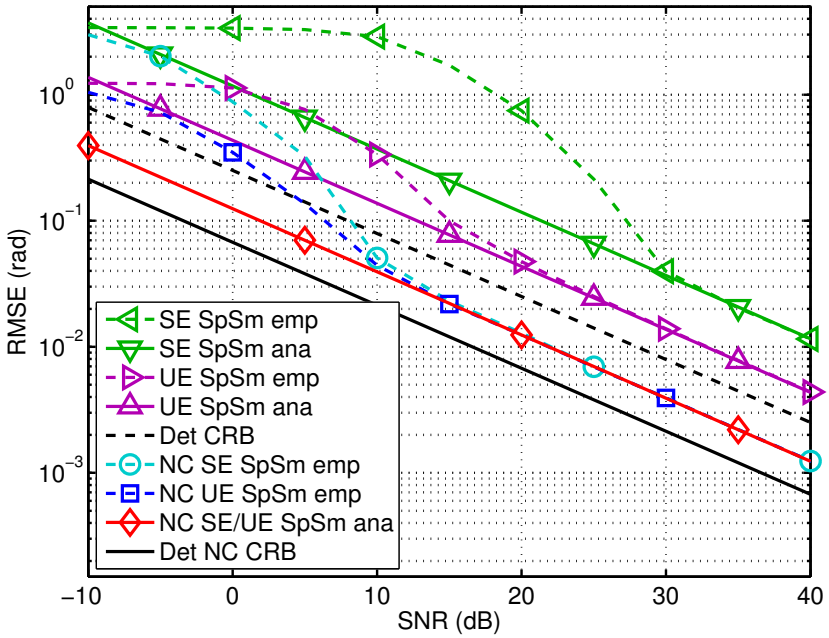


Figure 7.1.: RMSE versus SNR for a $6\times6\times6$ uniform cubic array ($R = 3$) and $N = 5$, $d = 2$ correlated sources ($\varrho = 0.9$) at $\mu_1^{(1)} = 0$, $\mu_2^{(1)} = 0.05$, $\mu_1^{(2)} = 0$, $\mu_2^{(2)} = 0.05$, $\mu_1^{(3)} = 0$, $\mu_2^{(3)} = 0.05$ with $\varphi_1 = 0$, $\varphi_2 = \pi/2$.

Monte Carlo trials.

In Fig. 7.1, we depict the total RMSE versus the SNR of $d = 2$ sources impinging on a $6 \times 6 \times 6$ uniform cubic array ($R = 3$) with $N = 5$. The sources are located at $\mu_1^{(1)} = 0$, $\mu_2^{(1)} = 0.05$, $\mu_1^{(2)} = 0$, $\mu_2^{(2)} = 0.05$, $\mu_1^{(3)} = 0$, and $\mu_2^{(3)} = 0.05$. They have a pair-wise correlation of $\varrho = 0.9$ and their rotation phases contained in $\mathbf{\Psi}$ are given by $\varphi_1 = 0$ and $\varphi_2 = \pi/2$. For L_r, we choose $L_r^{\text{opt}} = M_r/3 = 2$ in each mode, i.e., we have divided the array into a total of $L = 8$ subarrays. Fig. 7.2 investigates the total RMSE versus the number of snapshots N for a 6×6 uniform rectangular array (URA) ($R = 2$), where the SNR is 20 dB and $L_r = L_r^{\text{opt}} = M_r/3 = 2$. We have $d = 3$ uncorrelated ($\varrho = 0$) sources at $\mu_1^{(1)} = 0.25$, $\mu_2^{(1)} = 0.5$, $\mu_3^{(1)} = 0.75$, $\mu_1^{(2)} = 0.25$, $\mu_2^{(2)} = 0.5$, and $\mu_3^{(2)} = 0.75$. The rotation phases are given by $\varphi_1 = 0$, $\varphi_2 = \pi/4$, and $\varphi_3 = \pi/2$.

It is apparent from Fig. 7.1 and Fig. 7.2 that the analytical results agree well with the empirical results for high effective SNRs, i.e., either high SNRs or a large sample size. Furthermore, NC SE SpSm and NC UE SpSm provide the lowest estimation errors and perform asymptotically identical in the high effective SNR regime. However, NC UE SpSm should be preferred due to its lower complexity and its better performance at low SNRs.

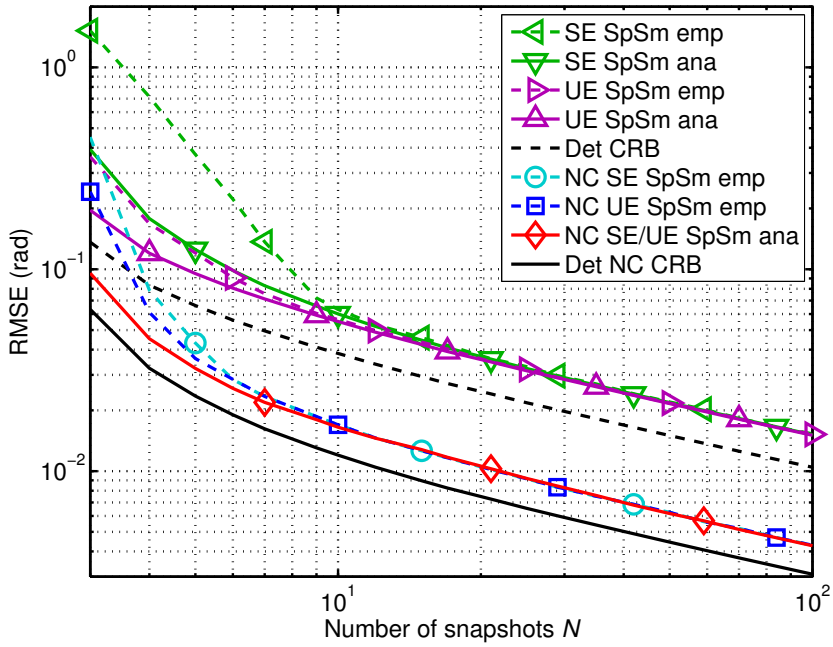


Figure 7.2.: RMSE versus N for a 6×6 URA ($R = 2$) and SNR = 20 dB, $d = 3$ uncorrelated sources ($\varrho = 0$) at $\mu_1^{(1)} = 0.25$, $\mu_2^{(1)} = 0.5$, $\mu_3^{(1)} = 0.75$, $\mu_1^{(2)} = 0.25$, $\mu_2^{(2)} = 0.5$, $\mu_3^{(2)} = 0.75$ with $\varphi_1 = 0$, $\varphi_2 = \pi/4$, $\varphi_3 = \pi/2$.

7.6.2. Analytical results for a single source

In this subsection, the derived analytical results ("ana") in (7.54) and (7.59) for a single source ($d = 1$) are compared to their empirical versions. We also include the analytical and empirical single source results from Section 4.5 and Section 6.6 without spatial smoothing. The source is located at $\mu^{(1)} = 0.1$ and $\mu^{(2)} = 0.5$, however, its location has no impact on the MSE. The effective SNR is $\rho = 46$ dB with $P = 1$, $N = 4$, and $\sigma_{\mathrm{n}}^2 = 10^{-4}$.

Fig. 7.3 illustrates the total RMSE using (7.54) as a function of the number of sensors $M_1 = M_2$ for a 2-D $M_1 \times M_2$ URA. We observe that the spatial smoothing based ESPRIT-type algorithms perform considerably closer to the CRB compared to the algorithms without spatial smoothing. Interestingly, there is no spatial smoothing gain up to $M_1 = M_2 = 4$ (a 4×4 URA). Then, as $M_1 = M_2$ grows, the spatial smoothing gain increases.

Fig. 7.4 presents the asymptotic efficiency (7.59) for $R = 1$ versus M of a ULA. The asymptotic efficiency for the non-spatial smoothing case, i.e., $L = 1$, is given by $\eta(L = 1) = \frac{6(M-1)}{M(M+1)}$. It is clear from Fig. 7.4 that all the algorithms are asymptotically efficient for $M = 2$ and $M = 3$. As M increases further, the efficiency of the algorithms with spatial smoothing approaches the value 8/9, while that of the non-spatial smoothing based algorithms becomes increasingly inefficient. Moreover, Fig. 7.4 confirms the observation from (7.59) that $\eta(L^{\mathrm{opt}})$ is slightly higher for values of M that are multiples of 3.

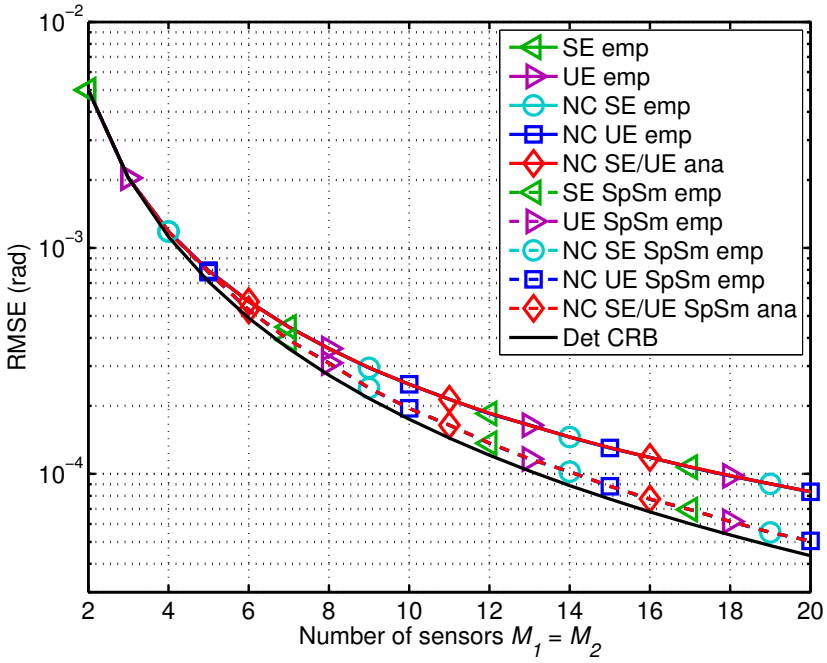


Figure 7.3.: RMSE versus $M_1 = M_2$ of a $M_1 \times M_2$ URA ($R = 2$) for a single source ($d = 1$) at $\mu^{(1)} = 0.1$, $\mu^{(2)} = 0.5$, and $\rho = 46$ dB ($P = 1$, $N = 4$, $\sigma_n^2 = 10^{-4}$).

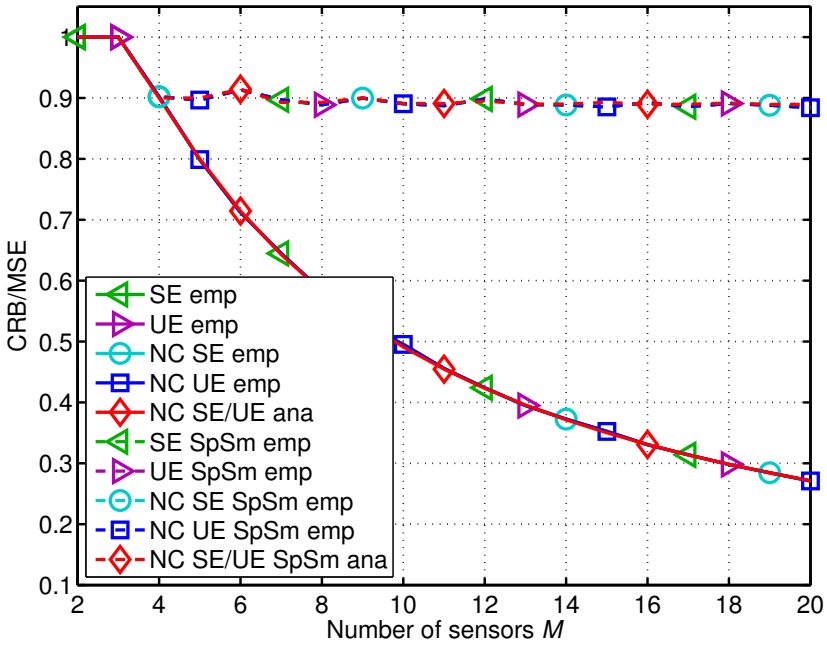


Figure 7.4.: Asymptotic efficiency versus M of a ULA ($R = 1$) for a single source ($d = 1$) at $\mu = 0$ and $\rho = 46$ dB ($P = 1$, $N = 4$, $\sigma_n^2 = 10^{-4}$).

7.7. Summary

In this chapter, we have presented a first-order performance analysis of the spatially smoothed versions of R-D Standard ESPRIT and R-D Unitary ESPRIT for arbitrary sources as well as R-D NC Standard ESPRIT and R-D NC Unitary ESPRIT for strictly SO non-circular sources. The derived expressions are asymptotic in the effective SNR and no assumptions on the noise statistics are required apart from a zero-mean and finite SO moments. We show that both spatially smoothed R-D NC ESPRIT-type algorithms perform asymptotically identical in the high effective SNR regime. As the performance generally depends on the number of subarrays, we have simplified the derived R-D MSE expressions for the special case of a single source, which allows to analytically compute the optimal number of subarrays for spatial smoothing by minimizing the MSE with respect to the number of subarrays. We have derived the optimal number of subarrays in the r-th mode to be $M_r/3$. Additionally, we have derived the asymptotic spatial smoothing gain and calculated the asymptotic efficiency for the single source case. The analytical results have been verified by numerical simulations.

8. Sparsity-based parameter estimation for strictly non-circular sources

In this chapter, we address the multi-dimensional harmonic retrieval problem by means of the relatively new concept of sparse signal reconstruction (SSR), which is based on the compressed sensing framework. According to this representation, the array response is modeled as the superposition of few wavefronts in an overcomplete basis, i.e., the received signal power is sparse in the angular domain. It has been observed that SSR-based parameter estimation algorithms exhibit super-resolution capabilities under challenging conditions [MCW05], e.g., a small number of snapshots, coherent sources, etc. In this chapter, we combine the advantages of SSR-based algorithms with the benefits achieved from exploiting the signals' strict non-circularity via sparse recovery and develop three NC SSR-based algorithms. In Section 8.1, we provide a literature overview of sparse signal recovery and its application to multi-dimensional harmonic retrieval, and outline our contributions. Section 8.2 introduces the main concept of SSR-based parameter estimation from multiple measurement vectors (MMV) and its further extensions. The main contribution of developing the SSR-based algorithms for strictly non-circular sources is given in Section 8.3. Numerical results as well as a summary are provided in Section 8.4 and Section 8.5.

8.1. Overview

Sparse signal reconstruction (SSR), which is also referred to as sparse recovery or compressed sensing, is a relatively new research area in signal processing. Due its applicability in a wide range of applications such as spectral analysis, image processing, and parameter estimation, it has attracted a considerable research interest in recent years. An overview of sparse signal reconstruction and their many signal processing applications is provided in [CW08, EK12]. The classical SSR problem is characterized by finding the solution to an underdetermined system of linear equations, which represents the reconstruction of a high-dimensional sparse signal vector in an overcomplete basis from a low-dimensional measurement vector [CW08]. By exploiting the prior knowledge on the sparsity of the signal vector, a unique solution to the underdetermined system of equations can be found. In the context of signal processing, this implies that perfect signal construction can be achieved by means of sub-Nyquist sampling [CW08, ME11], i.e., with fewer samples than required by the Nyquist sampling theorem. However, the reconstruction of the sparsest signal vector requires solving an ℓ_0-norm minimization problem, where the ℓ_0-norm of a vector denotes its cardinality, i.e., the number of its non-zero elements. Hence, the ℓ_0-norm minimization problem constitutes a combinatorial problem, which is NP-hard for large problem instances. Consequently,

a number of algorithms have been developed to find approximate solutions at a reasonable computational complexity. Examples include greedy algorithms such as matching pursuit [MZ93] or orthogonal matching pursuit [PRK93], which are fast but lack reconstruction guarantees and the convex ℓ_1-norm relaxation [CDS98], which requires a higher computational complexity but enables a better reconstruction performance. One of the key applications of sparse signal reconstruction is parameter estimation. In Section 8.1.1, we provide a literature review of SSR-based algorithms for parameter estimation and outline our contributions in Section 8.1.2.

8.1.1. State of the art

In recent years, a new perspective on the classical parameter estimation problem has been provided by employing sparse signal recovery (SSR) [Don92, Don06, CW08]. In this context, the array output is interpreted by means of a sparse power spectrum in the spatial domain, which represents the superposition of the received signal power from very few wavefronts in an overcomplete basis. Such a finite basis is obtained by sampling the continuous spatial domain with a predefined grid. The first SSR results have mainly considered the sparse signal reconstruction from a single measurement vector (SMV). However, in array processing, it is more common to consider the signal reconstruction from multiple measurement vectors (MMV), i.e., multiple snapshots. In stationary environments, the sparse representation even exhibits additional structure in form of the joint sparsity of the source signals vectors. Based on SSR, many sparsity-based DOA estimation algorithms [MCW05, HM10, SBL11, MZ06] have been developed. It has been observed that compared to the classical subspace-based parameter estimation algorithms [KV96, HPRE14], these sparsity-based algorithms can provide performance benefits and exhibit super-resolution in challenging scenarios such as a high source correlation, a low sample size or an unknown model order. These scenarios arise, for instance, in multipath environments, fast-changing tracking applications, and space-time adaptive processing applications. Despite the various benefits of the SSR-based parameter estimation algorithms, one common problem they all face is the required sampling of the continuous angular domain with a predefined grid in order to construct the overcomplete basis. As a consequence, the true DOAs mostly lie off the discretized grid, which results in a performance degradation due to the model mismatch [HS10, CSPC11]. Solutions to the off-grid problem include an adaptive refinement of the grid [MCW05], statistical modeling and fitting of the mismatch error [YXZ13], and a low-complexity analytical solution by explicitly estimating the grid offset [IRA+14].

A very recent approach to circumvent the off-grid problem is the super-resolution framework developed in [CFG14] and [TBSR13]. This framework provides a new perspective of SSR-based parameter estimation by proposing a gridless sparse recovery procedure based on atomic norm minimization (ANM) for the SMV case. Thereby, the sparse spatial line spectrum can be recovered in the continuous parameter domain with infinite precision in the noiseless case [CFG14, TBSR13]. Specifically, solving the ANM-equivalent semi-definite programming (SDP) problem provides a

Hermitian Toeplitz structured solution matrix, which admits a unique Vandermonde decomposition such that the set of DOAs can be uniquely recovered from the solution via conventional parameter estimation algorithms [KV96]. An extension to the MMV case is given in [YX16b] and a multi-dimensional extension of the ANM framework for multi-dimensional parameter estimation is provided in [CC15, YXS16]. In the latter, it is shown that the multi-dimensional ANM problem produces a multi-level Hermitian Toeplitz matrix, which still admits a unique Vandermonde decomposition involving a Kronecker structure. Therefore, the multi-dimensional spatial frequencies can be uniquely recovered from the obtained multi-level Hermitian Toeplitz matrix. A major drawback of the ANM approach is that it suffers from the so-called resolution limit, i.e., the spatial frequencies can only be recovered if they are sufficiently separated [CFG14, TBSR13].

In some of the applications involving parameter estimation, the signals possess specific statistical properties, i.e., a strictly non-circular (NC) signal structure [SS10]. NC signals result from real-valued modulation schemes such as BPSK, PAM, ASK, or Offset-QPSK (after a derotation) and are introduced in Section 2.2. Previous work has shown that the performance of the conventional DOA estimation algorithms [KV96] can be further improved if the NC structure of the received signals is exploited. Several subspace-based parameter estimation algorithms [AD06, CWS01, ZCW03, HR04, SRHD14] for NC signals have been developed. Matrix-based and tensor-based versions of R-D NC ESPRIT-type algorithms are presented in Chapter 6. The concept of exploiting the non-circularity property has recently been introduced for sparsity-based DOA estimation [LHZ12], [YLZ15]. While [LHZ12] proposes a sparse covariance matrix representation of the SO statistics of the non-circular data, in [YLZ15], the authors adopt a strategy, which relies on a sparsity-based fitting of the NC subspaces. However, both algorithms require a rather complex setting of the sparsity-inducing parameters depending on the scenario, are limited to the case of uncorrelated sources, and do not deal with the critical off-grid problem.

8.1.2. Contributions

In this chapter, we develop three different sparse recovery optimization frameworks [SRH16c, SSPH16, SRS+16] for exploiting the signals' strict non-circularity property introduced in Section 2.2, where the received complex symbols result from real-valued constellations rotated by an arbitrary phase φ. As the rotation phase φ is usually unknown, the estimation problem becomes a two-dimensional (2-D) sparse recovery problem, which requires estimating the support in the spatial domain as well as in the rotation phase domain.

In the first contribution [SRH16c], we introduce a combined 2-D finite dictionary for both dimensions and solve the resulting 2-D sparse recovery problem by $\ell_{2,1}$-mixed norm relaxation using MMV. Thereby, the known benefits associated with strictly non-circular (NC) sources (cf. Chapter 6), e.g., an improved estimation accuracy and a doubling of the number of resolvable signals, can also be achieved via sparse recovery. In order to handle the resulting 2-D off-grid problem, we

propose a 2-D off-grid estimation procedure based on [IRA+14] by means of local interpolation.

The second contribution [SSPH16] addresses the high computational complexity required for solving the 2-D mixed-norm problem in [SRH16c] as a result of sampling both dimensions, which significantly increases the problem size. Thus in [SSPH16], we propose a sparse optimization framework based on nuclear norm (rank) minimization after lifting the original optimization problem to a semidefinite programming (SDP) problem in a higher-dimensional space. To this end, the 2-D estimation problem is reduced to a 1-D estimation problem only in the sampled spatial domain, which automatically provides grid-less estimates of the rotation phases. As a result, the proposed method requires a significantly lower computational complexity while providing the same performance benefits. Additionally, we present a 1-D off-grid estimator for the spatial domain using the concept of [IRA+14].

In the third contribution [SRS+16], we present a grid-less sparse recovery algorithm for NC signals based on atomic norm minimization (ANM). After the NC preprocessing step, the ANM-equivalent SDP problem provides a solution matrix with a two-level Hermitian Toeplitz structure. We show that by using the multi-dimensional generalization of the Vandermonde decomposition, the desired direction estimates can be uniquely extracted from the two-level Hermitian Toeplitz matrix via NC ESPRIT-type algorithms (cf. Chapter 6) in closed-form. Analogously to the previous contributions, the proposed NC ANM procedure attains a superior estimation accuracy over the original methods for arbitrary signals and can resolve more sources than the number of sensors in the array.

8.2. Sparsity-based parameter estimation

In this section, we review the reconstruction of arbitrary sparse signals (without NC structure) using multiple measurement vectors in the context of parameter estimation from the literature. We consider the problem of 1-D DOA estimation, i.e., estimating the spatial frequencies corresponding to the azimuth angles of the impinging signals (cf. Example 2.1.1). We start the review with grid-based SSR algorithms in Section 8.2.1 and then introduce the gridless framework based on atomic norm minimization (ANM) in Section 8.2.2. For both techniques, we find a sparse representation of the data model in (2.15), which shows their applicability to parameter estimation.

8.2.1. Grid-based parameter estimation using sparse representation

Consider the data model in (2.15) for 1-D DOA estimation from Example 2.1.1. We suppose that d narrow-band signals from stationary sources in the far field are captured by an M-element uniform linear array (ULA). The noise-corrupted array output at N subsequent snapshots can be collected

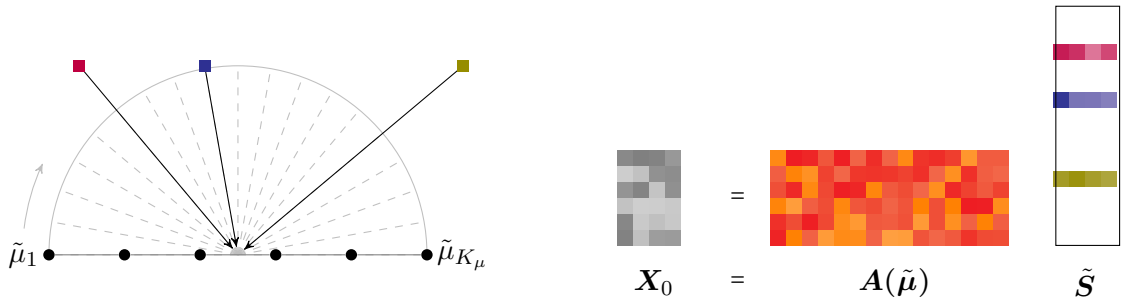


Figure 8.1.: Sparse representation of the noise-free measurement matrix $\boldsymbol{X}_0$.

in the measurement matrix

$$\boldsymbol{X} = \boldsymbol{A}(\boldsymbol{\mu}) \cdot \boldsymbol{S} + \boldsymbol{N} = \boldsymbol{X}_0 + \boldsymbol{N} \in \mathbb{C}^{M \times N}, \tag{8.1}$$

where $\boldsymbol{A}(\boldsymbol{\mu}) = [\boldsymbol{a}(\mu_1), \ldots, \boldsymbol{a}(\mu_d)] \in \mathbb{C}^{M \times d}$ contains the array steering vectors corresponding to the spatial frequencies $\boldsymbol{\mu} = [\mu_1, \ldots, \mu_d]^{\mathrm{T}}$, $\boldsymbol{S} \in \mathbb{C}^{d \times N}$ represents the symbols, and $\boldsymbol{N} \in \mathbb{C}^{M \times N}$ consists of additive zero-mean white complex Gaussian sensor noise samples with variance σ_{n}^2.

Considering the noise-free measurement matrix $\boldsymbol{X}_0 = \boldsymbol{A}(\boldsymbol{\mu}) \cdot \boldsymbol{S}$ in (8.1) from the compressed sensing perspective, $\boldsymbol{X}_0$ can be interpreted as a sparse spatial amplitude spectrum, where the received signal amplitudes concentrate at the locations of the d DOAs as visualized on the left hand side of Figure 8.1. Mathematically speaking, $\boldsymbol{X}_0$ is a set of multiple measurement vectors (MMV) that is d-row-sparse in an overcomplete basis obtained by discretizing the array manifold. Thus, an equivalent sparse signal representation of $\boldsymbol{X}_0$ in (8.1) is given by

$$\boldsymbol{X}_0 = \tilde{\boldsymbol{A}}(\tilde{\boldsymbol{\mu}}) \cdot \tilde{\boldsymbol{S}} \in \mathbb{C}^{M \times N}, \tag{8.2}$$

where $\tilde{\boldsymbol{A}}(\tilde{\boldsymbol{\mu}}) \in \mathbb{C}^{M \times K_\mu}$ contains the sampling of the spatial frequency range $[0, 2\pi]$ at the K_μ grid points $\tilde{\boldsymbol{\mu}} = [\tilde{\mu}_1, \ldots, \tilde{\mu}_{K_\mu}]^{\mathrm{T}}$. Typically, $K_\mu = MP_\mu$, where $P_\mu > 1$ is the oversampling factor such that $K_\mu > M > d$. For simplicity, we consider uniform sampling with $\tilde{\mu}_{n_\mu} = (n_\mu - 1)\Delta_\mu,\ n_\mu = 1, \ldots, K_\mu$, where $\Delta_\mu = 2\pi/K_\mu$ is the grid spacing. We assume that the grid is sufficiently fine such that the true spatial frequencies $\boldsymbol{\mu}$ are contained in the grid $\tilde{\boldsymbol{\mu}}$. This assumption is referred to as the on-grid assumption. In practice, the on-grid assumption is usually not fulfilled, which leads to a model mismatch [HS10, CSPC11]. Solutions to this off-grid problem are given in [MCW05, YXZ13, IRA+14] and considered in Section 8.3.1 and in Section 8.3.2. For now, we assume that the on-grid assumption is fulfilled. The matrix $\tilde{\boldsymbol{S}} \in \mathbb{C}^{K_\mu \times N}$ in (8.2) is the sparse symbol matrix of interest,

which contains the elements

$$[\tilde{\boldsymbol{S}}]_{k,n} = \begin{cases} [\boldsymbol{S}]_{i,n} & \text{if } \tilde{\mu}_k = \mu_i \\ \boldsymbol{0} & \text{else} \end{cases} \tag{8.3}$$

for $k = 1, \ldots, K_\mu$, $n = 1, \ldots, N$, and $i = 1, \ldots, d$. Hence, $\tilde{\boldsymbol{S}}$ is a row-sparse matrix, i.e., its columns share the same support as shown on the right hand side of Figure 8.1. The support of the non-zero rows of $\tilde{\boldsymbol{S}}$ then corresponds to the locations of the DOAs on the spatial grid.

In the literature, the problem of recovering $\tilde{\boldsymbol{S}}$ from $\boldsymbol{X}_0$ in (8.2) is termed linear inverse problem [Don92, Tib96, CT05, CRT06, Don06]. The underdetermined linear system of equations in (8.2) has infinitely many solutions. However, by exploiting the sparsity of $\tilde{\boldsymbol{S}}$, a unique solution can be obtained [CT05, Don06]. This approach is referred to as sparse signal reconstruction (SSR).

Considering the noisy measurement matrix $\boldsymbol{X}$, the joint sparse signal reconstruction problem can be formulated as the $\ell_{p,0}$-mixed-norm problem

$$\begin{aligned} \min_{\tilde{\boldsymbol{S}} \in \mathbb{C}^{K_\mu \times N}} \quad & \left\| \tilde{\boldsymbol{S}} \right\|_{p,0} \\ \text{s. t.} \quad & \left\| \boldsymbol{X} - \tilde{\boldsymbol{A}}(\tilde{\boldsymbol{\mu}}) \cdot \tilde{\boldsymbol{S}} \right\|_{\mathrm{F}}^2 \leq \beta, \end{aligned} \tag{8.4}$$

where β is a threshold parameter in relation to the noise power σ_{n}^2, which is usually selected according to [MCW05] as $\beta = \mathrm{Tr}\{\mathbb{E}\{\boldsymbol{N}\boldsymbol{N}^{\mathrm{H}}\}\} = MN\sigma_{\mathrm{n}}^2$. The $\ell_{p,0}$-norm is defined as the number of its non-zero rows $\boldsymbol{x}_k^{\mathrm{T}}$ according to

$$\|\boldsymbol{X}\|_{p,0} = \{k \mid \|\boldsymbol{x}_k\|_p \neq 0\} \tag{8.5}$$

for any ℓ_p-norm (cf. Appendix C.2.4). Due to the $\ell_{p,0}$-norm, the SSR problem in (8.4) requires combinatorial optimization and is therefore NP-hard. As a result, a number of algorithms that solve (8.4) approximately have been proposed. These include greedy methods such as simultaneous orthogonal matching pursuit (S-OMP) [TGS06] and convex relaxation to the $\ell_{p,1}$ mixed-norm minimization [MCW05, Kow09, HM10].

The $\ell_{p,1}$-mixed-norm is a common approach [MCW05, Kow09, HM10] to approximate the $\ell_{p,0}$-mixed-norm problem in (8.4). As described in Appendix C.2.4, the definition for general $\ell_{p,1}$ mixed-norms is given by $\|\boldsymbol{X}\|_{p,1} = \sum_{k=1}^{K} \|\boldsymbol{x}_k\|_p$. In practice, the $\ell_{2,1}$-mixed-norm is often used and defined by $\|\boldsymbol{X}\|_{2,1} = \sum_{k=1}^{K} \|\boldsymbol{x}_k\|_2$. Then, the convex $\ell_{2,1}$-mixed-norm relaxation of the MMV problem in (8.4) can be formulated as

$$\begin{aligned} \min_{\tilde{\boldsymbol{S}} \in \mathbb{C}^{K_\mu \times N}} \quad & \left\| \tilde{\boldsymbol{S}} \right\|_{2,1} \\ \text{s. t.} \quad & \left\| \boldsymbol{X} - \tilde{\boldsymbol{A}}(\tilde{\boldsymbol{\mu}}) \cdot \tilde{\boldsymbol{S}} \right\|_{\mathrm{F}}^2 \leq \beta, \end{aligned} \tag{8.6}$$

where β is typically chosen as discussed for problem (8.4). Problem (8.6) can be solved by any ℓ_1-type algorithm, e.g., the basis pursuit denoising (BPDN) algorithm [CDS98].

Given a solution $\tilde{\boldsymbol{S}}$ to the problem (8.6), the support set is then identified by the indices of the non-zero rows according to $\mathcal{S} = \{k \,\big|\, [\tilde{\boldsymbol{S}}]_{k,n} \neq \boldsymbol{0}\}$, which then corresponds to the locations of spatial frequencies on the spatial grid as a $\{\hat{\mu}_i\}_{i=1}^{\hat{d}} = \{\tilde{\mu}_k \,\big|\, k \in \mathcal{S}\}$.

The computational complexity of solving the mixed-norm minimization problem in (8.6) is mainly determined by the size of the sparse signal matrix $\tilde{\boldsymbol{S}}$, i.e., the number of grid points K_μ and the number of snapshots N. It is usually desirable to simultaneously have a large K_μ in order to achieve a high spatial resolution and a large sample size N to obtain more accurate estimates. This, however, significantly increases the computational complexity of solving (8.6). In order to reduce the computational cost associated with a large number of grid points K_μ, the work in [MCW05] proposes a heuristic approach of an iterative refinement of the grid around the estimated spatial frequencies. Moreover, [MCW05] also suggests to reduce the effective number of snapshots by operating on the signal subspace of the measurement matrix $\boldsymbol{X}$ rather than on $\boldsymbol{X}$ itself. This approach is referred to as the ℓ_1-SVD method.

Let the SVD of $\boldsymbol{X}$ be given by $\boldsymbol{X} = \boldsymbol{U} \cdot \boldsymbol{\Sigma} \cdot \boldsymbol{V}^{\mathrm{H}}$. Then, after estimating the model order d [WK85], we perform the following preprocessing step and replace $\boldsymbol{X}$ by

$$\begin{aligned} \boldsymbol{X}_{\mathrm{sv}} &= \boldsymbol{X} \cdot \boldsymbol{V} \cdot \boldsymbol{K} \\ &= \tilde{\boldsymbol{A}}(\tilde{\boldsymbol{\mu}}) \cdot \tilde{\boldsymbol{S}} \cdot \boldsymbol{V} \cdot \boldsymbol{K} + \boldsymbol{N} \cdot \boldsymbol{V} \cdot \boldsymbol{K} \\ &= \tilde{\boldsymbol{A}}(\tilde{\boldsymbol{\mu}}) \cdot \tilde{\boldsymbol{S}}_{\mathrm{sv}} + \boldsymbol{N}_{\mathrm{sv}} \in \mathbb{C}^{K_\mu \times r}, \end{aligned} \tag{8.7}$$

where $\tilde{\boldsymbol{S}}_{\mathrm{sv}} = \tilde{\boldsymbol{S}} \cdot \boldsymbol{V} \cdot \boldsymbol{K}$, $\boldsymbol{N}_{\mathrm{sv}} = \boldsymbol{N} \cdot \boldsymbol{V} \cdot \boldsymbol{K}$, and $\boldsymbol{K} = [\boldsymbol{I}_r,\ \boldsymbol{0}_{r\times(N-r)}]^{\mathrm{T}} \in \mathbb{R}^{N\times r}$ with $r = \min\{d, N, M\}$. Based on this model with a reduced dimensionality, the ℓ_1-SVD method solves the $\ell_{2,1}$-mixed-norm problem

$$\begin{aligned} \min_{\tilde{\boldsymbol{S}}_{\mathrm{sv}} \in \mathbb{C}^{K_\mu \times r}} \quad & \left\| \tilde{\boldsymbol{S}}_{\mathrm{sv}} \right\|_{2,1} \\ \text{s. t.} \quad & \left\| \boldsymbol{X}_{\mathrm{sv}} - \tilde{\boldsymbol{A}}(\tilde{\boldsymbol{\mu}}) \cdot \tilde{\boldsymbol{S}}_{\mathrm{sv}} \right\|_{\mathrm{F}}^2 \leq \beta_{\mathrm{sv}}, \end{aligned} \tag{8.8}$$

where the number of optimization parameters in $\tilde{\boldsymbol{S}}_{\mathrm{sv}} \in \mathbb{C}^{K_\mu \times r}$ has been reduced. A drawback of the ℓ_1-SVD method is that the selection of β_{sv} is more difficult [MCW05] as $\boldsymbol{N}_{\mathrm{sv}}$ depends on $\boldsymbol{V}$. An alternative method to reduce the number of snapshots N is proposed in [YLSX17].

8.2.2. Gridless parameter estimation using sparse representation

In the previous section, we have reasoned that a larger number of grid points is required to obtain a high spatial frequency resolution, which increases the computational complexity. Moreover, the

critical off-grid problem arises if the true spatial frequencies do not lie on the grid. Due to these reasons, it is highly desirable to develop gridless sparse recovery algorithms. Recently, two gridless sparse signal reconstruction (SSR) frameworks, namely, the total variation norm minimization framework [CFG13, CFG14] and the atomic norm minimization framework [CRPW12, BTR13, TBSR13, TBR15] have been proposed. Their equivalence has been shown in [TBSR13].

The concept of the atomic norm minimization (ANM) was introduced in [CRPW12] as a generalization of several norms commonly used for sparse recovery, such as the ℓ_1-norm minimization for sparse vector recovery or the nuclear norm minimization for low-rank matrix recovery. The advantage of the ANM framework is that it exploits the sparsity in the continuous parameter domain by means of the atomic norm metric. Thereby, the discretization of the spatial domain is not required such that the problem of choosing the grid spacing as well as the off-grid problem are completely avoided. A detailed discussion of the ANM concept and how to calculate it is provided in Appendix C.4. In [TBSR13, BTR13], the ANM framework was introduced for gridless line spectral estimation in the single measurement vector (SMV) case and under uniform sampling. In this section, we review the ANM framework derived in [YX15, LC16, YX16b] for the multiple measurement vector (MMV) case.

In the ANM framework [YX15, LC16, YX16b], the noise-free measurement matrix $\boldsymbol{X}_0$ in (8.1) is modeled as a linear combination of the atoms

$$\tilde{\boldsymbol{X}}_0(\tilde{\mu}, \tilde{\boldsymbol{s}}) = \boldsymbol{a}(\tilde{\mu}) \cdot \tilde{\boldsymbol{s}}^{\mathrm{T}}, \tag{8.9}$$

where $\tilde{\mu} \in [-\pi, \pi)$ and $\tilde{\boldsymbol{s}} \in \mathbb{C}^{N\times 1}$ with $\|\tilde{\boldsymbol{s}}\|_2 = 1$. Then, the continuous dictionary $\mathcal{A}$, also termed atomic set, is given by

$$\mathcal{A} = \left\{ \tilde{\boldsymbol{X}}_0(\tilde{\mu}, \tilde{\boldsymbol{s}}) \mid \tilde{\mu} \in [-\pi, \pi),\ \|\tilde{\boldsymbol{s}}\|_2 = 1 \right\}. \tag{8.10}$$

According to [YX15, LC16, YX16b], the atomic ℓ_0-norm of $\boldsymbol{X}_0$ is defined as

$$\left\|\boldsymbol{X}_0\right\|_{\mathcal{A},0} = \inf_{\{\tilde{\mu}_k, \tilde{\boldsymbol{s}}_k\}} \left\{ K \;\middle|\; \boldsymbol{X}_0 = \sum_{k=1}^{K} c_k \cdot \tilde{\boldsymbol{X}}_0(\tilde{\mu}_k, \tilde{\boldsymbol{s}}_k),\ c_k \geq 0 \right\} \tag{8.11}$$

and describes the smallest number of atoms to compose $\boldsymbol{X}_0$. A natural objective to obtain $\boldsymbol{X}_0$ is to minimize $\|\boldsymbol{X}_0\|_{\mathcal{A},0}$, i.e., to find the atomic decomposition of $\boldsymbol{X}_0$ with the minimal number of atoms while still being in line with the noise-corrupted measurement matrix $\boldsymbol{X}$. As minimizing the atomic ℓ_0-norm in (8.11) is non-convex and NP-hard, we instead consider the convex ℓ_1-relaxation (cf. Appendix C.4) of the atomic ℓ_0-norm, also referred to as the atomic ℓ_1-norm, which is defined

as

$$\|\boldsymbol{X}_0\|_{\mathcal{A}} = \inf_{\{\tilde{\mu}_k, \tilde{\boldsymbol{s}}_k\}} \left\{ \sum_k c_k \;\middle|\; \boldsymbol{X}_0 = \sum_k c_k \cdot \tilde{\boldsymbol{X}}_0(\tilde{\mu}_k, \tilde{\boldsymbol{s}}_k),\ c_k \geq 0 \right\}. \tag{8.12}$$

In the noiseless case and assuming a ULA, the atomic ℓ_1-norm minimization problem admits the following computationally efficient semi-definite programming (SDP) formulation [YX15, LC16, YX16b]

$$\begin{aligned} \min_{\boldsymbol{W},\boldsymbol{u}} \quad & \frac{1}{2} \cdot \mathrm{Tr}\{\boldsymbol{W}\} + \frac{1}{2M} \cdot \mathrm{Tr}\{\mathrm{Toep}\{\boldsymbol{u}\}\} \\ \text{s. t.} \quad & \begin{bmatrix} \boldsymbol{W} & \boldsymbol{X}_0^{\mathrm{H}} \\ \boldsymbol{X}_0 & \mathrm{Toep}\{\boldsymbol{u}\} \end{bmatrix} \succeq \boldsymbol{0}, \end{aligned} \tag{8.13}$$

where $\mathrm{Toep}\{\boldsymbol{u}\} \in \mathbb{C}^{M \times M}$ with $\boldsymbol{u} = [u_1, \ldots, u_M]^{\mathrm{T}} \in \mathbb{C}^{M \times 1}$ denotes the Hermitian Toeplitz matrix

$$\mathrm{Toep}\{\boldsymbol{u}\} = \begin{bmatrix} u_1 & u_2 & \cdots & u_M \\ u_2^* & u_1 & \cdots & u_{M-1} \\ \vdots & \vdots & \ddots & \vdots \\ u_M^* & u_{M-1}^* & \cdots & u_1 \end{bmatrix}. \tag{8.14}$$

A proof for the equivalence of the problems in (8.12) and in (8.13) for the SMV case is provided in Appendix C.4. The proof for the MMV case follows along the same lines.

Given the solution $\hat{\boldsymbol{u}}$ to the problem (8.13), the desired spatial frequencies μ_k and the magnitudes c_k for $k = 1, \ldots, d$ can be retrieved by means of the Vandermonde decomposition as shown in [TBSR13, YXS16, YLSX17]. The Vandermonde decomposition states the following result [TBSR13]:

Definition 8.2.1. In the noiseless case and for a ULA with M sensors, any positive semi-definite Hermitian Toeplitz matrix $\mathrm{Toep}\{\boldsymbol{u}\} \in \mathbb{C}^{M \times M}$ with $\mathrm{rank}\{\mathrm{Toep}\{\boldsymbol{u}\}\} = K \leq (M-1)$ can be decomposed according to

$$\mathrm{Toep}\{\boldsymbol{u}\} = \boldsymbol{A}(\tilde{\boldsymbol{\mu}}) \cdot \boldsymbol{C} \cdot \boldsymbol{A}^{\mathrm{H}}(\tilde{\boldsymbol{\mu}}) = \sum_{k=1}^{K} c_k \cdot \boldsymbol{a}(\tilde{\mu}_k) \cdot \boldsymbol{a}^{\mathrm{H}}(\tilde{\mu}_k), \tag{8.15}$$

where $\boldsymbol{A}(\tilde{\boldsymbol{\mu}}) = [\boldsymbol{a}(\tilde{\mu}_1), \ldots, \boldsymbol{a}(\tilde{\mu}_K)] \in \mathbb{C}^{M \times K}$ has a Vandermonde structure corresponding to the ULA and $\boldsymbol{C} = \mathrm{diag}\{[c_1, \ldots, c_K]\} \in \mathbb{R}^{K \times K}$ is a diagonal matrix that contains the non-negative coefficients $c_k > 0,\ k = 1, \ldots, K$, on its diagonal. By the Carathéodory theorem [Car11, CF11], the Vandermonde decomposition in (8.15) is unique for $\mathrm{rank}\{\mathrm{Toep}\{\boldsymbol{u}\}\} = K \leq (M-1)$.

The Vandermonde decomposition in (8.15) can be computed by estimating the spatial frequencies

$\tilde{\mu}_k$ from the Hermitian Toeplitz matrix $\text{Toep}\{\hat{\boldsymbol{u}}\}$ via Prony's method [dP95] or subspace-based parameter estimation algorithms such as ESPRIT-type algorithms introduced in Chapter 3. The corresponding coefficients $\boldsymbol{c} = [c_1, \ldots, c_K]^{\mathrm{T}} \in \mathbb{R}^{K \times 1}$ can be retrieved by solving the linear system of equations $\boldsymbol{A}(\tilde{\boldsymbol{\mu}}) \cdot \boldsymbol{c} = \boldsymbol{u}^*$. It should be emphasized that the reconstruction of $\tilde{\mu}_k$ and c_k is performed in a gridless fashion as the discretization of the spatial domain is not required.

In case of the noisy measurement data $\boldsymbol{X}$, the atomic ℓ_1-norm minimization problem can be formulated as

$$\begin{aligned} \min_{\boldsymbol{W},\boldsymbol{u}} \quad & \frac{1}{2} \cdot \text{Tr}\{\boldsymbol{W}\} + \frac{1}{2M} \cdot \text{Tr}\{\text{Toep}\{\boldsymbol{u}\}\} \\ \text{s. t.} \quad & \begin{bmatrix} \boldsymbol{W} & \boldsymbol{X}_0^{\mathrm{H}} \\ \boldsymbol{X}_0 & \text{Toep}\{\boldsymbol{u}\} \end{bmatrix} \succeq \boldsymbol{0}, \ \left\| \boldsymbol{X} - \boldsymbol{X}_0 \right\|_{\mathrm{F}}^2 \leq \beta. \end{aligned} \tag{8.16}$$

where the threshold parameter β is again chosen according to the noise statistics, e.g., $\beta = \mathrm{E}\{\|\boldsymbol{N}\|_{\mathrm{F}}^2\}$. In the noisy case, the Vandermonde decomposition in (8.15) only holds approximately. However, given a solution $\hat{\boldsymbol{u}}$ to (8.16), the spatial frequencies $\tilde{\mu}_k$ can still be extracted from $\text{Toep}\{\hat{\boldsymbol{u}}\}$ via, for instance, the ESPRIT-type algorithms presented in Chapter 3.

The computational complexity of solving (8.16) depends on the number of sensors M and the number of snapshots N. In order to reduce the computational complexity, the same dimensionality reduction techniques described in Section 8.2.1 for the $\ell_{2,1}$-mixed-norm minimization can be applied [YX16a]. Moreover, a very efficient implementation of the SDP in (8.16), which is based on the alternating direction method of multipliers (ADMM), has been proposed in [BTR13] for the SMV case and in [LC16] for the MMV case.

Note that a major drawback of the ANM algorithm is that it suffers from the Rayleigh resolution limit $2\pi/M$, i.e., the spatial frequencies can only be recovered if they are sufficiently separated [CFG14, TBSR13]. Moreover, due to the required Toeplitz structure, the ANM framework is only applicable to ULAs or ULAs with missing sensors.

8.3. Sparsity-based parameter estimation for strictly non-circular sources

In this section, we develop three different optimization algorithms published in [SRH16c, SSPH16, SRS+16] for exploiting the signals' strictly non-circular (NC) signal structure via sparse signal recovery. The concept and the properties of strictly non-circular sources are introduced in Section 2.2. We have seen for the R-D NC ESPRIT-type algorithms developed in Chapter 6 that exploiting the NC signal structure improves the estimation accuracy and doubles the number of resolvable sources. In this section, we show that these benefits associated with exploiting NC sources can also be achieved via sparse signal reconstruction (SSR) in the multiple measurement

vector (MMV) case. In Section 8.3.1 and in Section 8.3.2, we develop two grid-based SSR algorithms for NC sources based on $\ell_{2,1}$-mixed-norm minimization and nuclear norm minimization, respectively. To avoid the off-grid problem, we present a low-complexity grid offset estimation procedure for both algorithms for a single source and two closely-spaced sources using local interpolation. In Section 8.3.3, we propose a gridless SSR algorithm for NC sources based on the atomic norm minimization (ANM) framework.

8.3.1. Grid-based parameter estimation based on NC sparse recovery

We start the development with the model for strictly second-order (SO) non-circular (NC) sources from Section 2.2, which is restated again for convenience. In the case of strictly SO non-circular sources, the symbol matrix $\boldsymbol{S} \in \mathbb{C}^{d\times N}$ in (8.1) can be decomposed as (cf. Equation (2.36) in Section 2.2)

$$\boldsymbol{S} = \boldsymbol{\Psi}(\boldsymbol{\varphi}) \cdot \boldsymbol{S}_0, \tag{8.17}$$

where $\boldsymbol{S}_0 = [\boldsymbol{s}_{0,1}, \ldots, \boldsymbol{s}_{0,d}]^{\mathrm{T}} \in \mathbb{R}^{d\times N}$ is a real-valued symbol matrix with rows $\boldsymbol{s}_{0,i}^{\mathrm{T}}$ for $i = 1, \ldots, d$, and $\boldsymbol{\Psi}(\boldsymbol{\varphi}) = \operatorname{diag}\left\{[\mathrm{e}^{\mathrm{j}\varphi_1}, \ldots, \mathrm{e}^{\mathrm{j}\varphi_d}]\right\} \in \mathbb{C}^{d\times d}$ is a diagonal matrix that contains the rotation phase shifts corresponding to the phases $\boldsymbol{\varphi} = [\varphi_1, \ldots, \varphi_d]^{\mathrm{T}}$ on its diagonal, which can be arbitrary for each received signal as described in Section 2.2. Thus, the complex signal amplitudes of each source lie on a rotated line through the origin in the complex plane. Note that at the receiver, the rotation phase $\boldsymbol{\varphi}$ are usually unknown.

In Section 6.2.1, we have shown for subspace-based parameter estimation algorithms that the strictly non-circular signal structure can be exploited by applying the NC preprocessing step in (6.4) to the measurement matrix $\boldsymbol{X}$ in (8.1) to obtain the augmented measurement matrix $\boldsymbol{X}^{(\mathrm{nc})} \in \mathbb{C}^{2M\times N}$

$$\boldsymbol{X}^{(\mathrm{nc})} = \begin{bmatrix} \boldsymbol{X} \\ \boldsymbol{\Pi}_M \cdot \boldsymbol{X}^* \end{bmatrix} = \begin{bmatrix} \boldsymbol{A}(\boldsymbol{\mu}) \cdot \boldsymbol{\Psi}(\boldsymbol{\varphi}) \\ \boldsymbol{\Pi}_M \cdot \boldsymbol{A}^*(\boldsymbol{\mu}) \cdot \boldsymbol{\Psi}^*(\boldsymbol{\varphi}) \end{bmatrix} \cdot \boldsymbol{S}_0 + \begin{bmatrix} \boldsymbol{N} \\ \boldsymbol{\Pi}_M \cdot \boldsymbol{N}^* \end{bmatrix} \tag{8.18}$$

$$= \boldsymbol{A}^{(\mathrm{nc})}(\boldsymbol{\mu}, \boldsymbol{\varphi}) \cdot \boldsymbol{S}_0 + \boldsymbol{N}^{(\mathrm{nc})} = \boldsymbol{X}_0^{(\mathrm{nc})} + \boldsymbol{N}^{(\mathrm{nc})}, \tag{8.19}$$

where $\boldsymbol{A}^{(\mathrm{nc})}(\boldsymbol{\mu}, \boldsymbol{\varphi}) \in \mathbb{C}^{2M\times d}$ is the augmented array steering matrix with a virtually doubled number of sensors and $\boldsymbol{X}_0^{(\mathrm{nc})} \in \mathbb{C}^{2M\times N}$ is the noise-free augmented measurement matrix. It has been shown in Chapter 6 that processing $\boldsymbol{X}^{(\mathrm{nc})}$ instead of $\boldsymbol{X}$ reduces the estimation error and doubles the number of resolvable sources [SRHD14]. Note that (8.19) is a slight variation of the original model in (6.4) as it is expressed in terms of the real-valued symbol matrix $\boldsymbol{S}_0$ instead of $\boldsymbol{S}$.

For the derivation of grid-based sparse signal recovery algorithms for NC sources, we first find an equivalent sparse signal representation of the noise-free augmented measurement matrix $\boldsymbol{X}_0^{(\mathrm{nc})}$ from

(8.19) similar to that in (8.2). However, since the augmented array steering matrix in $\boldsymbol{A}^{(\mathrm{nc})}(\boldsymbol{\mu},\boldsymbol{\varphi})$ in (8.19) depends on both the spatial frequency parameters $\boldsymbol{\mu}$ and the unknown rotation phases $\boldsymbol{\varphi}$, we need to discritize the spatial domain as well as the rotation phase domain, resulting in a 2-D grid. Thus, a sparse representation of $\boldsymbol{X}_0^{(\mathrm{nc})}$ accounting for the NC structure is given by

$$\boldsymbol{X}_0^{(\mathrm{nc})} = \tilde{\boldsymbol{A}}^{(\mathrm{nc})}(\tilde{\boldsymbol{\mu}},\tilde{\boldsymbol{\varphi}}) \cdot \tilde{\boldsymbol{S}}_0, \tag{8.20}$$

where $\tilde{\boldsymbol{A}}^{(\mathrm{nc})}(\tilde{\boldsymbol{\mu}},\tilde{\boldsymbol{\varphi}}) \in \mathbb{C}^{2M\times K_\mu K_\varphi}$ is the overcomplete sensing matrix and $\tilde{\boldsymbol{S}}_0 \in \mathbb{R}^{K_\mu K_\varphi \times N}$ is the corresponding real-valued row-sparse matrix. The 2-D grid embedded in the sensing matrix $\tilde{\boldsymbol{A}}^{(\mathrm{nc})}(\tilde{\boldsymbol{\mu}},\tilde{\boldsymbol{\varphi}})$ is defined by the $K_\mu K_\varphi$ tuples $(\tilde{\mu}_{n_\mu},\tilde{\varphi}_{n_\varphi})$ with $n_\mu = 1,\dots,K_\mu$ and $n_\varphi = 1,\dots,K_\varphi$. Thus, similarly to (8.2), the spatial frequency range $[0,2\pi]$ is sampled at the $K_\mu = MP_\mu$ grid points $\tilde{\boldsymbol{\mu}} = [\tilde{\mu}_1,\dots,\tilde{\mu}_{K_\mu}]^{\mathrm{T}}$ with the uniform grid $\tilde{\mu}_{n_\mu} = (n_\mu - 1)\Delta_\mu$, where $\Delta_\mu = 2\pi/K_\mu$ is the grid spacing and $P_\mu > 1$ is the oversampling factor such that $K_\mu > M > d$. Additionally, the rotation phase range $[0,\pi]$ is sampled at $\tilde{\boldsymbol{\varphi}} = [\tilde{\varphi}_1,\dots,\tilde{\varphi}_{K_\varphi}]^{\mathrm{T}}$ with the uniform grid $\tilde{\varphi}_{n_\varphi} = (n_\varphi - 1)\Delta_\varphi$ with $n_\varphi = 1,\dots,K_\varphi$, where $\Delta_\varphi = \pi/K_\varphi$ and $K_\varphi = MP_\varphi$ with the oversampling factor $P_\varphi > 1$ such that $K_\varphi > M > d$. Then, the overcomplete sensing matrix $\tilde{\boldsymbol{A}}^{(\mathrm{nc})}(\tilde{\boldsymbol{\mu}},\tilde{\boldsymbol{\varphi}}) \in \mathbb{C}^{2M\times K_\mu K_\varphi}$ is defined as

$$\tilde{\boldsymbol{A}}^{(\mathrm{nc})}(\tilde{\boldsymbol{\mu}},\tilde{\boldsymbol{\varphi}}) = \left[\tilde{\boldsymbol{A}}^{(\mathrm{nc})}(\tilde{\boldsymbol{\mu}},\tilde{\varphi}_1) \quad \cdots \quad \tilde{\boldsymbol{A}}^{(\mathrm{nc})}(\tilde{\boldsymbol{\mu}},\tilde{\varphi}_{N_\varphi})\right], \tag{8.21}$$

where

$$\tilde{\boldsymbol{A}}^{(\mathrm{nc})}(\tilde{\boldsymbol{\mu}},\tilde{\varphi}_{n_\varphi}) = \begin{bmatrix} \boldsymbol{A}(\tilde{\boldsymbol{\mu}}) \cdot \mathrm{e}^{\mathrm{j}\tilde{\varphi}_{n_\varphi}} \\ \boldsymbol{\Pi}_M \cdot \boldsymbol{A}^*(\tilde{\boldsymbol{\mu}}) \cdot \mathrm{e}^{-\mathrm{j}\tilde{\varphi}_{n_\varphi}} \end{bmatrix} \in \mathbb{C}^{2M\times K_\mu}. \tag{8.22}$$

Therefore, the effective $N_\mu N_\varphi$-point sampling grid is given by the points $k = (n_\varphi - 1)K_\mu + n_\mu$. Note that the extended row dimensions of the dictionary $\tilde{\boldsymbol{A}}^{(\mathrm{nc})}(\tilde{\boldsymbol{\mu}},\tilde{\boldsymbol{\varphi}})$ can be interpreted as a virtual doubling of the number of sensor elements. Moreover, if the phase reference of the ULA is at the array centroid, i.e., $\boldsymbol{\Pi}_M \cdot \tilde{\boldsymbol{A}}^*(\tilde{\boldsymbol{\mu}}) = \tilde{\boldsymbol{A}}(\tilde{\boldsymbol{\mu}})$ holds, the 2-D sensing matrix $\tilde{\boldsymbol{A}}^{(\mathrm{nc})}(\tilde{\boldsymbol{\mu}},\tilde{\boldsymbol{\varphi}})$ in (8.21) can be compactly expressed as

$$\tilde{\boldsymbol{A}}^{(\mathrm{nc})}(\tilde{\boldsymbol{\mu}},\tilde{\boldsymbol{\varphi}}) = \boldsymbol{\Phi}(\tilde{\boldsymbol{\varphi}}) \otimes \tilde{\boldsymbol{A}}(\tilde{\boldsymbol{\mu}}), \tag{8.23}$$

where $\boldsymbol{\Phi}(\tilde{\boldsymbol{\varphi}}) = [\boldsymbol{\phi}(\tilde{\varphi}_1),\dots,\boldsymbol{\phi}(\tilde{\varphi}_{N_\varphi})] \in \mathbb{C}^{2\times N_\varphi}$ with $\boldsymbol{\phi}(\tilde{\varphi}_{n_\varphi}) = [\mathrm{e}^{\mathrm{j}\tilde{\varphi}_{n_\varphi}}, \mathrm{e}^{-\mathrm{j}\tilde{\varphi}_{n_\varphi}}]^{\mathrm{T}} \in \mathbb{C}^{2\times 1}$.

It is important to emphasize that due to the discretization of the spatial domain as well as the rotation phase domain, the estimation problem becomes a two-dimensional (2-D) sparse recovery problem, which requires estimating the support in the spatial domain as well as in the rotation phase domain. Assuming that the on-grid assumption discussed for (8.2) is satisfied, the real-valued

matrix $\tilde{\boldsymbol{S}}_0 \in \mathbb{C}^{K_\mu K_\varphi \times N}$ in (8.20) is has the row-sparse structure

$$[\tilde{\boldsymbol{S}}_0]_{k,n} = \begin{cases} [\boldsymbol{S}_0]_{i,n} & \text{if } \tilde{\mu}_k = \mu_i \\ \boldsymbol{0} & \text{else} \end{cases} \tag{8.24}$$

for $k = 1, \ldots, K_\mu K_\varphi$, $n = 1, \ldots, N$, and $i = 1, \ldots, d$. The support of the non-zero rows of $\tilde{\boldsymbol{S}}_0$ corresponds to the spatial frequencies on the spatial grid.

In analogy to (8.4) in Section 8.2.1, the joint sparse signal reconstruction problem can be formulated as the $\ell_{p,0}$-mixed-norm problem

$$\begin{aligned} \min_{\tilde{\boldsymbol{S}}_0 \in \mathbb{R}^{K_\mu K_\varphi \times N}} \quad & \left\| \tilde{\boldsymbol{S}}_0 \right\|_{p,0} \\ \text{s. t.} \quad & \left\| \boldsymbol{X}^{(\mathrm{nc})} - \tilde{\boldsymbol{A}}^{(\mathrm{nc})}(\tilde{\boldsymbol{\mu}}, \tilde{\boldsymbol{\varphi}}) \cdot \tilde{\boldsymbol{S}}_0 \right\|_{\mathrm{F}}^2 \leq \beta^{(\mathrm{nc})}, \end{aligned} \tag{8.25}$$

where the threshold parameter $\beta^{(\mathrm{nc})}$ related to the noise power is chosen according to [MCW05] as $\beta^{(\mathrm{nc})} = \mathrm{Tr}\left\{ \mathbb{E}\{ \boldsymbol{N}^{(\mathrm{nc})} \cdot \boldsymbol{N}^{(\mathrm{nc})^{\mathrm{H}}} \} \right\} = 2MN\sigma_{\mathrm{n}}^2$. As solving the $\ell_{p,0}$-mixed-norm problem in (8.25) is NP-hard, approximate solutions via simultaneous orthogonal matching pursuit (S-OMP) [TGS06] and the convex $\ell_{p,1}$ mixed-norm relaxation [MCW05, Kow09, HM10] can be obtained.

The corresponding $\ell_{2,1}$ mixed-norm minimization problem can be formulated as

$$\begin{aligned} \min_{\tilde{\boldsymbol{S}}_0 \in \mathbb{R}^{K_\mu K_\varphi \times N}} \quad & \left\| \tilde{\boldsymbol{S}}_0 \right\|_{2,1} \\ \text{s. t.} \quad & \left\| \boldsymbol{X}^{(\mathrm{nc})} - \tilde{\boldsymbol{A}}^{(\mathrm{nc})}(\tilde{\boldsymbol{\mu}}, \tilde{\boldsymbol{\varphi}}) \cdot \tilde{\boldsymbol{S}}_0 \right\|_{\mathrm{F}}^2 \leq \beta^{(\mathrm{nc})}, \end{aligned} \tag{8.26}$$

where $\beta^{(\mathrm{nc})}$ is chosen as above. Note that due to the effective sampling grid, closely-spaced NC sources (even if they are on the same grid point) are well-separated as long as they have a rotation phase discrimination. Therefore, not only the support estimation is usually improved, but also the estimated amplitudes of the sparse components in the spectrum. The support in both dimensions can be found by matching the effective grid into the 2-D grid.

The computational complexity of solving (8.26) depends on the number of grid points K_μ and K_φ in the spatial domain and the rotation phase domain, and the number of snapshots N. To reduce the computational complexity, the same dimensionality reduction techniques described in Section 8.2.1 for the $\ell_{2,1}$-mixed-norm minimization can be applied.

Note that similarly to the NC ESPRIT-type algorithms presented in Chapter 6, at most $2(M-1)$ NC sources can be uniquely resolved by the $\ell_{2,1}$ mixed-norm minimization for NC signals. This is due to the NC preprocessing step, which virtually doubles the number of antennas. Simulation results in Section 8.4 demonstrate this property.

8.3.1.1. 2-D Off-grid Estimation

So far, we have assumed that the NC sources lie on the sampling grid such that the model in (8.20) holds. However, this assumption is rather unrealistic in practice, which leads to a 2-D off-grid problem as a result of the 2-D sampling grid introduced for NC sources. In order to handle this model mismatch, we extend the previous work on the 1-D off-grid problem for the SMV case [IRA+14] to the NC case as well as the MMV case. The results are published in [SRH16c]. We present two efficient solutions for a single NC off-grid source and two closely-spaced NC off-grid sources that are used as a post-processing step after the support estimation via SSR. We show analytically for a single off-grid source that the atoms in the 2-D grid are separable, i.e., the off-grid estimation can be performed in both dimensions separately. Based on these findings, two simple low-complexity estimators are presented that require a considerably lower computational complexity compared to the SSR.

Let us first introduce the off-grid model for the i-th source located at the pair (μ_i, φ_i), $i = 1, \ldots, d$, in both dimensions as

$$\mu_i = \tilde{\mu}_{L_{\mu_i}} + \epsilon_i(\tilde{\mu}_{L_{\mu_i}+1} - \tilde{\mu}_{L_{\mu_i}}) = (L_{\mu_i} - 1 + \epsilon_i) \cdot \Delta_\mu, \tag{8.27}$$

$$\varphi_i = \tilde{\varphi}_{L_{\varphi_i}} + \delta_i(\tilde{\varphi}_{L_{\varphi_i}+1} - \tilde{\varphi}_{L_{\varphi_i}}) = (L_{\varphi_i} - 1 + \delta_i) \cdot \Delta_\varphi, \tag{8.28}$$

where we have used the defined uniform sampling grid on the right hand side. Moreover, L_{μ_i} and L_{φ_i} are the respective nearest left grid points obtained from the support estimation, and $\epsilon_i,\ \delta_i \in [0, 1]$ model the grid offset. It should be noted that for $\epsilon_i = \delta_i = 0$, (8.27) and (8.28) reduce to the on-grid model in (8.20).

For simplicity, we assume that the phase reference of the ULA is located at the array centroid. Then, the steering vector corresponding to the i-th source is given by

$$\boldsymbol{a}(\mu_i) = \left[\mathrm{e}^{-\mathrm{j}\frac{M-1}{2}\mu_i}, \ldots, \mathrm{e}^{\mathrm{j}\frac{M-1}{2}\mu_i}\right]^{\mathrm{T}} \in \mathbb{C}^{M\times 1} \tag{8.29}$$

Under this assumption, the augmented steering vector $\boldsymbol{a}^{(\mathrm{nc})}(\mu_i, \varphi_i)$ can be expressed according to (8.23) as $\boldsymbol{a}^{(\mathrm{nc})}(\mu_i, \varphi_i) = [\mathrm{e}^{\mathrm{j}\varphi_i}, \mathrm{e}^{-\mathrm{j}\varphi_i}]^{\mathrm{T}} \otimes \boldsymbol{a}(\mu_i)$.

To gain more insights into the 2-D off-grid problem, we first consider a single NC off-grid source. In [IRA+14], it has been shown that each 1-D off-grid source can be well approximated by the atoms corresponding to the two closest grid points. As a consequence, sparse recovery algorithms concentrate the signal power at exactly those grid points. This suggests that a similar approach can be used in the 2-D case for an NC off-grid source. Thus, based on the relative height of the peaks at the two neighboring atoms in both grid dimensions, we can estimate the corresponding two offsets to approximate the NC off-grid source.

In the noiseless case, the model (8.19) for a single NC source simplifies to

$$\boldsymbol{X}_0^{(\mathrm{nc})} = \boldsymbol{a}^{(\mathrm{nc})}(\epsilon,\delta)\cdot\boldsymbol{s}_0^{\mathrm{T}}, \tag{8.30}$$

where $\boldsymbol{a}^{(\mathrm{nc})}(\epsilon,\delta) = \boldsymbol{a}^{(\mathrm{nc})}(\tilde{\mu}_L+\epsilon\Delta_\mu, \tilde{\varphi}_L+\delta\Delta_\varphi) \in \mathbb{C}^{2M\times 1}$ is the true steering vector and $\boldsymbol{s}_0 \in \mathbb{R}^{N\times 1}$ is the real-valued symbol vector. Thus, we can represent $\boldsymbol{a}^{(\mathrm{nc})}(\epsilon,\delta)$ by the linear model

$$\boldsymbol{a}^{(\mathrm{nc})}(\epsilon,\delta) \approx \tilde{\boldsymbol{A}}^{(\mathrm{nc})}(\tilde{\mu}_L,\tilde{\mu}_{L+1},\tilde{\varphi}_L,\tilde{\varphi}_{L+1})\cdot\boldsymbol{\alpha}, \tag{8.31}$$

where the matrix $\tilde{\boldsymbol{A}}^{(\mathrm{nc})}(\tilde{\mu}_L,\tilde{\mu}_{L+1},\tilde{\varphi}_L,\tilde{\varphi}_{L+1}) = [\tilde{\boldsymbol{a}}^{(\mathrm{nc})}(\tilde{\mu}_L,\tilde{\varphi}_L), \tilde{\boldsymbol{a}}^{(\mathrm{nc})}(\tilde{\mu}_{L+1},\tilde{\varphi}_L), \tilde{\boldsymbol{a}}^{(\mathrm{nc})}(\tilde{\mu}_L,\tilde{\varphi}_{L+1}), \tilde{\boldsymbol{a}}^{(\mathrm{nc})}(\tilde{\mu}_{L+1},\tilde{\varphi}_{L+1})] \in \mathbb{C}^{2M\times 4}$ contains the neighboring grid points and $\boldsymbol{\alpha} = [\alpha_1,\ldots,\alpha_4]^{\mathrm{T}} \in \mathbb{R}^{4\times 1}$ contains the coefficients. Using $bmX_0^{(\mathrm{nc})}$ in (8.30), the coefficients $\boldsymbol{\alpha}$ can be found by solving the least squares problem

$$\min_{\boldsymbol{\alpha}} \quad \left\|\boldsymbol{X}_0^{(\mathrm{nc})} - \tilde{\boldsymbol{A}}^{(\mathrm{nc})}(\tilde{\mu}_L,\tilde{\mu}_{L+1},\tilde{\varphi}_L,\tilde{\varphi}_{L+1})\cdot\boldsymbol{\alpha}\cdot\boldsymbol{s}_0^{\mathrm{T}}\right\|_{\mathrm{F}}^2, \tag{8.32}$$

where $\boldsymbol{s}_0^{\mathrm{T}}$ on both sides can be neglected. The solution is given by

$$\hat{\boldsymbol{\alpha}}(\epsilon,\delta) = \tilde{\boldsymbol{A}}^{(\mathrm{nc})^+}(\tilde{\mu}_L,\tilde{\mu}_{L+1},\tilde{\varphi}_L,\tilde{\varphi}_{L+1})\cdot\boldsymbol{a}^{(\mathrm{nc})}(\epsilon,\delta). \tag{8.33}$$

After expanding the pseudo inverse in (8.33), we obtain terms such as for instance

$$\boldsymbol{a}_{\tilde{\mu}_L}^{\mathrm{H}}\cdot\boldsymbol{a}_{\tilde{\mu}_{L+1}} = \sum_{m=-(M-1)/2}^{(M-1)/2} \mathrm{e}^{\mathrm{j}m\Delta_\mu} = D(\Delta_\mu), \tag{8.34}$$

where we have defined

$$D(y) = \begin{cases} M & \text{if } y = 0 \\ \sin(yM/2)/\sin(y/2) & \text{otherwise.} \end{cases} \tag{8.35}$$

Similar terms to (8.34) can be obtained for all the combinations of $\tilde{\mu}_L, \tilde{\mu}_{L+1}$, and $\tilde{\mu}_L+\epsilon\Delta_\mu$.

Then, denoting $\boldsymbol{d}(x) = [D(x\Delta_\mu), D((x-1)\Delta_\mu)]^{\mathrm{T}} \in \mathbb{R}^{2\times 1}$ and $\boldsymbol{c}(x) = [\cos(x\Delta_\varphi), \cos((x-1)\Delta_\varphi)]^{\mathrm{T}} \in \mathbb{R}^{2\times 1}$ as well as $\boldsymbol{D}(x) = [\boldsymbol{d}(x), \boldsymbol{d}(x+1)] \in \mathbb{R}^{2\times 2}$ and $\boldsymbol{C}(x) = [\boldsymbol{c}(x), \boldsymbol{c}(x+1)] \in \mathbb{R}^{2\times 2}$, we obtain

$$\hat{\boldsymbol{\alpha}}(\epsilon,\delta) = \big(\boldsymbol{C}(0)\otimes\boldsymbol{D}(0)\big)^{-1}\big(\boldsymbol{c}(\delta)\otimes\boldsymbol{d}(\epsilon)\big) \tag{8.36}$$

$$= \boldsymbol{\alpha}(\delta)\otimes\boldsymbol{\beta}(\epsilon), \tag{8.37}$$

where $\boldsymbol{\alpha}(\delta) = [\alpha_1(\delta), \alpha_2(\delta)]^{\mathrm{T}}$ and $\boldsymbol{\beta}(\epsilon) = [\beta_1(\epsilon), \beta_2(\epsilon)]^{\mathrm{T}}$ with

$$\alpha_1(\delta) = \frac{\cos(\delta\Delta_\varphi) - \cos(\Delta_\varphi) \cdot \cos((\delta-1)\Delta_\varphi)}{1 - \cos^2(\Delta_\varphi)} \tag{8.38}$$

$$\alpha_2(\delta) = \frac{\cos((\delta-1)\Delta_\varphi) - \cos(\Delta_\varphi) \cdot \cos(\delta\Delta_\varphi)}{1 - \cos^2(\Delta_\varphi)} \tag{8.39}$$

$$\beta_1(\epsilon) = \frac{M \cdot D(\epsilon\Delta_\mu) - D(\Delta_\mu) \cdot D((\epsilon-1)\Delta_\mu)}{M^2 - D^2(\Delta_\mu)} \tag{8.40}$$

$$\beta_1(\epsilon) = \frac{M \cdot D((\epsilon-1)\Delta_\mu) - D(\Delta_\mu) \cdot D(\epsilon\Delta_\mu)}{M^2 - D^2(\Delta_\mu)}. \tag{8.41}$$

Thus, the 2-D offset estimation is separable in both grid dimensions. As it can be shown that $\alpha_n(\delta)$ and $\beta_n(\epsilon)$, $n = 1, 2$, become linear in ϵ and δ with increasing P_μ and P_φ, the simple 1-D estimator from [IRA+14]

$$\hat{\epsilon} = \frac{\beta_2(\epsilon)}{\beta_1(\epsilon) + \beta_2(\epsilon)}, \quad \hat{\delta} = \frac{\alpha_2(\delta)}{\alpha_1(\delta) + \alpha_2(\delta)} \tag{8.42}$$

can be applied in both dimensions independently. Arranging the elements of $\hat{\boldsymbol{\alpha}}$ in a matrix $\boldsymbol{B}$, we obtain $\boldsymbol{B} = \boldsymbol{\alpha}(\delta) \cdot \boldsymbol{\beta}(\epsilon)^{\mathrm{T}}$, which is of rank 1.

In the noisy case, the matrix $\boldsymbol{X}_0^{(\mathrm{nc})}$ in (8.32) needs to be replaced by $\boldsymbol{X}^{(\mathrm{nc})}$. As a consequence, $\boldsymbol{B}$ becomes an estimate that is not rank one anymore. However, a good rank-one approximation can be obtained from the SVD of $\hat{\boldsymbol{B}}$, i.e., $\hat{\boldsymbol{B}} = \boldsymbol{u} \cdot \boldsymbol{v}^{\mathrm{H}} = \hat{\boldsymbol{\alpha}}(\delta) \cdot \hat{\boldsymbol{\beta}}^{\mathrm{T}}(\epsilon)$. Based on $\hat{\boldsymbol{\alpha}}(\delta)$ and $\hat{\boldsymbol{\beta}}(\epsilon)$, the two estimators in (8.42) are used to determine the grid offsets.

In the case of two sources, their mutual influence depends on the correlation between the array steering vectors. If the source separation is much larger than P_μ grid points, which corresponds to the Rayleigh resolution limit $2\pi/M$, they are separable. Due to the effective grid, this is the case for two closely-spaced NC sources if their phases discriminate. Therefore, the two sources can be treated independently and the estimators (8.42) can be applied for each source separately.

However, this approach fails for closely-spaced sources with the same rotation phase. Thus, inspired by [IRA+14], we propose a numerical joint off-grid estimation procedure for two NC sources with the same rotation phase, i.e., $\varphi_1 = \varphi_2$. Assuming that the correct support has been estimated via SSR, we approximate each of the two sources located at (μ_i, φ_i), $i = 1, 2$, by the four respective neighboring grid points $L_{\mu_i}, L_{\mu_i}+1, L_{\varphi_i}$, and $L_{\varphi_i}+1$ as an extension of (8.31). Denote $\tilde{\boldsymbol{A}}_i^{(\mathrm{nc})} \in \mathbb{C}^{2M\times 4}$ as the matrix containing the neighboring grid points of the i-th source as in (8.31). Then, in analogy to the noiseless case in (8.33), we have

$$\boldsymbol{G}(\boldsymbol{\epsilon}, \boldsymbol{\delta}) = \tilde{\boldsymbol{A}}_{1,2}^{(\mathrm{nc})^+} \cdot \boldsymbol{X}_0^{(\mathrm{nc})} = \tilde{\boldsymbol{A}}_{1,2}^{(\mathrm{nc})^+} \cdot \boldsymbol{A}^{(\mathrm{nc})}(\boldsymbol{\epsilon}, \boldsymbol{\delta}) \cdot \boldsymbol{S}_0, \tag{8.43}$$

where $\tilde{\boldsymbol{A}}_{1,2}^{(\mathrm{nc})} = [\tilde{\boldsymbol{A}}_1^{(\mathrm{nc})}, \tilde{\boldsymbol{A}}_2^{(\mathrm{nc})}] \in \mathbb{C}^{2M\times 8}$ and $\boldsymbol{G}(\boldsymbol{\epsilon}, \boldsymbol{\delta}) \in \mathbb{R}^{8\times T}$ is the matrix of coefficients that depends

on $\boldsymbol{\epsilon} = [\epsilon_1, \epsilon_2]^{\mathrm{T}}$ and $\boldsymbol{\delta} = [\delta_1, \delta_2]^{\mathrm{T}}$. It can be shown that $\boldsymbol{G}(\boldsymbol{\epsilon}, \boldsymbol{\delta})$ can be expressed as

$$\boldsymbol{G}(\boldsymbol{\epsilon}, \boldsymbol{\delta}) = \boldsymbol{D}_0^{-1} \cdot \boldsymbol{D}(\boldsymbol{\epsilon}, \boldsymbol{\delta}) \cdot \boldsymbol{S}_0, \tag{8.44}$$

where

$$\boldsymbol{D}_0 = \begin{bmatrix} \boldsymbol{C}(0) \otimes \boldsymbol{D}(0) & \boldsymbol{C}(d_\varphi) \otimes \boldsymbol{D}(d_\mu) \\ \boldsymbol{C}(d_\varphi)^{\mathrm{T}} \otimes \boldsymbol{D}(d_\mu)^{\mathrm{T}} & \boldsymbol{C}(0) \otimes \boldsymbol{D}(0) \end{bmatrix} \in \mathbb{R}^{8\times 8},$$

$$\boldsymbol{D}(\boldsymbol{\epsilon}, \boldsymbol{\delta}) = \begin{bmatrix} \boldsymbol{c}(\delta_1) \otimes \boldsymbol{d}(\epsilon_1) & \boldsymbol{c}(\delta_2 + d_\varphi) \otimes \boldsymbol{d}(\epsilon_2 + d_\mu) \\ \boldsymbol{c}(\delta_1 - d_\varphi) \otimes \boldsymbol{d}(\epsilon_1 - 1) & \boldsymbol{c}(\delta_2) \otimes \boldsymbol{d}(\epsilon_2) \end{bmatrix} \in \mathbb{R}^{8\times 2},$$

where $d_\mu = L_{\mu_2} - L_{\mu_1}$ and $d_\varphi = L_{\varphi_2} - L_{\varphi_1}$ and the vectors $\boldsymbol{c}(x), \boldsymbol{d}(x)$ as well as the matrices $\boldsymbol{C}(x)$ and $\boldsymbol{D}(x)$ are defined as in (8.36).

Next, we consider the noisy case, which yields $\hat{\boldsymbol{G}} = \tilde{\boldsymbol{A}}_{1,2}^{(\mathrm{nc})^+} \cdot \boldsymbol{X}^{(\mathrm{nc})}$, and define $\bar{\boldsymbol{G}} = \boldsymbol{D}_0 \cdot \hat{\boldsymbol{G}}$. For the comparison of the coefficients $\bar{\boldsymbol{G}}$ obtained from the measurements with the analytical approximation $\boldsymbol{G}(\boldsymbol{\epsilon}, \boldsymbol{\delta})$, we note that the columns of $\boldsymbol{G}(\boldsymbol{\epsilon}, \boldsymbol{\delta})$ are a linear combination of the columns of $\boldsymbol{D}(\boldsymbol{\epsilon}, \boldsymbol{\delta})$. Thus, we aim at maximizing the overlap between $\bar{\boldsymbol{G}}$ and the column space of $\boldsymbol{D}(\boldsymbol{\epsilon}, \boldsymbol{\delta})$. To this end, we propose to estimate the off-grid parameters by minimizing the cost function

$$J(\boldsymbol{\epsilon}, \boldsymbol{\delta}) = \left\| \bar{\boldsymbol{G}} - \boldsymbol{D}(\boldsymbol{\epsilon}, \boldsymbol{\delta}) \cdot \boldsymbol{D}^{+}(\boldsymbol{\epsilon}, \boldsymbol{\delta}) \cdot \bar{\boldsymbol{G}} \right\|_{\mathrm{F}}^2. \tag{8.45}$$

We have observed that the function $J(\boldsymbol{\epsilon}, \boldsymbol{\delta})$ is smooth and convex in the parameter range $\epsilon_i, \delta_i \in [0, 1]$ with a unique minimum. Therefore, (8.45) can be minimized by any local optimization method, e.g., the gradient descent algorithm. Note again that the joint estimation procedure can resolve two sources with a separation below the Rayleigh resolution limit.

It is worth mentioning that the presented joint estimation procedure can be extended straightforwardly if a group of more than two closely-spaced sources with the same rotation phase is present.

8.3.2. Sparsity-based parameter estimation for strictly non-circular sources using nuclear norm minimization

The grid-based sparse signal reconstruction (SSR) algorithm for strictly non-circular (NC) signals based on $\ell_{2,1}$-mixed-norm minimization in Section 8.3.1 requires the sampling of both the spatial domain and the rotation phase domain. To achieve high resolution capabilities, a fine sampling is desirable. However, a large number of grid points in both domains significantly increases the dimensions of the joint overcomplete basis, which can lead to a prohibitive computational complexity. Therefore in this section and published in [SSPH16], we present an SSR algorithm for NC signals based on nuclear norm (rank) minimization after lifting the original bilinear optimization

problem to a linear optimization problem in a higher-dimensional space. Thereby, the 2-D estimation problem is reduced to a 1-D estimation problem only in the sampled spatial domain, which automatically provides gridless estimates of the rotation phases. As a result, the proposed method requires a significantly lower computational complexity, while providing the same performance benefits. Additionally, we present an low-complexity grid offset estimator for the spatial domain.

As in (8.19) in Section 8.3.1, we first account for the NC signal structure by applying the NC preprocessing step to the measurement matrix $\boldsymbol{X}$ from (8.1) to obtain the augmented measurement matrix $\boldsymbol{X}^{(\mathrm{nc})} \in \mathbb{C}^{2M \times N}$

$$\boldsymbol{X}^{(\mathrm{nc})} = \begin{bmatrix} \boldsymbol{X} \\ \boldsymbol{\Pi}_M \cdot \boldsymbol{X}^* \end{bmatrix} = \begin{bmatrix} \boldsymbol{A}(\boldsymbol{\mu}) \cdot \boldsymbol{\Psi}(\boldsymbol{\varphi}) \\ \boldsymbol{\Pi}_M \cdot \boldsymbol{A}^*(\boldsymbol{\mu}) \cdot \boldsymbol{\Psi}^*(\boldsymbol{\varphi}) \end{bmatrix} \cdot \boldsymbol{S}_0 + \begin{bmatrix} \boldsymbol{N} \\ \boldsymbol{\Pi}_M \cdot \boldsymbol{N}^* \end{bmatrix} \tag{8.46}$$

$$= \boldsymbol{A}^{(\mathrm{nc})}(\boldsymbol{\mu}, \boldsymbol{\varphi}) \cdot \boldsymbol{S}_0 + \boldsymbol{N}^{(\mathrm{nc})} = \boldsymbol{X}_0^{(\mathrm{nc})} + \boldsymbol{N}^{(\mathrm{nc})}, \tag{8.47}$$

where the virtually augmented array steering matrix $\boldsymbol{A}^{(\mathrm{nc})}(\boldsymbol{\mu}, \boldsymbol{\varphi}) \in \mathbb{C}^{2M \times d}$ has twice as many sensor elements M, such that processing $\boldsymbol{X}^{(\mathrm{nc})}$ instead of $\boldsymbol{X}$ can improve the accuracy of the parameter estimates and doubles the number of resolvable signals. The augmented steering matrix $\boldsymbol{A}^{(\mathrm{nc})}(\boldsymbol{\mu}, \boldsymbol{\varphi}) \in \mathbb{C}^{2M \times d}$ in (8.47) can be decomposed as

$$\boldsymbol{A}^{(\mathrm{nc})}(\boldsymbol{\mu}, \boldsymbol{\varphi}) = \bar{\boldsymbol{A}}^{(\mathrm{nc})}(\boldsymbol{\mu}) \cdot \boldsymbol{\Phi}^{(\mathrm{nc})}(\boldsymbol{\varphi}), \tag{8.48}$$

where $\bar{\boldsymbol{A}}^{(\mathrm{nc})}(\boldsymbol{\mu})$ and $\boldsymbol{\Phi}^{(\mathrm{nc})}(\boldsymbol{\varphi})$ are defined as

$$\bar{\boldsymbol{A}}^{(\mathrm{nc})}(\boldsymbol{\mu}) = \begin{bmatrix} \boldsymbol{A}(\boldsymbol{\mu}) & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{\Pi}_M \cdot \boldsymbol{A}^*(\boldsymbol{\mu}) \end{bmatrix}, \quad \boldsymbol{\Psi}^{(\mathrm{nc})}(\boldsymbol{\varphi}) = \begin{bmatrix} \boldsymbol{\Psi}(\boldsymbol{\varphi}) \\ \boldsymbol{\Psi}^*(\boldsymbol{\varphi}) \end{bmatrix}.$$

Then, following the idea in [SPP14, SPP18a], the model in (8.47) can be reformulated as

$$\boldsymbol{X}^{(\mathrm{nc})} = \boldsymbol{B}(\boldsymbol{\mu}) \cdot \boldsymbol{\Phi}(\boldsymbol{\varphi}) \cdot \boldsymbol{S}_0 + \boldsymbol{N}^{(\mathrm{nc})} \tag{8.49}$$

where $\boldsymbol{B}(\boldsymbol{\mu}) = \bar{\boldsymbol{A}}^{(\mathrm{nc})}(\boldsymbol{\mu}) \cdot \boldsymbol{J} \in \mathbb{C}^{2M \times 2d}$ is a column permuted version of the steering matrix $\bar{\boldsymbol{A}}^{(\mathrm{nc})}(\boldsymbol{\mu})$ with a permutation matrix $\boldsymbol{J} \in \mathbb{R}^{2d \times 2d}$ such that

$$\boldsymbol{B}(\boldsymbol{\mu}) = \begin{bmatrix} \boldsymbol{a}(\mu_1) & \boldsymbol{0} & \cdots & \boldsymbol{a}(\mu_d) & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{\Pi}_M \cdot \boldsymbol{a}^*(\mu_1) & & \boldsymbol{0} & \boldsymbol{\Pi}_M \cdot \boldsymbol{a}^*(\mu_d) \end{bmatrix}. \tag{8.50}$$

Similarly, we define the row permuted phase shift matrix $\boldsymbol{\Phi}(\boldsymbol{\varphi}) = \boldsymbol{J}^{\mathrm{T}} \cdot \boldsymbol{\Psi}^{(\mathrm{nc})}(\boldsymbol{\varphi}) \in \mathbb{C}^{2d \times d}$ that possesses the block-diagonal structure

$$\boldsymbol{\Phi}(\boldsymbol{\varphi}) = \mathrm{blkdiag}\left\{[\phi_1, \ldots, \phi_d]\right\}, \tag{8.51}$$

where the vector $\boldsymbol{\phi}_i = \left[\mathrm{e}^{\mathrm{j}\varphi_i},\ \mathrm{e}^{-\mathrm{j}\varphi_i}\right]^{\mathrm{T}} \in \mathbb{C}^{2\times 1}$ contains the rotation phase shift and its complex conjugate corresponding to the i-th source as defined in the model (8.17) in Section 8.3.1. Furthermore, we define the parametrization

$$\boldsymbol{Q}(\boldsymbol{\varphi}) = \boldsymbol{\Psi}(\boldsymbol{\varphi}) \cdot \boldsymbol{S}_0 = \begin{bmatrix} \boldsymbol{Q}_1^{\mathrm{T}}(\varphi_1) & \cdots & \boldsymbol{Q}_d^{\mathrm{T}}(\varphi_d) \end{bmatrix}^{\mathrm{T}} \in \mathbb{C}^{2d\times N}, \tag{8.52}$$

which is partitioned into d rank-one matrices

$$\boldsymbol{Q}_i(\varphi_i) = \boldsymbol{\phi}_i \cdot \boldsymbol{s}_{0,i}^{\mathrm{T}} \in \mathbb{C}^{2\times N} \tag{8.53}$$

for $i = 1, \ldots, d$. Additionally, due to the complex conjugate structure of the phase shift vectors $\boldsymbol{\phi}_i$ and the fact that $\boldsymbol{s}_{0,i}^{\mathrm{T}}$ is real-valued, each sub-matrix $\boldsymbol{Q}_i(\varphi_i)$ exhibits a conjugate row structure, i.e., $\boldsymbol{Q}_i(\varphi_i) = \boldsymbol{\Pi}_2 \cdot \boldsymbol{Q}_i^*(\varphi_i)$. Using the definition in (8.52), the signal models in (8.47) and (8.49) can be equivalently expressed as

$$\boldsymbol{X}^{(\mathrm{nc})} = \boldsymbol{B}(\boldsymbol{\mu}) \cdot \boldsymbol{Q}(\boldsymbol{\varphi}) + \boldsymbol{N}^{(\mathrm{nc})}. \tag{8.54}$$

Note that the parametrization $\boldsymbol{Q}(\boldsymbol{\varphi})$ in (8.54) contains $2Nd$ complex entries compared to the $(N+1)d$ real-valued entries when directly modeling $\boldsymbol{S}_0$ and $\boldsymbol{\varphi}$ in (8.47). Hence, the parametrization $\boldsymbol{Q}(\boldsymbol{\varphi})$ contains more redundancy and increases the number of estimation parameters. However, the advantage of the representation (8.54) is that, in contrast to (8.47), the bilinear nature is removed as the rotation phase matrix $\boldsymbol{\Psi}(\boldsymbol{\varphi})$ is embedded in the block signal matrix $\boldsymbol{Q}(\boldsymbol{\varphi})$. Therefore, in the context of optimization, the parameterization in (8.52) can be interpreted as lifting to a higher-dimensional space. Moreover, $\boldsymbol{Q}(\boldsymbol{\varphi})$ is composed of rank-one sub-matrices with conjugate row structure. Both properties of the sub-matrices will be exploited by the proposed SSR algorithm for NC signals.

The sparse representation of the model in (8.54) is given by

$$\boldsymbol{X}^{(\mathrm{nc})} = \boldsymbol{B}(\tilde{\boldsymbol{\mu}}) \cdot \tilde{\boldsymbol{Q}}(\boldsymbol{\varphi}) + \boldsymbol{N}^{(\mathrm{nc})}, \tag{8.55}$$

where the overcomplete sensing matrix $\boldsymbol{B}(\tilde{\boldsymbol{\mu}}) \in \mathbb{C}^{2M\times 2K_\mu}$ is computed according to (8.50) and constructed by discretizing the spatial frequency range $[0, 2\pi]$ at the K_μ grid points $\tilde{\boldsymbol{\mu}} = [\tilde{\mu}_1, \ldots, \tilde{\mu}_{K_\mu}]$. Typically, we have $K_\mu = MP_\mu$, where $P_\mu > 1$ is the oversampling factor such that $K_\mu > M > d$. For simplicity, we consider uniform sampling with $\tilde{\mu}_k = (k-1)\Delta_\mu$ with $k = 1, \ldots, K_\mu$, where $\Delta_\mu = 2\pi/K_\mu$ is the grid spacing, and assume that the true spatial frequencies lie exactly on the sampling grid, i.e., $\{\mu_i\}_{i=1}^{d} \in \{\tilde{\mu}_k\}_{k=1}^{K_\mu}$. Under the on-grid assumption, the block-sparse signal matrix

$\tilde{\boldsymbol{Q}}(\boldsymbol{\varphi}) = [\tilde{\boldsymbol{Q}}_1^{\mathrm{T}}(\varphi_1), \ldots, \tilde{\boldsymbol{Q}}_{K_\mu}^{\mathrm{T}}(\varphi_{K_\mu})]^{\mathrm{T}} \in \mathbb{C}^{2K_\mu \times N}$ contains the sub-matrices

$$\tilde{\boldsymbol{Q}}_k(\varphi_k) = \begin{cases} \boldsymbol{Q}_i(\varphi_i) & \text{if } \tilde{\mu}_k = \mu_i \\ \boldsymbol{0} & \text{else.} \end{cases} \tag{8.56}$$

It should be highlighted that the matrix $\tilde{\boldsymbol{Q}}(\boldsymbol{\varphi})$ in (8.55) exhibits two different levels of sparsity. Considering (8.56), it is block sparse, i.e., the elements of the blocks $\tilde{\boldsymbol{Q}}_k(\varphi_k)$, for $k = 1, \ldots, K_\mu$, are either jointly zero or jointly non-zero. Additionally, from (8.52), the blocks $\tilde{\boldsymbol{Q}}_k(\varphi_k)$ are rank-sparse, i.e., they are either of rank one or rank zero. In what follows, we drop the argument $\boldsymbol{\varphi}$ in $\tilde{\boldsymbol{Q}}(\boldsymbol{\varphi})$ for notational convenience. The support of the non-zero blocks of $\tilde{\boldsymbol{Q}}(\boldsymbol{\varphi})$ corresponds to the spatial frequencies on the spatial grid.

In case of the noisy measurement matrix $\boldsymbol{X}^{(\mathrm{nc})}$, the sparse reconstruction problem can be formulated by the rank minimization problem

$$\min_{\tilde{\boldsymbol{Q}}} \sum_{k=1}^{K_\mu} \operatorname{rank}\left\{\tilde{\boldsymbol{Q}}_k\right\} \tag{8.57a}$$

$$\text{s.t. } \left\|\boldsymbol{X}^{(\mathrm{nc})} - \boldsymbol{B}(\tilde{\boldsymbol{\mu}}) \cdot \tilde{\boldsymbol{Q}}\right\|_{\mathrm{F}}^2 \leq \beta^{(\mathrm{nc})} \tag{8.57b}$$

$$\tilde{\boldsymbol{Q}}_k = \boldsymbol{\Pi}_2 \cdot \tilde{\boldsymbol{Q}}_k^*, \text{ for } k = 1, \ldots, K_\mu, \tag{8.57c}$$

where the constraint (8.57c) enforces the conjugate row structure in each matrix block $\tilde{\boldsymbol{Q}}_k$ and the threshold parameter $\beta^{(\mathrm{nc})}$ can again be chosen [MCW05] as $\beta^{(\mathrm{nc})} = \operatorname{Tr}\left\{\mathbb{E}\{\boldsymbol{N}^{(\mathrm{nc})} \cdot \boldsymbol{N}^{(\mathrm{nc})^{\mathrm{H}}}\}\right\} = 2MN\sigma_{\mathrm{n}}^2$. The formulation in (8.57) takes advantage of the twofold sparsity structure discussed above. First, the minimization of the rank-terms promotes low-rank blocks $\tilde{\boldsymbol{Q}}_k$ and second, minimizing the sum-of-ranks enforces a block-sparse structure of $\tilde{\boldsymbol{Q}}$. However, solving the rank minimization problem in (8.57) is NP-hard [VB96] and an approximate solution can be obtained by means of convex relaxation [FHB01, CR09, CCS10, RFP10].

A prominent convex relaxation of the rank operator in (8.57) is the nuclear norm, also termed trace norm, which has been applied in several rank minimization problems [FHB01, CR09, CCS10, RFP10]. The nuclear norm is defined according to Appendix C.2.3 as

$$\left\|\tilde{\boldsymbol{Q}}_k\right\|_* = \operatorname{Tr}\left\{\left(\tilde{\boldsymbol{Q}}_k^{\mathrm{H}} \cdot \tilde{\boldsymbol{Q}}_k\right)^{(1/2)}\right\} = \sum_{r=1}^{\min(2,N)} \sigma_r(\tilde{\boldsymbol{Q}}_k), \tag{8.58}$$

where $\sigma_r(\tilde{\boldsymbol{Q}}_k)$ denotes the r-th singular value of the sub-matrix $\tilde{\boldsymbol{Q}}_k$. As described in Appendix C.2.3, the nuclear norm is equivalent to the convex ℓ_1-norm of the singular values of $\tilde{\boldsymbol{Q}}_k$, which enforces sparsity of the vector of singular values and thus, promotes solutions with low rank sub-matrices $\tilde{\boldsymbol{Q}}_k$.

Therefore, the rank minimization problem in (8.57) can be approximated by the convex nuclear

norm minimization problem

$$\min_{\tilde{\boldsymbol{Q}}} \sum_{k=1}^{K_\mu} \|\tilde{\boldsymbol{Q}}_k\|_* \tag{8.59a}$$

$$\text{s.t. } \left\|\boldsymbol{X}^{(\mathrm{nc})} - \boldsymbol{B}(\tilde{\boldsymbol{\mu}}) \cdot \tilde{\boldsymbol{Q}}\right\|_{\mathrm{F}}^2 \leq \beta^{(\mathrm{nc})} \tag{8.59b}$$

$$\tilde{\boldsymbol{Q}}_k = \boldsymbol{\Pi}_2 \cdot \tilde{\boldsymbol{Q}}_k^*, \text{ for } k = 1, \ldots, K_\mu. \tag{8.59c}$$

where $\beta^{(\mathrm{nc})}$ is chosen as in (8.57b) [MCW05]. Note that the problem in (8.59) can be interpreted as a lifted version of the $\ell_{2,1}$-mixed-norm problem in (8.26) due to the extended parametrization in (8.52).

The nuclear norm minimization problem in (8.59) can be efficiently implemented by its semi-definite programming (SDP) formulation. As shown in [FHB01], minimizing the nuclear norm is equivalent to an SDP and can be solved by standard solver such as CVX [GB14] or MOSEK [ApS15]. A proof is given in Appendix C.3.2. Consequently, the nuclear norm minimization problem in (8.59) can equivalently be expressed as

$$\min_{\tilde{\boldsymbol{Q}}, \boldsymbol{W}_{k,1}, \boldsymbol{W}_{k,2}} \frac{1}{2} \cdot \sum_{k=1}^{K_\mu} \mathrm{Tr}\left\{\boldsymbol{W}_{k,1}\right\} + \mathrm{Tr}\left\{\boldsymbol{W}_{k,2}\right\} \tag{8.60a}$$

$$\text{s.t. } \left\|\boldsymbol{X}^{(\mathrm{nc})} - \boldsymbol{B}(\tilde{\boldsymbol{\mu}}) \cdot \tilde{\boldsymbol{Q}}\right\|_{\mathrm{F}}^2 \leq \beta^{(\mathrm{nc})} \tag{8.60b}$$

$$\begin{bmatrix} \boldsymbol{W}_{k,1} & \tilde{\boldsymbol{Q}}_k \\ \tilde{\boldsymbol{Q}}_k^{\mathrm{H}} & \boldsymbol{W}_{k,2} \end{bmatrix} \succeq \boldsymbol{0}, \text{ for } k = 1, \ldots, K_\mu \tag{8.60c}$$

$$\tilde{\boldsymbol{Q}}_k = \boldsymbol{\Pi}_2 \cdot \tilde{\boldsymbol{Q}}_k^*, \text{ for } k = 1, \ldots, K_\mu, \tag{8.60d}$$

where the Hermitian matrices $\boldsymbol{W}_{k,1} \in \mathbb{C}^{2K_\mu \times 2K_\mu}$ and $\boldsymbol{W}_{k,2} \in \mathbb{C}^{N \times N}$ are auxiliary variables.

Given a solution $\hat{\tilde{\boldsymbol{Q}}}$ to the problem (8.60), the support set is identified from the indices of the non-zero sub-matrices according to

$$\mathcal{S} = \{k \,|\, \hat{\tilde{\boldsymbol{Q}}}_k \neq \boldsymbol{0}\} \tag{8.61}$$

and the spatial frequency estimates are extracted as $\{\hat{\mu}_i\}_{i=1}^{\hat{d}} = \{\tilde{\mu}_k \,|\, k \in \mathcal{S}\}$. Given a sub-matrix estimate $\hat{\tilde{\boldsymbol{Q}}}_k$, $k \in \mathcal{S}$, its singular value decomposition (SVD) is given by

$$\hat{\tilde{\boldsymbol{Q}}}_k = \hat{\boldsymbol{U}}_k \cdot \hat{\boldsymbol{\Sigma}}_k \cdot \hat{\boldsymbol{V}}_k^{\mathrm{H}} \quad \text{for } k = 1, \ldots, K_\mu. \tag{8.62}$$

Then, the corresponding real-valued signal vector $\hat{\boldsymbol{s}}_{0,k}$ and its rotation phase $\hat{\varphi}_k$ can be recovered

by means of a rank-one approximation of (8.62) via

$$\hat{\boldsymbol{s}}_{0,k} = \frac{\sigma_1\big(\hat{\tilde{\boldsymbol{Q}}}_k\big)}{\sqrt{2}} \cdot \hat{\boldsymbol{v}}_{k,1}, \qquad \hat{\varphi}_k = \arg\left\{[\hat{\boldsymbol{u}}_{k,1}]_1\right\}, \tag{8.63}$$

where $\hat{\boldsymbol{u}}_{k,1}$ and $\hat{\boldsymbol{v}}_{k,1}$ denote the principal left and right singular vectors of the sub-matrix $\hat{\tilde{\boldsymbol{Q}}}_k$, $[\boldsymbol{x}]_i$ denotes the i-th element of a vector $\boldsymbol{x}$ and the factor $\sqrt{2}$ is a normalization of $\hat{\boldsymbol{s}}_{0,k}$ to magnitude 1. Note that due to the conjugate row structure of the sub-matrices $\hat{\tilde{\boldsymbol{Q}}}_k = [\hat{\tilde{\boldsymbol{q}}}_k, \hat{\tilde{\boldsymbol{q}}}_k^*]^{\mathrm{T}} \in \mathbb{C}^{2\times N}$, there always exist real-valued right singular vectors $\hat{\boldsymbol{v}}_{k,1}$, since

$$\hat{\tilde{\boldsymbol{Q}}}_k^{\mathrm{H}} \cdot \hat{\tilde{\boldsymbol{Q}}}_k = \hat{\tilde{\boldsymbol{q}}}_k^* \cdot \hat{\tilde{\boldsymbol{q}}}_k^{\mathrm{T}} + \hat{\tilde{\boldsymbol{q}}}_k \cdot \hat{\tilde{\boldsymbol{q}}}_k^{\mathrm{H}} \in \mathbb{R}^{N\times N}. \tag{8.64}$$

It is important to note that (8.63) enables a gridless estimation of the rotation phases φ_i as in contrast to the $\ell_{2,1}$-mixed-norm minimization problem for NC signals in (8.26), the discretization of the rotation phase domain is not required.

The computational complexity of solving the nuclear norm minimization problem in (8.59) is mainly determined by the size of the sparse signal matrix $\tilde{\boldsymbol{Q}}$, i.e., the number of grid points K_μ and the number of snapshots N. For large N, we have seen in Section 8.2.1 that by applying the ℓ_1-SVD method from [MCW05], the number of optimization variables can be significantly reduced when operating on the signal subspace of the measurement matrix $\boldsymbol{X}^{(\mathrm{nc})}$ rather than on $\boldsymbol{X}^{(\mathrm{nc})}$ itself. Here, we present a similar preprocessing approach to reduce the computational complexity of solving the nuclear norm minimization problem in (8.59).

Let the SVD of the measurement matrix $\boldsymbol{X}^{(\mathrm{nc})}$ in the data model (8.49) be given by

$$\begin{aligned}\boldsymbol{X}^{(\mathrm{nc})} &= \boldsymbol{B}(\boldsymbol{\mu}) \cdot \boldsymbol{\Phi}(\boldsymbol{\varphi}) \cdot \boldsymbol{S}_0 + \boldsymbol{N}^{(\mathrm{nc})} \\ &= \boldsymbol{B}(\boldsymbol{\mu}) \cdot \boldsymbol{Q} + \boldsymbol{N}^{(\mathrm{nc})} = \boldsymbol{U} \cdot \boldsymbol{\Sigma} \cdot \boldsymbol{V}^{\mathrm{H}}.\end{aligned} \tag{8.65}$$

Since $\boldsymbol{X}^{(\mathrm{nc})^{\mathrm{H}}} \cdot \boldsymbol{X}^{(\mathrm{nc})} = \boldsymbol{X}^{\mathrm{H}} \cdot \boldsymbol{X} + \boldsymbol{X}^{\mathrm{T}} \cdot \boldsymbol{X}^* \in \mathbb{R}^{N\times N}$, there exists a real-valued unitary basis $\boldsymbol{V} \in \mathbb{R}^{N\times N}$ for the row space of $\boldsymbol{X}^{(\mathrm{nc})}$. Then, if $N > d$, we can define the reduced dimensional measurement matrix $\boldsymbol{X}_{\mathrm{sv}}^{(\mathrm{nc})} \in \mathbb{C}^{M\times d}$ by applying the following preprocessing step to $\boldsymbol{X}^{(\mathrm{nc})}$:

$$\begin{aligned}\boldsymbol{X}_{\mathrm{sv}}^{(\mathrm{nc})} &= \boldsymbol{X}^{(\mathrm{nc})} \cdot \boldsymbol{V} \cdot \boldsymbol{K} \\ &= \boldsymbol{B}(\boldsymbol{\mu}) \cdot \boldsymbol{\Phi}(\boldsymbol{\varphi}) \cdot \boldsymbol{S}_0 \cdot \boldsymbol{V} \cdot \boldsymbol{K} + \boldsymbol{N}^{(\mathrm{nc})} \cdot \boldsymbol{V} \cdot \boldsymbol{K} \\ &= \boldsymbol{B}(\boldsymbol{\mu}) \cdot \boldsymbol{\Phi}(\boldsymbol{\varphi}) \cdot \boldsymbol{S}_{\mathrm{sv}} + \boldsymbol{N}_{\mathrm{sv}}^{(\mathrm{nc})},\end{aligned} \tag{8.66}$$

where the selection matrix $\boldsymbol{K} = [\boldsymbol{I}_d,\ \boldsymbol{0}_{d\times(N-d)}]^{\mathrm{T}} \in \mathbb{R}^{N\times d}$ extracts the d dominant right singular vectors in $\boldsymbol{V}$. As both $\boldsymbol{S}_0$ and $\boldsymbol{V}$ are real-valued, so is the matrix product $\boldsymbol{S}_{\mathrm{sv}} = \boldsymbol{S}_0 \cdot \boldsymbol{V} \cdot \boldsymbol{K}$. Thus, the matrix $\boldsymbol{Q}_{\mathrm{sv}} = \boldsymbol{\Phi}(\boldsymbol{\varphi}) \cdot \boldsymbol{S}_{\mathrm{sv}} \in \mathbb{C}^{2K\times d}$ exhibits the same conjugate row structure as $\boldsymbol{Q} = \boldsymbol{\Phi}(\boldsymbol{\varphi}) \cdot \boldsymbol{S}_0 \in$

$\mathbb{C}^{2K\times N}$, but has a significantly reduced number of columns from N to d. Based on the model in (8.66) with reduced dimensionality, we can formulate the nuclear norm minimization problem

$$\begin{aligned} \min_{\tilde{\boldsymbol{Q}}_{\mathrm{sv}}} \ & \sum_{k=1}^{K_\mu} \|\tilde{\boldsymbol{Q}}_{\mathrm{sv},k}\|_* \\ \text{s.t.} \ & \left\|\boldsymbol{X}_{\mathrm{sv}}^{(\mathrm{nc})} - \boldsymbol{B}(\tilde{\boldsymbol{\mu}})\cdot\tilde{\boldsymbol{Q}}_{\mathrm{sv}}\right\|_{\mathrm{F}}^2 \le \beta_{\mathrm{sv}}^{(\mathrm{nc})} \\ & \tilde{\boldsymbol{Q}}_{\mathrm{sv},k} = \boldsymbol{\Pi}_2\cdot\tilde{\boldsymbol{Q}}_{\mathrm{sv},k}^*, \ \text{for} \ \ k=1,\ldots,K_\mu, \end{aligned} \tag{8.67}$$

which can be solved, for instance, via a similar SDP formulation as presented in (8.60).

Considering the selection of the regularization parameter $\beta_{\mathrm{sv}}^{(\mathrm{nc})}$, we follow the approach in [MCW05]. We choose $\beta_{\mathrm{sv}}^{(\mathrm{nc})}$ according to the noise statistics such that it provides an upper bound on the noise power with high probability γ, i.e.,

$$\mathrm{Prob}\left\{\|\boldsymbol{N}_{\mathrm{sv}}^{(\mathrm{nc})}\|_{\mathrm{F}}^2 \le \beta_{\mathrm{sv}}^{(\mathrm{nc})}\right\} = \gamma. \tag{8.68}$$

In the case that $\boldsymbol{N}$ contained in $\boldsymbol{N}^{(\mathrm{nc})} = [\boldsymbol{N}^{\mathrm{T}}, (\boldsymbol{\Pi}_M\cdot\boldsymbol{N})^{\mathrm{H}}]^{\mathrm{T}}$ has independent identically distributed (i.i.d.) Gaussian entries and for a moderate to a high SNR, $\|\boldsymbol{N}_{\mathrm{sv}}^{(\mathrm{nc})}\|_{\mathrm{F}}^2 = \|\boldsymbol{N}^{(\mathrm{nc})}\cdot\boldsymbol{V}\cdot\boldsymbol{K}\|_{\mathrm{F}}^2$ follows approximately a χ^2-distribution with Md degrees of freedom upon its normalization by the noise variance $2\sigma_{\mathrm{n}}^2$. The reason that this holds only approximately is that the SVD in (8.65) depends on the particular realization of the noise, and hence, the matrix $\boldsymbol{V}$ is a function of $\boldsymbol{N}^{(\mathrm{nc})}$. However, when the noise is small, the term $\boldsymbol{B}(\boldsymbol{\mu})\cdot\boldsymbol{Q}$ dominates the SVD, the effect of $\boldsymbol{N}^{(\mathrm{nc})}$ becomes small, and $\boldsymbol{N}_{\mathrm{sv}}^{(\mathrm{nc})}$ has a χ^2-distribution, such that, according to (8.68), we can compute $\beta_{\mathrm{sv}}^{(\mathrm{nc})}$ by an inverse χ^2-distribution for some predefined probability γ.

8.3.2.1. Off-grid Estimation

So far, we have assumed that the on-grid assumption is fulfilled, i.e., the true spatial frequencies are a subset of the sampling grid points such that $\{\mu_i\}_{i=1}^d \in \{\tilde{\mu}_k\}_{k=1}^{K_\mu}$. However, in practice, we generally have $\{\mu_i\}_{i=1}^d \notin \{\tilde{\mu}_k\}_{k=1}^{K_\mu}$, which leads to the well-known off-grid problem.

In Section 8.3.1.1, we have presented an efficient grid-offset estimation procedure based on the concept in [IRA+14] to address the emerging 2-D NC off-grid problem in the sparse recovery of NC signals via $\ell_{2,1}$-mixed-norm minimization. We have seen in (8.59) that by solving the sparse recovery problem of NC signals by means of nuclear norm minimization, we are left with only a 1-D off-grid problem in the spatial domain. In this section, we develop two analytical 1-D grid-offset estimation schemes for a single NC source and two closely-spaced NC sources, which can be applied after the support estimation in (8.61).

As in Section 8.3.1.1, we assume a ULA with centered phase reference such that the array

steering vectors are given by

$$\boldsymbol{a}(\mu) = \left[\mathrm{e}^{-\mathrm{j}\frac{M-1}{2}\mu} \quad \dots \quad \mathrm{e}^{\mathrm{j}\frac{M-1}{2}\mu}\right]^{\mathrm{T}} \in \mathbb{C}^{M\times 1}. \tag{8.69}$$

Then, the augmented NC array steering vector is given as $\boldsymbol{a}^{(\mathrm{nc})}(\mu,\varphi) = [\mathrm{e}^{\mathrm{j}\varphi}, \mathrm{e}^{-\mathrm{j}\varphi}]^{\mathrm{T}} \otimes \boldsymbol{a}(\mu) \in \mathbb{C}^{2M\times 1}$.

We first consider the case of a single NC source with spatial frequency μ_1. We assume that the support $\mathcal{S}$ defined in (8.61) consists of the indices $\mathcal{S} = \{k_1, k_1+1\}$ such that $\tilde{\mu}_{\mathcal{S}} = [\tilde{\mu}_{k_1}\ \tilde{\mu}_{k_1+1}]^{\mathrm{T}}$ contains the two neighboring grid points closest to μ_1, from the left and the right, respectively. Let ϵ_1 denote the grid offset with $0 \le \epsilon_1 \le 1$ such that the off-grid model

$$\mu_1 = \tilde{\mu}_{k_1} + \epsilon_1 \Delta_\mu \tag{8.70}$$

holds, where Δ_μ is the grid spacing. It is straightforward to see that in the case of a single NC source, the 1-D grid-offset estimator from [IRA+14] is also applicable in this case. This follows from the fact that a single NC source provides no performance benefits over a source with arbitrary symbols, which is proven based on the deterministic NC Cramér-Rao bound in Chapter 9. Thus, the rotation phase φ_1 of a single NC source is irrelevant. Consequently, in the absence of noise, the true steering vector $\boldsymbol{a}^{(\mathrm{nc})}(\mu_1,\varphi_1) = \boldsymbol{a}^{(\mathrm{nc})}(\tilde{\mu}_{k_1} + \epsilon_1\Delta_\mu, \varphi_1)$ can be approximated by the linear model

$$\boldsymbol{a}^{(\mathrm{nc})}(\mu_1,\varphi_1) \approx \left[\boldsymbol{a}^{(\mathrm{nc})}(\tilde{\mu}_{k_1}, \hat{\varphi}_{k_1}),\ \boldsymbol{a}^{(\mathrm{nc})}(\tilde{\mu}_{k_1+1}, \hat{\varphi}_{k_1+1})\right] \cdot \boldsymbol{\alpha}(\epsilon_1), \tag{8.71}$$

where the unknown coefficients in $\boldsymbol{\alpha}(\epsilon_1) = [\alpha_{k_1}(\epsilon_1), \alpha_{k_1+1}(\epsilon_1)]^{\mathrm{T}}$ provide a good representation of the true steering vector $\boldsymbol{a}^{(\mathrm{nc})}(\mu_1,\varphi_1)$, and $\hat{\varphi}_{k_1}$, $\hat{\varphi}_{k_1+1}$ are the two corresponding continuous rotation phase estimates reconstructed from (8.63). After computing the coefficients in $\boldsymbol{\alpha}(\epsilon)$ as presented in [IRA+14], the simple closed-form estimator from [IRA+14]

$$\hat{\epsilon}_1 = \frac{\alpha_{k_1+1}(\epsilon_1)}{\alpha_{k_1}(\epsilon_1) + \alpha_{k_1+1}(\epsilon_1)} \tag{8.72}$$

can also be applied in the NC case presented here. Note that in the presence of noise, $\boldsymbol{\alpha}(\epsilon)$ becomes $\hat{\boldsymbol{\alpha}}(\epsilon)$.

For the case of two closely-spaced NC off-grid sources, we present a numerical joint grid-offset estimation procedure, which is inspired by [IRA+14] and similar to that in Section 8.3.1.1. Let $\boldsymbol{\mu} = [\mu_1, \mu_2]^{\mathrm{T}}$ and $\boldsymbol{\varphi} = [\varphi_1, \varphi_2]^{\mathrm{T}}$ contain the true spatial frequencies and the rotation phases of the two impinging source signals. Furthermore, let the support set $\mathcal{S} = \{k_1, k_1+1, k_2, k_2+1\}$ found from (8.61) contain the grid indices such that $\tilde{\mu}_{k_i}$ and $\tilde{\mu}_{k_i+1}$ are the nearest grid points left and right of the source direction μ_i for $i = 1, 2$. We summarize the grid points as $\tilde{\boldsymbol{\mu}}_{\mathcal{S}} = [\tilde{\mu}_{k_1}, \tilde{\mu}_{k_1+1}, \tilde{\mu}_{k_2}, \tilde{\mu}_{k_2+1}]^{\mathrm{T}}$ and the corresponding gridless rotation phase estimates obtained from

(8.63) as $\hat{\boldsymbol{\varphi}}_{\mathcal{S}} = [\hat{\varphi}_{k_1}, \hat{\varphi}_{k_1+1}, \hat{\varphi}_{k_2}, \hat{\varphi}_{k_2+1}]^{\mathrm{T}}$. Moreover, we define the grid offsets $\boldsymbol{\epsilon} = [\epsilon_1, \epsilon_2]^{\mathrm{T}}$ in $\boldsymbol{\mu} = [\tilde{\mu}_{k_1}, \tilde{\mu}_{k_2}]^{\mathrm{T}} + \boldsymbol{\epsilon}\Delta_\mu$ and the free parameters $\boldsymbol{\delta} = [\delta_1, \delta_2]^{\mathrm{T}}$ in $\boldsymbol{\varphi} = [\bar{\varphi}_{k_1}, \bar{\varphi}_{k_2}]^{\mathrm{T}} + \boldsymbol{\delta}$, where $\bar{\varphi}_{k_i} = (\hat{\varphi}_{k_i} + \hat{\varphi}_{k_i+1})/2$ is the average of the two rotation phase estimates corresponding to $\tilde{\mu}_{k_i}$ $\tilde{\mu}_{k_i+1}$ for the i-th source.

Then, following the arguments for the noiseless case $\boldsymbol{X}_0^{(\mathrm{nc})} = \boldsymbol{A}^{(\mathrm{nc})}(\boldsymbol{\mu}, \boldsymbol{\varphi}) \cdot \boldsymbol{S}_0$ in Section 8.3.1.1, we define the matrix

$$\boldsymbol{G}(\boldsymbol{\epsilon}, \boldsymbol{\delta}) = \boldsymbol{A}^{(\mathrm{nc})^+}(\tilde{\boldsymbol{\mu}}_{\mathcal{S}}, \hat{\boldsymbol{\varphi}}_{\mathcal{S}}) \cdot \boldsymbol{A}^{(\mathrm{nc})}(\boldsymbol{\mu}, \boldsymbol{\varphi}) \cdot \boldsymbol{S}_0 \in \mathbb{R}^{4\times N}, \tag{8.73}$$

which is the correlation of the reconstructed steering matrix based on the spatial frequencies in $\tilde{\boldsymbol{\mu}}_{\mathcal{S}}$ and the phase estimates $\hat{\boldsymbol{\varphi}}_{\mathcal{S}}$ with the true noise-free measurement matrix. Furthermore, we make use of the fact that we can rewrite (8.73) as $\boldsymbol{G}(\boldsymbol{\epsilon}, \boldsymbol{\delta}) = \boldsymbol{D}_0^{-1} \cdot \boldsymbol{D}(\boldsymbol{\epsilon}, \boldsymbol{\delta}) \cdot \boldsymbol{S}_0$, where we define

$$\begin{aligned} \boldsymbol{D}_0 &= \boldsymbol{A}^{(\mathrm{nc})^{\mathrm{H}}}(\tilde{\boldsymbol{\mu}}_{\mathcal{S}}, \hat{\boldsymbol{\varphi}}_{\mathcal{S}}) \cdot \boldsymbol{A}^{(\mathrm{nc})}(\tilde{\boldsymbol{\mu}}_{\mathcal{S}}, \hat{\boldsymbol{\varphi}}_{\mathcal{S}}) \\ &= \begin{bmatrix} \boldsymbol{D}(0) & \cos(\Delta\hat{\varphi}) \cdot \boldsymbol{D}(d) \\ \cos(\Delta\hat{\varphi}) \cdot \boldsymbol{D}(d)^{\mathrm{T}} & \boldsymbol{D}(0) \end{bmatrix} \in \mathbb{R}^{4\times 4} \end{aligned} \tag{8.74}$$

and

$$\begin{aligned} \boldsymbol{D}(\boldsymbol{\epsilon}, \boldsymbol{\delta}) &= \boldsymbol{A}^{(\mathrm{nc})^{\mathrm{H}}}(\tilde{\boldsymbol{\mu}}_{\mathcal{S}}, \hat{\boldsymbol{\varphi}}_{\mathcal{S}}) \cdot \boldsymbol{A}^{(\mathrm{nc})}(\boldsymbol{\mu}, \boldsymbol{\varphi}) \\ &= \begin{bmatrix} \cos(\delta_1) \cdot \boldsymbol{d}(\epsilon_1) & \cos(\delta_2 + \Delta\hat{\varphi}) \cdot \boldsymbol{d}(\epsilon_2 + d) \\ \cos(\delta_1 - \Delta\hat{\varphi}) \cdot \boldsymbol{d}(\epsilon_1 - 1) & \cos(\delta_2) \cdot \boldsymbol{d}(\epsilon_2) \end{bmatrix} \in \mathbb{R}^{4\times 2} \end{aligned} \tag{8.75}$$

with $\boldsymbol{D}(x) = [\boldsymbol{d}(x), \boldsymbol{d}(x+1)] \in \mathbb{R}^{2\times 2}$, $\boldsymbol{d}(x) = [D(x\Delta_\mu), D((x-1)\Delta_\mu)]^{\mathrm{T}} \in \mathbb{R}^{2\times 1}$. The function $D(y)$ is defined as in (8.35) by

$$D(y) = \begin{cases} M & \text{if } y = 0 \\ \sin(yM/2)/\sin(y/2) & \text{else.} \end{cases} \tag{8.76}$$

Moreover, $d = k_2 - k_1$ denotes the grid distance of the estimated support and $\Delta\hat{\varphi} = |\bar{\varphi}_{k_2} - \bar{\varphi}_{k_1}|$ is the rotation phase difference.

In the noisy case, we obtain the estimate $\hat{\boldsymbol{G}} = \boldsymbol{A}^{(\mathrm{nc})^+}(\tilde{\boldsymbol{\mu}}_{\mathcal{S}}, \hat{\boldsymbol{\varphi}}_{\mathcal{S}}) \cdot \boldsymbol{X}^{(\mathrm{nc})}$, and define $\bar{\boldsymbol{G}} = \boldsymbol{D}_0 \cdot \hat{\boldsymbol{G}}$. As the columns of $\boldsymbol{G}(\boldsymbol{\epsilon}, \boldsymbol{\delta})$ are a linear combination of the columns of $\boldsymbol{D}(\boldsymbol{\epsilon}, \boldsymbol{\delta})$, we can estimate $\boldsymbol{\epsilon}$ and $\boldsymbol{\delta}$ by minimizing the projection of $\bar{\boldsymbol{G}}$ onto the complement of the column space spanned by $\boldsymbol{D}(\boldsymbol{\epsilon}, \boldsymbol{\delta})$, i.e., we minimize the cost function

$$J(\boldsymbol{\epsilon}, \boldsymbol{\delta}) = \left\| (\boldsymbol{I}_4 - \boldsymbol{D}(\boldsymbol{\epsilon}, \boldsymbol{\delta}) \cdot \boldsymbol{D}^+(\boldsymbol{\epsilon}, \boldsymbol{\delta})) \cdot \bar{\boldsymbol{G}} \right\|_{\mathrm{F}}^2, \tag{8.77}$$

which is smooth and convex for $0 \leq \epsilon_\ell, \delta_\ell \leq 1$. Thus, any local optimization algorithm, e.g., the gradient descent method, can be used to minimize (8.77). If more than two NC sources are present, the presented grid-offset estimation schemes can still be applied after clustering the sources into groups of single and two NC sources.

8.3.3. Gridless parameter estimation based on NC sparse recovery

In Section 8.3.1 and in Section 8.3.2, we have developed two sparse signal reconstruction (SSR) algorithms for NC signals based on $\ell_{2,1}$-mixed-norm minimization and nuclear norm minimization, respectively. As both algorithms are grid-based, a large number of grid points K_μ, which is associated with a high computational complexity, is required to achieved a sufficient frequency resolution. Moreover, in the case of off-grid sources, we have seen that postprocessing schemes for the grid-offset estimation as developed in Section 8.3.1.1 and in Section 8.3.2.1 have to be applied. In order to avoid the discretization of the spatial domain, the development of gridless sparse recovery algorithms for NC signals is highly desirable.

In this section, we present a gridless sparse recovery algorithm for NC signals based on atomic norm minimization (ANM) in the multiple measurement vector (MMV) case, which has been published in [SRS+16]. The general concept of ANM and its application to frequency estimation has been introduced in Section 8.2.2 and in Appendix C.4. The major advantage of the ANM framework is that it allows for the signal reconstruction in the continuous parameter domain and hence, does not require any discretization. As will be shown, the proposed NC ANM algorithm for NC signals relies on the 2-D extension of the ANM framework considered in [CC15], which is based on the 2-D Vandermonde decomposition for two-level Hermitian Block-Toeplitz matrices proposed in [GB02, Geo07, CC15]. The generalization of the ANM framework and the Vandermonde decomposition to the R-D case for $R \geq 2$ has recently been presented in [YXS16].

For the derivation of the ANM algorithm for NC signals, we first apply the common NC preprocessing step to the measurement matrix $\boldsymbol{X}$ from (8.1) to obtain the augmented measurement matrix $\boldsymbol{X}^{(\mathrm{nc})} \in \mathbb{C}^{2M \times N}$

$$\boldsymbol{X}^{(\mathrm{nc})} = \begin{bmatrix} \boldsymbol{X} \\ \boldsymbol{\Pi}_M \cdot \boldsymbol{X}^* \end{bmatrix} = \begin{bmatrix} \boldsymbol{A}(\boldsymbol{\mu}) \cdot \boldsymbol{\Psi}(\boldsymbol{\varphi}) \\ \boldsymbol{\Pi}_M \cdot \boldsymbol{A}^*(\boldsymbol{\mu}) \cdot \boldsymbol{\Psi}^*(\boldsymbol{\varphi}) \end{bmatrix} \cdot \boldsymbol{S}_0 + \begin{bmatrix} \boldsymbol{N} \\ \boldsymbol{\Pi}_M \cdot \boldsymbol{N}^* \end{bmatrix} \tag{8.78}$$

$$= \boldsymbol{A}^{(\mathrm{nc})}(\boldsymbol{\mu}, \boldsymbol{\varphi}) \cdot \boldsymbol{S}_0 + \boldsymbol{N}^{(\mathrm{nc})} = \boldsymbol{X}_0^{(\mathrm{nc})} + \boldsymbol{N}^{(\mathrm{nc})}, \tag{8.79}$$

where $\boldsymbol{A}^{(\mathrm{nc})}(\boldsymbol{\mu}, \boldsymbol{\varphi}) \in \mathbb{C}^{2M \times d}$ is the augmented array steering matrix. Due to the virtually doubled number of sensors in $\boldsymbol{A}^{(\mathrm{nc})}(\boldsymbol{\mu}, \boldsymbol{\varphi})$, processing $\boldsymbol{X}^{(\mathrm{nc})}$ instead of $\boldsymbol{X}$ can improve the parameter estimation accuracy and doubles the number of identifiable sources.

Next, we assume for simplicity that the phase reference of the original array is located at the array centroid as considered in Section 2.1.4.2, i.e., $\boldsymbol{\Delta} = \boldsymbol{I}_d$ such that $\boldsymbol{\Pi}_M \cdot \boldsymbol{A}^*(\boldsymbol{\mu}) = \boldsymbol{A}(\boldsymbol{\mu})$. Then,

the array steering vector $\boldsymbol{a}(\mu_i)$ corresponding to the i-th spatial frequency is given by

$$\boldsymbol{a}(\mu_i) = \left[\mathrm{e}^{-\mathrm{j}\frac{M-1}{2}\mu_i} \quad \dots \quad \mathrm{e}^{\mathrm{j}\frac{M-1}{2}\mu_i}\right]^{\mathrm{T}} \in \mathbb{C}^{M\times 1}. \tag{8.80}$$

With this assumption, the augmented steering matrix $\boldsymbol{A}^{(\mathrm{nc})}(\boldsymbol{\mu},\boldsymbol{\varphi})$ in (8.79) can be compactly expressed as

$$\boldsymbol{A}^{(\mathrm{nc})}(\boldsymbol{\mu},\boldsymbol{\varphi}) = \boldsymbol{\Phi}(\boldsymbol{\varphi}) \diamond \boldsymbol{A}(\boldsymbol{\mu}), \tag{8.81}$$

where $\boldsymbol{\Phi}(\boldsymbol{\varphi}) = [\boldsymbol{\phi}_1, \dots, \boldsymbol{\phi}_d] \in \mathbb{C}^{2\times d}$ with $\boldsymbol{\phi}_i = [\mathrm{e}^{\mathrm{j}\varphi_i}, \mathrm{e}^{-\mathrm{j}\varphi_i}]^{\mathrm{T}} \in \mathbb{C}^{2\times 1}$ and $\diamond$ denotes the Khatri-Rao product (the column-wise Kronecker product). Using (8.81), we can rewrite the noise-free augmented measurement matrix $\boldsymbol{X}_0^{(\mathrm{nc})}$ in (8.79) as

$$\boldsymbol{X}_0^{(\mathrm{nc})} = \sum_{i=1}^{d} \boldsymbol{a}^{(\mathrm{nc})}(\mu_i,\varphi_i) \cdot \boldsymbol{s}_{0_i}^{\mathrm{T}} = \sum_{i=1}^{d} \left(\begin{bmatrix} \mathrm{e}^{\mathrm{j}\varphi_i} \\ \mathrm{e}^{-\mathrm{j}\varphi_i} \end{bmatrix} \otimes \boldsymbol{a}(\mu_i) \right) \cdot \boldsymbol{s}_{0_i}^{\mathrm{T}}, \tag{8.82}$$

where $\boldsymbol{a}^{(\mathrm{nc})}(\mu_i,\varphi_i)$ represents the i-th column of $\boldsymbol{A}^{(\mathrm{nc})}(\boldsymbol{\mu},\boldsymbol{\varphi})$ in (8.81) and $\boldsymbol{s}_{0_i}^{\mathrm{T}}$ represents the i-th row of $\boldsymbol{S}_0$. The Kronecker structure of $\boldsymbol{a}^{(\mathrm{nc})}(\mu_i,\varphi_i)$ in (8.82) reveals the 2-D structure as both dimensions, i.e., the spatial domain and the rotation phase domain, are separable. Moreover, it is apparent that (8.82) is sparse in both dimensions if only a small number of signals d is present. Note that the Kronecker structure in (8.82) also admits a tensor formulation of the measurements contained in $\boldsymbol{X}_0^{(\mathrm{nc})}$, which enables the extension of the ANM framework to the tensor case. This extension, however, is not considered in this thesis and is left for future work.

In [CC15, YXS16], the authors have proposed an R-D extension of the ANM framework for multi-dimensional parameter estimation using a uniform separable R-D sampling grid as described in Section 2.1.4.1. From (8.82), we conclude that the NC case considered here corresponds to the 2-D case of [CC15, YXS16], where the spatial dimension and the rotation phase dimension are present. Therefore, similarly to [CC15, YXS16], we extend the 1-D ANM framework introduced in Section 8.2.2 to the 2-D case by modeling the noise-free augmented measurement matrix $\boldsymbol{X}_0^{(\mathrm{nc})}$ in (8.82) as a linear combination of the atoms

$$\tilde{\boldsymbol{X}}_0^{(\mathrm{nc})}(\tilde{\mu},\tilde{\varphi},\tilde{\boldsymbol{s}}_0) = \boldsymbol{a}^{(\mathrm{nc})}(\tilde{\mu},\tilde{\varphi}) \cdot \tilde{\boldsymbol{s}}_0^{\mathrm{T}} = \left(\begin{bmatrix} \mathrm{e}^{\mathrm{j}\tilde{\varphi}_i} \\ \mathrm{e}^{-\mathrm{j}\tilde{\varphi}_i} \end{bmatrix} \otimes \boldsymbol{a}(\tilde{\mu}_i) \right) \cdot \tilde{\boldsymbol{s}}_0^{\mathrm{T}}, \tag{8.83}$$

where $\tilde{\mu} \in [-\pi,\pi)$, $\tilde{\varphi} \in [0,2\pi)$ and $\tilde{\boldsymbol{s}}_0 \in \mathbb{R}^{N\times 1}$ with $\|\tilde{\boldsymbol{s}}_0\|_2 = 1$. Subsequently, the continuous atomic set $\mathcal{A}$ is given by

$$\mathcal{A} = \left\{ \tilde{\boldsymbol{X}}_0^{(\mathrm{nc})}(\tilde{\mu},\tilde{\varphi},\tilde{\boldsymbol{s}}_0) \mid \tilde{\mu} \in [-\pi,\pi),\ \tilde{\varphi} \in [0,2\pi),\ \|\tilde{\boldsymbol{s}}_0\|_2 = 1 \right\}. \tag{8.84}$$

According to [YX16b, YXS16], the atomic ℓ_0-norm of $\boldsymbol{X}_0^{(\mathrm{nc})}$ is defined as

$$\left\|\boldsymbol{X}_0^{(\mathrm{nc})}\right\|_{\mathcal{A},0} = \inf_{\substack{\{\tilde{\mu}_k,\\ \tilde{\varphi}_k,\tilde{\boldsymbol{s}}_{0_k}\}}} \left\{ K \;\middle|\; \boldsymbol{X}_0^{(\mathrm{nc})} = \sum_{k=1}^{K} c_k \cdot \tilde{\boldsymbol{X}}_0^{(\mathrm{nc})}(\tilde{\mu}_k, \tilde{\varphi}_k, \tilde{\boldsymbol{s}}_{0_k}),\ c_k \geq 0 \right\}, \tag{8.85}$$

which describes the smallest number of atoms to compose $\boldsymbol{X}_0^{(\mathrm{nc})}$. As discussed in Section 8.2.2, minimizing the atomic ℓ_0-norm in (8.85) is non-convex and NP-hard. The corresponding convex atomic ℓ_1-norm (cf. Appendix C.4) is defined as

$$\|\boldsymbol{X}_0^{(\mathrm{nc})}\|_{\mathcal{A}} = \inf_{\substack{\{\tilde{\mu}_k,\\ \tilde{\varphi}_k,\tilde{\boldsymbol{s}}_{0_k}\}}} \left\{ \sum_k c_k \;\middle|\; \boldsymbol{X}_0^{(\mathrm{nc})} = \sum_k c_k \cdot \tilde{\boldsymbol{X}}_0^{(\mathrm{nc})}(\tilde{\mu}_k, \tilde{\varphi}_k, \tilde{\boldsymbol{s}}_{0_k}),\ c_k \geq 0 \right\}. \tag{8.86}$$

As proven in [YX16b, YXS16] for the noiseless R-D case and assuming a uniform sampling grid, the general multi-dimensional atomic ℓ_1-norm minimization problem can be equivalently expressed as a semi-definite programming (SDP) problem. Applying this result to the 2-D atomic ℓ_1-norm minimization problem (8.86), we can equivalently formulate (8.86) as the SDP problem

$$\min_{\boldsymbol{W},\boldsymbol{u},\boldsymbol{v}} \quad \frac{1}{2} \cdot \mathrm{Tr}\{\boldsymbol{W}\} + \frac{1}{4M} \cdot \mathrm{Tr}\{\mathrm{Toep}\{\boldsymbol{u},\boldsymbol{v}\}\} \tag{8.87}$$

$$\text{s. t.} \quad \begin{bmatrix} \boldsymbol{W} & \boldsymbol{X}_0^{(\mathrm{nc})^{\mathrm{H}}} \\ \boldsymbol{X}_0^{(\mathrm{nc})} & \mathrm{Toep}\{\boldsymbol{u},\boldsymbol{v}\} \end{bmatrix} \succeq \boldsymbol{0}, \tag{8.88}$$

where $\mathrm{Toep}\{\boldsymbol{u},\boldsymbol{v}\} \in \mathbb{C}^{2M\times 2M}$ with $\boldsymbol{u} = [u_1,\ldots,u_M]^{\mathrm{T}} \in \mathbb{C}^{M\times 1}$ and $\boldsymbol{v} = [v_1,\ldots,v_M]^{\mathrm{T}} \in \mathbb{C}^{M\times 1}$ denotes the two-level Hermitian Toeplitz matrix, also termed Hermitian Toeplitz-Block-Toeplitz matrix

$$\mathrm{Toep}\{\boldsymbol{u},\boldsymbol{v}\} = \begin{bmatrix} \mathrm{Toep}\{\boldsymbol{u}\} & \mathrm{Toep}\{\boldsymbol{v}\} \\ \mathrm{Toep}\{\boldsymbol{v}^*\} & \mathrm{Toep}\{\boldsymbol{u}\} \end{bmatrix}, \tag{8.89}$$

where

$$\mathrm{Toep}\{\boldsymbol{u}\} = \begin{bmatrix} u_1 & u_2 & \cdots & u_M \\ u_2^* & u_1 & \cdots & u_{M-1} \\ \vdots & \vdots & \ddots & \vdots \\ u_M^* & u_{M-1}^* & \cdots & u_1 \end{bmatrix}, \quad \mathrm{Toep}\{\boldsymbol{v}\} = \begin{bmatrix} v_1 & v_2 & \cdots & v_M \\ v_2^* & v_1 & \cdots & v_{M-1} \\ \vdots & \vdots & \ddots & \vdots \\ v_M^* & v_{M-1}^* & \cdots & v_1 \end{bmatrix}. \tag{8.90}$$

In order to reconstruct the spatial frequencies μ_k and the magnitudes c_k for $k = 1,\ldots,d$ from a solution $\mathrm{Toep}\{\hat{\boldsymbol{u}},\hat{\boldsymbol{v}}\}$ to the problem (8.87), we apply the multi-dimensional Vandermonde decomposition and the multi-dimensional generalization of the Carathéodory theorem from [YXS16], which states the result:

Definition 8.3.1. In the noiseless case and for an M-element uniform R-D sampling grid with a ULA of M_r sensors in the r-th mode for $r = 1, \ldots, R$, any positive semi-definite R-level Hermitian Toeplitz matrix $\mathrm{Toep}\{\boldsymbol{u}_1, \ldots, \boldsymbol{u}_R\} \in \mathbb{C}^{M \times M}$ as defined in [YXS16] with $\mathrm{rank}\{\mathrm{Toep}\{\boldsymbol{u}_1, \ldots, \boldsymbol{u}_R\}\} = K \leq \min_r\{M_r - 1\}$ can be uniquely decomposed as

$$\mathrm{Toep}\{\boldsymbol{u}\} = \boldsymbol{A} \cdot \boldsymbol{C} \cdot \boldsymbol{A}^{\mathrm{H}} = \sum_{k=1}^{K} c_k \cdot \boldsymbol{a}(\tilde{\boldsymbol{\mu}}_k) \cdot \boldsymbol{a}^{\mathrm{H}}(\tilde{\boldsymbol{\mu}}_k), \tag{8.91}$$

where $\boldsymbol{A} = [\boldsymbol{a}(\tilde{\boldsymbol{\mu}}_1), \ldots, \boldsymbol{a}(\tilde{\boldsymbol{\mu}}_K)] \in \mathbb{C}^{M \times K}$ with $\boldsymbol{a}(\tilde{\boldsymbol{\mu}}_k) = \boldsymbol{a}\left(\tilde{\mu}_k^{(1)}\right) \otimes \cdots \otimes \boldsymbol{a}\left(\tilde{\mu}_k^{(R)}\right) \in \mathbb{C}^{M \times 1}$ and $\boldsymbol{a}\left(\tilde{\mu}_k^{(r)}\right)$ has a Vandermonde structure in the r-th mode. Moreover, $\boldsymbol{C} = \mathrm{diag}\{[c_1, \ldots, c_K]\} \in \mathbb{R}^{K \times K}$ is a diagonal matrix that contains the non-negative coefficients $c_k > 0,\ k = 1, \ldots, K$, on its diagonal.

Applying the multi-dimensional Vandermonde decomposition in (8.91) to the NC case in (8.87), we obtain the 2-D Vandermonde decomposition

$$\begin{aligned} \mathrm{Toep}\{\boldsymbol{u}, \boldsymbol{v}\} &= \boldsymbol{A}^{(\mathrm{nc})}(\tilde{\boldsymbol{\mu}}, \tilde{\boldsymbol{\varphi}}) \cdot \boldsymbol{C} \cdot \boldsymbol{A}^{(\mathrm{nc})^{\mathrm{H}}}(\tilde{\boldsymbol{\mu}}, \tilde{\boldsymbol{\varphi}}) && (8.92) \\ &= \sum_{k=1}^{K} c_k \cdot \boldsymbol{a}^{(\mathrm{nc})}(\tilde{\mu}_k, \tilde{\varphi}_k) \cdot \boldsymbol{a}^{(\mathrm{nc})^{\mathrm{H}}}(\tilde{\mu}_k, \tilde{\varphi}_k), && (8.93) \end{aligned}$$

where $\boldsymbol{C} = \mathrm{diag}\{[c_1, \ldots, c_K]\}$ contains the coefficients $c_k > 0,\ k = 1, \ldots, K$, on its diagonal,

$$\boldsymbol{a}^{(\mathrm{nc})}(\tilde{\mu}_k, \tilde{\varphi}_k) = \begin{bmatrix} \mathrm{e}^{\mathrm{j}\tilde{\varphi}_k} \\ \mathrm{e}^{-\mathrm{j}\tilde{\varphi}_k} \end{bmatrix} \otimes \boldsymbol{a}(\tilde{\mu}_k), \tag{8.94}$$

and $\mathrm{rank}\{\mathrm{Toep}\{\boldsymbol{u}, \boldsymbol{v}\}\} = K \leq 2(M-1)$. The 2-D Vandermonde decomposition in (8.93) can be computed by estimating the spatial frequencies $\tilde{\mu}_k$ from the two-level Hermitian Toeplitz matrix $\mathrm{Toep}\{\hat{\boldsymbol{u}}, \hat{\boldsymbol{v}}\}$ via subspace-based methods for NC signals such as NC Standard ESPRIT or NC Unitary ESPRIT presented in Section 6.2.

In the case of the noise-corrupted augmented measurement matrix $\boldsymbol{X}^{(\mathrm{nc})}$, the 2-D atomic ℓ_1-norm minimization problem can be formulated as

$$\begin{aligned} \min_{\boldsymbol{W}, \boldsymbol{u}, \boldsymbol{v}} \quad & \frac{1}{2} \cdot \mathrm{Tr}\{\boldsymbol{W}\} + \frac{1}{4M} \cdot \mathrm{Tr}\{\mathrm{Toep}\{\boldsymbol{u}, \boldsymbol{v}\}\} && (8.95) \\ \text{s. t.} \quad & \begin{bmatrix} \boldsymbol{W} & \boldsymbol{X}_0^{(\mathrm{nc})^{\mathrm{H}}} \\ \boldsymbol{X}_0^{(\mathrm{nc})} & \mathrm{Toep}\{\boldsymbol{u}, \boldsymbol{v}\} \end{bmatrix} \succeq \boldsymbol{0},\ \left\| \boldsymbol{X}^{(\mathrm{nc})} - \boldsymbol{X}_0^{(\mathrm{nc})} \right\|_{\mathrm{F}}^2 \leq \eta^{(\mathrm{nc})}, \end{aligned}$$

where the threshold parameter $\eta^{(\mathrm{nc})}$ is chosen according to the noise statistics [MCW05], e.g., $\eta^{(\mathrm{nc})} = \mathrm{E}\{\|\boldsymbol{N}^{(\mathrm{nc})}\|_{\mathrm{F}}^2\}$. In the noisy case, the Vandermonde decomposition in (8.93) holds approximately and the spatial frequencies $\tilde{\mu}_k$ can still be extracted from the solution $\mathrm{Toep}\{\hat{\boldsymbol{u}}, \hat{\boldsymbol{v}}\}$ to (8.95) via the NC ESPRIT-type algorithms from Section 6.2.

In the case $N > d$, solving the SDP problem in (8.95) may still require a substantial computational complexity. As shown in Section 8.3.2, the computational cost can be significantly decreased by operating on the signal subspace of $\boldsymbol{X}^{(\mathrm{nc})}$ instead of directly processing $\boldsymbol{X}^{(\mathrm{nc})}$. Thereby, the number of columns of the data can be reduced from N to d.

Following the steps in Section 8.3.2, we first compute the SVD of $\boldsymbol{X}^{(\mathrm{nc})}$ as $\boldsymbol{X}^{(\mathrm{nc})} = \boldsymbol{U} \cdot \boldsymbol{\Sigma} \cdot \boldsymbol{V}^{\mathrm{H}}$. Then, from the temporal covariance matrix

$$\boldsymbol{X}^{(\mathrm{nc})^{\mathrm{H}}} \cdot \boldsymbol{X}^{(\mathrm{nc})} = \boldsymbol{X}^{\mathrm{H}} \cdot \boldsymbol{X} + \boldsymbol{X}^{\mathrm{T}} \cdot \boldsymbol{X}^{*} = 2 \cdot \mathrm{Re}\left\{\boldsymbol{X}^{\mathrm{H}} \cdot \boldsymbol{X}\right\} \in \mathbb{R}^{N \times N},$$

there must be a real-valued unitary basis $\boldsymbol{V} \in \mathbb{R}^{N \times N}$ for the row space of $\boldsymbol{X}^{(\mathrm{nc})}$. Decomposing $\boldsymbol{V}$ into $\boldsymbol{V} = [\boldsymbol{V}_{\mathrm{s}}, \boldsymbol{V}_{\mathrm{n}}]$, where $\boldsymbol{V}_{\mathrm{s}} \in \mathbb{R}^{N \times d}$ contains the dominant right singular vectors of $\boldsymbol{V}$, we can reduce the dimensions of $\boldsymbol{X}_0^{(\mathrm{nc})} \in \mathbb{C}^{2M \times N}$ and $\boldsymbol{X}^{(\mathrm{nc})} \in \mathbb{C}^{2M \times N}$ to

$$\boldsymbol{X}_{\mathrm{sv}_0}^{(\mathrm{nc})} = \boldsymbol{X}_0^{(\mathrm{nc})} \cdot \boldsymbol{V}_{\mathrm{s}} \in \mathbb{C}^{2M \times d}, \quad \boldsymbol{X}_{\mathrm{sv}}^{(\mathrm{nc})} = \boldsymbol{X}^{(\mathrm{nc})} \cdot \boldsymbol{V}_{\mathrm{s}} \in \mathbb{C}^{2M \times d}. \tag{8.96}$$

Subsequently, the reduced dimensional version of the SDP problem in (8.95) can be reformulated as

$$\min_{\boldsymbol{W}, \boldsymbol{u}, \boldsymbol{v}} \quad \frac{1}{2} \cdot \mathrm{Tr}\left\{\boldsymbol{W}\right\} + \frac{1}{4M} \cdot \mathrm{Tr}\left\{\mathrm{Toep}\left\{\boldsymbol{u}, \boldsymbol{v}\right\}\right\} \tag{8.97}$$

$$\text{s. t.} \quad \begin{bmatrix} \boldsymbol{W} & \boldsymbol{X}_{\mathrm{sv}_0}^{(\mathrm{nc})^{\mathrm{H}}} \\ \boldsymbol{X}_{\mathrm{sv}_0}^{(\mathrm{nc})} & \mathrm{Toep}\left\{\boldsymbol{u}, \boldsymbol{v}\right\} \end{bmatrix} \succeq \boldsymbol{0}, \ \left\| \boldsymbol{X}_{\mathrm{sv}}^{(\mathrm{nc})} - \boldsymbol{X}_{\mathrm{sv}_0}^{(\mathrm{nc})} \right\|_{\mathrm{F}}^2 \leq \eta_{\mathrm{sv}}^{(\mathrm{nc})}, \tag{8.98}$$

whose solution can be obtained at a significantly lower computational complexity. The threshold parameter $\eta_{\mathrm{sv}}^{(\mathrm{nc})}$ is chosen according to (8.68) with the same reasoning.

Note that it was shown in [CC15] that the frequency separation condition for exact frequency recovery from the 1-D case in [CFG14] also holds in the 2-D case and therefore applies here as well. However, as demonstrated in the simulation section, in the specific case of two uncorrelated NC sources with maximum phase separation, these sources entirely decouple such that the resolution limit does not apply in this case.

An important feature of the presented NC ANM approach is that it can resolve more sources than the number of physical sensors. This is due to the NC preprocessing, which virtually doubles the number of sensor elements. This property is shown via simulation results in Section 8.4.3.

8.4. Numerical results

In this section, we provide simulation results to demonstrate the performance of the three sparse signal reconstruction (SSR) algorithms for strictly non-circular (NC) sources. In Section 8.4.1, we assess the performance of the NC SSR algorithm based on $\ell_{2,1}$-mixed-norm minimization, whereas

the performance of the NC SSR algorithm based on nuclear norm minimization is analyzed in Section 8.4.2. The gridless NC SSR algorithm based on atomic norm minimization is considered in Section 8.4.3 and a comparison of the computational complexity of the proposed NC SSR algorithms via the computation time is conducted in Section 8.4.4.

8.4.1. NC SSR based mixed-norm minimization

In this section, we present simulations to assess the performance of the proposed SSR algorithm for strictly non-circular sources. To this end, we use the OMP algorithm for the SSR step and compare the proposed offset estimation scheme "NC OMP Joint" according to (8.45) to its non-NC counterpart from [IRA+14]. For the stopping criteria of OMP, the sparsity level d is assumed known. Note that the OMP method can be replaced by any other SSR algorithm. We also consider the deterministic CRB "Det CRB" as well as the deterministic CRB for NC sources "Det NC CRB" derived in Section 9.2. For the computation of the mean square error (MSE), we only take the estimation error in the spatial frequency domain μ into account as the estimation of the rotation phases is not of primary interest, but can be added straightforwardly. For the numerical results, we adopt a ULA of $M = 8$ isotropic sensors with half-wavelength spacing. The phase reference is at the centroid. The symbols are drawn from a real-valued Gaussian distribution and the noise is circularly symmetric white complex Gaussian with $\sigma_{\mathrm{n}}^2 = 1/\mathrm{SNR}$. We have used 300 Monte Carlo trials.

In Figure 8.2, we display the MSE versus the SNR for a scenario, where $P_\mu = 8$, $P_\varphi = 6$, and $d = 2$ uncorrelated sources are located at $(15.1\Delta_\mu, 10.2\Delta_\varphi)$ and $(17.5\Delta_\mu, 34.2\Delta_\varphi)$, respectively, i.e., we have $\Delta\mu = \mu_2 - \mu_1 = 0.3P_\mu\Delta_\mu$ and $\Delta\varphi = \varphi_2 - \varphi_1 = 24\Delta_\varphi = \pi/2$. The number of snapshots is $N = 20$. Note that in such a setting of two closely-spaced uncorrelated NC sources with a phase discrimination of $\pi/2$, the maximum NC gain can be achieved as shown in Section 9.4.2. It can be seen from Figure 8.2 that the proposed algorithm for NC sources provides a significantly lower estimation error compared to its non-NC counterpart. The "NC OMP Joint" algorithm successfully estimates the grid offset and achieves the deterministic NC CRB.

Figure 8.3 illustrates the MSE versus the spatial separation $\Delta\mu$, where we have $d = 2$ uncorrelated sources at $\mu_1 = 20.2\Delta_\mu$ and $\mu_2 = \mu_1 + \Delta\mu$. The SNR is fixed to 40 dB and the remaining parameters are kept the same. We observe again that the NC scheme outperforms its counterpart as it is constant for all the distances.

In Figure 8.4, we show the MSE versus the rotation phase separation $\Delta\varphi$ for $d = 2$ sources at $\boldsymbol{\mu} = [2.1, 4.5]\Delta_\mu$ with $\varphi_1 = 5.1\Delta_\varphi$ and $\varphi_2 = \varphi_1 + \Delta\varphi$. All the parameters are kept the same as before. It can be seen that the NC method provides the best performance for a phase discrimination of $\Delta\mu = \pi/2$ while its non-NC counterpart remains constant.

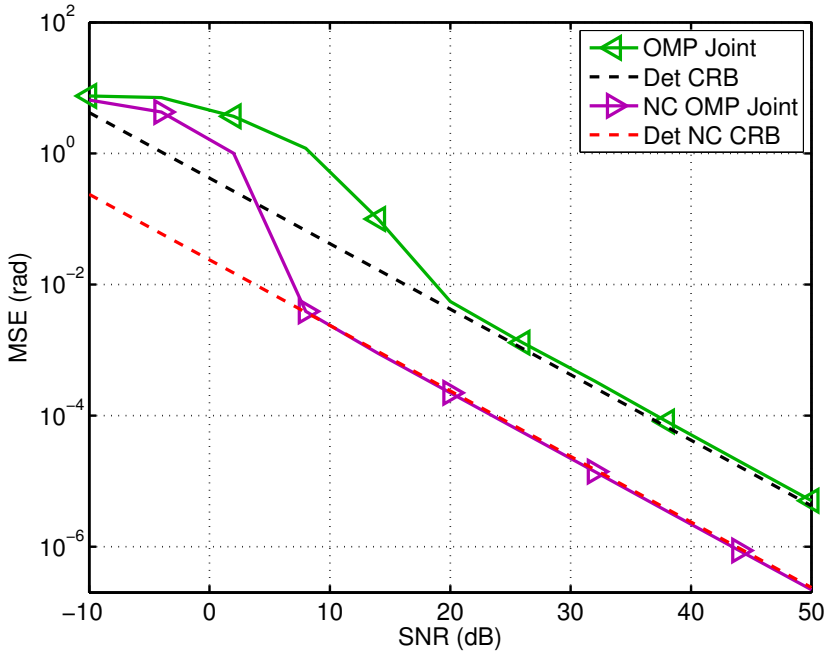


Figure 8.2.: MSE versus SNR for $M = 8$, $P_\mu = 8$, $P_\varphi = 6$, $N = 20$, $d = 2$ sources at $\boldsymbol{\mu} = [15.1, 17.5]\Delta_\mu$ with $\boldsymbol{\varphi} = [10.2, 34.2]\Delta_\varphi$.

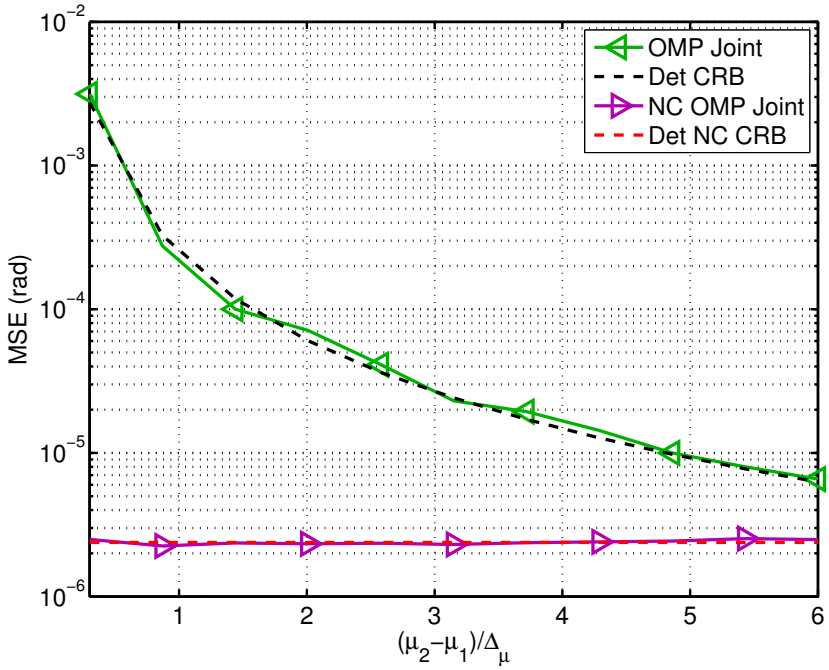


Figure 8.3.: MSE versus $\Delta\mu$ for $M = 8$, $P_\mu = 8$, $P_\varphi = 6$, $N = 20$, $d = 2$ sources at $\mu_1 = 20.2\Delta_\mu$ and $\mu_2 = \mu_1 + \Delta\mu$.

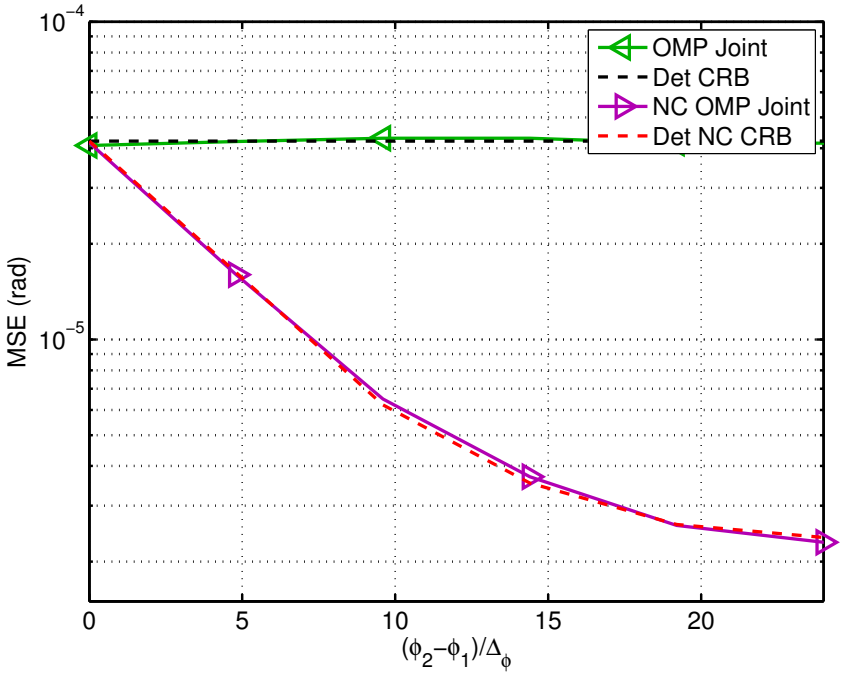


Figure 8.4.: MSE versus $\Delta\varphi$ for $M = 8$, $P_\mu = 8$, $P_\varphi = 6$, $N = 20$, $d = 2$ sources at $\boldsymbol{\mu} = [2.1, 4.5]\Delta_\mu$ with $\varphi_1 = 5.1\Delta_\varphi$ and $\varphi_2 = \varphi_1 + \Delta\varphi$.

8.4.2. NC SSR based nuclear norm minimization

This section provides simulation results to demonstrate the performance of the proposed NC SSR method that exploits the NC structure via nuclear norm minimization. Specifically, we compare the SSR solution of (8.67) combined with the proposed offset estimator (8.77) termed "NC NUC Joint" to its corresponding non-NC version (8.6) in combination with the offset estimator [IRA+14] termed "L1-SVD Joint". Moreover, we include the more complex NC SSR algorithm "NC L1-SVD Joint" from (8.26), Standard ESPRIT "SE", NC Standard ESPRIT "NC SE" [ZCW03], the deterministic Cramér-Rao bound (CRB) "Det CRB" [SN89], and the deterministic NC CRB "Det NC CRB" derived in Section 9.2. The regularization parameters for the SSR methods are chosen according to (8.68) for $\gamma = 0.99$. The mean square error (MSE) is only computed based on the spatial frequency estimates. The simulation setup consists of a ULA with $M = 8$ isotropic sensors half-wavelength spacing apart, where the phase reference is at the array centroid. We assume $d = 2$ sources that transmit symbols drawn from a real-valued Gaussian distribution, while the sensor noise is circularly symmetric white complex Gaussian. 1000 Monte Carlo trials have been used to generate the plots.

In Figure 8.5, the MSE versus the SNR is depicted for $P_\mu = 8$, $N = 10$, and the $d = 2$ uncorrelated sources are located at $\mu_1 = 15.3\Delta_\mu$ and $\mu_2 = 17.7\Delta_\mu$. As NC L1-SVD Joint in (8.26) requires a rotation phase grid, we choose a uniform grid spacing defined by $\Delta_\varphi = \pi/MP_\varphi$. Then, the

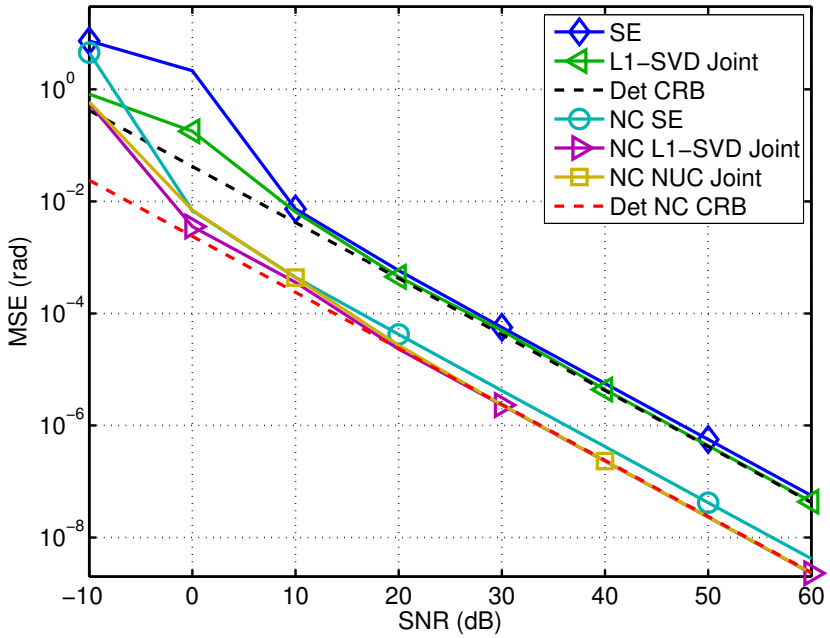


Figure 8.5.: MSE versus SNR for $M = 8$, $P_\mu = 8$, $N = 10$, $d = 2$ sources at $\mu_1 = 15.3\Delta_\mu$, $\mu_2 = 17.7\Delta_\mu$ with $\varphi_1 = 5.1\Delta_\varphi$, $\varphi_2 = 29.1\Delta_\varphi$ ($\Delta\varphi = \pi/2$) for $P_\varphi = 6$.

rotation phases are given by $\varphi_1 = 5.1\Delta_\varphi$ and $\varphi_2 = 29.1\Delta_\varphi$ with $P_\varphi = 6$. We emphasize again that the proposed algorithm NC NUC Joint does not require a rotation phase grid. Note that $\Delta\varphi = \varphi_2 - \varphi_1 = 24\Delta_\varphi = \pi/2$, which provides the maximum NC gain. It is apparent that NC NUC Joint and NC L1-SVD Joint both achieve the Det NC CRB and clearly outperform the L1-SVD Joint algorithm that does not exploit the NC structure.

Figure 8.6 shows the MSE versus the SNR for a different scenario, where $N = 5$ and $d = 2$ sources are positioned at $\mu_1 = 14.7\Delta_\mu$ and $\mu_2 = 17.1\Delta_\mu$ with $\varphi_1 = 10.3\Delta_\varphi$ and $\varphi_2 = 22.3\Delta_\varphi$ such that $\Delta\varphi = 12\Delta_\varphi = \pi/4$. The remaining parameters are kept the same. Again, the same behavior of the algorithms can be observed. NC NUC Joint and NC L1-SVD Joint perform identical and provide a clear gain over the L1-SVD Joint method.

8.4.3. NC SSR based atomic norm minimization

In this section, we present simulation results that demonstrate the performance of the NC ANM algorithm combined with NC Standard ESPRIT "NC ANM SE" in (8.97) in comparison to the NC SSR algorithm "NC L1-SVD Joint" from (8.26), and the NC SSR algorithm "NC NUC Joint" from (8.67) using the offset estimator (8.77). Moreover, we include the non-NC counterpart ANM with Standard ESPRIT "ANM SE" as well as the spatially smoothed versions of NC Standard ESPRIT "NC SE + SS" from Section 7.2.4 and Standard ESPRIT "SE + SS" using least squares. The algorithms are benchmarked by the deterministic Cramér-Rao bound (Det CRB) [SN89] and

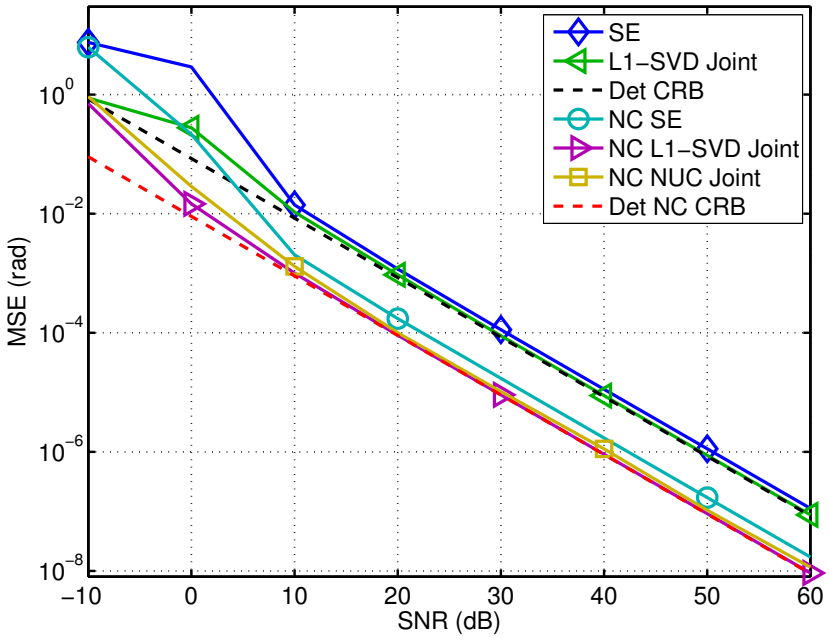


Figure 8.6.: MSE versus SNR for $M = 8$, $P_\mu = 8$, $N = 5$, $d = 2$ sources at $\mu_1 = 14.7\Delta_\mu$, $\mu_2 = 17.1\Delta_\mu$ with $\varphi_1 = 10.3\Delta_\varphi$, $\varphi_2 = 22.3\Delta_\varphi$ ($\Delta\varphi = \pi/4$) for $P_\varphi = 6$.

the deterministic NC CRB "Det NC CRB" derived in Section 9.2.

In the first experiment, we assume a uniform linear array (ULA) with $M = 12$ isotropic sensors. The phase reference of the array is located at the array centroid. The received signals are uncorrelated with unit power and their transmitted symbols are drawn from a real-valued Gaussian distribution. Moreover, we assume zero-mean circularly symmetric white sensor noise. The number of subarrays L for spatial smoothing is $L = 4$. The regularization parameters are chosen according to (8.68) for $\gamma = 0.99$. The curves are obtained by averaging over 1000 Monte Carlo trials. For the grid-based methods NC L1-SVD Joint and NC NUC Joint, we keep the parameters from the previous scenario.

Figure 8.7 illustrates the RMSE as a function of the SNR for $d = 2$ sources from the directions $\mu_1 = 1$ and $\mu_2 = 1.1$ with the rotation phases $\varphi_1 = 0$ and $\varphi_2 = \pi/2$. We assume $N = 5$ snapshots. It can be seen that while ANM SE produces a bias due to the violated resolution limit as $\Delta\mu = |\mu_2 - \mu_1| = 0.1$, the proposed NC ANM SE algorithm does not suffer from the resolution limit due to the entire decoupling of the two sources at $\Delta\varphi = |\varphi_2 - \varphi_1| = \pi/2$. Moreover, all the NC SSR methods achieve the Det NC CRB and have a superior performance over the NC SE + SS algorithm for higher SNRs.

In Figure 8.8, we analyze the RMSE as a function of the phase separation $\Delta\varphi = |\varphi_2 - \varphi_1|$. We again consider $d = 2$ sources located at $\mu_1 = 0.5$ and $\mu_2 = 0.9$. We have $N = 5$ snapshots and the SNR is set to SNR = 20 dB. It is apparent from Figure 8.8 that the RMSE of the NC algorithms reduces

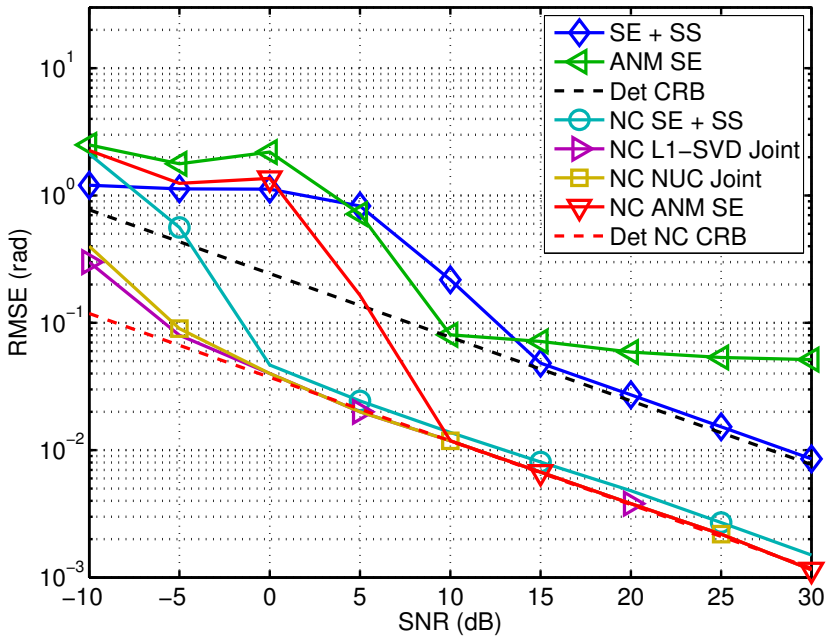


Figure 8.7.: RMSE versus the SNR for $d = 2$ at $\mu_1 = 1$ and $\mu_1 = 1.1$ with $M = 12$, $N = 5$, $L = 4$, and $\varphi_1 = 0$ and $\varphi_2 = \pi/2$.

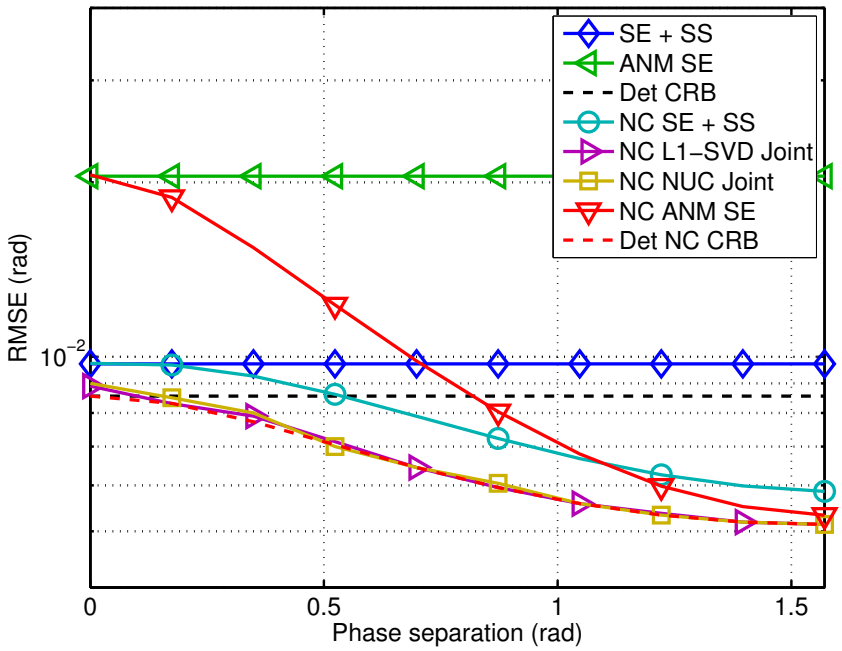


Figure 8.8.: RMSE versus the phase separation $\Delta\varphi = |\varphi_2 - \varphi_1|$ for $d = 2$ at $\mu_1 = 0.5$ and $\mu_1 = 0.9$ with $M = 12$, $N = 5$, $L = 4$, and SNR $= 20$ dB.

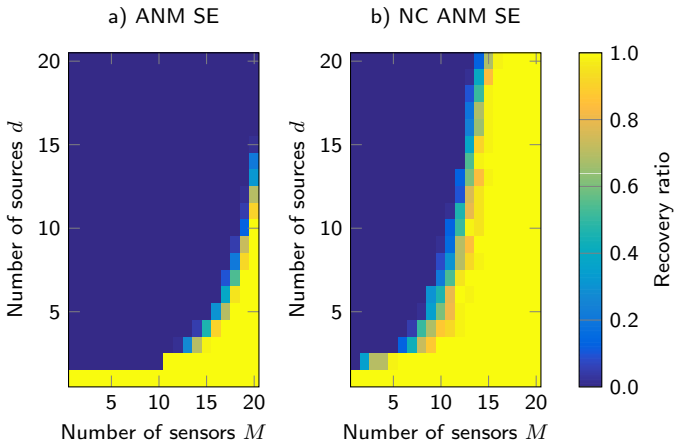


Figure 8.9.: Frequency recovery ratio of a) ANM with Standard ESPRIT and b) NC ANM with NC Standard ESPRIT.

for an increasing phase separation while the RMSE of the non-NC versions remains constant. The reason for this behavior in the NC case is again the spatial decorrelation of two closely-spaced sources due to the phase discrimination. While NC L1-SVD Joint and NC NUC Joint reach the Det NC CRB for the entire phase separation range, NC ANM SE only achieves the Det NC CRB close to $\Delta\varphi = \pi/2$, where the decoupling of the sources compensates the resolution limit.

In the second experiment, we compare the recovery performance of the proposed NC ANM SE algorithm to its non-NC counterpart ANM SE in the noise-free case, i.e., $\sigma_{\mathrm{n}}^2 = 0$. We assume $N = 10$ snapshots and vary the number of sensors M and the number of source signals d. The spatial frequency of the i-th source is given as $\mu_i = i \cdot \pi/10$ while its rotation phase is given as $\varphi_i = i \cdot \pi/2$, for $i = 1, \ldots, d$. Figure 8.9 shows the ratio of successful frequency recovery for 100 Monte Carlo trials with real-valued uncorrelated Gaussian source signals with the covariance matrix $\mathbb{E}\{\boldsymbol{S}_0 \cdot \boldsymbol{S}_0^{\mathrm{T}}\} = N \cdot \boldsymbol{I}_d$. From Figure 8.9a), it can be seen that ANM SE cannot recover the frequencies in most of the scenarios, which is due to the close frequency spacing of $\Delta\mu = 0.314$. On the other hand, Figure 8.9b) shows a significantly improved recovery ratio for the proposed NC ANM SE algorithm, which is caused by the optimal rotation phase difference of $\Delta\varphi = \pi/2$ between any two adjacent source signals. Furthermore, the plot demonstrates well that NC ANM SE admits the identifiability of more signals d than sensors M, e.g., $d = 20$ source signals can be resolved with $M = 16$ sensors.

8.4.4. Computational complexity of the NC SSR algorithms

In this section, we compare the computational complexity of the NC SSR based $\ell_{2,1}$-mixed-norm minimization "NC L1-SVD Joint" in (8.26), the NC SSR based nuclear norm minimization "NC

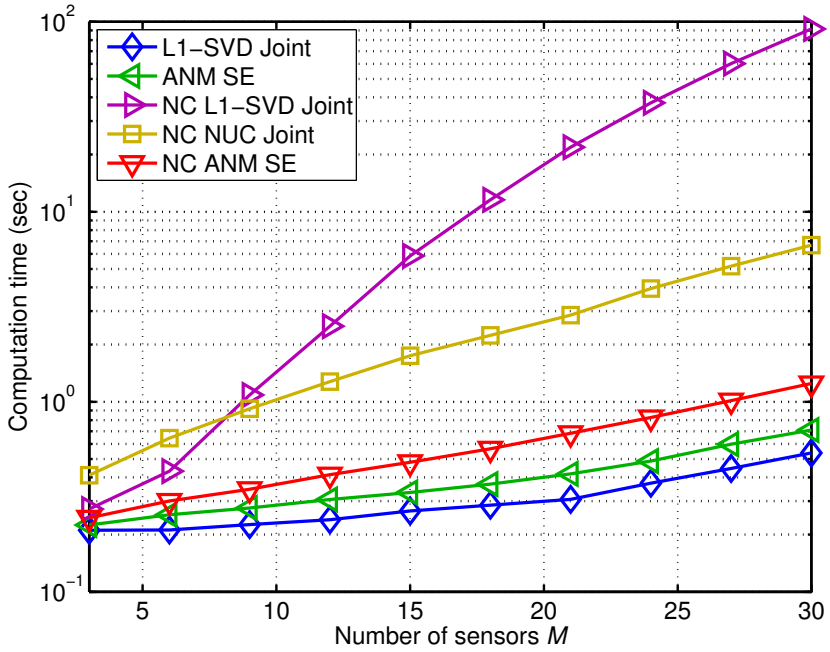


Figure 8.10.: Computation time versus M for $P_\mu = 8$, $N = 10$, $d = 2$ sources at $\mu_1 = 15.3\Delta_\mu$, $\mu_2 = 17.7\Delta_\mu$ with $\varphi_1 = 5.1\Delta_\varphi$, $\varphi_2 = 29.1\Delta_\varphi$ ($\Delta\varphi = \pi/2$) for $P_\varphi = 6$ and SNR $= 10$ dB.

NUC Joint" in (8.67), and the NC SSR based atomic norm minimization "NC ANM SE" using NC Standard ESPRIT from (8.97) by means of the average computation time in Matlab using an Intel Core i7-7700 CPU with 3,60 GHz and 16 GByte RAM. To reduce the computational cost of the considered NC SSR algorithms, we perform the corresponding dimensionality reduction procedures discussed in the Sections 8.3.1, 8.3.2, and 8.3.3. As a reference, we additionally consider the corresponding non-NC mixed-norm version "L1-SVD Joint" in (8.6) and the non-NC counterpart ANM from (8.16) with Standard ESPRIT "ANM SE".

The respective optimization problems are solved using the CVX framework [GB14]. For the simulation, we assume that $d = 2$ sources located at $\mu_1 = 15.3\Delta_\mu$ and $\mu_2 = 17.7\Delta_\mu$ with rotation phases $\varphi_1 = 5.1\Delta_\varphi$ and $\varphi_2 = 29.1\Delta_\varphi$, where $\Delta\varphi = |\varphi_2 - \varphi_1| = \pi/2$, impinge on a ULA, where the number of sensors M is varied. We have $P_\mu = 8$ and $P_\varphi = 6$, the SNR is fixed at 10 dB, and $N = 10$ snapshots are available.

Figure 8.10 displays the average computation time over 100 trials for the considered NC SSR algorithms. It is evident that the "NC L1-SVD Joint" algorithm requires the longest computation time, which is due to the grid-based 2-D support estimation. Moreover, the gridless "NC ANM SE" algorithm outperforms the "NC NUC Joint" algorithm, which requires the grid-based 1-D support estimation. Note that due to the additional NC preprocessing step, all the NC SSR based algorithms require a longer computation time compared to their non-NC counterparts.

8.5. Summary

In this chapter, we have presented three sparse signal reconstruction (SSR) algorithms for the multiple measurement vector (MMV) case that exploit the specific structure of strictly non-circular (NC) signals. The developed algorithms demonstrate that upon applying the NC preprocessing step, the performance benefits associated with NC signals in subspace-based algorithms, i.e., an improved estimation accuracy and a doubled number of resolvable signals, can also be achieved by sparse recovery algorithms.

The first NC SSR algorithm is based on $\ell_{2,1}$-mixed norm minimization using a 2-D finite dictionary to estimate the support in the spatial domain as well as in the rotation phase domain. In order to handle the resulting 2-D off-grid problem, we propose a 2-D grid-offset estimation procedure for a single source and two sources by means of local interpolation.

The second NC SSR algorithm is based on nuclear norm minimization after lifting the original optimization problem to a semi-definite programming (SDP) problem in a higher-dimensional space. Thereby, the 2-D estimation problem is reduced to a 1-D estimation problem only in the sampled spatial domain, which automatically provides gridless estimates of the rotation phases. As a result, the NC SSR algorithm is based on nuclear norm minimization requires a significantly lower computational complexity than the first algorithm, while providing the same performance benefits. Additionally, we present a 1-D grid-offset estimation procedure for a single source and two sources.

Finally, the third algorithm is a gridless sparse recovery algorithm for NC signals based on the atomic norm minimization (ANM). The ANM-equivalent SDP formulation provides a solution matrix with a two-level Hermitian Toeplitz structure. We show that by using the multi-dimensional extension of the Vandermonde decomposition, the desired direction estimates can be uniquely extracted from the two-level Hermitian Toeplitz matrix via NC ESPRIT-type algorithms in closed-form.

Numerical simulations have shown that all three NC SSR algorithms provide a superior estimation accuracy over their counterparts for arbitrary signals. Moreover, via an analysis of the computation time, we have found that the NC ANM algorithm requires the shortest computation time while the NC algorithm based on mixed-norm minimization requires the longest computation time.

9. Deterministic Cramér-Rao bound for strictly non-circular sources

In Chapter 6, we have developed matrix-based and tensor-based R-D NC ESPRIT-type algorithms to exploit the structure of strictly non-circular signals and provide an analytical performance framework to evaluate their performance in various scenarios. In Chapter 7, we have additionally incorporated spatial smoothing into the presented R-D NC ESPRIT-type algorithms and present the corresponding performance analysis to assess the performance in the case of coherent sources or a small number of snapshots. In this chapter, we derive a deterministic Cramér-Rao bound (CRB) for strictly non-circular (NC) R-D signals, termed deterministic R-D NC CRB. The deterministic data model assumes the signals to be a deterministic (non-random) and unknown sequence. This bound represents a lower limit on the achievable estimation error of any unbiased estimator and therefore provides a benchmark for the previously developed R-D NC ESPRIT-type algorithms. In Section 9.1, we give a brief literature review and summarize the contributions. The derivation of the deterministic R-D NC CRB is provided in Section 9.2 while its analysis and its comparison to the deterministic R-D CRB for arbitrary signals is presented in Section 9.3. The simplification for a single NC source and the NC gain for two closely-spaced sources are derived in Section 9.4.2. Section 9.5 illustrates the numerical results and summarizes the main properties, and concluding remarks are drawn in Section 9.6.

9.1. Overview

In this section, we provide a state of the art of Cramér-Rao bounds (CRB) for arbitrary signals and the existing CRBs for non-circular signals in Section 9.1.1. The specific contributions of this chapter are outlined in Section 9.1.2.

9.1.1. State of the art

The problem of estimating the parameters of R-D signals with $R \geq 1$, such as their directions of arrival, directions of departure, frequencies, and Doppler shifts, has been an extensive research area for a long time. Recently, various high-resolution parameter estimation algorithms such as NC MUSIC [AD06], NC Root-MUSIC [CWS01], NC Standard ESPRIT [ZCW03], and NC Unitary ESPRIT [HR04] and the matrix-based and tensor-based R-D NC ESPRIT-type algorithms presented in Chapter 6 have been developed to exploit the structure of strictly second-order (SO) non-circular (NC) signals introduced in Section 2.2. The term strictly SO NC (also termed rectilinear) is based on the fact that the non-circularity coefficient $|\zeta|$ of these signals is equal to

one (cf. Section 2.2.1.2) and [SS10]). The aforementioned NC algorithms that exploit the non-circularity property are known to achieve a higher estimation accuracy and can resolve up to twice as many sources [SRHD14] compared to the traditional methods for arbitrary signals [KV96]. In Section 6.4 and in Section 6.5, we have presented a first-order performance analysis framework for R-D NC ESPRIT-type algorithms and R-D NC Tensor-ESPRIT-type algorithms. Moreover, we have simplified the analytical MSE expressions for the special cases of a single NC source and two NC sources in Section 6.6. The resulting expressions only depend on the physical parameters of significance, e.g., the number of sensors M, the SNR, and the number of snapshots N. For a single NC source, we have found that no NC gain can be achieved in this case. For the special case of two NC sources, we have analytically computed the NC gain and analyzed its behavior as a function of the signal correlation and the signals' rotation phase separation. However, the obtained analytical expressions for the NC gain are specific to the matrix-based and tensor-based R-D NC ESPRIT-type algorithms.

The performance of high-resolution parameter estimation algorithms is often evaluated by comparing them to the Cramér-Rao bound (CRB), which provides a lower limit on the estimation error of any unbiased estimator. We can distinguish between the deterministic (conditional) and stochastic (unconditional) Cramér-Rao bounds (CRBs) derived in [SN89] and [SLG01], respectively. Whereas the stochastic data assumption requires both the signals and the noise to be complex Gaussian-distributed, the deterministic model assumes that the signals are arbitrary non-random sequences while only the noise follows a complex Gaussian distribution. Both CRBs have been of high relevance in the literature [SN90]. However, the deterministic CRB is easier to derive and still provides valuable engineering insight. For the data model used to describe weak-sense non-circular sources whose non-circularity coefficient $|\zeta|$ satisfies $0 \leq |\zeta| \leq 1$ (cf. Section 2.2.1.2), a stochastic NC CRB has been derived in [DA04]. The follow-up papers [AD05] and [DA06] consider further variations of the underlying stochastic model assumption. The stochastic NC CRB in [DA04] was derived by extending the original Slepian-Bangs formula for circular complex Gaussian distributions [SM05] to non-circular complex Gaussian distributions. Although this bound was derived for weak-sense non-circular signals with $0 \leq |\zeta| \leq 1$, it does not take the specific NC structure of strictly non-circular signals into account and therefore, this prior knowledge is not exploited. Specifically, in the weak-sense case [DA04], the real part and the imaginary part of the signals can be treated as independent random variables. However, this is not true for strictly non-circular signals, where the real and imaginary parts are linearly dependent as shown in Section 2.2.1.2. Hence, the parameter vector used to derive the CRB is different in both cases.

The special case of an arbitrary single source has been considered for the deterministic CRB in [SN89]. In [DA04], the authors have also derived an expression of the stochastic NC CRB for a single strictly non-circular source, which is discussed in more detail in Section 9.4.1.

9.1.2. Contributions

In this chapter, we derive a closed-form expression of the deterministic R-D NC CRB that exploits the prior knowledge that strictly non-circular signals impinge on an arbitrary R-D sensor array. The results are published in [SRHD16]. We resort to the deterministic case of the NC CRB as the NC signal model is a deterministic model, i.e., the stationary rotation phases are deterministic parameters. This allows to systematically study the influence of these parameters on the behavior of the bound. The derivation is based on the conventional Slepian-Bangs formula, which is still applicable due to the complex Gaussian noise assumption. Note that our initial contributions in [RH07a] and [SRHD14] only state the final R-D NC CRB result along with some special cases. However, therein, no proofs and further analysis are provided. Based on the devised R-D NC CRB and assuming the R-D array to be separable and centro-symmetric, we show that in the special cases of equal rotation phases or full coherence of all strictly non-circular signals (or a single snapshot) or for a single strictly non-circular source, the deterministic R-D NC CRB reduces to the deterministic R-D CRB for arbitrary signals in [SN89]. This suggests that no NC gain from strictly non-circular sources can be achieved under the deterministic data assumption in these special cases. Note that the single source case of the R-D NC CRB has already been analyzed by us in [SRHD14] for a uniform R-D array that contains a uniform linear array (ULA) in each mode. Here, we provide a generalization of this case to arbitrarily-formed (non-uniform) separable and centro-symmetric R-D arrays. Furthermore, the fact that twice as many sources can be resolved from the strictly non-circular data model is highlighted.

Furthermore, we assume 1-D parameter estimation and simplify the derived deterministic NC CRB and the deterministic CRB for the special case of two closely-spaced strictly non-circular sources captured by a uniform linear array (ULA). These simplified expressions are subsequently used to analytically compute the maximum achievable NC gain, which only depends on the physical parameters, e.g., the number of sensors, the SNR, the correlation, the phase separation, and the location of the phase reference of the array. The devised expression is based on a truncated Taylor series expansion for closely-spaced sources. This is, however, the scenario, where high-resolution algorithms are primarily applied. Due to the fact that the NC gain expression is very general, the properties of the NC gain are studied in terms of the above-mentioned physical parameters. For instance, it is shown that the NC gain is largest if the sources are uncorrelated, the phase separation is maximum, and the phase reference is at the array centroid. Under these conditions, the two sources entirely decouple and do not influence each other.

9.2. R-D deterministic NC CRB for strictly non-circular sources

In this section, we first review the R-D CRB for arbitrary multi-dimensional signals in Section 9.2.1 and then derive the R-D NC CRB for multi-dimensional strictly non-circular signals

in Section 9.2.2. Additionally, we provide simplified expressions of the respective CRBs for the 1-D parameter estimation case.

For the derivation, let us first recall the matrix-based R-D data model in Section 2.1.2. Following (2.3), the measurement matrix $\boldsymbol{X} \in \mathbb{C}^{M \times N}$ can be modeled as

$$\boldsymbol{X} = \boldsymbol{A} \cdot \boldsymbol{S} + \boldsymbol{N}, \tag{9.1}$$

where $\boldsymbol{S} \in \mathbb{C}^{d \times N}$ is the source symbol matrix and $\boldsymbol{N} \in \mathbb{C}^{M \times N}$ contains the noise samples. Furthermore, $\boldsymbol{A} = [\boldsymbol{a}(\boldsymbol{\mu}_1), \ldots, \boldsymbol{a}(\boldsymbol{\mu}_d)] \in \mathbb{C}^{M \times d}$ is referred to as the array steering matrix, which consists of the array steering vectors $\boldsymbol{a}(\boldsymbol{\mu}_i)$ defined by

$$\boldsymbol{a}(\boldsymbol{\mu}_i) = \boldsymbol{a}^{(1)}\left(\mu_i^{(1)}\right) \otimes \cdots \otimes \boldsymbol{a}^{(R)}\left(\mu_i^{(R)}\right) \in \mathbb{C}^{M \times 1}, \tag{9.2}$$

where $\boldsymbol{a}^{(r)}(\mu_i^{(r)}) \in \mathbb{C}^{M_r \times 1}$ is the array steering vector of the i-th spatial frequency in the r-th mode. An alternative expression of $\boldsymbol{A}$ is given by

$$\boldsymbol{A} = \boldsymbol{A}^{(1)} \diamond \boldsymbol{A}^{(2)} \diamond \cdots \diamond \boldsymbol{A}^{(R)}, \tag{9.3}$$

where $\boldsymbol{A}^{(r)} = [\boldsymbol{a}^{(r)}(\mu_1^{(r)}), \ldots, \boldsymbol{a}^{(r)}(\mu_d^{(r)})] \in \mathbb{C}^{M_r \times d}$ represents the array steering matrix in the r-th mode.

In the case of strictly non-circular signals, which is introduced in Section 2.2, the complex symbol amplitudes of each source lie on a rotated line in the complex plane. Therefore, according to (2.36), the symbol matrix $\boldsymbol{S}$ can be decomposed as

$$\boldsymbol{S} = \boldsymbol{\Psi} \cdot \boldsymbol{S}_0, \tag{9.4}$$

where $\boldsymbol{S}_0 \in \mathbb{R}^{d \times N}$ is a real-valued symbol matrix and $\boldsymbol{\Psi} = \operatorname{diag}\left\{\mathrm{e}^{\mathrm{j}\varphi_i}\right\}_{i=1}^{d}$ contains stationary complex phase shifts on its diagonal that can be different for each source.

Then, using (9.4), the model in (9.1) can be written as

$$\boldsymbol{X} = \boldsymbol{A} \cdot \boldsymbol{\Psi} \cdot \boldsymbol{S}_0 + \boldsymbol{N}. \tag{9.5}$$

9.2.1. Deterministic R-D Cramér-Rao bound

In the case of arbitrary signals, the set of parameters that needs to be considered for the deterministic R-D CRB is given by the angular parameters $\boldsymbol{\mu} = [\boldsymbol{\mu}^{(1)^{\mathrm{T}}}, \ldots, \boldsymbol{\mu}^{(R)^{\mathrm{T}}}]^{\mathrm{T}} \in \mathbb{R}^{Rd \times 1}$, the real part and the imaginary part of the symbols $\boldsymbol{s} = \operatorname{vec}\{\boldsymbol{S}\} \in \mathbb{C}^{Nd \times 1}$, and the noise power σ_{n}^2. For this parameter set that contains a total of $(2N + R)d + 1$ parameters, the deterministic CRB matrix in

the R-D parameter estimation case was derived in [SN89]. Its closed-form expression is given by

$$\boldsymbol{C} = \frac{\sigma_{\mathrm{n}}^2}{2N} \cdot \mathrm{Re}\left\{\left(\boldsymbol{D}^{\mathrm{H}} \cdot \boldsymbol{\Pi}_{\boldsymbol{A}}^{\perp} \cdot \boldsymbol{D}\right) \odot \hat{\boldsymbol{R}}_S^{(R)^{\mathrm{T}}}\right\}^{-1} \in \mathbb{R}^{Rd \times Rd}, \tag{9.6}$$

where

$$\boldsymbol{\Pi}_{\boldsymbol{A}}^{\perp} = \boldsymbol{I}_M - \boldsymbol{A} \cdot \left(\boldsymbol{A}^{\mathrm{H}} \cdot \boldsymbol{A}\right)^{-1} \cdot \boldsymbol{A}^{\mathrm{H}} \in \mathbb{C}^{M \times M} \tag{9.7}$$

and

$$\boldsymbol{D} = \left[\boldsymbol{D}^{(1)} \quad \dots \quad \boldsymbol{D}^{(R)}\right] \in \mathbb{C}^{M \times Rd} \tag{9.8}$$

with $\boldsymbol{D}^{(r)} = [\boldsymbol{d}_1^{(r)}, \dots, \boldsymbol{d}_d^{(r)}] \in \mathbb{C}^{M \times d},\ r = 1, \dots, R$, contains the partial derivatives of $\boldsymbol{A}$ with respect to the components of $\boldsymbol{\mu}_i,\ i = 1, \dots, d$, in the r-th mode. The vectors $\boldsymbol{d}_i^{(r)}$ are given by $\boldsymbol{d}_i^{(r)} = \partial \boldsymbol{a}(\boldsymbol{\mu}_i)/\partial \mu_i^{(r)},\ \forall i$. Writing $\boldsymbol{a}_i$ instead of $\boldsymbol{a}(\boldsymbol{\mu}_i)$ to simplify the notation and using (9.2), we obtain

$$\boldsymbol{d}_i^{(r)} = \boldsymbol{a}_i^{(1)} \otimes \cdots \otimes \boldsymbol{a}_i^{(r-1)} \otimes \tilde{\boldsymbol{d}}_i^{(r)} \otimes \boldsymbol{a}_i^{(r+1)} \otimes \cdots \otimes \boldsymbol{a}_i^{(R)}, \tag{9.9}$$

where $\tilde{\boldsymbol{d}}_i^{(r)} = \partial \boldsymbol{a}_i^{(r)}/\partial \mu_i^{(r)}$. Moreover, $\hat{\boldsymbol{R}}_S^{(R)} = \boldsymbol{1}_{R \times R} \otimes \hat{\boldsymbol{R}}_S$ contains the estimated signal covariance matrix $\hat{\boldsymbol{R}}_S = \boldsymbol{S} \cdot \boldsymbol{S}^{\mathrm{H}}/N$. Note that $\hat{\boldsymbol{R}}_S$ can be written in matrix form as

$$\hat{\boldsymbol{R}}_S = \begin{bmatrix} \hat{P}_1 & \hat{\rho}_{1,2}\sqrt{\hat{P}_1\hat{P}_2} & \dots & \hat{\rho}_{1,d}\sqrt{\hat{P}_1\hat{P}_d} \\ \hat{\rho}_{2,1}\sqrt{\hat{P}_1\hat{P}_2} & \hat{P}_2 & \dots & \hat{\rho}_{2,d}\sqrt{\hat{P}_2\hat{P}_d} \\ \vdots & \vdots & \ddots & \vdots \\ \hat{\rho}_{d,1}\sqrt{\hat{P}_1\hat{P}_d} & \hat{\rho}_{d,2}\sqrt{\hat{P}_2\hat{P}_d} & \dots & \hat{P}_d \end{bmatrix},$$

where $\hat{P}_i = \|\boldsymbol{s}_i\|_2^2/N$ is the empirical source power of the i-th source and $\boldsymbol{s}_i^{\mathrm{T}} \in \mathbb{R}^{1 \times N}$ is the i-th row of $\boldsymbol{S}$. Furthermore, the empirical correlation coefficients $\hat{\rho}_{i,j},\ \forall\ i \neq j,\ i, j = 1, \dots, d$ that represent the empirical correlation between the i-th and the j-th source vector are defined by

$$\hat{\rho}_{i,j} = |\hat{\rho}_{i,j}|\, \mathrm{e}^{\mathrm{j}\hat{\varphi}_{c_{i,j}}} = \frac{1}{N} \cdot \frac{\boldsymbol{s}_i^{\mathrm{H}} \cdot \boldsymbol{s}_j}{\sqrt{\hat{P}_i\hat{P}_j}},\quad \forall\ i \neq j, \tag{9.10}$$

where $|\hat{\rho}_{i,j}|$ is the magnitude and $\hat{\varphi}_{c_{i,j}}$ is the empirical correlation phase. Note that $\hat{\boldsymbol{R}}_S$ is Hermitian symmetric such that $\hat{\rho}_{i,j} = \hat{\rho}_{j,i}^*$.

In the special case of 1-D parameter estimation, the array steering matrix $\boldsymbol{A}$ reduces to $\boldsymbol{A} =$

$[\boldsymbol{a}(\mu_1),\ldots,\boldsymbol{a}(\mu_d)] \in \mathbb{C}^{M\times d}$ and the deterministic CRB matrix in (9.6) simplifies to

$$\boldsymbol{C} = \frac{\sigma_{\mathrm{n}}^2}{2N}\cdot \mathrm{Re}\left\{\left(\boldsymbol{D}^{\mathrm{H}}\cdot\boldsymbol{\Pi}_{\boldsymbol{A}}^{\perp}\cdot\boldsymbol{D}\right)\odot\hat{\boldsymbol{R}}_S^{\mathrm{T}}\right\}^{-1} \in \mathbb{R}^{d\times d}, \tag{9.11}$$

where $\boldsymbol{D}$ becomes

$$\boldsymbol{D} = \begin{bmatrix}\boldsymbol{d}_1 & \cdots & \boldsymbol{d}_d\end{bmatrix} \in \mathbb{C}^{M\times d} \tag{9.12}$$

with $\boldsymbol{d}_i = \partial\boldsymbol{a}(\mu_i)/\partial\mu_i,\ \forall\ i$.

9.2.2. Deterministic R-D NC Cramér-Rao bound for strictly non-circular signals

In contrast to the case of arbitrary signals, the set of parameters for the strictly non-circular source model in (9.5) is given by the angular parameters $\boldsymbol{\mu} \in \mathbb{R}^{Rd\times 1}$, the real-valued symbols $\boldsymbol{s}_0 = \mathrm{vec}\{\boldsymbol{S}_0\} \in \mathbb{R}^{Nd\times 1}$, the rotation phase angles $\boldsymbol{\varphi} \in \mathbb{R}^{d\times 1}$, and the noise power σ_{n}^2. Thus, the number of parameters is now equal to $(N+R+1)d+1$, which requires the derivation of a new CRB for this parameter set.

The resulting closed-form expression for the deterministic NC CRB matrix $\boldsymbol{C}^{(\mathrm{nc})}$ in the R-D case is stated in the following theorem:

Theorem 9.2.1. *The R-D deterministic NC CRB matrix $\boldsymbol{C}^{(\mathrm{nc})}$ for strictly non-circular sources is given by*

$$\begin{aligned}
\boldsymbol{C}^{(\mathrm{nc})} = \frac{\sigma_{\mathrm{n}}^2}{2N}\cdot\Big\{ &\left(\boldsymbol{G}_2-\boldsymbol{G}_1\cdot\boldsymbol{G}_0^{-1}\cdot\boldsymbol{G}_1^{\mathrm{T}}\right)\odot\hat{\boldsymbol{R}}_{S_0}^{(R)} + \left[\left(\boldsymbol{G}_1\cdot\boldsymbol{G}_0^{-1}\cdot\boldsymbol{H}_0\right)\odot\hat{\boldsymbol{R}}_{S_0}^{(R)}\right] \\
&\cdot\left[\left(\boldsymbol{G}_0-\boldsymbol{H}_0^{\mathrm{T}}\cdot\boldsymbol{G}_0^{-1}\cdot\boldsymbol{H}_0\right)\odot\hat{\boldsymbol{R}}_{S_0}^{(R)}\right]^{-1}\cdot\left[\left(\boldsymbol{H}_1^{\mathrm{T}}-\boldsymbol{H}_0^{\mathrm{T}}\cdot\boldsymbol{G}_0^{-1}\cdot\boldsymbol{G}_1^{\mathrm{T}}\right)\odot\hat{\boldsymbol{R}}_{S_0}^{(R)}\right] \\
&+\left[\boldsymbol{H}_1\odot\hat{\boldsymbol{R}}_{S_0}^{(R)}\right]\cdot\left[\boldsymbol{G}_0\odot\hat{\boldsymbol{R}}_{S_0}^{(R)}\right]^{-1}\cdot\left[\left(\boldsymbol{H}_0^{\mathrm{T}}\cdot\boldsymbol{G}_0^{-1}\cdot\boldsymbol{G}_1^{\mathrm{T}}\right)\odot\hat{\boldsymbol{R}}_{S_0}^{(R)}\right] \\
&+\left[\boldsymbol{H}_1\odot\hat{\boldsymbol{R}}_{S_0}^{(R)}\right]\cdot\left[\boldsymbol{G}_0\odot\hat{\boldsymbol{R}}_{S_0}^{(R)}\right]^{-1}\cdot\left[\left(\boldsymbol{H}_0^{\mathrm{T}}\cdot\boldsymbol{G}_0^{-1}\cdot\boldsymbol{H}_0\right)\odot\hat{\boldsymbol{R}}_{S_0}^{(R)}\right] \\
&\cdot\left[\left(\boldsymbol{G}_0-\boldsymbol{H}_0^{\mathrm{T}}\cdot\boldsymbol{G}_0^{-1}\cdot\boldsymbol{H}_0\right)\odot\hat{\boldsymbol{R}}_{S_0}^{(R)}\right]^{-1}\cdot\left[\left(\boldsymbol{H}_0^{\mathrm{T}}\cdot\boldsymbol{G}_0^{-1}\cdot\boldsymbol{G}_1^{\mathrm{T}}\right)\odot\hat{\boldsymbol{R}}_{S_0}^{(R)}\right] \\
&-\left[\boldsymbol{H}_1\odot\hat{\boldsymbol{R}}_{S_0}^{(R)}\right]\cdot\left[\left(\boldsymbol{G}_0-\boldsymbol{H}_0^{\mathrm{T}}\cdot\boldsymbol{G}_0^{-1}\cdot\boldsymbol{H}_0\right)\odot\hat{\boldsymbol{R}}_{S_0}^{(R)}\right]^{-1}\cdot\left[\boldsymbol{H}_1^{\mathrm{T}}\odot\hat{\boldsymbol{R}}_{S_0}^{(R)}\right]\Big\}^{-1} \in \mathbb{R}^{Rd\times Rd},
\end{aligned} \tag{9.13}$$

where $\hat{\boldsymbol{R}}_{S_0}^{(R)} = \mathbf{1}_{R\times R}\otimes\hat{\boldsymbol{R}}_{S_0}$ with $\hat{\boldsymbol{R}}_{S_0} = \boldsymbol{S}_0\cdot\boldsymbol{S}_0^{\mathrm{T}}/N$ and the matrices $\boldsymbol{G}_n$ and $\boldsymbol{H}_n$, $n = 0,1,2$, are defined as

$$\boldsymbol{G}_0 = \mathrm{Re}\left\{\boldsymbol{\Psi}^*\cdot\boldsymbol{A}^{\mathrm{H}}\cdot\boldsymbol{A}\cdot\boldsymbol{\Psi}\right\} \in \mathbb{R}^{d\times d}, \tag{9.14}$$

$$\boldsymbol{H}_0 = \mathrm{Im}\left\{\boldsymbol{\Psi}^*\cdot\boldsymbol{A}^{\mathrm{H}}\cdot\boldsymbol{A}\cdot\boldsymbol{\Psi}\right\} \in \mathbb{R}^{d\times d}, \tag{9.15}$$

$$\boldsymbol{G}_1 = \mathrm{Re}\left\{(\boldsymbol{I}_R\otimes\boldsymbol{\Psi}^*)\cdot\boldsymbol{D}^{\mathrm{H}}\cdot\boldsymbol{A}\cdot\boldsymbol{\Psi}\right\} \in \mathbb{R}^{Rd\times d}, \tag{9.16}$$

$$\boldsymbol{H}_1 = \mathrm{Im}\left\{(\boldsymbol{I}_R \otimes \boldsymbol{\Psi}^*) \cdot \boldsymbol{D}^{\mathrm{H}} \cdot \boldsymbol{A} \cdot \boldsymbol{\Psi}\right\} \in \mathbb{R}^{Rd \times d}, \tag{9.17}$$

$$\boldsymbol{G}_2 = \mathrm{Re}\left\{(\boldsymbol{I}_R \otimes \boldsymbol{\Psi}^*) \cdot \boldsymbol{D}^{\mathrm{H}} \cdot \boldsymbol{D} \cdot (\boldsymbol{I}_R \otimes \boldsymbol{\Psi})\right\} \in \mathbb{R}^{Rd \times Rd} \tag{9.18}$$

and $\boldsymbol{A}$ *and* $\boldsymbol{D}$ *are given by* (9.3) *and* (9.8)*, respectively.*

The proof is given in Appendix B.36. Note that the empirical source correlation (9.10) contained in $\hat{\boldsymbol{R}}_{S_0}$ becomes real-valued for NC sources, i.e., $\hat{\rho}_{i,j} = |\hat{\rho}_{i,j}|$.

It should be highlighted that the assumption of the R-D array to be separable is not required for the derivation of (9.13) in Appendix B.36. In fact, (9.13) is valid for arbitrarily formed R-D arrays[1], where the columns of $\boldsymbol{A}$ and $\boldsymbol{D}$ are represented accordingly. However, the separability assumption simplifies the further analysis and helps with the presentation of our results in the following sections.

In analogy to the 1-D parameter estimation case of the CRB for arbitrary signals in (9.11), the deterministic 1-D NC CRB matrix is stated in the corollary:

Corollary 9.2.2. *The deterministic 1-D NC CRB is given by* (9.13)*, where* $\hat{\boldsymbol{R}}_{S_0}^{(R)}$ *reduces to* $\hat{\boldsymbol{R}}_{S_0}$ *and* $\boldsymbol{G}_1$*,* $\boldsymbol{H}_1$*, and* $\boldsymbol{G}_2$ *simplify to*

$$\boldsymbol{G}_1 = \mathrm{Re}\left\{\boldsymbol{\Psi}^* \cdot \boldsymbol{D}^{\mathrm{H}} \cdot \boldsymbol{A} \cdot \boldsymbol{\Psi}\right\} \in \mathbb{R}^{d \times d}, \tag{9.19}$$

$$\boldsymbol{H}_1 = \mathrm{Im}\left\{\boldsymbol{\Psi}^* \cdot \boldsymbol{D}^{\mathrm{H}} \cdot \boldsymbol{A} \cdot \boldsymbol{\Psi}\right\} \in \mathbb{R}^{d \times d}, \tag{9.20}$$

$$\boldsymbol{G}_2 = \mathrm{Re}\left\{\boldsymbol{\Psi}^* \cdot \boldsymbol{D}^{\mathrm{H}} \cdot \boldsymbol{D} \cdot \boldsymbol{\Psi}\right\} \in \mathbb{R}^{d \times d} \tag{9.21}$$

with $\boldsymbol{A}$ *and* $\boldsymbol{D}$ *being defined in* (9.11) *and in* (9.12)*.*

9.3. Analysis of the deterministic R-D NC CRB for strictly non-circular signals

In this section, we discuss interesting special cases and properties of the derived R-D NC CRB, where the R-D array is assumed to be separable and centro-symmetric for simplicity. Specifically, we investigate the two cases of equal rotation phases and full coherence for an arbitrary number of strictly non-circular signals. It is shown that in these special cases, the deterministic R-D NC CRB reduces to the R-D CRB. Furthermore, we also analyze the maximum number of resolvable NC sources.

For our analysis, we first refine the model in (9.5) according to Section 2.1.4.2. Recall that an R-D array is termed centro-symmetric in the r-th mode if it satisfies (2.21). Further, review the

[1]These also include non-separable arrays such as cross-arrays and L-shaped arrays.

definition of the array centroid along the r-th mode as

$$\delta^{(r)} = \frac{1}{M_r} \sum_{m_r=1}^{M_r} k_{m_r}. \tag{9.22}$$

Then, $\boldsymbol{A}^{(r)}$ from (9.3) can be decomposed according to (2.24) as $\boldsymbol{A}^{(r)} = \boldsymbol{A}_{\mathrm{c}}^{(r)} \cdot \boldsymbol{\Delta}^{(r)}$, where $\boldsymbol{A}_{\mathrm{c}}^{(r)} = [\boldsymbol{a}_{\mathrm{c}}^{(r)}(\mu_1^{(r)}), \cdots, \boldsymbol{a}_{\mathrm{c}}^{(r)}(\mu_d^{(r)})] \in \mathbb{C}^{M_r \times d}$ satisfies $\boldsymbol{A}_{\mathrm{c}}^{(r)} = \boldsymbol{\Pi}_{M_r} \cdot \boldsymbol{A}_{\mathrm{c}}^{(r)^*}$. Furthermore, the diagonal matrix $\boldsymbol{\Delta}^{(r)} = \operatorname{diag}\left\{\mathrm{e}^{\mathrm{j}\delta^{(r)}\mu_i^{(r)}}\right\}_{i=1}^{d}$ contains the phase shifts of the phase reference $\delta^{(r)}$ for each $\mu_i^{(r)}$. If the phase reference is at the array centroid of the r-th mode, we have $\delta^{(r)} = 0$, $\boldsymbol{\Delta}^{(r)} = \boldsymbol{I}_d$, and consequently $\boldsymbol{A}^{(r)} = \boldsymbol{A}_{\mathrm{c}}^{(r)}$. Thus, we can rewrite $\boldsymbol{A}$ in (9.3) according to (2.25) as

$$\boldsymbol{A} = \boldsymbol{A}_{\mathrm{c}} \cdot \boldsymbol{\Delta}, \tag{9.23}$$

where $\boldsymbol{A}_{\mathrm{c}} = \boldsymbol{A}_{\mathrm{c}}^{(1)} \diamond \boldsymbol{A}_{\mathrm{c}}^{(2)} \diamond \cdots \diamond \boldsymbol{A}_{\mathrm{c}}^{(R)} \in \mathbb{C}^{M \times d}$ and $\boldsymbol{\Delta} = \boldsymbol{\Delta}^{(1)} \cdot \boldsymbol{\Delta}^{(2)} \cdot \ldots \cdot \boldsymbol{\Delta}^{(R)} \in \mathbb{C}^{d \times d}$ with $\delta_i = \sum_{r=1}^{R} \delta^{(r)} \mu_i^{(r)}$.

Inserting (9.23) into the model for strictly non-circular signals in (9.5), we obtain

$$\boldsymbol{X} = \boldsymbol{A}_{\mathrm{c}} \cdot \boldsymbol{\Delta} \cdot \boldsymbol{\Psi} \cdot \boldsymbol{S}_0 + \boldsymbol{N} = \boldsymbol{A}_{\mathrm{c}} \cdot \boldsymbol{\Phi} \cdot \boldsymbol{S}_0 + \boldsymbol{N}, \tag{9.24}$$

where we have defined $\boldsymbol{\Phi} = \boldsymbol{\Delta} \cdot \boldsymbol{\Psi} = \operatorname{diag}\left\{\mathrm{e}^{\mathrm{j}(\varphi_i + \delta_i)}\right\}_{i=1}^{d}$ with $\delta_i = \sum_{r=1}^{R} \delta^{(r)} \mu_i^{(r)}$.

9.3.1. Sources with equal phases

An interesting special case of the model (9.23) occurs when the phase references in each of the R modes coincide with the centroid of the R-D array, i.e., $\delta^{(r)} = 0 \ \forall \ r$ such that $\boldsymbol{\Delta} = \boldsymbol{I}_d$ and $\boldsymbol{A} = \boldsymbol{A}_{\mathrm{c}}$, and, at the same time, the rotation phase angles for all d sources are the same[2], i.e., $\varphi_i = \varphi \ \forall \ i$. Hence, we have

$$\boldsymbol{\Phi} = \boldsymbol{\Psi} = \mathrm{e}^{\mathrm{j}\varphi} \cdot \boldsymbol{I}_d. \tag{9.25}$$

Under these assumptions, the matrices $\boldsymbol{G}_n$, $n = 0, 1, 2$, can be expressed as

$$\boldsymbol{G}_0 = \operatorname{Re}\left\{\mathrm{e}^{-\mathrm{j}\varphi} \cdot \boldsymbol{I}_d \cdot \boldsymbol{A}^{\mathrm{H}} \cdot \boldsymbol{A} \cdot \boldsymbol{I}_d \cdot \mathrm{e}^{\mathrm{j}\varphi}\right\} = \operatorname{Re}\left\{\boldsymbol{A}^{\mathrm{H}} \cdot \boldsymbol{A}\right\} = \boldsymbol{A}^{\mathrm{H}} \cdot \boldsymbol{A} \tag{9.26}$$

$$\boldsymbol{G}_1 = \operatorname{Re}\left\{\mathrm{e}^{-\mathrm{j}\varphi} \cdot \boldsymbol{I}_{Rd} \cdot \boldsymbol{D}^{\mathrm{H}} \cdot \boldsymbol{A} \cdot \boldsymbol{I}_d \cdot \mathrm{e}^{\mathrm{j}\varphi}\right\} = \operatorname{Re}\left\{\boldsymbol{D}^{\mathrm{H}} \cdot \boldsymbol{A}\right\} = \boldsymbol{D}^{\mathrm{H}} \cdot \boldsymbol{A} \tag{9.27}$$

$$\boldsymbol{G}_2 = \operatorname{Re}\left\{\mathrm{e}^{-\mathrm{j}\varphi} \cdot \boldsymbol{I}_{Rd} \cdot \boldsymbol{D}^{\mathrm{H}} \cdot \boldsymbol{D} \cdot \boldsymbol{I}_{Rd} \cdot \mathrm{e}^{\mathrm{j}\varphi}\right\} = \operatorname{Re}\left\{\boldsymbol{D}^{\mathrm{H}} \cdot \boldsymbol{D}\right\} = \boldsymbol{D}^{\mathrm{H}} \cdot \boldsymbol{D} \tag{9.28}$$

while the matrices $\boldsymbol{H}_n$ evaluate to zero. The proof that the matrices $\boldsymbol{A}^{\mathrm{H}} \cdot \boldsymbol{A} \in \mathbb{R}^{d \times d}$, $\boldsymbol{D}^{\mathrm{H}} \cdot \boldsymbol{A} \in \mathbb{R}^{Rd \times d}$, and $\boldsymbol{D}^{\mathrm{H}} \cdot \boldsymbol{D} \in \mathbb{R}^{Rd \times Rd}$ are real-valued can be found in Appendix B.37.

[2] The same behavior applies to the more general case of equality modulo π, i.e., $\varphi_i = \varphi + k_i \cdot \pi$, $k_i \in \mathbb{Z}$ for $i = 1, \ldots, d$. For simplicity of presentation, we assume the angles to be equal, this generalization is however straightforward.

Using these observations, all terms in (9.13) containing $\boldsymbol{H}_0$ or $\boldsymbol{H}_1$ vanish and the R-D NC CRB matrix simplifies to

$$\begin{aligned}\boldsymbol{C}^{(\mathrm{nc})} &= \frac{\sigma_\mathrm{n}^2}{2N} \cdot \left\{ \left(\boldsymbol{G}_2 - \boldsymbol{G}_1 \cdot \boldsymbol{G}_0^{-1} \cdot \boldsymbol{G}_1^\mathrm{T} \right) \odot \hat{\boldsymbol{R}}_{S_0}^{(R)} \right\}^{-1} \\ &= \frac{\sigma_\mathrm{n}^2}{2N} \cdot \left\{ \left(\boldsymbol{D}^\mathrm{H} \cdot \boldsymbol{D} - \boldsymbol{D}^\mathrm{H} \cdot \boldsymbol{A} \cdot \left(\boldsymbol{A}^\mathrm{H} \cdot \boldsymbol{A} \right)^{-1} \cdot \boldsymbol{A}^\mathrm{H} \cdot \boldsymbol{D} \right) \odot \hat{\boldsymbol{R}}_{S_0}^{(R)} \right\}^{-1} \\ &= \frac{\sigma_\mathrm{n}^2}{2N} \cdot \left\{ \left(\boldsymbol{D}^\mathrm{H} \cdot \boldsymbol{\Pi}_{\boldsymbol{A}}^{\perp} \cdot \boldsymbol{D} \right) \odot \hat{\boldsymbol{R}}_{S}^{(R)^\mathrm{T}} \right\}^{-1} = \boldsymbol{C}, \end{aligned} \tag{9.29}$$

where we have used the fact that $\hat{\boldsymbol{R}}_{S}^{(R)} = \hat{\boldsymbol{R}}_{S_0}^{(R)} = \hat{\boldsymbol{R}}_{S_0}^{(R)^\mathrm{T}}$ for $\boldsymbol{\Psi} = \mathrm{e}^{\mathrm{j}\varphi} \cdot \boldsymbol{I}_d$. From (9.29), it is evident that the R-D NC CRB reduces to the R-D CRB if the phase reference is at the R-D array centroid and the rotation phase angles of the sources are equal. This suggests that no gain from strictly non-circular sources can be achieved in this case.

9.3.2. Coherent sources

First, it should be highlighted that the presented R-D NC CRB is derived for arbitrary source correlations ρ and prior knowledge such as uncorrelated ($\rho = 0$) or coherent ($\rho = 1$) sources is not considered. Nevertheless, in this section, we still investigate the special case of coherent sources for our final NC CRB expression, i.e., the correlation coefficients $\hat{\rho}_{i,j}$ between all pairs of sources are given by $|\hat{\rho}_{i,j}| = 1 \ \forall \ i,j$. For simplicity, we assume that all the sources have unit power, i.e., $\hat{P}_i = 1 \ \forall \ i$. Under these assumptions, the sample covariance matrix takes the form $\hat{\boldsymbol{R}}_{S_0} = \mathbf{1}_{d \times d}$ such that $\hat{\boldsymbol{R}}_{S_0}^{(R)} = \mathbf{1}_{Rd \times Rd}$. Hence, all the Hadamard products with $\hat{\boldsymbol{R}}_{S_0}^{(R)}$ in the R-D NC CRB matrix in (9.13) can be omitted and the remaining parts are arranged in the following form

$$\begin{aligned}\frac{\sigma_\mathrm{n}^2}{2N} \cdot \boldsymbol{C}^{(\mathrm{nc})^{-1}} &= \boldsymbol{G}_2 - \boldsymbol{G}_1 \cdot \left[\boldsymbol{G}_0^{-1} + \boldsymbol{G}_0^{-1} \cdot \boldsymbol{H}_0 \cdot \tilde{\boldsymbol{G}}^{-1} \cdot \boldsymbol{H}_0^\mathrm{T} \cdot \boldsymbol{G}_0^{-1} \right] \cdot \boldsymbol{G}_1^\mathrm{T} \\ &\quad + \boldsymbol{H}_1 \cdot \boldsymbol{G}_0^{-1} \cdot \boldsymbol{H}_0^\mathrm{T} \cdot \left[\boldsymbol{G}_0^{-1} + \boldsymbol{G}_0^{-1} \cdot \boldsymbol{H}_0 \cdot \tilde{\boldsymbol{G}}^{-1} \cdot \boldsymbol{H}_0^\mathrm{T} \cdot \boldsymbol{G}_0^{-1} \right] \cdot \boldsymbol{G}_1^\mathrm{T} \\ &\quad + \boldsymbol{G}_1 \cdot \boldsymbol{G}_0^{-1} \cdot \boldsymbol{H}_0 \cdot \tilde{\boldsymbol{G}}^{-1} \cdot \boldsymbol{H}_1^\mathrm{T} - \boldsymbol{H}_1 \cdot \tilde{\boldsymbol{G}}^{-1} \cdot \boldsymbol{H}_1^\mathrm{T} \qquad (9.30) \\ &= \boldsymbol{G}_2 - \left(\boldsymbol{G}_1 - \boldsymbol{H}_1 \cdot \boldsymbol{G}_0^{-1} \cdot \boldsymbol{H}_0^\mathrm{T} \right) \cdot \left(\boldsymbol{G}_0 - \boldsymbol{H}_0 \cdot \boldsymbol{G}_0^{-1} \cdot \boldsymbol{H}_0^\mathrm{T} \right)^{-1} \cdot \boldsymbol{G}_1^\mathrm{T} \\ &\quad - \left(\boldsymbol{H}_1 - \boldsymbol{G}_1 \cdot \boldsymbol{G}_0^{-1} \cdot \boldsymbol{H}_0 \right) \cdot \left(\boldsymbol{G}_0 - \boldsymbol{H}_0^\mathrm{T} \cdot \boldsymbol{G}_0^{-1} \cdot \boldsymbol{H}_0 \right)^{-1} \cdot \boldsymbol{H}_1^\mathrm{T}, \qquad (9.31) \end{aligned}$$

where in (9.30), we have defined $\tilde{\boldsymbol{G}} = \boldsymbol{G}_0 - \boldsymbol{H}_0^\mathrm{T} \cdot \boldsymbol{G}_0^{-1} \cdot \boldsymbol{H}_0$ and replaced the terms in the square brackets by applying the converse of the matrix inversion lemma, yielding the matrix $(\boldsymbol{G}_0 - \boldsymbol{H}_0 \cdot \boldsymbol{G}_0^{-1} \cdot \boldsymbol{H}_0^\mathrm{T})^{-1}$. Note that (9.31) can be transformed into the block matrix form

$$\frac{\sigma_\mathrm{n}^2}{2N} \cdot \boldsymbol{C}^{(\mathrm{nc})^{-1}} = \boldsymbol{G}_2 - \begin{bmatrix} \boldsymbol{G}_1 & \boldsymbol{H}_1 \end{bmatrix} \cdot \begin{bmatrix} \boldsymbol{G}_0 & \boldsymbol{H}_0 \\ \boldsymbol{H}_0^\mathrm{T} & \boldsymbol{G}_0 \end{bmatrix}^{-1} \cdot \begin{bmatrix} \boldsymbol{G}_1^\mathrm{T} \\ \boldsymbol{H}_1^\mathrm{T} \end{bmatrix},$$

which represents a very interesting simplification of the original expression.

In the next step, we rewrite the R-D CRB matrix for arbitrary signals in (9.6) for the case of coherent sources in a similar form. Under the aforementioned assumptions, the R-D sample covariance matrix is given by $\hat{\boldsymbol{R}}_S^{(R)} = \mathbf{1}_{R\times R} \otimes (\boldsymbol{\Psi}^* \cdot \mathbf{1}_{d\times d} \cdot \boldsymbol{\Psi}) = (\boldsymbol{I}_R \otimes \boldsymbol{\Psi}^*) \cdot \mathbf{1}_{Rd\times Rd} \cdot (\boldsymbol{I}_R \otimes \boldsymbol{\Psi})$. Hence, we simplify the original form of the CRB in (9.11) into

$$\frac{\sigma_\mathrm{n}^2}{2N} \cdot \boldsymbol{C}^{-1} = \mathrm{Re}\Big\{ \big(\boldsymbol{D}^\mathrm{H} \cdot \boldsymbol{\Pi}_{\boldsymbol{A}}^{\perp} \cdot \boldsymbol{D}\big) \odot \big((\boldsymbol{I}_R \otimes \boldsymbol{\Psi}^*) \cdot \mathbf{1}_{Rd\times Rd} \cdot (\boldsymbol{I}_R \otimes \boldsymbol{\Psi})\big) \Big\} \tag{9.32}$$

$$= \mathrm{Re}\Big\{ (\boldsymbol{I}_R \otimes \boldsymbol{\Psi}^*) \cdot \boldsymbol{D}^\mathrm{H} \cdot \boldsymbol{D} \cdot (\boldsymbol{I}_R \otimes \boldsymbol{\Psi}) - (\boldsymbol{I}_R \otimes \boldsymbol{\Psi}^*) \cdot \boldsymbol{D}^\mathrm{H} \cdot \boldsymbol{A} \cdot \boldsymbol{\Psi} \cdot \big(\boldsymbol{\Psi}^* \cdot \boldsymbol{A}^\mathrm{H} \cdot \boldsymbol{A} \cdot \boldsymbol{\Psi}\big)^{-1} \cdot \boldsymbol{\Psi}^* \cdot \boldsymbol{A}^\mathrm{H} \cdot \boldsymbol{D} \cdot (\boldsymbol{I}_R \otimes \boldsymbol{\Psi}) \Big\} \tag{9.33}$$

$$= \mathrm{Re}\Big\{ \boldsymbol{G}_2 + \mathrm{j}\boldsymbol{H}_2 - (\boldsymbol{G}_1 + \mathrm{j}\boldsymbol{H}_1) \cdot (\boldsymbol{G}_0 + \mathrm{j}\boldsymbol{H}_0)^{-1} \cdot (\boldsymbol{G}_1^\mathrm{T} - \mathrm{j}\boldsymbol{H}_1^\mathrm{T}) \Big\}, \tag{9.34}$$

where we have introduced additional matrices $\boldsymbol{\Psi}$ in (9.33) by noting that $\boldsymbol{\Psi} \cdot \boldsymbol{\Psi}^* = \boldsymbol{I}_d$. To proceed we require the following lemma:

Lemma 9.3.1. *The inverse of a full rank complex-valued matrix $\boldsymbol{C} = \boldsymbol{A} + \mathrm{j}\boldsymbol{B} \in \mathbb{C}^{n\times n}$ with the real part $\boldsymbol{A} \in \mathbb{R}^{n\times n}$ and the imaginary part $\boldsymbol{B} \in \mathbb{R}^{n\times n}$ can be split into its real part and its imaginary part as follows:*

$$\begin{aligned} \boldsymbol{C}^{-1} &= (\boldsymbol{A} + \mathrm{j}\boldsymbol{B})^{-1} \\ &= \big(\boldsymbol{A} + \boldsymbol{B} \cdot \boldsymbol{A}^{-1} \cdot \boldsymbol{B}\big)^{-1} - \mathrm{j}\boldsymbol{A}^{-1} \cdot \boldsymbol{B} \big(\boldsymbol{A} + \boldsymbol{B} \cdot \boldsymbol{A}^{-1} \cdot \boldsymbol{B}\big)^{-1} \end{aligned}$$

if $\boldsymbol{A}$ and $\big(\boldsymbol{A} + \boldsymbol{B} \cdot \boldsymbol{A}^{-1} \cdot \boldsymbol{B}\big)$ are invertible.

To prove this lemma, it is sufficient to multiply $\boldsymbol{C}$ with $\boldsymbol{C}^{-1}$ and show that the result is the identity matrix.

Applying Lemma 9.3.1 to (9.34), we split $\boldsymbol{C}^{-1}$ into its real and imaginary part. After some elementary operations and using the fact that $\boldsymbol{H}_0^\mathrm{T} = -\boldsymbol{H}_0$, we obtain equation (9.31) and consequently, we have $\boldsymbol{C}^{(\mathrm{nc})} = \boldsymbol{C}$. Thus, both R-D CRBs become equal if all the sources are coherent. Note that the case of a single snapshot, i.e., $N = 1$, can be considered a special case of coherent sources as this renders $\hat{\boldsymbol{R}}_{S_0}$ rank-one such that the same steps as above can be applied[3]. It should also be noted that the above result is valid for arbitrary R-D arrays as the assumptions of separability and centro-symmetry were not used in the derivation. Analogously to the special case considered in the previous subsection, our findings suggest that no NC gain can be achieved for coherent sources and a single snapshot.

[3] If the amplitude of the i-th source in the case $N = 1$ has a negative sign, it can be represented as $-s_{0_i} \mathrm{e}^{\mathrm{j}\varphi_i} = s_{0_i} \mathrm{e}^{\mathrm{j}(\varphi_i + \pi)}$ with $\hat{P}_i = 1$, such that $\hat{\boldsymbol{R}}_{S_0} = \mathbf{1}_{d\times d}$ still holds, which was assumed for the above derivation.

9.3.3. Maximum number of resolvable sources

In the case of arbitrary signals, it is well-known from [SN89] that the upper limit[4] of sources that can be resolved with M sensors is $d = M - 1$. However, if the sources are strictly non-circular, we can estimate the DOAs of even more sources than sensors available. In this section, we establish the conditions under which the deterministic NC CRB is valid for $d \geq M$.

Firstly, it is not difficult to see that the matrices $\boldsymbol{G}_n$ and $\boldsymbol{H}_n$, $n = 0, 1, 2$, can have a rank larger than M. For example, the matrix $\boldsymbol{G}_0$ can be rewritten as

$$\boldsymbol{G}_0 = \mathrm{Re}\left\{\boldsymbol{\Psi}^* \cdot \boldsymbol{A}^{\mathrm{H}} \cdot \boldsymbol{A} \cdot \boldsymbol{\Psi}\right\} = \begin{bmatrix} \boldsymbol{A} \cdot \mathrm{Re}\{\boldsymbol{\Psi}\} \\ \boldsymbol{A} \cdot \mathrm{Im}\{\boldsymbol{\Psi}\} \end{bmatrix}^{\mathrm{H}} \cdot \underbrace{\begin{bmatrix} \boldsymbol{A} \cdot \mathrm{Re}\{\boldsymbol{\Psi}\} \\ \boldsymbol{A} \cdot \mathrm{Im}\{\boldsymbol{\Psi}\} \end{bmatrix}}_{2M \times d}.$$

From this equation, it can be seen that unless the phase matrix $\boldsymbol{\Psi}$ is equal to $\boldsymbol{\Psi} = \mathrm{diag}\left\{\mathrm{e}^{\mathrm{j}\varphi_i}\right\}_{i=1}^{d}$ with $\varphi_i = \varphi + k_i \cdot \pi,\ k_i \in \mathbb{Z}$, i.e., all the rotation phases are equal modulo π, $\boldsymbol{G}_0$ has a rank larger than M if $d > M$. This result complies with the one from Subsection 9.3.1. For the matrices $\boldsymbol{G}_1$, $\boldsymbol{G}_2$ as well as $\boldsymbol{H}_n$, $n = 0, 1$, similar forms are easily found.

Secondly, regarding the additional dependence of the NC CRB on the sample covariance matrix $\hat{\boldsymbol{R}}_{S_0}$, we have proven in Subsection 9.3.2 that the NC CRB reduces to the CRB if the sources are coherent. This suggests that for non-coherent sources, the NC CRB is valid for $d \geq M$.

Consequently, we can infer for a uniform linear array that if the sources are non-coherent, i.e.,

$$|\hat{\rho}_{i,j}| < 1 \quad \forall i \neq j \ \text{ in } \ 1, 2, \ldots, d, \tag{9.35}$$

and the rotation phase angles are different, i.e.,

$$|\varphi_i - \varphi_j| \neq 0 \mod \pi \quad \forall i \neq j \ \text{ in } \ 1, 2, \ldots, d, \tag{9.36}$$

the Fisher information matrix has full rank and is invertible as long as $d \leq 2(M-1)$. To support our claim, we provide the numerical evaluation shown in Table 9.1, which suggests that the condition $d \leq 2(M-1)$ represents an upper limit on the number of sources that is resolvable. Therefore, up to twice as many signal sources can be resolved compared to the case of arbitrary signals.

9.4. Special NC source cases

In this section, we simplify the derived deterministic NC CRB for a single NC source and two NC sources and compare it the corresponding simplifications of the deterministic CRB for arbitrary

[4] This limit is not reached with all array geometries. An example for an array, which can achieve this limit is a ULA.

sources. In Section 9.4.1, we consider the single NC source case and in Section 9.4.2, we investigate the two NC source case.

9.4.1. Single NC source case

In Section 4.5.1 and in Section 6.6.1, we have already simplified the performance analysis expressions for the MSE of R-D ESPRIT-type algorithms and R-D NC ESPRIT-type algorithms for a single source ($d = 1$). In Section 7.5, we have also taken into account the spatial smoothing preprocessing step and derived analytical MSE expressions for a single source of R-D ESPRIT-type algorithms and R-D NC ESPRIT-type algorithms both with spatial smoothing. In this section, we derive a simplified expression of the deterministic R-D NC CRB for a single source, which provides a lower limit on the previously obtained single source expressions for the MSE of R-D NC ESPRIT-type algorithms both with and without spatial smoothing. The resulting expression only depends on the parameters of physical significance, e.g., the number of sensors M, the SNR, and the number of snapshots N. For the derivation, we assume a non-uniform R-D array, which is centro-symmetric and separable as introduced in Section 2.1.4.2.

We have shown in Section 9.3.1 that based on the deterministic R-D NC CRB, no NC gain can be obtained if the sources have the same rotation phase while the phase reference is at the array centroid or if the sources are coherent. As the single source case is a special case of each of these two properties, i.e., $\boldsymbol{\Psi} = \mathrm{e}^{\mathrm{j}\varphi}$ or $\hat{\boldsymbol{R}}_{S_0}^{(R)} = \mathbf{1}_{R\times R}$, we can directly conclude that the R-D NC CRB and the R-D CRB must be equal for this case as well.

The simplified expression of the deterministic R-D NC CRB for a single strictly non-circular source is shown in the next theorem:

Theorem 9.4.1. *For the case of an $M_1 \times \ldots \times M_R$ (M-element) separable R-D array with $\delta^{(r)} = 0\ \forall\ r$, i.e., the phase reference of the centro-symmetric array is at the centroid, and a single strictly non-circular source ($d = 1$), the deterministic R-D NC CRB can be simplified to*

$$\boldsymbol{C}^{(\mathrm{nc})} = \mathrm{diag}\left\{\left[C^{(\mathrm{nc})(1)}, \ldots, C^{(\mathrm{nc})(R)}\right]\right\} \in \mathbb{R}^{R\times R} \tag{9.37}$$

with

$$C^{(\mathrm{nc})(r)} = \frac{1}{\hat{\varrho}} \cdot \frac{M_r}{2M} \cdot \frac{1}{\sum_{m_r=1}^{M_r} k_{m_r}^2} \quad \forall\ r, \tag{9.38}$$

where $\hat{\varrho}$ represents the effective SNR $\hat{\varrho} = N\hat{P}/\sigma_{\mathrm{n}}^2$ with $\hat{P}$ being the empirical source power given by $\hat{P} = \|\boldsymbol{s}_0\|_2^2/N$ and $\boldsymbol{s}_0 \in \mathbb{R}^{N\times 1}$.

The proof is given in Appendix B.38. For the special case of a uniform R-D sampling grid, the R-D NC CRB expression from Theorem 9.4.1 is simplified in the following corollary:

Corollary 9.4.2. *For an M-element uniform R-D array with an M_r-element ULA in the r-th mode and a single strictly non-circular source ($d = 1$), the deterministic NC CRB for the r-th mode in* (9.38) *can be explicitly expressed as*

$$C^{(\mathrm{nc})(r)} = \frac{1}{\hat{\varrho}} \cdot \frac{6}{M(M_r^2 - 1)} \quad \forall\, r, \tag{9.39}$$

where $k_{m_r} = -(M_r - 1)/2, \ldots, (M_r - 1)/2$.

Note that the expression (9.39) is equivalent to the result for the single source case of the deterministic R-D CRB for arbitrary signals derived in [RH12]. This suggests that, based on the deterministic data assumption, no improvement in terms of the estimation accuracy can be achieved for a single strictly non-circular source. In connection to previous work, it is worth mentioning that the single source expression derived from the stochastic NC CRB in [DA04] does achieve a small gain for low SNRs. However, it can be shown that in the high SNR regime, the expression in [DA04] is equivalent to our result in (9.39).

9.4.2. Two NC source case

After establishing that according to the R-D NC CRB, no NC gain can be attained for a single source, the question to be studied is what is the maximum achievable NC gain if at least two sources are not fully coherent, their rotation phases are different, and the phase reference is arbitrary. In Section 6.6.2, we have already considered the special case of two NC sources for R-D NC ESPRIT-type algorithms and R-D NC Tensor-ESPRIT-type algorithms based on their performance analysis expressions for the MSE. We have analytically computed the NC again of these algorithms for the matrix case as well as the tensor case. In this section, in order to obtain a benchmark for these results, we analytically compute the maximum achievable NC gain for two closely-spaced NC sources based on the deterministic R-D NC CRB. The R-D CRB for arbitrary source constellations tends to infinity when the source separation approaches zero. This is not always true for the R-D NC CRB as under certain conditions, a finite value is reached. This observation motivates us to derive simplified expressions of the NC CRB and the CRB for the two source case, which are subsequently used to analytically compute the maximum achievable NC gain. To obtain generic expressions in terms of the physical parameters, the derivations are based on the model in (9.24).

For simplicity, we limit our analysis to the 1-D parameter estimation case from Example 2.1.1 and assume a ULA composed of M isotropic sensor elements, which is centro-symmetric. The phase reference is located at an arbitrary position. For this scenario, the array steering matrix $\boldsymbol{A}_{\mathrm{c}}$ in model (9.24) simplifies to

$$\boldsymbol{A}_{\mathrm{c}} = \begin{bmatrix} \boldsymbol{a}_{\mathrm{c}}(\mu_1) & \cdots & \boldsymbol{a}_{\mathrm{c}}(\mu_d) \end{bmatrix} \in \mathbb{C}^{M \times d}, \tag{9.40}$$

where the steering vectors $\boldsymbol{a}_{\mathrm{c}}(\mu_i),\ i = 1, \ldots, d$, are defined as

$$\boldsymbol{a}_{\mathrm{c}}(\mu_i) = \left[\mathrm{e}^{-\mathrm{j}\frac{(M-1)}{2}\mu_i} \quad \ldots \quad \mathrm{e}^{\mathrm{j}\frac{(M-1)}{2}\mu_i}\right] \in \mathbb{C}^{M\times 1}. \tag{9.41}$$

After inserting (9.40) into the expression (9.23), it is once more apparent that if the phase reference is at the array centroid, we have $\delta = 0$ and consequently $\boldsymbol{\Delta} = \boldsymbol{I}_d$. Moreover, if the phase reference is at the first element, we have $\delta = (M-1)/2$.

9.4.2.1. NC CRB for two closely-spaced sources

The result obtained by simplifying the NC CRB for two closely-spaced sources can be summarized in the following theorem:

Theorem 9.4.3. *For the case of an M-element ULA (1-D) and two closely-spaced strictly non-circular sources ($d = 2$), the deterministic NC Cramér-Rao bound can be simplified to*

$$\begin{aligned}\mathrm{Tr}\left\{\boldsymbol{C}^{(\mathrm{nc})}\right\} \approx 50400 \cdot \Bigg(&\hat{\rho}^2\Delta\mu^2 M(M-1)(M-2)(M+2)(M+1)\\ &\cdot\left(\Delta\mu^2(M-3)(M+3)\cdot\cos^2(\Delta\phi) + 140\cdot\sin^2(\Delta\phi)\right) + \left(1-\hat{\rho}^2\right)M(M-1)(M+1)\\ &\cdot\left(140\cdot\Delta\mu^2(M-2)(M+2)\cdot\cos^2(\Delta\phi) + 8400\cdot\sin^2(\Delta\phi)\right)\Bigg)^{-1}\cdot\frac{\hat{\varrho}_1+\hat{\varrho}_2}{\hat{\varrho}_1\hat{\varrho}_2}.\end{aligned} \tag{9.42}$$

In (9.42), we have defined $\Delta\mu = |\mu_2 - \mu_1|$ and $\Delta\phi = \Delta\varphi + \delta\Delta\mu$ with $\Delta\varphi = |\varphi_2 - \varphi_1|$. Moreover, $\hat{\varrho}_i = N\hat{P}_i/\sigma_{\mathrm{n}}^2,\ i = 1, 2$ represents the effective SNR of each of the two sources.

The proof is given in Appendix B.39.

It is worth highlighting that the analytical expression in (9.42) is only an approximate result as the derivation involves a Taylor series approximation for small $\Delta\mu$, where the higher order terms beyond $\mathcal{O}(\Delta\mu^4)$ have been neglected. Therefore, (9.42) becomes accurate if $\Delta\mu$ is small. The question of how small $\Delta\mu$ should be strongly depends on M. From (9.42), it is clear that $\Delta\mu$ and M are inversely proportional, i.e., the smaller the separation $\Delta\mu$, the larger the value of M that is required to maintain the same CRB. Therefore, the approximation becomes accurate if $\Delta\mu \ll \mathrm{const}\cdot\frac{1}{M}$. However, this condition is not a restriction as the case of closely-spaced sources is the scenario, where high-resolution parameter algorithms are primarily applied.

Also, note that the behavior of the simplified NC CRB in (9.42) is symmetric in $\Delta\varphi$ as the two sources can be interchanged. Moreover, as any real-valued data stream can be multiplied by the factor -1, which represents a phase shift of π, it is also π-periodic. Combining these two results, only the interval $\Delta\varphi \in [0, \pi/2]$ must be considered and the general behavior of the NC CRB can be extracted from this interval by mirroring and periodification. Consequently, the maximum phase

separation is given by $\Delta\varphi = \pi/2$.

Based on the result in (9.42), simplified expressions for several special cases can be deduced, e.g., for two uncorrelated ($\hat{\rho} = 0$) or coherent ($\hat{\rho} = 1$) sources as well as for $\Delta\phi = 0$ or $\Delta\phi = \pi/2$.

Remark 1: One specific case that is worth highlighting is the case $\hat{\rho} = 0$ and $\Delta\phi = \pi/2$, where $\Delta\varphi = \pi/2$ and $\delta = 0$. Under these conditions, the NC CRB for two sources in (9.42) simplifies to

$$\mathrm{Tr}\left\{\boldsymbol{C}^{(\mathrm{nc})}\right\} \approx \frac{6}{M(M^2-1)} \cdot \frac{\hat{\varrho}_1 + \hat{\varrho}_2}{\hat{\varrho}_1 \hat{\varrho}_2}, \tag{9.43}$$

which is independent of $\Delta\mu$. As (9.43) resembles the expression for a single source in (9.37), it is apparent that the individual NC CRB for each of the two sources represents the NC CRB for the single source case discussed in the previous section. Hence, the two sources entirely decouple as if each of them was present alone.

Remark 2: Another special case occurs when the two sources approach each other, i.e., $\Delta\mu$ approaches zero. In the CRB for arbitrary sources this always implies that the CRB tends to infinity. This is, however, not always true for the NC CRB. The limit can be computed as

$$\lim_{\Delta\mu\to 0} \mathrm{Tr}\left\{\boldsymbol{C}^{(\mathrm{nc})}\right\} = \frac{1}{1-\hat{\rho}^2} \cdot \frac{6}{M(M^2-1)} \cdot \frac{1}{\sin^2(\Delta\phi)} \cdot \frac{\hat{\varrho}_1 + \hat{\varrho}_2}{\hat{\varrho}_1 \hat{\varrho}_2}. \tag{9.44}$$

Thus, for $\hat{\rho} < 1$ and $\Delta\phi > 0$, a finite value is reached. If we have $\hat{\rho} = 0$ and $\Delta\phi = \pi/2$, the limit (9.44) corresponds to (9.43), and for $\hat{\rho} = 1$ and $\Delta\phi = 0$, the limit tends to infinity as the NC CRB matches the CRB.

9.4.2.2. CRB for two closely-spaced sources

The corresponding expression of the simplified CRB for two closely-spaced sources is stated as follows:

Theorem 9.4.4. *For the case of an M-element ULA (1-D) and two closely-spaced sources ($d = 2$), the deterministic Cramér-Rao bound can be simplified to*

$$\begin{aligned}\mathrm{Tr}\{\boldsymbol{C}\} \approx 50400 \cdot \Bigg(& \hat{\rho}^2 \Delta\mu^2 M(M-1)(M-2)(M+2)(M+1) \\ & \cdot \left(\Delta\mu^2 (M-3)(M+3) \cdot \cos^2(\Delta\phi) + 140 \cdot \sin^2(\Delta\phi)\right) \\ & + 140 \cdot \left(1-\hat{\rho}^2\right) \Delta\mu^2 M(M-1)(M-2)(M+2)(M+1)\Bigg)^{-1} \cdot \frac{\hat{\varrho}_1 + \hat{\varrho}_2}{\hat{\varrho}_1 \hat{\varrho}_2}.\end{aligned} \tag{9.45}$$

The proof is given in Appendix B.40.

In analogy to the result for the NC CRB, (9.45) becomes exact for small $\Delta\mu$ and the higher order terms beyond $\mathcal{O}(\Delta\mu^4)$ of the Taylor series expansion are negligible.

Again, more simplified expressions for several special cases can be derived from (9.45), e.g., $\hat{\rho} = 0$, $\hat{\rho} = 1$, $\Delta\phi = 0$, or $\Delta\phi = \pi/2$.

Remark 3: A very interesting property of the CRB can be shown for $\hat{\rho} = 1$ and $\Delta\phi = \pi/2$ with $\delta = 0$. For these parameters, we can reduce the CRB in (9.45) to

$$\mathrm{Tr}\{\boldsymbol{C}\} \approx \frac{1}{\Delta\mu^2} \cdot \frac{360}{M(M-1)(M-2)(M+2)(M+1)} \cdot \frac{\hat{\varrho}_1 + \hat{\varrho}_2}{\hat{\varrho}_1 \hat{\varrho}_2}, \tag{9.46}$$

which corresponds to the expression of the CRB for $\hat{\rho} = 0$ and arbitrary $\Delta\phi$. This implies that a rotation phase separation of $\pi/2$ decorrelates two coherent sources.

Remark 4: In contrast to the NC CRB, the limit for the CRB is given by

$$\lim_{\Delta\mu \to 0} \mathrm{Tr}\{\boldsymbol{C}\} = \infty \quad \forall\ \hat{\rho},\ \ \forall\ \Delta\phi. \tag{9.47}$$

Therefore, the NC CRB for strictly non-circular sources exhibits substantial benefits compared to the CRB if the sources are closely-spaced, incoherent, and have a non-vanishing phase discrimination $\Delta\phi$.

9.4.2.3. Analytical NC gain for two closely-spaced sources

Based on the simplified expressions for the two-source case of the NC CRB in (9.42) and the CRB in (9.45), we can explicitly compute the NC gain for two sources as

$$\begin{aligned}
\eta^{(\mathrm{nc})} &= \frac{\mathrm{Tr}\{\boldsymbol{C}\}}{\mathrm{Tr}\left\{\boldsymbol{C}^{(\mathrm{nc})}\right\}} \\
&\approx 1 + \Big(140 \cdot (1-\hat{\rho}^2) M(M-1)(M+1) \cdot \sin^2(\Delta\phi)\Big(60 - \Delta\mu^2(M-2)(M+2)\Big)\Big) \Big/ \\
&\Big(\Delta\mu^2 M(M-1)(M-2)(M+2)(M+1)\Big(\Delta\mu^2\hat{\rho}^2(M-3)(M+3) \cdot \cos^2(\Delta\phi) \\
&+ 140 \cdot (1 - \hat{\rho}^2 \cos^2(\Delta\phi))\Big)\Big)
\end{aligned} \tag{9.48}$$

As the derivation of (9.48) is based on (9.42) and (9.45), it becomes accurate for small source separations $\Delta\mu$ as well. We can now analyze the properties of the NC gain expression for different values of $\hat{\rho}$, $\Delta\varphi$, and δ.

Remark 5: As already established earlier for an arbitrary number of sources, the NC CRB becomes equal to the CRB if either $\hat{\rho} = 1$ or if $\Delta\phi = 0$, where $\Delta\varphi = 0$ and $\delta = 0$. This behavior also reflects in the NC gain computed for two strictly non-circular sources as it can easily be verified

that for these parameter values, the expression (9.48) evaluates to $\eta^{(\mathrm{nc})} = 1$. Hence, no NC gain is obtained in these cases. Note, however, that if $\delta \neq 0$, i.e., the phase reference is not at the array centroid, there may be an NC gain even if $\Delta\varphi = 0$.

Remark 6: By analyzing the NC CRB for two closely-spaced sources, we have found that for $\hat{\rho} = 0$ and $\Delta\phi = \pi/2$ with $\delta = 0$, the two sources entirely decouple. Evaluating the NC gain expression for these parameters leads to

$$\eta^{(\mathrm{nc})} \approx \frac{1}{\Delta\mu^2} \cdot \frac{60}{(M-2)(M+2)}. \tag{9.49}$$

Thus, this case represents the largest achievable gain for two closely-spaced strictly non-circular sources. It is apparent that the NC gain in (9.49) decays in proportion to M^{-2} but increases as $\Delta\mu$ decreases.

Remark 7: The limit of the NC gain for $\Delta\mu$ approaching zero is given by

$$\lim_{\Delta\mu \to 0} \eta^{(\mathrm{nc})} = \infty \quad \forall\ \hat{\rho},\ \ \forall\ \Delta\phi. \tag{9.50}$$

Therefore, the NC gain can theoretically approach infinity if the source separation tends to zero.

9.4.3. Two groups of equal phases

This subsection represents a generalization of the case of two uncorrelated strictly non-circular sources to two groups of equal phases. Let d mutually uncorrelated sources with unit power, i.e., $\hat{\boldsymbol{R}}_{S_0} = \boldsymbol{I}_d$, have the phase angles

$$\varphi_i = \varphi^{[1]} + k_i \cdot \pi \quad \text{or} \quad \varphi_i = \varphi^{[2]} + k_i \cdot \pi,\ \ i = 1, \ldots, d,$$

where $k_i \in \mathbb{Z}$, i.e., modulo π there are only two different phase angles: $\varphi^{[1]}$ and $\varphi^{[2]}$. Without loss of generality, we can reorder the sources such that the d_1 sources with phase $\varphi^{[1]}$ are the sources $1, 2, \ldots, d_1$ and the remaining $d - d_1$ sources $d_1 + 1, d_1 + 2, \ldots, d$ have phase $\varphi^{[2]}$. Thus, the sources fall into two groups, where the NC gain depends on the phase separation $|\varphi^{[2]} - \varphi^{[1]}|$ of the groups.

Now, in the special case $|\varphi^{[2]} - \varphi^{[1]}| = \pi/2$, i.e., the phase separation between the two groups is maximum, it is straightforward to see that the matrices $\boldsymbol{G}_0$, $\boldsymbol{G}_1$, and $\boldsymbol{G}_2$ are block diagonal, i.e., they are zero except for the upper left $d_1 \times d_1$ block matrix and the lower right $(d - d_1) \times (d - d_1)$ block. Combining these matrices and using the fact that the correlation coefficients are zero, we can show from the joint CRB that the two groups decouple, that is, the first d_1 sources are completely decoupled from the remaining $(d - d_1)$ sources. This case can provide a significant gain compared to the CRB for arbitrary sources if there are closely-spaced sources that belong to different groups.

9.5. Numerical results

In this section, we provide simulation results to evaluate the behavior of the R-D NC CRB and illustrate our analytical results.

9.5.1. Behavior of the deterministic R-D NC CRB

In this subsection, we compare the root mean squared error (RMSE) of the derived deterministic R-D NC CRB (Det NC CRB) to the deterministic R-D CRB (Det CRB) and the stochastic R-D NC CRB (Sto NC CRB) for weak-sense non-circular signals from [DA04]. Note that the R-D extension of the Sto NC CRB in [DA04] is obtained analogously to (9.6). Moreover, we include the R-D NC Standard ESPRIT (NC SE) and R-D NC Unitary ESPRIT (NC UE) algorithms [SRHD14] as well as their non-NC counterparts R-D Standard ESPRIT (SE) and R-D Unitary ESPRIT (UE) [HN95] into the comparison. It is assumed that a known number of signals with unit power and real-valued symbols ($\zeta = 1$) drawn from a Gaussian distribution impinge on the array.

Figure 9.1 illustrates the RMSE over all sources versus the SNR for the centro-symmetric 2-D array ($R = 2$) in Figure 9.2 with $M = 12$, where $N = 20$ available snapshots of $d = 3$ sources with the spatial frequencies $\mu_1^{(1)} = 0.25$, $\mu_2^{(1)} = 0.5$, $\mu_3^{(1)} = 0.75$, $\mu_1^{(2)} = 0.25$, $\mu_2^{(2)} = 0.5$, and $\mu_3^{(2)} = 0.75$, and a real-valued pair-wise correlation of $\rho = 0.9$. The rotation phases contained in $\boldsymbol{\Psi}$ are given by $\varphi_1 = 0$, $\varphi_2 = \pi/4$, and $\varphi_3 = \pi/2$. It is apparent from Figure 9.1 that the NC SE and NC UE algorithms perform close to the derived Det NC CRB and that all of these outperform the Sto NC CRB from [DA04].

In Table 9.1, we analyze the Det 1-D CRB and the Det 1-D NC CRB for a varying number of sources d in case of a ULA with $M = 4$, $N = 20$, and SNR = 10 dB. The spatial frequencies $\mu_i,\ \forall i$ are distributed equally in the interval $[-2, 2]$ and the rotation phases $\varphi_i,\ \forall i$ are drawn randomly. It can be seen that $d_{\max} = M - 1$ for the CRB and $d_{\max}^{(\mathrm{nc})} = 2(M - 1)$ for the NC CRB are the largest numbers of d that lead to an invertible Fisher matrix, otherwise, the problem is ill-posed. Therefore, twice as many sources can be resolved from the strictly non-circular data model.

9.5.2. Analytical results

In this subsection, we compare the analytical results "ana" in (9.42) and (9.45) to the empirical ones "emp" in (9.11) and Corollary 9.2.2 obtained by averaging over 1000 Monte-Carlo trials. We have $d = 2$ sources that impinge on a ULA (1-D) with the powers $P_1 = 0.5$ and $P_2 = 1.5$. The symbols $\boldsymbol{S}_0$ are randomly drawn from a real-valued Gaussian distribution.

In Figure 9.3, we display the RMSE of the Det 1-D NC CRB and the Det 1-D CRB for $d = 2$ sources as a function of the number of sensors M, where the square root of the analytical expressions is taken. The source separation is $\Delta\mu = 0.1$ rad with $\mu_1 = 0$ and $\mu_2 = 0.1$, however, the actual

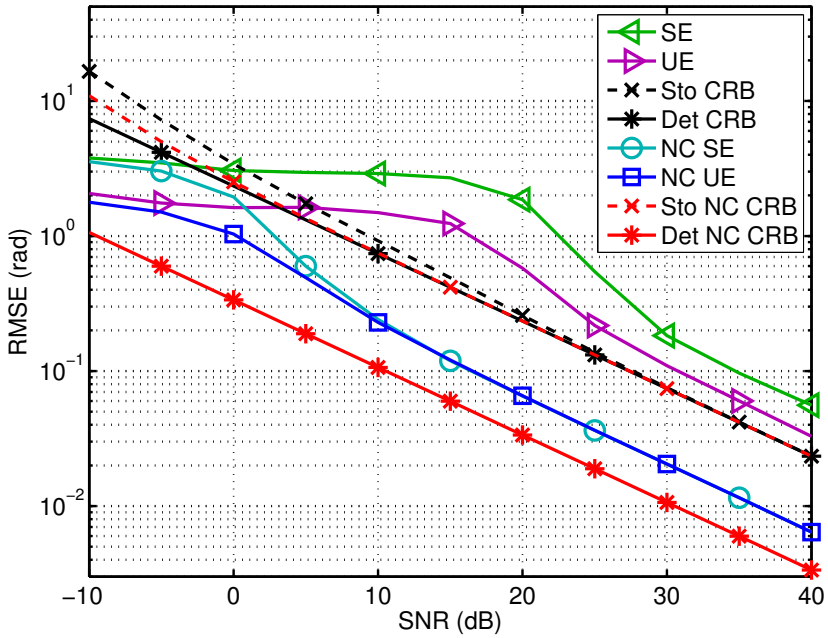


Figure 9.1.: Analytical and empirical RMSEs versus SNR for the 12-element 2-D array ($R = 2$) from Figure 9.2, and $N = 20$, $d = 3$ correlated sources ($\rho = 0.9$) at $\mu_1^{(1)} = 0.25$, $\mu_2^{(1)} = 0.5$, $\mu_3^{(1)} = 0.75$, $\mu_1^{(2)} = 0.25$, $\mu_2^{(2)} = 0.5$, $\mu_3^{(2)} = 0.75$ with rotation phases $\varphi_1 = 0$, $\varphi_2 = \pi/4$, and $\varphi_3 = \pi/2$.

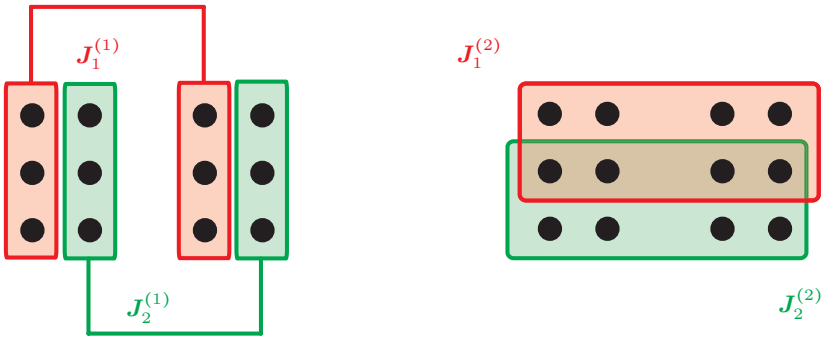


Figure 9.2.: 2-D shift invariance for the depicted centro-symmetric 4×3 sampling grid, left: subarrays for the first (horizontal) dimension, right: subarrays for the second (vertical) dimension.

positions are irrelevant and have no impact on the performance. The remaining parameters are given by $N = 10$, $\Delta\varphi = \pi/3$, $\delta = (M-1)/2$, i.e., the phase reference is located at the first sensor element, and $\sigma_n^2 = 0.032$. Moreover, the correlation coefficient ρ is set to $\rho = 0.8$. It is evident that the analytical results agree well with the empirical estimation errors and that both CRBs perform similarly for large M.

In Figure 9.4, we depict the RMSE as a function of the phase separation $\Delta\varphi \in [0, \pi]$. The results are shown for the phase references $\delta = (M-1)/2$, $\delta = 0$, and $\delta = -(M-1)/2$, which correspond to

Table 9.1.: RMSE for a varying number of sources with $M = 4$

RMSE	$d = 1$	$d = 2$	$d = 3$	$d = 4$	$d = 5$	$d = 6$	$d = 7$
CRB	0.02	0.13	0.80	∞	∞	∞	∞
NC CRB	0.02	0.11	0.12	0.14	0.35	2.93	∞

the phase reference at the first element, the center, and the last element of the array, respectively. Additionally, we set $M = 9$ and $\rho = 0$. The remaining parameters are kept the same. It is apparent from Figure 9.4 that the analytical curves match very well with the empirical ones. Furthermore, the Det NC CRB not only varies with the phase separation but also depends on the phase reference of the array, while the Det CRB is independent of the phase separation as $\rho = 0$. For instance, if $\delta = 0$, the improvement of the Det NC CRB over the Det CRB is maximal for $\Delta\varphi = \pi/2$ and vanishes for $\Delta\varphi = 0 + k \cdot \pi,\ k \in \mathbb{N}$.

Figure 9.5 shows the RMSE as a function of the source correlation ρ for the same phase reference values as in Figure 9.4. We set $\Delta\varphi = \pi/4$ and the remaining parameters are kept the same. It is clear from Figure 9.5 that in addition to the Det NC CRB, the Det CRB also depends on the phase reference for $\rho > 0$. Moreover, for all values of δ, the Det NC CRB provides the greatest improvement over the Det CRB if the two sources are uncorrelated ($\rho = 0$) and reduces to the Det CRB if the sources are coherent ($\rho = 1$). Again, the analytical curves coincide with the empirical ones.

Finally, Figure 9.6 illustrates the asymptotic NC gain in (9.48) for $d = 2$ sources as a function of $\Delta\mu$. The number of sensors is fixed to $M = 15$ and we set $\rho = 0$, $\Delta\varphi = \pi/2$, as well as $\delta = 0$. The remaining parameters are kept the same. For comparison purposes, we have also included the curves for the analytical NC gain of NC SE from [SRH14a]. It can be seen that the NC gain expression becomes accurate for small $\Delta\mu$ and that it is largest when $\Delta\mu$ goes to zero. Furthermore, the NC gain of NC SE is close to the maximum achievable NC gain computed from the Det NC CRB.

9.5.3. Summary of the results

In this subsection, we briefly summarize the main properties of the deterministic NC CRB:

- For coherent sources, equal rotation phases (equality modulo π), a single source ($d = 1$), or a single snapshot ($N = 1$), the deterministic NC CRB is equal to the CRB ($\boldsymbol{C}^{(\mathrm{nc})} = \boldsymbol{C}$), which suggests that there is no gain from non-circular sources in these cases.

- For $d = 2$ sources if $\hat{\rho} = 0$, $\Delta\varphi = \pi/2$, and $\delta = 0$, the deterministic NC CRB (9.43) entirely decouples, i.e., it is independent of $\Delta\mu$. Thus, for each source the CRB is the same as if the other source was not present, which provides the maximum NC gain (9.49).

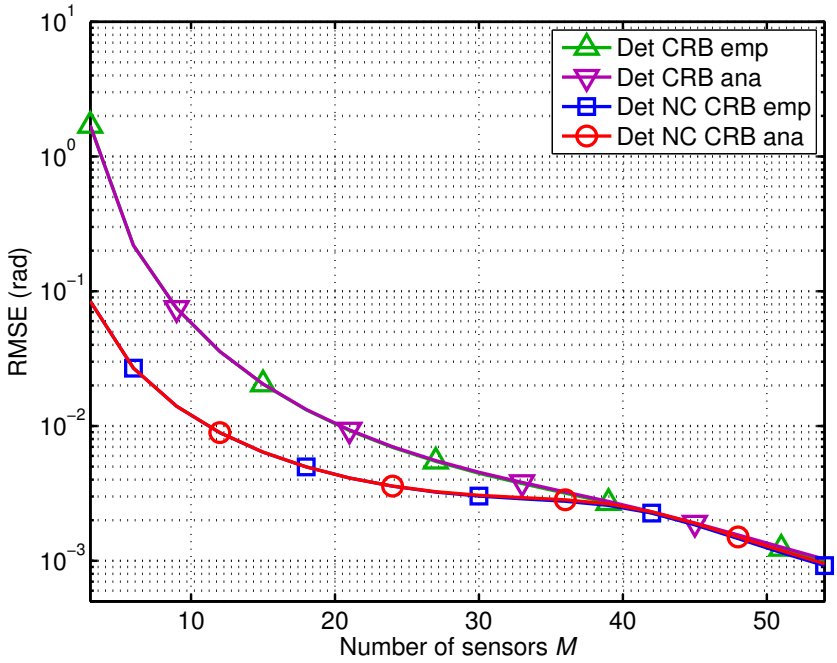


Figure 9.3.: Analytical and empirical RMSEs versus the number of sensors M for $d = 2$ correlated sources with $N = 10$, $\Delta\mu = 0.1$ rad, $\rho = 0.8$, $\Delta\varphi = \pi/3$, $\delta = (M-1)/2$, $P_1 = 1.5$, $P_2 = 0.5$, and $\sigma_n^2 = 0.032$.

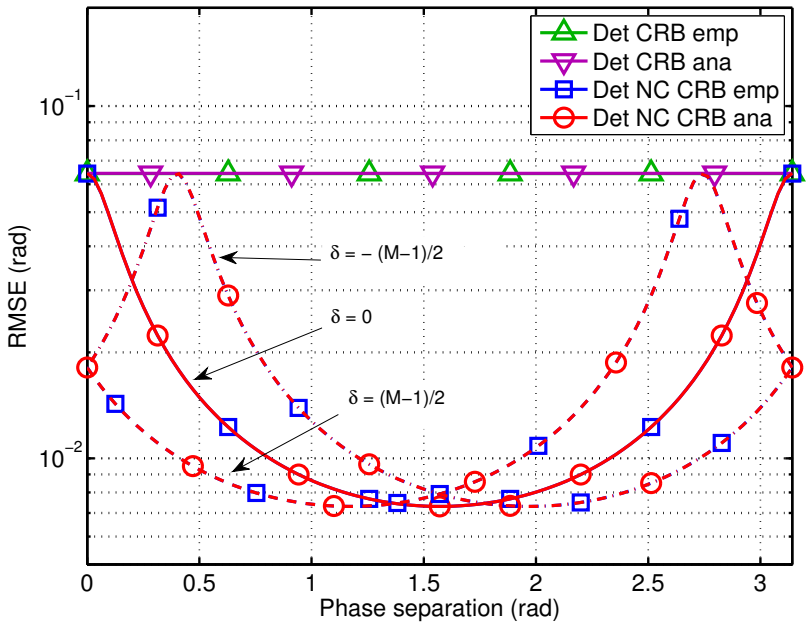


Figure 9.4.: RMSE versus the phase separation $\Delta\varphi$ for $d = 2$ sources and varying δ with $M = 9$, $N = 10$, $\Delta\mu = 0.1$ rad, $\rho = 0$, $P_1 = 1.5$, $P_2 = 0.5$, and $\sigma_n^2 = 0.032$.

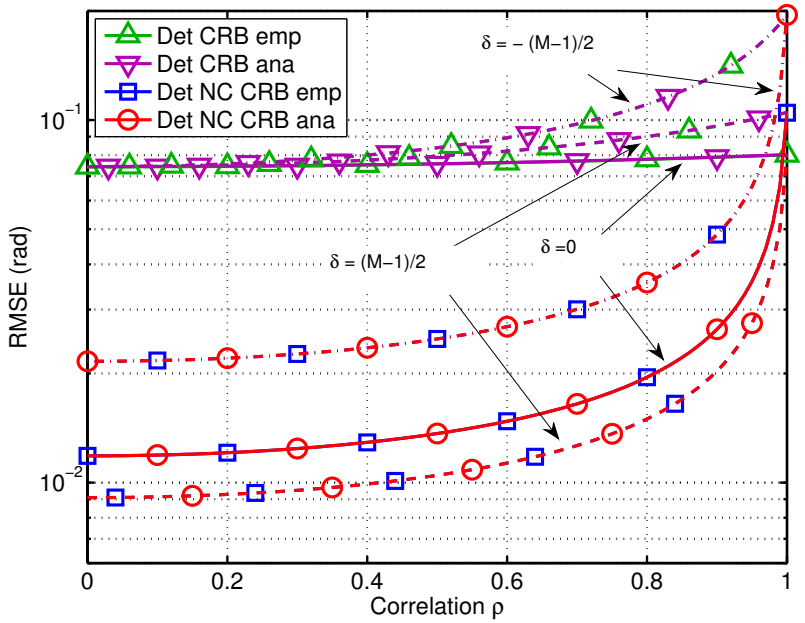


Figure 9.5.: RMSE versus the source correlation ρ for $d = 2$ sources and varying δ with $M = 9$, $N = 10$, $\Delta\mu = 0.1$ rad, $\Delta\varphi = \pi/4$, $P_1 = 1.5$, $P_2 = 0.5$, and $\sigma_n^2 = 0.032$.

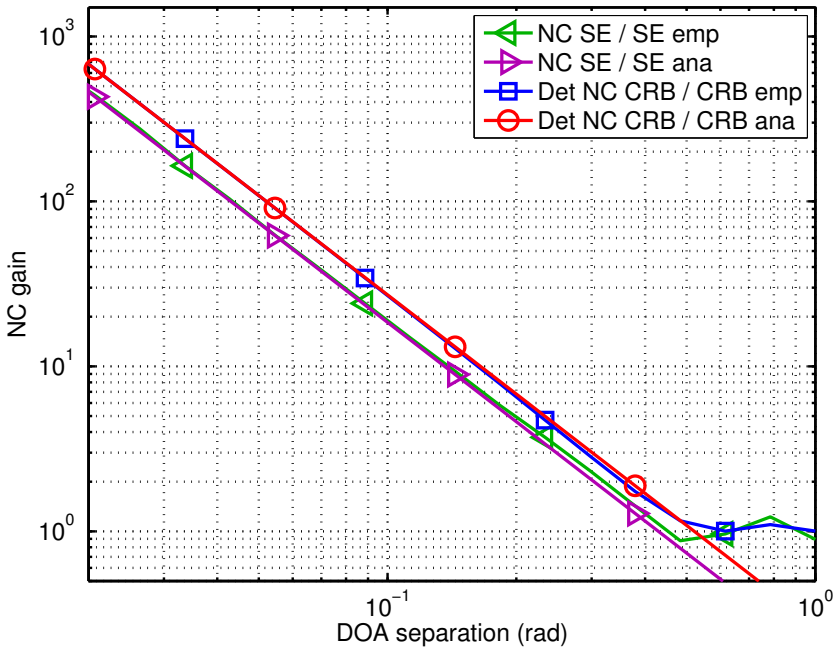


Figure 9.6.: Analytical and empirical NC gain versus the source separation $\Delta\mu$ for $d = 2$ uncorrelated sources with $M = 15$, $N = 10$, $\Delta\varphi = \pi/2$, $\delta = 0$, $P_1 = 1.5$, $P_2 = 0.5$, and $\sigma_n^2 = 0.032$.

- For $d = 2$ sources if $\hat{\rho} < 1$ and $\Delta\varphi > 0$, the deterministic NC CRB approaches a finite value (9.44) as the source separation $\Delta\mu$ approaches zero while the deterministic CRB tends to

infinity (9.47). Thus, the NC gain (9.50) increases when the source separation decreases.

- For $d = 2$ sources if $\hat{\rho} < 1$ and $\Delta\varphi > 0$, the NC gain also depends on the phase reference δ of the array (see Figure 9.5 and Figure 9.4). It is largest, for instance, for $\Delta\varphi = \pi/2$ when $\delta = 0$.
- The number of resolvable sources d may exceed the number of sensors M. If the sources are non-coherent and their rotation phases are different, $d \leq 2(M-1)$ represents the upper limit on the number of resolvable sources.

9.6. Summary

In this chapter, we have presented a closed-form expression of the deterministic R-D NC Cramér-Rao bound for multi-dimensional strictly non-circular (rectilinear) signals. This bound serves as a benchmark for the recently developed algorithms, e.g., R-D NC Standard ESPRIT and R-D NC Unitary ESPRIT, that exploit the NC structure of such strictly non-circular signals and thus outperform the traditional methods for arbitrary signals. Based on the resulting R-D NC CRB expression and assuming the R-D array to be separable and centro-symmetric, we have shown that in the special cases of equal phases, full coherence of the strictly non-circular signals, a single snapshot or for a single strictly non-circular source, the deterministic R-D NC CRB reduces to the existing deterministic R-D CRB for arbitrary signals. This suggests that no NC gain can be achieved in these specific cases. Furthermore, we have simplified the derived NC CRB and the existing CRB for the special case of two closely-spaced strictly non-circular signals captured by a uniform linear array (ULA). With these simplified CRB expressions, we have then analytically computed the maximum achievable asymptotic NC gain for this scenario. The resulting expression only depends on the various physical parameters, e.g., the number of sensors, the signal correlation, etc. Additionally, we have analyzed the dependence of the NC gain on these parameters to find that the largest NC gain is obtained if the two sources are closely-spaced, incoherent, and have a non-vanishing phase discrimination.

10. Conclusion and future work

In this thesis, we have addressed the multi-dimensional harmonic retrieval problem and show how exploiting signal structure can improve the estimation accuracy of ESPRIT-type and sparsity-based parameter estimation algorithms. Specifically, we have focused on the multi-dimensional signal structure and the strictly non-circular signal structure and developed matrix-based and tensor-based R-D NC ESPRIT-type algorithms with and without spatial smoothing. Note that the concepts demonstrated for ESPRIT-type algorithms can also be applied to other subspace-based parameter estimation algorithms, e.g., MUSIC [Sch79], MODE [Van02], or RARE [PGW02]. In generalized least squares, we have also developed an entirely new algorithm to solve the shift invariance equation in ESPRIT-type algorithms. Moreover, we have shown that the NC signal structure can also be exploited in parameter estimation based on sparse recovery algorithms. As benchmark for the ESPRIT-type and the sparsity-based algorithms for NC signals, we have derived the corresponding deterministic NC Cramér-Rao bound. In the conclusion of the thesis, we provide a summary of the specific contributions in this thesis in Section 10.1 and outline possible research directions for future work in Section 10.2.

10.1. Summary of contributions

In this section, we explicitly enumerate the various contributions in this thesis. We also give a literature overview of the various ESPRIT-type algorithms along with their performance analysis in Table 10.1. Table 10.2 provides a similar overview for the various least squares algorithms to solve the overdetermined shift invariance equation of ESPRIT-type algorithms and the corresponding performance analysis expressions. The columns "proposed" and "performance analysis" shows where the algorithm was originally published and where the corresponding analytical performance evaluation was first presented. Here "(open)" means that there are so far no results available in the literature. Additionally, we give an overview of the special cases of a single source ($d = 1$) and two sources ($d = 2$) for the performance analysis expressions of the various ESPRIT-type algorithms in Table 10.3. In Table 10.4, we provide a similar overview of the special cases of a single source ($d = 1$) and two sources ($d = 2$) for the performance analysis expressions of the various least squares algorithms. Our contributions [SRH14a], [SRH17a], [SRHD14], [SRHD17], [SRHD16] listed in these tables are presented in this thesis in the Chapters 4, 5, 6, 7, 9, respectively.

The main contributions in this thesis are:

- The simplification of the analytical MSE expressions from the performance analysis of the matrix-based and tensor-based R-D ESPRIT-type algorithms using LS in Section 4.5 for a single source and two sources. The resulting expressions only depend on the physical parameters,

Algorithm	Proposed	Performance analysis based on	
		[Bri75]	[LLV93]
1-D Standard ESPRIT	[RPK86]	[RH89a]	[LLV93], ...
1-D Unitary ESPRIT	[HN95]	[MHZ96]	[RBH10]
R-D Unitary ESPRIT	[HN98]	[MHZ96] (2-D)	[RBH10]
R-D Standard/Unitary Tensor-ESPRIT	[HRD08]		[RBHW09, RBH10]
R-D Standard/Unitary ESPRIT with spatial smoothing	[HN98]		[SRHD17]
1-D NC Standard ESPRIT	[ZCW03]		[SRHD14]
R-D NC Unitary ESPRIT	[HR04]		[SRHD14]
R-D NC Standard/Unitary Tensor-ESPRIT	[RH09]		cf. Sec. 6.5
R-D NC Standard/Unitary ESPRIT with spatial smoothing	[SRHD17]		[SRHD17]

Table 10.1.: Literature overview of ESPRIT-type algorithms and their performance analysis

Algorithm	Proposed	Performance analysis based on	
		[Bri75]	[LLV93]
Least Squares (LS)	[RPK86]	[RH89a]	[LLV93]
Total Least Squares (TLS)	[RK87]	[RH89a, OVK91]	
1-D Structured Least Squares (SLS)	[Haa97b]		[RH11]
R-D Structured Least Squares (SLS)	[Haa97b]	(open)	
R-D Tensor-Structure SLS (TS-SLS)	[RH07b]	(open)	
R-D Generalized Least Squares (GLS)	[SRH17a]		cf. Sec. 5.4

Table 10.2.: Literature overview of least squares algorithms to solve the shift invariance equation of ESPRIT-type algorithms and their performance analysis

Algorithm	Single source based on		Two sources based on	
	[Bri75]	[LLV93]	[Bri75]	[LLV93]
1-D Standard ESPRIT	[RH89a]			[SRH14a]
R-D Standard/Unitary ESPRIT		[RH12]		cf. Sec. 4.5.3
R-D Standard/Unitary Tensor-ESPRIT		[RH12]		cf. Sec. 4.5.4
R-D Standard/Unitary ESPRIT with spatial smoothing		[SRHD17]		
1-D NC Standard ESPRIT		[SRHD14]		[SRH14a]
R-D NC Standard/Unitary ESPRIT		[SRHD14]		[SRHD14]
R-D NC Standard/Unitary Tensor-ESPRIT		cf. Sec. 6.6.1		cf. Sec. 6.6.2
R-D NC Standard/Unitary ESPRIT with spatial smoothing		[SRHD17]		
R-D Det. Cramér-Rao bound (CRB)		[RH12]		[SRHD16]
R-D NC Det. Cramér-Rao bound (CRB)		[SRHD14]		[SRHD16]

Table 10.3.: Literature overview of the performance analysis of ESPRIT-type algorithms for the special cases of a single source and two sources

Algorithm	Single source based on		Two sources based on	
	[Bri75]	[LLV93]	[Bri75]	[LLV93]
Least Squares (LS)	[RH89a]			[SRH14a]
Total Least Squares (TLS)	[OVK91]			
1-D Structured Least Squares (SLS)		[RH11]		
1-D Generalized Least Squares (GLS)		cf. Sec. 5.5.1		cf. Sec. 5.5.2

Table 10.4.: Literature overview of the performance analysis of least squares algorithms to solve the shift invariance equation of ESPRIT-type algorithms for the special cases of a single source and two sources

e.g., the number of antennas, the SNR, etc. For a single source, all ESPRIT-type algorithms yield the same MSE and perform identical. For the two sources case of the matrix-based R-D ESPRIT-type algorithms, we analytically compute the gain from forward-backward averaging and analyze its dependence on the physical parameters. For the simplified MSE of the tensor-based R-D ESPRIT-type algorithms, we compute and analyze the tensor gain and the forward-backward averaging gain in the tensor case. Moreover, we compute the asymptotic efficiency, i.e., the ratio of the CRB and the MSE, for $R = 1$.

- The general least squares algorithm (GLS) to solve the shift invariance equation in R-D ESPRIT-type algorithms, which has been published in [SRH17a] and derived in Section 5.3. GLS takes the statistics of the subspace estimation error into account for the solution of the shift invariance equation, where we assume a ULA with maximum subarray overlap and circularly symmetric white noise. We show that in some cases, GLS provides an optimal solution to the shift invariance equation.

- The first-order performance analysis of GLS-based R-D ESPRIT-type algorithms presented in Section 5.4. We obtain analytical expressions for the parameter estimation error and the MSE of R-D Standard ESPRIT and R-D Unitary ESPRIT both with GLS. We prove that the real-valued transformation in GLS-based R-D Unitary ESPRIT has no effect on the performance in the asymptotic region of the high effective SNR.

- Simplified MSE expressions of the GLS-based R-D ESPRIT-type algorithms for a single source and two orthogonal sources (cf. Section 5.5), which only depend on the physical parameters of significance, e.g., the SNR, the number of sensors M, and the sample size N.

- We use the simplified MSE expressions for a single source as well as two orthogonal sources in Section 5.5 to analytically show that the GLS-based Standard ESPRIT algorithm is asymptotically efficient in both cases, i.e., they achieve the deterministic Cramér-Rao bound (CRB). Specifically, we show that the respective simplified MSE expressions can be reformulated into the corresponding CRB expressions for both cases.

- The matrix-based R-D NC Standard ESPRIT and R-D NC Unitary ESPRIT algorithms for strictly non-circular sources using least squares discussed in Section 6.2. We show that due to the NC preprocessing, the virtual array is always centro-symmetric, making R-D NC Unitary ESPRIT also applicable to non-centro-symmetric physical arrays.

- The R-D NC Tensor-ESPRIT-type algorithms for strictly non-circular sources using least squares introduced in Section 6.3. As in the matrix case, the virtual array steering tensor is always centro-symmetric after the NC tensor preprocessing such that R-D NC Unitary Tensor-ESPRIT is also applicable to R-D arrays that are not centro-symmetric.

- The first-order performance analysis of the matrix-based R-D NC ESPRIT-type algorithms, which is presented in Section 6.4. We derive analytical expressions for the parameter estimation error as well as the MSE of both algorithms. We also prove that R-D NC Standard ESPRIT and R-D NC Unitary ESPRIT provide the same asymptotic performance.

- The first-order performance analysis of the R-D NC Tensor-ESPRIT-type algorithms proposed in Section 6.5. We find analytical expressions for the parameter estimation error as well as the MSE of both algorithms. Similarly to the matrix case, R-D NC Standard Tensor-ESPRIT and R-D NC Unitary Tensor-ESPRIT are proven to perform asymptotically identical.

- The simplification of the MSE expressions for the matrix-based and tensor-based R-D NC ESPRIT-type algorithms using LS (cf. Section 6.6) for a single source and two sources to obtain compact formulas in terms of the physical parameters, e.g., the number of antennas M, the SNR, and the sample size N. For a single source, we obtain the same MSE as for the conventional ESPRIT-type algorithms, i.e., there is no gain from NC signals in both the matrix and the tensor case. In the case of two sources, we analytically compute the NC gain for the matrix-based and tensor-based R-D NC ESPRIT-type algorithms and analyze the dependence on the physical parameters. Preliminary results have been published in [SRH14a]. Note that in the tensor case, we additionally have the tensor gain.

- The performance analysis of R-D Standard ESPRIT and R-D Unitary ESPRIT as well as their NC versions with spatial smoothing, which has been published in [SRHD17] and described in Chapter 7. We show that both R-D NC ESPRIT-type algorithms with spatial smoothing perform asymptotically identical. Moreover, we simplify the analytical MSE expressions for a single source, where all the algorithms provide the same MSE. Finally, we analytically derive the optimal number of subarrays for spatial smoothing that minimize the MSE for a single source.

- Three different algorithms [SRH16c, SSPH16, SRS$^+$16] derived in Chapter 8 to exploit the strictly non-circular signal structure in sparsity-based parameter estimation using sparse signal recovery (SSR). While [SRH16c] solves a 2-D grid-based SSR problem, [SSPH16] only requires solving a 1-D grid-based SSR problem, and in [SRS$^+$16], we propose a gridless solution based on atomic norm minimization.

- A low-complexity grid-offset estimation procedure [SRH16c, SSPH16] for the two grid-based SSR algorithms for strictly non-circular signals is derived in Chapter 8. Thereby, the critical off-grid problem can be solved.

- The derivation of the deterministic R-D NC CRB for strictly non-circular signals published

in [SRHD17] and presented in Chapter 9. The deterministic R-D NC CRB provides a lower limit on the estimation error and constitutes a benchmark for the presented NC algorithms in this thesis. We show for the special cases of full coherence, a single snapshot, and a single strictly non-circular source that the deterministic R-D NC CRB reduces to the deterministic R-D CRB for arbitrary signals, which implies that no NC gain can be achieved in these cases.

- The simplification of both the 1-D NC CRB and the 1-D CRB for the special case of two closely-spaced NC signals and a uniform linear array (ULA) to obtain compact formulas in terms of the physical parameters. Based on these simplified expressions, we analytically compute the CRB-based NC gain and study its behavior in terms of the physical parameters.

10.2. Future work

From Table 10.1 and Table 10.2 as well as from Table 10.3 and Table 10.4, it is easy to identify the open aspects that are left for future work.

The performance analysis framework presented in the Chapters 4, 6, and 7 can still be extended in several ways. For instance, the tensor-based spatial smoothing (TBSS) technique from [THRG10, THG09b, THG09a] has so far not been incorporated into the performance analysis. The same applies to the R-D TS-SLS-based Tensor-ESPRIT-type algorithms from [RH07b]. Regarding the study of special cases on the number of sources to obtain simplified MSE expressions, there are a number of open cases as well. For instance, the respective two source cases of R-D ESPRIT-type algorithms with spatial smoothing and R-D ESPRIT-type algorithms with structured least squares are still to be considered. Especially, for scenarios, where forward-backward averaging (FBA) cannot be applied, the optimal number of subarrays for spatial smoothing for two sources can be derived analytically.

A very important direction for future work is the further advancement and analysis of the generalized least squares algorithm for ESPRIT-type parameter estimation developed in Chapter 5. Using the presented performance analysis of GLS for Standard ESPRIT, we have derived simplified MSE expressions for the special cases of a single source and two orthogonal sources assuming a ULA with maximum overlap. So far, we have only proven that GLS for 1-D Standard ESPRIT is asymptotically efficient, i.e., the ratio of the CRB and the MSE is equal to 1, in these cases. It would be interesting to generalize the proofs to an arbitrary number of signals. Moreover, we have observed via simulations that incorporating forward-backward averaging can achieve asymptotic efficiency even for correlated but incoherent signals. This is still to be shown analytically. Moreover, another crucial research direction would be the extension of the matrix-based R-D ESPRIT-type algorithms with GLS to the tensor case. It is suspected that asymptotic efficiency can also be achieved in the tensor case. However, this topic is still entirely open.

The presented R-D NC ESPRIT-type algorithms assume that all the signals are strictly non-circular. In [SRH15c], we have already relaxed this restriction to the case of coexisting circular and strictly non-circular signals and developed C-NC ESPRIT-type algorithms. However, it would be desirable to find a way for ESPRIT-type algorithms to handle the even more general case of weak-sense non-circular signals as defined in Section 2.2.1.2. This topic is still open in the literature.

Another promising research direction is to further extend our initial results on the second-order performance analysis for 1-D Standard ESPRIT in [SRH17c]. Second-order performance analysis expressions can be used to analytically describe the performance in lower SNR and snapshot regions. Moreover, they can be used to compute the threshold point as shown in [SRH17c]. Since this is an entirely new field, many ways for extensions are possible. First, a second-order performance analysis of 1-D Unitary ESPRIT can be derived. Then, extensions to the matrix-based and tensor-based R-D case of both algorithms are desirable. Moreover, additional preprocessing steps such as spatial smoothing and the NC preprocessing for strictly non-circular sources can be included. Finally, various least squares methods such as structured least squares or the novel generalized least squares algorithm derived in Chapter 5 can be incorporated.

Finally, another research direction is to extend the concept of exploiting the signals' non-circularity in sparsity-based parameter estimation algorithm. For instance, an interesting topic would be to consider the practical case of coexisting circular and non-circular signals. Another topic would be to develop sparsity-based algorithms that can exploit the weak-sense non-circularity of the signals. Furthermore, exploiting the strict non-circularity structure can be combined with additional signal structure. One example would be integrality constraints on the symbols, which exploit the fact that the real-valued symbols can only take certain points in the constellation diagram. Moreover, additional constraints on the measurement system can be combined with exploiting the NC structure such as magnitude-only measurements, also known as phase retrieval, or quantized measurements.

Appendix A.

Glossary of acronyms, symbols and notation

In this appendix, we provide a list of acronyms and summarize the frequently used symbols and notation in this thesis.

A.1. Acronyms

1-D	One-Dimensional
2-D	Two-Dimensional
3-D	Three-Dimensional
ANM	Atomic Norm Minimization
ASK	Amplitude Shift Keying
AWGN	Additive White Gaussian Noise
BLUE	Best Linear Unbiased Estimator
BPSK	Binary Phase Shift Keying
CDMA	Code Division Multiple Access
CRB	Cramér-Rao Bound
DFT	Discrete Fourier Transform
DOA	Direction Of Arrival
DOD	Direction Of Departure
DOF	Degrees Of Freedom
EADF	Effective Aperture Distribution Function
EEG	Electroencephalography
ESPRIT	Estimation of Signal Parameters via Rotational Invariance Techniques
EVD	EigenValue Decomposition
FBA	Forward Backward Averaging
FIM	Fisher Information Matrix
GLS	General Least Squares
GNSS	Global Navigation Satellite System
HOSVD	Higher-Order Singular Value Decomposition
LS	Least Squares
MEG	Magnetoencephalography
MIMO	Multiple Input Multiple Output
MLSVD	Multi-linear Singular Value Decomposition
MMV	Multiple Measurement Vectors
MODE	Method of Direction of Arrival Estimation
MUSIC	Multiple Signal Classification
MSE	Mean Squared Error
MSK	Minimum Shift Keying
MMSE	Minimum Mean Squared Error
NC	Non-Circular
NMR	Nuclear Magnetic Resonance
OFDM	Orthogonal Frequency Division Multiplexing

OFDMA	Orthogonal Frequency Division Multiple Access
OQPSK	Offset Quadrature Phase Shift Keying
OW	Optimally Weighted
PARAFAC	Parallel Factor Analysis
PCA	Principle Component Analysis
PDF	Probability Density Function
PRIME	Polynomial Root Intersection for Multi-dimensional Estimation
PUMA	Principle-Singular-Vector Utilization For Modal Analysis
QAM	Quadrature Amplitude Modulation
QPSK	Quadrature Phase Shift Keying
RARE	Rank-Reduction Estimator
R**-D**	R-Dimensional
RMSE	Root Mean Squared Error
SDS	Semi-Definite Programming
SLS	Structured Least Squares
SMD	Simultaneous Matrix Diagonalization
SMV	Single Measurement Vector
SNR	Signal-to-Noise Ratio
SO	Second-Order
SSR	Sparse Signal Reconstruction
SV	Singular Value
SVD	Singular Value Decomposition
TDOA	Time Delay Of Arrival
TLS	Total Least Squares
TS-SLS	Tensor-Structure Structured Least Squares
UCA	Uniform Circular Array
ULA	Uniform Linear Array
URA	Uniform Rectangular Array
WLS	Weighted Least Squares
ZMCSCG	Zero Mean Circularly Symmetric Complex Gaussian

A.2. Symbols and notation

$\mathbb{R}$	Set of real numbers
$\mathbb{C}$	Set of complex numbers
$\mathbb{Z}$	Set of integer numbers
e, π, j	Euler's number, π, and imaginary unit: $\mathrm{e}^{\mathrm{j}\pi} + 1 = 0$
$\doteq$	Definition
a, b, c	scalars
$\boldsymbol{a}$, $\boldsymbol{b}$, $\boldsymbol{c}$	column vectors
$\boldsymbol{A}$, $\boldsymbol{B}$, $\boldsymbol{C}$	matrices
$\boldsymbol{\mathcal{A}}$, $\boldsymbol{\mathcal{B}}$, $\boldsymbol{\mathcal{C}}$	tensors
$\mathrm{Re}\{x\}$	Real part of complex variable x
$\mathrm{Im}\{x\}$	Imaginary part of complex variable x

$\arg\{x\}$	Argument (phase) of complex variable x
x^*	Complex conjugate of x
$\mathbf{0}_{M\times N}$	Matrix of zeros of size $M \times N$
$\mathbf{1}_{M\times N}$	Matrix of ones of size $M \times N$
$\boldsymbol{I}_M$	Identity matrix of size $M \times M$
$\boldsymbol{\mathcal{I}}_{R,d}$	R-way identity tensor of size $d \times d \ldots \times d$
$\boldsymbol{\Pi}_M$	Exchange of size $M \times M$ with ones on its anti-diagonal and zeros elsewhere
$\boldsymbol{Q} \succcurlyeq \mathbf{0}$	$\boldsymbol{Q}$ is a positive-semidefinite matrix
$[\boldsymbol{A}]_{(i,j)}$	The (i,j)-element of the matrix $\boldsymbol{A}$
$[a_i]_{i=1,2,\ldots,I}$	An $I \times 1$ column vector $\boldsymbol{a}$ with i-th element a_i
$(\cdot)^{\mathrm{T}}$	matrix transpose
$(\cdot)^{\mathrm{H}}$	Hermitian transpose
$\|.\|_2$	Euclidean (two-) norm
$\|.\|_{\mathrm{F}}$	Frobenius norm
$\|.\|_{\mathrm{H}}$	Higher-Order (Frobenius) norm
$\boldsymbol{A} \otimes \boldsymbol{B}$	Kronecker product between $\boldsymbol{A} \in \mathbb{C}^{M\times N}$ and $\boldsymbol{B} \in \mathbb{C}^{P\times Q}$ defined as

$$\boldsymbol{A} \otimes \boldsymbol{B} = \begin{bmatrix} a_{1,1}\cdot\boldsymbol{B} & a_{1,2}\cdot\boldsymbol{B} & \cdots & a_{1,N}\cdot\boldsymbol{B} \\ a_{2,1}\cdot\boldsymbol{B} & a_{2,2}\cdot\boldsymbol{B} & \cdots & a_{2,N}\cdot\boldsymbol{B} \\ \vdots & \vdots & \vdots & \vdots \\ a_{M,1}\cdot\boldsymbol{B} & a_{M,2}\cdot\boldsymbol{B} & \cdots & a_{M,N}\cdot\boldsymbol{B} \end{bmatrix}.$$

$\boldsymbol{A} \diamond \boldsymbol{B}$	Khatri-Rao (column-wise Kronecker) product between $\boldsymbol{A} \in \mathbb{C}^{M\times N}$ and $\boldsymbol{B} \in \mathbb{C}^{P\times N}$
$\boldsymbol{A} \odot \boldsymbol{B}$	Schur (element-wise) product between $\boldsymbol{A} \in \mathbb{C}^{M\times N}$ and $\boldsymbol{B} \in \mathbb{C}^{M\times N}$
$\mathrm{vec}\{\cdot\}$	vec-operator: stack elements of a matrix/tensor into a column vector, begin with first (row) index, then proceed to second (column), third, etc.
$\mathrm{unvec}_{I\times J}\{\cdot\}$	inverse vec-operator: reshape elements of a vector back into a matrix/tensor of indicated size
$\mathrm{diag}\{\cdot\}$	transform a vector into a square diagonal matrix or extract main diagonal of a square matrix and place elements into a vector
$\mathrm{blkdiag}\{\cdot,\ldots,\cdot\}$	creates a block diagonal matrix from its (matrix) arguments
$\mathrm{Tr}\{\cdot\}$	trace of a matrix (sum of diagonal elements = sum of eigenvalues)
$\det\{\cdot\}$	determinant of a matrix (product of eigenvalues)
$\mathrm{rank}\{\cdot\}$	rank of a matrix
$\mathrm{EV}_k\{\cdot\}$	k-th eigenvalue of a matrix
	the same space as the columns of $\boldsymbol{A} \in \mathbb{C}^{M\times r}$ (assuming $r \leq M$)

$\boldsymbol{A}^{+}$ Moore-Penrose pseudo inverse [Moo20, Pen55] of a matrix $\boldsymbol{A} \in \mathbb{C}^{M \times N}$, which we can compute via

$\boldsymbol{A}^{+} = \boldsymbol{V}_{\mathrm{s}} \cdot \boldsymbol{\Sigma}_{\mathrm{s}}^{-1} \cdot \boldsymbol{U}_{\mathrm{s}}^{\mathrm{H}}$, where $\boldsymbol{A} = \boldsymbol{U}_{\mathrm{s}} \cdot \boldsymbol{\Sigma}_{\mathrm{s}} \cdot \boldsymbol{V}_{\mathrm{s}}^{\mathrm{H}}$ represents the economy-size SVD of $\boldsymbol{A}$.

$\boldsymbol{A}^{+} = (\boldsymbol{A}^{\mathrm{H}} \cdot \boldsymbol{A})^{-1} \cdot \boldsymbol{A}^{\mathrm{H}}$ if $\operatorname{rank}\{\boldsymbol{A}\} = N$ (full column rank)

$\boldsymbol{A}^{+} = \boldsymbol{A}^{\mathrm{H}} \cdot (\boldsymbol{A} \cdot \boldsymbol{A}^{\mathrm{H}})^{-1}$ if $\operatorname{rank}\{\boldsymbol{A}\} = M$ (full row rank).

$[\boldsymbol{\mathcal{X}}]_{(n)}$ n-mode unfolding of tensor $\boldsymbol{\mathcal{X}}$ in reverse cyclical column ordering

$\boldsymbol{\mathcal{X}} \times_n \boldsymbol{U}$ n-mode product between tensor $\boldsymbol{\mathcal{X}}$ and matrix $\boldsymbol{U}$

$\boldsymbol{\mathcal{X}} \bigtimes_{r=1}^{R} \boldsymbol{U}_r$ repeated n-mode products, short-hand notation for $\boldsymbol{\mathcal{X}} \times_1 \boldsymbol{U}_1 \ldots \times_R \boldsymbol{U}_R$

$[\boldsymbol{\mathcal{A}} \llcorner_n \boldsymbol{\mathcal{B}}]$ n-mode concatenation of tensors $\boldsymbol{\mathcal{A}}$ and $\boldsymbol{\mathcal{B}}$

$\mathbb{E}\{X\}$ Expectation operator, i.e., mean of the random variable X

$\mathcal{N}(\mu, \sigma^2)$ Gaussian distribution with mean μ, variance σ^2

$\mathcal{CN}(\mu, \sigma^2)$ circularly symmetric complex Gaussian distribution

$\boldsymbol{Q}_p$ Left-$\boldsymbol{\Pi}$-real matrix satisfying $\boldsymbol{\Pi}_p \cdot \boldsymbol{Q}_p^{*} = \boldsymbol{Q}_p$.

$\boldsymbol{Q}_p^{(\mathrm{s})}$ Unitary sparse left-$\boldsymbol{\Pi}$-real given by [HN95]

$$\boldsymbol{Q}_{2n}^{(\mathrm{s})} = \frac{1}{\sqrt{2}} \begin{bmatrix} \boldsymbol{I}_n & \mathrm{j}\boldsymbol{I}_n \\ \boldsymbol{\Pi}_n & -\mathrm{j}\boldsymbol{\Pi}_n \end{bmatrix} \quad \text{and} \quad \boldsymbol{Q}_{2n+1}^{(\mathrm{s})} = \frac{1}{\sqrt{2}} \begin{bmatrix} \boldsymbol{I}_n & \boldsymbol{0}_{n\times 1} & \mathrm{j}\boldsymbol{I}_n \\ \boldsymbol{0}_{n\times 1}^{\mathrm{T}} & \sqrt{2} & \boldsymbol{0}_{n\times 1}^{\mathrm{T}} \\ \boldsymbol{\Pi}_n & \boldsymbol{0}_{n\times 1} & -\mathrm{j}\boldsymbol{\Pi}_n \end{bmatrix},$$

for even and for odd order, respectively.

Appendix B.

Proofs and derivations

In this appendix, we provide a number of proofs and derivations, which are listed here to increase the readability in the main part of the thesis.

B.1. Proof of Theorem 2.2.1

We start the proof of Theorem 2.2.1 in Section 2.2.1.2 by expanding the non-circularity rate ζ of the complex random variable Z and insert the real part X and the imaginary part Y of Z.

$$\zeta = \frac{\mathbb{E}\left\{Z^2\right\}}{\mathbb{E}\left\{|Z|^2\right\}} = \frac{\mathbb{E}\left\{X^2 - Y^2 + 2 \cdot \mathrm{j} \cdot X \cdot Y\right\}}{\mathbb{E}\left\{X^2 + Y^2\right\}} = \frac{\mathbb{E}\left\{X^2\right\} - \mathbb{E}\left\{Y^2\right\} + 2 \cdot \mathrm{j} \cdot \mathbb{E}\left\{X \cdot Y\right\}}{\mathbb{E}\left\{X^2\right\} + \mathbb{E}\left\{Y^2\right\}}. \tag{B.1}$$

Let us introduce the short hand notations $\sigma_x^2 = \mathbb{E}\left\{X^2\right\}$, $\sigma_y^2 = \mathbb{E}\left\{Y^2\right\}$, and $\sigma_{xy} = \mathbb{E}\left\{X \cdot Y\right\}$. Then, the squared magnitude of ζ can be expressed as

$$\begin{aligned} |\zeta|^2 &= \left|\frac{\sigma_x^2 - \sigma_y^2}{\sigma_x^2 + \sigma_y^2}\right|^2 + \left|\frac{2\sigma_{xy}}{\sigma_x^2 + \sigma_y^2}\right|^2 = \frac{\left(\sigma_x^2 - \sigma_y^2\right)^2 + 4\sigma_{xy}^2}{\left(\sigma_x^2 + \sigma_y^2\right)^2} = \frac{\sigma_x^4 + \sigma_y^4 - 2\sigma_x^2 \cdot \sigma_y^2 + 4\sigma_{xy}^2}{\left(\sigma_x^2 + \sigma_y^2\right)^2} \\ &= \frac{\sigma_x^4 + \sigma_y^4 + 2\sigma_x^2 \cdot \sigma_y^2 - 4\sigma_x^2 \cdot \sigma_y^2 + 4\sigma_{xy}^2}{\left(\sigma_x^2 + \sigma_y^2\right)^2} = \frac{\left(\sigma_x^2 + \sigma_y^2\right)^2 + 4\left(\sigma_{xy}^2 - \sigma_x^2 \cdot \sigma_y^2\right)}{\left(\sigma_x^2 + \sigma_y^2\right)^2}. \end{aligned} \tag{B.2}$$

The Cauchy-Schwarz inequality for random variables X and Y states that

$$\mathbb{E}\left\{X \cdot Y\right\}^2 \leq \mathbb{E}\left\{X^2\right\} \cdot \mathbb{E}\left\{Y^2\right\}, \tag{B.3}$$

with equality iff X and Y are linearly dependent, i.e., $c_1 \cdot X = c_2 \cdot Y$ for some $c_1, c_2 \in \mathbb{R}$. In the short hand notation, the Cauchy-Schwarz inequality translates into $\sigma_{xy}^2 \leq \sigma_x^2 \cdot \sigma_y^2$. Applying this property, we can immediately conclude that

$$|\zeta|^2 = \frac{\left(\sigma_x^2 + \sigma_y^2\right)^2 + 4\left(\sigma_{xy}^2\right) - \sigma_x^2 \cdot \sigma_y^2}{\left(\sigma_x^2 + \sigma_y^2\right)^2} \leq \frac{\left(\sigma_x^2 + \sigma_y^2\right)^2 + 4\left(\sigma_x^2 \cdot \sigma_y^2 - \sigma_x^2 \cdot \sigma_y^2\right)}{\left(\sigma_x^2 + \sigma_y^2\right)^2} = 1, \tag{B.4}$$

which implies that $|\zeta| \leq 1$ and proves the theorem. □

B.2. Derivation of Equation (4.11)

For the derivation of (4.11) in Section 4.3.1, we generalize the first-order perturbation result from [LLV93] for 1-D Standard ESPRIT using least squares (LS) to the R-D case. The presented derivation is based on [RHD14].

The analytical expression for the first-order parameter estimation error of R-D Standard ESPRIT in (4.11) contains the following three steps: the perturbation of the spatial frequencies $\Delta\mu_i^{(r)}$ as a function of the eigenvalue perturbation $\Delta\lambda_i^{(r)}$, the perturbation of $\Delta\lambda_i^{(r)}$ in terms of the perturbation of the least squares (LS) solution $\Delta\boldsymbol{\Psi}^{(r)}$, and the perturbation $\Delta\boldsymbol{\Psi}^{(r)}$ as a function of the signal subspace estimation error $\Delta\boldsymbol{U}_\mathrm{s}$.

For the first step, we start with the eigenvalue perturbation, which is given by

$$\lambda_i^{(r)} + \Delta\lambda_i^{(r)} = \mathrm{e}^{\mathrm{j}(\mu_i^{(r)}+\Delta\mu_i^{(r)})}. \tag{B.5}$$

Taking the logarithm on both sides and expanding the left hand side using the Taylor series up to the first order, we obtain

$$\ln\left(\lambda_i^{(r)}\right) + \frac{\Delta\lambda_i^{(r)}}{\lambda_i^{(r)}} \approx \mathrm{j}\cdot\left(\mu_i^{(r)} + \Delta\mu_i^{(r)}\right). \tag{B.6}$$

Since $\ln\left(\lambda_i^{(r)}\right) = \ln\left(\mathrm{e}^{\mathrm{j}\mu_i^{(r)}}\right) = \mathrm{j}\cdot\mu_i^{(r)}$, we have

$$\frac{\Delta\lambda_i^{(r)}}{\lambda_i^{(r)}} \approx \mathrm{j}\cdot\Delta\mu_i^{(r)}. \tag{B.7}$$

Then, equating the imaginary parts of both sides, yields

$$\Delta\mu_i^{(r)} \approx \mathrm{Im}\left\{\frac{\Delta\lambda_i^{(r)}}{\lambda_i^{(r)}}\right\}. \tag{B.8}$$

For the second step, we first recall that in R-D Standard ESPRIT using LS, the R shift invariance equations are solved independently. Let the resulting matrices $\boldsymbol{\Psi}^{(r)}$ possess the eigendecomposition $\boldsymbol{\Psi}^{(r)} = \boldsymbol{Q}\cdot\boldsymbol{\Lambda}^{(r)}\cdot\boldsymbol{Q}^{-1}$, where $\boldsymbol{Q}$ denotes the matrix of eigenvectors common to all the matrices $\boldsymbol{\Psi}^{(r)}$ and $\boldsymbol{\Lambda}^{(r)} = \mathrm{diag}\left\{[\lambda_1^{(r)},\ldots,\lambda_d^{(r)}]\right\}$ is the diagonal matrix of eigenvalues in the r-th mode. Introducing the perturbation terms $\hat{\boldsymbol{Q}}^{(r)} = \boldsymbol{Q}+\Delta\boldsymbol{Q}^{(r)}$ and $\hat{\boldsymbol{\Psi}}^{(r)} = \boldsymbol{\Psi}^{(r)}+\Delta\boldsymbol{\Psi}^{(r)}$, it is shown in [LT78] that the first-order approximation of the eigendecomposition of a perturbed matrix $\boldsymbol{\Psi}^{(r)} + \Delta\boldsymbol{\Psi}^{(r)}$ can be expressed as

$$\boldsymbol{\Psi}^{(r)} + \Delta\boldsymbol{\Psi}^{(r)} \approx \boldsymbol{\Psi}^{(r)} + \boldsymbol{Q}\cdot\Delta\boldsymbol{\Lambda}^{(r)}\cdot\boldsymbol{Q}^{-1}, \tag{B.9}$$

which implies that considering the first order, the perturbation $\Delta\boldsymbol{\Psi}^{(r)}$ only affects the eigenvalues. Then, using the fact that $\boldsymbol{\Lambda}^{(r)}$ is diagonal and introducing $\boldsymbol{q}_i$ and $\boldsymbol{p}_i^{\mathrm{T}}$ as the i-th column and the i-th row of $\boldsymbol{Q}$ and $\boldsymbol{P} = \boldsymbol{Q}^{-1}$, respectively, we can express $\Delta\lambda_i^{(r)}$ as

$$\Delta\lambda_i^{(r)} \approx \boldsymbol{p}_i^{\mathrm{T}} \cdot \Delta\boldsymbol{\Psi}^{(r)} \cdot \boldsymbol{q}_i. \tag{B.10}$$

As an intermediate result, we can combine (B.8) and (B.10) to approximate the parameter estimation error $\Delta\mu_i^{(r)}$ as

$$\Delta\mu_i^{(r)} \approx \mathrm{Im}\left\{\boldsymbol{p}_i^{\mathrm{T}} \cdot \Delta\boldsymbol{\Psi}^{(r)} \cdot \boldsymbol{q}_i / \lambda_i^{(r)}\right\}. \tag{B.11}$$

To incorporate the third step, we write the perturbed shift invariance equation in terms of the estimated signal subspace $\hat{\boldsymbol{U}}_{\mathrm{s}} = \boldsymbol{U}_{\mathrm{s}} + \Delta\boldsymbol{U}_{\mathrm{s}}$ and the perturbed LS solution $\hat{\boldsymbol{\Psi}}_{\mathrm{LS}} = \boldsymbol{\Psi}^{(r)} + \Delta\boldsymbol{\Psi}^{(r)}$, where $\Delta\boldsymbol{U}_{\mathrm{s}}$ and $\Delta\boldsymbol{\Psi}^{(r)}$ represent the subspace estimation error and the LS perturbation error, respectively. In order to derive the first order approximation of $\Delta\boldsymbol{\Psi}^{(r)}$, we neglect the second-order terms and rearrange the shift invariance equation as follows [RH89a]:

$$\begin{aligned}
\tilde{\boldsymbol{J}}_1^{(r)} \cdot (\boldsymbol{U}_{\mathrm{s}} + \Delta\boldsymbol{U}_{\mathrm{s}}) \cdot \left(\boldsymbol{\Psi}^{(r)} + \Delta\boldsymbol{\Psi}^{(r)}\right) &\approx \tilde{\boldsymbol{J}}_2^{(r)} \cdot (\boldsymbol{U}_{\mathrm{s}} + \Delta\boldsymbol{U}_{\mathrm{s}}) \\
\tilde{\boldsymbol{J}}_1^{(r)} \cdot \boldsymbol{U}_{\mathrm{s}} \cdot \boldsymbol{\Psi}^{(r)} + \tilde{\boldsymbol{J}}_1^{(r)} \cdot \Delta\boldsymbol{U}_{\mathrm{s}} \cdot \boldsymbol{\Psi}^{(r)} + \tilde{\boldsymbol{J}}_1^{(r)} \cdot \boldsymbol{U}_{\mathrm{s}} \cdot \Delta\boldsymbol{\Psi}^{(r)} &\approx \tilde{\boldsymbol{J}}_2^{(r)} \cdot \boldsymbol{U}_{\mathrm{s}} + \tilde{\boldsymbol{J}}_2^{(r)} \cdot \Delta\boldsymbol{U}_{\mathrm{s}} \\
\tilde{\boldsymbol{J}}_1^{(r)} \cdot \Delta\boldsymbol{U}_{\mathrm{s}} \cdot \boldsymbol{\Psi}^{(r)} + \tilde{\boldsymbol{J}}_1^{(r)} \cdot \boldsymbol{U}_{\mathrm{s}} \cdot \Delta\boldsymbol{\Psi}^{(r)} &\approx \tilde{\boldsymbol{J}}_2^{(r)} \cdot \Delta\boldsymbol{U}_{\mathrm{s}}.
\end{aligned} \tag{B.12}$$

where in (B.12), we have taken into account the unperturbed shift invariance equation $\tilde{\boldsymbol{J}}_1^{(r)} \cdot \boldsymbol{U}_{\mathrm{s}} \cdot \boldsymbol{\Psi}^{(r)} = \tilde{\boldsymbol{J}}_2^{(r)} \cdot \boldsymbol{U}_{\mathrm{s}}$. Consequently, we obtain

$$\Delta\boldsymbol{\Psi}^{(r)} \approx \left(\tilde{\boldsymbol{J}}_1^{(r)} \cdot \boldsymbol{U}_{\mathrm{s}}\right)^{+} \cdot \left(\tilde{\boldsymbol{J}}_2^{(r)} \cdot \Delta\boldsymbol{U}_{\mathrm{s}} - \tilde{\boldsymbol{J}}_1^{(r)} \cdot \Delta\boldsymbol{U}_{\mathrm{s}} \cdot \boldsymbol{\Psi}^{(r)}\right). \tag{B.13}$$

Finally, combining the results in (B.11) and (B.13), the first-order approximation of the parameter estimation error $\Delta\mu_i^{(r)}$ can be expressed as

$$\Delta\mu_i^{(r)} \approx \mathrm{Im}\left\{\boldsymbol{p}_i^{\mathrm{T}} \cdot \left(\tilde{\boldsymbol{J}}_1^{(r)} \cdot \boldsymbol{U}_{\mathrm{s}}\right)^{+} \left(\tilde{\boldsymbol{J}}_2^{(r)} \cdot \Delta\boldsymbol{U}_{\mathrm{s}} - \tilde{\boldsymbol{J}}_1^{(r)} \cdot \Delta\boldsymbol{U}_{\mathrm{s}} \cdot \boldsymbol{\Psi}^{(r)}\right) \cdot \boldsymbol{q}_i / \lambda_i^{(r)}\right\} \tag{B.14}$$

$$\approx \mathrm{Im}\left\{\boldsymbol{p}_i^{\mathrm{T}} \cdot \left(\tilde{\boldsymbol{J}}_1^{(r)} \cdot \boldsymbol{U}_{\mathrm{s}}\right)^{+} \left(\tilde{\boldsymbol{J}}_2^{(r)} / \lambda_i^{(r)} - \tilde{\boldsymbol{J}}_1^{(r)}\right) \cdot \Delta\boldsymbol{U}_{\mathrm{s}} \cdot \boldsymbol{q}_i\right\}, \tag{B.15}$$

where in (B.14), we have multiplied out the last bracket and used the relation $\boldsymbol{\Psi}^{(r)} \cdot \boldsymbol{q}_i = \lambda_i^{(r)} \cdot \boldsymbol{q}_i$ to arrive at the desired result in (B.15). □

B.3. Proof of Theorem 4.3.1

In order to prove Theorem 4.3.1 in Section 4.3.1, we first apply a first-order Taylor-series expansion of $\hat{\mu}_i^{(1)}$ and $\hat{\mu}_i^{(2)}$ to obtain the perturbed azimuth and co-elevation angles $\Delta\theta_i$ and $\Delta\phi_i$ can be obtained as

$$\hat{\mu}_i^{(1)} = \mu_i^{(1)} + \Delta\mu_i^{(1)} = \mu_i^{(1)} + \frac{\partial \mu_i^{(1)}}{\partial \theta_i} \cdot \Delta\theta_i + \frac{\partial \mu_i^{(1)}}{\partial \phi_i} \cdot \Delta\phi_i, \tag{B.16}$$

$$\hat{\mu}_i^{(2)} = \mu_i^{(2)} + \Delta\mu_i^{(2)} = \mu_i^{(2)} + \frac{\partial \mu_i^{(2)}}{\partial \theta_i} \cdot \Delta\theta_i + \frac{\partial \mu_i^{(2)}}{\partial \phi_i} \cdot \Delta\phi_i. \tag{B.17}$$

Thus, the respective first-order perturbation can be conveniently expressed in matrix form as

$$\begin{bmatrix} \Delta\mu_i^{(1)} \\ \Delta\mu_i^{(2)} \end{bmatrix} = \begin{bmatrix} \frac{\partial \mu_i^{(1)}}{\partial \theta_i} & \frac{\partial \mu_i^{(1)}}{\partial \phi_i} \\ \frac{\partial \mu_i^{(2)}}{\partial \theta_i} & \frac{\partial \mu_i^{(2)}}{\partial \phi_i} \end{bmatrix} \cdot \begin{bmatrix} \Delta\theta_i \\ \Delta\phi_i \end{bmatrix}. \tag{B.18}$$

Solving (B.18) for the perturbations in the azimuth and co-elevation angles yields

$$\begin{aligned} \begin{bmatrix} \Delta\theta_i \\ \Delta\phi_i \end{bmatrix} &= \begin{bmatrix} \frac{\partial \mu_i^{(1)}}{\partial \theta_i} & \frac{\partial \mu_i^{(1)}}{\partial \phi_i} \\ \frac{\partial \mu_i^{(2)}}{\partial \theta_i} & \frac{\partial \mu_i^{(2)}}{\partial \phi_i} \end{bmatrix}^{-1} \cdot \begin{bmatrix} \Delta\mu_i^{(1)} \\ \Delta\mu_i^{(2)} \end{bmatrix} \\ &= \frac{1}{\frac{\partial \mu_i^{(1)}}{\partial \theta_i}\frac{\partial \mu_i^{(2)}}{\partial \phi_i} - \frac{\partial \mu_i^{(2)}}{\partial \theta_i}\frac{\partial \mu_i^{(1)}}{\partial \phi_i}} \cdot \begin{bmatrix} \frac{\partial \mu_i^{(2)}}{\partial \phi_i} & -\frac{\partial \mu_i^{(1)}}{\partial \phi_i} \\ -\frac{\partial \mu_i^{(2)}}{\partial \theta_i} & \frac{\partial \mu_i^{(1)}}{\partial \theta_i} \end{bmatrix} \cdot \begin{bmatrix} \Delta\mu_i^{(1)} \\ \Delta\mu_i^{(2)} \end{bmatrix}. \end{aligned} \tag{B.19}$$

Consequently, the estimation errors of the azimuth and co-elevation angles can be written as

$$\Delta\theta_i = \frac{\Delta\mu_i^{(1)} - \varrho_i \cdot \Delta\mu_i^{(2)}}{\frac{\partial \mu_i^{(1)}}{\partial \theta_i} - \varrho_i \cdot \frac{\partial \mu_i^{(2)}}{\partial \theta_i}}, \quad \Delta\phi_i = \frac{\Delta\mu_i^{(2)} - \kappa_i \cdot \Delta\mu_i^{(1)}}{\frac{\partial \mu_i^{(2)}}{\partial \phi_i} - \kappa_i \cdot \frac{\partial \mu_i^{(1)}}{\partial \phi_i}}, \tag{B.20}$$

where we have defined

$$\varrho_i = \frac{\frac{\partial \mu_i^{(1)}}{\partial \phi_i}}{\frac{\partial \mu_i^{(2)}}{\partial \phi_i}}, \qquad \kappa_i = \frac{\frac{\partial \mu_i^{(2)}}{\partial \theta_i}}{\frac{\partial \mu_i^{(1)}}{\partial \theta_i}}. \tag{B.21}$$

The partial derivatives are given by

$$\frac{\partial \mu_i^{(1)}}{\partial \theta_i} = -\frac{2\pi}{\lambda} \cdot \Delta \cdot \sin\theta_i \cdot \sin\phi_i, \qquad \frac{\partial \mu_i^{(1)}}{\partial \phi_i} = \frac{2\pi}{\lambda} \cdot \Delta \cdot \cos\theta_i \cdot \cos\phi_i, \tag{B.22}$$

$$\frac{\partial \mu_i^{(2)}}{\partial \theta_i} = \frac{2\pi}{\lambda} \cdot \Delta \cdot \cos\theta_i \cdot \sin\phi_i, \qquad \frac{\partial \mu_i^{(2)}}{\partial \phi_i} = \frac{2\pi}{\lambda} \cdot \Delta \cdot \sin\theta_i \cdot \cos\phi_i \tag{B.23}$$

and therefore,

$$\varrho_i = \frac{\cos\theta_i}{\sin\theta_i}, \qquad \kappa_i = -\frac{\cos\theta_i}{\sin\theta_i}. \tag{B.24}$$

Finally, the estimation errors in (B.20) are given by

$$\begin{aligned}
\Delta\theta_i &= \frac{\Delta\mu_i^{(1)} - \frac{\cos\theta_i}{\sin\theta_i} \cdot \Delta\mu_i^{(2)}}{-\frac{2\pi}{\lambda} \cdot \Delta \cdot \sin\theta_i \cdot \sin\phi_i - \frac{\cos\theta_i}{\sin\theta_i} \cdot \frac{2\pi}{\lambda} \cdot \Delta \cdot \cos\theta_i \cdot \sin\phi_i} \\
&= \frac{\Delta\mu_i^{(1)} - \frac{\cos\theta_i}{\sin\theta_i} \cdot \Delta\mu_i^{(2)}}{-\frac{2\pi}{\lambda} \cdot \Delta \cdot \sin\phi_i \cdot (\sin\theta_i + \frac{\cos\theta_i}{\sin\theta_i} \cdot \cos\theta_i)} \\
&= \frac{1}{\frac{2\pi}{\lambda} \cdot \Delta \cdot \sin\phi_i} \cdot \left(\cos\theta_i \cdot \Delta\mu_i^{(2)} - \sin\theta_i \cdot \Delta\mu_i^{(1)}\right)
\end{aligned} \tag{B.25}$$

and, similarly,

$$\Delta\phi_i = \frac{1}{\frac{2\pi}{\lambda} \cdot \Delta \cdot \cos\phi_i} \cdot \left(\sin\theta_i \cdot \Delta\mu_i^{(2)} + \cos\theta_i \cdot \Delta\mu_i^{(1)}\right). \tag{B.26}$$

Then, the first-order MSE expression of the azimuth estimation error based on (B.25) is given by

$$\begin{aligned}
\mathbb{E}\left\{(\Delta\theta_i)^2\right\} &\approx \left(\frac{1}{\frac{2\pi}{\lambda} \cdot \Delta \cdot \sin\phi_i}\right)^2 \cdot \mathbb{E}\left\{\left(\cos\theta_i \cdot \Delta\mu_i^{(2)} - \sin\theta_i \cdot \Delta\mu_i^{(1)}\right)^2\right\} \\
&= \left(\frac{1}{\frac{2\pi}{\lambda} \cdot \Delta \cdot \sin\phi_i}\right)^2 \cdot \mathbb{E}\left\{(\cos\theta_i)^2 \cdot \left(\Delta\mu_i^{(2)}\right)^2 + (\sin\theta_i)^2 \cdot \left(\Delta\mu_i^{(1)}\right)^2 \right. \\
&\qquad \left. -2 \cdot \sin\theta_i \cdot \Delta\mu_i^{(1)} \cos\theta_i \cdot \Delta\mu_i^{(2)}\right\} \\
&= \left(\frac{1}{\frac{2\pi}{\lambda} \cdot \Delta \cdot \sin\phi_i}\right)^2 \cdot \left((\cos\theta_i)^2 \cdot \mathbb{E}\left\{\left(\Delta\mu_i^{(2)}\right)^2\right\} + (\sin\theta_i)^2 \cdot \mathbb{E}\left\{\left(\Delta\mu_i^{(1)}\right)^2\right\} \right. \\
&\qquad \left. -2 \cdot \sin\theta_i \cdot \cos\theta_i \cdot \mathbb{E}\left\{\Delta\mu_i^{(1)} \cdot \Delta\mu_i^{(2)}\right\}\right),
\end{aligned} \tag{B.27}$$

where $\mathbb{E}\left\{\left(\Delta\mu_i^{(r)}\right)^2\right\}$ for $r = 1, 2$ is given in (4.17). Note that from [SRH16a], we can express $\mathbb{E}\left\{\Delta\mu_i^{(1)} \cdot \Delta\mu_i^{(2)}\right\}$ as

$$\mathbb{E}\left\{\Delta\mu_i^{(1)} \cdot \Delta\mu_i^{(2)}\right\} \approx \frac{1}{2} \cdot \left(\boldsymbol{z}_i^{(1)^{\mathrm{H}}} \cdot \boldsymbol{R}_{\mathrm{nn}}^{\mathrm{T}} \cdot \boldsymbol{z}_i^{(2)} - \mathrm{Re}\left\{\boldsymbol{z}_i^{(1)^{\mathrm{T}}} \cdot \boldsymbol{C}_{\mathrm{nn}}^{\mathrm{T}} \cdot \boldsymbol{z}_i^{(2)}\right\}\right), \tag{B.28}$$

where $\boldsymbol{z}_i^{(r)} = \boldsymbol{W}_{\mathrm{mat}}^{\mathrm{T}} \cdot \boldsymbol{r}_i^{(r)}$ for $r = 1, 2$ and $\boldsymbol{r}_i^{(r)}$ and $\boldsymbol{W}_{\mathrm{mat}}$ are given by (4.14) and (4.15), respectively.

Analogously, the first-order MSE expression of the co-elevation angle corresponding to (B.26) can be written as

$$\mathbb{E}\left\{(\Delta\phi_i)^2\right\} \approx \left(\frac{1}{\frac{2\pi}{\lambda_\mathrm{c}}\cdot\Delta\cdot\cos\phi_i}\right)^2 \cdot \Big((\sin\theta_i)^2\cdot\mathbb{E}\left\{(\Delta\mu_i^{(2)})^2\right\} + (\cos\theta_i)^2\cdot\mathbb{E}\left\{(\Delta\mu_i^{(1)})^2\right\} + 2\cdot\sin\theta_i\cdot\cos\theta_i\cdot\mathbb{E}\left\{\Delta\mu_i^{(1)}\cdot\Delta\mu_i^{(2)}\right\}\Big). \tag{B.29}$$

This completes the proof. □

B.4. Proof of Equation (4.35)

In order to derive the expression of $\boldsymbol{W}_\mathrm{ten}$ in Section 4.4.1 for arbitrary R, we generalize the steps provided in [RHD14] for the special case of $R = 2$.

To arrive at the MSE for R-D Standard Tensor-ESPRIT in (4.34), the first step is to express the estimation error in $\mu_i^{(r)}$ from (4.33) in terms of the perturbation $\boldsymbol{n} = \mathrm{vec}\{\boldsymbol{N}\} = \mathrm{vec}\left\{[\boldsymbol{\mathcal{N}}]_{(R+1)}^\mathrm{T}\right\}$. Starting with (4.33), we apply the vec-operator to obtain

$$\begin{aligned}\Delta\mu_i^{(r)} &\approx \mathrm{Im}\left\{\boldsymbol{p}_i^{(r)^\mathrm{T}}\cdot\left(\tilde{\boldsymbol{J}}_1^{(r)}\cdot\boldsymbol{U}_\mathrm{s}\right)^+\cdot\left[\tilde{\boldsymbol{J}}_2^{(r)}/\lambda_i^{(r)}-\tilde{\boldsymbol{J}}_1^{(r)}\right]\cdot\left[\Delta\hat{\boldsymbol{\mathcal{U}}}^{[\mathrm{s}]}\right]_{(R+1)}^\mathrm{T}\cdot\boldsymbol{q}_i^{(r)}\right\}\\ &= \mathrm{Im}\left\{\left[\boldsymbol{q}_i^{(r)^\mathrm{T}}\otimes\left(\boldsymbol{p}_i^{(r)^\mathrm{T}}\cdot\left(\tilde{\boldsymbol{J}}_1^{(r)}\cdot\boldsymbol{U}_\mathrm{s}\right)^+\cdot\left[\tilde{\boldsymbol{J}}_2^{(r)}/\lambda_i^{(r)}-\tilde{\boldsymbol{J}}_1^{(r)}\right]\right)\right]\cdot\mathrm{vec}\left\{\left[\Delta\hat{\boldsymbol{\mathcal{U}}}^{[\mathrm{s}]}\right]_{(R+1)}^\mathrm{T}\right\}\right\}\\ &= \mathrm{Im}\left\{\boldsymbol{r}_i^{(r)^\mathrm{T}}\cdot\mathrm{vec}\left\{\left[\Delta\hat{\boldsymbol{\mathcal{U}}}^{[\mathrm{s}]}\right]_{(R+1)}^\mathrm{T}\right\}\right\},\end{aligned} \tag{B.30}$$

where $\boldsymbol{r}_i^{(r)}$ is defined in (4.14). As $\mathrm{vec}\left\{\left[\Delta\hat{\boldsymbol{\mathcal{U}}}^{[\mathrm{s}]}\right]_{(R+1)}^\mathrm{T}\right\}$ depends linearly on $\boldsymbol{n}$, we can find an explicit expression for $\boldsymbol{W}_\mathrm{ten}$ such that

$$\mathrm{vec}\left\{\left[\Delta\hat{\boldsymbol{\mathcal{U}}}^{[\mathrm{s}]}\right]_{(R+1)}^\mathrm{T}\right\} = \boldsymbol{W}_\mathrm{ten}\cdot\boldsymbol{n}. \tag{B.31}$$

The HOSVD-based signal subspace error $\left[\Delta\hat{\boldsymbol{\mathcal{U}}}^{[\mathrm{s}]}\right]_{(R+1)}^\mathrm{T}$ was given in (4.10) and is restated again for convenience as

$$\left[\Delta\hat{\boldsymbol{\mathcal{U}}}^{[\mathrm{s}]}\right]_{(R+1)}^\mathrm{T} \approx \boldsymbol{T}_{1:R}^{\otimes}\cdot\Delta\boldsymbol{U}_\mathrm{s} + \sum_{r=1}^{R}\left(\boldsymbol{T}_{1:r-1}^{\otimes}\otimes\left[\Delta\boldsymbol{U}_r^{[\mathrm{s}]}\cdot\boldsymbol{U}_r^{[\mathrm{s}]^\mathrm{H}}\right]\otimes\boldsymbol{T}_{r+1:R}^{\otimes}\right)\cdot\boldsymbol{U}_\mathrm{s}, \tag{B.32}$$

where $\boldsymbol{T}^{\otimes}_{a:b}$ is defined as

$$\boldsymbol{T}^{\otimes}_{a:b} = \begin{cases} \boldsymbol{T}_a \otimes \ldots \otimes \boldsymbol{T}_b & a \leq b \\ 1 & a > b. \end{cases} \tag{B.33}$$

The first term of (B.32) is easily vectorized by applying property (1.14), which yields the first term of $\boldsymbol{W}_{\text{ten}}$ as

$$\begin{aligned} \operatorname{vec}\left\{\boldsymbol{T}^{\otimes}_{1:R} \cdot \Delta\boldsymbol{U}_{\text{s}}\right\} &= \operatorname{vec}\left\{\boldsymbol{T}^{\otimes}_{1:R} \cdot \boldsymbol{V}^{[\text{n}]^*}_{R+1} \cdot \boldsymbol{V}^{[\text{n}]^{\text{T}}}_{R+1} \cdot \boldsymbol{\mathcal{N}} \cdot \boldsymbol{U}^{[\text{s}]^*}_{R+1} \cdot \boldsymbol{\Sigma}^{[\text{s}]^{-1}}_{R+1}\right\} \\ &= \left(\boldsymbol{\Sigma}^{[\text{s}]^{-1}}_{R+1} \cdot \boldsymbol{U}^{[\text{s}]^{\text{H}}}_{R+1}\right) \otimes \left[\boldsymbol{T}^{\otimes}_{1:R} \cdot \boldsymbol{V}^{[\text{n}]^*}_{R+1} \cdot \boldsymbol{V}^{[\text{n}]^{\text{T}}}_{R+1}\right] \cdot \operatorname{vec}\{\boldsymbol{\mathcal{N}}\}. \end{aligned}$$

For the sum in (B.32), we first apply property (1.14) to obtain

$$\begin{aligned} &\sum_{r=1}^{R} \operatorname{vec}\left\{\left(\boldsymbol{T}^{\otimes}_{1:r-1} \otimes \left[\Delta\boldsymbol{U}^{[\text{s}]}_r \cdot \boldsymbol{U}^{[\text{s}]^{\text{H}}}_r\right] \otimes \boldsymbol{T}^{\otimes}_{r+1:R}\right) \cdot \boldsymbol{U}_{\text{s}}\right\} \\ &= \left(\boldsymbol{U}^{\text{T}}_{\text{s}} \otimes \boldsymbol{I}_M\right) \cdot \sum_{r=1}^{R} \operatorname{vec}\left\{\boldsymbol{T}^{\otimes}_{1:r-1} \otimes \left[\boldsymbol{U}^{[\text{n}]}_r \cdot \boldsymbol{U}^{[\text{n}]^{\text{H}}}_r \cdot [\boldsymbol{\mathcal{N}}]_{(r)} \cdot \boldsymbol{V}^{[\text{s}]}_r \cdot \boldsymbol{\Sigma}^{[\text{s}]^{-1}}_r \cdot \boldsymbol{U}^{[\text{s}]^{\text{H}}}_r\right] \otimes \boldsymbol{T}^{\otimes}_{r+1:R}\right\} \\ &= \left(\boldsymbol{U}^{\text{T}}_{\text{s}} \otimes \boldsymbol{I}_M\right) \cdot \sum_{r=1}^{R} \operatorname{vec}\left\{\boldsymbol{T}^{\otimes}_{1:r-1} \otimes \left[\Delta\boldsymbol{U}^{[\text{s}]}_r \cdot \boldsymbol{U}^{[\text{s}]^{\text{H}}}_r\right] \otimes \boldsymbol{T}^{\otimes}_{r+1:R}\right\}, \end{aligned} \tag{B.34}$$

where $\Delta\boldsymbol{U}^{[\text{s}]}_r$ is provided in (4.8). Then, in order to simplify the vectorization of a Kronecker product of three terms, we use the property in (1.20) such that

$$\begin{aligned} &\operatorname{vec}\left\{\boldsymbol{T}^{\otimes}_{1:r-1} \otimes \left[\Delta\boldsymbol{U}^{[\text{s}]}_r \cdot \boldsymbol{U}^{[\text{s}]^{\text{H}}}_r\right] \otimes \boldsymbol{T}^{\otimes}_{r+1:R}\right\} \\ &= \left(\bar{\boldsymbol{T}}_{1:r-1} \otimes \boldsymbol{I}_{M_{r:R}}\right) \cdot \left(\boldsymbol{I}_{M_r} \otimes \bar{\boldsymbol{T}}_{r+1:R}\right) \cdot \operatorname{vec}\left\{\left[\boldsymbol{U}^{[\text{n}]}_r \cdot \boldsymbol{U}^{[\text{n}]^{\text{H}}}_r \cdot [\boldsymbol{\mathcal{N}}]_{(r)} \cdot \boldsymbol{V}^{[\text{s}]}_r \cdot \boldsymbol{\Sigma}^{[\text{s}]^{-1}}_r \cdot \boldsymbol{U}^{[\text{s}]^{\text{H}}}_r\right]\right\} \\ &= \left(\bar{\boldsymbol{T}}_{1:r-1} \otimes \boldsymbol{I}_{M_{r:R}}\right) \cdot \left(\boldsymbol{I}_{M_r} \otimes \bar{\boldsymbol{T}}_{r+1:R}\right) \cdot \left[\left(\boldsymbol{V}^{[\text{s}]}_r \cdot \boldsymbol{\Sigma}^{[\text{s}]^{-1}}_r \cdot \boldsymbol{U}^{[\text{s}]^{\text{H}}}_r\right)^{\text{T}} \otimes \left(\boldsymbol{U}^{[\text{n}]}_r \cdot \boldsymbol{U}^{[\text{n}]^{\text{H}}}_r\right)\right] \cdot \operatorname{vec}\left\{[\boldsymbol{\mathcal{N}}]_{(r)}\right\}, \end{aligned} \tag{B.35}$$

where

$$\bar{\boldsymbol{T}}_{1:r-1} = \begin{bmatrix} \boldsymbol{I}_{M_{r:R}} \otimes \boldsymbol{t}_{1:r-1,1} \\ \vdots \\ \boldsymbol{I}_{M_{r:R}} \otimes \boldsymbol{t}_{1:r-1,M_{1:r-1}} \end{bmatrix}, \quad \bar{\boldsymbol{T}}_{r+1:R} = \begin{bmatrix} \boldsymbol{I}_{M_r} \otimes \boldsymbol{t}_{r+1:R,1} \\ \vdots \\ \boldsymbol{I}_{M_r} \otimes \boldsymbol{t}_{r+1:R,M_{r+1:R}} \end{bmatrix} \tag{B.36}$$

with the short-hand notation $\boldsymbol{t}_{a:b,n}$ as the n-th column of $\boldsymbol{T}^{\otimes}_{a:b}$ and

$$M_{a:b} = \begin{cases} \prod_{r=a}^{b} M_r & a \leq b \\ 1 & a > b \end{cases}. \tag{B.37}$$

To permute the vectors to be in consistent order with the vector $\boldsymbol{n} = \mathrm{vec}\{\boldsymbol{\mathcal{N}}\} = \mathrm{vec}\left\{[\boldsymbol{\mathcal{N}}]^{\mathrm{T}}_{(R+1)}\right\}$ we employ the permutation matrices defined in (1.38). By combining the previous results, the matrix $\boldsymbol{W}_{\mathrm{ten}}$ results in the expression stated in [RHD14]

$$\begin{aligned} \boldsymbol{W}_{\mathrm{ten}} = &\left(\boldsymbol{\Sigma}_{\mathrm{s}}^{-1} \cdot \boldsymbol{V}_{\mathrm{s}}^{\mathrm{T}}\right) \otimes \left(\boldsymbol{T}^{\otimes}_{1:R} \cdot \boldsymbol{U}_{\mathrm{n}} \cdot \boldsymbol{U}_{\mathrm{n}}^{\mathrm{H}}\right) + \sum_{r=1}^{R} \left(\boldsymbol{U}_{\mathrm{s}}^{\mathrm{T}} \otimes \boldsymbol{I}_{M}\right) \\ &\cdot \left(\bar{\boldsymbol{T}}_{1:r-1} \otimes \boldsymbol{I}_{M_{r:R}}\right) \cdot \left(\boldsymbol{I}_{M_r} \otimes \bar{\boldsymbol{T}}_{r+1:R}\right) \\ &\cdot \left[\left(\boldsymbol{V}_r^{[\mathrm{s}]} \cdot \boldsymbol{\Sigma}_r^{[\mathrm{s}]^{-1}} \cdot \boldsymbol{U}_r^{[\mathrm{s}]^{\mathrm{H}}}\right)^{\mathrm{T}} \otimes \left(\boldsymbol{U}_r^{[\mathrm{n}]} \cdot \boldsymbol{U}_r^{[\mathrm{n}]^{\mathrm{H}}}\right)\right] \cdot \boldsymbol{P}^{(r)^{\mathrm{T}}}_{M_1,\ldots,M_R,N} \cdot \boldsymbol{P}^{(R)}_{M_1,\ldots,M_R,N}, \end{aligned} \tag{B.38}$$

which can be compactly written as

$$\boldsymbol{W}_{\mathrm{ten}} = \boldsymbol{W}_0 + \sum_{r=1}^{R} \boldsymbol{W}_r \cdot \boldsymbol{P}^{(r)^{\mathrm{T}}}_{M_1,\ldots,M_R,N} \cdot \boldsymbol{P}^{(R)}_{M_1,\ldots,M_R,N},$$

where

$$\boldsymbol{W}_0 = \left(\boldsymbol{\Sigma}_{\mathrm{s}}^{-1} \cdot \boldsymbol{V}_{\mathrm{s}}^{\mathrm{T}}\right) \otimes \left(\boldsymbol{T}^{\otimes}_{1:R} \cdot \boldsymbol{U}_{\mathrm{n}} \cdot \boldsymbol{U}_{\mathrm{n}}^{\mathrm{H}}\right) \tag{B.39}$$

$$\begin{aligned} \boldsymbol{W}_r = &\left(\boldsymbol{U}_{\mathrm{s}}^{\mathrm{T}} \otimes \boldsymbol{I}_{M}\right) \cdot \left(\bar{\boldsymbol{T}}_{1:r-1} \otimes \boldsymbol{I}_{M_{r:R}}\right) \cdot \left(\boldsymbol{I}_{M_r} \otimes \bar{\boldsymbol{T}}_{r+1:R}\right) \\ &\cdot \left[\left(\boldsymbol{V}_r^{[\mathrm{s}]} \cdot \boldsymbol{\Sigma}_r^{[\mathrm{s}]^{-1}} \cdot \boldsymbol{U}_r^{[\mathrm{s}]^{\mathrm{H}}}\right)^{\mathrm{T}} \otimes \left(\boldsymbol{U}_r^{[\mathrm{n}]} \cdot \boldsymbol{U}_r^{[\mathrm{n}]^{\mathrm{H}}}\right)\right]. \end{aligned} \tag{B.40}$$

This concludes the proof. □

B.5. Proof of Theorem 4.5.1

For the proof of Theorem 4.5.1 in Section 4.5.3, we first recall the MSE expression of R-D Standard ESPRIT for circularly symmetric white noise in (4.18) as

$$\mathbb{E}\left\{\left(\Delta\mu_i^{(r)}\right)^2\right\} = \frac{\sigma_{\mathrm{n}}^2}{2} \cdot \left\|\boldsymbol{z}_i^{(r)}\right\|_2^2, \tag{B.41}$$

where $\boldsymbol{z}_i^{(r)} = \boldsymbol{W}_{\text{mat}}^{\text{T}} \cdot \boldsymbol{r}_i^{(r)}$ with

$$\boldsymbol{r}_i^{(r)} = \boldsymbol{q}_i \otimes \left(\left[\left(\tilde{\boldsymbol{J}}_1^{(r)} \boldsymbol{U}_{\text{s}} \right)^{+} \cdot \left(\tilde{\boldsymbol{J}}_2^{(r)} / \mathrm{e}^{\mathrm{j} \cdot \mu_i^{(r)}} - \tilde{\boldsymbol{J}}_1^{(r)} \right) \right]^{\text{T}} \cdot \boldsymbol{p}_i \right) \tag{B.42}$$

$$\boldsymbol{W}_{\text{mat}} = \left(\boldsymbol{\Sigma}_{\text{s}}^{-1} \cdot \boldsymbol{V}_{\text{s}}^{\text{T}} \right) \otimes \left(\boldsymbol{U}_{\text{n}} \cdot \boldsymbol{U}_{\text{n}}^{\text{H}} \right). \tag{B.43}$$

Then, we can equivalently express $\boldsymbol{z}_i^{(r)^{\text{T}}}$ as $\boldsymbol{z}_i^{(r)^{\text{T}}} = \boldsymbol{r}_i^{(r)^{\text{T}}} \cdot \boldsymbol{W}_{\text{mat}} = \tilde{\boldsymbol{s}}_i^{\text{T}} \otimes \boldsymbol{a}_i^{(r)^{\text{T}}}$, where

$$\tilde{\boldsymbol{s}}_i^{\text{T}} = \boldsymbol{q}_i^{\text{T}} \cdot \boldsymbol{\Sigma}_{\text{s}}^{-1} \cdot \boldsymbol{V}_{\text{s}}^{\text{T}} \tag{B.44}$$

$$\tilde{\boldsymbol{a}}_i^{(r)^{\text{T}}} = \boldsymbol{p}_i^{\text{T}} \cdot \left(\tilde{\boldsymbol{J}}_1^{(r)} \cdot \boldsymbol{U}_{\text{s}} \right)^{+} \cdot \left(\tilde{\boldsymbol{J}}_2^{(r)} / \lambda_i^{(r)} - \tilde{\boldsymbol{J}}_1^{(r)} \right) \cdot \boldsymbol{U}_{\text{n}} \cdot \boldsymbol{U}_{\text{n}}^{\text{H}}. \tag{B.45}$$

Using property (1.7), the MSE in (B.41) is given by

$$\mathbb{E}\left\{ \left(\Delta \mu_i^{(r)} \right)^2 \right\} = \frac{\sigma_{\text{n}}^2}{2} \cdot \left\| \boldsymbol{z}_i^{(r)^{\text{T}}} \right\|_2^2 = \frac{\sigma_{\text{n}}^2}{2} \cdot \left\| \tilde{\boldsymbol{s}}_i^{\text{T}} \otimes \tilde{\boldsymbol{a}}_i^{(r)^{\text{T}}} \right\|_2^2 = \frac{\sigma_{\text{n}}^2}{2} \cdot \left\| \tilde{\boldsymbol{s}}_i^{\text{T}} \right\|_2^2 \cdot \left\| \tilde{\boldsymbol{a}}_i^{(r)^{\text{T}}} \right\|_2^2. \tag{B.46}$$

In order to relate the subspaces in (B.44) and (B.45) to the array steering matrix $\boldsymbol{A}$ and the symbol matrix $\boldsymbol{S}$, we use the property $\boldsymbol{A} = \boldsymbol{U}_{\text{s}} \cdot \boldsymbol{T}$. Considering the eigenvectors $\boldsymbol{Q}$ from the exact solution $\boldsymbol{\Psi}^{(r)} = \boldsymbol{Q} \cdot \boldsymbol{\Lambda}^{(r)} \cdot \boldsymbol{P}$ with $\boldsymbol{P} = \boldsymbol{Q}^{-1}$, we can write $\boldsymbol{T} = \sqrt{M} \cdot \boldsymbol{Q}$. Thus, $\boldsymbol{U}_{\text{s}}$ is given by

$$\boldsymbol{U}_{\text{s}} = \frac{\boldsymbol{A}}{\sqrt{M}} \cdot \boldsymbol{P}. \tag{B.47}$$

Inserting (B.47) into the economy-size SVD of the noise-free measurement matrix $\boldsymbol{X}_0 = \boldsymbol{U}_{\text{s}} \cdot \boldsymbol{\Sigma}_{\text{s}} \cdot \boldsymbol{V}_{\text{s}}^{\text{H}}$ to obtain

$$\boldsymbol{X}_0 = \boldsymbol{U}_{\text{s}} \cdot \boldsymbol{\Sigma}_{\text{s}} \cdot \boldsymbol{V}_{\text{s}}^{\text{H}} = \boldsymbol{A} \cdot \frac{1}{\sqrt{M}} \cdot \boldsymbol{P} \cdot \boldsymbol{\Sigma}_{\text{s}} \cdot \boldsymbol{V}_{\text{s}}^{\text{H}} = \boldsymbol{A} \cdot \boldsymbol{S}. \tag{B.48}$$

Thus, we obtain

$$\boldsymbol{S} = \frac{1}{\sqrt{M}} \cdot \boldsymbol{P} \cdot \boldsymbol{\Sigma}_{\text{s}} \cdot \boldsymbol{V}_{\text{s}}^{\text{H}}. \tag{B.49}$$

Via the pseudo-inverse $\boldsymbol{S}^{+} = \sqrt{M} \cdot \boldsymbol{V}_{\text{s}} \cdot \boldsymbol{\Sigma}_{\text{s}}^{-1} \cdot \boldsymbol{Q}$, it is straightforward to see that

$$\boldsymbol{V}_{\text{s}} \cdot \boldsymbol{\Sigma}_{\text{s}}^{-1} = \frac{1}{\sqrt{M}} \cdot \boldsymbol{S}^{+} \cdot \boldsymbol{P}. \tag{B.50}$$

Taking the transpose on both sides, yields the required result[1]

$$\boldsymbol{\Sigma}_{\text{s}}^{-1} \cdot \boldsymbol{V}_{\text{s}}^{\text{T}} = \frac{1}{\sqrt{M}} \cdot \boldsymbol{P}^{\text{T}} \cdot \boldsymbol{S}^{+\text{T}}. \tag{B.51}$$

[1] Note that both terms $\boldsymbol{U}_{\text{s}} = \frac{\boldsymbol{A}}{\sqrt{M}} \cdot \boldsymbol{P}$ and $\boldsymbol{\Sigma}_{\text{s}}^{-1} \cdot \boldsymbol{V}_{\text{s}}^{\text{T}} = \frac{1}{\sqrt{M}} \cdot \boldsymbol{Q}^{-\text{T}} \cdot \boldsymbol{S}^{+\text{T}}$ contain a phase ambiguity due to the fact that the SVD and the EVD are only unique up to a unitary diagonal scaling matrix. However, this phase term cancels when inserting the results into (B.46), and is therefore omitted.

Moreover, $\boldsymbol{U}_\mathrm{n} \cdot \boldsymbol{U}_\mathrm{n}^\mathrm{H}$ is a projection matrix onto the orthogonal complement of the signal subspace spanned by $\boldsymbol{A}$. Thus, $\boldsymbol{U}_\mathrm{n} \cdot \boldsymbol{U}_\mathrm{n}^\mathrm{H}$ can be expressed in terms of $\boldsymbol{A}$ as

$$\boldsymbol{U}_\mathrm{n} \cdot \boldsymbol{U}_\mathrm{n}^\mathrm{H} = \boldsymbol{P}_{\boldsymbol{A}}^{\perp} = \boldsymbol{I}_M - \boldsymbol{A} \cdot \boldsymbol{A}^{+}. \tag{B.52}$$

Finally, inserting the results from (B.47), (B.51), and (B.52) into (B.44) and (B.45) yields

$$\tilde{\boldsymbol{s}}_i^\mathrm{T} = \boldsymbol{e}_i^\mathrm{T} \cdot \boldsymbol{S}^{+\mathrm{T}} \tag{B.53}$$

$$\tilde{\boldsymbol{a}}_i^{(r)^\mathrm{T}} = \boldsymbol{e}_i^\mathrm{T} \cdot \left(\tilde{\boldsymbol{J}}_1^{(r)} \cdot \boldsymbol{A}\right)^{+} \cdot \left(\tilde{\boldsymbol{J}}_2^{(r)} / \lambda_i^{(r)} - \tilde{\boldsymbol{J}}_1^{(r)}\right) \cdot \boldsymbol{P}_{\boldsymbol{A}}^{\perp}, \tag{B.54}$$

where $\boldsymbol{e}_i$ is the vector with a one at the i-th position and zeros elsewhere. Note that we have omitted the factor $\sqrt{M}$, which cancels. In the next step, we compute $\left\|\tilde{\boldsymbol{s}}_i^\mathrm{T}\right\|_2^2$ in (B.46). To this end, we first simplify $\tilde{\boldsymbol{s}}_i^\mathrm{T}$ by expanding the pseudo-inverse $\boldsymbol{S}^{+} = \boldsymbol{S}^\mathrm{H} \cdot \left(\boldsymbol{S} \cdot \boldsymbol{S}^\mathrm{H}\right)^{-1}$ as

$$\tilde{\boldsymbol{s}}_i^\mathrm{T} = \boldsymbol{e}_i^\mathrm{T} \cdot \boldsymbol{S}^{+\mathrm{T}} = \boldsymbol{e}_i^\mathrm{T} \cdot \left(\boldsymbol{S} \cdot \boldsymbol{S}^\mathrm{H}\right)^{-\mathrm{T}} \cdot \boldsymbol{S}^{*} = \frac{1}{N} \cdot \boldsymbol{e}_i^\mathrm{T} \cdot \hat{\boldsymbol{R}}_\mathrm{ss}^{-\mathrm{T}} \cdot \boldsymbol{S}^{*}, \tag{B.55}$$

where

$$\hat{\boldsymbol{R}}_\mathrm{ss} = \frac{1}{N} \cdot \boldsymbol{S} \cdot \boldsymbol{S}^\mathrm{H} \tag{B.56}$$

denotes the sample signal covariance matrix of the symbol matrix $\boldsymbol{S}$. Then, we can express the norm $\left\|\tilde{\boldsymbol{s}}_i^\mathrm{T}\right\|_2^2$ as $\left\|\tilde{\boldsymbol{s}}_i^\mathrm{T}\right\|_2^2 = \|\tilde{\boldsymbol{s}}_i\|_2^2 = \tilde{\boldsymbol{s}}_i^\mathrm{H} \cdot \tilde{\boldsymbol{s}}_i$, resulting in

$$\|\tilde{\boldsymbol{s}}_i\|_2^2 = \frac{1}{N^2} \cdot \boldsymbol{e}_i^\mathrm{T} \cdot \hat{\boldsymbol{R}}_\mathrm{ss}^{-\mathrm{H}} \cdot \underbrace{\boldsymbol{S} \cdot \boldsymbol{S}^\mathrm{H}}_{N \cdot \hat{\boldsymbol{R}}_\mathrm{ss}} \cdot \hat{\boldsymbol{R}}_\mathrm{ss}^{-1} \cdot \boldsymbol{e}_i = \frac{1}{N} \cdot \boldsymbol{e}_i^\mathrm{T} \cdot \hat{\boldsymbol{R}}_\mathrm{ss}^{-\mathrm{H}} \cdot \boldsymbol{e}_i. \tag{B.57}$$

Special case $d = 2$

As we can see from (B.57), the term $\|\tilde{\boldsymbol{s}}_i\|_2^2$ will be determined either by the top-left or by the bottom-right element of $\hat{\boldsymbol{R}}_\mathrm{ss}^{-1}$, depending on the considered source i. Note that for $d = 2$, $\hat{\boldsymbol{R}}_\mathrm{ss}$ can be expressed as

$$\hat{\boldsymbol{R}}_\mathrm{ss} = \frac{1}{N} \cdot \begin{bmatrix} \boldsymbol{s}_1^\mathrm{H} \cdot \boldsymbol{s}_1, & \boldsymbol{s}_1^\mathrm{H} \cdot \boldsymbol{s}_2 \\ \boldsymbol{s}_2^\mathrm{H} \cdot \boldsymbol{s}_1, & \boldsymbol{s}_2^\mathrm{H} \cdot \boldsymbol{s}_2 \end{bmatrix} = \begin{bmatrix} \hat{P}_1 & \hat{\rho} \cdot \sqrt{\hat{P}_1 \cdot \hat{P}_2} \\ \hat{\rho}^* \cdot \sqrt{\hat{P}_1 \cdot \hat{P}_2} & \hat{P}_2 \end{bmatrix}, \tag{B.58}$$

where $\hat{\rho}$ denotes the complex-valued empirical correlation between the two sources, given by

$$\hat{\rho} = \frac{1}{N} \cdot \frac{\boldsymbol{s}^{(1)^\mathrm{H}} \cdot \boldsymbol{s}^{(2)}}{\sqrt{\hat{P}_1 \cdot \hat{P}_2}} = |\hat{\rho}| \cdot \mathrm{e}^{\mathrm{j}\varphi_\mathrm{corr}}. \tag{B.59}$$

The required inverse in (B.57) is thus given by

$$\hat{\boldsymbol{R}}_{\mathrm{ss}}^{-1} = \frac{1}{\hat{P}_1 \cdot \hat{P}_2 \cdot \left(1 - |\hat{\rho}|^2\right)} \cdot \begin{bmatrix} \hat{P}_2, & -\hat{\rho} \cdot \sqrt{\hat{P}_1 \cdot \hat{P}_2} \\ -\hat{\rho}^* \cdot \sqrt{\hat{P}_1 \cdot \hat{P}_2}, & \hat{P}_1 \end{bmatrix}. \tag{B.60}$$

Finally, by inserting (B.60) in (B.57) the norm $\|\tilde{\boldsymbol{s}}_i\|_2^2$ evaluates to

$$\|\tilde{\boldsymbol{s}}_i\|_2^2 = \frac{1}{N} \cdot \frac{\hat{P}_{\bar{i}}}{\hat{P}_1 \cdot \hat{P}_2 \cdot \left(1 - |\hat{\rho}|^2\right)}, \tag{B.61}$$

where we use the over-lined letter $\bar{i}$ to indicate terms related to the interfering source.

For the term $\left\|\tilde{\boldsymbol{a}}_i^{(r)^{\mathrm{T}}}\right\|_2^2$, let us consider the first source $i = 1$. Note that for $d = 2$, the scenario is symmetric, i.e., the sources are interchangeable. Expanding (B.54) yields

$$\begin{aligned} \tilde{\boldsymbol{a}}_1^{(r)^{\mathrm{T}}} &= \boldsymbol{e}_1^{\mathrm{T}} \cdot \left(\tilde{\boldsymbol{J}}_1^{(r)} \cdot \boldsymbol{A}\right)^{+} \cdot \left(\tilde{\boldsymbol{J}}_2^{(r)} / \lambda_1^{(r)} - \tilde{\boldsymbol{J}}_1^{(r)}\right) \\ &= \boldsymbol{e}_1^{\mathrm{T}} \cdot \left(\boldsymbol{A}^{\mathrm{H}} \cdot \tilde{\boldsymbol{J}}_1^{(r)^{\mathrm{H}}} \cdot \tilde{\boldsymbol{J}}_1^{(r)} \cdot \boldsymbol{A}\right)^{-1} \cdot \boldsymbol{A}^{\mathrm{H}} \cdot \tilde{\boldsymbol{J}}_1^{(r)^{\mathrm{H}}} \cdot \left(\tilde{\boldsymbol{J}}_2^{(r)} / \lambda_1^{(r)} - \tilde{\boldsymbol{J}}_1^{(r)}\right). \end{aligned} \tag{B.62}$$

Recalling that the R-D selection matrices are given by

$$\begin{aligned} \tilde{\boldsymbol{J}}_k^{(r)} &= \boldsymbol{I}_{M_1} \otimes \ldots \otimes \boldsymbol{I}_{M_{r-1}} \otimes \boldsymbol{J}_k^{(r)} \otimes \boldsymbol{I}_{M_{r+1}} \otimes \ldots \otimes \boldsymbol{I}_{M_R} \\ &= \boldsymbol{I}_{\prod_{\ell=1}^{r-1} M_\ell} \otimes \boldsymbol{J}_k^{(r)} \otimes \boldsymbol{I}_{\prod_{\ell=r+1}^{R} M_\ell} \in \mathbb{C}^{\frac{M}{M_r} M_r^{(\mathrm{sel})} \times M}, \; k = 1, 2, \end{aligned} \tag{B.63}$$

it is straightforward to see that

$$\boldsymbol{A}^{\mathrm{H}} \cdot \tilde{\boldsymbol{J}}_1^{(r)^{\mathrm{H}}} \cdot \tilde{\boldsymbol{J}}_1^{(r)} \cdot \boldsymbol{A} = \begin{bmatrix} \frac{M}{M_r} \cdot (M_r - 1) & \tilde{\alpha}^{(r)} \\ \tilde{\alpha}^{(r)^*} & \frac{M}{M_r} \cdot (M_r - 1) \end{bmatrix}, \tag{B.64}$$

where

$$\tilde{\alpha}^{(r)} = \boldsymbol{a}_1^{\mathrm{H}} \cdot \tilde{\boldsymbol{J}}_1^{(r)^{\mathrm{H}}} \cdot \tilde{\boldsymbol{J}}_1^{(r)} \cdot \boldsymbol{a}_2 = \alpha_{\mathrm{sel}}^{(r)} \cdot \prod_{\substack{q=1 \\ q \neq r}}^{R} \alpha^{(q)} \tag{B.65}$$

with $\alpha_{\mathrm{sel}}^{(r)} = \boldsymbol{a}_1^{(r)^{\mathrm{H}}} \cdot \boldsymbol{J}_1^{(r)^{\mathrm{H}}} \cdot \boldsymbol{J}_1^{(r)} \cdot \boldsymbol{a}_2^{(r)}$ and $\alpha^{(r)} = \boldsymbol{a}_1^{(r)^{\mathrm{H}}} \cdot \boldsymbol{a}_2^{(r)}$. Note that $\prod_{\substack{q=1 \\ q \neq r}}^{R} \alpha^{(q)} = \frac{\alpha}{\alpha^{(r)}}$, where $\alpha = \prod_{q=1}^{R} \alpha^{(q)} = \boldsymbol{a}_1^{\mathrm{H}} \cdot \boldsymbol{a}_2$. Consequently, the inverse is given by

$$\left(\boldsymbol{A}^{\mathrm{H}} \cdot \tilde{\boldsymbol{J}}_1^{(r)^{\mathrm{H}}} \cdot \tilde{\boldsymbol{J}}_1^{(r)} \cdot \boldsymbol{A}\right)^{-1} = \frac{1}{D_{\mathrm{sel}}^{(r)}} \cdot \begin{bmatrix} \frac{M}{M_r} \cdot (M_r - 1) & -\tilde{\alpha}^{(r)} \\ -\tilde{\alpha}^{(r)^*} & \frac{M}{M_r} \cdot (M_r - 1) \end{bmatrix}, \tag{B.66}$$

where $D_{\text{sel}}^{(r)} = \left(\frac{M}{M_r}\right)^2 \cdot (M_r - 1)^2 - \frac{|\alpha|^2}{|\alpha^{(r)}|^2} \cdot \left|\alpha_{\text{sel}}^{(r)}\right|^2$. Inserting (B.66) into (B.62), we obtain

$$\tilde{\boldsymbol{a}}_1^{(r)^{\mathrm{T}}} = \frac{1}{D_{\text{sel}}^{(r)}} \cdot \left(\frac{M}{M_r} \cdot (M_r - 1) \cdot \breve{\boldsymbol{a}}_1^{(r)^{\mathrm{H}}} - \tilde{\alpha}^{(r)} \cdot \breve{\boldsymbol{a}}_2^{(r)^{\mathrm{H}}} \right),$$

where we have defined

$$\breve{\boldsymbol{a}}_1^{(r)} = \boldsymbol{a}_1^{(1)} \otimes \ldots \otimes \boldsymbol{a}_1^{(r-1)} \otimes \bar{\boldsymbol{a}}_1^{(r)} \otimes \boldsymbol{a}_1^{(r+1)} \ldots \otimes \boldsymbol{a}_1^{(R)} \tag{B.67}$$

$$\breve{\boldsymbol{a}}_2^{(r)} = \boldsymbol{a}_2^{(1)} \otimes \ldots \otimes \boldsymbol{a}_2^{(r-1)} \otimes \bar{\boldsymbol{a}}_2^{(r)} \otimes \boldsymbol{a}_2^{(r+1)} \ldots \otimes \boldsymbol{a}_2^{(R)} \tag{B.68}$$

with

$$\bar{\boldsymbol{a}}_1^{(r)^{\mathrm{H}}} = \boldsymbol{a}_1^{(r)^{\mathrm{H}}} \cdot \left(\boldsymbol{J}_2^{(r)^{\mathrm{H}}} \cdot \boldsymbol{J}_2^{(r)} - \boldsymbol{J}_1^{(r)^{\mathrm{H}}} \cdot \boldsymbol{J}_1^{(r)} \right) = \boldsymbol{a}_1^{(r)^{\mathrm{H}}} \cdot \breve{\boldsymbol{J}}_1^{(r)^{\mathrm{H}}} \tag{B.69}$$

$$\bar{\boldsymbol{a}}_2^{(r)^{\mathrm{H}}} = \boldsymbol{a}_2^{(r)^{\mathrm{H}}} \cdot \left(\boldsymbol{J}_2^{(r)^{\mathrm{H}}} \cdot \boldsymbol{J}_2^{(r)} \cdot \frac{\lambda_2^{(r)}}{\lambda_1^{(r)}} - \boldsymbol{J}_1^{(r)^{\mathrm{H}}} \cdot \boldsymbol{J}_1^{(r)} \right) = \boldsymbol{a}_2^{(r)^{\mathrm{H}}} \cdot \breve{\boldsymbol{J}}_2^{(r)^{\mathrm{H}}}. \tag{B.70}$$

Then, we can compute $\left\| \tilde{\boldsymbol{a}}_1^{(r)} \right\|_2^2$ as

$$\left\| \tilde{\boldsymbol{a}}_1^{(r)} \right\|_2^2 = \frac{1}{D_{\text{sel}}^{(r)^2}} \cdot \left\| \frac{M}{M_r} \cdot (M_r - 1) \cdot \breve{\boldsymbol{a}}_1^{(r)^{\mathrm{H}}} - \tilde{\alpha}^{(r)} \cdot \breve{\boldsymbol{a}}_2^{(r)^{\mathrm{H}}} \right\|_2^2 \tag{B.71}$$

$$\begin{aligned} &= \frac{1}{D_{\text{sel}}^{(r)^2}} \cdot \left(\frac{M}{M_r} \cdot (M_r - 1) \cdot \breve{\boldsymbol{a}}_1^{(r)^{\mathrm{H}}} - \tilde{\alpha}^{(r)} \cdot \breve{\boldsymbol{a}}_2^{(r)^{\mathrm{H}}} \right) \cdot \left(\frac{M}{M_r} \cdot (M_r - 1) \cdot \breve{\boldsymbol{a}}_1^{(r)} - \tilde{\alpha}^{(r)^*} \cdot \breve{\boldsymbol{a}}_2^{(r)} \right) \\ &= \frac{1}{D_{\text{sel}}^{(r)^2}} \cdot \left(\left(\frac{M}{M_r} \right) \cdot (M_r - 1) \right)^2 \cdot \left\| \breve{\boldsymbol{a}}_1^{(r)} \right\|_2^2 + \left| \tilde{\alpha}^{(r)} \right|^2 \cdot \left\| \breve{\boldsymbol{a}}_2^{(r)} \right\|_2^2 \\ &\quad - 2 \cdot \frac{M}{M_r} \cdot (M_r - 1) \cdot \mathrm{Re} \left\{ \tilde{\alpha}^{(r)^*} \cdot \breve{\boldsymbol{a}}_1^{(r)^{\mathrm{H}}} \cdot \breve{\boldsymbol{a}}_2^{(r)} \right\}. \end{aligned} \tag{B.72}$$

In the next step, we simplify the following terms:

$$\left\| \breve{\boldsymbol{a}}_1^{(r)} \right\|_2^2 = \frac{M}{M_r} \cdot \left\| \bar{\boldsymbol{a}}_1^{(r)} \right\|_2^2 = \frac{M}{M_r} \cdot \left\| \breve{\boldsymbol{J}}_1^{(r)} \cdot \boldsymbol{a}_1^{(r)} \right\|_2^2 \tag{B.73}$$

$$\left\| \breve{\boldsymbol{a}}_2^{(r)} \right\|_2^2 = \frac{M}{M_r} \cdot \left\| \bar{\boldsymbol{a}}_2^{(r)} \right\|_2^2 = \frac{M}{M_r} \cdot \left\| \breve{\boldsymbol{J}}_2^{(r)} \cdot \boldsymbol{a}_2^{(r)} \right\|_2^2 \tag{B.74}$$

$$\breve{\boldsymbol{a}}_1^{(r)^{\mathrm{H}}} \cdot \breve{\boldsymbol{a}}_2^{(r)} = \bar{\boldsymbol{a}}_1^{(r)^{\mathrm{H}}} \cdot \bar{\boldsymbol{a}}_2^{(r)} \cdot \prod_{\substack{q=1 \\ q \neq r}}^{R} \alpha^{(q)} = \bar{\boldsymbol{a}}_1^{(r)^{\mathrm{H}}} \cdot \bar{\boldsymbol{a}}_2^{(r)} \cdot \frac{\alpha}{\alpha^{(r)}}.$$

Assuming maximum subarray overlap for the selection matrices, we obtain

$$\breve{\boldsymbol{J}}_1^{(r)^{\mathrm{H}}} = \mathrm{diag} \left\{ \begin{bmatrix} -1, & 0, & \ldots, & 0, & 1 \end{bmatrix} \right\} \tag{B.75}$$

$$\breve{\boldsymbol{J}}_2^{(r)^{\mathrm{H}}} = \mathrm{diag}\left\{\left[\begin{array}{cccccc} -1, & \mathrm{e}^{\mathrm{j}\Delta\mu^{(r)}}-1, & \ldots, & \mathrm{e}^{\mathrm{j}\Delta\mu^{(r)}}-1, & \mathrm{e}^{\mathrm{j}\Delta\mu^{(r)}} \end{array}\right]\right\} \tag{B.76}$$

such that

$$\left\|\breve{\boldsymbol{J}}_1^{(r)} \cdot \boldsymbol{a}_1^{(r)}\right\|_2^2 = \boldsymbol{a}_1^{(r)^{\mathrm{H}}} \cdot \breve{\boldsymbol{J}}_1^{(r)^{\mathrm{H}}} \cdot \breve{\boldsymbol{J}}_1^{(r)} \cdot \boldsymbol{a}_1^{(r)} = \boldsymbol{a}_1^{(r)^{\mathrm{H}}} \cdot \mathrm{diag}\left\{\left[\begin{array}{cccccc} 1, & 0, & \ldots, & 0, & 1 \end{array}\right]\right\} \cdot \boldsymbol{a}_1^{(r)} = 2 \tag{B.77}$$

$$\begin{aligned}\left\|\breve{\boldsymbol{J}}_2^{(r)} \cdot \boldsymbol{a}_2^{(r)}\right\|_2^2 &= \boldsymbol{a}_2^{(r)^{\mathrm{H}}} \cdot \mathrm{diag}\left\{\left[\begin{array}{cccccc} 1, & \left|\mathrm{e}^{\mathrm{j}\Delta\mu^{(r)}}-1\right|^2, & \ldots, & \left|\mathrm{e}^{\mathrm{j}\Delta\mu^{(r)}}-1\right|^2, & 1 \end{array}\right]\right\} \cdot \boldsymbol{a}_2^{(r)} \\ &= 2 + (M_r - 2) \cdot \left|\mathrm{e}^{\mathrm{j}\Delta\mu^{(r)}}-1\right|^2 .\end{aligned} \tag{B.78}$$

Moreover, the term $\tilde{\alpha}^{(r)^*} \cdot \breve{\boldsymbol{a}}_1^{(r)^{\mathrm{H}}} \cdot \breve{\boldsymbol{a}}_2^{(r)}$ in (B.72) simplifies to

$$\alpha_{\mathrm{sel}}^{(r)^*} \cdot \frac{\alpha^*}{\alpha^{(r)^*}} \cdot \bar{\boldsymbol{a}}_1^{(r)^{\mathrm{H}}} \cdot \bar{\boldsymbol{a}}_2^{(r)} \cdot \prod_{\substack{q=1 \\ q\neq r}}^{R} \boldsymbol{a}_1^{(q)^{\mathrm{H}}} \cdot \boldsymbol{a}_2^{(q)} = \frac{|\alpha|^2}{\left|\alpha^{(r)}\right|^2} \cdot \alpha_{\mathrm{sel}}^{(r)^*} \cdot \bar{\boldsymbol{a}}_1^{(r)^{\mathrm{H}}} \cdot \bar{\boldsymbol{a}}_2^{(r)}. \tag{B.79}$$

Since we assume a uniform R-D array, we have $\boldsymbol{a}\left(\mu_i^{(r)}\right) = \left[\begin{array}{cccc} 1, & \mathrm{e}^{-\mathrm{j}\cdot\mu_i^{(r)}}, & \ldots, & \mathrm{e}^{-\mathrm{j}\cdot(M_r-1)\cdot\mu_i^{(r)}} \end{array}\right]$ we can write in case of maximum overlap

$$\begin{aligned}\alpha_{\mathrm{sel}}^{(r)^*} \cdot \bar{\boldsymbol{a}}_1^{(r)^{\mathrm{H}}} \cdot \bar{\boldsymbol{a}}_2^{(r)} &= \alpha_{\mathrm{sel}}^{(r)^*} \cdot \boldsymbol{a}_1^{(r)^{\mathrm{H}}} \cdot \mathrm{diag}\left\{\left[\begin{array}{cccccc} 1, & 0, & \ldots, & 0, & \mathrm{e}^{-\mathrm{j}\Delta\mu^{(r)}} \end{array}\right]\right\} \cdot \boldsymbol{a}_2^{(r)} \\ &= \alpha_{\mathrm{sel},0}^{(r)^*} \cdot \left(1 + \cdot \mathrm{e}^{\mathrm{j}\cdot(M_r-2)\cdot\Delta\mu^{(r)}}\right) \\ &= \alpha_{\mathrm{sel},0}^{(r)^*} + \alpha_{\mathrm{sel},0}^{(r)^*} \cdot \mathrm{e}^{\mathrm{j}\cdot(M_r-2)\cdot\Delta\mu^{(r)}} = 2 \cdot \mathrm{Re}\left\{\alpha_{\mathrm{sel},0}^{(r)}\right\},\end{aligned} \tag{B.80}$$

where $\alpha_{\mathrm{sel},0}^{(r)} = \sum_{k=0}^{M_r-2} \mathrm{e}^{\mathrm{j}\cdot k\cdot\Delta\mu^{(r)}}$ and we have used that $\mathrm{e}^{\mathrm{j}\cdot(M_r-2)\cdot\Delta\mu^{(r)}} \cdot \sum_{k=0}^{M_r-2} \mathrm{e}^{-\mathrm{j}\cdot k\cdot\Delta\mu^{(r)}} = \alpha_{\mathrm{sel},0}^{(r)}$. Then, combining the results in (B.77), (B.78), and (B.80) into (B.72), we have

$$\begin{aligned}\left\|\tilde{\boldsymbol{a}}_1^{(r)}\right\|_2^2 = \frac{1}{D_{\mathrm{sel}}^{(r)^2}} \cdot \Bigg(& 2 \cdot (M_r-1)^2 \cdot \left(\frac{M}{M_r}\right)^3 + \left|\tilde{\alpha}^{(r)}\right|^2 \cdot \frac{M}{M_r} \cdot \left(2 + (M_r-2) \cdot \left|\mathrm{e}^{\mathrm{j}\Delta\mu^{(r)}}-1\right|^2\right) \\ & - 4 \cdot \frac{M}{M_r} \cdot (M_r-1) \cdot \frac{|\alpha|^2}{\left|\alpha^{(r)}\right|^2} \cdot \mathrm{Re}\left\{\alpha_{\mathrm{sel},0}^{(r)}\right\} \Bigg)\end{aligned} \tag{B.81}$$

$$\begin{aligned}= \frac{M}{D_{\mathrm{sel}}^{(r)^2} \cdot M_r} \cdot \Bigg(& 2 \cdot (M_r-1)^2 \cdot \left(\frac{M}{M_r}\right)^2 + \left|\tilde{\alpha}^{(r)}\right|^2 \cdot \left(2 + (M_r-2) \cdot \left|\mathrm{e}^{\mathrm{j}\Delta\mu^{(r)}}-1\right|^2\right) \\ & - 4 \cdot (M_r-1) \cdot \frac{|\alpha|^2}{\left|\alpha^{(r)}\right|^2} \cdot \mathrm{Re}\left\{\alpha_{\mathrm{sel},0}^{(r)}\right\} \Bigg)\end{aligned} \tag{B.82}$$

In the last step, we insert (B.61) and (B.82) into the MSE expression in (B.46) for $i = 1$ and the

r-th mode to obtain

$$\mathbb{E}\left\{\left(\Delta\mu_1^{(r)}\right)^2\right\} = \frac{\sigma_\mathrm{n}^2}{2} \cdot \frac{\hat{P}_2}{N \cdot \hat{P}_1 \cdot \hat{P}_2} \cdot a_\mathrm{mat}^{(r)}, \tag{B.83}$$

where $a_\mathrm{mat}^{(r)}$ is given in (4.54). Note that due to the symmetry of the scenario, we arrive at the same result by exchanging both sources. Consequently, the total MSE over both sources and the R modes is given in (4.53). This completes the proof. □

B.6. Proof of Theorem 4.5.2

The proof of Theorem 4.5.2 in Section 4.5.3 follows the same steps as the one in Appendix B.5. For circularly symmetric white noise, the MSE for the i-th spatial frequency in the r-th mode from (4.32) is given by

$$\mathbb{E}\left\{\left(\Delta\mu_i^{(r)}\right)^2\right\} = \frac{\sigma_\mathrm{n}^2}{2} \cdot \left(\left\|\boldsymbol{z}_i^{(r)^{(\mathrm{fba})}}\right\|_2^2 - \mathrm{Re}\left\{\boldsymbol{z}_i^{(r)^{(\mathrm{fba})^\mathrm{T}}} \cdot \boldsymbol{\Pi}_{2MN} \cdot \boldsymbol{z}_i^{(r)^{(\mathrm{fba})}}\right\}\right), \tag{B.84}$$

where $\boldsymbol{z}_i^{(r)^{(\mathrm{fba})}} = \boldsymbol{W}_\mathrm{mat}^{(\mathrm{fba})^\mathrm{T}} \cdot \boldsymbol{r}_i^{(r)^{(\mathrm{fba})}}$ with

$$\boldsymbol{r}_i^{(r)^{(\mathrm{fba})}} = \boldsymbol{q}_i^{(\mathrm{fba})} \otimes \left(\left[\left(\tilde{\boldsymbol{J}}_1^{(r)} \boldsymbol{U}_\mathrm{s}^{(\mathrm{fba})}\right)^+ \cdot \left(\tilde{\boldsymbol{J}}_2^{(r)} / \mathrm{e}^{\mathrm{j}\cdot\mu_i^{(r)}} - \tilde{\boldsymbol{J}}_1^{(r)}\right)\right]^\mathrm{T} \cdot \boldsymbol{p}_i^{(\mathrm{fba})}\right) \tag{B.85}$$

$$\boldsymbol{W}_\mathrm{mat}^{(\mathrm{fba})} = \left(\boldsymbol{\Sigma}_\mathrm{s}^{(\mathrm{fba})^{-1}} \cdot \boldsymbol{V}_\mathrm{s}^{(\mathrm{fba})^\mathrm{T}}\right) \otimes \left(\boldsymbol{U}_\mathrm{n}^{(\mathrm{fba})} \cdot \boldsymbol{U}_\mathrm{n}^{(\mathrm{fba})^\mathrm{H}}\right). \tag{B.86}$$

Note that we can express the noise-free measurement matrix $\boldsymbol{X}_0$ as $\boldsymbol{X}_0 = \boldsymbol{A}_\mathrm{c} \cdot \tilde{\boldsymbol{S}}$, where $\tilde{\boldsymbol{S}} = \boldsymbol{\Delta} \cdot \boldsymbol{S}$. This way, we obtain $\boldsymbol{X}_0$ as

$$\boldsymbol{X}_0^{(\mathrm{fba})} = \left[\begin{array}{cc} \boldsymbol{A}_\mathrm{c} \cdot \tilde{\boldsymbol{S}}, & \boldsymbol{A}_\mathrm{c} \cdot \tilde{\boldsymbol{S}}^* \cdot \boldsymbol{\Pi}_N \end{array}\right] = \boldsymbol{A}_\mathrm{c} \cdot \left[\begin{array}{cc} \tilde{\boldsymbol{S}}, & \tilde{\boldsymbol{S}}^* \cdot \boldsymbol{\Pi}_N \end{array}\right] = \boldsymbol{A}_\mathrm{c} \cdot \overline{\boldsymbol{S}}, \tag{B.87}$$

where we define $\overline{\boldsymbol{S}} = \left[\begin{array}{cc} \tilde{\boldsymbol{S}}, & \tilde{\boldsymbol{S}}^* \cdot \boldsymbol{\Pi}_N \end{array}\right] \in \mathbb{C}^{d \times 2N}$ to contain the forward-backward averaged source symbols and the array phase reference. From the SVD of (B.87), we immediately see that $\boldsymbol{U}_\mathrm{s}^{(\mathrm{fba})} = \boldsymbol{U}_\mathrm{s}$ and $\boldsymbol{U}_\mathrm{n}^{(\mathrm{fba})} = \boldsymbol{U}_\mathrm{n}$, which can be replaced in (B.85) and (B.86). Then, we express the signal subspace $\boldsymbol{U}_\mathrm{s}$ in terms of the array steering matrix $\boldsymbol{A}_\mathrm{c}$ as

$$\boldsymbol{U}_\mathrm{s} = \frac{\boldsymbol{A}_\mathrm{c}}{\sqrt{M}} \cdot \boldsymbol{P}. \tag{B.88}$$

By applying the steps (B.48) - (B.50) from Appendix B.6 to the model $\boldsymbol{X}_0^{(\mathrm{fba})} = \boldsymbol{A}_\mathrm{c} \cdot \overline{\boldsymbol{S}}$, the term $\boldsymbol{V}_\mathrm{s}^{(\mathrm{fba})} \cdot \boldsymbol{\Sigma}_\mathrm{s}^{(\mathrm{fba})^{-1}}$ can be expressed as

$$\boldsymbol{V}_\mathrm{s}^{(\mathrm{fba})} \cdot \boldsymbol{\Sigma}_\mathrm{s}^{(\mathrm{fba})^{-1}} = \frac{1}{\sqrt{M}} \cdot \overline{\boldsymbol{S}}^{+} \cdot \boldsymbol{P}. \tag{B.89}$$

The projection matrix onto the orthogonal complement of the signal subspace $\boldsymbol{U}_\mathrm{n} \cdot \boldsymbol{U}_\mathrm{n}^\mathrm{H}$ can be written as

$$\boldsymbol{U}_\mathrm{n} \cdot \boldsymbol{U}_\mathrm{n}^\mathrm{H} = \boldsymbol{P}_{\boldsymbol{A}}^{\perp} = \boldsymbol{I}_M - \boldsymbol{A}_\mathrm{c} \cdot \boldsymbol{A}_\mathrm{c}^{+}. \tag{B.90}$$

Inserting the results from (B.88), (B.89), and (B.90) into (B.85) and (B.86), we can simplify $\boldsymbol{z}_i^{(\mathrm{fba})(r)}$ as

$$\boldsymbol{z}_i^{(\mathrm{fba})(r)^\mathrm{T}} = \tilde{\boldsymbol{s}}_i^{(\mathrm{fba})^\mathrm{T}} \otimes \tilde{\boldsymbol{a}}_i^{(\mathrm{fba})(r)^\mathrm{T}}, \tag{B.91}$$

where the terms $\tilde{\boldsymbol{s}}_i^{(\mathrm{fba})^\mathrm{T}}$ and $\tilde{\boldsymbol{a}}_i^{(\mathrm{fba})(r)^\mathrm{T}}$ are given by

$$\begin{aligned}
\tilde{\boldsymbol{s}}_i^{(\mathrm{fba})^\mathrm{T}} &= \boldsymbol{e}_i^\mathrm{T} \cdot \overline{\boldsymbol{S}}^{+\mathrm{T}} \\
\tilde{\boldsymbol{a}}_i^{(\mathrm{fba})(r)^\mathrm{T}} &= \boldsymbol{e}_i^\mathrm{T} \cdot \left(\tilde{\boldsymbol{J}}_1^{(r)} \cdot \boldsymbol{A}_\mathrm{c}\right)^{+} \cdot \left(\tilde{\boldsymbol{J}}_2^{(r)} / \lambda_i^{(r)} - \tilde{\boldsymbol{J}}_1^{(r)}\right) \cdot \boldsymbol{P}_{\boldsymbol{A}_\mathrm{c}}^{\perp},
\end{aligned}$$

where the factor $\sqrt{M}$ cancels. Using property (1.7), we can compute the MSE in (B.84) as

$$\begin{aligned}
\mathbb{E}\left\{\left(\Delta\mu_i^{(r)}\right)^2\right\} = \frac{\sigma_\mathrm{n}^2}{2} \cdot \Bigg(&\left\|\tilde{\boldsymbol{s}}_i^{(\mathrm{fba})^\mathrm{T}}\right\|_2^2 \cdot \left\|\tilde{\boldsymbol{a}}_i^{(\mathrm{fba})^\mathrm{T}}\right\|_2^2 \\
&- \mathrm{Re}\left\{\tilde{\boldsymbol{s}}_i^{(\mathrm{fba})^\mathrm{T}} \cdot \boldsymbol{\Pi}_{2N} \cdot \tilde{\boldsymbol{s}}_i^{(\mathrm{fba})} \cdot \boldsymbol{a}_i^{(\mathrm{fba})(r)^\mathrm{T}} \cdot \boldsymbol{\Pi}_M \cdot \boldsymbol{a}_i^{(\mathrm{fba})(r)}\right\}\Bigg).
\end{aligned} \tag{B.92}$$

Considering the term $\left\|\tilde{\boldsymbol{s}}_i^{(\mathrm{fba})^\mathrm{T}}\right\|_2^2$, we first expand the pseudo-inverse $\boldsymbol{S}^{+\mathrm{T}}$ in $\tilde{\boldsymbol{s}}_i^{(\mathrm{fba})^\mathrm{T}} = \boldsymbol{e}_i^\mathrm{T} \cdot \overline{\boldsymbol{S}}^{+\mathrm{T}}$ as

$$\overline{\boldsymbol{S}}^{+} = \overline{\boldsymbol{S}}^\mathrm{H} \cdot \left(\overline{\boldsymbol{S}} \cdot \overline{\boldsymbol{S}}^\mathrm{H}\right)^{-1} = \frac{1}{2 \cdot N} \cdot \overline{\boldsymbol{S}}^\mathrm{H} \cdot \hat{\boldsymbol{R}}_{\overline{\mathrm{ss}}}^{-1}. \tag{B.93}$$

Special case $d = 2$

The empirical correlation matrix for the forward-backward averaged symbols in $d = 2$ is given by

$$\begin{aligned}
\hat{\boldsymbol{R}}_{\overline{\mathrm{ss}}} &= \frac{1}{2 \cdot N} \cdot \overline{\boldsymbol{S}} \cdot \overline{\boldsymbol{S}}^\mathrm{H} = \frac{1}{2 \cdot N} \cdot \left(\begin{bmatrix} \tilde{\boldsymbol{S}}, & \tilde{\boldsymbol{S}}^* \cdot \boldsymbol{\Pi}_N \end{bmatrix} \cdot \begin{bmatrix} \tilde{\boldsymbol{S}}^\mathrm{H} \\ \boldsymbol{\Pi}_N \cdot \tilde{\boldsymbol{S}}^\mathrm{T} \end{bmatrix}\right) = \frac{1}{2} \cdot \left(\hat{\boldsymbol{R}}_{\tilde{\mathrm{ss}}} + \hat{\boldsymbol{R}}_{\tilde{\mathrm{ss}}}^\mathrm{T}\right) \\
&= \frac{1}{2} \cdot \left(\begin{bmatrix} \hat{P}_1, & \hat{\rho} \cdot \sqrt{\hat{P}_1 \cdot \hat{P}_2} \cdot \mathrm{e}^{-\mathrm{j}\Delta\varphi} \\ \hat{\rho}^* \cdot \sqrt{\hat{P}_1 \cdot \hat{P}_2} \cdot \mathrm{e}^{\mathrm{j}\Delta\varphi}, & \hat{P}_2 \end{bmatrix}\right.
\end{aligned}$$

$$+\begin{bmatrix} \hat{P}_1, & \hat{\rho}^* \cdot \sqrt{\hat{P}_1 \cdot \hat{P}_2} \cdot \mathrm{e}^{\mathrm{j}\Delta\varphi} \\ \hat{\rho} \cdot \sqrt{\hat{P}_1 \cdot \hat{P}_2} \cdot \mathrm{e}^{-\mathrm{j}\Delta\varphi}, & \hat{P}_2 \end{bmatrix}\Bigg)$$
$$= \begin{bmatrix} \hat{P}_1, & |\hat{\rho}| \cdot \sqrt{\hat{P}_1 \cdot \hat{P}_2} \cdot \cos(\Delta\varphi) \\ |\hat{\rho}| \cdot \sqrt{\hat{P}_1 \cdot \hat{P}_2} \cdot \cos(\Delta\varphi), & \hat{P}_2 \end{bmatrix}, \tag{B.94}$$

where $\Delta\varphi = \Delta\varphi_{\mathrm{ref}} + \varphi_{\mathrm{corr}}$ combines the phase terms for the empirical source correlation $\hat{\rho} = \mathrm{e}^{\mathrm{j}\cdot\varphi_{\mathrm{corr}}} \cdot |\hat{\rho}|$, and the array phase separation $\Delta\varphi_{\mathrm{ref}} = \varphi_{\mathrm{ref},2} - \varphi_{\mathrm{ref},1} = \sum_{r=1}^{R} \delta^{(r)} \cdot \Delta\mu^{(r)}$.
Thus, $\hat{\boldsymbol{R}}_{\overline{\mathrm{ss}}}^{-1}$ is given by

$$\hat{\boldsymbol{R}}_{\overline{\mathrm{ss}}}^{-1} = \frac{1}{D^{(\mathrm{fba})}} \cdot \begin{bmatrix} \hat{P}_2, & -|\hat{\rho}| \cdot \sqrt{\hat{P}_1 \cdot \hat{P}_2} \cdot \cos(\Delta\varphi) \\ -|\hat{\rho}| \cdot \sqrt{\hat{P}_1 \cdot \hat{P}_2} \cdot \cos(\Delta\varphi), & \hat{P}_1 \end{bmatrix}, \tag{B.95}$$

where $D^{(\mathrm{fba})} = \hat{P}_1 \cdot \hat{P}_2 \cdot \left(1 - |\hat{\rho}|^2 \cdot \cos^2(\Delta\varphi)\right)$. Subsequently, the term $\left\| \tilde{\boldsymbol{s}}_i^{(\mathrm{fba})^{\mathrm{T}}} \right\|_2^2$ is given by

$$\left\| \tilde{\boldsymbol{s}}_i^{(\mathrm{fba})} \right\|_2^2 = \frac{1}{2 \cdot N} \cdot \boldsymbol{e}_i^{\mathrm{T}} \cdot \hat{\boldsymbol{R}}_{\overline{\mathrm{ss}}}^{-1} \cdot \boldsymbol{e}_i = \frac{1}{2 \cdot N} \cdot \frac{\hat{P}_{\bar{i}}}{\hat{P}_1 \cdot \hat{P}_2 \cdot \left(1 - |\hat{\rho}|^2 \cdot \cos^2(\Delta\varphi)\right)}. \tag{B.96}$$

Note that the term $\left\| \tilde{\boldsymbol{a}}_i^{(\mathrm{fba})^{\mathrm{T}}} \right\|_2^2$ is identical to the term $\left\| \tilde{\boldsymbol{a}}_i^{\mathrm{T}} \right\|_2^2$ from (B.82) for $i = 1$. This is due to the fact that $\boldsymbol{U}_{\mathrm{s}}^{(\mathrm{fba})} = \boldsymbol{U}_{\mathrm{s}}$ and $\boldsymbol{U}_{\mathrm{n}}^{(\mathrm{fba})} = \boldsymbol{U}_{\mathrm{n}}$. Furthermore, it is easily verified that $\tilde{\boldsymbol{s}}_i^{(\mathrm{fba})^{\mathrm{T}}} \cdot \boldsymbol{\Pi}_{2N} \cdot \tilde{\boldsymbol{s}}_i^{(\mathrm{fba})} = \left\| \tilde{\boldsymbol{s}}_i^{(\mathrm{fba})^{\mathrm{T}}} \right\|_2^2$ and that $\boldsymbol{a}_i^{(\mathrm{fba})(r)^{\mathrm{T}}} \cdot \boldsymbol{\Pi}_M \cdot \boldsymbol{a}_i^{(\mathrm{fba})(r)} = -\left\| \tilde{\boldsymbol{a}}_i^{(\mathrm{fba})^{\mathrm{T}}} \right\|_2^2$. Combining all these terms, we obtain the MSE expression in (B.92) for $i = 1$ and the r-th mode as

$$\mathbb{E}\left\{ \left(\Delta\mu_1^{(r)} \right)^2 \right\} = \frac{\sigma_{\mathrm{n}}^2}{2} \cdot \frac{\hat{P}_2}{N \cdot \hat{P}_1 \cdot \hat{P}_2} \cdot a_{\mathrm{mat}}^{(\mathrm{fba})(r)} \tag{B.97}$$

where $a_{\mathrm{mat}}^{(\mathrm{fba})(r)}$ is given in (4.58). Due to the symmetry of the scenario, we arrive at the same result for both sources. Therefore, the total MSE over both sources and the R modes is given in (4.57). This completes the proof. □

B.7. Proof of Theorem 4.5.3

In this Appendix, we provide a sketch of the proof of Theorem 4.5.3 in Section 4.5.4. For the sake of brevity, we have left out some of the lengthy derivations, however, the full proof is provided by us in [Gra16]. For the simplification of the MSE expression of R-D Standard Tensor-ESPRIT assuming circularly symmetric white noise and a uniform R-D array geometry, we first recall the

MSE expression in (4.38) for the i-th spatial frequency in the r-th mode as

$$\mathrm{MSE}_{\mathrm{ten},i}^{(r)} = \mathbb{E}\left\{\left(\Delta\mu_i^{(r)}\right)^2\right\} = \frac{\sigma_\mathrm{n}^2}{2}\cdot\left\|\boldsymbol{z}_i^{(r)}\right\|_2^2, \tag{B.98}$$

where $\boldsymbol{z}_i^{(r)^\mathrm{T}} = \boldsymbol{r}_i^{(r)^\mathrm{T}}\cdot\boldsymbol{W}_\mathrm{ten}$. The vector $\boldsymbol{r}_i^{(r)}$ and the matrix $\boldsymbol{W}_\mathrm{ten}$ are given by

$$\boldsymbol{r}_i^{(r)} = \boldsymbol{q}_i^{(r)}\otimes\left(\left[\left(\tilde{\boldsymbol{J}}_1^{(r)}\cdot\boldsymbol{U}_\mathrm{s}\right)^+\left(\tilde{\boldsymbol{J}}_2^{(r)}/\mathrm{e}^{\mathrm{j}\cdot\mu_i^{(r)}}-\tilde{\boldsymbol{J}}_1^{(r)}\right)\right]^\mathrm{T}\cdot\boldsymbol{p}_i^{(r)}\right) \tag{B.99}$$

$$\boldsymbol{W}_\mathrm{ten} = \boldsymbol{W}_0 + \sum_{q=1}^{R}\boldsymbol{W}_q\cdot\boldsymbol{P}_{M_1,\ldots,M_R,N}^{(q)^\mathrm{T}}\cdot\boldsymbol{P}_{M_1,\ldots,M_R,N}^{(R)}, \tag{B.100}$$

where the permutation matrices $\boldsymbol{P}_{M_1,\ldots,M_R,N}^{(q)^\mathrm{T}}$ and $\boldsymbol{P}_{M_1,\ldots,M_R,N}^{(R)}$ are defined according to (1.38) and

$$\boldsymbol{W}_0 = \left(\boldsymbol{\Sigma}_\mathrm{s}^{-1}\cdot\boldsymbol{V}_\mathrm{s}^\mathrm{T}\right)\otimes\left(\boldsymbol{T}_{1:R}^{\otimes}\cdot\boldsymbol{U}_\mathrm{n}\cdot\boldsymbol{U}_\mathrm{n}^\mathrm{H}\right) \tag{B.101}$$

$$\begin{aligned}\boldsymbol{W}_q = &\left(\boldsymbol{U}_\mathrm{s}^\mathrm{T}\otimes\boldsymbol{I}_M\right)\cdot\left(\bar{\boldsymbol{T}}_{1:q-1}\otimes\boldsymbol{I}_{M_{q:R}}\right)\cdot\left(\boldsymbol{I}_{M_q}\otimes\bar{\boldsymbol{T}}_{q+1:R}\right)\\ &\cdot\left[\left(\boldsymbol{V}_q^{[\mathrm{s}]}\cdot\boldsymbol{\Sigma}_q^{[\mathrm{s}]^{-1}}\cdot\boldsymbol{U}_q^{[\mathrm{s}]^\mathrm{H}}\right)^\mathrm{T}\otimes\left(\boldsymbol{U}_q^{[\mathrm{n}]}\cdot\boldsymbol{U}_q^{[\mathrm{n}]^\mathrm{H}}\right)\right]\end{aligned} \tag{B.102}$$

with

$$\bar{\boldsymbol{T}}_{1:q-1} = \begin{bmatrix}\boldsymbol{I}_{M_{q:R}}\otimes\boldsymbol{t}_{1:q-1,1}\\ \vdots\\ \boldsymbol{I}_{M_{q:R}}\otimes\boldsymbol{t}_{1:q-1,M_{1:q-1}}\end{bmatrix},\quad \bar{\boldsymbol{T}}_{q+1:R} = \begin{bmatrix}\boldsymbol{I}_{M_q}\otimes\boldsymbol{t}_{q+1:R,1}\\ \vdots\\ \boldsymbol{I}_{M_q}\otimes\boldsymbol{t}_{q+1:R,M_{q+1:R}}\end{bmatrix}. \tag{B.103}$$

Moreover, we have

$$M_{a:b} = \begin{cases}\prod_{r=a}^{b} M_r & a\le b\\ 1 & a>b\end{cases} \tag{B.104}$$

and $\boldsymbol{t}_{a:b,n}$ denotes the n-th column of

$$\boldsymbol{T}_{a:b}^{\otimes} = \begin{cases}\boldsymbol{T}_a\otimes\ldots\otimes\boldsymbol{T}_b & a\le b\\ 1 & a>b\end{cases}, \tag{B.105}$$

where $\boldsymbol{T}_q = \boldsymbol{U}_q^{[\mathrm{s}]}\cdot\boldsymbol{U}_q^{[\mathrm{s}]^\mathrm{H}}$.

For the derivation, we first insert $\boldsymbol{W}_\mathrm{ten}$ from (B.100) into the vector $\boldsymbol{z}_i^{(r)^\mathrm{T}} = \boldsymbol{r}_i^{(r)^\mathrm{T}}\cdot\boldsymbol{W}_\mathrm{ten}$ to obtain

$$\boldsymbol{r}_i^{(r)^\mathrm{T}}\cdot\boldsymbol{W}_\mathrm{ten} = \boldsymbol{r}_i^{(r)^\mathrm{T}}\cdot\boldsymbol{W}_0 + \boldsymbol{r}_i^{(r)^\mathrm{T}}\cdot\sum_{q=1}^{R}\boldsymbol{W}_q\cdot\boldsymbol{P}_{M_1,\ldots,M_R,N}^{(q)^\mathrm{T}}\cdot\boldsymbol{P}_{M_1,\ldots,M_R,N}^{(R)}$$

$$= \boldsymbol{z}_{i,0}^{(r)^{\mathrm{T}}} + \boldsymbol{z}_{i,R}^{(r)^{\mathrm{T}}}. \tag{B.106}$$

Then, it can be shown that the terms $\boldsymbol{z}_{i,0}^{(r)}$ and $\boldsymbol{z}_{i,R}^{(r)}$ are orthogonal. Hence, we can compute $\left\|\boldsymbol{z}_i^{(r)}\right\|_2^2$ as

$$\left\|\boldsymbol{z}_i^{(r)}\right\|_2^2 = \left\|\boldsymbol{z}_{i,0}^{(r)}\right\|_2^2 + \left\|\boldsymbol{z}_{i,R}^{(r)}\right\|_2^2. \tag{B.107}$$

Thus, we proceed to simplify $\left\|\boldsymbol{z}_{i,0}^{(r)}\right\|_2^2$. To this end, we make use of the fact that $\boldsymbol{z}_{i,0}^{(r)^{\mathrm{T}}} = \boldsymbol{r}_i^{(r)^{\mathrm{T}}} \cdot \boldsymbol{W}_0$ can be expressed as the Kronecker product

$$\boldsymbol{z}_{i,0}^{(r)^{\mathrm{T}}} = \tilde{\boldsymbol{s}}_{i,0}^{\mathrm{T}} \otimes \tilde{\boldsymbol{a}}_{i,0}^{(r)^{\mathrm{T}}}, \tag{B.108}$$

where

$$\tilde{\boldsymbol{s}}_{i,0}^{\mathrm{T}} = \boldsymbol{e}_i^{\mathrm{T}} \cdot \boldsymbol{S}^{+\mathrm{T}} \tag{B.109}$$

$$\tilde{\boldsymbol{a}}_{i,0}^{(r)^{\mathrm{T}}} = \boldsymbol{e}_i^{\mathrm{T}} \cdot \left(\tilde{\boldsymbol{J}}_1^{(r)} \cdot \boldsymbol{A}\right)^{+} \cdot \left(\tilde{\boldsymbol{J}}_2^{(r)} / \lambda_i^{(r)} - \tilde{\boldsymbol{J}}_1^{(r)}\right) \cdot \boldsymbol{T}_{1:R}^{\otimes} \cdot \boldsymbol{P}_{\boldsymbol{A}}^{\perp}. \tag{B.110}$$

Thus, we can separate $\boldsymbol{z}_{i,0}^{(r)^{\mathrm{T}}}$ into $\tilde{\boldsymbol{s}}_{i,0}^{\mathrm{T}}$, which is related to the source symbols and $\tilde{\boldsymbol{a}}_{i,0}^{(r)^{\mathrm{T}}}$, which is related to the array steering matrix $\boldsymbol{A}$. Using property (1.7), we compute $\left\|\boldsymbol{z}_{i,0}^{(r)}\right\|_2^2$ as

$$\left\|\boldsymbol{z}_{i,0}^{(r)}\right\|_2^2 = \|\tilde{\boldsymbol{s}}_{i,0}\|_2^2 \cdot \left\|\tilde{\boldsymbol{a}}_{i,0}^{(r)}\right\|_2^2. \tag{B.111}$$

The norm $\|\tilde{\boldsymbol{s}}_{i,0}\|_2^2$ can be determined via straightforward calculations as

$$\|\tilde{\boldsymbol{s}}_{i,0}\|_2^2 = \frac{1}{N} \cdot \frac{\hat{P}_{\bar{i}}}{\hat{P}_1 \cdot \hat{P}_2 \cdot \left(1 - |\hat{\rho}|^2\right)}, \tag{B.112}$$

where $\hat{\rho}$ denotes the empirical source correlation $\hat{\rho} = \frac{1}{N} \cdot \frac{\boldsymbol{s}^{(1)^{\mathrm{H}}} \cdot \boldsymbol{s}^{(2)}}{\sqrt{\hat{P}_1 \cdot \hat{P}_2}} = |\hat{\rho}| \cdot \mathrm{e}^{\mathrm{j} \cdot \varphi_{\mathrm{corr}}}$. The over-lined letter $\bar{i}$ indicates a quantity related to the interfering source. For instance, $\hat{P}_{\bar{i}}$ denotes the empirical power of the interferer. Note that the norm $\|\tilde{\boldsymbol{s}}_{i,0}\|_2^2$ scales with $\hat{P}_{\bar{i}}$, which corresponds to the intuition that a strong interferer decreases the estimation accuracy for the desired source.

In order to compute $\left\|\tilde{\boldsymbol{a}}_{i,0}^{(r)}\right\|_2^2$, we assume maximally overlapping subarrays and obtain

$$\left\|\tilde{\boldsymbol{a}}_{i,0}^{(r)}\right\|_2^2 = \frac{b^{(r)}}{D_{\mathrm{sel}}^{(r)^2} \cdot D_r}, \tag{B.113}$$

where the determinants $D_{\text{sel}}^{(r)}$ and D_r can be expressed as

$$D_{\text{sel}}^{(r)} = \left(\frac{M}{M_r}\right)^2 \cdot (M_r - 1)^2 - \frac{|\alpha|^2}{\left|\alpha^{(r)}\right|^2} \cdot \left|\alpha_{\text{sel}}^{(r)}\right|^2 \quad \text{and} \quad D_r = M_r^2 - \left|\alpha^{(r)}\right|^2, \tag{B.114}$$

and the scalar $b^{(r)}$ is given by

$$b^{(r)} = M \cdot (M_r - 1)^2 \cdot \left|\mathrm{e}^{\mathrm{j}\Delta\mu^{(r)}} - 1\right|^2 \cdot \left|\alpha_{\text{sel}}^{(r)}\right|^2 \cdot \left[\left(\frac{M}{M_r}\right)^2 - \frac{|\alpha|^2}{\left|\alpha^{(r)}\right|^2}\right], \tag{B.115}$$

where $\alpha^{(r)} = \boldsymbol{a}_i^{(r)^{\mathrm{H}}} \cdot \boldsymbol{a}_{\bar{i}}^{(r)}$, $\alpha = \prod_{r=1}^{R} \alpha^{(r)}$, and $\alpha_{\text{sel}}^{(r)} = \boldsymbol{a}_i^{(r)^{\mathrm{H}}} \cdot \boldsymbol{J}_1^{(r)^{\mathrm{H}}} \cdot \boldsymbol{J}_1^{(r)} \cdot \boldsymbol{a}_{\bar{i}}^{(r)}$.

In the next step, we derive a simplified expression for the term $\left\|\boldsymbol{z}_{i,R}^{(r)}\right\|_2^2$. Again, we find that $\boldsymbol{z}_{i,R}^{(r)^{\mathrm{T}}} = \boldsymbol{r}_i^{(r)^{\mathrm{T}}} \cdot \sum_{q=1}^{R} \boldsymbol{W}_q \cdot \boldsymbol{P}_{M_1,\ldots,M_R,N}^{(q)^{\mathrm{T}}} \cdot \boldsymbol{P}_{M_1,\ldots,M_R,N}^{(R)}$ reduces to the term

$$\boldsymbol{z}_{i,R}^{(r)^{\mathrm{T}}} = \boldsymbol{z}_{i,r}^{(r)^{\mathrm{T}}} \cdot \boldsymbol{P}_{M_1,\ldots,M_R,N}^{(r)^{\mathrm{T}}} \cdot \boldsymbol{P}_{M_1,\ldots,M_R,N}^{(R)}, \tag{B.116}$$

where $\boldsymbol{z}_{i,r}^{(r)^{\mathrm{T}}} = \boldsymbol{r}_i^{(r)^{\mathrm{T}}} \cdot \boldsymbol{W}_r$ can again be expressed as the Kronecker product

$$\boldsymbol{z}_{i,r}^{(r)^{\mathrm{T}}} = \tilde{\boldsymbol{s}}_{i,r}^{(r)^{\mathrm{T}}} \otimes \tilde{\boldsymbol{a}}_{i,r}^{(r)^{\mathrm{T}}}, \tag{B.117}$$

where

$$\tilde{\boldsymbol{s}}_{i,r}^{(r)^{\mathrm{T}}} = \boldsymbol{e}_i^{\mathrm{T}} \cdot \left[\boldsymbol{A}_{r+1:R}^{\diamond} \diamond \boldsymbol{S}^{\mathrm{T}} \diamond \boldsymbol{A}_{1:r-1}^{\diamond}\right]^{+} \tag{B.118}$$

$$\tilde{\boldsymbol{a}}_{i,r}^{(r)^{\mathrm{T}}} = \boldsymbol{e}_i^{\mathrm{T}} \cdot \left(\tilde{\boldsymbol{J}}_1^{(r)} \cdot \boldsymbol{A}\right)^{+} \cdot \left(\tilde{\boldsymbol{J}}_2^{(r)} / \lambda_i^{(r)} - \tilde{\boldsymbol{J}}_1^{(r)}\right) \cdot \left(\boldsymbol{a}_i^{\otimes(1:r-1)} \otimes \boldsymbol{P}_{\boldsymbol{A}^{(r)}}^{\perp} \otimes \boldsymbol{a}_i^{\otimes(r+1:R)}\right). \tag{B.119}$$

As shown in property (1.41), the matrices $\boldsymbol{P}_{M_1,\ldots,M_R,N}^{(r)^{\mathrm{T}}}$ and $\boldsymbol{P}_{M_1,\ldots,M_R,N}^{(R)}$ in (B.116) perform a circular shift of the elements contained in $\boldsymbol{z}_{i,R}^{(r)}$. However, this does not affect the norm of $\boldsymbol{z}_{i,R}^{(r)}$ such that $\left\|\boldsymbol{z}_{i,R}^{(r)}\right\|_2^2 = \left\|\boldsymbol{z}_{i,r}^{(r)}\right\|_2^2$. Applying property (1.7), we can determine the norm $\left\|\boldsymbol{z}_{i,R}^{(r)}\right\|_2^2$ via

$$\left\|\boldsymbol{z}_{i,R}^{(r)}\right\|_2^2 = \left\|\tilde{\boldsymbol{s}}_{i,r}^{(r)}\right\|_2^2 \cdot \left\|\tilde{\boldsymbol{a}}_{i,r}^{(r)}\right\|_2^2. \tag{B.120}$$

We first compute the norm of $\tilde{\boldsymbol{s}}_{i,r}^{(r)}$ as

$$\left\|\tilde{\boldsymbol{s}}_{i,r}^{(r)}\right\|_2^2 = \frac{\frac{M}{M_r} \cdot \hat{P}_{\bar{i}}}{N \cdot \hat{P}_1 \cdot \hat{P}_2 \cdot \left[\left(\frac{M}{M_r}\right)^2 - |\hat{\rho}|^2 \cdot \left|\frac{\alpha}{\alpha^{(r)}}\right|^2\right]}, \tag{B.121}$$

where we have assumed maximum subarray overlap since $\tilde{\boldsymbol{s}}_{i,r}$ also contains the array steering matrices $\boldsymbol{A}^{(r)}$. Once more assuming maximum subarray overlap in the r-th mode, the term $\left\|\tilde{\boldsymbol{a}}_{i,r}^{(r)}\right\|_2^2$ can be simplified into

$$\left\|\tilde{\boldsymbol{a}}_{i,r}^{(r)}\right\|_2^2 = \frac{1}{D_{\text{sel}}^{(r)^2}} \cdot \left[c^{(r)} - \frac{1}{D_r} \cdot M_r \cdot (M_r - 1)^2 \cdot \left| \mathrm{e}^{\mathrm{j}\Delta\mu^{(r)}} - 1 \right|^2 \cdot \left| \alpha_{\text{sel}}^{(r)} \right|^2 \cdot \left[\left(\frac{M}{M_r} \right)^2 - \frac{|\alpha|^2}{\left|\alpha^{(r)}\right|^2} \right]^2 \right], \quad \text{(B.122)}$$

where

$$\begin{aligned} c^{(r)} = {} & \left(\frac{M}{M_r} \right)^4 \cdot (M_r - 1)^2 \cdot 2 + \left| \frac{\alpha}{\alpha^{(r)}} \right|^4 \cdot \left| \alpha_{\text{sel}}^{(r)} \right|^2 \cdot \left(2 + (M_r - 2) \cdot \left| \mathrm{e}^{\mathrm{j}\Delta\mu^{(r)}} - 1 \right|^2 \right) \\ & - 4 \cdot \left(\frac{M}{M_r} \right)^2 \cdot (M_r - 1) \cdot \left| \frac{\alpha}{\alpha^{(r)}} \right|^2 \cdot \mathrm{Re}\left\{ \alpha_{\text{sel},0}^{(r)} \right\}. \end{aligned} \quad \text{(B.123)}$$

Note that the term $c^{(r)}$ closely resembles the scalar $a_{\text{mat}}^{(r)}$ from the MSE expression for R-D Standard ESPRIT given in (4.53). Moreover, the term which is subtracted in (B.122) and which results from the multiplication with the projection matrix $\boldsymbol{P}_{\boldsymbol{A}^{(r)}}^{\perp}$ in (B.119) is very similar to the term $b^{(r)}$

$$b^{(r)} = M \cdot (M_r - 1)^2 \cdot \left| \mathrm{e}^{\mathrm{j}\Delta\mu^{(r)}} - 1 \right|^2 \cdot \left| \alpha_{\text{sel}}^{(r)} \right|^2 \cdot \left[\left(\frac{M}{M_r} \right)^2 - \frac{|\alpha|^2}{\left|\alpha^{(r)}\right|^2} \right] \quad \text{(B.124)}$$

that we defined for $\tilde{\boldsymbol{a}}_{i,0}^{(r)}$ in (B.115). In fact, we have

$$\begin{aligned} & \frac{1}{D_r} \cdot M_r \cdot (M_r - 1)^2 \cdot \left| \mathrm{e}^{\mathrm{j}\Delta\mu^{(r)}} - 1 \right|^2 \cdot \left| \alpha_{\text{sel}}^{(r)} \right|^2 \cdot \left[\left(\frac{M}{M_r} \right)^2 - \frac{|\alpha|^2}{\left|\alpha^{(r)}\right|^2} \right]^2 \Bigg] \\ & = b^{(r)} \cdot \left[\left(\frac{M}{M_r} \right)^2 - \frac{|\alpha|^2}{\left|\alpha^{(r)}\right|^2} \right] \cdot \frac{M_r}{M}. \end{aligned} \quad \text{(B.125)}$$

As we see, when multiplied with the signal part $\tilde{\boldsymbol{s}}_{i,r}^{(r)}$ from (B.121) the term completes to the term $b^{(r)}$, which in turn can be factored out to find a compact formulation for the MSE, as will be shown in (B.128).

We are now ready to combine the derived expressions into a simplified MSE expression for R-D Standard Tensor-ESPRIT for two sources. The expression for $\left\|\boldsymbol{z}_{i,0}^{(r)}\right\|_2^2 = \left\|\tilde{\boldsymbol{s}}_{i,0}\right\|_2^2 \cdot \left\|\tilde{\boldsymbol{a}}_{i,0}^{(r)}\right\|_2^2 = \left\|\boldsymbol{r}_i^{(r)^{\mathrm{T}}} \cdot \boldsymbol{W}_0\right\|_2^2$ results in

$$\left\|\boldsymbol{z}_{i,0}^{(r)}\right\|_2^2 = \frac{\hat{P}_{\tilde{i}}}{N \cdot \hat{P}_1 \cdot \hat{P}_2 \cdot \left(1 - |\hat{\rho}|^2\right)} \cdot \frac{b^{(r)}}{D_{\text{sel}}^{(r)^2} \cdot D_r} \quad \text{(B.126)}$$

and the expression for $\left\|\boldsymbol{z}_{i,r}^{(r)}\right\|_2^2 = \|\tilde{\boldsymbol{s}}_{i,r}\|_2^2 \cdot \left\|\tilde{\boldsymbol{a}}_{i,r}^{(r)}\right\|_2^2 = \left\|\boldsymbol{r}_i^{(r)^{\mathrm{T}}} \cdot \boldsymbol{W}_r\right\|_2^2$ results in

$$\left\|\boldsymbol{z}_{i,r}^{(r)}\right\|_2^2 = \frac{1}{D_{\mathrm{sel}}^{(r)^2}} \cdot \left[c^{(r)} - \frac{1}{D_r} \cdot M_r \cdot (M_r - 1)^2 \cdot \left|\mathrm{e}^{\mathrm{j}\Delta\mu^{(r)}} - 1\right|^2 \cdot \left|\alpha_{\mathrm{sel}}^{(r)}\right|^2 \cdot \left[\left(\frac{M}{M_r}\right)^2 - \frac{|\alpha|^2}{\left|\alpha^{(r)}\right|^2}\right]^2\right] \cdot \frac{\frac{M}{M_r} \cdot \hat{P}_{\bar{i}}}{N \cdot \hat{P}_1 \cdot \hat{P}_2 \cdot \left[\left(\frac{M}{M_r}\right)^2 - |\hat{\rho}|^2 \cdot \left|\frac{\alpha}{\alpha^{(r)}}\right|^2\right]}. \tag{B.127}$$

We can see from (B.126) and (B.127) that the terms for $\left\|\boldsymbol{z}_{i,0}^{(r)}\right\|_2^2$ and $\left\|\boldsymbol{z}_{i,r}^{(r)}\right\|_2^2$ share the expression $M \cdot (M_r - 1)^2 \cdot \left|\mathrm{e}^{\mathrm{j}\Delta\mu^{(r)}} - 1\right|^2 \cdot \left|\alpha_{\mathrm{sel}}^{(r)}\right|^2 \cdot \left[\left(\frac{M}{M_r}\right)^2 - \frac{|\alpha|^2}{\left|\alpha^{(r)}\right|^2}\right] = b^{(r)}$. Factoring it out yields

$$\begin{aligned}
&\frac{1}{D_r} \cdot M \cdot (M_r - 1)^2 \cdot \left|\mathrm{e}^{\mathrm{j}\Delta\mu^{(r)}} - 1\right|^2 \cdot \left|\alpha_{\mathrm{sel}}^{(r)}\right|^2 \cdot \left[\left(\frac{M}{M_r}\right)^2 - \frac{|\alpha|^2}{\left|\alpha^{(r)}\right|^2}\right] \cdot \frac{1}{1 - |\hat{\rho}|^2} \\
&\quad - \frac{1}{D_r} \cdot M_r \cdot (M_r - 1)^2 \cdot \left|\mathrm{e}^{\mathrm{j}\Delta\mu^{(r)}} - 1\right|^2 \cdot \left|\alpha_{\mathrm{sel}}^{(r)}\right|^2 \cdot \left[\left(\frac{M}{M_r}\right)^2 - \frac{|\alpha|^2}{\left|\alpha^{(r)}\right|^2}\right]^2 \cdot \frac{\frac{M}{M_r}}{\left(\frac{M}{M_r}\right)^2 - |\hat{\rho}|^2 \cdot \frac{|\alpha|^2}{\left|\alpha^{(r)}\right|^2}} \\
&= \frac{b^{(r)}}{D_r} \cdot \left[\frac{1}{1 - |\hat{\rho}|^2} - \frac{\left(\frac{M}{M_r}\right)^2 - \frac{|\alpha|^2}{\left|\alpha^{(r)}\right|^2}}{\left(\frac{M}{M_r}\right)^2 - |\hat{\rho}|^2 \cdot \frac{|\alpha|^2}{\left|\alpha^{(r)}\right|^2}}\right].
\end{aligned} \tag{B.128}$$

Making use of this factorization, we combine (B.126) and (B.127) into $\left\|\boldsymbol{z}_i^{(r)}\right\|_2^2 = \left\|\boldsymbol{z}_{i,0}^{(r)}\right\|_2^2 + \left\|\boldsymbol{z}_{i,R}^{(r)}\right\|_2^2$ to obtain

$$\left\|\boldsymbol{z}_i^{(r)}\right\|_2^2 = \frac{\hat{P}_{\bar{i}}}{N \cdot \hat{P}_1 \cdot \hat{P}_2 \cdot D_{\mathrm{sel}}^{(r)^2}} \cdot \left(\frac{b^{(r)}}{D_r} \cdot \left[\frac{1}{1 - |\hat{\rho}|^2} - \frac{\left(\frac{M}{M_r}\right)^2 - \frac{|\alpha|^2}{\left|\alpha^{(r)}\right|^2}}{\left(\frac{M}{M_r}\right)^2 - |\hat{\rho}|^2 \cdot \frac{|\alpha|^2}{\left|\alpha^{(r)}\right|^2}}\right] + c^{(r)} \cdot \frac{\frac{M}{M_r}}{\left(\frac{M}{M_r}\right)^2 - |\hat{\rho}|^2 \cdot \frac{|\alpha|^2}{\left|\alpha^{(r)}\right|^2}}\right). \tag{B.129}$$

Finally, combining these results, we have

$$\mathrm{MSE}_{\mathrm{ten},i}^{(r)} = \frac{\sigma_{\mathrm{n}}^2}{2} \cdot \frac{\hat{P}_{\bar{i}}}{N \cdot \hat{P}_1 \cdot \hat{P}_2} \cdot a_{\mathrm{ten}}^{(r)}, \tag{B.130}$$

where

$$a_{\mathrm{ten}}^{(r)} = \frac{1}{D_{\mathrm{sel}}^{(r)^2}} \cdot \left(\frac{b^{(r)}}{D_r} \cdot \left[\frac{1}{1 - |\hat{\rho}|^2} - \frac{\left(\frac{M}{M_r}\right)^2 - \frac{|\alpha|^2}{\left|\alpha^{(r)}\right|^2}}{\left(\frac{M}{M_r}\right)^2 - |\hat{\rho}|^2 \cdot \left|\frac{\alpha}{\alpha^{(r)}}\right|^2}\right] + c^{(r)} \cdot \frac{\frac{M}{M_r}}{\left(\frac{M}{M_r}\right)^2 - |\hat{\rho}|^2 \cdot \frac{|\alpha|^2}{\left|\alpha^{(r)}\right|^2}}\right).$$

The MSE for both sources in the r-th mode is obtained as $\mathrm{MSE}_{\mathrm{ten}}^{(r)} = \sum_{i=1}^{2} \mathrm{MSE}_{\mathrm{ten},i}^{(r)}$. Finally, we compute the sum over the estimation errors in the individual modes to obtain the total MSE for R-D Standard Tensor-ESPRIT as

$$\mathrm{MSE}_{\mathrm{ten}} = \sum_{r=1}^{R} \frac{\sigma_{\mathrm{n}}^2}{2} \cdot \frac{\hat{P}_i + \hat{P}_{\bar{i}}}{N \cdot \hat{P}_1 \cdot \hat{P}_2} \cdot a_{\mathrm{ten}}^{(r)}. \tag{B.131}$$

This completes the sketch of the proof.

B.8. Proof of Theorem 4.5.4

In this section, we provide sketch of the proof of Theorem 4.5.4 from Section 4.5.4. For the sake of brevity, we have left out some of the lengthy derivations, however, the full proof is provided by us in [Gra16]. Let us first recall the MSE expression of R-D Unitary Tensor-ESPRIT assuming circularly symmetric white noise in (4.41). The MSE for the k-th spatial frequency in the r-th mode can be expressed as

$$\mathbb{E}\left\{(\Delta\mu_i)^2\right\} = \frac{\sigma_{\mathrm{n}}^2}{2} \cdot \left(\left\| \boldsymbol{z}_i^{(\mathrm{fba})(r)} \right\|_2^2 - \mathrm{Re}\left\{ \boldsymbol{z}_i^{(\mathrm{fba})(r)^{\mathrm{T}}} \cdot \boldsymbol{\Pi}_{2\cdot M\cdot N} \cdot \boldsymbol{z}_i^{(\mathrm{fba})(r)} \right\} \right). \tag{B.132}$$

where $\boldsymbol{z}_i^{(\mathrm{fba})(r)^{\mathrm{T}}} = \boldsymbol{r}_i^{(\mathrm{fba})(r)^{\mathrm{T}}} \cdot \boldsymbol{W}_{\mathrm{ten}}^{(\mathrm{fba})}$. The vector $\boldsymbol{r}_i^{(\mathrm{fba})(r)}$ is given by

$$\boldsymbol{r}_i^{(\mathrm{fba})(r)} = \boldsymbol{q}_i^{(r)} \otimes \left(\left[\left(\tilde{\boldsymbol{J}}_1^{(r)} \cdot \boldsymbol{U}_{\mathrm{s}}^{(\mathrm{fba})} \right)^{+} \left(\tilde{\boldsymbol{J}}_2^{(r)} / \mathrm{e}^{\mathrm{j}\cdot\mu_i^{(r)}} - \tilde{\boldsymbol{J}}_1^{(r)} \right) \right]^{\mathrm{T}} \cdot \boldsymbol{p}_i^{(r)} \right). \tag{B.133}$$

and the matrix $\boldsymbol{W}_{\mathrm{ten}}^{(\mathrm{fba})}$ is given by

$$\boldsymbol{W}_{\mathrm{ten}}^{(\mathrm{fba})} = \boldsymbol{W}_0^{(\mathrm{fba})} + \sum_{q=1}^{R} \boldsymbol{W}_q^{(\mathrm{fba})} \cdot \boldsymbol{P}_{M_1,\ldots,M_R,2N}^{(q)^{\mathrm{T}}} \cdot \boldsymbol{P}_{M_1,\ldots,M_R,2N}^{(R)} \tag{B.134}$$

where

$$\boldsymbol{W}_0^{(\mathrm{fba})} = \left(\boldsymbol{\Sigma}_{\mathrm{s}}^{(\mathrm{fba})^{-1}} \cdot \boldsymbol{V}_{\mathrm{s}}^{(\mathrm{fba})^{\mathrm{T}}} \right) \otimes \left(\boldsymbol{T}_{1:R}^{(\mathrm{fba})\otimes} \cdot \boldsymbol{U}_{\mathrm{n}}^{(\mathrm{fba})} \cdot \boldsymbol{U}_{\mathrm{n}}^{(\mathrm{fba})^{\mathrm{H}}} \right) \tag{B.135}$$

$$\begin{aligned} \boldsymbol{W}_q^{(\mathrm{fba})} = &\left(\boldsymbol{U}_{\mathrm{s}}^{(\mathrm{fba})^{\mathrm{T}}} \otimes \boldsymbol{I}_M \right) \cdot \left(\bar{\boldsymbol{T}}_{1:q-1}^{(\mathrm{fba})} \otimes \boldsymbol{I}_{M_{q:R}} \right) \cdot \left(\boldsymbol{I}_{M_q} \otimes \bar{\boldsymbol{T}}_{q+1:R}^{(\mathrm{fba})} \right) \\ &\cdot \left[\left(\boldsymbol{V}_q^{[\mathrm{s}](\mathrm{fba})} \cdot \boldsymbol{\Sigma}_q^{[\mathrm{s}](\mathrm{fba})^{-1}} \cdot \boldsymbol{U}_q^{[\mathrm{s}](\mathrm{fba})^{\mathrm{H}}} \right)^{\mathrm{T}} \otimes \left(\boldsymbol{U}_q^{[\mathrm{n}](\mathrm{fba})} \cdot \boldsymbol{U}_q^{[\mathrm{n}](\mathrm{fba})^{\mathrm{H}}} \right) \right] \end{aligned} \tag{B.136}$$

and

$$\bar{\boldsymbol{T}}_{1:q-1}^{(\mathrm{fba})} = \begin{bmatrix} \boldsymbol{I}_{M_{q:R}} \otimes \boldsymbol{t}_{1:q-1,1}^{(\mathrm{fba})} \\ \vdots \\ \boldsymbol{I}_{M_{q:R}} \otimes \boldsymbol{t}_{1:q-1,M_{1:q-1}}^{(\mathrm{fba})} \end{bmatrix}, \quad \bar{\boldsymbol{T}}_{q+1:R}^{(\mathrm{fba})} = \begin{bmatrix} \boldsymbol{I}_{M_r} \otimes \boldsymbol{t}_{q+1:R,1}^{(\mathrm{fba})} \\ \vdots \\ \boldsymbol{I}_{M_r} \otimes \boldsymbol{t}_{q+1:R,M_{q+1:R}}^{(\mathrm{fba})} \end{bmatrix} \tag{B.137}$$

where $\boldsymbol{t}_{a:b,n}^{(\mathrm{fba})}$ denotes the n-th column of $\boldsymbol{T}_q^{(\mathrm{fba})} = \boldsymbol{U}_q^{[\mathrm{s}](\mathrm{fba})} \cdot \boldsymbol{U}_q^{[\mathrm{s}](\mathrm{fba})^{\mathrm{H}}}$.

In order to simplify (B.132), we first insert $\boldsymbol{W}_{\mathrm{ten}}^{(\mathrm{fba})}$ from (B.134) into the vector $\boldsymbol{z}_i^{(\mathrm{fba})(r)^{\mathrm{T}}} = \boldsymbol{r}_i^{(\mathrm{fba})(r)^{\mathrm{T}}} \cdot \boldsymbol{W}_{\mathrm{ten}}^{(\mathrm{fba})}$, which yields

$$\begin{aligned} \boldsymbol{z}_i^{(\mathrm{fba})(r)^{\mathrm{T}}} &= \boldsymbol{r}_i^{(\mathrm{fba})(r)^{\mathrm{T}}} \cdot \boldsymbol{W}_0^{(\mathrm{fba})} + \boldsymbol{r}_i^{(\mathrm{fba})(r)^{\mathrm{T}}} \cdot \sum_{q=1}^{R} \boldsymbol{W}_q^{(\mathrm{fba})} \cdot \boldsymbol{P}_{M_1,\ldots,M_R,2\cdot N}^{(q)^{\mathrm{T}}} \cdot \boldsymbol{P}_{M_1,\ldots,M_R,2\cdot N}^{(R)} \\ &:= \boldsymbol{z}_{i,0}^{(\mathrm{fba})(r)^{\mathrm{T}}} + \boldsymbol{z}_{i,R}^{(\mathrm{fba})(r)^{\mathrm{T}}}. \end{aligned} \tag{B.138}$$

Similarly to the proof in Appendix B.7 for R-D Standard Tensor-ESPRIT, it can be shown that the terms $\boldsymbol{z}_{i,0}^{(\mathrm{fba})(r)^{\mathrm{T}}}$ and $\boldsymbol{z}_{i,R}^{(\mathrm{fba})(r)^{\mathrm{T}}}$ are orthogonal and we can compute $\left\|\boldsymbol{z}_i^{(\mathrm{fba})(r)}\right\|_2^2 = \left\|\boldsymbol{z}_{i,0}^{(\mathrm{fba})(r)}\right\|_2^2 + \left\|\boldsymbol{z}_{i,R}^{(\mathrm{fba})(r)}\right\|_2^2$. Then, it can be shown that

$$\boldsymbol{z}_{i,0}^{(\mathrm{fba})(r)^*} = -\boldsymbol{\Pi}_{2\cdot M\cdot N} \cdot \boldsymbol{z}_{i,0}^{(\mathrm{fba})(r)} \tag{B.139}$$

$$\boldsymbol{z}_{i,R}^{(\mathrm{fba})(r)^*} = -\boldsymbol{\Pi}_{2\cdot M\cdot N} \cdot \boldsymbol{z}_{i,R}^{(\mathrm{fba})(r)} \tag{B.140}$$

such that

$$\boldsymbol{z}_i^{(\mathrm{fba})(r)*} = -\boldsymbol{\Pi}_{2\cdot M\cdot N} \cdot \boldsymbol{z}_i^{(\mathrm{fba})(r)}. \tag{B.141}$$

As a result, we obtain

$$\boldsymbol{z}_i^{(\mathrm{fba})(r)^{\mathrm{T}}} \cdot \boldsymbol{\Pi}_{2\cdot M\cdot N} \cdot \boldsymbol{z}_i^{(\mathrm{fba})(r)} = -\left\|\boldsymbol{z}_i^{(\mathrm{fba})(r)}\right\|_2^2. \tag{B.142}$$

Inserting (B.142) into (B.132), the MSE for R-D Unitary Tensor-ESPRIT can be simplified as

$$\mathrm{MSE}_i^{(\mathrm{fba})(r)} = \mathbb{E}\left\{(\Delta\mu_i)^2\right\} = \sigma_{\mathrm{n}}^2 \cdot \left\|\boldsymbol{z}_i^{(\mathrm{fba})(r)}\right\|_2^2. \tag{B.143}$$

Then, we follow the lines of the derivation for R-D Standard Tensor-ESPRIT to find $\left\|\boldsymbol{z}_{i,0}^{(\mathrm{fba})(r)}\right\|_2^2$. As a first step, we write $\boldsymbol{z}_{i,0}^{(\mathrm{fba})(r)^{\mathrm{T}}} = \boldsymbol{r}_i^{(\mathrm{fba})(r)^{\mathrm{T}}} \cdot \boldsymbol{W}_0^{(\mathrm{fba})}$ into the Kronecker product of an array and a signal part according to

$$\boldsymbol{z}_{i,0}^{(\mathrm{fba})(r)^{\mathrm{T}}} = \tilde{\boldsymbol{s}}_{i,0}^{(\mathrm{fba})^{\mathrm{T}}} \otimes \tilde{\boldsymbol{a}}_{i,0}^{(\mathrm{fba})(r)^{\mathrm{T}}}, \tag{B.144}$$

where

$$\tilde{\boldsymbol{s}}_{i,0}^{(\mathrm{fba})^{\mathrm{T}}} = \boldsymbol{e}_i^{\mathrm{T}} \cdot \overline{\boldsymbol{S}}^{+\mathrm{T}} \tag{B.145}$$

$$\tilde{\boldsymbol{a}}_{i,0}^{(\mathrm{fba})(r)^{\mathrm{T}}} = \boldsymbol{e}_i^{\mathrm{T}} \cdot \left(\tilde{\boldsymbol{J}}_1^{(r)} \cdot \boldsymbol{A}_{\mathrm{c}}\right)^{+} \cdot \left(\tilde{\boldsymbol{J}}_2^{(r)} / \lambda_i^{(r)} - \tilde{\boldsymbol{J}}_1^{(r)}\right) \cdot \boldsymbol{T}_{1:R}^{\otimes} \cdot \boldsymbol{P}_{\boldsymbol{A}_{\mathrm{c}}^{\perp}} \tag{B.146}$$

and $\overline{\boldsymbol{S}} = \left[\begin{array}{cc} \tilde{\boldsymbol{S}}, & \tilde{\boldsymbol{S}}^* \cdot \boldsymbol{\Pi}_N \end{array}\right] \in \mathbb{C}^{d \times 2N}$ with $\tilde{\boldsymbol{S}} = \boldsymbol{\Delta} \cdot \boldsymbol{S}$ is the matrix of forward-backward averaged source symbols. Note that the array part $\tilde{\boldsymbol{a}}_{i,0}^{(\mathrm{fba})(r)}$ in (B.146) is equal to the result obtained for R-D Standard Tensor-ESPRIT in (B.113), except for the matrix $\boldsymbol{\Delta}$ for the array phase reference, which has been moved towards the source symbols $\tilde{\boldsymbol{S}}$. Thus, we have

$$\tilde{\boldsymbol{a}}_{i,r}^{(\mathrm{fba})(r)} = \tilde{\boldsymbol{a}}_{i,r}^{(r)}\Big|_{\boldsymbol{A}=\boldsymbol{A}_{\mathrm{c}}}. \tag{B.147}$$

Using property (1.7), we take the squared norm $\left\|\boldsymbol{z}_{i,0}^{(r)}\right\|_2^2$ individually for $\tilde{\boldsymbol{s}}_{i,0}^{(\mathrm{fba})^{\mathrm{T}}}$ and $\tilde{\boldsymbol{a}}_{i,0}^{(\mathrm{fba})(r)^{\mathrm{T}}}$

$$\left\|\boldsymbol{z}_{i,0}^{(\mathrm{fba})(r)}\right\|_2^2 = \left\|\tilde{\boldsymbol{s}}_{i,0}^{(\mathrm{fba})}\right\|_2^2 \cdot \left\|\tilde{\boldsymbol{a}}_{i,0}^{(\mathrm{fba})(r)}\right\|_2^2. \tag{B.148}$$

In contrast to $\|\tilde{\boldsymbol{s}}_{i,0}\|_2^2$ from Appendix B.7, we find that the array phase reference and the correlation phase for the source symbols does not cancel for $\left\|\tilde{\boldsymbol{s}}_{i,0}^{(\mathrm{fba})}\right\|_2^2$. It can be shown that $\left\|\tilde{\boldsymbol{s}}_{i,0}^{(\mathrm{fba})}\right\|_2^2$ is given by

$$\left\|\tilde{\boldsymbol{s}}_{i,0}^{(\mathrm{fba})}\right\|_2^2 = \frac{1}{2 \cdot N} \cdot \frac{\hat{P}_i}{\hat{P}_1 \cdot \hat{P}_2 \cdot \left(1 - |\hat{\rho}|^2 \cdot \cos^2\left(\Delta\varphi\right)\right)}, \tag{B.149}$$

where

$$\Delta\varphi = \Delta\varphi_{\mathrm{ref}} + \varphi_{\mathrm{corr}} \tag{B.150}$$

combines the phase terms for the empirical source correlation $\hat{\rho} = \mathrm{e}^{\mathrm{j}\cdot\varphi_{\mathrm{corr}}} \cdot |\hat{\rho}|$, and the array phase separation $\Delta\varphi_{\mathrm{ref}} = \varphi_{\mathrm{ref},2} - \varphi_{\mathrm{ref},1} = \sum_{r=1}^{R} \delta^{(r)} \cdot \Delta\mu^{(r)}$ Hence, the array phase reference and the source correlation phase enter the MSE expression via the term $\cos^2\left(\Delta\varphi\right)$ and determines the decorrelation capability of the FBA preprocessing (cf. Section 4.5.5).

Next, we determine a simplified expression for $\left\|\tilde{\boldsymbol{a}}_{i,0}^{(\mathrm{fba})(r)}\right\|_2^2$. However, this is a straightforward task, since we have $\tilde{\boldsymbol{a}}_{i,0}^{(\mathrm{fba})(r)} = \tilde{\boldsymbol{a}}_{i,0}^{(r)}\Big|_{\boldsymbol{A}=\boldsymbol{A}_{\mathrm{c}}}$ as concluded in (B.147). Therefore, we obtain the same result for $\left\|\tilde{\boldsymbol{a}}_{i,0}^{(\mathrm{fba})(r)}\right\|_2^2$ as in Appendix B.7 for R-D Standard Tensor-ESPRIT, which is given by

$$\left\|\tilde{\boldsymbol{a}}_{i,0}^{(\mathrm{fba})(r)}\right\|_2^2 = \frac{M \cdot (M_r - 1)^2 \cdot \left|\mathrm{e}^{\mathrm{j}\Delta\mu^{(r)}} - 1\right|^2 \cdot \left|\alpha_{\mathrm{sel}}^{(r)}\right|^2 \cdot \left[\left(\frac{M}{M_r}\right)^2 - \frac{|\alpha|^2}{|\alpha^{(r)}|^2}\right]}{D_{\mathrm{sel}}^{(r)^2} \cdot D_r} \tag{B.151}$$

In the next step, we consider the term $\boldsymbol{z}_{i,R}^{(\mathrm{fba})(r)}$. Similarly to (B.116), we find that $\boldsymbol{z}_{i,R}^{(\mathrm{fba})(r)^{\mathrm{T}}} = \boldsymbol{r}_i^{(\mathrm{fba})(r)^{\mathrm{T}}} \cdot \sum_{q=1}^{R} \boldsymbol{W}_q^{(\mathrm{fba})} \cdot \boldsymbol{P}_{M_1,\ldots,M_R,2\cdot N}^{(q)^{\mathrm{T}}} \cdot \boldsymbol{P}_{M_1,\ldots,M_R,2\cdot N}^{(R)}$ simplifies to

$$\boldsymbol{z}_{i,r}^{(\mathrm{fba})(r)^{\mathrm{T}}} \cdot \boldsymbol{P}_{M_1,\ldots,M_R,2\cdot N}^{(r)^{\mathrm{T}}} \cdot \boldsymbol{P}_{M_1,\ldots,M_R,2\cdot N}^{(R)} = \boldsymbol{z}_{i,R}^{(\mathrm{fba})(r)^{\mathrm{T}}}, \tag{B.152}$$

where $\boldsymbol{z}_{i,r}^{(r)(\mathrm{fba})^{\mathrm{T}}} = \boldsymbol{r}_i^{(\mathrm{fba})(r)^{\mathrm{T}}} \cdot \boldsymbol{W}_r^{(\mathrm{fba})}$ can be expressed as the Kronecker product

$$\boldsymbol{z}_{i,r}^{(\mathrm{fba})(r)^{\mathrm{T}}} = \tilde{\boldsymbol{s}}_{i,r}^{(\mathrm{fba})^{\mathrm{T}}} \otimes \tilde{\boldsymbol{a}}_{i,r}^{(\mathrm{fba})(r)^{\mathrm{T}}} \tag{B.153}$$

with

$$\tilde{\boldsymbol{s}}_{i,r}^{(\mathrm{fba})(r)^{\mathrm{T}}} = \boldsymbol{e}_i^{\mathrm{T}} \cdot \left[\boldsymbol{A}_{\mathrm{c},r+1:R}^{\diamond} \diamond \overline{\boldsymbol{S}}^{\mathrm{T}} \diamond \boldsymbol{A}_{\mathrm{c},1:r-1}^{\diamond} \right]^{+} \tag{B.154}$$

$$\tilde{\boldsymbol{a}}_{i,r}^{(\mathrm{fba})(r)^{\mathrm{T}}} = \boldsymbol{e}_i^{\mathrm{T}} \cdot \left(\tilde{\boldsymbol{J}}_1^{(r)} \cdot \boldsymbol{A}_{\mathrm{c}} \right)^{+} \cdot \left(\tilde{\boldsymbol{J}}_2^{(r)} / \lambda_i^{(r)} - \tilde{\boldsymbol{J}}_1^{(r)} \right) \cdot \left(\boldsymbol{a}_{\mathrm{c},i}^{\otimes(1:r-1)} \otimes \boldsymbol{P}_{\boldsymbol{A}^{(r)}}^{\perp} \otimes \boldsymbol{a}_{\mathrm{c},i}^{\otimes(r+1:R)} \right). \tag{B.155}$$

Using property (1.7), we again have

$$\left\| \boldsymbol{z}_{i,r}^{(\mathrm{fba})(r)} \right\|_2^2 = \left\| \tilde{\boldsymbol{s}}_{i,r}^{(\mathrm{fba})(r)} \right\|_2^2 \cdot \left\| \tilde{\boldsymbol{a}}_{i,r}^{(\mathrm{fba})(r)} \right\|_2^2. \tag{B.156}$$

Assuming maximum subarray overlap, the signal part $\left\| \tilde{\boldsymbol{s}}_{i,r}^{(\mathrm{fba})(r)} \right\|_2^2$ can be computed as

$$\left\| \tilde{\boldsymbol{s}}_{i,r}^{(\mathrm{fba})(r)} \right\|_2^2 = \frac{\frac{M}{M_r} \cdot \hat{P}_i}{2 \cdot N \cdot \hat{P}_1 \cdot \hat{P}_2 \cdot \left[\left(\frac{M}{M_r} \right)^2 - \cos^2(\Delta\varphi) \cdot |\hat{\rho}|^2 \cdot \frac{|\alpha|^2}{|\alpha^{(r)}|^2} \right]}. \tag{B.157}$$

Similar to the considerations for the array part $\tilde{\boldsymbol{a}}_{i,0}^{(\mathrm{fba})(r)}$, the array part $\tilde{\boldsymbol{a}}_{i,r}^{(\mathrm{fba})(r)}$ is obtained as $\tilde{\boldsymbol{a}}_{i,r}^{(\mathrm{fba})(r)} = \tilde{\boldsymbol{a}}_{i,r}^{(r)}\big|_{\boldsymbol{A}=\boldsymbol{A}_{\mathrm{c}}}$, and, therefore, is equal to the result devised for R-D Standard Tensor-ESPRIT given by

$$\left\| \tilde{\boldsymbol{a}}_{i,r}^{(\mathrm{fba})(r)} \right\|_2^2 = \frac{1}{D_{\mathrm{sel}}^{(r)^2}} \cdot \left[c^{(r)} - \frac{1}{D_r} \cdot M_r \cdot (M_r - 1)^2 \cdot \left| \mathrm{e}^{\mathrm{j}\Delta\mu^{(r)}} - 1 \right|^2 \cdot \left| \alpha_{\mathrm{sel}}^{(r)} \right|^2 \cdot \left[\left(\frac{M}{M_r} \right)^2 - \frac{|\alpha|^2}{\left|\alpha^{(r)}\right|^2} \right]^2 \right], \tag{B.158}$$

where

$$\begin{aligned} c^{(r)} = {} & \left(\frac{M}{M_r} \right)^4 \cdot (M_r - 1)^2 \cdot 2 + \left| \frac{\alpha}{\alpha^{(r)}} \right|^4 \cdot \left| \alpha_{\mathrm{sel}}^{(r)} \right|^2 \cdot \left(2 + (M_r - 2) \cdot \left| \mathrm{e}^{\mathrm{j}\Delta\mu^{(r)}} - 1 \right|^2 \right) \\ & - 4 \cdot \left(\frac{M}{M_r} \right)^2 \cdot (M_r - 1) \cdot \left| \frac{\alpha}{\alpha^{(r)}} \right|^2 \cdot \mathrm{Re}\left\{ \alpha_{\mathrm{sel},0}^{(r)} \right\}. \end{aligned} \tag{B.159}$$

Finally, we are ready to combine the derived expressions into a simplified MSE expression for R-D Unitary Tensor-ESPRIT for two sources. The expression for $\left\|\boldsymbol{z}_{i,0}^{(\mathrm{fba})(r)}\right\|_2^2 = \left\|\tilde{\boldsymbol{s}}_{i,0}^{(\mathrm{fba})}\right\|_2^2 \cdot \left\|\tilde{\boldsymbol{a}}_{i,0}^{(\mathrm{fba})(r)}\right\|_2^2$ results in

$$\left\|\boldsymbol{z}_{i,0}^{(\mathrm{fba})(r)}\right\|_2^2 = \frac{1}{2\cdot N}\cdot\frac{\hat{P}_i}{\hat{P}_1\cdot\hat{P}_2\cdot\left(1-|\hat{\rho}|^2\cdot\cos^2\left(\Delta\varphi\right)\right)}\cdot\frac{b^{(r)}}{D_{\mathrm{sel}}^{(r)^2}\cdot D_r}. \tag{B.160}$$

Moreover, the expression for $\left\|\boldsymbol{z}_{i,r}^{(\mathrm{fba})(r)}\right\|_2^2 = \left\|\tilde{\boldsymbol{s}}_{i,r}^{(\mathrm{fba})}\right\|_2^2 \cdot \left\|\tilde{\boldsymbol{a}}_{i,r}^{(\mathrm{fba})(r)}\right\|_2^2$ results in

$$\begin{aligned}\left\|\boldsymbol{z}_{i,r}^{(\mathrm{fba})(r)}\right\|_2^2 = \frac{1}{D_{\mathrm{sel}}^{(r)^2}}\cdot\left[c^{(r)}-\frac{1}{D_r}\cdot b^{(r)}\cdot\left[\left(\frac{M}{M_r}\right)^2-\frac{|\alpha|^2}{\left|\alpha^{(r)}\right|^2}\right]^2\right]\cdot\frac{M_r}{M}\\ \cdot\frac{\frac{M}{M_r}\cdot\hat{P}_i}{2\cdot N\cdot\hat{P}_1\cdot\hat{P}_2\cdot\left[\left(\frac{M}{M_r}\right)^2-\cos^2\left(\Delta\varphi\right)\cdot|\hat{\rho}|^2\cdot\frac{|\alpha|^2}{|\alpha^{(r)}|^2}\right]}\end{aligned} \tag{B.161}$$

Finally, we obtain $\left\|\boldsymbol{z}_{i}^{(\mathrm{fba})(r)}\right\|_2^2$ as $\left\|\boldsymbol{z}_{i}^{(\mathrm{fba})(r)}\right\|_2^2 = \left\|\boldsymbol{z}_{i,0}^{(\mathrm{fba})(r)}\right\|_2^2 + \left\|\boldsymbol{z}_{i,r}^{(\mathrm{fba})(r)}\right\|_2^2$, resulting in

$$\begin{aligned}\left\|\boldsymbol{z}_{i}^{(\mathrm{fba})(r)}\right\|_2^2 = \frac{\hat{P}_i}{2\cdot N\cdot\hat{P}_1\cdot\hat{P}_2\cdot D_{\mathrm{sel}}^{(r)^2}}\cdot\Bigg\{\frac{1}{D_r}\cdot M\cdot(M_r-1)^2\cdot\left|\mathrm{e}^{\mathrm{j}\Delta\mu^{(r)}}-1\right|^2\cdot\left|\alpha_{\mathrm{sel}}^{(r)}\right|^2\cdot\left[\left(\frac{M}{M_r}\right)^2-\frac{|\alpha|^2}{\left|\alpha^{(r)}\right|^2}\right]\\ \cdot\left[\frac{1}{1-\cos^2\left(\Delta\varphi\right)\cdot|\hat{\rho}|^2}-\frac{\left(\frac{M}{M_r}\right)^2-\frac{|\alpha|^2}{\left|\alpha^{(r)}\right|^2}}{\left(\frac{M}{M_r}\right)^2-\cos^2\left(\Delta\varphi\right)\cdot|\hat{\rho}|^2\cdot\left|\frac{\alpha}{\alpha^{(r)}}\right|^2}\right]\\ +\Bigg[\left(\frac{M}{M_r}\right)^4\cdot(M_r-1)^2\cdot 2+\left|\frac{\alpha}{\alpha^{(r)}}\right|^4\cdot\left|\alpha_{\mathrm{sel}}^{(r)}\right|^2\cdot\left(2+(M_r-2)\cdot\left|\mathrm{e}^{\mathrm{j}\Delta\mu^{(r)}}-1\right|^2\right)\\ -4\cdot\left(\frac{M}{M_r}\right)^2\cdot(M_r-1)\cdot\frac{|\alpha|^2}{\left|\alpha^{(r)}\right|^2}\cdot\mathrm{Re}\left\{\alpha_{\mathrm{sel},0}^{(r)}\right\}\Bigg]\cdot\frac{\frac{M}{M_r}}{\left(\frac{M}{M_r}\right)^2-\cos^2\left(\Delta\varphi\right)\cdot|\hat{\rho}|^2\cdot\frac{|\alpha|^2}{\left|\alpha^{(r)}\right|^2}}\Bigg\}.\end{aligned} \tag{B.162}$$

Consequently, the total MSE over the sources $k = 1, 2$ and the dimensions $r = 1, \ldots, R$ of R-D Unitary Tensor-ESPRIT for two sources is given by

$$\mathrm{MSE}_{\mathrm{ten}}^{(\mathrm{fba})} = \sum_{r=1}^{R}\frac{\sigma_{\mathrm{n}}^2}{2}\cdot\frac{\hat{P}_i+\hat{P}_{\bar{i}}}{N\cdot\hat{P}_1\cdot\hat{P}_2}\cdot a_{\mathrm{ten}}^{(\mathrm{fba})(r)} \tag{B.163}$$

where $a_{\mathrm{ten}}^{(\mathrm{fba})(r)}$ is defined in (4.67). This completes the sketch of the proof.

B.9. Proof of Theorem 5.3.1

For the proof of Theorem 5.3.1 in Section 5.3.1.1, we replace $\boldsymbol{R}$ in the GLS expression in (5.18) by its regularized version $\bar{\boldsymbol{R}}$ from (5.26) and show that the resulting solution $\hat{\boldsymbol{\psi}}_{\mathrm{GLS}}$ is independent of λ. We start by expressing $\bar{\boldsymbol{R}}$ in (5.26) as

$$\bar{\boldsymbol{R}} = \boldsymbol{R} + \lambda \cdot \boldsymbol{R}_0, \tag{B.164}$$

where $\boldsymbol{R}$ and $\boldsymbol{R}_0$ are defined as

$$\boldsymbol{R} = \sigma_{\mathrm{n}}^2 \cdot \boldsymbol{F}_2 \cdot \left(\boldsymbol{\Sigma}_{\mathrm{s}}^{-2} \otimes \boldsymbol{U}_{\mathrm{n}} \cdot \boldsymbol{U}_{\mathrm{n}}^{\mathrm{H}}\right) \cdot \boldsymbol{F}_2^{\mathrm{H}} \tag{B.165}$$

$$\boldsymbol{R}_0 = \sigma_{\mathrm{n}}^2 \cdot \boldsymbol{F}_2 \cdot \left(\boldsymbol{\Sigma}_{\mathrm{s}}^{-2} \otimes \boldsymbol{I}_M\right) \cdot \boldsymbol{F}_2^{\mathrm{H}}, \tag{B.166}$$

where $\boldsymbol{F}_2$ is defined in (5.16) as $\boldsymbol{F}_2 = \left(\boldsymbol{\Psi}^{\mathrm{T}} \otimes \boldsymbol{J}_1\right) - (\boldsymbol{I}_d \otimes \boldsymbol{J}_2)$.

To compute $\bar{\boldsymbol{R}}^{-1}$, we first write $\bar{\boldsymbol{R}}^{-1}$ for $\lambda > 0$ as

$$\bar{\boldsymbol{R}}^{-1} = (\lambda \cdot \boldsymbol{R}_0 + \boldsymbol{R})^{-1} = \frac{1}{\lambda} \cdot \boldsymbol{R}_0^{-1} \cdot \left(\boldsymbol{I}_{(M-1)d} + \frac{1}{\lambda} \cdot \boldsymbol{R} \cdot \boldsymbol{R}_0^{-1}\right)^{-1}. \tag{B.167}$$

In order to proceed, we require the following two propositions:

Proposition B.9.1. *For any projection matrix $\boldsymbol{P} \in \mathbb{C}^{M \times M}$ and for any regularization parameter $\lambda \neq -1$, the following relation holds:*

$$(\boldsymbol{I}_M + \lambda \cdot \boldsymbol{P})^{-1} = \boldsymbol{I}_M - \frac{\lambda}{\lambda + 1} \cdot \boldsymbol{P}. \tag{B.168}$$

Proof: First, recall that any projection matrix $\boldsymbol{P} \in \mathbb{C}^{M \times M}$ satisfies the property

$$\boldsymbol{P}^2 = \boldsymbol{P}. \tag{B.169}$$

Define $r = \operatorname{rank}\{\boldsymbol{P}\}$ and let $\boldsymbol{P}$ have the eigendecomposition $\boldsymbol{P} = \boldsymbol{\Gamma} \cdot \boldsymbol{\Lambda} \cdot \boldsymbol{\Gamma}^{-1}$. From (B.169), it follows that $\boldsymbol{\Lambda} = \operatorname{blkdiag}\{\boldsymbol{I}_r,\ \boldsymbol{0}_{M-r}\}$, i.e., the eigenvalues of $\boldsymbol{P}$ are either 0 or 1. Then, we have

$$\begin{aligned}
(\boldsymbol{I}_M + \lambda \cdot \boldsymbol{P})^{-1} &= \left(\boldsymbol{I}_M + \lambda \cdot \boldsymbol{\Gamma} \cdot \boldsymbol{\Lambda} \cdot \boldsymbol{\Gamma}^{-1}\right)^{-1} = \boldsymbol{\Gamma} \cdot (\boldsymbol{I}_M + \lambda \cdot \boldsymbol{\Lambda})^{-1} \cdot \boldsymbol{\Gamma}^{-1} = \boldsymbol{\Gamma} \cdot \begin{bmatrix} (1+\lambda) \cdot \boldsymbol{I}_r & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{I}_{M-r} \end{bmatrix}^{-1} \cdot \boldsymbol{\Gamma}^{-1} \\
&= \boldsymbol{\Gamma} \cdot \begin{bmatrix} \frac{1}{1+\lambda} \cdot \boldsymbol{I}_r & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{I}_{M-r} \end{bmatrix} \cdot \boldsymbol{\Gamma}^{-1} = \frac{1}{1+\lambda} \cdot \boldsymbol{\Gamma} \cdot \begin{bmatrix} \boldsymbol{I}_r & \boldsymbol{0} \\ \boldsymbol{0} & (1+\lambda) \cdot \boldsymbol{I}_{M-r} \end{bmatrix} \cdot \boldsymbol{\Gamma}^{-1} \\
&= \frac{1}{1+\lambda} \cdot \boldsymbol{\Gamma} \cdot \left(\boldsymbol{I}_M + \lambda \cdot \begin{bmatrix} \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{I}_{M-r} \end{bmatrix}\right) \cdot \boldsymbol{\Gamma}^{-1} = \frac{1}{1+\lambda} \cdot \left(\boldsymbol{I}_M + \lambda \cdot \boldsymbol{\Gamma} \cdot (\boldsymbol{I}_M - \boldsymbol{\Lambda}) \cdot \boldsymbol{\Gamma}^{-1}\right)
\end{aligned}$$

$$= \frac{1}{1+\lambda} \cdot ((1+\lambda) \cdot \boldsymbol{I}_M - \lambda \cdot \boldsymbol{P}) = \boldsymbol{I}_M - \frac{\lambda}{1+\lambda} \cdot \boldsymbol{P}, \tag{B.170}$$

which is the result in (B.168). □

Proposition B.9.2. *Assuming a ULA and maximum subarray overlap, the matrix* $\boldsymbol{R} \cdot \boldsymbol{R}_0^{-1} \in \mathbb{C}^{(M-1)d \times (M-1)d}$ *is a projection matrix defining a non-orthogonal projection.*

Proof: See Appendix B.10.

Continuing from (B.167) and using (B.168), we have

$$\begin{aligned} \bar{\boldsymbol{R}}^{-1} &= \frac{1}{\lambda} \cdot \boldsymbol{R}_0^{-1} \cdot \left(\boldsymbol{I} - \frac{1}{\lambda+1} \cdot \boldsymbol{R} \cdot \boldsymbol{R}_0^{-1} \right) \\ &= \frac{1}{\lambda} \cdot \boldsymbol{R}_0^{-1} - \frac{1}{\lambda(\lambda+1)} \cdot \boldsymbol{R}_0^{-1} \cdot \boldsymbol{R} \cdot \boldsymbol{R}_0^{-1}. \end{aligned} \tag{B.171}$$

Next, we consider the term $(\boldsymbol{F}_1^{\mathrm{H}} \cdot \bar{\boldsymbol{R}}^{-1} \cdot \boldsymbol{F}_1)^{-1}$ in (5.18). Defining $\boldsymbol{G} = \boldsymbol{F}_1^{\mathrm{H}} \cdot \boldsymbol{R}_0^{-1} \cdot \boldsymbol{F}_1$ and inserting (B.171), we obtain

$$\begin{aligned} \left(\boldsymbol{F}_1^{\mathrm{H}} \cdot \bar{\boldsymbol{R}}^{-1} \cdot \boldsymbol{F}_1 \right)^{-1} &= \left(\frac{1}{\lambda} \cdot \left(\boldsymbol{G} - \frac{1}{\lambda+1} \cdot \boldsymbol{F}_1^{\mathrm{H}} \cdot \boldsymbol{R}_0^{-1} \cdot \boldsymbol{R} \cdot \boldsymbol{R}_0^{-1} \cdot \boldsymbol{F}_1 \right) \right)^{-1} \\ &= \lambda \cdot \boldsymbol{G}^{-1} \cdot \left(\boldsymbol{I} - \frac{1}{\lambda+1} \cdot \boldsymbol{F}_1^{\mathrm{H}} \cdot \boldsymbol{R}_0^{-1} \cdot \boldsymbol{R} \cdot \boldsymbol{R}_0^{-1} \cdot \boldsymbol{F}_1 \cdot \boldsymbol{G}^{-1} \right)^{-1} \\ &= \lambda \cdot \boldsymbol{G}^{-1} \cdot \left(\boldsymbol{I} + \bar{\lambda} \cdot \boldsymbol{F}_1^{\mathrm{H}} \cdot \boldsymbol{R}_0^{-1} \cdot \boldsymbol{R} \cdot \boldsymbol{R}_0^{-1} \cdot \boldsymbol{F}_1 \cdot \boldsymbol{G}^{-1} \right)^{-1}, \end{aligned} \tag{B.172}$$

where we have defined $\bar{\lambda} = -(\lambda+1)^{-1}$. Then, we require the following proposition:

Proposition B.9.3. *The matrix* $\boldsymbol{F}_1^{\mathrm{H}} \cdot \boldsymbol{R}_0^{-1} \cdot \boldsymbol{R} \cdot \boldsymbol{R}_0^{-1} \cdot \boldsymbol{F}_1 \cdot \boldsymbol{G}^{-1} \in \mathbb{C}^{(M-1)d \times (M-1)d}$ *is also a projection matrix that defines a non-orthogonal projection.*

Proof: See Appendix B.11.

Therefore, the inverse term in (B.172) can be simplified by again applying (B.168). After resubstituting the regularization parameter $\bar{\lambda}$, we arrive at

$$\begin{aligned} \left(\boldsymbol{F}_1^{\mathrm{H}} \cdot \bar{\boldsymbol{R}}^{-1} \cdot \boldsymbol{F}_1 \right)^{-1} &= \lambda \cdot \boldsymbol{G}^{-1} \cdot \left(\boldsymbol{I} + \frac{1}{\lambda} \cdot \boldsymbol{F}_1^{\mathrm{H}} \cdot \boldsymbol{R}_0^{-1} \cdot \boldsymbol{R} \cdot \boldsymbol{R}_0^{-1} \cdot \boldsymbol{F}_1 \cdot \boldsymbol{G}^{-1} \right) \\ &= \lambda \cdot \boldsymbol{G}^{-1} + \boldsymbol{G}^{-1} \cdot \boldsymbol{F}_1^{\mathrm{H}} \cdot \boldsymbol{R}_0^{-1} \cdot \boldsymbol{R} \cdot \boldsymbol{R}_0^{-1} \cdot \boldsymbol{F}_1 \cdot \boldsymbol{G}^{-1}. \end{aligned} \tag{B.173}$$

Finally, using the results (B.171) and (B.173), (5.18) can be expressed as

$$\begin{aligned} &\left(\boldsymbol{F}_1^{\mathrm{H}} \cdot \bar{\boldsymbol{R}}^{-1} \cdot \boldsymbol{F}_1 \right)^{-1} \cdot \boldsymbol{F}_1^{\mathrm{H}} \cdot \bar{\boldsymbol{R}}^{-1} \cdot \boldsymbol{b} \\ &= \left(\lambda \cdot \boldsymbol{G}^{-1} + \boldsymbol{G}^{-1} \cdot \boldsymbol{F}_1^{\mathrm{H}} \cdot \boldsymbol{R}_0^{-1} \cdot \boldsymbol{R} \cdot \boldsymbol{R}_0^{-1} \cdot \boldsymbol{F}_1 \cdot \boldsymbol{G}^{-1} \right) \cdot \left(\frac{1}{\lambda} \cdot \boldsymbol{F}_1^{\mathrm{H}} \cdot \boldsymbol{R}_0^{-1} - \frac{1}{\lambda(\lambda+1)} \cdot \boldsymbol{F}_1^{\mathrm{H}} \cdot \boldsymbol{R}_0^{-1} \cdot \boldsymbol{R} \cdot \boldsymbol{R}_0^{-1} \right) \cdot \boldsymbol{b} \end{aligned}$$

$$= \boldsymbol{G}^{-1} \cdot \boldsymbol{F}_1^{\mathrm{H}} \cdot \boldsymbol{R}_0^{-1} \cdot \boldsymbol{b} + \boldsymbol{G}^{-1} \cdot \boldsymbol{F}_1^{\mathrm{H}} \cdot \boldsymbol{R}_0^{-1} \cdot \boldsymbol{R} \cdot \boldsymbol{R}_0^{-1} \cdot \left(\boldsymbol{F}_1 \cdot \boldsymbol{G}^{-1} \cdot \boldsymbol{F}_1^{\mathrm{H}} \cdot \boldsymbol{R}_0^{-1} \cdot \left(\frac{1}{\lambda} \cdot \boldsymbol{I}_{(M-1)d} - \frac{1}{\lambda\,(\lambda+1)} \cdot \boldsymbol{R} \cdot \boldsymbol{R}_0^{-1} \right) - \frac{1}{\lambda+1} \cdot \boldsymbol{I}_{(M-1)d} \right) \cdot \boldsymbol{b}, \tag{B.174}$$

where $\boldsymbol{b} = \mathrm{vec}\,\{\boldsymbol{J}_2 \cdot \boldsymbol{U}_{\mathrm{s}}\}$. To proceed, we need the proposition:

Proposition B.9.4. *The matrix $\boldsymbol{F}_1 \cdot \boldsymbol{G}^{-1} \cdot \boldsymbol{F}_1^{\mathrm{H}} \cdot \boldsymbol{R}_0^{-1} \in \mathbb{C}^{(M-1)d \times (M-1)d}$ is also a projection matrix that defines a non-orthogonal projection.*

Proof: It is straightforward to verify that the matrix $\boldsymbol{F}_1 \cdot \boldsymbol{G}^{-1} \cdot \boldsymbol{F}_1^{\mathrm{H}} \cdot \boldsymbol{R}_0^{-1}$ satisfies (B.169) as

$$\boldsymbol{F}_1 \cdot \boldsymbol{G}^{-1} \cdot \boldsymbol{F}_1^{\mathrm{H}} \cdot \boldsymbol{R}_0^{-1} \cdot \boldsymbol{F}_1 \cdot \boldsymbol{G}^{-1} \cdot \boldsymbol{F}_1^{\mathrm{H}} \cdot \boldsymbol{R}_0^{-1} = \boldsymbol{F}_1 \cdot \boldsymbol{G}^{-1} \cdot \boldsymbol{F}_1^{\mathrm{H}} \cdot \boldsymbol{R}_0^{-1}, \tag{B.175}$$

where the middle term $\boldsymbol{G}^{-1} \cdot \boldsymbol{F}_1^{\mathrm{H}} \cdot \boldsymbol{R}_0^{-1} \cdot \boldsymbol{F}_1$ cancels due to the definition $\boldsymbol{G} = \boldsymbol{F}_1^{\mathrm{H}} \cdot \boldsymbol{R}_0^{-1} \cdot \boldsymbol{F}_1$.

Using Proposition B.9.4, we notice that $\boldsymbol{F}_1 \cdot \boldsymbol{G}^{-1} \cdot \boldsymbol{F}_1^{\mathrm{H}} \cdot \boldsymbol{R}_0^{-1}$ projects onto the subspace spanned by $\boldsymbol{F}_1 = (\boldsymbol{I}_d \otimes \boldsymbol{J}_1 \cdot \boldsymbol{U}_{\mathrm{s}})$. As $\boldsymbol{b} = \mathrm{vec}\,\{\boldsymbol{J}_2 \cdot \boldsymbol{U}_{\mathrm{s}}\} = (\boldsymbol{I}_d \otimes \boldsymbol{J}_2 \cdot \boldsymbol{U}_{\mathrm{s}}) \cdot \mathrm{vec}\,\{\boldsymbol{I}_d\}$ spans the same subspace, we have $\boldsymbol{F}_1 \cdot \boldsymbol{G}^{-1} \cdot \boldsymbol{F}_1^{\mathrm{H}} \cdot \boldsymbol{R}_0^{-1} \cdot \boldsymbol{b} = \boldsymbol{b}$. Moreover, we observe that $\boldsymbol{F}_1 \cdot \boldsymbol{G}^{-1} \cdot \boldsymbol{F}_1^{\mathrm{H}} \cdot \boldsymbol{R}_0^{-1} \cdot \boldsymbol{R} \cdot \boldsymbol{R}_0^{-1} \cdot \boldsymbol{b} = \boldsymbol{R} \cdot \boldsymbol{R}_0^{-1} \cdot \boldsymbol{b}$ since the projection onto $\boldsymbol{F}_1 \cdot \boldsymbol{G}^{-1} \cdot \boldsymbol{F}_1^{\mathrm{H}} \cdot \boldsymbol{R}_0^{-1}$ is redundant as it spans the same subspace as $\boldsymbol{b}$. Consequently, we obtain

$$\left(\boldsymbol{F}_1^{\mathrm{H}} \cdot \bar{\boldsymbol{R}}^{-1} \cdot \boldsymbol{F}_1 \right)^{-1} \cdot \boldsymbol{F}_1^{\mathrm{H}} \cdot \bar{\boldsymbol{R}}^{-1} \cdot \boldsymbol{b}$$

$$= \boldsymbol{\psi}_0 + \boldsymbol{G}^{-1} \cdot \boldsymbol{F}_1^{\mathrm{H}} \cdot \boldsymbol{R}_0^{-1} \cdot \boldsymbol{R} \cdot \boldsymbol{R}_0^{-1} \cdot \left(\frac{1}{\lambda} - \frac{1}{\lambda\,(\lambda+1)} - \frac{1}{\lambda+1} \right) \cdot \boldsymbol{b} \tag{B.176}$$

$$= \boldsymbol{\psi}_0, \tag{B.177}$$

where $\boldsymbol{\psi}_0 = (\boldsymbol{F}_1^{\mathrm{H}} \cdot \boldsymbol{R}_0^{-1} \cdot \boldsymbol{F}_1)^{-1} \cdot \boldsymbol{F}_1^{\mathrm{H}} \cdot \boldsymbol{R}_0^{-1} \cdot \boldsymbol{b}$ is the solution associated with $\boldsymbol{R}_0$. The last equality in (B.177) follows from the fact that

$$\frac{1}{\lambda} - \frac{1}{\lambda\,(\lambda+1)} - \frac{1}{\lambda+1} = 0 \tag{B.178}$$

such that the second term in (B.176) evaluates to zero. Thus, the solution is independent of the regularization parameter λ. □

B.10. Proof of Proposition B.9.2

The proposition that $\boldsymbol{R} \cdot \boldsymbol{R}_0^{-1}$ is a projection matrix can be proven by verifying the property (B.169), which states that

$$\boldsymbol{R} \cdot \boldsymbol{R}_0^{-1} \cdot \boldsymbol{R} \cdot \boldsymbol{R}_0^{-1} = \boldsymbol{R} \cdot \boldsymbol{R}_0^{-1} \tag{B.179}$$

must hold. To show (B.179), we assume a ULA and maximum subarray overlap. First, we recall the definitions of $\boldsymbol{R}$ and $\boldsymbol{R}_0$ in (B.165) and (B.166) as

$$\boldsymbol{R} = \boldsymbol{F}_2 \cdot \left(\boldsymbol{\Sigma}_{\mathrm{s}}^{-2} \otimes \boldsymbol{U}_{\mathrm{n}} \cdot \boldsymbol{U}_{\mathrm{n}}^{\mathrm{H}}\right) \cdot \boldsymbol{F}_2^{\mathrm{H}} \in \mathbb{C}^{(M-1)d \times (M-1)d} \tag{B.180}$$

$$\boldsymbol{R}_0 = \boldsymbol{F}_2 \cdot \left(\boldsymbol{\Sigma}_{\mathrm{s}}^{-2} \otimes \boldsymbol{I}_M\right) \cdot \boldsymbol{F}_2^{\mathrm{H}} \in \mathbb{C}^{(M-1)d \times (M-1)d}, \tag{B.181}$$

where we have omitted the factor σ_{n}^2, which cancels in (B.179). Moreover, $\boldsymbol{F}_2$ is given by

$$\boldsymbol{F}_2 = \left(\boldsymbol{\Psi}^{\mathrm{T}} \otimes \boldsymbol{J}_1\right) - \left(\boldsymbol{I}_d \otimes \boldsymbol{J}_2\right) \in \mathbb{C}^{(M-1)d \times Md}. \tag{B.182}$$

Using the definition

$$\tilde{\boldsymbol{F}} = \boldsymbol{F}_2 \cdot \left(\boldsymbol{\Sigma}_{\mathrm{s}}^{-1} \otimes \boldsymbol{I}_M\right) \in \mathbb{C}^{(M-1)d \times Md}, \tag{B.183}$$

we can simplify $\boldsymbol{R}$ and $\boldsymbol{R}_0$ in (B.180) and (B.181) as

$$\boldsymbol{R} = \tilde{\boldsymbol{F}} \cdot \left(\boldsymbol{I}_d \otimes \boldsymbol{U}_{\mathrm{n}} \cdot \boldsymbol{U}_{\mathrm{n}}^{\mathrm{H}}\right) \cdot \tilde{\boldsymbol{F}}^{\mathrm{H}} \tag{B.184}$$

$$\boldsymbol{R}_0 = \tilde{\boldsymbol{F}} \cdot \tilde{\boldsymbol{F}}^{\mathrm{H}}. \tag{B.185}$$

As a result, the right-hand-side of (B.179) evaluates to

$$\boldsymbol{R} \cdot \boldsymbol{R}_0^{-1} = \tilde{\boldsymbol{F}} \cdot \left(\boldsymbol{I}_d \otimes \boldsymbol{U}_{\mathrm{n}} \cdot \boldsymbol{U}_{\mathrm{n}}^{\mathrm{H}}\right) \cdot \tilde{\boldsymbol{F}}^{\mathrm{H}} \cdot \left(\tilde{\boldsymbol{F}} \cdot \tilde{\boldsymbol{F}}^{\mathrm{H}}\right)^{-1} = \tilde{\boldsymbol{F}} \cdot \left(\boldsymbol{I}_d \otimes \boldsymbol{U}_{\mathrm{n}} \cdot \boldsymbol{U}_{\mathrm{n}}^{\mathrm{H}}\right) \cdot \tilde{\boldsymbol{F}}^{+} \tag{B.186}$$

and the left-hand-side of (B.179) becomes

$$\boldsymbol{R} \cdot \boldsymbol{R}_0^{-1} \cdot \boldsymbol{R} \cdot \boldsymbol{R}_0^{-1} = \tilde{\boldsymbol{F}} \cdot \left(\boldsymbol{I}_d \otimes \boldsymbol{U}_{\mathrm{n}} \cdot \boldsymbol{U}_{\mathrm{n}}^{\mathrm{H}}\right) \cdot \tilde{\boldsymbol{F}}^{+} \cdot \tilde{\boldsymbol{F}} \cdot \left(\boldsymbol{I}_d \otimes \boldsymbol{U}_{\mathrm{n}} \cdot \boldsymbol{U}_{\mathrm{n}}^{\mathrm{H}}\right) \cdot \tilde{\boldsymbol{F}}^{+}. \tag{B.187}$$

Hence, proving (B.179) corresponds to showing that

$$\tilde{\boldsymbol{F}} \cdot \left(\boldsymbol{I}_d \otimes \boldsymbol{U}_{\mathrm{n}} \cdot \boldsymbol{U}_{\mathrm{n}}^{\mathrm{H}}\right) \cdot \tilde{\boldsymbol{F}}^{+} = \tilde{\boldsymbol{F}} \cdot \left(\boldsymbol{I}_d \otimes \boldsymbol{U}_{\mathrm{n}} \cdot \boldsymbol{U}_{\mathrm{n}}^{\mathrm{H}}\right) \cdot \tilde{\boldsymbol{F}}^{+} \cdot \tilde{\boldsymbol{F}} \cdot \left(\boldsymbol{I}_d \otimes \boldsymbol{U}_{\mathrm{n}} \cdot \boldsymbol{U}_{\mathrm{n}}^{\mathrm{H}}\right) \cdot \tilde{\boldsymbol{F}}^{+}. \tag{B.188}$$

In [BB02], the commutativity property $\boldsymbol{P}_1 \cdot \boldsymbol{P}_2 = \boldsymbol{P}_2 \cdot \boldsymbol{P}_1$ of two orthogonal projection matrices $\boldsymbol{P}_1$ and $\boldsymbol{P}_2$, i.e., $\boldsymbol{P}_1 = \boldsymbol{P}_1^{\mathrm{H}}$ and $\boldsymbol{P}_2 = \boldsymbol{P}_2^{\mathrm{H}}$, is studied. It is easily verified that both $\tilde{\boldsymbol{F}}^{+} \cdot \tilde{\boldsymbol{F}}$ and $\left(\boldsymbol{I}_d \otimes \boldsymbol{U}_{\mathrm{n}} \cdot \boldsymbol{U}_{\mathrm{n}}^{\mathrm{H}}\right)$ are orthogonal projection matrices by checking $\boldsymbol{P}^2 = \boldsymbol{P}$ and $\boldsymbol{P} = \boldsymbol{P}^{\mathrm{H}}$. In order for the equality in (B.188) to be satisfied, we require the following commutativity property to hold for $\tilde{\boldsymbol{F}}^{+} \cdot \tilde{\boldsymbol{F}}$ and $\left(\boldsymbol{I}_d \otimes \boldsymbol{U}_{\mathrm{n}} \cdot \boldsymbol{U}_{\mathrm{n}}^{\mathrm{H}}\right)$, i.e.,

$$\tilde{\boldsymbol{F}}^{+} \cdot \tilde{\boldsymbol{F}} \cdot \left(\boldsymbol{I}_d \otimes \boldsymbol{U}_{\mathrm{n}} \cdot \boldsymbol{U}_{\mathrm{n}}^{\mathrm{H}}\right) = \left(\boldsymbol{I}_d \otimes \boldsymbol{U}_{\mathrm{n}} \cdot \boldsymbol{U}_{\mathrm{n}}^{\mathrm{H}}\right) \cdot \tilde{\boldsymbol{F}}^{+} \cdot \tilde{\boldsymbol{F}}. \tag{B.189}$$

If (B.189) holds, we can write (B.188) as

$$\begin{aligned}\tilde{\boldsymbol{F}} \cdot \left(\boldsymbol{I}_d \otimes \boldsymbol{U}_{\mathrm{n}} \cdot \boldsymbol{U}_{\mathrm{n}}^{\mathrm{H}}\right) \cdot \tilde{\boldsymbol{F}}^{+} &= \tilde{\boldsymbol{F}} \cdot \left(\boldsymbol{I}_d \otimes \boldsymbol{U}_{\mathrm{n}} \cdot \boldsymbol{U}_{\mathrm{n}}^{\mathrm{H}}\right) \cdot \tilde{\boldsymbol{F}}^{+} \cdot \tilde{\boldsymbol{F}} \cdot \left(\boldsymbol{I}_d \otimes \boldsymbol{U}_{\mathrm{n}} \cdot \boldsymbol{U}_{\mathrm{n}}^{\mathrm{H}}\right) \cdot \tilde{\boldsymbol{F}}^{+} \\ &= \tilde{\boldsymbol{F}} \cdot \left(\boldsymbol{I}_d \otimes \boldsymbol{U}_{\mathrm{n}} \cdot \boldsymbol{U}_{\mathrm{n}}^{\mathrm{H}}\right) \cdot \left(\boldsymbol{I}_d \otimes \boldsymbol{U}_{\mathrm{n}} \cdot \boldsymbol{U}_{\mathrm{n}}^{\mathrm{H}}\right) \cdot \tilde{\boldsymbol{F}}^{+} \cdot \tilde{\boldsymbol{F}} \cdot \tilde{\boldsymbol{F}}^{+} \\ &= \tilde{\boldsymbol{F}} \cdot \left(\boldsymbol{I}_d \otimes \boldsymbol{U}_{\mathrm{n}} \cdot \boldsymbol{U}_{\mathrm{n}}^{\mathrm{H}}\right) \cdot \tilde{\boldsymbol{F}}^{+}. \end{aligned} \tag{B.190}$$

Thus, we need to show that (B.189) is satisfied. To do so, we first use the correspondence $\boldsymbol{U}_{\mathrm{n}} \cdot \boldsymbol{U}_{\mathrm{n}}^{\mathrm{H}} = \boldsymbol{I}_M - \boldsymbol{A} \cdot \boldsymbol{A}^{+}$, where $\boldsymbol{A}$ is the array steering matrix in (5.1) for the 1-D case and $\boldsymbol{A} \cdot \boldsymbol{A}^{+}$ is the projection matrix onto the signal subspace. Then, we expand both sides of (B.189) as

$$\begin{aligned}\tilde{\boldsymbol{F}}^{+} \cdot \tilde{\boldsymbol{F}} \cdot (\boldsymbol{I}_d \otimes (\boldsymbol{I}_M - \boldsymbol{A} \cdot \boldsymbol{A}^{+})) &= (\boldsymbol{I}_d \otimes (\boldsymbol{I}_M - \boldsymbol{A} \cdot \boldsymbol{A}^{+})) \cdot \tilde{\boldsymbol{F}}^{+} \cdot \tilde{\boldsymbol{F}} \\ \tilde{\boldsymbol{F}}^{+} \cdot \tilde{\boldsymbol{F}} - \tilde{\boldsymbol{F}}^{+} \cdot \tilde{\boldsymbol{F}} \cdot (\boldsymbol{I}_d \otimes \boldsymbol{A} \cdot \boldsymbol{A}^{+}) &= \tilde{\boldsymbol{F}}^{+} \cdot \tilde{\boldsymbol{F}} - (\boldsymbol{I}_d \otimes \boldsymbol{A} \cdot \boldsymbol{A}^{+}) \cdot \tilde{\boldsymbol{F}}^{+} \cdot \tilde{\boldsymbol{F}}. \end{aligned} \tag{B.191}$$

Since the term $\tilde{\boldsymbol{F}}^{+} \cdot \tilde{\boldsymbol{F}}$ appears on both sides of (B.191), it remains to be shown that

$$\tilde{\boldsymbol{F}}^{+} \cdot \tilde{\boldsymbol{F}} \cdot (\boldsymbol{I}_d \otimes \boldsymbol{A} \cdot \boldsymbol{A}^{+}) = (\boldsymbol{I}_d \otimes \boldsymbol{A} \cdot \boldsymbol{A}^{+}) \cdot \tilde{\boldsymbol{F}}^{+} \cdot \tilde{\boldsymbol{F}}. \tag{B.192}$$

In order to prove the equality of (B.192), we simplify both sides of (B.192) separately and show that their simplified expressions are equal. We start with the left hand side of (B.192) and expand it to

$$\tilde{\boldsymbol{F}}^{+} \cdot \tilde{\boldsymbol{F}} \cdot (\boldsymbol{I}_d \otimes \boldsymbol{A} \cdot \boldsymbol{A}^{+}) = \tilde{\boldsymbol{F}}^{\mathrm{H}} \cdot \left(\tilde{\boldsymbol{F}} \cdot \tilde{\boldsymbol{F}}^{\mathrm{H}}\right)^{-1} \cdot \tilde{\boldsymbol{F}} \cdot (\boldsymbol{I}_d \otimes \boldsymbol{A} \cdot \boldsymbol{A}^{+}). \tag{B.193}$$

Then, we compute the term $\tilde{\boldsymbol{F}} \cdot \tilde{\boldsymbol{F}}^{\mathrm{H}}$, where $\tilde{\boldsymbol{F}}$ is defined in (B.183) as

$$\tilde{\boldsymbol{F}} = \boldsymbol{F}_2 \cdot \left(\boldsymbol{\Sigma}_{\mathrm{s}}^{-1} \otimes \boldsymbol{I}_M\right) \tag{B.194}$$

and $\boldsymbol{F}_2$ is given in (B.182). Expanding $\boldsymbol{F}_2$ and using the eigendecomposition of $\boldsymbol{\Psi}$ given by $\boldsymbol{\Psi} = \boldsymbol{Q} \cdot \boldsymbol{\Lambda} \cdot \boldsymbol{Q}^{-1}$, we can write

$$\begin{aligned}\boldsymbol{F}_2 &= \left(\boldsymbol{\Psi}^{\mathrm{T}} \otimes \boldsymbol{J}_1\right) - (\boldsymbol{I}_d \otimes \boldsymbol{J}_2) \\ &= \left(\boldsymbol{Q}^{-\mathrm{T}} \cdot \boldsymbol{\Lambda} \cdot \boldsymbol{Q}^{\mathrm{T}} \otimes \boldsymbol{J}_1\right) - (\boldsymbol{I}_d \otimes \boldsymbol{J}_2) \\ &= \left(\boldsymbol{Q}^{-\mathrm{T}} \otimes \boldsymbol{I}_{M-1}\right) \cdot \bar{\boldsymbol{F}} \cdot \left(\boldsymbol{Q}^{\mathrm{T}} \otimes \boldsymbol{I}_M\right), \end{aligned} \tag{B.195}$$

where we have defined $\bar{\boldsymbol{F}} = (\boldsymbol{\Lambda} \otimes \boldsymbol{J}_1) - (\boldsymbol{I}_d \otimes \boldsymbol{J}_2)$. Inserting (B.195) into (B.194), we obtain

$$\tilde{\boldsymbol{F}} = \left(\boldsymbol{Q}^{-\mathrm{T}} \otimes \boldsymbol{I}_{M-1}\right) \cdot \bar{\boldsymbol{F}} \cdot \left(\boldsymbol{Q}^{\mathrm{T}} \cdot \boldsymbol{\Sigma}_{\mathrm{s}}^{-1} \otimes \boldsymbol{I}_M\right). \tag{B.196}$$

Then, we can express $\tilde{\boldsymbol{F}} \cdot \tilde{\boldsymbol{F}}^{\mathrm{H}}$ as

$$\begin{aligned}\tilde{\boldsymbol{F}} \cdot \tilde{\boldsymbol{F}}^{\mathrm{H}} &= \left(\boldsymbol{Q}^{-\mathrm{T}} \otimes \boldsymbol{I}_{M-1}\right) \cdot \bar{\boldsymbol{F}} \cdot \left(\boldsymbol{Q}^{\mathrm{T}} \cdot \boldsymbol{\Sigma}_{\mathrm{s}}^{-1} \otimes \boldsymbol{I}_M\right) \cdot \left(\left(\boldsymbol{Q}^{-\mathrm{T}} \otimes \boldsymbol{I}_{M-1}\right) \cdot \bar{\boldsymbol{F}} \cdot \left(\boldsymbol{Q}^{\mathrm{T}} \cdot \boldsymbol{\Sigma}_{\mathrm{s}}^{-1} \otimes \boldsymbol{I}_M\right)\right)^{\mathrm{H}} \\ &= \left(\boldsymbol{Q}^{-\mathrm{T}} \otimes \boldsymbol{I}_{M-1}\right) \cdot \bar{\boldsymbol{F}} \cdot \left(\boldsymbol{Q}^{\mathrm{T}} \cdot \boldsymbol{\Sigma}_{\mathrm{s}}^{-1} \otimes \boldsymbol{I}_M\right) \cdot \left(\boldsymbol{\Sigma}_{\mathrm{s}}^{-1} \cdot \boldsymbol{Q}^{*} \otimes \boldsymbol{I}_M\right) \cdot \bar{\boldsymbol{F}}^{\mathrm{H}} \cdot \left(\boldsymbol{Q}^{-*} \otimes \boldsymbol{I}_{M-1}\right) \\ &= \left(\boldsymbol{Q}^{-\mathrm{T}} \otimes \boldsymbol{I}_{M-1}\right) \cdot \bar{\boldsymbol{F}} \cdot \left(\boldsymbol{Q}^{\mathrm{T}} \cdot \boldsymbol{\Sigma}_{\mathrm{s}}^{-2} \cdot \boldsymbol{Q}^{*} \otimes \boldsymbol{I}_M\right) \cdot \bar{\boldsymbol{F}}^{\mathrm{H}} \cdot \left(\boldsymbol{Q}^{-*} \otimes \boldsymbol{I}_{M-1}\right). \end{aligned} \tag{B.197}$$

In the next step, we analyze the middle term $\left(\boldsymbol{Q}^{\mathrm{T}} \cdot \boldsymbol{\Sigma}_{\mathrm{s}}^{-2} \cdot \boldsymbol{Q}^{*} \otimes \boldsymbol{I}_M\right)$. To this end, we consider the subspace equivalence (cf. Equation (3.32) in the noise-free case)

$$\boldsymbol{A} = \boldsymbol{U}_{\mathrm{s}} \cdot \boldsymbol{T} = \sqrt{M} \cdot \boldsymbol{U}_{\mathrm{s}} \cdot \boldsymbol{Q}, \tag{B.198}$$

where we have decomposed $\boldsymbol{T}$ by $\boldsymbol{T} = \sqrt{M} \cdot \boldsymbol{Q}$. Rearranging (B.198) as $\boldsymbol{U}_{\mathrm{s}} = \frac{1}{\sqrt{M}} \cdot \boldsymbol{A} \cdot \boldsymbol{Q}^{-1}$, we obtain

$$\boldsymbol{X}_0 = \boldsymbol{U}_{\mathrm{s}} \cdot \boldsymbol{\Sigma}_{\mathrm{s}} \cdot \boldsymbol{V}_{\mathrm{s}}^{\mathrm{H}} = \frac{1}{\sqrt{M}} \cdot \boldsymbol{A} \cdot \boldsymbol{Q}^{-1} \cdot \boldsymbol{\Sigma}_{\mathrm{s}} \cdot \boldsymbol{V}_{\mathrm{s}}^{\mathrm{H}} = \boldsymbol{A} \cdot \boldsymbol{S}. \tag{B.199}$$

Hence, we have $\boldsymbol{S} = \frac{1}{\sqrt{M}} \cdot \boldsymbol{Q}^{-1} \cdot \boldsymbol{\Sigma}_{\mathrm{s}} \cdot \boldsymbol{V}_{\mathrm{s}}^{\mathrm{H}}$. Considering the sample signal covariance $\hat{\boldsymbol{R}}_{\mathrm{ss}} = \boldsymbol{S} \cdot \boldsymbol{S}^{\mathrm{H}}/N$, we can write

$$\hat{\boldsymbol{R}}_{\mathrm{ss}} = \frac{1}{N} \cdot \boldsymbol{S} \cdot \boldsymbol{S}^{\mathrm{H}} = \frac{1}{N \cdot M} \cdot \boldsymbol{Q}^{-1} \cdot \boldsymbol{\Sigma}_{\mathrm{s}}^{2} \cdot \boldsymbol{Q}^{-\mathrm{H}}. \tag{B.200}$$

Transposing and inverting (B.200), we arrive at

$$\boldsymbol{Q}^{\mathrm{T}} \cdot \boldsymbol{\Sigma}_{\mathrm{s}}^{-2} \cdot \boldsymbol{Q}^{*} = \frac{1}{M \cdot N} \cdot \hat{\boldsymbol{R}}_{\mathrm{ss}}^{-\mathrm{T}}. \tag{B.201}$$

Finally, we insert (B.201) into (B.197) to obtain

$$\tilde{\boldsymbol{F}} \cdot \tilde{\boldsymbol{F}}^{\mathrm{H}} = \left(\boldsymbol{Q}^{-\mathrm{T}} \otimes \boldsymbol{I}_{M-1}\right) \cdot \bar{\boldsymbol{F}} \cdot \left(\frac{1}{M \cdot N} \cdot \hat{\boldsymbol{R}}_{\mathrm{ss}}^{-\mathrm{T}} \otimes \boldsymbol{I}_M\right) \cdot \bar{\boldsymbol{F}}^{\mathrm{H}} \cdot \left(\boldsymbol{Q}^{-*} \otimes \boldsymbol{I}_{M-1}\right). \tag{B.202}$$

Now, we assume for simplicity that the signals are temporally orthogonal such that the sample signal covariance matrix $\hat{\boldsymbol{R}}_{\mathrm{ss}}$ becomes the diagonal matrix $\hat{\boldsymbol{R}}_{\mathrm{ss}} = \operatorname{diag}\left\{\hat{P}_i\right\}_{i=1}^{d}$, where $\hat{P}_i = \|\boldsymbol{s}_i\|_2^2/N$ is the empirical power of the i-th signal $\boldsymbol{s}_i$. Using the property that $\hat{\boldsymbol{R}}_{\mathrm{ss}}$ is diagonal, $\boldsymbol{Q}^{\mathrm{T}} \cdot \boldsymbol{\Sigma}_{\mathrm{s}}^{-2} \cdot \boldsymbol{Q}^{*}$ is a diagonal matrix as well due to (B.201) and we can express (B.197) as

$$\begin{aligned}\tilde{\boldsymbol{F}} \cdot \tilde{\boldsymbol{F}}^{\mathrm{H}} &= \left(\boldsymbol{Q}^{-\mathrm{T}} \otimes \boldsymbol{I}_{M-1}\right) \cdot \bar{\boldsymbol{F}} \cdot \left(\boldsymbol{Q}^{\mathrm{T}} \cdot \boldsymbol{\Sigma}_{\mathrm{s}}^{-2} \cdot \boldsymbol{Q}^{*} \otimes \boldsymbol{I}_M\right) \cdot \bar{\boldsymbol{F}}^{\mathrm{H}} \cdot \left(\boldsymbol{Q}^{-*} \otimes \boldsymbol{I}_{M-1}\right) \\ &= \left(\boldsymbol{Q}^{-\mathrm{T}} \otimes \boldsymbol{I}_{M-1}\right) \cdot \bar{\boldsymbol{F}} \cdot \bar{\boldsymbol{F}}^{\mathrm{H}} \cdot \left(\frac{1}{M \cdot N} \cdot \hat{\boldsymbol{R}}_{\mathrm{ss}}^{-\mathrm{T}} \otimes \boldsymbol{I}_{M-1}\right) \cdot \left(\boldsymbol{Q}^{-*} \otimes \boldsymbol{I}_{M-1}\right) \\ &= \left(\boldsymbol{Q}^{-\mathrm{T}} \otimes \boldsymbol{I}_{M-1}\right) \cdot \bar{\boldsymbol{F}} \cdot \bar{\boldsymbol{F}}^{\mathrm{H}} \cdot \left(\boldsymbol{Q}^{\mathrm{T}} \cdot \boldsymbol{\Sigma}_{\mathrm{s}}^{-2} \cdot \boldsymbol{Q}^{*} \otimes \boldsymbol{I}_{M-1}\right) \cdot \left(\boldsymbol{Q}^{-*} \otimes \boldsymbol{I}_{M-1}\right) \end{aligned} \tag{B.203}$$

$$= \left(\boldsymbol{Q}^{-\mathrm{T}} \otimes \boldsymbol{I}_{M-1}\right) \cdot \bar{\boldsymbol{F}} \cdot \bar{\boldsymbol{F}}^{\mathrm{H}} \cdot \left(\boldsymbol{Q}^{\mathrm{T}} \cdot \boldsymbol{\Sigma}_{\mathrm{s}}^{-2} \otimes \boldsymbol{I}_{M-1}\right), \tag{B.204}$$

where in (B.203), we have used the fact that diagonal matrices commute. Consequently, the term $(\tilde{\boldsymbol{F}} \cdot \tilde{\boldsymbol{F}}^{\mathrm{H}})^{-1}$ in (B.193) is given by

$$\boldsymbol{R}_0^{-1} = \left(\tilde{\boldsymbol{F}} \cdot \tilde{\boldsymbol{F}}^{\mathrm{H}}\right)^{-1} = \left(\boldsymbol{\Sigma}_{\mathrm{s}}^{2} \cdot \boldsymbol{Q}^{-\mathrm{T}} \otimes \boldsymbol{I}_{M-1}\right) \cdot \left(\bar{\boldsymbol{F}} \cdot \bar{\boldsymbol{F}}^{\mathrm{H}}\right)^{-1} \cdot \left(\boldsymbol{Q}^{\mathrm{T}} \otimes \boldsymbol{I}_{M-1}\right). \tag{B.205}$$

Next, we consider the term $\tilde{\boldsymbol{F}}^{\mathrm{H}} \cdot \left(\tilde{\boldsymbol{F}} \cdot \tilde{\boldsymbol{F}}^{\mathrm{H}}\right)^{-1} \cdot \tilde{\boldsymbol{F}}$ in (B.193), which can be written as

$$\tilde{\boldsymbol{F}}^{\mathrm{H}} \cdot \left(\tilde{\boldsymbol{F}} \cdot \tilde{\boldsymbol{F}}^{\mathrm{H}}\right)^{-1} \cdot \tilde{\boldsymbol{F}} = \tilde{\boldsymbol{F}}^{\mathrm{H}} \cdot \left(\boldsymbol{\Sigma}_{\mathrm{s}}^{2} \cdot \boldsymbol{Q}^{-\mathrm{T}} \otimes \boldsymbol{I}_{M-1}\right) \cdot \left(\bar{\boldsymbol{F}} \cdot \bar{\boldsymbol{F}}^{\mathrm{H}}\right)^{-1} \cdot \left(\boldsymbol{Q}^{\mathrm{T}} \otimes \boldsymbol{I}_{M-1}\right) \cdot \tilde{\boldsymbol{F}}. \tag{B.206}$$

We simplify the term $\left(\boldsymbol{Q}^{\mathrm{T}} \otimes \boldsymbol{I}_{M-1}\right) \cdot \tilde{\boldsymbol{F}}$ according to

$$\begin{aligned}
\left(\boldsymbol{Q}^{\mathrm{T}} \otimes \boldsymbol{I}_{M-1}\right) \cdot \tilde{\boldsymbol{F}} &= \left(\boldsymbol{Q}^{\mathrm{T}} \otimes \boldsymbol{I}_{M-1}\right) \cdot \boldsymbol{F}_2 \cdot \left(\boldsymbol{\Sigma}_{\mathrm{s}}^{-1} \otimes \boldsymbol{I}_{M}\right) && \text{(B.207)} \\
&= \left(\boldsymbol{Q}^{\mathrm{T}} \otimes \boldsymbol{I}_{M-1}\right) \cdot \left(\boldsymbol{Q}^{-\mathrm{T}} \otimes \boldsymbol{I}_{M-1}\right) \cdot \bar{\boldsymbol{F}} \cdot \left(\boldsymbol{Q}^{\mathrm{T}} \otimes \boldsymbol{I}_{M}\right) \left(\boldsymbol{\Sigma}_{\mathrm{s}}^{-1} \otimes \boldsymbol{I}_{M}\right) \\
&= \bar{\boldsymbol{F}} \cdot \left(\boldsymbol{Q}^{\mathrm{T}} \cdot \boldsymbol{\Sigma}_{\mathrm{s}}^{-1} \otimes \boldsymbol{I}_{M}\right) && \text{(B.208)}
\end{aligned}$$

and the term $\tilde{\boldsymbol{F}}^{\mathrm{H}} \cdot \left(\boldsymbol{\Sigma}_{\mathrm{s}}^{2} \cdot \boldsymbol{Q}^{-\mathrm{T}} \otimes \boldsymbol{I}_{M-1}\right)$ according to

$$\begin{aligned}
\tilde{\boldsymbol{F}}^{\mathrm{H}} \cdot \left(\boldsymbol{\Sigma}_{\mathrm{s}}^{2} \cdot \boldsymbol{Q}^{-\mathrm{T}} \otimes \boldsymbol{I}_{M-1}\right) &= \left(\boldsymbol{\Sigma}_{\mathrm{s}}^{-1} \otimes \boldsymbol{I}_{M}\right) \cdot \boldsymbol{F}_2^{\mathrm{H}} \cdot \left(\boldsymbol{\Sigma}_{\mathrm{s}}^{2} \cdot \boldsymbol{Q}^{-\mathrm{T}} \otimes \boldsymbol{I}_{M-1}\right) && \text{(B.209)} \\
&= \left(\boldsymbol{\Sigma}_{\mathrm{s}}^{-1} \otimes \boldsymbol{I}_{M}\right) \cdot \left(\boldsymbol{Q}^{*} \otimes \boldsymbol{I}_{M}\right) \cdot \bar{\boldsymbol{F}}^{\mathrm{H}} \cdot \left(\boldsymbol{Q}^{-*} \otimes \boldsymbol{I}_{M-1}\right) \cdot \left(\boldsymbol{\Sigma}_{\mathrm{s}}^{2} \cdot \boldsymbol{Q}^{-\mathrm{T}} \otimes \boldsymbol{I}_{M-1}\right) \\
&= \left(\boldsymbol{\Sigma}_{\mathrm{s}}^{-1} \cdot \boldsymbol{Q}^{*} \otimes \boldsymbol{I}_{M}\right) \cdot \bar{\boldsymbol{F}}^{\mathrm{H}} \cdot \left(\boldsymbol{Q}^{-*} \cdot \boldsymbol{\Sigma}_{\mathrm{s}}^{2} \cdot \boldsymbol{Q}^{-\mathrm{T}} \otimes \boldsymbol{I}_{M-1}\right) && \text{(B.210)} \\
&= \left(\boldsymbol{\Sigma}_{\mathrm{s}} \cdot \boldsymbol{Q}^{-\mathrm{T}} \otimes \boldsymbol{I}_{M}\right) \cdot \bar{\boldsymbol{F}}^{\mathrm{H}}, && \text{(B.211)}
\end{aligned}$$

where we have inserted $\boldsymbol{F}_2$ from (B.195) into (B.207) and (B.209). Moreover, we have used again the interchangeability of diagonal matrices in (B.210).

With the results from (B.208) and (B.211), (B.193) can be expanded as

$$\tilde{\boldsymbol{F}}^{\mathrm{H}} \cdot \left(\tilde{\boldsymbol{F}} \cdot \tilde{\boldsymbol{F}}^{\mathrm{H}}\right)^{-1} \cdot \tilde{\boldsymbol{F}} = \left(\boldsymbol{\Sigma}_{\mathrm{s}} \cdot \boldsymbol{Q}^{-\mathrm{T}} \otimes \boldsymbol{I}_{M}\right) \cdot \bar{\boldsymbol{F}}^{\mathrm{H}} \cdot \left(\bar{\boldsymbol{F}} \cdot \bar{\boldsymbol{F}}^{\mathrm{H}}\right)^{-1} \cdot \bar{\boldsymbol{F}} \cdot \left(\boldsymbol{Q}^{\mathrm{T}} \cdot \boldsymbol{\Sigma}_{\mathrm{s}}^{-1} \otimes \boldsymbol{I}_{M}\right). \tag{B.212}$$

Next, we simplify the expression $\bar{\boldsymbol{F}}^{\mathrm{H}} \cdot \left(\bar{\boldsymbol{F}} \cdot \bar{\boldsymbol{F}}^{\mathrm{H}}\right)^{-1} \cdot \bar{\boldsymbol{F}}$ and first consider the inverse term $\left(\bar{\boldsymbol{F}} \cdot \bar{\boldsymbol{F}}^{\mathrm{H}}\right)^{-1}$. To this end, we define the square matrix $\boldsymbol{Z} \in \mathbb{C}^{(M-1)d \times (M-1)d}$ according to

$$\boldsymbol{Z} = \bar{\boldsymbol{F}} \cdot \left(\boldsymbol{I}_d \otimes \boldsymbol{J}_1^{\mathrm{T}}\right), \tag{B.213}$$

which by taking into account the definition of $\bar{\boldsymbol{F}}$ in (B.195) can also be represented as

$$\boldsymbol{Z} = \mathrm{blkdiag}\left\{\boldsymbol{Z}_1, \ldots, \boldsymbol{Z}_d\right\} \tag{B.214}$$

with $\boldsymbol{Z}_i \in \mathbb{C}^{(M-1)\times(M-1)}$ for $i = 1, \ldots, d$ being given by

$$\boldsymbol{Z}_i = \begin{bmatrix} \mathrm{e}^{\mathrm{j}\mu_i} & -1 & 0 & \ldots & 0 \\ 0 & \mathrm{e}^{\mathrm{j}\mu_i} & -1 & \ldots & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & 0 & 0 & \ldots & -1 \\ 0 & 0 & 0 & \ldots & \mathrm{e}^{\mathrm{j}\mu_i} \end{bmatrix}, \tag{B.215}$$

where we assume a ULA and maximum subarray overlap (see (3.29)). Moreover, we let $\boldsymbol{u} = \begin{bmatrix} 0 & \cdots & 0 & 1 \end{bmatrix}^{\mathrm{T}} \in \mathbb{R}^{(M-1)\times 1}$. Using these definitions, the term $\bar{\boldsymbol{F}} \cdot \bar{\boldsymbol{F}}^{\mathrm{H}}$ can be expressed as

$$\bar{\boldsymbol{F}} \cdot \bar{\boldsymbol{F}}^{\mathrm{H}} = \boldsymbol{Z} \cdot \boldsymbol{Z}^{\mathrm{H}} + \boldsymbol{I}_d \otimes \boldsymbol{u} \cdot \boldsymbol{u}^{\mathrm{T}} \tag{B.216}$$

and upon applying the matrix inversion lemma [GvL96], we obtain

$$\begin{aligned} \left(\bar{\boldsymbol{F}} \cdot \bar{\boldsymbol{F}}^{\mathrm{H}}\right)^{-1} &= \left(\boldsymbol{Z} \cdot \boldsymbol{Z}^{\mathrm{H}}\right)^{-1} \\ &- \left(\boldsymbol{Z} \cdot \boldsymbol{Z}^{\mathrm{H}}\right)^{-1} \cdot (\boldsymbol{I}_d \otimes \boldsymbol{u}) \cdot \left(\boldsymbol{I}_d + \left(\boldsymbol{I}_d \otimes \boldsymbol{u}^{\mathrm{T}}\right) \cdot \left(\boldsymbol{Z} \cdot \boldsymbol{Z}^{\mathrm{H}}\right)^{-1} \cdot (\boldsymbol{I}_d \otimes \boldsymbol{u})\right)^{-1} \left(\boldsymbol{I}_d \otimes \boldsymbol{u}^{\mathrm{T}}\right) \cdot \left(\boldsymbol{Z} \cdot \boldsymbol{Z}^{\mathrm{H}}\right)^{-1}. \end{aligned} \tag{B.217}$$

Note that $\left(\boldsymbol{Z} \cdot \boldsymbol{Z}^{\mathrm{H}}\right)^{-1} = \boldsymbol{Z}^{-\mathrm{H}} \cdot \boldsymbol{Z}^{-1}$ as $\boldsymbol{Z}$ is a square matrix. Then, it is straightforward to establish that $\boldsymbol{Z}^{-1} \in \mathbb{C}^{(M-1)d\times(M-1)d}$ can be computed as

$$\boldsymbol{Z}^{-1} = \mathrm{blkdiag}\left\{\boldsymbol{Z}_1^{-1}, \ldots, \boldsymbol{Z}_d^{-1}\right\}, \tag{B.218}$$

where $\boldsymbol{Z}_i^{-1} \in \mathbb{C}^{(M-1)\times(M-1)}$ is given by

$$\boldsymbol{Z}_i^{-1} = \begin{bmatrix} \mathrm{e}^{-\mathrm{j}\mu_i} & \mathrm{e}^{-\mathrm{j}2\mu_i} & \ldots & \mathrm{e}^{-\mathrm{j}(M-1)\mu_i} \\ 0 & \mathrm{e}^{-\mathrm{j}\mu_i} & \ldots & \vdots \\ \vdots & \ddots & \ddots & \mathrm{e}^{-\mathrm{j}2\mu_i} \\ 0 & 0 & \ldots & \mathrm{e}^{-\mathrm{j}\mu_i} \end{bmatrix}. \tag{B.219}$$

Next, we find that the inverse term in (B.217) can be compactly expressed as

$$\boldsymbol{I}_d + \left(\boldsymbol{I}_d \otimes \boldsymbol{u}^{\mathrm{T}}\right) \cdot \boldsymbol{Z}^{-\mathrm{H}} \cdot \boldsymbol{Z}^{-1} \cdot (\boldsymbol{I}_d \otimes \boldsymbol{u}) = M \cdot \boldsymbol{I}_d. \tag{B.220}$$

To see this, notice that the i-th block $\boldsymbol{Z}_i^{-1} \cdot \boldsymbol{u}$ on the diagonal of $\boldsymbol{Z}^{-1} \cdot (\boldsymbol{I}_d \otimes \boldsymbol{u})$ denotes the last column of $\boldsymbol{Z}_i^{-1}$, which admits the formulation

$$\boldsymbol{Z}_i^{-1} \cdot \boldsymbol{u} = \boldsymbol{J}_1 \cdot \boldsymbol{a}_i \cdot \mathrm{e}^{-\mathrm{j}(M-1)\mu_i} \in \mathbb{C}^{(M-1)\times 1}, \tag{B.221}$$

where $\boldsymbol{a}_i \in \mathbb{C}^{M\times 1}$ with $i = 1,\ldots,d$ represents the i-th steering vector of the array steering matrix $\boldsymbol{A} = [\boldsymbol{a}_1,\ldots,\boldsymbol{a}_d] \in \mathbb{C}^{M\times d}$. Considering all blocks simultaneously, we obtain

$$\boldsymbol{Z}^{-1}\cdot(\boldsymbol{I}_d\otimes\boldsymbol{u}) = (\boldsymbol{I}_d\otimes\boldsymbol{J}_1)\cdot\operatorname{blkdiag}\left\{\boldsymbol{a}_1\cdot\mathrm{e}^{-\mathrm{j}(M-1)\mu_1},\ldots,\boldsymbol{a}_d\cdot\mathrm{e}^{-\mathrm{j}(M-1)\mu_d}\right\} \tag{B.222}$$

such that

$$\begin{aligned}\left(\boldsymbol{I}_d\otimes\boldsymbol{u}^{\mathrm{T}}\right)\cdot\boldsymbol{Z}^{-\mathrm{H}}\cdot\boldsymbol{Z}^{-1}\cdot(\boldsymbol{I}_d\otimes\boldsymbol{u}) &= \operatorname{blkdiag}\left\{\boldsymbol{a}_1^{\mathrm{H}}\cdot\boldsymbol{J}_1^{\mathrm{H}}\cdot\boldsymbol{J}_1\cdot\boldsymbol{a}_1,\ldots,\boldsymbol{a}_d^{\mathrm{H}}\cdot\boldsymbol{J}_1^{\mathrm{H}}\cdot\boldsymbol{J}_1\cdot\boldsymbol{a}_d\right\}\\ &= (M-1)\cdot\boldsymbol{I}_d.\end{aligned} \tag{B.223}$$

Hence, (B.220) is easily deduced and (B.217) can be written as

$$\begin{aligned}\left(\bar{\boldsymbol{F}}\cdot\bar{\boldsymbol{F}}^{\mathrm{H}}\right)^{-1} &= \boldsymbol{Z}^{-\mathrm{H}}\cdot\boldsymbol{Z}^{-1} - \frac{1}{M}\cdot\boldsymbol{Z}^{-\mathrm{H}}\cdot\boldsymbol{Z}^{-1}(\boldsymbol{I}_d\otimes\boldsymbol{u})\cdot\left(\boldsymbol{I}_d\otimes\boldsymbol{u}^{\mathrm{T}}\right)\cdot\boldsymbol{Z}^{-\mathrm{H}}\cdot\boldsymbol{Z}^{-1}\\ &= \boldsymbol{Z}^{-\mathrm{H}}\cdot\left(\boldsymbol{I}_{(M-1)d} - \frac{1}{M}\cdot\boldsymbol{Z}^{-1}\cdot(\boldsymbol{I}_d\otimes\boldsymbol{u})\cdot\left(\boldsymbol{I}_d\otimes\boldsymbol{u}^{\mathrm{T}}\right)\cdot\boldsymbol{Z}^{-\mathrm{H}}\right)\cdot\boldsymbol{Z}^{-1}.\end{aligned} \tag{B.224}$$

Inserting (B.222) into (B.224), we obtain

$$\begin{aligned}\left(\bar{\boldsymbol{F}}\cdot\bar{\boldsymbol{F}}^{\mathrm{H}}\right)^{-1} &= \boldsymbol{Z}^{-\mathrm{H}}\cdot\Bigg(\boldsymbol{I}_{(M-1)d} - \frac{1}{M}\cdot(\boldsymbol{I}_d\otimes\boldsymbol{J}_1)\cdot\operatorname{blkdiag}\left\{\boldsymbol{a}_1\cdot\mathrm{e}^{-\mathrm{j}(M-1)\mu_1},\ldots,\boldsymbol{a}_d\cdot\mathrm{e}^{-\mathrm{j}(M-1)\mu_d}\right\}\\ &\qquad\cdot\operatorname{blkdiag}\left\{\boldsymbol{a}_1\cdot\mathrm{e}^{-\mathrm{j}(M-1)\mu_1},\ldots,\boldsymbol{a}_d\cdot\mathrm{e}^{-\mathrm{j}(M-1)\mu_d}\right\}^{\mathrm{H}}\cdot\left(\boldsymbol{I}_d\otimes\boldsymbol{J}_1^{\mathrm{T}}\right)\Bigg)\cdot\boldsymbol{Z}^{-1}\\ &= \boldsymbol{Z}^{-\mathrm{H}}\cdot(\boldsymbol{I}_d\otimes\boldsymbol{J}_1)\cdot\operatorname{blkdiag}\left\{\boldsymbol{P}^{\perp}_{\boldsymbol{a}_1},\ldots,\boldsymbol{P}^{\perp}_{\boldsymbol{a}_d}\right\}\cdot\left(\boldsymbol{I}_d\otimes\boldsymbol{J}_1^{\mathrm{T}}\right)\cdot\boldsymbol{Z}^{-1},\end{aligned} \tag{B.225}$$

where $\boldsymbol{P}^{\perp}_{\boldsymbol{a}_i} = \boldsymbol{I}_M - \frac{1}{M}\cdot\boldsymbol{a}_i\cdot\boldsymbol{a}_i^{\mathrm{H}} \in \mathbb{C}^{M\times M}$ is the projection matrix onto the complement of the steering vector $\boldsymbol{a}_i$ of the i-th source. Then, the term $\bar{\boldsymbol{F}}^{\mathrm{H}}\cdot\left(\bar{\boldsymbol{F}}\cdot\bar{\boldsymbol{F}}^{\mathrm{H}}\right)^{-1}\cdot\bar{\boldsymbol{F}}$ in (B.212) is given by

$$\bar{\boldsymbol{F}}^{\mathrm{H}}\cdot\left(\bar{\boldsymbol{F}}\cdot\bar{\boldsymbol{F}}^{\mathrm{H}}\right)^{-1}\cdot\bar{\boldsymbol{F}} = \bar{\boldsymbol{F}}^{\mathrm{H}}\cdot\boldsymbol{Z}^{-\mathrm{H}}\cdot(\boldsymbol{I}_d\otimes\boldsymbol{J}_1)\cdot\operatorname{blkdiag}\left\{\boldsymbol{P}^{\perp}_{\boldsymbol{a}_1},\ldots,\boldsymbol{P}^{\perp}_{\boldsymbol{a}_d}\right\}\cdot\left(\boldsymbol{I}_d\otimes\boldsymbol{J}_1^{\mathrm{T}}\right)\cdot\boldsymbol{Z}^{-1}\cdot\bar{\boldsymbol{F}}. \tag{B.226}$$

Let us first analyze the last term $\left(\boldsymbol{I}_d\otimes\boldsymbol{J}_1^{\mathrm{T}}\right)\cdot\boldsymbol{Z}^{-1}\cdot\bar{\boldsymbol{F}}$. We can compute $\boldsymbol{Z}^{-1}\cdot\bar{\boldsymbol{F}}$ as

$$\begin{aligned}\boldsymbol{Z}^{-1}\cdot\bar{\boldsymbol{F}} &= \boldsymbol{Z}^{-1}\cdot((\boldsymbol{\Lambda}\otimes\boldsymbol{J}_1) - (\boldsymbol{I}_d\otimes\boldsymbol{J}_2))\\ &= (\boldsymbol{I}_d\otimes\boldsymbol{J}_1)\cdot\left(\boldsymbol{I}_{Md} - \left(\boldsymbol{I}_d\otimes\boldsymbol{J}_1^{\mathrm{T}}\right)\cdot\boldsymbol{Z}^{-1}\cdot(\boldsymbol{I}_d\otimes\boldsymbol{u})\cdot\left(\boldsymbol{I}_d\otimes\boldsymbol{u}^{\mathrm{T}}\right)\cdot(\boldsymbol{I}_d\otimes\boldsymbol{J}_2)\right) &\text{(B.227)}\\ &= (\boldsymbol{I}_d\otimes\boldsymbol{J}_1)\cdot\left(\boldsymbol{I}_{Md} - \operatorname{blkdiag}\left\{\boldsymbol{a}_1\cdot\mathrm{e}^{-\mathrm{j}(M-1)\mu_1},\ldots,\boldsymbol{a}_d\cdot\mathrm{e}^{-\mathrm{j}(M-1)\mu_d}\right\}\cdot\left(\boldsymbol{I}_d\otimes\boldsymbol{u}^{\mathrm{T}}\cdot\boldsymbol{J}_2\right)\right), &\text{(B.228)}\end{aligned}$$

where we have inserted (B.222) into (B.227). Hence, we can write

$$\left(\boldsymbol{I}_d\otimes\boldsymbol{J}_1^{\mathrm{T}}\right)\cdot\boldsymbol{Z}^{-1}\cdot\bar{\boldsymbol{F}} = \left(\boldsymbol{I}_{Md} - \operatorname{blkdiag}\left\{\boldsymbol{a}_1\cdot\mathrm{e}^{-\mathrm{j}(M-1)\mu_1},\ldots,\boldsymbol{a}_d\cdot\mathrm{e}^{-\mathrm{j}(M-1)\mu_d}\right\}\cdot\left(\boldsymbol{I}_d\otimes\boldsymbol{u}^{\mathrm{T}}\cdot\boldsymbol{J}_2\right)\right),$$

where the term $\left(\boldsymbol{I}_d \otimes \boldsymbol{J}_1^{\mathrm{T}} \cdot \boldsymbol{J}_1\right)$ can be omitted. Then, it is easily shown that

$$\begin{aligned}&\operatorname{blkdiag}\left\{\boldsymbol{P}_{\boldsymbol{a}_1}^{\perp}, \ldots, \boldsymbol{P}_{\boldsymbol{a}_d}^{\perp}\right\}\\&= \operatorname{blkdiag}\left\{\boldsymbol{P}_{\boldsymbol{a}_1}^{\perp}, \ldots, \boldsymbol{P}_{\boldsymbol{a}_d}^{\perp}\right\} \cdot \left(\boldsymbol{I}_{Md} - \operatorname{blkdiag}\left\{\boldsymbol{a}_1 \cdot \mathrm{e}^{-\mathrm{j}(M-1)\mu_1}, \ldots, \boldsymbol{a}_d \cdot \mathrm{e}^{-\mathrm{j}(M-1)\mu_d}\right\} \cdot \left(\boldsymbol{I}_d \otimes \boldsymbol{u}^{\mathrm{T}} \cdot \boldsymbol{J}_2\right)\right).\end{aligned}$$

Analogously, for the first part of the right hand side of (B.226), we obtain

$$\bar{\boldsymbol{F}}^{\mathrm{H}} \cdot \boldsymbol{Z}^{-\mathrm{H}} \cdot (\boldsymbol{I}_d \otimes \boldsymbol{J}_1) \cdot \operatorname{blkdiag}\left\{\boldsymbol{P}_{\boldsymbol{a}_1}^{\perp}, \ldots, \boldsymbol{P}_{\boldsymbol{a}_d}^{\perp}\right\} = \operatorname{blkdiag}\left\{\boldsymbol{P}_{\boldsymbol{a}_1}^{\perp}, \ldots, \boldsymbol{P}_{\boldsymbol{a}_d}^{\perp}\right\}.$$

As a result, the term $\bar{\boldsymbol{F}}^{\mathrm{H}} \cdot \left(\bar{\boldsymbol{F}} \cdot \bar{\boldsymbol{F}}^{\mathrm{H}}\right)^{-1} \cdot \bar{\boldsymbol{F}}$ in (B.212) is given by

$$\bar{\boldsymbol{F}}^{\mathrm{H}} \cdot \left(\bar{\boldsymbol{F}} \cdot \bar{\boldsymbol{F}}^{\mathrm{H}}\right)^{-1} \cdot \bar{\boldsymbol{F}} = \operatorname{blkdiag}\left\{\boldsymbol{P}_{\boldsymbol{a}_1}^{\perp}, \ldots, \boldsymbol{P}_{\boldsymbol{a}_d}^{\perp}\right\} \tag{B.229}$$

and we can express (B.212) as

$$\begin{aligned}&\tilde{\boldsymbol{F}}^{+} \cdot \tilde{\boldsymbol{F}} \cdot (\boldsymbol{I}_d \otimes \boldsymbol{A} \cdot \boldsymbol{A}^{+})\\&= \left(\boldsymbol{\Sigma}_{\mathrm{s}} \cdot \boldsymbol{Q}^{-\mathrm{T}} \otimes \boldsymbol{I}_M\right) \cdot \operatorname{blkdiag}\{\boldsymbol{P}_{\boldsymbol{a}_1}^{\perp}, \ldots, \boldsymbol{P}_{\boldsymbol{a}_d}^{\perp}\} \cdot \left(\boldsymbol{Q}^{\mathrm{T}} \cdot \boldsymbol{\Sigma}_{\mathrm{s}}^{-1} \otimes \boldsymbol{A} \cdot \boldsymbol{A}^{+}\right)\\&= \left(\boldsymbol{\Sigma}_{\mathrm{s}} \cdot \boldsymbol{Q}^{-\mathrm{T}} \otimes \boldsymbol{I}_M\right) \cdot (\boldsymbol{I}_{Md} - \operatorname{blkdiag}\{\boldsymbol{P}_{\boldsymbol{a}_1}, \ldots, \boldsymbol{P}_{\boldsymbol{a}_d}\}) \cdot (\boldsymbol{I}_d \otimes \boldsymbol{A} \cdot \boldsymbol{A}^{+}) \cdot \left(\boldsymbol{Q}^{\mathrm{T}} \cdot \boldsymbol{\Sigma}_{\mathrm{s}}^{-1} \otimes \boldsymbol{I}_M\right) \qquad \text{(B.230)}\\&= \left(\boldsymbol{\Sigma}_{\mathrm{s}} \cdot \boldsymbol{Q}^{-\mathrm{T}} \otimes \boldsymbol{I}_M\right) \cdot ((\boldsymbol{I}_d \otimes \boldsymbol{A} \cdot \boldsymbol{A}^{+}) - \operatorname{blkdiag}\{\boldsymbol{P}_{\boldsymbol{a}_1}, \ldots, \boldsymbol{P}_{\boldsymbol{a}_d}\}) \cdot \left(\boldsymbol{Q}^{\mathrm{T}} \cdot \boldsymbol{\Sigma}_{\mathrm{s}}^{-1} \otimes \boldsymbol{I}_M\right), \qquad \text{(B.231)}\end{aligned}$$

where in (B.230), we have rewritten $\boldsymbol{P}_{\boldsymbol{a}_i}^{\perp}$ as $\boldsymbol{P}_{\boldsymbol{a}_i}^{\perp} = \boldsymbol{I}_M - \boldsymbol{P}_{\boldsymbol{a}_i}$ with $\boldsymbol{P}_{\boldsymbol{a}_i} = \frac{1}{M} \cdot \boldsymbol{a}_i \cdot \boldsymbol{a}_i^{\mathrm{H}} \in \mathbb{C}^{M \times M}$, which is the projection matrix onto the steering vector $\boldsymbol{a}_i$. In the last step, we have used the fact that $\boldsymbol{P}_{\boldsymbol{a}_i} \cdot \boldsymbol{A} \cdot \boldsymbol{A}^{+} = \boldsymbol{P}_{\boldsymbol{a}_i}$ since $\boldsymbol{a}_i$ is contained in $\boldsymbol{A}$.

Next, we apply the same steps to the right hand side of (B.192) to arrive at

$$\begin{aligned}&(\boldsymbol{I}_d \otimes \boldsymbol{A} \cdot \boldsymbol{A}^{+}) \cdot \tilde{\boldsymbol{F}}^{+} \cdot \tilde{\boldsymbol{F}}\\&= \left(\boldsymbol{\Sigma}_{\mathrm{s}} \cdot \boldsymbol{Q}^{-\mathrm{T}} \otimes \boldsymbol{A} \cdot \boldsymbol{A}^{+}\right) \cdot \operatorname{blkdiag}\{\boldsymbol{P}_{\boldsymbol{a}_1}^{\perp}, \ldots, \boldsymbol{P}_{\boldsymbol{a}_d}^{\perp}\} \cdot \left(\boldsymbol{Q}^{\mathrm{T}} \cdot \boldsymbol{\Sigma}_{\mathrm{s}}^{-1} \otimes \boldsymbol{I}_M\right)\\&= \left(\boldsymbol{\Sigma}_{\mathrm{s}} \cdot \boldsymbol{Q}^{-\mathrm{T}} \otimes \boldsymbol{I}_M\right) \cdot (\boldsymbol{I}_d \otimes \boldsymbol{A} \cdot \boldsymbol{A}^{+}) \cdot (\boldsymbol{I}_{Md} - \operatorname{blkdiag}\{\boldsymbol{P}_{\boldsymbol{a}_1}, \ldots, \boldsymbol{P}_{\boldsymbol{a}_d}\}) \cdot \left(\boldsymbol{Q}^{\mathrm{T}} \cdot \boldsymbol{\Sigma}_{\mathrm{s}}^{-1} \otimes \boldsymbol{I}_M\right)\\&= \left(\boldsymbol{\Sigma}_{\mathrm{s}} \cdot \boldsymbol{Q}^{-\mathrm{T}} \otimes \boldsymbol{I}_M\right) \cdot ((\boldsymbol{I}_d \otimes \boldsymbol{A} \cdot \boldsymbol{A}^{+}) - \operatorname{blkdiag}\{\boldsymbol{P}_{\boldsymbol{a}_1}, \ldots, \boldsymbol{P}_{\boldsymbol{a}_d}\}) \cdot \left(\boldsymbol{Q}^{\mathrm{T}} \cdot \boldsymbol{\Sigma}_{\mathrm{s}}^{-1} \otimes \boldsymbol{I}_M\right), \qquad \text{(B.232)}\end{aligned}$$

where we have applied the property that $\boldsymbol{A} \cdot \boldsymbol{A}^{+} \cdot \boldsymbol{P}_{\boldsymbol{a}_i} = \boldsymbol{P}_{\boldsymbol{a}_i}$ holds as well. Finally, we see that (B.232) and (B.231) are equal and therefore, the equalities in (B.192) as well as in (B.189) hold true.

Thus, we have proven that $\boldsymbol{R}_1 \cdot \boldsymbol{R}_0^{-1}$ is a projection matrix. For a projection matrix to be orthogonal, it has to be a Hermitian matrix, i.e., $\boldsymbol{P} = \boldsymbol{P}^{\mathrm{H}}$. It can quickly be verified using (B.186) that this is not the case for $\boldsymbol{R}_1 \cdot \boldsymbol{R}_0^{-1}$, which is therefore a non-orthogonal projection matrix. □

B.11. Proof of Proposition B.9.3

For the proof of Proposition B.9.3, we again verify that property (B.169) for projection matrices is satisfied. Hence, we need to show that

$$\left(\boldsymbol{F}_1^{\mathrm{H}} \cdot \boldsymbol{R}_0^{-1} \cdot \boldsymbol{R} \cdot \boldsymbol{R}_0^{-1} \cdot \boldsymbol{F}_1 \cdot \boldsymbol{G}^{-1}\right)^2 = \boldsymbol{F}_1^{\mathrm{H}} \cdot \boldsymbol{R}_0^{-1} \cdot \boldsymbol{R} \cdot \boldsymbol{R}_0^{-1} \cdot \boldsymbol{F}_1 \cdot \boldsymbol{G}^{-1}, \tag{B.233}$$

where $\boldsymbol{F}_1 = \boldsymbol{I}_d \otimes \boldsymbol{J}_1 \cdot \boldsymbol{U}_{\mathrm{s}}$, $\boldsymbol{G} = \boldsymbol{F}_1^{\mathrm{H}} \cdot \boldsymbol{R}_0^{-1} \cdot \boldsymbol{F}_1$, and the matrices $\boldsymbol{R}$ and $\boldsymbol{R}_0$ are given in (B.180) and (B.181). Using the definition $\tilde{\boldsymbol{F}}$ from (B.183), we can simplify the right hand side of (B.233) as

$$\begin{aligned}\boldsymbol{F}_1^{\mathrm{H}} \cdot \boldsymbol{R}_0^{-1} \cdot \boldsymbol{R} \cdot \boldsymbol{R}_0^{-1} \cdot \boldsymbol{F}_1 \cdot \boldsymbol{G}^{-1} &= \boldsymbol{F}_1^{\mathrm{H}} \cdot \tilde{\boldsymbol{F}}^{+^{\mathrm{H}}} \cdot \left(\boldsymbol{I}_d \otimes \boldsymbol{U}_{\mathrm{n}} \cdot \boldsymbol{U}_{\mathrm{n}}^{\mathrm{H}}\right) \cdot \tilde{\boldsymbol{F}}^{+} \cdot \boldsymbol{F}_1 \cdot \left(\boldsymbol{F}_1^{\mathrm{H}} \cdot \left(\tilde{\boldsymbol{F}} \cdot \tilde{\boldsymbol{F}}^{\mathrm{H}}\right)^{-1} \cdot \boldsymbol{F}_1\right)^{-1} \\ &= \breve{\boldsymbol{F}}^{\mathrm{H}} \cdot \left(\boldsymbol{I}_d \otimes \boldsymbol{U}_{\mathrm{n}} \cdot \boldsymbol{U}_{\mathrm{n}}^{\mathrm{H}}\right) \cdot \breve{\boldsymbol{F}}^{+^{\mathrm{H}}}, \end{aligned} \tag{B.234}$$

where we have defined $\breve{\boldsymbol{F}} = \tilde{\boldsymbol{F}}^{+} \cdot \boldsymbol{F}_1$ and used the fact that $\breve{\boldsymbol{F}}^{+^{\mathrm{H}}} = \tilde{\boldsymbol{F}}^{+} \cdot \boldsymbol{F}_1 \left(\boldsymbol{F}_1^{\mathrm{H}} \cdot \left(\tilde{\boldsymbol{F}} \cdot \tilde{\boldsymbol{F}}^{\mathrm{H}}\right)^{-1} \cdot \boldsymbol{F}_1\right)^{-1}$. Similarly, the left hand side of (B.233) can be written as

$$\begin{aligned}&\left(\boldsymbol{F}_1^{\mathrm{H}} \cdot \boldsymbol{R}_0^{-1} \cdot \boldsymbol{R} \cdot \boldsymbol{R}_0^{-1} \cdot \boldsymbol{F}_1 \cdot \boldsymbol{G}^{-1}\right)^2 \\ &= \boldsymbol{F}_1^{\mathrm{H}} \cdot \tilde{\boldsymbol{F}}^{+^{\mathrm{H}}} \cdot \left(\boldsymbol{I}_d \otimes \boldsymbol{U}_{\mathrm{n}} \cdot \boldsymbol{U}_{\mathrm{n}}^{\mathrm{H}}\right) \cdot \tilde{\boldsymbol{F}}^{+} \cdot \boldsymbol{F}_1 \cdot \left(\boldsymbol{F}_1^{\mathrm{H}} \cdot \left(\tilde{\boldsymbol{F}} \cdot \tilde{\boldsymbol{F}}^{\mathrm{H}}\right)^{-1} \cdot \boldsymbol{F}_1\right)^{-1} \cdot \boldsymbol{F}_1^{\mathrm{H}} \cdot \tilde{\boldsymbol{F}}^{+^{\mathrm{H}}} \\ &\quad \cdot \left(\boldsymbol{I}_d \otimes \boldsymbol{U}_{\mathrm{n}} \cdot \boldsymbol{U}_{\mathrm{n}}^{\mathrm{H}}\right) \cdot \tilde{\boldsymbol{F}}^{+} \cdot \boldsymbol{F}_1 \cdot \left(\boldsymbol{F}_1^{\mathrm{H}} \cdot \left(\tilde{\boldsymbol{F}} \cdot \tilde{\boldsymbol{F}}^{\mathrm{H}}\right)^{-1} \cdot \boldsymbol{F}_1\right)^{-1} \qquad \text{(B.235)} \\ &= \breve{\boldsymbol{F}}^{\mathrm{H}} \cdot \left(\boldsymbol{I}_d \otimes \boldsymbol{U}_{\mathrm{n}} \cdot \boldsymbol{U}_{\mathrm{n}}^{\mathrm{H}}\right) \cdot \breve{\boldsymbol{F}} \cdot \breve{\boldsymbol{F}}^{+} \cdot \left(\boldsymbol{I}_d \otimes \boldsymbol{U}_{\mathrm{n}} \cdot \boldsymbol{U}_{\mathrm{n}}^{\mathrm{H}}\right) \cdot \breve{\boldsymbol{F}}^{+^{\mathrm{H}}}. \qquad \text{(B.236)}\end{aligned}$$

Consequently, the relation in (B.233) to be proven evaluates to

$$\breve{\boldsymbol{F}}^{\mathrm{H}} \cdot \left(\boldsymbol{I}_d \otimes \boldsymbol{U}_{\mathrm{n}} \cdot \boldsymbol{U}_{\mathrm{n}}^{\mathrm{H}}\right) \cdot \breve{\boldsymbol{F}} \cdot \breve{\boldsymbol{F}}^{+} \cdot \left(\boldsymbol{I}_d \otimes \boldsymbol{U}_{\mathrm{n}} \cdot \boldsymbol{U}_{\mathrm{n}}^{\mathrm{H}}\right) \cdot \breve{\boldsymbol{F}}^{+^{\mathrm{H}}} = \breve{\boldsymbol{F}}^{\mathrm{H}} \cdot \left(\boldsymbol{I}_d \otimes \boldsymbol{U}_{\mathrm{n}} \cdot \boldsymbol{U}_{\mathrm{n}}^{\mathrm{H}}\right) \cdot \breve{\boldsymbol{F}}^{+^{\mathrm{H}}}. \tag{B.237}$$

Using the same reasoning as in Appendix B.10, we need to show that the following commutativity property holds:

$$\breve{\boldsymbol{F}} \cdot \breve{\boldsymbol{F}}^{+} \cdot \left(\boldsymbol{I}_d \otimes \boldsymbol{U}_{\mathrm{n}} \cdot \boldsymbol{U}_{\mathrm{n}}^{\mathrm{H}}\right) = \left(\boldsymbol{I}_d \otimes \boldsymbol{U}_{\mathrm{n}} \cdot \boldsymbol{U}_{\mathrm{n}}^{\mathrm{H}}\right) \cdot \breve{\boldsymbol{F}} \cdot \breve{\boldsymbol{F}}^{+}. \tag{B.238}$$

Using the correspondence $\boldsymbol{U}_{\mathrm{n}} \cdot \boldsymbol{U}_{\mathrm{n}}^{\mathrm{H}} = \boldsymbol{I}_M - \boldsymbol{A} \cdot \boldsymbol{A}^{+}$, we expand the left hand side of (B.238) as

$$\breve{\boldsymbol{F}} \cdot \breve{\boldsymbol{F}}^{+} \cdot \left(\boldsymbol{I}_d \otimes \left(\boldsymbol{I}_M - \boldsymbol{A} \cdot \boldsymbol{A}^{+}\right)\right) = \breve{\boldsymbol{F}} \cdot \breve{\boldsymbol{F}}^{+} - \breve{\boldsymbol{F}} \cdot \breve{\boldsymbol{F}}^{+} \cdot \left(\boldsymbol{I}_d \otimes \boldsymbol{A} \cdot \boldsymbol{A}^{+}\right). \tag{B.239}$$

Consequently, (B.238) is equivalent to

$$\breve{\boldsymbol{F}} \cdot \breve{\boldsymbol{F}}^{+} \cdot \left(\boldsymbol{I}_d \otimes \boldsymbol{A} \cdot \boldsymbol{A}^{+}\right) = \left(\boldsymbol{I}_d \otimes \boldsymbol{A} \cdot \boldsymbol{A}^{+}\right) \cdot \breve{\boldsymbol{F}} \cdot \breve{\boldsymbol{F}}^{+}. \tag{B.240}$$

Then, applying the same steps as in (B.192) of Appendix B.10, it can be shown that (B.240) is true. □

B.12. Proof of Theorem 5.4.1

In this appendix, we provide the proof of Theorem 5.4.1 in Section 5.4.1. We start the proof by considering the GLS solution for R-D Standard ESPRIT in (5.54) given by

$$\hat{\boldsymbol{\psi}}_{\mathrm{GLS}}^{(r)} = \left(\hat{\boldsymbol{F}}_1^{(r)\mathrm{H}} \cdot \hat{\boldsymbol{R}}_0^{(r)^{-1}} \cdot \hat{\boldsymbol{F}}_1^{(r)}\right)^{-1} \cdot \hat{\boldsymbol{F}}_1^{(r)\mathrm{H}} \cdot \hat{\boldsymbol{R}}_0^{(r)^{-1}} \cdot \hat{\boldsymbol{b}}^{(r)}, \tag{B.241}$$

where $\hat{\boldsymbol{F}}_1^{(r)} = \boldsymbol{I}_d \otimes \tilde{\boldsymbol{J}}_1^{(r)} \cdot \hat{\boldsymbol{U}}_{\mathrm{s}} \in \mathbb{C}^{\frac{M}{M_r}(M_r-1)d \times d^2}$ and $\hat{\boldsymbol{b}}^{(r)} = \mathrm{vec}\left\{\tilde{\boldsymbol{J}}_2^{(r)} \cdot \hat{\boldsymbol{U}}_{\mathrm{s}}\right\} \in \mathbb{C}^{\frac{M}{M_r}(M_r-1)d \times 1}$. Defining the matrix[2] $\hat{\boldsymbol{K}}^{(r)} = \left(\hat{\boldsymbol{R}}_0^{(r)}\right)^{-1/2} \in \mathbb{C}^{\frac{M}{M_r}(M_r-1)d \times \frac{M}{M_r}(M_r-1)d}$, we can compactly write (B.241) as

$$\hat{\boldsymbol{\psi}}_{\mathrm{GLS}}^{(r)} = \left(\hat{\boldsymbol{K}}^{(r)} \cdot \hat{\boldsymbol{F}}_1^{(r)}\right)^{+} \cdot \hat{\boldsymbol{K}}^{(r)} \cdot \hat{\boldsymbol{b}}^{(r)}, \tag{B.242}$$

which is also the simple least squares solution to the linear equation

$$\hat{\boldsymbol{K}}^{(r)} \cdot \hat{\boldsymbol{F}}_1^{(r)} \cdot \hat{\boldsymbol{\psi}}^{(r)} \approx \hat{\boldsymbol{K}}^{(r)} \cdot \hat{\boldsymbol{b}}^{(r)} \tag{B.243}$$

$$\hat{\boldsymbol{K}}^{(r)} \cdot \left(\boldsymbol{I}_d \otimes \tilde{\boldsymbol{J}}_1^{(r)} \cdot \hat{\boldsymbol{U}}_{\mathrm{s}}\right) \cdot \hat{\boldsymbol{\psi}}^{(r)} \approx \hat{\boldsymbol{K}}^{(r)} \cdot \mathrm{vec}\left\{\tilde{\boldsymbol{J}}_2^{(r)} \cdot \hat{\boldsymbol{U}}_{\mathrm{s}}\right\}, \tag{B.244}$$

where we have inserted the definitions of $\hat{\boldsymbol{F}}_1^{(r)}$ and $\hat{\boldsymbol{b}}^{(r)}$ from above. Then, we express the estimated signal subspace $\hat{\boldsymbol{U}}_{\mathrm{s}}$ as $\hat{\boldsymbol{U}}_{\mathrm{s}} = \boldsymbol{U}_{\mathrm{s}} + \Delta\boldsymbol{U}_{\mathrm{s}}$, where $\Delta\boldsymbol{U}_{\mathrm{s}}$ is the signal subspace estimation error. Moreover, we write $\hat{\boldsymbol{K}}^{(r)} = \boldsymbol{K}^{(r)} + \Delta\boldsymbol{K}^{(r)}$ and $\hat{\boldsymbol{\psi}}^{(r)} = \mathrm{vec}\left\{\hat{\boldsymbol{\Psi}}^{(r)}\right\} = \boldsymbol{\psi}^{(r)} + \Delta\boldsymbol{\psi}^{(r)}$, where $\Delta\boldsymbol{K}^{(r)}$ and $\Delta\boldsymbol{\psi}^{(r)} = \mathrm{vec}\left\{\Delta\boldsymbol{\Psi}^{(r)}\right\}$ are the corresponding error terms. With these expressions, the linear equation in (B.244) can be written as

$$\begin{aligned}&\left(\boldsymbol{K}^{(r)} + \Delta\boldsymbol{K}^{(r)}\right) \cdot \left(\boldsymbol{I}_d \otimes \tilde{\boldsymbol{J}}_1^{(r)} \cdot (\boldsymbol{U}_{\mathrm{s}} + \Delta\boldsymbol{U}_{\mathrm{s}})\right) \cdot \left(\boldsymbol{\psi}^{(r)} + \Delta\boldsymbol{\psi}^{(r)}\right) \\ &\approx \left(\boldsymbol{K}^{(r)} + \Delta\boldsymbol{K}^{(r)}\right) \cdot \mathrm{vec}\left\{\tilde{\boldsymbol{J}}_2^{(r)} \cdot (\boldsymbol{U}_{\mathrm{s}} + \Delta\boldsymbol{U}_{\mathrm{s}})\right\}\end{aligned} \tag{B.245}$$

Expanding (B.245), neglecting the second-order and higher-order terms, and rearranging the terms, we obtain

$$\boldsymbol{K}^{(r)} \cdot \left(\boldsymbol{I}_d \otimes \tilde{\boldsymbol{J}}_1^{(r)} \cdot \boldsymbol{U}_{\mathrm{s}}\right) \cdot \Delta\boldsymbol{\psi}^{(r)} \approx \boldsymbol{K}^{(r)} \cdot \left(\mathrm{vec}\left\{\tilde{\boldsymbol{J}}_2^{(r)} \cdot \Delta\boldsymbol{U}_{\mathrm{s}}\right\} - \left(\boldsymbol{I}_d \otimes \tilde{\boldsymbol{J}}_1^{(r)} \cdot \Delta\boldsymbol{U}_{\mathrm{s}}\right) \cdot \boldsymbol{\psi}^{(r)}\right), \tag{B.246}$$

[2] We define the square root of a matrix $\boldsymbol{X}$ as $\boldsymbol{X}^{1/2} = \boldsymbol{Q} \cdot \boldsymbol{\Lambda}^{1/2} \cdot \boldsymbol{Q}^{-1}$ such that $\boldsymbol{X}^{1/2} \cdot \boldsymbol{X}^{1/2} = \boldsymbol{X}$.

where have used the fact that $\left(\boldsymbol{K}^{(r)} + \Delta\boldsymbol{K}^{(r)}\right) \cdot \left(\left(\boldsymbol{I}_d \otimes \tilde{\boldsymbol{J}}_1^{(r)} \cdot \boldsymbol{U}_\mathrm{s}\right) \cdot \boldsymbol{\psi}^{(r)} - \mathrm{vec}\left\{\tilde{\boldsymbol{J}}_2^{(r)} \cdot \boldsymbol{U}_\mathrm{s}\right\}\right) = \boldsymbol{0}$ since $\left(\boldsymbol{I}_d \otimes \tilde{\boldsymbol{J}}_1^{(r)} \cdot \boldsymbol{U}_\mathrm{s}\right) \cdot \boldsymbol{\psi}^{(r)} = \mathrm{vec}\left\{\tilde{\boldsymbol{J}}_2^{(r)} \cdot \boldsymbol{U}_\mathrm{s}\right\}$ is the vectorized version of the shift invariance equation $\tilde{\boldsymbol{J}}_1^{(r)} \cdot \boldsymbol{U}_\mathrm{s} \cdot \boldsymbol{\Psi}^{(r)} = \tilde{\boldsymbol{J}}_2^{(r)} \cdot \boldsymbol{U}_\mathrm{s}$ in the noiseless case. Then, we can express (B.246) in terms of the parameter estimation error $\Delta\boldsymbol{\psi}^{(r)}$ as

$$\begin{aligned}
\Delta\boldsymbol{\psi}^{(r)} &\approx \left(\boldsymbol{K}^{(r)} \cdot \left(\boldsymbol{I}_d \otimes \tilde{\boldsymbol{J}}_1^{(r)} \cdot \boldsymbol{U}_\mathrm{s}\right)\right)^{+} \cdot \boldsymbol{K}^{(r)} \cdot \left(\mathrm{vec}\left\{\tilde{\boldsymbol{J}}_2^{(r)} \cdot \Delta\boldsymbol{U}_\mathrm{s}\right\} - \mathrm{vec}\left\{\tilde{\boldsymbol{J}}_1^{(r)} \cdot \Delta\boldsymbol{U}_\mathrm{s} \cdot \boldsymbol{\Psi}^{(r)}\right\}\right) && \text{(B.247)}\\
&\approx \left(\boldsymbol{K}^{(r)} \cdot \left(\boldsymbol{I}_d \otimes \tilde{\boldsymbol{J}}_1^{(r)} \cdot \boldsymbol{U}_\mathrm{s}\right)\right)^{+} \cdot \boldsymbol{K}^{(r)} \cdot \left(\left(\boldsymbol{I}_d \otimes \tilde{\boldsymbol{J}}_2^{(r)}\right) \cdot \Delta\boldsymbol{u}_\mathrm{s} - \left(\boldsymbol{\Psi}^{(r)^\mathrm{T}} \otimes \tilde{\boldsymbol{J}}_1^{(r)}\right) \cdot \Delta\boldsymbol{u}_\mathrm{s}\right)\\
&\approx \left(\boldsymbol{K}^{(r)} \cdot \left(\boldsymbol{I}_d \otimes \tilde{\boldsymbol{J}}_1^{(r)} \cdot \boldsymbol{U}_\mathrm{s}\right)\right)^{+} \cdot \boldsymbol{K}^{(r)} \cdot \left(\boldsymbol{I}_d \otimes \tilde{\boldsymbol{J}}_2^{(r)} - \boldsymbol{\Psi}^{(r)^\mathrm{T}} \otimes \tilde{\boldsymbol{J}}_1^{(r)}\right) \cdot \Delta\boldsymbol{u}_\mathrm{s}\\
&\approx -\left(\boldsymbol{K}^{(r)} \cdot \boldsymbol{F}_1^{(r)}\right)^{+} \cdot \boldsymbol{K}^{(r)} \cdot \boldsymbol{F}_2^{(r)} \cdot \Delta\boldsymbol{u}_\mathrm{s}, && \text{(B.248)}
\end{aligned}$$

where we have applied property (1.14) to the last terms of (B.246) and (B.247), respectively. Moreover, in the last step, we have used the definitions of $\boldsymbol{F}_1^{(r)}$ in (B.241) and $\boldsymbol{F}_2^{(r)}$ in (5.54), and $\Delta\boldsymbol{u}_\mathrm{s} = \mathrm{vec}\{\Delta\boldsymbol{U}_\mathrm{s}\}$.

Recall from (B.11) in Appendix B.2 that the parameter estimation error $\Delta\mu_i^{(r)}$ can be expressed as

$$\Delta\mu_i^{(r)} \approx \mathrm{Im}\left\{\boldsymbol{p}_i^\mathrm{T} \cdot \Delta\boldsymbol{\Psi}^{(r)} \cdot \boldsymbol{q}_i / \lambda_i^{(r)}\right\} = \mathrm{Im}\left\{\left(\boldsymbol{q}_i^\mathrm{T} \otimes \boldsymbol{p}_i^\mathrm{T}\right) \cdot \Delta\boldsymbol{\psi}^{(r)} / \lambda_i^{(r)}\right\}, \tag{B.249}$$

where we have again applied property (1.14) to obtain the last equality. Finally, inserting (B.248) into (B.249), yields

$$\Delta\mu_i^{(r)} \approx \mathrm{Im}\left\{\boldsymbol{r}_{i,\mathrm{GLS}}^{(r)^\mathrm{T}} \cdot \Delta\boldsymbol{u}_\mathrm{s}\right\} = \mathrm{Im}\left\{\boldsymbol{r}_{i,\mathrm{GLS}}^{(r)^\mathrm{T}} \cdot \boldsymbol{W}_\mathrm{mat} \cdot \boldsymbol{n}\right\}, \tag{B.250}$$

where $\boldsymbol{W}_\mathrm{mat}$ is given as in (4.15) by

$$\boldsymbol{W}_\mathrm{mat} = \left(\boldsymbol{\Sigma}_\mathrm{s}^{-1} \cdot \boldsymbol{V}_\mathrm{s}^\mathrm{T} \otimes \boldsymbol{U}_\mathrm{n} \cdot \boldsymbol{U}_\mathrm{n}^\mathrm{H}\right) \tag{B.251}$$

and $\boldsymbol{r}_{i,\mathrm{GLS}}^{(r)^\mathrm{T}}$ can be computed as

$$\begin{aligned}
\boldsymbol{r}_{i,\mathrm{GLS}}^{(r)^\mathrm{T}} &= -\frac{1}{\lambda_i^{(r)}} \cdot \left(\boldsymbol{q}_i^\mathrm{T} \otimes \boldsymbol{p}_i^\mathrm{T}\right) \cdot \left(\boldsymbol{K}^{(r)} \cdot \boldsymbol{F}_1^{(r)}\right)^{+} \cdot \boldsymbol{K}^{(r)} \cdot \boldsymbol{F}_2^{(r)} && \text{(B.252)}\\
&= -\frac{1}{\lambda_i^{(r)}} \cdot \left(\boldsymbol{q}_i^\mathrm{T} \otimes \boldsymbol{p}_i^\mathrm{T}\right) \cdot \left(\left(\hat{\boldsymbol{R}}_0^{(r)}\right)^{-1/2} \cdot \boldsymbol{F}_1^{(r)}\right)^{+} \cdot \left(\hat{\boldsymbol{R}}_0^{(r)}\right)^{-1/2} \cdot \boldsymbol{F}_2^{(r)} && \text{(B.253)}\\
&= -\frac{1}{\lambda_i^{(r)}} \cdot \left(\boldsymbol{q}_i^\mathrm{T} \otimes \boldsymbol{p}_i^\mathrm{T}\right) \cdot \left(\boldsymbol{F}_1^{(r)^\mathrm{H}} \cdot \boldsymbol{R}_0^{(r)^{-1}} \cdot \boldsymbol{F}_1^{(r)}\right)^{-1} \cdot \boldsymbol{F}_1^{(r)^\mathrm{H}} \cdot \boldsymbol{R}_0^{(r)^{-1}} \cdot \boldsymbol{F}_2^{(r)}, && \text{(B.254)}
\end{aligned}$$

where we have inserted the definition $\hat{\boldsymbol{K}}^{(r)} = \left(\hat{\boldsymbol{R}}_0^{(r)}\right)^{-1/2}$ in line (B.252) and expanded the pseudo-inverse in line (B.253). This completes the proof. □

B.13. Proof of Theorem 5.4.2

For the proof of Theorem 5.4.2 in Section 5.4.2, we only present the derivation for the 1-D case, but the steps extend to the R-D case straightforwardly. The estimated parameters after the real-valued transformation for GLS-based Unitary ESPRIT are extracted via the arctangent function, which is different from the complex-valued case of GLS-based Standard ESPRIT with FBA. Hence, we develop a first-order perturbation expansion for the real-valued GLS-based Unitary ESPRIT algorithm and then show its equivalence to the complex-valued GLS-based Standard ESPRIT algorithm with FBA.

Recall that the SVD of $\boldsymbol{X}_0^{(\mathrm{fba})}$ in (5.34) can be expressed as

$$\boldsymbol{X}_0^{(\mathrm{fba})} = \begin{bmatrix} \boldsymbol{U}_{\mathrm{s}}^{(\mathrm{fba})} & \boldsymbol{U}_{\mathrm{n}}^{(\mathrm{fba})} \end{bmatrix} \cdot \begin{bmatrix} \boldsymbol{\Sigma}_{\mathrm{s}}^{(\mathrm{fba})} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} \end{bmatrix} \cdot \begin{bmatrix} \boldsymbol{V}_{\mathrm{s}}^{(\mathrm{fba})} & \boldsymbol{V}_{\mathrm{n}}^{(\mathrm{fba})} \end{bmatrix}^{\mathrm{H}}, \tag{B.255}$$

The complex-valued 1-D shift invariance equation after applying FBA has the form

$$\boldsymbol{J}_1 \cdot \boldsymbol{U}_{\mathrm{s}}^{(\mathrm{fba})} \cdot \boldsymbol{\Psi} = \boldsymbol{J}_2 \cdot \boldsymbol{U}_{\mathrm{s}}^{(\mathrm{fba})}, \tag{B.256}$$

where $\boldsymbol{\Psi} = \boldsymbol{Q}^{(\mathrm{fba})} \cdot \boldsymbol{\Lambda} \cdot \boldsymbol{Q}^{(\mathrm{fba})^{-1}}$ and $\boldsymbol{\Lambda} = \mathrm{diag}\left\{\begin{bmatrix}\lambda_1, \ldots, \lambda_d\end{bmatrix}\right\}$ with $\lambda_i = \mathrm{e}^{\mathrm{j}\mu_i}$, $i = 1, 2, \ldots, d$. For the GLS-based R-D Standard ESPRIT algorithm with FBA, the first-order approximation of the parameter estimation error is given in (5.68) (using (5.69)) as

$$\Delta\mu_i \approx \mathrm{Im}\left\{-\frac{1}{\lambda_i} \cdot \left(\boldsymbol{q}_i^{(\mathrm{fba})^{\mathrm{T}}} \otimes \boldsymbol{p}_i^{(\mathrm{fba})^{\mathrm{T}}}\right) \cdot \left(\boldsymbol{F}_1^{(\mathrm{fba})^{\mathrm{H}}} \cdot \boldsymbol{R}_0^{(\mathrm{fba})^{-1}} \cdot \boldsymbol{F}_1^{(\mathrm{fba})}\right)^{-1} \cdot \boldsymbol{F}_1^{(\mathrm{fba})^{\mathrm{H}}} \cdot \boldsymbol{R}_0^{(\mathrm{fba})^{-1}} \cdot \boldsymbol{F}_2^{(\mathrm{fba})} \cdot \Delta\boldsymbol{u}_{\mathrm{s}}^{(\mathrm{fba})}\right\}, \tag{B.257}$$

where we recall the definitions

$$\boldsymbol{F}_1^{(\mathrm{fba})} = \left(\boldsymbol{I}_d \otimes \boldsymbol{J}_1 \cdot \boldsymbol{U}_{\mathrm{s}}^{(\mathrm{fba})}\right) \tag{B.258}$$

$$\boldsymbol{F}_2^{(\mathrm{fba})} = \left(\boldsymbol{\Psi}^{\mathrm{T}} \otimes \boldsymbol{J}_1\right) - \left(\boldsymbol{I}_d \otimes \boldsymbol{J}_2\right) \tag{B.259}$$

$$\boldsymbol{R}_0^{(\mathrm{fba})} = \boldsymbol{F}_2^{(\mathrm{fba})} \cdot \left(\boldsymbol{\Sigma}_{\mathrm{s}}^{-2} \otimes \boldsymbol{I}_M\right) \cdot \boldsymbol{F}_2^{(\mathrm{fba})^{\mathrm{H}}}. \tag{B.260}$$

Moreover, the perturbation $\Delta\boldsymbol{u}_{\mathrm{s}}^{(\mathrm{fba})} = \mathrm{vec}\left\{\Delta\boldsymbol{U}_{\mathrm{s}}^{(\mathrm{fba})}\right\}$ is given by

$$\Delta\boldsymbol{u}_{\mathrm{s}} \approx \left(\boldsymbol{\Sigma}_{\mathrm{s}}^{(\mathrm{fba})^{-1}} \cdot \boldsymbol{V}_{\mathrm{s}}^{(\mathrm{fba})^{\mathrm{T}}} \otimes \boldsymbol{U}_{\mathrm{n}}^{(\mathrm{fba})} \cdot \boldsymbol{U}_{\mathrm{n}}^{(\mathrm{fba})^{\mathrm{H}}}\right) \cdot \boldsymbol{n}^{(\mathrm{fba})}, \tag{B.261}$$

where $\Delta\boldsymbol{U}_{\mathrm{s}}^{(\mathrm{fba})} = \boldsymbol{U}_{\mathrm{n}}^{(\mathrm{fba})} \cdot \boldsymbol{U}_{\mathrm{n}}^{(\mathrm{fba})^{\mathrm{H}}} \cdot \boldsymbol{N}^{(\mathrm{fba})} \cdot \boldsymbol{V}_{\mathrm{s}}^{(\mathrm{fba})} \cdot \boldsymbol{\Sigma}_{\mathrm{s}}^{(\mathrm{fba})^{-1}}$. Next, we show that the estimation error expansion for the real-valued GLS-based Unitary ESPRIT algorithm is equivalent to (B.257).

The SVD of $\varphi\left(\boldsymbol{X}_0^{(\mathrm{fba})}\right)$ after the real-valued transformation is given in (5.38). Then, the 1-D real-valued shift-invariance equation is given by

$$\boldsymbol{K}_1 \cdot \boldsymbol{E}_{\mathrm{s}} \cdot \boldsymbol{\Upsilon} = \boldsymbol{K}_2 \cdot \boldsymbol{E}_{\mathrm{s}}, \tag{B.262}$$

where $\boldsymbol{\Upsilon} = \boldsymbol{Q}^{(\varphi)} \cdot \boldsymbol{\Omega} \cdot \boldsymbol{Q}^{(\varphi)^{-1}}$ and $\boldsymbol{\Omega} = \mathrm{diag}\left\{\left[\omega_1, \ldots, \omega_d\right]\right\}$ with $\omega_i = \tan(\mu_i/2)$, $i = 1, \ldots, d$. As (B.262) has the same algebraic form as its complex-valued counterpart in (B.256), the same procedure from [LLV93] can be applied to develop a first-order perturbation expansion. In fact, following the three steps discussed in [LLV93], we find that the perturbation of ω_i in terms of $\boldsymbol{\Upsilon}$ and the perturbation of $\boldsymbol{\Upsilon}$ in terms of the signal subspace estimation error $\Delta\boldsymbol{E}_{\mathrm{s}}$ lead to the same result, where $\boldsymbol{J}_1, \boldsymbol{J}_2, \boldsymbol{U}_{\mathrm{s}}^{(\mathrm{fba})}$, and $\boldsymbol{\Psi}$ are consistently exchanged by $\boldsymbol{K}_1, \boldsymbol{K}_2, \boldsymbol{E}_{\mathrm{s}}$, and $\boldsymbol{\Upsilon}$, respectively. Thus, only the perturbation of μ_i in terms of $\omega_i = \tan(\mu_i/2)$ is to be derived. Therefore, we compute the Taylor series expansion of ω_i, which is given by

$$\begin{aligned} \omega_i + \Delta\omega &\approx \tan(\mu_i/2) + \Delta\mu \cdot \left(\frac{\tan^2(\mu_i/2)}{2} + \frac{1}{2}\right) = \omega_i + \Delta\mu \cdot \frac{\omega_i^2 + 1}{2} \\ \Delta\mu &\approx \Delta\omega \cdot \frac{2}{\omega_i^2 + 1}. \end{aligned} \tag{B.263}$$

Combining (B.263) with the corresponding real-valued expressions for the perturbations of ω_i and $\boldsymbol{\Upsilon}$, we obtain

$$\Delta\mu_i \approx -\frac{2}{\omega_i^2 + 1} \cdot \left(\boldsymbol{q}_i^{(\varphi)^{\mathrm{T}}} \otimes \boldsymbol{p}_i^{(\varphi)^{\mathrm{T}}}\right) \cdot \left(\boldsymbol{F}_1^{(\varphi)^{\mathrm{H}}} \cdot \boldsymbol{R}_0^{(\varphi)^{-1}} \cdot \boldsymbol{F}_1^{(\varphi)}\right)^{-1} \cdot \boldsymbol{F}_1^{(\varphi)^{\mathrm{H}}} \cdot \boldsymbol{R}_0^{(\varphi)^{-1}} \cdot \boldsymbol{F}_2^{(\varphi)} \cdot \Delta\boldsymbol{e}_{\mathrm{s}}, \tag{B.264}$$

where $\boldsymbol{q}_i^{(\varphi)}$ is the i-th column of $\boldsymbol{Q}^{(\varphi)}$ and $\boldsymbol{p}_i^{(\varphi)^{\mathrm{T}}}$ is the i-th row of $\boldsymbol{Q}^{(\varphi)^{-1}}$, and we have the definitions (cf. Section 5.3.1.2)

$$\boldsymbol{F}_1^{(\varphi)} = (\boldsymbol{I}_d \otimes \boldsymbol{K}_1 \cdot \boldsymbol{E}_{\mathrm{s}}) \tag{B.265}$$

$$\boldsymbol{F}_2^{(\varphi)} = \left(\boldsymbol{\Upsilon}^{\mathrm{T}} \otimes \boldsymbol{K}_1\right) - (\boldsymbol{I}_d \otimes \boldsymbol{K}_2) \tag{B.266}$$

$$\boldsymbol{R}_0^{(\varphi)} = \boldsymbol{F}_2^{(\varphi)} \cdot \left(\boldsymbol{\Sigma}_{\mathrm{s}}^{(\varphi)^{-2}} \otimes \boldsymbol{I}_M\right) \cdot \boldsymbol{F}_2^{(\varphi)^{\mathrm{H}}}. \tag{B.267}$$

Moreover, the perturbation $\Delta\boldsymbol{e}_{\mathrm{s}} = \mathrm{vec}\{\Delta\boldsymbol{E}_{\mathrm{s}}\}$ is given in (5.45) by

$$\Delta\boldsymbol{e}_{\mathrm{s}} \approx \left(\boldsymbol{\Sigma}_{\mathrm{s}}^{(\varphi)^{-1}} \cdot \boldsymbol{W}_{\mathrm{s}}^{\mathrm{T}} \cdot \boldsymbol{Q}_{2N}^{\mathrm{T}} \otimes \boldsymbol{E}_{\mathrm{n}} \cdot \boldsymbol{E}_{\mathrm{n}}^{\mathrm{H}} \cdot \boldsymbol{Q}_M^{\mathrm{H}}\right) \cdot \boldsymbol{n}^{(\mathrm{fba})}. \tag{B.268}$$

In order to show the equivalence of (B.264) and (B.257), we still require a number of identities

and tools. First, comparing the SVD of $\varphi\left(\boldsymbol{X}_0^{(\mathrm{fba})}\right) = \boldsymbol{Q}_M^{\mathrm{H}} \cdot \boldsymbol{X}_0^{(\mathrm{fba})} \cdot \boldsymbol{Q}_{2N}$ in (5.38) with the SVD of $\boldsymbol{X}_0^{(\mathrm{fba})}$ in (B.255) and using the fact that the matrices $\boldsymbol{Q}_p$ are unitary, the subspaces of $\varphi\left(\boldsymbol{X}_0^{(\mathrm{fba})}\right)$ are also given by choosing

$$\begin{aligned} \boldsymbol{E}_{\mathrm{s}} &= \boldsymbol{Q}_M^{\mathrm{H}} \cdot \boldsymbol{U}_{\mathrm{s}}^{(\mathrm{fba})}, \quad \boldsymbol{E}_{\mathrm{n}} = \boldsymbol{Q}_M^{\mathrm{H}} \cdot \boldsymbol{U}_{\mathrm{n}}^{(\mathrm{fba})}, \quad \boldsymbol{\Sigma}_{\mathrm{s}}{}^{(\varphi)} = \boldsymbol{\Sigma}_{\mathrm{s}}^{(\mathrm{fba})} \\ \boldsymbol{W}_{\mathrm{s}} &= \boldsymbol{Q}_{2N}^{\mathrm{H}} \cdot \boldsymbol{V}_{\mathrm{s}}^{(\mathrm{fba})}, \quad \boldsymbol{W}_{\mathrm{n}} = \boldsymbol{Q}_{2N}^{\mathrm{H}} \cdot \boldsymbol{V}_{\mathrm{n}}^{(\mathrm{fba})}. \end{aligned} \tag{B.269}$$

Second, the transformed selection matrices $\boldsymbol{K}_1$ and $\boldsymbol{K}_2$ defined in (5.40) can be reformulated as

$$\boldsymbol{K}_1 = 2 \cdot \mathrm{Re}\left\{\boldsymbol{Q}_{M^{(\mathrm{sel})}}^{\mathrm{H}} \cdot \boldsymbol{J}_2 \cdot \boldsymbol{Q}_M\right\} = \boldsymbol{Q}_{M^{(\mathrm{sel})}}^{\mathrm{H}} \cdot (\boldsymbol{J}_1 + \boldsymbol{J}_2) \cdot \boldsymbol{Q}_M \tag{B.270}$$

$$\boldsymbol{K}_2 = 2 \cdot \mathrm{Im}\left\{\boldsymbol{Q}_{M^{(\mathrm{sel})}}^{\mathrm{H}} \cdot \boldsymbol{J}_2 \cdot \boldsymbol{Q}_M\right\} = \mathrm{j} \cdot \boldsymbol{Q}_{M^{(\mathrm{sel})}}^{\mathrm{H}} \cdot (\boldsymbol{J}_1 - \boldsymbol{J}_2) \cdot \boldsymbol{Q}_M, \tag{B.271}$$

which follows from expanding the real part and the imaginary part according to $2 \cdot \mathrm{Re}\{x\} = x + x^*$ and $2 \cdot \mathrm{Im}\{x\} = -\mathrm{j}x + \mathrm{j}x^*$. The conjugated term $\boldsymbol{Q}_{M^{(\mathrm{sel})}}^{\mathrm{T}} \cdot \boldsymbol{J}_2 \cdot \boldsymbol{Q}^*$ can be simplified into $\boldsymbol{Q}_{M^{(\mathrm{sel})}}^{\mathrm{H}} \cdot \boldsymbol{J}_1 \cdot \boldsymbol{Q}$ using the fact that $\boldsymbol{J}_1 = \boldsymbol{\Pi}_{M^{(\mathrm{sel})}} \cdot \boldsymbol{J}_2 \cdot \boldsymbol{\Pi}_M$ holds since the array must be centro-symmetric for FBA to be applicable and the fact that $\boldsymbol{Q}_p$ is left-$\boldsymbol{\Pi}$-real.

Additionally, we require the following two lemmas:

Lemma B.13.1. *The following identities are satisfied*

$$(\boldsymbol{J}_1 + \boldsymbol{J}_2) \cdot \boldsymbol{U}_{\mathrm{s}}^{(\mathrm{fba})} = \boldsymbol{J}_1 \cdot \boldsymbol{U}_{\mathrm{s}}^{(\mathrm{fba})} \cdot \breve{\boldsymbol{\Psi}} \tag{B.272}$$

$$(\boldsymbol{J}_1 - \boldsymbol{J}_2) \cdot \boldsymbol{U}_{\mathrm{s}}^{(\mathrm{fba})} = \boldsymbol{J}_2 \cdot \boldsymbol{U}_{\mathrm{s}}^{(\mathrm{fba})} \cdot \mathring{\boldsymbol{\Psi}}, \tag{B.273}$$

where $\breve{\boldsymbol{\Psi}} = \boldsymbol{I}_d + \boldsymbol{\Psi} = \boldsymbol{Q}^{(\mathrm{fba})} \cdot (\boldsymbol{I}_d + \boldsymbol{\Lambda}) \cdot \boldsymbol{Q}^{(\mathrm{fba})^{-1}}$ *and* $\mathring{\boldsymbol{\Psi}} = -\boldsymbol{I}_d + \boldsymbol{\Psi}^{-1} = \boldsymbol{Q}^{(\mathrm{fba})} \cdot \left(-\boldsymbol{I}_d + \boldsymbol{\Lambda}^{-1}\right) \cdot \boldsymbol{Q}^{(\mathrm{fba})^{-1}}$.

Proof: These identities follow straightforwardly from $\boldsymbol{J}_1 \cdot \boldsymbol{U}_{\mathrm{s}}^{(\mathrm{fba})} \cdot \boldsymbol{\Psi} = \boldsymbol{J}_2 \cdot \boldsymbol{U}_{\mathrm{s}}^{(\mathrm{fba})}$ by adding $\boldsymbol{J}_1 \cdot \boldsymbol{U}_{\mathrm{s}}^{(\mathrm{fba})}$ to both sides of the equation for the first identity, and subtracting $\boldsymbol{J}_1 \cdot \boldsymbol{U}_{\mathrm{s}}^{(\mathrm{fba})}$ and substituting $\boldsymbol{J}_1 \cdot \boldsymbol{U}_{\mathrm{s}}^{(\mathrm{fba})}$ by $\boldsymbol{J}_2 \cdot \boldsymbol{U}_{\mathrm{s}}^{(\mathrm{fba})} \cdot \boldsymbol{\Psi}^{-1}$ for the second identity.

Lemma B.13.2. *In the noiseless case, the solution* $\boldsymbol{\Psi}$ *to* (B.256) *and the solution* $\boldsymbol{\Upsilon}$ *to* (B.262) *have the same eigenvectors, i.e.,* $\boldsymbol{Q}^{(\mathrm{fba})} = \boldsymbol{Q}^{(\varphi)}$. *Moreover, their eigenvalues are related as* $\omega_i = \mathrm{j} \cdot \frac{1-\lambda_i}{1+\lambda_i}$.

Proof: Starting from $\boldsymbol{\Upsilon} = (\boldsymbol{K}_1 \cdot \boldsymbol{E}_{\mathrm{s}})^+ \cdot \boldsymbol{K}_2 \cdot \boldsymbol{E}_{\mathrm{s}}$ and replacing $\boldsymbol{E}_{\mathrm{s}}$ with (B.269) and $\boldsymbol{K}_n^{(\mathrm{nc})}$ with (B.270) and (B.271), we get

$$\begin{aligned} \boldsymbol{\Upsilon} &= \left(\left(\boldsymbol{J}_1 + \boldsymbol{J}_2\right) \cdot \boldsymbol{U}_{\mathrm{s}}^{(\mathrm{fba})}\right)^+ \cdot \mathrm{j} \cdot \left(\boldsymbol{J}_1 - \boldsymbol{J}_2\right) \cdot \boldsymbol{U}_{\mathrm{s}}^{(\mathrm{fba})} \\ &= \mathrm{j} \cdot \breve{\boldsymbol{\Psi}}^{-1} \cdot \boldsymbol{\Psi} \cdot \mathring{\boldsymbol{\Psi}} = \mathrm{j} \cdot \boldsymbol{Q}^{(\mathrm{fba})} \cdot (\boldsymbol{I}_d + \boldsymbol{\Lambda})^{-1} (\boldsymbol{I}_d - \boldsymbol{\Lambda}) \cdot \boldsymbol{Q}^{(\mathrm{fba})^{-1}} \\ &= \boldsymbol{Q}^{(\mathrm{fba})} \cdot \boldsymbol{\Omega} \cdot \boldsymbol{Q}^{(\mathrm{fba})^{-1}}, \end{aligned} \tag{B.274}$$

where $\boldsymbol{\Omega} = \operatorname{diag}\left\{\mathrm{j}\cdot\left[\frac{1-\lambda_i}{1+\lambda_i}\right]\right\}_{i=1}^{d}$ and we have used Lemma B.13.1 in the first step.

Now, we are equipped with the tools to proof the equivalence of (B.264) and (B.257). To this end, we reformulate the terms $\Delta\boldsymbol{e}_{\mathrm{s}}$, $\boldsymbol{F}_1^{(\varphi)}$, $\boldsymbol{F}_2^{(\varphi)}$, and $\boldsymbol{R}_0^{(\varphi)}$ in (B.268) and (B.265)-(B.267).
For $\Delta\boldsymbol{e}_{\mathrm{s}}$ in (B.268), we apply (B.269) and easily obtain

$$\Delta\boldsymbol{e}_{\mathrm{s}} = (\boldsymbol{I}_d \otimes \boldsymbol{Q}_M^{\mathrm{H}}) \cdot \Delta\boldsymbol{u}_{\mathrm{s}}^{(\mathrm{fba})}, \tag{B.275}$$

where $\Delta\boldsymbol{u}_{\mathrm{s}}^{(\mathrm{fba})}$ is given in (B.261). For $\boldsymbol{F}_1^{(\varphi)}$ in (B.265), after using (B.270) and (B.272), we get

$$\begin{aligned}
\boldsymbol{F}_1^{(\varphi)} &= \left(\boldsymbol{I}_d \otimes \boldsymbol{Q}_{M^{(\mathrm{sel})}}^{\mathrm{H}} \cdot (\boldsymbol{J}_1 + \boldsymbol{J}_2) \cdot \boldsymbol{U}_{\mathrm{s}}^{(\mathrm{fba})}\right) \\
&= \left(\boldsymbol{I}_d \otimes \boldsymbol{Q}_{M^{(\mathrm{sel})}}^{\mathrm{H}}\right) \cdot \left(\boldsymbol{I}_d \otimes \boldsymbol{J}_1 \cdot \boldsymbol{U}_s^{(\mathrm{fba})} \cdot \breve{\boldsymbol{\Psi}}\right) \\
&= \left(\boldsymbol{I}_d \otimes \boldsymbol{Q}_{M^{(\mathrm{sel})}}^{\mathrm{H}}\right) \cdot \boldsymbol{F}_1^{(\mathrm{fba})} \cdot \left(\boldsymbol{I}_d \otimes \breve{\boldsymbol{\Psi}}\right),
\end{aligned} \tag{B.276}$$

where $\boldsymbol{F}_1^{(\mathrm{fba})}$ is given in (B.258). For $\boldsymbol{F}_2^{(\varphi)}$ in (B.266), we have

$$\begin{aligned}
\boldsymbol{F}_2^{(\varphi)} &= \left(\boldsymbol{\Upsilon}^{\mathrm{T}} \otimes \boldsymbol{K}_1\right) - (\boldsymbol{I}_d \otimes \boldsymbol{K}_2) \\
&= \left(\boldsymbol{\Upsilon}^{\mathrm{T}} \otimes \boldsymbol{Q}_{M^{(\mathrm{sel})}}^{\mathrm{H}} \cdot (\boldsymbol{J}_1 + \boldsymbol{J}_2) \cdot \boldsymbol{Q}_M\right) - \left(\boldsymbol{I}_d \otimes \mathrm{j}\cdot \boldsymbol{Q}_{M^{(\mathrm{sel})}}^{\mathrm{H}} \cdot (\boldsymbol{J}_1 - \boldsymbol{J}_2) \cdot \boldsymbol{Q}_M\right) \\
&= \left(\boldsymbol{I}_d \otimes \boldsymbol{Q}_{M^{(\mathrm{sel})}}^{\mathrm{H}}\right) \cdot \left(\boldsymbol{\Upsilon}^{\mathrm{T}} \otimes (\boldsymbol{J}_1 + \boldsymbol{J}_2) - \boldsymbol{I}_d \otimes \mathrm{j}\cdot (\boldsymbol{J}_1 - \boldsymbol{J}_2)\right) \cdot (\boldsymbol{I}_d \otimes \boldsymbol{Q}_M) \\
&= \left(\boldsymbol{Q}^{(\mathrm{fba})^{-\mathrm{T}}} \otimes \boldsymbol{Q}_{M^{(\mathrm{sel})}}^{\mathrm{H}}\right) \cdot (\boldsymbol{\Omega} \otimes (\boldsymbol{J}_1 + \boldsymbol{J}_2) - \boldsymbol{I}_d \otimes \mathrm{j}\cdot (\boldsymbol{J}_1 - \boldsymbol{J}_2)) \cdot \left(\boldsymbol{Q}^{(\mathrm{fba})^{\mathrm{T}}} \otimes \boldsymbol{Q}_M\right),
\end{aligned} \tag{B.277}$$

where we have applied (B.270) and (B.271) as well as $\boldsymbol{\Upsilon} = \boldsymbol{Q}^{(\mathrm{fba})} \cdot \boldsymbol{\Omega} \cdot \boldsymbol{Q}^{(\mathrm{fba})^{-1}}$. Next, we consider the middle term $(\boldsymbol{\Omega} \otimes (\boldsymbol{J}_1 + \boldsymbol{J}_2) - \boldsymbol{I}_d \otimes \mathrm{j}\cdot(\boldsymbol{J}_1 - \boldsymbol{J}_2))$ of (B.277), which can be expressed as

$$\begin{aligned}
&\boldsymbol{\Omega} \otimes (\boldsymbol{J}_1 + \boldsymbol{J}_2) - \boldsymbol{I}_d \otimes \mathrm{j}\cdot(\boldsymbol{J}_1 - \boldsymbol{J}_2) \\
&= \mathrm{j}\cdot(\boldsymbol{I}_d + \boldsymbol{\Lambda})^{-1} \cdot (\boldsymbol{I}_d - \boldsymbol{\Lambda}) \otimes (\boldsymbol{J}_1 + \boldsymbol{J}_2) - \boldsymbol{I}_d \otimes \mathrm{j}\cdot(\boldsymbol{J}_1 - \boldsymbol{J}_2) \\
&= \mathrm{j}\cdot\left(\left((\boldsymbol{I}_d + \boldsymbol{\Lambda})^{-1} \cdot (\boldsymbol{I}_d - \boldsymbol{\Lambda}) - \boldsymbol{I}_d\right) \otimes \boldsymbol{J}_1 + \left((\boldsymbol{I}_d + \boldsymbol{\Lambda})^{-1} \cdot (\boldsymbol{I}_d - \boldsymbol{\Lambda}) + \boldsymbol{I}_d\right) \otimes \boldsymbol{J}_2\right) \\
&= \mathrm{j}\cdot\left(\left((\boldsymbol{I}_d + \boldsymbol{\Lambda})^{-1} \cdot (-2) \cdot \boldsymbol{\Lambda}\right) \otimes \boldsymbol{J}_1 + \left((\boldsymbol{I}_d + \boldsymbol{\Lambda})^{-1} \cdot 2 \cdot \boldsymbol{I}_d\right) \otimes \boldsymbol{J}_2\right) \\
&= -\mathrm{j}\cdot 2 \cdot \left((\boldsymbol{I}_d + \boldsymbol{\Lambda})^{-1} \otimes \boldsymbol{I}_{M-1}\right) \cdot (\boldsymbol{\Lambda} \otimes \boldsymbol{J}_1 - \boldsymbol{I}_d \otimes \boldsymbol{J}_2)
\end{aligned} \tag{B.278}$$

Using this expression, $\boldsymbol{F}_2^{(\varphi)}$ can be written as

$$\begin{aligned}
\boldsymbol{F}_2^{(\varphi)} &= -\mathrm{j}\cdot 2 \cdot \left(\boldsymbol{Q}^{(\mathrm{fba})^{-\mathrm{T}}} \otimes \boldsymbol{Q}_{M^{(\mathrm{sel})}}^{\mathrm{H}}\right) \cdot \left((\boldsymbol{I}_d + \boldsymbol{\Lambda})^{-1} \otimes \boldsymbol{I}_{M-1}\right) \cdot (\boldsymbol{\Lambda} \otimes \boldsymbol{J}_1 - \boldsymbol{I}_d \otimes \boldsymbol{J}_2) \cdot \left(\boldsymbol{Q}^{(\mathrm{fba})^{\mathrm{T}}} \otimes \boldsymbol{Q}_M\right) \\
&= -\mathrm{j}\cdot 2 \cdot \left(\boldsymbol{Q}^{(\mathrm{fba})^{-\mathrm{T}}} \cdot (\boldsymbol{I}_d + \boldsymbol{\Lambda})^{-1} \otimes \boldsymbol{Q}_{M^{(\mathrm{sel})}}^{\mathrm{H}}\right) \cdot (\boldsymbol{\Lambda} \otimes \boldsymbol{J}_1 - \boldsymbol{I}_d \otimes \boldsymbol{J}_2) \cdot \left(\boldsymbol{Q}^{(\mathrm{fba})^{\mathrm{T}}} \otimes \boldsymbol{Q}_M\right) \\
&= -\mathrm{j}\cdot 2 \cdot \left(\boldsymbol{Q}^{(\mathrm{fba})^{-\mathrm{T}}} \cdot (\boldsymbol{I}_d + \boldsymbol{\Lambda})^{-1} \cdot \boldsymbol{Q}^{(\mathrm{fba})^{\mathrm{T}}} \otimes \boldsymbol{Q}_{M^{(\mathrm{sel})}}^{\mathrm{H}}\right) \cdot \left(\boldsymbol{\Psi}^{\mathrm{T}} \otimes \boldsymbol{J}_1 - \boldsymbol{I}_d \otimes \boldsymbol{J}_2\right) \cdot (\boldsymbol{I}_d \otimes \boldsymbol{Q}_M) \\
&= -\mathrm{j}\cdot 2 \cdot \left(\breve{\boldsymbol{\Psi}}^{-\mathrm{T}} \otimes \boldsymbol{Q}_{M^{(\mathrm{sel})}}^{\mathrm{H}}\right) \cdot \boldsymbol{F}_2^{(\mathrm{fba})} \cdot (\boldsymbol{I}_d \otimes \boldsymbol{Q}_M),
\end{aligned} \tag{B.279}$$

where $\boldsymbol{F}_2^{(\mathrm{fba})}$ is given in (B.259). Using (B.279), we can express $\boldsymbol{R}_0^{(\varphi)}$ in (B.267) as

$$\begin{aligned}\boldsymbol{R}_0^{(\varphi)} &= 4 \cdot \left(\breve{\boldsymbol{\Psi}}^{-\mathrm{T}} \otimes \boldsymbol{Q}_{M^{(\mathrm{sel})}}^{\mathrm{H}}\right) \cdot \boldsymbol{F}_2^{(\mathrm{fba})} \cdot \left(\boldsymbol{\Sigma}_{\mathrm{s}}^{-2} \otimes \boldsymbol{I}_M\right) \cdot \boldsymbol{F}_2^{(\mathrm{fba})^{\mathrm{H}}} \cdot \left(\breve{\boldsymbol{\Psi}}^{-\mathrm{T}} \otimes \boldsymbol{Q}_{M^{(\mathrm{sel})}}^{\mathrm{H}}\right)^{\mathrm{H}} \\ &= 4 \cdot \left(\breve{\boldsymbol{\Psi}}^{-\mathrm{T}} \otimes \boldsymbol{Q}_{M^{(\mathrm{sel})}}^{\mathrm{H}}\right) \cdot \boldsymbol{R}_0^{(\mathrm{fba})} \cdot \left(\breve{\boldsymbol{\Psi}}^{-*} \otimes \boldsymbol{Q}_{M^{(\mathrm{sel})}}\right). \end{aligned} \tag{B.280}$$

There, $\boldsymbol{R}_0^{(\varphi)^{-1}}$ is obtained as

$$\boldsymbol{R}_0^{(\varphi)^{-1}} = \frac{1}{4} \cdot \left(\breve{\boldsymbol{\Psi}}^{-*} \otimes \boldsymbol{Q}_{M^{(\mathrm{sel})}}\right)^{-1} \cdot \boldsymbol{R}_0^{(\mathrm{fba})^{-1}} \cdot \left(\breve{\boldsymbol{\Psi}}^{-\mathrm{T}} \otimes \boldsymbol{Q}_{M^{(\mathrm{sel})}}^{\mathrm{H}}\right)^{-1}. \tag{B.281}$$

Next, we use these results and compute the terms of (B.264). For the term $\boldsymbol{F}_1^{(\varphi)^{\mathrm{H}}} \cdot \boldsymbol{R}_0^{(\varphi)^{-1}} \cdot \boldsymbol{F}_1^{(\varphi)}$, we have

$$\begin{aligned}\boldsymbol{F}_1^{(\varphi)^{\mathrm{H}}} \boldsymbol{R}_0^{(\varphi)^{-1}} \boldsymbol{F}_1^{(\varphi)} &= \frac{1}{4} \cdot \left(\boldsymbol{I}_d \otimes \breve{\boldsymbol{\Psi}}\right)^{\mathrm{H}} \boldsymbol{F}_1^{(\mathrm{fba})^{\mathrm{H}}} \left(\breve{\boldsymbol{\Psi}}^{*} \otimes \boldsymbol{I}_{M-1}\right) \boldsymbol{R}_0^{(\mathrm{fba})^{-1}} \left(\breve{\boldsymbol{\Psi}}^{\mathrm{T}} \otimes \boldsymbol{I}_{M-1}\right) \boldsymbol{F}_1^{(\mathrm{fba})} \left(\boldsymbol{I}_d \otimes \breve{\boldsymbol{\Psi}}\right) \\ &= \frac{1}{4} \cdot \left(\breve{\boldsymbol{\Psi}}^{\mathrm{T}} \otimes \breve{\boldsymbol{\Psi}}\right)^{\mathrm{H}} \cdot \boldsymbol{F}_1^{(\mathrm{fba})^{\mathrm{H}}} \cdot \boldsymbol{R}_0^{(\mathrm{fba})^{-1}} \cdot \boldsymbol{F}_1^{(\mathrm{fba})} \cdot \left(\breve{\boldsymbol{\Psi}}^{\mathrm{T}} \otimes \breve{\boldsymbol{\Psi}}\right). \end{aligned} \tag{B.282}$$

Then, the inverse term $\left(\boldsymbol{F}_1^{(\varphi)^{\mathrm{H}}} \cdot \boldsymbol{R}_0^{(\varphi)^{-1}} \cdot \boldsymbol{F}_1^{(\varphi)}\right)^{-1}$ can be computed as

$$\left(\boldsymbol{F}_1^{(\varphi)^{\mathrm{H}}} \cdot \boldsymbol{R}_0^{(\varphi)^{-1}} \cdot \boldsymbol{F}_1^{(\varphi)}\right)^{-1} = 4 \cdot \left(\breve{\boldsymbol{\Psi}}^{\mathrm{T}} \otimes \breve{\boldsymbol{\Psi}}\right)^{-1} \cdot \left(\boldsymbol{F}_1^{(\mathrm{fba})^{\mathrm{H}}} \cdot \boldsymbol{R}_0^{(\mathrm{fba})^{-1}} \cdot \boldsymbol{F}_1^{(\mathrm{fba})}\right)^{-1} \cdot \left(\breve{\boldsymbol{\Psi}}^{\mathrm{T}} \otimes \breve{\boldsymbol{\Psi}}\right)^{-\mathrm{H}}. \tag{B.283}$$

Inserting (B.276), (B.281), and (B.283), we express the term $\left(\boldsymbol{F}_1^{(\varphi)^{\mathrm{H}}} \cdot \boldsymbol{R}_0^{(\varphi)^{-1}} \cdot \boldsymbol{F}_1^{(\varphi)}\right)^{-1} \cdot \boldsymbol{F}_1^{(\varphi)^{\mathrm{H}}} \cdot \boldsymbol{R}_0^{(\varphi)^{-1}}$ as

$$\begin{aligned}&\left(\boldsymbol{F}_1^{(\varphi)^{\mathrm{H}}} \cdot \boldsymbol{R}_0^{(\varphi)^{-1}} \cdot \boldsymbol{F}_1^{(\varphi)}\right)^{-1} \cdot \boldsymbol{F}_1^{(\varphi)^{\mathrm{H}}} \cdot \boldsymbol{R}_0^{(\varphi)^{-1}} \\ &= \left(\breve{\boldsymbol{\Psi}}^{\mathrm{T}} \otimes \breve{\boldsymbol{\Psi}}\right)^{-1} \cdot \left(\boldsymbol{F}_1^{(\mathrm{fba})^{\mathrm{H}}} \cdot \boldsymbol{R}_0^{(\mathrm{fba})^{-1}} \cdot \boldsymbol{F}_1^{(\mathrm{fba})}\right)^{-1} \cdot \left(\breve{\boldsymbol{\Psi}}^{\mathrm{T}} \otimes \breve{\boldsymbol{\Psi}}\right)^{-\mathrm{H}} \\ &\quad \cdot \left(\breve{\boldsymbol{\Psi}}^{\mathrm{T}} \otimes \breve{\boldsymbol{\Psi}}\right)^{\mathrm{H}} \cdot \boldsymbol{F}_1^{(\mathrm{fba})^{\mathrm{H}}} \cdot \boldsymbol{R}_0^{(\mathrm{fba})^{-1}} \cdot \left(\breve{\boldsymbol{\Psi}}^{-\mathrm{T}} \otimes \boldsymbol{Q}_{M^{(\mathrm{sel})}}^{\mathrm{H}}\right)^{-1} \\ &= \left(\breve{\boldsymbol{\Psi}}^{\mathrm{T}} \otimes \breve{\boldsymbol{\Psi}}\right)^{-1} \cdot \left(\boldsymbol{F}_1^{(\mathrm{fba})^{\mathrm{H}}} \cdot \boldsymbol{R}_0^{(\mathrm{fba})^{-1}} \cdot \boldsymbol{F}_1^{(\mathrm{fba})}\right)^{-1} \cdot \boldsymbol{F}_1^{(\mathrm{fba})^{\mathrm{H}}} \cdot \boldsymbol{R}_0^{(\mathrm{fba})^{-1}} \cdot \left(\breve{\boldsymbol{\Psi}}^{-\mathrm{T}} \otimes \boldsymbol{Q}_{M^{(\mathrm{sel})}}^{\mathrm{H}}\right)^{-1}. \end{aligned} \tag{B.284}$$

Upon combining (B.284) into (B.264), we can write $\Delta\mu_i$ as

$$\Delta\mu_i \approx \frac{\mathrm{j} \cdot 4}{\omega_i^2 + 1} \cdot \left(\boldsymbol{q}_i^{(\varphi)^{\mathrm{T}}} \breve{\boldsymbol{\Psi}}^{-\mathrm{T}} \otimes \boldsymbol{p}_i^{(\varphi)^{\mathrm{T}}} \breve{\boldsymbol{\Psi}}^{-1}\right) \left(\boldsymbol{F}_1^{(\mathrm{fba})^{\mathrm{H}}} \boldsymbol{R}_0^{(\mathrm{fba})^{-1}} \boldsymbol{F}_1^{(\mathrm{fba})}\right)^{-1} \boldsymbol{F}_1^{(\mathrm{fba})^{\mathrm{H}}} \boldsymbol{R}_0^{(\mathrm{fba})^{-1}} \boldsymbol{F}_2^{(\mathrm{fba})} \Delta\boldsymbol{u}_{\mathrm{s}}^{(\mathrm{fba})}. \tag{B.285}$$

Moreover, the term $\frac{4}{\omega_i^2+1}$ can be expressed in terms of λ_i via Lemma B.13.2. We obtain

$$\begin{aligned}\frac{4}{\omega_i^2+1} &= \frac{4}{\left(\mathrm{j}\frac{1-\lambda_i}{1+\lambda_i}\right)^2+1} = \frac{4\cdot(\lambda_i+1)^2}{-(\lambda_i-1)^2+(\lambda_i+1)^2} \\ &= \frac{4\cdot(\lambda_i+1)^2}{4\cdot\lambda_i} = \frac{(\lambda_i+1)^2}{\lambda_i}. \end{aligned} \tag{B.286}$$

Furthermore, the terms $\boldsymbol{q}_i^{(\varphi)^{\mathrm{T}}}\cdot\breve{\boldsymbol{\Psi}}^{-\mathrm{T}}$ and $\boldsymbol{p}_i^{(\varphi)^{\mathrm{T}}}\cdot\breve{\boldsymbol{\Psi}}^{-1}$ can be simplified by substituting $\boldsymbol{p}_i^{(\varphi)}=\boldsymbol{p}_i^{(\mathrm{fba})}$ and $\boldsymbol{q}_i^{(\varphi)}=\boldsymbol{q}_i^{(\mathrm{fba})}$ using Lemma B.13.2 and writing

$$\begin{aligned}\boldsymbol{p}_i^{(\mathrm{fba})^{\mathrm{T}}}\cdot\breve{\boldsymbol{\Psi}}^{-1} &= \boldsymbol{p}_i^{(\mathrm{fba})^{\mathrm{T}}}\cdot(1+\lambda_i)^{-1} \\ \boldsymbol{q}_i^{(\mathrm{fba})^{\mathrm{T}}}\cdot\breve{\boldsymbol{\Psi}}^{-\mathrm{T}} &= \boldsymbol{q}_i^{(\mathrm{fba})^{\mathrm{T}}}\cdot(1+\lambda_i)^{-1}, \end{aligned} \tag{B.287}$$

which follows from the eigendecomposition $\breve{\boldsymbol{\Psi}}=\boldsymbol{Q}^{(\mathrm{fba})}\cdot(\boldsymbol{I}_d+\boldsymbol{\Lambda})\cdot\boldsymbol{P}^{(\mathrm{fba})}$, where $\boldsymbol{P}^{(\mathrm{fba})}=\boldsymbol{Q}^{(\mathrm{fba})^{-1}}$ such that $\breve{\boldsymbol{\Psi}}^{-1}=\boldsymbol{Q}^{(\mathrm{fba})}\cdot(\boldsymbol{I}_d+\boldsymbol{\Lambda})^{-1}\cdot\boldsymbol{P}^{(\mathrm{fba})}$ and $\breve{\boldsymbol{\Psi}}^{-\mathrm{T}}=\boldsymbol{P}^{(\mathrm{fba})^{\mathrm{T}}}\cdot(\boldsymbol{I}_d+\boldsymbol{\Lambda})^{-1}\cdot\boldsymbol{Q}^{(\mathrm{fba})^{\mathrm{T}}}$.

Finally, we obtain

$$\Delta\mu_i \approx \mathrm{j}\cdot\frac{1}{\lambda_i}\cdot\left(\boldsymbol{q}_i^{(\mathrm{fba})^{\mathrm{T}}}\otimes\boldsymbol{p}_i^{(\mathrm{fba})^{\mathrm{T}}}\right)\left(\boldsymbol{F}_1^{(\mathrm{fba})^{\mathrm{H}}}\boldsymbol{R}_0^{(\mathrm{fba})^{-1}}\boldsymbol{F}_1^{(\mathrm{fba})}\right)^{-1}\boldsymbol{F}_1^{(\mathrm{fba})^{\mathrm{H}}}\boldsymbol{R}_0^{(\mathrm{fba})^{-1}}\boldsymbol{F}_2^{(\mathrm{fba})}\Delta\boldsymbol{u}_{\mathrm{s}}^{(\mathrm{fba})}, \tag{B.288}$$

where in the final step, we notice that (B.288) must be real-valued as we have started from the real-valued expansion (B.264) and only used equivalence transforms to arrive at (B.288). However, if $\mathrm{j}z\in\mathbb{R}$ for $z\in\mathbb{C}$ this implies that $\mathrm{Re}\{z\}=0$ and hence $\mathrm{j}z=\mathrm{Im}\{-z\}$. Consequently, (B.288) can also be written as (B.257), which concludes the proof of the theorem. □

B.14. Proof of Theorem 5.5.1

This theorem from Section 5.5.1 consists of several parts, which are addressed in separate subsections.

B.14.1. 1-D Standard ESPRIT with GLS

We start by simplifying the analytical MSE expression for 1-D Standard ESPRIT with GLS in (5.67) for circularly symmetric white noise. In the case of a single source, the noise-free measurement matrix is given by

$$\boldsymbol{X}_0=\boldsymbol{a}(\mu)\cdot\boldsymbol{s}^{\mathrm{T}}, \tag{B.289}$$

where $\boldsymbol{a} \in \mathbb{C}^{M\times 1}$ is the array steering vector and $\boldsymbol{s} \in \mathbb{C}^{N\times 1}$ is the source symbol vector. Let $\hat{P} = \|\boldsymbol{s}\|_2^2/N$ be the empirical source power. Furthermore, since we assume a uniform linear array (ULA) of isotropic elements, $\boldsymbol{a}(\mu)$ is given by $\boldsymbol{a}(\mu) = [1, \mathrm{e}^{\mathrm{j}\mu}, \mathrm{e}^{2\mathrm{j}\mu}, \ldots, \mathrm{e}^{(M-1)\mathrm{j}\mu}]$. Note that $\|\boldsymbol{a}(\mu)\|_2^2 = M$. For notational convenience, we drop the explicit dependence of $\boldsymbol{a}$ on μ and write $\boldsymbol{a}(\mu) = \boldsymbol{a}$ in what follows. The selection matrices $\boldsymbol{J}_1$ and $\boldsymbol{J}_2$ are then chosen as

$$\boldsymbol{J}_1 = \begin{bmatrix} \boldsymbol{I}_{(M-1)} & \boldsymbol{0}_{(M-1)\times 1} \end{bmatrix} \quad \boldsymbol{J}_2 = \begin{bmatrix} \boldsymbol{0}_{(M-1)\times 1} & \boldsymbol{I}_{M-1} \end{bmatrix} \tag{B.290}$$

for maximum overlap, i.e., $M^{(\mathrm{sel})} = M - 1$. Since (B.289) is a rank-one matrix, we can directly write the subspaces in terms of the array steering vector and the source symbol vector as

$$\boldsymbol{U}_\mathrm{s} = \boldsymbol{u}_\mathrm{s} = \frac{\boldsymbol{a}}{\|\boldsymbol{a}\|_2} = \frac{1}{\sqrt{M}} \cdot \boldsymbol{a} \tag{B.291}$$

$$\boldsymbol{V}_\mathrm{s} = \boldsymbol{v}_\mathrm{s} = \frac{\boldsymbol{s}^*}{\|\boldsymbol{s}\|_2} = \frac{1}{\sqrt{\hat{P} \cdot N}} \cdot \boldsymbol{s}^* \tag{B.292}$$

$$\boldsymbol{\Sigma}_\mathrm{s} = \sigma_\mathrm{s} = \sqrt{M \cdot N \cdot \hat{P}}. \tag{B.293}$$

For the MSE expression in (5.67), we also require $\boldsymbol{U}_\mathrm{n} \cdot \boldsymbol{U}_\mathrm{n}^\mathrm{H}$, which is a projection matrix on the noise subspace. However, since the signal subspace is spanned by $\boldsymbol{a}$ we can write $\boldsymbol{U}_\mathrm{n} \cdot \boldsymbol{U}_\mathrm{n}^\mathrm{H} = \boldsymbol{P}_{\boldsymbol{a}}^{\perp} = \boldsymbol{I}_M - \frac{1}{\|\boldsymbol{a}\|_2^2} \cdot \boldsymbol{a} \cdot \boldsymbol{a}^\mathrm{H} = \boldsymbol{I}_M - \frac{1}{M} \cdot \boldsymbol{a} \cdot \boldsymbol{a}^\mathrm{H}$. The MSE expression for 1-D Standard ESPRIT with GLS also includes the eigenvectors of $\boldsymbol{\Psi}$ denoted by $\boldsymbol{p}_i$ and $\boldsymbol{q}_i$. However, for the special case discussed here, $\boldsymbol{\Psi}$ is scalar and given by $\boldsymbol{\Psi} = \boldsymbol{\Phi} = \mathrm{e}^{\mathrm{j}\mu}$. Consequently, we have $\boldsymbol{p}_i = \boldsymbol{q}_i = 1$ for the eigenvectors.

Combining these expressions into the 1-D version of the analytical MSE expression (5.67) for circularly symmetric white noise, we obtain for $d = 1$

$$\mathbb{E}\left\{(\Delta\mu)^2\right\} = \frac{\sigma_\mathrm{n}^2}{2} \cdot \|\boldsymbol{z}_\mathrm{GLS}\|_2^2, \tag{B.294}$$

where $\boldsymbol{z}_\mathrm{GLS} = \boldsymbol{W}_\mathrm{mat}^\mathrm{T} \cdot \boldsymbol{r}_\mathrm{GLS}$. In the 1-D case, the vector $\boldsymbol{r}_\mathrm{GLS}$ in (5.65) simplifies to

$$\boldsymbol{r}_\mathrm{GLS}^\mathrm{T} = -\frac{1}{\lambda} \cdot \left(\boldsymbol{f}_1^\mathrm{H} \cdot \boldsymbol{R}_0^{-1} \cdot \boldsymbol{f}_1\right)^{-1} \cdot \boldsymbol{f}_1^\mathrm{H} \cdot \boldsymbol{R}_0^{-1} \cdot \boldsymbol{F}_2, \tag{B.295}$$

where the expressions for $\boldsymbol{F}_1$, $\boldsymbol{F}_2$, and $\boldsymbol{R}_0$ in (5.65) reduce to

$$\boldsymbol{F}_1 = \boldsymbol{f}_1 = \boldsymbol{J}_1 \cdot \frac{\boldsymbol{a}}{\sqrt{M}}, \quad \boldsymbol{F}_2 = \mathrm{e}^{\mathrm{j}\mu} \cdot \boldsymbol{J}_1 - \boldsymbol{J}_2, \quad \boldsymbol{R}_0 = \frac{1}{\sigma_\mathrm{s}^2} \cdot \boldsymbol{F}_2 \cdot \boldsymbol{F}_2^\mathrm{H} \tag{B.296}$$

Moreover, the matrix $\boldsymbol{W}_\mathrm{mat}$ in (5.64) simplifies to

$$\boldsymbol{W}_\mathrm{mat} = \left(\frac{1}{\sqrt{\hat{P}MN}} \cdot \frac{\boldsymbol{s}^\mathrm{H}}{\sqrt{\hat{P}N}}\right) \otimes \boldsymbol{P}_{\boldsymbol{a}}^{\perp}. \tag{B.297}$$

Note that we can write $\boldsymbol{z}_{\mathrm{GLS}}^{\mathrm{T}}$ as $\boldsymbol{z}_{\mathrm{GLS}}^{\mathrm{T}} = \boldsymbol{r}_{\mathrm{GLS}}^{\mathrm{T}} \cdot \boldsymbol{W}_{\mathrm{mat}} = \tilde{\boldsymbol{s}}^{\mathrm{T}} \otimes \tilde{\boldsymbol{a}}^{\mathrm{T}}$, where

$$\tilde{\boldsymbol{s}}^{\mathrm{T}} = \frac{1}{\sqrt{\hat{P}MN}} \cdot \frac{\boldsymbol{s}^{\mathrm{H}}}{\sqrt{\hat{P}N}} \tag{B.298}$$

$$\tilde{\boldsymbol{a}}^{\mathrm{T}} = -\frac{1}{\mathrm{e}^{\mathrm{j}\mu}} \cdot \left(\boldsymbol{f}_1^{\mathrm{H}} \cdot \boldsymbol{R}_0^{-1} \cdot \boldsymbol{f}_1\right)^{-1} \cdot \boldsymbol{f}_1^{\mathrm{H}} \cdot \boldsymbol{R}_0^{-1} \cdot \boldsymbol{F}_2 \cdot \boldsymbol{P}_{\boldsymbol{a}}^{\perp}. \tag{B.299}$$

Expanding $\boldsymbol{P}_{\boldsymbol{a}}^{\perp}$ in (B.299), we can simplify $\tilde{\boldsymbol{a}}^{\mathrm{T}}$ as

$$\begin{aligned}\tilde{\boldsymbol{a}}^{\mathrm{T}} &= -\frac{1}{\mathrm{e}^{\mathrm{j}\mu}} \cdot \left(\boldsymbol{f}_1^{\mathrm{H}} \cdot \boldsymbol{R}_0^{-1} \cdot \boldsymbol{f}_1\right)^{-1} \cdot \boldsymbol{f}_1^{\mathrm{H}} \cdot \boldsymbol{R}_0^{-1} \cdot \boldsymbol{F}_2 \cdot \left(\boldsymbol{I}_M - \frac{1}{M} \cdot \boldsymbol{a} \cdot \boldsymbol{a}^{\mathrm{H}}\right) \\ &= -\frac{1}{\mathrm{e}^{\mathrm{j}\mu}} \cdot \left(\boldsymbol{f}_1^{\mathrm{H}} \cdot \boldsymbol{R}_0^{-1} \cdot \boldsymbol{f}_1\right)^{-1} \cdot \boldsymbol{f}_1^{\mathrm{H}} \cdot \boldsymbol{R}_0^{-1} \cdot \boldsymbol{F}_2,\end{aligned} \tag{B.300}$$

where in the last step we have used the fact that $\boldsymbol{F}_2 \cdot \boldsymbol{a} = (\mathrm{e}^{\mathrm{j}\mu} \cdot \boldsymbol{J}_1 - \boldsymbol{J}_2) \cdot \boldsymbol{a} = \boldsymbol{0}$, which is the shift invariance equation.

Using (B.298) and (B.300), the MSE in (B.294) is given by

$$\mathbb{E}\left\{(\Delta\mu)^2\right\} = \frac{\sigma_{\mathrm{n}}^2}{2} \cdot \|\tilde{\boldsymbol{s}}^{\mathrm{T}} \otimes \tilde{\boldsymbol{a}}^{\mathrm{T}}\|_2^2 = \frac{\sigma_{\mathrm{n}}^2}{2} \cdot \|\tilde{\boldsymbol{s}}^{\mathrm{T}}\|_2^2 \cdot \|\tilde{\boldsymbol{a}}^{\mathrm{T}}\|_2^2, \tag{B.301}$$

where we have applied property (1.7). In order to compute $\|\tilde{\boldsymbol{s}}^{\mathrm{T}}\|_2^2$ in (B.301), we notice that $\tilde{\boldsymbol{s}}^{\mathrm{T}}$ is scaled version of $\boldsymbol{s}^{\mathrm{H}}$ and $\|\boldsymbol{s}^{\mathrm{H}}\|_2^2 = \hat{P}N$. Thus, we can simplify $\|\tilde{\boldsymbol{s}}^{\mathrm{T}}\|_2^2$ as

$$\|\tilde{\boldsymbol{s}}^{\mathrm{T}}\|_2^2 = \frac{1}{\hat{P}MN} \cdot \frac{\hat{P}N}{\hat{P}N} = \frac{1}{\hat{P}MN}. \tag{B.302}$$

Furthermore, expanding $\|\tilde{\boldsymbol{a}}^{\mathrm{T}}\|_2^2$ in (B.301) using (B.300) yields

$$\|\tilde{\boldsymbol{a}}^{\mathrm{T}}\|_2^2 = \left(\boldsymbol{f}_1^{\mathrm{H}} \cdot \boldsymbol{R}_0^{-1} \cdot \boldsymbol{f}_1\right)^{-1} \cdot \boldsymbol{f}_1^{\mathrm{H}} \cdot \boldsymbol{R}_0^{-1} \cdot \boldsymbol{F}_2 \cdot \boldsymbol{F}_2^{\mathrm{H}} \cdot \boldsymbol{R}_0^{-1} \cdot \boldsymbol{f}_1 \cdot \left(\boldsymbol{f}_1^{\mathrm{H}} \cdot \boldsymbol{R}_0^{-1} \cdot \boldsymbol{f}_1\right)^{-1} \tag{B.303}$$

$$= \sigma_{\mathrm{s}}^2 \left(\boldsymbol{f}_1^{\mathrm{H}} \cdot \boldsymbol{R}_0^{-1} \cdot \boldsymbol{f}_1\right)^{-1} \cdot \boldsymbol{f}_1^{\mathrm{H}} \cdot \boldsymbol{R}_0^{-1} \cdot \boldsymbol{f}_1 \cdot \left(\boldsymbol{f}_1^{\mathrm{H}} \cdot \boldsymbol{R}_0^{-1} \cdot \boldsymbol{f}_1\right)^{-1} \tag{B.304}$$

$$= \sigma_{\mathrm{s}}^2 \cdot \left(\boldsymbol{f}_1^{\mathrm{H}} \cdot \boldsymbol{R}_0^{-1} \cdot \boldsymbol{f}_1\right)^{-1} \tag{B.305}$$

$$= M \cdot \left(\boldsymbol{a}^{\mathrm{H}} \cdot \boldsymbol{J}_1^{\mathrm{H}} \cdot (\boldsymbol{F}_2 \cdot \boldsymbol{F}_2^{\mathrm{H}})^{-1} \cdot \boldsymbol{J}_1 \cdot \boldsymbol{a}\right)^{-1} \tag{B.306}$$

$$= M \cdot \gamma^{-1}, \tag{B.307}$$

where we have defined γ as

$$\gamma = \boldsymbol{a}^{\mathrm{H}} \cdot \boldsymbol{J}_1^{\mathrm{H}} \cdot (\boldsymbol{F}_2 \cdot \boldsymbol{F}_2^{\mathrm{H}})^{-1} \cdot \boldsymbol{J}_1 \cdot \boldsymbol{a} \tag{B.308}$$

and in (B.303) and (B.305) we have used the definition of $\boldsymbol{R}_0$ in (B.296). Consequently, the MSE

of 1-D Standard ESPRIT with GLS in (B.301) is given by

$$\begin{aligned}\mathbb{E}\left\{(\Delta\mu)^2\right\} &= \frac{\sigma_\mathrm{n}^2}{2}\cdot\|\tilde{\boldsymbol{s}}^\mathrm{T}\|_2^2\cdot\|\tilde{\boldsymbol{a}}^\mathrm{T}\|_2^2 \\ &= \frac{\sigma_\mathrm{n}^2}{2}\cdot\frac{1}{\hat{P}N}\cdot\gamma^{-1} = \frac{1}{2\hat{\varrho}}\cdot\gamma^{-1},\end{aligned} \tag{B.309}$$

where $\hat{\varrho} = N\hat{P}/\sigma_\mathrm{n}^2$ is the effective SNR.

Next, we simplify the deterministic Cramér-Rao Bound (CRB) from [SN89] for the special case of a single source. The deterministic CRB is introduced in Section 9.2.1. The closed-form expression for the deterministic CRB assuming Gaussian distributed noise is given in (9.6) as [SN89]

$$\boldsymbol{C} = \frac{\sigma_\mathrm{n}^2}{2\cdot N}\cdot\mathrm{Re}\left\{\left[\boldsymbol{D}^\mathrm{H}\cdot\left(\boldsymbol{I}_M - \boldsymbol{A}\cdot\left(\boldsymbol{A}^\mathrm{H}\cdot\boldsymbol{A}\right)^{-1}\cdot\boldsymbol{A}^\mathrm{H}\right)\cdot\boldsymbol{D}\right]\odot\hat{\boldsymbol{R}}_\mathrm{S}^\mathrm{T}\right\}^{-1}, \tag{B.310}$$

where $\hat{\boldsymbol{R}}_\mathrm{S} = \frac{1}{N}\cdot\boldsymbol{S}\cdot\boldsymbol{S}^\mathrm{H} \in \mathbb{C}^{d\times d}$ is the sample covariance matrix of the symbol matrix $\boldsymbol{S} \in \mathbb{C}^{d\times N}$ and $\boldsymbol{D} \in \mathbb{C}^{M\times d}$ is the matrix of partial derivatives of the array steering vectors $\boldsymbol{a}(\mu_i) \in \mathbb{C}^{M\times 1}$ for $i = 1,\ldots,d$ contained in $\boldsymbol{A} \in \mathbb{C}^{M\times d}$ with respect to the parameters of interest μ_i. In the case $d = 1$ considered in Section 9.4.1, we have $\hat{\boldsymbol{R}}_\mathrm{S} = \|\boldsymbol{s}\|_2^2/N = \hat{P}$ and the CRB expression in (B.310) simplifies into

$$C = \frac{\sigma_\mathrm{n}^2}{2N\hat{P}}\cdot\mathrm{Re}\left\{\boldsymbol{d}^\mathrm{H}\cdot\left(\boldsymbol{I}_M - \frac{\boldsymbol{a}\cdot\boldsymbol{a}^\mathrm{H}}{M}\right)\cdot\boldsymbol{d}\right\}^{-1} = \frac{1}{2\hat{\varrho}}\cdot\left[\boldsymbol{d}^\mathrm{H}\cdot\left(\boldsymbol{I}_M - \frac{\boldsymbol{a}\cdot\boldsymbol{a}^\mathrm{H}}{M}\right)\cdot\boldsymbol{d}\right]^{-1}. \tag{B.311}$$

For the derivatives contained in the vector $\boldsymbol{d}$, we obtain

$$\boldsymbol{d} = \frac{\partial\boldsymbol{a}}{\partial\mu} = \mathrm{j}\cdot\mathrm{diag}\left\{\boldsymbol{c}\right\}\cdot\boldsymbol{a}, \tag{B.312}$$

where $\boldsymbol{c} = \begin{bmatrix}0 & 1 & 2 & \ldots & M-1\end{bmatrix}^\mathrm{T} \in \mathbb{R}^{M\times 1}$. Then, (B.311) becomes

$$C = \frac{1}{2\hat{\varrho}}\cdot\left[\boldsymbol{a}^\mathrm{H}\cdot\mathrm{diag}\left\{\boldsymbol{c}\right\}\cdot\left(\boldsymbol{I}_M - \frac{\boldsymbol{a}\cdot\boldsymbol{a}^\mathrm{H}}{M}\right)\cdot\mathrm{diag}\left\{\boldsymbol{c}\right\}\cdot\boldsymbol{a}\right]^{-1}. \tag{B.313}$$

In order to prove that the MSE expression of Standard ESPRIT with GLS for a single source is equivalent to the CRB expression for a single source, we compare[3] (B.313) to (B.309). Conse-

[3] Although the assumption of the noise to be Gaussian distributed is not required for the GLS-based Standard ESPRIT algorithm, we need to make this assumption for the comparison to the CRB, which is derived for Gaussian noise.

quently, we have to show that the following relation holds:

$$\gamma = \boldsymbol{a}^{\mathrm{H}} \cdot \operatorname{diag}\{\boldsymbol{c}\} \cdot \left(\boldsymbol{I}_M - \frac{\boldsymbol{a} \cdot \boldsymbol{a}^{\mathrm{H}}}{M} \right) \cdot \operatorname{diag}\{\boldsymbol{c}\} \cdot \boldsymbol{a}. \tag{B.314}$$

To do so, we start with the expression $\gamma = \boldsymbol{a}^{\mathrm{H}} \cdot \boldsymbol{J}_1^{\mathrm{H}} \cdot (\boldsymbol{F}_2 \cdot \boldsymbol{F}_2^{\mathrm{H}})^{-1} \cdot \boldsymbol{J}_1 \cdot \boldsymbol{a}$ in (B.308) and first consider the term $(\boldsymbol{F}_2 \cdot \boldsymbol{F}_2^{\mathrm{H}})^{-1}$, where $\boldsymbol{F}_2$ is given in (B.296) as $\boldsymbol{F}_2 = \mathrm{e}^{\mathrm{j}\mu} \cdot \boldsymbol{J}_1 - \boldsymbol{J}_2$. In (B.213), we have introduced the block-diagonal matrix $\boldsymbol{Z} \in \mathbb{C}^{(M-1)d \times (M-1)d}$, which reduces for $d = 1$ to

$$\boldsymbol{Z} = \begin{bmatrix} \mathrm{e}^{\mathrm{j}\mu} & -1 & 0 & \dots & 0 \\ 0 & \mathrm{e}^{\mathrm{j}\mu} & -1 & \dots & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & -1 \\ 0 & 0 & 0 & \dots & \mathrm{e}^{\mathrm{j}\mu} \end{bmatrix} \in \mathbb{C}^{(M-1)\times(M-1)}. \tag{B.315}$$

Additionally, we let $\boldsymbol{u} = \begin{bmatrix} 0 & \cdots & 0 & 1 \end{bmatrix}^{\mathrm{T}} \in \mathbb{R}^{(M-1)\times 1}$ such $\boldsymbol{F}_2 = [\boldsymbol{Z},\ -\boldsymbol{u}]$. With these definitions, we can write $\boldsymbol{F}_2 \cdot \boldsymbol{F}_2^{\mathrm{H}}$ as

$$\boldsymbol{F}_2 \cdot \boldsymbol{F}_2^{\mathrm{H}} = \boldsymbol{Z} \cdot \boldsymbol{Z}^{\mathrm{H}} + \boldsymbol{u} \cdot \boldsymbol{u}^{\mathrm{T}}. \tag{B.316}$$

Applying the matrix inversion lemma [GvL96] in (1.23) to (B.316), we obtain

$$\begin{aligned} \left(\boldsymbol{F}_2 \cdot \boldsymbol{F}_2^{\mathrm{H}}\right)^{-1} &= \left(\boldsymbol{Z} \cdot \boldsymbol{Z}^{\mathrm{H}}\right)^{-1} - \frac{1}{1 + \boldsymbol{u}^{\mathrm{T}} \cdot \left(\boldsymbol{Z} \cdot \boldsymbol{Z}^{\mathrm{H}}\right)^{-1} \cdot \boldsymbol{u}} \cdot \left(\boldsymbol{Z} \cdot \boldsymbol{Z}^{\mathrm{H}}\right)^{-1} \cdot \boldsymbol{u} \cdot \boldsymbol{u}^{\mathrm{T}} \cdot \left(\boldsymbol{Z} \cdot \boldsymbol{Z}^{\mathrm{H}}\right)^{-1} \\ &= \boldsymbol{Z}^{-\mathrm{H}} \cdot \boldsymbol{Z}^{-1} - \frac{1}{1 + \boldsymbol{u}^{\mathrm{T}} \cdot \boldsymbol{Z}^{-\mathrm{H}} \cdot \boldsymbol{Z}^{-1} \cdot \boldsymbol{u}} \cdot \boldsymbol{Z}^{-\mathrm{H}} \cdot \boldsymbol{Z}^{-1} \cdot \boldsymbol{u} \cdot \boldsymbol{u}^{\mathrm{T}} \cdot \boldsymbol{Z}^{-\mathrm{H}} \cdot \boldsymbol{Z}^{-1}, \end{aligned} \tag{B.317}$$

where we have used the fact that $\left(\boldsymbol{Z} \cdot \boldsymbol{Z}^{\mathrm{H}}\right)^{-1} = \boldsymbol{Z}^{-\mathrm{H}} \cdot \boldsymbol{Z}^{-1}$ since $\boldsymbol{Z}$ is a square matrix. Then, from (B.218), it is easily verified that $\boldsymbol{Z}^{-1} \in \mathbb{C}^{(M-1)\times(M-1)}$ for $d = 1$ is given by

$$\boldsymbol{Z}^{-1} = \begin{bmatrix} \mathrm{e}^{-\mathrm{j}\mu} & \mathrm{e}^{-\mathrm{j}2\mu} & \dots & \mathrm{e}^{-\mathrm{j}(M-1)\mu} \\ 0 & \mathrm{e}^{-\mathrm{j}\mu} & \dots & \vdots \\ \vdots & \ddots & \ddots & \mathrm{e}^{-\mathrm{j}2\mu} \\ 0 & 0 & \dots & \mathrm{e}^{-\mathrm{j}\mu} \end{bmatrix}. \tag{B.318}$$

As the term $\boldsymbol{Z}^{-1} \cdot \boldsymbol{u}$ denotes the last column of $\boldsymbol{Z}^{-1}$, we find that

$$\boldsymbol{Z}^{-1} \cdot \boldsymbol{u} = \boldsymbol{J}_1 \cdot \boldsymbol{a} \cdot \mathrm{e}^{-\mathrm{j}(M-1)\mu} \in \mathbb{C}^{M-1\times 1}. \tag{B.319}$$

Using (B.319), we immediately see that

$$1 + \boldsymbol{u}^{\mathrm{T}} \cdot \boldsymbol{Z}^{-\mathrm{H}} \cdot \boldsymbol{Z}^{-1} \cdot \boldsymbol{u} = M. \tag{B.320}$$

Thus, we can simplify (B.317) into

$$\begin{aligned}
\left(\boldsymbol{F}_2 \cdot \boldsymbol{F}_2^{\mathrm{H}}\right)^{-1} &= \boldsymbol{Z}^{-\mathrm{H}}\boldsymbol{Z}^{-1} - \frac{1}{M} \cdot \boldsymbol{Z}^{-\mathrm{H}} \cdot \boldsymbol{Z}^{-1} \cdot \boldsymbol{u} \cdot \boldsymbol{u}^{\mathrm{T}} \cdot \boldsymbol{Z}^{-\mathrm{H}} \cdot \boldsymbol{Z}^{-1} \\
&= \boldsymbol{Z}^{-\mathrm{H}} \cdot \left(\boldsymbol{I}_{M-1} - \frac{1}{M} \cdot \boldsymbol{Z}^{-1} \cdot \boldsymbol{u} \cdot \boldsymbol{u}^{\mathrm{T}} \cdot \boldsymbol{Z}^{-\mathrm{H}}\right) \cdot \boldsymbol{Z}^{-1} \\
&= \boldsymbol{Z}^{-\mathrm{H}} \cdot \left(\boldsymbol{I}_{M-1} - \frac{1}{M} \cdot \boldsymbol{J}_1 \cdot \boldsymbol{a} \cdot \boldsymbol{a}^{\mathrm{H}} \cdot \boldsymbol{J}_1^{\mathrm{T}}\right) \cdot \boldsymbol{Z}^{-1} \\
&= \boldsymbol{Z}^{-\mathrm{H}} \cdot \boldsymbol{J}_1 \cdot \left(\boldsymbol{I}_M - \frac{\boldsymbol{a} \cdot \boldsymbol{a}^{\mathrm{H}}}{M}\right) \cdot \boldsymbol{J}_1^{\mathrm{T}} \cdot \boldsymbol{Z}^{-1},
\end{aligned} \tag{B.321}$$

where we have again used the relation (B.319). Inserting (B.321) into (B.308), γ is given by

$$\gamma = \boldsymbol{a}^{\mathrm{H}} \cdot \boldsymbol{J}_1^{\mathrm{T}} \cdot \boldsymbol{Z}^{-\mathrm{H}} \cdot \boldsymbol{J}_1 \cdot \left(\boldsymbol{I}_M - \frac{\boldsymbol{a} \cdot \boldsymbol{a}^{\mathrm{H}}}{M}\right) \cdot \boldsymbol{J}_1^{\mathrm{T}} \cdot \boldsymbol{Z}^{-1} \cdot \boldsymbol{J}_1 \cdot \boldsymbol{a}. \tag{B.322}$$

Using (B.315), we can easily establish that

$$\boldsymbol{J}_1^{\mathrm{T}} \cdot \boldsymbol{Z}^{-1} \cdot \boldsymbol{J}_1 \cdot \boldsymbol{a} = \mathrm{diag}\left\{\boldsymbol{\Pi}_{M-1} \cdot \boldsymbol{c}\right\} \cdot \boldsymbol{a} \cdot \mathrm{e}^{-\mathrm{j}\mu}, \tag{B.323}$$

where $\boldsymbol{c}$ is defined in (B.312). Consequently, we can write (B.322) as

$$\begin{aligned}
\gamma &= \mathrm{e}^{\mathrm{j}\mu} \cdot \boldsymbol{a}^{\mathrm{H}} \cdot \mathrm{diag}\left\{\boldsymbol{c} \cdot \boldsymbol{\Pi}_{M-1}\right\} \cdot \left(\boldsymbol{I}_M - \frac{\boldsymbol{a} \cdot \boldsymbol{a}^{\mathrm{H}}}{M}\right) \cdot \mathrm{diag}\left\{\boldsymbol{\Pi}_{M-1} \cdot \boldsymbol{c}\right\} \cdot \boldsymbol{a} \cdot \mathrm{e}^{-\mathrm{j}\mu} \\
&= \boldsymbol{a}^{\mathrm{H}} \cdot \mathrm{diag}\left\{\boldsymbol{c} \cdot \boldsymbol{\Pi}_{M-1}\right\} \cdot \left(\boldsymbol{I}_M - \frac{\boldsymbol{a} \cdot \boldsymbol{a}^{\mathrm{H}}}{M}\right) \cdot \mathrm{diag}\left\{\boldsymbol{\Pi}_{M-1} \cdot \boldsymbol{c}\right\} \cdot \boldsymbol{a} \\
&= \boldsymbol{a}^{\mathrm{T}} \cdot \mathrm{diag}\left\{\boldsymbol{c}\right\} \cdot \boldsymbol{\Pi}_{M-1} \cdot \left(\boldsymbol{I}_M - \frac{\boldsymbol{a} \cdot \boldsymbol{a}^{\mathrm{H}}}{M}\right) \cdot \boldsymbol{\Pi}_{M-1} \cdot \mathrm{diag}\left\{\boldsymbol{c}\right\} \cdot \boldsymbol{a}^*,
\end{aligned} \tag{B.324}$$

where we have used the fact that $\mathrm{diag}\left\{\boldsymbol{\Pi}_{M-1} \cdot \boldsymbol{c}\right\} = \boldsymbol{\Pi}_{M-1} \cdot \mathrm{diag}\left\{\boldsymbol{c}\right\} \cdot \boldsymbol{\Pi}_{M-1}$ and the centro-symmetry property $\boldsymbol{\Pi}_{M-1} \cdot \boldsymbol{a} = \boldsymbol{a}^* \cdot \mathrm{e}^{\mathrm{j}(M-1)\mu}$ of the ULA, which follows from rearranging (2.22). Note that the phase term $\mathrm{e}^{\mathrm{j}(M-1)\mu}$ cancels immediately. Then, transposing (B.324) yields

$$\begin{aligned}
\gamma &= \boldsymbol{a}^{\mathrm{H}} \cdot \mathrm{diag}\left\{\boldsymbol{c}\right\} \cdot \boldsymbol{\Pi}_{M-1} \cdot \left(\boldsymbol{I}_M - \frac{\boldsymbol{a} \cdot \boldsymbol{a}^{\mathrm{H}}}{M}\right)^{\mathrm{T}} \cdot \boldsymbol{\Pi}_{M-1} \cdot \mathrm{diag}\left\{\boldsymbol{c}\right\} \cdot \boldsymbol{a} \\
&= \boldsymbol{a}^{\mathrm{H}} \cdot \mathrm{diag}\left\{\boldsymbol{c}\right\} \cdot \left(\boldsymbol{I}_M - \frac{\boldsymbol{a} \cdot \boldsymbol{a}^{\mathrm{H}}}{M}\right) \cdot \mathrm{diag}\left\{\boldsymbol{c}\right\} \cdot \boldsymbol{a},
\end{aligned} \tag{B.325}$$

where we have again applied the centro-symmetry property of the ULA from above. The final expression for γ in (B.325) is equivalent to (B.314). Therefore, the analytical MSE of 1-D Standard ESPRIT with GLS for a single source corresponds exactly to the single source expression of the deterministic CRB. We conclude that the GLS-based 1-D Standard ESPRIT algorithm is asymptotically efficient for a single source, i.e., the ratio of the CRB to the MSE is equal to one. This completes the proof. □

B.14.2. 1-D Unitary ESPRIT with GLS

The second part of the theorem is to prove that the MSEs of 1-D Unitary ESPRIT with GLS and 1-D Standard ESPRIT with GLS are identical for a single source. However, it is shown in [RH12] that forward-backward averaging (FBA) only affects $\boldsymbol{v}_{\mathrm{s}}$ and has no effect on $\boldsymbol{u}_{\mathrm{s}}$ or $\boldsymbol{U}_{\mathrm{n}}$. Hence, applying the same steps as in Section B.14.1 for 1-D Standard ESPRIT with GLS immediately proves this part of the theorem. □

B.15. Proof of Theorem 5.5.2

For the proof of Theorem 5.5.2 in Section 5.5.2, we first consider the 1-D version of the analytical MSE expression in (5.67) for an arbitrary number of sources. The corresponding MSE for the i-th spatial frequency is given by

$$\mathbb{E}\left\{(\Delta\mu_i)^2\right\} = \frac{\sigma_{\mathrm{n}}^2}{2} \cdot \|\boldsymbol{z}_{i,\mathrm{GLS}}\|_2^2 \tag{B.326}$$

for $i = 1, \ldots, d$, where $\boldsymbol{z}_{i,\mathrm{GLS}} = \boldsymbol{W}_{\mathrm{mat}}^{\mathrm{T}} \cdot \boldsymbol{r}_{i,\mathrm{GLS}}$ and $\boldsymbol{r}_{i,\mathrm{GLS}}^{\mathrm{T}}$ and $\boldsymbol{W}_{\mathrm{mat}}$ are given by

$$\boldsymbol{r}_{i,\mathrm{GLS}}^{\mathrm{T}} = -\frac{1}{\lambda_i} \cdot \left(\boldsymbol{q}_i^{\mathrm{T}} \otimes \boldsymbol{p}_i^{\mathrm{T}}\right) \cdot \left(\boldsymbol{F}_1^{\mathrm{H}} \cdot \boldsymbol{R}_0^{-1} \cdot \boldsymbol{F}_1\right)^{-1} \cdot \boldsymbol{F}_1^{\mathrm{H}} \cdot \boldsymbol{R}_0^{-1} \cdot \boldsymbol{F}_2^{\mathrm{H}} \tag{B.327}$$

$$\boldsymbol{W}_{\mathrm{mat}} = \left(\boldsymbol{\Sigma}_{\mathrm{s}}^{-1} \cdot \boldsymbol{V}_{\mathrm{s}}^{\mathrm{T}}\right) \otimes \left(\boldsymbol{U}_{\mathrm{n}} \cdot \boldsymbol{U}_{\mathrm{n}}^{\mathrm{H}}\right). \tag{B.328}$$

Moreover, the matrices $\boldsymbol{F}_1$, $\boldsymbol{F}_2$, and $\boldsymbol{R}_0$ are given by

$$\boldsymbol{F}_1 = (\boldsymbol{I}_d \otimes \boldsymbol{J}_1 \cdot \boldsymbol{U}_{\mathrm{s}}) \tag{B.329}$$

$$\boldsymbol{F}_2 = \left(\boldsymbol{\Psi}^{\mathrm{T}} \otimes \boldsymbol{J}_1\right) - (\boldsymbol{I}_d \otimes \boldsymbol{J}_2) \tag{B.330}$$

$$\boldsymbol{R}_0 = \boldsymbol{F}_2 \cdot \left(\boldsymbol{\Sigma}_{\mathrm{s}}^{-2} \otimes \boldsymbol{I}_M\right) \cdot \boldsymbol{F}_2^{\mathrm{H}}. \tag{B.331}$$

In the first step, we simplify (B.326) by inserting the expression for $\boldsymbol{W}_{\mathrm{mat}}$ from (B.328). We obtain

$$\mathbb{E}\left\{(\Delta\mu_i)^2\right\} = \frac{\sigma_{\mathrm{n}}^2}{2} \cdot \boldsymbol{r}_{i,\mathrm{GLS}}^{\mathrm{T}} \cdot \boldsymbol{W}_{\mathrm{mat}} \cdot \boldsymbol{W}_{\mathrm{mat}}^{\mathrm{H}} \cdot \boldsymbol{r}_{i,\mathrm{GLS}}^{*} \tag{B.332}$$

$$= \frac{\sigma_{\mathrm{n}}^2}{2} \cdot \boldsymbol{r}_{i,\mathrm{GLS}}^{\mathrm{T}} \cdot \left(\boldsymbol{\Sigma}_{\mathrm{s}}^{-1} \cdot \boldsymbol{V}_{\mathrm{s}}^{\mathrm{T}} \otimes \boldsymbol{U}_{\mathrm{n}} \cdot \boldsymbol{U}_{\mathrm{n}}^{\mathrm{H}}\right) \cdot \left(\boldsymbol{\Sigma}_{\mathrm{s}}^{-1} \cdot \boldsymbol{V}_{\mathrm{s}}^{\mathrm{T}} \otimes \boldsymbol{U}_{\mathrm{n}} \cdot \boldsymbol{U}_{\mathrm{n}}^{\mathrm{H}}\right)^{\mathrm{H}} \cdot \boldsymbol{r}_{i,\mathrm{GLS}}^{*}$$
$$= \frac{\sigma_{\mathrm{n}}^2}{2} \cdot \boldsymbol{r}_{i,\mathrm{GLS}}^{\mathrm{T}} \cdot \left(\boldsymbol{\Sigma}_{\mathrm{s}}^{-2} \otimes \boldsymbol{U}_{\mathrm{n}} \cdot \boldsymbol{U}_{\mathrm{n}}^{\mathrm{H}}\right) \cdot \boldsymbol{r}_{i,\mathrm{GLS}}^{*}. \tag{B.333}$$

As $\boldsymbol{U}_{\mathrm{n}} \cdot \boldsymbol{U}_{\mathrm{n}}^{\mathrm{H}}$ is the projection matrix onto the noise subspace, we can write it in terms of the signal subspace spanned by $\boldsymbol{A}$ as $\boldsymbol{U}_{\mathrm{n}} \cdot \boldsymbol{U}_{\mathrm{n}}^{\mathrm{H}} = \boldsymbol{I}_M - \boldsymbol{A} \cdot \boldsymbol{A}^{+}$. Inserting this relation into (B.333) yields

$$\mathbb{E}\left\{(\Delta\mu_i)^2\right\} = \frac{\sigma_{\mathrm{n}}^2}{2} \cdot \boldsymbol{r}_{i,\mathrm{GLS}}^{\mathrm{T}} \cdot \left(\boldsymbol{\Sigma}_{\mathrm{s}}^{-2} \otimes \left(\boldsymbol{I}_M - \boldsymbol{A} \cdot \boldsymbol{A}^{+}\right)\right) \cdot \boldsymbol{r}_{i,\mathrm{GLS}}^{*}$$
$$= \frac{\sigma_{\mathrm{n}}^2}{2} \cdot \left(\boldsymbol{r}_{i,\mathrm{GLS}}^{\mathrm{T}} \cdot \left(\boldsymbol{\Sigma}_{\mathrm{s}}^{-2} \otimes \boldsymbol{I}_M\right) \cdot \boldsymbol{r}_{i,\mathrm{GLS}}^{*} - \boldsymbol{r}_{i,\mathrm{GLS}}^{\mathrm{T}} \cdot \left(\boldsymbol{\Sigma}_{\mathrm{s}}^{-2} \otimes \boldsymbol{A} \cdot \boldsymbol{A}^{+}\right) \cdot \boldsymbol{r}_{i,\mathrm{GLS}}^{*}\right). \tag{B.334}$$

As in the single source case considered in (B.300), it is straightforward to see by means of the noise-free shift invariance equation that the second term of (B.334) cancels. Therefore, the MSE of 1-D Standard ESPRIT with GLS for the i-th source is given by

$$\mathbb{E}\left\{(\Delta\mu_i)^2\right\} = \frac{\sigma_{\mathrm{n}}^2}{2} \cdot \boldsymbol{r}_{i,\mathrm{GLS}}^{\mathrm{T}} \cdot \left(\boldsymbol{\Sigma}_{\mathrm{s}}^{-2} \otimes \boldsymbol{I}_M\right) \cdot \boldsymbol{r}_{i,\mathrm{GLS}}^{*}, \tag{B.335}$$

which upon inserting $\boldsymbol{r}_{i,\mathrm{GLS}}$ from (B.327) and applying the same steps as in (B.303), simplifies into

$$\mathbb{E}\left\{(\Delta\mu_i)^2\right\} = \frac{\sigma_{\mathrm{n}}^2}{2} \cdot \left(\boldsymbol{q}_i^{\mathrm{T}} \otimes \boldsymbol{p}_i^{\mathrm{T}}\right) \cdot \left(\boldsymbol{F}_1^{\mathrm{H}} \cdot \boldsymbol{R}_0^{-1} \cdot \boldsymbol{F}_1\right)^{-1} \cdot \left(\boldsymbol{q}_i^{\mathrm{T}} \otimes \boldsymbol{p}_i^{\mathrm{T}}\right)^{\mathrm{H}}. \tag{B.336}$$

In Appendix B.10, we have simplified the term $\boldsymbol{R}_0$ in (B.331) by assuming temporally orthogonal signals with a diagonal sample signal covariance matrix $\hat{\boldsymbol{R}}_{\mathrm{ss}} = \mathrm{diag}\left\{\hat{P}_i\right\}_{i=1}^{d}$, where $\hat{P}_i = \|\boldsymbol{s}_i\|_2^2/N$ is the empirical power of the i-th symbol vector $\boldsymbol{s}_i$. In (B.205), we have found that $\boldsymbol{R}_0^{-1}$ can be reformulated as

$$\boldsymbol{R}_0^{-1} = \left(\boldsymbol{\Sigma}_{\mathrm{s}}^{2} \cdot \boldsymbol{Q}^{-\mathrm{T}} \otimes \boldsymbol{I}_{M-1}\right) \cdot \left(\bar{\boldsymbol{F}} \cdot \bar{\boldsymbol{F}}^{\mathrm{H}}\right)^{-1} \cdot \left(\boldsymbol{Q}^{\mathrm{T}} \otimes \boldsymbol{I}_{M-1}\right), \tag{B.337}$$

where $\bar{\boldsymbol{F}}$ is defined as $\bar{\boldsymbol{F}} = (\boldsymbol{\Lambda} \otimes \boldsymbol{J}_1) - (\boldsymbol{I}_d \otimes \boldsymbol{J}_2)$ according to (B.195). Then, the term $\boldsymbol{F}_1^{\mathrm{H}} \cdot \boldsymbol{R}_0^{-1} \cdot \boldsymbol{F}_1$ in (B.336), can be expressed as

$$\boldsymbol{F}_1^{\mathrm{H}} \cdot \boldsymbol{R}_0^{-1} \cdot \boldsymbol{F}_1 = (\boldsymbol{I}_d \otimes \boldsymbol{J}_1 \cdot \boldsymbol{U}_{\mathrm{s}})^{\mathrm{H}} \cdot \boldsymbol{R}_0^{-1} \cdot (\boldsymbol{I}_d \otimes \boldsymbol{J}_1 \cdot \boldsymbol{U}_{\mathrm{s}}) \tag{B.338}$$
$$= \frac{1}{M} \cdot \left(\boldsymbol{I}_d \otimes \boldsymbol{Q}^{-\mathrm{H}} \cdot \boldsymbol{A}^{\mathrm{H}} \cdot \boldsymbol{J}_1^{\mathrm{T}}\right) \cdot \boldsymbol{R}_0^{-1} \cdot \left(\boldsymbol{I}_d \otimes \boldsymbol{J}_1 \cdot \boldsymbol{A} \cdot \boldsymbol{Q}^{-1}\right), \tag{B.339}$$

where we have used the relation $\boldsymbol{U}_{\mathrm{s}} = \frac{1}{\sqrt{M}} \cdot \boldsymbol{A} \cdot \boldsymbol{Q}^{-1}$ from (B.198). Inserting (B.337) into (B.339), we get

$$\boldsymbol{F}_1^{\mathrm{H}} \cdot \boldsymbol{R}_0^{-1} \cdot \boldsymbol{F}_1 = \frac{1}{M} \cdot \left(\boldsymbol{\Sigma}_{\mathrm{s}}^{2} \cdot \boldsymbol{Q}^{-\mathrm{T}} \otimes \boldsymbol{Q}^{-\mathrm{H}} \cdot \boldsymbol{A}^{\mathrm{H}} \cdot \boldsymbol{J}_1^{\mathrm{T}}\right) \cdot \left(\bar{\boldsymbol{F}} \cdot \bar{\boldsymbol{F}}^{\mathrm{H}}\right)^{-1} \cdot \left(\boldsymbol{Q}^{\mathrm{T}} \otimes \boldsymbol{J}_1 \cdot \boldsymbol{A} \cdot \boldsymbol{Q}^{-1}\right)$$

$$= \frac{1}{M} \cdot \left(\boldsymbol{\Sigma}_{\mathrm{s}}^{2} \cdot \boldsymbol{Q}^{-\mathrm{T}} \otimes \boldsymbol{Q}^{-\mathrm{H}} \right) \cdot \boldsymbol{Y} \cdot \left(\boldsymbol{Q}^{\mathrm{T}} \otimes \boldsymbol{Q}^{-1} \right), \tag{B.340}$$

where we have defined the matrix $\boldsymbol{Y}$ as

$$\boldsymbol{Y} = \left(\boldsymbol{I}_{d} \otimes \boldsymbol{A}^{\mathrm{H}} \cdot \boldsymbol{J}_{1}^{\mathrm{T}} \right) \cdot \left(\bar{\boldsymbol{F}} \cdot \bar{\boldsymbol{F}}^{\mathrm{H}} \right)^{-1} \cdot \left(\boldsymbol{I}_{d} \otimes \boldsymbol{J}_{1} \cdot \boldsymbol{A} \right). \tag{B.341}$$

Hence, we can write the term $\left(\boldsymbol{F}_{1}^{\mathrm{H}} \cdot \boldsymbol{R}_{0}^{-1} \cdot \boldsymbol{F}_{1} \right)^{-1}$ in (B.336) as

$$\left(\boldsymbol{F}_{1}^{\mathrm{H}} \cdot \boldsymbol{R}_{0}^{-1} \cdot \boldsymbol{F}_{1} \right)^{-1} = M \cdot \left(\boldsymbol{Q}^{-\mathrm{T}} \otimes \boldsymbol{Q} \right) \cdot \boldsymbol{Y}^{-1} \cdot \left(\boldsymbol{Q}^{\mathrm{T}} \cdot \boldsymbol{\Sigma}_{\mathrm{s}}^{-2} \otimes \boldsymbol{Q}^{\mathrm{H}} \right). \tag{B.342}$$

Combining these results, the MSE expression in (B.336) is given by

$$\begin{aligned}
\mathbb{E}\left\{ (\Delta\mu_i)^2 \right\} &= \frac{\sigma_{\mathrm{n}}^{2} M}{2} \cdot \left(\boldsymbol{q}_i^{\mathrm{T}} \otimes \boldsymbol{p}_i^{\mathrm{T}} \right) \cdot \left(\boldsymbol{Q}^{-\mathrm{T}} \otimes \boldsymbol{Q} \right) \cdot \boldsymbol{Y}^{-1} \cdot \left(\boldsymbol{Q}^{\mathrm{T}} \cdot \boldsymbol{\Sigma}_{\mathrm{s}}^{-2} \otimes \boldsymbol{Q}^{\mathrm{H}} \right) \cdot \left(\boldsymbol{q}_i^{\mathrm{T}} \otimes \boldsymbol{p}_i^{\mathrm{T}} \right)^{\mathrm{H}} \\
&= \frac{\sigma_{\mathrm{n}}^{2} M}{2} \cdot \left(\boldsymbol{q}_i^{\mathrm{T}} \cdot \boldsymbol{Q}^{-\mathrm{T}} \otimes \boldsymbol{p}_i^{\mathrm{T}} \cdot \boldsymbol{Q} \right) \cdot \boldsymbol{Y}^{-1} \cdot \left(\boldsymbol{Q}^{\mathrm{T}} \cdot \boldsymbol{\Sigma}_{\mathrm{s}}^{-2} \cdot \boldsymbol{q}_i^{*} \otimes \boldsymbol{Q}^{\mathrm{H}} \cdot \boldsymbol{p}_i^{*} \right) \qquad \text{(B.343)} \\
&= \frac{\sigma_{\mathrm{n}}^{2} M}{2} \cdot \left(\boldsymbol{e}_i^{\mathrm{T}} \otimes \boldsymbol{e}_i^{\mathrm{T}} \right) \cdot \boldsymbol{Y}^{-1} \cdot \left(\boldsymbol{Q}^{\mathrm{T}} \cdot \boldsymbol{\Sigma}_{\mathrm{s}}^{-2} \cdot \boldsymbol{q}_i^{*} \otimes \boldsymbol{e}_i \right), \qquad \text{(B.344)}
\end{aligned}$$

where we have used the identities $\boldsymbol{q}_i^{\mathrm{T}} \cdot \boldsymbol{Q}^{-\mathrm{T}} = \boldsymbol{e}_i^{\mathrm{T}}$ and $\boldsymbol{p}_i^{\mathrm{T}} \cdot \boldsymbol{Q} = \boldsymbol{e}_i^{\mathrm{T}}$, where $\boldsymbol{e}_i$ denotes the vector with 1 at the i-th position and zeros elsewhere. In (B.201), for temporally orthogonal sources, we have established the relation

$$\boldsymbol{\Sigma}_{\mathrm{s}}^{-2} = \frac{1}{M \cdot N} \cdot \boldsymbol{Q}^{-\mathrm{T}} \cdot \hat{\boldsymbol{R}}_{\mathrm{ss}}^{-\mathrm{T}} \cdot \boldsymbol{Q}^{-*}. \tag{B.345}$$

Using (B.345) in (B.344), we obtain the MSE result

$$\mathbb{E}\left\{ (\Delta\mu_i)^2 \right\} = \frac{\sigma_{\mathrm{n}}^{2}}{2N} \cdot \left(\boldsymbol{e}_i^{\mathrm{T}} \otimes \boldsymbol{e}_i^{\mathrm{T}} \right) \cdot \boldsymbol{Y}^{-1} \cdot \left(\hat{\boldsymbol{R}}_{\mathrm{ss}}^{-\mathrm{T}} \cdot \boldsymbol{e}_i \otimes \boldsymbol{e}_i \right). \tag{B.346}$$

It should be highlighted that the MSE for 1-D Standard ESPRIT with GLS in (B.346) is valid for an arbitrary number of signals d. However, after inserting $\boldsymbol{Y}$ from (B.341) into (B.346), it is difficult to further simplify (B.346) for an arbitrary number of sources d due to the term $\boldsymbol{Y}^{-1}$. Nevertheless, special cases on the number of signals can be considered to gain further insights. In order to continue the proof of Theorem 5.5.2, we set $d = 2$ such that the array steering matrix becomes $\boldsymbol{A} = [\boldsymbol{a}_1, \boldsymbol{a}_2] \in \mathbb{C}^{M \times 2}$, where we have dropped the dependence of the array steering vectors $\boldsymbol{a}_i$, $i = 1, 2$, on μ_i for notational convenience. Notice that for the computation of the MSE for $d = 2$, we only need to derive the MSE for one of the sources as the second source has the same MSE due to the symmetry of the scenario for $d = 2$. To compute $\boldsymbol{Y}^{-1}$ in (B.346) for $d = 2$, we first consider the definition of $\boldsymbol{Y}$ in (B.341) and use the result from (B.225) to simplify the term

$\left(\bar{\boldsymbol{F}} \cdot \bar{\boldsymbol{F}}^{\mathrm{H}}\right)^{-1}$ contained in $\boldsymbol{Y}$ into

$$\left(\bar{\boldsymbol{F}} \cdot \bar{\boldsymbol{F}}^{\mathrm{H}}\right)^{-1} = \boldsymbol{Z}^{-\mathrm{H}} \cdot (\boldsymbol{I}_d \otimes \boldsymbol{J}_1) \cdot \mathrm{blkdiag}\left\{\boldsymbol{P}_{\boldsymbol{a}_1}^{\perp}, \boldsymbol{P}_{\boldsymbol{a}_2}^{\perp}\right\} \cdot \left(\boldsymbol{I}_d \otimes \boldsymbol{J}_1^{\mathrm{T}}\right) \cdot \boldsymbol{Z}^{-1}, \tag{B.347}$$

where $\boldsymbol{Z} = \mathrm{blkdiag}\{\boldsymbol{Z}_1, \boldsymbol{Z}_2\} \in \mathbb{C}^{2(M-1)\times 2(M-1)}$ is the block-diagonal matrix defined in (B.213) and $\boldsymbol{P}_{\boldsymbol{a}_i}^{\perp} = \boldsymbol{I}_M - \frac{1}{M} \cdot \boldsymbol{a}_i \cdot \boldsymbol{a}_i^{\mathrm{H}} \in \mathbb{C}^{M\times M}$ for $i = 1, 2$ is the projection matrix onto the complement of the signal subspace spanned by the steering vector $\boldsymbol{a}_i$. Upon inserting (B.347) into (B.341) for $d = 2$, we obtain

$$\boldsymbol{Y} = (\boldsymbol{I}_d \otimes \boldsymbol{J}_1 \cdot \boldsymbol{A})^{\mathrm{H}} \cdot \boldsymbol{Z}^{-\mathrm{H}} \cdot (\boldsymbol{I}_d \otimes \boldsymbol{J}_1) \cdot \mathrm{blkdiag}\left\{\boldsymbol{P}_{\boldsymbol{a}_1}^{\perp}, \boldsymbol{P}_{\boldsymbol{a}_2}^{\perp}\right\} \cdot \left(\boldsymbol{I}_d \otimes \boldsymbol{J}_1^{\mathrm{T}}\right) \cdot \boldsymbol{Z}^{-1} \cdot (\boldsymbol{I}_d \otimes \boldsymbol{J}_1 \cdot \boldsymbol{A}). \tag{B.348}$$

It is apparent that (B.348) is a block-diagonal matrix, which is to be inverted to compute $\boldsymbol{Y}^{-1}$. Hence, we can use the result that the inverse of a block-diagonal matrix can be computed by inverting the individual blocks separately [GvL96]. Note that due to the symmetry for $d = 2$, we only need to compute one of the two blocks as the other one is symmetric. Considering the first source ($i = 1$), the corresponding upper left block $\boldsymbol{Y}_1 \in \mathbb{C}^{2\times 2}$ of $\boldsymbol{Y} \in \mathbb{C}^{4\times 4}$ is given by

$$\boldsymbol{Y}_1 = \boldsymbol{A}^{\mathrm{H}} \cdot \boldsymbol{J}_1^{\mathrm{T}} \cdot \boldsymbol{Z}_1^{-\mathrm{H}} \cdot \boldsymbol{J}_1 \cdot \boldsymbol{P}_{\boldsymbol{a}_1}^{\perp} \cdot \boldsymbol{J}_1^{\mathrm{T}} \cdot \boldsymbol{Z}_1^{-1} \cdot \boldsymbol{J}_1 \cdot \boldsymbol{A}. \tag{B.349}$$

where $\boldsymbol{Z}_1$ is the first block of $\boldsymbol{Z}$. Using $\boldsymbol{A} = [\boldsymbol{a}_1, \boldsymbol{a}_2]$, we have found in (B.323) that

$$\boldsymbol{J}_1^{\mathrm{T}} \cdot \boldsymbol{Z}_1^{-1} \cdot \boldsymbol{J}_1 \cdot \boldsymbol{a}_1 = \mathrm{diag}\{\boldsymbol{\Pi}_{M-1} \cdot \boldsymbol{c}\} \cdot \boldsymbol{a}_1 \cdot \mathrm{e}^{-\mathrm{j}\mu_1}, \tag{B.350}$$

where $\boldsymbol{c} = \begin{bmatrix} 0 & 1 & 2 & \dots & M-1 \end{bmatrix}^{\mathrm{T}} \in \mathbb{R}^{M\times 1}$ and similarly, we can easily establish that

$$\boldsymbol{J}_1^{\mathrm{T}} \cdot \boldsymbol{Z}_1^{-1} \cdot \boldsymbol{J}_1 \cdot \boldsymbol{a}_2 = \frac{\mathrm{e}^{-\mathrm{j}\mu_1}}{1 - \mathrm{e}^{\mathrm{j}\Delta\mu}} \cdot \left(\boldsymbol{a}_2 - \mathrm{e}^{\mathrm{j}(M-1)\Delta\mu} \cdot \boldsymbol{a}_1\right). \tag{B.351}$$

With these results, we can express $\boldsymbol{Y}_1$ in (B.349) as

$$\begin{aligned} \boldsymbol{Y}_1 = \Bigg[\mathrm{diag}\{\boldsymbol{\Pi}_{M-1} \cdot \boldsymbol{c}\} \cdot \boldsymbol{a}_1,\ \frac{1}{1 - \mathrm{e}^{\mathrm{j}\Delta\mu}} \cdot \left(\boldsymbol{a}_2 - \mathrm{e}^{\mathrm{j}(M-1)\Delta\mu} \cdot \boldsymbol{a}_1\right)\Bigg]^{\mathrm{H}} \cdot \boldsymbol{P}_{\boldsymbol{a}_1}^{\perp} \\ \cdot \Bigg[\mathrm{diag}\{\boldsymbol{\Pi}_{M-1} \cdot \boldsymbol{c}\} \cdot \boldsymbol{a}_1,\ \frac{1}{1 - \mathrm{e}^{\mathrm{j}\Delta\mu}} \cdot \left(\boldsymbol{a}_2 - \mathrm{e}^{\mathrm{j}(M-1)\Delta\mu} \cdot \boldsymbol{a}_1\right)\Bigg], \end{aligned} \tag{B.352}$$

where the phase term $\mathrm{e}^{-\mathrm{j}\mu_1}$ from (B.350) and (B.351) cancels. Then, we use short hand notation and simplify (B.352) as

$$\boldsymbol{Y}_1 = \begin{bmatrix} y_1 & y_2 \\ y_3 & y_4 \end{bmatrix}, \tag{B.353}$$

we introduce the definitions

$$y_1 = \boldsymbol{a}_1^{\mathrm{H}} \cdot \operatorname{diag}\{\boldsymbol{c}\} \cdot \boldsymbol{P}_{\boldsymbol{a}_1}^{\perp} \cdot \operatorname{diag}\{\boldsymbol{c}\} \cdot \boldsymbol{a}_1 \tag{B.354}$$

$$y_2 = \boldsymbol{a}_1^{\mathrm{H}} \cdot \operatorname{diag}\left\{\boldsymbol{c} \cdot \boldsymbol{\Pi}_{(M-1)}\right\} \cdot \boldsymbol{P}_{\boldsymbol{a}_1}^{\perp} \cdot \frac{\left(\boldsymbol{a}_2 - \mathrm{e}^{\mathrm{j}(M-1)\Delta\mu} \cdot \boldsymbol{a}_1\right)}{1 - \mathrm{e}^{\mathrm{j}\Delta\mu}} \tag{B.355}$$

$$y_3 = \frac{\left(\boldsymbol{a}_2 - \mathrm{e}^{\mathrm{j}(M-1)\Delta\mu} \cdot \boldsymbol{a}_1\right)^{\mathrm{H}}}{1 - \mathrm{e}^{-\mathrm{j}\Delta\mu}} \cdot \boldsymbol{P}_{\boldsymbol{a}_1}^{\perp} \cdot \operatorname{diag}\{\boldsymbol{\Pi}_{M-1} \cdot \boldsymbol{c}\} \cdot \boldsymbol{a}_1 \tag{B.356}$$

$$y_4 = \frac{\left(\boldsymbol{a}_2 - \mathrm{e}^{\mathrm{j}(M-1)\Delta\mu} \cdot \boldsymbol{a}_1\right)^{\mathrm{H}}}{1 - \mathrm{e}^{-\mathrm{j}\Delta\mu}} \cdot \boldsymbol{P}_{\boldsymbol{a}_1}^{\perp} \cdot \frac{\left(\boldsymbol{a}_2 - \mathrm{e}^{\mathrm{j}(M-1)\Delta\mu} \cdot \boldsymbol{a}_1\right)}{1 - \mathrm{e}^{\mathrm{j}\Delta\mu}}. \tag{B.357}$$

Then, we can easily compute $\boldsymbol{Y}_1^{-1}$ by using the inversion rule for 2×2 matrices [GvL96] to obtain

$$\boldsymbol{Y}_1^{-1} = \frac{1}{D} \cdot \begin{bmatrix} y_4 & -y_2 \\ -y_3 & y_1 \end{bmatrix}. \tag{B.358}$$

where we define the determinant $D = y_1 \cdot y_4 - y_2 \cdot y_3$.

Finally, the MSE of 1-D Standard ESPRIT with GLS for the first source $i = 1$ is given by

$$\mathbb{E}\left\{(\Delta\mu_1)^2\right\} = \frac{\sigma_{\mathrm{n}}^2}{2N\hat{P}_1} \cdot \frac{y_4}{D}. \tag{B.359}$$

Similarly to the single source case in Appendix B.14, we next simplify the expression for the deterministic CRB [SN89] for $d = 2$ temporally orthogonal signals. The general expression for the CRB for arbitrary number of temporally orthogonal signals is given by

$$\boldsymbol{C} = \frac{\sigma_{\mathrm{n}}^2}{2N} \cdot \operatorname{Re}\left\{\left[\boldsymbol{D}^{\mathrm{H}} \cdot \left(\boldsymbol{I}_M - \boldsymbol{A} \cdot \left(\boldsymbol{A}^{\mathrm{H}} \cdot \boldsymbol{A}\right)^{-1} \cdot \boldsymbol{A}^{\mathrm{H}}\right) \cdot \boldsymbol{D}\right] \odot \hat{\boldsymbol{R}}_{\mathrm{S}}^{\mathrm{T}}\right\}^{-1}, \tag{B.360}$$

where $\hat{\boldsymbol{R}}_{\mathrm{S}} = \frac{1}{N} \cdot \boldsymbol{S} \cdot \boldsymbol{S}^{\mathrm{H}} = \operatorname{diag}\left\{\hat{P}_i\right\}_{i=1}^{d}$ is the diagonal sample signals covariance matrix containing the empirical powers of the sources and $\boldsymbol{D} \in \mathbb{C}^{M \times d}$ is the matrix of partial derivatives of the array steering vectors with respect to the parameters of interest.

Considering the i-th source, the general expression in (B.360) becomes

$$\begin{aligned} C_i &= \frac{\sigma_{\mathrm{n}}^2}{2N} \cdot \frac{1}{\hat{P}_i} \cdot \operatorname{Re}\left\{\boldsymbol{d}_i^{\mathrm{H}} \cdot \left(\boldsymbol{I}_M - \boldsymbol{A} \cdot \left(\boldsymbol{A}^{\mathrm{H}} \cdot \boldsymbol{A}\right)^{-1} \cdot \boldsymbol{A}^{\mathrm{H}}\right) \cdot \boldsymbol{d}_i\right\}^{-1} \\ &= \frac{\sigma_{\mathrm{n}}^2}{2N} \cdot \frac{1}{\hat{P}_i} \cdot \operatorname{Re}\left\{\boldsymbol{a}_i^{\mathrm{H}} \operatorname{diag}\{\boldsymbol{c}\} \cdot \left(\boldsymbol{I}_M - \boldsymbol{A} \cdot \left(\boldsymbol{A}^{\mathrm{H}} \cdot \boldsymbol{A}\right)^{-1} \cdot \boldsymbol{A}^{\mathrm{H}}\right) \cdot \operatorname{diag}\{\boldsymbol{c}\} \cdot \boldsymbol{a}_i\right\}^{-1}, \end{aligned} \tag{B.361}$$

where we have used the fact that

$$\boldsymbol{d} = \frac{\partial \boldsymbol{a}}{\partial \mu} = \mathrm{j} \cdot \operatorname{diag}\{\boldsymbol{c}\} \cdot \boldsymbol{a}, \tag{B.362}$$

where $\boldsymbol{c}$ is defined in (B.350).

In the special case of $d = 2$ and considering the first source ($i = 1$), we obtain

$$C_1 = \frac{\sigma_{\mathrm{n}}^2}{2N} \cdot \frac{1}{\hat{P}_1} \cdot \left(\boldsymbol{a}_1^{\mathrm{H}} \cdot \mathrm{diag}\left\{\boldsymbol{c}\right\}^2 \cdot \boldsymbol{a}_1 - \boldsymbol{a}_1^{\mathrm{H}} \cdot \mathrm{diag}\left\{\boldsymbol{c}\right\} \cdot \boldsymbol{A} \cdot \left(\boldsymbol{A}^{\mathrm{H}} \cdot \boldsymbol{A} \right)^{-1} \cdot \boldsymbol{A}^{\mathrm{H}} \cdot \mathrm{diag}\left\{\boldsymbol{c}\right\} \cdot \boldsymbol{a}_1 \right)^{-1}, \quad \text{(B.363)}$$

where the term $\left(\boldsymbol{A}^{\mathrm{H}} \cdot \boldsymbol{A} \right)^{-1}$ for $d = 2$ is given by

$$\left(\boldsymbol{A}^{\mathrm{H}} \cdot \boldsymbol{A} \right)^{-1} = \frac{1}{M^2 - |\alpha|^2} \cdot \begin{bmatrix} M & -\alpha \\ -\alpha^* & M \end{bmatrix} \quad \text{(B.364)}$$

and $\alpha = \boldsymbol{a}_1^{\mathrm{H}} \cdot \boldsymbol{a}_2$ is the spatial correlation of the two sources, which depends on their separation.

Comparing (B.363) to (B.359), we need to show that

$$\begin{aligned} \frac{D}{y_4} &= y_1 - \frac{y_2 \cdot y_3}{y_4} \\ &= \boldsymbol{a}_1^{\mathrm{H}} \cdot \mathrm{diag}\left\{\boldsymbol{c}\right\}^2 \cdot \boldsymbol{a}_1 - \boldsymbol{a}_1^{\mathrm{H}} \cdot \mathrm{diag}\left\{\boldsymbol{c}\right\} \cdot \boldsymbol{A} \cdot \left(\boldsymbol{A}^{\mathrm{H}} \cdot \boldsymbol{A} \right)^{-1} \cdot \boldsymbol{A}^{\mathrm{H}} \cdot \mathrm{diag}\left\{\boldsymbol{c}\right\} \cdot \boldsymbol{a}_1. \end{aligned} \quad \text{(B.365)}$$

Expanding y_1 as

$$y_1 = \boldsymbol{a}_1^{\mathrm{H}} \cdot \mathrm{diag}\left\{\boldsymbol{c}\right\} \cdot \boldsymbol{P}_{\boldsymbol{a}_1}^{\perp} \cdot \mathrm{diag}\left\{\boldsymbol{c}\right\} \cdot \boldsymbol{a}_1 \quad \text{(B.366)}$$

$$= \boldsymbol{a}_1^{\mathrm{H}} \cdot \mathrm{diag}\left\{\boldsymbol{c}\right\}^2 \cdot \boldsymbol{a}_1 - \boldsymbol{a}_1^{\mathrm{H}} \cdot \mathrm{diag}\left\{\boldsymbol{c}\right\} \cdot \frac{\boldsymbol{a}_1 \cdot \boldsymbol{a}_1^{\mathrm{H}}}{M} \cdot \mathrm{diag}\left\{\boldsymbol{c}\right\} \cdot \boldsymbol{a}_1. \quad \text{(B.367)}$$

It is straightforward to see that the first term of y_1 already corresponds to the first term of (B.365). Hence, we are left to show that

$$\boldsymbol{a}_1^{\mathrm{H}} \cdot \mathrm{diag}\left\{\boldsymbol{c}\right\} \cdot \frac{\boldsymbol{a}_1 \cdot \boldsymbol{a}_1^{\mathrm{H}}}{M} \cdot \mathrm{diag}\left\{\boldsymbol{c}\right\} \cdot \boldsymbol{a}_1 + \frac{y_2 \cdot y_3}{y_4} \quad \text{(B.368)}$$

$$= \boldsymbol{a}_1^{\mathrm{H}} \cdot \mathrm{diag}\left\{\boldsymbol{c}\right\} \cdot \boldsymbol{A} \cdot \left(\boldsymbol{A}^{\mathrm{H}} \cdot \boldsymbol{A} \right)^{-1} \cdot \boldsymbol{A}^{\mathrm{H}} \cdot \mathrm{diag}\left\{\boldsymbol{c}\right\} \cdot \boldsymbol{a}_1. \quad \text{(B.369)}$$

Starting with the left hand side in (B.368), we first simplify y_4 and we obtain

$$y_4 = \frac{1}{|1 - \mathrm{e}^{\mathrm{j}\Delta\mu}|^2} \cdot \left(\boldsymbol{a}_2 - \mathrm{e}^{\mathrm{j}(M-1)\Delta\mu} \cdot \boldsymbol{a}_1 \right)^{\mathrm{H}} \cdot \boldsymbol{P}_{\boldsymbol{a}_1}^{\perp} \cdot \left(\boldsymbol{a}_2 - \mathrm{e}^{\mathrm{j}(M-1)\Delta\mu} \cdot \boldsymbol{a}_1 \right) \quad \text{(B.370)}$$

$$= \frac{1}{|1 - \mathrm{e}^{\mathrm{j}\Delta\mu}|^2} \cdot \boldsymbol{a}_2^{\mathrm{H}} \cdot \left(\boldsymbol{I}_M - \frac{\boldsymbol{a}_1 \cdot \boldsymbol{a}_1^{\mathrm{H}}}{M} \right) \cdot \boldsymbol{a}_2$$

$$= \frac{1}{|1 - \mathrm{e}^{\mathrm{j}\Delta\mu}|^2} \cdot \left(M - \frac{|\alpha|^2}{M} \right) = \frac{1}{|1 - \mathrm{e}^{\mathrm{j}\Delta\mu}|^2} \cdot \frac{1}{M} \left(M^2 - |\alpha|^2 \right), \quad \text{(B.371)}$$

where $\alpha = \boldsymbol{a}_1^{\mathrm{H}} \cdot \boldsymbol{a}_2$ and in (B.370), we have used the fact $\boldsymbol{P}_{\boldsymbol{a}_1}^{\perp} \cdot \boldsymbol{a}_1 = \boldsymbol{0}$.

Next, we simplify the term $y_2 \cdot y_3$ in (B.368) and again use $\boldsymbol{P}_{\boldsymbol{a}_1}^{\perp} \cdot \boldsymbol{a}_1 = \boldsymbol{0}$ to obtain

$$\begin{aligned} y_2 \cdot y_3 &= \frac{1}{|1 - \mathrm{e}^{\mathrm{j}\Delta\mu}|^2} \cdot \boldsymbol{a}_1^{\mathrm{H}} \cdot \operatorname{diag}\left\{\boldsymbol{c} \cdot \boldsymbol{\Pi}_{(M-1)}\right\} \cdot \boldsymbol{P}_{\boldsymbol{a}_1}^{\perp} \cdot \boldsymbol{a}_2 \cdot \boldsymbol{a}_2^{\mathrm{H}} \cdot \boldsymbol{P}_{\boldsymbol{a}_1}^{\perp} \cdot \operatorname{diag}\{\boldsymbol{\Pi}_{M-1} \cdot \boldsymbol{c}\} \cdot \boldsymbol{a}_1 \\ &= \frac{1}{|1 - \mathrm{e}^{\mathrm{j}\Delta\mu}|^2} \cdot \boldsymbol{a}_1^{\mathrm{H}} \cdot \operatorname{diag}\{\boldsymbol{c}\} \cdot \boldsymbol{P}_{\boldsymbol{a}_1}^{\perp} \cdot \boldsymbol{a}_2 \cdot \boldsymbol{a}_2^{\mathrm{H}} \cdot \boldsymbol{P}_{\boldsymbol{a}_1}^{\perp} \cdot \operatorname{diag}\{\boldsymbol{c}\} \cdot \boldsymbol{a}_1. \end{aligned} \tag{B.372}$$

Then, we can express the last term in (B.368) have

$$\frac{y_2 \cdot y_3}{y_4} = \frac{M}{M^2 - |\alpha|^2} \cdot \boldsymbol{a}_1^{\mathrm{H}} \cdot \operatorname{diag}\{\boldsymbol{c}\} \cdot \boldsymbol{P}_{\boldsymbol{a}_1}^{\perp} \cdot \boldsymbol{a}_2 \cdot \boldsymbol{a}_2^{\mathrm{H}} \cdot \boldsymbol{P}_{\boldsymbol{a}_1}^{\perp} \cdot \operatorname{diag}\{\boldsymbol{c}\} \cdot \boldsymbol{a}_1. \tag{B.373}$$

Inserting (B.373) into (B.368), we can write (B.368) as

$$\boldsymbol{a}_1^{\mathrm{H}} \cdot \operatorname{diag}\{\boldsymbol{c}\} \cdot \frac{\boldsymbol{a}_1 \cdot \boldsymbol{a}_1^{\mathrm{H}}}{M} \cdot \operatorname{diag}\{\boldsymbol{c}\} \cdot \boldsymbol{a}_1 + \frac{y_2 \cdot y_3}{y_4} = \boldsymbol{a}_1^{\mathrm{H}} \cdot \operatorname{diag}\{\boldsymbol{c}\} \cdot \boldsymbol{W} \cdot \operatorname{diag}\{\boldsymbol{c}\} \cdot \boldsymbol{a}_1, \tag{B.374}$$

where

$$\boldsymbol{W} = \frac{\boldsymbol{a}_1 \cdot \boldsymbol{a}_1^{\mathrm{H}}}{M} + \frac{M}{M^2 - |\alpha|^2} \cdot \boldsymbol{P}_{\boldsymbol{a}_1}^{\perp} \cdot \boldsymbol{a}_2 \cdot \boldsymbol{a}_2^{\mathrm{H}} \cdot \boldsymbol{P}_{\boldsymbol{a}_1}^{\perp}. \tag{B.375}$$

The matrix $\boldsymbol{W}$ can be further simplified into

$$\boldsymbol{W} = \frac{1}{M\left(M^2 - |\alpha|^2\right)} \cdot \left(\left(M^2 - |\alpha|^2\right) \cdot \boldsymbol{a}_1 \cdot \boldsymbol{a}_1^{\mathrm{H}} + M^2 \cdot \boldsymbol{P}_{\boldsymbol{a}_1}^{\perp} \cdot \boldsymbol{a}_2 \cdot \boldsymbol{a}_2^{\mathrm{H}} \cdot \boldsymbol{P}_{\boldsymbol{a}_1}^{\perp}\right). \tag{B.376}$$

Considering the term $\boldsymbol{P}_{\boldsymbol{a}_1}^{\perp} \cdot \boldsymbol{a}_2 \cdot \boldsymbol{a}_2^{\mathrm{H}} \cdot \boldsymbol{P}_{\boldsymbol{a}_1}^{\perp}$ in (B.376) and expanding $\boldsymbol{P}_{\boldsymbol{a}_1}^{\perp}$, we get

$$\begin{aligned} \boldsymbol{P}_{\boldsymbol{a}_1}^{\perp} \cdot \boldsymbol{a}_2 \cdot \boldsymbol{a}_2^{\mathrm{H}} \cdot \boldsymbol{P}_{\boldsymbol{a}_1}^{\perp} &= \left(\boldsymbol{I}_M - \frac{1}{M} \cdot \boldsymbol{a}_1 \cdot \boldsymbol{a}_1^{\mathrm{H}}\right) \cdot \boldsymbol{a}_2 \cdot \boldsymbol{a}_2^{\mathrm{H}} \cdot \left(\boldsymbol{I}_M - \frac{1}{M} \cdot \boldsymbol{a}_1 \cdot \boldsymbol{a}_1^{\mathrm{H}}\right) \\ &= \left(\boldsymbol{a}_2 - \frac{1}{M} \cdot \boldsymbol{a}_1 \cdot \alpha\right) \cdot \left(\boldsymbol{a}_2^{\mathrm{H}} - \frac{1}{M} \cdot \alpha^* \cdot \boldsymbol{a}_1^{\mathrm{H}}\right) \\ &= \boldsymbol{a}_2 \cdot \boldsymbol{a}_2^{\mathrm{H}} - \frac{\alpha}{M} \cdot \boldsymbol{a}_1 \cdot \boldsymbol{a}_2^{\mathrm{H}} - \frac{\alpha^*}{M} \cdot \boldsymbol{a}_2 \cdot \boldsymbol{a}_1^{\mathrm{H}} + \frac{|\alpha|^2}{M^2} \cdot \boldsymbol{a}_1 \cdot \boldsymbol{a}_1^{\mathrm{H}} \\ &= \boldsymbol{A} \cdot \frac{1}{M^2} \cdot \begin{bmatrix} |\alpha|^2 & -\alpha M \\ -\alpha^* M & M^2 \end{bmatrix} \cdot \boldsymbol{A}^{\mathrm{H}}. \end{aligned} \tag{B.377}$$

Inserting (B.377) into (B.376), $\boldsymbol{W}$ becomes

$$\begin{aligned} \boldsymbol{W} &= \frac{1}{M\left(M^2 - |\alpha|^2\right)} \cdot \boldsymbol{A} \cdot \begin{bmatrix} M^2 & -\alpha M \\ -\alpha^* M & M^2 \end{bmatrix} \boldsymbol{A}^{\mathrm{H}} \\ &= \frac{1}{\left(M^2 - |\alpha|^2\right)} \cdot \boldsymbol{A} \cdot \begin{bmatrix} M & -\alpha \\ -\alpha^* & M \end{bmatrix} \boldsymbol{A}^{\mathrm{H}} \end{aligned}$$

$$= \boldsymbol{A} \cdot \left(\boldsymbol{A} \cdot \boldsymbol{A}^{\mathrm{H}} \right)^{-1} \boldsymbol{A}^{\mathrm{H}}. \tag{B.378}$$

Finally, we insert (B.378) into (B.373), we obtain

$$\boldsymbol{a}_1^{\mathrm{H}} \cdot \mathrm{diag}\left\{\boldsymbol{c}\right\} \cdot \frac{\boldsymbol{a}_1 \cdot \boldsymbol{a}_1^{\mathrm{H}}}{M} \cdot \mathrm{diag}\left\{\boldsymbol{c}\right\} \cdot \boldsymbol{a}_1 + \frac{y_2 \cdot y_3}{y_4} = \boldsymbol{a}_1^{\mathrm{H}} \cdot \mathrm{diag}\left\{\boldsymbol{c}\right\} \cdot \boldsymbol{A} \cdot \left(\boldsymbol{A} \cdot \boldsymbol{A}^{\mathrm{H}} \right)^{-1} \boldsymbol{A}^{\mathrm{H}} \cdot \mathrm{diag}\left\{\boldsymbol{c}\right\} \cdot \boldsymbol{a}_1, \tag{B.379}$$

which is exactly the right hand side (B.369) of the original equality to be shown. Hence, we have shown that the simplified analytical MSE expression of 1-D Standard ESPRIT with GLS for two temporally orthogonal sources coincides with the corresponding simplified deterministic CRB. This completes the proof. □

B.16. Proof of Theorem 6.2.1

For the proof of Theorem 6.2.1 in Section 6.2.1, we consider the 1-D case for simplicity and start by inserting $\boldsymbol{J}_1^{(\mathrm{nc})}$ and $\boldsymbol{J}_2^{(\mathrm{nc})}$ from (6.8) and (6.9) into (6.6), which yields

$$\begin{bmatrix} \boldsymbol{J}_1 \cdot \boldsymbol{A} \\ \boldsymbol{\Pi}_{M^{(\mathrm{sel})}} \cdot \boldsymbol{J}_2 \cdot \boldsymbol{\Pi}_M \cdot \boldsymbol{\Pi}_M \cdot \boldsymbol{A}^* \cdot \boldsymbol{\Psi}^* \cdot \boldsymbol{\Psi}^* \end{bmatrix} \cdot \boldsymbol{\Phi} = \begin{bmatrix} \boldsymbol{J}_2 \cdot \boldsymbol{A} \\ \boldsymbol{\Pi}_{M^{(\mathrm{sel})}} \cdot \boldsymbol{J}_1 \cdot \boldsymbol{\Pi}_M \cdot \boldsymbol{\Pi}_M \cdot \boldsymbol{A}^* \cdot \boldsymbol{\Psi}^* \cdot \boldsymbol{\Psi}^* \end{bmatrix}. \tag{B.380}$$

The first $M^{(\mathrm{sel})}$ rows are given by $\boldsymbol{J}_1 \cdot \boldsymbol{A} \cdot \boldsymbol{\Phi} = \boldsymbol{J}_2 \cdot \boldsymbol{A}$, which is the shift invariance equation of $\boldsymbol{A}$ that was assumed for the theorem. The second $M^{(\mathrm{sel})}$ rows can be simplified by multiplying both sides from the left by $\boldsymbol{\Pi}_{M^{(\mathrm{sel})}}$ and considering the fact that $\boldsymbol{\Pi}_M \cdot \boldsymbol{\Pi}_M = \boldsymbol{I}_M$. We have

$$\boldsymbol{J}_2 \cdot \boldsymbol{A}^* \cdot \boldsymbol{\Psi}^* \cdot \boldsymbol{\Psi}^* \cdot \boldsymbol{\Phi} = \boldsymbol{J}_1 \cdot \boldsymbol{A}^* \cdot \boldsymbol{\Psi}^* \cdot \boldsymbol{\Psi}^*. \tag{B.381}$$

As $\boldsymbol{\Psi}$ and $\boldsymbol{\Phi}$ are diagonal matrices, they commute. Then, multiplying both sides of (B.381) twice by $\boldsymbol{\Psi}$ from the right-hand side cancels $\boldsymbol{\Psi}$ as $\boldsymbol{\Psi}^* \cdot \boldsymbol{\Psi} = \boldsymbol{I}_d$ and we are left with

$$\begin{aligned} \boldsymbol{J}_2 \cdot \boldsymbol{A}^* \cdot \boldsymbol{\Phi} &= \boldsymbol{J}_1 \cdot \boldsymbol{A}^* \\ \boldsymbol{J}_2 \cdot \boldsymbol{A}^* &= \boldsymbol{J}_1 \cdot \boldsymbol{A}^* \cdot \boldsymbol{\Phi}^*, \end{aligned} \tag{B.382}$$

where in the last step, we have multiplied with $\boldsymbol{\Phi}^*$ from the right-hand side and used the fact that $\boldsymbol{\Phi}^* \cdot \boldsymbol{\Phi} = \boldsymbol{I}_d$.[4] Finally, conjugating (B.382) shows that this expression is equivalent to $\boldsymbol{J}_1 \cdot \boldsymbol{A} \cdot \boldsymbol{\Phi} = \boldsymbol{J}_2 \cdot \boldsymbol{A}$, which was again assumed for the theorem. This concludes the proof. □

[4] This equality only holds in the assumed case of undamped exponentials (cf. the model in (2.2)), where the spatial frequencies $\mu_i^{(r)}$ are real.

B.17. Proof of Theorem 6.2.2

In this appendix, we provide the proof of Theorem 6.2.2 from Section 6.2.1. Assuming that $\boldsymbol{A}$ does not necessarily satisfy (6.10), we have

$$\begin{aligned}\boldsymbol{\Pi}_{2M} \cdot \boldsymbol{A}^{(\mathrm{nc})^*} &= \begin{bmatrix} \mathbf{0} & \boldsymbol{\Pi}_M \\ \boldsymbol{\Pi}_M & \mathbf{0} \end{bmatrix} \cdot \begin{bmatrix} \boldsymbol{A}^* \\ \boldsymbol{\Pi}_M \cdot \boldsymbol{A} \cdot \boldsymbol{\Psi} \cdot \boldsymbol{\Psi} \end{bmatrix} = \begin{bmatrix} \boldsymbol{A} \cdot \boldsymbol{\Psi} \cdot \boldsymbol{\Psi} \\ \boldsymbol{\Pi}_M \cdot \boldsymbol{A}^* \end{bmatrix} \\ &= \begin{bmatrix} \boldsymbol{A} \\ \boldsymbol{\Pi}_M \cdot \boldsymbol{A}^* \cdot \boldsymbol{\Psi}^* \cdot \boldsymbol{\Psi}^* \end{bmatrix} \cdot \boldsymbol{\Psi} \cdot \boldsymbol{\Psi} = \boldsymbol{A}^{(\mathrm{nc})} \cdot \boldsymbol{\Delta}_{\mathrm{c}}, \end{aligned} \tag{B.383}$$

where $\boldsymbol{\Delta}_{\mathrm{c}}$ becomes $\boldsymbol{\Psi} \cdot \boldsymbol{\Psi}$, which is unitary and diagonal. Therefore, $\boldsymbol{A}^{(\mathrm{nc})}$ satisfies (6.10), which shows that it is centro-symmetric regardless of the centro-symmetry of $\boldsymbol{A}$. □

B.18. Proof of Equation (6.18)

In order to prove Equation (6.18) in Section 6.2.3, we notice that the real-valued transformation is carried out using sparse left $\boldsymbol{\Pi}$-real matrices of even order (cf. A.2). Expanding (6.17) yields

$$\begin{aligned}\varphi\left(\boldsymbol{X}^{(\mathrm{nc})^{(\mathrm{fba})}}\right) &= \boldsymbol{Q}_{2M}^{\mathrm{H}} \cdot \boldsymbol{X}^{(\mathrm{nc})^{(\mathrm{fba})}} \cdot \boldsymbol{Q}_{2N} \\ &= \frac{1}{2} \cdot \begin{bmatrix} \boldsymbol{I}_M & \boldsymbol{\Pi}_M \\ -\mathrm{j}\boldsymbol{I}_M & \mathrm{j}\boldsymbol{\Pi}_M \end{bmatrix} \cdot \begin{bmatrix} \boldsymbol{X}^{(\mathrm{nc})} & \boldsymbol{X}^{(\mathrm{nc})}\boldsymbol{\Pi}_N \end{bmatrix} \cdot \begin{bmatrix} \boldsymbol{I}_N & \mathrm{j}\boldsymbol{I}_N \\ \boldsymbol{\Pi}_N & -\mathrm{j}\boldsymbol{\Pi}_N \end{bmatrix} \\ &= \begin{bmatrix} \boldsymbol{I}_M & \boldsymbol{\Pi}_M \\ -\mathrm{j}\boldsymbol{I}_M & \mathrm{j}\boldsymbol{\Pi}_M \end{bmatrix} \cdot \begin{bmatrix} \boldsymbol{X} & \mathbf{0}_{M\times N} \\ \boldsymbol{\Pi}_M\boldsymbol{X}^* & \mathbf{0}_{M\times N} \end{bmatrix} \\ &= \begin{bmatrix} \boldsymbol{X} + \boldsymbol{X}^* & \mathbf{0}_{M\times N} \\ -\mathrm{j}\boldsymbol{X} + \mathrm{j}\boldsymbol{X}^* & \mathbf{0}_{M\times N} \end{bmatrix} = 2 \cdot \begin{bmatrix} \mathrm{Re}\{\boldsymbol{X}\} & \mathbf{0}_{M\times N} \\ \mathrm{Im}\{\boldsymbol{X}\} & \mathbf{0}_{M\times N} \end{bmatrix}, \end{aligned}$$

where we have used the fact that $-\mathrm{j}x + \mathrm{j}x^* = 2 \cdot \mathrm{Im}\{x\} \;\; \forall x \in \mathbb{C}$. This completes the proof. □

B.19. Proof of Theorem 6.3.1

In order to prove Theorem 6.3.1 in Section 6.3.1, we will follow the lines of the Proof B.16 in the matrix case and extend it to the tensor case. For notational convenience, we first introduce the short-hand notation

$$\boldsymbol{\mathcal{A}}^{(\mathrm{nc},r)} = \left[\boldsymbol{\mathcal{A}} \sqcup_r \tilde{\boldsymbol{\mathcal{A}}}\right] \quad \text{with} \quad \tilde{\boldsymbol{\mathcal{A}}} = \boldsymbol{\mathcal{A}}^* \bigtimes_{q=1}^{R}{}_q \boldsymbol{\Pi}_{M_q} \times_{R+1} (\boldsymbol{\Psi}^* \cdot \boldsymbol{\Psi}^*). \tag{B.384}$$

Then, the shift invariance equation in (6.27) can be written as

$$\left[\boldsymbol{\mathcal{A}} \llcorner_r \tilde{\boldsymbol{\mathcal{A}}}\right] \times_r \boldsymbol{J}_1^{(\mathrm{nc})(r)} \times_{R+1} \boldsymbol{\Phi}^{(r)} = \left[\boldsymbol{\mathcal{A}} \llcorner_r \tilde{\boldsymbol{\mathcal{A}}}\right] \times_r \boldsymbol{J}_2^{(\mathrm{nc})(r)}. \tag{B.385}$$

Using the properties (1.34) and (1.35), Equation (B.385) can be expressed as

$$\begin{aligned}&\left[\left(\boldsymbol{\mathcal{A}} \times_r \boldsymbol{J}_1^{(r)}\right) \llcorner_r \left(\tilde{\boldsymbol{\mathcal{A}}} \times_r \left(\boldsymbol{\Pi}_{M_r^{(\mathrm{sel})}} \cdot \boldsymbol{J}_2^{(r)} \cdot \boldsymbol{\Pi}_{M_r}\right)\right)\right] \times_{R+1} \boldsymbol{\Phi}^{(r)} \\ =&\left[\left(\boldsymbol{\mathcal{A}} \times_r \boldsymbol{J}_2^{(r)}\right) \llcorner_r \left(\tilde{\boldsymbol{\mathcal{A}}} \times_r \left(\boldsymbol{\Pi}_{M_r^{(\mathrm{sel})}} \cdot \boldsymbol{J}_1^{(r)} \cdot \boldsymbol{\Pi}_{M_r}\right)\right)\right].\end{aligned}$$

Therefore,

$$\begin{aligned}&\left[\left(\boldsymbol{\mathcal{A}} \times_r \boldsymbol{J}_1^{(r)} \times_{R+1} \boldsymbol{\Phi}^{(r)}\right) \llcorner_r \left(\tilde{\boldsymbol{\mathcal{A}}} \times_r \left(\boldsymbol{\Pi}_{M_r^{(\mathrm{sel})}} \cdot \boldsymbol{J}_2^{(r)} \cdot \boldsymbol{\Pi}_{M_r}\right) \times_{R+1} \boldsymbol{\Phi}^{(r)}\right)\right] \\ =&\left[\left(\boldsymbol{\mathcal{A}} \times_r \boldsymbol{J}_2^{(r)}\right) \llcorner_r \left(\tilde{\boldsymbol{\mathcal{A}}} \times_r \left(\boldsymbol{\Pi}_{M_r^{(\mathrm{sel})}} \cdot \boldsymbol{J}_1^{(r)} \cdot \boldsymbol{\Pi}_{M_r}\right)\right)\right],\end{aligned} \tag{B.386}$$

where the respective definitions of $\boldsymbol{J}_1^{(\mathrm{nc})(r)}$ and $\boldsymbol{J}_2^{(\mathrm{nc})(r)}$ from (6.8) and (6.9) have been inserted. Considering the left-hand side of the r-mode concatenation operator in (B.386), we have the shift invariance equation $\boldsymbol{\mathcal{A}} \times_r \boldsymbol{J}_1^{(r)} \times_{R+1} \boldsymbol{\Phi}^{(r)} = \boldsymbol{\mathcal{A}} \times_r \boldsymbol{J}_2^{(r)}$, which was assumed for the theorem. Hence, we are left to show that the right-hand side of the r-mode concatenation operator in (B.386) is equal, i.e.,

$$\tilde{\boldsymbol{\mathcal{A}}} \times_r \left(\boldsymbol{\Pi}_{M_r^{(\mathrm{sel})}} \cdot \boldsymbol{J}_2^{(r)} \cdot \boldsymbol{\Pi}_{M_r}\right) \times_{R+1} \boldsymbol{\Phi}^{(r)} = \tilde{\boldsymbol{\mathcal{A}}} \times_r \left(\boldsymbol{\Pi}_{M_r^{(\mathrm{sel})}} \cdot \boldsymbol{J}_1^{(r)} \cdot \boldsymbol{\Pi}_{M_r}\right). \tag{B.387}$$

Inserting the expression for $\tilde{\boldsymbol{\mathcal{A}}}$ from (B.384), (B.387) evaluates to

$$\begin{aligned}&\boldsymbol{\mathcal{A}}^* \bigtimes_{q=1}^{r-1}{}_q \boldsymbol{\Pi}_{M_q} \times_r \left(\boldsymbol{\Pi}_{M_r^{(\mathrm{sel})}} \cdot \boldsymbol{J}_2^{(r)}\right) \bigtimes_{q=r+1}^{R}{}_q \boldsymbol{\Pi}_{M_q} \times_{R+1} \left(\boldsymbol{\Phi}^{(r)} \cdot \boldsymbol{\Psi}^* \cdot \boldsymbol{\Psi}^*\right) \\ =&\boldsymbol{\mathcal{A}}^* \bigtimes_{q=1}^{r-1}{}_q \boldsymbol{\Pi}_{M_q} \times_r \left(\boldsymbol{\Pi}_{M_r^{(\mathrm{sel})}} \cdot \boldsymbol{J}_1^{(r)}\right) \bigtimes_{q=r+1}^{R}{}_q \boldsymbol{\Pi}_{M_q} \times_{R+1} \left(\boldsymbol{\Psi}^* \cdot \boldsymbol{\Psi}^*\right).\end{aligned} \tag{B.388}$$

Multiplying (B.388) with $\boldsymbol{\Pi}_{M_q}$ along all modes $q = 1, 2, \ldots, R$, $q \neq r$ and with $\boldsymbol{\Pi}_{M_r^{(\mathrm{sel})}}$ along mode r, we obtain

$$\boldsymbol{\mathcal{A}}^* \times_r \boldsymbol{J}_2^{(r)} \times_{R+1} \left(\boldsymbol{\Phi}^{(r)} \cdot \boldsymbol{\Psi}^* \cdot \boldsymbol{\Psi}^*\right) = \boldsymbol{\mathcal{A}} \times_r \boldsymbol{J}_1^{(r)} \times_{R+1} \left(\boldsymbol{\Psi}^* \cdot \boldsymbol{\Psi}^*\right). \tag{B.389}$$

Following Proof B.16, we multiply both sides of (B.389) by $\left(\boldsymbol{\Phi}^{(r)^*} \cdot \boldsymbol{\Psi} \cdot \boldsymbol{\Psi}\right)$ in the $(R+1)$-mode and conjugate the resulting expression to obtain the shift invariance equation $\boldsymbol{\mathcal{A}} \times_r \boldsymbol{J}_1^{(r)} \times_{R+1} \boldsymbol{\Phi}^{(r)} = \boldsymbol{\mathcal{A}} \times_r \boldsymbol{J}_2^{(r)}$, which was again assumed in the first place. This concludes the proof of the theorem. □

B.20. Proof of Theorem 6.3.2

To prove this theorem in Section 6.3.1, we need to show that the augmented array steering tensor $\boldsymbol{\mathcal{A}}^{(\mathrm{nc},r)}$ satisfies the tensor-based centry-symmetry property in centry-symmetry in (2.29). Inserting $\boldsymbol{\mathcal{A}}^{(\mathrm{nc},r)}$ into (2.29), it is to be shown that

$$\boldsymbol{\mathcal{A}}^{(\mathrm{nc},r)^*} \bigtimes_{q=1}^{r-1}{}_q \, \boldsymbol{\Pi}_{M_q} \times_r \boldsymbol{\Pi}_{2M_r} \bigtimes_{q=r+1}^{R}{}_q \, \boldsymbol{\Pi}_{M_q} = \boldsymbol{\mathcal{A}}^{(\mathrm{nc},r)} \times_{R+1} \boldsymbol{\Delta}_{\mathrm{c}}. \tag{B.390}$$

holds. Using the definition of $\boldsymbol{\mathcal{A}}^{(\mathrm{nc},r)}$ in (6.26), we have

$$\left[\boldsymbol{\mathcal{A}}^* \hookrightarrow_r \left(\boldsymbol{\mathcal{A}} \bigtimes_{q=1}^{R}{}_q \, \boldsymbol{\Pi}_{M_q} \times_{R+1} (\boldsymbol{\Psi} \cdot \boldsymbol{\Psi})\right)\right] \bigtimes_{q=1}^{r-1}{}_q \, \boldsymbol{\Pi}_{M_q} \times_r \boldsymbol{\Pi}_{2M_r} \bigtimes_{q=r+1}^{R}{}_q \, \boldsymbol{\Pi}_{M_q} = \boldsymbol{\mathcal{A}}^{(\mathrm{nc},r)} \times_{R+1} \boldsymbol{\Delta}_{\mathrm{c}}. \tag{B.391}$$

In order to simplify (B.391), we apply property (1.34), yielding

$$\begin{aligned}
&\left[\left(\boldsymbol{\mathcal{A}}^* \bigtimes_{\substack{q=1\\q\neq r}}^{R}{}_q \, \boldsymbol{\Pi}_{M_q}\right) \hookrightarrow_r \left(\boldsymbol{\mathcal{A}} \times_r \boldsymbol{\Pi}_{M_r} \times_{R+1} (\boldsymbol{\Psi} \cdot \boldsymbol{\Psi})\right)\right] \times_r \boldsymbol{\Pi}_{2M_r} = \boldsymbol{\mathcal{A}}^{(\mathrm{nc},r)} \times_{R+1} \boldsymbol{\Delta}_{\mathrm{c}} \\
&\left[\left(\boldsymbol{\mathcal{A}}^* \bigtimes_{\substack{q=1\\q\neq r}}^{R}{}_q \, \boldsymbol{\Pi}_{M_q}\right) \hookrightarrow_r \left(\boldsymbol{\mathcal{A}} \times_r \boldsymbol{\Pi}_{M_r} \times_{R+1} (\boldsymbol{\Psi} \cdot \boldsymbol{\Psi})\right)\right] \times_r \begin{bmatrix} \mathbf{0} & \boldsymbol{\Pi}_{M_r} \\ \boldsymbol{\Pi}_{M_r} & \mathbf{0} \end{bmatrix} = \boldsymbol{\mathcal{A}}^{(\mathrm{nc},r)} \times_{R+1} \boldsymbol{\Delta}_{\mathrm{c}}.
\end{aligned} \tag{B.392}$$

To proceed, we first combine the rules (1.35) and (1.36) to obtain the general rule

$$[\boldsymbol{\mathcal{A}} \hookrightarrow_r \boldsymbol{\mathcal{B}}] \times_r \begin{bmatrix} \mathbf{0} & \boldsymbol{\Pi}_{M_r} \\ \boldsymbol{\Pi}_{M_r} & \mathbf{0} \end{bmatrix} = [\boldsymbol{\mathcal{B}} \times_r \boldsymbol{\Pi}_{M_r} \hookrightarrow_r \boldsymbol{\mathcal{A}} \times_r \boldsymbol{\Pi}_{M_r}], \tag{B.393}$$

where the tensors have the dimensions $\boldsymbol{\mathcal{A}}, \boldsymbol{\mathcal{B}} \in \mathbb{C}^{M_1 \times \ldots \times M_R}$. Applying (B.393) to (B.392) gives

$$\begin{aligned}
&\left[(\boldsymbol{\mathcal{A}} \times_{R+1} (\boldsymbol{\Psi} \cdot \boldsymbol{\Psi})) \hookrightarrow_r \left(\boldsymbol{\mathcal{A}}^* \bigtimes_{q=1}^{R}{}_q \, \boldsymbol{\Pi}_{M_q}\right)\right] = \boldsymbol{\mathcal{A}}^{(\mathrm{nc},r)} \times_{R+1} \boldsymbol{\Delta}_{\mathrm{c}} \\
&\left[\boldsymbol{\mathcal{A}} \hookrightarrow_r \left(\boldsymbol{\mathcal{A}}^* \bigtimes_{q=1}^{R}{}_q \, \boldsymbol{\Pi}_{M_q} \times_{R+1} (\boldsymbol{\Psi}^* \cdot \boldsymbol{\Psi}^*)\right)\right] \times_{R+1} (\boldsymbol{\Psi} \cdot \boldsymbol{\Psi}) = \boldsymbol{\mathcal{A}}^{(\mathrm{nc},r)} \times_{R+1} \boldsymbol{\Delta}_{\mathrm{c}}.
\end{aligned} \tag{B.394}$$

Considering the definition of $\boldsymbol{\mathcal{A}}^{(\mathrm{nc},r)}$ in (6.26) and for $\boldsymbol{\Delta}_{\mathrm{c}} = \boldsymbol{\Psi} \cdot \boldsymbol{\Psi}$, (B.394) is a true statement and completes the proof. □

B.21. Proof of Theorem 6.3.3

In this appendix, we provide a proof of Theorem 6.3.3 in Section 6.3.2. When computing the truncated HOSVD of $\boldsymbol{\mathcal{X}}^{(\mathrm{nc},r)}$, special attention must be paid with respect to the n-ranks. While for R-D Tensor-ESPRIT-type algorithms the n-ranks of $\boldsymbol{\mathcal{X}}_0$ are always less than or equal to d, for R-D NC Tensor-ESPRIT the n-ranks can exceed d.

Consider the noise-free observation tensor $\boldsymbol{\mathcal{X}}^{(\mathrm{nc},r)} = \boldsymbol{\mathcal{A}}^{(\mathrm{nc},r)} \times_{R+1} \boldsymbol{S}^{\mathrm{T}}$ whose n-ranks are determined by the n-ranks of $\boldsymbol{\mathcal{A}}^{(\mathrm{nc},r)}$. Using the short-had notation $\boldsymbol{\mathcal{A}}^{(\mathrm{nc},r)} = \left[\boldsymbol{\mathcal{A}} \sqcup_r \tilde{\boldsymbol{\mathcal{A}}}\right]$ from Appendix B.19, we can write the r-mode unfolding $\left[\boldsymbol{\mathcal{A}}^{(\mathrm{nc},r)}\right]_{(r)}$ as

$$\left[\boldsymbol{\mathcal{A}}^{(\mathrm{nc},r)}\right]_{(r)} = \begin{bmatrix} [\boldsymbol{\mathcal{A}}]_{(r)} \\ \left[\tilde{\boldsymbol{\mathcal{A}}}\right]_{(r)} \end{bmatrix}. \tag{B.395}$$

where we have applied the property in (1.29). Next, we apply property (1.37) to the array steering tensor $\boldsymbol{\mathcal{A}}$ defined in (2.10), which yields for $[\boldsymbol{\mathcal{A}}]_{(r)}$ and $\left[\tilde{\boldsymbol{\mathcal{A}}}\right]_{(r)}$ the expressions

$$[\boldsymbol{\mathcal{A}}]_{(r)} = \boldsymbol{A}^{(r)} \cdot \underbrace{\left(\left(\bigdiamond_{q=r+1}^{R} {}^{q}\boldsymbol{A}^{(q)}\right) \diamond \boldsymbol{I}_d \diamond \left(\bigdiamond_{q=1}^{r-1} {}^{q}\boldsymbol{A}^{(q)}\right)\right)^{\mathrm{T}}}_{\boldsymbol{B}^{(r)\mathrm{T}}} \tag{B.396}$$

$$\left[\tilde{\boldsymbol{\mathcal{A}}}\right]_{(r)} = \boldsymbol{\Pi}_{M_r} \cdot \boldsymbol{A}^{(r)*} \cdot \underbrace{\left(\left(\bigdiamond_{q=r+1}^{R} {}^{q}\,\boldsymbol{\Pi}_{M_q} \cdot \boldsymbol{A}^{(q)*}\right) \diamond (\boldsymbol{\Psi}^* \cdot \boldsymbol{\Psi}^*) \diamond \left(\bigdiamond_{q=1}^{r-1} {}^{q}\,\boldsymbol{\Pi}_{M_q} \cdot \boldsymbol{A}^{(q)*}\right)\right)^{\mathrm{T}}}_{\tilde{\boldsymbol{B}}^{(r)\mathrm{T}}}. \tag{B.397}$$

With these relations, $\left[\boldsymbol{\mathcal{A}}^{(\mathrm{nc},r)}\right]_{(r)}$ in (B.395) becomes

$$\left[\boldsymbol{\mathcal{A}}^{(\mathrm{nc},r)}\right]_{(r)} = \begin{bmatrix} \boldsymbol{A}^{(r)} \cdot \boldsymbol{B}^{(r)\mathrm{T}} \\ \boldsymbol{\Pi}_{M_r} \cdot \boldsymbol{A}^{(r)*} \cdot \tilde{\boldsymbol{B}}^{(r)\mathrm{T}} \end{bmatrix} = \begin{bmatrix} \boldsymbol{A}^{(r)} & \boldsymbol{0}_{M_r \times d} \\ \boldsymbol{0}_{M_r \times d} & \boldsymbol{\Pi}_{M_r} \cdot \boldsymbol{A}^{(r)*} \end{bmatrix} \cdot \begin{bmatrix} \boldsymbol{B}^{(r)} & \tilde{\boldsymbol{B}}^{(r)} \end{bmatrix}^{\mathrm{T}}. \tag{B.398}$$

According to (B.398), $\left[\boldsymbol{\mathcal{A}}^{(\mathrm{nc},r)}\right]_{(r)}$ can be decomposed into the product of a matrix of size $2M_r \times 2d$ and a matrix of size $2d \times M \cdot d/M_r$. The former matrix has rank $2d$ if $\boldsymbol{A}^{(r)}$ has full column rank and the latter matrix has rank d if the array is centro-symmetric in all R modes as in this case, i.e., $\boldsymbol{B}^{(r)}$ and $\tilde{\boldsymbol{B}}^{(r)}$ have the same column space. However, if the array is not centro-symmetric in the q-th mode $q = 1, 2, \ldots, R$, $q \neq r$, $\boldsymbol{B}^{(r)}$ and $\tilde{\boldsymbol{B}}^{(r)}$ do not share the same column space and the latter matrix in (B.398) has a rank higher than d. Consequently, the n-ranks for $n = 1, 2, \ldots, R$ of the augmented array steering tensor can exceed d and even reach $2d$ in the worst case. □

B.22. Proof of Theorem 6.3.4

For the proof of Theorem 6.3.4 in Section 6.3.2, we consider the noise-free augmented measurement tensor from (6.25)

$$\boldsymbol{\mathcal{X}}_0^{(\mathrm{nc},r)} = \boldsymbol{\mathcal{A}}^{(\mathrm{nc},r)} \times_{R+1} \boldsymbol{S}^{\mathrm{T}} \tag{B.399}$$

and its HOSVD from (6.28) given by

$$\boldsymbol{\mathcal{X}}_0^{(\mathrm{nc},r)} = \boldsymbol{\mathcal{S}}^{[\mathrm{s}](r)} \times_1 \boldsymbol{U}_1^{[\mathrm{s}](r)} \ldots \times_R \boldsymbol{U}_R^{[\mathrm{s}](r)} \times_{R+1} \boldsymbol{U}_{R+1}^{[\mathrm{s}](r)}. \tag{B.400}$$

Then, the augmented signal subspace tensor $\boldsymbol{\mathcal{U}}^{[\mathrm{s}](r)}$ in (6.32) can be computed via

$$\boldsymbol{\mathcal{U}}^{[\mathrm{s}](r)} = \boldsymbol{\mathcal{X}}^{(\mathrm{nc},r)} \times_{R+1} \left(\boldsymbol{\Sigma}_{R+1}^{[\mathrm{s}](r)^{-1}} \cdot \left(\boldsymbol{U}_{R+1}^{[\mathrm{s}](r)} \right)^{\mathrm{H}} \right). \tag{B.401}$$

Inserting (B.399) into (B.401) yields

$$\begin{aligned} \boldsymbol{\mathcal{U}}^{[\mathrm{s}](r)} = \boldsymbol{\mathcal{A}}^{(\mathrm{nc},r)} \times_{R+1} \left(\boldsymbol{\Sigma}_{R+1}^{[\mathrm{s}](r)^{-1}} \cdot \left(\boldsymbol{U}_{R+1}^{[\mathrm{s}](r)} \right)^{\mathrm{H}} \cdot \boldsymbol{S}^{\mathrm{T}} \right) \\ \boldsymbol{\mathcal{A}}^{(\mathrm{nc},r)} \times_{R+1} \boldsymbol{T}_r, \end{aligned} \tag{B.402}$$

where we have defined $\boldsymbol{T}_r = \boldsymbol{\Sigma}_{R+1}^{[\mathrm{s}](r)^{-1}} \cdot \left(\boldsymbol{U}_{R+1}^{[\mathrm{s}](r)} \right)^{\mathrm{H}} \cdot \boldsymbol{S}^{\mathrm{T}}$.

In the next step, we need to show that $\boldsymbol{T}_r$ in (B.402) is independent of r. This property is required for the matrices $\boldsymbol{\Psi}^{(r)}$ to have common eigenvectors in order to enable the correct pairing of the parameter estimates in R-D ESPRIT-type algorithms. Note that $\boldsymbol{\Sigma}_{R+1}^{[\mathrm{s}](r)}$ and $\boldsymbol{U}_{R+1}^{[\mathrm{s}](r)}$ represent the right singular vectors and singular values of $\left[\boldsymbol{\mathcal{X}}^{(\mathrm{nc},r)}\right]_{(R+1)}$. Therefore, they can be computed via the eigenvalue decomposition of $\left[\boldsymbol{\mathcal{X}}^{(\mathrm{nc},r)}\right]_{(R+1)} \cdot \left[\boldsymbol{\mathcal{X}}^{(\mathrm{nc},r)}\right]_{(R+1)}^{\mathrm{H}}$.

To proceed, we require the following lemma:

Lemma B.22.1. *The augmented measurement tensors $\boldsymbol{\mathcal{X}}^{(\mathrm{nc},r)}$ in the r-th mode satisfy the property*

$$\left[\boldsymbol{\mathcal{X}}^{(\mathrm{nc},r)}\right]_{(R+1)} \cdot \left[\boldsymbol{\mathcal{X}}^{(\mathrm{nc},r)}\right]_{(R+1)}^{\mathrm{H}} = 2 \cdot \mathrm{Re}\left\{ [\boldsymbol{\mathcal{X}}]_{(R+1)} \cdot [\boldsymbol{\mathcal{X}}]_{(R+1)}^{\mathrm{H}} \right\}. \tag{B.403}$$

Proof. The $(R+1)$-mode unfolding of $\boldsymbol{\mathcal{X}}^{(\mathrm{nc},r)}$ contains all $(R+1)$-mode vectors of $\boldsymbol{\mathcal{X}}^{(\mathrm{nc},r)}$. As $\boldsymbol{\mathcal{X}}^{(\mathrm{nc},r)}$ is the r-mode concatenation of $\boldsymbol{\mathcal{X}}$ and $\boldsymbol{\mathcal{X}}^* \times_1 \boldsymbol{\Pi}_{M_1} \ldots \times_R \boldsymbol{\Pi}_{M_R}$, its $(R+1)$-mode unfolding contains the $(R+1)$-mode vectors of $\boldsymbol{\mathcal{X}}$ and the $(R+1)$-mode vectors of $\boldsymbol{\mathcal{X}}^*$ in a permuted order, i.e.,

$$\left[\boldsymbol{\mathcal{X}}^{(\mathrm{nc},r)}\right]_{(R+1)} = \begin{bmatrix} [\boldsymbol{\mathcal{X}}]_{(R+1)} & [\boldsymbol{\mathcal{X}}]_{(R+1)}^* \end{bmatrix} \cdot \boldsymbol{P}_r, \tag{B.404}$$

where $\boldsymbol{P}_r$ is a permutation matrix. Since all permutation matrices satisfy $\boldsymbol{P}_r^{\mathrm{H}} = \boldsymbol{P}_r^{-1}$, we can establish that $\left[\boldsymbol{\mathcal{X}}^{(\mathrm{nc},r)}\right]_{(R+1)} \cdot \left[\boldsymbol{\mathcal{X}}^{(\mathrm{nc},r)}\right]_{(R+1)}^{\mathrm{H}} = [\boldsymbol{\mathcal{X}}]_{(R+1)} \cdot [\boldsymbol{\mathcal{X}}]_{(R+1)}^{\mathrm{H}} + [\boldsymbol{\mathcal{X}}]_{(R+1)}^{*} \cdot [\boldsymbol{\mathcal{X}}]_{(R+1)}^{\mathrm{T}} = 2 \cdot \mathrm{Re}\left\{[\boldsymbol{\mathcal{X}}]_{(R+1)} \cdot [\boldsymbol{\mathcal{X}}]_{(R+1)}^{\mathrm{H}}\right\}$, which proves the lemma. □

From this lemma, we can conclude that $\left[\boldsymbol{\mathcal{X}}^{(\mathrm{nc},r)}\right]_{(R+1)} \cdot \left[\boldsymbol{\mathcal{X}}^{(\mathrm{nc},r)}\right]_{(R+1)}^{\mathrm{H}}$ is independent of r. As a result, $\boldsymbol{\Sigma}_{R+1}^{[\mathrm{s}](r)}$ and $\boldsymbol{U}_{R+1}^{[\mathrm{s}](r)}$ must be equal for all $r = 1, 2, \ldots, R$ such that $\boldsymbol{T}_r$ is also equal for all $r = 1, 2, \ldots, R$. Therefore, we have $\boldsymbol{\mathcal{U}}^{[\mathrm{s}](r)} = \boldsymbol{\mathcal{A}}^{(\mathrm{nc},r)} \times_{R+1} \boldsymbol{T}$, which proves the theorem. □

B.23. Proof of Equation (6.40)

For the proof of Equation (6.40) in Section 6.3.3, we start with the definition of the FBA-processed r-mode augmented measurement tensor $\boldsymbol{\mathcal{X}}^{(\mathrm{nc},r)^{(\mathrm{fba})}}$ in (6.38) and insert $\boldsymbol{\mathcal{X}}^{(\mathrm{nc},r)} = \left[\boldsymbol{\mathcal{X}} \hookrightarrow_r \tilde{\boldsymbol{\mathcal{X}}}\right]$ from (6.24), where we introduce the short-hand notation $\tilde{\boldsymbol{\mathcal{X}}} = \boldsymbol{\mathcal{X}}^* \bigtimes_{q=1}^{R}{}_q \boldsymbol{\Pi}_{M_q}$. We obtain

$$\boldsymbol{\mathcal{X}}^{(\mathrm{nc},r)^{(\mathrm{fba})}} = \left[\left[\boldsymbol{\mathcal{X}} \hookrightarrow_r \tilde{\boldsymbol{\mathcal{X}}}\right] \hookrightarrow_{R+1} \left[(\boldsymbol{\mathcal{X}} \times_{R+1} \boldsymbol{\Pi}_N) \hookrightarrow_r (\tilde{\boldsymbol{\mathcal{X}}} \times_{R+1} \boldsymbol{\Pi}_N)\right]\right], \tag{B.405}$$

where we have used the properties (1.34), (1.35), and (1.36). Then, we first consider the $(R+1)$-mode product of $\boldsymbol{\mathcal{X}}^{(\mathrm{nc},r)^{(\mathrm{fba})}}$ with $\boldsymbol{Q}_{2N}^{(\mathrm{s})}$, where the latter is an example of the sparse left-$\boldsymbol{\Pi}$-real matrices $\boldsymbol{Q}_p^{(\mathrm{s})}$ introduced in [HN95] (cf. Appendix A.2), which yields

$$\begin{aligned}
&\boldsymbol{\mathcal{X}}^{(\mathrm{nc},r)^{(\mathrm{fba})}} \times_{R+1} \frac{1}{\sqrt{2}} \begin{bmatrix} \boldsymbol{I}_N & \boldsymbol{\Pi}_N \\ -\mathrm{j}\boldsymbol{I}_N & \mathrm{j}\boldsymbol{\Pi}_N \end{bmatrix} \\
&= \left[\left[\boldsymbol{\mathcal{X}} \hookrightarrow_r \tilde{\boldsymbol{\mathcal{X}}}\right] \hookrightarrow_{R+1} \left[(\boldsymbol{\mathcal{X}} \times_{R+1} \boldsymbol{\Pi}_N) \hookrightarrow_r (\tilde{\boldsymbol{\mathcal{X}}} \times_{R+1} \boldsymbol{\Pi}_N)\right]\right] \times_{R+1} \frac{1}{\sqrt{2}} \begin{bmatrix} \boldsymbol{I}_N & \boldsymbol{\Pi}_N \\ -\mathrm{j}\boldsymbol{I}_N & \mathrm{j}\boldsymbol{\Pi}_N \end{bmatrix} \\
&= \frac{1}{\sqrt{2}} \cdot \left[\left(\boldsymbol{\mathcal{X}} \times_{R+1} \begin{bmatrix} \boldsymbol{I}_N \\ -\mathrm{j}\boldsymbol{I}_N \end{bmatrix}\right) \hookrightarrow_r \left(\tilde{\boldsymbol{\mathcal{X}}} \times_{R+1} \begin{bmatrix} \boldsymbol{I}_N \\ -\mathrm{j}\boldsymbol{I}_N \end{bmatrix}\right)\right] \\
&\quad + \frac{1}{\sqrt{2}} \cdot \left[\left(\boldsymbol{\mathcal{X}} \times_{R+1} \left(\begin{bmatrix} \boldsymbol{\Pi}_N \\ \mathrm{j}\boldsymbol{\Pi}_N \end{bmatrix} \cdot \boldsymbol{\Pi}_N\right)\right) \hookrightarrow_r \left(\tilde{\boldsymbol{\mathcal{X}}} \times_{R+1} \left(\begin{bmatrix} \boldsymbol{\Pi}_N \\ \mathrm{j}\boldsymbol{\Pi}_N \end{bmatrix} \cdot \boldsymbol{\Pi}_N\right)\right)\right] \\
&= \frac{1}{\sqrt{2}} \cdot \left[\left(\boldsymbol{\mathcal{X}} \times_{R+1} \begin{bmatrix} \boldsymbol{I}_N \\ -\mathrm{j}\boldsymbol{I}_N \end{bmatrix}\right) \hookrightarrow_r \left(\tilde{\boldsymbol{\mathcal{X}}} \times_{R+1} \begin{bmatrix} \boldsymbol{I}_N \\ -\mathrm{j}\boldsymbol{I}_N \end{bmatrix}\right)\right] + \frac{1}{\sqrt{2}} \cdot \left[\left(\boldsymbol{\mathcal{X}} \times_{R+1} \begin{bmatrix} \boldsymbol{I}_N \\ \mathrm{j}\boldsymbol{I}_N \end{bmatrix}\right) \hookrightarrow_r \left(\tilde{\boldsymbol{\mathcal{X}}} \times_{R+1} \begin{bmatrix} \boldsymbol{I}_N \\ \mathrm{j}\boldsymbol{I}_N \end{bmatrix}\right)\right] \\
&= \frac{1}{\sqrt{2}} \cdot \left[\left(\boldsymbol{\mathcal{X}} \times_{R+1} \begin{bmatrix} 2\boldsymbol{I}_N \\ \boldsymbol{0}_{N\times N} \end{bmatrix}\right) \hookrightarrow_r \left(\tilde{\boldsymbol{\mathcal{X}}} \times_{R+1} \begin{bmatrix} 2\boldsymbol{I}_N \\ \boldsymbol{0}_{N\times N} \end{bmatrix}\right)\right] \\
&= \frac{2}{\sqrt{2}} \cdot \left[\left[\boldsymbol{\mathcal{X}} \hookrightarrow_{R+1} \boldsymbol{\mathcal{O}}_{M_1\times\ldots\times M_R\times N}\right] \hookrightarrow_r \left[\tilde{\boldsymbol{\mathcal{X}}} \hookrightarrow_{R+1} \boldsymbol{\mathcal{O}}_{M_1\times\ldots\times M_R\times N}\right]\right],
\end{aligned} \tag{B.406}$$

where we have once again used the properties (1.34), (1.35), (1.36), and (1.33). Note that the last N blocks in (B.406) are zero and can be skipped. Thus, we further proceed with the non-zero part $\sqrt{2} \cdot \left[\boldsymbol{\mathcal{X}} \sqcup_r \tilde{\boldsymbol{\mathcal{X}}} \right] \in \mathbb{C}^{M_1 \times \ldots \times 2M_r \times \ldots \times M_R \times N}$. To this end, we require the following lemma:

Lemma B.23.1. *Defining* $\bar{\boldsymbol{\mathcal{X}}}^{(r)} = \boldsymbol{\mathcal{X}} \bigtimes_{\substack{q=1 \\ q \neq r}}^{R}{}_q \boldsymbol{Q}_{M_q}^{\mathrm{H}}$, *the identity* $\tilde{\boldsymbol{\mathcal{X}}} \bigtimes_{\substack{q=1 \\ q \neq r}}^{R}{}_q \boldsymbol{Q}_{M_q}^{\mathrm{H}} = \bar{\boldsymbol{\mathcal{X}}}^{(r)^*} \times_r \boldsymbol{\Pi}_{M_r}$ *holds true.*

Proof. This identity follows from inserting $\tilde{\boldsymbol{\mathcal{X}}}$ from above into the left hand side, which becomes

$$\begin{aligned} \tilde{\boldsymbol{\mathcal{X}}} \bigtimes_{\substack{q=1 \\ q \neq r}}^{R}{}_q \boldsymbol{Q}_{M_q}^{\mathrm{H}} &= \boldsymbol{\mathcal{X}}^* \bigtimes_{\substack{q=1 \\ q \neq r}}^{R}{}_q \left(\boldsymbol{Q}_{M_q}^{\mathrm{H}} \cdot \boldsymbol{\Pi}_{M_q} \right) \times_r \boldsymbol{\Pi}_{M_r} = \boldsymbol{\mathcal{X}}^* \bigtimes_{\substack{q=1 \\ q \neq r}}^{R}{}_q \boldsymbol{Q}_{M_q}^{\mathrm{T}} \times_r \boldsymbol{\Pi}_{M_r} \\ &= \left(\boldsymbol{\mathcal{X}} \bigtimes_{\substack{q=1 \\ q \neq r}}^{R}{}_q \boldsymbol{Q}_{M_q}^{\mathrm{H}} \right)^* \times_r \boldsymbol{\Pi}_{M_r} = \bar{\boldsymbol{\mathcal{X}}}^{(r)^*} \times_r \boldsymbol{\Pi}_{M_r}, \end{aligned} \tag{B.407}$$

where we have applied the property that $\boldsymbol{Q}_p^{\mathrm{H}} \cdot \boldsymbol{\Pi}_p = \boldsymbol{Q}_p^{\mathrm{T}}$ as $\boldsymbol{Q}_p$ is left-$\boldsymbol{\Pi}$-real. This concludes the proof of the lemma. □

Using Lemma B.23.1, we can express the product of the non-zero part with $\boldsymbol{Q}_{M_q}^{\mathrm{H}}$ for $q = 1, 2, \ldots, R$, $q \neq r$ as

$$\sqrt{2} \cdot \left[\boldsymbol{\mathcal{X}} \sqcup_r \tilde{\boldsymbol{\mathcal{X}}} \right] = \sqrt{2} \cdot \left[\bar{\boldsymbol{\mathcal{X}}} \sqcup_r \bar{\boldsymbol{\mathcal{X}}}^* \times_r \boldsymbol{\Pi}_{M_r} \right]. \tag{B.408}$$

In the final step to compute $\varphi\left(\boldsymbol{\mathcal{X}}^{(\mathrm{nc},r)^{(\mathrm{fba})}} \right)$, we turn to the product with $\boldsymbol{Q}_{2M_r}^{(\mathrm{s})^{\mathrm{H}}}$ in the r-th mode, which yields

$$\begin{aligned} \varphi\left(\boldsymbol{\mathcal{X}}^{(\mathrm{nc},r)^{(\mathrm{fba})}} \right) &= \sqrt{2} \cdot \left[\bar{\boldsymbol{\mathcal{X}}} \sqcup_r \bar{\boldsymbol{\mathcal{X}}}^* \times_r \boldsymbol{\Pi}_{M_r} \right] \times_r \frac{1}{\sqrt{2}} \cdot \begin{bmatrix} \boldsymbol{I}_{M_r} & \boldsymbol{\Pi}_{M_r} \\ -\mathrm{j}\boldsymbol{I}_{M_r} & \mathrm{j}\boldsymbol{\Pi}_{M_r} \end{bmatrix} \\ &= \left(\bar{\boldsymbol{\mathcal{X}}} \times_r \begin{bmatrix} \boldsymbol{I}_{M_r} \\ -\mathrm{j}\boldsymbol{I}_{M_r} \end{bmatrix} \right) + \left(\bar{\boldsymbol{\mathcal{X}}}^* \times_r \left(\begin{bmatrix} \boldsymbol{\Pi}_{M_r} \\ \mathrm{j}\boldsymbol{\Pi}_{M_r} \end{bmatrix} \cdot \boldsymbol{\Pi}_{M_r} \right) \right) \\ &= \left(\bar{\boldsymbol{\mathcal{X}}} \times_r \begin{bmatrix} \boldsymbol{I}_{M_r} \\ -\mathrm{j}\boldsymbol{I}_{M_r} \end{bmatrix} \right) + \left(\bar{\boldsymbol{\mathcal{X}}}^* \times_r \begin{bmatrix} \boldsymbol{I}_{M_r} \\ \mathrm{j}\boldsymbol{I}_{M_r} \end{bmatrix} \right) \\ &= \left[(\bar{\boldsymbol{\mathcal{X}}} + \bar{\boldsymbol{\mathcal{X}}}^*) \sqcup_r (-\mathrm{j}\bar{\boldsymbol{\mathcal{X}}} + \mathrm{j}\bar{\boldsymbol{\mathcal{X}}}^*) \right] \\ &= 2 \cdot \left[\mathrm{Re}\left\{ \bar{\boldsymbol{\mathcal{X}}} \right\} \sqcup_r \mathrm{Im}\left\{ \bar{\boldsymbol{\mathcal{X}}} \right\} \right] \end{aligned} \tag{B.409}$$

which proves the theorem. □

B.24. Proof of Theorem 6.4.1

In order to prove this theorem in Section 6.4.1, we first expand $\boldsymbol{n}^{(\mathrm{nc})} \in \mathbb{C}^{2MN\times 1}$ as

$$\boldsymbol{n}^{(\mathrm{nc})} = \mathrm{vec}\left\{\boldsymbol{N}^{(\mathrm{nc})}\right\} = \mathrm{vec}\left\{\begin{bmatrix}\boldsymbol{N}\\ \boldsymbol{\Pi}_M\cdot\boldsymbol{N}^*\end{bmatrix}\right\} = \boldsymbol{K}_{2M,N}^{\mathrm{T}}\cdot\begin{bmatrix}\mathrm{vec}\left\{\boldsymbol{N}^{\mathrm{T}}\right\}\\ \mathrm{vec}\left\{(\boldsymbol{\Pi}_M\cdot\boldsymbol{N}^*)^{\mathrm{T}}\right\}\end{bmatrix} \tag{B.410}$$

$$= \boldsymbol{K}_{2M,N}^{\mathrm{T}}\cdot\begin{bmatrix}\boldsymbol{K}_{M,N}\cdot\mathrm{vec}\{\boldsymbol{N}\}\\ \boldsymbol{K}_{M,N}\cdot\mathrm{vec}\{\boldsymbol{\Pi}_M\cdot\boldsymbol{N}^*\}\end{bmatrix} = \boldsymbol{K}_{2M,N}^{\mathrm{T}}\cdot(\boldsymbol{I}_2\otimes\boldsymbol{K}_{M,N})\cdot\begin{bmatrix}\mathrm{vec}\{\boldsymbol{N}\}\\ \mathrm{vec}\{\boldsymbol{\Pi}_M\cdot\boldsymbol{N}^*\}\end{bmatrix}, \tag{B.411}$$

where $\boldsymbol{K}_{M,N}$ is the commutation matrix of size $MN\times MN$ defined in (1.21) and we have applied property (1.21) to the last two identities in (B.410). Recalling that $\boldsymbol{n} = \mathrm{vec}\{\boldsymbol{N}\} \in \mathbb{C}^{MN\times 1}$ and using property (1.14), we can formulate (B.411) as

$$\boldsymbol{n}^{(\mathrm{nc})} = \boldsymbol{K}_{2M,N}^{\mathrm{T}}\cdot(\boldsymbol{I}_2\otimes\boldsymbol{K}_{M,N})\cdot\begin{bmatrix}\boldsymbol{n}\\ (\boldsymbol{I}_N\otimes\boldsymbol{\Pi}_M)\cdot\boldsymbol{n}^*\end{bmatrix} = \tilde{\boldsymbol{K}}\cdot\begin{bmatrix}\boldsymbol{n}\\ \boldsymbol{n}^*\end{bmatrix}, \tag{B.412}$$

where $\tilde{\boldsymbol{K}} = \boldsymbol{K}_{2M,N}^{\mathrm{T}}\cdot\mathrm{blkdiag}\{\boldsymbol{K}_{M,N}, \boldsymbol{K}_{M,N}\cdot(\boldsymbol{I}_N\otimes\boldsymbol{\Pi}_M)\} \in \mathbb{R}^{2MN\times 2MN}$. Hence, the SO statistics of $\boldsymbol{n}^{(\mathrm{nc})}$ can be expressed by means of the covariance matrix $\boldsymbol{R}_{\mathrm{nn}} = \mathbb{E}\left\{\boldsymbol{n}\cdot\boldsymbol{n}^{\mathrm{H}}\right\}$ and the pseudo-covariance matrix $\boldsymbol{C}_{\mathrm{nn}} = \mathbb{E}\left\{\boldsymbol{n}\cdot\boldsymbol{n}^{\mathrm{T}}\right\}$ of the physical noise $\boldsymbol{n}$. We obtain

$$\boldsymbol{R}_{\mathrm{nn}}^{(\mathrm{nc})} = \mathbb{E}\left\{\boldsymbol{n}^{(\mathrm{nc})}\cdot\boldsymbol{n}^{(\mathrm{nc})^{\mathrm{H}}}\right\} = \tilde{\boldsymbol{K}}\cdot\begin{bmatrix}\boldsymbol{R}_{\mathrm{nn}} & \boldsymbol{C}_{\mathrm{nn}}\\ \boldsymbol{C}_{\mathrm{nn}}^* & \boldsymbol{R}_{\mathrm{nn}}^*\end{bmatrix}\cdot\tilde{\boldsymbol{K}}^{\mathrm{H}} \tag{B.413}$$

$$\boldsymbol{C}_{\mathrm{nn}}^{(\mathrm{nc})} = \mathbb{E}\left\{\boldsymbol{n}^{(\mathrm{nc})}\cdot\boldsymbol{n}^{(\mathrm{nc})^{\mathrm{T}}}\right\} = \tilde{\boldsymbol{K}}\cdot\begin{bmatrix}\boldsymbol{C}_{\mathrm{nn}} & \boldsymbol{R}_{\mathrm{nn}}\\ \boldsymbol{R}_{\mathrm{nn}}^* & \boldsymbol{C}_{\mathrm{nn}}^*\end{bmatrix}\cdot\tilde{\boldsymbol{K}}^{\mathrm{T}}. \tag{B.414}$$

This completes the proof. □

B.25. Proof of Theorem 6.4.2

For the proof of Theorem 6.4.2 in Section 6.4.2, we follow the steps in Appendix B.13 and only consider the 1-D case, which can be extended to the R-D case straightforwardly. Again, we first develop a first-order error expansion for the real-valued NC Unitary ESPRIT algorithm and then show its equivalence to the complex-valued NC Standard ESPRIT algorithm with FBA. To this end, let $\boldsymbol{X}_0^{(\mathrm{nc})^{(\mathrm{fba})}} \in \mathbb{C}^{2M\times 2N}$ be the noise-free forward-backward averaged measurement matrix defined by decomposing (6.15) according to

$$\begin{aligned}\boldsymbol{X}^{(\mathrm{nc})^{(\mathrm{fba})}} &= \begin{bmatrix}\boldsymbol{X}_0^{(\mathrm{nc})} & \boldsymbol{X}_0^{(\mathrm{nc})}\cdot\boldsymbol{\Pi}_N\end{bmatrix} + \begin{bmatrix}\boldsymbol{N}^{(\mathrm{nc})} & \boldsymbol{N}^{(\mathrm{nc})}\cdot\boldsymbol{\Pi}_N\end{bmatrix}\\ &= \boldsymbol{X}_0^{(\mathrm{nc})^{(\mathrm{fba})}} + \boldsymbol{N}^{(\mathrm{nc})^{(\mathrm{fba})}}.\end{aligned} \tag{B.415}$$

Its SVD can be expressed as

$$\boldsymbol{X}_0^{(\mathrm{nc})^{(\mathrm{fba})}} = \begin{bmatrix} \boldsymbol{U}_\mathrm{s}^{(\mathrm{nc})^{(\mathrm{fba})}} & \boldsymbol{U}_\mathrm{n}^{(\mathrm{nc})^{(\mathrm{fba})}} \end{bmatrix} \cdot \begin{bmatrix} \boldsymbol{\Sigma}_\mathrm{s}^{(\mathrm{nc})^{(\mathrm{fba})}} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} \end{bmatrix} \cdot \begin{bmatrix} \boldsymbol{V}_\mathrm{s}^{(\mathrm{nc})^{(\mathrm{fba})}} & \boldsymbol{V}_\mathrm{n}^{(\mathrm{nc})^{(\mathrm{fba})}} \end{bmatrix}^\mathrm{H}$$

such that the complex-valued shift invariance equation for the forward-backward-averaged data has the form

$$\boldsymbol{J}_1^{(\mathrm{nc})} \cdot \boldsymbol{U}_\mathrm{s}^{(\mathrm{nc})^{(\mathrm{fba})}} \cdot \boldsymbol{\Psi} = \boldsymbol{J}_2^{(\mathrm{nc})} \cdot \boldsymbol{U}_\mathrm{s}^{(\mathrm{nc})^{(\mathrm{fba})}}, \tag{B.416}$$

where $\boldsymbol{\Psi} = \boldsymbol{Q}^{(\mathrm{fba})} \cdot \boldsymbol{\Lambda} \cdot \boldsymbol{Q}^{(\mathrm{fba})^{-1}}$ and $\boldsymbol{\Lambda} = \operatorname{diag}\left\{\begin{bmatrix}\lambda_1, \ldots, \lambda_d\end{bmatrix}\right\}$ with $\lambda_i = \mathrm{e}^{\mathrm{j}\mu_i}$, $i = 1, 2, \ldots, d$. Performing the same steps as in Section 6.4.1, the first-order approximation of the estimation error after the application of FBA is given by

$$\Delta\mu_i \approx \operatorname{Im}\left\{\boldsymbol{p}_i^{(\mathrm{fba})^\mathrm{T}} \cdot \left(\boldsymbol{J}_1^{(\mathrm{nc})} \cdot \boldsymbol{U}_\mathrm{s}^{(\mathrm{nc})^{(\mathrm{fba})}}\right)^+ \cdot \left[\boldsymbol{J}_2^{(\mathrm{nc})}/\lambda_i - \boldsymbol{J}_1^{(\mathrm{nc})}\right] \cdot \Delta\boldsymbol{U}_\mathrm{s}^{(\mathrm{nc})^{(\mathrm{fba})}} \cdot \boldsymbol{q}_i^{(\mathrm{fba})}\right\}, \tag{B.417}$$

where we have simply replaced the corresponding quantities in (6.47) by their FBA versions. Next, we show that the estimation error expansion for the real-valued case is equivalent to (B.417).

The 1-D real-valued shift-invariance equation is given by

$$\boldsymbol{K}_1^{(\mathrm{nc})} \cdot \boldsymbol{E}_\mathrm{s}^{(\mathrm{nc})} \cdot \boldsymbol{\Upsilon} = \boldsymbol{K}_2^{(\mathrm{nc})} \cdot \boldsymbol{E}_\mathrm{s}^{(\mathrm{nc})}, \tag{B.418}$$

where $\boldsymbol{\Upsilon} = \boldsymbol{V} \cdot \boldsymbol{\Omega} \cdot \boldsymbol{V}^{-1}$ and $\boldsymbol{\Omega} = \operatorname{diag}\left\{\begin{bmatrix}\omega_1, \ldots, \omega_d\end{bmatrix}\right\}$ with $\omega_i = \tan(\mu_i/2)$, $i = 1, 2, \ldots, d$. Following the reasoning in Appendix B.13 and using the result

$$\Delta\mu \approx \Delta\omega \cdot \frac{2}{\omega_i^2 + 1} \tag{B.419}$$

derived in (B.263), the real-valued expression for the parameter estimation error of 1-D NC Unitary ESPRIT is given by

$$\Delta\mu_i = \bar{\boldsymbol{p}}_i^\mathrm{T} \cdot \left(\boldsymbol{K}_1^{(\mathrm{nc})} \cdot \boldsymbol{E}_\mathrm{s}^{(\mathrm{nc})}\right)^+ \cdot \left(\boldsymbol{K}_2^{(\mathrm{nc})} - \omega_i \boldsymbol{K}_1^{(\mathrm{nc})}\right) \cdot \Delta\boldsymbol{E}_\mathrm{s}^{(\mathrm{nc})} \cdot \bar{\boldsymbol{q}}_i \cdot \frac{2}{\omega_i^2 + 1}, \tag{B.420}$$

where $\bar{\boldsymbol{q}}_i$ is the i-th column of $\boldsymbol{V}$ and $\bar{\boldsymbol{p}}_i^\mathrm{T}$ is the i-th row of $\boldsymbol{V}^{-1}$. Moreover, the perturbation of the real-valued subspace $\boldsymbol{E}_\mathrm{s}^{(\mathrm{nc})}$ is expanded in terms of the transformed noise contribution $\varphi\left(\boldsymbol{N}^{(\mathrm{nc})^{(\mathrm{fba})}}\right) = \boldsymbol{Q}_{2M}^\mathrm{H} \cdot \boldsymbol{N}^{(\mathrm{nc})^{(\mathrm{fba})}} \cdot \boldsymbol{Q}_{2N}$ as

$$\Delta\boldsymbol{E}_\mathrm{s}^{(\mathrm{nc})} = \boldsymbol{E}_\mathrm{n}^{(\mathrm{nc})} \cdot \boldsymbol{E}_\mathrm{n}^{(\mathrm{nc})^\mathrm{H}} \cdot \varphi\left(\boldsymbol{N}^{(\mathrm{nc})^{(\mathrm{fba})}}\right) \cdot \boldsymbol{W}_\mathrm{s}^{(\mathrm{nc})} \cdot \boldsymbol{\Sigma}_\mathrm{s}^{(\varphi)^{-1}}, \tag{B.421}$$

where the required subspaces are obtained from the SVD of the transformed real-valued measure-

ment matrix $\varphi\left(\boldsymbol{X}_0^{(\mathrm{nc})^{(\mathrm{fba})}}\right) = \boldsymbol{Q}_{2M}^{\mathrm{H}} \cdot \boldsymbol{X}_0^{(\mathrm{nc})^{(\mathrm{fba})}} \cdot \boldsymbol{Q}_{2N} \in \mathbb{R}^{2M \times 2N}$ expressed as

$$\varphi(\boldsymbol{X}_0^{(\mathrm{nc})^{(\mathrm{fba})}}) = \begin{bmatrix} \boldsymbol{E}_{\mathrm{s}}^{(\mathrm{nc})} & \boldsymbol{E}_{\mathrm{n}}^{(\mathrm{nc})} \end{bmatrix} \cdot \begin{bmatrix} \boldsymbol{\Sigma}_{\mathrm{s}}^{(\varphi)} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} \end{bmatrix} \cdot \begin{bmatrix} \boldsymbol{W}_{\mathrm{s}}^{(\mathrm{nc})} & \boldsymbol{W}_{\mathrm{n}}^{(\mathrm{nc})} \end{bmatrix}^{\mathrm{H}}.$$

Next, we use the property that the matrices $\boldsymbol{Q}_p$ are unitary and choose the subspaces of $\varphi\left(\boldsymbol{X}_0^{(\mathrm{nc})^{(\mathrm{fba})}}\right)$ as

$$\begin{aligned} \boldsymbol{E}_{\mathrm{s}}^{(\mathrm{nc})} &= \boldsymbol{Q}_{2M}^{\mathrm{H}} \cdot \boldsymbol{U}_{\mathrm{s}}^{(\mathrm{nc})^{(\mathrm{fba})}}, \ \boldsymbol{E}_{\mathrm{n}}^{(\mathrm{nc})} = \boldsymbol{Q}_{2M}^{\mathrm{H}} \cdot \boldsymbol{U}_{\mathrm{n}}^{(\mathrm{nc})^{(\mathrm{fba})}}, \ \boldsymbol{\Sigma}_{\mathrm{s}}^{(\varphi)} = \boldsymbol{\Sigma}_{\mathrm{s}}^{(\mathrm{nc})^{(\mathrm{fba})}} \\ \boldsymbol{W}_{\mathrm{s}}^{(\mathrm{nc})} &= \boldsymbol{Q}_{2N}^{\mathrm{H}} \cdot \boldsymbol{V}_{\mathrm{s}}^{(\mathrm{nc})^{(\mathrm{fba})}}, \ \boldsymbol{W}_{\mathrm{n}}^{(\mathrm{nc})} = \boldsymbol{Q}_{2N}^{\mathrm{H}} \cdot \boldsymbol{V}_{\mathrm{n}}^{(\mathrm{nc})^{(\mathrm{fba})}}. \end{aligned} \tag{B.422}$$

Additionally, we express the transformed selection matrices $\boldsymbol{K}_1^{(\mathrm{nc})}$ and $\boldsymbol{K}_2^{(\mathrm{nc})}$ defined in (6.20) and (6.21) as

$$\boldsymbol{K}_1^{(\mathrm{nc})} = \boldsymbol{Q}_{2M^{(\mathrm{sel})}}^{\mathrm{H}} \cdot \left(\boldsymbol{J}_1^{(\mathrm{nc})} + \boldsymbol{J}_2^{(\mathrm{nc})}\right) \cdot \boldsymbol{Q}_{2M} \tag{B.423}$$

$$\boldsymbol{K}_2^{(\mathrm{nc})} = \mathrm{j} \cdot \boldsymbol{Q}_{2M^{(\mathrm{sel})}}^{\mathrm{H}} \cdot \left(\boldsymbol{J}_1^{(\mathrm{nc})} - \boldsymbol{J}_2^{(\mathrm{nc})}\right) \cdot \boldsymbol{Q}_{2M}, \tag{B.424}$$

which can be established similarly to (B.270) and (B.271) by using the fact that the virtual array is always centro-symmetric as shown in Theorem 6.2.2.

Inserting (B.421) into (B.420) and applying the identities (B.422)-(B.424), we have

$$\begin{aligned} \Delta\mu_i \approx \bar{\boldsymbol{p}}_i^{\mathrm{T}} \cdot \left(\left(\boldsymbol{J}_1^{(\mathrm{nc})} + \boldsymbol{J}_2^{(\mathrm{nc})}\right) \cdot \boldsymbol{U}_{\mathrm{s}}^{(\mathrm{nc})^{(\mathrm{fba})}}\right)^{+} \cdot \left(\mathrm{j} \cdot \left(\boldsymbol{J}_1^{(\mathrm{nc})} - \boldsymbol{J}_2^{(\mathrm{nc})}\right) - \omega_i \cdot \left(\boldsymbol{J}_1^{(\mathrm{nc})} + \boldsymbol{J}_2^{(\mathrm{nc})}\right)\right) \\ \cdot \Delta\boldsymbol{U}_{\mathrm{s}}^{(\mathrm{nc})^{(\mathrm{fba})}} \cdot \bar{\boldsymbol{q}}_i \cdot \frac{2}{\omega_i^2 + 1}, \end{aligned} \tag{B.425}$$

where $\Delta\boldsymbol{U}_{\mathrm{s}}^{(\mathrm{nc})^{(\mathrm{fba})}} = \boldsymbol{U}_{\mathrm{n}}^{(\mathrm{nc})^{(\mathrm{fba})}} \cdot \boldsymbol{U}_{\mathrm{n}}^{(\mathrm{nc})^{(\mathrm{fba})^{\mathrm{H}}}} \cdot \boldsymbol{N}^{(\mathrm{nc})^{(\mathrm{fba})}} \cdot \boldsymbol{V}_s^{(\mathrm{nc})^{(\mathrm{fba})}} \cdot \boldsymbol{\Sigma}_{\mathrm{s}}^{(\mathrm{nc})^{(\mathrm{fba})^{-1}}}$.

In order to further simplify (B.425), we apply the following two lemmas, which represent a direct extension of Lemma B.13.1 and Lemma B.13.2 and can be proven in the same way.

Lemma B.25.1. *The following identities are satisfied*

$$\left(\boldsymbol{J}_1^{(\mathrm{nc})} + \boldsymbol{J}_2^{(\mathrm{nc})}\right) \cdot \boldsymbol{U}_{\mathrm{s}}^{(\mathrm{nc})^{(\mathrm{fba})}} = \boldsymbol{J}_1^{(\mathrm{nc})} \cdot \boldsymbol{U}_{\mathrm{s}}^{(\mathrm{nc})^{(\mathrm{fba})}} \cdot \breve{\boldsymbol{\Psi}} \tag{B.426}$$

$$\left(\boldsymbol{J}_1^{(\mathrm{nc})} - \boldsymbol{J}_2^{(\mathrm{nc})}\right) \cdot \boldsymbol{U}_{\mathrm{s}}^{(\mathrm{nc})^{(\mathrm{fba})}} = \boldsymbol{J}_2^{(\mathrm{nc})} \cdot \boldsymbol{U}_{\mathrm{s}}^{(\mathrm{nc})^{(\mathrm{fba})}} \cdot \mathring{\boldsymbol{\Psi}}, \tag{B.427}$$

where $\breve{\boldsymbol{\Psi}} = \boldsymbol{I}_d + \boldsymbol{\Psi} = \boldsymbol{Q}^{(\mathrm{fba})} \cdot (\boldsymbol{I}_d + \boldsymbol{\Lambda}) \cdot \boldsymbol{Q}^{(\mathrm{fba})^{-1}}$ *and* $\mathring{\boldsymbol{\Psi}} = -\boldsymbol{I}_d + \boldsymbol{\Psi}^{-1} = \boldsymbol{Q}^{(\mathrm{fba})} \cdot \left(-\boldsymbol{I}_d + \boldsymbol{\Lambda}^{-1}\right) \cdot \boldsymbol{Q}^{(\mathrm{fba})^{-1}}$.

Lemma B.25.2. *In the noiseless case, the solution* $\boldsymbol{\Psi}$ *to* (B.416) *and the solution* $\boldsymbol{\Upsilon}$ *to* (B.418) *have the same eigenvectors, i.e.,* $\boldsymbol{Q}^{(\mathrm{fba})} = \boldsymbol{V}$. *Moreover, their eigenvalues are related as* $\omega_i = \mathrm{j} \cdot \frac{1-\lambda_i}{1+\lambda_i}$.

Next, we consider the term $\left(\mathrm{j}\cdot\left(\boldsymbol{J}_1^{(\mathrm{nc})}-\boldsymbol{J}_2^{(\mathrm{nc})}\right)-\omega_i\cdot\left(\boldsymbol{J}_1^{(\mathrm{nc})}+\boldsymbol{J}_2^{(\mathrm{nc})}\right)\right)$ in (B.425) and apply the relation $\omega_i=\mathrm{j}\cdot\frac{1-\lambda_i}{1+\lambda_i}$ from Lemma B.25.2. We can then rewrite this term as $\mathrm{j}\cdot\left(\boldsymbol{J}_1^{(\mathrm{nc})}\cdot\lambda_i-\boldsymbol{J}_2^{(\mathrm{nc})}\right)\cdot\frac{2}{1+\lambda_i}$. Moreover, the term $\frac{2}{\omega_i^2+1}$ in (B.425) can be expressed in terms of λ_i as $\frac{2}{\omega_i^2+1}=\frac{(\lambda_i+1)^2}{2\lambda_i}$. Inserting these relations into (B.425), replacing $\left(\boldsymbol{J}_1^{(\mathrm{nc})}+\boldsymbol{J}_2^{(\mathrm{nc})}\right)\cdot\boldsymbol{U}_{\mathrm{s}}^{(\mathrm{nc})^{(\mathrm{fba})}}$ via (B.426), and substituting $\bar{\boldsymbol{p}}_i=\boldsymbol{p}_i^{(\mathrm{fba})}$ and $\bar{\boldsymbol{q}}_i=\boldsymbol{q}_i^{(\mathrm{fba})}$ using Lemma B.25.2, yields

$$\begin{aligned}\Delta\mu_i&=\mathrm{j}\cdot\boldsymbol{p}_i^{(\mathrm{fba})^{\mathrm{T}}}\cdot\breve{\boldsymbol{\Psi}}^{-1}\cdot\left(\boldsymbol{J}_1^{(\mathrm{nc})}\cdot\boldsymbol{U}_{\mathrm{s}}^{(\mathrm{nc})^{(\mathrm{fba})}}\right)^{+}\cdot\left(\boldsymbol{J}_1^{(\mathrm{nc})}\cdot\lambda_i-\boldsymbol{J}_2^{(\mathrm{nc})}\right)\\&\quad\cdot\Delta\boldsymbol{U}_{\mathrm{s}}^{(\mathrm{nc})^{(\mathrm{fba})}}\cdot\boldsymbol{q}_i^{(\mathrm{fba})}\cdot\frac{2}{1+\lambda_i}\cdot\frac{(\lambda_i+1)^2}{2\lambda_i}\\&=-\mathrm{j}\cdot\boldsymbol{p}_i^{(\mathrm{fba})^{\mathrm{T}}}\cdot\left(\boldsymbol{J}_1^{(\mathrm{nc})}\cdot\boldsymbol{U}_{\mathrm{s}}^{(\mathrm{nc})^{(\mathrm{fba})}}\right)^{+}\cdot\left(\boldsymbol{J}_2^{(\mathrm{nc})}/\lambda_i-\boldsymbol{J}_1^{(\mathrm{nc})}\right)\cdot\Delta\boldsymbol{U}_{\mathrm{s}}^{(\mathrm{nc})^{(\mathrm{fba})}}\cdot\boldsymbol{q}_i^{(\mathrm{fba})},\end{aligned}\tag{B.428}$$

where we used $\boldsymbol{p}_i^{(\mathrm{fba})^{\mathrm{T}}}\cdot\breve{\boldsymbol{\Psi}}^{-1}=\boldsymbol{p}_i^{(\mathrm{fba})^{\mathrm{T}}}\cdot(1+\lambda_i)^{-1}$ from Lemma B.25.1 in the first equation.

As a final step, we follow the reasoning in Appendix B.13 to find that $-\mathrm{j}z\in\mathbb{R}$ for $z\in\mathbb{C}$ implies that $\mathrm{Re}\{z\}=0$ and hence $-\mathrm{j}z=\mathrm{Im}\{z\}$. Consequently, (B.428) can also be written as (B.417) and is therefore equivalent to the first-order expansion for R-D NC Standard ESPRIT with FBA. This concludes the proof of the theorem. □

B.26. Proof of Theorem 6.5.1

For the proof of Theorem (6.5.1) in Section 6.5.1, we follow the steps in [RHD14] that are used to derive the result for the non-NC case in (3.15). Starting from (6.33), the truncated core tensor $\hat{\boldsymbol{\mathcal{S}}}^{[\mathrm{s}](r)}$ can be computed via

$$\hat{\boldsymbol{\mathcal{S}}}^{[\mathrm{s}](r)}=\boldsymbol{\mathcal{X}}^{(\mathrm{nc},r)}\times_1\hat{\boldsymbol{U}}_1^{[\mathrm{s}](r)^{\mathrm{H}}}\ldots\times_R\hat{\boldsymbol{U}}_R^{[\mathrm{s}](r)^{\mathrm{H}}}\times_{R+1}\hat{\boldsymbol{U}}_{R+1}^{[\mathrm{s}](r)^{\mathrm{H}}}.\tag{B.429}$$

Inserting (B.429) into (6.33), we obtain

$$\begin{aligned}\hat{\boldsymbol{\mathcal{U}}}^{[\mathrm{s}](r)}&=\boldsymbol{\mathcal{X}}^{(\mathrm{nc},r)}\times_1\left(\hat{\boldsymbol{U}}_1^{[\mathrm{s}](r)}\cdot\hat{\boldsymbol{U}}_1^{[\mathrm{s}](r)^{\mathrm{H}}}\right)\ldots\times_R\left(\hat{\boldsymbol{U}}_R^{[\mathrm{s}](r)}\cdot\hat{\boldsymbol{U}}_R^{[\mathrm{s}](r)^{\mathrm{H}}}\right)\times_{R+1}\left(\hat{\boldsymbol{\Sigma}}_{R+1}^{[\mathrm{s}](r)^{-1}}\cdot\hat{\boldsymbol{U}}_{R+1}^{[\mathrm{s}](r)^{\mathrm{H}}}\right)\\&=\boldsymbol{\mathcal{X}}^{(\mathrm{nc},r)}\times_1\hat{\boldsymbol{T}}_1^{(r)}\ldots\times_r\hat{\boldsymbol{T}}_r^{(\mathrm{nc},r)}\ldots\times_R\hat{\boldsymbol{T}}_R^{(r)}\times_{R+1}\left(\hat{\boldsymbol{\Sigma}}_{R+1}^{[\mathrm{s}](r)^{-1}}\cdot\hat{\boldsymbol{U}}_{R+1}^{[\mathrm{s}](r)^{\mathrm{H}}}\right)\end{aligned}$$

with the projection matrices $\hat{\boldsymbol{T}}_q^{(r)}=\hat{\boldsymbol{U}}_q^{[\mathrm{s}](r)}\cdot\hat{\boldsymbol{U}}_q^{[\mathrm{s}](r)^{\mathrm{H}}}\in\mathbb{C}^{M_q\times M_q}$, $q=1,\ldots,r-1,r+1,\ldots,R$ and $\hat{\boldsymbol{T}}_r^{(\mathrm{nc},r)}=\hat{\boldsymbol{U}}_r^{[\mathrm{s}](r)}\cdot\hat{\boldsymbol{U}}_r^{[\mathrm{s}](r)^{\mathrm{H}}}\in\mathbb{C}^{2M_r\times 2M_r}$. Then, applying the rule in (1.30), we can write

$$\left[\hat{\boldsymbol{\mathcal{U}}}^{[\mathrm{s}](r)}\right]_{(R+1)}^{\mathrm{T}}=\left[\left(\hat{\boldsymbol{\Sigma}}_{R+1}^{[\mathrm{s}](r)^{-1}}\cdot\hat{\boldsymbol{U}}_{R+1}^{[\mathrm{s}](r)^{\mathrm{H}}}\right)\cdot\left[\boldsymbol{\mathcal{X}}^{(\mathrm{nc},r)}\right]_{(R+1)}\cdot\left(\hat{\boldsymbol{T}}_{1:r-1}^{(r)\otimes}\otimes\hat{\boldsymbol{T}}_r^{(\mathrm{nc},r)}\otimes\hat{\boldsymbol{T}}_{r+1:R}^{(r)\otimes}\right)^{\mathrm{T}}\right]^{\mathrm{T}}$$

$$= \left(\hat{\boldsymbol{T}}_{1:r-1}^{(r)\otimes} \otimes \hat{\boldsymbol{T}}_{r}^{(\mathrm{nc},r)} \otimes \hat{\boldsymbol{T}}_{r+1:R}^{(r)\otimes} \right) \cdot \left[\boldsymbol{\mathcal{X}}^{(\mathrm{nc},r)} \right]_{(R+1)}^{\mathrm{T}} \cdot \hat{\boldsymbol{U}}_{R+1}^{[\mathrm{s}](r)^*} \cdot \hat{\boldsymbol{\Sigma}}_{R+1}^{[\mathrm{s}](r)^{-1}},$$

where $\hat{\boldsymbol{T}}_{a:b}^{(r)\otimes}$ is defined in (6.60). In Section 6.3.2, we have seen that due to the fact that $\boldsymbol{X}^{(\mathrm{nc},r)} = \left[\boldsymbol{\mathcal{X}}^{(\mathrm{nc},r)} \right]_{(R+1)}^{\mathrm{T}}$, the subspaces of $\boldsymbol{X}^{(\mathrm{nc},r)}$ and $\left[\boldsymbol{\mathcal{X}}^{(\mathrm{nc},r)} \right]_{(R+1)}^{\mathrm{T}}$ obtained from their SVDs are linked according to (6.35). Hence, we obtain

$$\left[\boldsymbol{\mathcal{X}}^{(\mathrm{nc},r)} \right]_{(R+1)}^{\mathrm{T}} \cdot \hat{\boldsymbol{U}}_{R+1}^{[\mathrm{s}](r)^*} \cdot \hat{\boldsymbol{\Sigma}}_{R+1}^{[\mathrm{s}](r)^{-1}} = \boldsymbol{X}^{(\mathrm{nc},r)} \cdot \hat{\boldsymbol{V}}_{\mathrm{s}}^{(\mathrm{nc},r)} \cdot \hat{\boldsymbol{\Sigma}}_{\mathrm{s}}^{(\mathrm{nc},r)^{-1}} = \hat{\boldsymbol{U}}_{\mathrm{s}}^{(\mathrm{nc},r)}, \tag{B.430}$$

where $\hat{\boldsymbol{U}}_{\mathrm{s}}^{(\mathrm{nc},r)} \in \mathbb{C}^{2M_r \times d}$, $\hat{\boldsymbol{V}}_{\mathrm{s}}^{(\mathrm{nc},r)} \in \mathbb{C}^{N \times d}$, and $\hat{\boldsymbol{\Sigma}}_{\mathrm{s}}^{(\mathrm{nc},r)} \in \mathbb{R}^{d \times d}$ denote the dominant left singular vectors, the dominant right singular vectors, and the diagonal matrix containing the dominant eigenvalues of $\boldsymbol{X}^{(\mathrm{nc},r)}$. Finally, we obtain the result

$$\left[\hat{\boldsymbol{\mathcal{U}}}^{[\mathrm{s}](r)} \right]_{(R+1)}^{\mathrm{T}} = \left(\hat{\boldsymbol{T}}_{1:r-1}^{(r)\otimes} \otimes \hat{\boldsymbol{T}}_{r}^{(\mathrm{nc},r)} \otimes \hat{\boldsymbol{T}}_{r+1:R}^{(r)\otimes} \right) \cdot \hat{\boldsymbol{U}}_{\mathrm{s}}^{(\mathrm{nc},r)} \tag{B.431}$$

for $r = 1, \ldots, R$. □

B.27. Proof of Theorem 6.5.2

We start the proof of Theorem 6.5.2 in Section 6.5.1 by inserting $\hat{\boldsymbol{U}}_{\mathrm{s}}^{(\mathrm{nc},r)} = \boldsymbol{U}_{\mathrm{s}}^{(\mathrm{nc},r)} + \Delta\boldsymbol{U}_{\mathrm{s}}^{(\mathrm{nc},r)}$ and $\hat{\boldsymbol{T}}_{q}^{(r)} = \boldsymbol{T}_{q}^{(r)} + \Delta\boldsymbol{T}_{q}^{(r)}$ as well as $\hat{\boldsymbol{T}}_{r}^{(\mathrm{nc},r)} = \boldsymbol{T}_{r}^{(\mathrm{nc},r)} + \Delta\boldsymbol{T}_{r}^{(\mathrm{nc},r)}$ into (6.59). Then, we have

$$\begin{aligned}
\left[\hat{\boldsymbol{\mathcal{U}}}^{[\mathrm{s}](r)} \right]_{(R+1)}^{\mathrm{T}} &= \left[\left(\boldsymbol{T}_1^{(r)} + \Delta\boldsymbol{T}_1^{(r)} \right) \otimes \ldots \left(\boldsymbol{T}_r^{(\mathrm{nc},r)} + \Delta\boldsymbol{T}_r^{(\mathrm{nc},r)} \right) \otimes \ldots \left(\boldsymbol{T}_R^{(r)} + \Delta\boldsymbol{T}_R^{(r)} \right) \right] \\
&\quad \cdot \left(\boldsymbol{U}_{\mathrm{s}}^{(\mathrm{nc},r)} + \Delta\boldsymbol{U}_{\mathrm{s}}^{(\mathrm{nc},r)} \right) \\
&\approx \underbrace{\left[\boldsymbol{T}_1^{(r)} \otimes \ldots \boldsymbol{T}_r^{(\mathrm{nc},r)} \otimes \ldots \boldsymbol{T}_R^{(r)} \right] \cdot \boldsymbol{U}_{\mathrm{s}}^{(\mathrm{nc},r)}}_{\boldsymbol{U}_{\mathrm{s}}^{(\mathrm{nc},r)}} + \left[\boldsymbol{T}_1^{(r)} \otimes \ldots \boldsymbol{T}_r^{(\mathrm{nc},r)} \otimes \ldots \boldsymbol{T}_R^{(r)} \right] \cdot \Delta\boldsymbol{U}_{\mathrm{s}}^{(\mathrm{nc},r)} \\
&\quad + \left[\Delta\boldsymbol{T}_1^{(r)} \otimes \ldots \boldsymbol{T}_r^{(\mathrm{nc},r)} \otimes \ldots \boldsymbol{T}_R^{(r)} \right] \cdot \boldsymbol{U}_{\mathrm{s}}^{(\mathrm{nc},r)} + \left[\boldsymbol{T}_1^{(r)} \otimes \Delta\boldsymbol{T}_2^{(r)} \otimes \ldots \boldsymbol{T}_r^{(\mathrm{nc},r)} \otimes \ldots \boldsymbol{T}_R^{(r)} \right] \cdot \boldsymbol{U}_{\mathrm{s}}^{(\mathrm{nc},r)} \ldots \\
&\quad + \left[\boldsymbol{T}_1^{(r)} \otimes \ldots \boldsymbol{T}_r^{(\mathrm{nc},r)} \otimes \ldots \Delta\boldsymbol{T}_R^{(r)} \right] \cdot \boldsymbol{U}_{\mathrm{s}}^{(\mathrm{nc},r)} \\
&= \boldsymbol{U}_{\mathrm{s}}^{(\mathrm{nc},r)} + \left[\Delta\hat{\boldsymbol{\mathcal{U}}}^{[\mathrm{s}](r)} \right]_{(R+1)}^{\mathrm{T}},
\end{aligned} \tag{B.432}$$

where the second-order perturbation terms have been neglected. The first term of (B.432) represents the true augmented signal subspace and the remaining terms form the first-order approximation of the HOSVD-based signal subspace estimation error $\left[\Delta\hat{\boldsymbol{\mathcal{U}}}^{[\mathrm{s}](r)} \right]_{(R+1)}^{\mathrm{T}}$. In order to show

the expression for $\left[\Delta\hat{\boldsymbol{\mathcal{U}}}^{[\mathrm{s}](r)}\right]^{\mathrm{T}}_{(R+1)}$ in Theorem 6.5.2, we expand

$$\begin{aligned}\left[\boldsymbol{T}_1^{(r)} \otimes \ldots \otimes \Delta\boldsymbol{T}_p^{(r)} \otimes \ldots \boldsymbol{T}_R^{(r)}\right] \cdot \boldsymbol{U}_{\mathrm{s}}^{(\mathrm{nc},r)} &= \left[\boldsymbol{T}_1^{(r)} \otimes \ldots \otimes \Delta\boldsymbol{T}_p^{(r)} \otimes \ldots \boldsymbol{T}_R^{(r)}\right] \\ &\quad \cdot \left[\boldsymbol{T}_1^{(r)} \otimes \ldots \otimes \boldsymbol{T}_p^{(r)} \otimes \ldots \boldsymbol{T}_R^{(r)}\right] \cdot \boldsymbol{U}_{\mathrm{s}}^{(\mathrm{nc},r)} \\ &= \left[\boldsymbol{T}_1^{(r)} \otimes \ldots \otimes (\Delta\boldsymbol{T}_p^{(r)} \cdot \boldsymbol{T}_p^{(r)}) \otimes \ldots \boldsymbol{T}_R^{(r)}\right] \cdot \boldsymbol{U}_{\mathrm{s}}^{(\mathrm{nc},r)}, \quad \text{(B.433)}\end{aligned}$$

where we have used the fact that the projection matrices $\boldsymbol{T}_p^{(r)}$ for $p = 1, \ldots, R$ are idempotent, i.e., $\boldsymbol{T}_p^{(r)} \cdot \boldsymbol{T}_p^{(r)} = \boldsymbol{T}_p^{(r)}$. Note that $\boldsymbol{T}_p^{(r)}$ also includes $\boldsymbol{T}_r^{(\mathrm{nc},r)}$ for $p = r$. Hence, we still need to show that $\Delta\boldsymbol{T}_p^{(r)} \cdot \boldsymbol{T}_p^{(r)} = \Delta\boldsymbol{U}_p^{[\mathrm{s}](r)} \cdot \boldsymbol{U}_p^{[\mathrm{s}](r)^{\mathrm{H}}}$. As $\hat{\boldsymbol{T}}_p^{(r)} = \boldsymbol{U}_p^{[\mathrm{s}](r)} \cdot \boldsymbol{U}_p^{[\mathrm{s}](r)^{\mathrm{H}}}$ and $\hat{\boldsymbol{U}}_p^{[\mathrm{s}](r)} = \boldsymbol{U}_p^{[\mathrm{s}](r)} + \Delta\boldsymbol{U}_p^{[\mathrm{s}](r)}$, we can compute $\Delta\boldsymbol{T}_p^{(r)}$ as

$$\begin{aligned}\hat{\boldsymbol{T}}_p^{(r)} &= \left(\boldsymbol{U}_p^{[\mathrm{s}](r)} + \Delta\boldsymbol{U}_p^{[\mathrm{s}](r)}\right) \cdot \left(\boldsymbol{U}_p^{[\mathrm{s}](r)^{\mathrm{H}}} + \Delta\boldsymbol{U}_p^{[\mathrm{s}](r)^{\mathrm{H}}}\right) \\ &\approx \boldsymbol{T}_p^{(r)} + \boldsymbol{U}_p^{[\mathrm{s}](r)} \cdot \Delta\boldsymbol{U}_p^{[\mathrm{s}](r)^{\mathrm{H}}} + \Delta\boldsymbol{U}_p^{[\mathrm{s}](r)} \cdot \boldsymbol{U}_p^{[\mathrm{s}](r)^{\mathrm{H}}} \\ \Rightarrow \Delta\boldsymbol{T}_p^{(r)} &= \boldsymbol{U}_p^{[\mathrm{s}](r)} \cdot \Delta\boldsymbol{U}_p^{[\mathrm{s}](r)^{\mathrm{H}}} + \Delta\boldsymbol{U}_p^{[\mathrm{s}](r)} \cdot \boldsymbol{U}_p^{[\mathrm{s}](r)^{\mathrm{H}}}, \quad \text{(B.434)}\end{aligned}$$

where $\Delta\boldsymbol{U}_p^{[\mathrm{s}](r)} \approx \boldsymbol{U}_p^{[\mathrm{n}](r)} \cdot \boldsymbol{\Gamma}_p^{[\mathrm{n}]}$ with $\boldsymbol{\Gamma}_p^{[\mathrm{n}]} = \boldsymbol{U}_p^{[\mathrm{n}]^{\mathrm{H}}} \cdot [\boldsymbol{\mathcal{N}}]_{(p)} \cdot \boldsymbol{V}_p^{[\mathrm{s}]} \cdot \boldsymbol{\Sigma}_p^{[\mathrm{s}]^{-1}}$ (cf. (4.8)). Using this relation in (B.434), yields

$$\begin{aligned}\Delta\boldsymbol{T}_p^{(r)} \cdot \boldsymbol{T}_p^{(r)} &\approx \boldsymbol{U}_p^{[\mathrm{s}](r)} \cdot \Delta\boldsymbol{U}_p^{[\mathrm{s}](r)^{\mathrm{H}}} \cdot \boldsymbol{T}_p^{(r)} + \Delta\boldsymbol{U}_p^{[\mathrm{s}](r)} \cdot \boldsymbol{U}_p^{[\mathrm{s}](r)^{\mathrm{H}}} \cdot \boldsymbol{T}_p^{(r)} \\ &= \boldsymbol{U}_p^{[\mathrm{s}](r)} \cdot \boldsymbol{\Gamma}_p^{[\mathrm{n}]^{\mathrm{H}}} \cdot \boldsymbol{U}_p^{[\mathrm{n}](r)^{\mathrm{H}}} \cdot \boldsymbol{T}_p^{(r)} + \Delta\boldsymbol{U}_p^{[\mathrm{s}](r)} \cdot \boldsymbol{U}_p^{[\mathrm{s}](r)^{\mathrm{H}}} \cdot \boldsymbol{T}_p^{(r)} \quad \text{(B.435)} \\ &= \Delta\boldsymbol{U}_p^{[\mathrm{s}](r)} \cdot \boldsymbol{U}_p^{[\mathrm{s}](r)^{\mathrm{H}}}, \quad \text{(B.436)}\end{aligned}$$

where in (B.435), we have used the fact that $\boldsymbol{U}_p^{[\mathrm{n}](r)^{\mathrm{H}}} \cdot \boldsymbol{T}_p^{(r)} = \boldsymbol{0}$ and that $\boldsymbol{U}_p^{[\mathrm{s}](r)^{\mathrm{H}}} \cdot \boldsymbol{T}_p^{(r)} = \boldsymbol{U}_p^{[\mathrm{s}](r)^{\mathrm{H}}}$. Introducing the notation $\boldsymbol{T}_{a:b}^{(r)\otimes}$ and $\boldsymbol{T}_{a:b}^{(\mathrm{nc},r)\otimes}$ from (6.60), we obtain the desired result. □

B.28. Proof of Equation (6.68)

In order to prove Equation (6.68) in Section 6.5.1, we first express the estimation error in $\mu_i^{(r)}$ from (6.65) in terms of the perturbation $\boldsymbol{n}^{(\mathrm{nc},r)} = \mathrm{vec}\left\{\boldsymbol{N}^{(\mathrm{nc},r)}\right\} = \mathrm{vec}\left\{\left[\boldsymbol{\mathcal{N}}^{(\mathrm{nc},r)}\right]^{\mathrm{T}}_{(R+1)}\right\}$. Vectorizing (6.65), we obtain

$$\begin{aligned}\Delta\mu_i^{(r)} &\approx \mathrm{Im}\left\{\boldsymbol{p}_i^{\mathrm{T}} \cdot \left(\tilde{\boldsymbol{J}}_1^{(\mathrm{nc})(r)} \cdot \boldsymbol{U}_{\mathrm{s}}^{(\mathrm{nc},r)}\right)^{+} \cdot \left[\tilde{\boldsymbol{J}}_2^{(\mathrm{nc})(r)}/\lambda_i^{(r)} - \tilde{\boldsymbol{J}}_1^{(\mathrm{nc})(r)}\right] \cdot \left[\Delta\boldsymbol{\mathcal{U}}^{[\mathrm{s}](r)}\right]^{\mathrm{T}}_{(R+1)} \cdot \boldsymbol{q}_i\right\} \\ &= \mathrm{Im}\left\{\left[\boldsymbol{q}_i^{\mathrm{T}} \otimes \left(\boldsymbol{p}_i^{\mathrm{T}} \cdot \left(\tilde{\boldsymbol{J}}_1^{(\mathrm{nc})(r)} \cdot \boldsymbol{U}_{\mathrm{s}}^{(\mathrm{nc},r)}\right)^{+} \cdot \left[\tilde{\boldsymbol{J}}_2^{(\mathrm{nc})(r)}/\lambda_i^{(r)} - \tilde{\boldsymbol{J}}_1^{(\mathrm{nc})(r)}\right]\right)\right] \cdot \mathrm{vec}\left\{\left[\Delta\boldsymbol{\mathcal{U}}^{[\mathrm{s}](r)}\right]^{\mathrm{T}}_{(R+1)}\right\}\right\}\end{aligned}$$

$$= \operatorname{Im}\left\{ \boldsymbol{r}_i^{(\mathrm{nc},r)^{\mathrm{T}}} \cdot \operatorname{vec}\left\{ \left[\Delta\boldsymbol{\mathcal{U}}^{[\mathrm{s}](r)} \right]_{(R+1)}^{\mathrm{T}} \right\} \right\}, \tag{B.437}$$

where $\boldsymbol{r}_i^{(\mathrm{nc},r)}$ is defined in (6.67). Due to the linear dependence of $\operatorname{vec}\left\{ \left[\Delta\boldsymbol{\mathcal{U}}^{[\mathrm{s}](r)} \right]_{(R+1)}^{\mathrm{T}} \right\}$ on $\boldsymbol{n}^{(\mathrm{nc},r)}$, we write

$$\operatorname{vec}\left\{ \left[\Delta\boldsymbol{\mathcal{U}}^{[\mathrm{s}](r)} \right]_{(R+1)}^{\mathrm{T}} \right\} = \boldsymbol{W}_{\mathrm{ten}}^{(\mathrm{nc},r)} \cdot \boldsymbol{n}^{(\mathrm{nc},r)}. \tag{B.438}$$

In order to find an explicit expression of $\boldsymbol{W}_{\mathrm{ten}}^{(\mathrm{nc},r)}$, we need to compute the left-hand side of (B.438). The expression for $\left[\Delta\boldsymbol{\mathcal{U}}^{[\mathrm{s}](r)} \right]_{(R+1)}^{\mathrm{T}}$ is given in (6.62) and restated here as

$$\begin{aligned}
\left[\Delta\boldsymbol{\mathcal{U}}^{[\mathrm{s}](r)} \right]_{(R+1)}^{\mathrm{T}} &= \left(\boldsymbol{T}_{1:r-1}^{(r)\otimes} \otimes \boldsymbol{T}_r^{(\mathrm{nc},r)} \otimes \boldsymbol{T}_{r+1:R}^{(r)\otimes} \right) \cdot \Delta\boldsymbol{U}_{\mathrm{s}}^{(\mathrm{nc},r)} \\
&+ \left[\sum_{q=1}^{r-1} \left(\boldsymbol{T}_{1:q-1}^{(r)\otimes} \otimes \left[\Delta\boldsymbol{U}_q^{[\mathrm{s}](r)} \cdot \boldsymbol{U}_q^{[\mathrm{s}](r)^{\mathrm{H}}} \right] \otimes \boldsymbol{T}_{q+1:R}^{(\mathrm{nc},r)\otimes} \right) + \boldsymbol{T}_{1:r-1}^{(r)\otimes} \otimes \left[\Delta\boldsymbol{U}_r^{[\mathrm{s}](r)} \cdot \boldsymbol{U}_r^{[\mathrm{s}](r)^{\mathrm{H}}} \right] \otimes \boldsymbol{T}_{r+1:R}^{(r)\otimes} \right. \\
&\left. + \sum_{q=r+1}^{R} \boldsymbol{T}_{1:q-1}^{(\mathrm{nc},r)\otimes} \otimes \left[\Delta\boldsymbol{U}_q^{[\mathrm{s}](r)} \cdot \boldsymbol{U}_q^{[\mathrm{s}](r)^{\mathrm{H}}} \right] \otimes \boldsymbol{T}_{q+1:R}^{(r)\otimes} \right] \cdot \boldsymbol{U}_{\mathrm{s}}^{(\mathrm{nc},r)}.
\end{aligned} \tag{B.439}$$

Vectorizing (B.439), the first term can be written as

$$\begin{aligned}
&\operatorname{vec}\left\{ \left(\boldsymbol{T}_{1:r-1}^{(r)\otimes} \otimes \boldsymbol{T}_r^{(\mathrm{nc},r)} \otimes \boldsymbol{T}_{r+1:R}^{(r)\otimes} \right) \cdot \Delta\boldsymbol{U}_{\mathrm{s}}^{(\mathrm{nc},r)} \right\} \\
&= \operatorname{vec}\left\{ \left(\boldsymbol{T}_{1:r-1}^{(r)\otimes} \otimes \boldsymbol{T}_r^{(\mathrm{nc},r)} \otimes \boldsymbol{T}_{r+1:R}^{(r)\otimes} \right) \cdot \boldsymbol{U}_{\mathrm{n}}^{(\mathrm{nc},r)} \cdot \boldsymbol{U}_{\mathrm{n}}^{(\mathrm{nc},r)^{\mathrm{H}}} \cdot \left[\boldsymbol{\mathcal{N}}^{(\mathrm{nc},r)} \right]_{(R+1)}^{\mathrm{T}} \cdot \boldsymbol{V}_{\mathrm{s}}^{(\mathrm{nc},r)} \cdot \boldsymbol{\Sigma}_{\mathrm{s}}^{(\mathrm{nc},r)^{-1}} \right\} \\
&= \left(\boldsymbol{\Sigma}_{\mathrm{s}}^{(\mathrm{nc},r)^{-1}} \cdot \boldsymbol{V}_{\mathrm{s}}^{(\mathrm{nc},r)^{\mathrm{T}}} \right) \otimes \left(\left(\hat{\boldsymbol{T}}_{1:r-1}^{(r)\otimes} \otimes \hat{\boldsymbol{T}}_r^{(\mathrm{nc},r)} \otimes \hat{\boldsymbol{T}}_{r+1:R}^{(r)\otimes} \right) \cdot \boldsymbol{U}_{\mathrm{n}}^{(\mathrm{nc},r)} \cdot \boldsymbol{U}_{\mathrm{n}}^{(\mathrm{nc},r)^{\mathrm{H}}} \right) \cdot \operatorname{vec}\left\{ \left[\boldsymbol{\mathcal{N}}^{(\mathrm{nc},r)} \right]_{(R+1)}^{\mathrm{T}} \right\} \\
&= \boldsymbol{W}_0^{(\mathrm{nc},r)} \cdot \operatorname{vec}\left\{ \left[\boldsymbol{\mathcal{N}}^{(\mathrm{nc},r)} \right]_{(R+1)}^{\mathrm{T}} \right\}.
\end{aligned} \tag{B.440}$$

where we have used the first-order perturbation of $\Delta\boldsymbol{U}_{\mathrm{s}}^{(\mathrm{nc},r)}$ in (6.63). For the vectorization of the first sum term in (B.439), we have $q < r$ such that $\boldsymbol{T}_{q+1:R}^{(\mathrm{nc},r)\otimes}$ contains the NC augmentation. We obtain

$$\begin{aligned}
&\sum_{q=1}^{r-1} \operatorname{vec}\left\{ \left(\boldsymbol{T}_{1:q-1}^{(r)\otimes} \otimes \left[\Delta\boldsymbol{U}_q^{[\mathrm{s}](r)} \cdot \boldsymbol{U}_q^{[\mathrm{s}](r)^{\mathrm{H}}} \right] \otimes \boldsymbol{T}_{q+1:R}^{(\mathrm{nc},r)\otimes} \right) \cdot \boldsymbol{U}_{\mathrm{s}}^{(\mathrm{nc},r)} \right\} \\
&= \left(\boldsymbol{U}_{\mathrm{s}}^{(\mathrm{nc},r)^{\mathrm{T}}} \otimes \boldsymbol{I}_{2\cdot M} \right) \cdot \sum_{q=1}^{r-1} \operatorname{vec}\left\{ \left(\boldsymbol{T}_{1:q-1}^{(r)\otimes} \otimes \left[\Delta\boldsymbol{U}_q^{[\mathrm{s}](r)} \cdot \boldsymbol{U}_q^{[\mathrm{s}](r)^{\mathrm{H}}} \right] \otimes \boldsymbol{T}_{q+1:R}^{(\mathrm{nc},r)\otimes} \right) \right\} \\
&= \left(\boldsymbol{U}_{\mathrm{s}}^{(\mathrm{nc},r)^{\mathrm{T}}} \otimes \boldsymbol{I}_{2\cdot M} \right) \cdot \sum_{q=1}^{r-1} \left(\bar{\boldsymbol{T}}_{1:q-1}^{(\mathrm{nc},r)} \otimes \boldsymbol{I}_{2\cdot M_{q:R}} \right) \cdot \left(\boldsymbol{I}_{M_q} \otimes \bar{\boldsymbol{T}}_{q+1:R}^{(\mathrm{nc},r)} \right) \cdot \operatorname{vec}\left\{ \Delta\boldsymbol{U}_q^{[\mathrm{s}](r)} \cdot \boldsymbol{U}_q^{[\mathrm{s}](r)^{\mathrm{H}}} \right\}
\end{aligned}$$

$$
\begin{aligned}
&= \Big(\boldsymbol{U}_{\mathrm{s}}^{(\mathrm{nc},r)^{\mathrm{T}}} \otimes \boldsymbol{I}_{2\cdot M}\Big) \cdot \sum_{q=1}^{r-1} \Big(\bar{\boldsymbol{T}}_{1:q-1}^{(\mathrm{nc},r)} \otimes \boldsymbol{I}_{2\cdot M_{q:R}}\Big) \cdot \Big(\boldsymbol{I}_{M_q} \otimes \bar{\boldsymbol{T}}_{q+1:R}^{(\mathrm{nc},r)}\Big) \\
&\quad \cdot \Big[\Big(\boldsymbol{V}_q^{[\mathrm{s}](r)} \cdot \boldsymbol{\Sigma}_q^{[\mathrm{s}](r)^{-1}} \cdot \boldsymbol{U}_q^{[\mathrm{s}](r)^{\mathrm{H}}}\Big)^{\mathrm{T}} \otimes \Big(\boldsymbol{U}_q^{[\mathrm{n}](r)} \cdot \boldsymbol{U}_q^{[\mathrm{n}](r)^{\mathrm{H}}}\Big)\Big] \cdot \operatorname{vec}\Big\{\Big[\boldsymbol{\mathcal{N}}^{(\mathrm{nc},r)}\Big]_{(q)}\Big\} \\
&= \sum_{q=1}^{r-1} \boldsymbol{W}_q^{(\mathrm{nc},r)} \cdot \operatorname{vec}\Big\{\Big[\boldsymbol{\mathcal{N}}^{(\mathrm{nc},r)}\Big]_{(q)}\Big\},
\end{aligned} \tag{B.441}
$$

where the matrices $\bar{\boldsymbol{T}}_{1:q-1}^{(\mathrm{nc},r)}$ and $\bar{\boldsymbol{T}}_{q+1:R}^{(\mathrm{nc},r)}$ are given in (6.72) and we have used the first-order error approximation of $\Delta\boldsymbol{U}_q^{[\mathrm{s}](r)}$ in (6.64).

For the term associated with $q = r$, we obtain

$$
\begin{aligned}
&\operatorname{vec}\Big\{\Big(\boldsymbol{T}_{1:r-1}^{(r)\otimes} \otimes \Big[\Delta\boldsymbol{U}_r^{[\mathrm{s}](r)} \cdot \boldsymbol{U}_r^{[\mathrm{s}](r)^{\mathrm{H}}}\Big] \otimes \boldsymbol{T}_{r+1:R}^{(r)\otimes}\Big) \cdot \boldsymbol{U}_{\mathrm{s}}^{(\mathrm{nc},r)}\Big\} \\
&= \Big(\boldsymbol{U}_{\mathrm{s}}^{(\mathrm{nc},r)^{\mathrm{T}}} \otimes \boldsymbol{I}_{2\cdot M}\Big) \cdot \operatorname{vec}\Big\{\boldsymbol{T}_{1:r-1}^{(r)\otimes} \otimes \Big[\Delta\boldsymbol{U}_r^{[\mathrm{s}](r)} \cdot \boldsymbol{U}_r^{[\mathrm{s}](r)^{\mathrm{H}}}\Big] \otimes \boldsymbol{T}_{r+1:R}^{(r)\otimes}\Big\} \\
&= \Big(\boldsymbol{U}_{\mathrm{s}}^{(\mathrm{nc},r)^{\mathrm{T}}} \otimes \boldsymbol{I}_{2\cdot M}\Big) \cdot \Big(\bar{\boldsymbol{T}}_{1:r-1}^{(\mathrm{nc},r)} \otimes \boldsymbol{I}_{2\cdot M_{r:R}}\Big) \cdot \Big(\boldsymbol{I}_{2\cdot M_r} \otimes \bar{\boldsymbol{T}}_{r+1:R}^{(\mathrm{nc},r)}\Big) \cdot \operatorname{vec}\Big\{\Delta\boldsymbol{U}_r^{[\mathrm{s}](r)} \cdot \boldsymbol{U}_r^{[\mathrm{s}](r)^{\mathrm{H}}}\Big\} \\
&= \Big(\boldsymbol{U}_{\mathrm{s}}^{(\mathrm{nc},r)^{\mathrm{T}}} \otimes \boldsymbol{I}_{2\cdot M}\Big) \cdot \Big(\bar{\boldsymbol{T}}_{1:r-1}^{(\mathrm{nc},r)} \otimes \boldsymbol{I}_{2\cdot M_{r:R}}\Big) \cdot \Big(\boldsymbol{I}_{2\cdot M_r} \otimes \bar{\boldsymbol{T}}_{r+1:R}^{(\mathrm{nc},r)}\Big) \qquad \text{(B.442)} \\
&\quad \cdot \Big[\Big(\boldsymbol{V}_r^{[\mathrm{s}](r)} \cdot \boldsymbol{\Sigma}_r^{[\mathrm{s}](r)^{-1}} \cdot \boldsymbol{U}_r^{[\mathrm{s}](r)^{\mathrm{H}}}\Big)^{\mathrm{T}} \otimes \Big(\boldsymbol{U}_r^{[\mathrm{n}](r)} \cdot \boldsymbol{U}_r^{[\mathrm{n}](r)^{\mathrm{H}}}\Big)\Big] \cdot \operatorname{vec}\Big\{\Big[\boldsymbol{\mathcal{N}}^{(\mathrm{nc},r)}\Big]_{(q)}\Big\} \\
&= \boldsymbol{W}_r^{(\mathrm{nc},r)} \cdot \operatorname{vec}\Big\{\Big[\boldsymbol{\mathcal{N}}^{(\mathrm{nc},r)}\Big]_{(q)}\Big\}, \qquad \text{(B.443)}
\end{aligned}
$$

where $\bar{\boldsymbol{T}}_{1:q-1}^{(\mathrm{nc},r)}$ and $\bar{\boldsymbol{T}}_{q+1:R}^{(\mathrm{nc},r)}$ are given in (6.74) and we have again used the first-order error approximation of $\Delta\boldsymbol{U}_q^{[\mathrm{s}](r)}$ in (6.64).

Finally, for the last term, we have $q > r$ such that the vectorization yields

$$
\begin{aligned}
&\sum_{q=r+1}^{R} \operatorname{vec}\Big\{\Big(\boldsymbol{T}_{1:q-1}^{(\mathrm{nc},r)\otimes} \otimes \Big[\Delta\boldsymbol{U}_q^{[\mathrm{s}](r)} \cdot \boldsymbol{U}_q^{[\mathrm{s}](r)^{\mathrm{H}}}\Big] \otimes \boldsymbol{T}_{q+1:R}^{(r)\otimes}\Big) \cdot \boldsymbol{U}_{\mathrm{s}}^{(\mathrm{nc},r)}\Big\} \\
&= \Big(\boldsymbol{U}_{\mathrm{s}}^{(\mathrm{nc},r)^{\mathrm{T}}} \otimes \boldsymbol{I}_{2\cdot M}\Big) \cdot \sum_{q=r+1}^{R} \operatorname{vec}\Big\{\boldsymbol{T}_{1:q-1}^{(\mathrm{nc},r)\otimes} \otimes \Big[\Delta\boldsymbol{U}_q^{[\mathrm{s}](r)} \cdot \boldsymbol{U}_q^{[\mathrm{s}](r)^{\mathrm{H}}}\Big] \otimes \boldsymbol{T}_{q+1:R}^{(r)\otimes}\Big\} \\
&= \Big(\boldsymbol{U}_{\mathrm{s}}^{(\mathrm{nc},r)^{\mathrm{T}}} \otimes \boldsymbol{I}_{2\cdot M}\Big) \cdot \sum_{q=r+1}^{R} \Big(\bar{\boldsymbol{T}}_{1:q-1}^{(\mathrm{nc},r)} \otimes \boldsymbol{I}_{M_{q:R}}\Big) \cdot \Big(\boldsymbol{I}_{M_q} \otimes \bar{\boldsymbol{T}}_{q+1:R}^{(\mathrm{nc},r)}\Big) \cdot \operatorname{vec}\Big\{\Delta\boldsymbol{U}_q^{[\mathrm{s}](r)} \cdot \boldsymbol{U}_q^{[\mathrm{s}](r)^{\mathrm{H}}}\Big\} \\
&= \Big(\boldsymbol{U}_{\mathrm{s}}^{(\mathrm{nc},r)^{\mathrm{T}}} \otimes \boldsymbol{I}_{2\cdot M}\Big) \cdot \sum_{q=r+1}^{R} \Big(\bar{\boldsymbol{T}}_{1:q-1}^{(\mathrm{nc},r)} \otimes \boldsymbol{I}_{M_{q:R}}\Big) \cdot \Big(\boldsymbol{I}_{M_q} \otimes \bar{\boldsymbol{T}}_{q+1:R}^{(\mathrm{nc},r)}\Big) \\
&\quad \cdot \Big[\Big(\boldsymbol{V}_q^{[\mathrm{s}](r)} \cdot \boldsymbol{\Sigma}_q^{[\mathrm{s}](r)^{-1}} \cdot \boldsymbol{U}_q^{[\mathrm{s}](r)^{\mathrm{H}}}\Big)^{\mathrm{T}} \otimes \Big(\boldsymbol{U}_q^{[\mathrm{n}](r)} \cdot \boldsymbol{U}_q^{[\mathrm{n}](r)^{\mathrm{H}}}\Big)\Big] \cdot \operatorname{vec}\Big\{\Big[\boldsymbol{\mathcal{N}}^{(\mathrm{nc},r)}\Big]_{(q)}\Big\} \\
&= \sum_{q=r+1}^{R} \boldsymbol{W}_q^{(\mathrm{nc},r)} \cdot \operatorname{vec}\Big\{\Big[\boldsymbol{\mathcal{N}}^{(\mathrm{nc},r)}\Big]_{(q)}\Big\}.
\end{aligned} \tag{B.444}
$$

Here, $\bar{\boldsymbol{T}}_{1:q-1}^{(\mathrm{nc},r)}$ and $\bar{\boldsymbol{T}}_{q+1:R}^{(\mathrm{nc},r)}$ are given in (6.77) and we have used (6.64) again.

In order to permute the vectorized noise contributions $\operatorname{vec}\left\{\left[\boldsymbol{\mathcal{N}}^{(\mathrm{nc},r)}\right]_{(q)}\right\}$ to be in consistent order with the noise vector $\boldsymbol{n}^{(\mathrm{nc},r)} = \operatorname{vec}\left\{\boldsymbol{N}^{(\mathrm{nc},r)}\right\} = \operatorname{vec}\left\{\left[\boldsymbol{\mathcal{N}}^{(\mathrm{nc},r)}\right]_{(R+1)}^{\mathrm{T}}\right\}$, we apply the permutation matrices defined in (1.38). Consequently, we obtain

$$\begin{aligned}
\operatorname{vec}\left\{\left[\boldsymbol{\mathcal{N}}^{(\mathrm{nc},r)}\right]_{(q)}\right\} &= \boldsymbol{P}^{(q)^{\mathrm{T}}}_{M_1,\ldots,M_{r-1},2M_r,M_{r+1},\ldots,M_R,N} \cdot \operatorname{vec}\left\{\boldsymbol{\mathcal{N}}^{(\mathrm{nc},r)}\right\} \\
&= \boldsymbol{P}^{(q)^{\mathrm{T}}}_{M_1,\ldots,M_{r-1},2M_r,M_{r+1},\ldots,M_R,N} \cdot \boldsymbol{P}^{(R)}_{M_1,\ldots,M_{r-1},2M_r,M_{r+1},\ldots,M_R,N} \cdot \operatorname{vec}\left\{\left[\boldsymbol{\mathcal{N}}^{(\mathrm{nc},r)}\right]_{(R)}\right\} \\
&= \boldsymbol{P}^{(q)^{\mathrm{T}}}_{M_1,\ldots,M_{r-1},2M_r,M_{r+1},\ldots,M_R,N} \cdot \boldsymbol{P}^{(R)}_{M_1,\ldots,M_{r-1},2M_r,M_{r+1},\ldots,M_R,N} \\
&\quad \cdot \operatorname{vec}\left\{\left[\boldsymbol{\mathcal{N}}^{(\mathrm{nc},r)}\right]_{(R+1)}^{\mathrm{T}}\right\},
\end{aligned} \tag{B.445}$$

where we have used property (1.42) in the last step.

Finally, we can express $\boldsymbol{W}_{\mathrm{ten}}^{(\mathrm{nc},r)}$ as

$$\boldsymbol{W}_{\mathrm{ten}}^{(\mathrm{nc},r)} = \boldsymbol{W}_0^{(\mathrm{nc},r)} + \sum_{q=1}^{R} \boldsymbol{W}_q^{(\mathrm{nc},r)} \cdot \boldsymbol{P}^{(q)^{\mathrm{T}}}_{M_1,\ldots,M_{r-1},2M_r,M_{r+1},\ldots,M_R,N} \cdot \boldsymbol{P}^{(R)}_{M_1,\ldots,M_{r-1},2M_r,M_{r+1},\ldots,M_R,N}, \tag{B.446}$$

which concludes the proof. □

B.29. Proof of Theorem 6.5.3

For the proof of Theorem 6.5.3 in Section 6.5.2, we simply extend the proof of Theorem 6.4.2 in Appendix B.25 to the tensor case. The procedure follows the same steps. Thus, we develop a first-order perturbation expansion for the parameter estimation error in the real-valued case for R-D NC Unitary Tensor-ESPRIT and then show the equivalence to the expression for the forward-backward averaged complex-valued case for R-D NC Standard Tensor-ESPRIT with FBA.

By consistently replacing all quantities in (6.65) by their forward-backward-averaged equivalents, we immediately obtain the first-order approximation of the estimation error after applying FBA as

$$\Delta\mu_i^{(r)} \approx \operatorname{Im}\left\{\boldsymbol{p}_i^{(\mathrm{fba})^{\mathrm{T}}} \cdot \left(\tilde{\boldsymbol{J}}_1^{(\mathrm{nc})(r)} \cdot \boldsymbol{U}_{\mathrm{s}}^{(\mathrm{nc},r)^{(\mathrm{fba})}}\right)^{+} \cdot \left[\tilde{\boldsymbol{J}}_2^{(\mathrm{nc})(r)}/\lambda_i^{(r)} - \tilde{\boldsymbol{J}}_1^{(\mathrm{nc})(r)}\right] \cdot \left[\Delta\boldsymbol{\mathcal{U}}^{[\mathrm{s}](r)^{(\mathrm{fba})}}\right]_{(R+1)}^{\mathrm{T}} \cdot \boldsymbol{q}_i^{(\mathrm{fba})}\right\}. \tag{B.447}$$

Then, performing the same steps as in the matrix case from Appendix B.25, we can express the

first-order approximation of the estimation error in the real-valued case as

$$\Delta\mu_i^{(r)} \approx \boldsymbol{p}_i^{(\varphi)^{\mathrm{T}}} \cdot \left(\tilde{\boldsymbol{K}}_1^{(\mathrm{nc})(r)} \cdot \boldsymbol{E}_{\mathrm{s}}^{(\mathrm{nc})(r)}\right)^{+} \cdot \left(\tilde{\boldsymbol{K}}_2^{(\mathrm{nc})(r)} - \omega_i^{(r)} \cdot \tilde{\boldsymbol{K}}_1^{(\mathrm{nc})(r)}\right) \cdot \left[\Delta\boldsymbol{\mathcal{E}}^{[\mathrm{s}](r)}\right]_{(3)}^{\mathrm{T}} \cdot \boldsymbol{q}_i^{(\varphi)} \cdot \frac{2}{{\omega_i^{(r)}}^2 + 1}. \tag{B.448}$$

Then, following the procedure from Appendix B.25, it is straightforward to show that (B.448) can be formulated into (B.447). This completes the proof. □

B.30. Proof of Theorem 6.6.1

This theorem in Section 6.6.1 consists of two parts, which are addressed in separate subsections.

B.30.1. MSE for R-D NC Standard ESPRIT for a single source

We start the proof by simplifying the MSE expression for R-D NC Standard ESPRIT in (6.52). In the single source case the noise-free NC measurement matrix can be written as

$$\boldsymbol{X}_0^{(\mathrm{nc})} = \boldsymbol{a}^{(\mathrm{nc})}(\boldsymbol{\mu}) \cdot \boldsymbol{s}^{\mathrm{T}}, \tag{B.449}$$

where $\boldsymbol{a}^{(\mathrm{nc})}(\boldsymbol{\mu}) = [\boldsymbol{a}^{\mathrm{T}}(\boldsymbol{\mu}), \tilde{\Psi} \cdot \boldsymbol{\Pi}_M \cdot \boldsymbol{a}^{\mathrm{H}}(\boldsymbol{\mu})]^{\mathrm{T}} \in \mathbb{C}^{2M\times 1}$ is the augmented array steering vector and $\boldsymbol{a}(\boldsymbol{\mu}) = \boldsymbol{a}^{(1)}(\mu^{(1)}) \otimes \cdots \otimes \boldsymbol{a}^{(R)}(\mu^{(R)}) \in \mathbb{C}^{M\times 1}$. Moreover, $\tilde{\Psi} = \Psi^* \cdot \Psi^* = \mathrm{e}^{-\mathrm{j}2\varphi}$, $\boldsymbol{s} \in \mathbb{C}^{N\times 1}$ contains the source symbols, and $\hat{P}_{\mathrm{s}} = \|\boldsymbol{s}\|_2^2/N$ is the empirical source power. In what follows, we drop the dependence of $\boldsymbol{a}^{(\mathrm{nc})}$ on $\boldsymbol{\mu}$ for notational convenience. If we assume a ULA of isotropic elements in each of the R modes, we have $\boldsymbol{a}^{(r)} = [1, \mathrm{e}^{\mathrm{j}\mu^{(r)}}, \ldots, \mathrm{e}^{\mathrm{j}(M_r-1)\mu^{(r)}}]^{\mathrm{T}}$ and $\left\|\boldsymbol{a}^{(\mathrm{nc})}\right\|_2^2 = 2M$. The selection matrices $\tilde{\boldsymbol{J}}_1^{(\mathrm{nc})(r)}$ and $\tilde{\boldsymbol{J}}_2^{(\mathrm{nc})(r)}$ are then chosen according to (6.7) with $\boldsymbol{J}_1^{(r)} = [\boldsymbol{I}_{M_r-1}, \boldsymbol{0}_{(M_r-1)\times 1}]$ and $\boldsymbol{J}_2^{(r)} = [\boldsymbol{0}_{(M_r-1)\times 1}, \boldsymbol{I}_{M_r-1}]$ for maximum overlap, i.e., $M_r^{(\mathrm{sel})} = M_r - 1$. Note that (B.449) is a rank-one matrix and we can directly determine the subspaces from the SVD as

$$\begin{aligned}
\boldsymbol{U}_{\mathrm{s}}^{(\mathrm{nc})} &= \boldsymbol{u}_{\mathrm{s}}^{(\mathrm{nc})} = \frac{\boldsymbol{a}^{(\mathrm{nc})}}{\left\|\boldsymbol{a}^{(\mathrm{nc})}\right\|_2} = \frac{\boldsymbol{a}^{(\mathrm{nc})}}{\sqrt{2M}} \\
\boldsymbol{\Sigma}_{\mathrm{s}}^{(\mathrm{nc})} &= \sigma_{\mathrm{s}}^{(\mathrm{nc})} = \sqrt{2MN\hat{P}_{\mathrm{s}}} \\
\boldsymbol{V}_{\mathrm{s}}^{(\mathrm{nc})} &= \boldsymbol{v}_{\mathrm{s}}^{(\mathrm{nc})} = \frac{\boldsymbol{s}^*}{\|\boldsymbol{s}\|_2} = \frac{\boldsymbol{s}^*}{\sqrt{N\hat{P}_{\mathrm{s}}}}.
\end{aligned}$$

For the MSE expression in (6.52), we also require $\boldsymbol{P}_{\boldsymbol{a}^{(\mathrm{nc})}}^{\perp} = \boldsymbol{U}_{\mathrm{n}}^{(\mathrm{nc})} \cdot \boldsymbol{U}_{\mathrm{n}}^{(\mathrm{nc})^{\mathrm{H}}} = \boldsymbol{I}_{2M} - \frac{1}{2M} \cdot \boldsymbol{a}^{(\mathrm{nc})} \cdot \boldsymbol{a}^{(\mathrm{nc})^{\mathrm{H}}}$, which is the projection matrix onto the noise subspace. Moreover, we have $\boldsymbol{\Psi}^{(r)} = \mathrm{e}^{\mathrm{j}\mu^{(r)}}$ and hence, the eigenvectors are $\boldsymbol{p}_i^{(r)} = \boldsymbol{q}_i^{(r)} = 1$. The SO moments $\boldsymbol{R}_{\mathrm{nn}}^{(\mathrm{nc})}$ and $\boldsymbol{C}_{\mathrm{nn}}^{(\mathrm{nc})}$ of the noise are given by (6.55).

Inserting these expressions into (6.56) for circularly symmetric white noise, we obtain

$$\mathbb{E}\left\{\left(\Delta\mu^{(r)}\right)^2\right\} \approx \frac{\sigma_\mathrm{n}^2}{2}\cdot\left(\left\|\boldsymbol{z}^{(\mathrm{nc})(r)}\right\|_2^2 - \mathrm{Re}\left\{\boldsymbol{z}^{(\mathrm{nc})(r)^\mathrm{T}}\cdot(\boldsymbol{I}_N\otimes\boldsymbol{\Pi}_{2M})\cdot\boldsymbol{z}^{(\mathrm{nc})(r)}\right\}\right), \tag{B.450}$$

where $\boldsymbol{z}^{(\mathrm{nc})(r)} = \boldsymbol{W}_\mathrm{mat}^{(\mathrm{nc})^\mathrm{T}}\cdot\boldsymbol{r}^{(\mathrm{nc})(r)}$ with

$$\boldsymbol{r}^{(\mathrm{nc})(r)} = \left[\left(\tilde{\boldsymbol{J}}_1^{(\mathrm{nc})(r)}\cdot\frac{\boldsymbol{a}^{(\mathrm{nc})}}{\sqrt{2M}}\right)^+\cdot\left(\tilde{\boldsymbol{J}}_2^{(\mathrm{nc})(r)}/\mathrm{e}^{\mathrm{j}\mu^{(r)}} - \tilde{\boldsymbol{J}}_1^{(\mathrm{nc})(r)}\right)\right]^\mathrm{T} \in \mathbb{C}^{2M\times 1} \tag{B.451}$$

$$\boldsymbol{W}_\mathrm{mat}^{(\mathrm{nc})} = \left(\frac{1}{\sqrt{2MN\hat{P}_\mathrm{s}}}\cdot\frac{\boldsymbol{s}^\mathrm{H}}{\sqrt{N\hat{P}_\mathrm{s}}}\right)\otimes\boldsymbol{P}_{\boldsymbol{a}^{(\mathrm{nc})}}^\perp \in \mathbb{C}^{2M\times 2MN}. \tag{B.452}$$

Note that the term $\boldsymbol{z}^{(\mathrm{nc})(r)}$ can also be written as $\boldsymbol{z}^{(\mathrm{nc})(r)} = \tilde{\boldsymbol{s}}^\mathrm{T}\otimes\tilde{\boldsymbol{a}}^{(r)^\mathrm{T}}$, where

$$\begin{aligned}\tilde{\boldsymbol{s}}^\mathrm{T} &= \frac{1}{\sqrt{2MN\hat{P}_\mathrm{s}}}\cdot\frac{\boldsymbol{s}^\mathrm{H}}{\sqrt{N\hat{P}_\mathrm{s}}},\\ \tilde{\boldsymbol{a}}^{(r)^\mathrm{T}} &= \left(\tilde{\boldsymbol{J}}_1^{(\mathrm{nc})(r)}\cdot\frac{\boldsymbol{a}^{(\mathrm{nc})}}{\sqrt{2M}}\right)^+\cdot\left(\tilde{\boldsymbol{J}}_2^{(\mathrm{nc})(r)}/\mathrm{e}^{\mathrm{j}\mu^{(r)}} - \tilde{\boldsymbol{J}}_1^{(\mathrm{nc})(r)}\right)\cdot\boldsymbol{P}_{\boldsymbol{a}^{(\mathrm{nc})}}^\perp.\end{aligned}$$

Thus, after straightforward calculations, the MSE in (B.450) is given by

$$\mathbb{E}\left\{\left(\Delta\mu^{(r)}\right)^2\right\} = \frac{\sigma_\mathrm{n}^2}{2}\cdot\left(\left\|\tilde{\boldsymbol{s}}^\mathrm{T}\right\|_2^2\cdot\left\|\tilde{\boldsymbol{a}}^{(r)^\mathrm{T}}\right\|_2^2 - \mathrm{Re}\left\{\tilde{\boldsymbol{s}}^\mathrm{T}\cdot\tilde{\boldsymbol{s}}\cdot\tilde{\boldsymbol{a}}^{(r)^\mathrm{T}}\cdot\boldsymbol{\Pi}_{2M}\cdot\tilde{\boldsymbol{a}}^{(r)}\right\}\right). \tag{B.453}$$

The first term $\left\|\tilde{\boldsymbol{s}}^\mathrm{T}\right\|_2^2$ of (B.453) can be conveniently expressed as $\left\|\tilde{\boldsymbol{s}}^\mathrm{T}\right\|_2^2 = \frac{1}{2MN\hat{P}_\mathrm{s}}$. For the second term $\left\|\tilde{\boldsymbol{a}}^{(r)^\mathrm{T}}\right\|_2^2$ of (B.453), we simplify $\tilde{\boldsymbol{a}}^{(r)^\mathrm{T}}$ and expand the pseudo-inverse of $\tilde{\boldsymbol{J}}_1^{(\mathrm{nc})(r)}\cdot\boldsymbol{a}^{(\mathrm{nc})(r)}$ using the relation $\boldsymbol{x}^+ = \boldsymbol{x}^\mathrm{H}/\left\|\boldsymbol{x}\right\|_2^2$. As $\tilde{\boldsymbol{J}}_1^{(\mathrm{nc})(r)}$ selects $2(M_r-1)$ out of $2M_r$ elements in the r-th mode, we have $\left\|\tilde{\boldsymbol{J}}_1^{(\mathrm{nc})(r)}\cdot\boldsymbol{a}^{(\mathrm{nc})(r)}\right\|_2^2 = \frac{M}{M_r}\cdot 2(M_r-1)$. Then, taking the shift invariance equation $\tilde{\boldsymbol{J}}_2^{(\mathrm{nc})(r)}\cdot\boldsymbol{a}^{(\mathrm{nc})}/\mathrm{e}^{\mathrm{j}\mu^{(r)}} - \tilde{\boldsymbol{J}}_1^{(\mathrm{nc})(r)}\cdot\boldsymbol{a}^{(\mathrm{nc})} = \boldsymbol{0}$ in the r-th mode into account, we obtain

$$\tilde{\boldsymbol{a}}^{(r)^\mathrm{T}} = \frac{\sqrt{2M}M_r}{2M(M_r-1)}\cdot\left(\boldsymbol{a}^{(\mathrm{nc})^\mathrm{H}}\cdot\tilde{\boldsymbol{J}}_1^{(\mathrm{nc})(r)^\mathrm{H}}\cdot\tilde{\boldsymbol{J}}_2^{(\mathrm{nc})(r)}/\mathrm{e}^{\mathrm{j}\mu^{(r)}} - \boldsymbol{a}^{(\mathrm{nc})^\mathrm{H}}\cdot\tilde{\boldsymbol{J}}_1^{(\mathrm{nc})(r)^\mathrm{H}}\cdot\tilde{\boldsymbol{J}}_1^{(\mathrm{nc})(r)}\right). \tag{B.454}$$

As a ULA is centro-symmetric, i.e., (6.10) holds, we can write $\boldsymbol{a}^{(\mathrm{nc})} = [1,\tilde{\Psi}]^\mathrm{T}\otimes\boldsymbol{a}$. Note that the phase term depending on the phase center in (6.10) cancels throughout the derivation and thus has been neglected. Since the vector $\boldsymbol{a}$ and the matrices $\tilde{\boldsymbol{J}}_k^{(\mathrm{nc})(r)}$, $k=1,2$, can be written as $\boldsymbol{a} = \boldsymbol{a}^{(1)}\otimes\cdots\otimes\boldsymbol{a}^{(R)}$ and $\tilde{\boldsymbol{J}}_k^{(\mathrm{nc})(r)} = \boldsymbol{I}_2\otimes\boldsymbol{I}_{\prod_{l=1}^{r-1}M_l}\otimes\boldsymbol{J}_k^{(r)}\otimes\boldsymbol{I}_{\prod_{l=r+1}^{R}M_l}$, all the unaffected modes can be

factored out of (B.454), yielding

$$\tilde{\boldsymbol{a}}^{(r)^{\mathrm{T}}} = \frac{\sqrt{2M}M_r}{2M(M_r-1)} \cdot \begin{bmatrix} 1 \\ \tilde{\Psi} \end{bmatrix}^{\mathrm{H}} \otimes \left(\boldsymbol{a}^{(1)} \otimes \cdots \otimes \boldsymbol{a}^{(r-1)}\right)^{\mathrm{H}} \otimes \left(\tilde{\boldsymbol{a}}_1^{(r)^{\mathrm{T}}} - \tilde{\boldsymbol{a}}_2^{(r)^{\mathrm{T}}}\right) \otimes \left(\boldsymbol{a}^{(r+1)} \otimes \cdots \otimes \boldsymbol{a}^{(R)}\right)^{\mathrm{H}}, \tag{B.455}$$

where

$$\begin{aligned} \tilde{\boldsymbol{a}}_1^{(r)^{\mathrm{T}}} &= \boldsymbol{a}^{(r)^{\mathrm{H}}} \cdot \boldsymbol{J}_1^{(r)^{\mathrm{H}}} \cdot \boldsymbol{J}_2^{(r)} / \mathrm{e}^{\mathrm{j}\mu^{(r)}} \quad \text{and} \\ \tilde{\boldsymbol{a}}_2^{(r)^{\mathrm{T}}} &= \boldsymbol{a}^{(r)^{\mathrm{H}}} \cdot \boldsymbol{J}_1^{(r)^{\mathrm{H}}} \cdot \boldsymbol{J}_1^{(r)}. \end{aligned}$$

Similarly to [RH12], it is easy to verify that

$$\begin{aligned} \tilde{\boldsymbol{a}}_1^{(r)^{\mathrm{T}}} &= \left[0, \mathrm{e}^{-\mathrm{j}\mu^{(r)}}, \ldots, \mathrm{e}^{-\mathrm{j}(M_r-2)\mu^{(r)}}, \mathrm{e}^{-\mathrm{j}(M_r-1)\mu^{(r)}}\right] \\ \tilde{\boldsymbol{a}}_2^{(r)^{\mathrm{T}}} &= \left[1, \mathrm{e}^{-\mathrm{j}\mu^{(r)}}, \ldots, \mathrm{e}^{-\mathrm{j}(M_r-2)\mu^{(r)}}, 0\right]. \end{aligned}$$

Consequently, we obtain

$$\begin{aligned} \left\|\tilde{\boldsymbol{a}}^{(r)^{\mathrm{T}}}\right\|_2^2 &= \frac{M_r^2}{2M(M_r-1)^2} \cdot 2 \cdot \prod_{n=1}^{r-1} \left\|\boldsymbol{a}^{(n)}\right\|_2^2 \cdot 2 \cdot \prod_{n=r+1}^{R} \left\|\boldsymbol{a}^{(n)}\right\|_2^2 \\ &= \frac{2M_r^2}{M(M_r-1)^2} \cdot \frac{M}{M_r} = \frac{2M_r}{(M_r-1)^2}. \end{aligned} \tag{B.456}$$

The third term $\tilde{\boldsymbol{s}}^{\mathrm{T}}\tilde{\boldsymbol{s}}$ of (B.453) can be simplified as $\tilde{\boldsymbol{s}}^{\mathrm{T}}\tilde{\boldsymbol{s}} = \frac{\tilde{\Psi}}{2MN\hat{P}_{\mathrm{s}}}$, where we have used the equality $\boldsymbol{s} = \Psi\boldsymbol{s}_0$ and the fact that $\boldsymbol{s}_0^{\mathrm{T}}\boldsymbol{s}_0 = N\hat{P}_{\mathrm{s}}$. Moreover, using (B.455), the last term of (B.453) can be reduced to $\tilde{\boldsymbol{a}}^{(r)^{\mathrm{T}}}\boldsymbol{\Pi}_{2M}\tilde{\boldsymbol{a}}^{(r)} = -\frac{2M_r\tilde{\Psi}^*}{(M_r-1)^2}$. Inserting these results into (B.453), we finally obtain for the MSE of R-D NC Standard ESPRIT

$$\mathbb{E}\left\{(\Delta\mu^{(r)})^2\right\} = \frac{\sigma_{\mathrm{n}}^2}{N\hat{P}_{\mathrm{s}}} \cdot \frac{M_r}{M(M_r-1)^2}, \tag{B.457}$$

which is the desired result. □

B.30.2. MSE for R-D NC Standard Tensor-ESPRIT for a single source

The second part of the theorem is to show that the MSE of R-D NC Standard Tensor-ESPRIT is the same as the MSE for R-D NC Standard ESPRIT for $d = 1$. For simplicity, we only show the proof for the special case $R = 2$. However, the expressions easily extend to the R-D case. Assuming circularly symmetric white noise, the MSE expression of 2-D NC Standard Tensor-ESPRIT for the

i-th spatial frequency in the r-th mode is given in (6.79) by

$$\mathbb{E}\left\{\left(\Delta\mu_i^{(r)}\right)^2\right\} \approx \frac{\sigma_{\mathrm{n}}^2}{2}\cdot\left(\left\|\boldsymbol{z}_i^{(\mathrm{nc},r)}\right\|_2^2 - \mathrm{Re}\left\{\boldsymbol{z}_i^{(\mathrm{nc},r)^{\mathrm{T}}}\cdot(\boldsymbol{I}_N\otimes\boldsymbol{\Pi}_{2M})\cdot\boldsymbol{z}_i^{(\mathrm{nc},r)}\right\}\right), \tag{B.458}$$

where $\boldsymbol{z}_i^{(\mathrm{nc},r)} = \boldsymbol{W}_{\mathrm{ten}}^{(\mathrm{nc},r)^{\mathrm{T}}}\cdot\boldsymbol{r}_i^{(\mathrm{nc},r)}$ for $r=1,2$ with

$$\boldsymbol{r}_i^{(r)} = \boldsymbol{q}_i\otimes\left(\left[\left(\tilde{\boldsymbol{J}}_1^{(\mathrm{nc})(r)}\boldsymbol{U}_{\mathrm{s}}^{(\mathrm{nc})(r)}\right)^{+}\cdot\left(\tilde{\boldsymbol{J}}_2^{(\mathrm{nc})(r)}/\lambda_i^{(r)} - \tilde{\boldsymbol{J}}_1^{(\mathrm{nc})(r)}\right)\right]^{\mathrm{T}}\boldsymbol{p}_i\right)$$

and

$$\begin{aligned}\boldsymbol{W}_{\mathrm{ten}}^{(r)} &= \left(\boldsymbol{\Sigma}_3^{[\mathrm{s}](r)^{-1}}\boldsymbol{U}_3^{[\mathrm{s}](r)^{\mathrm{H}}}\right)\otimes\left(\left(\boldsymbol{T}_1^{(r)}\otimes\boldsymbol{T}_2^{(r)}\right)\boldsymbol{V}_3^{[\mathrm{n}](r)^*}\boldsymbol{V}_3^{[\mathrm{n}](r)^{\mathrm{T}}}\right)\\ &\quad+\left(\boldsymbol{U}_{\mathrm{s}}^{(\mathrm{nc})(r)^{\mathrm{T}}}\otimes\boldsymbol{I}_{2M}\right)\bar{\boldsymbol{T}}_2^{(r)}\left(\boldsymbol{U}_1^{[\mathrm{s}](r)^*}\boldsymbol{\Sigma}_1^{[\mathrm{s}](r)^{-1}}\boldsymbol{V}_1^{[\mathrm{s}](r)^{\mathrm{T}}}\otimes\boldsymbol{U}_1^{[\mathrm{n}](r)}\boldsymbol{U}_1^{[\mathrm{n}](r)^{\mathrm{H}}}\right)\cdot\boldsymbol{K}_{M_2,M_1N}\\ &\quad+\left(\boldsymbol{U}_{\mathrm{s}}^{(\mathrm{nc})(r)^{\mathrm{T}}}\otimes\boldsymbol{I}_{2M}\right)\bar{\boldsymbol{T}}_1^{(r)}\left(\boldsymbol{U}_2^{[\mathrm{s}](r)^*}\boldsymbol{\Sigma}_2^{[\mathrm{s}](r)^{-1}}\boldsymbol{V}_2^{[\mathrm{s}](r)^{\mathrm{T}}}\otimes\boldsymbol{U}_2^{[\mathrm{n}](r)}\boldsymbol{U}_2^{[\mathrm{n}](r)^{\mathrm{H}}}\right).\end{aligned} \tag{B.459}$$

The matrices $\bar{\boldsymbol{T}}_\ell^{(r)}$, $\ell=1,2$, are defined as

$$\bar{\boldsymbol{T}}_1^{(r)} = \begin{bmatrix}\boldsymbol{I}_{M_2}\otimes\boldsymbol{t}_{1,1}^{(r)}\\ \vdots\\ \boldsymbol{I}_{M_2}\otimes\boldsymbol{t}_{1,M_1}^{(r)}\end{bmatrix}\otimes\boldsymbol{I}_{M_2},\quad \bar{\boldsymbol{T}}_2^{(r)} = \boldsymbol{I}_{M_1}\otimes\begin{bmatrix}\boldsymbol{I}_{M_1}\otimes\boldsymbol{t}_{2,1}^{(r)}\\ \vdots\\ \boldsymbol{I}_{M_1}\otimes\boldsymbol{t}_{2,M_2}^{(r)}\end{bmatrix} \tag{B.460}$$

with $\boldsymbol{t}_{\ell,m}^{(r)}$ denoting the m-th column of $\boldsymbol{T}_\ell^{(r)} = \boldsymbol{U}_\ell^{[\mathrm{s}](r)}\boldsymbol{U}_\ell^{[\mathrm{s}](r)^{\mathrm{H}}}$. Note that the MSE expression for R-D NC Standard Tensor-ESPRIT is in fact quite similar to the one for matrix-based R-D NC Standard ESPRIT with the only difference being that the matrix $\boldsymbol{W}_{\mathrm{mat}}^{(\mathrm{nc})}$ is replaced by the matrix $\boldsymbol{W}_{\mathrm{ten}}^{(r)}$. To simplify the MSE expression in (B.458), we first rewrite $\boldsymbol{\mathcal{X}}_0^{(\mathrm{nc},r)}$ into the symmetric model $\boldsymbol{\mathcal{X}}_0^{(\mathrm{nc},r)} = \boldsymbol{\mathcal{A}}_0^{(\mathrm{nc},r)}\times_{R+1}\boldsymbol{S}_0^{\mathrm{T}}$, where the symmetric array steering tensor $\boldsymbol{\mathcal{A}}_0^{(\mathrm{nc},r)}$ is given by

$$\boldsymbol{\mathcal{A}}_0^{(\mathrm{nc},r)} = \left[\boldsymbol{\mathcal{A}}\times_{R+1}\boldsymbol{\Psi}\underset{r}{\sqcup}\boldsymbol{\mathcal{A}}\times_{R+1}\boldsymbol{\Psi}^*\right]\in\mathbb{C}^{M_1\times\ldots\times M_{r-1}\times 2M_r\times M_{r+1}\times\ldots\times M_R\times d}, \tag{B.461}$$

where we have assumed the phase reference to be at the R-D array centroid. Note that this symmetric tensor model is the tensor extension of the symmetric matrix model in (8.19). Then, for the special case $d=1$ and the mode $r=1$ for the NC augmentation, the noise-free measurement tensor $\boldsymbol{\mathcal{X}}_0^{(\mathrm{nc},r)}$ simplifies to

$$\boldsymbol{\mathcal{X}}_0^{(\mathrm{nc},1)} = \left(\begin{bmatrix}\mathrm{e}^{\mathrm{j}\varphi}\\ \mathrm{e}^{-\mathrm{j}\varphi}\end{bmatrix}\otimes\boldsymbol{a}^{(1)}\right)\circ\boldsymbol{a}^{(2)}\circ\boldsymbol{s}_0 = \boldsymbol{a}^{(\mathrm{nc})(1)}\circ\boldsymbol{a}^{(2)}\circ\boldsymbol{s}_0\in\mathbb{C}^{2M_1\times M_2\times N}. \tag{B.462}$$

Then, the unfoldings of (B.462) are given by

$$\begin{aligned}\left[\boldsymbol{\mathcal{X}}_0^{(\mathrm{nc},1)}\right]_{(1)} &= \boldsymbol{a}^{(\mathrm{nc})(1)} \cdot \left(\boldsymbol{a}^{(2)} \otimes \boldsymbol{s}_0\right)^{\mathrm{T}}, \quad \left[\boldsymbol{\mathcal{X}}_0^{(\mathrm{nc},1)}\right]_{(2)} = \boldsymbol{a}^{(2)} \cdot \left(\boldsymbol{s}_0 \otimes \boldsymbol{a}^{(\mathrm{nc})(1)}\right)^{\mathrm{T}}, \\ \left[\boldsymbol{\mathcal{X}}_0^{(\mathrm{nc},1)}\right]_{(3)} &= \boldsymbol{s}_0 \cdot \left(\boldsymbol{a}^{(\mathrm{nc})(1)} \otimes \boldsymbol{a}^{(2)}\right)^{\mathrm{T}}.\end{aligned} \tag{B.463}$$

Consequently, we can relate the necessary subspaces of the unfoldings of $\boldsymbol{\mathcal{X}}_0^{(\mathrm{nc},1)}$ to $\boldsymbol{s}_0$ and $\boldsymbol{a}^{(r)}$ via

$$\begin{aligned}\boldsymbol{u}_1^{[\mathrm{s}](1)} &= \frac{\boldsymbol{a}^{(\mathrm{nc})(1)}}{\sqrt{2 \cdot M_1}}, \quad \boldsymbol{u}_2^{[\mathrm{s}](1)} = \frac{\boldsymbol{a}^{(2)}}{\sqrt{M_2}}, \quad \boldsymbol{u}_3^{[\mathrm{s}](1)} = \frac{\boldsymbol{s}_0}{\sqrt{N \cdot \hat{P}_{\mathrm{T}}}} \\ \boldsymbol{U}_1^{[\mathrm{n}](1)} &= \boldsymbol{I}_{2M_1} - \frac{\boldsymbol{a}^{(\mathrm{nc})(1)} \cdot \boldsymbol{a}^{(\mathrm{nc})(1)^{\mathrm{H}}}}{2 \cdot M_1}, \quad \boldsymbol{U}_2^{[\mathrm{n}](1)} = \boldsymbol{I}_{M_2} - \frac{\boldsymbol{a}^{(2)} \cdot \boldsymbol{a}^{(2)^{\mathrm{H}}}}{M_2} \\ \boldsymbol{\Sigma}_1^{[\mathrm{s}](1)} &= \boldsymbol{\Sigma}_2^{[\mathrm{s}](1)} = \boldsymbol{\Sigma}_3^{[\mathrm{s}](1)} = \sqrt{2 \cdot M \cdot N \cdot \hat{P}_{\mathrm{T}}} \\ \boldsymbol{v}_1^{[\mathrm{s}](1)} &= \frac{\left(\boldsymbol{a}^{(2)} \otimes \boldsymbol{s}_0\right)^*}{\sqrt{M_2 \cdot N \cdot \hat{P}_{\mathrm{T}}}}, \quad \boldsymbol{v}_2^{[\mathrm{s}](1)} = \frac{\left(\boldsymbol{s}_0 \otimes \boldsymbol{a}^{(\mathrm{nc})(1)}\right)^*}{\sqrt{2 \cdot M_1 \cdot N \cdot \hat{P}_{\mathrm{T}}}}, \quad \boldsymbol{v}_3^{[\mathrm{s}](1)} = \boldsymbol{u}_{\mathrm{s}}^{(1)} = \frac{\boldsymbol{a}^{(\mathrm{nc},1)}}{\sqrt{2 \cdot M}}\end{aligned} \tag{B.464}$$

$$\boldsymbol{V}_3^{[\mathrm{n}](1)^*} \cdot \boldsymbol{V}_3^{[\mathrm{n}](1)^{\mathrm{T}}} = \boldsymbol{U}_{\mathrm{n}}^{(\mathrm{nc},1)} \cdot \boldsymbol{U}_{\mathrm{n}}^{(\mathrm{nc},1)^{\mathrm{H}}} = \boldsymbol{I}_{2M} - \frac{\boldsymbol{a}^{(\mathrm{nc},1)} \cdot \boldsymbol{a}^{(\mathrm{nc},1)^{\mathrm{H}}}}{2 \cdot M}. \tag{B.465}$$

Moreover, we have for $\boldsymbol{T}_\ell^{(1)}$

$$\boldsymbol{T}_1^{(1)} = \boldsymbol{u}_1^{[\mathrm{s}](1)} \cdot \boldsymbol{u}_1^{[\mathrm{s}](1)^{\mathrm{H}}} = \frac{\boldsymbol{a}^{(\mathrm{nc})(1)} \cdot \boldsymbol{a}^{(\mathrm{nc})(1)^{\mathrm{H}}}}{2 \cdot M_1}, \quad \boldsymbol{T}_2^{(1)} = \boldsymbol{u}_2^{[\mathrm{s}](1)} \cdot \boldsymbol{u}_2^{[\mathrm{s}](1)^{\mathrm{H}}} = \frac{\boldsymbol{a}^{(2)} \cdot \boldsymbol{a}^{(2)^{\mathrm{H}}}}{M_2}$$

and thus

$$\begin{aligned}\boldsymbol{T}_1^{(1)} \otimes \boldsymbol{T}_2^{(1)} &= \frac{\boldsymbol{a}^{(\mathrm{nc})(1)} \cdot \boldsymbol{a}^{(\mathrm{nc})(1)^{\mathrm{H}}}}{2 \cdot M_1} \otimes \frac{\boldsymbol{a}^{(2)} \cdot \boldsymbol{a}^{(2)^{\mathrm{H}}}}{M_2} \\ &= \frac{(\boldsymbol{a}^{(\mathrm{nc})(1)} \otimes \boldsymbol{a}^{(2)}) \cdot (\boldsymbol{a}^{(\mathrm{nc})(1)} \otimes \boldsymbol{a}^{(2)})^{\mathrm{H}}}{2 \cdot M} = \frac{\boldsymbol{a}^{(\mathrm{nc},1)} \cdot \boldsymbol{a}^{(\mathrm{nc},1)^{\mathrm{H}}}}{2 \cdot M}\end{aligned} \tag{B.466}$$

From (B.465) and (B.466), it immediately follows that the first term in $\boldsymbol{W}_{\mathrm{ten}}^{(1)}$ cancels as it contains $\left[\boldsymbol{T}_1^{(1)} \otimes \boldsymbol{T}_2^{(1)}\right] \cdot \boldsymbol{V}_3^{[\mathrm{n}](1)^*} \cdot \boldsymbol{V}_3^{[\mathrm{n}](1)^{\mathrm{T}}}$. We also find $\boldsymbol{t}_{1,m_1}^{(1)} = [\mathrm{e}^{-\mathrm{j}\varphi}, \mathrm{e}^{\mathrm{j}\varphi}]^{\mathrm{T}} \otimes \boldsymbol{a}^{(\mathrm{nc})(1)} \cdot \mathrm{e}^{-\mathrm{j}\mu^{(1)} m_1}/2M_1$ for $m_1 = -(M_1-1)/2, \ldots, (M_1-1)/2$ and $\boldsymbol{t}_{2,m_2}^{(1)} = \boldsymbol{a}^{(2)} \cdot \mathrm{e}^{-\mathrm{j}\mu^{(1)} m_2}/M_2$ for $m_2 = -(M_2-1)/2, \ldots, (M_2-1)/2$. To simplify the remaining two terms in $\boldsymbol{W}_{\mathrm{ten}}^{(1)}$ we first look at some of their components. We start by $\left(\boldsymbol{u}_{\mathrm{s}}^{(1)^{\mathrm{T}}} \otimes \boldsymbol{I}_{2M}\right) \cdot \bar{\boldsymbol{T}}_1^{(1)}$ in (B.459). Using the identity $\boldsymbol{u}_{\mathrm{s}}^{(1)} = \boldsymbol{a}^{(\mathrm{nc},1)}/\sqrt{2 \cdot M}$ and the explicit expression for $\boldsymbol{t}_{r,m}$ we may write

$$\left(\boldsymbol{u}_{\mathrm{s}}^{(1)^{\mathrm{T}}} \otimes \boldsymbol{I}_{2M}\right) \cdot \bar{\boldsymbol{T}}_1^{(1)} = \frac{1}{\sqrt{2 \cdot M}} \left(\boldsymbol{a}^{(\mathrm{nc},1)^{\mathrm{T}}} \otimes \boldsymbol{I}_{2M}\right) \cdot \bar{\boldsymbol{T}}_1^{(1)}$$

$$
\begin{aligned}
&= \frac{1}{\sqrt{2 \cdot M}} \left(\boldsymbol{a}^{(\mathrm{nc})(1)^{\mathrm{T}}} \otimes \boldsymbol{a}^{(2)^{\mathrm{T}}} \otimes \boldsymbol{I}_{2M_1} \otimes \boldsymbol{I}_{M_2} \right) \cdot \bar{\boldsymbol{T}}_1^{(1)} \\
&= \frac{1}{\sqrt{2 \cdot M}} \left(\left[\mathrm{e}^{\mathrm{j}\varphi} \quad \mathrm{e}^{-\mathrm{j}\varphi} \right] \otimes \left[\mathrm{e}^{-\mathrm{j}\mu^{(1)} \frac{(M_1-1)}{2}}, \ldots, \mathrm{e}^{\mathrm{j}\mu^{(1)} \frac{(M_1-1)}{2}} \right] \otimes \boldsymbol{a}^{(2)^{\mathrm{T}}} \otimes \boldsymbol{I}_{2M_1} \otimes \boldsymbol{I}_{M_2} \right) \cdot \bar{\boldsymbol{T}}_1^{(1)} \\
&= \frac{1}{\sqrt{2 \cdot M}} \left(\left[\mathrm{e}^{\mathrm{j}\varphi} \quad \mathrm{e}^{-\mathrm{j}\varphi} \right] \otimes \left[\mathrm{e}^{-\mathrm{j}\mu^{(1)} \frac{(M_1-1)}{2}}, \ldots, \mathrm{e}^{\mathrm{j}\mu^{(1)} \frac{(M_1-1)}{2}} \right] \otimes \boldsymbol{a}^{(2)^{\mathrm{T}}} \otimes \boldsymbol{I}_{2M_1} \otimes \boldsymbol{I}_{M_2} \right) \\
&\quad \cdot \frac{1}{2 \cdot M_1} \cdot \begin{bmatrix} \mathrm{e}^{-\mathrm{j}\varphi} \\ \mathrm{e}^{\mathrm{j}\varphi} \end{bmatrix} \otimes \begin{bmatrix} \boldsymbol{I}_{M_2} \otimes \boldsymbol{a}^{(\mathrm{nc})(1)} \cdot \mathrm{e}^{\mathrm{j}\mu^{(1)} \frac{(M_1-1)}{2}} \otimes \boldsymbol{I}_{M_2} \\ \boldsymbol{I}_{M_2} \otimes \boldsymbol{a}^{(\mathrm{nc})(1)} \cdot \mathrm{e}^{\mathrm{j}\mu^{(1)} \frac{(M_1-3)}{2}} \otimes \boldsymbol{I}_{M_2} \\ \vdots \\ \boldsymbol{I}_{M_2} \otimes \boldsymbol{a}^{(\mathrm{nc})(1)} \cdot \mathrm{e}^{-\mathrm{j}\mu^{(1)} \frac{(M_1-3)}{2}} \otimes \boldsymbol{I}_{M_2} \\ \boldsymbol{I}_{M_2} \otimes \boldsymbol{a}^{(\mathrm{nc})(1)} \cdot \mathrm{e}^{-\mathrm{j}\mu^{(1)} \frac{(M_1-1)}{2}} \otimes \boldsymbol{I}_{M_2} \end{bmatrix} \\
&= \frac{1}{2 \cdot M_1 \sqrt{2 \cdot M}} \cdot 2 \cdot \sum_{m_1=-\frac{(M_1-1)}{2}}^{\frac{(M_1-1)}{2}} \mathrm{e}^{\mathrm{j}\mu^{(1)} m_1} \cdot \left(\boldsymbol{a}^{(2)^{\mathrm{T}}} \otimes \boldsymbol{I}_{2M_1} \otimes \boldsymbol{I}_{M_2} \right) \cdot \left(\boldsymbol{I}_{M_2} \otimes \boldsymbol{a}^{(\mathrm{nc})(1)} \otimes \boldsymbol{I}_{M_2} \right) \cdot \mathrm{e}^{-\mathrm{j}\mu^{(1)} m_1} \\
&= \frac{2}{2 \cdot M_1 \sqrt{2 \cdot M}} \left(\boldsymbol{a}^{(2)^{\mathrm{T}}} \otimes \boldsymbol{a}^{(\mathrm{nc})(1)} \otimes \boldsymbol{I}_{M_2} \right) \cdot \underbrace{\sum_{m_1=-\frac{(M_1-1)}{2}}^{\frac{(M_1-1)}{2}} 1}_{M_1} = \frac{1}{\sqrt{2 \cdot M}} \left(\boldsymbol{a}^{(2)^{\mathrm{T}}} \otimes \boldsymbol{a}^{(\mathrm{nc})(1)} \otimes \boldsymbol{I}_{M_2} \right), \qquad \text{(B.467)}
\end{aligned}
$$

where we have used the fact that $\boldsymbol{x}^{\mathrm{T}} \otimes \boldsymbol{Y} = \left[x_1 \cdot \boldsymbol{Y}, \ldots, x_N \cdot \boldsymbol{Y} \right]$, where x_n is the n-th element of $\boldsymbol{x} \in \mathbb{C}^{N \times 1}$, and $\boldsymbol{Y}$ is a matrix of arbitrary size. Note that in (B.467), we have assumed that M_1 is odd. However, the same result can be found for even M_1. By applying similar reasoning to $\left(\boldsymbol{u}_{\mathrm{s}}^{(1)^{\mathrm{T}}} \otimes \boldsymbol{I}_{2M} \right) \cdot \bar{\boldsymbol{T}}_2^{(1)}$ in (B.459), we can show

$$
\begin{aligned}
&\left(\boldsymbol{u}_{\mathrm{s}}^{(1)^{\mathrm{T}}} \otimes \boldsymbol{I}_{2M} \right) \cdot \bar{\boldsymbol{T}}_2^{(1)} = \frac{1}{\sqrt{2 \cdot M}} \left(\boldsymbol{a}^{(\mathrm{nc},1)^{\mathrm{T}}} \otimes \boldsymbol{I}_{2M} \right) \cdot \bar{\boldsymbol{T}}_2^{(1)} \\
&= \frac{1}{\sqrt{2 \cdot M}} \left(\boldsymbol{a}^{(\mathrm{nc})(1)^{\mathrm{T}}} \otimes \boldsymbol{a}^{(2)^{\mathrm{T}}} \otimes \boldsymbol{I}_{2M_1} \otimes \boldsymbol{I}_{M_2} \right) \cdot \bar{\boldsymbol{T}}_2^{(1)} \\
&= \frac{1}{\sqrt{2 \cdot M}} \left(\boldsymbol{a}^{(\mathrm{nc})(1)^{\mathrm{T}}} \otimes \left[\mathrm{e}^{-\mathrm{j}\mu^{(1)} \frac{(M_2-1)}{2}}, \ldots, \mathrm{e}^{\mathrm{j}\mu^{(1)} \frac{(M_2-1)}{2}} \right] \otimes \boldsymbol{I}_{2M_1} \otimes \boldsymbol{I}_{M_2} \right) \cdot \bar{\boldsymbol{T}}_2^{(1)} \\
&= \frac{1}{\sqrt{2 \cdot M}} \left(\boldsymbol{a}^{(\mathrm{nc})(1)^{\mathrm{T}}} \otimes \left[\mathrm{e}^{-\mathrm{j}\mu^{(1)} \frac{(M_2-1)}{2}}, \ldots, \mathrm{e}^{\mathrm{j}\mu^{(1)} \frac{(M_2-1)}{2}} \right] \otimes \boldsymbol{I}_{2M_1} \otimes \boldsymbol{I}_{M_2} \right) \\
&\quad \cdot \frac{1}{M_2} \cdot \boldsymbol{I}_{2M_1} \otimes \begin{bmatrix} \boldsymbol{I}_{2M_1} \otimes \boldsymbol{a}^{(2)} \cdot \mathrm{e}^{\mathrm{j}\mu^{(1)} \frac{(M_2-1)}{2}} \\ \vdots \\ \boldsymbol{I}_{2M_1} \otimes \boldsymbol{a}^{(2)} \cdot \mathrm{e}^{-\mathrm{j}\mu^{(1)} \frac{(M_2-1)}{2}} \end{bmatrix} \\
&= \frac{1}{M_2 \sqrt{2 \cdot M}} \sum_{m_2=-\frac{(M_2-1)}{2}}^{\frac{(M_2-1)}{2}} \mathrm{e}^{\mathrm{j}\mu^{(1)} m_2} \cdot \left(\boldsymbol{a}^{(\mathrm{nc})(1)^{\mathrm{T}}} \otimes \boldsymbol{I}_{2M_1} \otimes \boldsymbol{I}_{M_2} \right) \cdot \left(\boldsymbol{I}_{2M_1} \otimes \boldsymbol{I}_{2M_1} \otimes \boldsymbol{a}^{(2)} \right) \cdot \mathrm{e}^{-\mathrm{j}\mu^{(1)} m_2}
\end{aligned}
$$

$$= \frac{1}{M_2\sqrt{2\cdot M}} \left(\boldsymbol{a}^{(\mathrm{nc})(1)^{\mathrm{T}}} \otimes \boldsymbol{I}_{2M_1} \otimes \boldsymbol{a}^{(2)} \right) \cdot \underbrace{\sum_{m_2=-\frac{(M_2-1)}{2}}^{\frac{(M_2-1)}{2}} 1}_{M_2} = \frac{1}{\sqrt{2\cdot M}} \left(\boldsymbol{a}^{(\mathrm{nc})(1)^{\mathrm{T}}} \otimes \boldsymbol{I}_{2M_1} \otimes \boldsymbol{a}^{(2)} \right). \quad \text{(B.468)}$$

Again, we have used that M_2 is odd for simplicity. The remaining terms in $\boldsymbol{W}_{\mathrm{ten}}^{(r)}$ we have not yet associated with $\boldsymbol{s}$ and $\boldsymbol{a}^{(r)}$ are $\boldsymbol{U}_r^{[\mathrm{s}]^*} \boldsymbol{\Sigma}_r^{[\mathrm{s}]^{-1}} \boldsymbol{V}_r^{[\mathrm{s}]^{\mathrm{T}}} \otimes \boldsymbol{U}_r^{[\mathrm{n}](1)} \boldsymbol{U}_r^{[\mathrm{n}](1)^{\mathrm{H}}}$. Using the relations (B.464)-(B.465) we can write

$$\left(\boldsymbol{U}_1^{[\mathrm{s}](1)^*} \boldsymbol{\Sigma}_1^{[\mathrm{s}](1)^{-1}} \boldsymbol{V}_1^{[\mathrm{s}](1)^{\mathrm{T}}} \right) \otimes \left(\boldsymbol{U}_1^{[\mathrm{n}](1)} \boldsymbol{U}_1^{[\mathrm{n}](1)^{\mathrm{H}}} \right) = \frac{1}{2MN\hat{P}_{\mathrm{T}}} \cdot \left(\boldsymbol{a}^{(\mathrm{nc})(1)^*} \cdot \left(\boldsymbol{a}^{(2)} \otimes \boldsymbol{s}_0 \right)^{\mathrm{H}} \right) \otimes \boldsymbol{\Pi}^{\perp}_{\boldsymbol{a}^{(\mathrm{nc})(1)}}$$
$$\left(\boldsymbol{U}_2^{[\mathrm{s}](1)^*} \boldsymbol{\Sigma}_2^{[\mathrm{s}](1)^{-1}} \boldsymbol{V}_2^{[\mathrm{s}](1)^{\mathrm{T}}} \right) \otimes \left(\boldsymbol{U}_2^{[\mathrm{n}](1)} \boldsymbol{U}_2^{[\mathrm{n}](1)^{\mathrm{H}}} \right) = \frac{1}{2MN\hat{P}_{\mathrm{T}}} \cdot \left(\boldsymbol{a}^{(2)^*} \cdot \left(\boldsymbol{s}_0 \otimes \boldsymbol{a}^{(\mathrm{nc})(1)} \right)^{\mathrm{H}} \right) \otimes \boldsymbol{\Pi}^{\perp}_{\boldsymbol{a}^{(2)}}$$

where we have used the short-hand notation $\boldsymbol{\Pi}^{\perp}_{\boldsymbol{x}} = \boldsymbol{I}_N - \frac{\boldsymbol{x}\boldsymbol{x}^{\mathrm{H}}}{\boldsymbol{x}^{\mathrm{H}}\boldsymbol{x}}$ for the projection matrix onto the orthogonal complement of the vector $\boldsymbol{x} \in \mathbb{C}^{N\times 1}$. Combining these intermediate result, the last term in $\boldsymbol{W}_{\mathrm{ten}}^{(1)}$ can be expressed as

$$\begin{aligned}
&\left(\boldsymbol{u}_{\mathrm{s}}^{(1)^{\mathrm{T}}} \otimes \boldsymbol{I}_{2M} \right) \cdot \bar{\boldsymbol{T}}_1^{(1)} \cdot \left(\boldsymbol{U}_2^{[\mathrm{s}](1)^*} \boldsymbol{\Sigma}_2^{[\mathrm{s}](1)^{-1}} \boldsymbol{V}_2^{[\mathrm{s}](1)^{\mathrm{T}}} \right) \otimes \left(\boldsymbol{U}_2^{[\mathrm{n}](1)} \boldsymbol{U}_2^{[\mathrm{n}](1)^{\mathrm{H}}} \right) \\
=& \frac{1}{2MN\hat{P}_{\mathrm{T}}\sqrt{2M}} \left(\boldsymbol{a}^{(2)^{\mathrm{T}}} \otimes \boldsymbol{a}^{(\mathrm{nc})(1)} \otimes \boldsymbol{I}_{M_2} \right) \cdot \left[\left(\boldsymbol{a}^{(2)^*} \cdot \left(\boldsymbol{s}_0 \otimes \boldsymbol{a}^{(\mathrm{nc})(1)} \right)^{\mathrm{H}} \right) \otimes \boldsymbol{\Pi}^{\perp}_{\boldsymbol{a}^{(2)}} \right] \\
=& \frac{1}{2MN\hat{P}_{\mathrm{T}}\sqrt{2M}} \left\| \boldsymbol{a}^{(2)} \right\|_2^2 \cdot \left(\boldsymbol{s}_0 \otimes \boldsymbol{a}^{(\mathrm{nc})(1)} \right)^{\mathrm{H}} \otimes \left[\left(\boldsymbol{a}^{(\mathrm{nc})(1)} \otimes \boldsymbol{I}_{M_2} \right) \cdot \boldsymbol{\Pi}^{\perp}_{\boldsymbol{a}^{(2)}} \right] \\
=& \frac{1}{2MN\hat{P}_{\mathrm{T}}\sqrt{2M}} \left(M_2 \cdot \boldsymbol{s}_0^{\mathrm{H}} \otimes \boldsymbol{a}^{(\mathrm{nc})(1)^{\mathrm{H}}} \otimes \boldsymbol{a}^{(\mathrm{nc})(1)} \otimes \boldsymbol{\Pi}^{\perp}_{\boldsymbol{a}^{(2)}} \right)
\end{aligned} \quad \text{(B.469)}$$

With similar arguments, the second term in $\boldsymbol{W}_{\mathrm{ten}}^{(1)}$ can be simplified into

$$\begin{aligned}
&\left(\boldsymbol{u}_{\mathrm{s}}^{(1)^{\mathrm{T}}} \otimes \boldsymbol{I}_{2M} \right) \cdot \bar{\boldsymbol{T}}_2^{(1)} \cdot \left(\boldsymbol{U}_1^{[\mathrm{s}](1)^*} \boldsymbol{\Sigma}_1^{[\mathrm{s}](1)^{-1}} \boldsymbol{V}_1^{[\mathrm{s}](1)^{\mathrm{T}}} \right) \otimes \left(\boldsymbol{U}_1^{[\mathrm{n}](1)} \boldsymbol{U}_1^{[\mathrm{n}](1)^{\mathrm{H}}} \right) \cdot \boldsymbol{K}_{M_2\times(2M_1\cdot N)} \\
=& \frac{1}{2MN\hat{P}_{\mathrm{T}}\sqrt{2M}} \left(\boldsymbol{a}^{(\mathrm{nc})(1)^{\mathrm{T}}} \otimes \boldsymbol{I}_{2M_1} \otimes \boldsymbol{a}^{(2)} \right) \cdot \left[\left(\boldsymbol{a}^{(\mathrm{nc})(1)^*} \cdot \left(\boldsymbol{a}^{(2)} \otimes \boldsymbol{s}_0 \right)^{\mathrm{H}} \right) \otimes \boldsymbol{\Pi}^{\perp}_{\boldsymbol{a}^{(\mathrm{nc})(1)}} \right] \cdot \boldsymbol{K}_{M_2\times(2M_1\cdot N)} \\
=& \frac{1}{2MN\hat{P}_{\mathrm{T}}\sqrt{2M}} \left\| \boldsymbol{a}^{(\mathrm{nc})(1)} \right\|_2^2 \cdot \left(\boldsymbol{a}^{(2)} \otimes \boldsymbol{s}_0 \right)^{\mathrm{H}} \otimes \left[\left(\boldsymbol{I}_{2M_1} \otimes \boldsymbol{a}^{(2)} \right) \cdot \boldsymbol{\Pi}^{\perp}_{\boldsymbol{a}^{(\mathrm{nc})(1)}} \right] \\
=& \frac{1}{2MN\hat{P}_{\mathrm{T}}\sqrt{2M}} \left(2M_1 \cdot \boldsymbol{a}^{(2)^{\mathrm{H}}} \otimes \boldsymbol{s}_0^{\mathrm{H}} \otimes \boldsymbol{\Pi}^{\perp}_{\boldsymbol{a}^{(\mathrm{nc})(1)}} \otimes \boldsymbol{a}^{(2)} \right) \cdot \boldsymbol{K}_{M_2\times(2M_1\cdot N)} \\
=& \frac{1}{2MN\hat{P}_{\mathrm{T}}\sqrt{2M}} \left(2M_1 \cdot \boldsymbol{s}_0^{\mathrm{H}} \otimes \boldsymbol{\Pi}^{\perp}_{\boldsymbol{a}^{(\mathrm{nc})(1)}} \otimes \boldsymbol{a}^{(2)} \otimes \boldsymbol{a}^{(2)^{\mathrm{H}}} \right)
\end{aligned} \quad \text{(B.470)}$$

where the last step is a special case of property (1.22) for commutation matrices.

Using (B.469) and (B.470) in (B.459), we obtain

$$\begin{aligned}
\boldsymbol{W}_{\text{ten}}^{(1)} &= \frac{1}{2MN\hat{P}_{\text{T}}\sqrt{2M}}\left(2M_1 \cdot \boldsymbol{s}_0^{\text{H}} \otimes \boldsymbol{\Pi}^{\perp}_{\boldsymbol{a}^{(\text{nc})(1)}} \otimes \boldsymbol{a}^{(2)} \otimes \boldsymbol{a}^{(2)^{\text{H}}} + M_2 \cdot \boldsymbol{s}_0^{\text{H}} \otimes \boldsymbol{a}^{(\text{nc})(1)^{\text{H}}} \otimes \boldsymbol{a}^{(\text{nc})(1)} \otimes \boldsymbol{\Pi}^{\perp}_{\boldsymbol{a}^{(2)}}\right) \\
&= \frac{1}{2MN\hat{P}_{\text{T}}\sqrt{2M}} \cdot \boldsymbol{s}_0^{\text{H}} \otimes \left(2M_1 \cdot \boldsymbol{\Pi}^{\perp}_{\boldsymbol{a}^{(\text{nc})(1)}} \otimes \left(\boldsymbol{a}^{(2)} \cdot \boldsymbol{a}^{(2)^{\text{H}}}\right) + M_2 \cdot \left(\boldsymbol{a}^{(\text{nc})(1)} \cdot \boldsymbol{a}^{(\text{nc})(1)^{\text{H}}}\right) \otimes \boldsymbol{\Pi}^{\perp}_{\boldsymbol{a}^{(2)}}\right) \\
&= \frac{1}{N\hat{P}_{\text{T}}\sqrt{2M}} \cdot \boldsymbol{s}_0^{\text{H}} \otimes \left(\boldsymbol{\Pi}^{\perp}_{\boldsymbol{a}^{(\text{nc})(1)}} \otimes \frac{\boldsymbol{a}^{(2)} \cdot \boldsymbol{a}^{(2)^{\text{H}}}}{M_2} + \frac{\boldsymbol{a}^{(\text{nc})(1)} \cdot \boldsymbol{a}^{(\text{nc})(1)^{\text{H}}}}{2M_1} \otimes \boldsymbol{\Pi}^{\perp}_{\boldsymbol{a}^{(2)}}\right) \\
&= \frac{1}{N\hat{P}_{\text{T}}\sqrt{2M}} \cdot \boldsymbol{s}_0^{\text{H}} \otimes \left(\boldsymbol{\Pi}^{\perp}_{\boldsymbol{a}^{(\text{nc})(1)}} \otimes \boldsymbol{\Pi}_{\boldsymbol{a}^{(2)}} + \boldsymbol{\Pi}_{\boldsymbol{a}^{(\text{nc})(1)}} \otimes \boldsymbol{\Pi}^{\perp}_{\boldsymbol{a}^{(2)}}\right), \qquad \text{(B.471)}
\end{aligned}$$

where we have applied the rule $\boldsymbol{x} \otimes \boldsymbol{y}^{\text{T}} = \boldsymbol{x} \cdot \boldsymbol{y}^{\text{T}}$ and used the short-hand notation $\boldsymbol{\Pi}_{\boldsymbol{x}} = \frac{\boldsymbol{x}\boldsymbol{x}^{\text{H}}}{\boldsymbol{x}^{\text{H}}\boldsymbol{x}}$ for the projection matrix onto the vector $\boldsymbol{x} \in \mathbb{C}^{N\times 1}$. Comparing (B.471) with (B.452) from them matrix case in Appendix B.30.1 we find that for a single source, $\boldsymbol{W}_{\text{mat}}^{(\text{nc})}$ and $\boldsymbol{W}_{\text{ten}}^{(1)}$ are in fact quite similar, the only difference being that $\boldsymbol{\Pi}^{\perp}_{\boldsymbol{a}^{(\text{nc})}}$ is replaced by $\boldsymbol{\Pi}^{\perp}_{\boldsymbol{a}^{(\text{nc})(1)}} \otimes \boldsymbol{\Pi}_{\boldsymbol{a}^{(2)}} + \boldsymbol{\Pi}_{\boldsymbol{a}^{(\text{nc})(1)}} \otimes \boldsymbol{\Pi}^{\perp}_{\boldsymbol{a}^{(2)}}$. Therefore, to show the R-D NC Standard ESPRIT and R-D NC Standard Tensor-ESPRIT have the same MSE for $d = 1$, it is sufficient to show that the corresponding terms $\tilde{\boldsymbol{a}}^{(r)}$ are the same, i.e., that

$$\begin{aligned}
&\boldsymbol{a}^{(\text{nc})^{\text{H}}} \tilde{\boldsymbol{J}}_1^{(\text{nc})(r)^{\text{H}}} \left(\tilde{\boldsymbol{J}}_2^{(\text{nc})(r)}/\mathrm{e}^{\mathrm{j}\cdot\mu^{(r)}} - \tilde{\boldsymbol{J}}_1^{(\text{nc})(r)}\right) \cdot \boldsymbol{\Pi}^{\perp}_{\boldsymbol{a}^{(\text{nc})}} \\
&= \boldsymbol{a}^{(\text{nc},1)^{\text{H}}} \tilde{\boldsymbol{J}}_1^{(\text{nc})(r)^{\text{H}}} \left(\tilde{\boldsymbol{J}}_2^{(\text{nc})(r)}/\mathrm{e}^{\mathrm{j}\cdot\mu^{(r)}} - \tilde{\boldsymbol{J}}_1^{(\text{nc})(r)}\right) \cdot \left(\boldsymbol{\Pi}^{\perp}_{\boldsymbol{a}^{(\text{nc})(1)}} \otimes \boldsymbol{\Pi}_{\boldsymbol{a}^{(2)}} + \boldsymbol{\Pi}_{\boldsymbol{a}^{(\text{nc})(1)}} \otimes \boldsymbol{\Pi}^{\perp}_{\boldsymbol{a}^{(2)}}\right) \qquad \text{(B.472)}
\end{aligned}$$

for $r = 1, 2$. Note that the left-hand side of (B.472) was shown to be equal to (cf. equation (B.455))

$$\begin{bmatrix} \mathrm{e}^{\mathrm{j}\varphi} \\ \mathrm{e}^{-\mathrm{j}\varphi} \end{bmatrix}^{\text{H}} \otimes \left(\tilde{\boldsymbol{a}}_1^{(1)} - \tilde{\boldsymbol{a}}_2^{(1)}\right)^{\text{T}} \otimes \boldsymbol{a}^{(2)^{\text{T}}}, \quad \text{for } r = 1 \text{ and } \quad \begin{bmatrix} \mathrm{e}^{\mathrm{j}\varphi} \\ \mathrm{e}^{-\mathrm{j}\varphi} \end{bmatrix}^{\text{H}} \otimes \boldsymbol{a}^{(1)^{\text{T}}} \otimes \left(\tilde{\boldsymbol{a}}_1^{(2)} - \tilde{\boldsymbol{a}}_2^{(2)}\right)^{\text{T}} \quad \text{for } r = 2 \qquad \text{(B.473)}$$

where $\tilde{\boldsymbol{a}}_1^{(r)^{\text{T}}} = \boldsymbol{a}^{(r)^{\text{H}}} \cdot \boldsymbol{J}_1^{(r)^{\text{H}}} \cdot \boldsymbol{J}_2^{(r)}/\mathrm{e}^{\mathrm{j}\cdot\mu^{(r)}}$ and $\tilde{\boldsymbol{a}}_2^{(r)^{\text{T}}} = \boldsymbol{a}^{(r)^{\text{H}}} \cdot \boldsymbol{J}_1^{(r)^{\text{H}}} \cdot \boldsymbol{J}_1^{(r)}$.

Expanding the corresponding right-hand side of (B.472) we have for $r = 1$

$$\begin{aligned}
&\boldsymbol{a}^{(\text{nc},1)^{\text{H}}} \tilde{\boldsymbol{J}}_1^{(\text{nc})(r)^{\text{H}}} \left(\tilde{\boldsymbol{J}}_2^{(\text{nc})(r)}/\mathrm{e}^{\mathrm{j}\cdot\mu^{(r)}} - \tilde{\boldsymbol{J}}_1^{(\text{nc})(r)}\right) \cdot \left(\boldsymbol{\Pi}^{\perp}_{\boldsymbol{a}^{(\text{nc})(1)}} \otimes \boldsymbol{\Pi}_{\boldsymbol{a}^{(2)}} + \boldsymbol{\Pi}_{\boldsymbol{a}^{(\text{nc})(1)}} \otimes \boldsymbol{\Pi}^{\perp}_{\boldsymbol{a}^{(2)}}\right) \\
&= \left[\begin{bmatrix} \mathrm{e}^{\mathrm{j}\varphi} \\ \mathrm{e}^{-\mathrm{j}\varphi} \end{bmatrix}^{\text{H}} \otimes \left(\boldsymbol{a}^{(1)^{\text{H}}} \boldsymbol{J}_1^{(1)^{\text{H}}} \left(\boldsymbol{J}_2^{(1)}/\mathrm{e}^{\mathrm{j}\cdot\mu^{(1)}} - \boldsymbol{J}_1^{(1)}\right)\right) \otimes \boldsymbol{a}^{(2)^{\text{H}}}\right] \cdot \left(\boldsymbol{\Pi}^{\perp}_{\boldsymbol{a}^{(\text{nc})(1)}} \otimes \boldsymbol{\Pi}_{\boldsymbol{a}^{(2)}} + \boldsymbol{\Pi}_{\boldsymbol{a}^{(\text{nc})(1)}} \otimes \boldsymbol{\Pi}^{\perp}_{\boldsymbol{a}^{(2)}}\right) \\
&= \left(\left(\begin{bmatrix} \mathrm{e}^{\mathrm{j}\varphi} \\ \mathrm{e}^{-\mathrm{j}\varphi} \end{bmatrix}^{\text{H}} \otimes \boldsymbol{a}^{(1)^{\text{H}}} \boldsymbol{J}_1^{(1)^{\text{H}}} \left(\boldsymbol{J}_2^{(1)}/\mathrm{e}^{\mathrm{j}\cdot\mu^{(1)}} - \boldsymbol{J}_1^{(1)}\right)\right) \cdot \boldsymbol{\Pi}^{\perp}_{\boldsymbol{a}^{(\text{nc})(1)}}\right) \otimes \Big(\underbrace{\boldsymbol{a}^{(2)^{\text{H}}} \cdot \boldsymbol{\Pi}_{\boldsymbol{a}^{(2)}}}_{\boldsymbol{a}^{(2)^{\text{H}}}}\Big)
\end{aligned}$$

$$+\left(\left(\begin{bmatrix} \mathrm{e}^{\mathrm{j}\varphi} \\ \mathrm{e}^{-\mathrm{j}\varphi} \end{bmatrix}^{\mathrm{H}} \otimes \boldsymbol{a}^{(1)^{\mathrm{H}}} \boldsymbol{J}_1^{(1)^{\mathrm{H}}} \left(\boldsymbol{J}_2^{(1)} / \mathrm{e}^{\mathrm{j}\cdot\mu^{(1)}} - \boldsymbol{J}_1^{(1)}\right)\right) \cdot \boldsymbol{\Pi}_{\boldsymbol{a}^{(\mathrm{nc})(1)}}\right) \otimes \Big(\underbrace{\boldsymbol{a}^{(2)^{\mathrm{H}}} \cdot \boldsymbol{\Pi}^{\perp}_{\boldsymbol{a}^{(2)}}}_{\mathbf{0}_{1\times M_2}}\Big) \tag{B.474}$$

where we have used the fact that $\boldsymbol{a}^{(\mathrm{nc},1)} = \begin{bmatrix} \mathrm{e}^{\mathrm{j}\varphi} \\ \mathrm{e}^{-\mathrm{j}\varphi} \end{bmatrix} \otimes \boldsymbol{a}^{(1)} \otimes \boldsymbol{a}^{(2)}$ and $\tilde{\boldsymbol{J}}_\ell^{(1)} = \boldsymbol{I}_2 \otimes \boldsymbol{J}_\ell^{(1)} \otimes \boldsymbol{I}_{M_2}$ for $\ell = 1, 2$. Moreover, $\boldsymbol{a}^{(2)^{\mathrm{H}}} \cdot \boldsymbol{\Pi}_{\boldsymbol{a}^{(2)}} = \boldsymbol{a}^{(2)^{\mathrm{H}}}$ and $\boldsymbol{a}^{(2)^{\mathrm{H}}} \cdot \boldsymbol{\Pi}^{\perp}_{\boldsymbol{a}^{(2)}} = \mathbf{0}_{1\times M_2}$ follow from the fact that $\boldsymbol{\Pi}_{\boldsymbol{a}^{(2)}}$ and $\boldsymbol{\Pi}^{\perp}_{\boldsymbol{a}^{(2)}}$ represent projection matrices onto the vector $\boldsymbol{a}^{(2)}$ and its orthogonal complement space, respectively. Therefore, the right-hand side of (B.472) can be written as

$$\begin{aligned}
&\left(\left(\begin{bmatrix} \mathrm{e}^{\mathrm{j}\varphi} \\ \mathrm{e}^{-\mathrm{j}\varphi} \end{bmatrix}^{\mathrm{H}} \otimes \boldsymbol{a}^{(1)^{\mathrm{H}}} \boldsymbol{J}_1^{(1)^{\mathrm{H}}} \left(\boldsymbol{J}_2^{(1)} / \mathrm{e}^{\mathrm{j}\mu^{(1)}} - \boldsymbol{J}_1^{(1)}\right)\right) \cdot \left(\boldsymbol{I}_{2M_1} - \frac{1}{2M_1} \boldsymbol{a}^{(\mathrm{nc})(1)} \cdot \boldsymbol{a}^{(\mathrm{nc})(1)^{\mathrm{H}}}\right)\right) \otimes \boldsymbol{a}^{(2)^{\mathrm{H}}} \\
&= \left(\begin{bmatrix} \mathrm{e}^{\mathrm{j}\varphi} \\ \mathrm{e}^{-\mathrm{j}\varphi} \end{bmatrix}^{\mathrm{H}} \otimes \boldsymbol{a}^{(1)^{\mathrm{H}}} \boldsymbol{J}_1^{(1)^{\mathrm{H}}} \left(\boldsymbol{J}_2^{(1)} / \mathrm{e}^{\mathrm{j}\mu^{(1)}} - \boldsymbol{J}_1^{(1)}\right)\right) \otimes \boldsymbol{a}^{(2)^{\mathrm{H}}} \\
&- \frac{1}{2M_1} \left(\left(\begin{bmatrix} \mathrm{e}^{\mathrm{j}\varphi} \\ \mathrm{e}^{-\mathrm{j}\varphi} \end{bmatrix}^{\mathrm{H}} \otimes \boldsymbol{a}^{(1)^{\mathrm{H}}} \boldsymbol{J}_1^{(1)^{\mathrm{H}}} \left(\boldsymbol{J}_2^{(1)} / \mathrm{e}^{\mathrm{j}\mu^{(1)}} - \boldsymbol{J}_1^{(1)}\right)\right) \cdot \boldsymbol{a}^{(\mathrm{nc})(1)} \cdot \boldsymbol{a}^{(\mathrm{nc})(1)^{\mathrm{H}}}\right) \otimes \boldsymbol{a}^{(2)^{\mathrm{H}}} \\
&= \left(\begin{bmatrix} \mathrm{e}^{\mathrm{j}\varphi} \\ \mathrm{e}^{-\mathrm{j}\varphi} \end{bmatrix}^{\mathrm{H}} \otimes \tilde{\boldsymbol{a}}^{(1)^{\mathrm{T}}} - \tilde{\boldsymbol{a}}^{(2)^{\mathrm{T}}}\right) \otimes \boldsymbol{a}^{(2)^{\mathrm{H}}} - \frac{1}{2M_1} \Big(2 \cdot \boldsymbol{a}^{(1)^{\mathrm{H}}} \boldsymbol{J}_1^{(1)^{\mathrm{H}}} \underbrace{\left(\boldsymbol{J}_2^{(1)} / \mathrm{e}^{\mathrm{j}\mu^{(1)}} \boldsymbol{a}^{(1)} - \boldsymbol{J}_1^{(1)} \boldsymbol{a}^{(1)}\right)}_{\mathbf{0}_{(M_1-1)\times 1}} \boldsymbol{a}^{(\mathrm{nc})(1)^{\mathrm{H}}}\Big) \otimes \boldsymbol{a}^{(2)^{\mathrm{H}}} \\
&= \begin{bmatrix} \mathrm{e}^{\mathrm{j}\varphi} \\ \mathrm{e}^{-\mathrm{j}\varphi} \end{bmatrix}^{\mathrm{H}} \otimes \left(\tilde{\boldsymbol{a}}^{(1)^{\mathrm{T}}} - \tilde{\boldsymbol{a}}^{(2)^{\mathrm{T}}}\right) \otimes \boldsymbol{a}^{(2)^{\mathrm{H}}}
\end{aligned} \tag{B.475}$$

where $\boldsymbol{J}_2^{(1)} / \mathrm{e}^{\mathrm{j}\cdot\mu^{(1)}} \boldsymbol{a}^{(1)} - \boldsymbol{J}_1^{(1)} \boldsymbol{a}^{(1)} = \mathbf{0}_{(M_1-1)\times 1}$ follows from the fact that $\boldsymbol{a}^{(1)}$ satisfies the shift invariance equation for $r = 1$. Note that (B.475) is similar to the matrix case considered in Appendix B.30.1. Thus, we have shown that the left-hand side and the right-hand side of (B.472) are equal for $r = 1$. The same result can be found for $r = 2$. Consequently, we have shown that for $d = 1$

$$\boldsymbol{r}^{(r)^{\mathrm{T}}} \cdot \boldsymbol{W}_{\mathrm{mat}}^{(\mathrm{nc})} = \boldsymbol{r}^{(r)^{\mathrm{T}}} \cdot \boldsymbol{W}_{\mathrm{ten}}^{(r)}, \quad \text{for } r = 1, 2 \tag{B.476}$$

and hence, the MSE for 2-D NC Standard ESPRIT and 2-D NC Standard Tensor-ESPRIT are indeed equal. □

B.31. Proof of Theorem 6.6.2

The proof of Theorem 6.6.2 in Section 6.6.2.1 follows analogously the steps for the proof of Theorem 4.5.1 in Appendix B.5 for R-D Standard ESPRIT. Therefore, we can start with the MSE expression of R-D NC Standard ESPRIT for circularly symmetric white noise in (6.56) and straightforwardly apply the same steps as in Appendix B.5 to prove the desired result. □

B.32. Proof of Theorem 6.6.3

In this appendix, we provide a sketch of the proof of Theorem 6.6.3 from Section 6.6.2.2. For the sake of brevity, we have left out some of the lengthy derivations, however, the full proof is provided by us in [Gra16]. We start with the general MSE expression for R-D NC Standard Tensor-ESPRIT for circularly symmetric white sensor noise in (6.79). The MSE for the i-th spatial frequency for $i = 1, 2$ in the r-th mode is given by

$$\mathbb{E}\left\{\left(\Delta\mu_i^{(r)}\right)^2\right\} = \frac{\sigma_{\mathrm{n}}^2}{2} \cdot \left(\left\|\boldsymbol{z}_i^{(\mathrm{nc},r)}\right\|_2^2 - \mathrm{Re}\left\{\boldsymbol{z}_i^{(\mathrm{nc},r)^{\mathrm{T}}} \cdot (\boldsymbol{I}_N \otimes \boldsymbol{\Pi}_{2M}) \cdot \boldsymbol{z}_i^{(\mathrm{nc},r)}\right\}\right), \tag{B.477}$$

where $\boldsymbol{z}_i^{(\mathrm{nc},r)} = \boldsymbol{W}_{\mathrm{ten}}^{(\mathrm{nc},r)^{\mathrm{T}}} \cdot \boldsymbol{r}_i^{(\mathrm{nc},r)}$ with

$$\boldsymbol{r}_i^{(\mathrm{nc},r)} = \boldsymbol{q}_i \otimes \left(\left[\left(\tilde{\boldsymbol{J}}_1^{(\mathrm{nc})(r)} \cdot \boldsymbol{U}_{\mathrm{s}}^{(\mathrm{nc},r)}\right)^+ \cdot \left(\tilde{\boldsymbol{J}}_2^{(\mathrm{nc})(r)}/\lambda_i^{(r)} - \tilde{\boldsymbol{J}}_1^{(\mathrm{nc})(r)}\right)\right]^{\mathrm{T}} \cdot \boldsymbol{p}_i\right) \tag{B.478}$$

$$\boldsymbol{W}_{\mathrm{ten}}^{(\mathrm{nc},r)} = \boldsymbol{W}_0^{(\mathrm{nc},r)} + \sum_{q=1}^{R} \boldsymbol{W}_q^{(\mathrm{nc},r)} \cdot \boldsymbol{P}_{M_1,\ldots,M_{r-1},2M_r,M_{r+1},\ldots,M_R,N}^{(q)^{\mathrm{T}}} \cdot \boldsymbol{P}_{M_1,\ldots,M_{r-1},2M_r,M_{r+1},\ldots,M_R,N}^{(R)}. \tag{B.479}$$

Then, inserting $\boldsymbol{W}_{\mathrm{ten}}^{(\mathrm{nc},r)}$ from (B.479) into the vector $\boldsymbol{z}_i^{(\mathrm{nc},r)^{\mathrm{T}}} = \boldsymbol{r}_i^{(\mathrm{nc},r)^{\mathrm{T}}} \cdot \boldsymbol{W}_{\mathrm{ten}}^{(\mathrm{nc},r)}$, we obtain

$$\begin{aligned}\boldsymbol{z}_i^{(\mathrm{nc},r)^{\mathrm{T}}} &= \boldsymbol{r}_i^{(\mathrm{nc},r)^{\mathrm{T}}} \cdot \boldsymbol{W}_0^{(\mathrm{nc},r)} + \boldsymbol{r}_i^{(\mathrm{nc},r)^{\mathrm{T}}} \cdot \sum_{q=1}^{R} \boldsymbol{W}_q^{(\mathrm{nc},r)} \cdot \boldsymbol{P}_{M_1,\ldots,M_{r-1},2M_r,M_{r+1},\ldots,M_R,N}^{(q)^{\mathrm{T}}} \cdot \boldsymbol{P}_{M_1,\ldots,M_{r-1},2M_r,M_{r+1},\ldots,M_R,N}^{(R)} \\ &= \boldsymbol{z}_{i,0}^{(\mathrm{nc},r)^{\mathrm{T}}} + \boldsymbol{z}_{i,R}^{(\mathrm{nc},r)^{\mathrm{T}}}.\end{aligned} \tag{B.480}$$

Similar to the derivation of R-D Unitary Tensor-ESPRIT for two sources in Appendix B.8, it can be shown that

$$(\boldsymbol{I}_N \otimes \boldsymbol{\Pi}_{2M}) \cdot \boldsymbol{z}_{i,0}^{(\mathrm{nc},r)} = -\boldsymbol{z}_{i,0}^{(\mathrm{nc},r)^*} \tag{B.481}$$

$$(\boldsymbol{I}_N \otimes \boldsymbol{\Pi}_{2M}) \cdot \boldsymbol{z}_{i,R}^{(\mathrm{nc},r)} = -\boldsymbol{z}_{i,R}^{(\mathrm{nc},r)^*}. \tag{B.482}$$

Then, since

$$\begin{aligned}(\boldsymbol{I}_N \otimes \boldsymbol{\Pi}_{2M}) \cdot \boldsymbol{z}_i^{(\mathrm{nc},r)} &= (\boldsymbol{I}_N \otimes \boldsymbol{\Pi}_{2M}) \cdot \left(\boldsymbol{z}_{i,0}^{(\mathrm{nc},r)} + \boldsymbol{z}_{i,R}^{(\mathrm{nc},r)}\right) = (\boldsymbol{I}_N \otimes \boldsymbol{\Pi}_{2M}) \cdot \boldsymbol{z}_{i,0}^{(\mathrm{nc},r)} + (\boldsymbol{I}_N \otimes \boldsymbol{\Pi}_{2M}) \cdot \boldsymbol{z}_{i,R}^{(\mathrm{nc},r)} \\ &= -\boldsymbol{z}_{i,0}^{(\mathrm{nc},r)^*} - \boldsymbol{z}_{i,R}^{(\mathrm{nc},r)^*} = -\boldsymbol{z}_i^{(\mathrm{nc},r)^*}, \end{aligned} \tag{B.483}$$

it follows immediately that

$$\boldsymbol{z}_i^{(\mathrm{nc},r)^{\mathrm{T}}} \cdot (\boldsymbol{I}_N \otimes \boldsymbol{\Pi}_{2M}) \cdot \boldsymbol{z}_i^{(\mathrm{nc},r)} = -\left\| \boldsymbol{z}_i^{(\mathrm{nc},r)} \right\|_2^2. \tag{B.484}$$

Therefore, the MSE for R-D NC Standard Tensor-ESPRIT in (B.477) simplifies to

$$\mathrm{MSE}_i^{(\mathrm{nc},r)} = \mathbb{E}\left\{ \left(\Delta\mu_i^{(r)} \right)^2 \right\} = \sigma_{\mathrm{n}}^2 \cdot \left\| \boldsymbol{z}_i^{(\mathrm{nc},r)} \right\|_2^2. \tag{B.485}$$

In order to further simplify (B.485) for $d = 2$, we introduce an equivalent symmetric model $\boldsymbol{\mathcal{A}}_0^{(\mathrm{nc},r)}$ for the array steering tensor $\boldsymbol{\mathcal{A}}^{(\mathrm{nc},r)}$ in (6.25). This is achieved by merging the rotation phase matrix $\boldsymbol{\Psi}$ with the array phase reference $\boldsymbol{\Delta}$ into a matrix $\boldsymbol{\Xi} = \boldsymbol{\Delta} \cdot \boldsymbol{\Psi}$ according to (2.38). Then, we can write

$$\boldsymbol{\mathcal{A}}_0^{(\mathrm{nc},r)} = (\boldsymbol{\mathcal{A}}_{\mathrm{c}} \times_{R+1} \boldsymbol{\Phi}) \llcorner_r (\boldsymbol{\mathcal{A}}_{\mathrm{c}} \times_{R+1} \boldsymbol{\Phi}^*) \tag{B.486}$$

such that the unperturbed measurement tensor $\boldsymbol{\mathcal{X}}_0^{(\mathrm{nc},r)}$ can be expressed as

$$\boldsymbol{\mathcal{X}}_0^{(\mathrm{nc},r)} = \boldsymbol{\mathcal{A}}_0^{(\mathrm{nc},r)} \times_{R+1} \boldsymbol{S}_0^{\mathrm{T}}. \tag{B.487}$$

An alternative formulation for $\boldsymbol{A}_0^{(\mathrm{nc},r)}$ is given by

$$\boldsymbol{\mathcal{A}}_0^{(\mathrm{nc},r)} = \boldsymbol{\mathcal{I}}_{R+1,d} \times_1 \boldsymbol{A}_{\mathrm{c}}^{(1)} \ldots \times_{r-1} \boldsymbol{A}_{\mathrm{c}}^{(r-1)} \times_r \boldsymbol{A}_0^{(r)(\mathrm{nc})} \times_{r+1} \boldsymbol{A}_{\mathrm{c}}^{(r+1)} \ldots \times_R \boldsymbol{A}_{\mathrm{c}}^{(R)}, \tag{B.488}$$

where $\boldsymbol{A}_0^{(r)(\mathrm{nc})} = \left[\boldsymbol{a}_0^{(r)(\mathrm{nc})}(\mu_1^{(r)}) \quad \ldots \quad \boldsymbol{a}_0^{(r)(\mathrm{nc})}(\mu_d^{(r)}) \right] \in \mathbb{C}^{2 \cdot M_r \times d}$ is the augmented r-mode array steering matrix where the array centroid is chosen as phase reference and the array steering matrices in the unaffected modes $q = 1, \ldots, R$, $q \neq r$ are chosen as $\boldsymbol{A}_{\mathrm{c}}^{(q)} = \left[\boldsymbol{a}_{\mathrm{c}}^{(q)}(\mu_1^{(q)}) \quad \ldots \quad \boldsymbol{a}_{\mathrm{c}}^{(q)}(\mu_d^{(q)}) \right] \in \mathbb{C}^{M_q \times d}$. Then, the matrix version of $\boldsymbol{\mathcal{A}}_0^{(\mathrm{nc},r)}$ is obtained via $\boldsymbol{A}_0^{(\mathrm{nc},r)} = \left[\boldsymbol{\mathcal{A}}_0^{(\mathrm{nc},r)} \right]_{R+1}^{\mathrm{T}} \in \mathbb{C}^{2 \cdot M \times d}$ as

$$\boldsymbol{A}_0^{(\mathrm{nc},r)} = \left[\boldsymbol{\mathcal{A}}_0^{(\mathrm{nc},r)} \right]_{R+1}^{\mathrm{T}} = \boldsymbol{A}_{\mathrm{c},1:r-1}^{\diamond} \diamond \begin{bmatrix} \boldsymbol{A}_{\mathrm{c}}^{(r)} \cdot \boldsymbol{\Phi} \\ \boldsymbol{A}_{\mathrm{c}}^{(r)} \cdot \boldsymbol{\Phi}^* \end{bmatrix} \diamond \boldsymbol{A}_{\mathrm{c},r+1:R}^{\diamond}. \tag{B.489}$$

With these modifications, we proceed to determine the norm $\left\| \boldsymbol{z}_i^{(\mathrm{nc},r)} \right\|_2^2$ in (B.485). Again, it can be shown that the terms $\boldsymbol{z}_{i,0}^{(\mathrm{nc},r)}$ and $\boldsymbol{z}_{i,R}^{(\mathrm{nc},r)}$ are orthogonal. Hence, the norm for $\boldsymbol{z}_i^{(\mathrm{nc},r)}$ is given by

the sum of the norms of $\boldsymbol{z}_{i,0}^{(\mathrm{nc},r)}$ and $\boldsymbol{z}_{i,R}^{(\mathrm{nc},r)}$

$$\left\|\boldsymbol{z}_{i}^{(\mathrm{nc},r)}\right\|_2^2 = \left\|\boldsymbol{z}_{i,0}^{(\mathrm{nc},r)}\right\|_2^2 + \left\|\boldsymbol{z}_{i,R}^{(\mathrm{nc},r)}\right\|_2^2. \tag{B.490}$$

Similar to the steps of the derivations for the R-D Tensor-ESPRIT-type algorithms in Appendix B.7 and Appendix B.8, we will devise $\left\|\boldsymbol{z}_{i,0}^{(\mathrm{nc},r)}\right\|_2^2$ by writing $\boldsymbol{z}_{i,0}^{(\mathrm{nc},r)^{\mathrm{T}}} = \boldsymbol{r}_i^{(\mathrm{nc},r)^{\mathrm{T}}} \cdot \boldsymbol{W}_0^{(\mathrm{nc},r)}$ into the Kronecker product

$$\boldsymbol{z}_{i,0}^{(\mathrm{nc},r)^{\mathrm{T}}} = \tilde{\boldsymbol{s}}_{i,0}^{(\mathrm{nc},r)^{\mathrm{T}}} \otimes \tilde{\boldsymbol{a}}_{i,0}^{(\mathrm{nc},r)^{\mathrm{T}}}, \tag{B.491}$$

where

$$\tilde{\boldsymbol{s}}_{i,0}^{(\mathrm{nc},r)^{\mathrm{T}}} = \boldsymbol{e}_i^{\mathrm{T}} \cdot \boldsymbol{S}_0^{+\mathrm{T}} \tag{B.492}$$

$$\tilde{\boldsymbol{a}}_{i,0}^{(\mathrm{nc},r)^{\mathrm{T}}} = \boldsymbol{e}_i^{\mathrm{T}} \cdot \left(\tilde{\boldsymbol{J}}_1^{(\mathrm{nc},r)} \cdot \boldsymbol{A}_0^{(\mathrm{nc},r)}\right)^{+} \cdot \left(\tilde{\boldsymbol{J}}_2^{(\mathrm{nc},r)}/\lambda_i^{(r)} - \tilde{\boldsymbol{J}}_1^{(\mathrm{nc},r)}\right) \cdot \left(\boldsymbol{T}_{1:r-1}^{\otimes} \otimes \boldsymbol{T}_r^{(\mathrm{nc},r)} \otimes \boldsymbol{T}_{r+1:R}^{\otimes}\right) \cdot \boldsymbol{P}_{\boldsymbol{A}_0^{(\mathrm{nc},r)}}^{\perp}. \tag{B.493}$$

To obtain this result, we have used the symmetric data model in (B.487) Consequently, the signal part $\tilde{\boldsymbol{s}}_{i,0}^{(\mathrm{nc},r)}$ only contains the real-valued symbols $\boldsymbol{S}_0$, and the rotation phase $\boldsymbol{\Psi} = \operatorname{diag}\left\{\mathrm{e}^{\mathrm{j}\cdot\varphi_{\mathrm{rot,i}}}\right\}_{i=1}^{d} \in \mathbb{C}^{d\times d}$ has been included in the symmetric augmented array steering matrix $\boldsymbol{A}_0^{(\mathrm{nc},r)}$ together with the array phase $\boldsymbol{\Delta} = \operatorname{diag}\left\{\mathrm{e}^{\mathrm{j}\cdot\varphi_{\mathrm{ref,i}}}\right\}_{i=1}^{d} \in \mathbb{C}^{d\times d}$.

By making use of property (1.7) we obtain the squared norm $\left\|\boldsymbol{z}_{i,0}^{(\mathrm{nc},r)}\right\|_2^2$ as

$$\left\|\boldsymbol{z}_{i,0}^{(\mathrm{nc},r)}\right\|_2^2 = \left\|\tilde{\boldsymbol{s}}_{i,0}^{(\mathrm{nc},r)}\right\|_2^2 \cdot \left\|\tilde{\boldsymbol{a}}_{i,0}^{(\mathrm{nc},r)}\right\|_2^2. \tag{B.494}$$

To compute $\left\|\tilde{\boldsymbol{s}}_{i,0}^{(\mathrm{nc},r)}\right\|_2^2$, we make use of the fact that $\tilde{\boldsymbol{s}}_{i,0}^{(\mathrm{nc},r)}$ is a special case of $\tilde{\boldsymbol{s}}_{i,0}$ from the derivation for R-D Standard Tensor-ESPRIT in Appendix B.7. Since $\tilde{\boldsymbol{s}}_{i,0}^{(\mathrm{nc},r)}$ differs from $\tilde{\boldsymbol{s}}_{i,0}$ only by the real-valued source symbols, it is clear that $\left\|\tilde{\boldsymbol{s}}_{i,0}^{(\mathrm{nc},r)}\right\|_2^2 = \|\tilde{\boldsymbol{s}}_{i,0}\|_2^2$ such that

$$\left\|\tilde{\boldsymbol{s}}_{i,0}^{(\mathrm{nc},r)}\right\|_2^2 = \frac{1}{N} \cdot \frac{\hat{P}_i}{\hat{P}_1 \cdot \hat{P}_2 \cdot \left(1-\hat{\rho}^2\right)} = \|\tilde{\boldsymbol{s}}_{i,0}\|_2^2, \tag{B.495}$$

where $\hat{\rho}$ denotes the empirical correlation of the real-valued source symbols $\boldsymbol{S}_0$. This result corresponds to the observation that the signal part remains unaffected, since the NC preprocessing only affects the array steering vectors and not the source symbols.

Then, we proceed with the term $\left\|\tilde{\boldsymbol{a}}_{i,0}^{(\mathrm{nc},r)}\right\|_2^2$. By assuming maximally overlapping subarrays, we

can simplify $\left\|\tilde{\boldsymbol{a}}_{i,0}^{(\mathrm{nc},r)}\right\|_2^2$ into

$$\left\|\tilde{\boldsymbol{a}}_{i,0}^{(\mathrm{nc},r)}\right\|_2^2 = \frac{b^{(r)} \cdot \cos^2(\Delta\varphi)}{2 \cdot {D_{\mathrm{sel}}^{(\mathrm{nc},r)}}^2 \cdot D_r^{(\mathrm{nc},r)}} \tag{B.496}$$

where the determinants $D_{\mathrm{sel}}^{(\mathrm{nc},r)}$ and $D_r^{(\mathrm{nc},r)}$ can be expressed as

$$D_{\mathrm{sel}}^{(\mathrm{nc},r)} = \left(\frac{M}{M_r}\right)^2 \cdot (M_r - 1)^2 - \cos^2(\Delta\varphi) \cdot \left|\tilde{\alpha}^{(r)}\right|^2 \text{ and } D_r^{(\mathrm{nc},r)} = M_r^2 - \cos^2(\Delta\varphi) \cdot \left|\alpha^{(r)}\right|^2. \tag{B.497}$$

In contrast to the result for $\left\|\tilde{\boldsymbol{a}}_{i,0}^{(r)}\right\|_2^2$ in Appendix B.7 for R-D Standard Tensor-ESPRIT, the term (B.496) depends on $\Delta\varphi = \Delta\varphi_{\mathrm{ref}} + \Delta\varphi_{\mathrm{rot}}$.

In the next step, we move on to the term $\left\|\boldsymbol{z}_{i,R}^{(\mathrm{nc},r)}\right\|_2^2$ and find that ${\boldsymbol{z}_{i,R}^{(\mathrm{nc},r)}}^{\mathrm{T}} = {\boldsymbol{r}_i^{(\mathrm{nc},r)}}^{\mathrm{T}} \cdot \sum_{q=1}^{R} \boldsymbol{W}_q^{(\mathrm{nc},r)} \cdot {\boldsymbol{P}_{M_1,\ldots,2M_r,\ldots,M_R,N}^{(q)}}^{\mathrm{T}} \cdot \boldsymbol{P}_{M_1,\ldots,2M_r,\ldots,M_R,N}^{(R)}$ reduces to

$${\boldsymbol{z}_{i,R}^{(\mathrm{nc},r)}}^{\mathrm{T}} = {\boldsymbol{z}_{i,r}^{(\mathrm{nc},r)}}^{\mathrm{T}} \cdot {\boldsymbol{P}_{M_1,\ldots,2M_r,\ldots,M_R,N}^{(r)}}^{\mathrm{T}} \cdot \boldsymbol{P}_{M_1,\ldots,2M_r,\ldots,M_R,N}^{(R)}, \tag{B.498}$$

where ${\boldsymbol{z}_{i,r}^{(\mathrm{nc},r)}}^{\mathrm{T}} = {\boldsymbol{r}_i^{(\mathrm{nc},r)}}^{\mathrm{T}} \cdot \boldsymbol{W}_r^{(\mathrm{nc},r)}$ can be expressed as the Kronecker product

$${\boldsymbol{z}_{i,r}^{(\mathrm{nc},r)}}^{\mathrm{T}} = {\tilde{\boldsymbol{s}}_{i,r}^{(\mathrm{nc},r)}}^{\mathrm{T}} \otimes {\tilde{\boldsymbol{a}}_{i,r}^{(\mathrm{nc},r)}}^{\mathrm{T}}, \tag{B.499}$$

with

$${\tilde{\boldsymbol{s}}_{i,r}^{(\mathrm{nc},r)}}^{\mathrm{T}} = \boldsymbol{e}_i^{\mathrm{T}} \cdot \left[\boldsymbol{A}_{\mathrm{c},r+1:R}^{\diamond} \diamond \boldsymbol{S}_0^{\mathrm{T}} \diamond \boldsymbol{A}_{\mathrm{c},1:r-1}^{\diamond}\right]^{+} \tag{B.500}$$

$${\tilde{\boldsymbol{a}}_{i,r}^{(\mathrm{nc},r)}}^{\mathrm{T}} = \boldsymbol{e}_i^{\mathrm{T}} \cdot \left(\tilde{\boldsymbol{J}}_1^{(r)} \cdot \boldsymbol{A}_{\mathrm{c}}\right)^{+} \cdot \left(\tilde{\boldsymbol{J}}_2^{(r)} / \lambda_i^{(r)} - \tilde{\boldsymbol{J}}_1^{(r)}\right) \cdot \left(\boldsymbol{a}_{\mathrm{c},i}^{\otimes(1:r-1)} \otimes \boldsymbol{P}_{\boldsymbol{A}_r}^{\perp} \otimes \boldsymbol{a}_{\mathrm{c},i}^{\otimes(r+1:R)}\right). \tag{B.501}$$

Note that the permutation matrices ${\boldsymbol{P}_{M_1,\ldots,2M_r,\ldots,M_R,N}^{(r)}}^{\mathrm{T}}$ and $\boldsymbol{P}_{M_1,\ldots,2M_r,\ldots,M_R,N}^{(R)}$, do not affect the result when taking the norm. Making use of property (1.7), the term $\left\|\boldsymbol{z}_{i,R}^{(\mathrm{nc},r)}\right\|_2^2$ can be computed as

$$\left\|\boldsymbol{z}_{i,R}^{(\mathrm{nc},r)}\right\|_2^2 = \left\|\tilde{\boldsymbol{s}}_{i,r}^{(\mathrm{nc},r)}\right\|_2^2 \cdot \left\|\tilde{\boldsymbol{a}}_{i,r}^{(\mathrm{nc},r)}\right\|_2^2. \tag{B.502}$$

Assuming maximum subarray overlap, we can write $\left\|\tilde{\boldsymbol{s}}_{i,r}^{(\mathrm{nc},r)}\right\|_2^2$ as

$$\left\|\tilde{\boldsymbol{s}}_{i,r}^{(\mathrm{nc},r)}\right\|_2^2 = \frac{\frac{M}{M_r} \cdot \hat{P}_i}{N \cdot \hat{P}_1 \cdot \hat{P}_2 \cdot \left[\left(\frac{M}{M_r}\right)^2 - |\hat{\rho}|^2 \cdot \frac{|\alpha|^2}{|\alpha^{(r)}|^2}\right]}. \tag{B.503}$$

The result for the array part $\tilde{\boldsymbol{a}}_{i,r}^{(\mathrm{nc},r)}$ for maximum overlap is given by

$$\begin{aligned}\left\|\tilde{\boldsymbol{a}}_{i,r}^{(\mathrm{nc},r)}\right\|_2^2 = \frac{1}{2\cdot D_{\mathrm{sel}}^{(\mathrm{nc},r)^2}}\cdot\Bigg[&\left(\frac{M}{M_r}\right)^4\cdot(M_r-1)^2\cdot 2\\ &+\cos^2(\Delta\varphi)\cdot\left\{\left|\frac{\alpha}{\alpha^{(r)}}\right|^4\cdot\left|\alpha_{\mathrm{sel}}^{(r)}\right|^2\cdot\left(2+(M_r-2)\cdot\left|\mathrm{e}^{\mathrm{j}\Delta\mu^{(r)}}-1\right|^2\right)\right.\\ &-4\cdot\left(\frac{M}{M_r}\right)^2\cdot(M_r-1)\cdot\left|\frac{\alpha}{\alpha^{(r)}}\right|^2\cdot\mathrm{Re}\left\{\alpha_{\mathrm{sel},0}^{(r)}\right\}\\ &\left.-\frac{1}{D_r^{(\mathrm{nc},r)}}\cdot M_r\cdot(M_r-1)^2\cdot\left|\mathrm{e}^{\mathrm{j}\Delta\mu^{(r)}}-1\right|^2\cdot\left|\alpha_{\mathrm{sel}}^{(r)}\right|^2\cdot\left[\left(\frac{M}{M_r}\right)^2-\frac{|\alpha|^2}{\left|\alpha^{(r)}\right|^2}\right]^2\right\}\Bigg].\end{aligned}$$

At this point, we are ready to combine the derived expressions into a simplified MSE expression for R-D NC Standard Tensor-ESPRIT for two sources. The expression for $\left\|\boldsymbol{z}_{i,0}^{(\mathrm{nc},r)}\right\|_2^2 = \|\tilde{\boldsymbol{s}}_{i,0}\|_2^2\cdot\left\|\tilde{\boldsymbol{a}}_{i,0}^{(\mathrm{nc},r)}\right\|_2^2$ results in

$$\begin{aligned}&\left\|\boldsymbol{z}_{i,0}^{(\mathrm{nc},r)}\right\|_2^2\\ &=\frac{1}{N}\cdot\frac{\hat{P}_{\bar{i}}}{\hat{P}_1\cdot\hat{P}_2\cdot\left(1-|\hat{\rho}|^2\right)}\cdot\frac{M\cdot(M_r-1)^2\cdot\cos^2(\Delta\varphi)\cdot\left|\mathrm{e}^{\mathrm{j}\Delta\mu^{(r)}}-1\right|^2\cdot\left|\alpha_{\mathrm{sel}}^{(r)}\right|^2\cdot\left[\left(\frac{M}{M_r}\right)^2-\frac{|\alpha|^2}{\left|\alpha^{(r)}\right|^2}\right]}{2\cdot D_{\mathrm{sel}}^{(\mathrm{nc},r)^2}\cdot D_r^{(\mathrm{nc},r)}}\end{aligned}\tag{B.504}$$

The expression for $\left\|\boldsymbol{z}_{i,r}^{(\mathrm{nc},r)}\right\|_2^2 = \left\|\tilde{\boldsymbol{s}}_{i,r}^{(\mathrm{nc},r)}\right\|_2^2\cdot\left\|\tilde{\boldsymbol{a}}_{i,r}^{(\mathrm{nc},r)}\right\|_2^2$ results in

$$\begin{aligned}\left\|\boldsymbol{z}_{i,r}^{(\mathrm{nc},r)}\right\|_2^2 = &\frac{1}{2\cdot D_{\mathrm{sel}}^{(\mathrm{nc},r)^2}}\cdot\frac{\frac{M}{M_r}\cdot\hat{P}_{\bar{i}}}{N\cdot\hat{P}_1\cdot\hat{P}_2\cdot\left[\left(\frac{M}{M_r}\right)^2-|\hat{\rho}|^2\cdot\frac{|\alpha|^2}{\left|\alpha^{(r)}\right|^2}\right]}\\ &\cdot\Bigg[\left(\frac{M}{M_r}\right)^4\cdot(M_r-1)^2\cdot 2+\cos^2(\Delta\varphi)\cdot\left\{\left|\frac{\alpha}{\alpha^{(r)}}\right|^4\cdot\left|\alpha_{\mathrm{sel}}^{(r)}\right|^2\cdot\left(2+(M_r-2)\cdot\left|\mathrm{e}^{\mathrm{j}\Delta\mu^{(r)}}-1\right|^2\right)\right.\\ &-4\cdot\left(\frac{M}{M_r}\right)^2\cdot(M_r-1)\cdot\left|\frac{\alpha}{\alpha^{(r)}}\right|^2\cdot\mathrm{Re}\left\{\alpha_{\mathrm{sel},0}^{(r)}\right\}\\ &\left.\left.-\frac{1}{D_r^{(\mathrm{nc},r)}}\cdot M_r\cdot(M_r-1)^2\cdot\left|\mathrm{e}^{\mathrm{j}\Delta\mu^{(r)}}-1\right|^2\cdot\left|\alpha_{\mathrm{sel}}^{(r)}\right|^2\cdot\left[\left(\frac{M}{M_r}\right)^2-\frac{|\alpha|^2}{\left|\alpha^{(r)}\right|^2}\right]^2\right\}\right]\end{aligned}\tag{B.505}$$

Finally, we compute $\left\|\boldsymbol{z}_i^{(\mathrm{nc},r)}\right\|_2^2 = \left\|\boldsymbol{z}_{i,0}^{(\mathrm{nc},r)}\right\|_2^2 + \left\|\boldsymbol{z}_{i,r}^{(\mathrm{nc},r)}\right\|_2^2$ as

$$\begin{aligned}\left\|\boldsymbol{z}_i^{(\mathrm{nc},r)}\right\|_2^2 = \frac{\hat{P}_i}{2\cdot N\cdot \hat{P}_1\cdot \hat{P}_2\cdot D_{\mathrm{sel}}^{(\mathrm{nc},r)^2}}\cdot\Bigg\{\Bigg[\frac{1}{1-|\hat{\rho}|^2}-\frac{\left(\frac{M}{M_r}\right)^2-\frac{|\alpha|^2}{|\alpha^{(r)}|^2}}{\left(\frac{M}{M_r}\right)^2-|\hat{\rho}|^2\cdot\frac{|\alpha|^2}{|\alpha^{(r)}|^2}}\Bigg]\\ \cdot\frac{1}{D_r^{(\mathrm{nc},r)}}\cdot M\cdot(M_r-1)^2\cdot\cos^2(\Delta\varphi)\cdot\left|\mathrm{e}^{\mathrm{j}\Delta\mu^{(r)}}-1\right|^2\cdot\left|\alpha_{\mathrm{sel}}^{(r)}\right|^2\cdot\Bigg[\left(\frac{M}{M_r}\right)^2-\frac{|\alpha|^2}{|\alpha^{(r)}|^2}\Bigg]\\ +\Bigg[\left(\frac{M}{M_r}\right)^4\cdot(M_r-1)^2\cdot 2+\cos^2(\Delta\varphi)\cdot\Bigg\{\left|\frac{\alpha}{\alpha^{(r)}}\right|^4\cdot\left|\alpha_{\mathrm{sel}}^{(r)}\right|^2\cdot\left(2+(M_r-2)\cdot\left|\mathrm{e}^{\mathrm{j}\Delta\mu^{(r)}}-1\right|^2\right)\\ -4\cdot\left(\frac{M}{M_r}\right)^2\cdot(M_r-1)\cdot\left|\frac{\alpha}{\alpha^{(r)}}\right|^2\cdot\mathrm{Re}\left\{\alpha_{\mathrm{sel},0}^{(r)}\right\}\Bigg\}\Bigg]\cdot\frac{\frac{M}{M_r}}{\left(\frac{M}{M_r}\right)^2-|\hat{\rho}|^2\cdot\frac{|\alpha|^2}{|\alpha^{(r)}|^2}}\Bigg\}.\end{aligned}$$

Eventually, we obtain the MSE of R-D NC Standard Tensor-ESPRIT for the k-th source in the r-th dimension as

$$\mathrm{MSE}_{\mathrm{ten},i}^{(\mathrm{nc},r)} = \frac{\sigma_{\mathrm{n}}^2}{2}\cdot\frac{\hat{P}_{\bar{i}}}{N\cdot\hat{P}_1\cdot\hat{P}_2}\cdot a_{\mathrm{ten}}^{(\mathrm{nc},r)}. \tag{B.506}$$

Thus, the total MSE of R-D NC Standard Tensor-ESPRIT for two sources is given by

$$\mathrm{MSE}_{\mathrm{ten}}^{(\mathrm{nc})} = \sum_{r=1}^{R}\frac{\sigma_{\mathrm{n}}^2}{2}\cdot\frac{\hat{P}_i+\hat{P}_{\bar{i}}}{N\cdot\hat{P}_1\cdot\hat{P}_2}\cdot a_{\mathrm{ten}}^{(\mathrm{nc},r)}, \tag{B.507}$$

where $a_{\mathrm{ten}}^{(\mathrm{nc},r)}$ can be formulated into (6.88). This completes the sketch of the proof. □

B.33. Proof of Theorem 7.4.1

To show this result in Section 7.4.2, we simply use the FBA-processed and spatially smoothed augmented measurement matrix

$$\tilde{\boldsymbol{X}}_{\mathrm{SS}}^{(\mathrm{nc})} = \left[\boldsymbol{X}_{\mathrm{SS}}^{(\mathrm{nc})} \quad \boldsymbol{\Pi}_{2M_{\mathrm{sub}}}\cdot\boldsymbol{X}_{\mathrm{SS}}^{(\mathrm{nc})^*}\cdot\boldsymbol{\Pi}_{NL}\right]\in\mathbb{C}^{2M_{\mathrm{sub}}\times 2NL} \tag{B.508}$$

and compute the Gram matrix $\boldsymbol{G} = \tilde{\boldsymbol{X}}_{\mathrm{SS}}^{(\mathrm{nc})}\cdot\tilde{\boldsymbol{X}}_{\mathrm{SS}}^{(\mathrm{nc})^{\mathrm{H}}}$, which yields

$$\boldsymbol{G} = \boldsymbol{X}_{\mathrm{SS}}^{(\mathrm{nc})}\cdot\boldsymbol{X}_{\mathrm{SS}}^{(\mathrm{nc})^{\mathrm{H}}} + \boldsymbol{\Pi}_{2M_{\mathrm{sub}}}\cdot\boldsymbol{X}_{\mathrm{SS}}^{(\mathrm{nc})^*}\cdot\boldsymbol{X}_{\mathrm{SS}}^{(\mathrm{nc})^{\mathrm{T}}}\cdot\boldsymbol{\Pi}_{2M_{\mathrm{sub}}}. \tag{B.509}$$

Expanding the second term of (B.509) using (7.24), we obtain

$$\begin{aligned}
&\boldsymbol{\Pi}_{2M_{\mathrm{sub}}} \cdot \left(\sum_{\underline{\boldsymbol{\ell}}=\underline{\mathbf{1}}}^{\underline{\boldsymbol{L}}} (\mathrm{nc}) \boldsymbol{J}_{\underline{\boldsymbol{\ell}}} \cdot \boldsymbol{X}^{(\mathrm{nc})^*} \cdot \boldsymbol{X}^{(\mathrm{nc})^{\mathrm{T}}} \cdot \boldsymbol{J}_{\underline{\boldsymbol{\ell}}}^{(\mathrm{nc})^{\mathrm{T}}} \right) \cdot \boldsymbol{\Pi}_{2M_{\mathrm{sub}}} \\
&= \sum_{\underline{\boldsymbol{\ell}}=\underline{\mathbf{1}}}^{\underline{\boldsymbol{L}}} \begin{bmatrix} \boldsymbol{\Pi}_{M_{\mathrm{sub}}} \cdot \boldsymbol{J}_{\underline{\boldsymbol{\ell}}} \cdot \boldsymbol{\Pi}_M \cdot \boldsymbol{X} \cdot \boldsymbol{X}^{\mathrm{H}} \cdot \boldsymbol{\Pi}_M \cdot \boldsymbol{J}_{\underline{\boldsymbol{\ell}}}^{\mathrm{T}} \cdot \boldsymbol{\Pi}_{M_{\mathrm{sub}}} & \boldsymbol{\Pi}_{M_{\mathrm{sub}}} \cdot \boldsymbol{J}_{\underline{\boldsymbol{\ell}}} \cdot \boldsymbol{\Pi}_M \cdot \boldsymbol{X} \cdot \boldsymbol{X}^{\mathrm{T}} \cdot \boldsymbol{J}_{\underline{\boldsymbol{\ell}}}^{\mathrm{T}} \cdot \boldsymbol{\Pi}_{M_{\mathrm{sub}}} \\ \boldsymbol{\Pi}_{M_{\mathrm{sub}}} \cdot \boldsymbol{J}_{\underline{\boldsymbol{\ell}}} \cdot \boldsymbol{X}^* \cdot \boldsymbol{X}^{\mathrm{H}} \cdot \boldsymbol{\Pi}_M \cdot \boldsymbol{J}_{\underline{\boldsymbol{\ell}}}^{\mathrm{T}} \cdot \boldsymbol{\Pi}_{M_{\mathrm{sub}}} & \boldsymbol{\Pi}_{M_{\mathrm{sub}}} \cdot \boldsymbol{J}_{\underline{\boldsymbol{\ell}}} \cdot \boldsymbol{X}^* \cdot \boldsymbol{X}^{\mathrm{T}} \cdot \boldsymbol{J}_{\underline{\boldsymbol{\ell}}}^{\mathrm{T}} \cdot \boldsymbol{\Pi}_{M_{\mathrm{sub}}} \end{bmatrix}.
\end{aligned} \tag{B.510}$$

Next, we observe the symmetries $\boldsymbol{\Pi}_{M_{\mathrm{sub}}} \cdot \boldsymbol{J}_{\underline{\boldsymbol{\ell}}} \cdot \boldsymbol{\Pi}_M = \boldsymbol{J}_{\underline{\boldsymbol{L}}-\underline{\boldsymbol{\ell}}+\underline{\mathbf{1}}}$ and $\boldsymbol{\Pi}_{M_{\mathrm{sub}}} \cdot \boldsymbol{J}_{\underline{\boldsymbol{\ell}}} = \boldsymbol{J}_{\underline{\boldsymbol{L}}-\underline{\boldsymbol{\ell}}+\underline{\mathbf{1}}} \cdot \boldsymbol{\Pi}_M$. Hence, we perform a change of variables to $\underline{\boldsymbol{m}} = \underline{\boldsymbol{L}} - \underline{\boldsymbol{\ell}} + \underline{\mathbf{1}}$, which simplifies (B.510) to

$$\sum_{\underline{\boldsymbol{m}}=\underline{\mathbf{1}}}^{\underline{\boldsymbol{L}}} \begin{bmatrix} \boldsymbol{J}_{\underline{\boldsymbol{m}}} \cdot \boldsymbol{X} \cdot \boldsymbol{X}^{\mathrm{H}} \cdot \boldsymbol{J}_{\underline{\boldsymbol{m}}}^{\mathrm{T}} & \boldsymbol{J}_{\underline{\boldsymbol{m}}} \cdot \boldsymbol{X} \cdot \boldsymbol{X}^{\mathrm{T}} \cdot \boldsymbol{\Pi}_M \cdot \boldsymbol{J}_{\underline{\boldsymbol{m}}}^{\mathrm{T}} \\ \boldsymbol{J}_{\underline{\boldsymbol{m}}} \cdot \boldsymbol{\Pi}_M \cdot \boldsymbol{X}^* \cdot \boldsymbol{X}^{\mathrm{H}} \cdot \boldsymbol{J}_{\underline{\boldsymbol{m}}}^{\mathrm{T}} & \boldsymbol{J}_{\underline{\boldsymbol{m}}} \cdot \boldsymbol{\Pi}_M \cdot \boldsymbol{X}^* \cdot \boldsymbol{X}^{\mathrm{T}} \cdot \boldsymbol{\Pi}_M \boldsymbol{J}_{\underline{\boldsymbol{m}}}^{\mathrm{T}} \end{bmatrix} = \boldsymbol{X}_{\mathrm{SS}}^{(\mathrm{nc})} \cdot \boldsymbol{X}_{\mathrm{SS}}^{(\mathrm{nc})^{\mathrm{H}}}. \tag{B.511}$$

Replacing the second term of (B.509) by (B.511), we have $\boldsymbol{G} = 2 \cdot \boldsymbol{X}_{\mathrm{SS}}^{(\mathrm{nc})} \cdot \boldsymbol{X}_{\mathrm{SS}}^{(\mathrm{nc})^{\mathrm{H}}}$. Thus, the matrix $\boldsymbol{G}$ reduces to the scaled Gram matrix of $\boldsymbol{X}_{\mathrm{SS}}^{(\mathrm{nc})}$, i.e., the column space of $\boldsymbol{X}_{\mathrm{SS}}^{(\mathrm{nc})}$ is the same as the column space of the Gram matrix of $\boldsymbol{X}_{\mathrm{SS}}^{(\mathrm{nc})}$. Consequently, FBA has no effect on the column space of $\boldsymbol{X}_{\mathrm{SS}}^{(\mathrm{nc})}$. This completes the proof. □

B.34. Proof of Theorem 7.5.1

This theorem from Section 7.5 consists of several parts, which we address in separate subsections.

B.34.1. MSE for R-D Standard ESPRIT with Spatial Smoothing

We start the proof by simplifying the MSE expression for R-D Standard ESPRIT with spatial smoothing in (7.32) and for $d = 1$. In the single source case the noise-free spatially smoothed measurement matrix $\boldsymbol{X}_{\mathrm{SS}_0} \in \mathbb{C}^{M_{\mathrm{sub}} \times NL}$ can be written as

$$\begin{aligned}
\boldsymbol{X}_{\mathrm{SS}_0} &= \bar{\boldsymbol{a}}_{\mathrm{SS}}(\boldsymbol{\mu}) \cdot \boldsymbol{\phi}^{\mathrm{T}} \cdot \left(\boldsymbol{I}_L \otimes \bar{\boldsymbol{s}}^{\mathrm{T}} \right) = \bar{\boldsymbol{a}}_{\mathrm{SS}}(\boldsymbol{\mu}) \cdot \boldsymbol{a}_L^{\mathrm{T}} \cdot \left(\boldsymbol{I}_L \otimes \bar{\boldsymbol{s}}^{\mathrm{T}} \right) \\
&= \bar{\boldsymbol{a}}_{\mathrm{SS}}(\boldsymbol{\mu}) \cdot \left(\boldsymbol{a}_L \otimes \bar{\boldsymbol{s}} \right)^{\mathrm{T}} = \bar{\boldsymbol{a}}_{\mathrm{SS}}(\boldsymbol{\mu}) \cdot \bar{\boldsymbol{s}}_L^{\mathrm{T}},
\end{aligned} \tag{B.512}$$

where

$$\bar{\boldsymbol{a}}_{\mathrm{SS}}(\boldsymbol{\mu}) = \bar{\boldsymbol{a}}_1^{(1)}(\mu^{(1)}) \otimes \cdots \otimes \bar{\boldsymbol{a}}_1^{(R)}(\mu^{(R)}) \in \mathbb{C}^{M_{\mathrm{sub}} \times 1} \tag{B.513}$$

is the spatially smoothed array steering vector in all R modes with $\bar{\boldsymbol{a}}_1^{(r)}(\mu^{(r)}) = \boldsymbol{J}_{1_r}^{(M_r)} \cdot \bar{\boldsymbol{a}}^{(r)}(\mu^{(r)}) \in \mathbb{C}^{M_{\mathrm{sub}_r} \times 1}$, $r = 1, \ldots, R$ and

$$\boldsymbol{\phi} = \boldsymbol{a}_L = \boldsymbol{a}_{L_1}^{(1)}(\mu^{(1)}) \otimes \cdots \otimes \boldsymbol{a}_{L_R}^{(R)}(\mu^{(R)}) \in \mathbb{C}^{L \times 1} \tag{B.514}$$

with $\boldsymbol{a}_{L_r}^{(r)}(\mu^{(r)}) = [1, \mathrm{e}^{\mathrm{j}\mu^{(r)}}, \ldots, \mathrm{e}^{\mathrm{j}\mu^{(r)}(L_r-1)}]^{\mathrm{T}} \in \mathbb{C}^{L_r \times 1}$, $\forall\, r$. Moreover, $\bar{\boldsymbol{s}} = \mathrm{e}^{\mathrm{j}\delta} \cdot \boldsymbol{s} \in \mathbb{C}^{N \times 1}$ with $\delta = \sum_{r=1}^{R} \delta^{(r)} \mu^{(r)}$ contains the source symbols with the empirical source power $\hat{P}_{\mathrm{s}} = \|\bar{\boldsymbol{s}}\|_2^2 / N$ and we have $\bar{\boldsymbol{s}}_L^{\mathrm{H}} \cdot \bar{\boldsymbol{s}}_L = NL\hat{P}_{\mathrm{s}}$. In what follows, we drop the dependence of $\bar{\boldsymbol{a}}_{\mathrm{SS}}(\boldsymbol{\mu})$ on $\boldsymbol{\mu}$ for notational convenience. For a ULA of isotropic elements in each of the R modes, $\bar{\boldsymbol{a}}^{(r)}$ is given by (cf. (2.26))

$$\bar{\boldsymbol{a}}^{(r)} = \left[\mathrm{e}^{-\mathrm{j}\frac{M_r-1}{2}\mu_i^{(r)}} \quad \mathrm{e}^{-\mathrm{j}\frac{M_r-3}{2}\mu_i^{(r)}} \quad \ldots \quad \mathrm{e}^{\mathrm{j}\frac{M_r-3}{2}\mu_i^{(r)}} \quad \mathrm{e}^{\mathrm{j}\frac{M_r-1}{2}\mu_i^{(r)}} \right] \tag{B.515}$$

and $\|\bar{\boldsymbol{a}}_{\mathrm{SS}}\|_2^2 = M_{\mathrm{sub}} = M - L + 1$. The selection matrices $\tilde{\boldsymbol{J}}_{\mathrm{SS}_1}^{(r)}$ and $\tilde{\boldsymbol{J}}_{\mathrm{SS}_2}^{(r)}$ are chosen as $\tilde{\boldsymbol{J}}_{\mathrm{SS}_1}^{(r)} = [\boldsymbol{I}_{M_{\mathrm{sub}_r}-1}, \boldsymbol{0}_{(M_{\mathrm{sub}_r}-1)\times 1}]$ and $\tilde{\boldsymbol{J}}_{\mathrm{SS}_2}^{(r)} = [\boldsymbol{0}_{(M_{\mathrm{sub}_r}-1)\times 1}, \boldsymbol{I}_{M_{\mathrm{sub}_r}-1}]$ for maximum overlap, i.e., $M_{\mathrm{sub}_r}^{(\mathrm{sel})} = M_{\mathrm{sub}_r} - 1$. Note that (B.512) is a rank-one matrix and we can directly determine the subspaces from the SVD as

$$\begin{aligned}
\boldsymbol{U}_{\mathrm{SS_s}} &= \boldsymbol{u}_{\mathrm{SS_s}} = \frac{\bar{\boldsymbol{a}}_{\mathrm{SS}}}{\|\bar{\boldsymbol{a}}_{\mathrm{SS}}\|_2} = \frac{\bar{\boldsymbol{a}}_{\mathrm{SS}}}{\sqrt{M_{\mathrm{sub}}}}, \\
\boldsymbol{\Sigma}_{\mathrm{SS_s}} &= \sigma_{\mathrm{SS_s}} = \sqrt{M_{\mathrm{sub}} NL\hat{P}_{\mathrm{s}}}, \\
\boldsymbol{V}_{\mathrm{SS_s}} &= \boldsymbol{v}_{\mathrm{SS_s}} = \frac{\bar{\boldsymbol{s}}_L^*}{\|\bar{\boldsymbol{s}}_L\|_2} = \frac{\bar{\boldsymbol{s}}_L^*}{\sqrt{NL\hat{P}_{\mathrm{s}}}}.
\end{aligned}$$

For the MSE expression in (7.32), we also require $\boldsymbol{P}_{\bar{\boldsymbol{a}}_{\mathrm{SS}}}^{\perp} = \boldsymbol{U}_{\mathrm{SS_n}} \cdot \boldsymbol{U}_{\mathrm{SS_n}}^{\mathrm{H}} = \boldsymbol{I}_{M_{\mathrm{sub}}} - \frac{1}{M_{\mathrm{sub}}} \cdot \bar{\boldsymbol{a}}_{\mathrm{SS}} \cdot \bar{\boldsymbol{a}}_{\mathrm{SS}}^{\mathrm{H}}$, which is the projection matrix onto the noise subspace. Moreover, we have $\boldsymbol{\Phi}^{(r)} = \mathrm{e}^{\mathrm{j}\mu^{(r)}}$ and hence, the eigenvectors are $\boldsymbol{p}_i^{(r)} = \boldsymbol{q}_i^{(r)} = 1$. The SO moments $\boldsymbol{R}_{\mathrm{SS}}$ and $\boldsymbol{C}_{\mathrm{SS}}$ of the noise are given by (7.36) with $\boldsymbol{R}_{\mathrm{nn}} = \sigma_{\mathrm{n}}^2 \cdot \boldsymbol{I}_M$ and $\boldsymbol{C}_{\mathrm{nn}} = \boldsymbol{0}$.

Inserting these expressions into (7.32), we get

$$\mathbb{E}\left\{ (\Delta\mu^{(r)})^2 \right\} = \frac{1}{2} \cdot \boldsymbol{z}_{\mathrm{SS}}^{(r)^{\mathrm{H}}} \cdot \boldsymbol{R}_{\mathrm{SS}}^{\mathrm{T}} \cdot \boldsymbol{z}_{\mathrm{SS}}^{(r)} = \frac{1}{2} \cdot \boldsymbol{z}_{\mathrm{SS}}^{(r)^{\mathrm{T}}} \cdot \boldsymbol{R}_{\mathrm{SS}} \cdot \boldsymbol{z}_{\mathrm{SS}}^{(r)^*} \tag{B.516}$$

with $\boldsymbol{z}_{\mathrm{SS}}^{(r)} = \boldsymbol{W}_{\mathrm{SS}}^{\mathrm{T}} \cdot \boldsymbol{r}_{\mathrm{SS}}^{(r)}$ and

$$\begin{aligned}
\boldsymbol{r}_{\mathrm{SS}}^{(r)} &= \left[\left(\tilde{\boldsymbol{J}}_{\mathrm{SS}_1}^{(r)} \cdot \frac{\bar{\boldsymbol{a}}_{\mathrm{SS}}}{\sqrt{M_{\mathrm{sub}}}} \right)^{+} \cdot \left(\tilde{\boldsymbol{J}}_{\mathrm{SS}_2}^{(r)} / \mathrm{e}^{\mathrm{j}\mu^{(r)}} - \tilde{\boldsymbol{J}}_{\mathrm{SS}_1}^{(r)} \right) \right]^{\mathrm{T}} \in \mathbb{C}^{M_{\mathrm{sub}} \times 1}, \\
\boldsymbol{W}_{\mathrm{SS}} &= \left(\frac{1}{\sqrt{M_{\mathrm{sub}} NL\hat{P}_{\mathrm{s}}}} \cdot \frac{\bar{\boldsymbol{s}}_L^{\mathrm{H}}}{\sqrt{NL\hat{P}_{\mathrm{s}}}} \right) \otimes \boldsymbol{P}_{\bar{\boldsymbol{a}}_{\mathrm{SS}}}^{\perp} \in \mathbb{C}^{M_{\mathrm{sub}} \times M_{\mathrm{sub}} NL}.
\end{aligned}$$

Note that the term $\boldsymbol{z}_{\mathrm{SS}}^{(r)^{\mathrm{T}}}$ can also be written as $\boldsymbol{z}_{\mathrm{SS}}^{(r)^{\mathrm{T}}} = \tilde{\boldsymbol{s}}^{\mathrm{T}} \otimes \tilde{\boldsymbol{a}}^{(r)^{\mathrm{T}}}$, where

$$\tilde{\boldsymbol{s}}^{\mathrm{T}} = \frac{1}{\sqrt{M_{\mathrm{sub}} N L \hat{P}_{\mathrm{s}}}} \cdot \frac{(\boldsymbol{a}_L \otimes \bar{\boldsymbol{s}})^{\mathrm{H}}}{\sqrt{N L \hat{P}_{\mathrm{s}}}}, \tag{B.517}$$

$$\tilde{\boldsymbol{a}}^{(r)^{\mathrm{T}}} = \left(\tilde{\boldsymbol{J}}_{\mathrm{SS}_1}^{(r)} \cdot \frac{\bar{\boldsymbol{a}}_{\mathrm{SS}}}{\sqrt{M_{\mathrm{sub}}}} \right)^{+} \cdot \left(\tilde{\boldsymbol{J}}_{\mathrm{SS}_2}^{(r)} / \mathrm{e}^{\mathrm{j}\mu^{(r)}} - \tilde{\boldsymbol{J}}_{\mathrm{SS}_1}^{(r)} \right) \cdot \boldsymbol{P}_{\bar{\boldsymbol{a}}_{\mathrm{SS}}}^{\perp}. \tag{B.518}$$

Next, we further simplify the expression $\tilde{\boldsymbol{a}}^{(r)^{\mathrm{T}}}$ and expand the pseudo-inverse of $\tilde{\boldsymbol{J}}_{\mathrm{SS}_1}^{(r)} \cdot \bar{\boldsymbol{a}}_{\mathrm{SS}}$ using the relation $\boldsymbol{x}^{+} = \boldsymbol{x}^{\mathrm{H}} / \|\boldsymbol{x}\|_2^2$. As $\tilde{\boldsymbol{J}}_{\mathrm{SS}_1}^{(r)}$ selects $M_{\mathrm{sub}_r} - 1$ out of M_{sub_r} elements in the r-th mode, we have $\left\| \tilde{\boldsymbol{J}}_{\mathrm{SS}_1}^{(r)} \cdot \bar{\boldsymbol{a}}_{\mathrm{SS}} \right\|_2^2 = \frac{M_{\mathrm{sub}}}{M_{\mathrm{sub}_r}} \cdot (M_{\mathrm{sub}_r} - 1)$. Then, taking the shift invariance equation $\tilde{\boldsymbol{J}}_{\mathrm{SS}_2}^{(r)} \cdot \bar{\boldsymbol{a}}_{\mathrm{SS}} / \mathrm{e}^{\mathrm{j}\mu^{(r)}} - \tilde{\boldsymbol{J}}_{\mathrm{SS}_1}^{(r)} \cdot \bar{\boldsymbol{a}}_{\mathrm{SS}} = \boldsymbol{0}$ in the r-th mode into account, we obtain

$$\tilde{\boldsymbol{a}}^{(r)^{\mathrm{T}}} = \frac{\sqrt{M_{\mathrm{sub}}} M_{\mathrm{sub}_r}}{M_{\mathrm{sub}} (M_{\mathrm{sub}_r} - 1)} \cdot \breve{\boldsymbol{a}}^{(r)^{\mathrm{T}}}, \tag{B.519}$$

$$\breve{\boldsymbol{a}}^{(r)^{\mathrm{T}}} = \bar{\boldsymbol{a}}_{\mathrm{SS}}^{\mathrm{H}} \cdot \left(\tilde{\boldsymbol{J}}_{\mathrm{SS}_2}^{(r)^{\mathrm{H}}} \cdot \tilde{\boldsymbol{J}}_{\mathrm{SS}_2}^{(r)} - \tilde{\boldsymbol{J}}_{\mathrm{SS}_1}^{(r)^{\mathrm{H}}} \cdot \tilde{\boldsymbol{J}}_{\mathrm{SS}_1}^{(r)} \right). \tag{B.520}$$

Since the vector $\bar{\boldsymbol{a}}_{\mathrm{SS}}$ and the matrices $\tilde{\boldsymbol{J}}_{\mathrm{SS}_k}^{(r)}$, $k = 1, 2$, contained in $\bar{\boldsymbol{a}}^{(r)^{\mathrm{T}}}$ can be written as $\bar{\boldsymbol{a}}_{\mathrm{SS}} = \bar{\boldsymbol{a}}_1^{(1)} \otimes \cdots \otimes \bar{\boldsymbol{a}}_1^{(R)}$ and $\tilde{\boldsymbol{J}}_{\mathrm{SS}_k}^{(r)} = \boldsymbol{I}_{\prod_{l=1}^{r-1} M_{\mathrm{sub}_l}} \otimes \boldsymbol{J}_{\mathrm{SS}_k}^{(r)} \otimes \boldsymbol{I}_{\prod_{l=r+1}^{R} M_{\mathrm{sub}_l}}$, all the unaffected modes can be factored out of $\breve{\boldsymbol{a}}^{(r)^{\mathrm{T}}}$, yielding

$$\breve{\boldsymbol{a}}^{(r)^{\mathrm{T}}} = \left(\bar{\boldsymbol{a}}_1^{(1)} \otimes \cdots \otimes \bar{\boldsymbol{a}}_1^{(r-1)} \right)^{\mathrm{H}} \otimes \breve{\boldsymbol{a}}_1^{(r)^{\mathrm{T}}} \otimes \left(\bar{\boldsymbol{a}}_1^{(r+1)} \otimes \cdots \otimes \bar{\boldsymbol{a}}_1^{(R)} \right)^{\mathrm{H}}, \tag{B.521}$$

where we have $\breve{\boldsymbol{a}}_1^{(r)^{\mathrm{T}}} = \bar{\boldsymbol{a}}_1^{(r)^{\mathrm{H}}} \cdot \left(\boldsymbol{J}_{\mathrm{SS}_2}^{(r)^{\mathrm{H}}} \cdot \boldsymbol{J}_{\mathrm{SS}_2}^{(r)} - \boldsymbol{J}_{\mathrm{SS}_1}^{(r)^{\mathrm{H}}} \cdot \boldsymbol{J}_{\mathrm{SS}_1}^{(r)} \right)$ with

$$\bar{\boldsymbol{a}}_1^{(r)^{\mathrm{H}}} = \left[\mathrm{e}^{\mathrm{j}\frac{(M_r - 1)}{2}\mu^{(r)}} \quad \cdots \quad \mathrm{e}^{-\mathrm{j}\frac{(M_r - 2L_r - 1)}{2}\mu^{(r)}} \quad \mathrm{e}^{-\mathrm{j}\frac{(M_r - 2L_r + 1)}{2}\mu^{(r)}} \right]. \tag{B.522}$$

Then, it is easy to verify that

$$\breve{\boldsymbol{a}}_1^{(r)^{\mathrm{T}}} = \left[-\mathrm{e}^{\mathrm{j}\frac{(M_r - 1)}{2}\mu^{(r)}}, 0, \ldots, 0, \mathrm{e}^{-\mathrm{j}\frac{(M_r - 2L_r + 1)}{2}\mu^{(r)}} \right].$$

Thus, the MSE expression in (B.516) is given by

$$\mathbb{E}\left\{ (\Delta\mu^{(r)})^2 \right\} = \frac{{k^{(r)}}^2}{2} \cdot \boldsymbol{v}^{(r)^{\mathrm{T}}} \cdot \boldsymbol{R}_{\mathrm{SS}} \cdot \boldsymbol{v}^{(r)^{*}}, \tag{B.523}$$

where we have used $\boldsymbol{z}_{\mathrm{SS}}^{(r)^{\mathrm{T}}} = k^{(r)} \cdot \boldsymbol{v}^{(r)^{\mathrm{T}}}$ with $\boldsymbol{v}^{(r)^{\mathrm{T}}} = \boldsymbol{a}_L^{\mathrm{H}} \otimes \bar{\boldsymbol{s}}^{\mathrm{H}} \otimes \breve{\boldsymbol{a}}^{(r)^{\mathrm{T}}}$ and $k^{(r)} = \frac{1}{N L \hat{P}_{\mathrm{s}}} \cdot \frac{M_{\mathrm{sub}_r}}{M_{\mathrm{sub}} (M_{\mathrm{sub}_r} - 1)}$.

After straightforward calculations, we further simplify (B.523) to obtain

$$\begin{aligned}
\mathbb{E}\left\{(\Delta\mu^{(r)})^2\right\} &= \frac{{k^{(r)}}^2}{2}\cdot\sigma_{\mathrm{n}}^2\cdot\bar{\boldsymbol{s}}^{\mathrm{H}}\cdot\bar{\boldsymbol{s}} \\
&\quad\cdot\sum_{\underline{\boldsymbol{\ell}}=\underline{\mathbf{1}}}^{\underline{\boldsymbol{L}}}\sum_{\underline{\boldsymbol{m}}=\underline{\mathbf{1}}}^{\underline{\boldsymbol{L}}}\left(\left(\prod_{\substack{p=1\\p\neq r}}^{R}\boldsymbol{a}_1^{(p)\mathrm{H}}\cdot\boldsymbol{J}_{\ell_p}^{(M_p)}\cdot{\boldsymbol{J}_{m_p}^{(M_p)}}^{\mathrm{T}}\cdot\boldsymbol{a}_1^{(p)}\right)\cdot\breve{\boldsymbol{a}}_1^{(r)\mathrm{T}}\cdot\boldsymbol{J}_{\ell_r}^{(M_r)}\cdot{\boldsymbol{J}_{m_r}^{(M_r)}}^{\mathrm{T}}\cdot\breve{\boldsymbol{a}}_1^{(r)*}\cdot\mathrm{e}^{\mathrm{j}\sum_{s=1}^{R}\mu^{(s)}(\ell_s-m_s)}\right) \\
&= \frac{{k^{(r)}}^2}{2}\cdot\sigma_{\mathrm{n}}^2\cdot N\hat{P}_{\mathrm{s}}\cdot\left(\prod_{\substack{p=1\\p\neq r}}^{R}\boldsymbol{a}_1^{(p)\mathrm{H}}\cdot\left(\sum_{\ell_p=1}^{L_p}\sum_{m_p=1}^{L_p}\boldsymbol{J}_{\ell_p}^{(M_p)}\cdot{\boldsymbol{J}_{m_p}^{(M_p)}}^{\mathrm{T}}\right)\cdot\boldsymbol{a}_1^{(p)}\right) \\
&\qquad\qquad\cdot\breve{\boldsymbol{a}}_1^{(r)\mathrm{T}}\cdot\left(\sum_{\ell_r=1}^{L_r}\sum_{m_r=1}^{L_r}\boldsymbol{J}_{\ell_r}^{(M_r)}\cdot{\boldsymbol{J}_{m_r}^{(M_r)}}^{\mathrm{T}}\cdot\mathrm{e}^{\mathrm{j}\sum_{s=1}^{R}\mu^{(s)}(\ell_s-m_s)}\right)\cdot\breve{\boldsymbol{a}}_1^{(r)*} \qquad \text{(B.524)} \\
&= \frac{{k^{(r)}}^2}{2}\cdot\sigma_{\mathrm{n}}^2\cdot N\hat{P}_{\mathrm{s}}\cdot\left(\prod_{\substack{p=1\\p\neq r}}^{R}c_p\right)\cdot 2\cdot\min\{L_r, M_r-L_r\} \qquad \text{(B.525)}
\end{aligned}$$

where c_p in (B.525) is defined in (7.50) and it can be shown that the last term in (B.524) evaluates to $2\cdot(L_r-\max\{2\cdot L_r-M_r,0\}) = 2\cdot\min\{L_r, M_r-L_r\}$. Consequently, the MSE of R-D Standard ESPRIT with spatial smoothing is given by

$$\begin{aligned}
\mathbb{E}\left\{(\Delta\mu^{(r)})^2\right\} &= \frac{\sigma_{\mathrm{n}}^2}{N\hat{P}_{\mathrm{s}}}\cdot\frac{M_{\mathrm{sub}_r}^2\cdot\min\{L_r, M_r-L_r\}}{L^2M_{\mathrm{sub}}^2(M_{\mathrm{sub}_r}-1)^2}\cdot\prod_{\substack{p=1\\p\neq r}}^{R}c_p \\
&= \frac{\sigma_{\mathrm{n}}^2}{N\hat{P}_{\mathrm{s}}}\cdot\frac{\min\{L_r, M_r-L_r\}}{(M_r-L_r)^2L_r^2}\cdot\prod_{\substack{p=1\\p\neq r}}^{R}\frac{c_p}{M_{\mathrm{sub}_p}^2L_p^2}, \qquad \text{(B.526)}
\end{aligned}$$

where we have used the fact that $M_{\mathrm{sub}} = M_{\mathrm{sub}_r}\cdot\prod_{\substack{p=1\\p\neq r}}^{R}M_{\mathrm{sub}_p}$ and $L = L_r\cdot\prod_{\substack{p=1\\p\neq r}}^{R}L_p$. Equation (B.526) is the desired result. □

B.34.2. MSE for R-D Unitary ESPRIT with Spatial Smoothing

The second part of the theorem is to show that for a single source, the MSE of R-D Unitary ESPRIT with spatial smoothing in (7.38) is the same as the MSE of R-D Standard ESPRIT with spatial smoothing in (7.32). Firstly, we simplify $\tilde{\boldsymbol{X}}_{\mathrm{SS}_0}$ from (7.37) and find

$$\begin{aligned}
\tilde{\boldsymbol{X}}_{\mathrm{SS}_0} &= \left[\bar{\boldsymbol{a}}_{\mathrm{SS}}\cdot\bar{\boldsymbol{s}}_L^{\mathrm{T}} \quad \boldsymbol{\Pi}_{M_{\mathrm{sub}}}\cdot\bar{\boldsymbol{a}}_{\mathrm{SS}}^{*}\cdot\bar{\boldsymbol{s}}_L^{\mathrm{H}}\cdot\boldsymbol{\Pi}_{NL}\right] \qquad \text{(B.527)} \\
&= \bar{\boldsymbol{a}}_{\mathrm{SS}}\cdot\left[\bar{\boldsymbol{s}}_L^{\mathrm{T}} \quad \mathrm{e}^{\mathrm{j}\sum_{r=1}^{R}(L_r-1)\mu^{(r)}}\cdot\bar{\boldsymbol{s}}_L^{\mathrm{H}}\cdot\boldsymbol{\Pi}_{NL}\right] \\
&= \bar{\boldsymbol{a}}_{\mathrm{SS}}\cdot\bar{\tilde{\boldsymbol{s}}}_L^{\mathrm{T}}, \qquad \text{(B.528)}
\end{aligned}$$

where in (B.527), we have used the fact that $\boldsymbol{\Pi}_{M_{\mathrm{sub}_r}} \cdot \bar{\boldsymbol{a}}_1^{(r)^*}(\mu^{(r)}) = \bar{\boldsymbol{a}}_1^{(r)}(\mu^{(r)}) \cdot \mathrm{e}^{\mathrm{j}(L_r-1)\mu^{(r)}}$ holds for a ULA in the r-th mode. Moreover, we have defined

$$\bar{\bar{\boldsymbol{s}}}_L = \begin{bmatrix} \bar{\boldsymbol{s}}_L \\ \mathrm{e}^{\mathrm{j}\sum_{r=1}^{R}(L_r-1)\mu^{(r)}} \cdot \boldsymbol{\Pi}_{NL} \cdot \bar{\boldsymbol{s}}_L^* \end{bmatrix} = \begin{bmatrix} \boldsymbol{a}_L \otimes \bar{\boldsymbol{s}} \\ \boldsymbol{a}_L \otimes \boldsymbol{\Pi}_N \cdot \bar{\boldsymbol{s}}^* \end{bmatrix}. \tag{B.529}$$

Note that $\|\bar{\bar{\boldsymbol{s}}}_L\|_2^2 = 2NL\hat{P}_{\mathrm{s}}$. The subspaces from the SVD of $\tilde{\boldsymbol{X}}_{\mathrm{SS}_0}$ are obtained as

$$\tilde{\boldsymbol{u}}_{\mathrm{SS_s}} = \frac{\bar{\boldsymbol{a}}_{\mathrm{SS}}}{\sqrt{M_{\mathrm{sub}}}} = \boldsymbol{u}_{\mathrm{SS_s}}, \quad \tilde{\sigma}_{\mathrm{SS_s}} = \sqrt{2M_{\mathrm{sub}}NL\hat{P}_{\mathrm{s}}}, \quad \tilde{\boldsymbol{v}}_{\mathrm{SS_s}} = \frac{\bar{\bar{\boldsymbol{s}}}_L^*}{\sqrt{2NL\hat{P}_{\mathrm{s}}}}.$$

Compared to the previous subsection, it is apparent that FBA does not affect the column space $\boldsymbol{u}_{\mathrm{SS_s}}$, such that $\tilde{\boldsymbol{U}}_{\mathrm{SS_n}} = \boldsymbol{U}_{\mathrm{SS_n}}$ and thus $\tilde{\boldsymbol{P}}_{\bar{\boldsymbol{a}}_{\mathrm{SS}}}^{\perp} = \boldsymbol{P}_{\bar{\boldsymbol{a}}_{\mathrm{SS}}}^{\perp}$. However, FBA destroys the circular symmetry of the noise, resulting in an additional term in the MSE expression. Following the derivation for R-D Standard ESPRIT with spatial smoothing, it can be shown that $\tilde{\boldsymbol{z}}_{\mathrm{SS}}^{(r)^{\mathrm{T}}} = \tilde{\boldsymbol{r}}_{\mathrm{SS}}^{(r)^{\mathrm{T}}} \cdot \tilde{\boldsymbol{W}}_{\mathrm{SS}} = \tilde{\bar{\boldsymbol{s}}}^{\mathrm{T}} \otimes \tilde{\boldsymbol{a}}^{(r)^{\mathrm{T}}}$, where

$$\tilde{\bar{\boldsymbol{s}}}^{\mathrm{T}} = \frac{1}{\sqrt{2M_{\mathrm{sub}}NL\hat{P}_{\mathrm{s}}}} \cdot \frac{\bar{\bar{\boldsymbol{s}}}_L^{\mathrm{H}}}{\sqrt{2NL\hat{P}_{\mathrm{s}}}} \tag{B.530}$$

and $\tilde{\boldsymbol{a}}^{(r)^{\mathrm{T}}}$ is given as in (B.518). Thus, the MSE expression for R-D Unitary ESPRIT with spatial smoothing in (7.38) can be written as

$$\mathbb{E}\left\{(\Delta\mu^{(r)})^2\right\} = \frac{1}{2} \cdot \left(\tilde{\boldsymbol{z}}_{\mathrm{SS}}^{(r)^{\mathrm{T}}} \cdot \tilde{\boldsymbol{R}}_{\mathrm{SS}} \cdot \tilde{\boldsymbol{z}}_{\mathrm{SS}}^{(r)^*} - \mathrm{Re}\left\{\tilde{\boldsymbol{z}}_{\mathrm{SS}}^{(r)^{\mathrm{T}}} \cdot \tilde{\boldsymbol{C}}_{\mathrm{SS}} \cdot \tilde{\boldsymbol{z}}_{\mathrm{SS}}^{(r)}\right\}\right), \tag{B.531}$$

where $\tilde{\boldsymbol{R}}_{\mathrm{SS}} = \boldsymbol{I}_2 \otimes \boldsymbol{R}_{\mathrm{SS}}$ and $\tilde{\boldsymbol{C}}_{\mathrm{SS}} = \boldsymbol{\Pi}_2 \otimes (\boldsymbol{\Pi}_{M_{\mathrm{sub}}NL} \cdot \boldsymbol{R}_{\mathrm{SS}})$. Expanding (B.531), we have

$$\begin{aligned}\mathbb{E}\left\{(\Delta\mu^{(r)})^2\right\} = \frac{\tilde{k}^{(r)^2}}{2} \cdot \Big(&\boldsymbol{v}^{(r)^{\mathrm{T}}} \cdot \boldsymbol{R}_{\mathrm{SS}} \cdot \boldsymbol{v}^{(r)^*} + \bar{\boldsymbol{v}}^{(r)^{\mathrm{T}}} \cdot \boldsymbol{R}_{\mathrm{SS}} \cdot \bar{\boldsymbol{v}}^{(r)^*} \\ &- \mathrm{Re}\left\{\boldsymbol{v}^{(r)^{\mathrm{T}}} \cdot \boldsymbol{\Pi}_{M_{\mathrm{sub}}NL} \cdot \boldsymbol{R}_{\mathrm{SS}} \cdot \bar{\boldsymbol{v}}^{(r)} + \bar{\boldsymbol{v}}^{(r)^{\mathrm{T}}} \cdot \boldsymbol{\Pi}_{M_{\mathrm{sub}}NL} \cdot \boldsymbol{R}_{\mathrm{SS}} \cdot \boldsymbol{v}^{(r)^*}\right\}\Big),\end{aligned} \tag{B.532}$$

where $\tilde{\boldsymbol{z}}_{\mathrm{SS}}^{(r)^{\mathrm{T}}} = \tilde{k}^{(r)} \cdot \tilde{\boldsymbol{v}}^{(r)^{\mathrm{T}}}$ with $\tilde{\boldsymbol{v}}^{(r)^{\mathrm{T}}} = [\boldsymbol{v}^{(r)^{\mathrm{T}}}, \bar{\boldsymbol{v}}^{(r)^{\mathrm{T}}}]$, $\bar{\boldsymbol{v}}^{(r)^{\mathrm{T}}} = \boldsymbol{a}_L^{\mathrm{H}} \otimes \bar{\boldsymbol{s}}^{\mathrm{T}} \cdot \boldsymbol{\Pi}_N \otimes \breve{\boldsymbol{a}}^{(r)^{\mathrm{T}}}$, and $\tilde{k}^{(r)} = \frac{1}{2NL\hat{P}_{\mathrm{s}}} \cdot \frac{M_{\mathrm{sub}_r}}{M_{\mathrm{sub}}(M_{\mathrm{sub}_r}-1)}$. Note that the first term of (B.532) was already computed in (B.525) as $2 \cdot \sigma_{\mathrm{n}}^2 \cdot N\hat{P}_{\mathrm{s}} \cdot \left(\prod_{\substack{p=1 \\ p\neq r}}^{R} c_p\right) \cdot \min\{L_r, M_r - L_r\}$. The remaining terms can be computed accordingly, where for the second term, we also obtain $2 \cdot \sigma_{\mathrm{n}}^2 \cdot N\hat{P}_{\mathrm{s}} \cdot \left(\prod_{\substack{p=1 \\ p\neq r}}^{R} c_p\right) \cdot \min\{L_r, M_r - L_r\}$ while the third and fourth terms both evaluate to $-2 \cdot \sigma_{\mathrm{n}}^2 \cdot N\hat{P}_{\mathrm{s}} \cdot \left(\prod_{\substack{p=1 \\ p\neq r}}^{R} c_p\right) \cdot \min\{L_r, M_r - L_r\}$. Inserting these intermediate

results into (B.532), we obtain for the MSE of R-D Unitary ESPRIT with spatial smoothing

$$\mathbb{E}\left\{(\Delta\mu^{(r)})^2\right\} = \frac{\sigma_{\mathrm{n}}^2}{N\hat{P}_{\mathrm{s}}} \cdot \frac{\min\{L_r, M_r - L_r\}}{(M_r - L_r)^2 L_r^2} \cdot \prod_{\substack{p=1 \\ p\neq r}}^{R} \frac{c_p}{M_{\mathrm{sub}_p}^2 L_p^2}, \tag{B.533}$$

which is equal to (B.533) and hence proves this part.

B.34.3. MSE for R-D NC Standard ESPRIT and R-D NC Unitary ESPRIT with Spatial Smoothing

The third part of the theorem is to show that the MSE of the spatially smoothed versions of R-D NC Standard ESPRIT and R-D NC Unitary ESPRIT is the same as the MSE for R-D Standard ESPRIT and Unitary ESPRIT. As we have already proven that the performance of R-D NC Standard and R-D NC Unitary ESPRIT with spatial smoothing is identical in the high effective SNR in general, this must also hold true for the case $d = 1$. Hence, it is sufficient to simplify the MSE of R-D NC Standard ESPRIT in (7.41) for this special case.

We start by writing $\boldsymbol{X}_{\mathrm{SS}_0}^{(\mathrm{nc})}$ in (7.25) as

$$\boldsymbol{X}_{\mathrm{SS}_0} = \bar{\boldsymbol{a}}_{\mathrm{SS}}^{(\mathrm{nc})} \cdot \bar{\boldsymbol{s}}_L^{\mathrm{T}}, \tag{B.534}$$

where $\bar{\boldsymbol{s}}_L$ was defined in (B.512) and $\bar{\boldsymbol{a}}_{\mathrm{SS}}^{(\mathrm{nc})} = [1, \tilde{\Psi}]^{\mathrm{T}} \otimes \bar{\boldsymbol{a}}_{\mathrm{SS}}$ with $\tilde{\Psi} = \Xi^* \Xi^* = \mathrm{e}^{-\mathrm{j}2(\varphi+\delta)}$. This follows from (7.25) and the fact that $\bar{\boldsymbol{a}}^{(\mathrm{nc})} = [1, \tilde{\Psi}]^{\mathrm{T}} \otimes \bar{\boldsymbol{a}}$ for a uniform R-D array whose phase reference is at the centroid, i.e. $\boldsymbol{\Pi}_M \cdot \bar{\boldsymbol{a}}^* = \bar{\boldsymbol{a}}$ holds. Therefore, we have $\left\|\bar{\boldsymbol{a}}_{\mathrm{SS}}^{(\mathrm{nc})}\right\|_2^2 = 2M_{\mathrm{sub}}$. The selection matrices $\tilde{\boldsymbol{J}}_{\mathrm{SS}_k}^{(\mathrm{nc})(r)}$, $k = 1, 2$ are given by $\tilde{\boldsymbol{J}}_{\mathrm{SS}_k}^{(\mathrm{nc})(r)} = \boldsymbol{I}_2 \otimes \tilde{\boldsymbol{J}}_{\mathrm{SS}_k}^{(r)}$. The SVD of (B.534) can be explicitly expressed as

$$\boldsymbol{u}_{\mathrm{SS_s}}^{(\mathrm{nc})} = \frac{\bar{\boldsymbol{a}}_{\mathrm{SS}}^{(\mathrm{nc})}}{\sqrt{2M_{\mathrm{sub}}}}, \quad \sigma_{\mathrm{SS_s}}^{(\mathrm{nc})} = \sqrt{2M_{\mathrm{sub}} N L \hat{P}_{\mathrm{s}}}, \quad \boldsymbol{v}_{\mathrm{SS_s}}^{(\mathrm{nc})} = \frac{\bar{\boldsymbol{s}}_L^*}{\sqrt{NL\hat{P}_{\mathrm{s}}}} = \boldsymbol{v}_{\mathrm{SS_s}}.$$

It is evident that the NC preprocessing only affects the column space $\boldsymbol{u}_{\mathrm{SS_s}}^{(\mathrm{nc})}$ while the row space $\boldsymbol{v}_{\mathrm{SS_s}}$ of R-D Standard ESPRIT remains unaffected. Therefore, we have $\boldsymbol{P}_{\bar{\boldsymbol{a}}_{\mathrm{SS}}^{(\mathrm{nc})}}^{\perp} = \boldsymbol{U}_{\mathrm{SS_n}}^{(\mathrm{nc})} \cdot \boldsymbol{U}_{\mathrm{SS_n}}^{(\mathrm{nc})^{\mathrm{H}}} = \boldsymbol{I}_{M_{\mathrm{sub}}} - \frac{1}{M_{\mathrm{sub}}} \cdot \bar{\boldsymbol{a}}_{\mathrm{SS}}^{(\mathrm{nc})} \cdot \bar{\boldsymbol{a}}_{\mathrm{SS}}^{(\mathrm{nc})^{\mathrm{H}}}$. Similarly to FBA, the circular symmetry of the noise is destroyed by the NC preprocessing step. In the NC case, it can be shown that $\boldsymbol{z}_{\mathrm{SS}}^{(\mathrm{nc})(r)^{\mathrm{T}}} = \boldsymbol{r}_{\mathrm{SS}}^{(\mathrm{nc})(r)^{\mathrm{T}}} \cdot \boldsymbol{W}_{\mathrm{SS}}^{(\mathrm{nc})} = \tilde{\boldsymbol{s}}^{(\mathrm{nc})^{\mathrm{T}}} \otimes \tilde{\boldsymbol{a}}^{(\mathrm{nc})(r)^{\mathrm{T}}}$, where

$$\tilde{\boldsymbol{s}}^{(\mathrm{nc})^{\mathrm{T}}} = \frac{1}{\sqrt{2M_{\mathrm{sub}} N L \hat{P}_{\mathrm{s}}}} \cdot \frac{(\boldsymbol{a}_L \otimes \bar{\boldsymbol{s}})^{\mathrm{H}}}{\sqrt{NL\hat{P}_{\mathrm{s}}}} \tag{B.535}$$

$$\tilde{\boldsymbol{a}}^{(\mathrm{nc})(r)^{\mathrm{T}}} = \left(\tilde{\boldsymbol{J}}_{\mathrm{SS}_1}^{(\mathrm{nc})(r)} \cdot \frac{\bar{\boldsymbol{a}}_{\mathrm{SS}}^{(\mathrm{nc})}}{\sqrt{2M_{\mathrm{sub}}}} \right)^{+} \cdot \left(\tilde{\boldsymbol{J}}_{\mathrm{SS}_2}^{(\mathrm{nc})(r)} / \mathrm{e}^{\mathrm{j}\mu^{(r)}} - \tilde{\boldsymbol{J}}_{\mathrm{SS}_1}^{(\mathrm{nc})(r)} \right) \cdot \boldsymbol{P}_{\bar{\boldsymbol{a}}_{\mathrm{SS}}^{(\mathrm{nc})}}^{\perp}. \tag{B.536}$$

Following the lines of the derivation of R-D Standard ESPRIT with spatial smoothing, $\tilde{\boldsymbol{a}}^{(\mathrm{nc})(r)^{\mathrm{T}}}$ can be simplified as

$$\tilde{\boldsymbol{a}}^{(\mathrm{nc})(r)^{\mathrm{T}}} = \frac{\sqrt{2M_{\mathrm{sub}}} M_{\mathrm{sub}_r}}{2M_{\mathrm{sub}}(M_{\mathrm{sub}_r} - 1)} \cdot \begin{bmatrix} 1 \\ \tilde{\Psi} \end{bmatrix}^{\mathrm{H}} \otimes \breve{\boldsymbol{a}}^{(r)^{\mathrm{T}}}, \tag{B.537}$$

where $\bar{\boldsymbol{a}}^{(r)^{\mathrm{T}}}$ is given in (B.521). Consequently, the MSE for R-D NC Standard ESPRIT with spatial smoothing in (7.41) can be written as

$$\mathbb{E}\left\{ (\Delta\mu^{(r)})^2 \right\} = \frac{1}{2} \cdot \left(\boldsymbol{z}_{\mathrm{SS}}^{(\mathrm{nc})(r)^{\mathrm{T}}} \cdot \boldsymbol{R}_{\mathrm{SS}}^{(\mathrm{nc})} \cdot \boldsymbol{z}_{\mathrm{SS}}^{(\mathrm{nc})(r)^{*}} - \mathrm{Re}\left\{ \boldsymbol{z}_{\mathrm{SS}}^{(\mathrm{nc})(r)^{\mathrm{T}}} \cdot \boldsymbol{C}_{\mathrm{SS}}^{(\mathrm{nc})} \cdot \boldsymbol{z}_{\mathrm{SS}}^{(\mathrm{nc})(r)} \right\} \right), \tag{B.538}$$

where $\boldsymbol{R}_{\mathrm{SS}}^{(\mathrm{nc})}$ and $\boldsymbol{C}_{\mathrm{SS}}^{(\mathrm{nc})}$ are given according to (7.46). Next, we use (B.535) and (B.537) to express (B.538) as

$$\mathbb{E}\left\{ (\Delta\mu^{(r)})^2 \right\} = \frac{{k^{(\mathrm{nc})}}^2}{2} \cdot \left(\boldsymbol{v}^{(\mathrm{nc})^{\mathrm{T}}} \cdot \boldsymbol{R}_{\mathrm{SS}}^{(\mathrm{nc})} \cdot \boldsymbol{v}^{(\mathrm{nc})^{*}} - \mathrm{Re}\left\{ \boldsymbol{v}^{(\mathrm{nc})^{\mathrm{T}}} \cdot \boldsymbol{C}_{\mathrm{SS}}^{(\mathrm{nc})} \cdot \boldsymbol{v}^{(\mathrm{nc})} \right\} \right), \tag{B.539}$$

where again $\boldsymbol{z}_{\mathrm{SS}}^{(\mathrm{nc})(r)^{\mathrm{T}}} = k^{(\mathrm{nc})(r)} \cdot \boldsymbol{v}^{(\mathrm{nc})(r)^{\mathrm{T}}}$ with $\boldsymbol{v}^{(\mathrm{nc})(r)^{\mathrm{T}}} = \boldsymbol{a}_L^{\mathrm{H}} \otimes \bar{\boldsymbol{s}}^{\mathrm{H}} \otimes \begin{bmatrix} 1 \\ \tilde{\Psi} \end{bmatrix}^{\mathrm{H}} \otimes \breve{\boldsymbol{a}}^{(r)^{\mathrm{T}}}$ and $k^{(\mathrm{nc})(r)} = \frac{1}{2NL\hat{P}_{\mathrm{s}}} \cdot \frac{M_{\mathrm{sub}_r}}{M_{\mathrm{sub}}(M_{\mathrm{sub}_r}-1)}$. Considering the first term of (B.539) and expanding $\boldsymbol{R}_{\mathrm{SS}}^{(\mathrm{nc})}$, we apply the same steps as in (B.525). As a result, the first term reduces to $4 \cdot \sigma_{\mathrm{n}}^2 \cdot N\hat{P}_{\mathrm{s}} \cdot \left(\prod_{\substack{p=1 \\ p \neq r}}^{R} c_p \right) \cdot \min\{L_r, M_r - L_r\}$. The second term of (B.539) can be computed accordingly to obtain $-4 \cdot \sigma_{\mathrm{n}}^2 \cdot N\hat{P}_{\mathrm{s}} \cdot \left(\prod_{\substack{p=1 \\ p \neq r}}^{R} c_p \right) \cdot \min\{L_r, M_r - L_r\}$.

Using these expressions in (B.539), the MSE of R-D NC Standard ESPRIT with spatial smoothing is given by

$$\mathbb{E}\left\{ (\Delta\mu^{(r)})^2 \right\} = \frac{\sigma_{\mathrm{n}}^2}{N\hat{P}_{\mathrm{s}}} \cdot \frac{\min\{L_r, M_r - L_r\}}{(M_r - L_r)^2 L_r^2} \cdot \prod_{\substack{p=1 \\ p \neq r}}^{R} \frac{c_p}{M_{\mathrm{sub}_p}^2 L_p^2}. \tag{B.540}$$

As this result is equal to (B.526) and (B.533), the theorem has been proven. □

B.35. Proof of Equation (7.52)

For the proof of Equation (7.52) in Section 7.5.2, we consider the case $L_r \leq \frac{M_r}{2}$, however, the derivation for $L_r > \frac{M_r}{2}$ follows the same steps. The MSE in (B.541) is given by

$$\mathrm{MSE}_{\mathrm{SS}}^{(r)} \approx \frac{1}{\hat{\rho}} \cdot \frac{a}{(M_r - L_r)^2 L_r} \quad \text{for} \quad L_r \leq \frac{M_r}{2}, \tag{B.541}$$

where we have defined $a = \prod_{\substack{p=1 \\ p \neq r}}^{R} \frac{c_p}{M_{\mathrm{sub}_p}^2 L_p^2}$. In order to determine the optimal number of subarrays L_r in the r-th mode, we minimize (B.541) with respect to L_r. That is, we first compute the derivative of (B.541) with respect to L_r given by

$$\frac{\partial \mathrm{MSE}_{\mathrm{SS}}^{(r)}}{\partial L_r} = \frac{1}{\hat{\rho}} \cdot \frac{a(M_r - 3L_r)}{(L_r - M_r)^3 L_r^2} \tag{B.542}$$

and then equate (B.542) to zero and solve for L_r, yielding

$$L_r^{\mathrm{opt}} = \frac{1}{3} \cdot M_r, \tag{B.543}$$

which is the desired result in (7.52). □

B.36. Proof of Theorem 9.2.1

For convenience, we start the proof of Theorem 9.2.1 in Section 9.2.2 by vectorizing the R-D NC data model in (9.5) by using property (1.14). We obtain

$$\boldsymbol{x} = \mathrm{vec}\{\boldsymbol{X}\} = (\boldsymbol{I}_N \otimes \boldsymbol{A} \cdot \boldsymbol{\Psi}) \cdot \boldsymbol{s}_0 + \boldsymbol{n} \in \mathbb{C}^{MN \times 1}, \tag{B.544}$$

where $\boldsymbol{s}_0 = \mathrm{vec}\{\boldsymbol{S}_0\} = [\boldsymbol{s}_0^{\mathrm{T}}(1), \ldots, \boldsymbol{s}_0^{\mathrm{T}}(N)]^{\mathrm{T}} \in \mathbb{R}^{Nd \times 1}$ with $\boldsymbol{s}_0(t),\ t = 1, \ldots, N$, being the t-th column of $\boldsymbol{S}_0$, and $\boldsymbol{n} = \mathrm{vec}\{\boldsymbol{N}\} \in \mathbb{C}^{MN \times 1}$. To suit the deterministic data assumption, the signal vector $\boldsymbol{s}_0$ is assumed to be deterministic and unknown to the receiver, while the sensor noise $\boldsymbol{n}$ is zero-mean circularly symmetric white complex Gaussian distributed, i.e., $\mathbb{E}\{\boldsymbol{n} \cdot \boldsymbol{n}^{\mathrm{T}}\} = \boldsymbol{0}$. Hence, the observations $\boldsymbol{x}$ satisfy the model

$$\boldsymbol{x} \sim \mathcal{CN}(\boldsymbol{\nu}, \boldsymbol{\Sigma}), \tag{B.545}$$

where $\boldsymbol{\nu} = (\boldsymbol{I}_N \otimes \boldsymbol{A} \cdot \boldsymbol{\Psi}) \cdot \boldsymbol{s}_0$ and $\boldsymbol{\Sigma} = \sigma_{\mathrm{n}}^2 \cdot \boldsymbol{I}_{MN}$ are the mean and the covariance of the array output vector $\boldsymbol{x}$.

Let us now define the real-valued vector of unknown parameters as

$$\boldsymbol{\xi} = \begin{bmatrix} \boldsymbol{\mu}^{\mathrm{T}} & \boldsymbol{s}_0^{\mathrm{T}} & \boldsymbol{\varphi}^{\mathrm{T}} & \sigma_{\mathrm{n}}^2 \end{bmatrix}^{\mathrm{T}} \in \mathbb{R}^{[(R+N+1)d+1] \times 1}. \tag{B.546}$$

Here, $\boldsymbol{\mu} = [\boldsymbol{\mu}^{(1)^{\mathrm{T}}}, \ldots, \boldsymbol{\mu}^{(R)^{\mathrm{T}}}]^{\mathrm{T}} \in \mathbb{R}^{Rd\times 1}$ is the principal parameter vector of interest and $\boldsymbol{s}_0 \in \mathbb{R}^{Nd\times 1}$, $\boldsymbol{\varphi} \in \mathbb{R}^{d\times 1}$, and σ_{n}^2 are the nuisance parameters. As the CRB matrix is usually computed by taking the inverse of the Fisher information matrix (FIM) $\boldsymbol{J}$, we first need to calculate $\boldsymbol{J}$. Due to (B.545), i.e., $\boldsymbol{x}$ is Gaussian distributed, the Slepian-Bangs formulation [SM05] of the FIM is still valid for the strictly non-circular data model in (B.544). Hence, the Slepian-Bangs formulation of $\boldsymbol{J}$ for the parameter vector $\boldsymbol{\xi}$ is given by [SM05]

$$\boldsymbol{J}_{p,q} = \mathrm{Tr}\left\{\boldsymbol{\Sigma}^{-1}\cdot\frac{\partial\boldsymbol{\Sigma}}{\partial\boldsymbol{\xi}_p}\cdot\boldsymbol{\Sigma}^{-1}\cdot\frac{\partial\boldsymbol{\Sigma}}{\partial\boldsymbol{\xi}_q}\right\} + 2\cdot\mathrm{Re}\left\{\left(\frac{\partial\boldsymbol{\nu}}{\partial\boldsymbol{\xi}_p}\right)^{\mathrm{H}}\cdot\boldsymbol{\Sigma}^{-1}\cdot\frac{\partial\boldsymbol{\nu}}{\partial\boldsymbol{\xi}_q}\right\}, \tag{B.547}$$

for $p,q = 1,\ldots,(R+N+1)d+1$. Note that we are only interested in the CRB for $\boldsymbol{\mu}$, denoted as $\boldsymbol{C}^{(\mathrm{nc})}$. Therefore, it is sufficient to compute the upper left block of $\boldsymbol{J}^{-1}$. In order to find $\boldsymbol{J}$ from (B.547), the partial derivatives of $\boldsymbol{\nu}$ with respect to the parameters of $\boldsymbol{\xi}$ can be calculated straightforwardly. We have

$$\frac{\partial\boldsymbol{\nu}}{\partial\boldsymbol{\mu}^{\mathrm{T}}} = (\boldsymbol{I}_N\otimes(\boldsymbol{D}\cdot(\boldsymbol{I}_R\otimes\boldsymbol{\Psi})))\cdot\tilde{\boldsymbol{S}}_0^{(R)} \in \mathbb{C}^{MN\times Rd}, \tag{B.548}$$

where $\boldsymbol{D}$ is given in (9.8) and $\tilde{\boldsymbol{S}}_0^{(R)} = [(\boldsymbol{I}_R\otimes\tilde{\boldsymbol{S}}_0(1)),\ldots,(\boldsymbol{I}_R\otimes\tilde{\boldsymbol{S}}_0(N))]^{\mathrm{T}} \in \mathbb{R}^{NRd\times Rd}$ with $\tilde{\boldsymbol{S}}_0(t) = \mathrm{diag}\{\boldsymbol{s}_0(t)\} \in \mathbb{R}^{d\times d}$. For the remaining parameters, we get

$$\frac{\partial\boldsymbol{\nu}}{\partial\boldsymbol{s}_0^{\mathrm{T}}} = (\boldsymbol{I}_N\otimes\boldsymbol{A}\cdot\boldsymbol{\Psi}) \in \mathbb{C}^{MN\times Nd},\quad \frac{\partial\boldsymbol{\nu}}{\partial\boldsymbol{\varphi}^{\mathrm{T}}} = \mathrm{j}\cdot(\boldsymbol{I}_N\otimes\boldsymbol{A}\cdot\boldsymbol{\Psi})\cdot\tilde{\boldsymbol{S}}_0 \in \mathbb{C}^{MN\times d},\quad \frac{\partial\boldsymbol{\nu}}{\partial\sigma_{\mathrm{n}}^2} = \boldsymbol{0} \in \mathbb{R}^{MN\times 1},$$

where $\tilde{\boldsymbol{S}}_0 = [\tilde{\boldsymbol{S}}_0(1),\ldots,\tilde{\boldsymbol{S}}_0(N)]^{\mathrm{T}} \in \mathbb{R}^{Nd\times d}$. Next, these results are combined to obtain

$$\begin{aligned}\frac{d\boldsymbol{\nu}}{d\boldsymbol{\xi}^{\mathrm{T}}} = \Big[&(\boldsymbol{I}_N\otimes(\boldsymbol{D}\cdot(\boldsymbol{I}_R\otimes\boldsymbol{\Psi})))\cdot\tilde{\boldsymbol{S}}_0^{(R)},\quad(\boldsymbol{I}_N\otimes\boldsymbol{A}\cdot\boldsymbol{\Psi}),\\ &\mathrm{j}\cdot(\boldsymbol{I}_N\otimes\boldsymbol{A}\cdot\boldsymbol{\Psi})\cdot\tilde{\boldsymbol{S}}_0,\quad\boldsymbol{0}\Big] \in \mathbb{C}^{MN\times[(R+N+1)d+1]}.\end{aligned} \tag{B.549}$$

As for the derivative of $\boldsymbol{\Sigma}$ with respect to $\boldsymbol{\xi}$, the only non-zero term is

$$\frac{d\boldsymbol{\Sigma}}{d\sigma_{\mathrm{n}}^2} = \boldsymbol{I}_{MN}, \tag{B.550}$$

such that

$$\frac{d\boldsymbol{\Sigma}}{d\boldsymbol{\xi}^{\mathrm{T}}} = \begin{bmatrix}\boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{I}_{MN}\end{bmatrix} \in \mathbb{C}^{MN\times[(R+N+1)d+MN]}. \tag{B.551}$$

Inserting (B.549) and (B.551) into (B.547) and bearing in mind that we are interested in the $\boldsymbol{\mu}$-block of $\boldsymbol{J}$, only the second term of (B.547) is of concern. Therefore, we only consider the non-zero

block $\tilde{\boldsymbol{J}}$ of $\boldsymbol{J}$, which is given by

$$\tilde{\boldsymbol{J}} = \begin{bmatrix} \boldsymbol{J}_{\boldsymbol{\mu},\boldsymbol{\mu}} & \boldsymbol{J}_{\boldsymbol{\mu},\boldsymbol{s}_0} & \boldsymbol{J}_{\boldsymbol{\mu},\boldsymbol{\varphi}} \\ \boldsymbol{J}_{\boldsymbol{s}_0,\boldsymbol{\mu}} & \boldsymbol{J}_{\boldsymbol{s}_0,\boldsymbol{s}_0} & \boldsymbol{J}_{\boldsymbol{s}_0,\boldsymbol{\varphi}} \\ \boldsymbol{J}_{\boldsymbol{\varphi},\boldsymbol{\mu}} & \boldsymbol{J}_{\boldsymbol{\varphi},\boldsymbol{s}_0} & \boldsymbol{J}_{\boldsymbol{\varphi},\boldsymbol{\varphi}} \end{bmatrix} = \frac{2}{\sigma_\mathrm{n}^2} \cdot \mathrm{Re}\left\{\boldsymbol{G}^\mathrm{H} \cdot \boldsymbol{G}\right\}, \tag{B.552}$$

where

$$\boldsymbol{G} = \left[(\boldsymbol{I}_N \otimes (\boldsymbol{D} \cdot (\boldsymbol{I}_R \otimes \boldsymbol{\Psi}))) \cdot \tilde{\boldsymbol{S}}_0^{(R)}, \quad (\boldsymbol{I}_N \otimes \boldsymbol{A} \cdot \boldsymbol{\Psi}), \quad \mathrm{j} \cdot (\boldsymbol{I}_N \otimes \boldsymbol{A} \cdot \boldsymbol{\Psi}) \cdot \tilde{\boldsymbol{S}}_0\right] \in \mathbb{C}^{MN \times (R+N+1)d}. \tag{B.553}$$

It is easy to see that $\tilde{\boldsymbol{J}} = \tilde{\boldsymbol{J}}^\mathrm{T}$. Consequently, only the block matrices on and above the diagonal of $\tilde{\boldsymbol{J}}$ need to be computed. For the block matrix $\boldsymbol{J}_{\boldsymbol{\mu},\boldsymbol{\mu}}$, we obtain

$$\begin{aligned}
\boldsymbol{J}_{\boldsymbol{\mu},\boldsymbol{\mu}} &= \frac{2}{\sigma_\mathrm{n}^2} \cdot \sum_{t=1}^{N} \mathrm{Re}\left\{(\boldsymbol{I}_R \otimes \tilde{\boldsymbol{S}}_0(t)) \cdot (\boldsymbol{I}_R \otimes \boldsymbol{\Psi}^*) \cdot \boldsymbol{D}^\mathrm{H} \cdot \boldsymbol{D} \cdot (\boldsymbol{I}_R \otimes \boldsymbol{\Psi}) \cdot (\boldsymbol{I}_R \otimes \tilde{\boldsymbol{S}}_0(t))\right\} \\
&= \frac{2}{\sigma_\mathrm{n}^2} \cdot \mathrm{Re}\left\{\left((\boldsymbol{I}_R \otimes \boldsymbol{\Psi}^*) \cdot \boldsymbol{D}^\mathrm{H} \cdot \boldsymbol{D} \cdot (\boldsymbol{I}_R \otimes \boldsymbol{\Psi})\right) \odot \sum_{t=1}^{N} (\mathbf{1}_R \otimes \boldsymbol{s}_0(t)) \cdot (\mathbf{1}_R \otimes \boldsymbol{s}_0(t))^\mathrm{T}\right\} \\
&= \frac{2}{\sigma_\mathrm{n}^2} \cdot \mathrm{Re}\left\{\left((\boldsymbol{I}_R \otimes \boldsymbol{\Psi}^*) \cdot \boldsymbol{D}^\mathrm{H} \cdot \boldsymbol{D} \cdot (\boldsymbol{I}_R \otimes \boldsymbol{\Psi})\right) \odot \left(\mathbf{1}_{R\times R} \otimes \sum_{t=1}^{N} \boldsymbol{s}_0(t) \cdot \boldsymbol{s}_0^\mathrm{T}(t)\right)\right\} \\
&= \frac{2N}{\sigma_\mathrm{n}^2} \cdot \mathrm{Re}\left\{\left((\boldsymbol{I}_R \otimes \boldsymbol{\Psi}^*) \cdot \boldsymbol{D}^\mathrm{H} \cdot \boldsymbol{D} \cdot (\boldsymbol{I}_R \otimes \boldsymbol{\Psi})\right) \odot \left(\mathbf{1}_{R\times R} \otimes \frac{1}{N} \cdot \boldsymbol{S}_0 \cdot \boldsymbol{S}_0^\mathrm{T}\right)\right\} \\
&= \frac{2N}{\sigma_\mathrm{n}^2} \cdot \mathrm{Re}\left\{(\boldsymbol{I}_R \otimes \boldsymbol{\Psi}^*) \cdot \boldsymbol{D}^\mathrm{H} \cdot \boldsymbol{D} \cdot (\boldsymbol{I}_R \otimes \boldsymbol{\Psi})\right\} \odot \hat{\boldsymbol{R}}_{S_0}^{(R)} \\
&= \frac{2N}{\sigma_\mathrm{n}^2} \cdot \boldsymbol{G}_2 \odot \hat{\boldsymbol{R}}_{S_0}^{(R)} \in \mathbb{R}^{Rd \times Rd},
\end{aligned} \tag{B.554}$$

where $\boldsymbol{G}_2$ is defined according to (9.18) and we have used the fact that $\mathrm{diag}\{\boldsymbol{a}\} \cdot \boldsymbol{C} \cdot \mathrm{diag}\{\boldsymbol{b}\} = \boldsymbol{C} \odot (\boldsymbol{a} \cdot \boldsymbol{b}^\mathrm{T})$ for arbitrary vectors $\boldsymbol{a} \in \mathbb{C}^M, \boldsymbol{b} \in \mathbb{C}^N$, and a matrix $\boldsymbol{C} \in \mathbb{C}^{M\times N}$. In a similar manner, the other blocks of $\tilde{\boldsymbol{J}}$ can be computed. The results are given by

$$\boldsymbol{J}_{\boldsymbol{s}_0,\boldsymbol{s}_0} = \frac{2}{\sigma_\mathrm{n}^2} \cdot \boldsymbol{I}_N \otimes \boldsymbol{G}_0 \in \mathbb{R}^{Nd \times Nd} \tag{B.555}$$

$$\boldsymbol{J}_{\boldsymbol{\varphi},\boldsymbol{\varphi}} = \frac{2N}{\sigma_\mathrm{n}^2} \cdot \boldsymbol{G}_0 \odot \hat{\boldsymbol{R}}_{S_0} \in \mathbb{R}^{d \times d} \tag{B.556}$$

$$\boldsymbol{J}_{\boldsymbol{\mu},\boldsymbol{s}_0} = \frac{2}{\sigma_\mathrm{n}^2} \cdot \tilde{\boldsymbol{S}}_0^{(R)^\mathrm{T}} \cdot (\boldsymbol{I}_N \otimes \boldsymbol{G}_1) \in \mathbb{R}^{Rd \times Nd} \tag{B.557}$$

$$\boldsymbol{J}_{\boldsymbol{s}_0,\boldsymbol{\varphi}} = -\frac{2}{\sigma_\mathrm{n}^2} \cdot (\boldsymbol{I}_N \otimes \boldsymbol{H}_0) \cdot \tilde{\boldsymbol{S}}_0 \in \mathbb{R}^{Nd \times d} \tag{B.558}$$

$$\boldsymbol{J}_{\boldsymbol{\mu},\boldsymbol{\varphi}} = -\frac{2N}{\sigma_\mathrm{n}^2} \cdot \boldsymbol{H}_1 \odot (\mathbf{1}_R \otimes \hat{\boldsymbol{R}}_{S_0}) \in \mathbb{R}^{Rd \times d}, \tag{B.559}$$

where the matrices $\boldsymbol{G}_n$ and $\boldsymbol{H}_n$, $n = 0, 1, 2$ are given in (9.14)-(9.17). Note that we have the symmetries $\boldsymbol{G}_0 = \boldsymbol{G}_0^{\mathrm{T}}$, $\boldsymbol{G}_2 = \boldsymbol{G}_2^{\mathrm{T}}$, and $\boldsymbol{H}_0 = -\boldsymbol{H}_0^{\mathrm{T}}$.

In the next step, we need to extract the upper left block of $\tilde{\boldsymbol{J}}^{-1}$. To this end, we make use of the following lemma:

Lemma B.36.1. *For matrices $\boldsymbol{A} \in \mathbb{C}^{p\times p}$, $\boldsymbol{B} \in \mathbb{C}^{p\times q}$, $\boldsymbol{C} \in \mathbb{C}^{p\times r}$, $\boldsymbol{D} \in \mathbb{C}^{q\times p}$, $\boldsymbol{E} \in \mathbb{C}^{q\times q}$, $\boldsymbol{F} \in \mathbb{C}^{q\times r}$, $\boldsymbol{G} \in \mathbb{C}^{r\times p}$, $\boldsymbol{H} \in \mathbb{C}^{r\times q}$, and $\boldsymbol{J} \in \mathbb{C}^{r\times r}$ the upper left $p \times p$ block of the matrix*

$$\boldsymbol{K} = \begin{bmatrix} \boldsymbol{A} & \boldsymbol{B} & \boldsymbol{C} \\ \boldsymbol{D} & \boldsymbol{E} & \boldsymbol{F} \\ \boldsymbol{G} & \boldsymbol{H} & \boldsymbol{J} \end{bmatrix}^{-1} \tag{B.560}$$

is given by

$$\begin{aligned} \boldsymbol{K}_{1:p,1:p} = \Big(& \boldsymbol{A} - \boldsymbol{B}\boldsymbol{E}^{-1}\boldsymbol{D} - \boldsymbol{B}\boldsymbol{E}^{-1}\boldsymbol{F}\boldsymbol{S}_E^{-1}\boldsymbol{H}\boldsymbol{E}^{-1}\boldsymbol{D} + \boldsymbol{B}\boldsymbol{E}^{-1}\boldsymbol{F}\boldsymbol{S}_E^{-1}\boldsymbol{G} + \boldsymbol{C}\boldsymbol{J}^{-1}\boldsymbol{H}\boldsymbol{E}^{-1}\boldsymbol{D} \\ & + \boldsymbol{C}\boldsymbol{J}^{-1}\boldsymbol{H}\boldsymbol{E}^{-1}\boldsymbol{F}\boldsymbol{S}_E^{-1}\boldsymbol{H}\boldsymbol{E}^{-1}\boldsymbol{D} - \boldsymbol{C}\boldsymbol{S}_E^{-1}\boldsymbol{G} \Big)^{-1}, \end{aligned}$$

where $\boldsymbol{S}_E = \boldsymbol{J} - \boldsymbol{H}\boldsymbol{E}^{-1}\boldsymbol{F}$.

The proof of Lemma B.36.1 can easily be constructed by applying the inversion formula for a 2×2 block-partitioned matrix [Lüt96] to the 3×3 block matrix in (B.560) twice.

Applying Lemma B.36.1 to compute the upper left block of $\tilde{\boldsymbol{J}}^{-1}$, it is straightforward to obtain the expression in (9.13), where we have

$$\begin{aligned} \boldsymbol{S}_E &= \frac{2N}{\sigma_{\mathrm{n}}^2} \cdot \left(\boldsymbol{G}_0 - \boldsymbol{H}_0^{\mathrm{T}} \boldsymbol{G}_0^{-1} \boldsymbol{H}_0 \right) \odot \hat{\boldsymbol{R}}_{S_0}^{(R)} \\ \boldsymbol{B}\boldsymbol{E}^{-1}\boldsymbol{D} &= \frac{2N}{\sigma_{\mathrm{n}}^2} \cdot \left(\boldsymbol{G}_1 \boldsymbol{G}_0^{-1} \boldsymbol{G}_1^{\mathrm{T}} \right) \odot \hat{\boldsymbol{R}}_{S_0}^{(R)} \\ \boldsymbol{B}\boldsymbol{E}^{-1}\boldsymbol{F} &= -\frac{2N}{\sigma_{\mathrm{n}}^2} \cdot \left(\boldsymbol{G}_1 \boldsymbol{G}_0^{-1} \boldsymbol{H}_0 \right) \odot \hat{\boldsymbol{R}}_{S_0}^{(R)} \\ \boldsymbol{H}\boldsymbol{E}^{-1}\boldsymbol{D} &= -\frac{2N}{\sigma_{\mathrm{n}}^2} \cdot \left(\boldsymbol{H}_0^{\mathrm{T}} \boldsymbol{G}_0^{-1} \boldsymbol{G}_1^{\mathrm{T}} \right) \odot \hat{\boldsymbol{R}}_{S_0}^{(R)} \\ \boldsymbol{H}\boldsymbol{E}^{-1}\boldsymbol{F} &= \frac{2N}{\sigma_{\mathrm{n}}^2} \cdot \left(\boldsymbol{H}_0^{\mathrm{T}} \boldsymbol{G}_0^{-1} \boldsymbol{H}_0 \right) \odot \hat{\boldsymbol{R}}_{S_0}^{(R)}. \end{aligned}$$

This concludes the proof. □

B.37. Proof of Equation (9.28)

For the proof of Equation (9.28) in Section 9.3.1, we show that for $\delta^{(r)} = 0 \ \forall \ r$ and subsequently $\boldsymbol{A} = \boldsymbol{A}_{\mathrm{c}}$, the matrices $\boldsymbol{A}^{\mathrm{H}} \cdot \boldsymbol{A} \in \mathbb{R}^{d\times d}$, $\boldsymbol{D}^{\mathrm{H}} \cdot \boldsymbol{A} \in \mathbb{R}^{Rd\times d}$, and $\boldsymbol{D}^{\mathrm{H}} \cdot \boldsymbol{D} \in \mathbb{R}^{Rd\times Rd}$ are real-valued. To this

end, we make use of the following lemma:

Lemma B.37.1. *For two arbitrary non-singular left* $\boldsymbol{\Pi}$*-real matrices* $\boldsymbol{X} \in \mathbb{C}^{M\times N}$ *and* $\boldsymbol{Y} \in \mathbb{C}^{M\times N}$ *satisfying* $\boldsymbol{\Pi}\cdot\boldsymbol{X}^* = \boldsymbol{X}$ *and* $\boldsymbol{\Pi}\cdot\boldsymbol{Y}^* = \boldsymbol{Y}$*, respectively, the following identity holds:*

$$\begin{aligned}\boldsymbol{Y}^{\mathrm{H}}\cdot\boldsymbol{X} &= (\boldsymbol{\Pi}\cdot\boldsymbol{Y}^*)^{\mathrm{H}}\cdot\boldsymbol{\Pi}\cdot\boldsymbol{X}^* = \boldsymbol{Y}^{\mathrm{T}}\cdot\boldsymbol{\Pi}\cdot\boldsymbol{\Pi}\cdot\boldsymbol{X}^* \\ &= (\boldsymbol{Y}^{\mathrm{H}}\cdot\boldsymbol{X})^* \in \mathbb{R}^{N\times N}.\end{aligned} \tag{B.561}$$

Therefore, to prove that the aforementioned matrices are real-valued, we simply show that the matrices $\boldsymbol{A}$ and $\boldsymbol{D}$ are left $\boldsymbol{\Pi}$-real. It is straightforward to see that due to $\delta^{(r)} = 0\ \forall\ r$, this is the case for $\boldsymbol{A}$ since $\boldsymbol{A} = \boldsymbol{A}_{\mathrm{c}}$ and $\boldsymbol{A}_{\mathrm{c}}^{(r)} = \boldsymbol{\Pi}_{M_r}\cdot\boldsymbol{A}_{\mathrm{c}}^{(r)^*}$ holds. As for the matrix $\boldsymbol{D}$, we utilize the linearity of the differentiation operator and obtain

$$\boldsymbol{\Pi}\cdot\boldsymbol{D}^* = \boldsymbol{\Pi}\cdot\left(\frac{\partial\boldsymbol{A}}{\partial\boldsymbol{\mu}}\right)^* = \frac{\partial\boldsymbol{\Pi}\cdot\boldsymbol{A}^*}{\partial\boldsymbol{\mu}} = \frac{\partial\boldsymbol{A}}{\partial\boldsymbol{\mu}} = \boldsymbol{D}, \tag{B.562}$$

which also renders $\boldsymbol{D}$ left $\boldsymbol{\Pi}$-real and concludes the proof. □

B.38. Proof of Theorem 9.4.1

In this appendix, we show the proof of Theorem 9.4.1 in Section 9.4.1. Evaluating the R-D NC CRB in (9.13) for the special case $d = 1$, the array steering matrix $\boldsymbol{A}$ reduces to $\boldsymbol{a}(\boldsymbol{\mu})$, $\boldsymbol{D} = [\boldsymbol{d}^{(1)},\ldots,\boldsymbol{d}^{(R)}] \in \mathbb{C}^{M\times R}$, $\boldsymbol{\Psi} = \mathrm{e}^{\mathrm{j}\varphi}$, and $\hat{\boldsymbol{R}}_{S_0} = \boldsymbol{s}_0^{\mathrm{T}}\cdot\boldsymbol{s}_0/N = \hat{P}$, where $\boldsymbol{s}_0 \in \mathbb{R}^{N\times 1}$. Moreover, we choose $\delta^{(r)} = 0\ \forall\ r$ for simplicity. Then, dropping the dependence of $\boldsymbol{a}$ on $\boldsymbol{\mu}$ and utilizing the definitions in (9.2) and (9.9), respectively, we have

$$\boldsymbol{a}^{\mathrm{H}}\cdot\boldsymbol{a} = \prod_{r=1}^{R}\boldsymbol{a}^{(r)^{\mathrm{H}}}\cdot\boldsymbol{a}^{(r)} = \prod_{r=1}^{R}M_r = M, \tag{B.563}$$

$$\boldsymbol{d}^{(r)^{\mathrm{H}}}\cdot\boldsymbol{a} = \prod_{\substack{p=1\\p\neq r}}^{R}\boldsymbol{a}^{(p)^{\mathrm{H}}}\cdot\boldsymbol{a}^{(p)}\cdot\tilde{\boldsymbol{d}}^{(r)^{\mathrm{H}}}\cdot\boldsymbol{a}^{(r)} = \prod_{\substack{p=1\\p\neq r}}^{R}\boldsymbol{a}^{(p)^{\mathrm{H}}}\cdot\boldsymbol{a}^{(p)}\cdot\left(-\mathrm{j}\sum_{m_r=1}^{M_r}k_{m_r}\right) = 0 \quad \forall\ r, \tag{B.564}$$

$$\begin{aligned}\boldsymbol{d}^{(r)^{\mathrm{H}}}\cdot\boldsymbol{d}^{(r)} &= \prod_{\substack{p=1\\p\neq r}}^{R}\boldsymbol{a}^{(p)^{\mathrm{H}}}\cdot\boldsymbol{a}^{(p)}\cdot\tilde{\boldsymbol{d}}^{(r)^{\mathrm{H}}}\cdot\tilde{\boldsymbol{d}}^{(r)} = \prod_{\substack{p=1\\p\neq r}}^{R}\boldsymbol{a}^{(p)^{\mathrm{H}}}\cdot\boldsymbol{a}^{(p)}\cdot\left(\sum_{m_r=1}^{M_r}k_{m_r}^2\right) \\ &= \frac{M}{M_r}\cdot\sum_{m_r=1}^{M_r}k_{m_r}^2 = \Gamma^{(r)} \quad \forall\ r.\end{aligned} \tag{B.565}$$

Using the results in (B.563)-(B.565), the matrices $\boldsymbol{G}_n$ and $\boldsymbol{H}_n$, $n = 0,1,2$, simplify to

$$\boldsymbol{G}_0 = M, \qquad \boldsymbol{G}_1 = \boldsymbol{H}_0 = \boldsymbol{H}_1 = 0, \tag{B.566}$$

$$\boldsymbol{G}_2 = \boldsymbol{D}^{\mathrm{H}} \cdot \boldsymbol{D} = \operatorname{diag}\left\{\left[\Gamma^{(1)}, \ldots, \Gamma^{(R)}\right]\right\}, \tag{B.567}$$

where in $\boldsymbol{G}_2$, the terms $\boldsymbol{d}^{(r_1)^{\mathrm{H}}} \cdot \boldsymbol{d}^{(r_2)}$ for $r_1 \neq r_2$ evaluate to zero due to (B.564). Inserting these expressions into (9.13), the remaining part of the R-D NC CRB matrix is given by

$$\boldsymbol{C}^{(\mathrm{nc})} = \frac{\sigma_{\mathrm{n}}^2}{2N\hat{P}} \cdot \left\{\operatorname{diag}\left\{\left[\Gamma^{(1)}, \ldots, \Gamma^{(R)}\right]\right\}\right\}^{-1} \tag{B.568}$$

$$= \operatorname{diag}\left\{\left[C^{(\mathrm{nc})(1)}, \ldots, C^{(\mathrm{nc})(R)}\right]\right\} \in \mathbb{R}^{R \times R}, \tag{B.569}$$

where

$$C^{(\mathrm{nc})(r)} = \frac{\sigma_{\mathrm{n}}^2}{2N\hat{P}} \cdot \frac{M_r}{M} \cdot \frac{1}{\sum_{m_r=1}^{M_r} k_{m_r}^2} \;\; \forall \; r, \tag{B.570}$$

which is the desired result. □

B.39. Proof of Theorem 9.4.3

Based on the model in (9.24) after inserting (9.40), we start the proof of Theorem 9.4.3 in Section 9.4.2.1 by assuming without loss of generality that the phase reference is at the array centroid, i.e., $\boldsymbol{\Delta} = \boldsymbol{I}_d$ such that $\boldsymbol{A} = \bar{\boldsymbol{A}}$ and $\boldsymbol{\Phi} = \boldsymbol{\Psi}$. Using the results from Appendix B.37, we can write the real-valued matrices $\boldsymbol{A}^{\mathrm{H}} \cdot \boldsymbol{A}$, $\boldsymbol{D}^{\mathrm{H}} \cdot \boldsymbol{A}$, and $\boldsymbol{D}^{\mathrm{H}} \cdot \boldsymbol{D}$ as

$$\boldsymbol{A}^{\mathrm{H}} \cdot \boldsymbol{A} = \begin{bmatrix} M & \alpha \\ \alpha & M \end{bmatrix}, \quad \boldsymbol{D}^{\mathrm{H}} \cdot \boldsymbol{A} = \begin{bmatrix} 0 & \beta \\ -\beta & 0 \end{bmatrix}, \quad \boldsymbol{D}^{\mathrm{H}} \cdot \boldsymbol{D} = \begin{bmatrix} \Gamma & \gamma \\ \gamma & \Gamma \end{bmatrix},$$

where we have defined $\alpha = \boldsymbol{a}_1^{\mathrm{H}} \;\; \boldsymbol{a}_2 = \boldsymbol{a}_2^{\mathrm{H}} \;\; \boldsymbol{a}_1$, $\beta = \boldsymbol{d}_1^{\mathrm{H}} \;\; \boldsymbol{a}_2 = -\boldsymbol{d}_2^{\mathrm{H}} \;\; \boldsymbol{a}_1$, and $\gamma = \boldsymbol{d}_1^{\mathrm{H}} \;\; \boldsymbol{d}_2 = \boldsymbol{d}_2^{\mathrm{H}} \;\; \boldsymbol{d}_1$. Then, the matrices $\boldsymbol{G}_0$ and $\boldsymbol{H}_0$ can be written as

$$\boldsymbol{G}_0 = \operatorname{Re}\left\{\boldsymbol{\Psi}^* \cdot \boldsymbol{A}^{\mathrm{H}} \cdot \boldsymbol{A} \cdot \boldsymbol{\Psi}\right\} = \begin{bmatrix} M & \alpha \cdot \cos(\Delta\varphi) \\ \alpha \cdot \cos(\Delta\varphi) & M \end{bmatrix}$$

$$\boldsymbol{H}_0 = \operatorname{Im}\left\{\boldsymbol{\Psi}^* \cdot \boldsymbol{A}^{\mathrm{H}} \cdot \boldsymbol{A} \cdot \boldsymbol{\Psi}\right\} = \begin{bmatrix} 0 & \alpha \cdot \sin(\Delta\varphi) \\ -\alpha \cdot \sin(\Delta\varphi) & 0 \end{bmatrix}.$$

The matrices $\boldsymbol{G}_1$, $\boldsymbol{H}_1$, and $\boldsymbol{G}_2$ can be expressed in a similar manner. In order to obtain an expression of the 1-D NC CRB that only depends on the physical parameters, e.g, M, ρ, $\Delta\varphi$, etc., we approximate the scalars α, β, and γ using a Taylor series expansion for small source separations $\Delta\mu = |\mu_2 - \mu_1|$. Hence, these approximations become accurate for a small $\Delta\mu$. Therefore, for α, we

have

$$\alpha = \sum_{m=-\frac{(M-1)}{2}}^{\frac{(M-1)}{2}} \mathrm{e}^{\mathrm{j}m\Delta\mu} \approx M + \mathrm{j}\Delta\mu \cdot \sum_{m=-\frac{(M-1)}{2}}^{\frac{(M-1)}{2}} m - \frac{\Delta\mu^2}{2} \cdot \sum_{m=-\frac{(M-1)}{2}}^{\frac{(M-1)}{2}} m^2 - \cdots$$
$$\approx M - \frac{M}{24} \cdot \Delta\mu^2(M^2-1) + \mathcal{O}(\Delta\mu^4).$$

Note that the terms containing odd powers of m evaluate to zero. Similarly, in case of a small $\Delta\mu$, we get for β and γ the expressions

$$\beta = -\mathrm{j} \cdot \sum_{m=-\frac{(M-1)}{2}}^{\frac{(M-1)}{2}} m \cdot \mathrm{e}^{\mathrm{j}m\Delta\mu} \approx -\mathrm{j} \cdot \sum_{m=-\frac{(M-1)}{2}}^{\frac{(M-1)}{2}} m \cdot \left(1 + \mathrm{j}m\Delta\mu - \frac{\Delta\mu^2}{2}m^2 - \cdots\right)$$
$$\approx \frac{M}{12} \cdot \Delta\mu(M^2-1) - \frac{M}{1440} \cdot \Delta\mu^3(3M^4 - 10M^2 + 7) + \mathcal{O}(\Delta\mu^5),$$
$$\gamma = \sum_{m=-\frac{(M-1)}{2}}^{\frac{(M-1)}{2}} m^2 \cdot \mathrm{e}^{\mathrm{j}m\Delta\mu} \approx -\mathrm{j} \cdot \sum_{m=-\frac{(M-1)}{2}}^{\frac{(M-1)}{2}} m^2 \cdot \left(1 + \mathrm{j}m\Delta\mu - \frac{\Delta\mu^2}{2}m^2 - \cdots\right)$$
$$\approx \frac{M}{12} \cdot (M^2-1) - \frac{M}{480} \cdot \Delta\mu^2(3M^4 - 10M^2 + 7) + \mathcal{O}(\Delta\mu^4).$$

Finally, with the sample covariance matrix

$$\hat{\boldsymbol{R}}_{S_0} = \begin{bmatrix} \hat{P}_1 & \hat{\rho}\sqrt{\hat{P}_1\hat{P}_2} \\ \hat{\rho}\sqrt{\hat{P}_1\hat{P}_2} & \hat{P}_2 \end{bmatrix} \tag{B.571}$$

and the help of the Taylor approximation terms above, we can evaluate the 1-D NC CRB expression in Corollary 9.2.2 for two closely-spaced strictly non-circular sources. Due to the cancellation of relevant terms when using only approximation terms of lower order, we also need to consider higher-order Taylor approximation terms[5] for α, β, and γ. After some tedious calculations, we obtain

$$\mathrm{Tr}\left\{\boldsymbol{C}^{(\mathrm{nc})}\right\} = \frac{\sigma_{\mathrm{n}}^2}{2N} \cdot z \cdot \frac{\hat{P}_1 + \hat{P}_2}{\hat{P}_1\hat{P}_2}, \tag{B.572}$$

where

$$z = \frac{x_0 + x_1\Delta\mu^2 + x_2\Delta\mu^4 + \mathcal{O}(\Delta\mu^6)}{y_1\Delta\mu^2 + y_2\Delta\mu^4 + y_3\Delta\mu^6 + \mathcal{O}(\Delta\mu^8)}. \tag{B.573}$$

[5]Here, we used Taylor approximation terms up to the 6th order.

It is apparent that the first term in the numerator and the first two terms in the denominator of (B.572) are dominant. Neglecting the non-relevant higher-order terms in the numerator and denominator of (B.572) and applying some algebraic manipulations, an expression in the form of (9.42) can be deduced. Finally, to make the result more general, we consider an arbitrary phase reference and substitute $\Delta\varphi$ by $\Delta\phi$ to obtain (9.42). This concludes the proof. □

B.40. Proof of Theorem 9.4.4

The proof of Theorem 9.4.4 in Section 9.4.2.2 follows the same steps as the proof in Appendix B.39. Under the same assumptions, we compute the matrices $\boldsymbol{A}^{\mathrm{H}} \cdot \boldsymbol{A}$, $\boldsymbol{D}^{\mathrm{H}} \cdot \boldsymbol{A}$, and $\boldsymbol{D}^{\mathrm{H}} \cdot \boldsymbol{D}$ in the same way. The difference is, however, that we evaluate the 1-D CRB expression given in (9.6). Using the same Taylor series approximations as before, we obtain a similar expression as (B.572). Finally, neglecting the non-dominant terms in the numerator and the denominator, and substituting $\Delta\phi$ for $\Delta\varphi$, we arrive at the expression in (9.45) to prove this theorem. □

Appendix C.

Norms

This appendix introduces various vector and matrix norms as well as some important results that are of particular relevance to the compressed sensing and sparse signal recovery algorithms discussed in Chapter 8. For more details, we refer the reader to [BV04] and [RFP10, CW08].

C.1. Vector norms

A vector norm denoted by $\|\boldsymbol{x}\|$ of a vector $\boldsymbol{x} \in \mathbb{C}^{m\times 1}$ is a function from $\mathbb{C}^{m\times 1}$ to $\mathbb{R}$ that satisfies the following properties [BV04]:

1. $\|\boldsymbol{x}\| \geq 0$ for any vector $\boldsymbol{x} \in \mathbb{C}^{m\times 1}$, and $\|\boldsymbol{x}\| = 0$ if and only if $\boldsymbol{x} = 0$
2. $\|t\boldsymbol{x}\| = |t|\|\boldsymbol{x}\|$ for any vector $\boldsymbol{x} \in \mathbb{C}^{m\times 1}$ and any scalar $t \in \mathbb{C}$
3. the triangle inequality $\|\boldsymbol{x} + \boldsymbol{y}\| \leq \|\boldsymbol{x}\| + \|\boldsymbol{y}\|$ for any vectors $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{C}^{m\times 1}$.

Note that in the scalar case $m = 1$, the vector norm reduces to the absolute value function.

The most commonly used vector norms are the so-called ℓ_p-norms, which are defined by

$$\|\boldsymbol{x}\|_p = \left(\sum_{i=1}^{m} |\boldsymbol{x}_i|^p\right)^{1/p}. \tag{C.1}$$

It can be shown that the ℓ_p-norm defines a vector norm for any $p > 0$. The following ℓ_p-norms are of particular interest:

The ℓ_1-norm of a vector $\boldsymbol{x}$ is defined as

$$\|\boldsymbol{x}\|_1 = \sum_{i=1}^{m} |\boldsymbol{x}_i|. \tag{C.2}$$

The ℓ_2-norm or the Euclidian norm of a vector $\boldsymbol{x}$ is defined as

$$\|\boldsymbol{x}\|_2 = \sqrt{\boldsymbol{x}^{\mathrm{H}}\boldsymbol{x}} = \sqrt{\sum_{i=1}^{m} |\boldsymbol{x}_i|^2}. \tag{C.3}$$

The ℓ_∞-norm of a vector $\boldsymbol{x}$ is defined as

$$\|\boldsymbol{x}\|_\infty = \max_{1\leq i\leq m} |x_i|. \tag{C.4}$$

Another special norm is the ℓ_0-norm of a vector $\boldsymbol{x}$, which denotes the number of non-zero elements of $\boldsymbol{x}$. It is defined as

$$\|\boldsymbol{x}\|_0 = \{k \mid x_k \neq 0\}\,. \tag{C.5}$$

The ℓ_0-norm is not a norm but a pseudo-norm as it does not satisfy the homogeneity property (the second property) of a norm, i.e., the ℓ_0-norm satisfies $\|t\boldsymbol{x}\|_0 = \|\boldsymbol{x}\|_0$ instead of $\|t\boldsymbol{x}\|_0 = |t|\|\boldsymbol{x}\|_0$. The ℓ_0-norm is popular in compressive sensing, where the goal is to find the sparsest solution to an underdetermined linear system of equations. As minimizing the ℓ_0-norm of a vector is a non-convex problem, which is usually NP-hard to solve, it is often approximated by the convex ℓ_1-norm minimization.

C.2. Matrix norms

The concept of a vector norm can be extended to the matrix case. The matrix norm $\|\boldsymbol{X}\|$ of a matrix $\boldsymbol{X} \in \mathbb{C}^{m\times n}$ is a function from $\mathbb{C}^{m\times n}$ to $\mathbb{R}$ that satisfies the above mentioned properties for vector norms. The following matrix norms are commonly used.

C.2.1. Frobenius norm

The Frobenius norm of a matrix $\boldsymbol{X} \in \mathbb{C}^{m\times n}$ of rank r is defined by [BV04]

$$\|\boldsymbol{X}\|_{\mathrm{F}} = \sqrt{\mathrm{Tr}\left\{\boldsymbol{X}^{\mathrm{H}}\boldsymbol{X}\right\}} = \sqrt{\mathrm{Tr}\left\{\mathrm{vec}\left\{\boldsymbol{X}\right\}^{\mathrm{H}}\mathrm{vec}\left\{\boldsymbol{X}\right\}\right\}} = \sqrt{\sum_{i=1}^{m}\sum_{j=1}^{n}|\boldsymbol{X}_{ij}|^2} \tag{C.6}$$

$$= \sqrt{\sum_{i=1}^{r}\sigma_i^2(\boldsymbol{X})}, \tag{C.7}$$

where $\sigma_i(\boldsymbol{X})$ denotes the i-th singular value of $\boldsymbol{X}$. The last equality can be shown as follows: Let the economy size SVD of $\boldsymbol{X}$ be given by $\boldsymbol{X} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^{\mathrm{H}}$, where $\boldsymbol{U} \in \mathbb{C}^{m\times r}$ and $\boldsymbol{V} \in \mathbb{C}^{n\times r}$ have complex orthonormal columns, i.e., $\boldsymbol{U}^{\mathrm{H}}\boldsymbol{U} = \boldsymbol{I}_r$ and $\boldsymbol{V}^{\mathrm{H}}\boldsymbol{V} = \boldsymbol{I}_r$, and $\boldsymbol{\Sigma} \in \mathbb{R}^{r\times r}$ contains the real-valued singular values on its diagonal. Inserting this result into (C.6), we obtain

$$\|\boldsymbol{X}\|_{\mathrm{F}} = \sqrt{\mathrm{Tr}\left\{\boldsymbol{V}\boldsymbol{\Sigma}\boldsymbol{U}^{\mathrm{H}}\boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^{\mathrm{H}}\right\}} = \sqrt{\mathrm{Tr}\left\{\boldsymbol{V}\boldsymbol{\Sigma}^2\boldsymbol{V}^{\mathrm{H}}\right\}} = \sqrt{\mathrm{Tr}\left\{\boldsymbol{V}^{\mathrm{H}}\boldsymbol{V}\boldsymbol{\Sigma}^2\right\}}$$

$$= \sqrt{\mathrm{Tr}\left\{\boldsymbol{\Sigma}^2\right\}} = \sqrt{\sum_{i=1}^{r}\sigma_i^2(\boldsymbol{X})}. \tag{C.8}$$

Thus, the Frobenius norm of a matrix $\boldsymbol{X}$ is equivalent to the ℓ_2-norm of its vector of singular values.

C.2.2. Spectral norm

The spectral norm (also induced 2-norm or operator norm) of a $\boldsymbol{X} \in \mathbb{C}^{m\times n}$ is defined as [BV04]

$$\|\boldsymbol{X}\|_2 = \max_{\boldsymbol{z}\neq\boldsymbol{0}} \frac{\|\boldsymbol{X}\boldsymbol{z}\|_2}{\|\boldsymbol{z}\|_2} = \sigma_1(\boldsymbol{X}). \tag{C.9}$$

To show the last equality, we first consider $\|\boldsymbol{X}\|_2^2$ and obtain

$$\|\boldsymbol{X}\|_2^2 = \left(\max_{\boldsymbol{z}} \frac{\|\boldsymbol{X}\boldsymbol{z}\|_2}{\|\boldsymbol{z}\|_2}\right)^2 = \max_{\boldsymbol{z}} \frac{\|\boldsymbol{X}\boldsymbol{z}\|_2^2}{\|\boldsymbol{z}\|_2^2} = \max_{\boldsymbol{z}} \frac{\boldsymbol{z}^{\mathrm{H}}\boldsymbol{X}^{\mathrm{H}}\boldsymbol{X}\boldsymbol{z}}{\boldsymbol{z}^{\mathrm{H}}\boldsymbol{z}} = \lambda_1(\boldsymbol{X}^{\mathrm{H}}\boldsymbol{X}), \tag{C.10}$$

which follows from the Rayleigh quotient rule, where $\lambda_1(\boldsymbol{X}^{\mathrm{H}}\boldsymbol{X})$ denotes the maximum eigenvalue of $\boldsymbol{X}^{\mathrm{H}}\boldsymbol{X}$. Then, taking the square root of $\|\boldsymbol{X}\|_2^2$, yields

$$\|\boldsymbol{X}\|_2 = \sqrt{\lambda_1(\boldsymbol{X}^{\mathrm{H}}\boldsymbol{X})} = \sigma_1(\boldsymbol{X}). \tag{C.11}$$

C.2.3. Nuclear norm

The nuclear norm (also trace norm) of a matrix $\boldsymbol{X} \in \mathbb{C}^{m\times n}$ of rank r is defined by [BV04]

$$\|\boldsymbol{X}\|_* = \mathrm{Tr}\left\{(\boldsymbol{X}^{\mathrm{H}}\boldsymbol{X})^{1/2}\right\} = \sum_{i=1}^{r} \sigma_i(\boldsymbol{X}). \tag{C.12}$$

To show the last equality, we again insert the SVD of $\boldsymbol{X}$ into the definition (C.12), which yields

$$\begin{aligned}\|\boldsymbol{X}\|_* &= \mathrm{Tr}\left\{(\boldsymbol{V}\boldsymbol{\Sigma}\boldsymbol{U}^{\mathrm{H}}\boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^{\mathrm{H}})^{1/2}\right\} = \mathrm{Tr}\left\{(\boldsymbol{V}\boldsymbol{\Sigma}^2\boldsymbol{V}^{\mathrm{H}})^{1/2}\right\} = \mathrm{Tr}\left\{(\boldsymbol{V}\boldsymbol{\Sigma}\boldsymbol{V}^{\mathrm{H}}\boldsymbol{V}\boldsymbol{\Sigma}\boldsymbol{V}^{\mathrm{H}})^{1/2}\right\} \\ &= \mathrm{Tr}\left\{\left(\left(\boldsymbol{V}\boldsymbol{\Sigma}\boldsymbol{V}^{\mathrm{H}}\right)^2\right)^{1/2}\right\} = \mathrm{Tr}\left\{\boldsymbol{V}\boldsymbol{\Sigma}\boldsymbol{V}^{\mathrm{H}}\right\} = \mathrm{Tr}\left\{\boldsymbol{V}^{\mathrm{H}}\boldsymbol{V}\boldsymbol{\Sigma}\right\} = \mathrm{Tr}\left\{\boldsymbol{\Sigma}\right\} = \sum_{i=1}^{r}\sigma_i(\boldsymbol{X}).\end{aligned} \tag{C.13}$$

The nuclear norm of a matrix $\boldsymbol{X}$ is equivalent to the ℓ_1-norm of its singular values, since these are all positive.

The nuclear norm can be understood as a convex relaxation of the number of non-zero singular values (i.e., the rank). Thus, by enforcing sparsity of the vector of singular values, the rank of the matrix is minimized.

The matrix norms considered above are related via the following inequalities, which hold for any matrix $\boldsymbol{X}$ of rank r:

$$\|\boldsymbol{X}\|_2 \leq \|\boldsymbol{X}\|_{\mathrm{F}} \leq \|\boldsymbol{X}\|_*. \tag{C.14}$$

The equality only holds if $\boldsymbol{X}$ is a rank-one matrix or a zero matrix.

C.2.4. Mixed norms

The $\ell_{p,q}$-mixed norm of a matrix $\boldsymbol{X}$ is defined by [BV04]

$$\|\boldsymbol{X}\|_{\ell_{p,q}} = \left(\sum_{i=1}^{m} \left(\sum_{j=1}^{n} |\boldsymbol{X}_{ij}|^p \right)^{q/p} \right)^{1/q}. \tag{C.15}$$

More insights into the $\ell_{p,q}$-mixed norm are given by the alternative definition

$$\|\boldsymbol{X}\|_{\ell_{p,q}} = \left(\sum_{i=1}^{m} \|\boldsymbol{x}_i^{\mathrm{T}}\|_p^q \right)^{1/q} \tag{C.16}$$

where $\boldsymbol{x}_i^{\mathrm{T}}$ denotes the i-th row of $\boldsymbol{X} = [\boldsymbol{x}_1, \ldots, \boldsymbol{x}_m]^{\mathrm{T}}$. The $\ell_{p,q}$-mixed norm applies an inner ℓ_p-norm on the rows $\boldsymbol{x}_i^{\mathrm{T}}$ and an outer ℓ_q-norm on the ℓ_p row-norms. The following $\ell_{p,q}$-mixed norms are most commonly used.

For example, if $p = q = 2$, the $\ell_{p,q}$-mixed norm is precisely the Frobenius norm.

In compressed sensing, the $\ell_{2,1}$-mixed norm of $\boldsymbol{X}$ plays an important role as shall be seen. The $\ell_{2,1}$-mixed norm of $\boldsymbol{X}$ is defined as

$$\|\boldsymbol{X}\|_{\ell_{2,1}} = \sum_{i=1}^{m} \left(\sum_{j=1}^{n} |\boldsymbol{X}_{ij}|^2 \right)^{1/2} = \sum_{i=1}^{m} \|\boldsymbol{x}_i^{\mathrm{T}}\|_2, \tag{C.17}$$

Thus, the $\ell_{2,1}$-mixed norm of $\boldsymbol{X}$ is the ℓ_1-norm, i.e., the summation, of the ℓ_2-norms of the rows of $\boldsymbol{X}$. Therefore, minimizing the $\ell_{2,1}$-mixed norm of $\boldsymbol{X}$ promotes row-sparsity of $\boldsymbol{X}$, i.e., the columns of $\boldsymbol{X}$ have the same row support. Note that if $\boldsymbol{X} = \boldsymbol{x}$, $\|\boldsymbol{x}\|_{\ell_{2,1}}$ reduces to the ℓ_1-norm of $\boldsymbol{x}$ and if $\boldsymbol{X} = \boldsymbol{x}^{\mathrm{T}}$, $\|\boldsymbol{x}^{\mathrm{T}}\|_{\ell_{2,1}}$ reduces to the ℓ_2-norm of $\boldsymbol{x}^{\mathrm{T}}$. Thus, the $\ell_{2,1}$-mixed norm can be seen as a generalization of the ℓ_1-norm.

A generalization of the ℓ_0-norm to the matrix case is given by the $\ell_{p,0}$-mixed norm of $\boldsymbol{X} = [\boldsymbol{x}_1, \ldots, \boldsymbol{x}_m]^{\mathrm{T}}$. It is defined as the number of its non-zero rows x_i^{T} according to

$$\|\boldsymbol{X}\|_{p,0} = \{k \mid \|x_k\|_p \neq 0\} \tag{C.18}$$

for any ℓ_p-norm. As for the ℓ_0-vector norm, the $\ell_{p,0}$-mixed norm is a pseudo-norm. In compressed sensing, minimizing the $\ell_{p,0}$-mixed norm promotes joint sparsity in the rows of $\boldsymbol{X}$, i.e., the rows are either jointly zero or non-zero. However, the minimization of the $\ell_{p,0}$-pseudo-norm is non-convex and NP hard. Consequently, the convex $\ell_{p,1}$-mixed norm relaxation is usually applied in practice, which is computationally more tractable.

C.3. Dual norms

C.3.1. Dual vector norm

For any vector norm $\|\cdot\|$ on $\mathbb{C}^{m\times 1}$, its associated dual norm $\|\cdot\|_d$ is defined as [BV04]

$$\|\boldsymbol{z}\|_d = \max_{\boldsymbol{x}} \left\{ |\boldsymbol{z}^{\mathrm{H}}\boldsymbol{x}| \;\middle|\; \|\boldsymbol{x}\| \leq 1 \right\}, \tag{C.19}$$

which can be shown to be a norm. The dual norm can be interpreted as the operator norm of the linear function $f(\boldsymbol{x}) = |\boldsymbol{z}^{\mathrm{T}}\boldsymbol{x}|$, i.e., how large can $f(\boldsymbol{x})$ maximally be relative to the norm of $\boldsymbol{x}$. Therefore, the dual norm is the largest number

$$\|\boldsymbol{z}\|_d = \max_{\boldsymbol{x}\neq\boldsymbol{0}} \frac{|\boldsymbol{z}^{\mathrm{H}}\boldsymbol{x}|}{\|\boldsymbol{x}\|}. \tag{C.20}$$

The dual norm of the dual norm $\|\cdot\|_d$ is again the original norm $\|\cdot\|$.

As an example, the dual norm of the ℓ_2-norm is the ℓ_2-norm itself, i.e.,

$$\|\boldsymbol{z}\|_2 = \max_{\boldsymbol{x}} \left\{ |\boldsymbol{z}^{\mathrm{H}}\boldsymbol{x}| \;\middle|\; \|\boldsymbol{x}\|_2 \leq 1 \right\}, \tag{C.21}$$

which follows from the Cauchy-Schwartz inequality $|\boldsymbol{z}^{\mathrm{H}}\boldsymbol{x}| \leq \|\boldsymbol{z}\|_2\|\boldsymbol{x}\|_2$. Thus, the value of $\boldsymbol{x}$ that maximizes $|\boldsymbol{z}^{\mathrm{H}}\boldsymbol{x}|$ for $\|\boldsymbol{x}\|_2 \leq 1$ is $\boldsymbol{x} = \boldsymbol{z}/\|\boldsymbol{z}\|_2$.

C.3.2. Dual matrix norm

For any matrix norm $\|\cdot\|$ on $\mathbb{C}^{m\times n}$, its dual norm $\|\cdot\|_d$ is given by [BV04]

$$\|\boldsymbol{Z}\|_d = \max_{\boldsymbol{X}} \left\{ \mathrm{Tr}\left\{\boldsymbol{Z}^{\mathrm{H}}\boldsymbol{X}\right\} \;\middle|\; \|\boldsymbol{X}\| \leq 1 \right\}. \tag{C.22}$$

The example from (C.21) also extends to the matrix case. That is the dual norm of the Frobenius norm is again the Frobenius norm, since

$$\|\boldsymbol{Z}\|_F = \max_{\boldsymbol{X}} \left\{ \mathrm{Tr}\left\{\boldsymbol{Z}^{\mathrm{H}}\boldsymbol{X}\right\} \;\middle|\; \|\boldsymbol{X}\|_F \leq 1 \right\}, \tag{C.23}$$

where the maximizing $\boldsymbol{X}$ is equal to $\boldsymbol{X} = \boldsymbol{Z}/\|\boldsymbol{Z}\|_F$.

As a second example, the dual norm of the spectral norm $\|\cdot\|_2$ is the nuclear norm $\|\cdot\|_*$, i.e.,

$$\|\boldsymbol{Z}\|_* = \max_{\boldsymbol{X}} \left\{ \mathrm{Tr}\left\{\boldsymbol{Z}^{\mathrm{H}}\boldsymbol{X}\right\} \;\middle|\; \|\boldsymbol{X}\|_2 \leq 1 \right\}. \tag{C.24}$$

To show this result, we follow [RFP10] and first analyze the constraint $\|\boldsymbol{X}\|_2 \leq 1$. A general bound t on the spectral norm can be written in terms of a linear matrix inequality. $\|\boldsymbol{X}\|_2 \leq t$ corresponds

to

$$\boldsymbol{X}^{\mathrm{H}}\boldsymbol{X} \preceq t^2\boldsymbol{I}_n \quad \text{for} \quad t \geq 0, \tag{C.25}$$

which can be expressed a linear matrix inequality by means of the Schur complement lemma [BV04], i.e.,

$$\begin{bmatrix} t\boldsymbol{I}_m & \boldsymbol{X} \\ \boldsymbol{X}^{\mathrm{H}} & t\boldsymbol{I}_n \end{bmatrix} \succeq \boldsymbol{0}. \tag{C.26}$$

Then, the right hand side of (C.24) for $t = 1$ can be written as the semidefinite program (SDP)

$$\begin{aligned} \max_{\boldsymbol{X}} \quad & \mathrm{Tr}\left\{\boldsymbol{Z}^{\mathrm{H}}\boldsymbol{X}\right\} \\ \text{s.t.} \quad & \begin{bmatrix} \boldsymbol{I}_m & \boldsymbol{X} \\ \boldsymbol{X}^{\mathrm{H}} & \boldsymbol{I}_n \end{bmatrix} \succeq \boldsymbol{0}. \end{aligned} \tag{C.27}$$

The associated dual problem of (C.27) is given by [RFP10]

$$\begin{aligned} \min_{\boldsymbol{W}_1,\boldsymbol{W}_2} \quad & \frac{1}{2}\cdot\mathrm{Tr}\left\{\boldsymbol{W}_1\right\} + \frac{1}{2}\cdot\mathrm{Tr}\left\{\boldsymbol{W}_2\right\} \\ \text{s.t.} \quad & \begin{bmatrix} \boldsymbol{W}_1 & \boldsymbol{X} \\ \boldsymbol{X}^{\mathrm{H}} & \boldsymbol{W}_2 \end{bmatrix} \succeq \boldsymbol{0}. \end{aligned} \tag{C.28}$$

Let the economy size SVD of $\boldsymbol{X}$ be given by $\boldsymbol{X} = \boldsymbol{U}\cdot\boldsymbol{\Sigma}\cdot\boldsymbol{V}^{\mathrm{H}}$ and set $\boldsymbol{W}_1 = \boldsymbol{U}\cdot\boldsymbol{\Sigma}\cdot\boldsymbol{U}^{\mathrm{H}}$ and $\boldsymbol{W}_2 = \boldsymbol{V}\cdot\boldsymbol{\Sigma}\cdot\boldsymbol{V}^{\mathrm{H}}$. Then, it can be shown that the constraint of (C.28) is indeed feasible to (C.27), i.e.,

$$\begin{bmatrix} \boldsymbol{W}_1 & \boldsymbol{X} \\ \boldsymbol{X}^{\mathrm{H}} & \boldsymbol{W}_2 \end{bmatrix} = \begin{bmatrix} \boldsymbol{U} \\ \boldsymbol{V} \end{bmatrix}\cdot\boldsymbol{\Sigma}\cdot\begin{bmatrix} \boldsymbol{U} \\ \boldsymbol{V} \end{bmatrix}^{\mathrm{H}} \succeq \boldsymbol{0}$$

and that the objective function evaluates to $\frac{1}{2}\cdot\mathrm{Tr}\left\{\boldsymbol{W}_1\right\} + \frac{1}{2}\cdot\mathrm{Tr}\left\{\boldsymbol{W}_2\right\} = \mathrm{Tr}\left\{\boldsymbol{\Sigma}\right\} = \|\boldsymbol{X}\|_*$. Hence, we have that the dual norm of the spectral norm is upper-bounded by the nuclear norm [RFP10].

This also shows that the nuclear norm can either be computed by the SDP in (C.27) or by its dual problem in (C.28).

C.4. Atomic norm

In compressed sensing, the vector to be recovered is often a non-negative combination of a small number of atoms from an arbitrary, possibly infinite set $\mathcal{A} \in \mathbb{C}^{m\times 1}$, which is termed the atomic set.

This assumption generalizes the notion of sparsity as shall be seen. Suppose the vector $\boldsymbol{x} \in \mathbb{C}^{m\times 1}$ is formed as

$$\boldsymbol{x} = \sum_{i=1}^{k} c_i \cdot \boldsymbol{a}_i \quad \text{for} \quad \boldsymbol{a}_i \in \mathcal{A}, c_i > 0. \tag{C.29}$$

A metric, which promotes the sparse recovery of $\boldsymbol{x}$ in this general sense is the atomic norm. The atomic norm was first proposed in [CRPW12] and is defined as

$$\|\boldsymbol{x}\|_{\mathcal{A}} = \inf\left\{t > 0 \mid \boldsymbol{x} \in t \cdot \operatorname{conv}(\mathcal{A})\right\}, \tag{C.30}$$

where $\operatorname{conv}(\mathcal{A})$ denotes the convex hull of $\mathcal{A}$, i.e., the convex combination of all points contained in $\mathcal{A}$. The atomic norm $\|\boldsymbol{x}\|_{\mathcal{A}}$ provides a measure of how much we need to scale the convex hull of $\mathcal{A}$ so that it contains $\boldsymbol{x}$. The concept of the atomic norm can be understood as a generalization of several norms commonly used for sparse recovery. For instance, when $\mathcal{A}$ is the set of unit norm 1-sparse elements in $\mathbb{C}^{m\times 1}$, the atomic norm is the ℓ_1-norm [CRT06]. Similarly, when $\mathcal{A}$ is the set of unit norm rank-1 matrices, the atomic norm is the nuclear norm [RFP10]. Due to its notion of an infinite atomic set $\mathcal{A}$, the atomic norm can be viewed as a continuous counterpart of the ℓ_1-norm and the nuclear norm used in the discrete setting. In [CRPW12], it is shown that minimizing the atomic norm provides exact solutions of a variety of linear inverse problems. In practice, the atomic norm minimization problem can be reformulated as a computationally efficient semi-definite programming problem (SDP). To elaborate this, we consider the following example.

Consider Example 2.1.1 and suppose a noise-free single snapshot vector $\boldsymbol{x}_0 \in \mathbb{C}^{M\times 1}$ is given by

$$\boldsymbol{x}_0(t) = \sum_{i=1}^{d} \boldsymbol{a}(\mu_i) \cdot s_i(t), \tag{C.31}$$

where $s_i(t) \in \mathbb{C}$ is the symbol received at time t and $\boldsymbol{a}(\mu_i)$ is the steering vector corresponding to the i-th frequency parameter μ_i given by

$$\boldsymbol{a}(\mu_i) = \begin{bmatrix} 1 & \mathrm{e}^{\mathrm{j}\mu_i} & \mathrm{e}^{\mathrm{j}2\mu_i} & \dots & \mathrm{e}^{\mathrm{j}(M-1)\mu_i} \end{bmatrix}^{\mathrm{T}} \in \mathbb{C}^{M\times 1}. \tag{C.32}$$

Using the atomic norm minimization framework, the vector $\boldsymbol{x}_0(t)$ is considered as a non-negative linear combination of the atoms $\boldsymbol{a}(\nu)\cdot b$, where $\nu \in [-\pi, \pi]$ and $b \in \mathbb{C}, |b| = 1$ represents a phase shift corresponding to the sign of $s_i(t)$. Omitting the time instance t, the atomic norm of $\boldsymbol{x}_0$ can be written as [TBSR13]

$$\|\boldsymbol{x}_0\|_{\mathcal{A}} = \inf_{\{\nu_k, c_k, b_k\}} \left\{ \sum_k c_k \;\middle|\; \boldsymbol{x}_0 = \sum_k c_k \cdot \boldsymbol{a}(\nu_k) \cdot b_k,\ c_k \geq 0 \right\}. \tag{C.33}$$

The atomic norm minimization problem in (C.33) is equivalent to the SDP formulation [TBSR13]

$$\min_{t,\boldsymbol{u}} \quad \frac{1}{2}\cdot t + \frac{1}{2M}\cdot \mathrm{Tr}\left\{\mathrm{Toep}\left\{\boldsymbol{u}\right\}\right\} \tag{C.34}$$

$$\text{s. t.} \quad \begin{bmatrix} t & \boldsymbol{x}_0^{\mathrm{H}} \\ \boldsymbol{x}_0 & \mathrm{Toep}\left\{\boldsymbol{u}\right\} \end{bmatrix} \succeq \boldsymbol{0}, \tag{C.35}$$

where $\mathrm{Toep}\left\{\boldsymbol{u}\right\} \in \mathbb{C}^{M\times M}$ with $\boldsymbol{u} = [u_1, \ldots, u_M]^{\mathrm{T}} \in \mathbb{C}^{M\times 1}$ denotes the Hermitian Toeplitz matrix

$$\mathrm{Toep}\left\{\boldsymbol{u}\right\} = \begin{bmatrix} u_1 & u_2 & \cdots & u_M \\ u_2^* & u_1 & \cdots & u_{M-1} \\ \vdots & \vdots & \ddots & \vdots \\ u_M^* & u_{M-1}^* & \cdots & u_1 \end{bmatrix}. \tag{C.36}$$

For the proof of the equivalence of (C.33) and (C.34), we follow the derivation in [YLSX17]. We need to show that $p^\star = \|\boldsymbol{x}_0\|_{\mathcal{A}}$, where $p^\star$ denotes the optimal value of the objective function in (C.34).

We start by showing that $p^\star \leq \|\boldsymbol{x}_0\|_{\mathcal{A}}$. Let $\boldsymbol{x}_0 = \sum_k c_k \cdot \boldsymbol{a}(\nu_k)$ be an atomic decomposition of $\boldsymbol{x}_0$ with $c_k > 0$, where we have omitted b_k for simplicity. Then, we define $\boldsymbol{u}$ such that $\mathrm{Toep}\left\{\boldsymbol{u}\right\} = \sum_k c_k \cdot \boldsymbol{a}(\nu_k)\cdot \boldsymbol{a}^{\mathrm{H}}(\nu_k)$ and let $t = \sum_k c_k$. It follows that

$$\begin{bmatrix} t & \boldsymbol{x}_0^{\mathrm{H}} \\ \boldsymbol{x}_0 & \mathrm{Toep}\left\{\boldsymbol{u}\right\} \end{bmatrix} = \sum_k c_k \cdot \begin{bmatrix} 1 \\ \boldsymbol{a}(\nu_k) \end{bmatrix} \cdot \begin{bmatrix} 1 \\ \boldsymbol{a}(\nu_k) \end{bmatrix}^{\mathrm{H}} \succeq \boldsymbol{0} \tag{C.37}$$

Consequently, t and $\boldsymbol{u}$ form a feasible solution to (C.34) with the objective value $\frac{1}{2}\cdot t + \frac{1}{2M}\cdot \mathrm{Tr}\left\{\mathrm{Toep}\left\{\boldsymbol{u}\right\}\right\} = \sum_k c_k$. Therefore, we have $p^\star \leq \sum_k c_k$. As this is true for any atomic decomposition of $\boldsymbol{x}_0$, we have that $p^\star \leq \|\boldsymbol{x}_0\|_{\mathcal{A}}$.

Now, suppose $(t, \boldsymbol{u})$ is an optimal solution to (C.34). Since $\mathrm{Toep}\left\{\boldsymbol{u}\right\} \succeq \boldsymbol{0}$, we can apply the Vandermonde decomposition $\mathrm{Toep}\left\{\boldsymbol{u}\right\} = \boldsymbol{A}\cdot\boldsymbol{C}\cdot\boldsymbol{A}^{\mathrm{H}}$ in (8.15) to obtain a set (K, ν_k, c_k). Due to (C.35), $\boldsymbol{x}_0$ lies in the range of $\mathrm{Toep}\left\{\boldsymbol{u}\right\}$ and admits the atomic decomposition $\boldsymbol{x}_0 = \sum_{k=1}^{K} c_k \cdot \boldsymbol{a}(\nu_k) = \boldsymbol{A}\cdot\boldsymbol{c}$. By the Schur complement lemma [BV04], we have that

$$\begin{aligned} t &\geq \boldsymbol{x}_0^{\mathrm{H}} \cdot \mathrm{Toep}\left\{\boldsymbol{u}\right\}^{-1} \cdot \boldsymbol{x}_0 = \boldsymbol{c}^{\mathrm{H}} \cdot \boldsymbol{A}^{\mathrm{H}} \cdot (\boldsymbol{A}\cdot\boldsymbol{C}\cdot\boldsymbol{A}^{\mathrm{H}})^{-1} \cdot \boldsymbol{A}\cdot\boldsymbol{c} \\ &= \boldsymbol{c}^{\mathrm{H}} \cdot \boldsymbol{C}^{-1} \cdot \boldsymbol{c} = \sum_{k=1}^{K} c_k. \end{aligned} \tag{C.38}$$

Then, it follows that the objective value $p^\star = \frac{1}{2}\cdot t + \frac{1}{2M}\cdot \mathrm{Tr}\left\{\mathrm{Toep}\left\{\boldsymbol{u}\right\}\right\} \geq \sum_k c_k \geq \|\boldsymbol{x}_0\|_{\mathcal{A}}$. Thus, we have shown that $p^\star \leq \|\boldsymbol{x}_0\|_{\mathcal{A}}$ and that $p^\star \geq \|\boldsymbol{x}_0\|_{\mathcal{A}}$, from which we conclude that $p^\star = \|\boldsymbol{x}_0\|_{\mathcal{A}}$. This completes the proof.

Bibliography

Publications as First Author or Co-Author

[BCS+16] E. R. Balda, S. A. Cheema, J. Steinwandt, M. Haardt, A. Weiss, and A. Yeredor, "First-order perturbation analysis of low-rank tensor approximations based on the truncated HOSVD," in *Proc. Asilomar Conference on Signals, Systems, and Computers*, Pacific Grove, CA, Nov. 2016, pp. 1723 – 1727.

[SdH11a] J. Steinwandt, R. C. de Lamare, and M. Haardt, "Beamspace direction finding based on the conjugate gradient algorithm," in *Proc. International ITG Workshop on Smart Antennas (WSA)*, Aachen, Germany, Feb. 2011.

[SdH11b] ——, "Knowledge-aided direction finding based on Unitary ESPRIT," in *Proc. Asilomar Conference on Signals, Systems, and Computers*, Pacific Grove, CA, Nov. 2011, pp. 613 – 617.

[SdH13] ——, "Beamspace direction finding based on the conjugate gradient and the auxiliary vector filtering algorithms," *Signal Processing*, vol. 93, pp. 641 – 651, Apr. 2013.

[SdW+11] J. Steinwandt, R. C. de Lamare, L. Wang, N. Song, and M. Haardt, "Widely linear adaptive beamforming algorithm based on the conjugate gradient method," in *Proc. International ITG Workshop on Smart Antennas (WSA)*, Aachen, Germany, Feb. 2011.

[SH13a] J. Steinwandt and M. Haardt, "Optimal widely-linear distributed beamforming for relay networks," in *Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, Vancouver, Canada, May 2013, pp. 4241 – 4245.

[SH13b] ——, "Reduced-complexity distributed beamforming algorithm for individual relay power constraints," in *Proc. IEEE International Workshop on Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP)*, Saint Martin, French Antilles, Dec. 2013, pp. 496 – 499.

[SH13c] ——, "Widely-linear distributed beamforming for weak-sense non-circular sources based on relay power minimization," in *Proc. International Symposium on Wireless Communications Systems (ISWCS)*, Ilmenau, Germany, Aug. 2013, invited paper.

[SRH13a] J. Steinwandt, F. Roemer, and M. Haardt, "Performance analysis of ESPRIT-type algorithms for non-circular sources," in *Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, Vancouver, Canada, May 2013, pp. 3986 – 3990.

[SRH13b] ——, "Performance analysis of ESPRIT-type algorithms for strictly non-circular sources using structured least squares," in *Proc. IEEE International Workshop on Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP)*, Saint Martin, French Antilles, Dec. 2013, pp. 316 – 319.

[SRH14a] ——, "Analytical ESPRIT-based performance study: What can we gain from non-circular sources?" in *Proc. IEEE Sensor Array and Multichannel Signal Processing Workshop (SAM)*, A Coruña, Spain, Jun. 2014, pp. 17 – 20.

[SRH14b] ——, "Asymptotic performance analysis of ESPRIT-type algorithms for circular and strictly non-circular sources with spatial smoothing," in *Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, Florence, Italy, May 2014, pp. 2277 – 2281.

[SRH15a] J. Steinwandt, V. Radhakrishnan, and M. Haardt, "Distributed beamforming for cooperative networks with widely-linear processing at the relays and the receiver," in *Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, Brisbane, Australia, Apr. 2015, pp. 2714 – 2718.

[SRH15b] J. Steinwandt, F. Roemer, and M. Haardt, "Deterministic Cramér-Rao bound for a mixture of circular and strictly non-circular signals," in *Proc. International Symposium on Wireless Communications Systems (ISWCS)*, Brussels, Belgium, Aug. 2015.

[SRH15c] ——, "ESPRIT-type algorithms for a received mixture of circular and strictly non-circular signals," in *Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, Brisbane, Australia, Apr. 2015, pp. 2809 – 2813.

[SRH16a] ——, "Analytical performance assessment of ESPRIT-type algorithms for coexisting circular and strictly non-circular signals," in *Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, Shanghai, China, Mar. 2016, pp. 2931 – 2935.

[SRH16b] ——, "Analytical performance evaluation of multi-dimensional Tensor-ESPRIT-based algorithms for strictly non-circular sources," in *Proc. IEEE Sensor Array and Multichannel Signal Processing Workshop (SAM)*, Rio de Janeiro, Brazil, Jul. 2016, invited paper.

[SRH16c] ——, "Sparsity-based direction-of-arrival estimation for strictly non-circular sources," in *Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, Shanghai, China, Mar. 2016, pp. 3246 – 3250.

[SRH17a] ——, "Generalized least squares for ESPRIT-type direction of arrival estimation," *IEEE Signal Processing Letters*, vol. 24, no. 11, pp. 1681 – 1685, Nov. 2017.

[SRH17b] ——, "Performance analysis of ESPRIT-type algorithms for co-array structures," in *Proc. IEEE International Workshop on Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP)*, Curacao, Dutch Antilles, Dec. 2017, pp. 482 – 486.

[SRH17c] ——, "Second-order performance analysis of Standard ESPRIT," in *Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, New Orleans, LA, Mar. 2017, pp. 3051 – 3055.

[SRHD14] J. Steinwandt, F. Roemer, M. Haardt, and G. Del Galdo, "R-dimensional ESPRIT-type algorithms for strictly second-order non-circular sources and their performance analysis," *IEEE Transactions on Signal Processing*, vol. 62, no. 18, pp. 4824 – 4838, Sep. 2014, received the TU Ilmenau Best Paper Award in the area of 'Engineering Sciences'.

[SRHD16] ——, "Deterministic Cramér-Rao bound for strictly non-circular sources and analytical analysis of the achievable gains," *IEEE Transactions on Signal Processing*, vol. 64, no. 17, pp. 4417 – 4431, Sep. 2016.

[SRHD17] ——, "Performance analysis of multi-dimensional ESPRIT-type algorithms for arbitrary and strictly non-circular sources with spatial smoothing," *IEEE Transactions on Signal Processing*, vol. 65, no. 9, pp. 2262 – 2276, May 2017.

[SRS+16] J. Steinwandt, F. Roemer, C. Steffens, M. Haardt, and M. Pesavento, "Gridless super-resolution direction finding for strictly non-circular sources based on atomic norm minimization," in *Proc. Asilomar Conference on Signals, Systems, and Computers*, Pacific Grove, CA, Nov. 2016, pp. 1518 – 1522.

[SSPH16] J. Steinwandt, C. Steffens, M. Pesavento, and M. Haardt, "Sparsity-aware direction finding for strictly non-circular sources based on rank minimization," in *Proc. IEEE Sensor Array and Multichannel Signal Processing Workshop (SAM)*, Rio de Janeiro, Brazil, Jul. 2016, invited paper.

[SSW+11] N. Song, J. Steinwandt, L. Wang, R. C. de Lamare, and M. Haardt, "Non-data-aided adaptive beamforming algorithm based on the widely linear auxiliary vector filter," in *Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, Prague, Czech Republic, May 2011, pp. 2636 – 2639.

[SVH14a] J. Steinwandt, S. A. Vorobyov, and M. Haardt, "Joint beamforming and transmit design for the non-regenerative MIMO broadcast relay channel," in *Proc. IEEE Sensor Array and Multichannel Signal Processing Workshop (SAM)*, A Coruña, Spain, Jun. 2014, pp. 169 – 172.

[SVH14b] ——, "Secrecy rate maximization for information and energy transfer in MIMO beamforming networks," in *Proc. Asilomar Conference on Signals, Systems, and Computers*, Pacific Grove, CA, Nov. 2014, pp. 1989 – 1993.

[SVH14c] ——, "Secrecy rate maximization for MIMO Gaussian wiretap channels with multiple eavesdroppers via alternating matrix POTDC," in *Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, Florence, Italy, May 2014, pp. 5686 – 5690.

References by Other Authors

[AD05] H. Abeida and J. P. Delmas, "Gaussian Cramer-Rao bound for direction estimation of noncircular signals in unknown noise fields," *IEEE Transactions on Signal Processing*, vol. 53, no. 12, pp. 4610–4618, Dec. 2005.

[AD06] H. Abeida and J.-P. Delmas, "MUSIC-like estimation of direction of arrival for noncircular sources," *IEEE Transactions on Signal Processing*, vol. 54, no. 7, pp. 2678–2690, Jul. 2006.

[AD08] H. Abeida and J. P. Delmas, "Statistical performance of MUSIC-like algorithms in resolving noncircular sources," *IEEE Transactions on Signal Processing*, vol. 56, no. 9, pp. 4317–4329, Sep. 2008.

[Ame85] T. Amemiya, "Generalized least squares theory," in *Advanced Econometrics*. Harvard University Press, Cambridge, MA, 1985.

[ApS15] M. ApS, *The MOSEK optimization toolbox for MATLAB manual. Version 7.1 (Revision 28).*, 2015. [Online]. Available: http://docs.mosek.com/7.1/toolbox/index.html

[AR88] K. S. Arun and B. D. Rao, "An improved Toeplitz approximation method," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing*, vol. IV, New York, NY, Apr. 1988, pp. 2352–2355.

[Bar83] A. J. Barabell, "Improving the resolution performance of eigenstructure-based direction-finding algorithms," in *Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, Boston, MA, May 1983, pp. 336–339.

[BB02] J. K. Baksalary and O. M. Baksalary, "Commutativity of projectors," *Linear Algebra and its Applications*, vol. 341, no. 1, pp. 129–142, Jan. 2002.

[BCW+17] E. R. Balda, S. A. Cheema, A. Weiss, A. Yeredor, and M. Haardt, "Perturbation analysis of joint eigenvalue decomposition algorithms," in *Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, New Orleans, LA, Mar. 2017.

[BGP+13] M. Boizard, G. Ginolhac, F. Pascal, S. Miron, and P. Forster, "Numerical performance of a tensor MUSIC algorithm based on HOSVD for a mixture of polarized sources," in *Proc. European Signal Processing Conference (EUSIPCO)*, Marrakech, Marocco, Sep. 2013.

[BL86] A. Bax and L. Lerner, "Two-dimensional NMR spectroscopy," *J. Amer. Assoc. for the Advancement of Science*, vol. 232, pp. 960–967, May 1986.

[Boy08] R. Boyer, "Decoupled root-MUSIC algorithm for multidimensional harmonic retrieval," in *Proc. of the 9th Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*, Recife, Brazil, Jul. 2008.

[Bre78] J. W. Brewer, "Kronecker products and matrix calculus in system theory," *IEEE Transactions on Circuits and Systems*, vol. CAS-25, no. 9, pp. 772–781, Sep. 1978.

[Bri75] D. R. Brillinger, *Time Series: Data Analysis and Theory.* New York: Holt, Rhinehart and Winston, 1975.

[BTR13] B. N. Bhaskar, G. Tang, and B. Recht, "Atomic norm denoising with applications to line spectral estimation," *IEEE Transactions on Signal Processing*, vol. 61, no. 23, pp. 5987–5999, Dec. 2013.

[BV04] S. Boyd and L. Vandenberghe, *Convex Optimization.* Cambridge University Press, 2004.

[Car11] C. Carathéodory, "Über den Variabilitätsbereich der Fourierschen Konstanten von positiven harmonischen Funktionen," *Rendiconti del Circolo Matematico di Palermo (1884-1940)*, vol. 32, no. 1, pp. 193–217, 1911.

[CB07] P. Chevalier and A. Blin, "Widely linear MVDR beamformers for the reception of an unknown signal corrupted by noncircular interferences," *IEEE Transactions on Signal Processing*, vol. 55, no. 11, pp. 5323–5336, Nov. 2007.

[CC15] Y. Chi and Y. Chen, "Compressive two-dimensional harmonic retrieval via atomic norm minimization," *IEEE Transactions on Signal Processing*, vol. 63, no. 4, pp. 1030–1042, Feb. 2015.

[CCS10] J. Cai, E. Candés, and Z. Shen, "A singular value thresholding algorithm for matrix completion," *SIAM Journal on Optimization*, vol. 20, no. 4, pp. 1956–1982, Mar. 2010.

[CDS98] S. S. Chen, D. L. Donoho, and M. A. Saunders, "Atomic decomposition by basis pursuit," *SIAM Journal On Scientific Computing*, vol. 20, no. 1, pp. 33–61, Jan. 1998.

[CF11] C. Carathéodory and L. Fejér, "Über den Zusammenhang der extremen von harmonischen Funktionen mit ihren Koeffizienten und über den Picard-Landauschen Satz," *Rendiconti del Circolo Matematico di Palermo (1884-1940)*, vol. 32, no. 1, pp. 218–239, 1911.

[CFG13] E. J. Candés and C. Fernandez-Granda, "Super-resolution from noisy data," *Journal of Fourier Analysis and Applications*, vol. 19, no. 6, pp. 1229–1254, Dec. 2013.

[CFG14] ——, "Towards a mathematical theory of super-resolution," *Communications on Pure and Applied Mathematics*, vol. 67, no. 6, pp. 906–956, Jan. 2014.

[CG90] P. Comon and G. H. Golub, "Tracking a few extreme singular values and vectors in signal processing," *Proceedings of the IEEE*, vol. 78, no. 8, pp. 1327–1343, Sep. 1990.

[CGN98] W. Chen, G. B. Giannakis, and N. Nandhakumar, "A harmonic retrieval framework for discontinuous motion estimation," *IEEE Transactions on Image Processing*, vol. 7, no. 9, pp. 1242–1257, Sep. 1998.

[CP06] P. Chevalier and F. Pipon, "New insights into optimal widely linear array receivers for the demodulation of BPSK, MSK and GMSK signals corrupted by non circular interferences - application to SAIC," *IEEE Transactions on Signal Processing*, vol. 54, no. 3, pp. 870–883, Mar. 2006.

[CR09] E. Candés and B. Recht, "Exact matrix completion via convex optimization," *Foundations of Computational Mathematics*, vol. 9, pp. 717–772, Dec. 2009.

[CRKH14] Y. Cheng, F. Roemer, O. Khatib, and M. Haardt, "Tensor subspace tracking via Kronecker structured projections (TeTraKron) for time-varying multidimensional harmonic retrieval," *EURASIP Journal on Advances in Signal Processing*, vol. 123, no. 1, Dec. 2014.

[CRPW12] V. Chandrasekaran, B. Recht, P. A. Parrilo, and A. S. Willsky, "The convex geometry of linear inverse problems," *Foundations of Computational Mathematics*, vol. 12, no. 6, pp. 805–849, 2012.

[CRT06] E. Candés, J. K. Romberg, and T. Tao, "Stable signal recovery from incomplete and inaccurate measurements," *Comm. on Pure and Applied Mathematics*, vol. 59, no. 8, pp. 1207–1223, Aug. 2006.

[CSPC11] Y. Chi, L. Scharf, A. Pezeshki, and A. Calderbank, "Sensitivity to basis mismatch in compressed sensing," *IEEE Transactions on Signal Processing*, vol. 59, no. 5, pp. 2182–2195, May 2011.

[CSS12] F. K. W. Chan, H. C. So, and W. Sun, "Subspace approach for two-dimensional parameter estimation of multiple damped sinusoids," *Signal Processing*, vol. 92, no. 9, pp. 2172–2179, Feb. 2012.

[CT05] E. Candés and T. Tao, "Decoding by linear programming," *IEEE Transactions on Information Theory*, vol. 51, no. 12, pp. 4203–4215, Dec. 2005.

[CW08] E. Candés and M. Wakin, "An introduction to compressive sampling," *IEEE Signal Processing Magazine*, vol. 25, no. 2, pp. 21–30, Aug. 2008.

[CWS01] P. Chargé, Y. Wang, and J. Saillard, "A non circular sources direction finding method using polynomial rooting," *Signal Processing*, no. 81, pp. 1765–1770, 2001.

[DA04] J. P. Delmas and H. Abeida, "Stochastic Cramér-Rao bound for noncircular signals with application to DOA estimation," *IEEE Transactions on Signal Processing*, vol. 52, no. 11, pp. 3192–3199, Nov. 2004.

[DA06] ——, "Cramer-Rao bounds of DOA estimates for BPSK and QPSK modulated signals," *IEEE Transactions on Signal Processing*, vol. 54, no. 1, pp. 117–126, Jan. 2006.

[dLdMV00a] L. de Lathauwer, B. de Moor, and J. Vandewalle, "A multilinear singular value decomposition," *SIAM J. Matrix Anal. Appl.*, vol. 21, no. 4, pp. 1253–1278, 2000.

[dLdMV00b] ——, "On the best rank-1 and rank-$(r_1, r_2, \ldots, r_n)$ approximation of higher-order tensors," *SIAM J. Matrix Anal. Appl.*, vol. 21, p. 1324–1342, 2000.

[Don92] D. Donoho, "Super-resolution via sparsity constraints," *SIAM Journal on Mathematical Analysis*, vol. 23, pp. 1309–1331, Sep. 1992.

[Don06] ——, "Compressed sensing," *IEEE Transactions on Information Theory*, vol. 52, no. 4, pp. 1289–1306, Apr. 2006.

[dP95] G. de Prony, "Essai expérimental et analytique: sur les lois de la dilatabilité des fluides élastiques et sur celles de la force expansive de la vapeur de l'eau et de la vapeur de l'alcool á differentes températures," *Journal de l'École Polytechnique*, vol. 1, no. 22, pp. 24–76, 1795.

[EJS82] J. Evans, J. Johnson, and D. Sun, "Application of advanced signal processing techniques to angle of arrival estimation in ATC navigation and surveillance systems," Tech. Rep. 582, M.I.T. Lincoln Laboratory, Cambridge, MA, Tech. Rep., Jun. 1982.

[EK06] J. Eriksson and V. Koivunen, "Complex random vectors and ICA models: Identifiability, uniqueness, and separability," *IEEE Transactions on Information Theory*, vol. 52, no. 3, pp. 1017–1029, Mar. 2006.

[EK12] Y. C. Eldar and G. Kutyniok, *Compressed Sensing: Theory and Applications*. Cambridge University Press, May 2012.

[ES94] A. Eriksson and P. Stoica, "Optimally weighted ESPRIT for direction estimation," *Signal Processing*, vol. 38, no. 2, pp. 223–229, Jul. 1994.

[ESS93] A. Eriksson, P. Stoica, and T. Söderström, "Second-order properties of MUSIC and ESPRIT estimates of sinusoidal frequencies in high SNR scenarios," *IEE Proceedings-F*, vol. 140, no. 4, Aug. 1993.

[EY36] C. Eckart and G. Young, "The approximation of one matrix by another one of lower rank," *Psychometrika*, vol. 1, pp. 211–218, 1936.

[FC10] A. Ferreol and P. Chevalier, "Higher order direction finding for arbitrary noncircular sources: The NC-2q-MUSIC algorithm," in *Proc. European Signal Processing Conference (EUSIPCO)*, Aalborg, Denmark, Aug. 2010.

[FC14] ——, "New insights into second and fourth-order direction finding for non-circular sources," in *Proc. IEEE Sensor Array and Multichannel Signal Processing Workshop (SAM)*, A Coruna, Spain, Jun. 2014.

[FG06] T. Fu and X. Gao, "Simultaneous diagonalization with similarity transformation for non-defective matrices," in *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP 2006)*, Toulouse, France, May 2006.

[FHB01] M. Fazel, H. Hindi, and S. Boyd, "A rank minimization heuristic with application to minimum order system approximation," in *American Control Conference*, Arlington, VA, Jun. 2001.

[FLB04] P. Forster, P. Larzabal, and E. Boyer, "Threshold performance analysis of maximum likelihood DOA estimation," *IEEE Transactions on Signal Processing*, vol. 52, no. 11, pp. 3183–3191, Nov. 2004.

[Fri90] B. Friedlander, "A sensitivity analysis of MUSIC algorithm," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 38, no. 10, pp. 1740–1751, Oct. 1990.

[GB02] L. Gurvits and H. Barnum, "Largest separable balls around the maximally mixed bipartite quantum state," *Physical Review A*, vol. 66, no. 6, Dec. 2002.

[GB14] M. Grant and S. Boyd, "Cvx: Matlab software for disciplined convex programming, version 2.1," http://cvxr.com/cvx, Mar. 2014.

[Geo07] T. T. Georgiou, "The carathéodory–fejér–pisarenko decomposition and its multivariable counterpart," *IEEE Transactions on Automatic Control*, vol. 52, no. 2, pp. 212–228, Feb. 2007.

[GNW08] F. Gao, A. Nallanathan, and Y. Wang, "Improved MUSIC under the coexistence of both circular and noncircular sources," *IEEE Transactions on Signal Processing*, vol. 56, no. 7, pp. 3033–3038, Jul. 2008.

[Gra15] M. J. Grasis, "Analytical performance assessment of ESPRIT-type and NC-ESPRIT-type algorithms for a single source and two sources," TU Ilmenau, Student project, Tech. Rep., Jun. 2015.

[Gra16] ——, "Analytical performance assessment of Tensor-ESPRIT-type algorithms for two sources," TU Ilmenau, Master's thesis, Tech. Rep., Nov. 2016.

[GRP10] A. B. Gershman, M. Rübsamen, and M. Pesavento, "One- and two-dimensional direction-of-arrival estimation: An overview of search-free techniques," *Signal Processing*, vol. 90, pp. 1338–1349, 2010.

[GS05] A. Gershman and N. Sidiropoulos, *Space-Time Processing for MIMO Communications*. New York: Wiley, 2005.

[Gue03] J. R. Guerci, *Space-Time Adaptive Processing for Radar*. Norwood, MA: Artech House, 2003.

[GvL96] G. H. Golub and C. F. van Loan, *Matrix computations*, 3rd ed. John Hopkins University Press, Oct. 1996.

[Haa97a] M. Haardt, "Efficient One-, Two-, and Multidimensional High-Resolution Array Signal Processing," Ph.D. dissertation, Technische Universität München, Shaker Verlag, Aachen, 1997.

[Haa97b] ——, "Structured least squares to improve the performance of ESPRIT-type algorithms," *IEEE Transactions on Signal Processing*, vol. 45, no. 3, pp. 792–799, Mar. 1997.

[HC08] C. J. Hegarty and E. Chatre, "Evolution of the Global Navigation Satellite System (GNSS)," *Proceedings of the IEEE*, vol. 96, no. 12, pp. 1902–1917,, Dec. 2008.

[HF94] G. F. Hatke and K. W. Forsythe, "A class of polynomial rooting algorithm for joint azimuth/elevation estimation using multidimensional arrays," in *Proc. Asilomar Conference on Signals, Systems, and Computers*, Pacific Grove, CA, Nov. 1994, pp. 694–699.

[HLLZ10] Z. T. Huang, Z. M. Liu, J. Liu, and Y. Y. Zhou, "Performance analysis of MUSIC for non-circular signals in the presence of mutual coupling," *IET Radar Sonar Navigation*, vol. 4, no. 5, pp. 703–711, Oct. 2010.

[HM10] M. M. Hyder and K. Mahata, "Direction-of-arrival estimation using a mixed L2,0 norm approximation," *IEEE Transactions on Signal Processing*, vol. 58, no. 9, pp. 4646–4655, Sep. 2010.

[HN95] M. Haardt and J. A. Nossek, "Unitary ESPRIT: How to obtain increased estimation accuracy with a reduced computational burden," *IEEE Transactions on Signal Processing*, vol. 43, no. 5, pp. 1232–1242, May 1995.

[HN98] ——, "Simultaneous Schur decomposition of several non-symmetric matrices to achieve automatic pairing in multidimensional harmonic retrieval problems," *IEEE Transactions on Signal Processing*, vol. 46, no. 1, pp. 161–169, Jan. 1998.

[HPRE14] M. Haardt, M. Pesavento, F. Roemer, and M. N. El Korso, "Subspace methods and exploitation of special array structures," in *Academic Press Library in Signal Processing: Volume 3 - Array and Statistical Signal Processing*, A. M. Zoubir, M. Viberg, R. Chellappa, and S. Theodoridis, Eds., vol. 3. Elsevier Ltd., 2014, pp. 651 – 717, chapter 15, ISBN 978-0-12-411597-2.

[HR99] K. V. S. Hari and B. V. Ramakrishnan, "Performance analysis of modified spatial smoothing technique for direction estimation," *Signal Processing*, vol. 79, pp. 73–85, Nov. 1999.

[HR04] M. Haardt and F. Roemer, "Enhancements of unitary ESPRIT for non-circular sources," in *Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, Montreal, Canada, May 2004.

[HRD08] M. Haardt, F. Roemer, and G. Del Galdo, "Higher-order SVD based subspace estimation to improve the parameter estimation accuracy in multi-dimensional harmonic retrieval problems," *IEEE Transactions on Signal Processing*, vol. 56, no. 7, pp. 3198–3213, Jul. 2008.

[HS90] Y. Hua and T. K. Sarkar, "Matrix pencil method for estimating parameters of exponentially damped/undamped sinusoids in noise," *IEEE Transactions on Signal Processing*, vol. 38, no. 5, pp. 814–824, May 1990.

[HS91] ——, "On SVD for estimating generalized eigenvalues of singular matrix pencil in noise," *IEEE Transactions on Signal Processing*, vol. 39, pp. 892–900, Apr. 1991.

[HS10] M. A. Herman and T. Strohmer, "General deviants: An analysis of perturbations in compressed sensing," *IEEE Journal on Selected Topics in Signal Processing*, vol. 4, no. 2, pp. 342–349, Apr. 2010.

[HTR04] M. Haardt, R. S. Thomä, and A. Richter, "Multidimensional high-resolution parameter estimation with applications to channel sounding," in *High-Resolution and Robust Signal Processing*, Y. Hua, A. Gershman, and Q. Chen, Eds. New York, NY: Marcel Dekker, 2004, pp. 255–338, chapter 5.

[HZN95] M. Haardt, T. Zeitler, and J. A. Nossek, "Spatial separation of wavefronts via antenna arrays using Unitary ESPRIT and spatial smoothing," in *Proc. IEEE International Conference on Telecommunications*, Bali, Indonesia, Apr. 1995.

[IRA+14] M. Ibrahim, F. Roemer, R. Alieiev, G. Del Galdo, and R. S. Thomae, "On the estimation of grid offsets in CS-based direction of arrival estimation," in *Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, Florence, Italy, May 2014.

[JLL09] M. Jin, G. Liao, and J. Li, "Joint DOD and DOA estimation for bistatic MIMO radar," *Signal Processing*, vol. 89, no. 2, pp. 244–251, Feb. 2009.

[KAB83] S. Y. Kung, K. S. Arun, and D. V. Bhaskar Rao, "State space and SVD based approximation methods for the harmonic retrieval problem," *Journal of the Optical Society of America*, vol. 73, pp. 1799–1811, Dec. 1983.

[Kay93] S. M. Kay, *Fundamentals of statistical signal processing: Estimation theory.* Prentice-Hall, Englewood Cliffs, NJ, 1993.

[KB86] M. Kaveh and A. J. Barabell, "The statistical performance of the MUSIC and the minimum-norm algorithms in resolving plane waves in noise," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 34, no. 2, pp. 331–341, Apr. 1986.

[KB09] T. G. Kolda and B. W. Bader, "Tensor decompositions and applications," *SIAM Review*, vol. 51, no. 3, pp. 455–500, Sep. 2009.

[Kow09] M. Kowalski, "Sparse regression using mixed norms," *Applied and Computational Harmonic Analysis*, vol. 27, no. 3, pp. 303–324, 2009.

[KT83] R. Kumaresan and D. W. Tufts, "Estimating the angles of arrival of multiple plane waves," *IEEE Transactions on Aerospace and Electronic Systems*, pp. 134–139, Jan. 1983.

[KV96] H. Krim and M. Viberg, "Two decades of array signal processing research: parametric approach," *IEEE Signal Processing Magazine*, vol. 13, no. 4, pp. 67–94, Jul. 1996.

[LC16] Y. Li and Y. Chi, "Off-the-grid line spectrum denoising and estimation with multiple measurement vectors," *IEEE Transactions on Signal Processing*, vol. 64, no. 5, pp. 1257–1269, Mar. 2016.

[Lee80] A. Lee, "Centrohermitian and skew-centrohermitian matrices," *Linear Algebra and its Applications*, vol. 29, pp. 205–210, Feb. 1980.

[LHZ12] Z. M. Liu, Z. T. Huang, and Y. Y. Zhou, "Direction-of-arrival estimation of noncircular signals via sparse representation," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 48, no. 3, pp. 2690–2698, Jul. 2012.

[LLM08] J. Liu, X. Liu, and X. Ma, "First-order perturbation analysis of singular vectors in singular value decomposition," *IEEE Transactions on Signal Processing*, vol. 56, no. 7, pp. 3044–3049, Jul. 2008.

[LLV93] F. Li, H. Liu, and R. J. Vaccaro, "Performance analysis for DOA estimation algorithms: Unification, simplifications, and observations," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 29, no. 4, pp. 1170–1184, Oct. 1993.

[LLXZ12] A. Liu, G. Liao, Q. Xu, and C. Zeng, "A circularity-based DOA estimation method under coexistence of noncircular and circular signals," in *Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, Kyoto, Japan, Mar. 2012.

[LM15] S. Lee and X. Ma, "Eigenvector-based initial ranging process for OFDMA uplink systems," *EURASIP Journal on Advances in Signal Processing*, vol. 2015:1, Jan. 2015.

[LRL98] Y. Li, J. Razavilar, and K. J. R. Liu, "A high-resolution technique for multidimensional NMR spectroscopy," *IEEE Transactions on Biomedical Engineering*, vol. 45, no. 1, pp. 78–86, Jan. 1998.

[LSJ05] X. Liu, N. D. Sidiropoulos, and T. Jiang, "Multidimensional harmonic retrieval with applications in MIMO wireless channel sounding," in *Space-Time Processing for MIMO Communications*. New York: Wiley, 2005, pp. 41–75, a. Gershman and N. Sidiropoulos, Eds.

[LSS02] X. Liu, N. D. Sidiropoulos, and A. Swami, "Blind high-resolution localization and tracking of multiple frequency hopped signals," *IEEE Transactions on Signal Processing*, vol. 50, no. 4, pp. 889–901, Apr. 2002.

[LT78] P. Lancaster and M. Tismentsky, *The Theory of Matrices*, 2nd ed. New York Acadmic Press, 1978.

[Lüt96] H. Lütkepohl, *Handbook of Matrices*. John Wiley and Sons, 1996.

[LvdVD03] A. N. Lemma, A.-J. van der Veen, and E. F. Deprettere, "Analysis of joint angle-frequency estimation using ESPRIT," *IEEE Transactions on Signal Processing*, vol. 51, no. 5, pp. 1264–1283, May 2003.

[MCW05] D. Malioutov, M. Cetin, and A. S. Willsky, "A sparse signal reconstruction perspective for source localization with sensor arrays," *IEEE Transactions on Signal Processing*, vol. 53, no. 8, pp. 3010–3022, Aug. 2005.

[ME11] M. Mishali and Y. C. Eldar, "Sub-nyquist sampling: Bridging theory and practice," *IEEE Signal Processing Magazine*, vol. 28, no. 6, pp. 98–124, Nov. 2011.

[MHZ96] C. P. Mathews, M. Haardt, and M. D. Zoltowski, "Performance analysis of closed-form, ESPRIT based 2-D angle estimator for rectangular arrays," *IEEE Signal Processing Letters*, vol. 3, no. 4, pp. 124–126, Apr. 1996.

[MLM05] S. Miron, N. Le Bihan, and J. I. Mars, "Vector-sensor MUSIC for polarized seismic sources localization," *EURASIP Journal on Applied Signal Processing*, vol. 10, pp. 74–84, 2005.

[MN95] J. R. Magnus and H. Neudecker, *Matrix differential calculus with applications in statistics and econometrics*. John Wiley and Sons, 1995.

[Moo20] E. H. Moore, "On the reciprocal of the general algebraic matrix," *Bulletin of the American Mathematical Society*, vol. 26, pp. 394–395, 1920.

[MSPM04] K. N. Mokios, N. D. Sidiropoulos, M. Pesavento, and C. E. Mecklenbräuker, "On 3-D harmonic retrieval for wireless channel sounding," in *Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, vol. 2, Toulouse, Canada, May 2004, pp. 89–92.

[MZ93] S. Mallat and Z. Zhang, "Matching pursuits with time-frequency dictionaries," *IEEE Transactions on Signal Processing*, vol. 41, no. 12, pp. 3397–3415, Dec. 1993.

[MZ06] D. Model and M. Zibulevsky, "Signal reconstruction in sensor arrays using sparse representations," *Signal Processing*, vol. 86, no. 3, pp. 624–638, Mar. 2006.

[NM93] F. D. Neeser and J. L. Massey, "Proper complex random processes with applications to information theory," *IEEE Transactions on Information Theory*, vol. 39, no. 4, pp. 1293–1302, Jul. 1993.

[NOH+07] T. Nara, J. Oohama, M. Hashimoto, T. Takeda, and S. Ando, "Direct reconstruction algorithm of current dipoles for vector magnetoencephalography and electroencephalography," *Physics in Medicine and Biology*, vol. 52, pp. 3859–3879, Jun. 2007.

[NS10] D. Nion and N. D. Sidiropoulos, "Tensor algebra and multidimensional harmonic retrieval in signal processing for MIMO radar," *IEEE Transactions on Signal Processing*, vol. 58, no. 11, pp. 5693–5705, Nov. 2010.

[NSA06] T. Nara, S. Suzuki, and S. Ando, "A closed-form formula for magnetic dipole localization by measurement of its magnetic field and spatial gradients," *IEEE Transactions on Magnetics*, vol. 42, no. 10, pp. 3291–3293, Oct. 2006.

[Oll08] E. Ollila, "On the circularity of a complex random variable," *IEEE Signal Processing Letters*, vol. 15, pp. 841–844, 2008.

[OVK91] B. Ottersten, M. Viberg, and T. Kailath, "Performance analysis of the Total Least Squares ESPRIT algorithm," *IEEE Transactions on Signal Processing*, vol. 39, no. 5, pp. 1122–1135, May 1991.

[PB97] B. Picinbono and P. Bondon, "Second-order statistics of complex signals," *IEEE Transactions on Signal Processing*, vol. 45, no. 2, pp. 411–420, Feb. 1997.

[PC95] B. Picinbono and P. Chevalier, "Widely linear estimation with complex data," *IEEE Transactions on Signal Processing*, vol. 43, no. 8, Aug. 1995.

[Pen55] R. Penrose, "A generalized inverse for matrices," *Proceedings of the Cambridge Philosophical Society*, vol. 51, pp. 406–413, 1955.

[PF88] B. Porat and B. Friedlander, "Analysis of the asymptotic relative efficiency of the MUSIC algorithm," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 36, no. 4, pp. 532–544, Apr. 1988.

[PGH00] M. Pesavento, A. B. Gershman, and M. Haardt, "Unitary Root-MUSIC with a real-valued eigendecomposition: A theoretical and experimental performance study," *IEEE Transactions on Signal Processing*, vol. 48, no. 5, pp. 1306–1314, May 2000.

[PGW02] M. Pesavento, A. B. Gershman, and K. M. Wong, "Direction finding in partly calibrated sensor arrays composed of multiple subarrays," *IEEE Transactions on Signal Processing*, vol. 50, no. 9, pp. 2103–2115, Sep. 2002.

[Pic94] B. Picinbono, "On circularity," *IEEE Transactions on Signal Processing*, vol. 42, no. 12, pp. 3473–3482, Dec. 1994.

[Pic96] ——, "Second-order complex random vectors and normal distributions," *IEEE Transactions on Signal Processing*, vol. 44, no. 10, Oct. 1996.

[Pis73] V. F. Pisarenko, "The retrieval of harmonics from a covariance function," *Geophys. J. Royal Astronomical Soc.*, vol. 33, pp. 347–366, 1973.

[PK89a] S. U. Pillai and B. H. Kwon, "Forward/backward spatial smoothing techniques for coherent signal identification," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 37, no. 1, pp. 8–15, Jan. 1989.

[PK89b] ——, "Performance analysis of MUSIC-type high resolution estimators for direction finding in correlated and coherent scenes," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 37, no. 8, pp. 1176–1189, Aug. 1989.

[PMB04] M. Pesavento, C. F. Mecklenbräuker, and J. F. Böhme, "Multi-dimensional rank reduction estimator for parametric MIMO channel estimation," *EURASIP J. Appl. Signal Proc.*, pp. 1354–1363, Aug. 2004.

[PP08] K. B. Petersen and M. S. Pedersen, "The matrix cookbook," http://matrixcookbook.com (until 2010), Nov. 2008.

[PRK93] Y. C. Pati, R. Rezaiifar, and P. S. Krishnaprasad, "Orthogonal matching pursuit: Recursive function approximation with applications to wavelet decomposition," in *Proc. Asilomar Conference on Signals, Systems, and Computers*, Pacific Grove, CA, Nov. 1993.

[PV10] P. Pal and P. P. Vaidyanathan, "Nested arrays: A novel approach to array processing with enhanced degrees of freedom," *IEEE Transactions on Signal Processing*, vol. 58, no. 8, pp. 4167–4181, Aug. 2010.

[QHS+15] C. Qian, L. Huang, H. C. So, N. D. Sidiropoulos, and J. Xie, "Unitary PUMA algorithm for estimating the frequency of a complex sinusoid," *IEEE Transactions on Signal Processing*, vol. 63, no. 20, pp. 5358–5368, Oct. 2015.

[QHSS16] C. Qian, L. Huang, N. Sidiropoulos, and H. C. So, "Enhanced PUMA for direction-of-arrival estimation and its performance analysis," *IEEE Transactions on Signal Processing*, vol. 64, no. 16, pp. 4127–4137, Aug. 2016.

[RA92] B. D. Rao and K. S. Arun, "Model based processing of signals: A state space approach," *Proceedings of the IEEE*, vol. 80, pp. 283–309, Feb. 1992.

[RBH10] F. Roemer, H. Becker, and M. Haardt, "Analytical performance analysis for multi-dimensional Tensor-ESPRIT-type parameter estimation algorithms," in *Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, Dallas, TX, Mar. 2010.

[RBHW09] F. Roemer, H. Becker, M. Haardt, and M. Weis, "Analytical performance evaluation for HOSVD-based parameter estimation schemes," in *Proc. IEEE International Workshop on Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP)*, Aruba, Dutch Antilles, Dec. 2009.

[RFP10] B. Recht, M. Fazel, and P. A. Parrilo, "Guaranteed minimum-rank solutions of linear matrix equations via nuclear norm minimization," *SIAM Review*, vol. 52, no. 3, pp. 471–501, Aug. 2010.

[RG09] M. Rübsamen and A. Gershman, "Direction-of-arrival estimation for non-uniform sensor arrays: from manifold separation to Fourier domain MUSIC methods," *IEEE Transactions on Signal Processing*, vol. 57, no. 2, pp. 588–599, Feb. 2009.

[RH89a] B. D. Rao and K. V. S. Hari, "Performance analysis of ESPRIT and TAM in determining the direction of arrival of plane waves in noise," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 37, no. 12, pp. 1990–1995, Dec. 1989.

[RH89b] ——, "Performance analysis of Root-MUSIC," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 37, no. 12, pp. 1939–1948, Dec. 1989.

[RH90] ——, "Effect of spatial smoothing on the performance of MUSIC and minimum-norm method," *Proc. Inst. Elect. Eng. F*, vol. 137, pp. 449–458, Dec. 1990.

[RH93] ——, "Weighted subspace methods and spatial smoothing: Analysis and comparison," *IEEE Transactions on Signal Processing*, vol. 41, no. 2, pp. 788–803, Feb. 1993.

[RH07a] F. Roemer and M. Haardt, "Deterministic Cramér-Rao bounds for strict sense non-circular sources," in *Proc. International ITG Workshop on Smart Antennas (WSA)*, Vienna, Austria, Feb. 2007.

[RH07b] ——, "Tensor-structure structured least squares (TS-SLS) to improve the performance of multi-dimensional ESPRIT-type algorithms," in *Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, Honolulu, HI, Apr. 2007, pp. 893–896.

[RH09] ——, "Multidimensional Unitary Tensor-ESPRIT for non-circular sources," in *Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, Taipei, Taiwan, Apr. 2009.

[RH11] ——, "Analytical performance assessment of 1-D structured least squares," in *Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, Prague, Czech Republic, May 2011.

[RH12] ——, "A framework for the analytical performance assessment of matrix and tensor-based ESPRIT-type algorithms," *pre-print*, Sep. 2012, arXiv:1209.3253.

[RHD06] F. Roemer, M. Haardt, and G. Del Galdo, "Higher order SVD based subspace estimation to improve multi-dimensional parameter estimation algorithms," in *Proc. Asilomar Conference on Signals, Systems, and Computers*, Pacific Grove, CA, Nov. 2006, pp. 961–965.

[RHD14] ——, "Analytical performance assessment of multi-dimensional matrix- and tensor-based ESPRIT-type algorithms," *IEEE Transactions on Signal Processing*, vol. 62, no. 10, pp. 2611–2625, May 2014.

[RHST01] A. Richter, D. Hampicke, G. Sommerkorn, and R. S. Thomä, "MIMO measurement and joint M-D parameter estimation of mobile radio channels," in *Proc. IEEE Vehicular Technology Conference (VTC Spring 2001)*, Rhodes, Greece, May 2001.

[RK87] R. Roy and T. Kailath, "Total least-squares ESPRIT," in *Proc. Asilomar Conference on Signals, Systems, and Computers*, Pacific Grove, CA, Nov. 1987.

[RK89] R. H. Roy and T. Kailath, "ESPRIT-estimation of signal parameters via rotational invariance techniques," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 37, no. 7, pp. 984–995, Jul. 1989.

[Roe13] F. Roemer, "Advanced algebraic concepts for efficient multi-channel signal processing," PhD thesis, Ilmenau University of Technology, Communications Research Laboratory, Germany, Jan. 2013.

[RPK86] R. H. Roy, A. Paulraj, and T. Kailath, "ESPRIT - a subspace rotation approach to estimation of parameters of cisoids in noise," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. AASP-34, no. 5, pp. 1340–1342, Oct. 1986.

[RZZ93] C. R. Rao, L. C. Zhao, and B. Zhou, "A novel algorithm for 2-dimensional frequency estimation," in *Proc. 27th Asilomar Conf. Circuits, Syst., Computing*, Pacific Grove, CA, Nov. 1993, pp. 199–202.

[SBG00] N. D. Sidiropoulos, R. Bro, and G. B. Giannakis, "Parallel factor analysis in sensor array processing," *IEEE Transactions on Signal Processing*, vol. 48, no. 8, pp. 2377–2388, Aug. 2000.

[SBL11] P. Stoica, P. Babu, and J. Li, "SPICE: A sparse covariance-based estimation method for array processing," *IEEE Transactions on Signal Processing*, vol. 59, no. 2, pp. 629–638, Feb. 2011.

[Sch79] R. O. Schmidt, "Multiple emitter location and signal parameter estimation," in *in Proc. RADC Spectrum Estimation Workshop*, Oct. 1979.

[Sch86] ——, "Multiple emitter location and signal parameter estimation," *IEEE Transactions on Antennas and Propagation*, vol. AP-34, no. 3, pp. 243–258, Mar. 1986.

[Sch91] L. L. Scharf, *Statistical Signal Processing.* Reading, MA: Addison-Wesley Publishing Company, 1991.

[SCLC10] H. C. So, F. K. W. Chan, W. H. Lau, and C. F. Chan, "An efficient approach for two-dimensional parameter estimation of a single-tone," *IEEE Transactions on Signal Processing*, vol. 58, no. 4, pp. 1999–2009, Apr. 2010.

[SdHW12] N. Song, R. C. de Lamare, M. Haardt, and M. Wolf, "Adaptive widely linear reduced-rank interference suppression based on the multi-stage Wiener filter," *IEEE Transactions on Signal Processing*, vol. 60, no. 8, pp. 4003–4016, Aug. 2012.

[SK93] A. L. Swindlehurst and T. Kailath, "Azimuth/elevation direction finding using regular array geometries," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 29, no. 1, pp. 145–156, Jan. 1993.

[SLG01] P. Stoica, E. G. Larsson, and A. B. Gershman, "The stochastic CRB for array processing: A textbook derivation," *IEEE Signal Processing Letters*, vol. 8, no. 5, pp. 148–150, May 2001.

[SM05] P. Stoica and R. L. Moses, *Spectral Analysis of Signals, 1st Edition.* Upper Saddle River, NJ: Prentice-Hall, 2005.

[SMB01] M. Steinbauer, A. F. Molisch, and E. Bonek, "The double-directional radio channel," *IEEE Transactions on Antennas and Propagation*, vol. 43, no. 4, Aug. 2001.

[SN89] P. Stoica and A. Nehorai, "MUSIC, maximum likelihood, and Cramér-Rao bound," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 37, no. 5, pp. 720–741, May 1989.

[SN90] ——, "Performance study of conditional and unconditional direction of arrival estimation," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 38, no. 10, pp. 1783–1795, Oct. 1990.

[SN91] ——, "Performance comparison of subspace rotation and MUSIC methods for direction estimation," *IEEE Transactions on Signal Processing*, vol. 39, no. 2, pp. 446–453, Feb. 1991.

[SORK92] A. L. Swindlehurst, B. Ottersten, R. Roy, and T. Kailath, "Multiple invariance ESPRIT," *IEEE Transactions on Signal Processing*, vol. 40, no. 4, pp. 867–881, Apr. 1992.

[SPP14] C. Steffens, P. Parvazi, and M. Pesavento, "Direction finding and array calibration based on sparse reconstruction in partly calibrated arrays," in *Proc. IEEE Sensor Array and Multichannel Signal Processing Workshop (SAM)*, A Coruna, Spain, Jun. 2014.

[SPP17] C. Steffens, M. Pesavento, and M. Pfetsch, "A compact formulation for the $\ell_{2,1}$ mixed-norm minimization problem," in *Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, New Orleans, LA, Mar. 2017.

[SPP18a] ——, "Block- and rank-sparse recovery for direction finding in partly calibrated arrays," *IEEE Transactions on Signal Processing*, vol. 66, no. 2, pp. 384–399, Jan. 2018.

[SPP18b] ——, "A compact formulation for the $\ell_{2,1}$ mixed-norm minimization problem," *IEEE Transactions on Signal Processing*, vol. 66, no. 6, pp. 1483–1497, Mar. 2018.

[SS91] P. Stoica and T. Söderström, "Statistical analysis of MUSIC and subspace rotation estimates of sinusoidal frequencies," *IEEE Transactions on Signal Processing*, vol. 39, no. 8, Aug. 1991.

[SS03] P. J. Schreier and L. L. Scharf, "Second-order analysis of improper complex random vectors and processes," *IEEE Transactions on Signal Processing*, vol. 51, no. 3, pp. 714–725, Mar. 2003.

[SS05] ——, "Detection and estimation of improper complex random signals," *IEEE Transactions on Information Theory*, vol. 51, no. 1, pp. 306–312, Jan. 2005.

[SS10] ——, *Statistical Signal Processing of Complex-Valued Data: The Theory of Improper and Noncircular Signals*. Cambridge University Press, 2010.

[SS12] W. Sun and H. C. So, "Accurate and computationally efficient tensor-based subspace approach for multi-dimensional harmonic retrieval," *IEEE Transactions on Signal Processing*, vol. 60, no. 10, pp. 5077–5088, Oct. 2012.

[Ste07] F. Sterle, "Widely linear MMSE transceivers for MIMO channels," *IEEE Transactions on Signal Processing*, vol. 55, no. 8, pp. 4258–427, Aug. 2007.

[SV15] M. Shaghaghi and S. A. Vorobyov, "Subspace leakage analysis and improved DOA estimation with small sample size," *IEEE Transactions on Signal Processing*, vol. 63, no. 12, pp. 3251–3265, Jun. 2015.

[SW01] M. K. Schneider and A. S. Willsky, "Krylov subspace estimation," *SIAM Journal on Scientific Computing*, vol. 22, no. 5, pp. 1840–1864, 2001.

[SWK85] T. J. Shan, M. Way, and T. Kailath, "On spatial smoothing for direction-of-arrival estimation of coherent signals," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. AASP-33, no. 4, pp. 806–811, Aug. 1985.

[TBR15] G. Tang, B. N. Bhaskar, and B. Recht, "Near minimax line spectral estimation," *IEEE Transactions on Information Theory*, vol. 61, no. 1, pp. 499–512, Jan. 2015.

[TBSR13] G. Tang, B. N. Bhaskar, P. Shah, and B. Recht, "Compressed sensing off the grid," *IEEE Transactions on Information Theory*, vol. 59, no. 11, pp. 7465–7490, Nov. 2013.

[TGS06] J. A. Tropp, A. C. Gilbert, and M. J. Strauss, "Algorithms for simultaneous sparse approximation. Part I: Greedy pursuit," *Signal Processing*, vol. 86, no. 3, pp. 572–588, Mar. 2006.

[THG09a] A. Thakre, M. Haardt, and K. Giridhar, "Single snapshot R-D Unitary ESPRIT using an augmentation of the tensor order," in *Proc. IEEE International Workshop on Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP)*, Dec. 2009.

[THG09b] ——, "Single snapshot spatial smoothing with improved effective array aperture," *IEEE Signal Processing Letters*, vol. 16, no. 6, pp. 505–509, Jun. 2009.

[THRG10] A. Thakre, M. Haardt, F. Roemer, and K. Giridhar, "Tensor-based spatial smoothing (TB-SS) using multiple snapshots," *IEEE Transactions on Signal Processing*, vol. 25, no. 5, pp. 2715–2728, May 2010.

[Tib96] R. Tibshirani, "Regression shrinkage and selection via the LASSO," *Journal of the Royal Statistical Society. Series B (Methodological)*, vol. 58, no. 1, pp. 267–288, 1996.

[TV05] D. Tse and P. Viswanath, *Fundamentals of wireless communications.* Cambridge University Press, 2005.

[Van02] H. L. Van Trees, *Optimum Array Processing: Detection, Estimation, and Modulation Theory, Part IV, Edition I.* New York: Wiley, 2002.

[VB96] L. Vandenberghe and S. Boyd, "Semidefinite programming," *SIAM review*, vol. 38, no. 1, pp. 49–95, Mar. 1996.

[VOK91] M. Viberg, B. Ottersten, and T. Kailath, "Detection and estimation in sensor arrays using weighted subspace fitting," *IEEE Transactions on Signal Processing*, vol. 39, no. 11, pp. 2436– 2449, Sep. 1991.

[VP11] P. P. Vaidyanathan and P. Pal, "Sparse sensing with co-prime samplers and arrays," *IEEE Transactions on Signal Processing*, vol. 59, no. 2, pp. 573–586, Feb. 2011.

[WF93] A. J. Weiss and B. Friedlander, "Performance analysis of spatial smoothing with interpolated data," *IEEE Transactions on Signal Processing*, vol. 41, no. 5, pp. 1881–1892, May 1993.

[WK85] M. Wax and T. Kailath, "Detection of signals by information theoretic criteria," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 33, no. 3, pp. 387–392, Apr. 1985.

[WS15] F. Wen and H. C. So, "Tensor MODE for multi-dimensional harmonic retrieval with coherent sources," *Signal Processing*, vol. 108, no. 3, pp. 530–534, Mar. 2015.

[XRK94] G. Xu, R. H. Roy, and T. Kailath, "Detection of number of sources via exploitation of centro-symmetry property," *IEEE Transactions on Signal Processing*, vol. 42, no. 1, pp. 102–112, Jan. 1994.

[Xu02] Z. Xu, "Perturbation analysis for subspace decomposition with applications in subspace-based algorithms," *IEEE Transactions on Signal Processing*, vol. 50, no. 11, pp. 2820–2830, Nov. 2002.

[Yan95] B. Yang, "Projection approximation subspace tracking," *IEEE Transactions on Signal Processing*, vol. 43, no. 1, pp. 95–107, Jan. 1995.

[YLSX17] Z. Yang, J. Li, P. Stoica, and L. Xie, "Sparse methods for direction-of-arrival estimation," in *Academic Press Library in Signal Processing - Array, Radar and Communications Engineering*, S. Theodoridis and R. Chellappa, Eds., vol. 7. Eds. Academic Press, 2017, chapter 11.

[YLZ15] X. Yang, G. Li, and Z. Zheng, "DOA estimation of noncircular signal based on sparse representation," *Springer Wireless Personal Communications*, vol. 82, no. 4, pp. 2363–2375, Jun. 2015.

[YX15] Z. Yang and L. Xie, "On gridless sparse methods for line spectral estimation from complete and incomplete data," *IEEE Transactions on Signal Processing*, vol. 63, no. 12, pp. 3139–3153, Jun. 2015.

[YX16a] ——, "Enhancing sparsity and resolution via reweighted atomic norm minimization," *IEEE Transactions on Signal Processing*, vol. 64, no. 4, pp. 995–1006, Feb. 2016.

[YX16b] ——, "Exact joint sparse frequency recovery via optimization methods," *IEEE Transactions on Signal Processing*, vol. 64, no. 19, pp. 5145–5157, Oct. 2016.

[YXS16] Z. Yang, L. Xie, and P. Stoica, "Vandermonde decomposition of multilevel toeplitz matrices with application to multidimensional super-resolution," *IEEE Transactions on Information Theory*, vol. 62, no. 6, pp. 3685–3701, Jun. 2016.

[YXZ13] Z. Yang, L. Xie, and C. Zhang, "Off-grid direction of arrival estimation using sparse bayesian inference," *IEEE Transactions on Signal Processing*, vol. 61, no. 1, pp. 38–43, Jan. 2013.

[ZCW03] A. Zoubir, P. Chargé, and Y. Wang, "Non circular sources localization with ESPRIT," in *Proc. European Conference on Wireless Technology (ECWT 2003)*, Munich, Germany, Oct. 2003.

[ZFDW00] T. Zwick, C. Fischer, D. Didascalou, and W. Wiesbeck, "A stochastic spatial channel model based on wave-propagation modeling," *IEEE Journal on Selected Areas in Communications*, vol. 18, no. 1, pp. 6–15, Jan. 2000.

[ZHM96] M. D. Zoltowski, M. Haardt, and C. P. Mathews, "Closed-form 2-D angle estimation with rectangular arrays in element space or beamspace via Unitary ESPRIT," *IEEE Transactions on Signal Processing*, vol. 44, no. 2, pp. 316–328, Feb. 1996.

[ZHM+00] T. Zwick, D. Hampicke, J. Maurer, A. Richter, G. Sommerkorn, R. Thomä, and W. Wiesbeck, "Results of double-directional channel sounding measurements," in *Proc. IEEE Vehicular Technology Conference (VTC Spring 2000)*, Tokyo, Japan, May 2000, pp. 2497–2501.

[ZSJ17] D. Zachariah, P. Stoica, and M. Jansson, "Comments on "Enhanced PUMA for direction-of-arrival estimation and its performance analysis"," *IEEE Transactions on Signal Processing*, vol. 65, no. 22, pp. 6113–6114, Nov. 2017.

Erklärung

Ich versichere, dass ich die vorliegende Arbeit ohne unzulässige Hilfe Dritter und ohne Benutzung anderer als der angegebenen Hilfsmittel angefertigt habe. Die aus anderen Quellen direkt oder indirekt übernommenen Daten und Konzepte sind unter Angabe der Quelle gekennzeichnet.

Bei der Auswahl und Auswertung folgenden Materials haben mir die nachstehend aufgeführten Personen in der jeweils beschriebenen Weise entgeltlich/unentgeltlich geholfen:

1. .....
2. .....
3. .....

Weitere Personen waren an der inhaltlich-materiellen Erstellung der vorliegenden Arbeit nicht beteiligt. Insbesondere habe ich hierfür nicht die entgeltliche Hilfe von Vermittlungs- bzw. Beratungsdiensten (Promotionsberater oder anderer Personen) in Anspruch genommen. Niemand hat von mir unmittelbar oder mittelbar geldwerte Leistungen für Arbeiten erhalten, die im Zusammenhang mit dem Inhalte der vorgelegten Dissertation stehen.

Die Arbeit wurde bisher weder im In- noch im Ausland in gleicher oder ähnlicher Form einer Prüfungsbehörde vorgelegt.

Ich bin darauf hingewiesen worden, dass die Unrichtigkeit dieser Erklärung als Täuschungsversuch bewertet wird und gemäß 7 Abs. 10 der Promotionsordnung den Abbruch des Promotionsverfahrens zur Folge hat.

(Ort, Datum) (Unterschrift)