



FRIEDRICH-SCHILLER-
UNIVERSITÄT
JENA

Improved correction methods for symmetry-free systems

Dissertation

for the acquisition of the academic title

Doctor Rerum Naturalium (Dr. rer. nat.)

submitted to

the Council of the Faculty of Physics and Astronomy

of Friedrich-Schiller-Universität Jena

By M. Sc. Ziyao Tang

born in Harbin, Heilongjiang Province, China on 21.02.1994

Gutachter:

1. Prof. Dr. Herbert Gross, Friedrich-Schiller-Universität Jena
2. Prof. Dr. Alois Herkommer, Universität Stuttgart
3. Prof. Dr. Rongguang Liang, University of Arizona

Date of the Disputation: 01. 09. 2022

Zusammenfassung

Da moderne optische Systeme heutzutage oft kompaktere und leistungsfähigere Abbildungssysteme erfordern, werden nicht-sphärische Oberflächen immer häufiger in optischen Design-Prozessen eingesetzt. Mit einer relativ kompakten Struktur können solche Systeme eine gute Aberrationskorrektur für große numerische Aperturen und Sichtfelder realisieren. Die steigende Nachfrage nach solchen Systemen bringt die Notwendigkeit nach fortschrittlicheren Methoden für die optische Auslegung mit sich.

Um in der Praxis eine sehr gute Abbildungsqualität von Systemen zu erreichen, sind die Bewertung der Aberration sowie die Sensitivitätsanalyse von großer Bedeutung. Insbesondere für symmetriefreie optische Systeme wird eine umfassende Methode zur Bewertung der Abbildungsleistung während der Korrektur angestrebt. Daher liegt ein Schwerpunkt dieser Arbeit auf einer neuen quantitativen Analysemethode für transversale Aberration in symmetriefreien Systemen, die eine Zerlegung in die Anteile der optischen Flächen erlaubt. Basierend auf einer gemischten paraxialen/realen Ray-Tracing-Berechnung ermöglicht die Methode die Berechnung der Bildfehler beliebig hoher Ordnung sowie eine Auftrennung in intrinsische und induzierte Anteile. Dabei unterstützt die Methode die Darstellung der Aberrationen in Zernikepolynome. Diese neue Methode ermöglicht es, die Abbildungsqualität unter Berücksichtigung relativ kritischer Oberflächen in einem beliebigen nicht-symmetrischen System zu bewerten.

Der am weitesten verbreitete Optimierungsalgorithmus zur optischen Auslegung, die konventionelle Methode der kleinsten Fehlerquadrate, bietet eine schnelle Konvergenz und einen deterministischen Optimierungspfad. Neben diesen Vorteilen hat dieser Algorithmus jedoch Schwächen bei komplizierten Topologien der Gütefunktion mit vielen Variablen und erlaubt keine globale Optimierung oder Strukturänderungen. Daher zielt das zweite Thema dieser Arbeit darauf ab, die Fähigkeit des Algorithmus zur Suche globaler Lösungen zu verbessern. Zu diesem Zweck wird ein biologisch-inspirierter Algorithmus verwendet und erweitert, welcher auf dem sogenannten Ameisenalgorithmus (Ant Colony Optimization) basiert. Im Verlauf dieser Arbeit wird nachgewiesen, dass ein auf physikalischen Erkenntnissen und Erfahrungen beruhender Algorithmus dazu geeignet ist, unabhängig vom Startdesign nicht zu komplexe Optimierungsprobleme mittels sinnvoller struktureller Änderungen zu lösen. Als Ergebnis liefert der Algorithmus dem Benutzer eine Vielzahl möglicher Lösungen mit unterschiedlichen Strukturen, so dass der Designer die

Möglichkeit hat, die besten Lösungen für den jeweiligen Zweck auszuwählen. Darüber hinaus erzeugt der Algorithmus auch bei der Bearbeitung von hochdimensionalen Optimierungsaufgaben zufriedenstellende Ergebnisse. Das Verfahren basiert auf einem Wahrscheinlichkeitsprinzip und hat die Fähigkeit zur Suche des globalen Optimums von komplizierten Systemen mit nicht-sphärischen Oberflächen/Geometrien und geknickten optischen Achsen. Die erzielten Ergebnisse deuten darauf hin, dass die Kombination eines Optimierungsalgorithmus und physikalischen Grundprinzipien der Korrektur großes Potenzial für zukünftige intelligente Designmethoden optischer Systeme mit hohem Automatisierungsgrad bietet.

Abstract

As the modern optical setups nowadays often require a more compact and well-performed imaging system, non-spherical surfaces are more and more applied in optical design tasks. With a relatively compact structure, such systems can realize a good aberration correction for a large numerical aperture or a wide field of view. Consequently, the increasing demand for such systems brings the request for more advanced optical design methods.

For successful optical design, the aberration assessment, as well as the sensitivity analysis, is of great importance. Particularly for symmetry-free optical systems, a comprehensive imaging performance evaluation method is desired during the aberration correction process. Therefore, one major topic of the work focuses on the quantitative analysis method for surface-decomposed transverse aberration in symmetry-free systems is proposed. Based on a mixed paraxial/real ray-tracing calculation, the method can be applied for the calculation of full-order total, intrinsic, and induced aberration. In addition, the method supports surface-additive Zernike coefficient representation for the assessment of specific aberrations. The implementations of this novel method help to assess the correction performance considering the relatively critical surfaces in an arbitrary system.

Besides, as the most widely used optimization algorithm in optical design, the conventional damped least square method is advantageous with fast convergence and its deterministic optimization path, but weak in complex merit function topologies in case of many variables and not able to perform global optimization or structural changes. Thus, the second topic of this work aims to enhance the global searching ability of the optimization algorithm. A bio-inspired algorithm based on the ant colony optimization is used and extended for this purpose. Guided by physical knowledge, the algorithm is proved feasible to solve simple optimization problems with proper structural changes, regardless of the initial design. The algorithm outputs a large solution database so that the user can gain an overview of the optional solutions with various structures and out select the best fitting ones according to the specific purpose. In addition, the algorithm also provides satisfactory results when dealing with high-dimensional optimization tasks. The strong global searching ability based on the probabilistic feature supports the optimization of complicated systems with non-spherical surfaces. The obtained results indicate that the combination of optimization algorithm and physical considerations is of great potential in optical system design with a high level of automation for the future outlook.

This page is left blank intentionally.

Contents

Zusammenfassung	i
Abstract	iii
Contents	1
1. Introduction and motivation.....	4
2. State of the art	8
2.1 Fundamentals of optics	8
2.2 Matrix calculus for paraxial imaging.....	10
2.3 Traditional aberration analysis methods	11
2.3.1 Seidel aberration theory	12
2.3.2 Wave aberrations with Zernike Fringe coefficient representation.....	13
2.3.3 Intrinsic and induced aberration	15
2.3.4 Aldis theory.....	15
2.3.5 Surface-decomposed aberration with the phase space method.....	16
2.4 General optical design procedure	20
2.5 Conventional optimization method.....	21
2.6 Asphere surfaces in optical design.....	23
2.6.1 Aspherical surface descriptions	23
2.6.2 Performance enhancement with non-spherical surfaces	24
2.7 Bio-inspired global optimization methods.....	26
2.8 Fundamental of ACOR algorithm for optical design.....	28
2.9 Problems in symmetry-free system design	30
3. New method for aberration analysis	33
3.1 Surface contribution of total transverse aberration.....	33
3.1.1 Additive surface contribution Δy	34
3.1.2 Chief ray referred surface contribution ΔY	36
3.1.3 Discussion: Δy_{Img} , ΔY_{Img} , and spot diagram	36
3.2 Surface-decomposed intrinsic and induced aberration	37
3.2.1 Calculation method	37
3.2.2 Approximation in the calculation	38
3.3 Surface-decomposed Zernike coefficient representation.....	40
3.4 Discussion.....	43
4 Improved global optimization algorithm	44
4.1 GACOR global optimization with structural changes	47

4.1.1 General workflow.....	47
4.1.2 Global exploration.....	49
4.1.3 Local exploration	53
4.2 GACOR algorithm for final improvement phase.....	56
4.2.1 General final improvement strategy of the optical designer	57
4.2.2 Extended final improvement method of the GACOR algorithm	58
5 Examples and applications.....	61
5.1 Comprehensive aberration analysis with the MRT method.....	61
5.2 Quasi-automatic global optimization	65
5.2.1 Optimization strategy for retro-focus systems	67
5.2.2 Local exploration for one ant group in one main iteration.....	70
5.2.3 Solution evolution	73
5.2.4 Analysis of the output solutions.....	78
5.2.5 Successful solution analysis.....	80
5.2.6 Discussion	89
5.3 Freeform system optimization.....	90
5.3.1 Final improvement of an anamorphic system	90
5.3.2 Successful solution analysis.....	93
5.3.3 Discussion	97
6 Conclusion and outlook	98
Appendix A: Verification of the MRT method.....	100
A.1 Transverse aberration calculation results	100
A.2 Intrinsic/induced aberration calculation results.....	101
A.3 Surface-additive Zernike coefficient fitting	103
Appendix B: Further discussion of the MRT method.....	106
B.1 Approximation of intrinsic/induced aberration calculation.....	106
B.2 Distortion removal.....	109
Appendix C: Parameterization of the GACOR algorithm	112
Appendix D: Searching for the best structural change.....	114
D.1 Choice of the structural change option.....	114
D.2 Decision of the structural change surface	114
D.3 Lens splitting option.....	116
D.4 Aspherization	117
Appendix E: Switch from RI to RII	119
Appendix F: Variable and MF adaption.....	121

F.1 MF adaption in various cases	121
F.2 Boundary condition control	123
Appendix G: Similarity and lens shape check	125
Appendix H: ACOR local search	126
H.1 Artificial initial deviation	126
H.2 Prevention of infinite loops	128
Appendix I: Further optimization examples	129
I.1 Tele-system optimization	129
I.2 High NA collimator system optimization	132
Appendix J: Freeform surfaces for distortion correction	134
J.1 Distortion correction of spectrometer systems	134
J.2 Example: Modified Offner system optimization with freeforms	136
References	141
List of figures	146
List of tables	151
List of symbols	152
List of abbreviations	158
Acknowledgment	159
Ehrenwörtliche Erklärung	160
Publications	161

1. Introduction and motivation

Since several centuries, optical systems play a crucial role in a broad field of applications. From Abbe's microscopes to the modern astronomical telescopes, the topic of optical system design has been developed rapidly and the corresponding research is always of great interest to physicists. With the highly advanced computational assistance, the essential tool of optical system design – ray tracing – no longer impedes the design and analysis of complicated optical systems. In addition, the modern diamond turning technology with high accuracy supports the application of non-spherical surfaces in the optical system. The large degree of freedom brought by such surfaces is greatly beneficial for aberration correction compared to conventional spherical surfaces. Such systems can realize superior imaging performance for a large field of view with a more compact structure. However, the higher complexity of the modern optical systems with non-spherical surfaces also generates higher demands on the system design task [1].

Concerning the theoretical development process of an optical system, there are three important general research directions, namely the initial system design, optimization, and aberration analysis. Corresponding to the specific purposes, various methods were proposed to support an efficient optical system design systematically in the past decades. Due to the large variety of possible setups, the complexity of optical design is strongly dependent on the system structure. Generally, the existence of a common straight mechanical axis and surface shapes determines the rotational symmetry of the system, and the path of the optical axis ray (OAR) through the field and pupil center determines the symmetry of the ray bundle. These two kinds of symmetry have a great impact on the special features of the system during the design process.

As an overview, Figure 1.1 summarizes the current completeness level of the important research in the main directions presented with the cell color, in case of different system structure symmetry and OAR bending situations. As one of the first steps of optical design, establishing a meaningful initial system has been investigated and understood at the best level among the research directions mentioned in Fig.1.1. Especially for systems with a straight OAR, the geometrical tilt and decenter of the optical components are of no concern, making the initial design less challenging. However, due to the lack of understanding of aberrations and the large number of degrees of freedom, it is not realistic to explore a universal initial system design method for all the system types, particularly considering the

large geometrical variety of the symmetry-free systems. Thus, due to the relatively higher completeness, the initial system design topic is not the focus of this work [2-4].

Research directions	Important research	Straight OAR		Bent OAR	
		Rotational symmetric system structure	Non-symmetric system structure	Rotational symmetric system structure	Non-symmetric system structure
Initial structure design	First-order property				
	Tilt and decenter				
Optimization	Global searching	X	X		
	Automation	X	X		
Surface-decomposed aberration	Primary	X	X	X	X
	Higher-order	X	X	X	X
	Full-order	X	X	X	X
Correction with freeform	Resolution				X
	Distortion				X

Figure 1.1. An overview of the optical design tasks. Green-marked cells are those which are understood or solved completely. Yellow-marked cells correspond to those not completely understood or solved, and red-marked cells mean being far from satisfaction. 'X' marks the corresponding topic investigated and addressed in this work.

During the optical design process, aberration analysis plays a significant role in understanding the system. Particularly, a clear comparison among the surface contributions of aberrations is desirable for the understanding of the imaging performance of the system, as well as the corresponding sensitivity and ease of manufacturing. For the systems with rotational symmetry, the surface-resolved primary aberration analysis method based on the paraxial optics, such as Seidel theory, is applicable [5,6]. In comparison, for the symmetry-free systems, the assumption of paraxial optics is not valid anymore, leading to more challenges in the aberration assessment. The currently available methods, such as Nodal aberration theory (NAT), can solve the problem to some extent, but due to the complexity and limitations, the higher-order aberrations in the system cannot be well analyzed and are hard to control. Particularly for modern optical system design with non-spherical surfaces, the understanding and assessment of the higher-order aberrations are more critical, and the problems in finding out the optimal solutions are still not very well solved yet [7-9].

Furthermore, as the key to a successful design, the optimization process requires a profound understanding of aberration theory and sufficient experience. However, although the conventional optimization algorithms, such as the Damped Least Square (DLS) method, perform well at a fast converge speed to reach the local minimum, they are not satisfactory in global searching, especially concerning the symmetry-free systems with high complexity. On one hand, due to the high dimension of the optimization problem and the lack of experience in this new field, the global optimal solution always remains unknown.

The strong influence of the starting point limits the chance of finding globally optimal solutions. On the other hand, as the algorithm is based on only mathematical principles, a successful optimization result demands a great effort from the optical designer concerning the necessary optimization strategy, such as a structural change. The inferior automation greatly degrades the efficiency and limits the variety of the resulted solutions [10,11]. To enhance the global searching ability, some genetic optimization methods have been investigated and proved with promising results. However, the lack of physical guidance remains a problem in developing the automation of the algorithm, which limits the application for high-dimensional optimization tasks [12].

Therefore, the necessity of improving the aberration analysis and optimization method indicates the direction of the research in this work. The modern optical system design methods concerning aberration analysis and optimization are the focus. Intending to solve the problems and improve the modern optical design method, this work discusses a new method of comprehensive aberration analysis concerning the resolution of symmetry-free systems. With the obtained results, a bio-inspired optimization method with improved global searching ability is proposed, proved by some test optimization problems. In a summary, the corresponding research topics discussed in this dissertation are marked with ‘X’ in Figure 1.1.

Chapter 2 briefly introduces the currently available methods concerning optical system design. Starting from the fundamental concepts, both the classical and modern aberration analysis methods are introduced. Then, the general optical system design process and the basic system evaluation methods are illustrated. In addition, an overview of the current development of bio-inspired optimization methods is given.

In Chapter 3, a novel method based on the mixed ray-tracing method for aberration calculation is proposed, which is capable of calculating the surface-decomposed total, intrinsic, and induced aberrations in full order. Based on the additivity of the surface contributions, the surface additive Zernike coefficient representation of the aberrations can also be obtained, which is of great help in assessing the specific aberrations.

With the powerful tool for aberration calculation, a new optimization method based on the ant colony optimization (ACO) method is developed and explained in Chapter 4, with the goal to enhance the global searching ability. Concerning the potential in the global searching ability of the bio-inspired method and the current drawback of lacking

automation, a new optimization algorithm is investigated to improve the optimization performance. Specifically, the new algorithm intends to solve the problems in the structural change limitations by combining the mathematic algorithm and physical knowledge, while keeping the bio-inspired methodology. Compared to the conventional algorithms, the optimization process can be of higher automation with physical guidance which demands less effort from the optical designer. Therefore, as the user of the algorithm, the optical designer will get a large database of the possible solutions, providing an overview of the solution landscape, so that the best solution can be selected concerning the practical issues, such as the manufacturability and the cost. Concerning the large variety of optical setups, the development of the general physical guidance cannot be fully accomplished within a limited time. However, the first steps towards this far goal are investigated in this work, and the algorithm is proved applicable and promising for automatized optimization.

As a collection of the practical study of modern optical system design, Chapter 5 presents the application of the proposed methods with several concrete optical design tasks. The results are assessed to illustrate the advantages of the methods during the design process.

Finally, a conclusion is drawn in Chapter 6, together with the outlook for future research.

In addition to the main content, the appendices are provided for a better illustration of the main content. Appendix A verifies the calculation accuracy of the MRT method, and Appendix B discusses the impact of the approximation and distortion involved in the calculation results. As a supplementation to the general working principle of the algorithm in Chapter 4, the detailed physical considerations corresponding to the optimization strategies are provided in Appendix C-H. These optimization rules are essential for successful results as the guidance of the quasi-automatic optimization process. In addition to the optimization results shown in Chapter 5, two more optimization tasks and the corresponding optimization results are collected in Appendix I to illustrate the universality of the algorithm. Furthermore, together with the resolution-oriented optical design methods, the work also covers the research of distortion for a comprehensive understanding of the correction potential of non-spherical surfaces. For this purpose, a case study about distortion correction performance of freeform surfaces is presented in Appendix J.

2. State of the art

2.1 Fundamentals of optics

In general, an optical system should be modeled according to practical purposes. If the results are required with high accuracy, the physical approaches considering the diffraction effect should be applied. Usually, for optical imaging system design and simulation, only the geometrical feature of optics regardless of diffraction is considered. For a fast and effective calculation, the geometrical ray-tracing method is used to describe the ray path through the system, which is based on Snell's law of refraction, written as [5]

$$n \sin i_0 = n' \sin i_0', \quad (2.1)$$

where $\sin i_0$ and $\sin i_0'$ correspond to the incidence angle of the incoming and outgoing ray, n and n' are the refractive indices of the respective optical media. During the optical design process, most of the geometrical analysis provided by the optical design software is based on the ray-tracing results of a sequential set of optical surfaces.

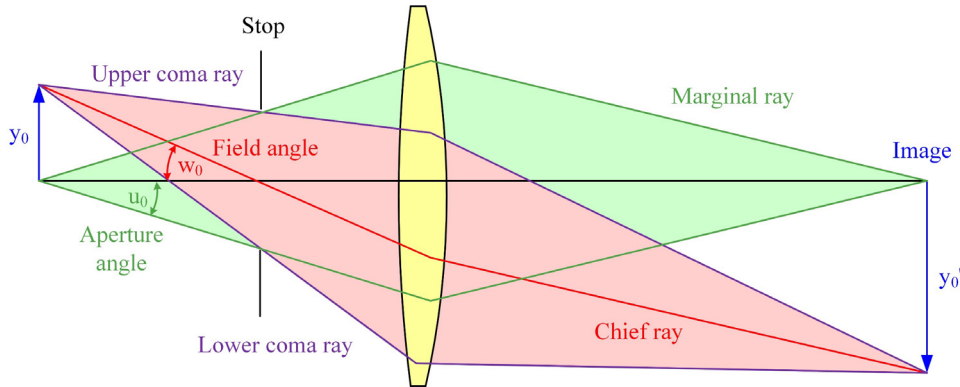


Figure 2.1. Definition of pupil, chief ray, and marginal ray in the optical system [13].

Besides the optical elements, the pupil and field of view (FoV) are the most important system parameters, as shown in Figure 2.1. The pupil is controlled by the stop defining the size of the light cone going through the optical system, which can be a real diaphragm inside the system or one of the lenses. The stop is conjugate to the entrance pupil (EnP) and exit pupil (ExP), which are the corresponding images of the stop formed by the optical system before and after the stop. The numerical aperture (NA) is defined as

$$NA = n \sin u_0, \quad (2.2)$$

where the aperture angle u_0 is the half light cone opening. In comparison, the FoV is determined by the size of the object, which will be magnified by the system and finally

seen on the image plane. Dependent on the finite or infinite location of the object, the FoV can be described by the object height y_0 or the field angle w_0 . Referring to the size and location of the stop and object, some characteristic rays can be defined. The chief ray (CR) comes from the outermost field point and goes through the center of the stop, while the marginal ray (MR) comes from the center of the object and goes through the edge of the pupil. And the upper or lower coma rays are defined as going through the off-axis field point and the upper or lower boundary of the pupil [13].

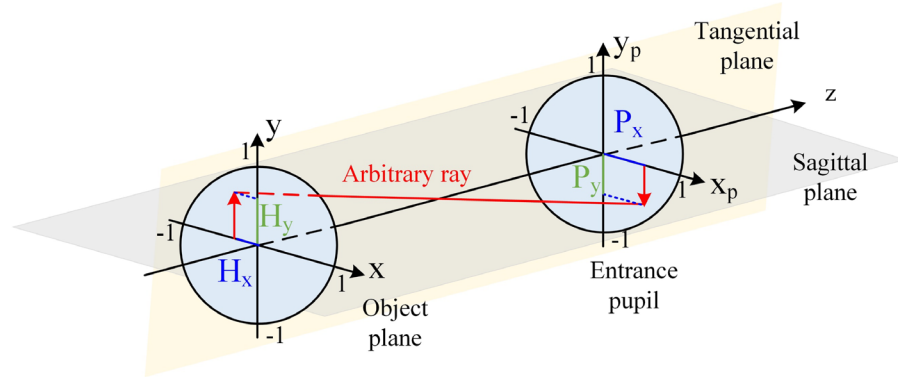


Figure 2.2. Normalized field and pupil coordinate of an arbitrary ray [13].

Due to the nonlinearity of the sine function in the law of refraction and the non-quadratic shape of the spherical surfaces, the imaging of the object point contains higher-order errors. For simplification, the paraxial optics is used for investigating the first-order properties, considering only the linear part of the sine function for the law of refraction, written as

$$n \cdot i_0 = n' \cdot i_0' \quad (2.3)$$

The paraxial approximation is only valid if both the aperture and the FoV are small enough near the optical axis for a rotationally symmetric system structure. In this case, the imaging condition is linear without any aberration, and the sag of optical surfaces can be neglected. Thus, the first-order properties of the system, such as the focal length, magnification, and EnP location, can be calculated accordingly. As a consequence of the linear light propagation, an invariant is valid before and after the refraction of a surface, denoted as the Helmholtz-Lagrange invariant, given by

$$n \cdot y_0 \cdot u_0 = n' \cdot y_0' \cdot u_0', \quad (2.4)$$

where y_0 and y_0' represent the object and Gaussian image height. Considering the 3D modeling of the system, a single ray can be defined by the normalized field and pupil coordinates, denoted as (H_x, H_y, P_x, P_y) , which is used for ray tracing in the optical design

software, as Figure 2.2 shows. The normalized coordinates therefore all have the range of $[-1, +1]$. Assuming the optical axis is along the z-axis, the paraxial optics is defined in the tangential plane composed of the y- and z-axis, while for an arbitrary ray, the coordinates also consist of the components in the x-z plane, denoted as the sagittal plane [13].

2.2 Matrix calculus for paraxial imaging

As mentioned above, the optical system has linear behavior when the paraxial approximation is valid. Thus, paraxial imaging only considers the linear propagation of rays without aberrations, which can be described by matrices. For rotationally symmetric systems, the ABCD matrix is used for the paraxial calculation to obtain the first-order properties. Specifically, any paraxial ray can be represented with a 2×1 vector and propagated by multiplying the system ABCD matrix, written as [13]

$$\begin{pmatrix} y' \\ v' \end{pmatrix} = \begin{pmatrix} A & B \\ C & D \end{pmatrix} \begin{pmatrix} y \\ v \end{pmatrix}, \quad (2.5)$$

where y is the paraxial ray height coordinate, and v is the direction angle. However, for an arbitrary ray in the 3D space, the calculation concerning only the tangential plane is not enough. Instead, the coordinates and the projected direction angles in the sagittal plane are also necessary. Therefore, the ABCD matrix needs to be expanded to a 4×4 matrix to include the components of both tangential and sagittal planes, and the 2×1 ray vector for the rotationally symmetric case should also be modified to a 4×1 vector. Thus, the paraxial propagation can be written as [14]

$$\begin{pmatrix} x' \\ y' \\ u' \\ v' \end{pmatrix} = \begin{pmatrix} A_{11} & A_{12} & B_{11} & B_{12} \\ A_{21} & A_{22} & B_{21} & B_{22} \\ C_{11} & C_{12} & D_{11} & D_{12} \\ C_{21} & C_{22} & D_{21} & D_{22} \end{pmatrix} \begin{pmatrix} x \\ y \\ u \\ v \end{pmatrix}, \quad (2.6)$$

where the ray vector elements x and u represent the global ray coordinate and projected angle in the sagittal plane, and y and v refer to the tangential plane accordingly. If the 4×4 matrix is applied for the paraxial calculation in the symmetry-free system, it is only valid around each component, which has local rotational symmetry. Furthermore, if the broken symmetry in the global frame due to the misalignment components or the off-axis system structure is considered, the 4×4 matrix should be generalized with the decenter and tilt components. Therefore, for an arbitrary ray in the 3D space, a 5×5 matrix is generated [15, 16], written as

$$\begin{pmatrix} x' \\ y' \\ u' \\ v' \\ 1 \end{pmatrix} = \begin{pmatrix} A_{11} & A_{12} & B_{11} & B_{12} & E_{11} \\ A_{21} & A_{22} & B_{21} & B_{22} & E_{21} \\ C_{11} & C_{12} & D_{11} & D_{12} & F_{11} \\ C_{21} & C_{22} & D_{21} & D_{22} & F_{21} \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \\ u \\ v \\ 1 \end{pmatrix}, \quad (2.7)$$

where the projected tilt angle and decenter are indicated in the decomposed E and F components in both planes. Such a matrix is used for the generalized paraxial propagation of rays in an arbitrary system, where the common axis for all the optical components may not exist anymore. In this case, the global nominal axis of the system should be taken as the reference.

2.3 Traditional aberration analysis methods

In reality, due to the nonlinearity of the refraction law, aberrations occur in the optical system, causing degradation of the imaging performance. Therefore, the main task for imaging optical system design is the correction of aberrations. Ideally, a point object should form a perfect image point at the ideal position in the image plane, but the existence of aberrations makes it blurred and displaced. From the geometrical viewpoint, there are three kinds of aberration descriptions, as shown in Figure 2.3. The longitudinal aberration, ΔS_{Img} , is defined as the displacement of the real ray intersection point from the ideal image plane along the axis, while the transverse aberration, Δy_{Img} , is the displacement from the ideal position in the image plane. As the wavefront is always perpendicular to the ray cone for optical systems, accordingly, the difference between the real wavefront and the ideal reference sphere at the ExP of the optical system is denoted as the wave aberration [13]. Due to the equivalence of wavefront and rays, the three kinds of aberration descriptions can be converted to each other. Mathematically, concerning the tangential plane, with the approximation of small NA and limited aberrations, the relations between them write as

$$\Delta y_{Img} = -\frac{R_{ref}}{n_l} \frac{\partial W}{\partial y_p}, \quad (2.8)$$

and

$$\Delta S_{Img} = -\frac{R_{ref}}{y_p} \Delta y_{Img}, \quad (2.9)$$

where R_{ref} represents the reference sphere radius, n_l is the image space refractive index, y_p is the ExP coordinate in the tangential plane. The transverse aberration in the sagittal

plane, denoted as Δx_{Img} , shares the same format as Δy_{Img} . Eq. (2.8) and Eq. (2.9) indicate that the wave aberration is one order higher in pupil power dependence than transverse aberration, and transverse aberration is one order higher than the longitudinal aberration.

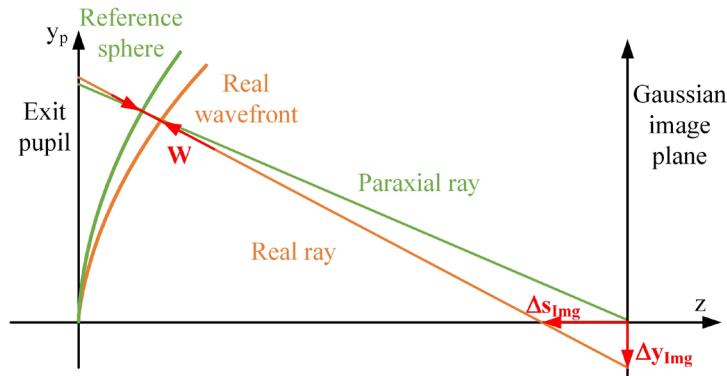


Figure 2.3. Longitudinal, transverse, and wave aberrations [13].

2.3.1 Seidel aberration theory

Considering rotationally symmetric systems, the primary aberrations can be understood as the lowest-order perturbation from the paraxial rays. Therefore, only the CR and the MR are involved in primary aberration calculation, as illustrated in Figure 2.4. The primary aberration theory is named after Ludwig von Seidel. With Taylor expansion of the geometrical perturbation, and concerning the rotation invariance, there are five kinds of monochromatic primary aberrations, namely spherical aberration, coma, astigmatism, field curvature, and distortion, which are of the third order for transverse aberration [5,6].

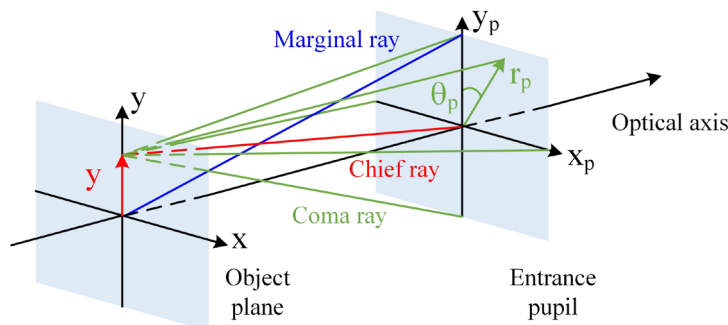


Figure 2.4. Chief ray and marginal ray from an off-axis field [13].

For refractive optical components, the refractive index is dependent on the wavelength. Therefore, the primary chromatic aberrations are of the first order, proportional to the derivative of the refractive index. The chromatic aberration of the chief ray causes a change of magnification, denoted as transverse chromatic aberration. And for marginal rays, the dispersion introduces a chromatic defocus along the optical axis, called longitudinal chromatic aberration. Due to the lowest-order perturbation, the most important feature of

the Seidel coefficients is the additivity among the surfaces. As the primary aberrations are dominant for most of the systems, the additive coefficients give clear information about the critical surfaces, which need further correction with structural change during optical design. Assuming the parameters $A_j = n_j(h_j c_j + u_j)$, and $\bar{A}_j = n_j(\bar{h}_j c_j + \bar{u}_j)$, the corresponding Seidel coefficients of the j^{th} surface are listed in Table 2.1.

Table 2.1. Seidel coefficients of primary aberrations.

Aberration	Coefficient
Spherical aberration	$S_{Ij} = A_j^2 h_j \left(\frac{u'_j}{n_j} - \frac{u_j}{n_{j-1}} \right)$
Coma	$S_{IIj} = \bar{A}_j A_j h_j \left(\frac{u'_j}{n_j} - \frac{u_j}{n_{j-1}} \right)$
Astigmatism	$S_{IIIj} = \bar{A}_j^2 h_j \left(\frac{u'_j}{n_j} - \frac{u_j}{n_{j-1}} \right)$
Field curvature	$S_{IVj} = L_j^2 c_j \left(\frac{1}{n_j} - \frac{1}{n_{j-1}} \right)$
Distortion	$S_{Vj} = \frac{\bar{A}_j}{A_j} (S_{IIIj} + S_{IVj})$
Longitudinal aberration	$C_{Ij} = A_j h_j \Delta \left(\frac{\partial n_j}{n_j} \right)$
Transverse aberration	$C_{IIj} = \bar{A}_j h_j \Delta \left(\frac{\partial n_j}{n_j} \right)$

where h_j and u_j are the object side ray height and angle of MR, while \bar{h}_j and \bar{u}_j are the ones of CR, and u'_j represents the image side angle. c_j is the radius of curvature of the surface. L_j is the Lagrange invariant, written as

$$L_j = n_j (\bar{h}_j u_j - h_j \bar{u}_j). \quad (2.10)$$

2.3.2 Wave aberrations with Zernike Fringe coefficient representation

Considering the wave aberration, defined as the deformed real wavefront compared to the reference sphere, the most common analysis method is the decomposition with Zernike polynomials. The wavefront aberration is defined by polar coordinate at the ExP, where ρ is the normalized radial aperture height, and θ is the azimuthal angle, as Figure 2.5 shows. Measured at the ExP, W can be fit as the sum of all the Zernike polynomials, written as

$$W(\rho, \theta) = \sum_{i=1}^Q \gamma_i Z_i(\rho, \theta), \quad (2.11)$$

where Z_i is the i^{th} Zernike term, and γ_i is the coefficient of the term. Q is the total number of Zernike terms. Figure 2.5 illustrates the plots of the Zernike polynomial terms from the lowest to higher orders, where n_z and m_z are the radial and azimuthal order.

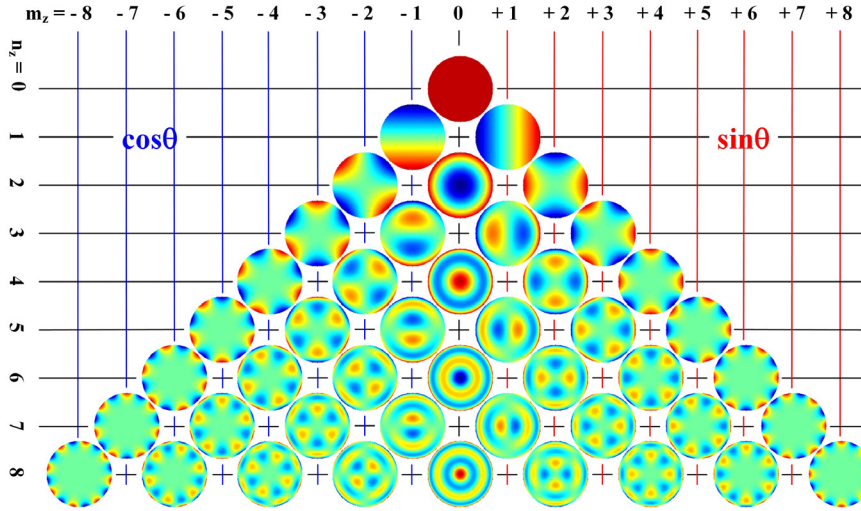


Figure 2.5. Plots of the Zernike terms according to the radial and azimuthal order [13].

Compared to Seidel's simple power expansion, Zernike polynomials are orthogonal and can be expanded to higher-order terms, which corresponds to the higher-order aberrations. Besides, the Zernike representation of the wave aberration can be applied to the symmetry-free systems regardless of the validation of the paraxial calculation. With the help of the least-square data fitting methods, the coefficients of Zernike polynomials give a clear clue of the critical aberrations in the system. However, the interpretation of the fitting results can be more complicated, if the orthogonality is perturbed. Specifically, the non-circular pupil boundary, apodization, and discretization due to a discrete finite sampling ray set can all bring errors in the fitting results.

In addition, when evaluating the wave aberration contributions of an arbitrary surface in the system with Zernike polynomials, the results are not very comfortable for some reasons: the fitting method implemented by the available software for the surface-resolved wave aberration is usually based on the approximated calculation of the corresponding intermediate image plane and the reference sphere parameter. This method always assumes a different reference sphere after each optical surface due to the changing magnification throughout the system. Specifically for mirror systems, the intermediate image may suffer from large astigmatism, and a toric reference surface makes more sense from a numerical

point of view. Consequently, the additivity of the coefficients among the surfaces is not fully valid, although the optical path difference along every ray is always additive, so the errors due to the approximations cannot be avoided. Furthermore, for the systems without proper ray aiming, the pupil aberrations cause a displacement of the sampling points from the entrance to the ExP. But as the Zernike polynomials are usually fitted based on the undistorted EnP coordinates, there are usually errors occurring in the results [17].

2.3.3 Intrinsic and induced aberration

As mentioned, Seidel theory describes the lowest order perturbation from the paraxial optics. Compared to the additive surface contributions of the aberrations, the higher-order perturbation involves the non-linear accumulated superposition of the primary aberrations.

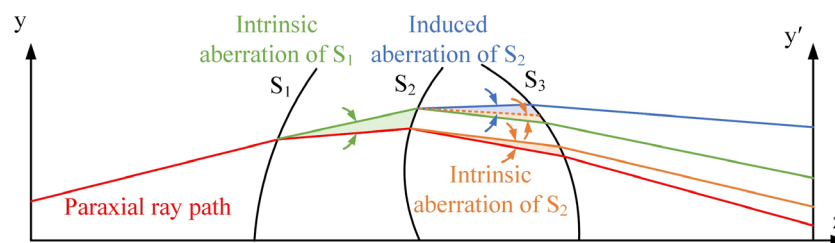


Figure 2.6. Intrinsic and induced aberration [13].

As Figure 2.6 illustrates, a paraxial ray coming from the object plane hits the first refractive surface S_1 . Due to the non-linear refraction and the non-quadratic surface shape, the aberration occurs, as the green zone shows. As for the aberration caused by S_2 , if the refraction of the paraxial ray is considered, the aberration caused by S_2 is not influenced by the aberrations of S_1 , but only caused by S_2 alone, which is defined as the intrinsic aberration. In contrast, if the perturbed ray after S_1 is further refracted at S_2 , the aberration includes the higher-order perturbation. Thus, the additional aberration subtracted from the total aberration is denoted as induced aberration.

Due to the complicated superposition of the aberration of various orders, the full-order induced aberration is hard to be expressed with analytical formulas. However, the understanding of induced effect is important for many applications, especially the complicated optical systems with large NA or wide FoV.

2.3.4 Aldis theory

Aldis theory was proposed for the full-order surface-additive transverse aberration calculation, which considers only one ray. The calculation is based on the paraxial and real

tracing data of the tested ray. Combining the ray-tracing results of both cases, the transverse aberration contribution of the j^{th} surface in the tangential and sagittal plane can be obtained for a ray in the tangential plane with [18]

$$\Delta x_j = \frac{n' X_j}{u'_I N_I} \left(\frac{u'_j N_j}{n_j} - \frac{u'_{j-1} N_{j-1}}{n_{j-1}} \right) - \frac{n' Z_j}{u'_I N_I} \left(\frac{u'_j L_{rj}}{n_j} - \frac{u'_{j-1} L_{j-1}}{n_{j-1}} \right) - \frac{n' h'_j}{u'_I N_I} (L_j - L_{j-1}), \quad (2.12)$$

$$\Delta y_j = \frac{n' Y_j}{u'_I N_I} \left(\frac{u'_j N_j}{n_j} - \frac{u'_{j-1} N_{j-1}}{n_{j-1}} \right) - \frac{n' Z_j}{u'_I N_I} \left(\frac{u'_j M_j}{n_j} - \frac{u'_{j-1} M_{j-1}}{n_{j-1}} \right) - \frac{n' h'_j}{u'_I N_I} (M_j - M_{j-1}) - \frac{n' H_I}{N_I} \left(\frac{N_i}{n_i} - \frac{N_{i-1}}{n_{i-1}} \right), \quad (2.13)$$

where (X_j, Y_j, Z_j) is the coordinate of the real ray intersection point on the j^{th} surface, and (L_j, M_j, N_j) is the corresponding optical direction cosine. u'_j and h'_j are denoted as the angle and height of the corresponding paraxial ray at the j^{th} surface. n_j and n' are the refractive index after the surface and in the image space. The subscript I denotes the corresponding parameters in the image plane. H_I represents the Gaussian image height. The surface contributions sum up to the exact total transverse aberration of the ray

$$\Delta x_{img} = \sum_{j=1}^N \Delta x_j, \quad (2.14)$$

$$\Delta y_{img} = \sum_{j=1}^N y_j. \quad (2.15)$$

The advantage of Aldis calculation is the additivity among the surface, but as the calculation is only valid for one ray in the rotational-symmetric systems, the application for a total quality assessment is still limited.

2.3.5 Surface-decomposed aberration with the phase space method

To solve the problem of surface-decomposed aberration calculation in the off-axis systems, a new higher-order aberration analysis tool was built based on the phase space (PS) method as an alternative option. It was an extended description in comparison to what was already proposed and sketched in [19, 20]. Inspired by the description method with position and velocity of a particle in mechanics, the optical PS method is first applied in illumination design, where an arbitrary ray can be geometrically represented analogously by the coordinate and direction angle [21]. With the same idea, the method was later also

implemented on imaging optical design tasks to help analyze full-order surface aberration contribution.

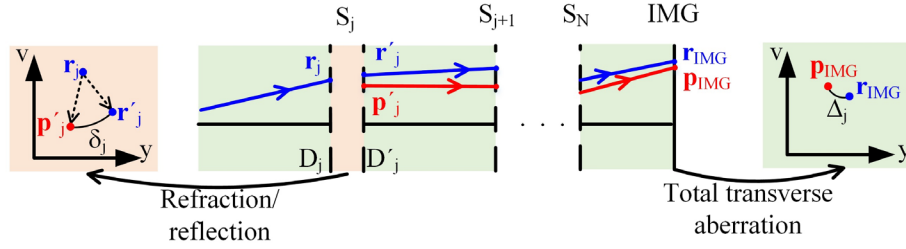


Figure 2.7. Propagation of the OAR in an arbitrary refractive system. [19].

Figure 2.7 illustrates the idea of the PS method for surface-decomposed transverse aberration calculation. Assuming a tested ray through an arbitrary optical system with N surfaces, the paraxial propagation of this ray can be calculated, considering the refraction inside an optical component and the propagation between the adjacent optical components separately. With the generalized 4×4 matrix introduced in Section 2.2, the refraction can be calculated between the dummy surfaces D and D' defined for paraxial calculation (orange zone), while the ray path between the two optical surfaces (green zone) can be simply obtained by free-space propagation. \mathbf{r} and \mathbf{p} indicate the ray vectors on the corresponding dummy surfaces, where the subscript denotes the surface number, and the primed values are the vectors after refraction. Therefore, concerning the transverse aberration contribution of S_j , the incoming real ray vector \mathbf{r}_j is propagated by both paraxial and real ray tracing, resulting in \mathbf{p}'_j and \mathbf{r}'_j respectively, the vector components of which in the tangential plane are shown in the y - v map. Thus, the distance between them, δ_j , can be calculated and further propagated paraxially until the image plane. Finally, the resulted total transverse aberration caused by S_j is denoted as Δ_j . It has been proved mathematically that all the surface contributions sum up to the total transverse aberration of this ray [19], written as

$$\Delta_{IMG} = \sum_{j=1}^N \Delta_j. \quad (2.16)$$

As an extension of Aldis theory, the PS method can distinguish the surface contribution of full-order transverse aberrations, and the calculation results show a good match to the exact Aldis results in the case of rotationally symmetric systems.

In addition, the method can be also applied to the symmetry-free system aberration

analysis tasks. As the real ray tracing data can be easily obtained from the optical design software, the challenge lies in the paraxial calculation with the generalized matrices in such cases. Since it physically takes over the paraxial calculation method with the ABCD matrix, it is considered the ‘generalized paraxial calculation’ for arbitrary system structures in this dissertation. For a symmetry-free system, the generalized paraxial calculation should follow the same definitions as the original paraxial calculation method, which can be explained in Figure 2.8.

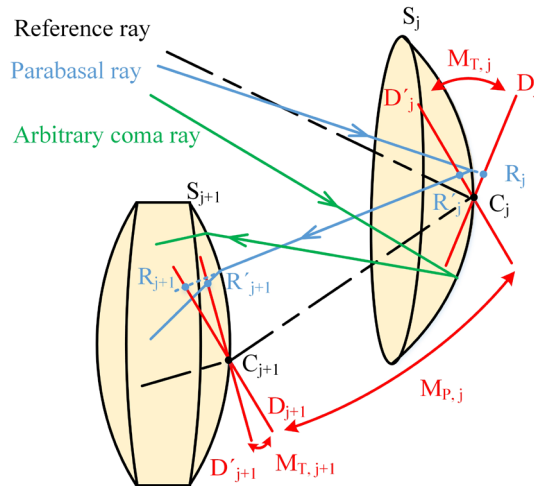


Figure 2.8. Propagation of the RR and a parabasal ray in an off-axis system, where S_j is a reflective surface and S_{j+1} is a refractive surface.

Similar to the rotationally symmetric system, where the ABCD matrix determines the path of paraxial rays around the optical axis, in the off-axis system, the 4×4 matrices are defined in a ‘parabasal zone’ around a real reference ray (RR) of the system. Figure 2.8 illustrates a part of an off-axis optical system with the j^{th} reflective surface S_j and the subsequent refractive surface S_{j+1} . The incoming RR marked in black first intersects both surfaces at C_j and C_{j+1} . Analogously, the generalized matrix determines the propagation of the rays slightly deviated from the RR, which are defined as ‘parabasal rays’. Usually, the so-called optical axis ray (OAR) is considered as the RR, which plays the same role as the optical axis in rotationally symmetric systems but is calculated by real ray-tracing. Different from the co-axial case, the ray paths of OAR and parabasal rays are not aberration-free due to the real propagation through the system with finite-sized tilts and decenters [17].

With the definition of the OAR as the generalized optical axis, the dummy planes can be determined at each C point shown in Figure 2.8, which are perpendicular to the incoming

and outgoing RR directions respectively. Specifically, D_j and D_{j+1} are dummy surfaces of the incoming RR, while D'_j and D'_{j+1} are responsible for the outgoing RR. Since the ray propagates in the free space from S_j to S_{j+1} , D'_j must always be parallel to D_{j+1} . Thus, the 4×4 matrices can be generally further categorized: the transfer matrix $M_{P,j}$ indicates either the refraction or the reflection effect due to the j^{th} optical component, and the propagation matrix $M_{T,j}$ only records the ray path between the j^{th} and $(j+1)^{\text{th}}$ surface in the free space. Similar to the conventional ABCD matrix, the refraction/reflection of the optical components in the system is automatically embedded in the paraxial matrices, so that the generalized paraxial propagation of the rays is only determined by the matrices.

If the off-axis system structure and arbitrary non-spherical surface shape are considered, the generalized paraxial calculation is complicated. Specifically, the coordinate transformation between the optical components is needed because the generalized 4×4 matrices are based on the local reference, and the matrix calculation of a general non-spherical surface is hard to formulate. Therefore, the authors proposed two methods: one can either rely on certain optical design software capable of such matrix calculation, or calculate the local 4×4 matrix by tracing a set of paraxial rays and solving linear equations. As the advanced function of automatic matrix calculation is not always available, it is meaningful to briefly illustrate the latter option.

Given a paraxial ray marked in blue in Figure 2.8, its ray path can be traced, and the intersection points at the corresponding dummy surfaces are denoted as R_j , R'_j , R_{j+1} , and R'_{j+1} . Therefore, the 4×1 ray vectors as given in Eq. (2.6) at each intersection point R can be written in the same format. Therefore, a linear system of equations with 16 unknown numbers can be established, and at least 16 equations are necessary to obtain a unique set of solutions. Thus, with each ray providing four linear functions, only four non-parallel paraxial rays need to be traced, so that the generalized matrices can be deduced from the ray data. Finally, for each two sequential optical surfaces, the transfer matrices are calculated simply by solving the linear equations [22, 23]. And the propagation matrix in the free space can be also easily obtained as long as the distance between the corresponding parallel dummy surfaces is known. It is important to mention that as the RR and the neighboring paraxial rays are traced in the real case, the aberrations existing in the paraxial zone are consequently included in the linear propagation calculation.

In this way, the generalized paraxial calculation can be realized, combining the

coordinate transformation due to the off-axis system structure. The PS method has been proved feasible for distortion analysis in the off-axis systems with freeform surfaces.

2.4 General optical design procedure

The goal of optical design is to create a system, which fulfills all the specifications [24]. For imaging systems, typically the first-order properties, the constraints, and the imaging quality are required. The general optical design procedure is shown in Figure 2.9.

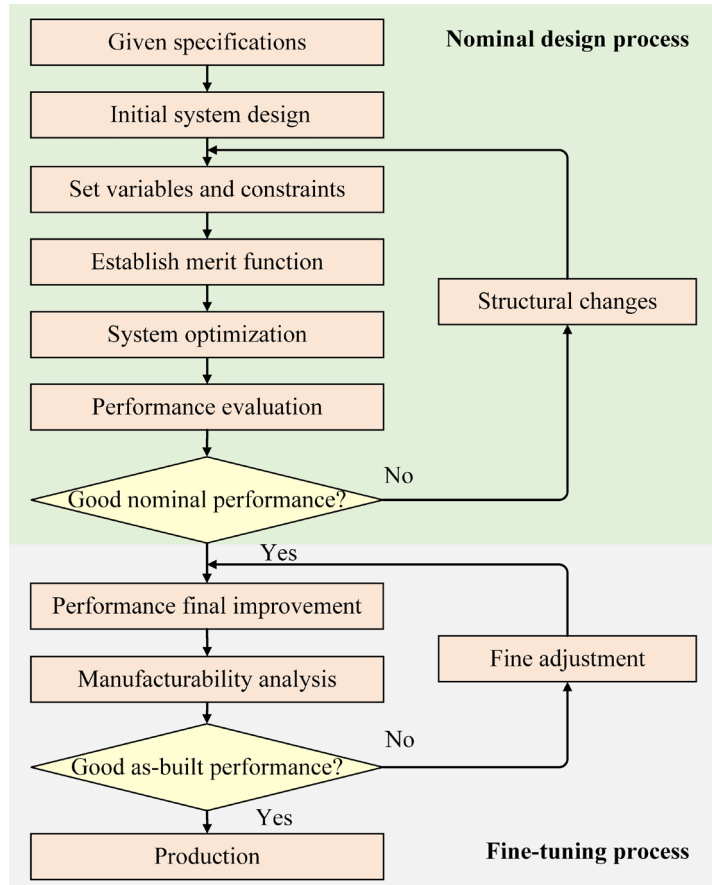


Figure 2.9. General optical design process [2].

First, a proper initial system should be chosen carefully as the starting point. The initial system is essential for a smooth and effective optimization path, and it determines the potential to obtain a well-performing final system. Therefore, it is meaningful to find out a proper initial system that roughly fulfills the requirements with limited aberrations. The easiest way to build an initial system is to modify an existing system, the data of which can be found in a patent or literature [24]. With several iterations, the structure of such an initial system can be adjusted closer to the specifications. Besides, a preliminary design with only paraxial lenses is also an option, as the ideal lenses determine the basic structure of the system considering the first properties [3]. Then the stepwise replacement with real lenses

guarantees a stable switch from the ideal to the real system for further optimization. As one of the decisive factors for a successful optical design, the initial system design should be worked out carefully.

Second, the system should be optimized based on a merit function (MF) that defines all the specifications and boundary conditions if the variables are set properly. The performance is then evaluated after each iteration to determine if any structural change is needed. In case of very challenging specifications, usually, the system should be adapted step-by-step to reach the requirements. With the help of aberration theory and experience as guidance, the system can be improved gradually until all the requirements are at least roughly fulfilled. Such a nominal system should indicate the general structure of the final design with adequate imaging performance. This procedure is denoted as the nominal design process.

Particularly for complicated systems, the nominal system may still not reach the ideal imaging performance, requesting further adjustments, for instance, with non-spherical surfaces. Also, the as-built performance evaluation concerning manufacturability should be considered, including the tolerance, mechanical design, environmental stability, and cost. Thus, the fine-tuning process of the system is necessary to further improve the system performance, until the system is realistic enough for production [24].

Therefore, optical design is a comprehensive process with a balance of many aspects. A successful result requires a deep understanding of aberration theory, practical manufacturing knowledge, and much experience as an optical designer, which is particularly challenging for beginners. Thus, nowadays, the development of a more efficient and automatic optical design method has become more and more attractive.

2.5 Conventional optimization method

The optimization process is the key to the success of optical system design. In most cases, the optimization is a complicated problem with a high-dimensional searching space which brings optical designers much difficulty in meeting all the requirements. Mathematically, the optimization problem of an optical system can be described as follows: assuming \vec{x} is a vector representing all the variables in the system with the dimension of m . The merit function $F(\vec{x})$ is denoted as a weighted sum of the squared differences between the target values $f_{tar,i}$ and MF operand values $f_i(\vec{x})$ [24]

$$F(\vec{x}) = \sum_{i=1}^m w_i \cdot [f_{tar,i} - f_i(\vec{x})]^2, \quad (2.17)$$

where m is the target number, and w_i is the weighting for each target. The main task is to minimize the MF value as much as possible while keeping the system realistic. In other words, a successful optimization should bring the system to the bottom of the deepest ‘valley’ of the searching space topology within the boundaries, understood as the global minimum. Ideally, $F(\vec{x}) = 0$ means that the current system fulfills all the targets and boundary conditions.

Besides the initial system introduced above, other three factors also influence the optimization results. First, the variables should be chosen carefully among all the lens data, as the number of the variables directly determines the degrees of freedom during optimization. Usually, different types of variables also contribute to the optimization to a different extent. For instance, the curvatures of the lenses usually have a stronger impact on the system performance compared to the thicknesses. If the system contains aspherical surfaces, the additional asphere terms of surface data bring much more degrees of freedom for correction. But meanwhile, the additional dimensions of the searching space due to the aspherical terms also increase the difficulty.

Second, the formulation of the MF is essential for the final results, as mathematically it determines the geography of the searching space topology. As Eq. (2.17) illustrates, the targets are translated from the practical specifications, which describe the ideal values of the specific parameters. As the most important components of the MF, the constraints can be handled as soft or hard targets. The soft targets do not need to be fulfilled exactly, while the hard ones require an exact fulfillment typically using Lagrange multipliers [24]. The weighting indicates the priority of each target that also determines the optimization path. With a well-formulated MF, the system structure should be controlled with the corresponding boundary conditions to guarantee a physically realistic system, and the system performance is improved concerning the appropriate criteria. In the usual cases, the number of the variables is much smaller than the number of targets, which means there can be no perfect solution, but the best compromise among all the constraints can be found.

Besides, the optimization algorithm also influences the optimization. Among all the available algorithms, the DLS algorithm is commonly applied in most commercial optical design software. As a derivative-based optimization algorithm, the direction of the damped

step Δd_j is described by the Jacobian matrix J . The solution in index form is

$$\Delta d_j = -(J_{ij} \cdot J_{ij} + \lambda \cdot I_{kk})^{-1} \cdot J_{jk} \cdot f_k, \quad (2.18)$$

with a damping factor $\lambda > 0$. $J_{ij} = \partial f_i / \partial x_j$ is the Jacobian matrix in index notation, and I_{kk} is the diagonal component of diagonal unity matrix I [24].

The mathematical principle makes it a very effective algorithm for local optimization, which searches for the short way deterministically to reach a local minimum with a fast converging speed. However, as it is a purely mathematical algorithm without any physical consideration, an inappropriate formulation of the variables or the MF may cause an unphysical result. And the optimization performance with the DLS algorithm is also not satisfactory due to two reasons: On one hand, due to the complexity of the high-dimensional searching space, there exist plenty of local minima, while the converging behavior of the algorithm brings no chance to escape automatically from the local minimum. Thus, the optimization stops before reaching the global minimum. On the other hand, as the starting point strongly influences the optimization results, the global search is hard to realize with only local optimization. And the experience-dependent initial system design is still a challenging task, especially for beginners with insufficient skills [25].

Combining with mathematical modifications and other algorithms, the DLS algorithm can be also implemented for global optimization. However, the inferior efficiency and large consumed time due to a tremendous amount of calculation are also not desired. In addition, the global optimization performance of optical systems with a large number of parameters, particularly for freeform systems, is still far from satisfactory enough due to the extremely high complexity of the searching space.

2.6 Asphere surfaces in optical design

2.6.1 Aspherical surface descriptions

During the optimization of the complicated optical systems, the correction of higher-order aberrations is important for the desired imaging quality. The spherical surfaces have only limited corrective power for the induced higher-order aberrations due to the limited degrees of freedom. To solve the problem, one of the most powerful methods is to introduce non-spherical surfaces into the system. The additional deviation from the basic spherical shape makes it effective for balancing the high-order aberrations. The general surface sag can be written as [13]

$$z_s = \frac{cr^2}{1 + \sqrt{1 - (1 + \kappa)c^2 r^2}} + \Delta z, \quad (2.19)$$

where c is the curvature, and r is the radial coordinate. The first part of the formula describes the basic shape of the surface, which is determined by the conic constant κ . Specifically, the various cases of surface shape dependent on κ are listed in Table 2.2.

Table 2.2 Surface shape against conic constant.

Surface shape	Conic constant	Surface shape	Conic constant
Hyperboloid	$\kappa < -1$	Paraboloid	$\kappa = -1$
Prolate ellipsoid	$-1 < \kappa < 0$	Sphere	$\kappa = 0$
Oblate ellipsoid	$\kappa > 0$		

Besides, the second part of the formula is the deviation from the basic shape, denoted as Δz . The rotationally symmetric non-spherical surface is an asphere, of which the simplest type is ‘even asphere’. The surface sag is described as

$$z_s = \frac{cr^2}{1 + \sqrt{1 - (1 + \kappa)c^2 r^2}} + a_1 r^2 + a_2 r^4 + a_3 r^6 + a_4 r^8 + \dots, \quad (2.20)$$

where a represents the coefficients of the exponential asphere terms with increasing even orders. Due to the non-orthogonality of the terms, the optimization is usually not stable. To solve the problem, the so-called ‘Q-type asphere’ description is proposed [26], which benefits both the stabilization of the optimization and the surface manufacture evaluation.

Furthermore, if the rotational symmetry of the surface is broken, it becomes a freeform, which introduces even more degrees of freeform to enhance the correction power. There are various description methods for freeform surfaces. In this dissertation, only the surface type ‘Zernike Fringe sag’ is used for freeform, which applies Zernike polynomials to describe the surface deviation, corresponding to the introduction in Section 2.6.2.

2.6.2 Performance enhancement with non-spherical surfaces

As mentioned above, the nominal design determines the general system, such as the lens number and the relative lens group positions. However, particularly for the complicated systems, the imaging performance can be close to, but still not reach the expectation. In this case, it is meaningful to consider adding aspherical surfaces or even other more complicated surface shapes like freeforms into the system to further improve the system performance while maintaining the general structure. Therefore, in the fine-tuning process, the imaging

performance should be improved toward the target. Compared to the nominal design process, the optimization with structural changes is no more necessary, and the MF operands concerning the original specifications only need to remain activated to ensure the fulfillment. Only those for the boundary conditions still need dynamic adjustment according to the varied lens data to keep the acceptable manufacturability. As only one of the tasks during the fine-tuning process, the final performance enhancement of the system is called ‘final improvement’ in this dissertation for distinguishment. This process only aims for better academic imaging performance, regardless of manufacturability.

Therefore, compared to the early optimization phase with structural changes, the final improvement phase can be less challenging. In addition, as the starting point of the final improvement, the system with a large degree of freedom indicates larger flexibility for further optimization, and the very complicated MF topology in the last phase of the optimization process also contains a huge number of local minima. In consequence, the more relaxing condition and the high dimension of the optimization problem together result in many different optimization paths that finally reach the goals. However, the optimization path strongly influences the result, and the best final improvement strategy cannot be easily predicted. In principle, the strategy can be either rougher, which may trigger a large jump over the MF landscape, or more conservative, which ensures a smooth improvement along the optimization path. Depending on the specific situation, sometimes a rough strategy can bring a large improvement in the performance, but it may also cause an unstable optimization path due to the extremely complicated MF topology, finally leading to an undesired result. In comparison, a conservative method is safer, but might block the chance to find out potentially optimal solutions.

In case the current design cannot reach the required imaging performance, the optical designer should analyze the current system to decide if, for example, the application of a non-spherical surface is necessary. Particularly for the system with broken rotational symmetry, such as the co-axial anamorphic system or off-axis systems, the rotationally symmetric surfaces have only limited power to correct the strongly asymmetric aberration distribution in the tangential and sagittal planes simultaneously. In this case, the freeform surfaces are of great help in improving the imaging performance. However, concerning the balance between the nominal and as-built performance, the determination of the location and type of the non-spherical surface is not trivial. Very often a compromise is necessary in the end, depending on the practical situation.

Generally, a freeform which allows the separation of individual ray bundles has a better performance for field-dependent aberrations, and a freeform near the pupil with coincident ray bundles is better for field-independent aberrations. This empirical conclusion can be explained by NAT to some extent [9], but because the current development of NAT still cannot reach so high aberration order as freeforms introduce to the system, it cannot explain or completely prove the application of freeforms. Therefore, as for our research, the choices of freeform locations still follow this general principle when necessary [27, 28].

Furthermore, when dealing with the optimization with freeforms, to ensure a smooth optimization path, the additional terms should be included in the system step by step. As for the rotationally symmetric aspheres, usually the aspherical terms are added gradually with increasing orders. In terms of freeform surfaces, the Zernike Fringe sag terms of the 4th order (Z5-Z9), 6th order (Z10-Z16), 8th order (Z17-Z25), and 10th order (Z26-Z36) are added sequentially to freeform surfaces. It is also important to reoptimize the system after each addition, using the same MF.

2.7 Bio-inspired global optimization methods

To overcome the drawbacks of conventional mathematics-based optimization algorithms, genetic algorithms have been proposed for global optimization and some are already implemented in commercial software. Inspired by nature, various ideas help escape from the local minima and improve global optimization efficiency. For instance, particle swarm optimization (PSO) and ant colony optimization (ACO) are both proved feasible for optical design [28, 29].

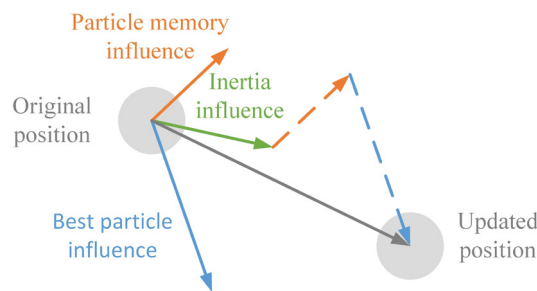


Figure 2.10. Particle movement principle of PSO [30].

PSO for optimization imitates the social behavior of bird flocking or fish schooling, the basic principle of which is illustrated in Figure 2.10. During each iteration, the movement of each member of the group is determined by three factors: the inertia velocity of the particle itself, the best position ever found by the particle, and the best position among the

whole group. These components add an updating learning and memory mechanism to the particle, which greatly improves the efficiency, and avoids being stuck in a local minimum. The method has been successfully implemented for some optimization problems with satisfactory results [29].

In contrast, ACO algorithm is inspired by the food foraging behavior of the ant colony. The ‘ants’ randomly explore the circumstances around the starting position and leave a pheromone trail along their walking route. If an ant finds an attraction area promising for a local minimum corresponding to the food sources, the quantity of pheromone it leaves on the way back will be changed according to its assessment of the solution. As the optimal solution can be blocked by some boundary conditions, the random exploration of the ‘ants’ makes it possible to find the solution via various paths on the landscape, as shown in Figure 2.11(a). Since the pheromone evaporates with time, it can be accumulated faster in shorter ways and the fellow ants tend to follow the more ‘attractive’ ways (Figure 2.11(b)). In the end, all the ants should converge to the best way to the attraction area, as Figure 2.11(c) illustrates.

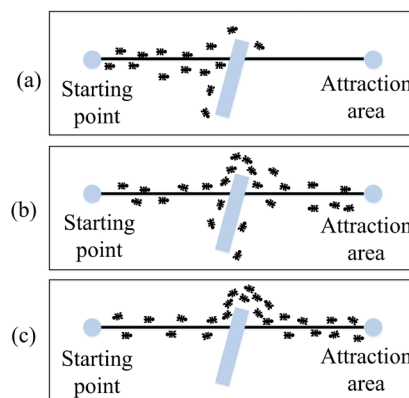


Figure 2.11. Concept of ACO. (a) the ants start to explore various paths; (b) the following ants choose a path to follow according to the pheromone trail; (c) ants all converge to the best way to the goal [14].

Regardless of the specific algorithm, such bio-inspired algorithms share some similarities. The optimization principle can be summarized by two special features: the stigmergy which is explicitly described by a deterministic mathematical model which controls the general direction of optimization; and the metaheuristic character represented by a probabilistic construction of solutions. In terms of the ACO algorithm, the pheromone model determines the rules of the communication among all the ‘working ants’ and indirectly ‘teaches’ the ants to make decisions by assigning probabilities to various paths. Meanwhile, the metaheuristic character of ACO encourages the ants to explore further

unfamiliar regions to potentially bring up more creative solutions. In general, the mixed probabilistic and deterministic features of such genetic algorithms greatly enhance the global searching ability.

Compared to the traditional DLS method, the convergence speed of such a bio-inspired optimization algorithm is mostly slower in local searching due to the probabilistic feature. Therefore, based on the global searching results, DLS algorithm can be used to enhance the convergence speed. As there are many aspects to consider while evaluating the global searching speed, such as the parameters involved in the algorithm and the programming software, it is hard to determine the best optimization methods. However, the bio-inspired algorithm offers a promising option to improve the optical design efficiency [29].

2.8 Fundamental of ACOR algorithm for optical design

The ACO algorithm was first applied to solve the problem with discrete variables, such as the traveling salesman problem [31], and later the algorithm was adapted for continuous problem optimization with various ideas. As for the optical systems, the optimization mostly deals with continuous variables like curvatures and thicknesses, but the discrete lens material variable is an exception, due to the limited availability of glasses. In addition, the range of the lens parameters cannot be estimated before the optimization, making the searching space topology much more complicated compared to the practical optimization tasks. Therefore, to find the most suitable continuous ACO algorithm for optical design, some of the available ideas were implemented for to investigate the feasibility of the idea, and it has been proved that the so-called ‘ACO in the field of real numbers’ (ACOR) algorithm is of the largest potential for optical design tasks [32].

The ACOR algorithm is implemented based on a solution archive of a capacity of K , the structure of which is shown in Figure 2.12 together with the algorithm outline. During the initialization, all the variables are determined and K solutions are randomly generated. They are all evaluated and ranked according to the MF value $F(\vec{x}_j)$. In the archive, \vec{x}_j is an m -dimensional variable vector of a solution with the j^{th} ranking. In each iteration, each ant in the family with a population P chooses a solution as a starting point. The possibility of each solution to be chosen is weighted according to the ranking. The weight ω of the j^{th} ranked solution is calculated by

$$\omega_j = \frac{1}{qK\sqrt{2\pi}} \exp\left(-\frac{(j-1)^2}{2q^2K^2}\right), \quad (2.21)$$

where q is a parameter for the bias towards the best-ranked solutions. With the guide solution, the ant explores a new solution nearby, the position of which is determined by a continuous probability density function (PDF). It is a Gaussian kernel function defined as the sum of different weighted individual Gaussian functions with the following form [32]

$$G^i(x) = \sum_{j=1}^K \omega_j g_j^i(x) = \sum_{j=1}^K \omega_j \frac{1}{\sigma_j^i \sqrt{2\pi}} \exp\left(-\frac{(x - \mu_j^i)^2}{2\sigma_j^{i2}}\right), \quad (2.22)$$

where $j = 1, 2, \dots, m$ indicates a certain component of the vector. μ is the mean value of this dimension among all the archive solutions and σ is the corresponding standard deviation which is presented as

$$\sigma_j^i = \zeta \sum_{h=1}^K [|s_h^i - s_j^i| / (K - 1)], \quad (2.23)$$

where s_j^i represents the i^{th} variable value of the j^{th} solution, and ζ is a parameter for convergence speed as an imitation of the evaporation of pheromone in nature. PDF is calculated only once for each component per iteration. In other words, for the construction of the same parameter of the solutions, the PDF does not change. After all the ants have moved, the new solutions are merged with the archive solutions and all these total ($P + K$) solutions are ranked again. The archive is updated with the top-ranked K solutions and the rest are deleted, followed by the next iteration, until the final stop criteria are reached [32].

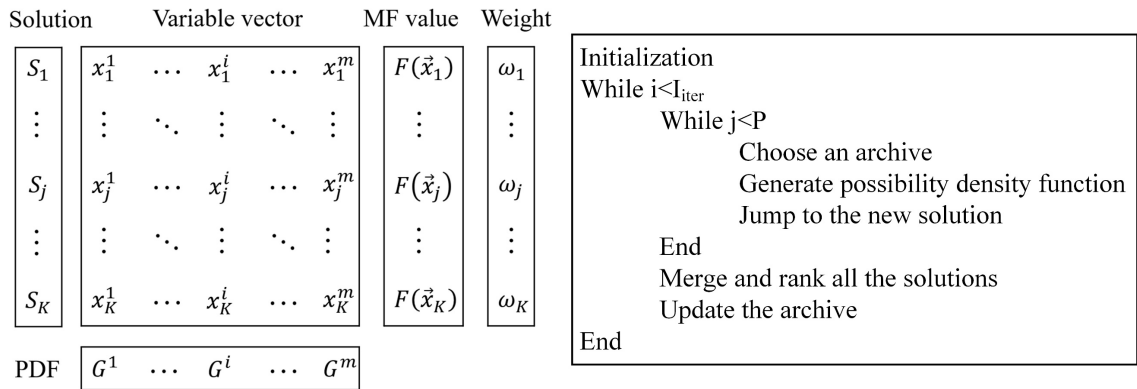


Figure 2.12. Structure of the solution archive of ACOR, and the algorithm outline [32].

The ACOR algorithm was proved to be feasible for simple optimization tasks with a good global searching ability. But as the ants only jump randomly according to the PDF without any physical consideration, the solution is very often not physically meaningful. Thus, the large number of failures degrade the efficiency of the optimization.

In addition, the simple ACOR method can only deal with optimization problems with a consistent number of variables and fixed MF. As a result, the optimization process still requires a large effort from the optical designer for determining the initial system and structural changes, so the automation is far from satisfactory.

2.9 Problems in symmetry-free system design

Nowadays, as applications based on symmetry-free optical systems have become more common and desirable in many fields, optical design tasks with freeform surfaces are in higher demand. Compared to conventional optical design methods, the limited understanding of freeform surfaces complicates the analysis and correction of aberrations. Concerning the practical design process, it is still challenging to summarize a reliable systematic design method for a symmetry-free optical system due to the following reasons:

1) Surface-decomposed resolution-related aberration calculation

For most rotationally symmetric systems with spherical surfaces, the primary aberrations are usually dominant. Thus, Seidel aberration theory with surface decomposition is generally enough to evaluate the system performance. In the case of the symmetry-free systems, although the classical analysis of aberrations based on paraxial optics is not valid anymore for off-axis systems, the lowest-order aberrations can be still calculated according to the theory developed by Araki for instance[33-35]. However, the higher-order aberration evaluation cannot be tackled by either of these methods. To solve this problem, various theories are developed, such as the well-known NAT [4, 9, 36-39], which is also surface resolved. Currently, the NAT is usually restricted to the 6th order, and various literature [40-43] have presented the further extension of NAT up to the 8th order concerning some specific aberrations, which is of great help for understanding the dependency on the field and pupil to some extent. However, the mathematical formulas grow strongly in length and complexity with the increasing order. Thus, the practical interpretations and assessment of higher-order aberrations are complicated. Furthermore, the complete understanding of the aberrations over the 8th order is still missing. For the freeform surfaces with much higher-order deviations in the optical system for achieving superior imaging performance, it is necessary to take the wavefront aberration analysis of the same order into consideration. In this case, NAT for freeform analysis is often not enough. [17]

As mentioned in Section 2.3.2, the surface-resolved Zernike coefficient fitting results of

the wave aberrations still suffer from inaccuracy due to the assumptions and approximations used in most optical design software. Concerning the ray bundle from an arbitrary field, as the wavefront changes in radius and shape during propagation through the system, the Zernike coefficients of the wave aberration measured after a certain surface cannot be added up with the ones of the successive surface. Besides, very often in the 3-dimensional space, the pupil distortion also causes errors in the fitting results, which may influence the aberration analysis. In addition to the full-order aberration evaluation, a better understanding of the critical aberrations is meaningful for a more effective correction strategy. Therefore, a reliable calculation method of the surface-additive Zernike coefficient representation should be investigated.

Besides, for a deeper understanding of the system performance, as well as the qualitative prediction of the manufacturing sensitivity, the induced effect evaluation of optical surfaces, particularly for off-axis systems, is also of great importance. However, the systematic study of intrinsic and induced aberration is also restricted in the range of circular symmetric systems until the 6th order [6], and the surface additivity cannot be clearly visualized because of the complicated superposition of aberrations among the surfaces. In addition, a numerical calculation method [44] for surface contributions of intrinsic and induced aberrations has been proposed to investigate the full-order aberration. But the assumptions and approximations hidden in the method cause error concerning the generalized system structure, therefore limiting the applications. As the understanding of induced aberration is essential for analyzing the high order aberration and sensitivity, the calculation method should be further improved.

2) Higher-order distortion correction

Despite the aberrations that degrade the resolution of the image, the symmetry-free structure also introduces higher-order distortion, such as keystone or bow-type distortion. [45] Although the existence of distortion does not influence the resolution, it still deforms the image, which can be a severe problem, especially for optical metrology applications. As many investigations were focused on the correction ability of freeform in resolution, the potential of higher-order distortion correction was not much illustrated in many application fields. Therefore, the correction performance of freeform considering distortion is meaningful to be discussed.

3) Improvement of the optimization method

With a better understanding of aberration correction, the system can be optimized according to the assessment result. The optimization method for rotationally symmetric systems has been discussed a lot considering the boundary conditions and structural changes [10, 11, 24]. Dependent on the actual purpose, the optical designers may have different criteria for a successful task. A globally optimal solution among the whole searching space is, in any case, desirable, while more creative solutions with various system structures can be also of interest. However, as mentioned above, the drawbacks of traditional DLS algorithm for local and global optimization still have limited ability in the global searching for outputting a large variety of solutions. Therefore, it is meaningful to investigate an optimization method, which improves the global searching performance and enhances the possibility to find out the global optimal solution. Furthermore, it is also desirable for the optimization method to be automatic to some extent as a great help for the optical designer without sufficient experience.

To reach the goal, one possibility is to equip the mathematical algorithm with physical knowledge and the experience for lens design, imitating the optimization strategy that a real optical designer would follow to find the optimal solution. For example, the intelligent aberration analysis and the execution of structural changes could be programmed in the optimization. In addition, the bio-inspired optimization idea could be also implemented in the algorithm, so that the global searching ability is enhanced. Thus, the optimization results are both deterministic and probabilistic with the advantage of accelerating the convergence speed of fully random searching schemes.

Furthermore, due to the high-order terms of the non-spherical surfaces, a large number of degrees of freedom are involved in the optimization problem of the freeform systems. In addition, the geometrical structure of the symmetry-free system also brings more degrees of freedom concerning the tilt and decenter of the lenses. Thus, the searching space of such optimization tasks can be extremely complicated with a high dimension and considerable local minima. Currently, the traditional optimization method is limited in global searching ability when many variables are involved, and there is not yet a general optimization method valid for all the system structures. However, with the help of an advanced aberration analysis tool, the global optimization of the symmetry-free systems can be developed to improve the system performance in a more systematic and promising way.

3. New method for aberration analysis

Following the idea introduced in Section 2.3.5, with the local refraction/reflection 4×4 matrices, the generalized paraxial ray-tracing can be calculated for each optical surface, combined with the free-space propagation between every two sequential surfaces. As the transformation of the ray vectors takes the local coordinate system as the reference, the paraxial calculation always needs to take the coordinate transformation into account. For the optical design software without the automatic matrix calculation, such calculation in the 3D space can be complicated. Therefore, inspired by the great potential of the method, a novel comprehensive higher-order aberration analysis tool is investigated, attempting to solve the problems in aberration analysis.

To simplify the mathematical approach of the generalized paraxial ray-tracing calculation, the first task of this study is to describe the propagation of rays with global reference in the 3D space. Additionally, the flexible switch between real and paraxial ray-tracing can be applied to develop more advanced functions for aberration evaluation. Therefore, in this chapter, the improved mixed ray-tracing (MRT) method is investigated to develop a multi-functional aberration analysis tool for surface-decomposed aberration evaluation in symmetry-free systems. In this study, the real ray-tracing is performed by Zemax, and Matlab is used for the main body of the program [17].

3.1 Surface contribution of total transverse aberration

To enhance the calculation efficiency, the generalized paraxial calculation method is first extended and improved in this work in comparison to what was suggested in [19,20]. Instead of the 4×4 matrices with the local reference, the global 5×5 matrices are calculated, so that the refraction/reflection and the free-space propagation including the tilt/decenter information are considered together for simplification, instead of the separate calculation. The calculation method follows that of the 4×4 matrix calculation with the linear equation set, while the number of unknown components becomes 20 in this case due to the additional non-zero components in the 5×5 matrices. In this dissertation, the 5×5 matrix is denoted as the ‘parabasal matrix’. Thus, the propagation of an arbitrary ray can always be described in the global reference frame with the parabasal matrices.

As introduced in Section 2.3.5, the phase space method has already been investigated and proved feasible for the system performance assessment concerning the surface-resolved

full-order aberrations [19, 20]. The additivity of the surface contributions is also the essential feature of the MRT method, and all the extended aberration evaluation implementations mentioned in this work are based on this idea. However, in comparison to the PS method, which only takes the paraxial calculation as the reference, the MRT method introduces more definitions of aberration with other references for extended functions. Correspondingly, the working principle of the MRT method is generalized. Therefore, due to the improved calculation method of the global paraxial matrices, the modified working principle, and the extended calculations in this work, the basic idea of the full-order total transverse aberration calculation method is still illustrated here, so that the core of the MRT method and the important complementary notations are clarified. This changed calculation scheme allows for extended analysis options, which are a great benefit to analyzing the systems and the functionality of the surfaces in a generalized system.

3.1.1 Additive surface contribution Δy

Figure 3.1 shows the tangential cross-section of an arbitrary optical system with a given RR, whose path is drawn in purple. Specifically, the RR here can be the CR of any finite field, in the special case of which the OAR is simply the CR of the central field. The notations of the dummy surfaces and intersection points still all follow Figure 2.8.

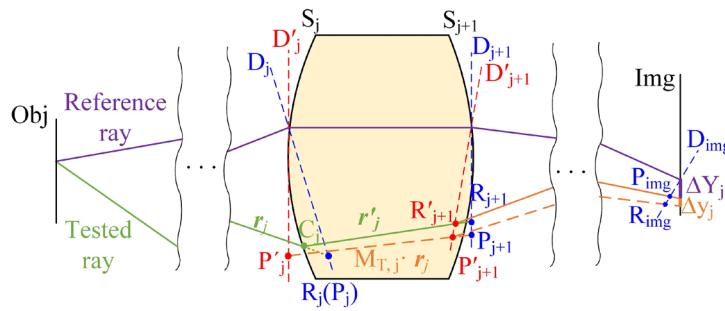


Figure 3.1. Surface contribution of transverse aberration projected in the tangential plane and transformation to the image plane.

After the paraxial matrices of the whole system are calculated, for any tested ray from the same object field point, marked in green in Figure 3.1, the transverse aberration contribution of an arbitrary surface can be calculated as follows: Assume first the tested ray starts from an arbitrary field point and hits an optical surface S_j at C_j in the real case, denoted as vector r_j . To calculate the transverse aberration of the tested ray caused only by S_j , the tested ray needs to be propagated through S_j in two different ways for the comparison. First, it is refracted by S_j in the real case from C_j , until it hits the next optical

surface S_{j+1} , with R_{j+1} being the intersection point at D_{j+1} . The ray path of this case is drawn in green between S_j and S_{j+1} ; Meanwhile, the extension of the tested ray intersects on the front dummy surface D_j at R_j , which is also considered as the starting point of paraxial refraction P_j . With the coordinates of P_j and angle of \mathbf{r}_j , the ray is refracted paraxially through S_j by multiplying the transfer matrix $M_{T,j}$, resulting in P'_j on the rear dummy surface D'_j . Then, the paraxial ray is further propagated in the homogeneous medium to the next front dummy surface D_{j+1} with the propagation matrix $M_{P,j}$. Therefore, the ray coming out of S_j already is divided into real and paraxial paths, denoted by the vectors \mathbf{r}'_j and $M_{T,j} \cdot \mathbf{r}_j$ respectively, which finally intersect D_{j+1} with different direction angles and coordinates. Furthermore, the two ray vectors from the two propagation methods are both paraxially propagated until the dummy image plane D_{Img} with the following paraxial matrices. Finally, the ray coordinates on the real image plane can be obtained simply by calculating the intersection points of the image plane and the ray vectors. Therefore, in the tangential/sagittal plane, the projected distance between the two ray vectors on the image plane, denoted as Δy_j and Δx_j , are the transverse aberration contribution components of S_j , as the only difference between the two ray vectors is the refraction method at S_j . For a clear comparison specifically on the image plane, Δx_j and Δy_j are both transferred back to the local coordinate on the image plane. The definition of these two parameters are corresponding to Δ_{IMG} in Eq. (2.16), which is the total aberration in the image plane. Despite the slightly adapted working principle, the additivity property proved for Δ_{IMG} is not violated. Similarly, the additivity of the surface contributions can be written as

$$\Delta x_{Img} = \sum_{j=1}^N \Delta x_j, \quad (3.1)$$

and

$$\Delta y_{Img} = \sum_{j=1}^N \Delta y_j, \quad (3.2)$$

where Δx_{Img} and Δy_{Img} are the total transverse aberrations of the tested ray projected in the sagittal and tangential planes, and N is the total number of the optical surfaces in the system. As the x- and y-components of transverse aberration contribution always share the same formulation and are only different in the cross-sections, for the rest of the chapter we

only consider the tangential plane, meaning only the y -components of the aberration descriptions are given. Due to the change in the paraxial matrix definition and calculation method, the accuracy and reliability of the calculation results of Δx_j and Δy_j still needs to be proved. Thus, a brief verification is provided in Appendix A.1.

3.1.2 Chief ray referred surface contribution ΔY

Except for the additive surface contribution Δy_j referred to the paraxial ray vector, the MRT method also supports the aberration calculation with another reference for extended functions. In some cases, paraxial propagation is not of practical concern in optical design. Instead, the evaluation of the system performance may require the comparison between the chief ray and other coma rays from this field, such as the evaluation of the real spot diagram and wavefront aberration. Thus, we define the distance between the intersection positions of the CR and the test coma ray from the same field as ΔY_j , as shown in Figure 3.1. In this case, the CR is considered as the RR here, and the tested coma ray coming from the same field point goes through a finite pupil position. In other words, the paraxial propagation at S_j is no more of concern for ΔY_j calculation. Instead, only the real ray vector \mathbf{r}'_j is paraxially transferred to the image plane as introduced before to compare to the real ray vector of the CR. According to the propagation method, ΔY_j here includes all the surface contributions of transverse aberration from S_1 to S_j . Consequently, ΔY_j is not additive among all the surfaces, as it compares only to the real CR position. Thus, for the N^{th} optical surfaces in the system, we have

$$\Delta Y_{img} = \Delta Y_N. \quad (3.3)$$

In general, the definition of Δy_{img} and ΔY_{img} provide alternative options for system performance analysis concerning different purposes. If the additive surface contribution to the transverse aberration is of concern, Δy_j is required for determining the relative contribution and balance among all the surfaces. However, if a visualized transverse aberration regarding the real spot diagram is required, ΔY_j can better illustrate the spread intersection positions of the rays from the same field on the image plane.

3.1.3 Discussion: Δy_{img} , ΔY_{img} , and spot diagram

The definitions of Δy_{img} and ΔY_{img} share similarities and distinctions because of the different references. To better illustrate the relationship between them, Figure 3.2 shows

two arbitrary rays coming from the ExP, marked in blue and red respectively. For simplification, we assume the two rays both come from the on-axis FoV, so that they should intersect the axis again in the ideal paraxial image plane, as the dashed lines illustrate. However, in the real case, their ray paths with aberrations are different, as the solid lines indicate. The Δy_{Img} for each ray is defined as the distance between its real and paraxial intersection point in the real image plane, while ΔY_{Img} shows the distance between the real intersection points of the tested ray and the CR of this field in the image plane. Specifically in this case, the CR is the optical axis ray, so that ΔY_{Img} in Figure 3.2 refers directly to the optical axis. Thus, if the image plane locates at the ideal image position, ΔY_{Img} will coincide with Δy_{Img} . However, if the real image plane is moved away from the paraxial position, according to the same definition, we observe separated Δy_{Img_1} and ΔY_{Img_1} , as well as Δy_{Img_2} and ΔY_{Img_2} as marked in Figure 3.2.

In consequence, due to the different definitions, the scale of Δy_{Img} is not necessarily corresponding to the spot diagram of the same field, but the scale of ΔY_{Img} must be.

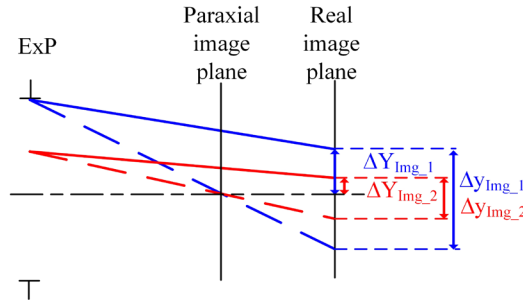


Figure 3.2. Geometrical illustration of Δy_{Img} and ΔY_{Img} .

3.2 Surface-decomposed intrinsic and induced aberration

3.2.1 Calculation method

As the core idea of the MRT method, the switch between real and paraxial ray-tracing at any surface realizes the surface-decomposed aberration calculation. Furthermore, if the definition of the full-order intrinsic and induced aberration is considered, the MRT method can be easily adapted for the extended application. As an overview, Figure 3.3 illustrates the extension of intrinsic/induced aberration calculation.

As mentioned above, if the ray vector \mathbf{r}_{obj} from the object reaches S_j in the real case, with both real and paraxial calculations of \mathbf{r}_j and the paraxial propagation afterward, we can finally calculate Δy_j caused by S_j alone by calculating the distance between $\mathbf{r}_{Img,R}$ and

$\mathbf{r}_{Img,P}$. Similarly, if \mathbf{r}_{obj} is propagated to S_j only by paraxial calculation instead of real ray-tracing, the resulted paraxial ray vector \mathbf{p}_j will be still divided into two paths after refraction/reflection at S_j , denoted as $M_{T,j} \cdot \mathbf{p}_j$ and \mathbf{p}'_j for the paraxial and real ray tracing respectively. Following the same procedure explained in Section 3.1.1, finally, the distance between the resulted ray vectors $\mathbf{p}_{Img,R}$ and $\mathbf{p}_{Img,P}$ in the image plane is the surface contribution of the intrinsic aberration, denoted as $\Delta y_{int,j}$.

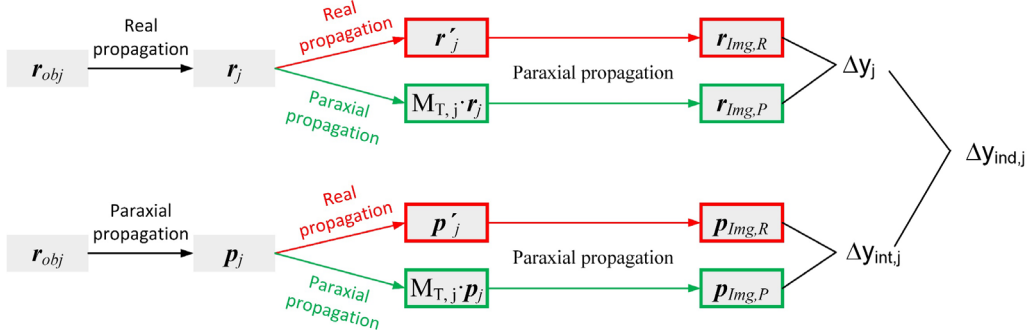


Figure 3.3. Mixed ray-tracing process for calculating the surface-decomposed total, intrinsic, and induced transverse aberration of S_j concerning only the y -component.

As $\Delta y_{int,j}$ of each surface is independent according to the calculation method, it is easy to clear that the additivity $\Delta y_{int,j}$ is still valid, written as

$$\Delta y_{Img} = \sum_{j=1}^N \Delta y_{int,j}. \quad (3.4)$$

Then, the corresponding surface contribution of the induced aberration $\Delta y_{ind,j}$ can be easily obtained by subtracting the intrinsic aberration from the total value. As for a single surface, we have

$$\Delta y_{ind,j} = \Delta y_j - \Delta y_{int,j}. \quad (3.5)$$

Consequently, the induced aberrations of the surfaces are also additive, written as

$$\Delta y_{ind,Img} = \sum_{j=1}^N \Delta y_{ind,j}. \quad (3.6)$$

3.2.2 Approximation in the calculation

As the essential point of the switch between real and paraxial propagation of the rays, it should be emphasized, that the real ray tracing always starts from the intersection point at the real surface, while the generalized paraxial ray tracing considers always the intersection

distance between C_{pj} and C'_{pj} is no more neglectable. In addition, if the surface deviation from the basic spherical shape is too large, the strongly curved surface sag also causes a pitfall in the approximation. The corresponding detailed illustrations of the two cases can be found in Appendix B, together with the analytical estimation of the error. The evaluation of the error and a guideline concerning the practical applications are also introduced there. In general, for most of the well-corrected systems with moderate freeform surfaces, the error caused by this approximation can always be considered acceptable. The approximation here simplifies and therefore speeds up the whole calculation procedure.

Among the current methods for calculating the intrinsic or induced aberrations, there is no available tool that also calculates the full-order intrinsic/induced aberrations based on the same definitions. Therefore, it is hard to find an appropriate reference for proving the reliability of the MRT method concerning full-order intrinsic/induced aberration calculation. Thus, one of the best ways to evaluate the results with the MRT method is to make a test on a simple on-axis system, whose intrinsic aberrations can be easily predicted, so that the results of intrinsic aberration calculation with the MRT method can be verified. The corresponding illustration is given in Appendix A.2. The results show that considering the qualitative analysis of the intrinsic/induced aberration, the calculation is accurate enough with minor error caused by the approximation [17].

3.3 Surface-decomposed Zernike coefficient representation

With the full-order aberration calculation results with the MRT method, the application can be further extended. For many purposes, the analysis for specific types of aberration is of great help to better correct the system, particularly for the symmetry-free systems. To solve the problems in the accuracy and the additivity of the conventional surface-decomposed Zernike coefficients fitting method mentioned in Section 2.3.2, the improved Zernike representation based on the MRT calculation is illustrated in this section.

The relation between wave and the transverse aberration in the tangential plane has been given by Eq. (2.8). In the fitting procedure, the real wavefront is calculated with the optical path length of the sampling rays and the center of the wavefront reference sphere is determined by the real CR of this field, regardless of the paraxial rays. Thus, assuming the image space is in the air and considering the two kinds of transverse aberration notation introduced in Section 3.1.2, the transverse aberration here should be written as

$$\Delta Y = -R_{ref} \frac{\partial W}{\partial y_p}, \quad (3.7)$$

where ΔY here is regarding the CR intersection position of this field, instead of that of the paraxial ray. Furthermore, if the corresponding wave aberration after S_j is denoted as W_j , it can be fitted to the Zernike polynomials terms with a number of Q , written as

$$W_j = \sum_{i=1}^Q \gamma_{i,j} Z_i, \quad (3.8)$$

where $\gamma_{i,j}$ is the coefficient of the i^{th} Zernike polynomial Z_i . Considering ΔY of a certain surface, together with Eq. (3.7), we have

$$\Delta Y_j = -R_{ref} \left(\gamma_{1,j} \frac{\partial Z_1}{\partial y_p} + \gamma_{2,j} \frac{\partial Z_2}{\partial y_p} + \dots + \gamma_{Q,j} \frac{\partial Z_Q}{\partial y_p} \right). \quad (3.9)$$

This relation in terms of ΔX_j in sagittal plane shares the same format. It can be seen, that if ΔX_j and ΔY_j are directly used to fit the derivatives of Zernike polynomial terms, the corresponding coefficients do not change. Thus, given sufficient sampling rays M , and using the first 36 Zernike Fringe terms for fitting, we can rewrite Eq. (3.9) with matrices to express the wave aberration after the j^{th} surface, as

$$-R_{ref} \begin{pmatrix} \frac{\partial Z_1(x,y)}{\partial x_{p1}} & \frac{\partial Z_2(x,y)}{\partial x_{p1}} & \dots & \frac{\partial Z_{36}(x,y)}{\partial x_{p1}} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial Z_1(x,y)}{\partial x_{pM}} & \frac{\partial Z_2(x,y)}{\partial x_{pM}} & \dots & \frac{\partial Z_{36}(x,y)}{\partial x_{pM}} \\ \frac{\partial Z_1(x,y)}{\partial y_{p1}} & \frac{\partial Z_2(x,y)}{\partial y_{p1}} & \dots & \frac{\partial Z_{36}(x,y)}{\partial y_{p1}} \\ \vdots & \dots & \ddots & \vdots \\ \frac{\partial Z_1(x,y)}{\partial y_{pM}} & \frac{\partial Z_2(x,y)}{\partial y_{pM}} & \dots & \frac{\partial Z_{36}(x,y)}{\partial y_{pM}} \end{pmatrix}_j \begin{pmatrix} \gamma_1 \\ \gamma_2 \\ \vdots \\ \gamma_{35} \\ \gamma_{36} \end{pmatrix}_j = \begin{pmatrix} \Delta X_1 \\ \vdots \\ \Delta X_M \\ \Delta Y_1 \\ \vdots \\ \Delta Y_M \end{pmatrix}_j. \quad (3.10)$$

For simplification, this expression can be also written as

$$-R_{ref} (\nabla Z \cdot \gamma)_j = \Delta \eta_j. \quad (3.11)$$

where $\Delta \eta_j$ represents both ΔX_j and ΔY_j . According to the definition of the least-square data fitting method, we have

$$\left[-R_{ref} (\nabla Z \cdot \gamma^*) - \Delta \eta \right]_j^2 \rightarrow \min. \quad (3.12)$$

where γ^* is supposed as the best fitting results of Zernike coefficients. By solving the equation of matrices, finally, we have the solution for γ^*

$$\gamma^*_j = \left\{ \left[(\nabla Z)^T \cdot \nabla Z \right]^{-1} \cdot (\nabla Z)^T \cdot \left(-\frac{\Delta \eta}{R_{ref}} \right) \right\}_j. \quad (3.13)$$

This means, by collecting the additive ΔY_j obtained by the methods mentioned above, we can also finally calculate the Zernike coefficients after S_j which correspond to the accumulated wave aberrations after this surface.

Furthermore, as for the ‘pure’ surface contribution of S_j , denoted as ΔY_{j0} , we have

$$\Delta Y_{j0} = \Delta Y_j - \Delta Y_{j-1}, \quad (3.14)$$

inserting in Eq. (3.9), we have

$$\begin{aligned} \Delta Y_{j0} &= -R_{ref} \left[(\gamma_{1,j} - \gamma_{1,j-1}) \frac{\partial Z_1}{\partial y_p} + (\gamma_{2,j} - \gamma_{2,j-1}) \frac{\partial Z_2}{\partial y_p} + \cdots + (\gamma_{n,j} - \gamma_{n,j-1}) \frac{\partial Z_n}{\partial y_p} \right] \\ &= -R_{ref} \sum_{i=1}^n (\gamma_{i,j} - \gamma_{i,j-1}) \frac{\partial Z_i}{\partial y_p}. \end{aligned} \quad (3.15)$$

This means the single surface contribution of Zernike coefficients can be finally obtained by iteratively subtracting the coefficients of the surface before. In other words, the coefficient of a certain term fitted directly after S_j is the sum of all the coefficients until S_j . Compared to the definition of ΔY_j , the coefficients are additive among the surfaces, which finally add up to the image space wave aberration.

Due to the strict additivity of the surface contributions of the MRT method, the calculation results cannot be compared to those based on approximations and assumptions. However, it still makes sense to verify the results with Zemax wavefront fitting results. The details are given in Appendix A1.3.

Considering the change of the wavefront during the free-space transformation between the adjacent optical surfaces, the calculation results of the MRT method already indirectly include the propagation into the results. Compared to other numerical calculation methods with intermediate image calculation [44], the MRT method calculation for the Zernike coefficient after a certain surface is not impacted by the corresponding approximations

involved in the calculation. Specifically, an individual surface contribution of Zernike coefficients automatically contains the propagation to this surface and the refraction/reflection through it.

As the transverse aberrations after each surface are all finally transferred to the image plane, the corresponding wavefront is calculated at the same ExP position with the same radius. Thus, the changing of the shape and radius during the propagation through the system is no longer of concern. Consequently, the problem in the wavefront propagation variance and the normalized radius is solved. In addition, as the sampling rays are all defined by the ideal pupil coordinates, the integral wavefront is fitted also based on the ideal pupil, which means the problem of pupil distortion also does not exist. Instead, all the fitted coefficients share the same ExP parameters which are determined by the chief and coma rays. Therefore, the problem in the conventional Zernike fitting algorithm is avoided by this method.

So far, Nodal aberration theory usually investigates the first 16 Zernike terms. For higher-order aberrations, despite a well understanding of some specific aberrations [41], a complete investigation and comparison among all the aberrations in the same order of freeform deviation is still missing. In comparison, the orders of aberrations calculated with the MRT method are not limited, as long as the derivatives of the Zernike polynomial terms are given.

Furthermore, we could also evaluate the intrinsic and induced surface contribution by fitting only the corresponding transverse aberration to Zernike polynomial derivatives, which is a powerful analysis tool for aberration correction together with sensitivity [17].

3.4 Discussion

So far, the working principle of the MRT is introduced. The core of the method is the switch between the real ray tracing and the generalized paraxial calculation based on the global parabal matrices. The verification results prove the accuracy of the method.

As the reference of the parabal matrices, the choice of the RR has a great impact on the aberration calculation results. Specifically, if the RR and the tested rays come from different FoVs, distortion will be included in the transverse aberrations. For some purposes, the distortion may need to be removed. The various cases concerning the distortion are discussed in more detail in Appendix C.

4 Improved global optimization algorithm

As introduced in Chapter 2, the conventional optimization methods still have limitations in various aspects, which demand much effort from the user concerning the optimization strategy. In this chapter, an improved optimization method for quasi-automatic global searching is discussed. As an overview of the research directions and motivations of this work, Table 4.1 summarizes the challenges in improving the optimization algorithm and the corresponding solutions and development.

Table 4.1. Overview of the optimization method research in this work.

Challenges in optimization		Corresponding development	System with only spherical surfaces	System including aspheres	System including freeforms
i.	Structural change	Lens splitting Aspherization	X	X X	
ii.	Higher-order aberration evaluation	MRT method	X	X	X
iii.	Handling a large degree of freedom	Final improvement			X
iv.	Automation		Improved	Improved	Improved
v.	Global searching ability		Improved	Improved	Improved
vi.	Manufacturability / sensitivity		Critical cases avoided	Critical cases avoided	Critical cases avoided
Applications			Retro-focus system		Anamorphic system

The first column of the table lists the general goals considered for the improved optimization algorithm development, which are currently not fully reached by the conventional optimization methods. The first challenge is the structural change limitations, as (i) shows. Starting from the initial system, appropriate structural changes are essential for aberration correction during the whole nominal design process. Second, as the imaging performance is improved, the higher complexity of the system structure usually brings a larger degree of freedom. Thus, as (iii) illustrates, the algorithm should be capable of dealing with very high-dimensional optimization problems, particularly the optical systems with aspherical or even freeform surfaces. In addition, an improved higher-order aberration analysis method is also beneficial for improving the system performance during the optimization process, referring to challenge (ii).

Therefore, the final goal of the research is to complete all these challenges by introducing physical knowledge into the optimization algorithm. Ideally, it should be

feasible for a fully automatic optimization from the initial design until the final output for all kinds of optical systems (challenge (iv)). Simultaneously, the algorithm should be strong in global searching to provide the user with a large number of output solutions with acceptable manufacturability, referring to challenges (v) and (vi). Considering the large variety of optical systems, the complete development of such an algorithm is a very broad research topic requesting a huge amount of programming, making it impossible to include all the detailed physical issues in this work. Therefore, this work concentrates on the development of an improved optimization algorithm to investigate the benefit of physical guidance in optimization.

The original ACOR algorithm is already introduced in Section 2.8 as a promising global optimization method. Based on the archiving mechanism, each ‘ant’ individually explores for new local minima, so that the global searching ability can be greatly enhanced. However, the drawbacks of this method still limit the application. Considering its great potential in global searching ability, an improved ACOR algorithm based on the ‘ant group’ is developed, denoted as the GACOR algorithm. Intending to tackle a successful optimization with a high level of automation, its working principle mimics the optical design process performed manually by experienced optical designers, who manage the task with both physical and empirical knowledge. Correspondingly, the ‘ants’ are trained to have their own ‘judgment’ during the optimization process, so that the whole ‘ant family’ can have better overall control of the results. Therefore, the algorithm is investigated as a first step to completing the challenges mentioned in Table 4.1.

As mentioned above, the structural changes are mainly considered in the nominal procedure, and the high-dimensional optimization problems refer to the complicated systems in the later optimization stage, whilst the complete development of the algorithm is not realistic. Therefore, to include the research concerning both challenge (i) and (iii) in the work within the limited time frame, the GACOR algorithm development is divided into two subtopics: In the first part, the algorithm is developed for quasi-automatic global optimization with structural changes guided by physical knowledge concerning simplified nominal design tasks. The algorithm only focuses on the spherical aberration correction for rotationally symmetric systems with only an on-axis field. Correspondingly, lens splitting and aspherization are developed in the algorithm as the concrete structural change options, marked with blue ‘X’s in Table 4.1. The usage of these two options covers both systems with and without aspheres, which will be discussed in the following sections.

Correspondingly, a retro-focus system design task will be illustrated in Section 5.1 as an application example of the method, presented in the bottom row in Table 4.1.

In the second part, only the high-dimensional optimization problem is considered to prove the capability of the GACOR algorithm when dealing with a large number of variables. As mentioned in Section 2.4, the obtained nominal system only indicates the general structure of the system, while the lens parameters might not be optimal concerning ideal imaging quality, particularly for complicated systems. Thus, the final improvement process is considered as the second part of the algorithm development, marked with red 'X', which further optimizes the system performance without changing the general structure. As the very late phase of the optical design process, the starting system can be much more complicated with aspherical or even freeform surfaces, which is ideal for modeling a high-dimension optimization problem. For this purpose, an anamorphic system with freeform is chosen to test the optimization performance, presented in Section 5.2.

In addition, the aberration analysis of the algorithm is supported by the MRT method introduced in Chapter 3, as marked with black 'X's. It is of great help for the optimization with non-spherical surfaces.

Besides all the challenges mentioned above, the manufacturability of the output solutions is also of great concern, as challenge (v) shows. Depending on the various purposes of the design task and the actual manufacturing conditions, the criteria of sensitivity and tolerance analysis can be very different. However, the GACOR algorithm is capable of filtering out the solutions with critical manufacturability based on general criteria, which will be discussed in the following section. Thus the critical cases are always avoided in the output.

As the extension of the original ACOR algorithm, the GACOR algorithm is designed to output a solution collection with various system structures at a higher level of automation, all fulfilling the original design specifications with moderate manufacturability. However, it is important to mention that the output solutions collected in the database found by the algorithm are only pre-selected according to the imaging specifications. The manufacture-related issues are not taken into consideration, because the evaluation and ranking of all the solution 'candidates' cannot be easily determined. The large database provides the user with enough choices, so that the best fitting solution can be finally filtered out by the user according to the specific purpose.

In this chapter, the working principles of both two parts of the GACOR algorithm are introduced, and the brief mathematical modeling method and the necessary evaluation criteria are illustrated. The output systems in the database are denoted as ‘successful solutions’ in the dissertation, regardless of which part of the algorithm.

4.1 GACOR global optimization with structural changes

For nominal design, in practice, the optical designer often decomposes the challenging final specifications into smaller steps and improves the system from a simple initial system gradually until all the requirements are fulfilled. Correspondingly, the original ACOR algorithm should also be modified with such a stepwise optimization idea to realize the automation. In addition, during the stepwise optimization, some structural change options according to the physical considerations are necessary, requesting the algorithm of adaptive adjustment of the structure, MF, and variables.

Additionally, optical designers mostly determine the optimization strategy according to only qualitative analysis provided by the design software with their physical knowledge and experience. In comparison, as a mathematics-based optimization algorithm, all the optimization rules involved in the program can only be executed by the ‘ants’ when they are ‘translated’ to concrete quantitative mathematical models. In other words, the ants need very concrete ‘lectures’ to know how to work on optimization. Therefore, in this section, the translation of the physical guidance for the algorithm is explained.

4.1.1 General workflow

As an overview of the GACOR algorithm, the general workflow is illustrated in Figure 4.1. Before starting the optimization, the initialization of the algorithm includes three aspects, as marked in the blue zone: First, the manual settings of all the algorithm parameters can be adjusted according to the specific optimization problem. Second, the system requirements are denoted as the ‘original specifications’ in contrast to the intermediate goals during the optimization. Most importantly, a starting system should be given in the optical design software, for which Zemax is used. As mentioned in Chapter 2, the initial system strongly influences the final result when the conventional DLS algorithm is applied for optimization. With the help of the structural change, the initial system design becomes easier for optimization with the GACOR algorithm. Thus, assuming the algorithm is completed with the full nominal design procedure, it is sufficient to set up a simple singlet

imaging system as the initial system during the initialization. However, the initial system should in any case contain system parameters like the EnP diameter, wavelength, and FoV, which are not changed during the whole optimization process.

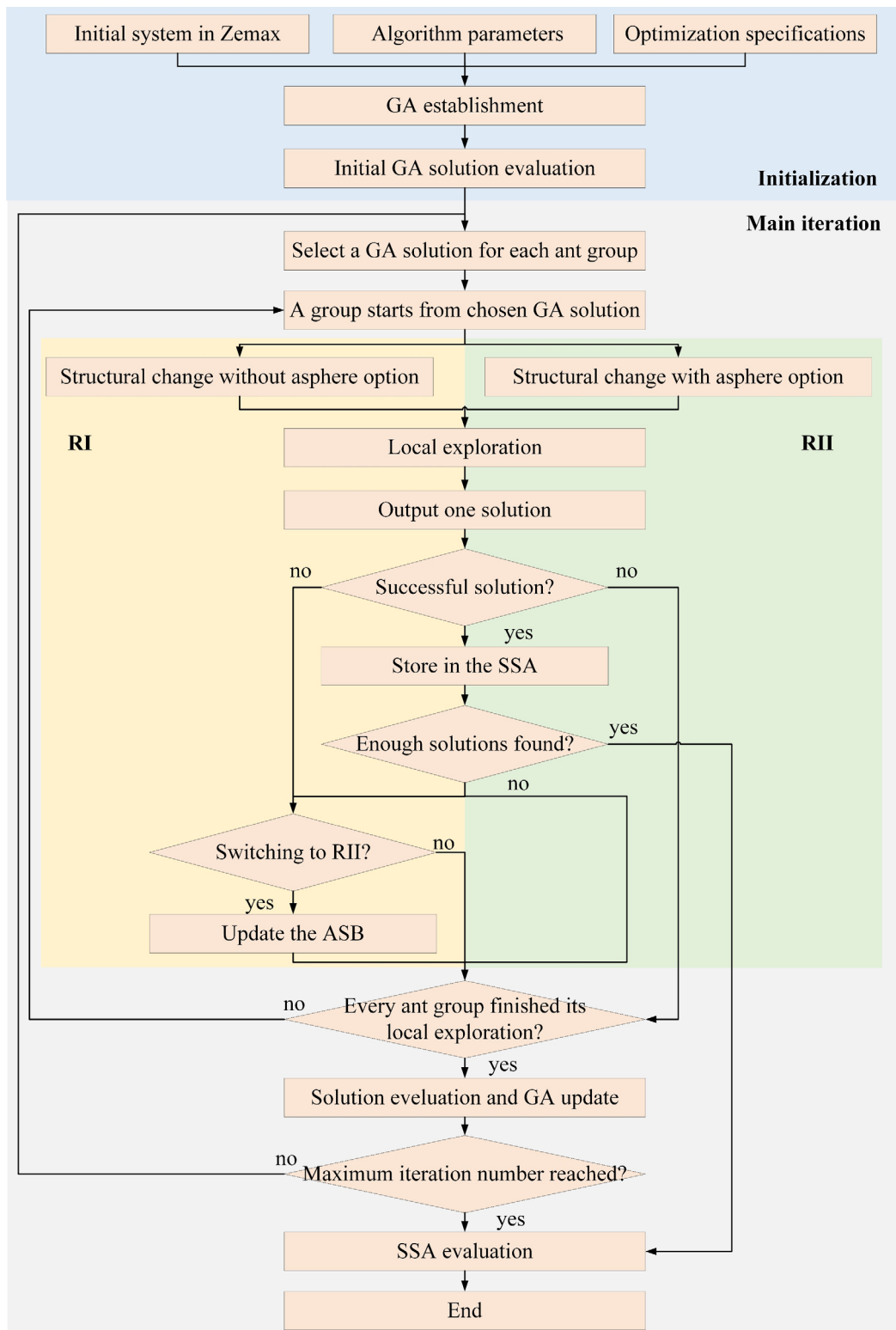


Figure 4.1 General workflow of the GACOR algorithm.

In addition, before the main iteration starts, an archive similar to the ACOR algorithm

is established with the capacity of K_g , named as the ‘global archive’ (GA), and the solutions stored in the GA are called ‘global archive solution’. As for the initialization, only the initial system is stored in the GA and evaluated.

The basic working principle of the GACOR algorithm still follows the ACOR idea. In the main iterations marked in the gray zone, the GA solutions are chosen by ant ‘groups’ instead of ant ‘individuals’. The capacity of GA and ant group number are also adapted during the iterations to improve the efficiency, as given in Appendix C. The first important task of the ant groups is to operate the structural change based on the dynamic physical evaluation. Then, each ant group is responsible for exploring new solutions after the corresponding structural change in the format of an embedded ‘local exploration’ carried out by all the members of this ant group. Each ant group should output one solution after the exploration, which is immediately evaluated. If a successful solution is obtained, it is stored in the so-called ‘successful solution archive’(SSA), otherwise, it is either stored in the GA for further evaluation or deleted. Thus, at the end of each main iteration, the GA includes both the new solutions found by all the ant groups and those obtained before. All these solutions are merged and evaluated according to the original specifications, and the ones which fail to place in the top K_g will be eliminated from the GA. During the main iterations, if enough successful solutions are found, the algorithm stops and outputs the SSA, otherwise, the algorithm keeps running until all the iterations of the pre-set number are finished.

In general, the main iteration process performed by the ‘ant group’ is called ‘global exploration, compared to the embedded ‘local exploration’. To control the optional structural changes more practically, the global exploration is executed in two rounds, namely RI and RII, which will be introduced in the following section. It can be seen from the general workflow, that better cooperation within the whole ‘ant family’ can boost the global searching ability.

4.1.2 Global exploration

4.1.2.1 Structural change options

Compared to the traditional optimization algorithm, one of the essential improvements of the GACOR algorithm is the ‘training’ of the ants to make structural changes. Considering the simplification of this work, only two representative structural change options for spherical aberration correction are included in the algorithm. First, as one of the most

common methods for rotationally symmetric system optimization, lens splitting (called the ‘splitting option’ in the following discussion) is included in the algorithm. Second, as a more powerful correction method for higher-order spherical aberrations, turning a spherical surface into an asphere (denoted as the ‘asphere option’ in this dissertation) is also included. Certainly, changing the glass material is also meaningful in this case. But the glass optimization is discrete due to the limited available glass types, distinguished from the optimization of other continuous lens parameters. Therefore, in this research, this option is not considered for simplification reasons. Furthermore, concerning the main purpose of proving the feasibility of the algorithm, other options such as adding/removing a lens do not greatly influence the results. Thus, they are not included in the research in the current stage as well.

In general, these two structural change options applied in the algorithm either increase the total lens number or the complexity of the surface, which may enlarge the system volume or increase the cost of the system. Therefore, it makes sense to prevent the algorithm from endlessly making structural changes for avoiding unnecessary system complexity. Thus, a parameter of ‘equivalent lens number’ is defined, denoted by $\langle N_L \rangle$, which qualitatively represents the system complexity. Empirically, an aspherical lens can be considered as three spherical lenses referring to the general cost in practice. Following this estimation, $\langle N_L \rangle$ can be calculated as

$$\langle N_L \rangle = N_{Ls} + 3N_{La}. \quad (4.1)$$

where N_{Ls} is the total spherical lens number, and N_{La} is the total aspherical lens number. Thus, for each optimization task, a maximum allowed equivalent lens number, denoted by $\langle N_L \rangle_{max}$, is set by the user as a rough estimation to limit the system complexity.

The same as what an optical designer needs to know about these structural change options, the ants should also learn the important lectures about ‘how to choose from the options’, and ‘when and where to apply the option’. Regardless of the specific structural change option, the modification of the system should be carried out step by step to avoid any undesired large change in the MF topology. Therefore, the GACOR algorithm only allows one kind of structural change on one lens per ‘ant group’ in each main iteration during the global exploration. Consequently, the ant group needs to consider the best location for applying the corresponding structural change. Regardless of the splitting option worked on the ‘lens’ or the asphere option involving only the ‘surface’, the algorithm

always determines the location referring to the optical surfaces. Such a surface where the structural change is finally operated is denoted as the ‘structural change surface’. Automatically, the splitting option is applied to the lens to which the structural change surface belongs. The choice of the structural change option, the rules of the structural change surface determination, as well as the operations of the structural changes developed in the GACOR algorithm, are all explained in Appendix D in more detail.

4.1.2.2 Switch from RI to RII

As one of the common structural change options, the asphere option is helpful to reach the final goal with a simpler system structure instead of introducing too many lenses in the system. However, considering the practical issues, the optical designer usually avoids applying aspherical surfaces in the system in the very early stage of optical design, but prefers first trying out the structural change options involving only spherical surfaces to roughly reach the original specifications. Therefore, as an imitation of such preferences, in RI, the algorithm only allows the optional structural changes while keeping all the optical surfaces in the spherical shape, and the algorithm frequently analyzes the possibility of obtaining a successful solution with only spherical surfaces. If it is assessed that any further improvement with only spherical surfaces is minor, or if a successful system is already obtained in the current situation, it makes sense to allow the applications of aspherical surfaces. The former case indicates a further imaging enhancement with the asphere, and the latter case implies the application of aspheres to explore more alternative solutions. In addition, RI also helps to estimate the minimum number of spherical lenses needed for a successful solution. According to the estimation, a so-called ‘archive solution bank’ (ASB) will be established based on the intermediate solutions obtained in RI as a database to avoid the redundant optimization in RII. In comparison, all kinds of structural changes are allowed in RII, as the asphere option is also included. The specific rules for switching from RI to RII are given in Appendix E.

4.1.2.3 Quantitative Performance evaluation

When the ants find out new solutions, in contrast to the optical designer, they are not able to qualitatively assess the solution flexibly from the most reasonable perspective. Therefore, the performance evaluation of the solution needs to be quantified. However, due to the different optimization stages and individual system structures, it is hard to fix a universal rule for all the solutions. Thus, as a more fair method, the performance of each solution can be assessed from two viewpoints. First, if the solution is optimized only for intermediate

targets as a smaller step to the final goals, the performance according to the intermediate targets should be analyzed, called ‘stepwise performance’. Second, considering the final goals of the optimization, the solution is assessed according to the fulfillment of the original specifications, denoted as the ‘overall performance. In this way, these two kinds of performance can both be quantified with MF calculation simply by adjusting the MF operands, resulting in ‘stepwise performance cost’ (SPC) and ‘overall performance cost’ (OPC). The settings of the MF will be clarified in Section 5.2 combined with an example and Appendix F.

4.1.2.4 Solution evaluation

As the result of the local exploration, each ant group outputs one solution. The quantitative performance assessment helps to determine the destination of the solutions, either being stored or eliminated. The general assessment process is illustrated in Figure 4.2.

As the most important criterion, the performance should be checked first by calculating the OPC value instead of the SPC value, so that the original specifications are emphasized. Only the solution with an OPC value smaller than 10 is considered appropriate for the next evaluation step. The criterion originates from the empirical evaluation of the MF calculation, which is considered meaningful for this purpose. As the next step, the fulfillment of all the original specifications will be checked. If the solution meets all the final goals concerning the imaging performance, it is denoted as a ‘qualified solution’. The ‘qualified solutions must be distinguished from the ‘successful solution’ in the final output, as it is not ensured to be output.

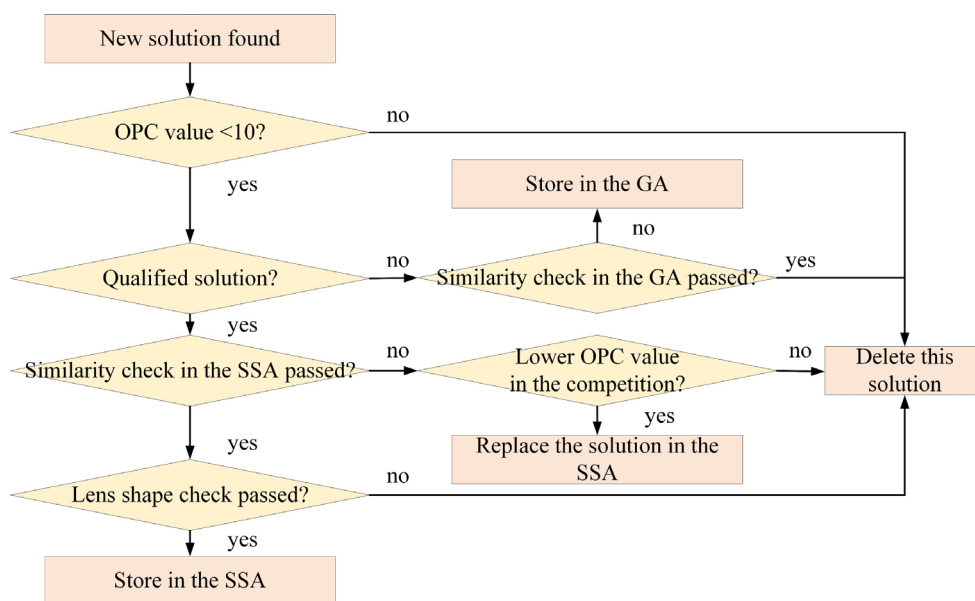


Figure 4.2 Evaluation criteria of the output solution from the local exploration.

Furthermore, to keep the variety of the archive solutions, every solution in the GA or SSA should be unique. Therefore, a so-called ‘similarity check’ is executed for the new solution, which examines if there is any similar solution already existing in the target archive. The qualified solutions should be checked among the SSA, while a not qualified one is checked among the GA. Regardless of the target archive, if a similar solution has already been stored, the two solutions need to compete. Only the one with a relatively lower OPC value is stored in the corresponding archive, while the other is eliminated.

In addition, if a qualified solution is unique in the SSA, before being stored in the SSA, the solution needs to pass the ‘lens shape check’ including two items. First, the lens should not be strongly bent concerning manufacturability. Specifically, a bending parameter is defined as

$$X_j = \frac{r_{1j} + r_{2j}}{r_{2j} - r_{1j}}. \quad (4.2)$$

where r_{1j} and r_{2j} are the radii of curvature of S_j of the two solutions. Second, regardless of some special application design tasks like the cellphone camera, the asphere surface should not be in a waved shape with turning points which can be referred to Figure B.2. The methods of the similarity check and the lens shape check are introduced in Appendix G.

4.1.3 Local exploration

In both RI and RII of the global exploration, after the ant group has performed the structural change for the chosen GA solution, the MF topology mostly also changes due to the varied dimension and boundary conditions. Consequently, the system immediately relocates to a new relative position on the MF landscape. Thus, the local exploration is carried out to search for the local minima around the new position.

4.1.3.1 General local exploration workflow

The local exploration is carried out by each ant group in each main iteration, starting after the structural change is made for the chosen GA solution. The purpose of the local exploration is to find out a local minimum around the solution, which considers the stepwise optimization targets. Figure 4.3 shows the workflow of the local exploration.

First, the system is read from the optical design software directly after the structural change to record the necessary lens parameters. Due to the new structure, the system variables and the MF should be adapted, denoted with index ‘i’ for distinguishment. To

obtain a new solution, there are two essential steps in the process, marked in red. The first step of the local exploration is a simple ACOR process embedded in the global exploration by the ant group, denoted as the ‘ACOR local search’. All the ant group members directly start after the structural change and search around the new position individually following the simple ACOR algorithm, which helps to enhance the solution variety. As the embedded local search process in the general frame of global exploration, only the best solution found by the whole group will be kept.

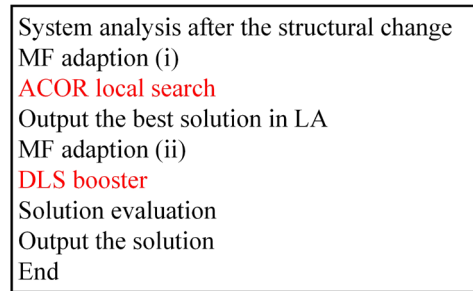


Figure 4.3. General algorithm workflow of the local exploration.

It has been proved that the ants are good at locating the local minimum area, but weak in reaching the minimum point. Thus, as the second important step of the local exploration, the DLS algorithm will be applied to this solution again to further improve the solution until it reaches the local minimum position. This post-processing after the embedded ACOR procedure is called the ‘DLS booster’. Before the process starts, the variables and the MF should be adapted again, represented by index ‘ii’, following the MF adaption rules.

After evaluation, the final solution is considered as the only output of the whole ant group in this main iteration whose performance will be further evaluated. In general, the combination of the probabilistic searching by ACOR and local optimization by DLS method can greatly enhance the efficiency of the optimization by taking advantage of both algorithms. Section 5.2.2 illustrates this process with a concrete optimization example.

4.1.3.2 ACOR local search

As mentioned above, the purpose of the implementation of the ACOR local search is to enhance the global searching ability of the GACOR algorithm. Particularly, even though one GA solution might be chosen by more than one ant group in the main iteration and these ant groups choose the same structural change option, the ACOR local search process based on the probabilistic feature is still beneficial in finding various local minima around the same starting point.

When the ACOR local search starts, following the basic ACOR idea, another archive

should be established for storing the top-ranking solutions ever found, which is denoted as a ‘local archive’ (LA) distinguished from the GA. However, after the exploration, only the best of all the solutions in the LA can be output for the further process. Compared to the original ACOR method introduced in Chapter 2, the ACOR local search algorithm should be adapted to the fixed starting point, of which the details are given in Appendix H.

4.1.3.3 DLS booster

As introduced, the goal of the ACOR local search is to look for different system structures with a fixed lens number after a structural change, which may still vary in focal power distribution, lens shape, or lens position. In comparison, the DLS booster aims to bring the system to the lowest position in the ‘valley’ of the MF topology, so that the best performance in the local minimum area found by the ants can be evaluated.

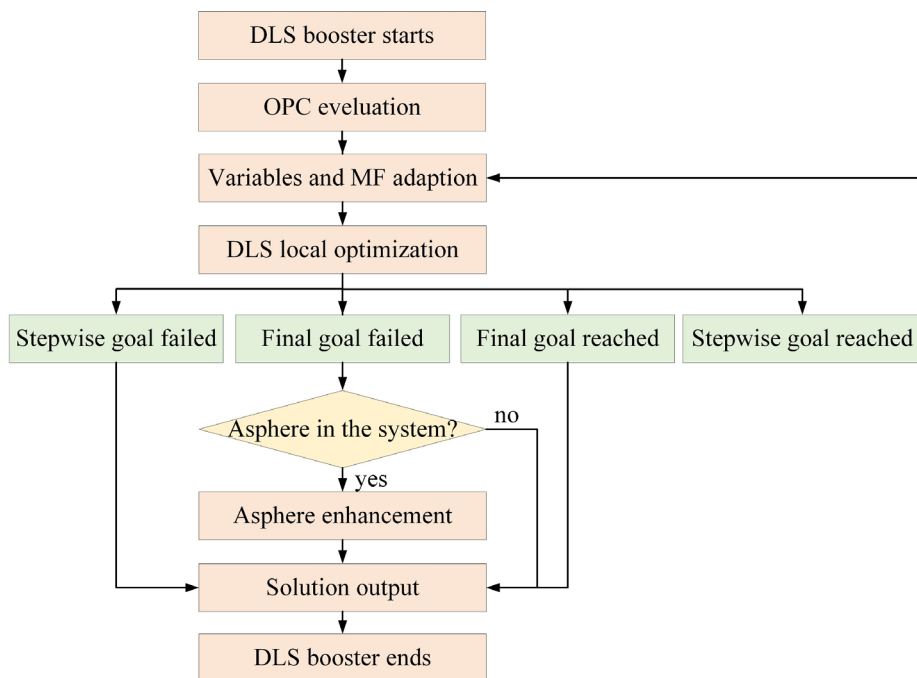


Figure 4.4. Workflow of the DLS booster.

Figure 4.4 illustrates the workflow of the DLS booster process. The solution output by the ACOR local search will be evaluated first. As mentioned above, usually the solution is not yet the local minimum and needs further improvement. Therefore, the OPC of the solution is calculated to decide the adjustment of the MF and variables. In general, there are two possibilities concerning the fulfillment of the specifications: If the system cannot meet the original specifications yet, it needs to stick to the stepwise intermediate targets. In the better case, if the optimization is almost finished, the original requirements of the system as the final targets are set as the current optimization goal. For each case, the DLS

local optimization supported by the optical design software is then applied to the system. The details are explained in Section 5.2.1 combined with the example for better understanding. After optimization, the system is evaluated concerning the SPC, and there are four various possibilities, as marked in green in Figure 4.4.

If the stepwise goal is failed, it is impossible to fulfill the current optimization targets with the system structure, which implies either further structural change is needed, or the structure is not promising. Thus, the system should be output in the GA for evaluation during the main iterations.

If the stepwise goal is reached with the system structure, we can learn that the solution is qualified for the stepwise targets, and potential for higher optimization goals. Therefore, it makes sense to immediately adapt the MF again to level the targets up, so that the local optimization can be applied again with the current system structure before making further unnecessary structural changes in the next main iteration.

If the system is already optimized under the final goals but fails to fulfill all the requirements, further optimization is necessary for the current system. For the system with at least one asphere, in the last phase of the whole optimization process, its performance may still reach the final goals by simply tuning the asphere terms without any additional structural change. In this case, if the algorithm finds any asphere in the system, the so-called ‘asphere enhancement’ will be started. If so, the fine adjustment for aspherical surface terms will be executed to further improve the imaging performance, the details of which are mentioned in Appendix D.

Finally, if the system is successful after the local optimization, the system will be output immediately for further overall performance evaluation.

4.2 GACOR algorithm for final improvement phase

In Section 4.1, the optimization method with structural changes is introduced. As for the second goal of the research to prove the capability of the GACOR algorithm when dealing with very high-dimensional optimization problems, the algorithm developed only for the simplified design tasks is not enough. Therefore, an extension of the main body of the algorithm is developed in the program, concerning the performance final improvement in the fine-tuning phase of optical design.

In this section, the general manual final improvement method applied by the optical

designer is first introduced. Based on that, an extension of the GACOR system is explained in detail, in order to verify the feasibility of the algorithm concerning the very high dimension of the optimization problem.

4.2.1 General final improvement strategy of the optical designer

Considering the specific final improvement methods, each optical designer has his preference and habit in practice. As for the extension of the algorithm in this work, the working principle originates from the author's education, which can differ from other optical designers. Thus, before the introduction of the extended algorithm, the manual final improvement method applied by the author is first explained, so that the essential idea can be better understood. Regardless of the specific method, given the best imaging performance as the final goal, the method shown in this section is only an example.

Despite the variety of final improvement strategies, the optical designer usually first divides the lens parameters into some categories, so that the variables involved in the optimization can be organized more systematically. The lens parameters are divided according to the corresponding type, such as curvature and thickness, and the parameters of the same type are also categorized by the main lens groups indicated by the system structure. Based on the grouping situation, the lens parameter categories can be included step by step as variables. If the final improvement method is rougher, then many of the categories can be turned into variables simultaneously, and in contrast, a more conservative method allows only a few more variable groups at each optimization step. Specifically, the general grouping rules are as follows:

- 1) Within each lens group, all the curvature parameters are in the same category;
- 2) Within each lens group, all the lens thicknesses and the air gaps in between are considered as the same category;
- 3) Each air gap between the two main lens groups, as well as the object and image distance alone, is independent as one category;
- 4) Each curvature or conic constant of the aspherical (including cylindrical) lenses is considered individually as a category;
- 5) All the aspherical surface sag terms of each asphere are together in one group.

During the final improvement process, all the lens parameter categories are included in the variables step by step, and the optimization is carried out after each round of variable

adding until the system imaging performance is qualified. Corresponding to the symmetry of the system, the corresponding Zernike terms with the same symmetry are successively allowed as variables with increasing orders, in the sequence of Zernike terms until Z9, Z16, Z25, and Z36. Given a proper sequence when adding the lens parameter categories as variables, the system can be systematically improved.

4.2.2 Extended final improvement method of the GACOR algorithm

Following the basic idea of the final improvement method introduced above, an extension of the GACOR algorithm is developed. As mentioned, there is a large variety of possible solutions due to the different strategies. Thus, in comparison to only one single output solution in the manual process, the automatic program is capable of creating a large output collection of possible solutions by implementing various final improvement strategies and repeating the process. The output solutions fulfilling the imaging requirements are still denoted as the ‘successful solution’, and the desired number of them is denoted by K_{max} .

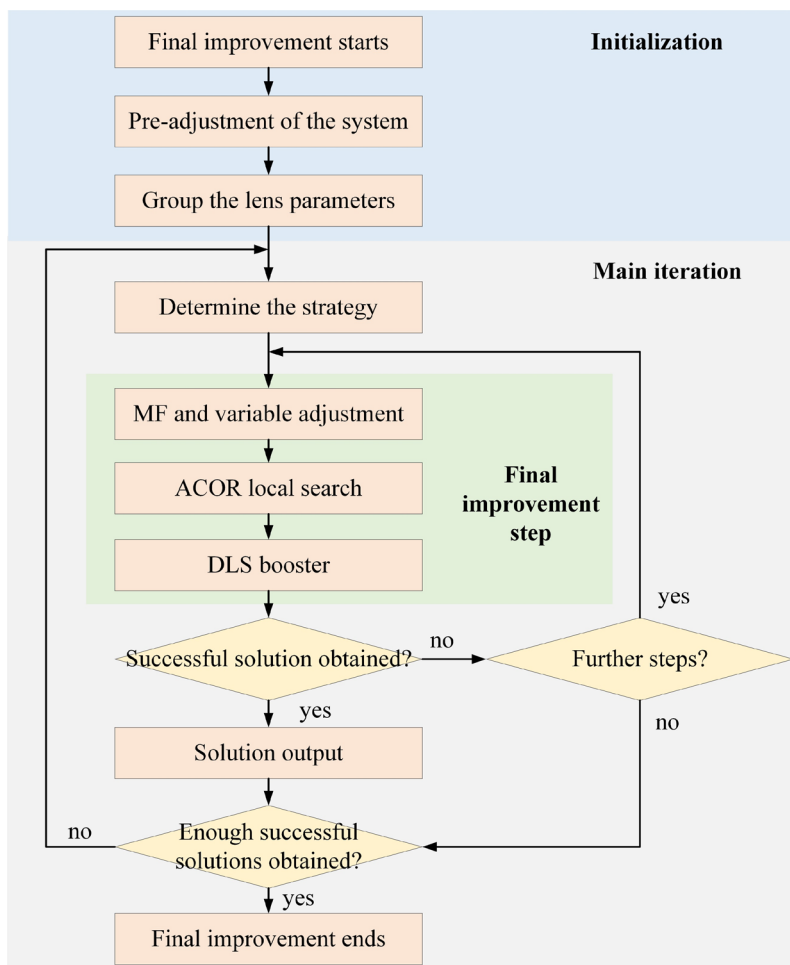


Figure 4.5. Workflow of the final improvement process.

Figure 4.5 illustrates the workflow of the final improvement process. Given the solution from the early optimization phase as the starting point, the system is first pre-adjusted according to the same boundary condition adjustment rules, and the adjusted system is denoted as the ‘starting system’. Then the algorithm categorizes all the lens parameters to generate the final improvement strategy as the essential step of the process. The total number of the lens parameter categories is denoted as M_{g0} . During each main iteration, the algorithm executes a complete final improvement process from the starting system until the final solution, following the strategy specially generated for this process alone.

Similar to the manual final improvement process by the optical designers, the program also optimizes the system by including more variables step by step. In this section, each optimization step regarding the new variable settings is denoted as a ‘final improvement step’, marked in green in Figure 4.5. The generation of the strategy for the starting system refers to the settings of the variables concerning the lens parameter categories. Specifically, the algorithm determines how many, and which lens parameter categories are included as variables in each final improvement step.

Corresponding to whether the final improvement method is more conservative or coarse, the algorithm first decides how many lens parameter categories are included for each final improvement step. The mathematical approach is similar to the approach of choosing the archive of the ant group introduced in Section 4.1, which depends on the probability of all the options. Specifically, assuming M_g is the number of all the available lens parameter categories, which has not been included in the optimization as variables, then the number of those being added as variables at this final improvement step ranges from 1 to M_g . Thus, the probability of the algorithm choosing each possible option is determined by Eq. (2.21), for which the option ‘only adding one more group as variable’ has the first ranking with the highest probability, and ‘including all the M_g lens parameter groups as variables’ option ranks the last corresponding to the lowest probability. With this calculation, the number of newly added variable groups in each step is determined, denoted as ΔM_g . For the next fine-tuning step, the remaining available lens parameter group number M_g is updated with $M_g - \Delta M_g$, and the process is repeated until all the variable groups are turned into variables in the end. Table 4.2 presents an example of the strategy sketch. Assuming there are 12 lens parameter categories, they are represented by G1 to G12. Each column shows the strategy of each final improvement step, and the second row gives the calculation results of ΔM_g in

each final improvement step.

Table 4.2. An example of the final improvement strategy.

	Step 1	Step 2	Step 3	Step 4	Step 5	Step 6
	$\Delta M_g = 3$	$\Delta M_g = 1$	$\Delta M_g = 2$	$\Delta M_g = 1$	$\Delta M_g = 4$	$\Delta M_g = 1$
G1			X	X	X	X
G2	X	X	X	X	X	X
G3	X	X	X	X	X	X
G4					X	X
G5			X	X	X	X
G6						X
G7					X	X
G8				X	X	X
G9	X	X	X	X	X	X
G10					X	X
G11		X	X	X	X	X
G12					X	X

When ΔM_g is determined, the algorithm chooses the lens parameter categories randomly from the available ones with the number of ΔM_g . The ‘X’'s marked red are the newly added variable groups, while the black ‘X’ indicates that in the corresponding step, the corresponding group remains variable. In this example, there are in total 6 steps until all the possible lens parameter groups are included in this example, but it can be different due to the probabilistic generation of ΔM_g .

Following the final improvement strategy, the variables are added, and the boundary conditions are updated in the MF before the optimization starts. Then, similar to the ACOR local search introduced in Section 4.1, a group of ants first locate a local minimum area to enhance the global searching ability. Then the DLS booster is applied for further improvement towards the local minimum position. In this way, the system is optimized step by step, until in principle all the necessary parameters are involved in the optimization. Exceptionally, if a successful system is obtained before all the steps are executed, the final improvement process ends immediately, and the system is stored in the output collection. Furthermore, due to the same practical consideration as mentioned in Section 4.1, the lens shape should be checked before the solution is output.

5 Examples and applications

In the past decades, the rising demand for advanced optical systems motivates the research in modern optical design methods. Therefore, the novel methods introduced in Chapter 3 and 4 are applied in some application-oriented research, and the results are presented in this chapter. In Section 5.1, the practical application of the MRT method for symmetry-free optical system evaluation is illustrated with an example of a freeform lithographic system, where the advantage of the assessment tool is clearly visualized. Then, based on the reliability of the MRT method and the GACOR algorithm with physical guidance, the global optimization results of two example systems with the GACOR algorithm are shown in Section 5.2 and 5.3.

Besides the lithography system illustrated in Section 5.1, the MRT method is also applied to other systems with various structures and complexity, which can be referred to [46] and [47]. In addition, concerning the practical purpose, the imaging performance of a system can be assessed from the viewpoint of both the resolution and distortion. Therefore, as the MRT method is currently only applied for the resolution-related aberration assessment, an additional case study of the distortion correction potential of freeform surfaces is also conducted in this work to cover the field of distortion analysis. The corresponding results of this supplementary research are illustrated in Appendix I.

5.1 Comprehensive aberration analysis with the MRT method

In Chapter 3, the working principle of the MRT method is introduced for surface-decomposed transverse aberration analysis. The calculation results have been verified with Zemax calculation results, proving that the method is reliable for comprehensive aberration analysis concerning symmetry-free systems. To further illustrate the practical usage, the aberration analysis results of a complicated reflective freeform lithographic obtained with the MRT method are presented in this section as an example.

Lithography systems are well known for the difficulty in aberration correction due to the large etendue and superior imaging performance. Many designs can be found using only reflective elements for EUV wavelengths according to the growing request for large NA and field size. Thus, the complicated off-axis structure with non-spherical surfaces of such pure reflective systems cannot be avoided. It has been investigated and proved in present publications, that the as-built performance based on the sensitivity analysis in the optical

systems is strongly determined by the induced effect [48-50]. Specifically for lithographic systems, high-order freeform surfaces are commonly applied to achieve the system specifications. Consequently, the misalignment of the components could introduce considerable induced aberrations into the system, greatly degrading the system performance. Therefore, the sensitivity analysis of lithographic systems is of great concern during the optical design process [51, 52].

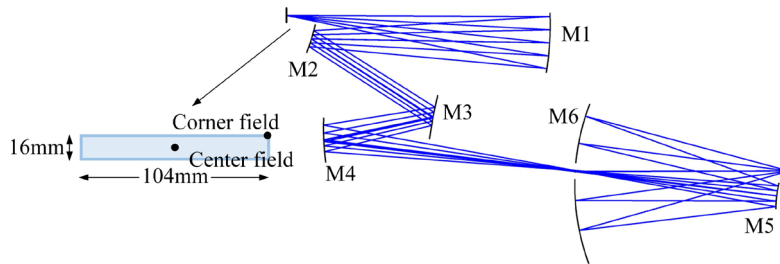


Figure 5.1. The layout of the test lithography system and the object field.

The example lithographic system comprises six freeform mirrors. The layout is shown in Figure 5.1, where M6 has a central obscuration. The working wavelength is 13.5nm with an image space NA of 0.27. The system is chosen from the available patent [53] where lens data can be found. The object is a $16 \times 104\text{mm}^2$ off-axis rectangle. The central field CR is taken as the RR for parabalas matrix calculation, and the upper right vertex of the object is chosen as a representative of the corner fields, marked also in Figure 5.1.

By applying the MRT method, the surface-decomposed full-order total, intrinsic, and induced transverse aberrations can be calculated. In order to gain an overview of the surface contributions of the aberrations all over the pupil concerning one field point, it is beneficial to illustrate the surface-resolved aberrations of multiple sampling rays in one plot. Therefore, the so-called Kingslake plot is applied here, as shown in Figure 5.2 and Figure 5.3. The circle with a radius of 1 in the plot represents the ideal normalized circular pupil, on which the starting position of each arrow has exactly the pupil coordinates of the tested ray. The length and the direction of the arrows illustrate the scaled values and directions of ΔY or Δy of the sampled coma rays. The various colors of the arrows represent different surface contributions. In other words, the plots show the aberration performance of a ray cone coming from the same field. The bar in the lower-left corner indicates the scale for the actual values of ΔY or Δy , which is equal to the largest arrow's length [46].

According to the Kingslake plots in Figure 5.2, for the center field, M5 is dominant for intrinsic aberration among all the surfaces, making M6 suffer from large induced aberrations correspondingly, while the total aberrations of them form a good balance.

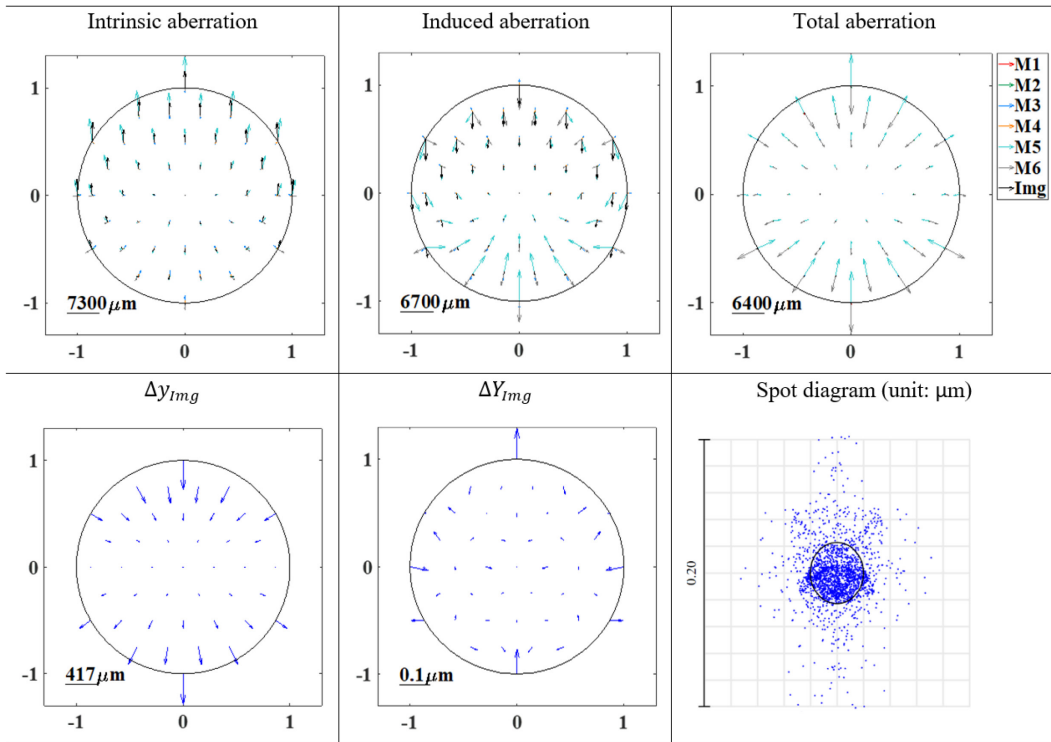


Figure 5.2. Kingslake plots of intrinsic, induced, and total aberration with decomposed surface contributions of the lithography system concerning the center field.

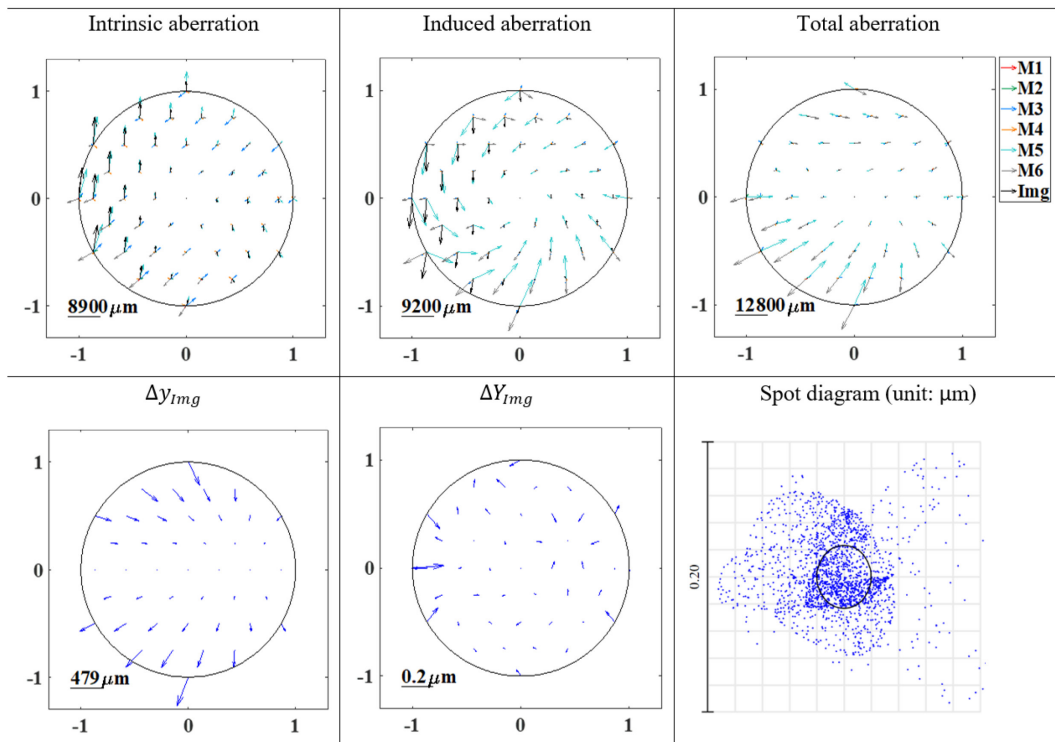


Figure 5.3. Kingslake plots of intrinsic, induced, and total aberration with decomposed surface contributions of the lithography system concerning the corner field.

As for the aberration distributions of the corner field illustrated in Figure 5.3, the asymmetry can be easily seen, and M5 is still dominant. Due to the quite large extension of

the field in the sagittal plane and the large aperture size of M6, its intrinsic aberration becomes larger, but the balance against M5 remains. The results indicate that M5 should bring most trouble when calibrating the position due to the high sensitivity, while M6 is also critical as a compensator with a large aperture.

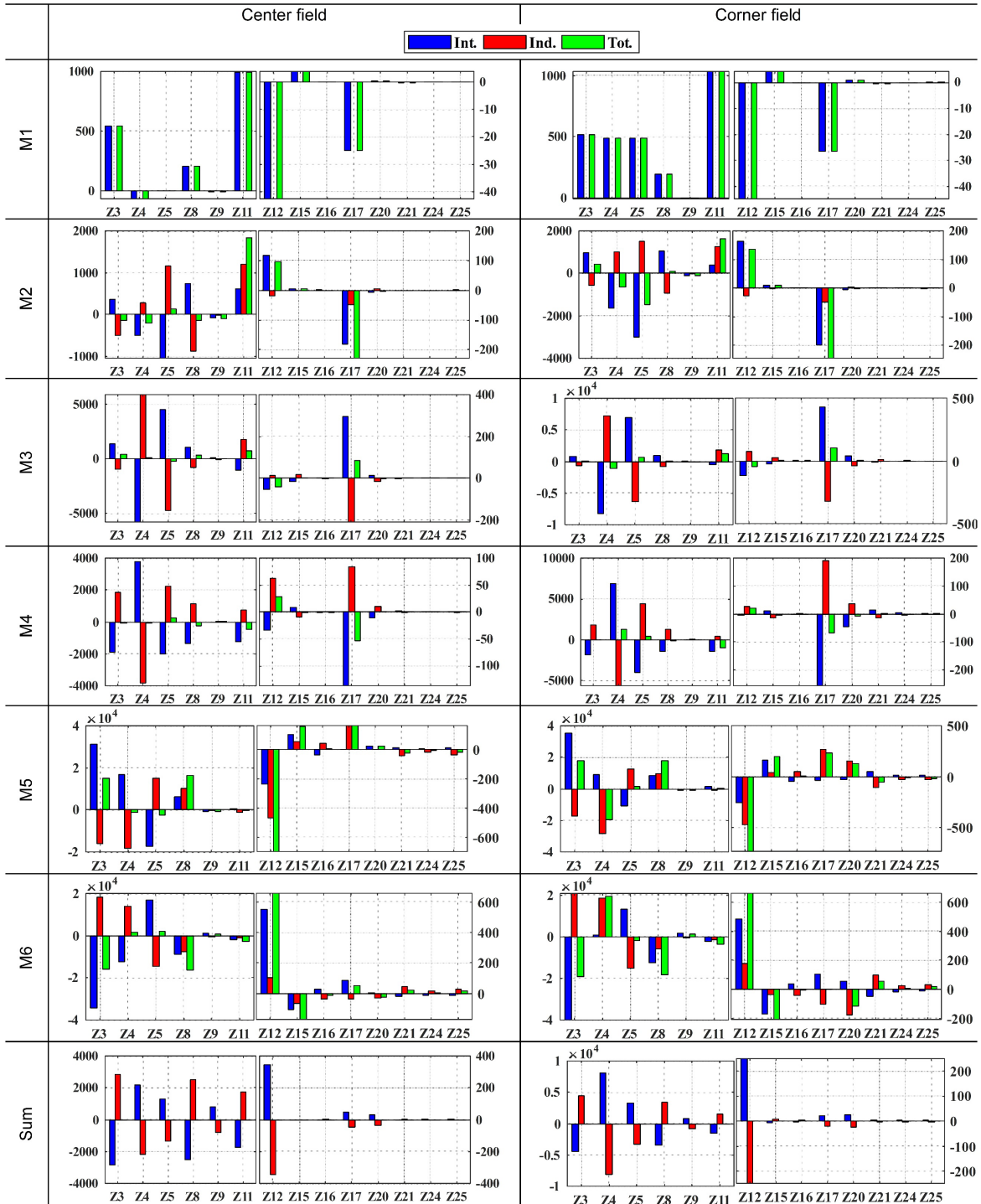


Figure 5.4. Surface-additive Zernike coefficients until Z25 for the lithography system as well as the sum value (unit: waves) [46].

Besides, the surface-additive Zernike coefficients of the total, intrinsic, and induced

aberrations can be also obtained from the transverse aberration with the MRT method, which illustrates clearly the critical aberrations concerning the sensitivity in the system. Considering all the six freeform mirrors of the system, such analysis with additive Zernike coefficients helps a lot to understand the correction of the system. Figure 5.4 collects all the surface contributions of Zernike coefficients, decomposed by intrinsic and induced aberrations. As mentioned, Zernike Fringe terms are used for the fitting. Only the terms until Z25 are illustrated due to the limited size, but in principle, the Zernike terms can be further extended. Since the coefficients of the first six plane-symmetric terms until Z11 are relatively larger than those of higher orders, for better visualization, the lower and higher-order terms are separated with two sets of scales. Besides, for a better comparison between the two selected fields, only the plane-symmetric terms of Zernike polynomials are illustrated in the plots. The other aberrations in the corner field are non-zero, but not shown in the plot due to the limited size.

For each field, the surface contributions are comparable as the fitting procedure considers the same pupil parameters. Consistent with the conclusion drawn from the Kingslake plots, M5 and M6 suffer from the largest aberrations among all the surfaces. As the relation among the total, intrinsic, and induced aberration still exists, the sum of intrinsic and induced aberrations for each term coefficient always equals the total value. Specifically for M1, all the total values originate from the intrinsic aberrations because there is no induced aberration for the first optical component.

5.2 Quasi-automatic global optimization

In Section 4.1, the global optimization method with structural changes of the GACOR algorithm is introduced. In order to test the performance, the GACOR algorithm is implemented to design a retro-focus system starting from an arbitrary simple system.

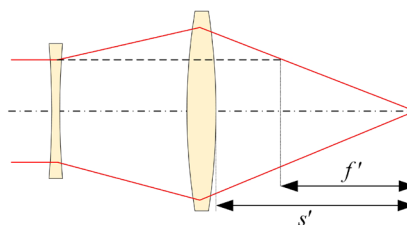


Figure 5.5. Layout of the retro-focus system [54].

The retro-focus structure is one of the typical system structures applied for camera objectives. The basic layout is shown in Figure 5.5, which contains a negative lens group

in front and a positive lens group behind. The collimated incoming beam is first expanded by the negative lens and then converged by the positive lens. Thus, according to the definition of the principal plane, the focal length f' is shorter than the free working distance s' , which is the distance between the center of the last optical surface and the image plane. To describe the relationship between these two parameters, a retro-focus factor R_F is defined as

$$R_F = \frac{s'}{f'}. \quad (5.1)$$

As for the optimization task, the system specifications are listed in Table 5.1. Among all the system parameters, the image space NA and the retro-focus factor R_F are considered as the structure requirements which need to be optimized step by step. Together with the spot size requirement, there are in total three requirements in need of dynamic adjustment and optimization during the whole design process, while all the others are fixed either in the system properties or by the MF.

Table 5.1. System specifications of the retro-focus system.

Entrance pupil diameter	20mm	Image space NA	0.4
Wavelength	550nm	Retro-focus factor	3
Stop position	L1 front surface	Field of view	On-axis field only
Total length	Maximum 200mm	Image performance	Diffraction limited

To be diffraction limited for the imaging performance, the RMS spot size should be smaller than the Airy diameter, calculated by

$$D_{Airy} = \frac{1.22\lambda}{NA}. \quad (5.2)$$

As mentioned in Chapter 4, it is hard to formulate the universal manufacturability criteria independent of the specific situations. Therefore, in this work, for an algorithm aiming at the nominal optimization phase, the detailed optimization targets for controlling the sensitivity are not included in the MF.

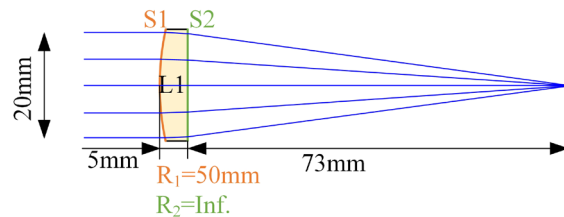


Figure 5.6. The layout and the parameters of the initial system.

The initial system is a single lens system with a collimated incoming beam, as illustrated in Figure 5.6. The system parameters are fixed in the optical design software. The lens is a plano-convex lens made of SF12. As glass optimization is not investigated in this work, for simplification reasons, it is assumed that all lenses are made of the same glass material during the optimization. The image distance is simply optimized for the smallest RMS spot size.

As there are a lot of systems involved in this section to present the optimization results, it should be clarified for the following sections, that the specific surfaces of the system are always denoted as ‘S’ + ‘optical surface number’, and the specific lenses are always denoted as ‘L’ + ‘lens number’. Both the optical surface number and lens number are counted from the left to the right. As an example, for the system shown in Figure 5.6, the lens and all the surfaces are marked following the notation rules.

5.2.1 Optimization strategy for retro-focus systems

Given an optical system design task, the optical designer usually figures out a general strategy before starting the optimization process concerning the realization of all the specifications with his experience. But because the optimization is extremely nonlinear with considerable local minima, the results obtained during the process cannot be predicted. Thus, an experienced optical designer usually adjusts the strategy dynamically according to the current results so that the optimization process is kept in the right direction. In comparison, the GACOR program is completely executed by the ants, and training all the ants to be as flexible and experienced as real optical designers is not realistic. Therefore, corresponding to the optical designer’s experience, the GACOR algorithm is developed with the archiving mechanism which makes the ants in the wrong direction ‘disappear’.

The general optimization strategy is the most important ‘lecture’ for the ants and is essential for the success of the optimization. In Chapter 4, the general optimization workflow of the GACOR algorithm has been introduced, but concerning the specific system type, the detailed optimization strategy can differ a lot. Therefore, only the optimization strategy for the retro-focus system applied by the author is introduced in this section. It should be mentioned that the appropriate strategy may not be unique.

5.2.1.1 MF adjustment rules

The optimization strategy is mainly reflected by the adjustment of the MF. Specifically for the retro-focus system, in the MF, the three main requirements can be represented by the

corresponding operands, which are specified with weighting and target values. The weighting and target for the NA operands are denoted as W_{NA} and T_{NA} , and the ones for the R_F operands are W_{RF} and T_{RF} . The corresponding operands for the spot size in the MF are taken from the default settings controlling the transverse aberration of each sampled ray, so that the weightings and targets are also automatically set. Considering the possible aspherical surfaces in the system, the MF samples the rays with 6 arms and 20 rings over the pupil with Gaussian quadrature. With the systematically formulated MF, the dynamic assessment of the system can be tackled easily when necessary, according to the current spot size, NA, and R_F values, denoted as V_{RMS} , V_{NA} and V_{RF} . Among all the requirements for the retro-focus system, the spot size is the most important criterion concerning the imaging performance. Thus, the corresponding default MF operands for the spot size are not changed and are always activated during the whole optimization process, while T_{NA} and T_{RF} are decomposed into several stepwise values. In this task, the NA targets should be increased from 0 to 0.4 with an interval of 0.1, and the R_F target is increased from 0 to 3 with an interval of 1.

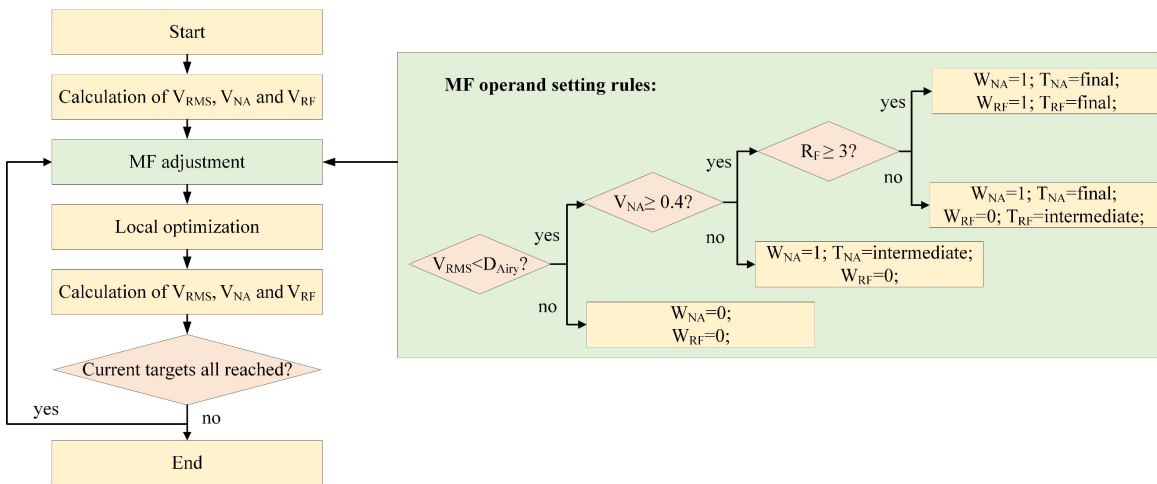


Figure 5.7. Workflow of the DLS local optimization (left) and the setting rules of the MF (right). The ‘final’ targets represent the corresponding value of the original specification, and ‘intermediate’ refers to the intermediate target values before the system reaches the original specification.

Experience also shows, that the corresponding operands of the three requirements should not be always activated simultaneously during the optimization. Instead, the better strategy is to assign priority to them for keeping a smooth improvement of the system. Especially for the automatic optimization program, the proper MF adjustment rules are essential to ensure successful solutions, which must be very clear and rigid for the program. Therefore, the corresponding expressions in the program should be formulated carefully. The right

part of Figure 5.7 illustrates the MF setting rules of the corresponding operands of the specifications.

As mentioned, before the system becomes diffraction limited, the algorithm only optimizes the spot size with the help of structural changes. When the system is diffraction limited, the NA requirement as the second priority is included in the MF by switching the weighting value to $W_{NA} = 1$. According to the V_{NA} value, the GACOR algorithm sets the T_{NA} value as the next intermediate value that the system has not reached. Mathematically, it is represented as

$$T_{NA} = \begin{cases} 0.1 & 0 < V_{NA} < 0.1 \\ 0.2 & 0.1 \leq V_{NA} < 0.2 \\ 0.3 & 0.2 \leq V_{NA} < 0.3 \\ 0.4 & 0.3 \leq V_{NA} \leq 0.4 \end{cases} \quad (5.3)$$

As the last requirement included in the MF, the R_F operand is not activated until the NA and spot size reach the requirements simultaneously, and the target should be set according to the current V_{RF} value as

$$T_{RF} = \begin{cases} 1 & 0 < V_{RF} < 1 \\ 2 & 1 \leq V_{RF} < 2, \\ 3 & 2 \leq V_{RF} \leq 3 \end{cases} \quad (5.4)$$

It is important to mention that the MF setting rules are universal for all the retro-focus system optimization tasks with various NA and R_F . Only the stepwise target values may vary. The MF setting rules introduced in this section are corresponding to the general MF adjustment rules introduced in Appendix F, and the target values and weightings of the operands of the three requirements are regarding the T_p and W_p in Figure F.1.

5.2.1.2 DLS Local optimization

The DLS local optimization is an important step for the DLS booster introduced in Section 4.1.3.3, as it explores the best solution in the current local minimum area. The workflow is also shown in Figure 5.7. The GACOR algorithm operates the program, while the local optimization of the system is executed by the optical design software.

Before the local optimization starts, the system is assessed according to the specifications, so that the MF can be adjusted correspondingly. Then the optical design software starts the local optimization with DLS algorithm until the stopping criteria are reached. To check the improvement of the system, the same assessment is performed again

after the optimization. If all the current targets are reached, it makes only limited sense to stop the optimization process and directly output the system for further structural changes in the next main iteration. Thus, imitating the optical designer, the algorithm adjusts the MF again to the next optimization step and immediately starts the local optimization to improve the system towards the final goals. This loop ends if either the system reaches all the original specifications, or any one of the currently activated targets is not reached, as such a case indicates that the current system structure is not good enough to reach all the goals. More concrete explanations will be illustrated in Section 5.2.2 combined with the system evolution.

5.2.2 Local exploration for one ant group in one main iteration

As introduced in Section 4.1, each ant group chooses one GA solution as the starting point, and only outputs one solution after the ACOR local search for the DLS booster to further optimize it. Because of the probabilistic feature of the algorithm, it happens often that the best-ranked GA solution is chosen by many ant groups, while the less well-performing solution has a lower chance to be selected. Even though more than one ant groups select one GA solution, the ACOR local exploration can still lead them to different local minimum areas by the individual search of each member in the ant group. Consequently, the final output solutions are completely different so that the global searching ability is greatly enhanced. In this section, such a case with various optimization paths is illustrated.

To avoid the large diversity of the optimization paths and the complexity of the optimization process, the first main iteration starting from the initial system is chosen as an example to illustrate all the intermediate solutions during a main iteration. Due to the simple MF topology of the initial system, it turned out after sufficient times of execution, that there are only two possible solutions as the final output of the ant group, when only the splitting option is allowed for structural change. Thus, Figure 5.8 illustrates the evolution of the solutions in different optimization paths as a comparison, where each solution is specified with a name. But it should be clarified that each ant group can only follow one evolution path during one main iteration.

As shown in Figure 5.8, in the first main iteration, the optimization starts from the initial solution **a**, the lens data of which have been given in Figure 5.6. Due to the limited lens number, the splitting option is the only allowed structural change. Following the corresponding rules for lens splitting explained in Appendix D, solution **b** is obtained. Then,

the program executes the simple ACOR local search to find out more possible local minimum areas. Due to the strong probabilistic feature of the ACOR algorithm, the output solutions can be similar, but always slightly different in lens data. To illustrate the possible solutions, five various solutions with visible differences found by five different ant groups are shown, all denoted with **c**, and they are considered as the starting points of the following DLS booster process for further improvement.

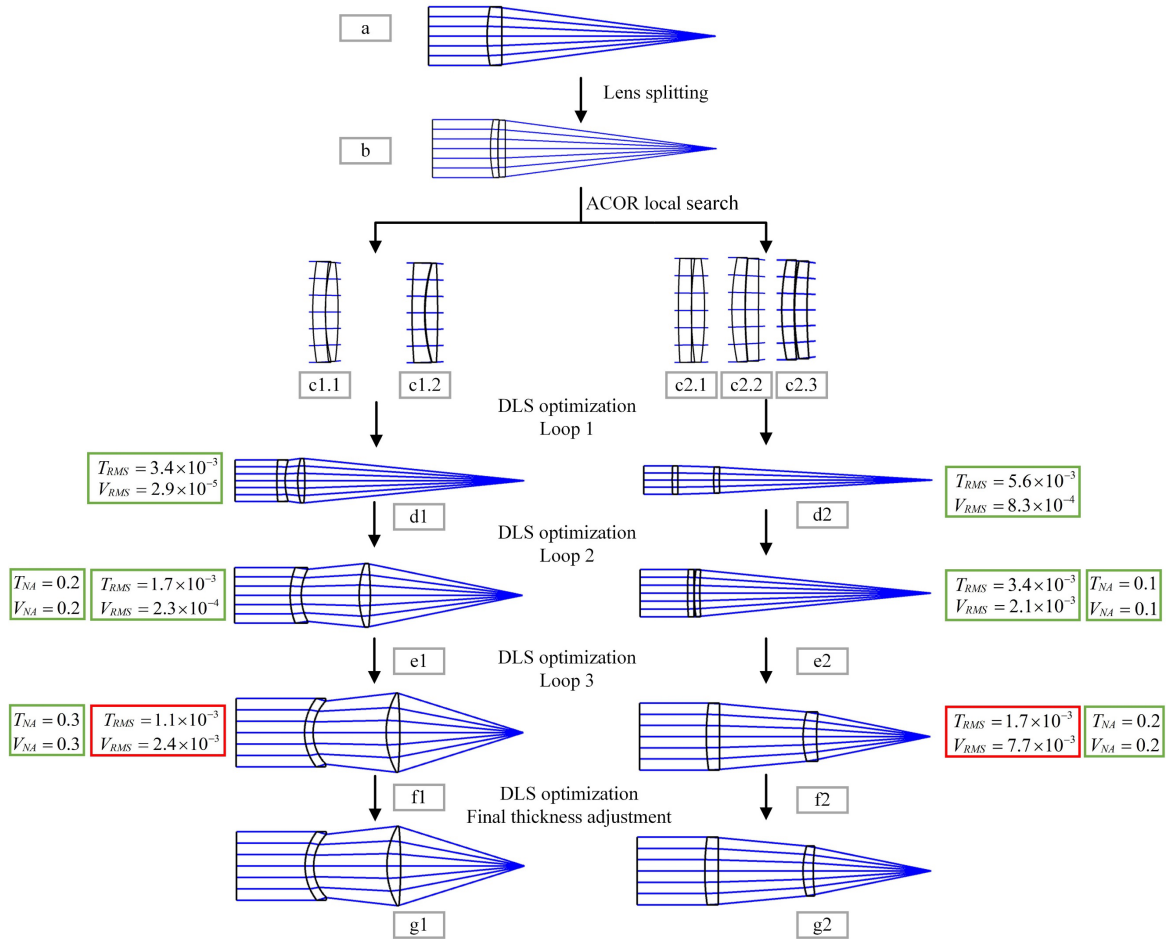


Figure 5.8. Intermediate solutions during the first main iteration. The current targets and values of the spot size (unit: μm) and NA in the MF are also given.

Although the five output solutions of ACOR local exploration are all different, they converge to only two optimization paths during the DLS booster process which are from **c1** to **g1** (left) and from **c2** to **g2** (right). As for this specific case, the large difference in the optimization paths is due to the different focal power distribution of the two lenses, which is either a ‘negative + positive’ combination (solution **c1.1** and **c1.2**), or a ‘positive + positive’ combination (solution **c2.1**, **c2.2**, and **c2.3**). If the first lens is negative, the DLS optimization keeps the negative value, and the final output solution can only be **g1**, while in the other situation only **g2** can be obtained. This result also corresponds to the

understanding of the drawback of the DLS local optimization algorithm, that the sign of the focal power cannot be easily changed. In the case of more complicated systems, there can be many more different optimization paths, which finally result in a larger variety of output solutions.

In addition, it should be clarified that considering the various aberration distributions, there should be other possible solutions with different lens bendings besides **g1** and **g2**. However, in this case, such solutions cannot be easily found by the ants, corresponding to very ‘hidden’ local minimum areas on the MF topology from the perspective of the starting position for the local search. Such a feature of the ACOR searching is considered as a part of the algorithm, which cannot be avoided. However, if necessary, more lens splitting methods can be added to the algorithm to enhance the deviation, so that the ants have a higher chance to find more local minimum areas leading to other solutions.

Taking a closer look at the process of the DLS booster, **g1** and **g2** are obtained in very different routes. According to the optimization strategy, the MF is first adjusted before the DLS local optimization starts. Among the three main requirements, the activated targets and the actual values after optimization are also given in Figure 5.8 to better explain the workflow introduced in Figure 5.7. As the system is usually not diffraction limited due to the weak local optimization ability of the ACOR algorithm, the first DLS optimization loop only focuses on minimizing the spot size. In other words, neither the NA nor R_F operands are activated. After the intensification, both **d1** and **d2** are diffraction limited, meaning the only activated MF target is fulfilled. Thus, the DLS optimization will keep proceeding to a second loop, where the NA operands are included. Due to the different V_{NA} values of **d1** and **d2**, the T_{NA} values for them in the second loop are different. After the optimization, both **e1** and **e2** fulfill both two targets, so the process still goes on. In the third loop, the NA target is further increased, leading to the solutions **f1** and **f2** after optimization. The performance evaluation indicates that both of them succeed in reaching the NA target, but the system is not diffraction limited anymore. In this case, the DLS local optimization for imaging performance is stopped here, but only the lens thicknesses and air gaps are finely adjusted according to the updated boundary conditions to prevent the inappropriate thickness due to the changing diameter after the optimization.

5.2.3 Solution evolution

During the optimization process, as the complexity of the system increases, the number of local minima in the MF topology grows greatly in the later stages. In addition, the evolution of the systems during the optimization by the GACOR algorithm cannot be reproduced, as the algorithm is strongly based on probability and random decisions. Consequently, it is impossible to collect all the solution evolution cases, as the number of optimization paths is considerable. Therefore, in this section, a part of the evolution among the whole picture is chosen to illustrate the large variety of the solution and the impact of each main step of the algorithm. All the results are obtained during one execution of the program with the GACOR algorithm.

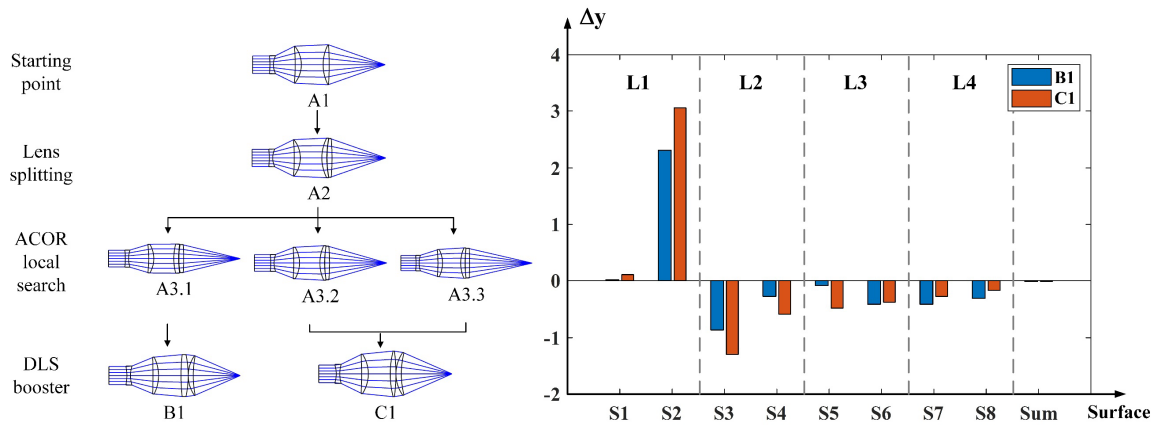


Figure 5.9. First partial solution evolution map (left), and the comparison of surface-decomposed transverse aberrations (unit: mm) calculated with the MRT method between solution B1 and C1 (right).

The left part of Figure 5.9 shows the partial evolution starting from 3-lens solution **A1** to two 4-lens systems **B1** and **C1** in RI. **A1** is directly evolved from system **g1** in Figure 5.8, but the detailed optimization path is not shown here. In this path, the algorithm chooses L3 to split, leading to **A2**. During the whole program, the path from **A1** to **A2** occurred three times, ending in three different output solutions after the execution of the ACOR local exploration, denoted as **A3.1**, **A3.2**, and **A3.3**. In contrast to the solutions **c1.1** - **c1.5** in Figure 5.8, these solutions all have the same focal power distribution, and the layouts still do not show a big difference. However, after the further improvement by the DLS optimization, **A3.1** finally evolves to **B1**, while the optimization of **A3.2** and **A3.3** both converge to the output solution **C1**.

It can be seen in this example, that although the locations of the similar solutions found by the ACOR local search may be very close on the MF topology, they may belong to

different local minimum areas. Consequently, the final output solution can be completely different after the DLS booster. Concerning plenty of ant groups, the repetition of the searching around one starting system is the essential reason for the global searching ability of the algorithm.

As the optimization task only concerns the correction of the spherical aberration and there is no off-axis field point, the surface-decomposed aberration distributions of the solutions can be represented by the MRT method calculation results considering only the MR. The transverse aberrations of **B1** and **C1** are plotted in the right part of Figure 5.9 for comparison. The general surface contributions are comparable, while the deviation among the surface contributions of **B1** is clearly smaller than **C1**. The comparison implies that the sensitivity of **B1** is less critical concerning further optimization.

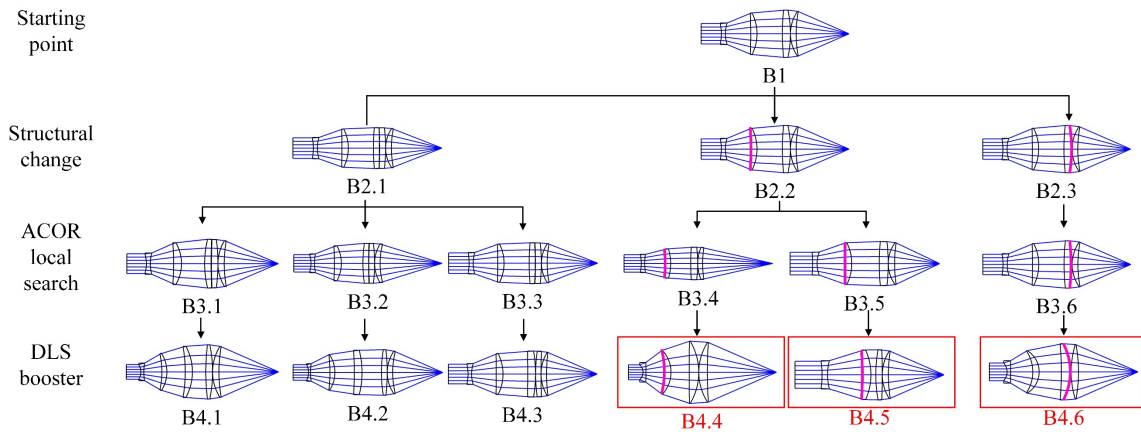


Figure 5.10. Second partial evolution map of solutions. The aspherical surfaces are marked pink and the successful solutions are marked red.

As the obtained solutions **B1** and **C1** are still not diffraction limited, they will be considered as the starting points for further optimization. Figure 5.10 shows the further evolution of **B1** recorded in the later optimization process. During the program execution, the further evolution only happened in RII, when both of the structural change options were allowed. Therefore, the splitting option applied for L3 in **B1** leads to system **B2.1**. The systems **B2.2** and **B2.3** are obtained because of the asphere option applied on S3 and S6, marked in pink. The same as already mentioned, these three solutions after the structural change cannot represent all the possibilities. The reason behind this is again the probabilistic feature of the algorithm. Given another execution of the program, the obtained solutions can be completely different.

At the end of these optimization paths illustrated in Figure 5.10, **B4.1**, **B4.2**, and **B4.3** are still not successful, and they may have some clear drawbacks. For example, L3 in **B4.3**

seems redundant without much correction contribution. However, the algorithm still keeps them, as they still can be promising starting points for further optimization. In addition, the asphere can be also added earlier in the system in principle to reduce the number of spherical lenses, compared to the structural change in **B2.2** and **B2.3**. Such an optimization path will be discussed in Section 5.2.4.

As introduced in Appendix D, the algorithm first decides the structural change option and then determines the surface for operating the structural changes. According to the rules, the splitting option considers Seidel contribution, incidence angle, and the MR height simultaneously, while the asphere option for **B1** only considers the Seidel contribution to determine the structural change surface. Dependent on the structural change option, the critical indices for all the candidate surfaces are calculated and plotted in Figure 5.11.

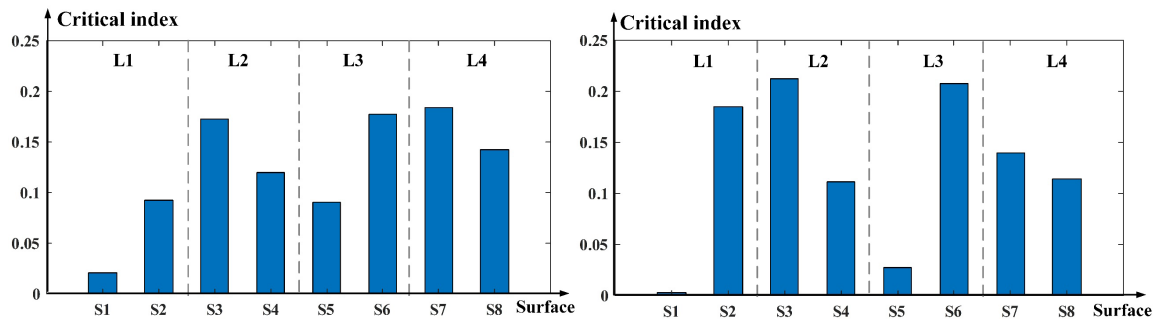


Figure 5.11. Critical index (u.a.) of B1 for splitting option (left) and asphere option (right).

According to Figure 5.11, the surface S2 of solution B1 clearly has the largest aberration contribution because of the large divergence of the rays made by the only negative lens in the system. In this case, the author would preferentially choose the lenses with a larger diameter for lens splitting, so that the more critical positive lens group can be optimized better. Therefore, for this optimization task, the critical indices are scaled with the squared lens diameters. Consequently, when **B1** is considered as the starting point for further structural change, despite the largest aberration contribution of S2, its critical impact should be weakened. Otherwise, the algorithm tends to only split the negative lenses in the front group during the whole process because of the much larger possibility to be chosen. In this case, the algorithm would keep splitting the negative lenses until the large ray divergence is shared by a greatly expanded negative lens group. Thus, restricted by the lens number and total length, the final solutions are not so satisfactory concerning the variety and the system structure. Therefore, in this work, the strong scaling of the lens diameter makes a certain sense. However, in general, the mathematical modeling of the critical index

calculation is not unique.

As for the splitting option corresponding to the left plot, S3, S6, and S7 are the most critical among all the surfaces, which belong to L2, L3, and L4 respectively. In this case, the algorithm has a larger possibility to choose them for splitting. In comparison, if the algorithm chooses the asphere option, according to the corresponding calculation rules of the critical index, S2, S3, and S6 become the most critical surfaces that are of higher probability to be turned into an asphere. Consequently, although each surface has the chance to be chosen, the relatively more critical surfaces would more likely be chosen due to the probabilistic feature of the GACOR algorithm. In this example, although the solutions with split L1 are not found, in principle it is possible, and such structural change is found in other optimization paths during the program execution.

Comparable to the cases mentioned above, the solutions **B3.1**, **B3.2**, and **B3.3** after the ACOR local search are also similar, but the three different solutions **B4.1**, **B4.2**, and **B4.3** after the DLS booster are very different. The results prove again that the embedded ACOR algorithm is very helpful for locating various local minimum areas, and the variety among them indicates that the MF topology is already highly complicated even though there are only 5 lenses in the system.

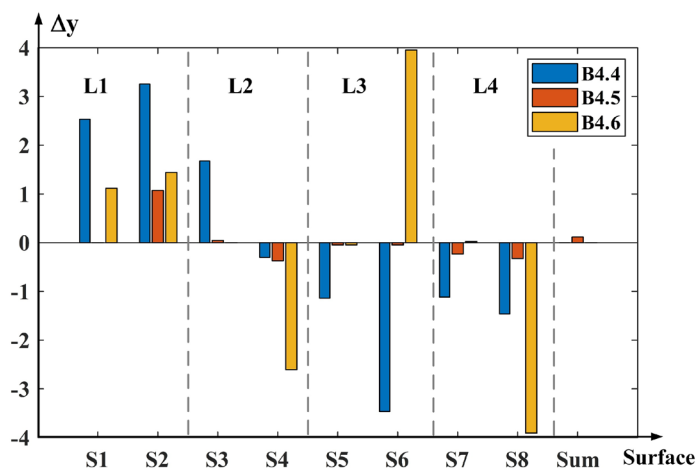


Figure 5.12. Comparison of the surface-decomposed transverse aberrations calculated with the MRT method among solutions B4.4, B4.5, and B4.6 (unit: mm).

Besides, when the asphere locates at S3, the ACOR local exploration also finds out two local minimum areas resulting in two different successful solutions with the same asphere position. When the asphere locates at S6, the algorithm finally finds out a successful solution **B4.6**. The results prove that the algorithm can deal with the higher-dimensional optimization problems with aspherical parameters to some extent. Figure 5.12 again

illustrates the surface-decomposed transverse aberration of the MR calculated with the MRT method. As a good indicator of the sensitivity of the system, the deviation of the aberration distribution of **B4.6** and **B4.4** is relatively higher. The results can be also predicted according to the layouts, where the rays are unevenly distributed over the lenses in **B4.4**, and L2 in **B4.6** is strongly bent. In comparison, the rays clearly have a smaller change in height and angle in **B4.5**, corresponding to a much better balance of the aberration distribution. As it is hard to realize a fast and accurate surface contribution analysis for an aspherical system with the conventional methods, this example illustrates a good combination of the MRT method and the GACOR algorithm concerning the aspherical system optimization.

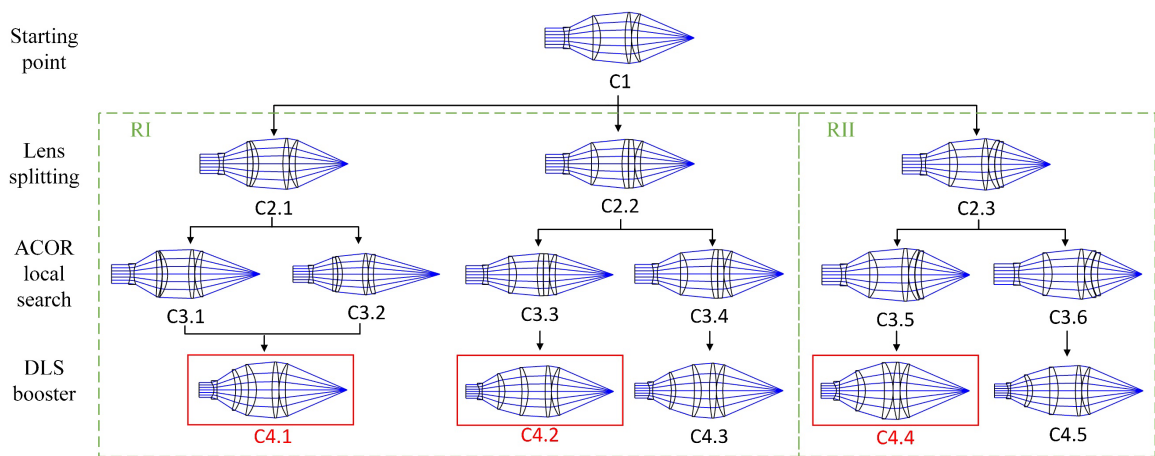


Figure 5.13. Third partial solution evolution map. The aspherical surfaces are marked pink and the successful solutions are marked in red

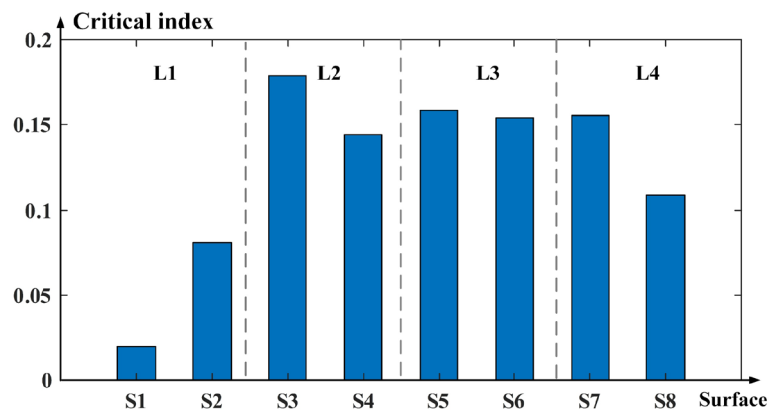


Figure 5.14. Critical index (u.a.) of C1 for splitting option.

The last example illustrated in Figure 5.13 shows the optimization paths starting from solution **C1**, also obtained from **A1** in Figure 5.9. L2, L3, and L4 of **C1** were chosen for the splitting option during the whole program. According to the critical index calculation results plotted in Figure 5.14, except for S1 and S2, all other surfaces are quite comparable,

indicating similar probabilities for them to be chosen for lens splitting, which well explains why the three lenses are all found split.

It is worth mentioning that **C2.1** and **C2.2** are found in RI, while **C2.3** is found in RII. Specifically, **C4.1** is the first successful system ever found in RI during the program, which marks the time point of the switch from RI to RII. The details of this switch are discussed in Appendix E.

Concerning the final solutions, both **C3.1** and **C3.2** converge to **C4.1** after the DLS booster although the difference in the layout can be easily observed. From **C3.3** to **C3.6**, each local minimum area found by the ACOR local search leads to a unique solution after the DLS booster, but only **C4.2** and **C4.4** are successful. Figure 5.15 compares the surface contributions of the transverse aberration again among these three successful solutions. **C4.1** clearly has the largest absolute aberration values among all, implying the most critical manufacturability, while **C4.2** and **C4.4** are both less critical. However, as there is no asphere in these solutions, the general distributions of the additive aberrations are comparable, which can be understood with the same starting system **C1**.

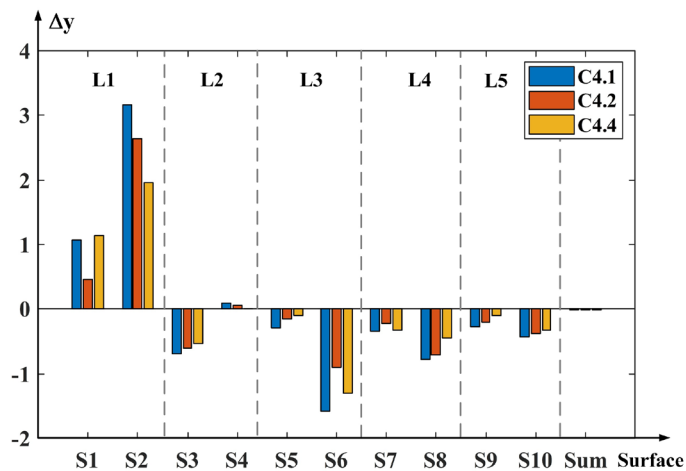


Figure 5.15. Comparison of the surface-decomposed transverse aberration calculated with the MRT method among the solutions C4.1, C4.2, and C4.3 (unit: mm).

5.2.4 Analysis of the output solutions

As mentioned in section 4.1.3, although the qualified solutions fulfill all the performance requirements, they still need further checks on the lens shape to ensure acceptable manufacturability. In this section, two examples of the evaluation process are illustrated.

Figure 5.16 shows a partial evolution map, where the starting system **h1** also evolved from **A1** with the asphere option. By splitting L3, **j1** is obtained, while on the other

optimization path, S5 is turned into an asphere in addition to aspherical S4, resulting in two aspheres in **j2**. After the local exploration process, **k1** and **k2** are obtained. First, the bending parameter X of both solutions is calculated according to Eq. (4.2) and plotted in Figure 5.16. Considering the limit of the bending parameter $|X| = 10$, solution **k1** has a very moderate bending, while the bending of L2 in **k2** reaches almost -20. It can be also seen from the layout, that the strongly bent L2 is critical for manufacture. In addition, considering the surface shape of the aspheres, the second asphere of **k2**, S5, has a clear turning point corresponding to the surface sag plot lower right in Figure 5.16. Therefore, concerning the solution evaluation rules, **k2** is neither appropriate in bending nor qualified in asphere surface shape. Consequently, in this case, **k1** is qualified as a final output successful solution, and **k2** will be eliminated.

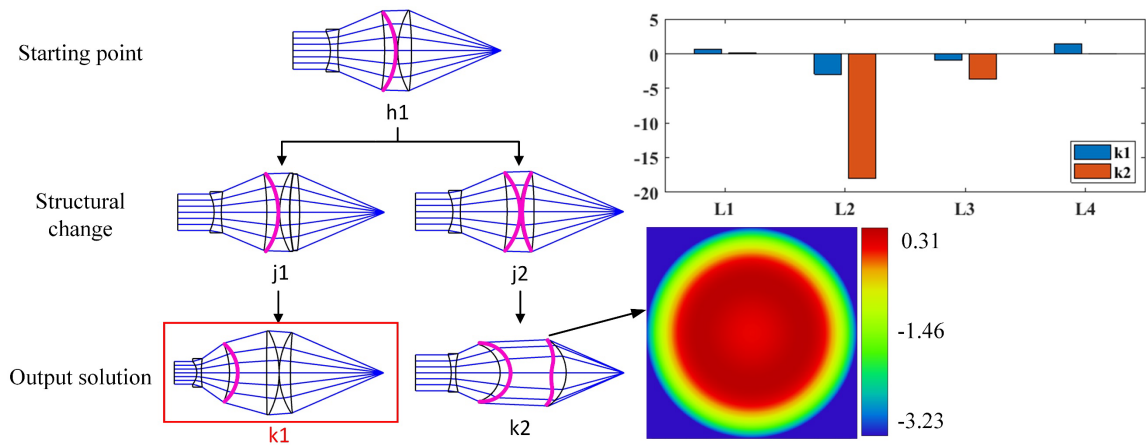


Figure 5.16. Fourth partial solution evolution map (left). Comparison of lens bending parameters of **k1** and **k2** (upper-right). Surface sag plot of S5 in **k2** in mm (lower-right).

It should be mentioned again that the optimization paths leading to **k1** and **k2** are parallel and only one path is possible for each ant group in one iteration. This example shows that even though starting from the same starting system (**h1**), the choice of the structural change also has a huge impact on the final output solutions. Furthermore, it is proved by sufficient executions of the program, that a system with two aspherical lenses and one spherical lens like **k2** is not possible to simultaneously fulfill all the requirements, boundary conditions, and restrictions. In other words, a successful system for this optimization task at least needs to have two spherical lenses to balance the aberrations. The other failure examples are not shown in this dissertation.

In addition to the example shown above, there is another situation for the solution being eliminated due to inappropriate bending. Figure 5.17 gives a whole optimization path starting from the solution **g1** obtained as shown in Figure 5.8, which finally leads to **g1.5**

which fulfills all the imaging requirements. Here only the solutions saved in the GA are shown, and the structural change methods applied to the intermediate solutions are marked. According to the layout of **g1**, the two curvatures of L1 are nearly the same, and after the optimization with the splitting option on L1, the new L1 is flipped over in **g1.2**, but the lens still looks parallel in shape. Since then, the shape of L1 remains similar until the final solution **g1.5** is obtained. According to the transverse aberration plot also in Figure 5.17, the total contribution of L1 is only very small, and deleting it will not cause a large degradation of the performance. If L1 is removed properly, it is found that the system will finally turn into the solution **S3A1(4.1)** illustrated in the next section. Concerning the bending parameter, although the lens is not strongly curved, the close values of the two curvatures result in a very large X value. Therefore, such a system is finally eliminated because of the bending criteria, but the reason behind is different from the case of solution **k2** in Figure 5.16. As mentioned, the automatic removal of a lens according to the aberration contribution is not included in the current program.

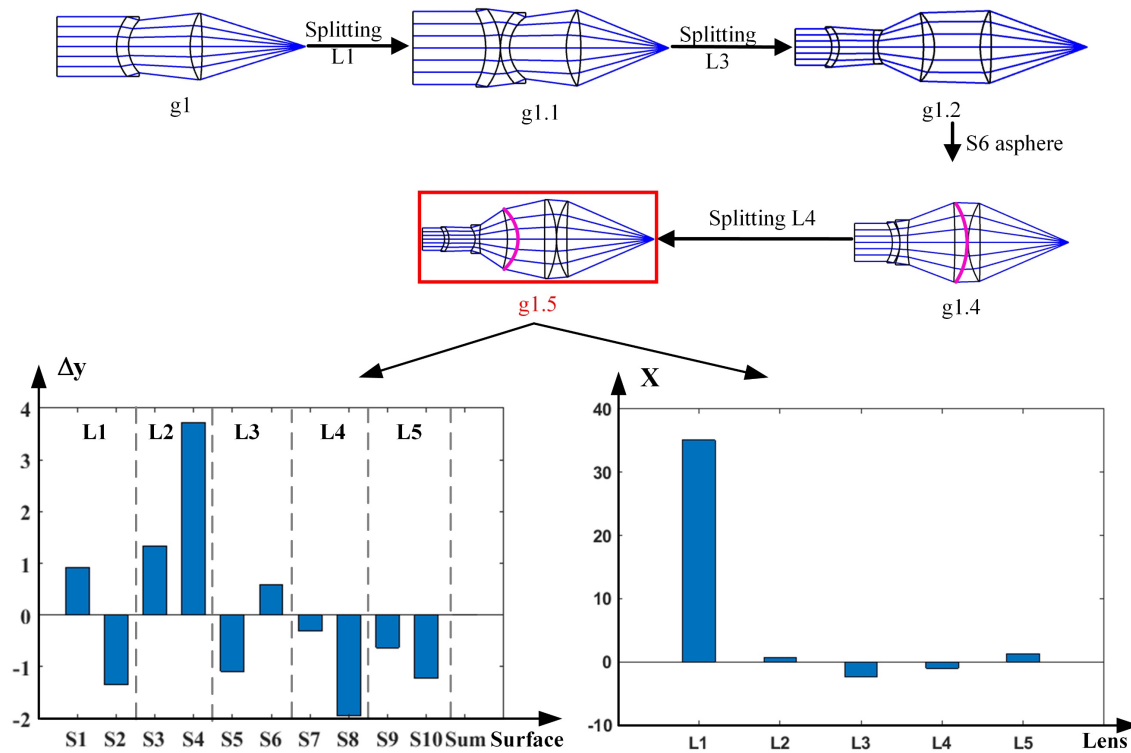


Figure 5.17. Evolution of a solution where L1 has nearly parallel surfaces. The aberration analysis (unit: mm) and the lens bending analysis of solution **g1.5** are plotted below.

5.2.5 Successful solution analysis

In addition to the successful solutions illustrated in the examples of the evolution maps shown above, there are more successful solutions found by the GACOR algorithm.

Concerning the large number of local minima of this optimization task, it is impossible to find all the possible solutions. Thus, the whole program is executed only two times, and 20 solutions are asked to be output for each execution. In addition, a maximum equivalent lens number $\langle N_L \rangle_{max} = 7$ is given as a restriction. Finally, all the solutions are collected and the repetitive solutions coming from the two parallel executions are only listed once in the collection. Furthermore, for a better analysis, the solutions are all categorized and denoted according to their spherical and aspherical lens numbers. As the whole program produces a lot of intermediate solutions, only the final successful results in the output are listed.

The successful solutions with five spherical lenses are listed in Figure 5.18, which includes the name, the layout, and the spot diagram. In addition, the surface-decomposed transverse aberrations of each solution are calculated with the MRT method and plotted in Figure 5.19. Based on the surface contributions, the standard deviation of the aberration distributions, Δy_σ , is also calculated and given in Figure 5.18 in the ‘comments’ column (same for the following figures).

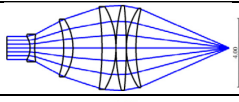
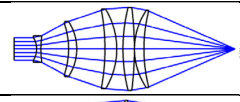
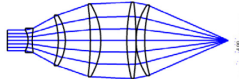
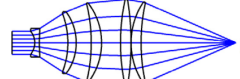
Name	Layout and spot diagram (scale: 4 μ m)	Comments	Name	Layout and spot diagram (scale: 4 μ m)	Comments
S5.1		$\Delta y_\sigma = 0.62$	S5.3		$\Delta y_\sigma = 0.62$
S5.2		$\Delta y_\sigma = 0.73$	S5.4		$\Delta y_\sigma = 0.83$

Figure 5.18. Collection of successful solutions with five spherical lenses.

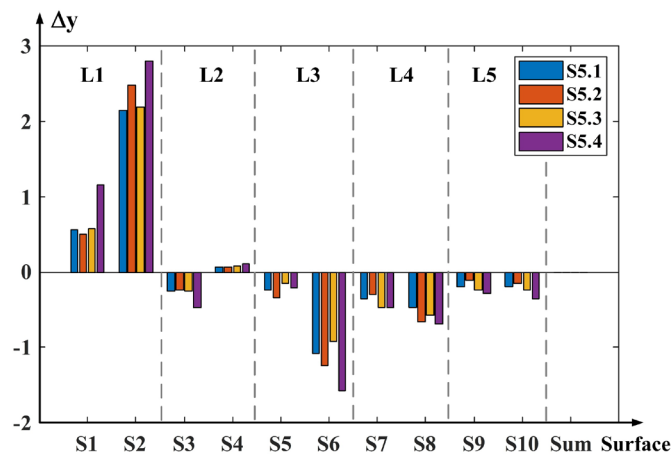


Figure 5.19. Comparison of the surface-decomposed transverse aberration calculated with the MRT method of the ‘S5’ solution category (unit: mm).

According to the spot diagram, all the solutions are similar and just reach the diffraction limit. The difference among the layouts of them mostly lies in the lens positions, while the focal power distributions look similar. Therefore, it can be understood, that such a system

structure is the only possibility for meeting the imaging requirements with only five lenses. From the aberration plot in Figure 5.19, it can be seen that **S5.4** has the highest surface contributions, indicating more sensitive system manufacturability, while the general distributions among the surfaces among them are very similar. As the negative lens surface diverges the ray bundle, S2 contributes the largest full-order transverse aberration among all the surfaces.

Name	Layout and spot diagram (scale: 4μm)	Comments	Name	Layout and spot diagram (scale: 4μm)	Comments
S6.1		$\Delta y_{\sigma} = 0.45$	S6.3		$\Delta y_{\sigma} = 0.45$
S6.2		$\Delta y_{\sigma} = 0.40$	S6.4		$\Delta y_{\sigma} = 0.59$

Figure 5.20. Collection of successful solutions with six spherical lenses.

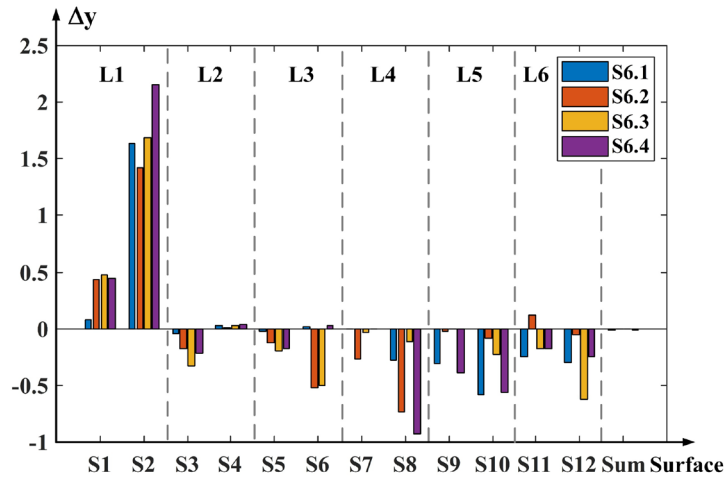


Figure 5.21. Comparison of the surface-decomposed transverse aberration calculated with the MRT method of the ‘S6’ solution category (unit: mm).

In addition to the 5-lens solutions, the algorithm can also find 6-lens solutions with only spherical lenses, as illustrated in Figure 5.20. Although they share quite similar focal power distributions, the imaging spots show a larger difference in size. According to the spot size, **S6.2** has the best imaging performance among all, while the imaging performance of **S6.1** and **S6.4** is not much improved compared to the 5-lens solutions. Concerning the aberration distribution shown in Figure 5.21, the negative lens L1 still contributes the largest positive aberration value, while the positive lenses have very different aberration contributions in each solution. According to the analysis of the aberrations, this group of solutions has smaller Δy_{σ} values in general compared to the 5-lens solutions, which indicates less critical sensitivity due to the additional lens. **S6.4** is the most sensitive because of the largest aberration deviation.

For the solutions with seven spherical lenses, the spherical aberration can be even better corrected. Figure 5.22 lists all the 7-lens solutions found by the algorithm. It can be seen from the layouts, that these solutions show a larger diversity concerning focal power distribution and the spot becomes even smaller in general.

The MRT calculation results of the aberration distribution are shown in Figure 5.23. Specifically concerning the aberration contributions, as **S7.1** and **S7.4** both contain two negative lenses, the large positive transverse aberration are shared by two lenses. Consequently, the standard deviations of the aberrations of **S7.1** and **S7.4** are smaller.

Name	Layout and spot diagram (scale: 2 μ m)	Comments	Name	Layout and spot diagram (scale: 2 μ m)	Comments
S7.1		$\Delta y_\sigma = 0.39$	S7.5		$\Delta y_\sigma = 0.58$
S7.2		$\Delta y_\sigma = 0.27$	S7.6		$\Delta y_\sigma = 0.70$
S7.3		$\Delta y_\sigma = 0.29$	S7.7		$\Delta y_\sigma = 0.79$
S7.4		$\Delta y_\sigma = 0.28$			

Figure 5.22. Collection of successful solutions with seven spherical lenses.

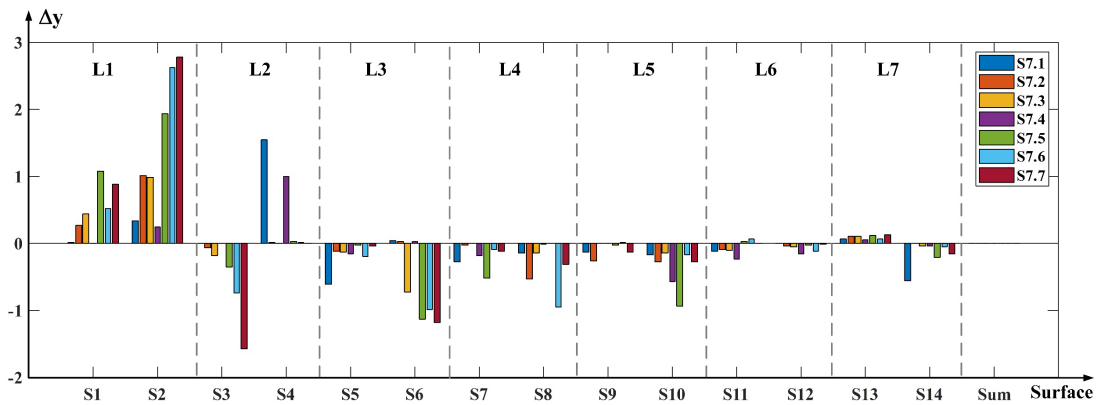


Figure 5.23. Comparison of the surface-decomposed transverse aberration calculated with the MRT method of the ‘S6’ solution category (unit: mm).

The algorithm finds that the solutions with aspherical lenses should contain at least three spherical lenses (3S) and one aspherical lens (1A), namely $\langle N_L \rangle_0 = 6$, and such solutions are denoted as the category ‘S3A1’. Different from the notation of the spherical solutions, the solutions with aspherical lenses are distinguished with the asphere surface location as the first value in the parentheses. The corresponding solutions are listed in Figure 5.24. In addition to the Δy_σ values, the coefficient of the corresponding highest asphere term is also given in the comments column.

According to the layouts of the solutions, regardless of the asphere location, almost all the successful solutions have a similar system structure, where the first negative lens greatly diverges the ray bundle, and the three positive lenses converge it. In addition, the air gaps between the lenses are comparable. The only exception is **S3A1(3.2)**, which is longer than the other solutions along the z-axis, and the positive lenses are closer to each other. Due to the smoother ray path, the Δy_σ values of this solution are much smaller than the others. In the output results, the successful solutions with aspherical surface locations at S2, S5, and S8 were not found. The reason behind this can be the probabilistic feature of the algorithm, or these asphere locations do not help succeed in fulfilling all the specifications.

Name	Layout and spot diagram (scale: 2 μ m)	Comments	Name	Layout and spot diagram (scale: 2 μ m)	Comments
S3A1 (6.1)		A1 $\Delta y_\sigma = 0.78$	S3A1 (7.1)		A1 $\Delta y_\sigma = 1.67$
S3A1 (6.2)		A1 $\Delta y_\sigma = 1.19$	S3A1 (4.1)		A3 $\Delta y_\sigma = 1.34$
S3A1 (6.3)		A1 $\Delta y_\sigma = 0.88$	S3A1 (3.1)		A3 $\Delta y_\sigma = 1.23$
S3A1 (6.4)		κ $\Delta y_\sigma = 1.16$	S3A1 (3.2)		A1 $\Delta y_\sigma = 0.34$
S3A1 (1.1)		A3 $\Delta y_\sigma = 1.95$			

Figure 5.24. Collection of successful solutions with three spherical lenses and one aspherical lens.

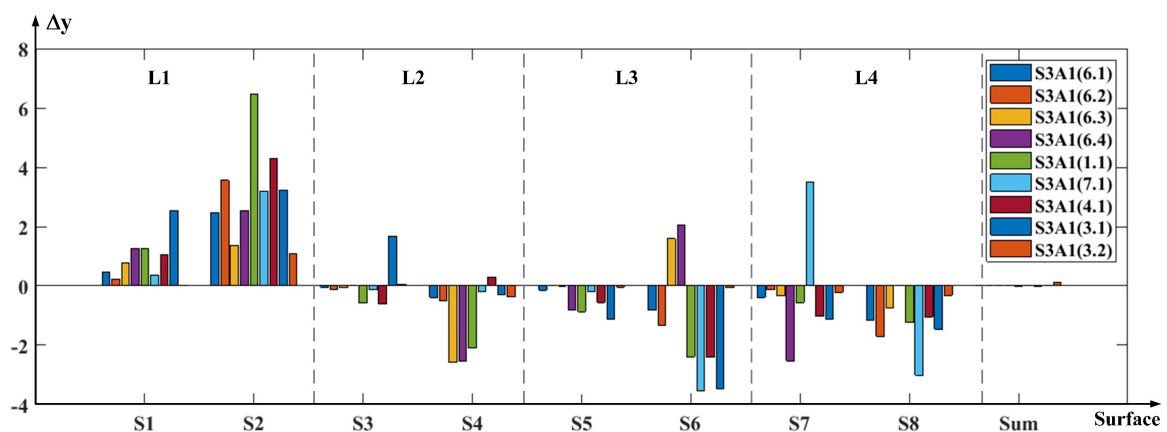


Figure 5.25. Comparison of the surface-decomposed transverse aberration calculated with the MRT method of the ‘S3A1’ solution category (unit: mm).

Correspondingly, Figure 5.25 compares the aberration distributions of all the solutions. In some cases, the aspherical surface contributes a positive value together with the negative

lens to balance the aberration, such as **S3A1(3.1)**, **S3A1(6.3)**, and **S3A1(6.4)**, while in the other cases, the asphere simply reduces the negative aberration values for the same purpose. Due to the application of the aspheres, the aberration distributions among all the solutions show a greater variety, compared to the pure spherical lens systems. As the aspherical surfaces introduce higher-order aberrations into the system, their contributions to the spherical aberration correction can differ a lot in these successful solutions. Thus, it is hard to summarize the overall aberration balance of the whole solution category. However, **S3A1(3.2)** indicates a great potential of aspherical surfaces for finding highly tolerant solutions, with the great global searching ability of the algorithm.

Name	Layout and spot diagram (scale: 2 μ m)	Comments	Name	Layout and spot diagram (scale: 2 μ m)	Comments
S4A1 (8.1)		κ $\Delta y_{\sigma} = 0.38$	S4A1 (8.2)		A1 $\Delta y_{\sigma} = 0.39$
S4A1 (8.3)		κ $\Delta y_{\sigma} = 0.58$	S4A1 (8.4)		κ $\Delta y_{\sigma} = 0.70$
S4A1 (8.5)		A1 $\Delta y_{\sigma} = 0.41$	S4A1 (8.6)		κ $\Delta y_{\sigma} = 1.12$
S4A1 (8.7)		A1 $\Delta y_{\sigma} = 0.66$	S4A1 (8.8)		κ $\Delta y_{\sigma} = 0.66$
S4A1 (1.1)		A1 $\Delta y_{\sigma} = 1.46$	S4A1 (2.1)		A1 $\Delta y_{\sigma} = 0.74$
S4A1 (3.1)		κ $\Delta y_{\sigma} = 0.52$	S4A1 (3.2)		A1 $\Delta y_{\sigma} = 2.016$
S4A1 (3.3)		A1 $\Delta y_{\sigma} = 3.06$	S4A1 (4.1)		A3 $\Delta y_{\sigma} = 0.99$
S4A1 (4.2)		Conic $\Delta y_{\sigma} = 0.81$	S4A1 (4.3)		Conic $\Delta y_{\sigma} = 1.49$
S4A1 (4.4)		Conic $\Delta y_{\sigma} = 1.09$	S4A1 (6.1)		Conic $\Delta y_{\sigma} = 0.43$
S4A1 (6.2)		Conic $\Delta y_{\sigma} = 0.42$	S4A1 (6.3)		Conic $\Delta y_{\sigma} = 0.73$
S4A1 (7.1)		Conic $\Delta y_{\sigma} = 0.85$	S4A1 (10.1)		Conic $\Delta y_{\sigma} = 0.36$
S4A1 (10.2)		Conic $\Delta y_{\sigma} = 1.30$	S4A1 (10.3)		Conic $\Delta y_{\sigma} = 0.46$

Figure 5.26. Collection of successful solutions with four spherical lenses and one aspherical lens.

Finally, the spherical and aspherical lens combination for the solution can be either

‘S1A2’ or ‘S4A1’, namely $\langle N_L \rangle = 7$. As illustrated in Figure 5.16, the combination of ‘S1A2’ cannot meet all the requirements with appropriate lens shapes. Therefore, only the successful solutions in the ‘S4A1’ category are illustrated in Figure 5.26. More spherical lense in this solution category relaxe the optimization, so the program found many various solutions. Among all the asphere locations, the solutions with aspherical S8 are the most often found, up to eight times, followed by S4, S3, S6, and S10. All the solutions in this category show the greatest variety of aberration distribution conditions compared to the former ones. Furthermore, due to the additional spherical lens in the ‘S4A1’ category compared to the ‘S3A1’ category, most of the solutions only need an asphere with only the conic constant term to fulfill all the requirements.

Considering the large number of solutions, the aberration distributions are divided into three parts plotted in Figure 5.27, Figure 5.28, and Figure 5.29. The first part only plots the aberrations of the solutions with the asphere located at S8, so that the different performances only caused by various system structures can be compared. **S4A1(8.6)** has the largest deviation, indicating the highest sensitivity among all, while the aberrations of solution **S4A1(8.1)** are balanced the best.

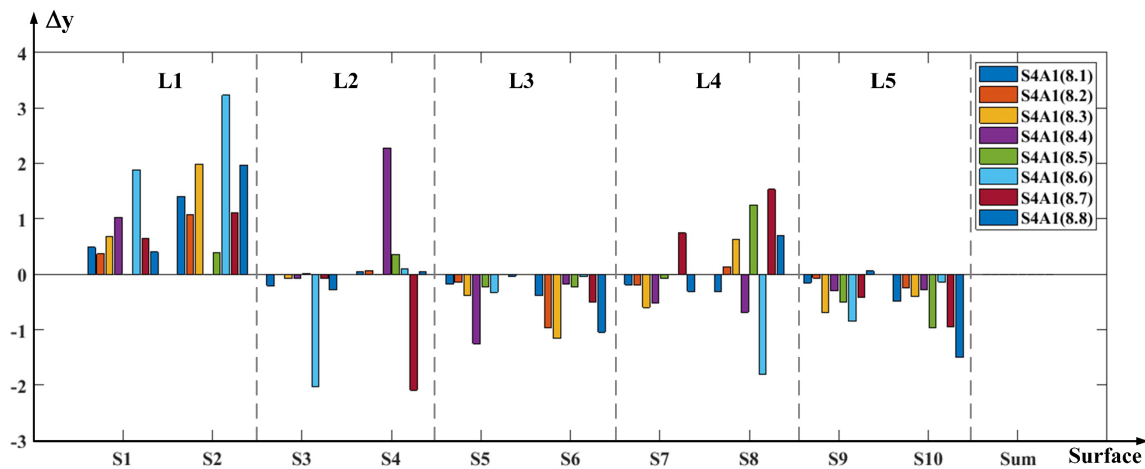


Figure 5.27. Comparison of the surface-decomposed transverse aberration calculated with the MRT method of the ‘S4A1’ solution category (part 1) (unit: mm)

As for the solutions with asphere surfaces located in the front part of the system from S1 to S4, the difference in the aberration distribution is even larger. Especially concerning the three solutions with S3 being aspherical, the aberration balance of the solution **S4A1(3.1)** is much more comfortable than **S4A1(3.3)** due to the different system structures. The results prove again that the aspherical surface behavior in the system is complicated, and has a great impact on the system tolerance.

In the last part, the aberrations of the solutions with asphere locations at S6, S7, and S10 are compared in Figure 5.29. Among all, the solution **S4A1(10.2)** shows the largest deviation, meaning the more critical sensitivity for manufacture. In comparison, solution **S4A1(10.3)** is considered to have a better aberration correction concerning the smaller aberration deviation among the surfaces, although the asphere is also located at S10.

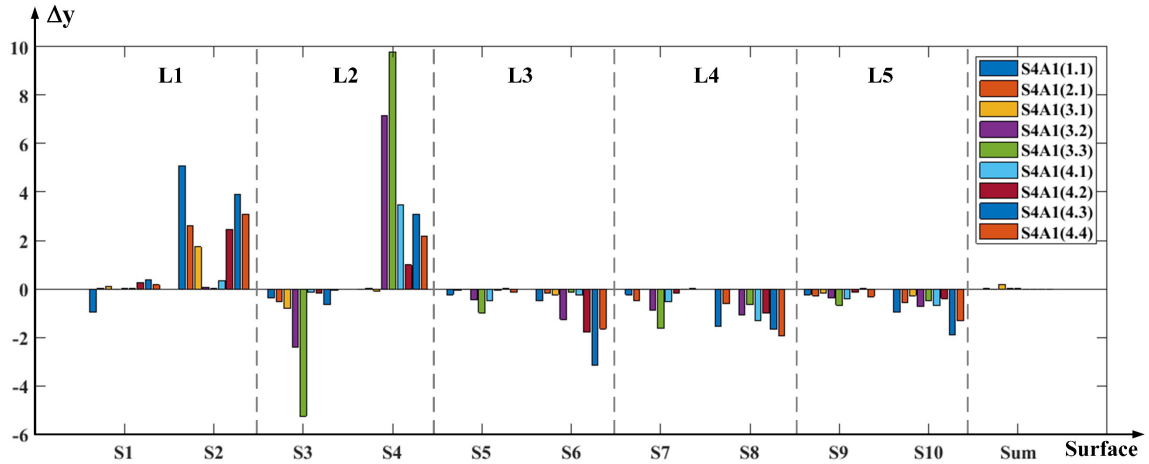


Figure 5.28. Comparison of the surface-decomposed transverse aberration calculated with the MRT of the 'S4A1' solution category (part 2) (unit: mm).

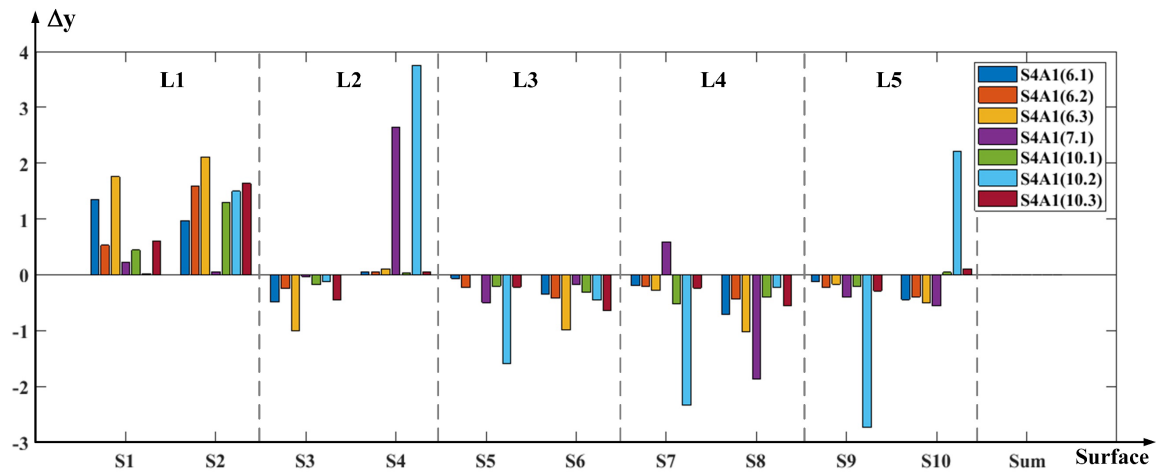


Figure 5.29. Comparison of the surface-decomposed transverse aberration calculated with the MRT of the 'S4A1' solution category (part 3) (unit: mm).

To obtain an overview of the aberration correction and imaging performance among all the solutions in various categories, the aberration deviation Δy_σ values against the equivalent lens number are plotted in Figure 5.30. In this plot, each solution is represented by a 'bubble', and all the solutions from the same category share the same color. Concerning the overall performance evaluation, the MF values of all the solutions are not fully comparable, because different lens numbers of the systems lead to a different amount of boundary constraints in the MF, influencing the OPC calculation. Therefore, as a rough

indicator of the imaging performance, the bubble size is determined by the relative spot size of the system, compared to all the other output solutions.

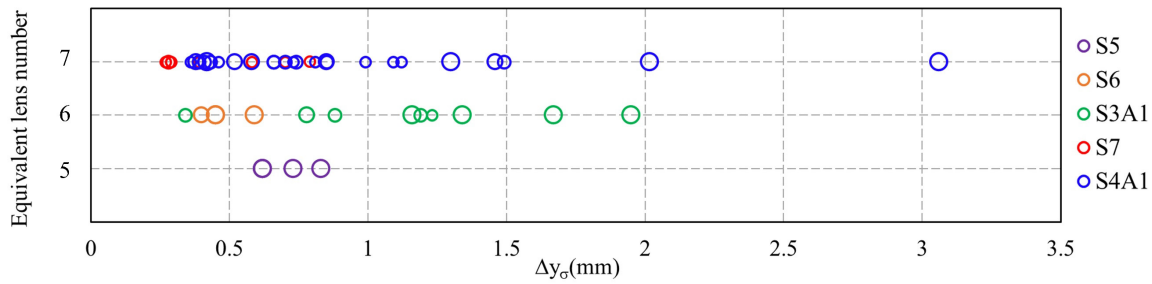


Figure 5.30. Aberration standard deviation of all the solutions (unit: mm). Each solution category is marked with a different color, and the bubble size represents the relative spot size of the solutions.

Concerning the solutions composed of only spherical lenses, as the equivalent lens number increases, the Δy_σ value can reach a lower level in general. In accordance with the physical knowledge, all the solutions in the ‘S6’ category have a smaller aberration deviation, compared to those in the ‘S5’ category. However, although the aberration correction of the best solution found in the category ‘S7’ is further improved, there are still solutions in the same category suffering from a larger aberration deviation. The larger variety of the ‘S7’ solution category can be understood by the very different focal power distributions of the solutions. In addition, the plot clearly shows that the spot size becomes smaller from ‘S5’ to ‘S7’, due to the larger equivalent lens number in the system. Consequently, Figure 5.30 clearly indicates that for the systems with only spherical lenses, the aberration correction and manufacturability can be predicted according to the lens number to some extent. However, the nominal imaging performance and the system volume should reach a compromise.

As mentioned above, the aspherical surfaces can strongly influence Δy_σ due to the higher-order aberration. Considering the two solution categories with different aspherical lens numbers, namely ‘S3A1’ and ‘S4A1’, the aberration balance performances are hard to compare. On one hand, both categories include solutions with superior aberration correction compared to the ‘S5’, ‘S6’, ‘S7’ categories. On the other hand, the large difference in the system performance among the solutions also indicates the uncertainty of the optimization path and the difficulty in controlling the higher-order aberrations, especially concerning the ‘S4A1’ category. In addition, the spot size is also hard to predict in a system with aspherical surfaces, according to the irregular bubble sizes of these solution categories in the plot. Consequently, due to the much more complicated full-order

aberration distribution, the system imaging performance and sensitivity cannot be simply predicted according to the equivalent lens number.

5.2.6 Discussion

With all the discussions and illustrations above concerning the optimization results, the GACOR algorithm is considered strong in global searching ability. It is capable of guiding the optimization path and assessing the imaging performance of the output solutions with a high degree of automation for the proposed retro-focus system design task. Given the general optimization strategy based on the physical knowledge and the corresponding algorithm parameters, the algorithm succeeds in the design task from the very beginning until a large output solution database with great diversity.

Due to the probabilistic feature of the algorithm, there may be always new solutions being found every time the whole program is executed. Particularly, the different asphere locations bring much more degrees of freedom in the optimization process. However, concerning the essential purpose of optical design, it is only of limited sense to find out all the possible solutions, especially for a complicated high-dimensional optical design task. Instead of a considerably large number of successful solutions, gaining an overview of the solutions, understanding the tendency of the solution evolution, and having enough choices for a final design regarding the practical considerations are the most essential intentions of the investigation.

The execution time of a full running of the program consists of three parts according to the program platforms. First, the main workflow, the analysis of the system, and the execution of the ACOR local search process are all done by Matlab where the program is written. Among all, the ACOR local search costs most of the execution time. Although the exact execution time is not predictable due to the probabilistic feature of the algorithm, the general execution time is strongly dependent on some of the algorithm parameters, such as the output solution number and the iteration number. Second, the DLS local optimization is only operated by Zemax, which cannot be improved or modified. Third, as the connection tool between Zemax and Matlab, the ZOS-API toolbox is always needed during the whole program, and it is found that the DLS optimization called by ZOS-API tool takes much longer than directly operating in Zemax.

Consequently, concerning only the most time-consuming process of the whole program, only the ACOR local exploration time and the DLS booster running time are calculated

accurately as an overview. In general, given the retro-focus system design task and 20 output solutions in 10 main iterations, the ACOR local exploration time ranges from 10-70min, and the DLS booster takes 4-6 hours.

Finally, the algorithm can be also applied for other optimization tasks, if the optimization strategy is adapted according to the different system specifications. Besides the optimization of the retro-focus system, a tele-system and a high NA collimator are also chosen to test the optimization performance of the GACOR algorithm. The additional experiments prove that the GACOR algorithm can output also satisfactory results by simply adjusting the optimization strategy. And considering the lower complexity of the tasks, the execution time is much shorter (0.5-2 hours) for the whole process with enough output solutions, the details of which are given in Appendix J.

5.3 Freeform system optimization

With the simple optimization tasks, the GACOR algorithm has been proved feasible for the quasi-automatic optimization with structural changes, if the ants are ‘trained’ with proper physical knowledge about optical design. The results confirmed the great potential of the global searching ability. Concerning the second goal of the research, the algorithm should also be tested for feasibility when high-dimensional optimization problems are given.

5.3.1 Final improvement of an anamorphic system

Concerning the GACOR algorithm developed in the current stage, a complicated system with straight OAR but broken rotational symmetry is ideal as a test example. Anamorphic systems are applied in many fields such as camera objects and 3D scanners [45]. However, the broken symmetry due to the differently stretched NA in both tangential and sagittal planes brings troubles in correcting the aberrations with only spherical lenses. Therefore, an anamorphic system is chosen as an example to test the final improvement ability of the GACOR algorithm. The common axis system structure with complicated freeform surfaces makes the anamorphic system appropriate for the purpose.

The layout of the starting anamorphic system for the final improvement is shown in Figure 5.31. The system contains five lenses, marked as L1 to L5. L1, L4, and L5 compose a retro-focus system comparable to the example introduced in Section 5.2. L2 and L3 are two cylindrical lenses placed between L1 and L4, working only in the x-z cross section. The two cylindrical surfaces are marked in red. Due to technical reasons, the other surface

of both cylindrical lenses always remains plane. In comparison, the aspherical surface location is marked in orange, which will be discussed in the following paragraphs. Furthermore, as discussed before, the optimization of the glass materials is not considered in this work, and all the lenses are simply made of BK7 in this example system. In addition, the selected object fields of view of maximum 2° are also given in Figure 5.31.

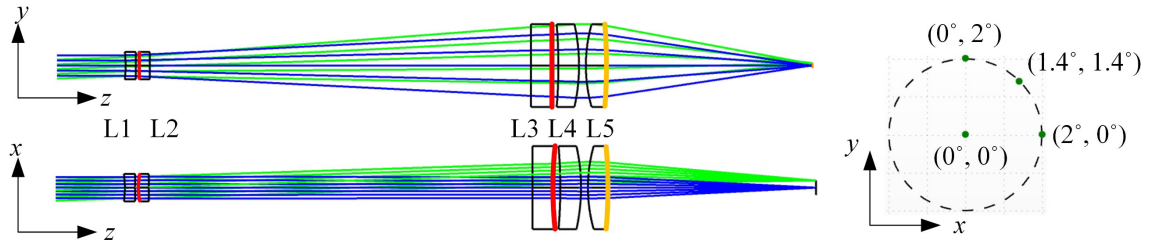


Figure 5.31. Layout of the anamorphic system in both cross-sections and the sampled FoV. The cylindrical surface locations are marked in red, and the aspherical surface locations are marked in orange.

The specifications of the system are listed in Table 5.2. The focal length and the retro-focus factor R_F determine the image distance together. The image space NA in the tangential plane, denoted by NA_y , can be also calculated with the focal length and the incoming beam diameter. Two cylindrical lenses are applied to reach an anamorphic stretching factor of 3, resulting in a smaller NA_x in the sagittal plane. The goal of the final improvement is to obtain diffraction limited spots all over the objective field.

Table 5.2 Specification of the anamorphic system.

Required specifications	Target values	Required specifications	Target values
Focal length in y	20mm	Free working distance	60mm
Focal length in x	60 mm	Stop position	L1 rear surface
NA_x	0.05	Total length	200mm
NA_y	0.15	Retro focus factor	3
Incoming beam diameter	6mm	Field of view	2°

The given starting system currently has only spherical surfaces. Therefore, it can be predicted that the aberration correction is poor. Figure 5.32 shows the spot diagrams of all the selected FoV. The large scale indicates a large improvement space for the system.

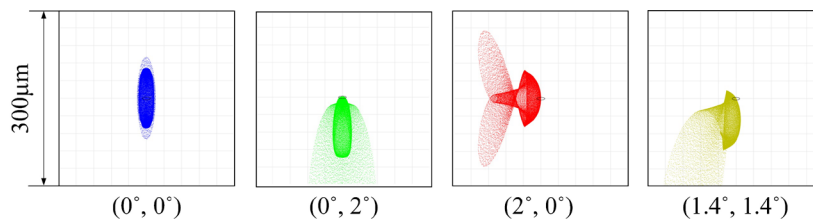


Figure 5.32. Spot diagrams of all the selected fields.

Approximately, we can assume a decoupling of the sagittal and tangential plane for this system. As a result, the aberration correction is difficult if only spherical lenses are used. Therefore, the application of aspherical surfaces is considered. Concerning the choice of the position and type of the non-spherical surface, the necessary criteria are not trivial. Compared to rotationally symmetric aspherical surfaces, the freeform surface provides better correction ability, but is also challenging for manufacturing and increases the cost. Instead, if more rotationally symmetric lenses are used for the same correction purpose, the general cost and sensitivity problems due to the higher complexity of the system structure are also hard to control. Therefore, according to the specific system structure and the manufacturing considerations, there is no simple indicator for the best decision of the asphere. As the purpose of this research is to illustrate the potential ability of the algorithm concerning a large number of degrees of freedom, the decision of the position and surface type of the aspherical surface is not the focus.

Concerning the final improvement of the example anamorphic system, L2 and L3 are allowed for a conic constant deviation from the cylindrical surface sag, while only one of the other three lenses can be non-spherical for this optimization task. With many tryouts, it is found that one rotationally symmetric asphere in the type of ‘even asphere’ and ‘Q-type aspheres’ is not able to reach the diffraction limited goal without changing the current structure. Therefore, a freeform surface with the type of ‘Zernike fringe sag’ is applied in the system for the final improvement.

Following the empirical rules for the freeform position as mentioned in Section 2.6, the freeform surface location of the anamorphic system is fixed at the rear surface of L5. For the correction all over the object field, it is appropriate as the ray bundles are best separated at this surface, so that the correction of the spots all over the field can reach the best. The location is marked in pink in Figure 5.31.

Table 5.3. Categorization of lens parameters.

	Curvature (front surface)	Curvature (rear surface)	Lens thickness	Air gap	Conic constant	Zernike term
L1	r_{1f}	r_{1r}	t_{1g}	t_{1a}	0	0
L2	r_{2x}	0	t_{2g}	t_{2a}	κ_{2x}	0
L3	0	r_{3x}	t_{3g}	t_{3a}	κ_{3x}	0
L4	r_{4f}	r_{4r}	t_{4g}	t_{4a}	0	0
L5	r_{5f}	r_{5r}	t_{5g}	t_{5a}	κ_5	Z1 – Z36

As explained in Section 4.2, the various optimization paths of the final improvement

process originate from the variable settings, dependent on the lens parameter categories. Following the categorization rules, Table 5.3 summarizes the lens parameter categories of the example anamorphic system. The corresponding lens parameters in the same category are marked by a unique colored background with black font color, while the categories of an individual component are marked in blue background with blue font color. As the system structure clearly indicates two main groups of the lenses divided between L2 and L3, the air gap t_{2a} is considered individually. For the same reason, both the curvatures and thicknesses are divided into two main groups correspondingly. Thus, there are in total 13 categories of parameters in this system.

During the final improvement process, assuming all the lens parameters are allowed for being variables, except for t_{5a} fixed at 60mm, all the 12 parameter categories are included in the variables step by step. As more variables are added, the optimization is carried out until the system is diffraction limited. Especially for the Zernike terms, only plane-symmetric Zernike terms are added step by step during the optimization, following the instruction given in Figure D.2 in Appendix D.

5.3.2 Successful solution analysis

Concerning the output solution number of the optimization task, the desired number of $K_{max} = 10$ is given. The final improvement program is executed for one time and all the solutions are listed in Figure 5.33, where the layout in both cross-sections, the spot diagram of all the fields, and the freeform surface sag (basic sphere shape removed) are given. As for the layout, only the rays from the on-axis field are plotted to better announce the lens shape. In addition, as the data for analysis in the last column, the information about the Zernike polynomial order, as well as the standard deviation values Δy_σ is given. The standard deviation is again calculated with the MRT method concerning the full-order transverse aberrations of one ray. Due to the off-axis FoV and the asymmetric structure of the anamorphic system, a skew ray with the sampling coordinate (0.7, 0.7, 0, 1) is chosen here for the aberration calculation as a representative of the general performance of the whole system. It has been checked, that the results are similar if any other ray is chosen due to the similar system structure of the solutions, but only on different scales. As the solutions have almost the same system structure, the MF values regarding the overall performance are comparable, indicating the small difference in the aberration correction. Therefore, the MF calculation values are also given in the comments column, denoted as V_m .



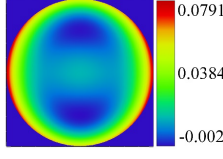
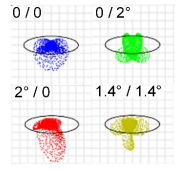


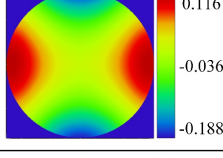
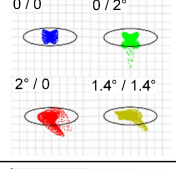


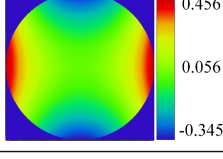
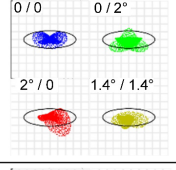


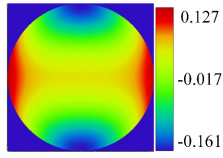
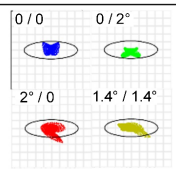


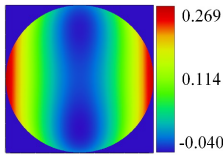
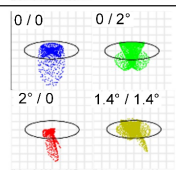


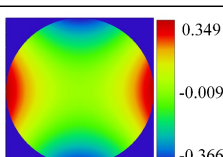
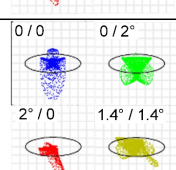


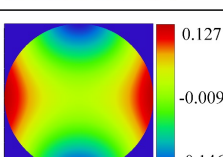
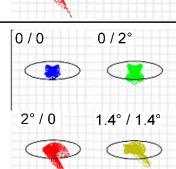


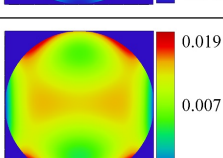
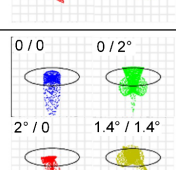


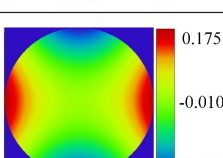
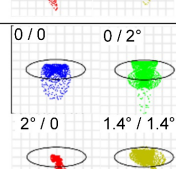


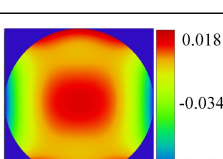
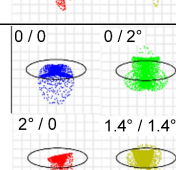
	Layout	Zernike surface sag	Spot diagram	Comments
1	y-z layout:  x-z layout: 			$Z_{max} = 16$ $\Delta y_{\sigma} = 0.22mm$ $\Delta x_{\sigma} = 0.06mm$ $V_m = 6.7 \times 10^{-5}$
2	y-z layout:  x-z layout: 			$Z_{max} = 36$ $\Delta y_{\sigma} = 0.05mm$ $\Delta x_{\sigma} = 0.01mm$ $V_m = 4.3 \times 10^{-5}$
3	y-z layout:  x-z layout: 			$Z_{max} = 16$ $\Delta y_{\sigma} = 0.25mm$ $\Delta x_{\sigma} = 0.05mm$ $V_m = 5.6 \times 10^{-5}$
4	y-z layout:  x-z layout: 			$Z_{max} = 36$ $\Delta y_{\sigma} = 0.51mm$ $\Delta x_{\sigma} = 0.12mm$ $V_m = 3.6 \times 10^{-5}$
5	y-z layout:  x-z layout: 			$Z_{max} = 36$ $\Delta y_{\sigma} = 0.31mm$ $\Delta x_{\sigma} = 0.09mm$ $V_m = 5.8 \times 10^{-5}$
6	y-z layout:  x-z layout: 			$Z_{max} = 36$ $\Delta y_{\sigma} = 0.35mm$ $\Delta x_{\sigma} = 0.09mm$ $V_m = 6.0 \times 10^{-5}$
7	y-z layout:  x-z layout: 			$Z_{max} = 36$ $\Delta y_{\sigma} = 0.09mm$ $\Delta x_{\sigma} = 0.02mm$ $V_m = 4.5 \times 10^{-5}$
8	y-z layout:  x-z layout: 			$Z_{max} = 36$ $\Delta y_{\sigma} = 0.25mm$ $\Delta x_{\sigma} = 0.07mm$ $V_m = 5.0 \times 10^{-5}$
9	y-z layout:  x-z layout: 			$Z_{max} = 36$ $\Delta y_{\sigma} = 0.33mm$ $\Delta x_{\sigma} = 0.11mm$ $V_m = 3.9 \times 10^{-5}$
10	y-z layout:  x-z layout: 			$Z_{max} = 25$ $\Delta y_{\sigma} = 0.50mm$ $\Delta x_{\sigma} = 0.13mm$ $V_m = 4.3 \times 10^{-5}$

Figure 5.33. Collection of the output solutions of the anamorphic system. The surface sag is in the unit of mm, and the spot diagrams are scaled in 300 μ m.

Due to the different image space NA values in the two cross-sections, the Airy pattern looks elliptical. Therefore, for a clarification of the diffraction limited criteria for this optimization task, the generalized definition of the Airy diameter according to Zemax is applied, which is approximated with the squared NA regarding four MRs, written as [55]

$$D_{Airy} = 2 \times \frac{1.22\lambda}{\sqrt{n^2 (\sin^2 \theta_t + \sin^2 \theta_b + \sin^2 \theta_l + \sin^2 \theta_r)}}, \quad (6.1)$$

where θ_t , θ_b , θ_l , and θ_r are the real MR angle at the top, bottom, left, and right side of the pupil edge. Thus, the RMS spot sizes of all the sampled fields can be compared to the nominal D_{Airy} value, and the system is considered diffraction limited only when all the field spot sizes are smaller than D_{Airy} .

Compared to the nominal design shown in Figure 5.31, for most of the final improvement results, the lens bending does not show a significant change. The only exception is solution 2, where L4 is flipped after fine-tuning. This result proves that the global searching ability of the algorithm is good enough to find some potentially better lens bending structures even during the final improvement phase of optical design, but such creative solutions are always probability-based and therefore cannot be predicted. However, with the help of the ACOR local search, the chance can be greatly enhanced.

Despite the similar layouts of the different solutions, the freeform surface sag can be completely different. Considering the technical issues of freeform manufacturing, solution 1, 8, and 10 have the smallest deviation from the basic sphere, while solutions 3 and 6 might be critical due to the large deviation. As the acceptable manufacturability may vary in different situations, in this work, the large collection of different solutions are all presented here to illustrate their diversity. The large database provides the user with a large choice according to the specific purpose.

To better compare the aberration distributions among the surfaces, the standard deviations of the surface-decomposed transverse aberration projected in both sagittal and tangential planes, Δx_σ and Δy_σ , are plotted in Figure 5.34. The values indicate the performance in aberration correction and sensitivity. As the only solution with reversed L4 found by the program, solution 2 has the smallest Δx_σ and Δy_σ , implying the best tolerance for manufacture. Although solution 7 keeps the same structure as the nominal system, it still reaches a very well-balanced aberration correction compared to the other. Therefore, it can be concluded that the bending orientation of L4 is not the essential reason for the

final performance, while the considerable local minima make it harder for the optical designer to find the optimal solution.

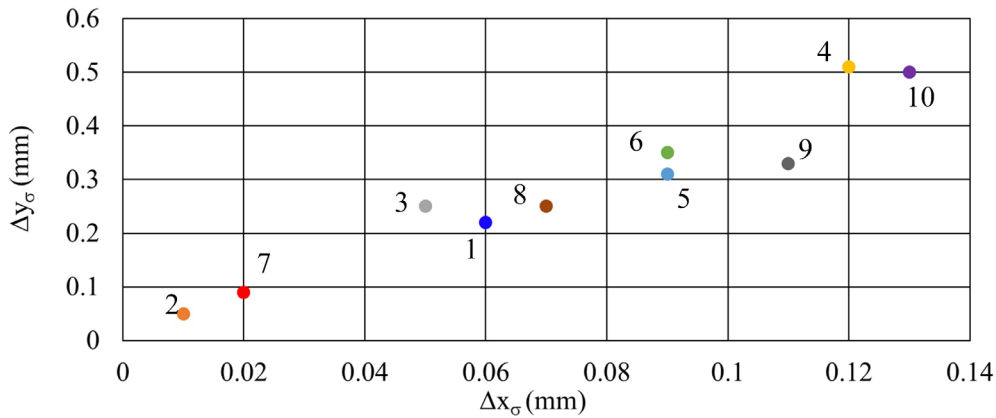


Figure 5.34. Aberration standard deviation of the transverse aberration surface contribution in x- and y-direction of all the final improvement solutions.

With this example, the benefit of the MRT aberration calculation tool can be illustrated again. Among all the solutions shown in Figure 5.33, solution 2 and 10 respectively show the best and worst aberration balance. Thus, they are selected for analysis with the MRT method concerning the additive Zernike coefficient representation, as shown in Figure 5.35.

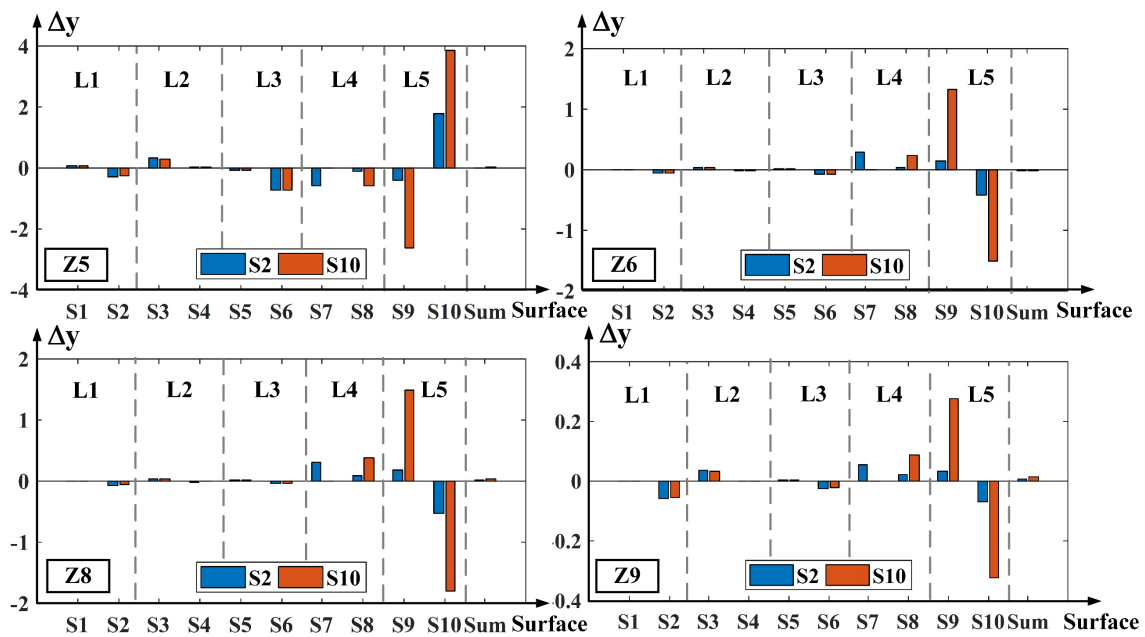


Figure 5.35. Comparison between solution 2 and 10 concerning the surface-additive Zernike coefficients in the case of Z5, Z6, Z8, and Z9 (unit: waves). ‘S2’ and ‘S10’ refer to solution 2 and 10.

Among all the Zernike terms, astigmatism in both cross-sections (Z5 and Z6), tangential coma (Z8), and spherical aberration (Z9) are compared respectively between these two

solutions for comparison. For all the coefficients, the surface contributions of the front lens group show no differences between the two solutions. But due to the different lens bendings and freeform surface shapes, the contributions of the rear group of solution 10, especially the last freeform surface, are much higher than solution 2. The results are consistent with the calculation results of Δx_σ and Δy_σ , and more explicitly proves that the sensitivity of the solution 10 is far more critical concerning the manufacturability.

5.3.3 Discussion

Concerning the large degree of freedom involved in the final improvement phase of optimization, the local exploration based on the ACOR algorithm is still capable of locating the local minimum area so that the variety of the output solutions is enhanced. Different from the nominal optimization phase with structural changes, there is almost no chance to obtain the same solution, given a limited number of the output solutions. The reason behind this is the high dimension of the MF topology and the large variety of the final improvement strategy, especially when the system contains freeform surfaces. Therefore, it is not necessary to check the similarity in this program. Thus, the large variety of the results offers a big database of the possible solutions, from which the user can set their filter considering the specific criteria to finally decide on the more appropriate system as the final design. Compared to the manual optimization influenced by the individual preference of the optical designer, the automatic program includes various strategies, which bring a higher probability to find ‘new’ solutions that are hard to be found with the optical designer’s habitual methods.

Furthermore, due to the lower correction complexity compared to the nominal system design, the final improvement optimization program is found much faster than the nominal design program. With sufficient executions of the program, it is found that the running time ranges from 30 to 90min, given 10 output solutions. Because of the connection between Matlab and Zemax, the execution time is still mostly occupied by the DLS local optimization.

6 Conclusion and outlook

The main goal of this work is to seek new methods of modern optical system design with a special focus on systems without symmetry, with the desire that the design process could be directed to greater productivity. The research discussed in this dissertation makes certain contributions regarding a more powerful aberration analysis tool and the improved global optimization methods at a higher automation level.

First, a novel mixed ray-tracing (MRT) method for comprehensively analyzing the off-axis system surface contribution of transverse aberration is introduced, applicable for the calculation of surface-decomposed total, intrinsic, and induced aberrations. Thanks to the additivity among the surface contributions, the surface-additive Zernike coefficient representation is proposed for the first time as an extension of the method. The large flexibility concerning the choice of the reference ray supports the aberration evaluation with various purposes for symmetry-free systems. The accuracy and reliability of the results are proved to the best extent despite the lack of perfectly comparable references. The application of an example lithographic system proves that the MRT method is of great help to evaluate the critical surfaces with large aberration contributions, and the higher-order aberration balancing effects among the surfaces can be better understood. The quantitative calculation results also qualitatively indicate the sensitivity of the surfaces during manufacture from the viewpoint of the high-order aberrations. The modified Kingslake plots showing the contributions of all the surfaces can clearly visualize the aberrations of arbitrary systems, and therefore help for better analysis for practical purposes. Besides, the surface-additive Zernike coefficient fitting for separated intrinsic and induced aberration also breaks the limits in the high and low order aberration assessment, so that the specific aberration orders can be analyzed separately. Furthermore, if the finite object field instead of the pupil is sampled, the MRT method is also capable of analyzing distortion. In addition, the direction angle components of the ray vectors can be directly used for analyzing the angle aberrations specifically for generalized afocal systems.

Besides, a bio-inspired algorithm referring to the ant colony behavior is further developed for an improved global optimization method in lens design, and the MRT method is applied for aberration analysis in the algorithm. It breaks the limitations in making structural changes compared to the conventional optimization algorithm, so that the strong impact and restriction of the initial design is eliminated, and a higher level of automation

of the optimization process can be achieved. The large database of the output systems provides the user with an overview of the solution landscape. The large variety of the solutions without ranking also support the user to fix the best-fitting solution according to the specific practical purposes. Although the sensitivity is not quantitatively formulated during the optimization, the output solutions are all ensured with acceptable manufacturability to avoid the critical cases. With the simple optimization task of the retro-focus system, the optimization algorithm is proved feasible for operating necessary structural changes, given proper physical guidance. Furthermore, the optimization results of the anamorphic system with freeform surfaces also verify the reliability of the algorithm concerning the handling of large degree of freedom. Concerning the far goal of a fully automatic global optimization algorithm, the research in this work only accomplished the first steps. Within the limited time frame of this work, it was nearly impossible to complete the development of the algorithm. However, the satisfactory results prove that the combination of genetic methodology and physical guidance is on the right direction and promising towards the final goal. It can be predicted that given enough physical guidance, the optimization algorithm could reach a great global searching ability with a high level of automation and efficiency.

Besides the resolution-related aberration analysis and correction topics, the distortion correction potential of freeforms is also investigated with a case study of selected spectrometer systems. The results of all the research topics in this work indicate the powerful aberration correction ability of non-spherical surfaces, which proves again the necessity of modern optical design method research. As an outlook of this work, there are several points for future search. The MRT method can be combined with the aberration of the real reference ray for an off-axis system, so that the absolute aberration calculation can be more precise. The further development of the GACOR algorithm is also desired to cover the whole optimization process, from the singlet system to the final nominal design, especially for complicated systems with non-spherical surfaces. When this goal is fulfilled, a more generalized optimization method applicable for more optical system types can be considered, which requires considerable detailed physical guidance and experience in the program. Ideally, the algorithm can be extended to the full optimization process of the off-axis systems with more advanced design methods for the 3D structure.

Appendix A: Verification of the MRT method

In this appendix, all the new functions of the MRT methods are verified with concrete example optical systems. The total aberration surface contributions are compared to the results in [19]. Due to the unavailability of comparable calculation methods for full-order intrinsic/induced aberrations and additive Zernike representation, the MRT calculation results are verified to the best extent with limited references.

A.1 Transverse aberration calculation results

It is illustrated in [19], the calculation results of Δx_j and Δy_j with the 4×4 matrices are comparable to the exact values resulted from Aldis theory. Due to the change from local to global reference, it is necessary to first verify the calculation with the MRT with a reproduction of such comparison. Therefore, the same triplet system, first proposed by Brewer [28] and later chosen as a test system in [19], is selected again for the purpose. The system layout is shown in Figure A.1, as well as the corresponding spot diagram of the outermost y-direction field of 16.5° , the lens data of which can be found in [19].

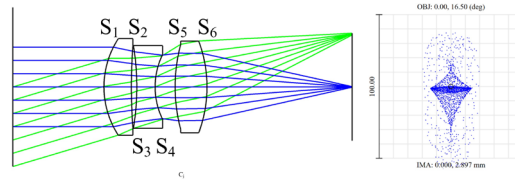


Figure A.1. Layout of the triplet system proposed by Brewer, and the spot diagram of the outermost field in the y-direction (scale: $100\mu\text{m}$).

The comparison between the MRT results and Aldis results is shown in Figure A.2. As Aldis theory takes the paraxial calculation of the system, the calculation also takes the OAR as the RR to be comparable. In total three representative rays are chosen here to illustrate the results, namely the MR, CR, and a skew ray in the 3D space. For the two rays only in the tangential plane, the Δy_j results are very close to the exact values from Aldis formulas. For the skew ray, the Δx_j and Δy_j results with the MRT method have a relatively larger difference compared to Aldis results, but still can be considered as a good approximation, as the chosen ray should suffer from the largest error compared to other rays in the system with smaller field and pupil coordinates.

Considering the sum values, both methods give almost the same value, which is also the same as the transverse aberration value obtained from Zemax real ray-tracing. In addition, it is not visualized in the figure but proved during calculation, that the transverse aberration

calculated directly in the image plane is exactly the same compared to simply adding up all the surface contributions.

It needs to be noticed that if the CR is chosen as the RR, the surface contributions, as well as the sum value, are changed, as the paraxial matrices for paraxial calculation are changed. However, the comparison of the contributions among the surfaces remains the same.

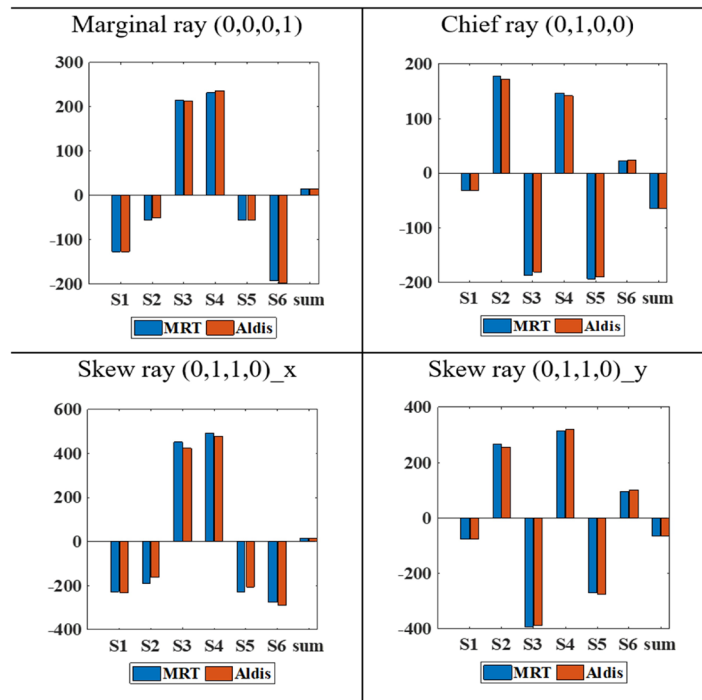


Figure A.2. Calculation results of Δx_j and Δy_j (in μm) by the MRT method and Aldis formulas, concerning different rays in the system.

A.2 Intrinsic/induced aberration calculation results

To demonstrate the reliability of the full-order intrinsic/induced aberration calculation with the MRT method, here a system with a symmetric structure is considered, the layout of which is shown in Figure A.3.

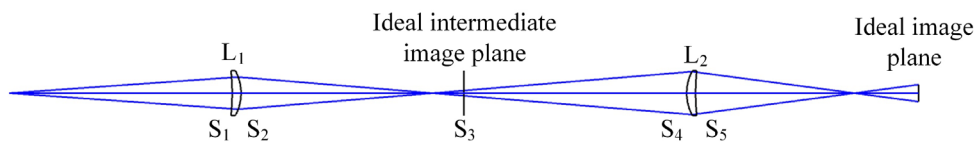


Figure A.3. Layout of the test symmetric system.

The system consists of two identical but reversed meniscus lenses with an objective space NA of 0.07. The system data are listed in Table A.1. The object and image distance are the same, and S1, S5 are both concentric. Concerning the radii of curvature, S2 and S4

are determined so that the ideal intermediate image locates exactly in the middle of the system at the same distance as the object. Therefore, in the case of paraxial calculation, the system is symmetric about S3, and both lenses have symmetric object and image.

Table A.1. Lens data of the symmetric system.

Surface number	Radius (mm)	Thickness (mm)	Material
Object	Infinity	100	
1	-100	4	BK7
2	-21.1	100	
3	Infinity	100	
4	21.1	4	BK7
5	100	100	
Image	Infinity	-	

Considering only the on-axis object point of the system, the system should suffer from pure spherical aberration, independent of the stop position. Concerning the special symmetry along the z-direction, it is easy to understand that the $\Delta y_{int,1}$ and $\Delta y_{int,5}$ should be zero, and $\Delta y_{int,2}$ and $\Delta y_{int,4}$ should be the same according to the definition. With such prediction, Δy_j , $\Delta y_{int,j}$ and $\Delta y_{ind,j}$ are calculated, and Table A.1 shows the calculation results of the surface-decomposed total, intrinsic and induced aberration of this system.

Table A.2. Transverse aberration calculation (unit: mm).

Surface number	Total aberration	Intrinsic aberration	Induced aberration
S1	0	0	0
S2	1.2011	1.2011	0
S4	2.5885	1.2004	1.3880
S5	0.0061	-0.0192	0.0253
Sum	3.7958	2.3823	1.4134

From Table A.2, the corresponding intrinsic aberration contributions can be considered as a fulfillment of these predictions except for S5. The non-zero value of $\Delta y_{int,5}$ comes from the approximation of C'_{pj} illustrated in Figure 3.5, which introduces a small error at S5. As the ideal image plane is far from the best image plane, the error is scaled, and finally leads to an obvious non-zero transverse aberration, but compared to the total intrinsic aberration, this error only contributes an additional 0.8%. In the case of induced aberration, $\Delta y_{ind,4}$ contributes 98.2% of the total value, which also makes sense concerning the system structure. $\Delta y_{ind,5}$ also shares the same error as the intrinsic result. Therefore, the results of the MRT method for full-order intrinsic or induced aberration calculation can be considered

accurate enough for determining the critical surfaces in the system and indicating the balances between them.

For a more complicated system, it is hard to predict the intrinsic aberrations by simple ray-tracing calculations. However, based on the test results, the method is considered also reliable for symmetry-free system analysis.

A.3 Surface-additive Zernike coefficient fitting

To evaluate the reliability of the Zernike coefficient fitting method, it is also meaningful to test the results of a rotationally symmetric system due to two reasons. First, due to the approximations and assumptions used in the surface-decomposed wavefront fitting method of Zemax, the results are scaled and not fully comparable to the additive surface contributions calculated with the MRT method. And the pupil distortion in the system also has an impact on the accuracy of the fitting results in Zemax. Second, the evaluation with an on-axis example system brings another convincing analysis option concerning the additive surface contributions of the primary aberrations, which is the qualitative comparison to Seidel coefficients, since Seidel coefficients are exact additive values serving as a reliable reference. Therefore, the same triplet system as shown in Figure A.1 is again taken as the test system for the verification. With the method introduced in Section 3.3, Zernike coefficient calculation takes ΔY_j instead of the additive Δy_j . Practically, the large number of sampling rays helps improve the accuracy of fitting, especially for non-spherical surfaces with high-order aberrations. The normalization radius of Zernike polynomials also needs to be calculated in advance, which can be obtained by tracing the CR and coma ray for the considered field point. Furthermore, as the piston term Z_1 is constant, it is neglected in the Zernike coefficient fitting with derivative terms.

Figure A.4 shows the Zernike coefficient fitting results of the outermost field of the triplet system corresponding to the total wave aberration of the system at the ExP. The comparison is between the MRT calculation and Zemax fitting results. The results show a very good match between each other with a maximal difference in the absolute value of 0.0176λ . The RMS of the fitting error compared to Zemax fitting results is 0.0061λ in this case. The choice of the RR for Zernike polynomial fitting also has an impact on the coefficients, as the propagation of ΔY_j is determined by the paraxial matrices. For the triplet system, if the OAR is considered as the reference, the wave aberration of the outermost field includes all the information of the wave aberration. In this case, although

Zemax calculates the Zernike coefficients with different assumptions and approximations, the results are still comparable.

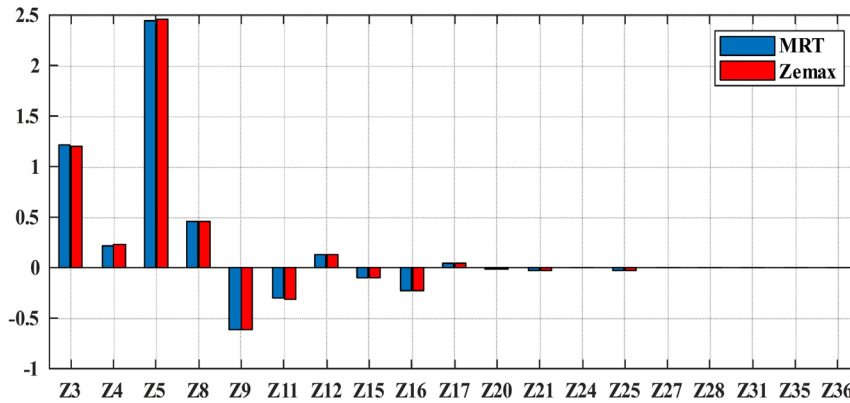


Figure A.4. Comparison of Zernike coefficient fitting results between Zemax and the MRT method calculation results. The unit is in waves.

If an arbitrary optical surface is investigated, the surface contribution of Zernike coefficients can be also calculated and plotted. Figure A.5 shows the additive Zernike coefficients among all the surfaces under some different circumstances. For better visualization, only Z9 (primary spherical aberration), Z8 (sagittal coma), and Z5 (primary astigmatism) are compared.

Plot (a) demonstrates the Zernike coefficients fitted with the MRT method with OAR as the reference, and plot (b) shows the ones calculated by Zemax. Generally, the absolute values of these coefficients in the two plots are slightly scaled, because Zemax assumes a different normalization radius for each surface, while the MRT method keeps the additivity among the coefficients. However, the relations among the three coefficients are almost the same in both two cases.

For qualitative comparison, the additive surface contributions of Seidel coefficients are also plotted as (d), which proves that both (a) and (b) match the Seidel calculation of primary aberrations. Furthermore, as the aperture dependence of primary astigmatism is quadratic, if the MRT method takes the CR as the reference, the paraxial matrices will include the information of the local toric surface in the paraxial zone during the paraxial calculation. Consequently, the astigmatism fitting results, in this case, can only be seen as defocusing. Concerning other kinds of pupil dependence for spherical aberration and coma, the problem does not exist. Thus, as (c) indicates, the astigmatism Z5 shows a completely different distribution, while the other two coefficients remain almost unchanged and still match the tendency of additive Seidel coefficients.

Therefore, if considering the ray cone and the asymmetrically located optical components together as an off-axis, the Zernike coefficient fitting results still provide an additive surface contribution of any specific aberration when Seidel is not valid anymore, except for primary astigmatism. With the help of the Coddington equation, primary astigmatism caused by the finite field can be calculated if necessary.

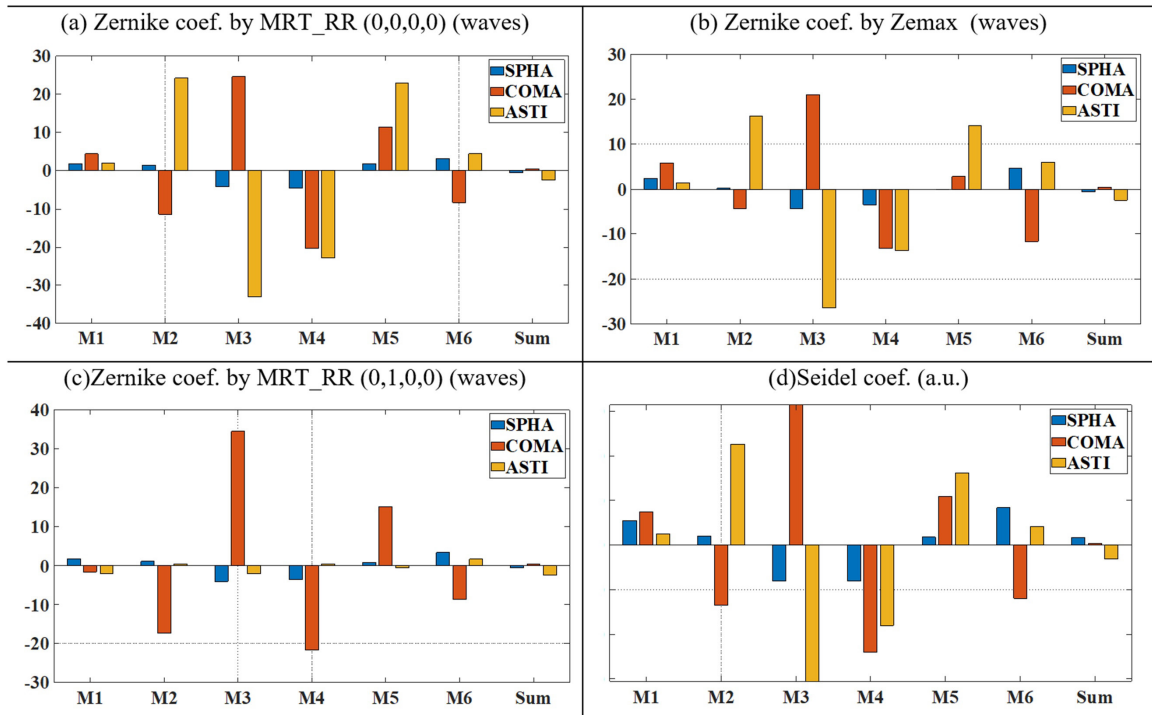


Figure A.5. Spherical aberration (Z9), coma (Z8), and astigmatism (Z5) calculated with different methods.

Appendix B: Further discussion of the MRT method

B.1 Approximation of intrinsic/induced aberration calculation

Here the approximation when calculating the intrinsic/induced aberration is discussed in more detail to get an idea about the validity of this simplification. To simplify the problem, the discussion here considers only the y - z cross section, and the corresponding coordinates all refer to the local coordinate system of the surface with the origin located at the surface vertex. As introduced in Section 3.2, the real ray vector $\mathbf{r}_j = (x_r, y_r, u_r, v_r, 1)'$ coming from the object intersects S_j at C_j . The tangent plane T is drawn at C_j , of which the projected line is illustrated in Figure B.1, together with the coordinates of the intersection points.

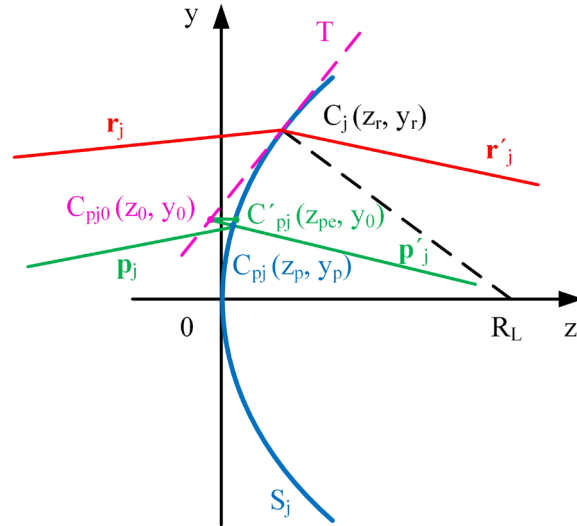


Figure B.1. Surface sag in the tangential plane and the tangent plane T .

Assuming S_j is an arbitrary freeform surface with the local radius of curvature R_L at C_j , the slope of the projected line of T (marked red) can be obtained as

$$k_T = \frac{R_L - z_r}{y_r}. \quad (\text{B.1})$$

Inserting the coordinate of C_j , the projected line of T writes as

$$y_1 = k_T z_1 + (y_r - k_T z_r), \quad (\text{B.2})$$

where z_1 and y_1 are the independent and dependent variables. As the paraxial ray vector \mathbf{p}_j coming from the object is refracted with the paraxial matrix, a ray vector \mathbf{p}'_j can be calculated after refraction by S_j . Clearly, the projected slope of \mathbf{p}'_j is the direction angle,

denoted as k_p . Assuming \mathbf{p}'_j intersects the surface at C_{pj} , the projected line function of \mathbf{p}'_j is

$$y_2 = k_p z_2 + (y_p - k_p z_p). \quad (\text{B.3})$$

Therefore, the intersection point C_{pj0} (z_0 , y_0) can be obtained by Eq. (20) and Eq. (21)

$$z_0 = \frac{k_T z_r - k_p z_p + y_p - y_r}{k_T - k_p}, \quad (\text{B.4})$$

$$y_0 = \frac{k_T k_p (z_r - z_p) + k_T y_p - k_p y_r}{k_T - k_p}. \quad (\text{B.5})$$

Consequently, the error in the y-direction is

$$E_y = |y_0 - y_p| = \left| \frac{k_T k_p (z_r - z_p) + k_p (y_p - y_r)}{k_T - k_p} \right|. \quad (\text{B.6})$$

As the surface sag of an arbitrary surface can be described as the basic spherical shape and the derivative of freeform term along the z-axis, with Tylor expansion of spherical basic shape, it can be written as

$$z_s = \frac{y^2}{2R} + \frac{y^4}{8R^3} + \dots + \Delta z_{FF}, \quad (\text{B.7})$$

where R is the surface radius, and Δz_{FF} is the freeform deviation. Thus, the total error of the surface sag in the z-direction is

$$E_z = |z_{pe} - z_p| = \left| \frac{y_0^2 - y_p^2}{2R} + \frac{y_0^4 - y_p^4}{8R^3} + \dots + (\Delta z_{FF,pe} - \Delta z_{FF,p}) \right|. \quad (\text{B.8})$$

Therefore, Eq. (B.6) shows that both y- and z- components of the distance between C_j and C_{pj} have an impact on E_y . Physically, such pitfalls can be visualized in Figure B.2. On one hand, if \mathbf{p}_j and \mathbf{r}_j are far from each other, which means the surface suffers from a considerable induced effect from the front groups of the system, then despite the small freeform deviation, the large distance in the y direction would cause a large error in the results; On the other hand, if the surface sag has a strong deviation from the basic spherical shape, such as a turning point illustrated in Figure B-2 (right), a large distance between C_{pj} and C'_{pj} in z-direction will occur, which also finally leads to an unneglectable error.

However, as for the error finally considered in the image plane, the analytical error

cannot be analytically calculated due to the following reasons: Firstly, the exact coordinates of the intersection points of the paraxial ray on the real freeform surface cannot be obtained with simple geometrical methods. Consequently, the true distance between C_{pj} and C'_{pj} where the error origins remain unknown. Secondly, as \mathbf{p}_j and \mathbf{r}_j need to be transferred to the image plane for the transverse aberration calculation, the error included in the vector is also magnified or demagnified by the following optical components, which cannot be predicted due to the huge variety of the system structure, the surface shape, and the correction performance. In addition, according to Eq. (B.8), the full-order error E_z can be strongly influenced by the additional freeform deviation, and the order of the surface sag expansion also impacts the analytical evaluation of E_z .

Consequently, there are three factors that determine the final error in the image plane: the splitting of C_{pj} and C'_{pj} caused by the induced effect from the front group, the local curvature of the surface, and the magnification of the error due to the following part of the optical system. For most optical systems, whose surfaces are smooth without turning points, the error caused by the approximation can be neglected. In practice, the compact cellphones system with strong aspheres can be an exception. The error can be critical, if the system structure is complicated with various groups but the front groups suffer from the inferior correction. In this case, the large induced effect could cause a considerable error in the image plane, when calculating the intrinsic/induced aberration with the MRT method for the rear group surfaces. Thus, the complicated microscopy systems and lithography systems should be paid more attention when considering the performance of the front groups, when applying the MRT method. Otherwise, the error can be always considered acceptable.

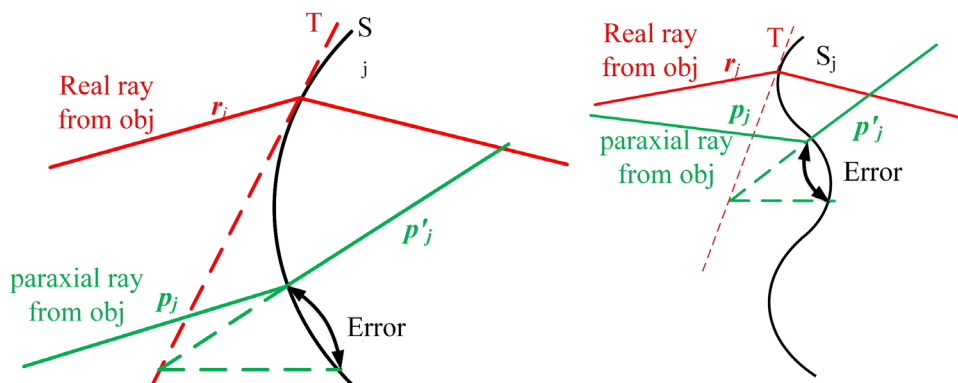


Figure B.2. Possible pitfalls of the approximation: the surfaces suffering from considerable induced effect (left), or freeform surfaces with extremely large deviations from the basic spherical shape (right).

B.2 Distortion removal

As the RR determines the paraxial zone of the system, the paraxial matrices are strongly dependent on the choice of the RR. Figure B.3 shows two different paraxial zones when the OAR or the chief ray of an off-axis field are considered as the RR respectively, assuming the OAR is not folded. It is well known that for systems with rotational or double plane symmetry, the RR is usually the common optical axis, which is also the reference for paraxial ray-tracing. In this case, if a finite field is under investigation, the transverse aberration of an arbitrary coma ray from this field consists of two parts: the distortion brought by the chief ray of this field Δy_{Dist} , and other aberrations, denoted as Δy_0 . Then for the surface contribution Δy_j , we have

$$\Delta y_j = \Delta y_{Dist,j} + \Delta y_{0,j}. \quad (\text{B.9})$$

Therefore, if the ray-tracing data of a rotationally symmetric system is considered as the reference to verify the calculation of the MRT method, the OAR needs to be the RR.

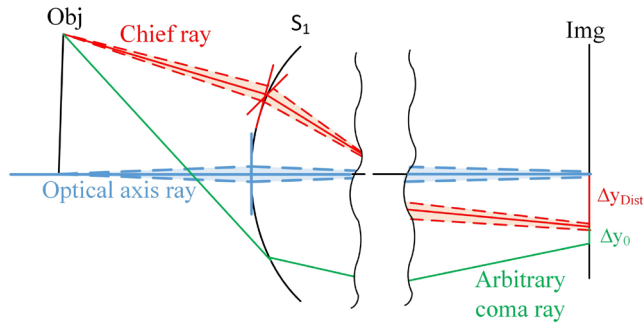


Figure B.3. Different choices of RR and the corresponding paraxial zone.

Sometimes it is necessary to evaluate the aberrations without distortion for a better observation of the resolution-related aberrations, especially when the distortion is critical. In this case, we want to remove $\Delta y_{Dist,j}$ from Δy_j . Combining Eq. (3.3) and (B.9), we have

$$\Delta y_{Img} = \sum_{j=1}^N (\Delta y_{Dist,j} + \Delta y_{0,j}). \quad (\text{B.10})$$

In the image plane, we have correspondingly

$$\Delta y_{0,Img} = \sum_{j=1}^N \Delta y_{Dist,j} - \Delta y_{Dist,Img} + \sum_{j=1}^N \Delta y_{0,j}. \quad (\text{B.11})$$

For the coma rays of the finite field, $\Delta y_{Dist,j}$ here is exactly the surface contribution of

the chief ray distortion, and it adds up to the distortion on the image plane, which can be finally eliminated to obtain

$$\Delta y_{0,Img} = \sum_{j=1}^N (\Delta y_j - \Delta y_{Dist,j}). \quad (\text{B.12})$$

Therefore, by simply subtracting the surface contribution of the CR from one of the coma rays, we can still keep the additivity of the transverse aberration surface contribution excluding distortion.

Concerning the RR choices for different field points, there are various situations where the distortion is included in the results or automatically excluded. In addition, the cases are also not identical for rotationally symmetric and symmetry-free systems. The possible situations are concluded in Table B.1 as a summary. Following the definition of the normalized field and pupil coordinates, the RR can be represented by a 4-dimensional coordinate. Then by definition, OAR is described by (0,0,0,0), and the other case of RR represented by (0,1,0,0) refers to the CR of any off-axis FoV. The field coordinate (0,1) here also represents any finite field point away from the center of the object. The comparison illustrates if $\Delta y_{Dist,j}$ exists in the coma ray transverse aberration, and if the transverse aberration in the image plane is comparable to Aldis surface-decomposed results calculated with the ray-tracing data.

Table B.1. Possible situations with RR and tested fields.

	No.	RR	Tested field	$\Delta y_{Dist,j}$ included?	Δy_j comparable to Aldis results?
Rotationally symmetry system	(a)	(0,0,0,0)	(0,0)	No	Yes
	(b)	(0,0,0,0)	(0,1)	Yes	Yes
	(c)	(0,1,0,0)	(0,1)	No	No
Symmetry-free system	(d)	(0,0,0,0)	(0,0)	No	No
	(e)	(0,0,0,0)	(0,1)	Yes	No
	(f)	(0,1,0,0)	(0,1)	No	No

As the original Aldis theory is limited to rotationally symmetry systems, only when the RR is the unfolded OAR, the results of the MRT method are comparable to Aldis calculation considering the surface contribution of the full-order transverse aberration, as in case (a) and (b).

if the CR of a finite field is chosen as the RR in a rotationally symmetric system (case (c)), the ray cone from the field together with the optical components of this system can be

considered as an off-axis system, as marked by red in Figure B.1. Therefore, case (c) is comparable to case (f). Then automatically, the matrices already contain all the field-related aberrations that exist in the paraxial zone during propagation. Therefore, if Δy_j of an arbitrary ray from this FoV is calculated with the paraxial matrices, the final result will not contain separated distortion.

Furthermore, if the OAR is taken as the RR, the existence of distortion shows the same tendency regardless of the symmetry of the system. As long as the tested ray comes from an off-axis FoV, the aberration calculation results include distortion (case (b) and case (e)), otherwise, distortion does not exist (case (a) and case (f)).

Appendix C: Parameterization of the GACOR algorithm

The GACOR algorithm is an improved optimization algorithm inspired by the great potential of the basic ACOR algorithm for optical design. Therefore, the important parameters involved in the GACOR algorithm can be referred to Section 2.8, and [25] discusses the physical meaning and the impact on the ACOR algorithm results in more detail. Concerning the GACOR program introduced in this dissertation, there are two kinds of parameters. The first parameter category includes the fixed parameter during the optimization process, such as the bias parameter q and the convergence speed parameter ζ . The second category consists of parameters that are adaptive during the program corresponding to the physical guidance. As introduced in [25], the fixed parameters do not have a great impact on the optimization result if they stay in the meaningful range, particularly for high-dimensional optimization problems. Therefore, it is not necessary to fix the values of such parameters here. In this appendix, only the important adaptive parameters, namely the GA capacity and ant group number, are discussed.

First, serving as the database of the GACOR algorithm, the GA plays an important role during the optimization. At the beginning of the global exploration, all the ant groups start from the initial system, as it is the only given system stored in the GA before the main iteration starts. Therefore, we have $K_{g0} = 1$, where the numerical subscript indicates the main iteration index, and 0 here indicates the initial index.

Very often, for the nominal optimization process, where the system starts from a singlet initial system, the system only has very few variables and quite relaxing boundary conditions. Thus, the MF landscape is not complicated due to the low dimension and boundary complexity, which contains only a limited number of local minima. As the system structure becomes more and more complicated with more lenses during the optimization process, the variety of the new solutions found by the ant groups strongly grows due to the higher dimension of the MF landscape. In other words, in the first several iterations, it may happen that the ant groups have to pick the same GA solution and output the same solution after local exploration, resulting in redundant calculations. Therefore, the size of the GA should be adapted dynamically according to the increasing variety of the solutions to avoid such efficiency degradation in the beginning. Additionally, the maximum allowed capacity K_{gm} is also given to limit the size of the GA. Dependent on the options of structural

changes, the capacity will finally reach the maximum allowed value sooner or later.

After all the ant groups have finished their local exploration in the i^{th} main iteration, the newly generated various solutions are merged together with the original solutions in the GA. Then, each solution is compared with every other in the GA, and only the unique ones can remain, the process of which can be referred to Section 4.1.2.4. Finally, the total number of those unique solutions is counted and denoted as K'_{gi} . Concerning the maximum allowed capacity, the GA capacity for the next iteration is determined as

$$K_{gi} = \min\{K'_{gi}, K_{gm}\}. \quad (C.1)$$

Besides the GA capacity, the ant group number P_g should also be adapted dynamically due to the same reason of the increasing possible solutions. In the beginning, P_g is more limited to avoid the repetitive local exploration process, but increases with the GA capacity. Specifically, P_g should be adapted according to the estimated number of unique new solutions. However, the assessment is not trivial because the number of local minima of the MF topology always remains unknown, and the structural change options also scale the possible solutions. Thus, as a part of the probabilistic feature of the original ACOR idea, the ant group number is simply determined by a rigorous mathematical model designed differently for RI and RII.

In RI, as the splitting option is the only structural change option, the pre-estimated ant group number P'_{gl} is determined only by the iteration index i_g , written as

$$P'_{gl} = \max\{k \in Z \mid k \leq (i_g + 2) / 2\}. \quad (C.2)$$

And in RII, the rough number of the expected unique new solutions, denoted as P'_{gII} , is represented as

$$P'_{gII} = 2K_{gi} \cdot \langle N_L \rangle, \quad (C.3)$$

where the scaling factor of 2 corresponds to the structural change options, and P'_{gII} is determined by the GA capacity size K_g and equivalent lens number $\langle N_L \rangle$. For both RI and RII, a maximum ant group number P_{gm} is also given to prevent the unnecessary execution time in each main iteration. Finally, the ant group number writes generally as

$$P_g = \min\{P'_{gl}, P_{gm}\}. \quad (C.4)$$

Appendix D: Searching for the best structural change

In order to answer the questions raised in Section 4.1.2 regarding the choice of the structural change option, the rules of the structural change surface determination, as well as the operations of the structural changes developed in the GACOR algorithm, the basic rules for structural changes are introduced in this appendix.

D.1 Choice of the structural change option

As mentioned, the algorithm only makes structural changes with spherical surfaces in RI. Therefore, considering the two options investigated in the study, only the split option is allowed in RI, while both options are allowed in RII. Therefore, each ant group can choose these options randomly in RII.

As mentioned in Section 4.1.2, when choosing from the structural change options, the algorithm considers the current $\langle N_L \rangle$ of the system, and the asphere option is only allowed if $\langle N_L \rangle \leq \langle N_L \rangle_{max}$ is still valid after the structural change.

If the current system already has $\langle N_L \rangle = \langle N_L \rangle_{max}$, then neither of the structural changes can be chosen, the ant group simply skips this step and directly executes the further local exploration.

D.2 Decision of the structural change surface

Empirically, the surface which contributes a large aberration should be considered as the structural change surface. However, although Seidel aberration theory can be a good indicator of the critical surfaces, the best choice of the structural change surface is still dependent on other criteria, such as the higher-order aberration or manufacturability issues. Thus, there is no fixed universal rule for the choice, as the importance of different criteria may vary under different conditions. Concerning the spherical aberration alone in this work, in addition to the surface contribution, the MR height and the incident ray angle on this surface should also be considered as impact factors of the surface sensitivity. Therefore, the comprehensive criteria for the determination of the critical surface should include all three aspects.

Considering the surface-decomposed spherical aberration contribution, Seidel coefficient S_{Ij} is sufficient for systems with only spherical surfaces under relaxed conditions. However, for systems with aspherical surfaces, the higher-order aberrations

cannot be ignored. Thus, for the rotationally symmetric systems with aspheres, the MRT method is used for analyzing the surface-resolved total transverse aberration. Here the aberration contributions for the j^{th} surface are denoted as α_{Sj} for Seidel coefficients and α_{Aj} for the transverse aberration with the MRT method. In addition, the MR height and the MR incidence angle at every surface are also calculated, the parameters of which are denoted as α_{Hj} and α_{Ij} respectively.

With the parameters, the overall evaluation of the critical surface can be modeled, indicated by a ‘critical index’, denoted as A_j . Considering the different units of all the parameters, it is impossible to represent the A_j by simply adding up all the involved parameters, but it is still possible to interpret the sum of them if they are all normalized before, written as

$$A_j = \sum_k \bar{\alpha}_j, \quad k \in \mathbf{Z} \quad (\text{D.1})$$

where $\bar{\alpha}_j$ generally represents any of the normalized value $\bar{\alpha}_{Sj}$, $\bar{\alpha}_{Aj}$, $\bar{\alpha}_{Hj}$, and $\bar{\alpha}_{Ij}$. Concerning the calculation of $\bar{\alpha}_j$, the surface diameter D_{Sj} should also be included as a scaling factor, as a larger surface aperture with a larger MR height usually causes a large degradation of the imaging performance. Therefore, the modeling of $\bar{\alpha}_j$ can be written as

$$\bar{\alpha}_j = \frac{|\alpha_j|^\beta \cdot D_{Sj}^\varepsilon}{\sum_{j=1}^N |\alpha_j|^\beta \cdot D_{Sj}^\varepsilon}, \quad (\text{D.2})$$

where α_j is the general representation of α_{Sj} , α_{Aj} , α_{Hj} , and α_{Ij} . N is the total optical surface number of the system. β and ε are the exponential scaling factors of the j^{th} surface, as the parameters should also be scaled according to their different impacts on A_j . For example, compared to the incidence angle, the Seidel coefficient has a greater influence on the surface aberration contribution, therefore α_{Sj} should be scaled with a higher weighting. Table D.1 lists all the scaling factor values set in the GACOR algorithm.

In the different cases according to the existence of asphere in the system, the most relevant parameters are not always the same. For instance, if there are already aspheres in the system, due to the unavoidable higher-order aberrations, A_j cannot be represented by α_{Hj} or α_{Ij} as well as by α_{Aj} . Thus, Table D.1 also illustrates the various cases, where only the most relevant parameters are included in the mathematical model, marked by ‘Yes’.

Table D.1. Parameterization of α_j and the involved ones in different cases.

	β	ε	Case a*	Case b*	Case c*
α_{Sj}	1	2	Yes	Yes	
α_{Aj}	1	2			Yes
α_{Hj}	1/2	2	Yes		
α_{Ij}	1/4	2	Yes		

*Case a: splitting option in the system; Case b: asphere option when the system currently has no asphere; Case c: asphere option when the system currently has already an asphere;

Although the surface with the highest A_j value can be considered a good candidate for the structural change, it is not ensured that this choice is the only best way to find a successful solution. Therefore, to enlarge the variety of solutions with a balance of the deterministic and the probabilistic feature, the algorithm chooses a surface in the same way as the ant chooses an archive solution. According to the A_j values, all the surfaces are ranked in descending order and each one is assigned to a weighting as Eq. (2.21) indicates. Then, the probability of a surface being chosen as the final critical surface for the structural change is determined according to the weightings.

In practice, it is only of limited benefit to use a lens with both surfaces aspherical in the case of thin lenses. Therefore, in the case of the asphere option, both surfaces of the aspherical lens in the system are deleted from the ranked candidates before choosing the critical surface.

D.3 Lens splitting option

If the splitting option is decided for the lens where the critical surface locates, the lens will be split into two by inserting a surface inside the lens. Figure E.1 illustrates the process of splitting the lens with the radii of curvature of R_1 and R_2 . If the lens is too thin, the thickness of the two split lenses may not be physically meaningful. Thus, dependent on the diameter D_j of the lens, the lens should be first thickened to $D_j/5$, if the current thickness is smaller than this value. Then, two adjacent surfaces with a radius of curvature R_3 are inserted in the middle of the lens, so that the thickness is evenly divided. Concerning the choice of the R_3 value, there are different options. Usually, the inserted surface can be a plane plate or has the average value of the two curvatures. In this work, the radius of R_3 is calculated as the average value, written as

$$\frac{1}{R_3} = \frac{1}{2} \left(\frac{1}{R_1} + \frac{1}{R_2} \right). \quad (\text{D.3})$$

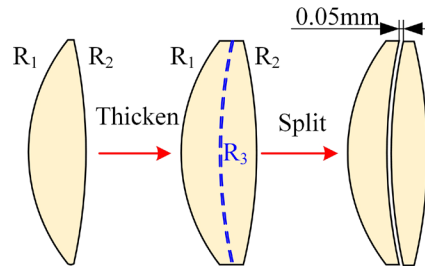


Figure D.1. Splitting process of the lens.

And finally, the two lenses are separated with a small air gap of 0.05mm in between, which is an empirically appropriate value in the normal cases. Due to the change in the thickness and the small air gap, the MF topology will be changed after the lens splitting. To prevent a large sudden change of the MF, the image distance is immediately optimized again for the smallest RMS spot after the splitting.

D.4 Aspherization

As one of the structural options included in this work, the asphere option is an important step for the system to switch from spherical to aspherical system classification. Due to the application of the aspheres, the system structure can be varied a lot, and consequently, the manufacturability assessment might require very different criteria. Therefore, the settings of the aspherical surfaces should be clarified for different purposes during the optimization.

The asphere option is allowed from the start of RII. It is defined in the algorithm that only one kind of asphere surface is allowed for one optimization task regardless of the aspherical surface number, and the specific type should be determined by the user before the optimization starts. Thus, when the asphere option is chosen, the algorithm will switch the surface type from spherical to this specific asphere type for the structural change surface. For the nominal design function of the GACOR algorithm, the asphere option only allows ‘even asphere’ and ‘Q-type asphere’ as the aspherical surface type, while for the final improvement function, the non-spherical surface type ‘Zernike fringe sag’ is allowed. Following the optimization method with non-spherical surfaces introduced in Section 2.6, the coefficients of the non-spherical terms are set as variables step by step from the lower to higher order during the optimization. However, to avoid the critical tolerance in manufacture when dealing with high-order non-spherical surfaces, the maximum order of the non-spherical terms should be also limited.

As for the asphere option programmed in the nominal design function of the GACOR

algorithm, when performing the structural change, only κ is included as the asphere-related variable for the local exploration to prevent a large sudden increment of the MF dimension. The higher-order additional aspherical terms are not increased unless the system optimization comes to the asphere enhancement process. In addition, simultaneous with any structural change option of each main iteration, the variables of all the surfaces which are already aspheres in the system are adjusted by adding the next higher-order aspherical terms as variables. In this way, the aspherical terms of the aspherical surfaces are always of higher-order as the following iterations are executed until reaching the limited order. To illustrate the new terms involved in each new execution time point,

Table D.2 shows the term coefficients that are added as variables when they are supposed to be adjusted. In the GACOR algorithm, for each kind of non-spherical surface, the maximum allowed term can be seen in the ‘final’ column, which means no further terms can be included for the surfaces. Particularly for the Zernike fringe surface, as the algorithm currently only deals with systems with plane-symmetric structure, only the plane-symmetric Zernike terms within each order are included as variables. When the non-spherical terms are added to the system, all the already existing variables in the system remain unchanged.

Table D.2 Addition of the asphere coefficients.

Surface type	1 st round	2 nd round	3 rd round	...	Final round
Even asphere	κ	a_2, a_3	a_4, a_5	...	a_8
Q-type asphere	κ	a_1, a_2	a_3, a_4	...	a_{15}, a_{16}
Zernike fringe sag	κ	$Z_4 - Z_9$	$Z_{10} - Z_{16}$...	$Z_{26} - Z_{36}$

Regardless of the nominal design function or the final improvement function, when the asphere enhancement is executed, the algorithm keeps the same MF, and reoptimizes the system step by step. During this process, the addition of the variables also follows the same rules of the increment of the order. And after each addition round, the system is optimized with the DLS algorithm again, and then evaluated. If the system meets all the original specifications, the process immediately stops and outputs the solution; If not, the next higher-order terms are added and the same process will be executed again until the system is successful, or all the last allowed terms of the non-spherical surfaces are included in the system.

Appendix E: Switch from RI to RII

As introduced in Section 4.1, RI helps to estimate the number of spherical lenses needed for a successful solution. As the lens number increases, there can be two possibilities after a certain time point: First, if a successful solution is obtained, the current equivalent lens number $\langle N_L \rangle_0$ in the system is estimated as the necessary lens number for a successful solution. Second, if it is determined that the chance to have a successful solution with only spherical surfaces is very low, which means the splitting option is not enough anymore, then the asphere option should be included for further optimization. In this case, $\langle N_L \rangle_0$ is also recorded and the algorithm switches to RII from the next main iteration.

In the later iterations, it can be predicted that the systems with more lenses are evaluated as better performing compared to those with fewer lenses in the earlier iterations. Consequently, it may happen that only solutions with more lenses remain in GA when RII starts. Therefore, it makes no sense to keep using the current GA for the ant groups, as the asphere option may introduce more lenses than needed based on the current GA solutions. Therefore, a new GA is needed as the beginning database for the ant groups. To distinguish the GA in the two rounds, the GA is denoted as GAI for RI and GAII for RII.

The purpose of the establishment of the ASB in RI is to prevent the calculation performed at the beginning of RI from being repeated in RII. The ASB collects all the various solutions ever found in RI and categorizes them according to the lens number, so that they can be directly recalled from the GAII for the ant groups. In the ASB, all the solutions are categorized according to the lens number N_L , and all the solutions in the GAII are directly taken from the ASB when RII starts. Figure F.1 explains the relation which solutions stored in the ASB are to be included in the GAII when the algorithm switches to RII. The number of solutions in each category is qualitatively shown with the blue bar, and the categories are partly listed along the horizontal axis indicated by the increasing N_L . When the GAII is requested, the ASB collects all the solutions in the categories with the lens number of $\{N_{LII} - 1, N_{LII}, N_{LII} + 1\}$, where N_{LII} is the nominal lens number of RII, calculated as

$$N_{LII} = \begin{cases} 2, & \text{if } 3 \leq \langle N_L \rangle_0 \leq 5 \\ \max\{k \in Z \mid k \leq \langle N_L \rangle_0 / 3\}, & \text{if } \langle N_L \rangle_0 > 5 \end{cases} \quad (\text{E.1})$$

In Figure E.1, the solution categories included in the GAII are marked in the red zone in comparison to the whole ASB in the yellow zone. It should be mentioned that, the GACOR

algorithm considers only the optimization task for which asphere surfaces are meaningful for improving the performance. In other words, if the optimization task is very simple and can be successful with only two spherical lenses, correspondingly $\langle N_L \rangle_0 = 2$, the application of the GACOR algorithm is not considered as necessary.

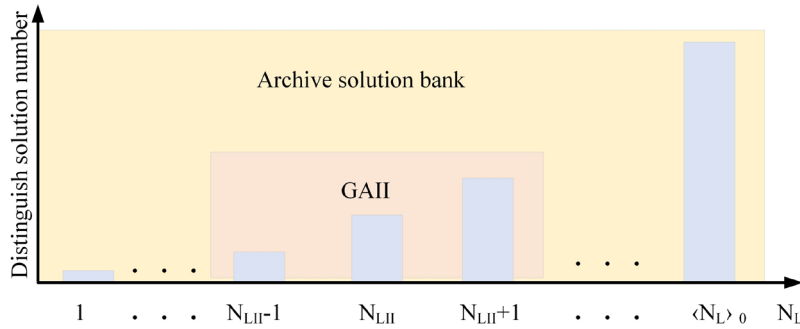


Figure E.1 Categories of solutions in GAI and ASB.

As for the analysis of the chance of obtaining a successful solution with only spherical surfaces, the criteria are given as follows: After the local exploration is finished in each main iteration, the solution categories with the lens number of $\{N_{L0} - 1, N_{L0}\}$ from the ASB are checked. For each category, the best solution with the lowest OPC value is picked out, and the improvement in percentage is calculated referring to these two solutions. This value is considered as the ‘pure’ improvement brought by introducing one more lens in the system. If the improvement reaches more than 10%, it is considered meaningful to further split the lens for the optimization. Otherwise, the algorithm stops RI and switches to RII.

For instance, Figure 5.14 illustrates the first successful system ever found by the program since the start of running, which is solution **C1**. Therefore, the estimated necessary lens number needed for a successful solution is determined as 5, namely $\langle N_L \rangle_0 = 5$. Therefore, $N_{LII} = 2$ is also obtained according to Eq. (E.1), and the starting GAI can be established. According to the rules mentioned above, **C1** is not included in GAI, while the program places all the intermediate solutions ever found with the lens number of 1, 2, and 3 in the GAI. On one hand, these solutions in the GAI help to prevent redundant calculations from the initial system in RII to improve the efficiency, while the intermediate solutions with fewer lenses are more appropriate for performing the asphere option. On the other hand, for instance, if one ant group follows the same optimization path shown in Figure 5.9, **C1** is obtained again in RII. This repetition is also meaningful, as more successful solutions originating from C1 can be found in the program, such as **C4.4** and **C4.5** in Figure 5.14. In conclusion, the switch between RI and RII and the establishment of the ASB are helpful for the algorithm to find various successful solutions.

Appendix F: Variable and MF adaption

For the optimization procedure, the MF is essential as it indicates the stepwise optimization targets, and it is the most important performance evaluation reference of the GACOR algorithm. The MF adaption rules for optimization concerning the specific task is discussed in Section 5.2 and 5.3. In this appendix, the general MF and variable adaption rules are introduced.

F.1 MF adaption in various cases

As introduced in Section 4.1, to simplify the editing of the MF in the optical software via the algorithm programmed in Matlab, the operands corresponding to the system specifications are set in the MF before the optimization starts, and the indices of the operands are all recorded in the algorithm to help the program locate the operands when editing the MF. Thus, during the optimization, the algorithm only edits the necessary parameters, weightings, and targets of the operands in the MF. During the local exploration, the MF should be adapted according to the optimization strategy, and the variables should also be adapted to fit the stepwise optimization goal.

First, when the ACOR local search or the DLS booster is executed, the MF needs to be edited for the optimization goals as the instruction of the optimization. Second, as for the solution evaluation, the performance is quantified with the MF calculation results. The performance is either evaluated for the SPC which focuses only on the stepwise optimization goals, or the OPC regarding the original specifications. With the corresponding resulted values, the solutions can be ranked in the GA or SSA. Therefore, the MF adjustment is necessary for both the optimization and evaluation purposes.

Considering these two main functions of the MF, the detailed methods about the MF and variable adaption in various cases are given in Table F.1, where ‘ W ’ and ‘ T ’ represent the weighting and targets, and the subscripts ‘ p ’, ‘ g ’, and ‘ a ’ indicate the imaging performance-related operands, and those for the glass or air thicknesses. The target values of the air and glass thicknesses are marked with ‘Yes’ if they are adapted. As for the specific performance-related operands, if the adjustment rule for T_p is described by ‘TBD’ (to be determined) in the table, the actual target values of them are determined according to the optimization strategy for the specific optimization task, which can be referred to Section 5.2 or Appendix J. If T_p is described by ‘final’, the final goal values are set as the target

values of the corresponding operands for the overall performance evaluation for example.

In addition, V_g and V_a represent the variable conditions of the glass and air thicknesses, for which ‘F’ means ‘fixed’ and ‘V’ means ‘variable’ in the settings. Following Figure 4.3, the two different cases when the MF is adjusted for optimization are marked with indices ‘i’ and ‘ii’. The index ‘iii’ is defined for the situation when MF is adjusted for SPC evaluation, and index ‘iv’ refers to the OPC evaluation case.

Table F.1. MF and variable adjustment rules in different situations.

Adaption case	Wp	Tp	Wg	Tg	Wa	Ta	Vg	Va
i	1	TBD	0		0		F	V
ii	1	TBD	1	Yes	1	Yes	V	V
iii	1	TBD	1	Yes	1	Yes	/	/
iv	1	Final	10	Yes	10	Yes	/	/

Regarding the rules, the specific considerations for these cases are discussed:

(i) is for the local ACOR search. As the glass thickness is of weaker impact on the performance compared to other lens parameters, they are not allowed variable, so the dimension of the problem is reduced. This helps to enhance the success probability of the ACOR process, and the efficiency can be also improved. In contrast to glass thicknesses, the air thickness as a variable allows for more possible structures of the system combined with the change of the curvatures found by the ants. Furthermore, with the sequential ray-tracing tool, both air and glass thicknesses are not directly relevant to the imaging performance. Thus, it is enough for the algorithm to focus only on finding the possible solutions with good imaging performance. Concerning the possible problems of the geometrical boundary conditions, the next step with the DLS booster can be applied for a subsequent correction.

(ii) is performed by the algorithm before the DLS booster process. Considering the possible improvement of the solution found during the local ACOR search, the stepwise performance-related operands should be updated for the DLS booster. Different from (i), the DLS booster also optimizes the thicknesses to ensure the physical boundary conditions. Thus, the weighting and targets of both air and glass thickness operands are adapted and the corresponding lens data parameters are set as variables.

(iii) and (iv) both refer to performance evaluation criteria. For (iii), the MF adaption is only for checking the fulfillment of the current stepwise goals during the optimization,

while for (iv), the MF is adapted according to the final goals of the system. Therefore, the original specifications are set as the targets of the performance-related operand targets with a weighting of 1. Specifically, to rank the solutions according to the OPC, the physical considerations should be emphasized, so that the systems with better manufacturability can be recommended. Thus, the weighting of the thicknesses is set as 10, so that even small violations of the boundary conditions can be ‘punished’ in the ranking. For this purpose, the variables in the system are irrelevant for both cases.

F.2 Boundary condition control

Among the MF operands, those for controlling the boundary conditions are independent of the performance requirements but are set according to the practical issues during manufacture. The thickness of the lens and the distance between the lenses should be controlled for acceptable manufacturability to avoid possible problems during the manufacturing process. When the MF is adapted, the boundary conditions of each lens should be considered individually to ensure the feasible geometry, as the diameter of the lens also has an impact on the constraints as a scaling factor.

Table F.2 Target value setting rules of thickness constraints.

Operand	Meaning	Target value
MXCA	Maximum air gap center thickness	Preset in Zemax
MNCA ₁	Minimum air gap center thickness	0.05mm for air gaps between lenses
MNCA ₂	Minimum image distance thickness	Preset in Zemax for image distance
MXCG	Maximum glass center thickness	$t_j = \frac{D_{sj}}{6}$ for all lenses
MNCG	Minimum glass center thickness	$t_j = \begin{cases} \max\left\{\frac{D_{sj}}{8}, 1\right\}, & \text{if } F_j > 0 \\ \max\left\{\frac{D_{sj}}{8}, 0.3\right\}, & \text{if } F_j \leq 0 \end{cases}$
MNEG	Minimum glass edge thickness	$t_j = \begin{cases} \frac{D_{sj}}{20}, & \text{if } D_{sj} \geq 10 \\ 0.2, & \text{if } 2 \leq D_{sj} < 10. \\ 0.1, & \text{if } 1 \leq D_{sj} < 2 \\ 0.05, & \text{if } D_{sj} < 1 \end{cases}$
MNEA	Minimum air gap edge thickness	$t_j = \begin{cases} 0.1, & \text{if } D_{sj} \geq 10 \\ 0.05, & \text{if } 2 \leq D_{sj} < 10. \\ 0.02, & \text{if } D_{sj} < 2 \end{cases}$

Table F.2 lists the target setting rules for all the necessary constraints in the MF for the edge and center thicknesses, which originate from the experience considering the production process. The rules might not be optimal and generic for all kinds of optical systems, while it is applied in the algorithm to ensure the acceptable manufacturability in the normal cases. As the MF is set in Zemax, the notations of the specific operands all correspond to the definitions in Zemax [55], and they will be used in the following discussion for simplification.

In the table, t_j generally means the target value of the corresponding constraints. D_j and F_j are the diameter and the focal power of the j^{th} lens in the system respectively.

For the MXCA constraint in the system, the target is not as important as the other constraints, as the air gaps in the system are always of a finite distance limited by the total volume. Therefore, the target value can be predicted by the user according to the overall scale of the system to prevent the system from an infinite image distance. And concerning the MNCA₂ constraint, the target value can also be estimated to ensure a feasible free working distance.

Appendix G: Similarity and lens shape check

Before a solution is considered successful and output by the GACOR algorithm, it needs to first pass both similarity checks to ensure every output solution is unique. Then, the lens shape of the solution is checked to avoid any output system with unrealistic manufacturability. In this appendix, these two checking methods are introduced.

To judge if the solution is unique, it is necessary to compare the structure of this solution to all the others in the corresponding archive. As the key factors of the structure, the thicknesses between all the lenses and the radii of curvature of them should be compared. Because of the stop criteria of the DLS local optimization algorithm, it may happen that the solutions are slightly different in the lens data, although they look almost the same. Consequently, the criteria for the similarity check should be adjusted, so that the slight differences between the two solutions can be ignored to some extent. Thus, the algorithm calculates the standard deviation values δ between the two solutions in terms of the curvature and the thicknesses, the one of the curvatures is written as

$$\delta_c = \sqrt{\frac{\sum_{j=1}^N (c_{1j} - c_{2j})^2}{N}}, \quad (\text{G.1})$$

where c_{1j} is the curvature of S_j of the first solution, and c_{2j} is obtained from the second solution. The δ_t value for the thickness shares the same format. Thus, the criteria according to empirical evaluation are determined. Specifically, if

$$\delta_c < 0.05 \cdot c_{max} \text{ or } \delta_t < 0.5\text{mm}, \quad (\text{G.2})$$

then these two solutions are considered similar. c_{max} is the maximum curvature value among all the surfaces as an indicator of the scale of the system. If the two solutions have different non-spherical surface locations, or only one of them contains non-spherical surfaces, they are immediately considered not similar, and the further check is skipped.

Besides, the lens shape also has an impact on the manufacturability of the system. Thus, for the overall performance evaluation, the lenses with strong bending should be avoided. During the lens bending check, the bending parameter X of each lens is calculated, following Eq. (4.2). Empirically, the limit of the absolute value of X_j is determined as 10 for a general assessment. Correspondingly, the system is considered appropriate and passes the lens bending check if $|X_j|$ of every lens in the system is smaller than this value.

Appendix H: ACOR local search

Due to the improved workflow of the GACOR algorithm compared to the simple ACOR algorithm, some modifications for the local ACOR search embedded in the GACOR program are necessary for effective optimization. In this appendix, the adjusted rules for the local ACOR search process are clarified.

H.1 Artificial initial deviation

According to the principle of the simple ACOR algorithm, the new position found by the ant members during the local ACOR exploration is determined by the PDF. As an important parameter for the PDF, the standard deviation σ is determined by all the solutions in the LA. A larger σ physically describes a higher probability that the ants can reach further from the initial value of the variables. However, because of the same starting system shared by all the ants, we have the initial value $\sigma_0 = 0$, which blocks the necessary deviation to generate the new solutions. Therefore, an initial value of σ_0 should be artificially set to trigger the dispersion of the ants around the starting solution.

Particularly for optical system optimization problems, the values of different lens parameter types are on different scales. For example, the surface curvature interval is limited in $[-1, 1]$, while the range of the air gap is $[0, +\infty)$ in principle. The big difference between them would cause a higher chance of failure, if σ_0 is set the same for them all. Thus, σ_0 should be dependent on the different variable types. Specifically, all the variables currently involved in the GACOR algorithm are categorized in curvature, thickness, conic section, and the coefficients of the non-spherical terms, so that they can be assigned to various setting rules. Furthermore, within one category, due to the different scales of the lens parameters and thicknesses, the impact of σ_0 on the searching performance is different. Therefore, determining the σ_0 individually for each parameter also helps to improve the global searching ability and reduce the failure. Therefore, the rules should be dependent on both the category and the initial values of the variables.

In addition, it is learned from the experience that, when non-spherical surface sag terms are involved in the system, the MF topology is more complicated with a much higher dimension, where the local minimum area can be smaller and steeper. Consequently, if σ_0 is set large in this case, the chance to find a local minimum area is lower. Therefore, to enhance the global searching ability, the σ_0 values are defined differently for the systems

with and without aspheres.

Table. H.1 lists the calculation of σ_0 according to different categories, where c_0, t_0, κ_0, z_0 are the initial values of the starting system variables. Correspondingly, $\sigma_{c0}, \sigma_{t0}, \sigma_{\kappa0}, \sigma_{z0}$ are the initial deviation values. N_L is the total number of lenses, and N_z is the total number of aspherical surface terms in the system. It should be noticed that the modeling method here is proved feasible in the program, but it may not be optimal and unique.

Table H.1. Calculation of σ_0 dependent on the existence of asphere and the variable category.

Parameter	Starting system without asphere	Starting system with asphere
Curvature c	$\sigma_{c0} = \begin{cases} \sigma_{c0}, & \text{if } c_0 = 0 \\ 5 \cdot \frac{ \sigma_{c0} \cdot c_0 }{N_L^2}, & \text{if } c_0 \neq 0 \end{cases}$	$\sigma_{c0} = \begin{cases} \frac{\sigma_{c0}}{N_z}, & \text{if } c_0 = 0 \\ 5 \cdot \frac{ \sigma_{c0} \cdot c_0 }{N_L^4 \cdot N_z}, & \text{if } c_0 \neq 0 \end{cases}$
Thickness t	$\sigma_{t0} = \begin{cases} \frac{ \sigma_{t0} \cdot \bar{t} }{N_L^2}, & \text{if } t_0 < \frac{\bar{t}}{10} \\ \frac{ \sigma_{t0} \cdot \bar{t} }{2N_L^2}, & \text{if } t_0 \geq \frac{\bar{t}}{10} \end{cases}$	$\sigma_{t0} = \begin{cases} \frac{ \sigma_{t0} \cdot \bar{t} }{N_L^2 \cdot N_z}, & \text{if } t_0 < \frac{\bar{t}}{10} \\ \frac{ \sigma_{t0} \cdot \bar{t} }{2N_L^2 \cdot N_z}, & \text{if } t_0 \geq \frac{\bar{t}}{10} \end{cases}$
Conic constant κ	$\sigma_{\kappa0} = \begin{cases} \sigma_{\kappa0}, & \text{if } \kappa_0 = 0 \\ 5 \cdot \frac{ \sigma_{\kappa0} \cdot \kappa_0 }{N_L^2}, & \text{if } \kappa_0 \neq 0 \end{cases}$	$\sigma_{\kappa0} = \begin{cases} \frac{\sigma_{\kappa0}}{N_z}, & \text{if } \kappa_0 = 0 \\ 5 \cdot \frac{ \sigma_{\kappa0} \cdot \kappa_0 }{N_L^4 \cdot N_z}, & \text{if } \kappa_0 \neq 0 \end{cases}$
Aspherical term coef. z	$\sigma_{z0} = \begin{cases} \sigma_{z0}, & \text{if } z_0 = 0 \\ 5 \cdot \frac{ \sigma_{z0} \cdot z_0 }{N_L^2}, & \text{if } z_0 \neq 0 \end{cases}$	$\sigma_{z0} = \begin{cases} \frac{\sigma_{z0}}{N_z}, & \text{if } z_0 = 0 \\ 5 \cdot \frac{ \sigma_{z0} \cdot z_0 }{N_L^4 \cdot N_z}, & \text{if } z_0 \neq 0 \end{cases}$

The difference between the existence of non-spherical surfaces is described by the scaling with N_z . And among the categories, the thickness has a different expression from all the others, as the initial value is non-zero after the structural change. Instead of t_0 , the average occupied thickness per lens \bar{t} is used to calculate σ_0 , written as

$$\bar{t} = T_s / N_L, \quad (\text{H.1})$$

where T_s is the distance from the first to the last optical surface. Thus, the deviation can homogeneously stimulate the lens variables for a beneficial deviation. As the impact of thickness changing is relatively weak compared to other types of variables, a large σ_0 is not critical, but helpful for finding more different types of solutions.

H.2 Prevention of infinite loops

The local ACOR exploration follows the basic workflow introduced in Section 2.8, which is formulated as a loop until each ant member has found a solution. When an ant finds out a new solution, the SPC of the solution will be calculated. Empirically, only those with an SPC value lower than 5 can be considered acceptable, otherwise, it is denoted as a failure. Although the setting rules for σ_0 introduced above are determined carefully, it still might happen that the ant cannot find any meaningful solution around the starting point. Particularly, the situation can be understood for two reasons: either the current initial deviation σ_0 is still too large due to the very high dimension of the MF, or the GA solution locates at a very sensitive position on the MF landscape, so that each step of the ants around the current starting point leads to great growth of the SPC. In this case, a mechanism is developed for the ants to jump out of the loop to prevent being stuck in the infinite loop.

Considering the two various reasons, the GACOR algorithm includes two methods to stop the infinite loop. First, if all the ants fail to find any new solution for more than F_{m1} times directly in the beginning of the whole process, it can be predicted that the problem is in σ_0 . Thus, all the σ_0 values are reduced to a half, and the algorithm starts the loop again. This process is repeated until each ant can find a new solution. On the other hand, a record is established for each GA solution, which counts the total failure time among all the individual ants starting from this solution. The failures due to the large σ_0 values are not counted in this record. If it is found out that the ants have made in total over F_{m2} times of failures when a specific GA solution is chosen, then this GA solution is considered as sensitive in the MF topology. In practice, concerning the tolerance for the manufacture, the system can be predicted to have a low as-built performance. Therefore, this GA solution is deleted from the GA and no more ant groups will choose this solution as its starting point of the local ACOR exploration, similar to the ‘group memory’ in nature.

In general, these two methods help to keep the local ACOR exploration in an acceptable execution time, which finally improves the efficiency of the GACOR algorithm.

Appendix I: Further optimization examples

In Chapter 5, the optimization task of a retro-focus system is introduced as one of the applications of the GACOR algorithm. The output solutions have a large variety with appropriate physical considerations, which proves that the GACOR algorithm is helpful for the users in optical design. Besides, the GACOR algorithm can also optimize other kinds of systems, if the corresponding optimization strategy can be properly adjusted. In this appendix, the results for another two optimization problems are illustrated as a supplement to the test results of the GACOR algorithm.

I.1 Tele-system optimization

The first example is the optimization of a tele-system, the basic layout of which is shown in Figure I.1. A tele-system consists of two main lens groups, namely the front positive group and rear negative groups. Due to the divergence of the negative lens group, the free working distance s' is shorter than the focal length f' . Meanwhile, the negative group also reduces the image space NA, which makes the system less critical for aberration correction, compared to a retro-focus system with the same entrance pupil size.

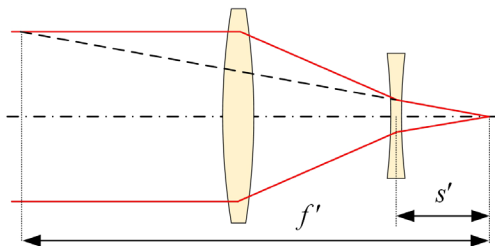


Figure I.1. Layout of a tele-system.

As for the optimization task, the system specifications are listed in Table I.1. The system only has an on-axis field, and the corresponding spot should be diffraction limited. During the optimization, the focal length is fixed in the MF as a hard constraint. Similar to the system structure specifications of the retro-focus system optimization task, the free-working distance and the total length should not be considered from the beginning, but in the later phase.

Table I.1. System specifications of the tele-system.

Entrance pupil diameter	20mm	Wavelength	550nm
Free-working distance	20mm	Total length	50mm
Focal length	100mm	Image performance	Diffraction limited
Stop position	L1 front surface	Field of view	On axis only

According to the specific system type, the optimization strategy should be adjusted to ensure the optimization path is appropriate for obtaining the tele-system solutions. Thus, the dynamic MF settings are modified, as illustrated in Figure I.2. The total length and the free working distance are denoted as V_{LT} and V_{LI} . The weighting and target values of them in the MF are correspondingly W_{LT} , W_{LI} , T_{LT} , and T_{LI} . As for this optimization task, it makes only limited sense to optimize the system with the two length constraints step by step, as the thickness parameters have a relatively weak impact on the system performance. Therefore, during the optimization process, once the system is diffraction limited, the L_I constraint is included in the MF immediately with the original specification target. Compared to V_{LT} , V_{LI} has the lowest priority, which is only activated in the MF when the system has fulfilled all the other targets.

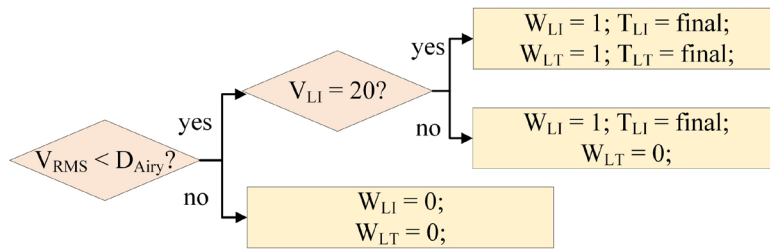


Figure I.2. MF setting rules for the tele-system optimization.

Given the maximum equivalent lens number, namely $\langle N_L \rangle_{max} = 5$, and expecting 20 output solutions, the program is executed for one time, and the obtained solutions are listed in Figure I.3. The same as Chapter 5, the solutions are also categorized according to the number of spherical and aspherical lens numbers, and each of them is named according to the category and the aspherical surface location.

Among the solutions, there are three solutions found with only three lenses. The solutions **S3.1** and **S3.2** with two meniscus lenses in the positive lens group have larger aberration deviations, but smaller spot sizes. In comparison, solution **S3.3** has a different lens bending, leading to the best aberration balance but a slightly worse spot diagram. If more spherical lenses are included in the system, the aberrations can be better balanced by all the lenses, so that the standard deviation represented by Δy_σ becomes smaller, such as **S4.3** and **S5.1**. If an asphere is applied in the system, it is harder to control the aberration distributions due to the higher-order aberration brought by the asphere. Besides, according to the calculation rules of the critical index, the probability of the negative lenses being chosen for structural changes is lower due to their smaller diameter. As a result, there is only one solution with an aspherical negative lens found by the program. The results show

again that the algorithm is capable of finding out solutions with different lens numbers and asphere locations.

Name	Layout and spot diagram (scale: 2 μ m)	Comments	Name	Layout and spot diagram (scale: 2 μ m)	Comments
S3.1		$\Delta y_\sigma = 0.47$	S3.2		$\Delta y_\sigma = 0.61$
S3.3		$\Delta y_\sigma = 0.12$	S4.1		$\Delta y_\sigma = 0.40$
S4.2		$\Delta y_\sigma = 0.34$	S4.3		$\Delta y_\sigma = 0.07$
S4.4		$\Delta y_\sigma = 0.18$	S4.5		$\Delta y_\sigma = 0.50$
S5.1		$\Delta y_\sigma = 0.04$	S1A1 (1.1)		κ $\Delta y_\sigma = 0.11$
S1A1 (2.1)		κ $\Delta y_\sigma = 0.90$	S2A1 (1.1)		κ $\Delta y_\sigma = 0.26$
S2A1 (1.2)		κ $\Delta y_\sigma = 0.51$	S2A1 (1.3)		κ $\Delta y_\sigma = 0.14$
S2A1 (2.1)		κ $\Delta y_\sigma = 0.13$	S2A1 (3.1)		κ $\Delta y_\sigma = 0.51$
S2A1 (3.2)		κ $\Delta y_\sigma = 0.25$	S2A1 (4.1)		κ $\Delta y_\sigma = 0.14$
S2A1 (4.2)		κ $\Delta y_\sigma = 0.24$	S2A1 (6.1)		κ $\Delta y_\sigma = 0.46$

Figure I.3. Output solutions of the tele-system optimization.

The same as Figure 5.30, the system performance of all the obtained solutions is summarized in Figure I.4. The standard deviation Δy_σ is calculated with the MRT method, representing the full-order transverse aberrations of the MR among the surfaces. The size of the bubbles indicates the relative spot sizes of the solution. The different solution categories are marked by different colors.

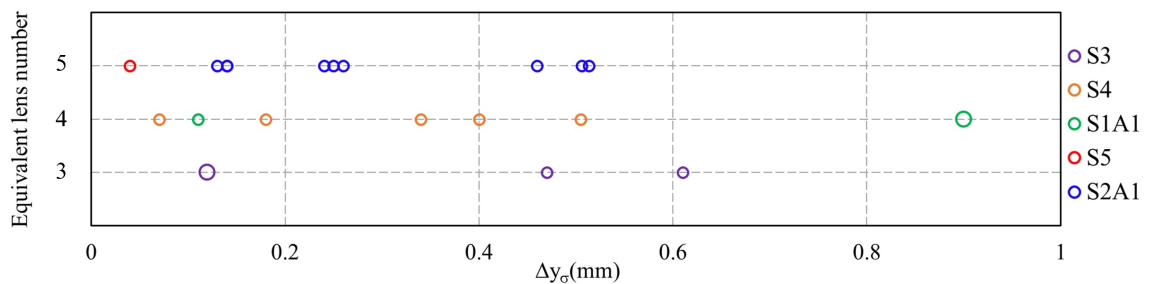


Figure I.4. Aberration standard deviation of all the solutions (unit: mm).

As for the solutions with only spherical lenses, namely ‘S3’, ‘S4’, and ‘S5’, the aberrations are better balanced as the lens number increases. However, an aspherical

surface in the surface could cause a large Δy_σ , indicating a higher sensitivity. In general, an aspherical surface is helpful for a smaller spot size.

I.2 High NA collimator system optimization

Besides the tele-system, another nominal design task for testing the GACOR algorithm is the optimization of a high NA collimator. The system structure is simple, comprising only one converging lens group to obtain the high image space NA. The initial system remains a singlet system, and the system specifications are listed in Table. I.2.

Table I.2. System specifications of the NA collimator system.

Entrance pupil diameter	20mm	Image-space NA	0.6
Wavelength	550nm	Image performance	Diffraction limited
Field of view	On-axis only	Stop position	L1 front surface

The optimization strategy of the collimator system is also simple, concerning only the image space NA and the spot size during the optimization process. Similar to the other two example systems, the diffraction limited criterion is still the first priority, while the NA value should be improved step by step. The corresponding MF setting rules are shown in Figure I.5, where W_{NA} and T_{NA} are the weighting and the target value of the image space NA operand. In this program, the NA optimization is divided into six steps, each of which refers to increasing NA of 0.1, until finally the system reaches an image space NA of 0.6.

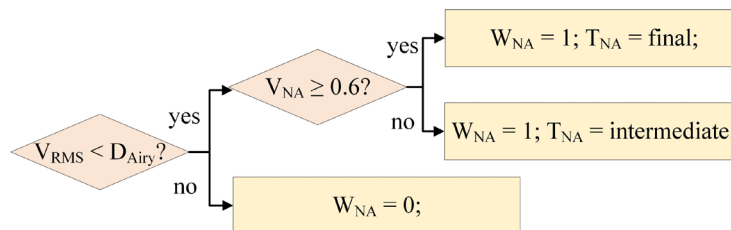


Figure I.5. MF setting rules for the high NA collimator system optimization.

Following the optimization strategy, the program is executed for one time, and the resulted solutions are listed in Figure I.6. Considering the simpler structure of the system, only 10 output solutions are desired and $\langle N_L \rangle_{max} = 5$ is again set for the maximum equivalent lens number for them.

As the high NA specification requires a strong convergence of the incoming collimated rays, almost all the lenses are positive and bent towards the image. Such a structure is more helpful for the balance of the spherical aberration distribution among all the surfaces concerning the limited lens number. Among all the solutions, only two spherical lens

systems are found. Solution **S4.2** is the only exception with a negative lens. Compared to **S4.1**, the spot is better corrected, but the aberration balance indicated by the standard deviation Δy_σ seems much worse. As for the solutions with an aspherical surface, it is found that for such a high NA system, a diffraction limited system cannot be obtained by using only one aspherical lens. Thus, in the solution collection, only the systems with both one aspherical and one spherical lens can be found. Among the solution category ‘S1A1’, the various aspherical surface locations and the lens bending of the systems result in very different aberration correction performances.

Name	Layout and spot diagram (scale: 2 μ m)	Comments	Name	Layout and spot diagram (scale: 2 μ m)	Comments
S4.1		$\Delta y_\sigma = 0.0066$	S4.2		$\Delta y_\sigma = 0.1608$
S1A1 (1.1)		κ $\Delta y_\sigma = 0.0003$	S1A1 (1.2)		κ $\Delta y_\sigma = 0.0276$
S1A1 (1.3)		κ $\Delta y_\sigma = 0.0127$	S1A1 (2.1)		κ $\Delta y_\sigma = 0.0290$
S1A1 (3.1)		κ $\Delta y_\sigma = 0.2141$	S1A1 (4.1)		κ $\Delta y_\sigma = 0.0568$
S2A1 (1.1)		κ $\Delta y_\sigma = 0.0069$	S2A1 (2.1)		κ $\Delta y_\sigma = 0.1165$

Figure I.6. Output solutions of the high NA system optimization.

As an overview of the aberration surface distributions, Figure I.7 illustrates the Δy_σ of all the output solutions against the equivalent lens number and the relative spot size. The application of an aspherical lens for this optimization task helps to reduce Δy_σ , but compared to the correction ability of **S4.1**, the improvement is not obvious. However, these two examples prove the universality of the application of the GACOR algorithm.

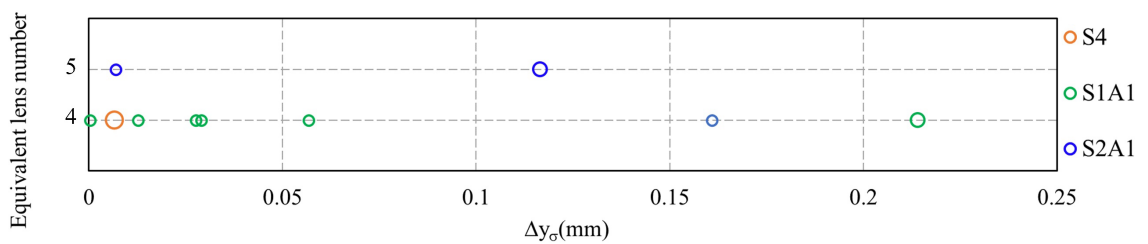


Figure I.7. Aberration standard deviation of all the solutions (unit: mm)

Appendix J: Freeform surfaces for distortion correction

For understanding the corrective power of freeform surfaces we analyze the corrective power of freeform surfaces concerning the applications in the refractive spectrometer systems for high-performance requirements.

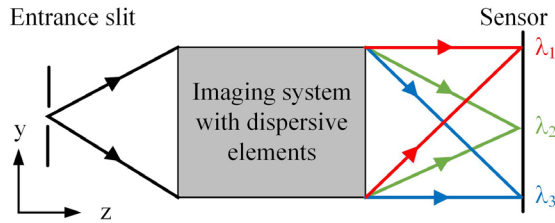


Figure J.1. Basic imaging spectrometer layout.

The correction of spatial resolution and distortion in imaging spectrometer systems is of great importance due to their significant impact on efficiency and quality. A traditional imaging spectrometer comprises an entrance slit, an imaging optical system with dispersive elements, and the detector, as shown in Figure J.1. The slit is usually rectangular, being very narrow in the y direction to limit the FoV in the dispersive direction, and wide in the x -direction (perpendicular to the y - z plane). In this way, the limited FoV in the y direction helps to separate the images formed by different wavelengths on the detector. Therefore, the slit is often simulated as a line object. The light beam with a broad spectral range enters the slit and propagates through the system to be dispersed. The dispersive elements are located at or near the pupil of the system. As the spectral components are dispersed only along the y direction, the imaging system maps them on the detector according to wavelength. Ideally, the spectral components are separated as parallel straight lines at the image plane for further scanning and analysis. In such a system, both gratings and prisms can be used as dispersive elements. The spectral separation is performed by an exit slit or a spectrally resolving detector [56].

J.1 Distortion correction of spectrometer systems

Due to the dispersive behavior, two kinds of distortion require special attention when evaluating the optical system performance. These are spectral ('smile') distortion, which means a bent monochromatic line image (assumed along the x -axis), and spatial (keystone) distortion, referring to a wavelength-dependent magnification of the entrance slit [57-60], as illustrated in Figure J.2. As for this study, they are both defined quantitatively for the evaluation. Smile distortion is defined as [61]

$$D_{\text{smile},\lambda} = y_{\text{max},\lambda} - y_{\text{min},\lambda}, \quad (\text{J.1})$$

where $y_{\text{max},\lambda}$ represents the centroid image position of the outermost field, and $y_{\text{min},\lambda}$ regards to the one of the central field. Keystone distortion is defined as [62]

$$D_{\text{keystone},\lambda} = \frac{x_{\text{max},\lambda} - \frac{1}{2}L}{\frac{1}{2}L} \times 100\%, \quad (\text{J.2})$$

where $x_{\text{max},\lambda}$ is the centroid image position of the outer field point along the x -axis and L is the length of the entrance slit along the x -axis.

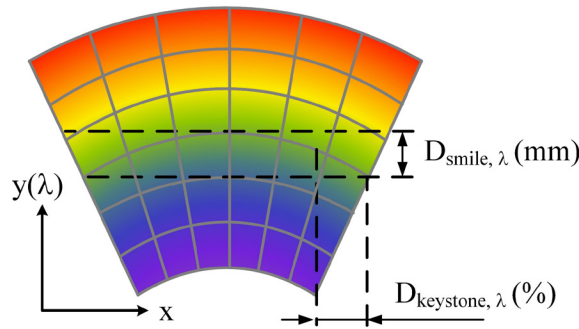


Figure J.2. Distorted image with spectral and spatial distortion.

Concerning the general performance of an imaging spectrometer, resolution and efficiency are both of great importance. Given a specific entrance slit length and sensor pixel size, the entire area of each image line should ideally be captured on the sensor array. However, since the spot size of a single field point determines the width of the image line, poor correction can lead to overlapping adjacent image lines, which cannot be fully captured or spectrally resolved by the sensor. The spot size also determines how critical distortion is, as a broader image line means more energy loss with the same bending. Furthermore, both smile and keystone distortion cause a loss of captured signal, leading to decreased resolution and efficiency. Consequently, distortion common to imaging spectrometers can complicate camera calibration, reduce efficiency and resolution performance, as well as impede data processing, making its correction of great concern.

In imaging spectrometers, the dispersive elements usually break the rotational symmetry of the whole system, making it difficult to correct higher-order aberrations with only rotationally symmetric surfaces. Therefore, freeforms can be introduced in such systems to specifically correct the aberrations caused by asymmetry. The benefit of freeforms in grating spectrometer systems has been already discussed in the literature [27], but a

systematic investigation of prism-based systems is still lacking to the best knowledge of the authors. Therefore, we investigate the use of freeforms to improve imaging performance in common prism spectrometer systems. It is also important to mention that our investigation is not intended to deal with any particular application but to exhaust the potential of freeforms in a case study to illustrate the improvement of the system performance with the higher-order correction of freeforms in general prism imaging spectrometers. Thus, the results of this study can be projected to any practical application with such spectrometers as instructive guidance for distortion correction.

J.2 Example: Modified Offner system optimization with freeforms

In order to draw a generalized conclusion from this study, several typical refractive spectrometers are chosen for the case study. As an example for the dissertation, the results of a refractive spectrometer of modified Offner structure [63] are illustrated here.

Distinguished from the typical Offner spectrometer with reflective gratings, the modified Offner structure spectrometer here is composed of 3 mirrors and 2 prisms. The original image space F-number is kept as 3 with an original entrance slit of 30 mm. The spectral range is 400-900 nm and the dispersion distance is 3 mm. The stop is located at M2, as illustrated in Figure J.3. It is interesting to mention that, according to the symmetry principle, an original Offner system without prisms generally enjoys the benefits of small field-related aberrations including distortion due to its symmetric structure about the stop (M2). However, as two prisms are added to the system, the symmetry is broken due to the separation of dispersive angles, which reduces the automatic distortion correction brought by the symmetry principle. Besides, the large final image distance, compared to the total length along the z-axis, further magnifies the distortion at the image plane. Thus, distortion becomes a critical problem in this system despite the partial fulfillment of the symmetry principle.

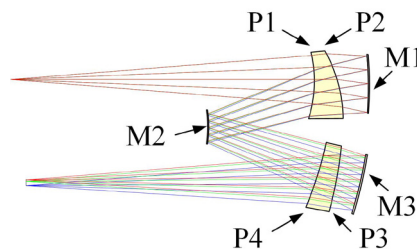


Figure J.3. Modified Offner spectrometer system layout, where S1, S2, S3, S4, and M2 are potential freeform surface locations.

First, a systematic optimization procedure should be established. Regarding the general rules mentioned above, all the prism and mirror surfaces are possible freeform locations. The investigation should include all possible combinations of freeform surfaces, so that the cases in which freeforms are added at various locations in different numbers can be compared.

For the purpose of this study, a systematic optimization rule should be fixed. Concerning the optimization, the distortion correction should be based on the essential criterion that the original spot size is not degraded. Within the original spectral range, the system uses five sampling wavelengths for optimization and RMS distortion value calculation.

During the actual spectrometer system design, the spot size should be optimized to match the detector pixel size. As the pixel size of the spectrometer system detector are various according to the specific application, and the ideal system image size is also determined by focal length, magnification, spectral range, and the scale of the system, the system performance cannot be simply analyzed with only pixel size. Therefore, according to the purpose of the study, only the improvement of the system performance by the introduction of freeforms is of concern, the spot size and distortion will not be compared to the pixel size. Therefore, with the precondition of maintaining the original spot size, the optimization prioritizes minimizing the distortion. In addition, the optimized freeform surfaces are limited to a peak-valley value of 6 mm based on the current state of manufacturing technology.

Since there are three mirrors and four prism surfaces in the system, the possible freeform combinations can be categorized into many different cases. As we focus on comparing the imaging performance without and with freeforms, so not all cases need to be presented. Among all the mirrors, only M2 is considered as a possible freeform location because it is the system stop, which better illustrates the impact of the freeform with overlapped ray bundles. Meanwhile, all four prism surfaces will be considered, since each can contribute quite differently to aberration correction, and any relative advantage is difficult to predict. Therefore, the investigated categories are: single prism surface (P); two prism surfaces ($2\times P$); three prism surfaces ($3\times P$); all four prism surfaces ($4\times P$); the mirror and single prism surface (M+P); the mirror and two prism surfaces (M+ $2\times P$); the mirror and three prism surfaces (M+ $3\times P$), and the mirror and all prism surfaces (M+ $4\times P$). After optimization, the performances of the various freeform groups and locations are compared using the RMS spot sizes at the edge and central wavelengths, as well as the RMS values of smile and

keystone distortion. The distortion values are calculated using all five sampling wavelengths, and ignoring spot size.

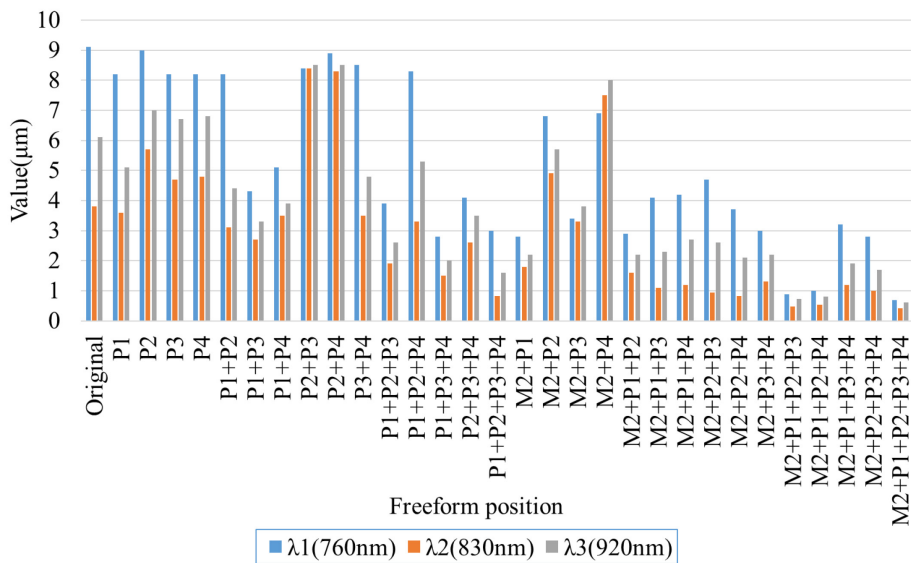


Figure J.4. System RMS spot size comparison with different freeform positions after optimization.

Using only one freeform surface, position P1 provides the best distortion correction and also offers a slight spot size reduction, as shown in Figure J.4 and Figure J.5. Considering only spot size, among the categories using two freeforms ($2\times P$ and $M+P$), those that include P1 typically show a smaller spot size than the others. Similarly, as shown in Figure 5.5, for the categories P, $2\times P$, $3\times P$, and $M+P$, the freeform groups including P1 all show a significantly better distortion correction in their respective categories. Being the first refractive surface of the system, its pre-correction of higher aberrations is important for preventing aberration magnification through the system. From the performance comparison of the $2\times P$ cases, P1+P3 and P1+P4 perform the best, indicating again that proximity to the image enables good prevention of induced aberrations.

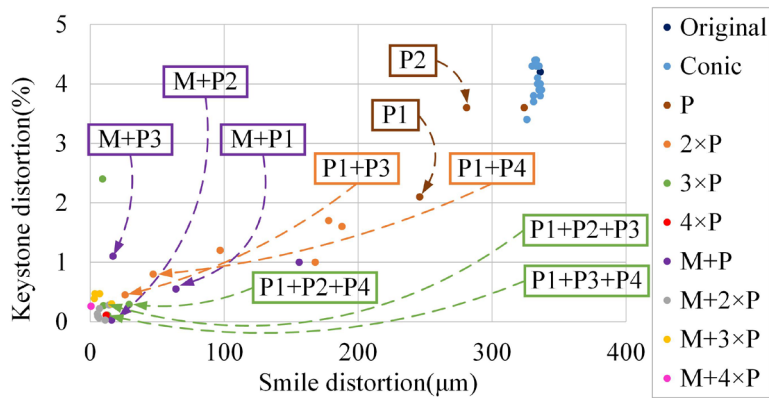


Figure J.5. Distortion illustration of the modified Offner spectrometer.

For the $4\times P$, $M+2\times P$, $M+3\times P$, and $M+4\times P$ categories, all combinations show similar, very good distortion correction. Therefore, we conclude that using M2 and any two prism surfaces can guarantee great performance in typical Offner systems. In this particular setup, the combination of M2 and P1 alone can achieve this top-level correction performance after optimization. The corresponding surface sag and gradient are shown in Figure J.6, where only a small deviation from the basic shape can be observed.

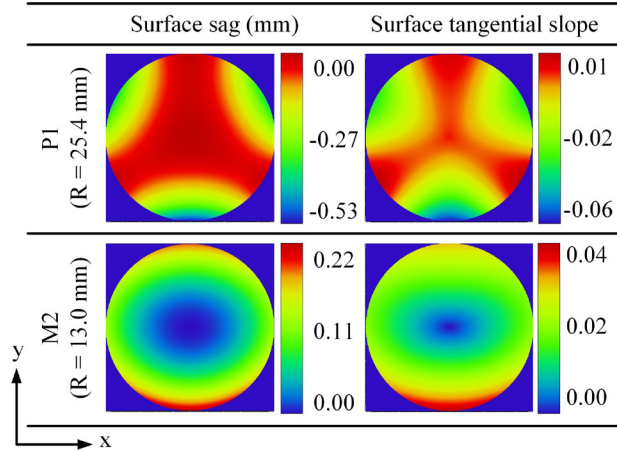


Figure J.6. Freeform surface illustrations of the modified Offner spectrometer.

Furthermore, the large number of freeform group combinations in this system offers an interesting opportunity to examine the effect of available degrees of freedom on the aberration correction. Every freeform surface introduces 37 degrees of freedom, coming from the 36 Zernike fringe sag terms plus the corresponding normalized radius, so we can simply use the number of freeforms (1 to 5) as an independent variable. For a better illustration of the general improvement with a certain number of degrees of freedom, an averaged improvement factor is defined as

$$F = \frac{D_{original}}{\sum_1^{\tau} D_{freeform} / \tau}, \quad (J. 3)$$

where D represents the RMS value of any of the three evaluation criteria—keystone distortion, smile distortion, and spot size—calculated over the three wavelengths., which are keystone distortion, smile distortion, or spot size. $D_{original}$ means the original value and $D_{freeform}$ is the value after optimization with freeforms. Here n indicates the total number of cases in each category. Therefore, the improvement factor F calculates the change of the average RMS values for each of the three criteria. The results are shown in Figure J.7, where the blue and orange lines represent the logarithmic improvement factor for the two kinds of distortion, and the gray line represents that of the spot size. It can be

clearly observed that all three lines grow exponentially as more freeforms are added. Distortion correction improves greatly up to three freeforms, but adding a 4th does not offer significant improvement. Adding a 5th considerably improves mainly smile distortion. In contrast, the spot size shows constant, significant improvement with the addition of further freeforms. In general, we observe a roughly exponential improvement in corrective power with increasing degrees of freedom [28].

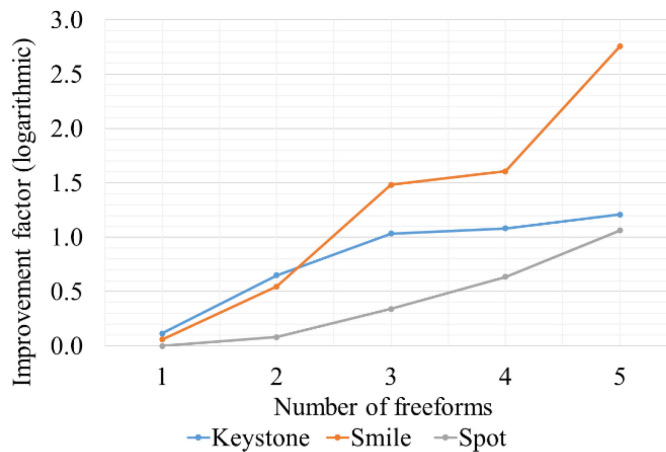


Figure J.7. Improvement factor against the number of freeforms in logarithmic scale.

With the aim to improve the smile and keystone distortion correction and potentially the resolution of line imaging refractive spectrometers, we discuss the correction of a refractive imaging spectrometer with the modified Offner structure by introducing freeforms at various locations. In general, as more freeforms are introduced, the improvement of the system performance can be greatly enhanced due to the additional degrees of freedom. For the folded structure of the system with more critical higher-order aberrations, the freeform should be located in the front part of the system to prevent any induced aberrations as early as possible. Based on the limits of manufacturing and cost, the number and optimal locations of the freeforms should be taken into consideration to reach the best balance with the imaging performance. Through this study, the corrective power of freeform surfaces is better understood, and the critical impact of the induced aberration is revealed, which emphasizes the necessity of the analysis of intrinsic/induced aberrations.

References

1. W.J. Smith, E. Betensky, D. Williamson, J.C. Miñano, R. J. Koschel, "The past, present, and future of optical design," Proc. SPIE 6342, 63422Y (2006).
2. R.E. Fischer, B. Tadic-Galeb, P.R. Yoder, R. Galeb, B.C. Kress, S.C. McClain, T. Baur, R. Plympton, B. Wiederhold and Bob Grant Alastair J, *Optical system design* (New York: McGraw Hill, 2000).
3. R. Kingslake, and R. B. Johnson, *Lens Design Fundamentals* (SPIE, 2010).
4. Y. Zhong and H. Gross, "Initial system design method for non-rotationally symmetric systems based on Gaussian brackets and Nodal Aberration Theory," Opt. Express 25, 10016-10030 (2017).
5. W. T. Welford, *Aberrations of optical systems* (CRC Press, 1986).
6. J. Sasian, *Introduction to Aberrations in Optical Imaging Systems* (Cambridge University, 2013).
7. K. P. Thompson, "Aberrations fields in tilted and decentered optical systems," Ph.D. dissertation (University of Arizona, Tucson, Arizona, 1980).
8. E. M. Schiesser, A. Bauer, and J. P. Rolland, "Estimating field-dependent nodal aberration theory coefficients from Zernike full-field displays by utilizing eighth-order astigmatism," J. Opt. Soc. Am. A 36, 2115-2128 (2019).
9. K. Fuerschbach, J. P. Rolland, and K. P. Thompson, "Theory of aberration fields for general optical systems with freeform surfaces," Opt. Express 22, 26585-26606 (2014).
10. T. H. Jamieson, *Optimization techniques in lens design*. (London: A. Hilger, 1971).
11. A. Yabe, *Optimization in Lens Design*. (SPIE Press, 2018).
12. D. Vasiljevic, *Classical and evolutionary algorithms in the optimization of optical systems*, (Springer Science & Business Media, 2012).
13. H. Gross, *Handbook of Optical Systems, Volume 1: Fundamentals of Technical Optics*, (Wiley-VCH, 2005).
14. S. Wang and D. Zhao, *Matrix optics* (CHEP and Springer, 2000).
15. M. Nazarathy, A. Hardy and J. Shamir, "Misaligned first-order optics: canonical operator theory," JOSA A, 3, 1360-1369 (1986).

16. M. Harrigan, "General Beam Propagation through non-orthogonal Optical Systems," International Optical Design Conference, OSA Technical Digest Series, IWA4 (2002).
17. Z. Tang, and H. Gross, " Extended aberration analysis in symmetry-free optical systems – part I: Method of calculation," *Opt. Express*, 29, 39967-39982 (2021).
18. S. H. Brewer, "Aberration Analysis Via Exact Surface Contributions," Proc. SPIE 0294, New Methods for Optical, Quasi-Optical, Acoustic, and Electromagnetic Synthesis (1982).
19. B. Chen and A. M. Herkommer, "High order surface aberration contributions from phase space analysis of differential rays, " *Opt. Express* 24, 5934-5945 (2016).
20. B. Chen and A. M. Herkommer, "Generalized Aldis theorem for calculating aberration contributions in freeform systems," *Opt. Express*, 24, 26999-27008 (2016).
21. M. Testorf, B. Hennelly and J. Ojeda-Castaneda, *Phase-space optics* (McGraw-Hill Companies, 2010).
22. J. D. Mansell, R. Suizu, R. Praus, B. Strickler, A. Seward, and S. Coy, "Integrating Wave-Optics and 5x5 Ray Matrices for More Accurate Optical System Modeling," DEPS Fourth Directed Energy Modeling & Simulation Conference (2006).
23. P. D. Colbourne, "Generally Astigmatic Gaussian Beam Representation and Optimization Using Skew Rays," Proc. SPIE 9293, International Optical Design Conference (2014).
24. H. Gross, *Handbook of Optical Systems, Volume 3: Aberration Theory and Correction of Optical Systems*, (Wiley-VCH, 2005).
25. Z. Tang, M. Sonntag, and H. Gross, "Ant colony optimization in lens design," *Appl. Opt.* 58, 6357-6364 (2019).
26. G. W. Forbes, "Shape specification for axially symmetric optical surfaces," *Opt. Express*, 15, 5218-5226, (2007).
27. C. Liu, C. Straif, T. Flügel-Paul, U. D. Zeitner, and H. Gross, "Comparison of hyperspectral imaging spectrometer designs and the improvement of system performance with freeform surfaces," *Appl. Opt.* 56, 6894-6901 (2017).
28. Z. Tang, and H. Gross, "Improved correction by freeform surfaces in prism spectrometer concepts," *Appl. Opt.* 60, 333-341 (2021).

29. C. Menke, "Application of particle swarm optimization to the automatic design of optical systems." *Optical Design and Engineering VII*, Vol. 10690 (International Society for Optics and Photonics, 2018).
30. D. Sedighizadeh, E. Masehiam, M. Sedighizadeh, and H. Akbaripour, "A new generalized particle swarm optimization algorithm." *Mathematics and Computers in Simulation*, 179, 194-212 (2021).
31. C. Blum, "Ant Colony Optimization: Introduction and Recent Trends," *Physics of Life reviews*, 2(4), 353-373 (2005).
32. K. Socha, and M. Dorigo, "Ant Colony Optimization for Continuous Domains," *European Journal of Operational Research*, 185, 1155-1173 (2008).
33. K. Araki, "Paraxial and aberration analysis of off-axial optical systems." *Optical review* 12.3, 219-222 (2005).
34. K. Araki, "Analysis of off-axial optical systems (1)", *Optical Review* 7, 221-229 (2000).
35. K. Araki, "Analysis of off-axial optical systems (2)", *Optical Review* 7, 326-336 (2000).
36. K. P. Thompson, "Description of the third-order optical aberrations of near-circular pupil optical systems without symmetry," *J. Opt. Soc. Am. A* 22, 1389-1401 (2005).
37. K. P. Thompson, "Multinodal fifth-order optical aberrations of optical systems without rotational symmetry: spherical aberration," *J. Opt. Soc. Am. A* 26, 1090-1100 (2009).
38. K. P. Thompson, "Multinodal fifth-order optical aberrations of optical systems without rotational symmetry: the comatic aberrations," *J. Opt. Soc. Am. A* 27, 1490-1504 (2010).
39. K. P. Thompson, "Multinodal fifth-order optical aberrations of optical systems without rotational symmetry: the astigmatic aberrations," *J. Opt. Soc. Am. A* 28, 821-836 (2011).
40. T. Yang, D. Cheng, and Y. Wang, "Aberration analysis for freeform surface terms overlay on general decentered and tilted optical surfaces," *Optics Express* 26, 7751-7770 (2018).

41. E. M. Schiesser, A. Bauer, and J. P. Rolland, "Estimating field-dependent nodal aberration theory coefficients from Zernike full-field displays by utilizing eighth-order astigmatism," *J. Opt. Soc. Am. A* 36, 2115-2128 (2019).
42. A. García-Moreno, R. Restrepo, T. Belenguer-Dávila, and L. M. González-Fernández, "Field aberrations in terms of the Q-polynomial basis and its relationship to the Zernike basis," *OSA Continuum* 4, 542-555 (2021).
43. A. Bauer, J. P. Rolland, and K. P. Thompson, "Ray-based optical design tool for freeform optics: coma full-field display," *Opt. Express* 24, 459-472 (2016).
44. M. Oleszko, R. Hambach and H. Gross, "Decomposition of the total wave aberration in generalized optical systems," *J. Opt. Soc. Am. A* 34, 1856-1864 (2017).
45. Y. Zhong, Z. Tang, and H. Gross, "Correction of 2D-telecentric scan systems with freeform surfaces," *Opt. Express* 28, 3041-3056 (2020).
46. Z. Tang, and H. Gross, "Extended aberration analysis in symmetry-free optical systems – part II: evaluation and application," *Opt. Express*, 29, 42020-42036 (2021).
47. Z. Tang, and H. Gross, "Higher-order aberration analysis in symmetry-free optical systems," *Proc. SPIE* 11871 (2021).
48. Z. Pang, X. Fan, Z. Ma, and G. Zou, "Misalignment induced aberration off-axis optical system," *Proc. SPIE*, 8th International Symposium on Advanced Optical Manufacturing and Testing Technologies: Large Mirrors and Telescopes, 96820U (2016).
49. B. G. Crowther, J. Sasián, and J. M. Hoffman, "Enhancements and applications of induced aberration theory," *Proc. SPIE*, Roland V. Shack Memorial Session: A Celebration of One of the Great Teachers of Optical Aberration Theory, 114790F (2020).
50. Yan Liu, Yanqiu Li, and Zhen Cao, "Design of anamorphic magnification high-numerical aperture objective for extreme ultraviolet lithography by curvatures combination method," *Appl. Opt.* 55, 4917-4923 (2016).
51. Y. Bai, T. Xing, and Y. Jiang, "Applying Q-type aspheres in the ultraviolet lithography objective lens," *Proc. SPIE*, Novel Optical Systems Design and Optimization XIX, 994804 (2016).

-
52. L. Hsiao and H. Lin, "Extreme ultra violet lithographic optical projection system design method using Code V lens module and generalized Gaussian constants," Proc. SPIE 10450, International Conference on Extreme Ultraviolet Lithography (2017).
 53. H. Mann, "Imaging optical system," U.S. patent, US010007187B2 (2018).
 54. H. Gross, *Handbook of Optical Systems, Volume 4: Survey of Optical Components and Systems*, (Wiley-VCH, 2005).
 55. Zemax OpticStudio, *User Manual*, (2020).
 56. M. T. Eismann, *Hyperspectral Remote Sensing* (SPIE, 2012).
 57. L. Yuan, J. Xie, Z. He, Y. Wang, and J. Wang, "Optical design and evaluation of airborne prism-grating imaging spectrometer," Opt. Express 27, 17686-17700 (2019).
 58. P. Mouroulis, and M. M. McKerns, "Pushbroom imaging spectrometer with high spectroscopic data fidelity: experimental demonstration," Opt. Eng. 39, 808-816 (1999).
 59. J. F. Silny, and T. G. Chrien, "Large format imaging spectrometers for future hyperspectral Landsat mission," Proc. SPIE, Imaging Spectrometry XVI, 815803 (2011).
 60. T. Skauli, "Quantifying coregistration errors in spectral imaging," Proc. SPIE, Imaging Spectrometry XVI, 81580A (2011).
 61. J. Hong, Y. Kim, B. Choi, S. Hwang, D. Jeong, J. Lee, Y. Kim, and H. Kim, "Efficient method to measure the spectral distortions using periodically distributed slit in hyperspectral imager," Opt. Express 25, 20340-20351 (2017).
 62. V. Deppo, E. Simioni, G. Naletto, and G. Cremonese "Distortion definition and correction in off-axis systems," Proc. SPIE, Optical Design and Engineering VI, 962634 (2015).
 63. L. Feng, J. Zhou, L. Wei, X. He, Y. Li, J. Jing, and B. Xiangli, "Design of a compact wide-spectrum double-channel prism imaging spectrometer with freeform surface," Applied Optics 57, 9512-9522 (2018).

List of figures

Figure 1.1. An overview of the optical design tasks.....	5
Figure 2.1. Definition of pupil, chief ray and marginal ray in the optical system.....	8
Figure 2.2. Normalized field and pupil coordinate of an arbitrary ray.....	9
Figure 2.3. Longitudinal, transverse, and wave aberrations.....	12
Figure 2.4. Chief ray and marginal ray from an off-axis field.....	12
Figure 2.5. Plots of the Zernike terms according to the radial and azimuthal order.....	14
Figure 2.6. Intrinsic and induced aberration.....	15
Figure 2.7. Propagation of the OAR in an arbitrary refractive system.....	17
Figure 2.8. Propagation of the RR and a paraxial ray in an off-axis system, where S_j is a reflective surface and S_{j+1} is a refractive surface.....	18
Figure 2.9. General optical design process.....	20
Figure 2.10. Particle movement principle of PSO.....	26
Figure 2.11. Concept of ACO. (a) the ants start to explore various paths; (b) the following ants choose a path to follow according to the pheromone trail; (c) ants all converge to the best way to the goal.....	27
Figure 2.12. The structure of solution archive of ACOR, and the algorithm outline.....	29
Figure 3.1. Surface contribution of transverse aberration projected in the tangential plane and transformation to the image plane.....	34
Figure 3.2. Geometrical illustration of Δy_{img} and ΔY_{img}	37
Figure 3.3. Mixed ray-tracing process for calculating the surface-decomposed total, intrinsic, and induced transverse aberration of S_j concerning only the y-component.....	38
Figure 3.4. Approximation in the intrinsic/induced aberration calculation concerning S_j	39
Figure 4.1 General workflow of the GACOR algorithm.....	48
Figure 4.2 Evaluation criteria of the output solution from the local exploration.....	52
Figure 4.3. General algorithm workflow of the local exploration.....	54
Figure 4.4. Workflow of the DLS booster.....	55

Figure 4.5. Workflow of the final improvement process.....	58
Figure 5.1. The layout of the 6-M lithography system and the object field.....	62
Figure 5.2. Kingslake plots of intrinsic, induced, and total aberration with decomposed surface contributions of the lithography system concerning the center field.....	63
Figure 5.3. Kingslake plots of intrinsic, induced, and total aberration with decomposed surface contributions of the lithography system concerning the corner field.....	63
Figure 5.4. Surface-additive Zernike coefficients until Z25 for the lithography system as well as the sum value (unit: waves)	64
Figure 5.5. Layout of the retro-focus system.....	65
Figure 5.6. The layout and the parameters of the initial system.....	66
Figure 5.7. Workflow of the DLS local optimization (left) and the setting rules of the MF (right).....	68
Figure 5.8. Intermediate solutions during the first main iteration. The current targets and values in the MF of the spot size (unit: μm) and NA are also given.....	71
Figure 5.9. First partial solution evolution map (left), and the comparison of surface-decomposed transverse aberrations (unit: mm) calculated with the MRT method between solution B1 and C1 (right).....	73
Figure 5.10. Second partial evolution map of solutions. The aspherical surfaces are marked pink and the successful solutions are marked in red.....	74
Figure 5.11. Critical index (u.a.) of B1 for splitting option (left) and asphere option (right).....	75
Figure 5.12. Comparison of the surface-decomposed transverse aberrations calculated with the MRT method among solutions B4.4, B4.5, and B4.6 (unit: mm).....	76
Figure 5.13. Third partial solution evolution map. The aspherical surfaces are marked pink and the successful solutions are marked in red.....	77
Figure 5.14. Critical index (u.a.) of C1 for splitting option.....	77
Figure 5.15. Comparison of the surface-decomposed transverse aberration calculated with the MRT method among the solutions C4.1, C4.2, and C4.3 (unit: mm).....	78

Figure 5.16. Fourth partial solution evolution map (left). Comparison of lens bending parameters of k_1 and k_2 (upper-right). Surface sag plot of S5 in k_2 in mm (lower-right)...79

Figure 5.17. Evolution of a solution where L1 has nearly parallel surfaces. The aberration analysis (unit: mm) and the lens bending analysis of solution **g1.5** are plotted below.....80

Figure 5.18. Collection of successful solutions with five spherical lenses.....81

Figure 5.19. Comparison of the surface-decomposed transverse aberration calculated with the MRT method of the ‘S5’ solution category (unit: mm).....81

Figure 5.20. Collection of successful solutions with six spherical lenses.....82

Figure 5.21. Comparison of the surface-decomposed transverse aberration calculated with the MRT method of the ‘S6’ solution category (unit: mm).....82

Figure 5.22. Collection of successful solutions with seven spherical lenses.....83

Figure 5.23. Comparison of the surface-decomposed transverse aberration calculated with the MRT method of the ‘S6’ solution category (unit: mm).....83

Figure 5.24. Collection of successful solutions with three spherical lenses and one aspherical lens.....84

Figure 5.25. Comparison of the surface-decomposed transverse aberration calculated with the MRT method of the ‘S3A1’ solution category (unit: mm).....84

Figure 5.26. Collection of successful solutions with four spherical lenses and one aspherical lens.85

Figure 5.27. Comparison of the surface-decomposed transverse aberration calculated with the MRT method of the ‘S4A1’ solution category (part 1) (unit: mm).....86

Figure 5.28. Comparison of the surface-decomposed transverse aberration calculated with the MRT of the ‘S4A1’ solution category (part 2) (unit: mm).....87

Figure 5.29. Comparison of the surface-decomposed transverse aberration calculated with the MRT of the ‘S4A1’ solution category (part 3) (unit: mm).....87

Figure 5.30. Aberration standard deviation of all the solutions (unit: mm). Each solution category is marked with a different color, and the bubble size represents the relative spot size of the solutions.88

Figure 5.31. Layout of the anamorphic system in both cross-sections and the sampled FoV.....	91
Figure 5.32. Spot diagrams of all the selected fields.....	91
Figure 5.33. Collection of the output solutions of the anamorphic system. The surface sag is in the unit of mm, and the spot diagrams are scaled in 300 μ m.....	94
Figure 5.34. Aberration standard deviation of the transverse aberration surface contribution in x- and y-direction of all the final improvement solutions (unit: mm).....	96
Figure 5.35. Comparison between solution 2 and 10 concerning the surface-additive Zernike coefficients in the case of Z5, Z6, Z8, and Z9 (unit: waves). ‘S2’ and ‘S10’ refer to the solution 2 and 10.....	96
Figure A.1. Layout of the triplet system proposed by Brewer, and the spot diagram of the outermost field in the y-direction (scale: 100 μ m).....	100
Figure A.2. Calculation results of Δx_j and Δy_j (in μ m) by the MRT method and Aldis formulas, concerning different rays in the system.....	101
Figure A.3. Layout of the test symmetric system.....	101
Figure A.4. Comparison of Zernike coefficient fitting results between Zemax and the MRT method calculation results. The unit is in waves.....	104
Figure A.5. Spherical aberration (Z9), coma (Z8), and astigmatism (Z5) calculated with different methods.....	105
Figure B.1. Surface sag in the tangential plane and the tangent plane T.....	106
Figure B.2. Possible pitfalls of the approximation: the surfaces suffering from considerable induced effect (left), or freeform surfaces with extremely large deviations from the basic spherical shape (right).....	108
Figure B.3. Different choices of RR and the corresponding parabolal zone.....	109
Figure D.1. Splitting process of the lens.....	117
Figure E.1 Categories of solutions in GAII and ASB.....	120
Figure I.1. Layout of a tele-system.....	129
Figure I.2. MF setting rules for the tele-system optimization.....	130
Figure I.3. Output solutions of the tele-system optimization.....	131

Figure I.4. Aberration standard deviation of all the solutions (unit: mm).....131

Figure I.5. MF setting rules for the high NA collimator system optimization.....132

Figure I.6. Output solutions of the high NA system optimization.....133

Figure I.7. Aberration standard deviation of all the solutions (unit: mm).....133

Figure J.1. Basic imaging spectrometer layout.....134

Figure J.2. Distorted image with spectral and spatial distortion.....135

Figure J.3. Modified Offner spectrometer system layout, where S1, S2, S3, S4, and M2 are potential freeform surface locations.....136

Figure J.4. System RMS spot size comparison with different freeform positions after optimization.....136

Figure J.5. Distortion illustration of the modified Offner spectrometer.....138

Figure J.6. Freeform surface illustrations of the modified Offner spectrometer.....139

Figure J.7. Improvement factor against the number of freeforms in logarithmic scale...140

List of tables

Table 2.1. Seidel coefficients of primary aberrations.....	13
Table 2.2. Surface shape against conic constant.....	24
Table 4.1. Overview of the optimization method research in this work.....	44
Table 4.2. An example of the final improvement strategy.....	60
Table 5.1. System specifications of the retro-focus system.....	66
Table 5.2 Specification of the anamorphic system.....	91
Table 5.3. Categorization of lens parameters.....	92
Table A.1. Lens data of the symmetric system.....	102
Table A.2. Transverse aberration calculation (unit: mm).....	102
Table B.1. Possible situations with RR and tested fields.....	110
Table D.1. Parameterization of α_j and the involved ones in different cases.....	116
Table D.2 Addition of the asphere coefficients.....	118
Table F.1. MF and variable adjustment rules in different situations.	122
Table F.2 Target value setting rules of thickness constraints.....	123
Table H.1. Calculation of σ_0 dependent on the existence of asphere and the variable category.....	127
Table I.1. System specifications of the tele-system.....	129
Table I.2. System specifications of the NA collimator system.....	132

List of symbols

S_j	The j^{th} surface
R	Surface radius of curvature
n_l	Image-space refractive index
n, n'	Refractive index before and after refraction
L_j	Lagrange invariant of S_j
r	Radial aperture height
c	Radius of curvature of the surface
t	Thickness after the surface
κ	Surface conic constant
ρ	Normalized radial aperture height
θ	Azimuthal angle of the pupil coordinate
D_{Sj}	Surface diameter of S_j
s'	Free working distance
f'	Focal length
D_{Airy}	Airy diameter
X	Bending parameter
i_0, i_0'	Incidence angle of the incoming and outgoing ray
u_0, u_0'	Object and image space ray angle referred to the optical axis
y_0, y_0'	Object and image size
x_p, y_p	Pupil coordinate
H_x, H_y	Normalized field coordinates in x- and y-direction
P_x, P_y	Normalized pupil coordinates in x- and y-direction
h_j, \bar{h}_j	Object side ray height of MR/CR
u_j, \bar{u}_j	Object side ray angle of MR/CR

u'_j, h'_j	Angle and height of an arbitrary paraxial ray at S_j
$S_I \sim S_V$	Seidel coefficients of monochromatic aberrations
C_I, C_{II}	Seidel coefficients of chromatic aberrations
(X_j, Y_j, Z_j)	Coordinate of the real ray intersection point on S_j
(L_j, M_j, N_j)	Optical direction cosine of the real ray intersection point on S_j
H_I	Gaussian image height
n_j	Refractive index after S_j
R_{ref}	Radius of the reference sphere
W_j	Wave aberration after S_j
a_i	i^{th} asphere polynomial (even asphere, Q-type asphere)
γ_i	i^{th} Coefficient of the Zernike polynomial
Z_i	i^{th} Zernike polynomial
Q	Total number of Zernike terms
z_s	Surface sag
ΔZ_{FF}	Surface sag freeform deviation
N	Total number of optical surfaces
N_L	Total number of lenses in the system
N_{Ls}	Total number of spherical lenses in the system
N_{La}	Total number of aspherical lenses in the system
N_{L0}	Current largest lens number among the solutions ever found
N_{LII}	Nominal lens number of RII
N_z	Total number of aspherical terms involved in the system
Δ_j	Total transverse aberration contribution of S_j
Δ_{Img}	Total transverse aberration contribution in the image plane
$\Delta x_j, \Delta y_j$	Transverse aberration contribution of S_j in sagittal/tangential plane

List of symbols

$\Delta x_{Img}, \Delta y_{Img}$	Total transverse aberration in the sagittal, tangential plane
$\Delta y_{Dist,j}$	The distortion part of Δy_j
$\Delta y_{Dist,Img}$	The distortion part of Δy_{Img}
$\Delta y_{0,j}$	The remaining aberration of Δy_j with distortion subtracted
$\Delta y_{0,Img}$	The remaining aberration of Δy_{Img} with distortion subtracted
ΔS_{Img}	Longitudinal aberration referring to the image plane position
ΔX_j	CR referred transverse aberration contribution of S_j in the sagittal plane
ΔY_j	CR referred transverse aberration contribution of S_j in the tangential plane
ΔY_{j0}	Single surface contribution of ΔY_j
ΔY_{Img}	Chief ray-referred total transverse aberration in the tangential image plane
$\Delta y_{int,j}$	Transverse intrinsic aberration contribution of S_j in the tangential plane
$\Delta y_{ind,j}$	Transverse induced aberration contribution of S_j in the tangential plane
$\mathbf{r}_j, \mathbf{p}_j$	Real /paraxial ray vector on S_j
x, x'	Global ray intersection coordinate x-component with the dummy surface
y, y'	Global ray intersection coordinate y-component with the dummy surface
u, u'	Projected direction cosine of the ray in the sagittal plane
v, v'	Projected direction cosine of the ray in the tangential plane
A, B, C, D	(Generalized) components of parabal/paraxial matrix
E, F	Tilt/decenter components of a parabal matrix
D_j, D'_j	Front/rear dummy surface of the incoming and outgoing RR on S_j
D_{Img}	Dummy surface of the incoming RR in the image plane
$M_{P,j}, M_{T,j}$	Reflection/refraction and propagation parabal matrix
R_j, R'_j	Intersection point of the real ray on the front/rear dummy surfaces
P_j, P'_j	Intersection point of the paraxial ray on the front/rear dummy surfaces
P_{Img}, R_{Img}	Intersection point of $\mathbf{p}_j/\mathbf{r}_j$ on the front dummy surfaces of the image plane

T	Tangent plane to the intersection point C_j
C_j	Intersection point of \mathbf{r}_j and S_j
C_{pj}	Intersection point of \mathbf{p}_j and S_j
C_{pj0}	Intersection point of the extension of vector $M_{Tj} \cdot \mathbf{p}_j$ on the tangent plane
C'_{pj}	Intersection point of the parallel line of the local optical axis of S_j through
R_L	Local radius of curvature
M	Sampling ray number
(y_r, z_r)	Coordinate of C_j in the tangential plane
(z_p, y_p)	Coordinate of C_{pj} in the tangential plane
(z_{pe}, y_0)	Coordinate of C'_{pj} in the tangential plane
(z_0, y_0)	Coordinate of C_{pj0} in the tangential plane
k_T	Slope of the projected line of T
k_P	Slope of the projected line of \mathbf{p}'_j
E_y, E_z	Surface sag error in the y- and z- direction
m	Dimension of the optimization problem
\vec{x}	An m-dimensional vector representing all the variables during optimization
$F(\vec{x})$	Merit function value
$f_{tar,i}$	i^{th} target value in the merit function
$f_i(\vec{x})$	i^{th} operand function value of the merit function
w_i	Weighting for the i^{th} target of the merit function
J	Jacobi matrix
λ	Damping factor of DLS algorithm
I_{kk}	Diagonal component of diagonal unity matrix
Δd_j	Solution of damped step in DLS iteration
K	Capacity of the solution archive of basic ACOR algorithm

P	Number of ants of basic ACOR algorithm
ω	Weight of the ranked solution of basic ACOR algorithm
q	Parameter for the bias towards the best-ranked solutions
G^i	Probability density function
μ	Mean value of the variables among archive solutions
σ	Standard deviation of the variables among archive solutions
σ_0	Initial standard deviation of the variables among archive solutions
T_s	Distance from the first to the last optical surface
δ	Standard deviation between the two solutions for similarity check
ζ	Parameter for convergence speed
K_g	Capacity of the GA
K_{g0}	Initial capacity of the GA
K_{gm}	Maximum allowed capacity of the GA
K'_{gi}	Capacity of the GA when merging with new GA solutions
P_g	Number of ant groups
i_g	Iteration index
$\bar{\alpha}_{Sj}$	Seidel coefficients for critical index calculation
$\bar{\alpha}_{Aj}$	Aldis contributions for critical index calculation
$\bar{\alpha}_{Hj}$	Marginal ray height for critical index calculation
$\bar{\alpha}_{Ij}$	Marginal ray incidence angle for critical index calculation
A_j	Critical index
β, ε	Exponential scaling factors for critical index calculation
K_{max}	Desired number of output optimization solutions
M_{g0}	Total number of the lens parameter categories
M_g	Number of the lens parameter categories not yet turned into variables

ΔM_g	Number of lens parameter categories not yet newly turned into variables
R_F	Retro-focus factor
W_a, T_a, V_a	Weighting, target, and the current value of the glass thickness operands
W_a, T_a, V_a	Weighting, target, and the current value of the air thickness operands
W_{NA}, T_{NA}, V_{NA}	Weighting, target, and the current value of the NA operands
W_{RF}, T_{RF}, V_{RF}	Weighting, target, and the current value of the R_F operands
W_{LT}, T_{LT}, V_{LT}	Weighting, target, and the current value of the system total length operands
W_{LI}, T_{LI}, V_{LI}	Weighting, target, and the current value of the image distance operands
V_{RMS}	Current value of the RMS spot size
V_m	Current MF value
$\langle N_L \rangle$	equivalent lens number
$\langle N_L \rangle_{max}$	maximum allowed equivalent lens number
F_{m1}	maximum allowed failure number of ACOR local search in the beginning
F_{m2}	maximum allowed total failure number of ACOR local search after starting
$\Delta y_\sigma, \Delta x_\sigma$	Standard deviation of the aberration distribution in y- and x- direction
$\theta_t, \theta_b, \theta_l, \theta_r$	Real MR angle at the top, bottom, left, and right side of the pupil edge
$D_{smile,\lambda}$	Smile distortion
$y_{max,\lambda}$	Centroid image position of the outermost field
$y_{min,\lambda}$	Centroid image position of the central field
$D_{keystone,\lambda}$	Keystone distortion
$x_{max,\lambda}$	Centroid image position of the outer field point along the x -axis
L	Length of the entrance slit along the x -axis
F	Averaged improvement factor of the spectrometer
D_{ori}	Original value of evaluation criteria
D_{ff}	Value of evaluation criteria after optimization
τ	Total number of cases in each category of freeform location

List of abbreviations

OAR	Optical axis ray
EnP	Entrance pupil
ExP	Exit pupil
NAT	Nodal aberration theory
DLS	Damp least square
MF	Merit function
RMS	Root mean square
FoV	Field of view
NA	Numerical aperture
CR	Chief ray
MR	Marginal ray
RR	Reference ray
MRT	Mixed ray-tracing
PDF	Probability density function
GA	Global archive
ACO	Ant colony optimization
ACOR	ACO in the field of real numbers
GACOR	Group ACOR
PS	Phase space
ASB	Archive solution bank
SSA	Successful solution archive
RI	First, second global exploration round
GAI, GAII	Global archive for RI, RII
SPC	Stepwise performance cost
OPC	Overall performance cost

Acknowledgment

First of all, I would like to express my most sincere gratitude to my supervisor, Prof. Dr. Herbert Gross. In the past 5 years, I was always supported by his generous help and patient coaching in my study and research work, which enabled me to make continuous progress. He opened the door of optical system design for me, and guided me to grow up step by step as a scientific researcher in this field. His kind encouragement kept me always optimistic when facing difficulties, and his faith in me built up my self-confidence, which made me a person I like more. I always appreciate the open-minded conversation with him. He is not only an academic supervisor of mine, but also a venerable elder who enlightened me a lot in my life. It is a great honor knowing him, learning from him, and working with him.

I would also like to thank my parents who are always concerned about me overseas for five years. It is their support and understanding that made me study and live in Germany without worries. Although I haven't seen them in person for almost three years, I always feel surrounded by their love and trust. They are the most precious treasure in my life, and I am proud of being their child.

My appreciation also goes to my boyfriend, Christian, who was a great companion to me in those happy and sad moments. His generous support brought me strength and bravery to face the challenges in my study and life. Thanks to him, I adapted myself better to German Culture. Looking back to the past years, I am grateful to have so many precious memories with him 8000km away from my motherland.

In addition, I would also like to thank all my group members. Especially at the beginning of my research, they were always there offering me instructive suggestions when I encountered problems. They all witnessed my progress, and I greatly enjoyed the time working with them, although it was only a short period.

In the end, I am grateful to all my lovely friends, whom I met both in China and Germany. It is because of them that I did not feel lonely on the cold nights. They always stand by me and listen to me whenever I face troubles in my life. I greatly appreciate the time we spent together. Their inspiring ideas and warm-hearted encouragement make my life more colorful and delightful.

Ehrenwörtliche Erklärung

Ich erkläre hiermit ehrenwörtlich, dass ich die vorliegende Arbeit selbständig, ohne unzulässige Hilfe Dritter und ohne Benutzung anderer als der angegebenen Hilfsmittel und Literatur 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:

Herbert Gross, Betreuer.

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 andere Personen) in Anspruch genommen. Niemand hat von mir unmittelbar oder mittelbar geldwerte Leistungen für Arbeiten erhalten, die im Zusammenhang mit dem Inhalt der vorgelegten Dissertation stehen.

Die Arbeit wurde bisher weder im In- noch im Ausland in gleicher oder ähnlicher Form einer anderen Prüfungsbehörde vorgelegt.

Die geltende Promotionsordnung der Physikalisch-Astronomischen Fakultät ist mir bekannt.

Ich versichere ehrenwörtlich, dass ich nach bestem Wissen die reine Wahrheit gesagt und nichts verschwiegen habe.

Jena, 17. 10. 2022

Ort, Datum

Unterschrift d. Verfassers

Publications

Journal articles

Ziyao Tang, Matthias Sonntag, and Herbert Gross

"Ant colony optimization in lens design"

Appl. Opt. 58, 6357-6364 (2019).

Yi Zhong, Ziyao Tang, and Herbert Gross

"Correction of 2D-telecentric scan systems with freeform surfaces"

Opt. Express 28, 3041-3056 (2020).

Ziyao Tang, and Herbert Gross

"Improved correction by freeform surfaces in prism spectrometer concepts"

Appl. Opt. 60, 333-341 (2021).

Ziyao Tang, and Herbert Gross

"Extended aberration analysis in symmetry-free optical systems – part I: Method of calculation"

Opt. Express, 29(24), 39967-39982 (2021).

Ziyao Tang, and Herbert Gross

"Extended aberration analysis in symmetry-free optical systems – part II: Evaluation and application,"

Opt. Express, 29(25), 42020-42036 (2021).

Conference proceedings

Ziyao Tang, and Herbert Gross

"Higher-order aberration analysis in symmetry-free optical systems,"

Proc. SPIE 11871 (2021).