

# Dynamische Optimierung zur Identifikation von Regulationsstrategien des Stoffwechsels

## DISSERTATION

zur Erlangung des akademischen Grades Doktoringenieur (Dr.-Ing.)

vorgelegt der  
**Fakultät für Informatik und Automatisierung  
der Technische Universität Ilmenau**

von  
**Dipl.-Ing. Martin Bartl**  
geboren am 18. Juli 1980 in Hildburghausen

Gutachter:

1. Herr Univ.-Prof. Dr.-Ing. habil. Pu Li, Technische Universität Ilmenau
2. Herr Jun.-Prof. Dr. Christoph Kaleta, Friedrich Schiller Universität Jena
3. Herr Prof. Dr.-Ing. Michael Pfaff, Ernst-Abbe-Fachhochschule Jena

Tag der Einreichung: 01.07.2013

Tag der wissenschaftlichen Aussprache: 26.03.2014

urn:nbn:de:gbv:ilm1-2014000075

*Ich widme diese Arbeit  
meiner Familie, insbesondere meiner Frau Katrin und unseren wundervollen Kindern.*

# Inhaltsverzeichnis

<b>Abkürzungen und mathematische Symbole</b>	<b>vi</b>
<b>Zusammenfassung</b>	<b>x</b>
<b>Abstract</b>	<b>xi</b>
<b>1 Einleitung und Motivation</b>	<b>1</b>
1.1 Einordnung und Zielstellung der Arbeit . . . . .	2
1.2 Aufbau . . . . .	4
<b>2 Grundlagen des Stoffwechsels und dessen Regulation</b>	<b>6</b>
2.1 Dynamische Modellierung von Stoffwechselnetzwerken . . . . .	6
2.2 Regulation des Stoffwechsels . . . . .	9
2.3 Optimalitätsprinzipien und Stoffwechselregulation . . . . .	12
<b>3 Grundlagen der optimalen Steuerung</b>	<b>17</b>
3.1 Die Problemdarstellung . . . . .	17
3.2 Die Lösungsmethoden der dynamischen Optimierung . . . . .	20
3.3 Die quasi-sequentielle Methode . . . . .	24
3.3.1 Die Diskretisierung . . . . .	24
3.3.2 Die Gradientenberechnung . . . . .	28
<b>4 Erweiterungen der quasi-sequentiellen Methode</b>	<b>29</b>
4.1 Motivation . . . . .	29
4.2 Approximationskontrolle der Zustände . . . . .	30
4.2.1 Erfahrungen aus der simultanen Methode . . . . .	30
4.2.2 Varianten der Fehlerabschätzung . . . . .	32
4.2.3 Approximationskontrolle in der quasi-sequentiellen Methode . . . . .	41
4.2.4 Die Gradientenberechnung . . . . .	45

4.3	Bewegte Elemente . . . . .	51
4.3.1	Das Verfahren in der quasi-sequentiellen Methode . . . . .	51
4.3.2	Die Gradientenberechnung . . . . .	52
4.4	Anwendung . . . . .	59
4.4.1	Dynamische Optimierung eines Fermentierungsprozesses . . . . .	59
4.4.2	Die Lösung . . . . .	61
4.4.3	Der Berechnungsaufwand . . . . .	64
4.5	Die C++ Bibliothek „QuasiOPT“ . . . . .	65
4.5.1	Module . . . . .	66
4.5.2	Konfiguration . . . . .	67
4.5.3	Eingabepuffer und Ausgabepuffer . . . . .	68
4.5.4	Eigenschaften und Dokumentation . . . . .	70
4.6	Diskussion . . . . .	70
<b>5</b>	<b>Optimale Regulation metabolischer Netzwerke in <i>Escherichia coli</i></b>	<b>72</b>
5.1	Motivation . . . . .	72
5.2	Minimale Regulation . . . . .	74
5.2.1	Das untersuchte Stoffwechselnetzwerk . . . . .	75
5.2.2	Zielstellungen und Anforderungen an die Regulationsstrategie . . . . .	77
5.2.3	Formulierung des Optimierungsproblems . . . . .	78
5.3	Optimale Regulationsstrategien . . . . .	80
5.3.1	Die minimale transkriptionelle Regulation . . . . .	80
5.3.2	Rolle der Enzyme am Netzwerkanfang und -ende . . . . .	86
5.3.3	Rolle der inneren Enzyme . . . . .	90
5.4	Konsequenzen für die Regulation in <i>Escherichia coli</i> . . . . .	91
5.4.1	Ableitung von Vorhersagen . . . . .	92
5.4.2	Verhalten in <i>Escherichia coli</i> . . . . .	93
5.5	Diskussion . . . . .	100
<b>6</b>	<b>Signaturen optimaler Aktivierung von Stoffwechselnetzwerken in Prokaryoten</b>	<b>102</b>
6.1	Motivation . . . . .	102
6.2	Aktivierung von Stoffwechselnetzwerken . . . . .	103
6.2.1	Das untersuchte Stoffwechselnetzwerk . . . . .	104
6.2.2	Zielstellungen und Anforderungen an die Aktivierung . . . . .	105
6.2.3	Formulierung des Optimierungsproblems . . . . .	106

6.3	Optimale Aktivierungsstrategien . . . . .	108
6.3.1	Zeitoptimale Aktivierung . . . . .	108
6.3.2	Wachstumsoptimale Aktivierung . . . . .	112
6.4	Konsequenzen für die Aktivierungsstrategien in Prokaryoten . . . . .	114
6.4.1	Ableitung von Vorhersagen . . . . .	114
6.4.2	Verhalten in Prokaryoten . . . . .	117
6.5	Diskussion . . . . .	122
<b>7</b>	<b>Schlussbemerkungen und Ausblick</b>	<b>124</b>
	<b>Anhang</b>	<b>129</b>
A.1	Anhang zum Kapitel 4 . . . . .	129
A.2	Anhang zum Kapitel 5 . . . . .	132
A.3	Anhang zum Kapitel 6 . . . . .	133

# Abkürzungen und mathematische Symbole

## Abkürzungen

<b>DAE</b>	differential algebraisches Gleichungssystem
<b>DGL</b>	differentielles Gleichungssystem
<b>DNA</b>	„deoxyribonucleic acid“
<b>HTML</b>	„Hypertext Markup Language“
<b>IP</b>	Innere Punkt Methode
<b>IPOPT</b>	„interior point optimizer“
<b>LP</b>	lineare Programmierung
<b>MFE</b>	„moving finite elements“
<b>mRNA</b>	„messenger RNA“
<b>NLP</b>	nichtlineare Programmierung
<b>qMFE</b>	„moving finite elements“ in der quasi-sequentiellen Methode
<b>RNA</b>	„ribonucleic acid“
<b>rRNA</b>	„ribosomal RNA“
<b>SNOPT</b>	„sparse nonlinear optimizer“
<b>SQP</b>	sequentielle quadratische Programmierung
<b>tRNA</b>	„transfer RNA“

## Liste ausgewählter Symbole

$C_n$	algebraisches Gleichungssystem des Modelles nach Diskretisierung
$C_j$	algebraisches Gleichungssystem in einer Untersimulation

<b>D</b>	Projektionsmatrix für die Anfangsbedingungen
$d_i(t)$	Anstieg der Enzymprofile durch Proteinsynthese
$d_{i,max}$	maximaler Anstieg der Enzymprofile durch Proteinsynthese
$d_{max}$	maximaler Anstieg aller Enzymprofile durch Gesamtproteinsynthese
$d_{\mu,max}$	maximaler Anstieg der Wachstumsrate
<b>E</b>	Enzym
<b>ES</b>	Enzymsubstratkomplex
$e_{tot}$	totale Enzymkonzentration
$e_i(t)$	Enzymkonzentrationen im Zeitverlauf
$\mathbf{e}(t)$	Vektor aller Enzymkonzentrationen
<b>F</b>	Vektorfunktion der differentiellen Zustände
<b>G</b>	Vektorfunktion der algebraischen Zustände
$k, k_i$	Parameter der Massenwirkungskinetik
$k_{cat}, k_{cat,i}$	„Turnover number“ der Michaelis-Menten-Kinetik
$K_M, K_{M,i}$	Halbsättigungskonstante der Michaelis-Menten-Kinetik
$\ell_j(t)$	Lagrange-Polynome
<b>N</b>	stöchiometrische Matrix
<b>NA</b>	Anzahl der algebraischen Zustände
<b>NC</b>	Kollokationsordnung
<b>NL</b>	Anzahl der finiten Elemente
<b>ND</b>	Anzahl der differentiellen Zustände
<b>P</b>	Produkt
$p(t)$	Produktkonzentration
$R_{n,i}^m(t_{n,i})$	Fehlerfunktionen der Kollokation

## Abkürzungen und mathematische Symbole

$S$	Substrat
$s(t)$	Substratkonzentration
$\mathbf{S}, \mathbf{S}_j$	Zusammenfassung diskretisierten Zustandswerte in einer Untersimulation
$t_f$	Produktionszeit bzw. Simulationszeithorizont
$t_{n,i}$	Kollokationspunkt
$t_{n,nonc}$	Nicht-Kollokationspunkt
$tol_i^m$	erlaubte Toleranz des Testkriteriums
$\mathbf{u}(t)$	Steuerungsvektor eines Systems
$\mathbf{u}(t)_{min}, \mathbf{u}(t)_{max}$	untere bzw. obere Schranke des Steuerungsvektors
$\mathbf{U}$	Zusammenfassung aller Steuerungsparametrierungen
$\mathbf{U}_j^S$	Zusammenfassung der Steuerungsparametrierungen in einer Untersimulation
$\mathbf{U}_{min}, \mathbf{U}_{max}$	untere und obere Schranken (diskretisiert)
$V_{max}, V_{max,i}$	maximale Reaktionsrate der Michaelis-Menten-Kinetik
$\mathbf{v}$	Reaktionsratenvektor der Michaelis-Menten-Kinetik
$v_i(t)$	Reaktionsraten der Michaelis-Menten-Kinetik
$v_{growth}(t)$	Ausflussrate zur Simulation verschiedener Szenarien
$\mathbf{x}(t)$	Zustandsvektor differentiell und algebraisch
$\mathbf{X}$	Zusammenfassung aller diskretisierten Zustandswerte (differentiell und algebraisch)
$\mathbf{X}_{min}, \mathbf{X}_{max}$	untere und obere Schranken (diskretisiert)
$\mathbf{X}(\mathbf{U})$	Simulationsergebnis aller diskretisierten Zustandswerte (differentiell und algebraisch)
$Y, Y_i$	Intermediate
$y(t), y_i(t)$	Intermediatkonzentrationen
$\mathbf{y}(t)$	algebraischer Zustandsvektor

$\mathbf{y}(t)_{min}, \mathbf{y}(t)_{max}$	untere bzw. obere Schranke
$\mathbf{Y}_{min}, \mathbf{Y}_{max}$	untere und obere Schranken (diskretisiert)
$\mathbf{z}(t)$	differentieller Zustandsvektor
$\mathbf{z}(t)_{min}, \mathbf{z}(t)_{max}$	untere bzw. obere Schranke
$z_n^m(t)$	polynomiale Approximation des differentiellen Zustandes
$z_{n,j}^m$	Kollokationspunkt des differentiellen Zustandes
$\dot{z}_{n,i}^m$	Ableitung am Kollokationspunkt des differentiellen Zustandes
$\mathbf{Z}_{min}, \mathbf{Z}_{max}$	untere und obere Schranken (diskretisiert)
$\alpha_{n,i}^m$	Testkriterium eines differentiellen Zustandes
$\Delta t_n$	finites Element
$\Delta \mathbf{T}$	Zusammenfassung aller finiten Element
$\varepsilon_{n,nonc}^m$	absolute Fehler der diskretisierten, differentiellen Zustandsgleichungen
$\mu(t)$	Wachstumsrate
$\sigma$	Gewichtung von Anfangskonzentrationen der Enzyme
$\tau_i$	Kollokationspunkt des normalisierten Zeitbereiches
$\tau_{nonc}$	Nicht-Kollokationspunkt des normalisierten Zeitbereiches
$\varphi(\mathbf{z}(t), \mathbf{y}(t), \mathbf{u}(t), t_f)$	allgemeine Darstellung der Zielfunktion
$\Psi_{rel}$	Operator zur Auswertung des Approximationsfehlers
$\Omega$	obere Schranke der Intermediatkonzentrationen

# Zusammenfassung

In dieser Arbeit werden optimale Regulationsstrategien von Stoffwechselnetzwerken unter Anwendung der quasi-sequentiellen Methode identifiziert und analysiert. Diese Vorgehensweise setzt Erweiterungen der quasi-sequentiellen Methode voraus, welche im ersten Teil der Arbeit beschrieben werden. Die Erweiterungen betreffen eine Kontrolle der Approximationsgenauigkeit der Zustandsverläufe und eine adaptive Diskretisierung während der Lösung des Optimierungsproblems. Hierbei wurde die Approximationskontrolle in der Simulationsschicht der quasi-sequentiellen Methode realisiert. Dies führt dazu, dass gradienten-basierte Lösungsalgorithmen in ihrer iterativen Lösungsstrategie ungehindert fortfahren können und somit keine Neustarts notwendig sind. Weiterhin verbessert die gesicherte Approximationsgenauigkeit die Konvergenzeigenschaften und erhöht die Robustheit gegenüber den Startwertschätzungen der zu optimierenden Steuerungsprofile.

Im zweiten Teil wird die Identifikation von Regulationsstrategien des Stoffwechsels mit den obigen Erweiterungen der quasi-sequentiellen Methode für zwei verschiedene Regulationsszenarien durchgeführt. In Szenario A wird durch die Formulierung von Optimalsteuerungsproblemen untersucht, welche Aufgabe die transkriptionelle Regulation bei der Kontrolle von Stoffwechselnetzwerken übernimmt. Es ergibt sich für steigende Kosten der Enzyme ein Umschalten des optimalen regulatorischen Programms von einer dünn verteilten, transkriptionellen Regulation zu einer umfassenden, transkriptionellen Regulation. Die Vorhersagen dieser regulatorischen Strategien wurden durch eine teilsystem-bezogene Datenanalyse in *Escherichia coli* überprüft und können durch einen Kompromiss zwischen zu minimierenden Kosten für Proteine und einer optimalen Antwortzeit auf veränderte Umweltbedingungen erklärt werden. In Szenario B wird die Situation untersucht, wo alle Enzyme transkriptionell reguliert werden und sich somit der Fokus auf optimale Aktivierungsstrategien verändert. Dabei ergeben sich, in Abhängigkeit der Proteinmassen und der Proteinsynthesekapazität verschiedene Ausprägungen von optimalen Aktivierungsstrategien. Weiterhin ergeben sich für große Unterschiede in den benötigten Proteinmassen komplexe Aktivierungsstrategien. Die Signaturen dieser Aktivierungsstrategien und auch der Einfluss der Beschränkungen wurden in den Regulationen von vielen Stoffwechselnetzwerken in hunderten Prokaryoten nachgewiesen.

# Abstract

This work analyzes regulatory strategies of metabolic pathways by using a dynamic optimization framework with the quasi-sequential approach. For this purpose it was in a first part necessary to improve the accuracy of state profile approximation and to derive a finite element placement strategy for the quasi-sequential approach. Since the accuracy control is only done in the simulation layer, the nonlinear programming solver need not be restarted and the resulting enhanced accuracy accelerates convergence performance and increases the robustness of the solution to initialization of the parameterised control profiles.

The second part consider two studies for the identification of regulatory strategies of unbranched metabolic pathways with the extended quasi-sequential approach: Study A investigates the extent to which transcriptional regulation controls metabolism. With increasing enzyme costs, the optimal regulatory program shifts from a sparse regulation of initial and terminal reactions to a pervasive regulation of all reactions within a pathway. The predicted regulatory strategies were confirmed by a subsystem dependent data analysis of *Escherichia coli* and can be explained by a trade-off between protein cost minimization and response time optimization due to changes in environmental conditions. Study B investigates the cases where all enzymes are transcriptional regulated. Here it is shown that protein abundance and protein synthesis capacity are key factors that determine the optimal activation strategies. Furthermore, in case of pathways with large differences in protein abundance, complex pathway activation strategies are optimal. Signatures of these pathway activation strategies as well as their dependence on the proposed constraints were confirmed by data analysis for a large number of metabolic pathways in several hundred prokaryotes.



# 1 Einleitung und Motivation

Die Biologie ist die Lehre vom Leben bzw. die Wissenschaft der belebten Natur und unterteilt sich je nach betrachteten lebendem System in verschiedenste Disziplinen. Historisch gesehen wurde das erworbene Wissen meist mit einer bildhaften Darstellung und der qualitativen Beschreibung von prinzipiellen Interaktion einzelner Teile dargestellt. Durch die Entwicklung und weiter anhaltende Verbesserung von Messtechnologien der letzten Jahrzehnte, entstand hierbei das Forschungsgebiet der Molekularbiologie. Die Molekularbiologie ist Teil der modernen Biologie und ermöglicht zelluläre Vorgänge auf molekularer Ebene zu untersuchen. Sie erweitert so bedeutend das erworbene biologische Wissen, insbesondere aber auch die Anzahl möglicher Interaktionen und deren Dynamiken. Für die Beschreibung und Untersuchung von diesen Interaktionen können Methoden der Systemwissenschaft angewendet werden. Die Systemwissenschaft beschäftigt sich mit den Wechselwirkungen und Zusammenhängen von komplexen Systemen. Es werden dabei die Strukturen und die Eigenschaften, wie beispielsweise Dynamiken oder Unsicherheiten der Systeme analysiert, um so das Verhalten zu verstehen und Vorhersagen über das zukünftige Verhalten machen zu können. Die Methoden werden stets unabhängig von den betrachteten Systemen (technisch, ökologisch, sozial, usw.) entwickelt und sind somit auf andere Systeme übertragbar.

Die Idee, die Systemwissenschaft und die Biologie zu verbinden, wurde bereits von Norbert Wiener vor ca. 60 Jahren in seinem bekannten Werk „Cybernetics: Or Control and Communication in the Animal and the Machine“ [155] aufgezeigt. Neuerungen in den Messtechniken zellulärer Prozesse (z.B. „high-throughput“, Transkriptionsstärken, Proteinmassen, Stoffwechselkonzentrationen, Stoffwechselflüsse) und deren fallende Kosten ermöglichten in den letzten Jahren einen starken Aufschwung von Wieners Gedanken, welcher die Forschungsrichtung der Systembiologie [78] mit zusätzlichen Methoden aus der Datenanalyse und Molekularbiologie etablierte. Dabei ist der Kreis der Systembiologie [79] als ein anzustrebendes Ideal anzusehen, welcher die verschiedenen Wissenschaften verbindet, damit diese sich gegenseitig inspirieren und so biologisches Wissen erweitern oder gar erneuern können. Im Segment der Systemwissenschaften entstehen so Beschreibungsformen des biologischen Verhaltens (Modelle), welche nicht nur in der Lage sind biologische Phänomene abzubilden, sondern diese auf Systemebene

erklären zu können. Letztendlich bietet dies auch Potential das beschriebene Objekt gezielt einzusetzen bzw. zu beeinflussen. Dies wird z.B. als „metabolic engineering“ bezeichnet und steht für die Manipulation der zellulären Produktion zur Erzeugung von gewünschten Stoffen.

Die Systembiologie beschäftigt sich mit komplexen und dynamischen, biologischen Prozessen auf zellulärer Ebene und untersucht dabei hauptsächlich drei Bereiche. Zur Einführung werden diese Bereiche im Folgendem kurz erläutert und für eine detaillierte Darstellung sei hier auf [3, 81, 105] oder in deutscher Sprache auf [z.B. 86] verwiesen. Der erste Bereich untersucht den reinen Stoffwechsel, welcher dazu dient, bestimmte Edukte über Zwischenprodukte in benötigte Stoffe umzusetzen und die Zelle mit Energie zu versorgen. Die Interaktionen innerhalb des Stoffwechsels werden in sogenannten Netzwerken dargestellt, welche die möglichen biochemischen Reaktionen beschreiben. Der zweite Bereich betrifft die Signalübertragung, wie z.B. von der Zellmembran zum Zellkern oder die externe Kommunikation mit anderen Zellen. Anhand dieser Signalübertragung und des aktuellen Zellzustandes steuert die Zelle den Stoffwechsel. Eine wesentliche Komponente der Kontrolle zellulärer Netzwerke ist die transkriptionelle Regulation und entspricht dem dritten Bereich.

Die Signalübertragung und insbesondere die dynamische Betrachtung der Regulation stellen die Schwerpunkte mit dem momentanen höchsten Forschungsbedarf dar. Die Untersuchungen fokussierten sich dabei meist auf die Modellorganismen *Escherichia coli* und *Saccharomyces cerevisiae*. *Escherichia coli* ist ein Bakterium, welches sich in der Darmflora befindet und es gehört zu der Gruppe der Prokaryoten (ohne Zellkern). *Saccharomyces cerevisiae* wird auch als Backhefe bezeichnet und ist ein einzelliger Pilz, welcher als Modellorganismus für Eukaryoten (mit Zellkern) oft betrachtet wird.

### 1.1 Einordnung und Zielstellung der Arbeit

Im Fokus dieser Arbeit steht die Regulation des Stoffwechsels unter Beachtung des dynamischen Vorhaltens. Eine weit verbreitete Modellierung der Regulation besteht in der Formulierung fester Regulationsregeln in Form von Wenn-Dann-Anweisungen [3, 81, 105]. Es wurden nur vereinzelt einfache dynamische Modelle erstellt, die eine Simulation des zeitlichen Verhaltens spezieller Regulationsnetzwerke und die Vorhersage der Netzwerkantwort auf Änderung der Randbedingungen ermöglichen [119]. Diese basieren auf der Verwendung von Differenzen- und Differentialgleichungen [30, 119, 138] und hybriden Modellen unter Einbindung von diskreten Regulationsnetzwerken [119, 127].

Bei der Analyse des Metabolismus (Synonym für Stoffwechsel) spielten bisher evolutionäre Betrachtungen eine wichtige Rolle [66, 80, 152, 158]. Dabei wird angenommen, dass die ausge-

bildeten Regulationsstrategien während der Entwicklungszeit durch Mutation und Selektion angepasst wurden, um das Überleben der Zelle durch gezielte Synthese von benötigten Stoffen zu sichern. Zur Einführung zeigt hier Abbildung 1.1 eine Illustration, wie die Bestandteile einer Zelle mit einem industriellen Prozess gegenüber gestellt werden könnten.

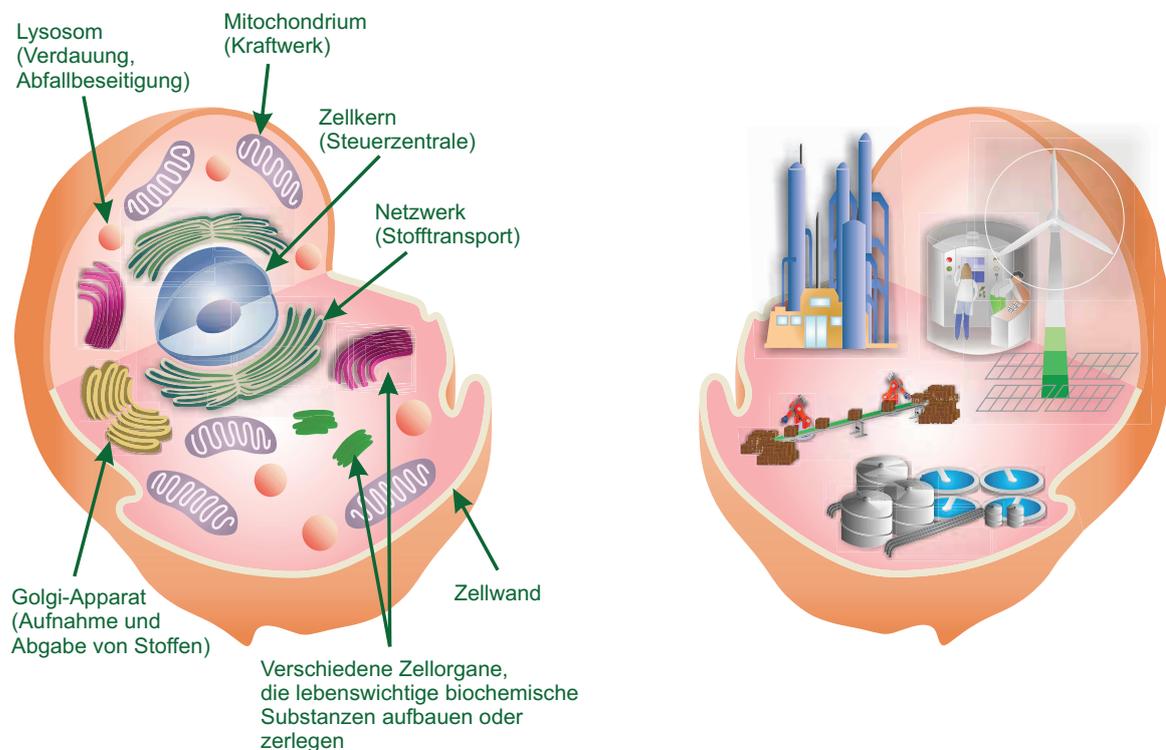


Abbildung 1.1: Die Zelle aus Sicht der industriellen Produktion

Derartige Vorstellungen werden im besonderen Maße in der sogenannten Flussbilanzanalyse (engl. Flux balance analysis) verwendet [49, 70]. Grundgedanke ist hier, dass Flüsse (oder auch Reaktionsraten) in einem stöchiometrischen Modell des Metabolismus derart gesteuert werden, dass die Wachstumsausbeute (definiert durch produzierte Biomasse pro aufgenommenem Substrat) unter gegebenen zellulären Beschränkungen maximiert wird. Auch wenn die grundlegende Annahme einer Optimalität dieser Wachstumsausbeute nicht überschätzt werden sollte [siehe 107, 120, 124], hat die Flussbilanzanalyse in einem wichtigem Maße zum heutigen Verständnis des Metabolismus beigetragen. So wird z.B. das Prinzip der Flussbilanzanalyse auf das stationäre Verhalten von Mikroorganismen angewendet und erfolgreich eingesetzt [7, 42, 89]. Die Flussbilanzanalyse ermöglicht somit eine Aussagekraft auf die stationäre Flussverteilung im Metabolismus, jedoch sind Aussagen zu einer dynamischen Regulation nicht möglich. Eine Erweiterung der Flussbilanzanalyse für dynamische Betrachtungen wurde in [96] aufgezeigt. Diese ist bereits zu der Untersuchung von verschiedenste Mikroorganismen und deren Lebensgemeinschaften eingesetzt worden. Übersichten hierfür sind in den aktuellen Arbeiten [89, 160] zu finden.

In der Arbeit von Klipp et al. [80] wurden erstmals optimale Steuerungsstrategien von dynamischen Stoffwechselnetzwerken untersucht. Dieser Ansatz basiert auf der Betrachtung von Reaktionswegen des Stoffwechsels und geht davon aus, dass sich dessen Regulation an Kriterien, wie minimale Übergangszeit oder maximale Überlebenszeit durch evolutionären Druck angepasst hat. Die Modellierung des Stoffwechsels wurde dabei durch Reaktionskinetiken realisiert und so eine direkte Aussage über dynamische Regulationsstrategien ermöglicht. Später führte die Gruppe von Alon [158] Messungen in *Escherichia coli* durch und konnte nachweisen, dass ein ähnliches Verhalten, wie bei den theoretischen Untersuchungen von [80] *in vivo* beobachtet werden kann. Der Begriff *in vivo* bezeichnet dabei Prozesse die im lebendigen Organismus ablaufen. Es wurde hierfür in der Literatur der Begriff „just-in-time activation“ geprägt, welcher auch in der Industrie verwendet wird („just-in-time production“). Vergleichbare Aktivitätsmuster wurden auch in verschiedenen Teilsystemen des Metabolismus von *Saccharomyces cerevisiae* beobachtet [29].

Dies veranschaulicht das Potential der optimalen Steuerung für die Identifikation und die Erklärung von dynamischen Regulationsstrategien im Stoffwechsel und stellt den Gegenstand der vorliegende Arbeit dar. Weiterhin beschäftigt sich die vorliegende Arbeit nicht mit einer ausschließlichen Anwendung von Methoden aus der Optimalsteuerungstheorie auf die Identifikation von dynamischen Regulationsstrategien im Stoffwechsel, sondern leistet auch hier einen Beitrag in Form von Erweiterungen einer bestehenden Methode. Dabei liegt der Fokus auf der Entwicklung von effizienten Algorithmen und der Implementierung von benutzerfreundlichen Formulierungsmöglichkeiten zur Lösung von optimal Steuerungsaufgaben.

## 1.2 Aufbau

Die vorliegende Arbeit leistet Forschungsbeiträge bei der Identifikation von Regulationsstrategien des Stoffwechsels und der Methodenentwicklung zur Lösung von optimalen Steuerungsproblemen. Die Grundlagen des Stoffwechsels und dessen Regulation werden in Kapitel 2 dargestellt. Dabei wird mit den notwendigen Grundlagen für die Modellierung von Stoffwechselwegen (Beschreibung durch Differentialgleichungen) begonnen und anschließend bekannte Mechanismen der Regulation von Stoffwechselwegen dargestellt. Da die Optimalitätsprinzipien in der Regulation von Stoffwechselwegen bereits Anwendung gefunden haben (siehe obige Diskussionen), werden auch diese hier grundlegend aufgezeigt. Die notwendigen Grundlagen der optimalen Steuerung sind entsprechend in Kapitel 3 zu finden. Hierbei wird das Prinzip der optimalen Steuerung erklärt und die Auswahl einer Lösungsmethode (quasi-sequentielle Methode) mit deren Eigenschaften aufgezeigt.

In Kapitel 4 wird die Notwendigkeit von Erweiterungen in der quasi-sequentielle Methode diskutiert und entsprechend eine Approximationskontrolle der Zustände und die Möglichkeit der Umsetzung einer adaptiven Diskretisierung aufgezeigt. Hierbei wird der Einfluss der Erweiterungen an Fallbeispielen jeweils erläutert und durch ein Anwendungsbeispiel der positive Einfluss der Erweiterungen gezeigt. Anschließend wird in Kurzform eine während der Arbeit entwickelte C++ Bibliothek vorgestellt und deren Eigenschaften diskutiert.

Die Anwendungen der entwickelten Bibliothek auf die Identifikation von Regulationsstrategien des Stoffwechsels sind Gegenstand der folgenden Kapitel. In Kapitel 5 wird die optimale Regulation metabolischer Netzwerke für den Modellorganismus *Escherichia coli* untersucht. Dabei wird eine minimale Regulation zur Kontrolle der Netzwerke (Flusskontrolle) anhand netzwerk-abhängiger Beobachtungen als optimales Steuerungsproblem formuliert und die Rolle der Regulation einzelner Enzyme aufgezeigt. Dabei werden Vorhersagen aus den optimalen Regulationsstrategien abgeleitet, welche durch Datenanalysen verifiziert werden bzw. eine zielgerichtete Datenanalyse erlauben. In Kapitel 6 werden Signaturen von optimalen Aktivierungsstrategien weiterführend in Prokaryoten untersucht. Dabei werden bisher nicht beachtete Kapazitäten in der Enzymsynthese analysiert und deren Einfluss durch verschiedene Formulierungen von optimalen Steuerungsproblemen aufgezeigt. Anschließend wird wiederum durch Ableitung von Vorhersagen eine zielgerichtete Datenanalyse durchgeführt und somit Signaturen der optimalen Aktivierungsstrategien in den regulatorischen Mechanismen von Prokaryoten untersucht.

Die vorliegende Arbeit wird in Kapitel 7 mit Schlussbemerkungen und Ausblick auf weiterführende Forschungsthemen für die Weiterentwicklung der quasi-sequentiellen Methode und die Untersuchung von optimalen Regulationsstrategien des Stoffwechsels abgeschlossen.

## 2 Grundlagen des Stoffwechsels und dessen Regulation

Auf zellulärer Ebene laufen verschiedenste chemische Reaktionen ab, welche für die Energiegewinnung, das Wachstum, die Zellteilung und somit zum Überleben notwendig sind. Hierbei wird z.B. in aerobe und anaerobe Atmung, Gärung und Photosynthese unterschieden. Die Gesamtheit all dieser biochemischen Reaktionen wird als Stoffwechsel oder Metabolismus bezeichnet.

Der Stoffwechsel verwendet Nährstoffe aus der Umgebung und nutzt diese zur Energiegewinnung, sowie zum Aufbau von Zellbestandteilen. Die Stoffwechselreaktionen laufen dabei meist nur durch Biokatalysatoren ab. Diese Katalysatoren werden als Enzyme bezeichnet und setzen somit die Aktivierungsenergie der biochemischen Reaktion herab und ermöglichen eine „portionierte“ Energiegewinnung und eine stückweise Synthese von Zellbestandteilen. Prinzipiell werden dabei die Nährstoffe durch aufeinander folgende Enzymreaktionen in unterschiedlichen Stoffwechselwegen in Moleküle für Zellbestandteile und in notwendige Energie umgesetzt. Hervorzuheben ist hier, dass die beteiligten Enzyme substrat-abhängig sind und somit ein Enzym meist nur eine biochemische Reaktion katalysieren kann. Dies ermöglicht der Zelle Stoffwechselwege bzw. -netzwerke explizit über diese Enzyme kontrollieren zu können. In den folgenden Kapiteln werden die zum Verständnis der Arbeit benötigten Grundlagen der dynamischen Stoffwechselmodellierung und der Stoffwechselkontrolle dargestellt.

### 2.1 Dynamische Modellierung von Stoffwechselnetzwerken

Die Beschreibung von biochemischen Reaktionsnetzwerken kann durch deterministische und stochastische Ansätze erfolgen. Die deterministische Beschreibungsform bedient sich der Bilanzierung von Stoff- und Energieströmen, welche durch differential und differential-algebraische Gleichungssysteme dargestellt werden. Die stochastische Modellierung basiert auf der „Master-Gleichung“ zur Simulation von Reaktionen mit einer geringen Zahl von Reaktionspartnern [siehe hierfür 59, 60, 61, 111]. In der vorliegenden Arbeit werden ausschließlich

deterministische Modelle von enzymkatalysierten Reaktionen betrachtet. Eine der einfachsten Beschreibungsformen ist die Modellierung durch bilineare Reaktionskinetiken. Hier wird die Reaktionsgleichung zur Überführung des Substrates  $S$  durch das Enzym  $E$  in das Produkt  $P$  wie folgt angenommen



welche in folgende Reaktionskinetik resultiert:

$$\begin{aligned} \dot{s}(t) &= -e_{tot} \cdot k \cdot s(t); & s(t_0) &= s_0 \\ \dot{p}(t) &= +e_{tot} \cdot k \cdot s(t); & p(t_0) &= p_0. \end{aligned} \tag{2.1}$$

Hierbei stellt  $s(t)$  die Konzentration des Substrats,  $e_{tot}$  die totale Enzymkonzentration und  $p(t)$  die Konzentration des Produktes dar. Diese bilineare Reaktionskinetik besitzt keine Sättigung der Reaktionsrate bezüglich des Substrates, wie es *in vivo* zu beobachten ist und gilt somit nur für kleine Substratkonzentrationen als Näherung. Wird die plausible Vorstellung aus Abbildung 2.1 betrachtet, wobei Enzym und Substrat erst aneinander binden (im Folgenden dargestellt durch  $ES$ ) und diese Bindung auch wieder zerfallen kann, ohne dass eine Substratmenge in Produkt umgesetzt wird, entsteht die Reaktionsgleichung

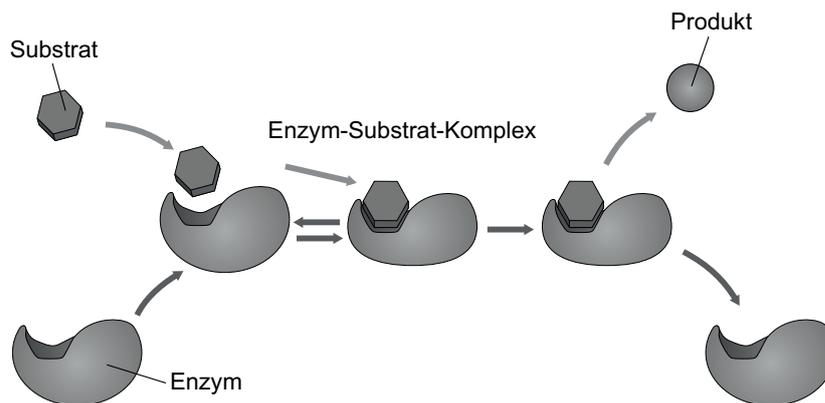
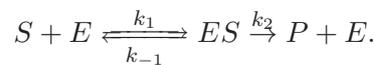


Abbildung 2.1: Der katalytische Zyklus einer Enzymreaktion

Werden die entsprechenden Gleichungen (durch Massenwirkungskinetik) für die Reaktionsraten aufgestellt, entsteht ein differentielles Gleichungssystem. Da dies für nur einen Stoffwechselschritt steht, liegt es nahe mögliche Vereinfachungen vorzunehmen und so einen kurzen geschlossenen Ausdruck abzuleiten. Hierbei sei auf [22, 97] verwiesen, welche unter den darin

## 2 Grundlagen des Stoffwechsels und dessen Regulation

begründeten Annahmen  $k_1, k_{-1} \gg k_2$  oder einer Gleichgewichtsreaktion von  $S + E \xrightleftharpoons[k_{-1}]{k_1} ES$  die Michaelis-Menten-Kinetik mit

$$\begin{aligned} \dot{s}(t) &= -v; & s(t_0) &= s_0 \\ \dot{p}(t) &= +v; & p(t_0) &= p_0 \end{aligned} \quad (2.2)$$

$$v = \frac{V_{max} \cdot s(t)}{K_M + s(t)}$$

ableiteten. Dabei stellt  $V_{max} = e_{tot} \cdot k_2$  die maximale Reaktionsrate von  $v$  und  $K_M = \frac{k_{-1} + k_2}{k_1}$  oder  $K_M = \frac{k_{-1}}{k_1}$  die sogenannte Halbsättigungskonstante dar. In der englischen Literatur findet auch der Begriff „Turnover number“ Verwendung, welche sich durch  $k_{cat} = \frac{V_{max}}{e_{tot}}$  definiert und somit gilt  $k_{cat} = k_2$ . Für eine detaillierte Herleitung und grafische Auswertung der Kinetik, sei hier auf die einschlägige Literatur verwiesen [z.B. 81, 86].

Modelle für einzelne Stoffwechselabschnitte und deren Regulation sind beispielsweise für *Escherichia coli* in [18, 23, 26, 27, 28, 40, 84, 146, 149] oder für *Saccharomyces cerevisiae* in [82, 101, 115, 140, 143, 147, 148] zu finden. Dabei werden verschiedenartige Reaktionskinetiken wie Hill, linlog oder Derivate von Gleichung (2.2) benutzt. Ziel dieser Arbeit ist es jedoch generelle Regulationsprinzipien zu identifizieren, welche nicht nur auf einzelne konkretisierte Netzwerke zutreffen. Für diese Untersuchungen wird hier die Michaelis-Menten-Kinetik gewählt, um das nichtlineare Sättigungsverhalten von biochemischen Reaktionen zu beschreiben. In Abbildung 2.2 wird eine typische Abfolge von nacheinander stattfindenden Stoffwechselreaktionen dargestellt. Hierbei ist das Produkt eines Stoffwechselnetzwerkes das Ergebnis mehrere Schritte (Intermediate).

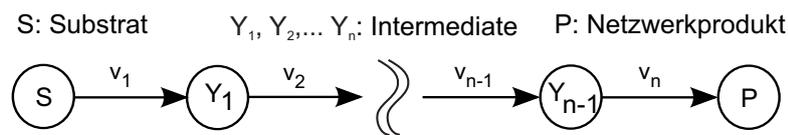


Abbildung 2.2: Modellstruktur der hier verwendeten Stoffwechselnetzwerke

Zur mathematischen Beschreibung wird in der Literatur eine stöchiometrische Schreibweise bevorzugt, welche auch hier verwendet werden soll. Diese fasst die stöchiometrischen Koeffizienten und die Netzwerkstruktur in der stöchiometrischen Matrix  $\mathbf{N}$ , alle Metabolite in einem Zustandsvektor  $\mathbf{z}(t)$  und alle Reaktionsraten im Vektor  $\mathbf{v}$  zusammen:

$$\dot{\mathbf{z}}(t) = \mathbf{N} \cdot \mathbf{v} ; \mathbf{z}(t_0) = \mathbf{z}_0. \quad (2.3)$$

Jede einzelnen Reaktion wird durch ein spezifisches Enzym katalysiert und durch die Michaelis-

Mentenkinetik modelliert. Es ergibt sich somit

$$\mathbf{v} = \mathbf{v}(\mathbf{z}(t)), \quad (2.4)$$

$$v_i(z_i(t)) = \frac{V_{max,i} \cdot z_i(t)}{K_{M,i} + z_i(t)}$$

und

$$V_{max,i} = e_{tot,i} \cdot k_{cat,i}.$$

Die Nachführung des Substrates durch vorhergehende Stoffwechselwege entspricht einer oft vorhandene Situation *in vivo*, welche hier motiviert das Substrat als gepuffert zu modellieren. Für die Stöchiometrie wird angenommen, dass aus einem Mol Substrat auch nur ein Mol Produkt gebildet werden kann. Es ergibt sich somit für den in Abbildung 2.2 gezeigten Stoffwechselweg:

$$\mathbf{N} = \begin{bmatrix} 0 & 0 & \cdots & 0 \\ 1 & -1 & \cdots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ 0 & 1 & -1 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}$$

mit

$$\mathbf{z}(t) = [s(t), y_1(t), y_2(t), \dots, y_{n-1}(t), p(t)]^T$$

und den in Gl. (2.3) und (2.4) dargestellten Beziehungen.

## 2.2 Regulation des Stoffwechsels

Der Stoffwechsel in Zellen basiert auf sehr vielen biochemisch katalysierten Reaktionen, die gleichzeitig ablaufen. Diese finden *in vivo* in verschiedenster Stärke bzw. unterschiedlichem Umfang statt und unterliegen einer Kontrolle bzw. Regulation. Werden Stoffwechselprodukte in größerer Menge benötigt, z.B. als Substrat für mehrere darauffolgende Reaktionen oder einfach aus erhöhtem Bedarf, müssen die Stoffwechselreaktionen einen höheren Fluss realisieren - im umgekehrten Fall gilt es die Flussrate gezielt zu senken. Hierbei ermöglicht die Substratabhängigkeit der Enzyme eine gezielte Regulation des Stoffwechsels. Dabei stellen prinzipiell alle Enzyme mögliche Kontrollpunkte des Stoffwechsels dar. Dies kann über die Enzymkonzentration oder über eine Beeinflussung deren katalytischen Eigenschaften erfolgen. Die dabei bekannten Prozesse sollen im Folgendem kurz dargestellt werden.

Die Synthese von Enzymen ermöglicht eine gezielte Konzentrationsanpassung im Zellmedium und somit eine Kontrolle über die Reaktionsgeschwindigkeiten (z.B. Veränderung von  $V_{max}$  aus Gl. 2.4 über die Enzymmenge). Dieser Weg beginnt bei der Expression eines Genes der DNA, wobei der entsprechende Sequenzbereich der DNA durch „ribonucleic acid“ (RNA)-Polymerase in mRNA umgeschrieben wird. Hierbei steht mRNA für die Abkürzung von „messenger RNA“, welche die genetische Information der DNA zum Ribosom überträgt. Der prinzipielle Ablauf ist in Abbildung 2.3 dargestellt. Dabei bindet die RNA-Polymerase an einen Promoter der DNA (Initiation), welcher als Startpunkt für einen DNA-Sequenzbereich von Genen dient. Die RNA-Polymerase läuft dann entlang des DNA-Stranges (Elongation) und bildet dabei das mRNA-Transkript. Die hierbei entstandene mRNA ist kurzlebig und beeinflusst somit die Enzymsynthese nur für einen kurzen Zeitraum. Dieser Weg von der DNA zur mRNA wird als Transkription bezeichnet, welcher nun die Translation folgt.

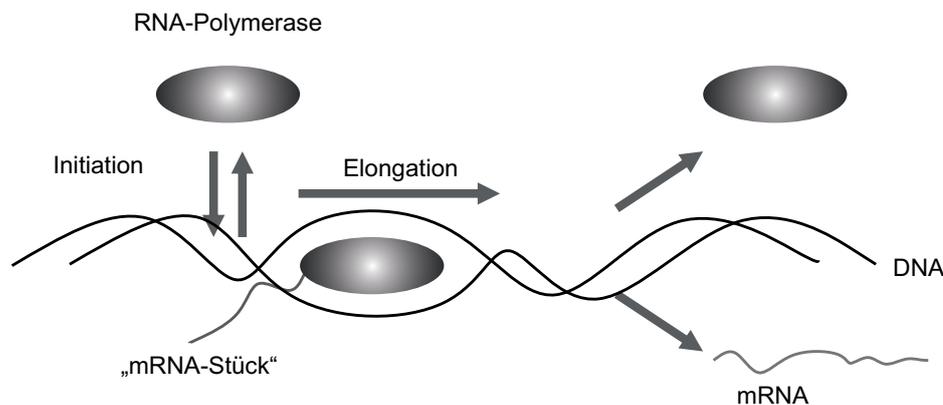


Abbildung 2.3: Von der DNA zur mRNA - die Transkription

Die Translation beschreibt die Proteinbiosynthese aus der transkribierten mRNA durch die Ribosomen (siehe auch Abbildung 2.4). Die Ribosomen stellen somit die Proteinbiosynthesemaschine der Zelle dar, welche sich durch ein katalytisches und strukturelles Element, die ribosomale RNA (rRNA), auszeichnen. Dabei bindet die mRNA an die Ribosombindestellen an und die Transfer-RNA (tRNA) wirkt als Adaptermolekül, welches die Bausteine (Aminosäuren) für die zu synthetisierenden Proteine bereitstellt. Das Binden der mRNA an das Ribosom wird als Initiation bezeichnet, dieser folgt ein Abwandern der mRNA durch das Ribosom beginnend beim Start-Codon, wobei nun die dazugehörigen Aminosäuren der wachsenden Polypeptidkette durch die tRNA verknüpft werden (Elongation). Wird nun ein Codon erreicht, für welches keine „beladene“ tRNA zu Verfügung steht (Stopp-Codon) wird die Translation beendet (Termination), die Polypeptidkette durch andere Proteine abgespalten und das Ribosom von der mRNA gelöst. Die richtige Faltung des zu synthetisierenden Proteins aus der Polypeptidkette wird in vielen Fällen wiederum durch andere Enzyme (Cha-

perone) realisiert. Der dargestellte Prozess der Translation ist für Prokaryoten spezifisch, wohingegen die Translation für Eukaryoten mehrere Zwischenschritte beinhaltet und wird aus Gründen der Übersichtlichkeit nicht dargestellt. An dieser Stelle sei für eine sehr detaillierte molekulare Darstellung des Weges von der DNA zum Protein auf [1] verwiesen oder auch die Übersetzung in Deutscher Sprache empfohlen [2].

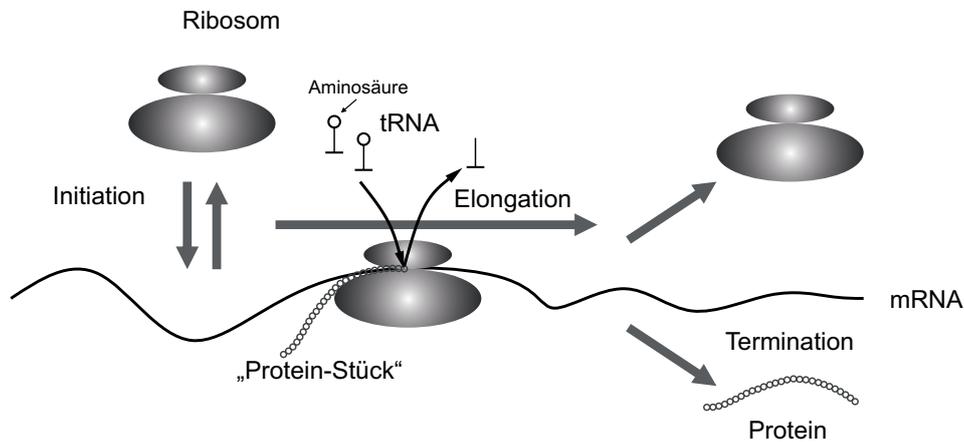


Abbildung 2.4: Die Translation - von der mRNA zum Protein

Die Regulation des Stoffwechsels kann durch eine gezielte Enzymsynthese (d.h. der Enzymmenge) auf der Ebene der Transkription und Translation ansetzen. Diese werden dann entsprechend als transkriptionelle Regulation und translationale Regulation bezeichnet. Auf der Ebene der transkriptionellen Regulation kann bestimmt werden, wann bzw. ob ein Transkript entsteht und wie viel mRNA transkribiert wird, welche sich entsprechend auf die Proteinbiosynthese durch die Ribosomen auswirkt. Die translationale Regulation bestimmt wie oft eine mRNA durch ein Ribosom übersetzt wird und beeinflusst somit direkt die Proteinbiosynthese. Darüber hinaus können RNA und Proteine nach ihrer Synthese gezielt abgebaut werden. Diese Kontrollmöglichkeiten werden dementsprechend einer post-transkriptionellen und post-translationalen Regulation zugeordnet. Zusammenfassend haben alle Regulationsarten der Enzymsynthese eine Veränderung der Enzymkonzentration in der Zelle zu Folge.

In dieser Arbeit werden hauptsächlich die Eigenschaften der Regulation von prokaryotischen Zellen untersucht. Wie oben beschrieben kann prinzipiell jeder Schritt der Enzymsynthese ein möglicher Angriffspunkt für regulatorische Eingriffe sein, um den Stoffwechsel gezielt zu beeinflussen. Hierbei tritt die transkriptionelle Regulation (zumindest bei Bakterien wie *Escherichia coli*) am häufigsten auf, um die Enzymsynthese zu kontrollieren. Dies erscheint auch plausibel, da es energetisch erhöhten Aufwand bedarf RNA zu transkribieren, wenn das dazugehörige Protein nicht benötigt wird. In Prokaryoten kann dies z.B. über Aktivierung und Repression auf transkriptioneller Ebene durch sogenannte Operons (siehe Abbildung 2.5)

realisiert werden. Ein Operon beinhalten eine Gruppe von Genen, welche unter der Kontrolle eines Promoters und einer Kontrollsequenz (oder auch Operator genannt) stehen. Dabei binden Regulatorproteine an den Operator an und ermöglichen die Transkription durch die RNA-Polymerase (Aktivierung) oder verhindern diese (Repression). Bekannte Regulationen durch Operons sind in der Tryptophan-Biosynthese (*trp*-Operon) oder bei der Aufnahme und dem Abbau von Laktose (*lac*-Operon) *in vivo* umgesetzt. Erste Untersuchungen sind hier auf 1961 und 1953 zu datieren [73, 99] und wurden mit den Nobelpreis für Physiologie und Medizin ausgezeichnet.

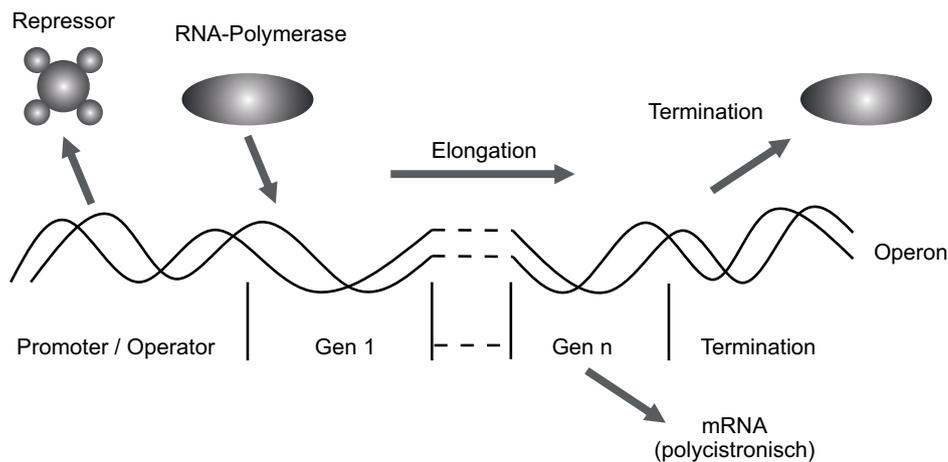


Abbildung 2.5: Transkriptionelle Regulation für Operonstrukturen

Die Regulation des Stoffwechsels kann auch über eine gezielte Modifikation der Enzymaktivität realisiert werden, also nach dem das Enzym bereits vollständig synthetisiert wurde. Hier kann z.B. eine direkte Produktinhibition oder eine End-Produkthemmung des Stoffwechselproduktes vorliegen. Die Regulation der Aktivität eines Enzyms beeinflusst im Vergleich zur Regulation der Enzymsynthese den betreffenden Stoffwechselabschnitt sehr schnell. Umfangreiche Beispiele für alle Regulationsarten und die Kontrolle der Genexpression sind in [1, 2] zu finden.

### 2.3 Optimalitätsprinzipien und Stoffwechselregulation

In Kapitel 1.1 wurde die Vorgehensweise für die Modellierung der Regulation in Form von festen Regulationsregeln durch Wenn-Dann-Anweisungen diskutiert, sowie die Modellierung des zeitlichen Verhaltens des Regulationsnetzwerkes durch Differenzen- oder Differentialgleichungen dargestellt und auf hybride Modelle unter Einbindung von diskreten Regulationsnetzwerken eingegangen [3, 30, 81, 105, 119, 127, 138]. Doch basiert die dynamische Modellierung dieser genregulatorischen Mechanismen auf der Verwendung von Messwerten, welche

nur schwer oder bisher nicht zugänglich sind. Hinzukommt hier, dass ein derartiges Vorgehen keinen expliziten Hinweis darauf geben kann, warum verschiedenste Regulationsstrategien vorkommen. Deshalb werden hier alternative Möglichkeiten diskutiert, welche auf Optimalitätskriterien in der Regulation von Stoffwechselnetzwerken basieren. Im Folgendem werden die wichtigsten Ansätze mit den Grundlagen aus dem vorherigen Abschnitt in Beziehung gebracht und ihr mögliches Leistungspotential diskutiert.

Die Flussbilanzanalyse [70] basiert dabei auf der Verwendung von Gleichung (2.3) mit  $\dot{\mathbf{z}} = 0$ :

$$0 = \mathbf{N} \cdot \mathbf{v}. \quad (2.5)$$

Dieses Gleichungssystem besitzt im allgemeinen keine eindeutige Lösung, da es unterbestimmt ist ( $\text{Rang}(N) \ll \text{Dim}(v)$ ). Um dieses Problem zu umgehen, kann eine Zielfunktion definiert werden, welche z.B. die Wachstumsausbeute des betreffenden Netzwerkes darstellt (ein Vergleich verschiedenster Zielfunktionen ist in [120] zu finden). Dies führt zur plausiblen Maximierung der Wachstumseffizienz des untersuchten Netzwerkes und kann im Allgemeinen ausgedrückt werden durch

$$\min_{\mathbf{v}} f(\mathbf{v}).$$

Dabei werden die Flussraten  $v$  direkt als Optimierungsvariablen verwendet, welche somit durch lineare Programmierung (LP) ermittelt werden können. Zudem müssen Flussrichtungen und maximale Flussraten beachtet werden (in der Literatur wird auch der Begriff „Constraint based model“ [105] verwendet). Die dynamische Flussbilanzanalyse [96] erweitert diese Betrachtungen auf eine Reihe von Zeitpunkten, wobei sich die Flussraten zwischen diesen Zeitpunkten nur in einem definierbarem Maße ändern dürfen.

Als Vorteil der Verfahren hebt sich hier hervor, dass durch die optimale Flussratenverteilung die Kinetiken dahinter nicht benötigt werden, sondern einzig die stöchiometrischen Informationen aus  $N$ , welche in Datenbanken zu finden sind (siehe unter anderem z.B. [76]). Die Verfahren vernachlässigen jedoch die Eigenschaften der katalysierenden Enzyme, da ausschließlich die Flüsse zwischen den Stoffen ermittelt werden. Im Umkehrschluss lassen sich somit bedingt generelle Aussagen treffen, welche den Einfluss der Enzyme mit einschließen. Dies betrifft entsprechend auch die Regulation der Enzymsynthese und Enzymaktivität.

Die Pionierarbeit von Klipp et al. [80] untersuchte zum ersten Mal optimale Steuerungsstrategien von Stoffwechselwegen. In derartigen Untersuchungen werden dynamische Modelle des Stoffwechsels betrachtet, welche durch Kinetiken entsprechend Gleichung (2.1) modelliert wurden. Das untersuchte System befindet sich somit nicht im stationärem Zustand wie die

Flussbilanzanalyse aus Gleichung (2.5). In [80] dienen die Enzymprofile (zeitlicher Verlauf der Enzyme) als Regulationsmechanismen, womit die totalen Enzymmengen Funktionen der Zeit darstellen ( $e_{tot,i} = e_i(t)$ ) und es ergibt sich in Vektorschreibweise

$$\mathbf{e}(t) = [e_1(t), e_1(t), \dots, e_n(t)]^T; e_i(t) \geq 0$$

und somit

$$\dot{\mathbf{z}}(t) = \mathbf{N} \cdot \mathbf{v}(\mathbf{z}(t), \mathbf{e}(t)); \mathbf{z}(t_0) = \mathbf{z}_0 \quad (2.6)$$

für die untersuchten Fragestellungen. Der Ansatz basiert nun auf der Annahme, dass Zellen Anpassungsstrategien an wechselnde Umweltbedingungen entwickelt haben, welche einem bestimmten Ziel folgen. Untersucht wurde z.B. die Aktivierung eines unverzweigten Stoffwechselnetzwerkes aus Abbildung 2.2 unter der Annahme, dass die Übergangszeit [93]

$$\min_{\mathbf{e}(t)} \int_0^{\infty} \frac{1}{s(0)} (s(0) - p(t)) dt$$

mit einer limitierten Gesamtproteinmenge nach

$$\sum_{i=1}^n (e_i(t)) \leq E_{gesamt}$$

für den betrachteten Reaktionsweg minimiert wurde. Hierbei wurde eine sequentielle Aktivierung von Enzymen entsprechend ihrer Reihenfolge in einem Reaktionsweg ermittelt („just-in-time activation“) und durch [29, 158] in *Saccharomyces cerevisiae* und *Escherichia coli* beobachtet.

Aufbauend auf Klipp et al. wurden in [103] und auch durch eigene Vorarbeiten [10] zeitoptimale Aktivierungsstrategien durch

$$\min_{\mathbf{e}(t), t_f} t_f$$

untersucht ( $t_f$  entspricht der Produktionszeit). Wobei in [103] eine Verallgemeinerung auf weitere Enzymkinetiken abgeleitet und durch einen zusätzlichen Strafterm verbrauchsorientierte Strategien bezüglich der investierten Enzymmenge untersucht wurden. Die in [10, 103] aufgezeigten optimalen Regulationsstrategien beschreiben im Vergleich zu [80] qualitativ besser, dass in [158] beobachtete Verhalten. Dies liegt letztendlich an einer zielgerichteten Problemformulierung anhand bekannter Regulationsmuster und reduziert nicht das durch [80] gezeigte prädiktive Potential von Optimalitätskriterien. Die Untersuchung von dynamischer Optimie-

regung für die Identifikation und Erklärung von Regulationsstrategien des Stoffwechsels wird somit in der vorliegenden Arbeit weiter verfolgt und offene Fragen im Folgendem diskutiert.

Die Regulation des Stoffwechsels geschieht über eine gezielte Enzymsynthese und Modifikation der Enzymaktivität (siehe Kapitel 2.2). Die Verwendung der Enzymprofile als Freiheitsgrade schafft hier einen Zugang zu beiden Regulationsprozessen. So kann angenommen werden, dass folglich der Aussagen aus Kapitel 2.2 langsame zeitliche Veränderungen in der Regulationsvariable die Enzymsynthese und schnelle Veränderungen die Modifikation der Enzymaktivität darstellen. Vorgreifend sei hier erwähnt, dass diese sich im Verlauf der Arbeit als wichtige Einflussfaktoren für die Entwicklung verschiedenster Regulationsstrategien herausstellen. So kann z.B. die instantane Verteilung einer Proteinmenge zwischen katalysierten Reaktionen aus [80], durch eine Enzymsynthese (hier auch der Abbau von Enzymen) von der Zelle nicht realisiert werden. Dies wurde in [103] nur oberflächlich untersucht, wobei in fortführenden Arbeiten sogar eine Vernachlässigung von biologisch notwendigen Bedingungen vorgenommen wurde [104].

Eine weitere wichtige Bedingung in zellulären Systemen, besteht darin, dass in der Regel Stoffwechselprodukte nicht zu stark angehäuft werden, da dies eine Gefahr für die Zelle darstellt (osmotischer Druck). Für den stationären Fall wurden in [123] bereits Untersuchungen durchgeführt, welche hier für die dynamischen Betrachtungen als Beschränkung nach

$$\sum_{i=1}^n (z_i(t)) \leq \Omega$$

eingeführt werden können.

Durch den dargestellten Ansatz und die in Kapitel 2.2 aufgezeigten biologischen Zusammenhänge, kann nun zusammenfassend die illustrative Darstellung aus Abbildung 1.1 in Bezug auf die Regulationsmechanismen und des Optimalsteuerungsproblems in Abbildung 2.6 veranschaulicht werden. Hier ist zu erkennen, dass der Optimalitätsansatz die regulatorischen Mechanismen in Zellen identifizieren oder gar modellieren kann und welche biologischen Vorgänge einen Einfluss auf die Teile der Problemformulierung haben. Es wurden in den bisherigen Ansätzen analytische Lösungen des Optimierungsproblems angestrebt, was einerseits die Untersuchung von komplexen Fragestellungen verhindert, aber andererseits die gefundenen Zusammenhänge verallgemeinert. In der vorliegenden Arbeit gilt es beide Sachverhalte zu bedienen und dementsprechend die Details der Fragestellung zu erhöhen und die gefundenen Zusammenhänge zu generalisieren. Die Grundlagen des hierfür gewählten Verfahrens werden im nächsten Kapitel dargestellt.

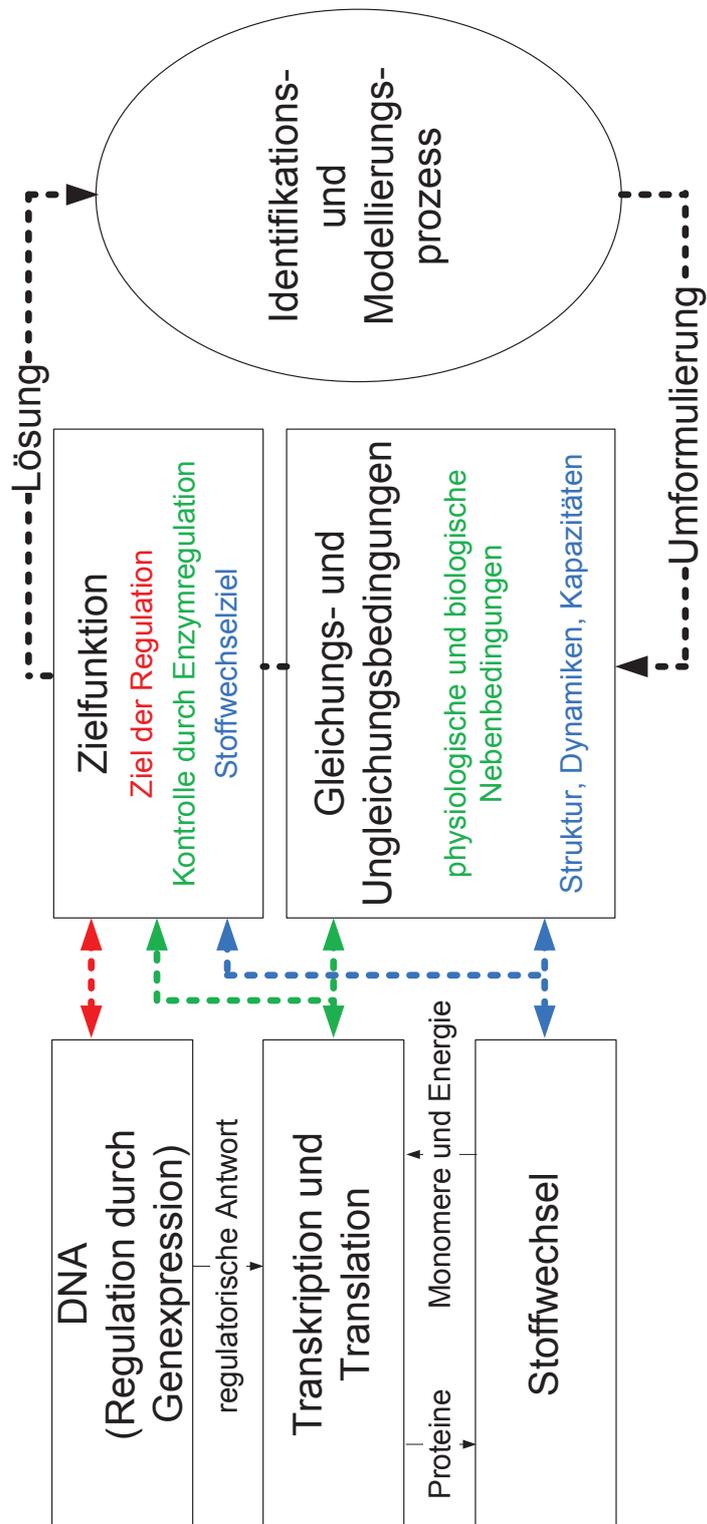


Abbildung 2.6: Optimalitätsprinzipien in der Stoffwechselregulation

## 3 Grundlagen der optimalen Steuerung

Die optimale Steuerung basiert im Allgemeinen auf der Lösung eines nichtlinearen dynamischen Optimierungsproblems zur Berechnung der gesuchten optimalen Steuerstrategie eines technischen Prozesses. Die effiziente Lösung nichtlinearer dynamischer Optimierungsprobleme stellt bis heute eine wichtige Aufgabe der Forschung dar. Ein Überblick über die Wirkungsweise der optimalen Steuerung und der Problemdarstellung wird in Kapitel 3.1 dargestellt und in Kapitel 3.2 mögliche Lösungsmethoden diskutiert, sowie die Auswahl einer Methode getroffen. Die Vorgehensweise der ausgewählten Methode wird anschließend in 3.3 im Detail aufgezeigt.

### 3.1 Die Problemdarstellung

Hier wird die grundlegende Vorstellung der Ein- und Ausgangsbeschreibung von Prozessen eingeführt und im Folgendem als Systeme bezeichnet [153]. Zur Einführung ist eine vereinfachte Darstellung in Abbildung 3.1 gezeigt. Die untersuchten Systeme sind nichtlinearer und dynamischer Natur und besitzen mehrere Ein- und Ausgangsgrößen, wobei eine Rückwirkungsfreiheit herrscht (die Eingangsgrößen beeinflussen ausschließlich die Ausgangsgrößen). Dabei stellt  $\mathbf{u}(t)$  die vektorielle Zusammenfassung aller system-beeinflussenden Größen und  $\mathbf{x}(t)$  die vektorielle Zusammenfassung aller Zustände des Systems dar. Für technische Prozesse werden meist Größen definiert, welche einer Teilmenge von  $\mathbf{x}(t)$  entsprechen (Regelgrößen oder zusteuernde Größen), doch wird aufgrund der biologischen Untersuchungen in der vorliegenden Arbeit bewusst darauf verzichtet.



Abbildung 3.1: Eine einfache Vorstellung des Systemgedankens

Die zu untersuchenden Systeme entsprechen dem Stoffwechselmodell aus Gleichung (2.6). Hinzu kommt, dass algebraische Gleichungen z.B. in Form von Bilanzgleichungen auftreten können und eine Beschreibung durch ein differential algebraisches Gleichungssystem (DAE)

nach

$$\frac{d\mathbf{z}(t)}{dt} = \mathbf{F}(\mathbf{z}(t), \mathbf{y}(t), \mathbf{u}(t), t) ; \mathbf{z}_0 = \mathbf{z}(t_0)$$

$$\mathbf{G}(\mathbf{z}(t), \mathbf{y}(t), \mathbf{u}(t), t) = \mathbf{0}$$

erfolgen kann, welches den Index 1 nicht überschreitet [siehe z.B. 24]. Hierbei stellt nun  $\mathbf{z}(t) \in \mathfrak{R}^{ND}$  den differentiellen Zustandsvektor beschrieben durch  $\mathbf{F} \in \mathfrak{R}^{ND}$  mit  $\mathbf{z}_0$  als Anfangsbedingung dar und  $\mathbf{y}(t) \in \mathfrak{R}^{NA}$  entspricht dem algebraischen Zustandsvektor beschrieben durch  $\mathbf{G} \in \mathfrak{R}^{NA}$ . Diese Trennung von algebraischen und differentiellen Zuständen ist für spätere Betrachtungen notwendig und wird deshalb hier eingeführt. Für die Darstellung in Abbildung 3.1 gilt  $\mathbf{x}(t) = [\mathbf{z}^T(t), \mathbf{y}^T(t)]^T$ .

Die optimale Steuerung derartiger Systeme basiert nun auf der Formulierung einer Zielstellung in mathematischer Form. Dies wird als Zielfunktion bezeichnet und kann im Allgemeinen wie folgt dargestellt werden

$$\min_{\mathbf{u}(t), \mathbf{z}(t), \mathbf{y}(t), t_f} \varphi(\mathbf{z}(t), \mathbf{y}(t), \mathbf{u}(t), t_f).$$

Weiterhin können Anforderungen bestehen bestimmte Begrenzungen einzuhalten, welche sich z.B. für technische Prozesse durch Prozessbeschränkungen wie sicherheitsrelevante Grenzen und/oder Qualitätsanforderungen definieren. Für die zu untersuchenden Stoffwechselregulationen können dies Leistungskapazitäten der Zelle oder Stoffwechsellanforderungen sein. Eine Formulierung in allgemeiner Form lässt sich für die Steuergrößen durch

$$\mathbf{u}_{min}(t) \leq \mathbf{u}(t) \leq \mathbf{u}_{max}(t)$$

und für die Zustandsgrößen durch

$$\mathbf{z}_{min}(t) \leq \mathbf{z}(t) \leq \mathbf{z}_{max}(t) ; \mathbf{y}_{min}(t) \leq \mathbf{y}(t) \leq \mathbf{y}_{max}(t)$$

darstellen. Werden nun die gezeigten Sachverhalte zusammenfassend dargestellt, ergibt sich das nichtlineare dynamische Optimierungsproblem zu

$$\min_{\mathbf{u}(t), \mathbf{z}(t), \mathbf{y}(t), t_f} \varphi(\mathbf{z}(t), \mathbf{y}(t), \mathbf{u}(t), t_f) \tag{3.1}$$

$$\frac{d\mathbf{z}(t)}{dt} = \mathbf{F}(\mathbf{z}(t), \mathbf{y}(t), \mathbf{u}(t), t) ; \mathbf{z}_0 = \mathbf{z}(t_0) \tag{3.2}$$

$$\mathbf{G}(\mathbf{z}(t), \mathbf{y}(t), \mathbf{u}(t), t) = \mathbf{0} \tag{3.3}$$

$$\mathbf{u}_{\min}(t) \leq \mathbf{u}(t) \leq \mathbf{u}_{\max}(t) \quad (3.4)$$

$$\mathbf{z}_{\min}(t) \leq \mathbf{z}(t) \leq \mathbf{z}_{\max}(t) \quad (3.5)$$

$$\tilde{\mathbf{y}}_{\min}(t) \leq \mathbf{y}(t) \leq \tilde{\mathbf{y}}_{\max}(t). \quad (3.6)$$

Dies entspricht einem Optimierungsproblem mit Ungleichungsnebenbedingungen für Steuer- und Zustandsverläufe. Die Behandlung von nichtlinearen Nebenbedingungen  $\mathbf{H} \in \mathfrak{R}^{NO}$  wie

$$\mathbf{h}_{\min}(t) \leq \mathbf{H}(\mathbf{z}(t), \mathbf{y}(t), \mathbf{u}(t), t) \leq \mathbf{h}_{\max}(t)$$

wird durch die Verwendung von Schlupfvariablen  $\mathbf{s}(t) \in \mathfrak{R}^{NO}$  erreicht. Dabei gilt

$$\tilde{\mathbf{G}}(\mathbf{z}(t), \tilde{\mathbf{y}}(t), \mathbf{u}(t), t) = \begin{bmatrix} \mathbf{G}(\mathbf{z}(t), \mathbf{y}(t), \mathbf{u}(t), t) \\ \mathbf{s}(t) - \mathbf{H}(\mathbf{z}(t), \mathbf{y}(t), \mathbf{u}(t), t) \end{bmatrix} = \mathbf{0}$$

mit

$$\tilde{\mathbf{y}}(t) = \begin{bmatrix} \mathbf{y}(t) \\ \mathbf{s}(t) \end{bmatrix}$$

und

$$\begin{bmatrix} \mathbf{y}_{\min}(t) \\ \mathbf{h}_{\min}(t) \end{bmatrix} \leq \tilde{\mathbf{y}}(t) \leq \begin{bmatrix} \mathbf{y}_{\max}(t) \\ \mathbf{h}_{\max}(t) \end{bmatrix}$$

und entsprechend  $\tilde{\mathbf{G}} \in \mathfrak{R}^{NA+NO}$  und  $\tilde{\mathbf{y}} \in \mathfrak{R}^{NA+NO}$ . Weiterhin können die untersuchten Zielfunktionen verschiedenste Charaktere aufweisen. Die optimale Steuerung von technischen Prozessen betrachtet hier den Mayer- und Lagrange-Term. Treten beide Terme gleichzeitig auf, wird dies als Bolza-Funktional bezeichnet. Der Mayer-Term entspricht einem rein algebraischen Ausdruck in Abhängigkeit der Zustände und kann somit berechnet werden. Jedoch stellt der Lagrange-Term ein bestimmtes Integral in Abhängigkeit der Zustände und Steuerungen dar und wird im Allgemeinen dargestellt durch

$$\int_{t_0}^{t_f} L(\mathbf{z}(t), \mathbf{y}(t), \mathbf{u}(t), t) dt.$$

Dies kann über eine zusätzliche Zustandsgleichung entsprechend behandelt werden nach

$$\frac{dl(t)}{dt} = L(\mathbf{z}(t), \mathbf{y}(t), \mathbf{u}(t), t); l(t_0) = 0,$$

$$\tilde{\mathbf{F}}(\tilde{\mathbf{z}}(t), \mathbf{y}(t), \mathbf{u}(t), t) = \begin{bmatrix} \mathbf{F}(\mathbf{z}(t), \mathbf{y}(t), \mathbf{u}(t), t) \\ L(\mathbf{z}(t), \mathbf{y}(t), \mathbf{u}(t), t) \end{bmatrix},$$

mit

$$\tilde{\mathbf{z}}(t) = \begin{bmatrix} \mathbf{z}(t) \\ l(t) \end{bmatrix}; \tilde{\mathbf{z}}(t_0) = \begin{bmatrix} \mathbf{z}_0 \\ 0 \end{bmatrix}$$

und

$$\begin{bmatrix} \mathbf{z}_{\min}(t) \\ -\infty \end{bmatrix} \leq \tilde{\mathbf{z}}(t) \leq \begin{bmatrix} \mathbf{z}_{\max}(t) \\ \infty \end{bmatrix}.$$

Die optimale Steuerung  $\mathbf{u}^*(t)$  basiert entsprechend auf der Minimierung der Zielfunktion unter Einhaltung der Modellgleichungen und Ungleichungsnebenbedingungen (siehe auch Abbildung 3.2). Dies führt dazu, dass  $\mathbf{u}^*(t)$  durch die definierte Zielstellung, durch das zugrunde liegende Modell und durch den erlaubten Freiheitsgrad geprägt ist. Die Berechnung der optimalen Steuerung geschieht offline und es ergibt sich die gezeigte „open loop“ Betriebsart des Systems mit dem gewünschten optimalem Verhalten (veranschaulicht durch  $\mathbf{x}^*(t)$ ). Eine Erweiterung der optimalen Steuerung ist diese repetierend anzuwenden, um so einen „closed loop“-Ansatz zu erhalten (Modell prädikative Regelung). Die Grundlage aller Verfahren ist dabei die Lösung von dynamischen Optimierungsproblemen, für welche im nächsten Kapitel Lösungsmethoden vorgestellt werden.

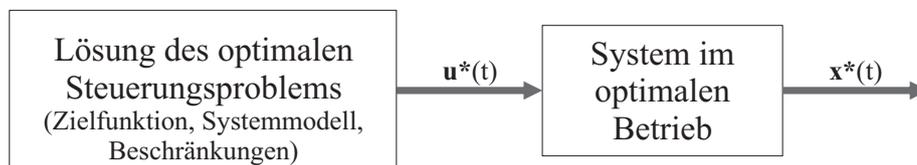


Abbildung 3.2: Wirkungsweise der optimalen Steuerung

## 3.2 Die Lösungsmethoden der dynamischen Optimierung

Dynamische Optimierungsprobleme lassen sich als Aufgabenstellungen in vielen technischen Prozessen wieder finden. Dazu zählen der Betrieb von Batchprozessen, Anfahr- sowie Abschaltvorgänge und verschiedenste Strategien zum Wechseln von Betriebspunkten während eines laufenden Prozesses [141]. Die historische Entwicklung und die breite Anwendungsvielfalt führte zu einer Vielzahl von Ansätzen zur Lösung dieser Optimierungsprobleme, welche sich prinzipiell in indirekte und direkte Methoden unterteilen lassen [106]. Ein Überblick der wichtigsten Lösungsmethoden ist in Abbildung 3.3 gegeben. Zu den indirekten Methoden zählt z.B. die Dynamische Programmierung nach Bellman [16], die bekannte Vorgehensweise

des Maximum- oder Minimumprinzips nach Pontryagin [108]. Die indirekten Methoden sind analytische Lösungsverfahren bei denen Nichtlinearitäten dazu führen, dass die betrachteten Optimierungsprobleme geschlossen nicht lösbar sind. So stellen z.B. im systemtechnischen Sinne die einfachen bilinearen Reaktionskinetiken aus Gleichung (2.1) Nichtlinearitäten dar. Für diese wurde mit hohem Aufwand auch nur eine fast geschlossene Lösung in [10] unter Anwendung des Maximumprinzips erzielt.

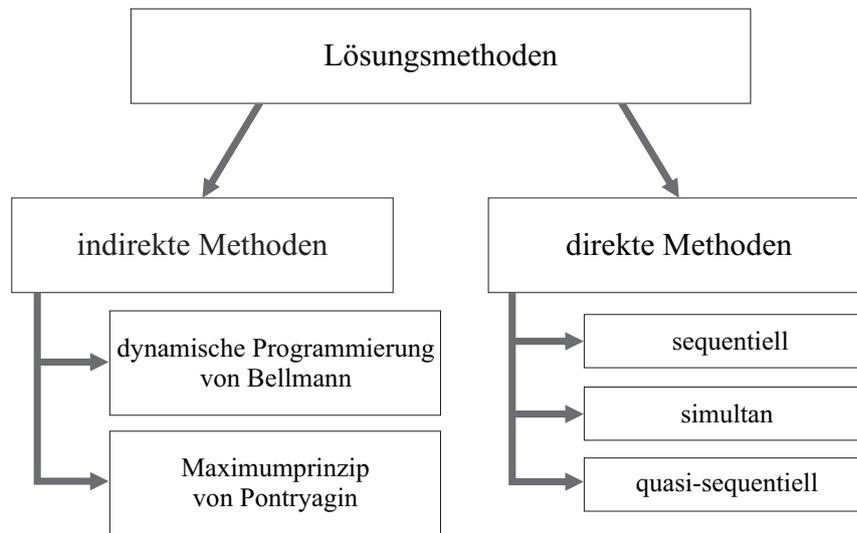


Abbildung 3.3: Übersicht über Lösungsverfahren der dynamischen Optimierung

Die direkten Methoden basieren auf der Transformation kontinuierlicher Optimierungsprobleme unendlicher Dimension aus den Gleichungen (3.1) - (3.6) in Optimierungsprobleme endlicher Dimension. Dies wird auch als Kollokation, Diskretisierung oder Parametrisierung bezeichnet, nach welcher nun Verfahren Anwendung finden, welche prinzipiell durch iterativen Charakter das erzielte endliche Optimierungsproblem lösen. Hierbei werden numerische Lösungsverfahren, wie die nichtlineare Programmierung (NLP) in Form von sequentielle quadratische Programmierung (SQP) (z.B. „sparse nonlinear optimizer“ (SNOPT) [58]) oder die Innere Punkt Methode (IP) (z.B. „interior point optimizer“ (IPOPT) [150]) eingesetzt. Diese numerischen Verfahren finden bei der optimalen Steuerung von Prozessen Anwendung, da diese für hoch-dimensionale und beschränkte Probleme entwickelt worden. Weiterhin benötigen diese Verfahren die Ableitungen der Zielfunktion und aller betrachteten Nebenbedingungen nach den freien Optimierungsvariablen. An dieser Stelle sei [110] genannt, worin ein Überblick der direkten Methoden und die Leistung von NLP-Lösern gegeben wird. Andere Lösungsverfahren sind heuristischer Natur, wie die simulierte Abkühlung, oder entsprechen stochastischen Suchverfahren, wie die evolutionären Algorithmen (für einen Überblick siehe [34]). Diese sind für Probleme mit einer geringen Anzahl von Beschränkungen geeignet und

finden Anwendung bei der Schätzung von Modellparametern. Außerdem bedingt die Verwendung der heuristischen Lösungsverfahren empirische Einstellungen von problemspezifischen Parametern für die Lösung des Optimierungsproblems. In der hier vorliegenden Arbeit werden verschiedenste Probleme mit Beschränkungen betrachtet und die Wahl fällt somit auf das NLP-Verfahren. Der Vorteil der direkten Methoden liegt in der Möglichkeit nichtlineare Problemstellungen untersuchen zu können, weshalb im Folgendem die Wirkungsweisen dieser Methoden vorgestellt werden. Prinzipiell lassen sich die direkten Methoden durch die Behandlung der Steuergrößen und Zustandsvariablen (und somit auch des Prozessmodells) aus Gleichung (3.4),(3.5) und (3.6) unterscheiden.

Die sequentielle Methode basiert auf einer ausschließlichen Parametrisierung der Steuergrößen und einer Simulation des Modells. Dabei wird Simulationssoftware, wie z.B. Matlab/Simulink und Dymola zur Simulation der Modellgleichungen (3.2) und (3.3) verwendet. Um die Wahl eines passenden Algorithmus zu erleichtern, ist z.B. in [45] ein Vergleich der Simulationsalgorithmen für die Merkmale wie Genauigkeit, Geschwindigkeit, Effizienz und Kompaktheit dargestellt. Durch die Vorgabe einer anfänglichen Steuerungsparametrisierung werden nun durch Simulation die Zustandsgrößen berechnet. Die betrachteten freien Variablen entsprechen dabei ausschließlich den Steuerungen. Hier kann durch einen NLP-Löser iterativ die optimale Steuerungsparametrisierung ermittelt werden und es entsteht ein 2-Schichtenansatz aus Optimierungs- und Simulationsschicht. Bekannte Nachteile der sequentiellen Methode sind die Schwierigkeiten bei der Gradientenberechnung und die Behandlung von Beschränkungen der Zustandsverläufe.

Die simultane Methode [19] basiert auf der gleichzeitigen Diskretisierung der Steuergrößen (auch als Steuerungsparametrisierung bezeichnet), Zustandsgrößen und Modellgleichungen durch orthogonale Kollokation auf finiten Elementen [siehe z.B. 47, 145]. Ähnlich sind hierzu die sogenannten Pseudo-Spektrale-Methoden [siehe z.B. 25], welche die orthogonale Kollokation global (nur ein Element) verwenden und in [68] eine Zusammenfassung dargestellt wird. Die diskreten Steuergrößen und Zustandsgrößen sind hier zusätzliche freie Variablen des aufgestellten Optimierungsproblems. Die diskretisierten Modellgleichungen, entsprechen algebraischen Ausdrücken, welche als Gleichungsnebenbedingungen hinzugefügt werden. Dies resultiert in einer Vielzahl von Freiheitsgraden mit Gleichungsnebenbedingungen und ergibt somit ein großskaliges stationäres Optimierungsproblem, welches durch die NLP iterativ gelöst wird (entspricht also einem Einschichtverfahren). Diesbezüglich ist hervorzuheben, dass im Vergleich zur sequentiellen Methode die Gradientenberechnung durch automatische Differentiation des Quellcodes oder durch Operatorüberladung umgesetzt werden kann. Hinzu-

kommt die einfache Handhabung von Beschränkungen der diskretisierten Zustandsverläufe, da diese von der NLP direkt eingehalten werden können. Der Nachteil der simultanen Methode besteht in der hohen Problemdimension, welche entsprechenden Rechenaufwand der iterativen NLP verursacht und die Anzahl von lokalen Optima erhöhen kann.

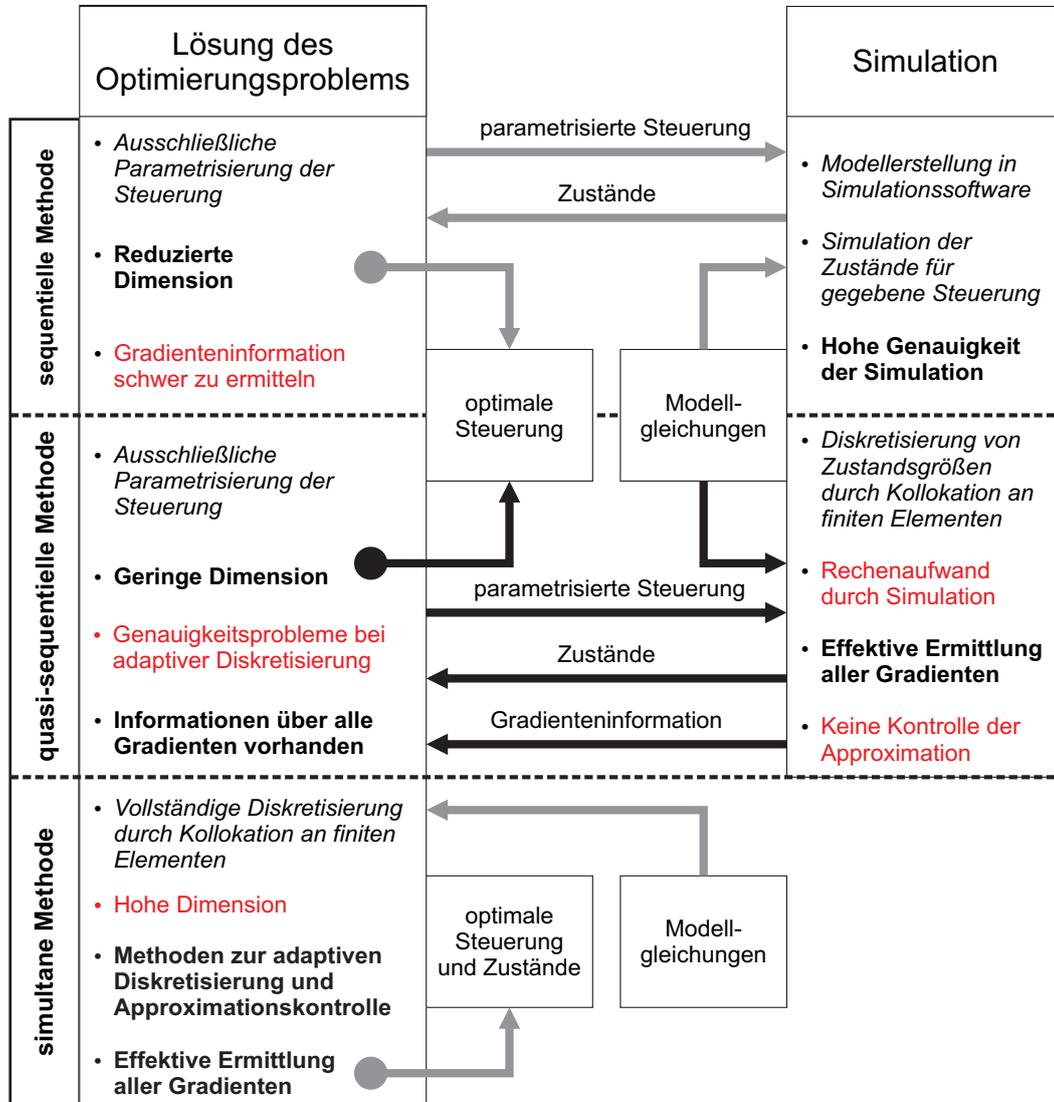


Abbildung 3.4: Direkte Methoden im Vergleich

Die quasi-sequentielle Methode [67, 91, 90] basiert auf der 2-Schichtenstruktur der sequentiellen Methode und der Vorgehensweise bei der Diskretisierung in der simultanen Methode und kann somit als Mischform gesehen werden. Durch eine ausschließliche Steuerungsparametrisierung in der Optimierungsschicht ergibt sich durch die quasi-sequentielle Methode eine Reduzierung der Problemdimension im Vergleich zur simultanen Methode. Die Simulationsschicht basiert auf der Finiten-Element-Kollokation für Zustandsgrößen und Modellgleichungen, wobei das resultierende algebraische Gleichungssystem z.B. durch ein Newton-Verfahren gelöst werden kann. Es ergeben sich die Zustandsgrößen, eine effiziente Berechnung

der Gradienten und die Möglichkeit die Begrenzung von Zustandsverläufen zu behandeln. Der Vereinigung von Vorteilen beider Methoden steht nachteilig der erhöhte Rechenaufwand des Newton-Verfahrens in der Simulationsschicht gegenüber. Dieser Nachteil reduziert sich, wenn das Newton-Verfahren schnell konvergiert und ist somit abhängig vom untersuchten System. Eine schematische Darstellung der diskutierten, direkten Methoden ist in Abbildung 3.4 zusammenfassend veranschaulicht.

Abschließend soll nun die Auswahl einer Methode anhand der obigen Diskussionen erfolgen. Es werden in dieser Arbeit die indirekten Methoden nicht weiter verfolgt, da diese die Untersuchungen von nichtlinearen Systemen erschweren. Wie angedeutet, treten auch Beschränkungen bei den betrachteten Optimierungsproblemen auf, was die Auswahl auf die simultane und die quasi-sequentielle Methode beschränkt. Zudem kann angenommen werden, dass die Regulationen des Stoffwechsels innerhalb eines langen Entwicklungszeitraums an die zu untersuchenden Regulationskriterien angepasst sind. Es gilt entsprechend die globale Lösung der optimalen Regulation zu identifizieren, was durch den reduzierten Suchraum der quasi-sequentiellen Methode erleichtert werden kann. Somit fällt hier die Wahl (nicht nur aus Erfahrungsgründen) auf die quasi-sequentielle Methode.

## 3.3 Die quasi-sequentielle Methode

Die quasi-sequentielle Methode basiert auf einer Diskretisierung der Steuerung und der Zustandsgrößen auf finiten Elementen durch polynomiale Approximation. Eine detaillierte Vorstellung der Methode und alle benötigten Informationen sind bereits in [91, 90] abgeleitet bzw. dargestellt. Um die Notationsform für die vorliegende Arbeit zu definieren, wird im Folgendem die prinzipielle Vorgehensweise aufgezeigt.

### 3.3.1 Die Diskretisierung

Das zu untersuchende dynamische Optimierungsproblem entspricht den Gleichungen (3.1) bis (3.6) und sei hier aus Gründen der besseren Lesbarkeit nochmals dargestellt

$$\min_{\mathbf{u}(t), \mathbf{z}(t), \mathbf{y}(t), t_f} \varphi(\mathbf{z}(t), \mathbf{y}(t), \mathbf{u}(t), t_f) \quad (3.7)$$

$$\frac{d\mathbf{z}(t)}{dt} = \mathbf{F}(\mathbf{z}(t), \mathbf{y}(t), \mathbf{u}(t), t) ; \mathbf{z}_0 = \mathbf{z}(t_0) \quad (3.8)$$

$$\mathbf{G}(\mathbf{z}(t), \mathbf{y}(t), \mathbf{u}(t), t) = \mathbf{0} \quad (3.9)$$

$$\mathbf{u}_{min}(t) \leq \mathbf{u}(t) \leq \mathbf{u}_{max}(t) \quad (3.10)$$

$$\mathbf{z}_{min}(t) \leq \mathbf{z}(t) \leq \mathbf{z}_{max}(t) \quad (3.11)$$

$$\tilde{\mathbf{y}}_{min}(t) \leq \mathbf{y}(t) \leq \tilde{\mathbf{y}}_{max}(t) \quad (3.12)$$

mit  $\mathbf{z}(t) \in \mathfrak{R}^{ND}$ ,  $\mathbf{F} \in \mathfrak{R}^{ND}$ ,  $\mathbf{y}(t) \in \mathfrak{R}^{NA}$ ,  $\mathbf{G} \in \mathfrak{R}^{NA}$  und  $\mathbf{u}(t) \in \mathfrak{R}^{NU}$ . In einem ersten Schritt wird der betrachtete Zeithorizont  $[t_0, t_f]$  in  $NL$  finite Elemente zerlegt und es gilt

$$\Delta t_n = t_{n+1} - t_n ; n = 1, \dots, NL$$

mit  $t_n$  als Startzeitpunkt und  $t_{n+1}$  als Endzeitpunkt des  $n'ten$  Intervalls. Die Approximation der differentiellen Zustandsgrößen wird durch Linearkombination von Lagrange-Polynomen nach

$$z_n^m(t) = \sum_{j=0}^{NC} \ell_j(t) \cdot z_{n,j}^m ; m = 1, \dots, ND ; n = 1, \dots, NL$$

mit

$$\ell_j(t) = \prod_{\substack{i=0 \\ i \neq j}}^{NC} \frac{t - t_{n,i}}{t_{n,j} - t_{n,i}}$$

vorgenommen. Dabei stellt  $NC$  die Anzahl der Kollokationspunkte (bzw. die diskretisierten Werte) eines finiten Elementes,  $z_n^m(t)$  die polynomiale Approximation des  $m'ten$  differentiellen Zustandes im  $n'ten$  Element,  $z_{n,j}^m$  der Wert des  $m'ten$  differentiellen Zustandes am  $j'ten$  Kollokationspunkt und  $t_{n,j}$  der  $j'te$  Kollokationspunkt im  $n'ten$  Element dar (wobei  $t_{n,0}$  und  $z_{n,0}^m$  für die Anfangszeit und die Anfangsbedingung des  $n'ten$  Elementes steht). Folglich gilt für einen differentiellen Zustand an den Kollokationspunkten eines Elementes

$$z_n^m(t_{n,i}) = \sum_{j=0}^{NC} \ell_j(t_{n,i}) \cdot z_{n,j}^m = z_{n,i}^m ; i = 1, \dots, NC \quad (3.13)$$

und durch Differentiation

$$\frac{dz_n^m(t_{n,i})}{dt} = \sum_{j=0}^{NC} \frac{d\ell_j(t_{n,i})}{dt} \cdot z_{n,j}^m ; i = 1, \dots, NC. \quad (3.14)$$

Zudem wird in der quasi-sequentiellen Methode die Kontinuität der differentiellen Zustandsgrößen gewährt durch:

$$z_{n-1,NC}^m = z_{n,0}^m.$$

In dieser Arbeit werden die Steuergrößen, ohne die Alleingültigkeit zu verlieren, in jedem Element als stückweise konstant angenommen. Dadurch lassen sich alle diskretisierten Steuergrößen eines Elementes im Vektor  $\mathbf{U}_n \in \mathfrak{R}^{NU}$  zusammenfassend darstellen. Wird nun ein

### 3 Grundlagen der optimalen Steuerung

normalisierter Zeitbereich  $\tau \in [0, 1]$  für die Diskretisierung betrachtet, ergibt sich für die Gleichung (3.8) durch (3.14) in jedem finiten Element  $\Delta t_n$  die sogenannte Fehlerfunktionen zu

$$R_{n,i}^m(t_{n,i}) = \sum_{j=0}^{NC} \frac{d\ell_j(\tau_i)}{d\tau} \cdot z_{n,j}^m - \Delta t_n \cdot F_m(\mathbf{Z}_{n,i}, \mathbf{Y}_{n,i}, \mathbf{U}_n, t_{n,i}) = 0 \quad (3.15)$$

$$i = 1, \dots, NC; m = 1, \dots, ND; n = 1, \dots, NL.$$

Hierbei steht  $\mathbf{Z}_{n,i} \in \mathfrak{R}^{ND}$  und  $\mathbf{Y}_{n,i} \in \mathfrak{R}^{NA}$  für die Zusammenfassung der differentiellen und algebraischen Zustandswerte im  $n$ 'ten Element und am  $i$ 'ten Kollokationspunkt, wobei  $\mathbf{Y}_{n,i}$  durch Gleichung (3.9) definiert wird. Es kann dann für alle Kollokationspunkte und finiten Elemente

$$\mathbf{U} = [\mathbf{U}_1, \dots, \mathbf{U}_{NL}]^T \in \mathfrak{R}^{NU*NL}$$

$$\mathbf{Z} = [\mathbf{Z}_{1,1}, \dots, \mathbf{Z}_{1,NC}, \dots, \mathbf{Z}_{NL,NC}]^T \in \mathfrak{R}^{NC*ND*NL}$$

$$\mathbf{Y} = [\mathbf{Y}_{1,1}, \dots, \mathbf{Y}_{1,NC}, \dots, \mathbf{Y}_{NL,NC}]^T \in \mathfrak{R}^{NC*NA*NL}$$

als Notation eingeführt werden, wobei es gilt die Gleichungen (3.10), (3.11) und (3.12) einzuhalten. Wird nun die vorgestellte Diskretisierung auf die infinite Problemstellung angewendet, ergeben sich aus Gleichungen (3.7)-(3.12) das folgende diskretisierte Optimierungsproblem

$$\min_{\mathbf{U}, \mathbf{Z}, \mathbf{Y}} \varphi(\mathbf{U}, \mathbf{Z}, \mathbf{Y})$$

$$R_{n,i}^m(t_{n,i}) = 0; z_{n-1,NC}^m = z_{n,0}^m; z_{1,0}^m = z^m(t_o) \quad (3.16)$$

$$\mathbf{G}_n(\mathbf{Z}_{n,i}, \mathbf{Y}_{n,i}, \mathbf{U}_n, t_{n,i}) = \mathbf{0} \quad (3.17)$$

$$i = 1, \dots, NC; m = 1, \dots, ND; n = 1, \dots, NL$$

$$\mathbf{U}_{min} \leq \mathbf{U} \leq \mathbf{U}_{max}; \mathbf{Z}_{min} \leq \mathbf{Z} \leq \mathbf{Z}_{max}; \mathbf{Y}_{min} \leq \mathbf{Y} \leq \mathbf{Y}_{max}.$$

In einem nächsten Schritt werden nun alle differentiellen und algebraischen Zustände eines finiten Elementes in  $\mathbf{X}_n = [\mathbf{Z}_n, \mathbf{Y}_n]^T$  zusammengefasst und zur besseren Darstellung  $\mathbf{X} = [\mathbf{Z}, \mathbf{Y}]^T$  gewählt. Somit können die aus der Diskretisierung der Modellgleichungen entstandenen algebraischen Gleichungen (3.16) und (3.17) durch die Vektorfunktion  $\mathbf{C}_n = \mathbf{0}$  für ein finites Element dargestellt werden und es ergibt sich

$$\min_{\mathbf{U}, \mathbf{X}} \varphi(\mathbf{U}, \mathbf{X})$$

$$\mathbf{C}_n(\mathbf{X}_{n,0}, \mathbf{X}_n, \mathbf{U}_n) = \mathbf{0}; \mathbf{X}_{n,0} = \mathbf{D} \cdot \mathbf{X}_{n-1}; n = 1, \dots, NL \quad (3.18)$$

$$\mathbf{U}_{min} \leq \mathbf{U} \leq \mathbf{U}_{max} ; \mathbf{X}_{min} \leq \mathbf{X} \leq \mathbf{X}_{max}$$

mit  $\mathbf{X}_{n,0}$  als Anfangsbedingung des  $n'$ ten Intervalls und der Projektionsmatrix  $\mathbf{D}$ . Die Projektionsmatrix  $\mathbf{D}$  selektiert die Anfangsbedingungen  $\mathbf{X}_{n,0}$  für die differentiellen Zustände aus der zusammengefassten Darstellung des vorhergehenden Elementes  $\mathbf{X}_{n-1}$ . Die quasi-sequentielle Methode führt an dieser Stelle einen Simulationsschritt durch und berechnet, z.B. mit einem Newton-Verfahren, aus Gleichung (3.18) die diskretisierten Zustandsgrößen in Abhängigkeit der Steuergrößen. Dies zerlegt die Problemstellung in die Ermittlung der optimalen Steuerung mit einer separaten Berechnung der Zustandsgrößen. Folglich ergibt sich das reduzierte Optimierungsproblem in diskreter Form zu

$$\min_{\mathbf{U}} \varphi(\mathbf{U}, \mathbf{X}(\mathbf{U}))$$

$$s.t. \mathbf{U}_{min} \leq \mathbf{U} \leq \mathbf{U}_{max} ; \mathbf{X}_{min} \leq \mathbf{X}(\mathbf{U}) \leq \mathbf{X}_{max}$$

mit  $\mathbf{X}(\mathbf{U}) \in \mathfrak{R}^{NC*(NA+ND)*NL}$  als vektorielle Zusammenfassung aller berechneten differentiellen und algebraischen Zustandsgrößen in Abhängigkeit von  $\mathbf{U}$ . In Abbildung 3.5 wird zusammenfassend das allgemeine Lösungsschema der quasi-sequentuellen Methode dargestellt. Dabei wurde die Verwendung eines NLP-Löser [siehe z.B. 58, 150] veranschaulicht, die Wahl eines anderen oder hybriden Lösungsverfahrens ist jedoch prinzipiell möglich. Hier ist die bereits angesprochene Zerlegung in Optimierungs- und Simulationsschicht ersichtlich. Die notwendigen Gradienten lassen sich effektiv ermitteln und werden im Folgendem aufgezeigt.

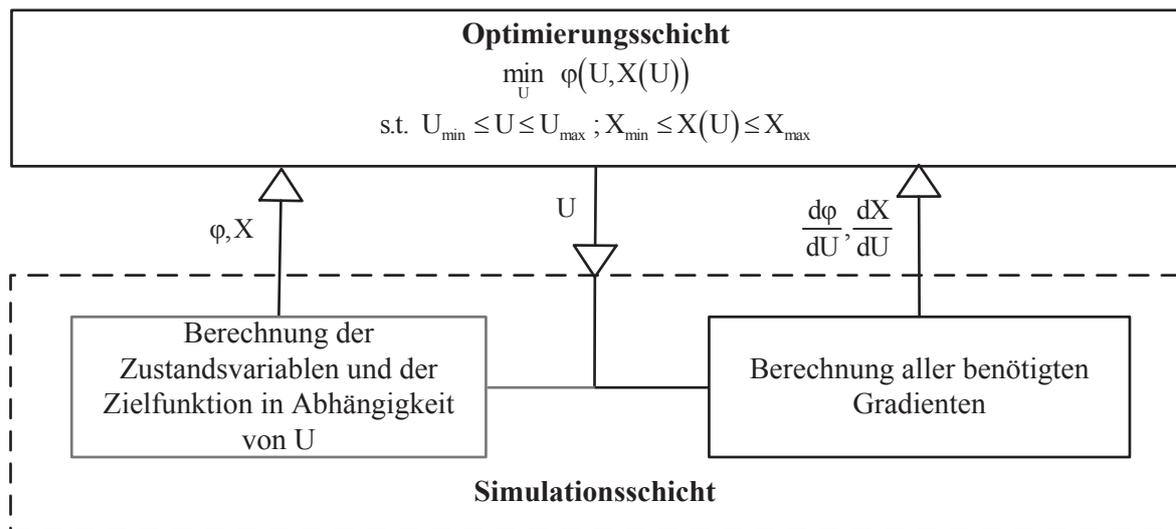


Abbildung 3.5: Lösungsschema der quasi-sequentuellen Methode

### 3.3.2 Die Gradientenberechnung

In der quasi-sequentiellen Methode wird das Gleichungssystem (3.18) von einem finiten Element zum nächsten gelöst. Die benötigten Gradienten nach den Steuergrößen werden parallel zur Lösung der Modellgleichungen ermittelt. Dies wird für jede Iteration des NLP-Lösers durchgeführt [91] und basiert auf dem gelösten Gleichungssystem mit der Taylor-Entwicklung

$$\frac{d\mathbf{C}_n}{d\mathbf{X}_{n,0}} \cdot \Delta\mathbf{X}_{n,0} + \frac{d\mathbf{C}_n}{d\mathbf{X}_n} \cdot \Delta\mathbf{X}_n + \frac{d\mathbf{C}_n}{d\mathbf{U}_n} \cdot \Delta\mathbf{U}_n = \mathbf{0},$$

bzw.

$$\mathbf{K}_n \cdot \Delta\mathbf{X}_{n,0} + \mathbf{L}_n \cdot \Delta\mathbf{X}_n + \mathbf{M}_n \cdot \Delta\mathbf{U}_n = \mathbf{0}.$$

Dementsprechend können die Gradienten der Zustände nach der parametrisierten Steuerung und nach den Anfangsbedingungen eines finiten Elementes berechnet werden durch

$$\mathbf{L}_n \cdot \frac{d\mathbf{X}_n}{d\mathbf{U}_n} = -\mathbf{M}_n \quad (3.19)$$

$$\mathbf{L}_n \cdot \frac{d\mathbf{X}_n}{d\mathbf{X}_{n,0}} = -\mathbf{K}_n. \quad (3.20)$$

Hierbei kann die Matrix  $\mathbf{L}_n$  aus den Newton-Verfahren entnommen oder eine vorliegende LU-Zerlegung verwendet werden. In einem nächsten Schritt ermöglicht die Kontinuitätsbedingung der Zustände die Gradienten von einem Element zum nächsten durch die „chain rule“ [67] zu transformieren. Dadurch werden die benötigten Gradienten der Zustände eines  $n'$ ten Elementes nach der parametrisierten Steuerung der vorhergehenden Elemente berechnet nach

$$\frac{d\mathbf{X}_n}{d\mathbf{U}_i} = \frac{d\mathbf{X}_n}{d\mathbf{D} \cdot \mathbf{X}_{n-1}} \cdot \frac{d\mathbf{D} \cdot \mathbf{X}_{n-1}}{d\mathbf{U}_i} = \frac{d\mathbf{X}_n}{d\mathbf{X}_{n,0}} \frac{d\mathbf{D} \cdot \mathbf{X}_{n-1}}{d\mathbf{U}_i}. \quad (3.21)$$

Für  $i = n$  sind die benötigten Gradienten durch Gleichung (3.19) gegeben und für  $i > n$  liegt keine Abhängigkeit vor (diese Gradienten ergeben sich zu Null). Die Berechnung von (3.20) ist für die Evaluierung von Gleichung (3.21) notwendig. Durch dieses Vorgehen sind alle Gradienteninformationen aus Abbildung 3.5 ermittelt und ergeben sich im Allgemeinen

zu

$$\frac{d\mathbf{X}}{d\mathbf{U}} = \begin{bmatrix} \frac{d\mathbf{X}_1}{d\mathbf{U}_1} & & & & \mathbf{0} \\ \vdots & \ddots & & & \\ \frac{d\mathbf{X}_n}{d\mathbf{U}_1} & \dots & \frac{d\mathbf{X}_n}{d\mathbf{U}_n} & & \\ \vdots & \ddots & \vdots & \ddots & \\ \frac{d\mathbf{X}_{NL}}{d\mathbf{U}_1} & \dots & \frac{d\mathbf{X}_{NL}}{d\mathbf{U}_n} & \dots & \frac{d\mathbf{X}_{NL}}{d\mathbf{U}_{NL}} \end{bmatrix}.$$

# 4 Erweiterungen der quasi-sequentiellen Methode

Die quasi-sequentielle Methode stellt eine effiziente und flexible Strategie zur Lösung von dynamischen Optimierungsproblemen dar. Das prinzipielle Vorgehen wurde bereits in Kapitel 3.3 dargestellt und auch die resultierenden Eigenschaften diskutiert. Dabei wird eine Diskretisierung der Zeit (finite Elemente), der Zustände (orthogonale Kollokation) und der Steuerungen (Parametrisierung der Steuerung) durchgeführt. Da dies letztendlich eine Approximation, des originalen Optimierungsproblems darstellt, können hierbei Fehler entstehen, welche die Qualität der Lösung beeinflussen oder gar die Lösbarkeit verschlechtern. Dies dient hier als Motivation für die angestrebten Erweiterungen der quasi-sequentiellen Methode und wird im Folgendem im Detail diskutiert.

## 4.1 Motivation

Wie einführend dargestellt, kann die Diskretisierung des dynamischen Optimierungsproblems durch Kollokation an finiten Elementen letztendlich eine ungenaue Approximation bzw. nicht exakte Darstellung des kontinuierlichen Optimierungsproblems verursachen. Dies wirkt sich nachteilig auf die Lösung des tatsächlichen dynamischen Optimierungsproblems aus. So beeinflusst z.B. die Approximationsgüte der Zustände bzw. der Modellgleichungen des diskretisierten Optimierungsproblems den Lösungsweg und die Lösungsgenauigkeit des NLP-Verfahrens. Weiterhin beschränkt die Lage der finiten Elemente (die Diskretisierung in zeitliche Abschnitte), die Ausprägung des zeitlichen Verlaufes der Steuergrößen, da für den gesamten Zeitraum eines finiten Elementes die Parametrisierung einer Steuergröße gilt. Dies ergibt somit für größere finite Elemente eine Einschränkung der Veränderbarkeit des Steuerungsprofils im Vergleich zu kleineren finiten Elementen und erschwert bzw. verhindert die Ermittlung von Umschaltzeitpunkten, wie z.B. plötzliche Veränderung der optimalen Steuergrößen in der Bang-Bang-Steuerung, welche typischerweise bei technischen Prozessen und auch in vorherigen Untersuchungen der Stoffwechselregulation [10] auftreten.

Diese beiden Nachteile sind direkt an die Anzahl von finiten Elementen gekoppelt. Durch eine Erhöhung der finiten Elemente wird die Genauigkeit der Approximation von Zuständen (bzw. Modellgleichungen) verbessert und die Veränderbarkeit der Steuerungsprofile positiv beeinflusst. Die Detektion von Umschaltzeitpunkten (im optimalen Steuergrößenverlauf) kann durch eine Vielzahl von finiten Elementen verbessert werden oder durch eine empirische Platzierung der Elemente an die Umschaltzeitpunkte bedient werden.

Auch wenn eine simple Erhöhung der Elementanzahl sich prinzipiell positiv auf die diskutierten Nachteile auswirkt, können hier Situationen entstehen, welche die Lösung des tatsächlichen dynamischen Optimierungsproblems erschweren. Werden z.B. ausschließlich Elemente benötigt um die Approximation der Zustände (bzw. Modellgleichungen) zu gewähren, zieht dies direkt auch eine Erhöhung des diskretisierten Optimierungsproblems nach sich. Stellt diese Erhöhung keine Verbesserung der optimalen Steuerungsprofile dar, dann bedeutet dies letztendlich nur erhöhten, numerischen Aufwand für die Ermittlung der Lösung durch die NLP und entsprechend schlechteres Konvergenzverhalten. Hier kommt erschwerend hinzu, dass die Ermittlung von Umschaltzeitpunkten weiterhin eine hohe Anzahl von finiten Elementen oder eine empirische Elementplatzierung erfordert. Dies gilt hier als Motivation für die Entwicklung eines direkten und effizienten Vorgehens in Bezug auf die Anwendung der quasi-sequentiellen Methode aus Kapitel 3.3.

## 4.2 Approximationskontrolle der Zustände

Da die Diskretisierung der simultanen Methode Anwendung findet, liegt es nahe, aus den bisherigen Erfahrungen dieser Methode zu lernen. Wie nun eine Approximationskontrolle der simultanen Methode in der quasi-sequentiellen Methode Anwendung finden kann, ist Gegenstand dieses Kapitels.

### 4.2.1 Erfahrungen aus der simultanen Methode

Während der Entwicklung der simultanen Methode wurden verschiedenste Verfahren für die Kontrolle des Approximationsfehlers und der Ermittlung von Umschaltvorgängen untersucht. Die einfachsten Untersuchungen galten der Anzahl von Kollokationspunkten und der Anzahl von finiten Elementen [69, 136], was dem Charakter eines empirischen Verfahrens entspricht.

In [36] wurde ein diskretisiertes Optimierungsproblem betrachtet, welches die Elementlängen direkt als Freiheitsgrade beinhaltet. Dies ermöglicht die Detektion von Umschaltvorgängen im optimalen Steuerungsprofil, wobei Approximationsfehler durch zu große Elementlängen mittels einer Gleichverteilungsstrategie des Fehlers [37, 38, 39] berücksichtigt wurden.

Dieser Ansatz wurde in [142] durch zwei Verfahren erweitert, welche sich in der Behandlung von Approximationsfehlern der Zustandsgrößen unterscheiden. Hier wurde im ersten Verfahren die Zielfunktion so modifiziert, dass eine Gleichverteilung des Approximationsfehlers erzielt werden kann. In einem zweiten Verfahren wurden Nebenbedingungen für jedes finite Element eingeführt, welche die Approximationsgüte der Zustandsgrößen definieren lassen (z.B. durch den Benutzer). Dieses zweite Verfahren benötigt ein adaptives Hinzufügen von finiten Elementen, wenn die Genauigkeitsanforderungen nicht erfüllbar sind und somit die Lösung des Optimierungsproblems keinen gültigen Bereich besitzen würde. Dies wurde durch sogenannte „dummy elements“ ermöglicht, welche in das betreffende finite Element eingefügt werden. Dieses Verfahren wurde z.B. erfolgreich von anderen Gruppen auf die optimale Steuerung von Bioreaktoren angewendet [114] und stellt auch das Verfahren mit dem höchsten Anwendungspotential dar. Jedoch hat das Einfügen von zusätzlichen Elementen und dazugehörigen nichtlinearen Nebenbedingungen der Genauigkeitsanforderung eine Erhöhung der Problemdimension zur Folge. Daraus resultieren wiederholte Neustarts des NLP-Lösers, was sich nachteilig auf die benötigte Zeit zum Lösen des Optimierungsproblems auswirkt.

Eine Erweiterung der automatischen Elementplatzierung wurde von [113, 134, 135] aufgezeigt. Dabei wurden zusätzlich Information aus der optimalen Steuerungstheorie der indirekten Verfahren (diskretisierte Darstellung der Hamilton-Funktion) und die Genauigkeitsanforderungen der Zustandsgrößen gleichzeitig bei der Lösung des diskretisierten Optimierungsproblems durch die NLP betrachtet. Das gleichzeitige Lösen des originalen diskretisierten Optimierungsproblems, die Elementplatzierung und die zusätzlichen nichtlinearen Nebenbedingungen (zur Sicherung der Approximation) resultieren letztendlich in einem nur schwer lösbarem nichtlinearem Optimierungsproblem [134, 135, 142]. Aus den obigen Grund wurde in [134, 135] eine Zwei-Schichten-Struktur zur Lösung des Gesamtproblems aufgezeigt, wobei es in jeder Schicht ein Optimierungsproblem zu lösen gilt. Dabei wird die Elementplatzierung und ihre zusätzlichen nichtlinearen Nebenbedingungen in einem äußeren Problem gelöst. Ein inneres Problem löst dann entsprechend das diskretisierte, dynamische Optimierungsproblem. Nachteilig wirkt sich hier aus, dass für jede Iteration des äußeren Problems das innere Problem gelöst werden muss. Hinzu kommt, dass diese Struktur ein nicht-glattes Optimierungsproblem darstellt, welches z.B. durch ein Schichtebenenverfahren gelöst werden muss [134, 135].

In einer neueren Arbeit [20] wurde ein heuristisches Verfahren entwickelt, welches die Differenzierbarkeit der Problemstellung aus [134, 135] ermöglicht und die Konvergenzeigenschaften verbessert. Hierbei ist der Begriff „moving finite elements“ (MFE) geprägt worden und

das Verfahren wurde auch bereits erfolgreich eingesetzt [88]. Jedoch findet stets eine Zwei-Schichten-Struktur Anwendung, wobei jede Schicht ein Optimierungsproblem darstellt und das eigentliche diskretisierte Optimierungsproblem durch nichtlineare Nebenbedingungen erweitert wurde.

### 4.2.2 Varianten der Fehlerabschätzung

In diesem Kapitel werden zwei Varianten dargestellt, wie eine Fehlerabschätzung und Approximationskontrolle für die quasi-sequentielle Methode realisiert werden kann. Die Variante 1 stellt ein Verfahren dar, das schon in der simultanen Methode verwendet wurde und die Variante 2 entspricht einer empirisch ermittelten Approximationskontrolle.

#### Ein Verfahren zur direkten Fehlerabschätzung (Variante 1)

Die Schätzung des polynomialen Approximationsfehlers an Gausspunkten (oder Kollokationspunkt) wurde erstmals von De Boor und Schwartz [39] untersucht. Daraufhin wurde in [37] eine Strategie entwickelt, welche den Approximationsfehler über die finiten Elemente gleich verteilt, für welche in [38] die Berechnung des entsprechenden Grids (die Lage der finiten Elemente) durchgeführt wurde. Eine Übersicht von Fehlerabschätzungsverfahren für die Kollokation auf finiten Elementen und deren entsprechende Platzierung, wurde in [32, 117] dargestellt. Basierend auf [117] und [6] wurde in [142] angewendet, dass für eine ausreichend glatte Funktion der absolute Fehler der diskretisierten, differentiellen Zustandsgleichungen abgeschätzt werden kann, durch

$$\left| c_{nonc} \cdot R_{n,nonc}^m(t_{n,nonc}) \right| \leq \varepsilon_{n,nonc}^m ; m = 1, \dots, ND ; n = 1, \dots, NL, \quad (4.1)$$

wobei  $R_{n,nonc}^m(t_{n,nonc})$  der Wert der Fehlerfunktion (3.15) für die  $m$ 'te differentielle approximierte Zustandsgröße (3.14) an einem Nicht-Kollokationspunkt  $t_{n,nonc}$  (veranschaulicht durch  $nonc$ ) des  $n$ 'ten Elementes darstellt. Es ergibt sich somit

$$R_{n,nonc}^m(t_{n,nonc}) = \sum_{j=0}^{NC} \frac{dl_j(\tau_{nonc})}{d\tau} \cdot z_{n,j}^m - \Delta t_n \cdot F_m(\mathbf{Z}_{n,nonc}, \mathbf{Y}_{n,nonc}, \mathbf{U}_{n,nonc}, t_{n,nonc})$$

mit  $\mathbf{Z}_{n,nonc} \in \mathfrak{R}^{ND}$  und  $\mathbf{Y}_{n,nonc} \in \mathfrak{R}^{NA}$  als approximierte Werte der differentiellen und algebraischen Zustandsgrößen am Nicht-Kollokationspunkt. Hierbei wird der Nicht-Kollokationspunkt  $\tau_{nonc}$  im normierten Intervall  $\tau = [0, 1]$  von  $t_{n,nonc}$  betrachtet. Die Steuergröße fließt ebenfalls am Nicht-Kollokationspunkt  $\mathbf{U}_{n,nonc}$  in die Berechnung mit ein, bei stückweise konstanter Steuerungsparametrisierung gilt  $\mathbf{U}_{n,nonc} = \mathbf{U}_n$ . Die approximierten Zustandsgrößen

$\mathbf{Z}_{n,nonc}$  können durch die Interpolation aus Gleichung (3.13) am Nicht-Kollokationspunkt berechnet werden nach

$$z_n^m(t_{n,nonc}) = \sum_{j=0}^{NC} \ell_j(t_{n,nonc}) \cdot z_{n,j}^m; m = 1, \dots, ND. \quad (4.2)$$

Die benötigten algebraischen Zustandswerte ergeben sich aus Umformulierung von Gleichung (3.17) zu

$$\mathbf{G}_n(\mathbf{Z}_{n,nonc}, \mathbf{Y}_{n,nonc}, \mathbf{U}_{n,nonc}, t_{n,nonc}) = \mathbf{0}$$

und können durch Lösung dieser umformulierten Gleichung entsprechend berechnet werden. Die noch benötigten Konstanten  $c_{nonc}$  auf Gleichung (4.1) sind abhängig von der Wahl von  $\tau_{nonc}$  und der Kollokationsordnung  $NC$ :

$$c_{nonc} = \frac{1}{a_{nonc}} \int_0^{\tau_{nonc}} \left( \prod_{i=1}^{NC} (s - \tau_i) \right) ds \quad (4.3)$$

mit

$$a_{nonc} = \prod_{i=1}^{NC} (\tau_{nonc} - \tau_i).$$

Für eine exakte Kontrolle des Approximationsfehlers über ein gesamtes Element wäre eine hohe Abtastrate zwischen den Kollokationspunkten nötig. Dies resultiert in einer hohen Anzahl der Evaluierungen von Gleichung (4.1) und die dafür notwendigen Berechnungen für jeden Nicht-Kollokationspunkt und jede Zustandsgröße. Dieses Vorgehen ist rechenintensiv und es wird in [20] vorgeschlagen die Approximation an nur wenigen Nicht-Kollokationspunkten zu überprüfen. Doch es gibt diesbezüglich in der Literatur keine konkreten Vorschläge für die Wahl der Nicht-Kollokationspunkte. Aus diesem Grund wird untersucht, ob sich prinzipielle Richtlinien für diese Wahl durch ein Beispiel ableiten lassen.

### Fallbeispiel (zu Variante 1)

Die Approximation wird durch fein gerasterte Nicht-Kollokationspunkte (in 0.011 Schritten) für  $NC = 3$  anhand einer einfachen differentiellen Gleichung (DGL) der Form

$$\dot{x}(t) = u \cdot x(t); x(t = t_0) = x_0 \quad (4.4)$$

untersucht, welche folgende analytische Lösung besitzt

$$x(t) = x_0 \cdot e^{u \cdot t}. \quad (4.5)$$

#### 4 Erweiterungen der quasi-sequentiellen Methode

Für die angestrebten Untersuchungen sei beispielhaft,  $x_0 = 1$ ,  $t_0 = 0s$  und der Integrationszeitraum zu  $t_f = 10s$  gewählt. Um die Charakteristika von fallenden und steigenden Funktionen zu beachten, wurde  $u = \mp 0,25s^{-1}$  angenommen.

Abbildung 4.1 stellt das Verhalten des Approximationsfehlers über den Integrationszeitraum dar. Hierfür wurden drei finite Elemente verwendet, um  $t_f$  in drei gleiche Abschnitte zu unterteilen. Wie in den Grafiken für die analytische Lösung von  $x(t)$  und deren diskretisierter Approximation zu erkennen ist, sind diese fast identisch, was hier zusätzlich die Leistungsfähigkeit der orthogonalen Kollokation zeigt. Um hier weitere Aussagen treffen zu können wird jeweils der geschätzte Approximationsfehler mit dem bekannten Fehler der polynomialen Approximation nach Gleichung (3.13) zu der analytischen Lösung untersucht (dabei werden relative Fehler betrachtet). Für den exponentiellen Zerfall (zweite Grafik von oben mit  $u = -0,25s^{-1}$ ) ergibt sich für die Umgebung eines Kollokationspunktes eine Überschätzung des Approximationsfehlers, wobei sich für die anderen Bereiche gute Näherungen ergeben. Werden die einzelnen finiten Elemente betrachtet, ergibt sich eine Tendenz zur Verschlechterung der Approximation zum Ende eines Intervalls. Für  $u = +0,25s^{-1}$  (zweite Grafik von unten) ergibt sich, abgesehen von der typische Fehlerfortpflanzung bei der Verwendung von Integrationsverfahren, in allen Bereichen eine gute Abschätzung des Approximationsfehlers. Werden an dieser Stelle wieder die einzelnen finiten Elemente betrachtet, ergibt sich eine Tendenz zur Verschlechterung der Approximation zum Anfang eines Intervalls.

Zusammenfassend lässt sich die Wahl auf Bereiche zwischen den Kollokationspunkten eingrenzen. Durch die verschiedenartige Fehlerentwicklung über die Intervalle, lassen sich an dieser Stelle keine Aussagen bezüglich des Endes oder des Anfangs eines finiten Elementes treffen. Abgesehen davon, dass sich noch weitere unzählige Lösungsverläufe (nicht nur exponentielle) für nichtlineare DGLs oder DEAs ergeben, muss hier doch eine Festlegung für die Wahl der Nicht-Kollokationspunkte getroffen werden. Es ist zu beachten, dass Gleichung (4.3) sich nicht zu Null ergibt und die Fehlerabschätzung somit verfälscht wird. Um nun den aufgezeigten Eigenschaften zu entsprechen, werden in dieser Arbeit die Mittelpunkte zwischen den Kollokationspunkten betrachtet. Somit ergeben sich für ein Kollokationsverfahren der Ordnung  $NC = 3$  auch drei Nicht-Kollokationspunkte. Die hierfür notwendigen Parameter sind zur Reproduktion in Tabelle A.1 des Anhangs aufgelistet, wobei die benötigten Interpolationsparameter an den Nicht-Kollokationspunkten in Tabelle A.2 zusammengefasst sind. Diese Parameter sind notwendig für die Evaluierung von Gleichung (4.1).

Die Wirkungsweise der Fehlerabschätzung soll nun im Folgendem kurz veranschaulicht werden. Dazu wird das obige Beispiel durch eine gröbere Approximation erneut betrachtet,

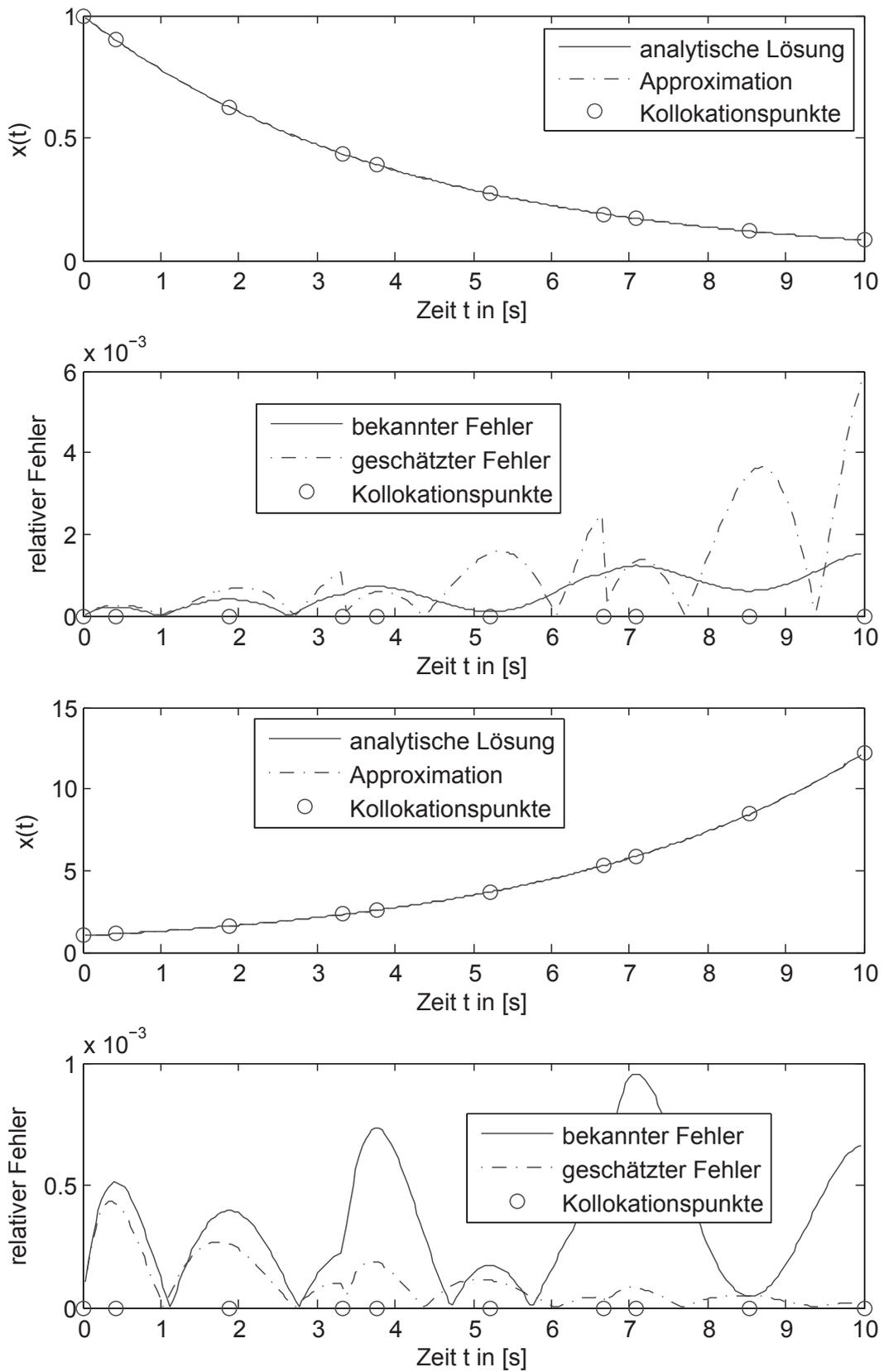


Abbildung 4.1: Analyse des Approximationsfehlers bei ausreichender Approximation (oben:  $u = -0,25s^{-1}$ , unten:  $u = +0,25s^{-1}$ ).

wobei für die Integration nur ein finites Element gewählt wird (die Kollokationspunkte werden entsprechend auf  $t_f = 10s$  verschoben). Die hierbei entstehenden Approximationsfehler und deren Schätzung sind in Abbildung 4.2 dargestellt. Die erzielten Genauigkeiten der Approximation fallen im Vergleich zu der Wahl von drei finiten Elementen im Allgemeinen schlecht aus und die Genauigkeit ist für eine Simulation eher nicht ausreichend. Jedoch stellt die implementierte Fehlerabschätzung hierbei eine erfolgreiche Detektion des Approximationsfehlers dar, anhand welcher eine Fehlerbehandlung erfolgen kann.

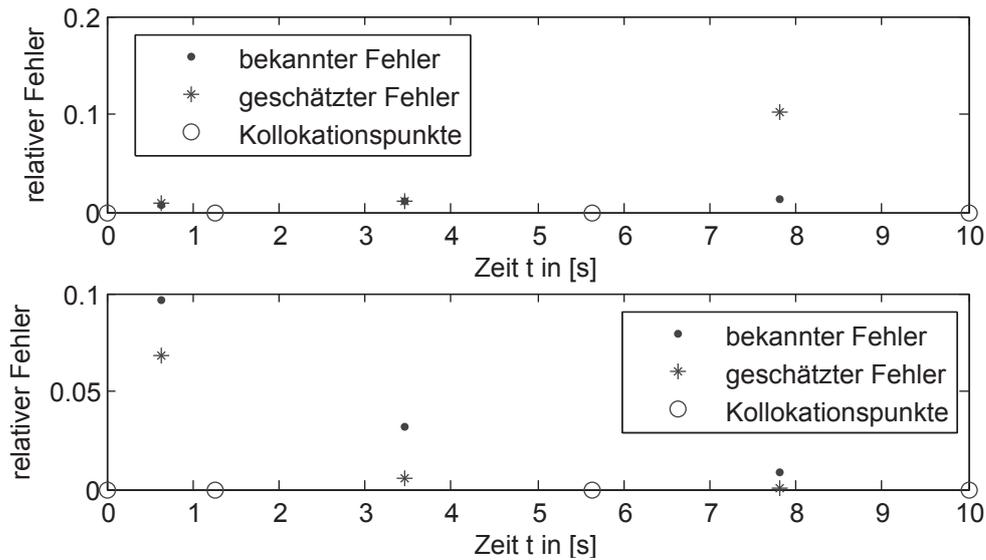


Abbildung 4.2: Analyse des Approximationsfehlers bei unzureichender Approximation (oben:  $u = -0,25s^{-1}$ , unten:  $u = +0,25s^{-1}$ ).

### Ein empirisches Verfahren zur Fehlerdetektion (Variante 2)

Wie aus der obigen Vorgehensweise entnommen werden kann, stellt das Verfahren zur direkten Fehlerabschätzung der Approximation in den Zuständen einen erhöhten Rechenaufwand dar. Dies betrifft die Evaluierung des Approximationsfehlers an hier gewählten drei Nicht-Kollokationspunkten für jedes Intervall und jeden differentiellen Zustand. Hinzu kommt die Ermittlung der algebraischen Zustandsgrößen an den Nicht-Kollokationspunkten.

Es gilt nun ein Verfahren zu entwickeln, welches nur bereits vorhandene Informationen verarbeitet, um eine Aussage über die Güte der polynomialen Approximation treffen zu können. Im Vordergrund steht hierbei den rechenintensiven Aufwand möglichst klein zu halten. In [32, 117] wurde der Approximationsfehler durch höhere zeitliche Ableitungen abgeschätzt. Während der Lösung des diskretisierten Optimierungsproblems sind jedoch nur die Information des NLP-Lösers (Steuerungsparametrisierung) und der Simulationsschicht (Zustandsgrößen mit erster zeitlicher Ableitung) ohne weiteren Rechenaufwand zugänglich.

Die Kollokation basiert auf der Gauß-Quadratur, für welche bekannt ist, dass eine exakte Approximation der Integration erreicht wird, wenn die zu integrierende Funktion eine polynomiale Ordnung nicht überschreitet (hier maximal  $2NC$ ). Eine Abschätzung dieser Ordnung, ist mit den obigen Informationen nicht möglich und soll auch in dieser Arbeit nicht vorgenommen werden.

Das hier entwickelte **empirische** Verfahren basiert auf der Untersuchung der ersten zeitlichen Ableitungen, um eine Detektion von Nichtlinearitäten entlang eines Zustandsverlaufes zu ermöglichen, da nur diese aus der Simulationsschicht bekannt sind.

Im Folgendem werden zur vereinfachten Darstellung vorerst benachbarte Kollokationspunkte ( $t_1$  und  $t_2$ ) und ein Zustand  $z(t)$  betrachtet. Der Grundgedanke basiert auf dem Verhalten einer konstanten zeitlichen Ableitung, die in einem linearen zeitlichen Verlauf resultiert und entsprechend problemlos approximiert werden kann. Anhand dieser Überlegungen wird folgendes Testkriterium abgeleitet

$$\alpha = \left| |\dot{z}(t_1)| - |\dot{z}(t_2)| \right|,$$

dass sich für einen linearen zeitlichen Verlauf zu Null ergibt. Ist  $\alpha \neq 0$  liegt ein Profil mit zeitlich nichtlinearem Verhalten vor.

Die Änderungen der zeitlichen Ableitung stellen somit verschiedenste  $\alpha$  dar, welche wiederum einen Hinweis auf die Stärke einer Nichtlinearität geben. Um diese Werte im Intervall  $[0, 1]$  zu beschreiben, wird eine Normierung von  $\alpha$  nach

$$\alpha_{rel} = \begin{cases} \frac{\alpha}{\max\{|\dot{z}(t_1)|, |\dot{z}(t_2)|\}}, & \text{wenn } \max\{|\dot{z}(t_1)|, |\dot{z}(t_2)|\} \neq 0 \\ \alpha, & \text{sonst} \end{cases}$$

vorgenommen. Dies stellt eine Distanz zu dem problemlos approximierbaren linearen Verhalten dar, wobei ein erlaubter Bereich nach

$$\alpha_{rel} \leq tol$$

untersucht wird. Dieser Bereich definiert entsprechend ein ausreichend genaue Approximation der Zustände und kann durch den Anwender gewählt werden.

Da dem Kollokationsverfahren Information zu mehreren Punkten ( $NC + 1$ ) in einem finiten Element vorliegen, kann die Krümmung eines Zeitverlaufes als Hinweis auf Nichtlinearitäten untersucht werden. Dafür ergibt sich die relative Änderung der Zustände (an den Kollokati-

onspunkten) bezüglich der Anfangsbedingungen eines Intervalls als Testkriterium

$$\alpha_{n,i}^m = \left\{ \begin{array}{l} \frac{||\dot{z}_{n,0}^m| - |\dot{z}_{n,i}^m||}{\max\{|\dot{z}_{n,0}^m|, |\dot{z}_{n,i}^m|\}}, \quad \text{wenn } \max\{|\dot{z}_{n,0}^m|, |\dot{z}_{n,i}^m|\} \neq 0 \\ ||\dot{z}_{n,0}^m| - |\dot{z}_{n,i}^m||, \quad \text{sonst} \end{array} \right\} \leq tol_i^m \quad (4.6)$$

$$i = 1, \dots, NC ; m = 1, \dots, ND ; n = 1, \dots, NL$$

wobei  $\alpha_{n,i}^m$  für das  $i$ 'te Kriterium des  $m$ 'ten differentiellen Zustandes im  $n$ 'ten Element,  $\dot{z}_{n,i}^m$  für die erste Ableitung am  $k$ 'ten Kollokationspunkt des  $m$ 'ten differentiellen Zustandes im  $n$ 'ten Element und  $tol_i^m$  für die  $i$ 'te erlaubte Toleranz der  $m$ 'ten differentiellen Zustandsgröße jeweils steht.

Durch eine Vorgabe von  $tol_i^m$  aus Gleichung (4.6) kann entsprechend durch den Anwender eine erlaubte Krümmung vorgegeben werden. Dieses Kriterium ist empirisch abgeleitet worden und es existiert (zumindest bisher) kein direkter, mathematischer Zusammenhang bezüglich des vorliegenden Approximationsfehlers und dem Wert von  $tol_i^m$ . Doch kann die Wahl der anzugebenden Toleranzen durch eine Simulation für ein konkretes Beispiel unterstützt und an einem Fallbeispiel verdeutlicht werden.

### Fallbeispiel (zu Variante 2)

Zur Veranschaulichung befinden sich in Abbildung 4.3 und 4.4 Simulationsergebnisse für die Betrachtungen aus Gleichung (4.4) mit  $u = \pm 0.25$  und die dazugehörigen Approximationsfehler. Hierbei wurde die Anzahl von finiten Elementen stückweise reduziert und die entsprechend unterschiedliche Genauigkeiten für die Approximation erreicht. Wird beispielsweise die Approximation mit zwei finiten Elementen (d.h.  $NL = 2$ ) als ausreichend betrachtet, können die dabei vorliegenden Werte von Gleichung (4.6) für die Wahl hilfreich sein. Dabei ergeben sich für  $u = -0.25s^{-1}$  die Kriterien in Reihenfolge der Kollokationspunkte zu  $\alpha_- = [0, 146 \ 0, 507 \ 0, 713]$  und für  $u = +0.25s^{-1}$  zu  $\alpha_+ = [0, 149 \ 0, 505 \ 0, 714]$ . Hieraus kann nun für beide DGL eine Wahl der Toleranz zu  $tol = [0, 15 \ 0, 51 \ 0, 72]$  getroffen werden. Wird diese Fehlerdetektion an den übrigen Beispielen untersucht, ergeben sich für  $NL = 1$  mit  $\alpha_+ = [0, 35 \ 0, 75 \ 0, 92]$  und  $\alpha_- = [0, 27 \ 0, 76 \ 0, 91]$  und für  $NL = 3$  mit  $\alpha_{\pm} = [0, 10 \ 0, 37 \ 0, 56]$  die entsprechende Hinweise zur Wahl der Toleranz.

Wird die Einhaltung eines konkreten, relativen Fehlers beabsichtigt, können die Toleranzen aus Simulationsszenarien bestimmt werden. Die hierbei ermittelte Toleranz ergibt sich z.B. für einen relativen Fehler im Bereich von  $1 \cdot 10^{-4}$  zu  $tol = [0, 15 \ 0, 44 \ 0, 71]$ , für welche  $NL = 3, 2, 1$  als fehlerhafte Approximationen richtig detektiert werden. Die dabei zu berechnenden

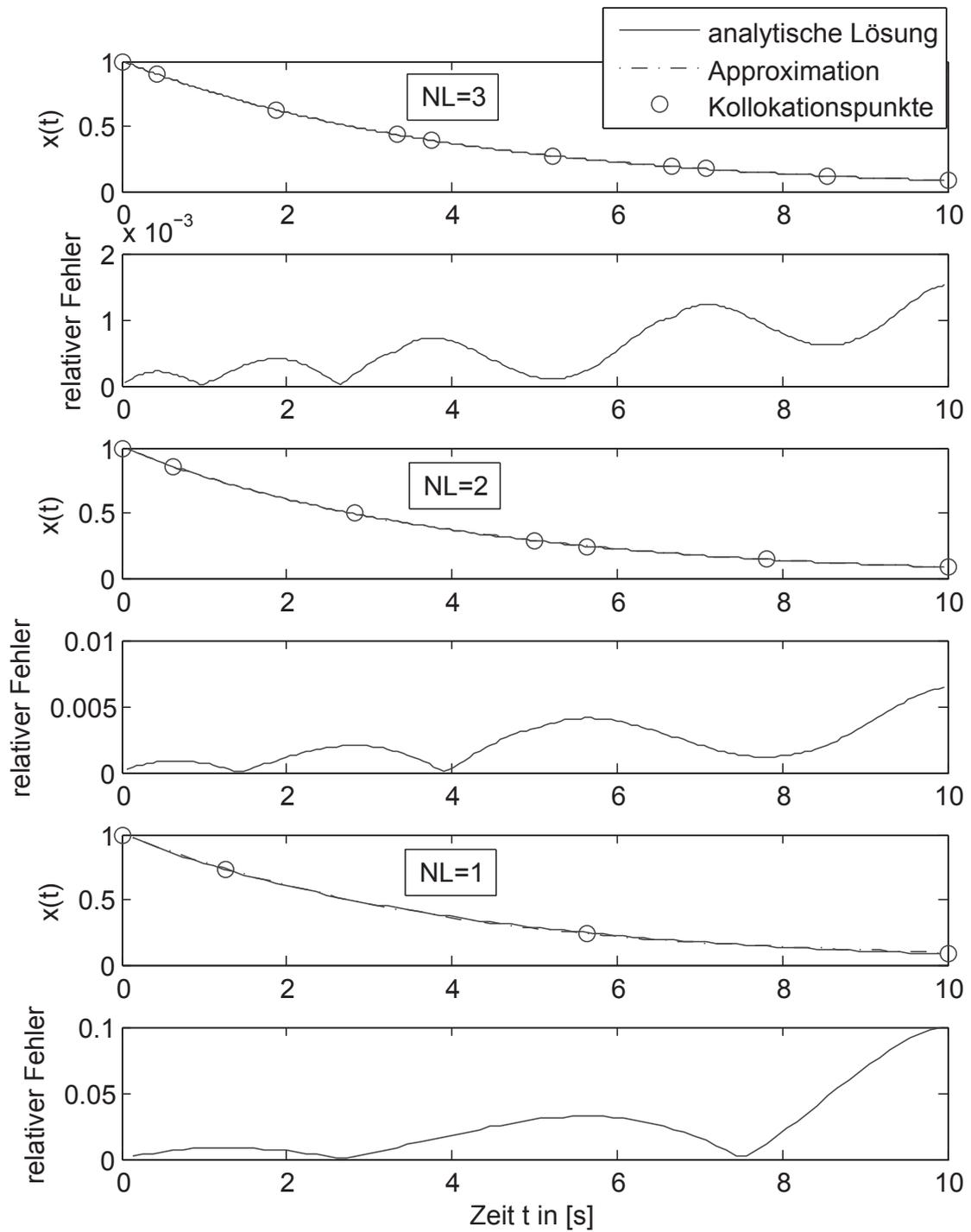


Abbildung 4.3: Approximation in Abhängigkeit der finiten Elemente ( $u = -0,25s^{-1}$ ). Dargestellt werden die relativen Fehler zur analytischen Lösung.

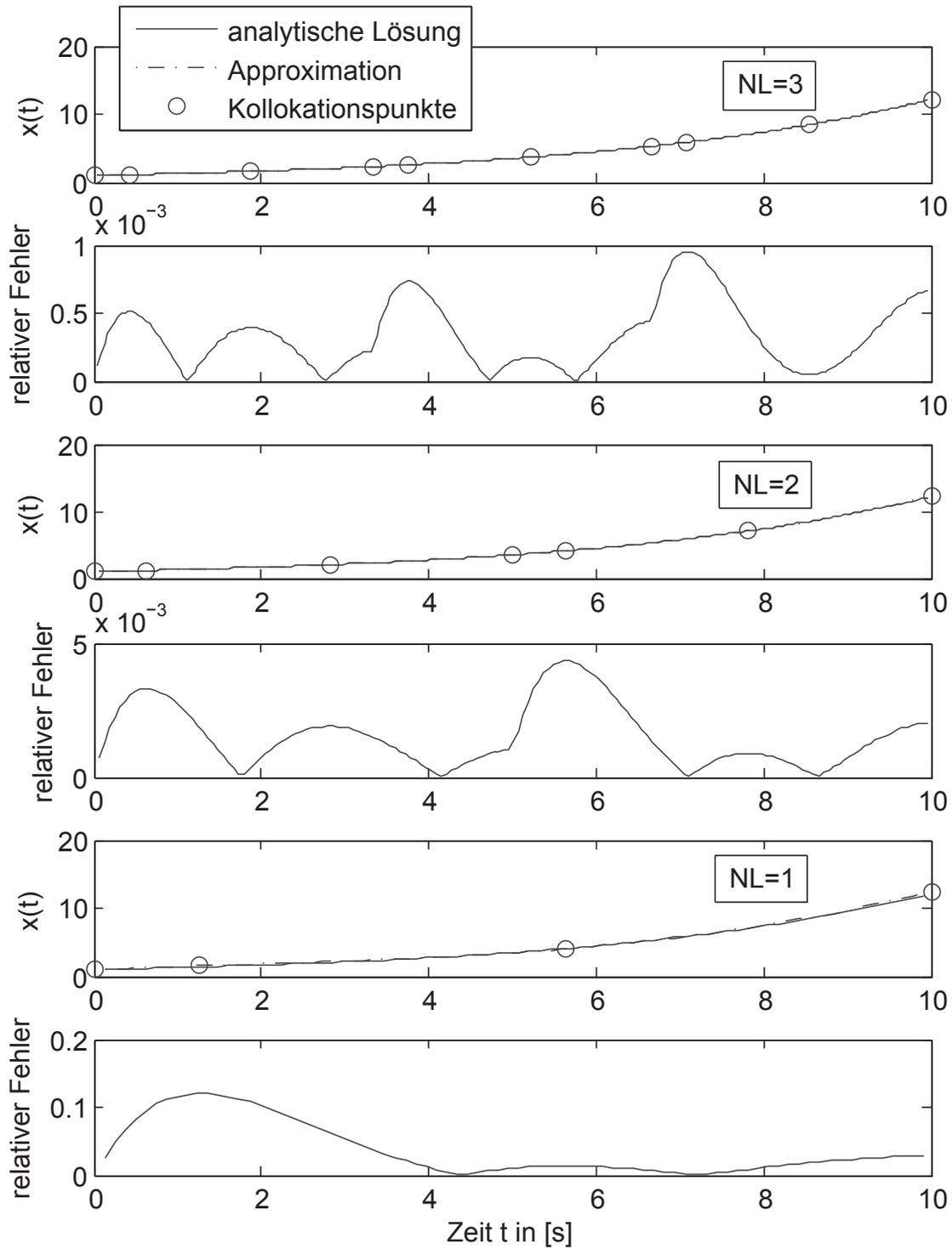


Abbildung 4.4: Approximation in Abhängigkeit der finiten Elemente ( $u = +0,25s^{-1}$ ). Dargestellt werden die relativen Fehler zur analytischen Lösung.

Approximationsfehler können einfach durch eine Simulationssoftware oder durch Kollokation mit einer hohen Anzahl von finiten Elementen, oder hier auch aus der Lösung der DGL (diese liegt im Allgemeinen nicht vor) oder aus der oben vorgestellten direkten Fehlerabschätzung (Variante 1) ermittelt werden.

### 4.2.3 Approximationskontrolle in der quasi-sequentiellen Methode

In diesem Kapitel wird die konkrete Erweiterung der quasi-sequentiellen Methode unter Beachtung der Approximationsgüte in den Zuständen vorgestellt. Das zu lösende Optimierungsproblem lässt sich wie folgt in Abhängigkeit der Approximationskontrolle formulieren:

$$\min_{\mathbf{U}} \varphi(\mathbf{U}, \mathbf{X}(\mathbf{U}))$$

$$s.t. \Psi_{rel}\{\mathbf{X}(\mathbf{U}), \varepsilon_{rel}\} \leq \mathbf{0} \quad (4.7)$$

$$\mathbf{U}_{min} \leq \mathbf{U} \leq \mathbf{U}_{max} ; \mathbf{X}_{min} \leq \mathbf{X}(\mathbf{U}) \leq \mathbf{X}_{max}.$$

Dabei stellt  $\Psi_{rel}$  einen Operator dar, welcher den Approximationsfehler auswertet. Wird die direkte Fehlerabschätzung an Nicht-Kollokationspunkten gewählt (Variante 1 aus Kapitel 4.2.2), evaluiert  $\Psi_{rel}$  Gleichung (4.1) und ermittelt den relativen Fehler  $\varepsilon_{rel}$  unter Anwendung der Interpolation aus (4.2). Wird die Analyse der zeitlichen Ableitung (Variante 2 aus Kapitel 4.2.2) gewählt, berechnet  $\Psi_{rel}$  die Krümmung der Zustandsverläufe nach Gleichung (4.6) und vergleicht diese mit den erlaubten vorgegebenen Toleranzen. Die Behandlung der Approximationsgenauigkeit kann prinzipiell durch die vorliegende Zweischichtenstruktur der quasi-sequentiellen Methode (siehe Abbildung 3.5) in jeder Schicht gelöst werden.

Wird eine Erweiterung des Optimierungsproblems vorgenommen, um die Approximation sicher zu stellen, ergeben sich zusätzliche nichtlineare Nebenbedingungen für die Einhaltung der Approximationsgenauigkeit in der Optimierungsschicht nach Gleichung (4.7). Dieses Vorgehen ist in Abbildung 4.5 dargestellt und würde zudem zusätzliche Verfahren benötigen, um die Approximation im Falle einer Ungenauigkeit während der Lösung wieder herzustellen. Zum Einen betrifft dies entsprechend eine gezielte Längen Anpassung der finiten Elemente und zum Anderen das Einfügen von Elementen, wenn die Längen Anpassungen nicht ausreichend wären. Wie schon aufgezeigt, werden ähnliche Problemstellung in der simultanen Methode durch eine aufwendige Zwei-Schichten-Optimierung gelöst [134, 135, 20], in welcher sich die wiederholte Lösung des inneren Optimierungsproblems als nachteilig herausstellt [20]. Zusammenfassend ergibt dies ein schwer zu lösendes Optimierungsproblem mit zusätzlichen nichtlinearen Nebenbedingungen, welche durch verschiedenste Verfahren eingehalten werden.

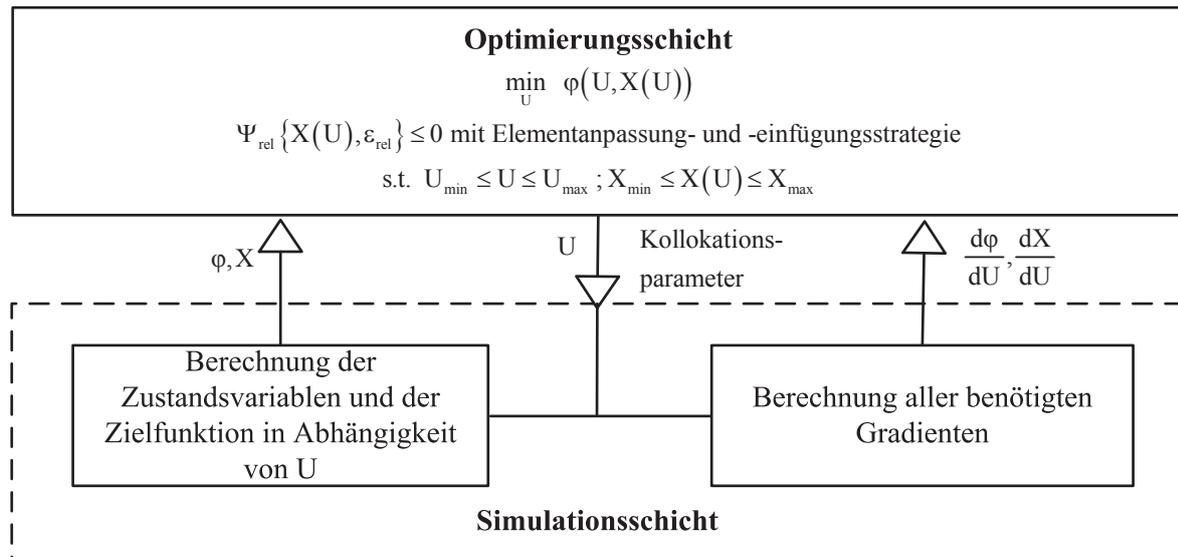


Abbildung 4.5: Approximationskontrolle der Zustände in der Optimierungsschicht

Somit ergeben sich für die Approximationskontrolle der Zustände in der Optimierungsschicht, unter Anwendung der quasi-sequentiiellen Methode, keine Vorteile der bereits vorhandene Zwei-Schichten-Struktur. Wird die Simulationsschicht der quasi-sequentiiellen Methode als Möglichkeit genutzt um die Approximation der Zustände sicherzustellen, reduziert sich das betrachtete Optimierungsproblem wieder zu

$$\begin{aligned} \min_{\mathbf{U}} \varphi(\mathbf{U}, \mathbf{X}(\mathbf{U})) & \tag{4.8} \\ \text{s.t. } \mathbf{U}_{min} \leq \mathbf{U} \leq \mathbf{U}_{max} ; \mathbf{X}_{min} \leq \mathbf{X}(\mathbf{U}) \leq \mathbf{X}_{max}. \end{aligned}$$

Hierbei ist das Optimierungsproblem in Verbindung mit dem dazugehörigen erweiterten Lösungsschema aus Abbildung 4.6 zu sehen. Nach der Simulation eines finiten Elementes (Lösung des algebraischen Gleichungssystems aus der Diskretisierung) wird eine Fehlerbehandlung durchgeführt. Dabei werden die Genauigkeitsanforderungen je nach gewähltem Verfahren durch den Operator  $\Psi_{rel}$  überprüft.

Besteht eine zu hohe Approximationsungenauigkeit werden Unterintervalle eingefügt, um von einem Kollokationspunkt des originalen finiten Elementes zum nächsten Kollokationspunkt zu simulieren. Dabei wird wieder entsprechend die Approximation überprüft und eventuelle weitere Untersimulationen eingefügt. Die notwendigen Untersimulationen werden ebenfalls durch orthogonale Kollokation realisiert und fortgeführt bis der letzte Kollokationspunkt erreicht wird. Ist die Approximation in einem finiten Element ausreichend genau, wird entsprechend keine Untersimulation durchgeführt.

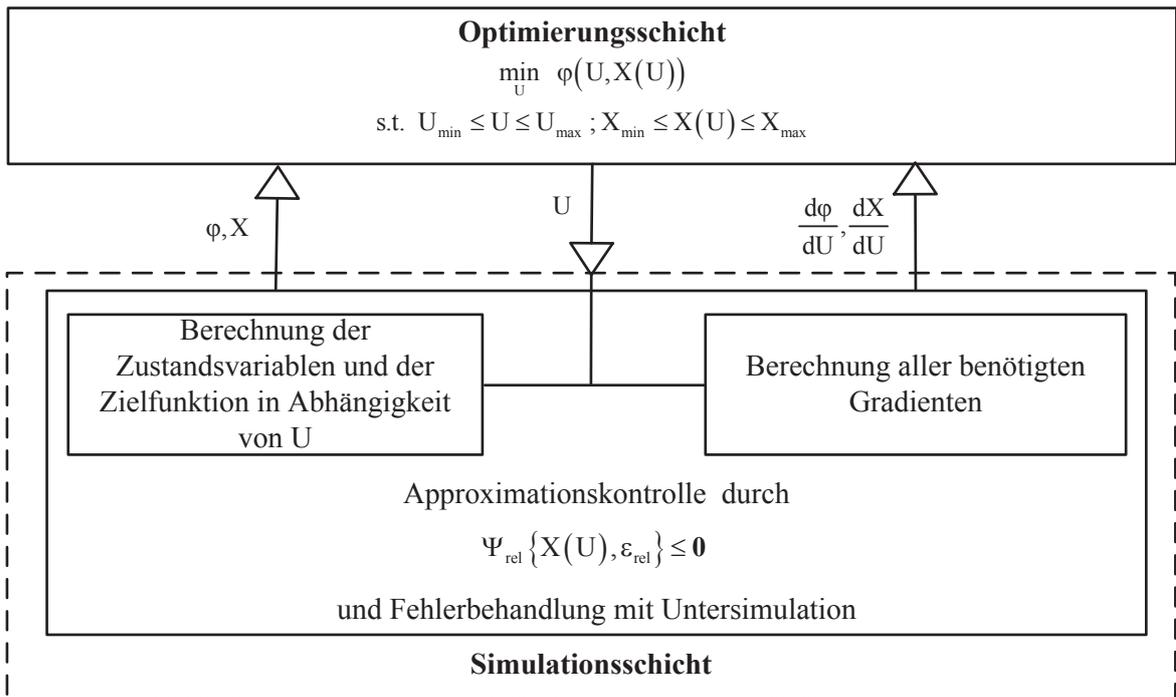


Abbildung 4.6: Approximationskontrolle der Zustände in der quasi-sequentiellen Methode

Zur Veranschaulichung einer Untersimulation dient Abbildung 4.7. Die Kollokationspunkte der originalen Diskretisierung werden durch „o“ und die der Untersimulationen durch „x“ markiert. Wobei hier eine Kollokationsordnung von  $NC = 3$  für die originale Diskretisierung und für die Untersimulation angenommen wurde (prinzipiell können diese auch ohne weitere Einschränkungen unterschiedlich sein). Wie zu erkennen ist, werden für die Untersimulation

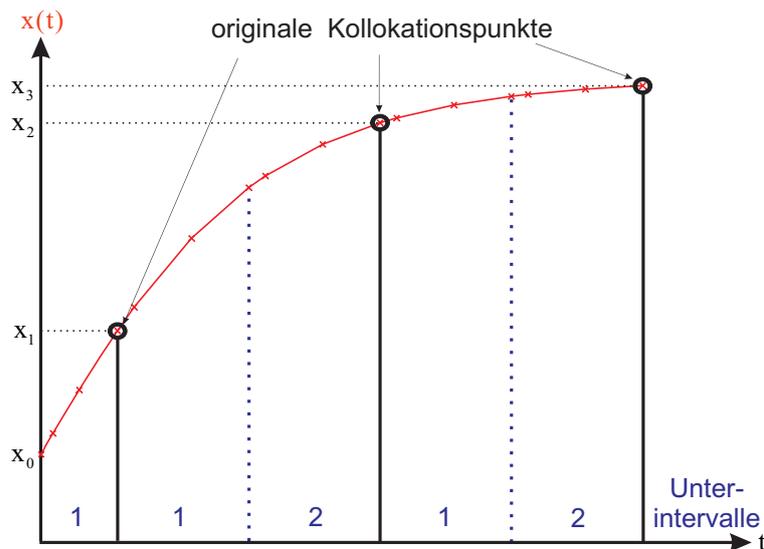


Abbildung 4.7: Die Untersimulation in der quasi-sequentiellen Methode

zum ersten originalen Kollokationspunkt ein Unterintervall und für die anderen originalen Kollokationspunkte jeweils zwei Unterintervalle beispielhaft benötigt. Diese Untersimulation

und das Lösungsschema aus Abbildung 4.6 ergibt eine Berechnung der Zustände an den originalen Diskretisierungspunkten unter Beachtung der Approximation ohne die Dimension des originalen diskretisierten Optimierungsproblems zu verändern. Dies hat positiv zur Folge, dass ein NLP-Löser ohne Neustarts die Iterationen fortführen kann und die Approximation der Zustände ohne MFE-Verfahren gesichert ist.

### Fallbeispiel (Approximationskontrolle)

Hier wird das Vorgehen der Approximationskontrolle mit Untersimulation auf das Beispiel aus Gleichung (4.4) angewendet und eine unzureichende Approximation mit  $NL = 1$  (siehe Abbildung 4.3 und 4.4 unten) untersucht. Dabei wird für die Variante 1 ein relativer Fehler im Bereich von  $\epsilon_{rel} = 10^{-4}$  und für die Variante 2 eine Toleranz von  $tol = [0, 15 \ 0, 44 \ 0, 71]$  aus Kapitel 4.2.2 gewählt. Es ergeben sich die relativen Fehler an den originalen Kollokationspunkten (aus der Sicht der Optimierungsschicht) aus Abbildung 4.8.

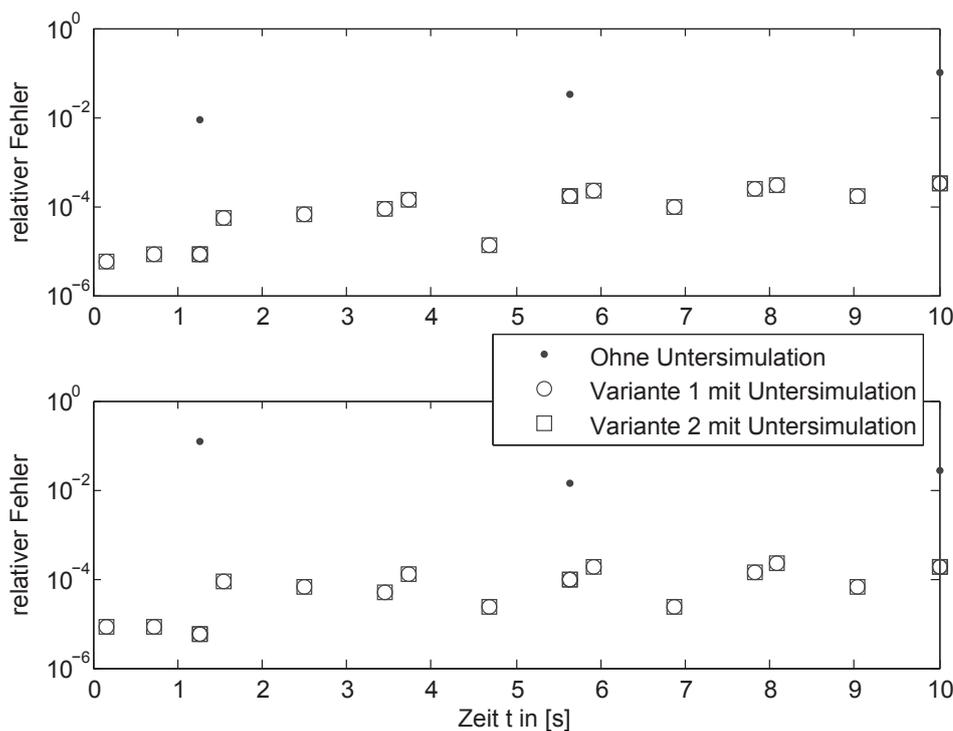


Abbildung 4.8: Approximationskontrolle durch Untersimulation (oben:  $u = -0,25s^{-1}$ , unten:  $u = 0,25s^{-1}$ ). Dargestellt werden zusätzlich die relativen Fehler zur analytischen Lösung für die Kollokationspunkte der Untersimulationen.

In beiden Fällen verläuft die Fehlerdetektion erfolgreich und es ergibt sich ein relativer Fehler im Bereich von  $10^{-4}$  zur analytischen Lösung der DGL. Damit sichert dieses Vorgehen den Approximationsfehler des Zustandsverlaufes an den originalen Kollokationspunkten. Für beide Varianten werden jeweils 1 Unterintervall zum ersten originalen Kollokationspunkt

und jeweils 2 für die anderen originalen Kollokationspunkte benötigt. Dies veranschaulicht zusätzlich, dass eine zu grobe originale Diskretisierung und hohe Genauigkeitsanforderungen auch eine hohe Anzahl von Untersimulationen zur Folge haben. So werden z.B.  $NL = 5$  gleich verteilte, originale finite Elemente benötigt um eine Genauigkeitsanforderung von  $\epsilon_{rel} = 10^{-4}$  zu realisieren (dies entspricht jedoch auch der Anzahl der vorherigen Untersimulationen). Es muss dementsprechend abgewogen werden, wie hoch die Genauigkeit und welche größte Diskretisierung benötigt wird, um das kontinuierliche Problem darzustellen.

#### 4.2.4 Die Gradientenberechnung

Die Grundlagen der Gradientenberechnung in der quasi-sequentiellen Methode wurden in Kapitel 3.3 dargestellt. Dabei werden während der sukzessiven Lösung des Gleichungssystems (3.18) die benötigten Gradienten nach den Steuergrößen parallel zur Lösung der Modellgleichungen ermittelt und durch die „chain rule“ zwischen den Elementen verkettet. Dieses Verfahren bleibt entsprechend in den Elementen unberührt, wo die Approximationsgenauigkeit erfüllt ist. Wird jedoch durch einen Approximationsfehler die Untersimulation ausgeführt, muss hierfür die Gradientenberechnung angepasst werden. Hierbei sind ausschließlich die Gradienten der originalen Kollokationspunkte von Interesse. Dies betrifft für ein entsprechendes  $n'tes$  finites Element

$$\frac{d\mathbf{X}_n}{d\mathbf{U}_n}$$

und

$$\frac{d\mathbf{X}_n}{d\mathbf{X}_{n,0}}.$$

Im Folgendem wird die Gradientenberechnung im  $n'ten$  Element für  $NL_{AC}$  Unterintervalle beginnend beim ersten Kollokationspunkt betrachtet. Für diese Berechnungen muss entsprechend Gleichung (3.18) für eine Untersimulation nach

$$\mathbf{C}_j \left( \mathbf{S}_{j,0}, \mathbf{S}_j, \mathbf{U}_j^S \right) = \mathbf{0} ; \mathbf{S}_{0,0} = \mathbf{X}_{n,0} ; \mathbf{S}_{j,0} = \mathbf{D} \cdot \mathbf{S}_{j-1} ; j = 1, \dots, NL_{AC}$$

angepasst werden. Wobei  $\mathbf{S}$  die diskretisierten Zustandsgrößen einer Untersimulation darstellt (veranschaulicht wurden diese durch „x“ in Abbildung 4.9).

Es wird nun ähnlich zu Kapitel 3.3 verfahren und die gelöste Gleichung durch Taylor-Entwicklung betrachtet:

$$\frac{d\mathbf{C}_j}{d\mathbf{S}_{j,0}} \cdot \Delta\mathbf{S}_{j,0} + \frac{d\mathbf{C}_j}{d\mathbf{S}_j} \cdot \Delta\mathbf{S}_j + \frac{d\mathbf{C}_j}{d\mathbf{U}_j} \cdot \Delta\mathbf{U}_j = \mathbf{0},$$

bzw.

$$\mathbf{K}_j \cdot \Delta \mathbf{S}_{j,0} + \mathbf{L}_j \cdot \Delta \mathbf{S}_j + \mathbf{M}_j \cdot \Delta \mathbf{U}_j = \mathbf{0}.$$

Dementsprechend können die Gradienten der Zustände nach der parametrisierten Steuerung und nach den Anfangsbedingungen eines Intervalls der Untersimulation berechnet werden durch

$$\mathbf{L}_j \cdot \frac{d\mathbf{S}_j}{d\mathbf{U}_j} = -\mathbf{M}_j \quad (4.9)$$

und

$$\mathbf{L}_j \cdot \frac{d\mathbf{S}_j}{d\mathbf{S}_{j,0}} = -\mathbf{K}_j. \quad (4.10)$$

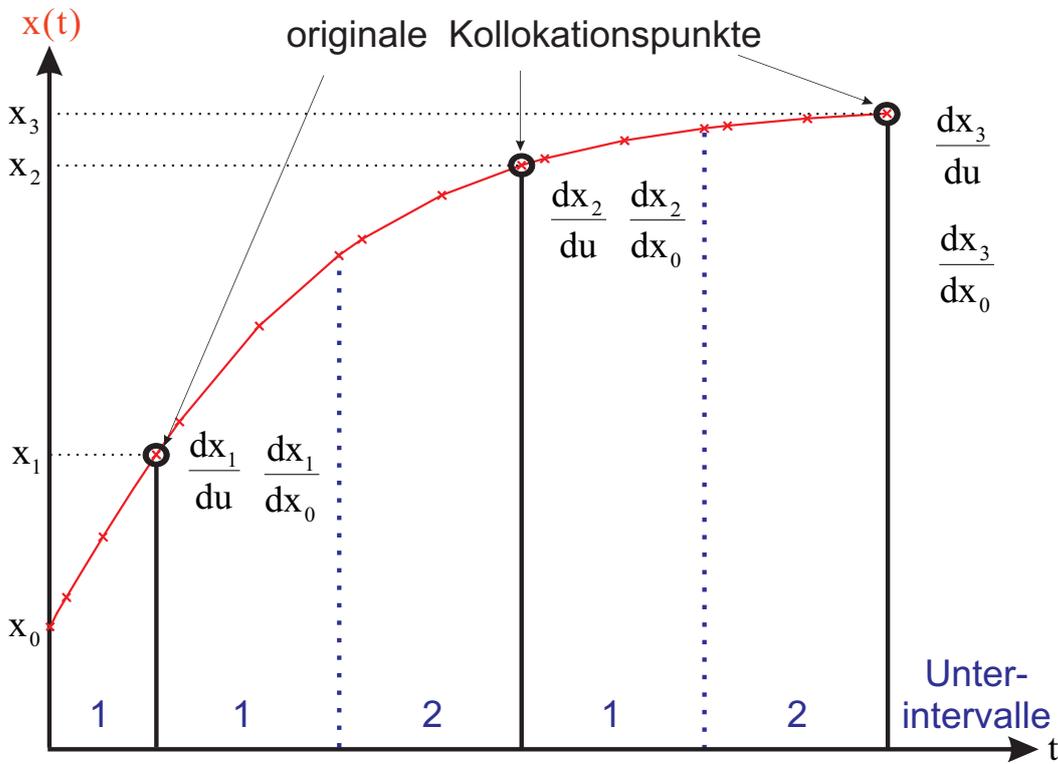


Abbildung 4.9: Gradienten bei einer Untersimulation in der quasi-sequentiellen Methode

Hierbei kann die Matrix  $\mathbf{L}_j$  wieder aus den Newton-Verfahren entnommen oder eine vorliegende LU-Zerlegung verwendet werden. Es gilt die Kontinuitätsbedingung der Zustände auch während einer Untersimulation. Somit können die Gradienten von einem Unterintervall zum nächsten durch die „chain rule“ [67] für  $i < j$  transformiert werden nach:

$$\frac{d\mathbf{S}_j}{d\mathbf{U}_i} = \frac{d\mathbf{S}_j}{d\mathbf{D} \cdot \mathbf{S}_{j-1}} \cdot \frac{d\mathbf{D} \cdot \mathbf{S}_{j-1}}{d\mathbf{U}_i} = \frac{d\mathbf{S}_j}{d\mathbf{S}_{j,0}} \frac{d\mathbf{D} \cdot \mathbf{S}_{j-1}}{d\mathbf{U}_i}. \quad (4.11)$$

Für  $i = j$  sind die benötigten Gradienten durch Gleichung (4.9) gegeben und für  $i > j$  ergeben

sich die Gradienten zu Null. Die Berechnung von (4.10) ist wieder für die Evaluierung von Gleichung (4.11) notwendig. Dadurch ergeben sich die Gradienten für eine Untersimulation mit  $NL_{AC}$  finiten Unterelementen im Allgemeinen zu

$$\frac{d\mathbf{S}}{d\mathbf{U}^S} = \begin{bmatrix} \frac{d\mathbf{S}_1}{d\mathbf{U}_1^S} & & & & \mathbf{0} \\ \vdots & \ddots & & & \\ \frac{d\mathbf{S}_j}{d\mathbf{U}_1^S} & \cdots & \frac{d\mathbf{S}_j}{d\mathbf{U}_j^S} & & \\ \vdots & \ddots & \vdots & \ddots & \\ \frac{d\mathbf{S}_{NL_{AC}}}{d\mathbf{U}_1^S} & \cdots & \frac{d\mathbf{S}_{NL_{AC}}}{d\mathbf{U}_j^S} & \cdots & \frac{d\mathbf{S}_{NL_{AC}}}{d\mathbf{U}_{NL_{AC}}^S} \end{bmatrix}. \quad (4.12)$$

Während einer Untersimulation sind ausschließlich die Gradienten nach der Steuerungsparametrisierung der originalen Kollokationspunkte von Interesse und entsprechen den Gradienten der letzten Kollokationspunkte in einer Untersimulation. Somit befinden sich die gesuchten Informationen für alle Zustände des originalen Elementes in der letzten Zeile aus Gleichung (4.12). Hier kommt erschwerend hinzu, dass der Zusammenhang zur Steuerungsparametrisierung des originalen finiten Elementes ermittelt werden muss. Da eine Untersimulation die Diskretisierung der Zeit verfeinert, kann der jeweilige Einfluss der Steuerungsparametrisierung  $\mathbf{U}_n$  im Unterintervall  $t_j^S = [t_{j-1}^S, t_j^S]$  durch  $\mathbf{U}_n^j$  symbolisiert werden und es ergibt sich

$$\frac{d\mathbf{S}}{d\mathbf{U}^S} = \begin{bmatrix} \frac{d\mathbf{S}_1}{d\mathbf{U}_n^j} & & & & \mathbf{0} \\ \vdots & \ddots & & & \\ \frac{d\mathbf{S}_j}{d\mathbf{U}_n^j} & \cdots & \frac{d\mathbf{S}_j}{d\mathbf{U}_n^j} & & \\ \vdots & \ddots & \vdots & \ddots & \\ \frac{d\mathbf{S}_{NL_{AC}}}{d\mathbf{U}_n^j} & \cdots & \frac{d\mathbf{S}_{NL_{AC}}}{d\mathbf{U}_n^j} & \cdots & \frac{d\mathbf{S}_{NL_{AC}}}{d\mathbf{U}_n^{NL_{AC}}} \end{bmatrix}. \quad (4.13)$$

Somit sind die berechneten Gradienten aus Gleichung (4.12) alle von der Steuerungsparametrisierung des originalen finiten Elementes abhängig (vergl. Gleichung (4.13)). Deshalb kann für die Gradienten am ersten Kollokationspunkt des originalen finiten Elementes

$$\frac{d\mathbf{X}_{n,1}}{d\mathbf{U}_n} = \frac{d\mathbf{S}_{NL_{AC},NC}}{d\mathbf{U}_n}$$

durch eine Differentiation der zusammengesetzten Funktionen nach

$$\frac{d\mathbf{X}_{n,1}}{d\mathbf{U}_n} = \sum_{j=1}^{NL_{AC}} \frac{d\mathbf{S}_{NL_{AC},NC}}{d\mathbf{U}_j^S} \cdot \frac{d\mathbf{U}_j^S}{d\mathbf{U}_n}$$

#### 4 Erweiterungen der quasi-sequentiellen Methode

und einer konstanter Steuerungsparametrisierung

$$\mathbf{U}_j^S = \mathbf{U}_n ; j = 1, \dots, NLAC$$

der Zusammenhang

$$\frac{d\mathbf{X}_{n,1}}{d\mathbf{U}_n} = \sum_{j=1}^{NLAC} \frac{d\mathbf{S}_{NLAC,NC}}{d\mathbf{U}_j^S}$$

abgeleitet werden. Zudem kann nun die Verkettung vereinfacht werden zu

$$\frac{d\hat{\mathbf{S}}_j}{d\mathbf{U}_n} = \frac{d\hat{\mathbf{S}}_j}{d\hat{\mathbf{S}}_{j,0}} \frac{d\mathbf{D} \cdot \hat{\mathbf{S}}_{j-1}}{d\mathbf{U}_n} + \frac{d\mathbf{S}_j}{d\mathbf{U}_j^S} \text{ mit } j = 1, \dots, NLAC$$

und

$$\frac{d\hat{\mathbf{S}}_0}{d\mathbf{U}_n} = \mathbf{0}$$

und entsprechend

$$\frac{d\mathbf{X}_{n,1}}{d\mathbf{U}_n} = \frac{d\hat{\mathbf{S}}_{NLAC,NC}}{d\mathbf{U}_n}.$$

Die Berechnung der Gradienten am ersten Kollokationspunkt des originalen finiten Elementes nach den Anfangsbedingungen (nötigt für die spätere Verkettung über das finite Element hinweg) kann einfach ermittelt werden nach

$$\mathbf{L}_1 \cdot \frac{d\mathbf{S}_1}{d\mathbf{X}_{n,0}} = -\mathbf{K}_1$$

und anschließender Verkettung

$$\frac{d\hat{\mathbf{S}}_j}{d\mathbf{X}_{n,0}} = \frac{d\hat{\mathbf{S}}_j}{d\hat{\mathbf{S}}_{j,0}} \frac{d\mathbf{D} \cdot \hat{\mathbf{S}}_{j-1}}{d\mathbf{X}_{n,0}} \text{ mit } j = 2, \dots, NLAC$$

mit

$$\frac{d\mathbf{X}_{n,1}}{d\mathbf{X}_{n,0}} = \frac{d\hat{\mathbf{S}}_{NLAC,NC}}{d\mathbf{X}_{n,0}}.$$

Durch diese Vorgehen können die geforderten Gradienten (siehe z.B. Abbildung 4.9) aller Zustände des ersten originalen Kollokationspunktes im Allgemeinen effektiv ermittelt werden. Die restlichen Gradienten der originalen Kollokationspunkte (2, 3, ..., NC) können unter einer modifizierten Startbedingung der Verkettung berechnet werden. Dies ergibt für die Gradienten nach der parametrisierten Steuerung des originalen Elementes in Kurzform:

$$\frac{d\hat{\mathbf{S}}_j}{d\mathbf{U}_n} = \frac{d\hat{\mathbf{S}}_j}{d\hat{\mathbf{S}}_{j,0}} \frac{d\mathbf{D} \cdot \hat{\mathbf{S}}_{j-1}}{d\mathbf{U}_n} + \frac{d\mathbf{S}_j}{d\mathbf{U}_j^S} \text{ mit } j = 2, \dots, NLAC,$$

$$\frac{d\hat{\mathbf{S}}_0}{d\mathbf{U}_n} = \frac{d\mathbf{X}_{n,(1,2,\dots,NC-1)}}{d\mathbf{U}_n}$$

und

$$\frac{d\mathbf{X}_{n,(2,3,\dots,NC)}}{d\mathbf{U}_n} = \frac{d\hat{\mathbf{S}}_{NLAC,NC}}{d\mathbf{U}_n}.$$

Die Berechnung der Gradienten an den restlichen Kollokationspunkten des originalen finiten Elementes nach den Anfangsbedingungen ergibt sich wie folgt

$$\frac{d\mathbf{S}_1}{d\mathbf{X}_{n,0}} = \frac{d\mathbf{X}_{n,(1,2,\dots,NC-1)}}{d\mathbf{X}_{n,0}}$$

und anschließender Verkettung

$$\frac{d\hat{\mathbf{S}}_j}{d\mathbf{X}_{n,0}} = \frac{d\hat{\mathbf{S}}_j}{d\hat{\mathbf{S}}_{j,0}} \frac{d\mathbf{D} \cdot \hat{\mathbf{S}}_{j-1}}{d\mathbf{X}_{n,0}} \text{ mit } j = 2, \dots, NLAC$$

und

$$\frac{d\mathbf{X}_{n,(2,3,\dots,NC)}}{d\mathbf{X}_{n,0}} = \frac{d\hat{\mathbf{S}}_{NLAC,NC}}{d\mathbf{X}_{n,0}}.$$

Was dieses für die Gradientenapproximation bedeutet, wird im Folgendem untersucht.

### Fallbeispiel (Gradienten und Approximationskontrolle)

Hier wird der Einfluss der Approximationskontrolle durch die Untersimulation auf die Genauigkeit der Gradienten an einem Beispiel aus Gleichung (4.4) untersucht. Wird die analytische Lösung aus Gleichung (4.5) nach  $u$  differenziert, ergibt sich

$$\frac{dx(t)}{du} = t \cdot x_0 \cdot e^{u \cdot t}$$

und nach  $x_0$

$$\frac{dx(t)}{dx_0} = e^{u \cdot t}.$$

Dies kann zum Vergleich der Gradientenapproximation genutzt werden. Dabei wird eine unzureichender Approximation mit  $NL = 1$  (siehe Abbildung 4.3 und 4.4 unten) aus Kapitel 4.2.3 erneut angewendet (Variante 1 ein relativer Fehler im Bereich von  $\epsilon_{rel} = 10^{-4}$  und Variante 2 eine Toleranz von  $tol = [0, 15 \ 0, 44 \ 0, 71]$ ). Es ergeben sich die relativen Fehler der Gradienten an den original Kollokationspunkten aus Abbildung 4.10 und 4.11. An dieser Stelle erhöht die Approximationsgenauigkeit auch die Genauigkeit der Gradienten. Dies wirkt sich grundsätzlich positiv auf die Konvergenzeigenschaften eines NLP-Lösers aus.

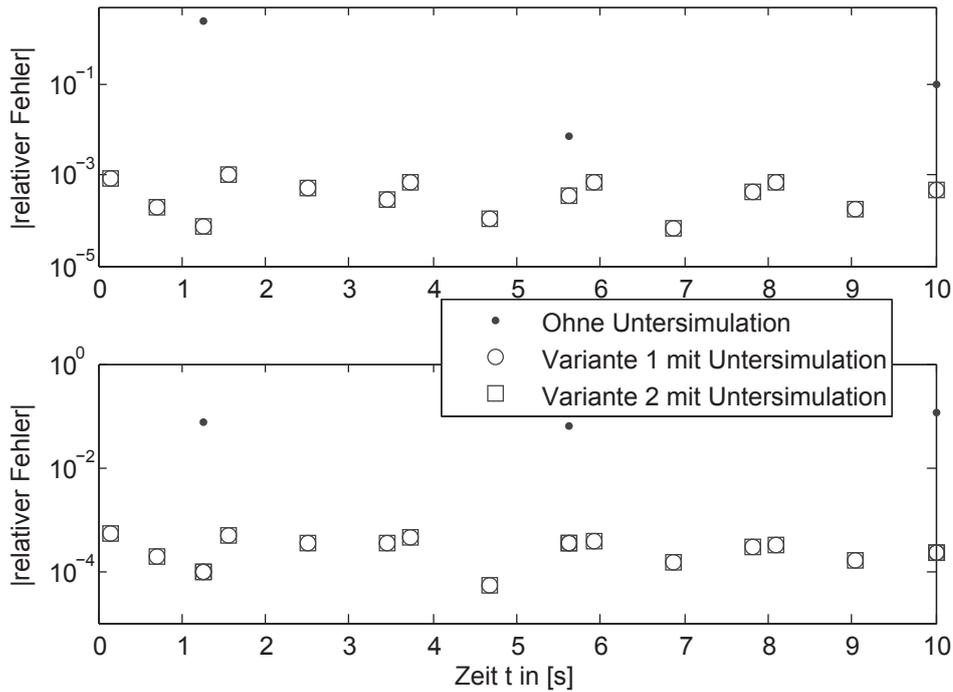


Abbildung 4.10: Einfluss der Untersimulation auf die Approximationsgüte der Gradienten nach  $u$  (oben:  $u = -0,25s^{-1}$ , unten:  $u = 0,25s^{-1}$ ). Dargestellt werden zusätzlich die relativen Fehler zur analytischen Lösung für die Kollokationspunkte der Untersimulationen.

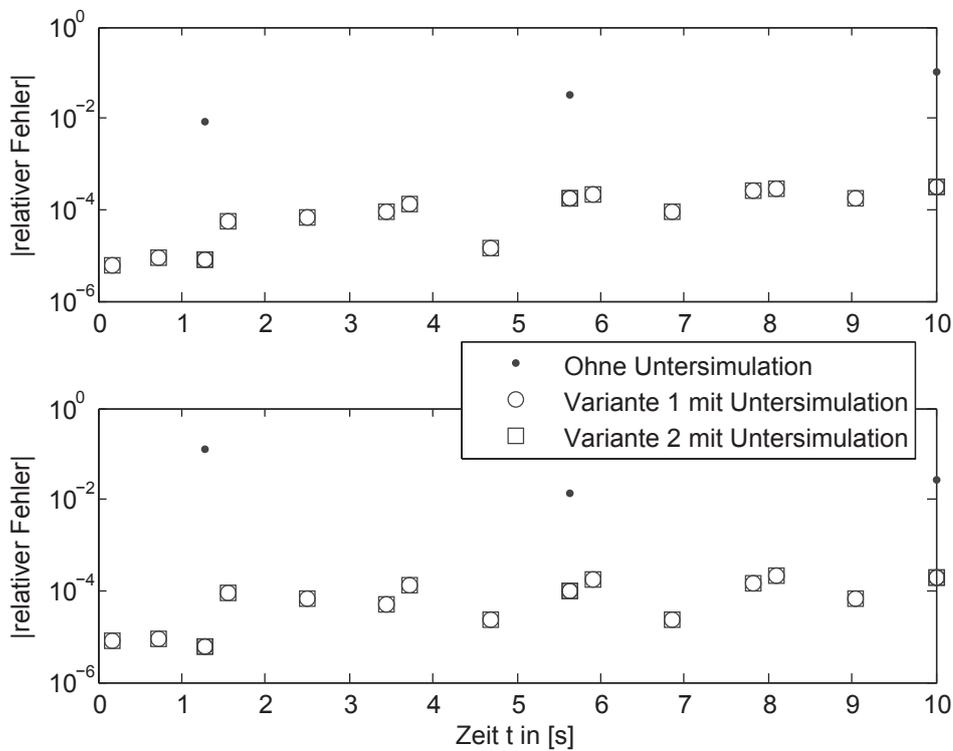


Abbildung 4.11: Einfluss der Untersimulation auf die Approximationsgüte der Gradienten nach  $x_0$  (oben:  $u = -0,25s^{-1}$ , unten:  $u = 0,25s^{-1}$ ). Dargestellt werden zusätzlich die relativen Fehler zur analytischen Lösung für die Kollokationspunkte der Untersimulationen.

## 4.3 Bewegte Elemente

Dieses Kapitel untersucht die Detektion von Umschaltvorgängen in der optimalen Steuerung unter Anwendung der quasi-sequentiellen Methode. Zu dieser Kategorie gehört auch die sogenannte Bang-Bang-Steuerung, die in vorherigen Untersuchungen der Stoffwechselregulation zahlreich aufgetreten ist [10, 80, 103].

### 4.3.1 Das Verfahren in der quasi-sequentiellen Methode

Eine Detektion von optimalen Umschaltvorgängen ist in der Simulationsschicht der quasi-sequentiellen Methode nicht zu realisieren, jedoch kann diese in der Optimierungsschicht umgesetzt werden. Der grundlegende Ansatz besteht hierbei in einer gezielten Anpassung der finiten Elemente. Dies entspricht einer adaptiven Diskretisierung, wobei große Elemente einen entsprechend negativen Einfluss auf die Approximationsgüte der Zustandsverläufe haben. Somit ist bei einem derartigen Verfahren eine Approximationskontrolle unumgänglich und wird durch das in Kapitel 4.2.3 dargestellte Verfahren umgesetzt (vergl. Lösungsschema aus Abbildung 4.6 und Gleichung (4.8)).

In der simultanen Methode wird die Elementanpassung durch eine Zwei-Schichten-Optimierung realisiert, um auch eine Approximationskontrolle zu ermöglichen. Wie oben erwähnt wird die Approximationskontrolle in der quasi-sequentiellen Methode durch eine Erweiterung in der Simulationsschicht bereits garantiert und die Elementanpassung kann somit in der Optimierungsschicht ohne weitere Schichtenmodelle realisiert werden. Dies wird auch als „moving finite elements“ MFE (bewegte finite Elemente) bezeichnet, um die dynamische Anpassung der Elementlängen durch einen NLP-Lösers zum Ausdruck zu bringen. Es ergibt sich somit das betrachtete Optimierungsproblem zu

$$\min_{\mathbf{U}, \Delta \mathbf{T}} \varphi(\mathbf{U}, \Delta \mathbf{T}, \mathbf{X}(\mathbf{U}, \Delta \mathbf{T})) \quad (4.14)$$

$$s.t. \mathbf{U}_{min} \leq \mathbf{U} \leq \mathbf{U}_{max} ; \Delta \mathbf{T}_{min} \leq \Delta \mathbf{T} \leq \Delta \mathbf{T}_{max} ; \mathbf{X}_{min} \leq \mathbf{X}(\mathbf{U}, \Delta \mathbf{T}) \leq \mathbf{X}_{max}$$

mit dem in Abbildung 4.12 dargestelltem Lösungsschema und wird in dieser Arbeit auch als qMFE bezeichnet. Hierbei werden die finiten Elemente als zusätzliche Freiheitsgrade eingesetzt und durch

$$\Delta \mathbf{T} = [\Delta t_1, \dots, \Delta t_{NL}]^T \in \Re^{NL}$$

vektoriell dargestellt. Diese Erweiterung ermöglicht dem NLP-Löser die Längen der Zeitintervalle ausschließlich durch die Optimalitätsbedingungen des diskretisierten Optimierungs-

problems zu bestimmen. In diesem Zusammenhang hat entsprechend die Approximationskontrolle und deren nichtlineare Nebenbedingungen keinen Einfluss, da diese in der Simulationsschicht behandelt werden.

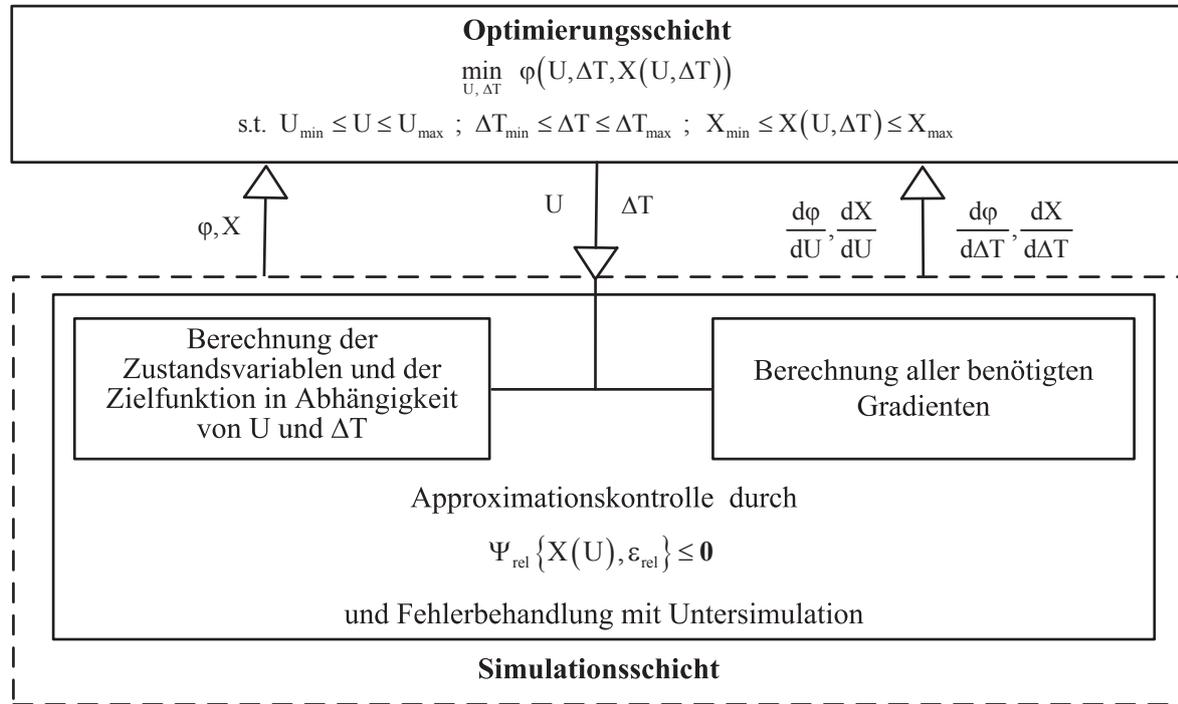


Abbildung 4.12: Lösungsschema für bewegte Elemente in der quasi-sequentiellen Methode

### 4.3.2 Die Gradientenberechnung

Die Berechnungen basierten bisher auf den Gradienten nach der Steuerungsparametrisierung und im Falle einer Untersimulation zusätzlich nach den Anfangsbedingungen der finiten Elemente mit detektiertem Approximationsfehler. Dazu wird im Folgendem die Gradientenberechnung nach den Elementlängen ohne Untersimulation und nach den Elementlängen mit einer Untersimulation erweitert.

#### Gradienten ohne Untersimulation

Die bisherige Gradientenberechnung basiert auf Gleichung (3.18), wobei die Elementlängen bei den Parametern des Diskretisierungsverfahrens eingeordnet wurden. Die Abhängigkeit von den Elementlängen ist in Gleichung (3.15) ersichtlich und kann entsprechend über Gleichung (3.16), (3.17) zu (3.18) dargestellt werden durch

$$\mathbf{C}_n(\mathbf{X}_{n,0}, \mathbf{X}_n, \mathbf{U}_n, \Delta t_n) = \mathbf{0} ; \mathbf{X}_{n,0} = \mathbf{D} \cdot \mathbf{X}_{n-1} ; n = 1, \dots, NL. \quad (4.15)$$

Die benötigten Gradienten werden durch die Taylor-Entwicklung des gelösten Gleichungssystems ermittelt:

$$\frac{d\mathbf{C}_n}{d\mathbf{X}_{n,0}} \cdot \Delta\mathbf{X}_{n,0} + \frac{d\mathbf{C}_n}{d\mathbf{X}_n} \cdot \Delta\mathbf{X}_n + \frac{d\mathbf{C}_n}{d\mathbf{U}_n} \cdot \Delta\mathbf{U}_n + \frac{d\mathbf{C}_n}{d(\Delta t_n)} \cdot \Delta(\Delta t_n) = \mathbf{0}$$

oder

$$\mathbf{K}_n \cdot \Delta\mathbf{X}_{n,0} + \mathbf{L}_n \cdot \Delta\mathbf{X}_n + \mathbf{M}_n \cdot \Delta\mathbf{U}_n + \mathbf{N}_n \cdot \Delta(\Delta t_n) = \mathbf{0}.$$

Somit ergeben sich für die Gradienten der Zustände nach der parametrisierten Steuerung, nach den Anfangsbedingungen eines finiten Elementes und nach der Elementlänge zu

$$\mathbf{L}_n \cdot \frac{d\mathbf{X}_n}{d\mathbf{U}_n} = -\mathbf{M}_n,$$

$$\mathbf{L}_n \cdot \frac{d\mathbf{X}_n}{d\mathbf{X}_{n,0}} = -\mathbf{K}_n$$

und

$$\mathbf{L}_n \cdot \frac{d\mathbf{X}_n}{d(\Delta t_n)} = -\mathbf{N}_n. \quad (4.16)$$

Dementsprechend können die Gradienten nach den Steuergrößen weiterhin wie in Kapitel 3.3 ermittelt werden. Die Kontinuitätsbedingung der Zustände ermöglicht es die Gradienten von einem Element zum nächsten durch die „chain rule“ [67] zu transformieren. Dadurch werden die benötigten Gradienten der Zustände eines  $n'$ ten Elementes nach den Elementlängen der vorhergehenden Elemente (für  $i < n$ ) berechnet nach

$$\frac{d\mathbf{X}_n}{d(\Delta t_i)} = \frac{d\mathbf{X}_n}{d\mathbf{D} \cdot \mathbf{X}_{n-1}} \cdot \frac{d\mathbf{D} \cdot \mathbf{X}_{n-1}}{d(\Delta t_i)} = \frac{d\mathbf{X}_n}{d\mathbf{X}_{n,0}} \frac{d\mathbf{D} \cdot \mathbf{X}_{n-1}}{d(\Delta t_i)}.$$

Für  $i = n$  sind die benötigten Gradienten durch Gleichung (4.16) gegeben und für  $i > n$  liegt keine Abhängigkeit vor (diese Gradienten ergeben sich zu Null). Durch dieses Vorgehen sind nun die benötigten Gradienten nach den Elementlängen ermittelt und ergeben sich im Allgemeinen zu

$$\frac{d\mathbf{X}}{d(\Delta \mathbf{T})} = \begin{bmatrix} \frac{d\mathbf{X}_1}{d(\Delta t_1)} & & & & & \mathbf{0} \\ \vdots & \ddots & & & & \\ \frac{d\mathbf{X}_n}{d(\Delta t_1)} & \cdots & \frac{d\mathbf{X}_n}{d(\Delta t_n)} & & & \\ \vdots & \ddots & \vdots & \ddots & & \\ \frac{d\mathbf{X}_{NL}}{d(\Delta t_1)} & \cdots & \frac{d\mathbf{X}_{NL}}{d(\Delta t_n)} & \cdots & \frac{d\mathbf{X}_{NL}}{d(\Delta t_{NL})} & \end{bmatrix}.$$

### Gradienten mit Untersimulation

Wird durch einen Approximationsfehler die Untersimulation ausgeführt, muss auch hier die Gradientenberechnung nach den Elementlängen angepasst werden. Wie bereits bekannt, sind diesbezüglich ausschließlich die Gradienten der originalen Kollokationspunkte von Interesse. Dies betrifft für ein entsprechendes  $n'tes$  finites Element

$$\frac{d\mathbf{X}_n}{d\mathbf{U}_n}$$

und

$$\frac{d\mathbf{X}_n}{d\mathbf{X}_{n,0}}$$

und

$$\frac{d\mathbf{X}_n}{d\Delta t_n}.$$

Der Zusammenhang zwischen Untersimulation und dieser gesuchten Gradienten ist in Abbildung 4.13 dargestellt. Die Gradientenberechnung im  $n'ten$  Element für  $NL_{AC}$  Unterintervalle wird beginnend beim ersten Kollokationspunkt betrachtet. Dementsprechend ergibt sich eine Anpassung von Gleichung (4.15) bei bewegten Elementen mit Untersimulation nach

$$\mathbf{C}_j(\mathbf{S}_{j,0}, \mathbf{S}_j, \mathbf{U}_j^S, \Delta t_j^S) = \mathbf{0}; \mathbf{S}_{0,0} = \mathbf{X}_{n,0}; \mathbf{S}_{j,0} = \mathbf{D} \cdot \mathbf{S}_{j-1}; j = 1, \dots, NL_{AC}.$$

Hierbei stellt  $\mathbf{S}$  wieder die diskretisierten Zustandsgrößen einer Untersimulation dar (veranschaulicht durch „x“ in Abbildung 4.13). Hier wird ähnlich zu Kapitel 4.2.4 verfahren und die gelöste Gleichung durch Taylor-Entwicklung betrachtet:

$$\frac{d\mathbf{C}_j}{d\mathbf{S}_{j,0}} \cdot \Delta \mathbf{S}_{j,0} + \frac{d\mathbf{C}_j}{d\mathbf{S}_j} \cdot \Delta \mathbf{S}_j + \frac{d\mathbf{C}_j}{d\mathbf{U}_j} \cdot \Delta \mathbf{U}_j + \frac{d\mathbf{C}_j}{d(\Delta t_j^S)} \cdot \Delta (\Delta t_j^S) = \mathbf{0}$$

und

$$\mathbf{K}_j \cdot \Delta \mathbf{S}_{j,0} + \mathbf{L}_j \cdot \Delta \mathbf{S}_j + \mathbf{M}_j \cdot \Delta \mathbf{U}_j + \mathbf{N}_j \cdot \Delta (\Delta t_j^S) = \mathbf{0}.$$

Es ergibt sich somit

$$\mathbf{L}_j \cdot \frac{d\mathbf{S}_j}{d\mathbf{U}_j} = -\mathbf{M}_j,$$

$$\mathbf{L}_j \cdot \frac{d\mathbf{S}_j}{d\mathbf{S}_{j,0}} = -\mathbf{K}_j$$

und

$$\mathbf{L}_j \cdot \frac{d\mathbf{S}_j}{d(\Delta t_j^S)} = -\mathbf{N}_j. \quad (4.17)$$

Dies ermöglicht, die Gradienten nach den Steuergrößen und Anfangsbedingungen weiterhin wie in Kapitel 4.2.4 zu berechnen. Es gilt außerdem durch die Verkettungsregel für  $i < j$ :

$$\frac{d\mathbf{S}_j}{d(\Delta t_i^S)} = \frac{d\mathbf{S}_j}{d\mathbf{D} \cdot \mathbf{S}_{j-1}} \cdot \frac{d\mathbf{D} \cdot \mathbf{S}_{j-1}}{d(\Delta t_i^S)} = \frac{d\mathbf{S}_j}{d\mathbf{S}_{j,0}} \frac{d\mathbf{D} \cdot \mathbf{S}_{j-1}}{d(\Delta t_i^S)}.$$

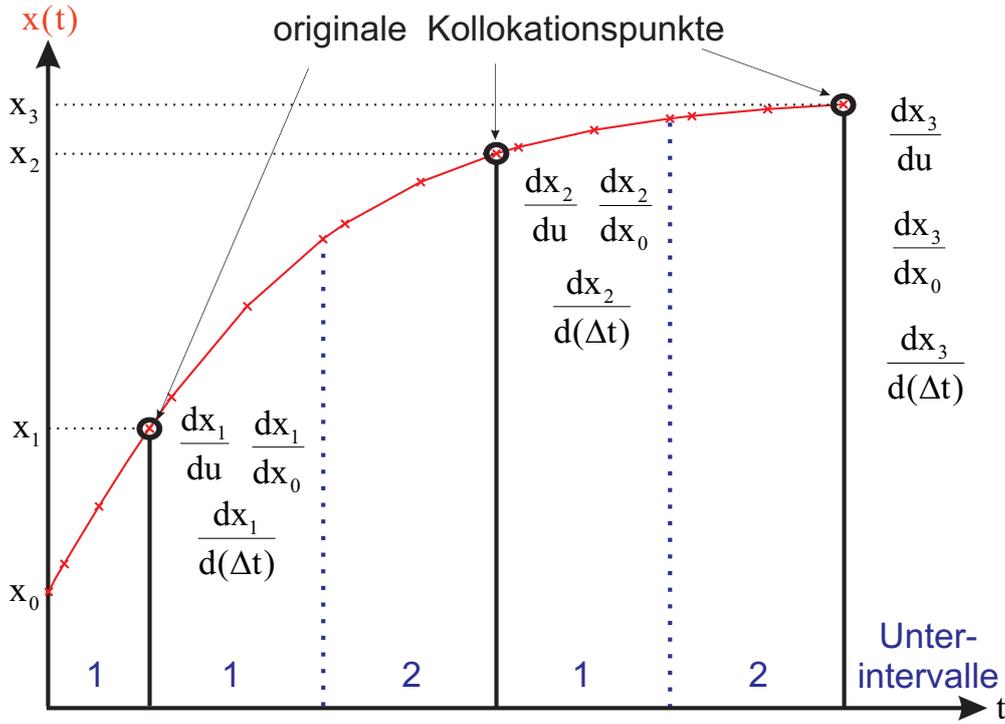


Abbildung 4.13: Die notwendigen Gradienten für bewegte Elemente mit einer Untersimulation in der quasi-sequentiellen Methode

Für  $i = j$  sind die benötigten Gradienten durch Gleichung (4.17) gegeben und für  $i > j$  ergeben sich die Gradienten zu Null. Dadurch resultieren die Gradienten für eine Untersimulation mit  $NL_{AC}$  finiten Unter-elementen im Allgemeinen zu

$$\frac{d\mathbf{S}}{d(\Delta \mathbf{T}^S)} = \begin{bmatrix} \frac{d\mathbf{S}_1}{d(\Delta t_1^S)} & & & & & \mathbf{0} \\ \vdots & \ddots & & & & \\ \frac{d\mathbf{S}_j}{d(\Delta t_1^S)} & \cdots & \frac{d\mathbf{S}_j}{d(\Delta t_j^S)} & & & \\ \vdots & \ddots & \vdots & \ddots & & \\ \frac{d\mathbf{S}_{NL_{AC}}}{d(\Delta t_1^S)} & \cdots & \frac{d\mathbf{S}_{NL_{AC}}}{d(\Delta t_j^S)} & \cdots & \frac{d\mathbf{S}_{NL_{AC}}}{d(\Delta t_{NL_{AC}}^S)} & \end{bmatrix}. \quad (4.18)$$

An dieser Stelle sind ausschließlich die Gradienten nach der Elementlänge der originalen Kollokationspunkte von Interesse. Die entsprechenden Gradienteninformationen befinden sich

#### 4 Erweiterungen der quasi-sequentiellen Methode

für alle Zustände des originalen Elementes in der letzten Zeile aus Gleichung (4.18). Die Gleichung (4.18) stellt jedoch nur die Abhängigkeiten nach den Unterintervallen  $\Delta t_j^S$  dar. Da eine Untersimulation die Diskretisierung der Zeit verfeinert, kann der jeweilige Einfluss des originalen finiten Elementes  $\Delta t_n$  im Unterintervall  $t_j^S = [t_{j-1}^S, t_j^S]$  durch  $\Delta t_n^j$  symbolisiert werden und es ergibt sich

$$\frac{d\mathbf{S}}{d(\Delta \mathbf{T}^S)} = \begin{bmatrix} \frac{d\mathbf{S}_1}{d(\Delta t_n^1)} & & & & \mathbf{0} \\ \vdots & \ddots & & & \\ \frac{d\mathbf{S}_j}{d(\Delta t_n^j)} & \cdots & \frac{d\mathbf{S}_j}{d(\Delta t_n^j)} & & \\ \vdots & \ddots & \vdots & \ddots & \\ \frac{d\mathbf{S}_{NLAC}}{d(\Delta t_n^1)} & \cdots & \frac{d\mathbf{S}_{NLAC}}{d(\Delta t_n^j)} & \cdots & \frac{d\mathbf{S}_{NLAC}}{d(\Delta t_n^{NLAC})} \end{bmatrix}. \quad (4.19)$$

Somit sind die berechneten Gradienten aus Gleichung (4.18) alle von der Länge des originalen finiten Elementes abhängig (vergl. Gleichung (4.19)). Deshalb kann für die Gradienten am ersten Kollokationspunkt des originalen finiten Elementes

$$\frac{d\mathbf{X}_{n,1}}{d(\Delta t_n)} = \frac{d\mathbf{S}_{NLAC,NC}}{d(\Delta t_n)}$$

durch eine Differentiation der zusammengesetzten Funktionen nach

$$\frac{d\mathbf{X}_{n,1}}{d(\Delta t_n)} = \sum_{j=1}^{NLAC} \frac{d\mathbf{S}_{NLAC,NC}}{d(\Delta t_j^S)} \cdot \frac{d(\Delta t_j^S)}{d(\Delta t_n)}$$

und unter Anwendung einer gleichmäßigen Aufteilung in Unterintervalle nach

$$\Delta t_j^S = \frac{\Delta t_n \cdot \tau_1}{NLAC} ; j = 1, \dots, NLAC$$

der Zusammenhang

$$\frac{d\mathbf{X}_{n,1}}{d(\Delta t_n)} = \sum_{j=1}^{NLAC} \frac{d\mathbf{S}_{NLAC,NC}}{d(\Delta t_j^S)} \cdot \frac{\tau_1}{NLAC}$$

abgeleitet werden. Dies ermöglicht zudem die Verkettung zu vereinfachen mit

$$\frac{d\hat{\mathbf{S}}_j}{d(\Delta t_n)} = \frac{d\hat{\mathbf{S}}_j}{d\hat{\mathbf{S}}_{j,0}} \frac{d\mathbf{D} \cdot \hat{\mathbf{S}}_{j-1}}{d(\Delta t_n)} + \frac{d\mathbf{S}_j}{d(\Delta t_j^S)} \cdot \frac{\tau_1}{NLAC} \text{ mit } j = 1, \dots, NLAC$$

und

$$\frac{d\hat{\mathbf{S}}_0}{d(\Delta t_n)} = \mathbf{0}$$

und entsprechend

$$\frac{d\mathbf{X}_{n,1}}{d(\Delta t_n)} = \frac{d\hat{\mathbf{S}}_{NLAC,NC}}{d(\Delta t_n)}.$$

Die restlichen Gradienten der originalen Kollokationspunkte  $(2, 3, \dots, NC)$  können unter modifizierter Startbedingung der Verkettung berechnet werden. Zusätzlich gilt für die Beziehungen der Unterintervalle zu der originalen Elementlänge

$$\Delta t_j^S = \frac{\Delta t_n \cdot (\tau_{(2,3,\dots,NC)} - \tau_{(1,2,\dots,NC-1)})}{NLAC}; j = 1, \dots, NLAC.$$

Dies ergibt für die Gradienten nach der Elementlänge des originalen Elementes in Kurzform:

$$\begin{aligned} \frac{d\hat{\mathbf{S}}_j}{d(\Delta t_n)} &= \frac{d\hat{\mathbf{S}}_j}{d\hat{\mathbf{S}}_{j,0}} \frac{d\mathbf{D} \cdot \hat{\mathbf{S}}_{j-1}}{d(\Delta t_n)} \\ &+ \frac{d\mathbf{S}_j}{d(\Delta t_j^S)} \cdot \frac{\tau_{(2,3,\dots,NC)} - \tau_{(1,2,\dots,NC-1)}}{NLAC} \text{ mit } j = 1, \dots, NLAC, \end{aligned}$$

$$\frac{d\hat{\mathbf{S}}_0}{d(\Delta t_n)} = \frac{d\mathbf{X}_{n,(1,2,\dots,NC-1)}}{d(\Delta t_n)}$$

und

$$\frac{d\mathbf{X}_{n,(2,3,\dots,NC)}}{d(\Delta t_n)} = \frac{d\hat{\mathbf{S}}_{NLAC,NC}}{d(\Delta t_n)}.$$

Hierbei ist anzumerken, dass die Unterintervallanzahl  $NLAC$  zwischen den originalen Kollokationspunkten je nach Approximationsgenauigkeit unterschiedlich sein kann. Dies ist einfach umzusetzen, indem  $NLAC$  in den Untersimulationen als variabel definiert wird. Im Folgenden wird der Einfluss einer solchen Untersimulation auf die Gradientenapproximation nach den Elementlängen veranschaulicht.

### Fallbeispiel (Gradientenberechnung für bewegte Elemente)

In diesem Beispiel wird die Gradientenberechnung für die Erweiterung der bewegten Elemente (qMFE) in Bezug auf die Approximationsgenauigkeit der Zustandsgrößen an dem Beispiel aus Gleichung (4.4) untersucht. Wird die analytische Lösung aus Gleichung (4.5) betrachtet, ergibt sich wieder für die Differenzierung nach  $u$

$$\frac{dx(t)}{du} = t \cdot x_0 \cdot e^{u \cdot t}$$

und nach  $x_0$

$$\frac{dx(t)}{dx_0} = e^{u \cdot t}.$$

#### 4 Erweiterungen der quasi-sequentiiellen Methode

Der Gradient nach einem finiten Elemente bedarf einer Modifizierung von Gleichung (4.5) nach

$$x(t_i) = x_0 \cdot e^{u \cdot t_i} ; i = 1, 2, \dots, NC$$

mit

$$t_i = \Delta t \cdot \tau_i$$

für  $t_0 = 0$  um die Abhängigkeit von  $\Delta t$  aufzuzeigen. Nach Differenzierung ergibt sich

$$\frac{dx(t_i)}{d(\Delta t)} = u \cdot \tau_i \cdot x_0 \cdot e^{u \cdot \Delta t \cdot \tau_i}.$$

Diese analytischen Lösungen können zum Vergleich genutzt werden. Dabei wird eine unzureichende Approximation mit  $NL = 1$  (siehe Abbildung 4.3 und 4.4 unten) aus Kapitel 4.2.3 erneut angewendet (für Variante 1 ein relativer Fehler im Bereich von  $\epsilon_{rel} = 10^{-4}$  und Variante 2 eine Toleranz von  $tol = [0, 15 \ 0, 44 \ 0, 71]$ ). Es ergeben sich die relativen Fehler der Gradienten an den original Kollokationspunkten, wie aus Abbildung 4.10 nach  $u$ , 4.11 nach  $x_0$  und 4.14 nach  $\Delta t$ . Die erhöhte Gradientenapproximation wirkt sich hierbei wieder grundsätzlich positiv auf die Konvergenzeigenschaften eines NLP-Lösers aus.

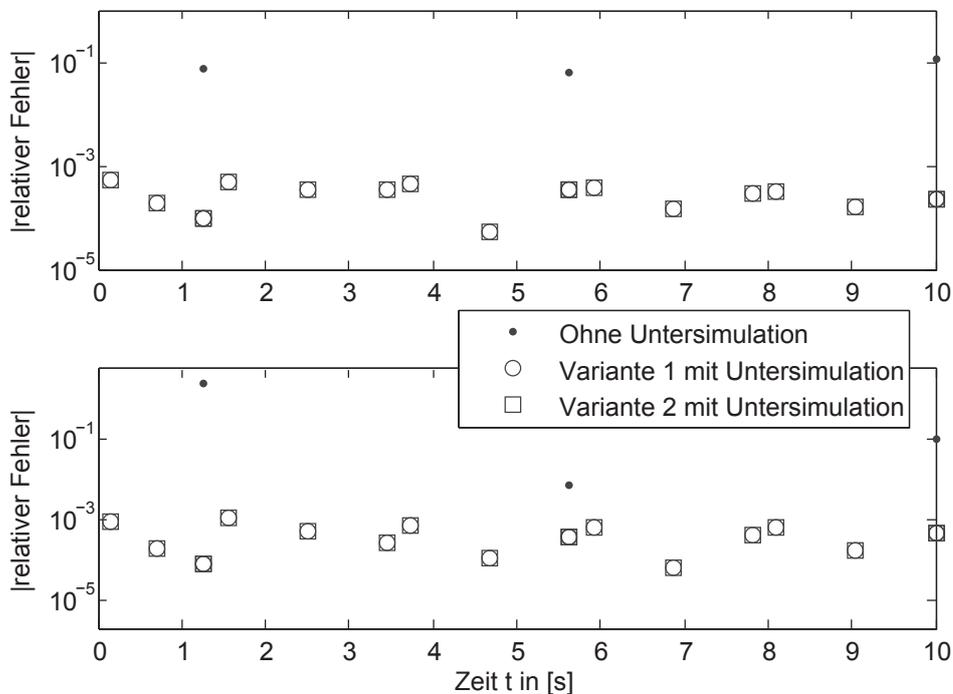


Abbildung 4.14: Approximationsgüte  $\frac{dx}{d(\Delta t)}$  (oben:  $u = -0,25s^{-1}$ , unten:  $u = 0,25s^{-1}$ ). Dargestellt werden zusätzlich die relativen Fehler zur analytischen Lösung für die Kollokationspunkte der Untersimulationen.

## 4.4 Anwendung

In diesem Kapitel werden die Eigenschaften der Erweiterungen an einer konkreten Anwendung untersucht. Dabei wird die Lösung in Bezug auf die kontinuierliche Problemstellung qualitativ beurteilt, die Konvergenzeigenschaften verglichen und der Berechnungsaufwand aufgezeigt.

### 4.4.1 Dynamische Optimierung eines Fermentierungsprozesses

Das betrachtete Optimalsteuerungsproblem stellt eine optimale Fahrweise eines Fermentierungsprozesses (oder auch Bierbrauprozesses) im Batchbetrieb dar und wurde aus [53] entnommen. Hier wird die Maximierung der Ethanolkonzentration im fertigen Sud und die Minimierung der benötigten Batchzeit gleichzeitig betrachtet.

#### Das kontinuierliche Optimierungsproblem

Im Medium des Fermenters befinden sich die Zuckerarten Glukose  $g(t)$ , Maltose  $m(t)$  und Maltotriose  $n(t)$ . Die dynamischen Zusammenhänge werden durch ein nichtlineares DAE-System für Reaktionsraten, für das Temperaturverhalten  $temp(t)$  und für die Relationen der Biomasse  $xb(t)$  und zum Ethanolgehalt  $e(t)$  beschreiben durch:

$$\begin{aligned}
 \frac{dg(t)}{dt} &= -\mu_1(g(t)) xb(t) \\
 \frac{dm(t)}{dt} &= -\mu_2(m(t), g(t)) xb(t) \\
 \frac{dn(t)}{dt} &= -\mu_3(n(t), m(t), g(t)) xb(t) \\
 \frac{dtemp(t)}{dt} &= \frac{1}{\rho C_P} \left[ \Delta H_{FG} \frac{dg(t)}{dt} + \Delta H_{FM} \frac{dm(t)}{dt} + \Delta H_{FN} \frac{dn(t)}{dt} - u(t) (temp(t) - TEMP_C) \right] \\
 xb_0 + R_{X_g}(g(t_0) - g(t)) + R_{X_m}(m(t_0) - m(t)) + R_{X_n}(n(t_0) - n(t)) - xb(t) &= 0 \\
 R_{E_g}(g(t_0) - g(t)) + R_{E_m}(m(t_0) - m(t)) + R_{E_n}(n(t_0) - n(t)) - e(t) &= 0.
 \end{aligned} \tag{4.20}$$

Für eine detaillierte Darstellung des Modelles sei aus Gründen der Übersichtlichkeit an dieser Stelle auf den Anhang A.1 verwiesen. Die Prozesskontrolle wird über eine Kühlungsrate  $u(t)$  realisiert, für welche eine untere Schranke ( $u_{min} = 0 kJ/h/m^3 / ^\circ C$ ) und eine obere Schranke ( $u_{max} = 30 kJ/h/m^3 / ^\circ C$ ) nach

$$u_{min} \leq u(t) \leq u_{max} \tag{4.21}$$

gilt. Während des Fermentierungsprozesses soll weiterhin der Zeitverlauf der Temperatur (aus Gründen der Produktqualität) nach

$$temp_{min} \leq temp(t) \leq temp_{max} \tag{4.22}$$

mit  $temp_{min} = 0^\circ C$  und  $temp_{max} = 12^\circ C$  begrenzt sein. In [53] wird die Zielfunktion aus einer Gewichtung  $w$  von maximaler Ethanolkonzentration im fertigen Sud  $e(t_f)$  und minimaler Batchzeit  $t_f$  gleichzeitig durch

$$\min_{u(t), t_f} \varphi = \min_{u(t), t_f} \{-e(t_f) + w \cdot t_f\} \quad (4.23)$$

und mit  $w = 3mol/h/m^3$  betrachtet. Das Optimalsteuerungsproblem besteht dementsprechend aus Gleichung (4.20), (4.21), (4.22) und (4.23). Dies stellt ein nichtlineares dynamisches Optimierungsproblem mit Beschränkungen des Zeitverlaufes einer Zustandsgröße dar, welches durch das Maximumprinzip in [53] und auch bereits numerisch ohne Kontrolle der Approximationsgenauigkeit gelöst wurde [92].

### Das diskretisierte Optimierungsproblem

In der Lösung von [53], spielen Umschaltvorgänge in der optimalen Steuergröße eine entscheidende Rolle (Bang-Bang-Steuerung mit singulärem Abschnitt), für welche die Eigenschaften der bewegten Elementstrategie und die dafür notwendige Approximationskontrolle untersucht werden sollen.

Die Diskretisierung des Optimierungsproblems basiert auf der Approximation der Zustände bzw. der Modellgleichungen (4.20) durch die orthogonale Kollokation aus Kapitel 3.3 und den Erweiterungen der quasi-sequentiellen Methode aus Kapitel 4.2.3 und 4.3.1 mit der entsprechenden abgeleiteten Gradientenberechnung (Kapitel 4.2.4 und 4.3.2). Hierfür werden die Kollokationsordnung  $NC = 3$  und für die Diskretisierung der Zeitachse  $NL = 40$  finite Elemente gewählt. Es ergibt sich für stückweise konstante Steuerungsparametrisierung

$$\mathbf{U} = [u_1, u_2, \dots, u_{40}]^T$$

und für die finiten Elemente:

$$\Delta \mathbf{T} = [\Delta t_1, \Delta t_2, \dots, \Delta t_{40}]^T.$$

Die Identifikation von Umschaltzeitpunkten und die Behandlung von Zeitoptimalität erfordert eine bewegte Elementstrategie. Dieses Vorgehen führt zu dem reduzierten NLP-Problem

$$\min_{\mathbf{U}, \Delta \mathbf{T}} \left\{ -e_{t_f} + w \sum_{i=1}^{NL} \Delta t_i \right\}$$

unter den Nebenbedingungen

$$u_{min} \leq \mathbf{U} \leq u_{max} ; \Delta \mathbf{T}_{min} \leq \Delta \mathbf{T} \leq \Delta \mathbf{T}_{max} ; temp_{min} \leq \mathbf{TEMP} \leq temp_{max},$$

welches in der Optimierungsschicht zu lösen gilt (vergl. Lösungsschema aus Abbildung 4.12 und Gleichung (4.14)). Dabei entspricht  $e_{t_f}$  dem letzten Kollokationspunkt des Zustandes für die Ethanolkonzentration und  $TEMP$  dem diskreten Zustandsvektor der Temperatur. Die konkreten Werte und deren benötigten Gradienten nach der parametrisierten Steuerung und den finiten Elementlängen werden in der Simulationsschicht entsprechend berechnet. Die Softwareumsetzung erfolgte in Form einer Testumgebung in MATLAB mit automatischer Differenziation durch INTLAB [116] für die Gradientenberechnung und SNOPT [58] zur Lösung des NLP (Intel Core 2 duo mit 2.4GHz und Microsoft Windows XP ohne Parallelisierung). Die Anfangswerte in der Optimierungsschicht wurden zu

$$\mathbf{U}^0 = [30, \dots, 30]^T \quad (4.24)$$

und

$$\Delta \mathbf{T}^0 = [400/NL, \dots, 400/NL]^T \quad (4.25)$$

angenommen (hier ist entsprechend  $t_{f,0} = 400h$ ) oder die Vektoren werden vollständig zufällig erzeugt. Die bewegte Elementmethode bedingt eine Definition von Grenzen für die freien Zeitintervalle, es sei hierbei  $\Delta \mathbf{T}_{min} = \Delta \mathbf{T}^0/3$  und  $\Delta \mathbf{T}_{max} = 5\Delta \mathbf{T}^0/3$  angenommen.

#### 4.4.2 Die Lösung

Dieses diskretisierte Optimierungsproblem wird durch bewegte Elemente in der Optimierungsschicht gelöst. Dabei werden in der Simulationsschicht mittels Variante 1 und Variante 2 die Approximationskontrolle gesichert. An dieser Stelle soll zusätzlich die Lösung des Optimierungsproblems ohne Approximationskontrolle zum Vergleich dienen bzw. zur Motivation eine Kontrolle zu implementieren. Die Annahme der Startbedingungen aus Gleichung (4.24) und (4.25) führt zu den Lösungen aus Abbildung 4.15. Es sei hier die prozess-bezogene Deutung der Ergebnisse vernachlässigt und somit die optimalen Zustandsverläufe nicht gezeigt (siehe hierfür Anhang A.1).

Die obere Grafik in Abbildung 4.15 zeigt die berechnete optimale Kühlungsstrategie in Abhängigkeit der Approximationskontrolle. Durch die Variante 1 und Variante 2 wird eine nahezu gleiche Kühlungsstrategie berechnet, welche zudem der in [53] dargestellten Küh-

lungsstrategie entsprechen. Die Lösung ohne Approximationskontrolle ermittelt jedoch einen verfrühten Umschaltzeitpunkt  $t_1$ .

Die Elementlängen stellen einen Bestandteil der numerischen Lösungsstrategie dar und haben keine direkte Bedeutung für die Prozessführung. Diese beeinflussen direkt die Approximationsgenauigkeit der orthogonalen Kollokation eines finiten Elementes mit dem entsprechend erhöhtem (lange Elemente) oder verbessertem (kurze Elementen) Rechenaufwand in der Simulationsschicht. Wie die mittlere Grafik in Abbildung 4.15 zeigt, sind die optimalen Elementlängen wiederum für die Variante 1 und Variante 2 nahezu identisch. Auffällig sind hierbei die langen Elementlängen ohne Anwendung einer Approximationskontrolle, was entsprechend die Approximation zusätzlich verschlechtert. Ein derartiger Effekt ist bei der Variante 1 und Variante 2 nicht zu beobachten, was an dieser Stelle die Approximation zusätzlich unterstützt.

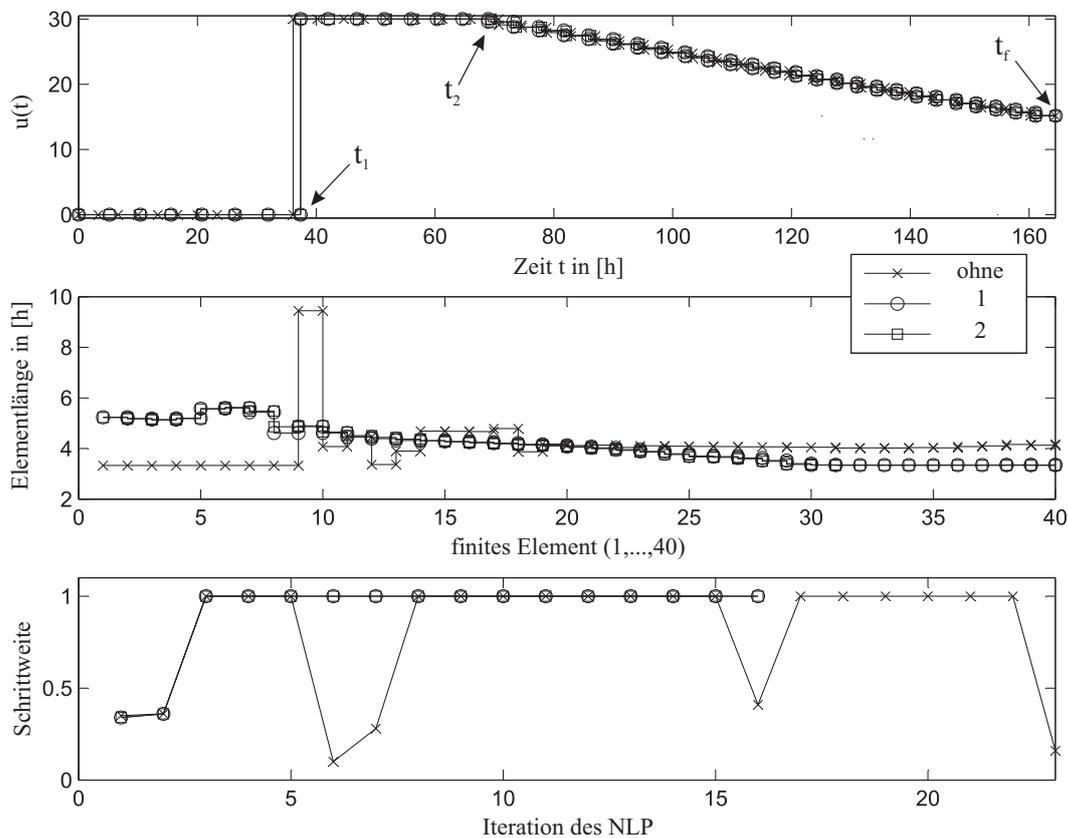


Abbildung 4.15: Vergleich der Lösungsstrategien

Die Schrittweiten während der Lösung des diskretisierten Optimierungsproblems werden in jeder Iteration durch den NLP-Algorithmus bestimmt. Dies geschieht anhand des momentan zu lösenden Problems und beeinflusst die Konvergenzgeschwindigkeit. Im Prinzip benötigen kurze Schrittweiten (im Vergleich zu langen Schrittweiten) mehr Iterationen, um in das gesuchte Optimum zu gelangen und benötigen zusätzlichen Rechenaufwand für die Ermittlung

der jeweiligen Schrittweiten. Die während der iterativen Lösung berechneten Schrittweiten sind im unterem Teil von Abbildung 4.15 jeweils in Abhängigkeit der Approximationskontrolle dargestellt. Interessant ist hierbei der Unterschied in den Konvergenzeigenschaften des NLP-Lösers mit und ohne Approximationskontrolle. Hierbei benötigt das Verfahren ohne Approximationskontrolle konvergenzverschlechternde, kleinere Schrittweiten (3 zusätzliche) und weitaus mehr Iterationen (23 zu 16) im Vergleich zu Variante 1 und Variante 2.

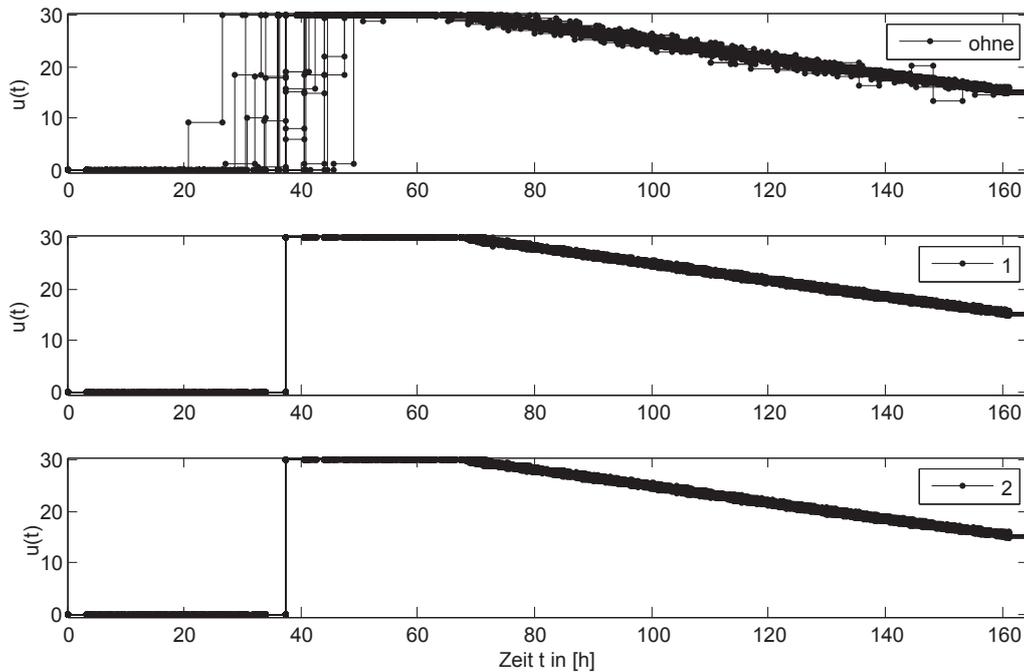


Abbildung 4.16: Vergleich der Lösungsstrategien für zufällige Startwerte

Der Lösungsweg des diskretisierten Optimierungsproblems durch SNOPT ist grundsätzlich abhängig von den gegebenen Startwerten. Dies betrifft prinzipiell alle gradienten-basierte Lösungsverfahren. Im Folgendem wird untersucht, ob die oben gezeigten Eigenschaften durch die Wahl der Startwerte bedingt sind. Um diesen Einfluss zu reduzieren, werden zufällige Startwertsets gleichverteilt zwischen 0 und 30 für Gleichung (4.24) und zwischen 5 und 10 für Gleichung (4.25) erzeugt. Dabei werden entsprechend 100 Sets mit jeweils  $NL \cdot 2$  Zufallsrealisierungen erzeugt und anschließend das Optimierungsproblem gelöst. Dabei wird für die einzelnen Varianten das selbe Set verwendet. Hierbei stellt Abbildung 4.16 die optimalen Kühlungsstrategien dieser Startwertsets und der jeweiligen Approximationskontrolle dar (Oben ohne Kontrolle, Mitte Variante 1, Unten Variante 2). Es ist leicht zu erkennen, dass eine Approximationskontrolle für qualitativ gute Lösungen des untersuchten Beispiels unumgänglich ist (es wird an dieser Stelle von einem Vergleich der Elementlängen und Schrittweiten abgesehen). Es ergibt sich eine erhöhte Robustheit des Lösungsverfahrens bezüglich der Startwertwahl unter Approximationskontrolle. So wird z.B. der Umschaltzeitpunkt für jedes

zufällige Startwertset mit Variante 1 und Variante 2 richtig berechnet und zudem wird in 26 Fällen (diese sind nicht in Abbildung 4.16 eingezeichnet) ohne Approximationskontrolle keine Konvergenz des NLP-Lösers erzielt.

Es werden nun zum Abschluss die Ergebnisse rekapituliert: Die Einführung von Unterintervallen in einer NLP-Iteration führt unausweichlich zu einer strukturellen Änderung der Kollokationsgleichungen und der Gradientenberechnung. Auch wenn dabei das betrachtete NLP-Problem die gleiche Anzahl von Variablen enthält, folgt aus einer Untersimulation verschiedene Funktionswerte, Gradienten und entsprechende KKT (Karush-Kuhn-Tucker) Bedingungen, als wenn keine Untersimulation durchgeführt wird. Wie die hier gezeigten Beispiele verdeutlichen, stört dies nicht die Konvergenz des NLP-Algorithmus. Sondern die Problemstellung wird in jeder Iteration auf einen höherem Approximationsgrad angehoben, auf welchem der NLP-Löser arbeiten kann, bis die Konvergenz erfolgt ist. Dieses Vorgehen hat eine starke Verbesserung der Konvergenz zur Folge, welche jedoch durch einen erhöhten Rechenaufwand in der Simulationsschicht erkaufte wird. Unabhängig von der diskutierten Genauigkeit der optimalen Ergebnisse soll im Folgendem auch der benötigte Rechenaufwand verglichen werden.

#### 4.4.3 Der Berechnungsaufwand

Die Berechnungszeit (oder auch CPU-Zeit) für die Ergebnisse aus Abbildung 4.15 betragen ohne Approximationskontrolle:  $t_{CPU,ohne} = 195,74s$ , für Variante 1:  $t_{CPU,1} = 125,26s$  und Variante 2:  $t_{CPU,2} = 123,41s$ . Die Methoden der Approximationskontrolle resultieren somit in ca. 37% und 35% weniger Rechenaufwand als das Vorgehen ohne Approximationskontrolle. Trotz der Untersimulationen und der dazugehörigen aufwendigeren Gradientenberechnung wird dieser Geschwindigkeitsgewinn über wenige Iterationen und geringerem Aufwand in der Schrittweitensuche des NLP-Lösers erzielt.

An dieser Stelle soll im Folgendem ein detaillierter Einblick in den Rechenaufwand der Approximationskontrolle gegeben werden. Die totale CPU-Zeit für die Fehlervorhersage oder Toleranzberechnung (Zeit für die Detektion) während der Lösung des Optimierungsproblems beträgt für Variante 1:  $t_{detekt,1} = 2,1s$  und Variante 2:  $t_{detekt,2} = 0,38s$ . Dementsprechend kann durch die Verwendung von Variante 2, ca. 82% der Zeit für die Fehlerdetektion eingespart werden. Die Detektionszeit ist jedoch relativ gering zu der in Summe benötigten Rechenzeit und entsprechend ergibt sich ein geringer Unterschied von ca. 1,5% für CPU-Zeit pro Iteration ( $t_{iter,1} = 7,83s$  zu  $t_{iter,2} = 7,71s$ ) der beiden Varianten. Dieser Unterschied kann jedoch für große Systeme entsprechend höher ausfallen.

Die Analyse zufälliger Startwerte ermöglicht eine Aussage über die vom Startwert unabhängige Performance. Die dazugehörigen Rechenzeiten sind in Abbildung 4.17 in Form von Histogrammen dargestellt. Der obere Teil illustriert die Häufigkeit der gesamten CPU-Zeit in Abhängigkeit der Approximationskontrolle. Hierbei zeigt das Verfahren ohne Approximationskontrolle eine weitaus schlechtere Rechenzeitperformance (150s – 1250s) als Variante 1 und Variante 2. Die Verteilungen von Variante 1 und Variante 2 sind zudem stark linksschief und zeigen damit zusätzlich ein besseres Verhalten. Die beiden unteren Teile aus Abbildung 4.17 vergleichen die Verfahren der Approximationskontrolle. Hierbei ist links die Häufigkeit der CPU-Zeit der Untersimulation (mit dazugehöriger Gradientenberechnung) und rechts die CPU-Zeit der Fehlerdetektion dargestellt. In Summe erzielt Variante 2 durch die kurze Detektionszeit einen Vorteil im Rechenaufwand, im Vergleich zu Variante 1.

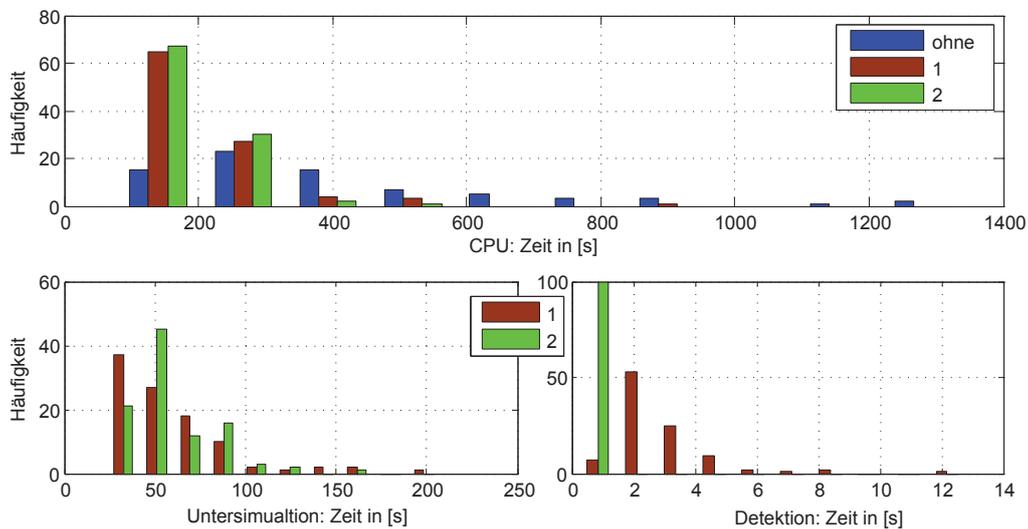


Abbildung 4.17: Rechenaufwand für zufällige Startwerte

Werden die Ergebnisse auch hier rekapituliert, ergeben sich trotz erhöhtem Aufwand in der Simulationsschicht Vorteile in der Gesamtrechenzeit. Damit ist auch von Seiten der Rechenzeit eine Approximationskontrolle für das untersuchte Beispiel zu empfehlen.

## 4.5 Die C++ Bibliothek „QuasiOPT“

Die in Kapitel 4.4.3 gezeigten Rechenzeiten resultieren größtenteils aus der Implementierung der Testumgebung in MATLAB und der entsprechenden Verwendung von INTLAB. Um die Berechnung zu beschleunigen, wurde deshalb während der Bearbeitung der vorliegenden Arbeit eine Umsetzung der Erweiterungen in einer C/C++ Bibliothek angestrebt (dabei dient der Prototyp in MATLAB als Grundlage). Hierbei ist der Fokus während der Implementierung der Bibliothek auf die Identifikation von dynamischen Regulationsstrategien im Stoffwechsel

gelegt (betrifft die nachfolgenden Kapitel). Die programmiertechnischen Entwicklungen sind in den daran aktiv beteiligten Diplomarbeiten [87, 94, 156] zu finden. In der Diplomarbeit [94] wurde die Konfiguration des Bibliotheksaufrufs grundlegend überarbeitet und eine ressourcenschonende Berechnung der Gradienten umgesetzt. Zusätzlich wurde eine Vielzahl von Implementierungsbeispielen einfacher Probleme der Arbeit beigelegt. Dieses Kapitel stellt den Aufbau und momentanen Leistungsstand der entstandenen Bibliothek in Kurzform dar.

##### 4.5.1 Module

Eine Aufteilung in Module dient die Übersichtlichkeit des umgesetzten Projektes zu bewahren. Dies vereinfacht eine spätere Softwarepflege und reduziert den Aufwand für Softwareerweiterungen.

**collocation.cpp** Hier wird die Schnittstelle zur Simulation umgesetzt und die Berechnungen der einzelnen Intervalle ausgeführt. Die Approximationskontrolle wird dabei überprüft und falls nötig eine Untersimulation durchgeführt.

**solver.cpp** Dies beinhaltet die Berechnungen der Zustände und die Gradientenberechnung nach der parametrisierten Steuerung, nach der finiten Elementlänge und nach den Anfangszuständen eines Elementes. Dies wird auch während einer Untersimulation genutzt.

**error.cpp** In diesem Modul sind die Algorithmen der Fehlerabschätzungen implementiert. Dabei liegt der Schwerpunkt auf der Untersuchung von Stoffwechselwegen. Diese weisen verschiedenste Strukturen und unterschiedlichste kinetische Parameter auf, weshalb ausschließlich die Fehlerabschätzung (Variante 1) umgesetzt wurde.

**chainrule.cpp, chainrule\_dXdX0.cpp, chainrule\_const.cpp** In diesen Dateien sind die verschiedenen Verkettungen hinterlegt. Hierbei beinhaltet die Datei **chainrule.cpp** die Verkettung der Ableitungen über alle Intervalle, die Datei **chainrule\_dXdX0.cpp** nur die Algorithmen für die Gradienten nach den Anfangszuständen und die Datei **chainrule\_const.cpp** die angepasste Gradientenberechnung für konstante parametrisierte Steuerung (z.B. während einer Untersimulation).

**matrixoperations.cpp, buffers.cpp, lagrange.cpp** Hier wurde Quelltext hinterlegt, welcher häufig Anwendung findet. Dies betrifft Matrizenmultiplikationen in **matrixoperations.cpp** (ermöglicht schnelle Modifikation der Berechnungsverfahren) und Speicherallokationen in **buffers.cpp**. Die Parameter der orthogonalen Kollokation sind in **lagrange.cpp** zu finden. Dies ermöglicht einen leichten Austausch der Diskretisierung durch z.B. in [68] dargestellte Verfahren.

### 4.5.2 Konfiguration

Die Implementierung der Konfiguration wird durch eine C-Struktur realisiert. Diese Struktur bestimmt nicht nur die Definition der Problemstellung, sondern beinhaltet auch gleichzeitig die berechneten Simulationsergebnisse und gesuchten Gradienteninformationen. Es lassen sich hierbei, durch ein einfaches Vorgehen, komplexe Anforderungen formulieren. Dies betrifft z.B. die Art der freien Variable (zeitvariant oder zeitinvariant), die Art der Beschränkungen (Zeitverlauf oder Endbedingungen) und die Definition der Gradientenberechnung (keine, die der Steuervariablen, bewegte Elemente, freie Anfangsbedingung, Auswahl bestimmter Zustände). Die Konfiguration der Simulation wird durch verschiedene Funktionen durchgeführt, welche im Folgendem kurz beschrieben werden:

**new\_Simulation(...)** Diese Funktion erzeugt eine neue Konfiguration in Form einer Struktur.

Hier wird als minimale Angabe ein Zeiger auf die Funktion für die Modellgleichungen, die Anzahl der Systemzustände und die Anzahl der Intervalle gefordert. Hier können bereits existierende Puffer für die Anfangsbedingungen und Elementlängen angegeben werden (ohne Übergabe werden diese automatisch alloziert).

**set\_control\_variables(...)** Hier sind Angaben zu den freien Variablen gefordert (oder auch Steuervariablen). Die Funktion benötigt die Anzahl an Steuervariablen und ob diese zeitvariant oder zeitinvariant sind. Sind beide gleichzeitig von Interesse kann dies über einen zusätzlichen Zustand umgesetzt werden. Es kann auch hier wieder ein Puffer mit den jeweiligen Werten übergeben werden, wenn nicht wird dieser automatisch alloziert.

**enable\_gradients(...)** Diese Funktion aktiviert die Gradientenberechnung bei Simulationsaufruf und definiert die Art der Begrenzung, je nachdem ob dies Zeitverläufe und/oder Endbedingungen (auch für Zielfunktionsberechnung geeignet) betrifft. Weiterhin wird definiert, welche Gradienteninformationen (die der Steuervariablen, der bewegte Elemente und die von freien Anfangsbedingung) von Interesse sind. Es besteht wieder die Möglichkeit vorhandene Puffer für die berechneten Gradienten zu übergeben oder diesen automatisch allozieren zu lassen.

**set\_gradients\_system\_states(...)** Sind die Gradienten nur für eine Teilmenge der Systemzustände von Interesse (dies tritt sehr häufig auf), dann liegt es nahe, aus Gründen der Rechenzeit und der Speicherverwaltung, die betreffenden Berechnungen darauf zu begrenzen. Dies wird durch die Funktion *set\_gradients\_system\_states(...)* konfiguriert. Hierbei ist wieder die Angabe einer Begrenzung des Zeitverlaufes oder das Interesse auf Endbedingungen gefordert.

**set\_variable\_x0(...)** Durch diese Funktion lassen sich Probleme mit freien Anfangsbedingungen der Systemzustände untersuchen. Dies kann auch gleichzeitig mit den oben genannten Konfigurationsmöglichkeiten betrachtet werden. Es lässt sich definieren, welche Gradienten nach den Anfangsbedingungen (auch eine Teilmenge der Zustände) von Interesse sind.

**delete\_Simulation(...)** Sind alle Simulationsdurchläufe beendet und die gesuchten Ergebnisse gesichert, gibt diese Funktion alle automatisch allozierten Speicherbereiche wieder frei. Dies dient einer effektiven Fürsorge zur Vermeidung von Speicherlecks und sollte dementsprechend auch ausgeführt werden. Hinweis: manuell allozierte Speicher muss auch manuell wieder freigegeben werden.

### 4.5.3 Eingabepuffer und Ausgabepuffer

Eingabepuffer stellen Daten dar, welche für einen Simulationsaufruf benötigt werden, wobei Ausgabepuffer die Daten der Simulationsergebnisse bereitstellen. Diese Puffer werden im Folgendem in Kurzform beschrieben.

#### Eingabepuffer

**X0** Das Feld beinhaltet die Anfangszustände  $\mathbf{x}_0$  der Modellgleichungen nach:

$$\mathbf{X0} = \begin{bmatrix} x_{0,1} & x_{0,2} & \cdots & x_{0,NX} \end{bmatrix}.$$

**$\Delta\mathbf{T}$**  Hier werden alle Elementlängen übergeben:

$$\Delta\mathbf{T} = \begin{bmatrix} \Delta t_1 & \Delta t_2 & \cdots & \Delta t_{NL} \end{bmatrix}.$$

**U** Dieser Puffer beinhaltet die Steuerungsparametrisierung für jedes Element nacheinander:

$$\mathbf{U} = \begin{bmatrix} u_{1,1} & u_{1,2} & \cdots & u_{1,NU} & u_{2,1} & \cdots & u_{2,NU} & \cdots & u_{NL,NU} \end{bmatrix}.$$

**dae** Dies entspricht einem Puffer aus Flags, welche einen Systemzustand als algebraisch (true oder 1) oder differenziell (false oder 0, entspricht der standardmäßigen Initialisierung) definieren nach:

$$\mathbf{dae} = \begin{bmatrix} \text{dae}_1 & \text{dae}_2 & \cdots & \text{dae}_{NX} \end{bmatrix}.$$

**Ausgabepuffer**

**X** Dieser Puffer beinhaltet das Simulationsergebnis der diskretisierten Zustände nach:

$$X = \begin{bmatrix} \mathbf{x}_{1,1} & \mathbf{x}_{1,2} & \cdots & \mathbf{x}_{1,NC} & \mathbf{x}_{2,1} & \cdots & \mathbf{x}_{2,NC} & \cdots & \mathbf{x}_{NL,NC} \end{bmatrix}.$$

**dXdU, dXddt** Diese Puffer beinhalten die Ableitungen der diskretisierten Zustandstrajektorien nach der parametrisierten Steuerung bzw. Elementlängen:

$$\begin{aligned} dXdU &= \begin{bmatrix} \frac{d\mathbf{X}_{1,1}}{d\mathbf{U}_1} & \frac{d\mathbf{X}_{1,2}}{d\mathbf{U}_1} & \cdots & \frac{d\mathbf{X}_{1,NC}}{d\mathbf{U}_1} & \frac{d\mathbf{X}_{2,1}}{d\mathbf{U}_1} & \cdots & \frac{d\mathbf{X}_{2,NC}}{d\mathbf{U}_1} & \cdots & \frac{d\mathbf{X}_{NL,NC}}{d\mathbf{U}_1} & \cdots \\ \frac{d\mathbf{X}_{2,1}}{d\mathbf{U}_2} & \cdots & \frac{d\mathbf{X}_{NL,NC}}{d\mathbf{U}_2} & \frac{d\mathbf{X}_{3,1}}{d\mathbf{U}_3} & \cdots & \frac{d\mathbf{X}_{NL,NC}}{d\mathbf{U}_3} & \cdots & \frac{d\mathbf{X}_{NL,NC}}{d\mathbf{U}_{NL}} \end{bmatrix} \\ dXddt &= \begin{bmatrix} \frac{d\mathbf{X}_{1,1}}{d\Delta t_1} & \frac{d\mathbf{X}_{1,2}}{d\Delta t_1} & \cdots & \frac{d\mathbf{X}_{1,NC}}{d\Delta t_1} & \frac{d\mathbf{X}_{2,1}}{d\Delta t_1} & \cdots & \frac{d\mathbf{X}_{2,NC}}{d\Delta t_1} & \cdots & \frac{d\mathbf{X}_{NL,NC}}{d\Delta t_1} & \cdots \\ \frac{d\mathbf{X}_{2,1}}{d\Delta t_2} & \cdots & \frac{d\mathbf{X}_{NL,NC}}{d\Delta t_2} & \frac{d\mathbf{X}_{3,1}}{d\Delta t_3} & \cdots & \frac{d\mathbf{X}_{NL,NC}}{d\Delta t_3} & \cdots & \frac{d\mathbf{X}_{NL,NC}}{d\Delta t_{NL}} \end{bmatrix}. \end{aligned}$$

Sind die Steuerparameter über alle Intervalle konstant ergibt sich:

$$dXdU = \begin{bmatrix} \frac{d\mathbf{X}_{1,1}}{d\mathbf{U}} & \frac{d\mathbf{X}_{1,2}}{d\mathbf{U}} & \cdots & \frac{d\mathbf{X}_{1,NC}}{d\mathbf{U}} & \frac{d\mathbf{X}_{2,1}}{d\mathbf{U}} & \cdots & \frac{d\mathbf{X}_{2,NC}}{d\mathbf{U}} & \cdots & \frac{d\mathbf{X}_{NL,NC}}{d\mathbf{U}} \end{bmatrix}.$$

**dXdX0** Der Puffer beinhaltet die Gradienten der diskretisierten Zustandstrajektorien nach den Anfangsbedingungen:

$$dXdX0 = \begin{bmatrix} \frac{d\mathbf{X}_{1,1}}{d\mathbf{x}_0} & \frac{d\mathbf{X}_{1,2}}{d\mathbf{x}_0} & \cdots & \frac{d\mathbf{X}_{1,NC}}{d\mathbf{x}_0} & \frac{d\mathbf{X}_{2,1}}{d\mathbf{x}_0} & \cdots & \frac{d\mathbf{X}_{2,NC}}{d\mathbf{x}_0} & \cdots & \frac{d\mathbf{X}_{NL,NC}}{d\mathbf{x}_0} \end{bmatrix}.$$

**dXdU\_last, dXddt\_last** Diese Puffer beinhalten die Ableitungen der Zustandsendwerte nach der parametrisierten Steuerung bzw. Elementlängen. Es gilt hier:

$$\begin{aligned} dXdU\_last &= \begin{bmatrix} \frac{d\mathbf{X}_{NL,NC,1}}{d\mathbf{U}_{1,1}} & \frac{d\mathbf{X}_{NL,NC,2}}{d\mathbf{U}_{1,1}} & \cdots & \frac{d\mathbf{X}_{NL,NC,NX}}{d\mathbf{U}_{1,1}} & \frac{d\mathbf{X}_{NL,NC,1}}{d\mathbf{U}_{1,2}} & \cdots \end{bmatrix}, \\ dXddt\_last &= \begin{bmatrix} \frac{d\mathbf{X}_{NL,NC,1}}{d\Delta t_1} & \frac{d\mathbf{X}_{NL,NC,2}}{d\Delta t_1} & \cdots & \frac{d\mathbf{X}_{NL,NC,NX}}{d\Delta t_1} & \frac{d\mathbf{X}_{NL,NC,1}}{d\Delta t_2} & \cdots \end{bmatrix}. \end{aligned}$$

**dXdX0\_last** Der Puffer beinhaltet die Gradienten der Zustandsendwerte nach den Anfangsbedingungen:

$$dXdX0\_last = \begin{bmatrix} \frac{d\mathbf{X}_{NL,NC,1}}{d\mathbf{x}_{0,1}} & \frac{d\mathbf{X}_{NL,NC,2}}{d\mathbf{x}_{0,1}} & \cdots & \frac{d\mathbf{X}_{NL,NC,NX}}{d\mathbf{x}_{0,1}} & \cdots & \frac{d\mathbf{X}_{NL,NC,NX}}{d\mathbf{x}_{0,NX_s}} \end{bmatrix}.$$

#### 4.5.4 Eigenschaften und Dokumentation

Die Konfigurationsmöglichkeiten geben einen einfachen Zugang auf eine Vielzahl von Problemstellungen, welche durch die Bibliothek untersucht werden können. Somit ist die entwickelte Software nicht auf eine ausschließliche Betrachtung von Stoffwechselregulationen oder auf Probleme der Optimalen Steuerung beschränkt und kann auf eine Vielzahl von dynamischen Optimierungsproblemen angewendet werden.

Der Bedienkomfort wird ebenfalls durch die einfache Konfiguration unterstützt. Zusätzlich erhöht die Verwendung von ADOL-C für die automatische Berechnung der Gradienten den Komfort enorm (die Software kann unter [128] bezogen werden). Weiterhin wirkt sich hier positiv aus, dass die Modellgleichungen außerhalb der Bibliothek erstellt und durch einen Zeiger an die Bibliothek übergeben werden. Ein wiederholtes Kompilieren der Bibliothek nach der Neudefinition einzelner Modellgleichungen (oder einzelner Parameter) ist somit nicht notwendig. Der Nutzer benötigt zudem nicht zwingend detaillierte Kenntnisse der orthogonalen Kollokation und eine Abwärtskompatibilität durch Weiterentwicklungen der Bibliothek ist auch gegeben. Eine Plattformunabhängigkeit (Windows, Linux, Unix) wird durch das Buildsystem CMake realisiert. Durch CMake besteht die Möglichkeit Projektdateien für unterschiedliche Entwicklungsumgebungen (zum Beispiel *Eclipse* oder *Visual Studio*) zu erstellen. Die Entwickler haben somit die Möglichkeit leicht auf eine ihnen bekannte Programmierungsumgebung umzusteigen. Weitere Informationen zu CMake befinden sich auf der Homepage des Projektes [129] und der dortigen Dokumentation. Die Dokumentation des Projektes wurde mittels *doxygen* erstellt. Dieses freie Werkzeug der Projektdokumentation übernimmt anhand von Anweisungen im Quelltext (einfach ein Ausrufezeichen hinter einem Kommentar) den darauf folgenden Kommentartext in die Projektdokumentation. Informationen über *doxygen* sind auf der Internetseite des Projektes [130] und der Dokumentation zu finden. Die erstellte Dokumentation des Projektes liegt im „Hypertext Markup Language“ (HTML)-Format vor.

#### 4.6 Diskussion

In der vorliegenden Arbeit wurde die quasi-sequentielle Methode durch eine Approximationskontrolle erweitert (vergl. Kapitel 4.2) und ein Ansatz für bewegte Elemente vorgestellt (qMFE genannt, vergl. Kapitel 4.3). Dabei wurde die Approximationskontrolle durch eine Fehlerdetektion und eine entsprechende Untersimulation gesichert. Für die Lösung des Optimierungsproblems wurde eine effiziente Berechnung von notwendigen Sensitivitäten nach den parametrisierten Steuerungen und den bewegten Elementen abgeleitet und umgesetzt.

Das Anwendungsbeispiel (siehe Kapitel 4.4) verdeutlicht die Notwendigkeit einer Approximationskontrolle bei einer bewegten Elementstrategie. Dabei wurde die Konvergenz verbessert, entsprechend die Rechenzeit verringert und zusätzlich ein robusteres Verhalten auf zu schätzende Starttrajektorien (hier im Speziellen sogar auf zufällige) bei der Lösung eines dynamischen Optimierungsproblems erzielt. Dies kann damit erklärt werden, dass der NLP-Löser auf einem hohen Approximationsniveau (gesichert in der Simulationsschicht) die parametrisierte Steuerung und Elemente anpasst, um nur das Optimierungsproblem zu lösen.

Die beiden vorgestellten Methoden zur Approximationskontrolle aus Kapitel 4.2.2 ergeben ein nahezu identisches Verhalten und dazugehörige optimale Steuerungsverläufe. Von Seiten der Rechenzeit ist das empirische Verfahren (Variante 2) der numerischen Abschätzung (Variante 1) durch eine geringere Detektionszeit vorzuziehen. Da jedoch keine analytische Beziehung des empirischen Verfahrens zum vorliegenden Approximationsfehler abgeleitet werden konnte, wird für sich verändernde Modelle des dynamischen Systems (z.B. zeitvariante Systeme bzw. Veränderungen von Modellparametern) die Wahl der Fehlerabschätzung empfohlen.

Weiterhin wurde basierend auf dem Prototyp in MATLAB eine C++ Bibliothek entwickelt, welche funktionell identisch zu den Leistungen der Implementierung in Matlab ist. Jedoch wurde dabei ausschließlich die Fehlerabschätzung (Variante 1) umgesetzt, da in den weiteren Untersuchungen der vorliegenden Arbeit hohe Variationen in den Modellen zu erwarten sind. Zum Abschluss werden die Rechenzeiten dieser C++ Bibliothek im Vergleich zur Implementierung des Prototyps in MATLAB untersucht. Dabei wurden ausschließlich die Berechnungen in der Simulationsschicht durch die Bibliothek mittels eines MATLAB-Interface in C++ ausgelagert. Die Untersuchung für zufällige Startwerte aus Kapitel 4.4 benötigte eine Gesamtzeit von ca. 19623s, eine Auslagerung der Simulationsschicht reduziert dies auf ca. 188s. Somit ergibt sich ein Geschwindigkeitsgewinn vom Faktor  $10^2$ . Dies ermöglicht unter Anwendung der Erweiterungen aus Kapitel 4.2.3 und 4.3.1 die Untersuchungen der Stoffwechselregulation in einer absehbaren Rechenzeit.

*Teile dieses Kapitels wurden in „American Institute of Chemical Engineers Journal“*

*M. Bartl, P. Li and L.T. Biegler. „Improvement of State Profile Accuracy in Nonlinear Dynamic Optimization with the Quasi-Sequential Approach“, 57(8):2185-2197,2011*

*publiziert (im nachfolgendem Text gekennzeichnet mit [9]). In den vorhergehenden Kapiteln wurden somit Ergebnisse und weiterführende Untersuchungen bzw. Softwareentwicklungen aufgezeigt, welche auf diesem Artikel basieren.*

# 5 Optimale Regulation metabolischer Netzwerke in *Escherichia coli*

Die Untersuchung von Regulationen des Stoffwechsels stellt einen wichtigen Schwerpunkt in der momentanen Forschung der Systembiologie dar. In Kapitel 2.2 wurden bereits die grundlegenden Begriffe eingeführt und in Kapitel 2.3 der Stand der Forschung bezüglich der Anwendung von Optimalitätsprinzipien für Stoffwechselanalysen dargestellt. Vorhergehende Arbeiten haben sich mit dem Zusammenhang zwischen transkriptioneller Regulation und der Struktur von Stoffwechselnetzwerken befasst [siehe z.B. 71, 102, 131] oder wie bereits erwähnt, wurden Aktivierungsmuster anhand der Netzwerkposition und der sich daraus ergebenden Distanz erklärt [80, 158]. Die Ausprägung der Regulation metabolischer Netzwerke bezogen auf die Netzwerkstruktur wird in den Arbeiten [102, 131] als nicht sehr stark angegeben. Hierbei wurde entsprechend der Einfluss des dynamischen Verhaltens bisher nur unzureichend untersucht.

In der hier vorliegenden Arbeit wird analysiert, ob sich ein Zusammenhang zwischen transkriptioneller Regulation und der dynamischen Kontrolle (technisch gesehen der Regelung) metabolischer Netzwerke ergibt. Für die Untersuchung dieser Fragestellung gibt es nur wenige Arbeiten - theoretische Ansätze zur Kontrolle metabolischer Netzwerke sind allerdings bereits in [65] zu finden. Eine derartige Kernfrage erneut aufzugreifen, wurde durch einen Gastvortrag von Christoph Kaleta (Friedrich-Schiller-Universität Jena) in Ilmenau initiiert.

## 5.1 Motivation

Es sei ausdrücklich darauf hingewiesen, dass die Datenanalysen dieses Kapitels durch Kollegen von der Friedrich Schiller Universität (Christoph Kaleta, Frank Wessely, Stefan Schuster) und dem Hans Knöll Institut Jena (Reinhard Guthke) durchgeführt bzw. angeleitet wurden. Das hierbei untersuchte Modell von *Escherichia coli* wurde aus [44, iAF1260] entnommen. Es basiert auf Informationen zum Stoffwechsel (z.B. Stöchiometrie) unter Einbindung genetischer Informationen und wird auch als „genome scale“ Stoffwechselnetzwerk bezeichnet.

Die Untersuchung der transkriptionellen Regulation basiert auf Berechnung von elementaren Pfaden mit physiologisch, realisierbaren Flüssen (sogenannte „Elementary flux patterns“ [75]) durch die in [44] angegebenen metabolischen Teilsysteme (siehe oberen Teil der Abbildung 5.1). Hinzu kommen Datenanalysen aus experimentellen Datensätzen für Zeitreihen der Genexpression in *Escherichia coli* der „Many Microbe Microarrays Database“ [43], Proteinabundanzen [95], Informationen zum transkriptionellen Regulationsnetzwerk des Stoffwechsels aus RegulonDB [51] und EcoCyc [77] und Daten zur post-translationalen Regulation von Enzymen aus EcoCyc und Phosida [62]. Die betreffenden Grundlagen der genannten Regulationsarten wurden bereits in Kapitel 2.2 angesprochen. Die zur Motivation dienenden Zusammenhänge der Datenanalyse werden im Folgendem aufgezeigt.

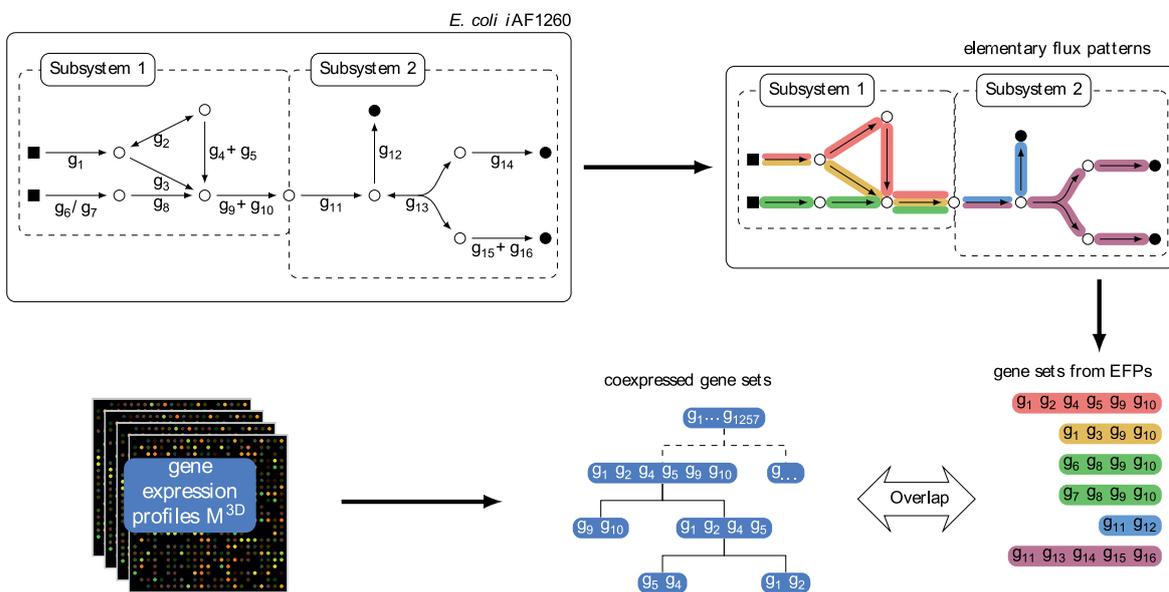


Abbildung 5.1: Das prinzipielle Vorgehen während der Analyse von *Escherichia coli*

Die ermittelten „Elementary flux patterns“ des Stoffwechsels von *Escherichia coli* wurden auf ihre regulatorischen Eigenschaften untersucht (hier die Koexpression der beteiligten Enzyme, siehe dazu unterer Teil von Abbildung 5.1). Dabei wurde festgestellt, dass nur 14,5% der berechneten Pfade einer transkriptionellen Regulation unterliegen. Dies entspricht einer geringen Koordination in Bezug auf die klassischen Vorstellungen der Stoffwechselregulation. Jedoch passt dies in das derzeitig entstehende Bild einer nicht allgegenwärtigen, transkriptionellen Regulation einzelner Stoffwechselschritte [63]. Die Frage warum für die „Elementary flux patterns“ eine derartig geringe Koordination der Expression von Enzymen festgestellt wurde, wird hier über eine Analyse der einzelnen metabolischen Teilsysteme näher betrachtet. Dabei ergeben sich die in Abbildung 5.2 dargestellten Koordinationen der Teilsysteme. Eine stark ausgeprägte Kontrolle wird hierbei durch den Wert 1 beschrieben, wobei dies einer Ko-

expression aller Enzyme in allen Pfaden des Teilsystems entspricht. Hierbei wird ersichtlich, dass die ermittelte Koordination zwischen den Teilsystemen variiert.

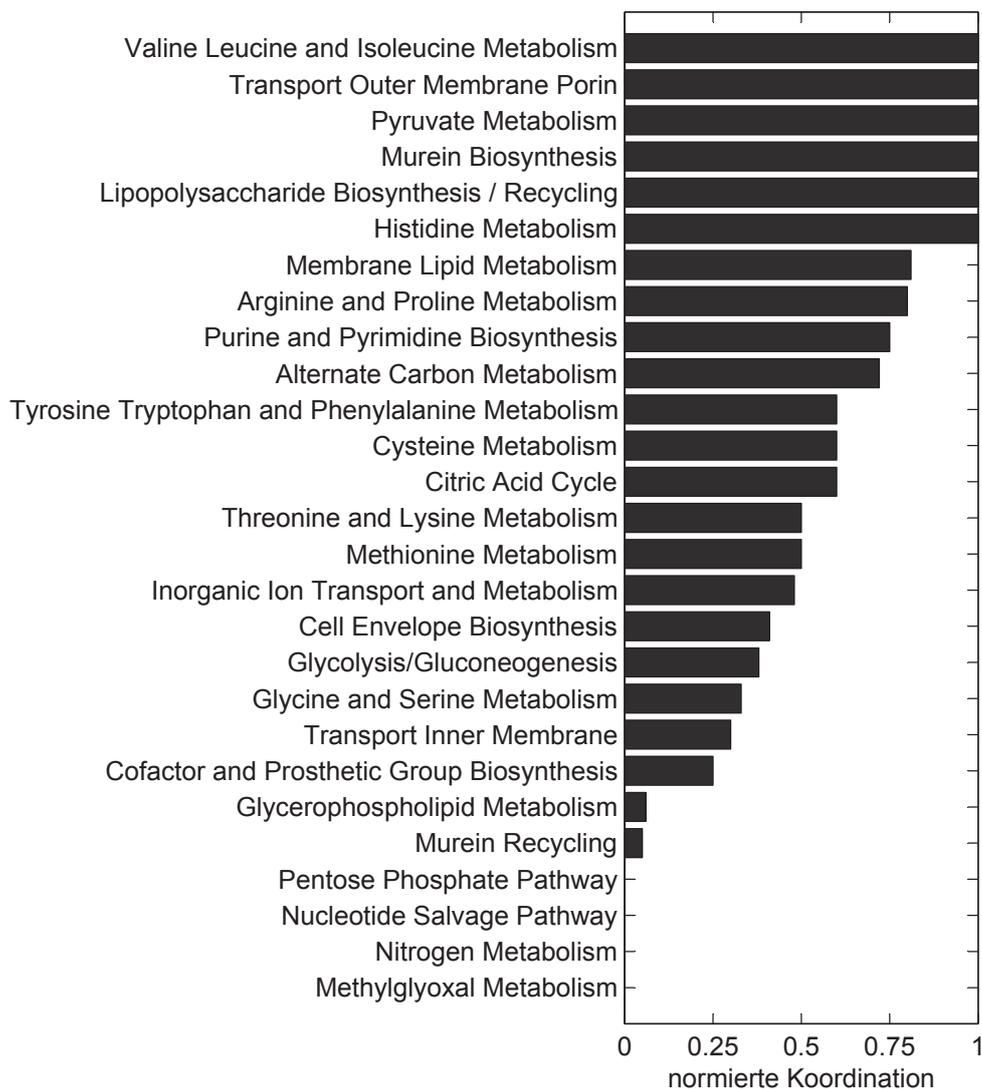


Abbildung 5.2: Übersicht der Koordination der Teilsysteme in *Escherichia coli*

## 5.2 Minimale Regulation

Die in Kapitel 5.1 gezeigten Zusammenhänge veranschaulichen, dass sich die Regulationsstrategien in *Escherichia coli* in Abhängigkeit der untersuchten Teilsysteme stark unterscheiden. Es ist auffällig, dass eine geringe Koexpression der berechneten „Elementary flux patterns“ auftritt, welche einer geringen transkriptionellen Regulation der betreffenden Netzwerke entspricht und somit nicht alle Enzyme des Netzwerkes transkriptionell reguliert werden. Da diese geringe Regulation (bzw. Koordination) dem klassischen Bild der Stoffwechselregulation widerspricht, gilt es diese, in der vorliegenden Arbeit, näher zu untersuchen.

Die angestrebten Untersuchungen können prinzipiell durch eine empirische Wahl oder zufällige Erzeugung von Simulationsszenarien erfolgen. Dieses Vorgehen ergibt allerdings nur schwer einen Rückschluss auf den Grund der ermittelten Aussagen, erschwert somit weitere Analysen und reduziert entsprechend die Aussagekraft der Ergebnisse. Hierbei eignet sich die Formulierung eines dynamischen Optimierungsproblems, da die Regulation anhand des formulierten Problems ermittelt wird. Die berechneten Ergebnisse stellen somit eine zielgerichtete Beeinflussung (Steuerung) des dynamischen Systems dar, welche der Lösung des Optimierungsproblems entspricht. Die Eigenschaften der optimalen Steuerung eines Systems wurden bereits in Kapitel 3 ausführlich erläutert. Durch die Lösung einer optimalen Steuerung, geben die Eigenschaften des Optimierungsproblems Anhaltspunkte für weitere Untersuchungen (wie z.B. eine zielgerichtete Datenanalyse). Dementsprechend wird die Aussagekraft der Ergebnisse gefestigt und es können zusätzliche Information gewonnen werden, um das Verhalten des untersuchten Systems besser zu verstehen.

### 5.2.1 Das untersuchte Stoffwechselnetzwerk

Die hier angewendeten Methoden sind modell-basiert und benötigen somit eine vollständige Beschreibung des dynamischen Systems. Dies gilt für eine Simulation und auch für die Anwendung der dynamischer Optimierung. Hierbei stellt sich die begrenzte Verfügbarkeit von kinetischen Parametern der Stoffwechselreaktionen als problematisch heraus. Demzufolge ist es nicht möglich eine vollständige, dynamische Beschreibung der Modelle für die zu untersuchenden Teilsysteme in *Escherichia coli* (siehe z.B. Abbildung 5.2) zu ermitteln. Um die angestrebten Analysen trotzdem durchführen zu können, wird hier ein theoretisches Stoffwechselnetzwerk (siehe Kapitel 2.1) ohne Verzweigungen untersucht. Auch wenn dies einer Vereinfachung entspricht, kann hier jedoch eine prinzipielle Analyse der Regulation durchgeführt werden.

S: Substrat Y<sub>1</sub>, Y<sub>2</sub>, Y<sub>3</sub>, Y<sub>4</sub>: Intermediate P: Netzwerkprodukt

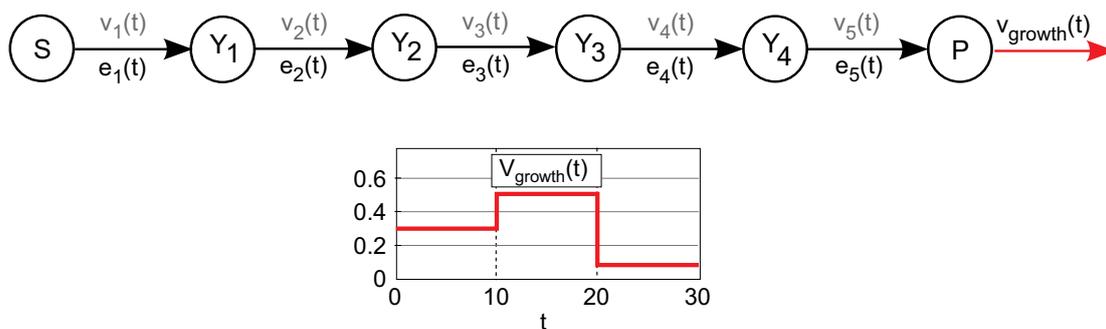


Abbildung 5.3: Das zu regulierende Stoffwechselnetzwerk

Für das betrachtete Stoffwechselnetzwerk wird die Umsetzung eines Substrat in ein Netzwerkprodukt durch fünf enzymatisch katalysierte Schritte untersucht. Dabei steht die Synthese eines Netzwerkproduktes unter veränderlichem Verbrauch des Produktes im Vordergrund. Abbildung 5.3 dient zur Veranschaulichung des zu regulierenden Stoffwechselnetzwerkes. Die zeitveränderlichen Ausflussraten stellen ein Szenario für Variationen der Umgebungsbedingungen dar, hierbei wurde

$$v_{growth}(t) = \begin{cases} 0,3 & \text{für } 0 \leq t < 10 \\ 0,5 & \text{für } 10 \leq t < 20 \\ 0,1 & \text{für } 20 \leq t \leq 30 \end{cases} \quad (5.1)$$

für die Ausflussrate beispielhaft angenommen. Der Zeithorizont für die Untersuchungen wird hier einheitslos betrachtet und auf  $t_f = 30$  (unskalierte Zeiteinheit) gewählt. Diese einheitslose Betrachtung gilt ebenfalls für Konzentrationen, Flüsse und Modellparameter. Für das zu untersuchende Stoffwechselnetzwerk kann durch entsprechende Regulation der Enzyme

$$\mathbf{e}(t) = [e_1(t), \dots, e_5(t)]^T$$

der Fluss durch die einzelnen Stoffwechselreaktion kontrolliert werden. Dies resultiert in einer Beschreibungsform für zeitlich veränderliche Enzyme des stöchiometrischen Modells. Dabei wird Gleichung (2.3) entsprechend modifiziert durch

$$\dot{\mathbf{z}}(t) = \mathbf{N} \cdot \mathbf{v} ; \mathbf{z}(t_0) = \mathbf{z}_0$$

mit

$$v_i(z_i(t), e_i(t)) = \frac{e_i(t) \cdot k_{cat,i} \cdot z_i(t)}{K_{M,i} + z_i(t)}, \quad ; i = 1, \dots, 5$$

und

$$k_{cat,i} = K_{M,i} = 1 ; i = 1, \dots, 5$$

und

$$\mathbf{z}(t) = [s(t), y_1(t), y_2(t), \dots, y_4(t)]^T$$

und

$$\mathbf{z}(t_0) = \mathbf{z}_0 = [1, 1, 1, 1, 1]^T$$

und

$$\dot{p}(t) = v_5(t) - v_{growth}(t) ; p(t_0) = p_0 = 1$$

und

$$\mathbf{N} = \begin{bmatrix} 0 & 0 & 0 & \cdots & 0 \\ 1 & -1 & 0 & \cdots & 0 \\ 0 & 1 & -1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & 0 & 1 & -1 \end{bmatrix}.$$

Das Substrat wird dabei als gepuffert angenommen. Dies entspricht einer häufig *in vivo* vorkommenden Situation, wobei die Netzwerksubstrate durch vorhergehende Netzwerke nachgeführt werden. Hier kommt hinzu, dass die Realisierung der Ausflussrate  $v_{growth}(t)$  eine ausreichende Substratmenge zusätzlich bedingt.

### 5.2.2 Zielstellungen und Anforderungen an die Regulationsstrategie

Das in Kapitel 5.2.1 dargestellte Stoffwechselnetzwerk beinhaltet die Schwierigkeit den Stofffluss derart aufrecht zu erhalten, dass eine gewünschte Rate des Produktausflusses gewährt werden kann. Dies entspricht einer Kontrolle des Stoffwechselnetzwerkes durch ein regulatorisches Programm, welches letztendlich über die Anpassung der Enzyme  $e_1(t), \dots, e_5(t)$  realisiert wird.

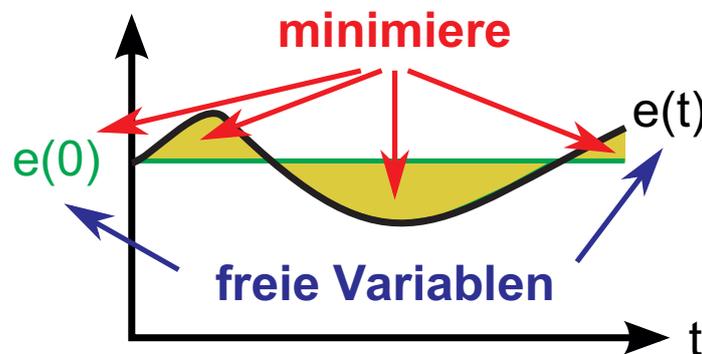


Abbildung 5.4: Die minimale Regulation

Die Zielstellung des regulatorischen Programms wird hier aus den unerwarteten Ergebnissen von Kapitel 5.1 abgeleitet. Die darin ermittelte, transkriptionelle Regulation stellt sich für die verschiedenen Teilsysteme stark unterschiedlich dar, sogar bis eine Koexpression nahezu entfällt. Dies lässt als Zielstellung ein regulatorisches Programm der Teilsysteme vermuten, welches die regulatorischen Interaktionen auf ein Minimum reduziert. Die hierbei nicht regulierten Enzyme müssen jedoch im Zellmedium vorhanden sein, um die Katalyse der entsprechenden Stoffwechselreaktion realisieren zu können. Diese Möglichkeit wird durch eine freie Bestimmung der Anfangskonzentrationen der Enzyme in der Formulierung der Zielstel-

lung gegeben. Zusammenfassend wird nun in Abbildung 5.4 die Regulation mit minimalen Interaktionen und minimaler Enzymmenge vereinfacht dargestellt.

Zusätzlich soll ein Leerlaufen des Netzwerkes für darauf folgende Zeitabschnitte bzw. Wachstumsphasen vermieden werden. Zudem gilt es eine Akkumulation der Intermediate des Stoffwechselweges auf toxische Level zu verhindern (siehe z.B. [122]). Prinzipiell können diese Anforderungen auch als Strafterme in der Zielfunktion behandelt werden. Jedoch wird hier eine Formulierung dieser Anforderungen über Nebenbedingungen gewählt.

### 5.2.3 Formulierung des Optimierungsproblems

Das Ziel dieser dynamischen Optimierungsaufgabe ist es eine transkriptionelle, regulatorische Strategie zu identifizieren, indem die Zeitverläufe  $e_1(t), \dots, e_5(t)$  und deren Anfangswerte angepasst werden, um zwei Zielfunktionsanteile zu minimieren:

- die Minimierung der Gesamtkosten der Enzyme und
- die Minimierung der Veränderlichkeit der Enzymverläufe bezogen auf deren Anfangswert.

Die Anfangskonzentrationen der Enzyme können somit entsprechend als optimale Arbeitspunkte angesehen werden. Es ergibt sich für diese Formulierungen folgende Zielfunktion

$$\min_{\mathbf{e}(t=0), \mathbf{e}(t)} \left\{ \sum_{i=1}^5 \int_{t=0}^{t=30} \left( \sigma \cdot e_i(0) + (e_i(t) - e_i(0))^2 \right) dt \right\} \quad (5.2)$$

mit

$$\sigma = \frac{1}{30}.$$

Hierbei stellt der erste Term aus Gleichung (5.2) eine Gesamtkostenminimierung für die Enzymbereitstellung dar und der zweite Term beschreibt die geforderte Minimierung der Veränderlichkeit der Zeitverläufe der Enzyme bezogen auf deren Anfangswert. Der Einfluss der verschiedenen Anteile kann durch einen Wichtungsfaktor  $\sigma$  angepasst werden.

Um ein Leerlaufen des Stoffwechselnetzwerkes zu vermeiden, wird während der Lösung des dynamischen Optimierungsproblems die Produktkonzentration  $p(t)$  in einem definierbaren Bereich gehalten. Dies geschieht hier durch Toleranzbereiche abhängig vom Anfangswert  $p_0$ , welche beschrieben werden nach

$$(1 - \epsilon) \cdot p_0 \leq p(t) \leq (1 + \epsilon) \cdot p_0 \quad (5.3)$$

mit der Wahl des Toleranzbereiches auf  $\pm 20\%$  ergibt sich z.B.

$$\epsilon = 0, 2.$$

Weiterhin wird die Anforderung einer beschränkten Akkumulation der Intermediate beschrieben durch

$$\sum_{j=2}^5 z_j(t) \leq \Omega \quad (5.4)$$

mit

$$\Omega = \sum_{j=2}^5 z_j(t = 0).$$

Die angestrebte numerische Lösung bedingt zudem noch

$$\mathbf{e}(t) \geq \mathbf{0}$$

für physikalisch richtige und biologisch bedeutsame Ergebnisse. Für die Einhaltung der diskretisierten Gleichungsnebenbedingungen aus Kapitel 5.2.1 wird an dieser Stelle auf die Methoden aus Kapitel 4 verwiesen.

Als problematisch hat sich hierbei die Ermittlung von kontinuierlichen Enzymverläufen, wie sie *in vivo* auftreten, ergeben. Eine stückweise konstante Approximation erweist sich als ungünstig, da diese in einer sehr hohen Anzahl von benötigten finiten Elementen resultiert. Eine mögliche Strategie für die bessere Approximation von charakteristisch *in vivo* auftretenden Enzymverläufen kann z.B. durch eine stückweise lineare oder polynomiale Approximation der Enzymverläufe realisiert werden. Weiterhin gilt es die Kontinuität der Enzymverläufe zu garantieren, welche durch die in der Steuerungsparametrierung auftretende Diskontinuität zwischen den finiten Elementen gefährdet ist.

Durch eine Ausnutzung der softwaretechnischen Realisierung der quasi-sequentiellen Methode aus Kapitel 4.5 kann die lineare Approximation der kontinuierlichen Enzymverläufe durch Modifikation des Optimierungsproblems mit den freien Variablen  $d_i$  nach

$$\frac{de_i(t)}{dt} = d_i(t)$$

geschickt realisiert werden (wie schon erwähnt sind hier die Anfangsbedingungen ebenfalls frei). Durch den Integrationsansatz der Differentialgleichungen von der quasi-sequentiellen Methode, garantiert dieses Vorgehen, zudem auf einfachem Wege, die Einhaltung der Kontinuität zwischen den Intervallen.

## 5.3 Optimale Regulationsstrategien

### 5.3.1 Die minimale transkriptionelle Regulation

Wird das in Kapitel 5.2 dargestellte Problem einer optimalen Steuerung des Stoffwechselnetzwerkes mit der in Kapitel 4.2 erweiterten Methode gelöst, ergeben sich die in Abbildung 5.5 dargestellten optimalen Enzymverläufe und optimalen metabolische Konzentrationsverläufe. Wie anhand der Dynamik in den Enzymverläufen aus Abbildung 5.5 zu erkennen ist, lässt sich hierbei ein prinzipielles regulatorisches Programm ableiten:

- Die Konzentrationsverläufe des ersten und letzten Enzyms unterliegen einer transkriptionellen Regulation, wobei die Konzentrationsverläufe der anderen (inneren) Enzyme nur einer geringen bzw. keiner transkriptionellen Regulation unterliegen.

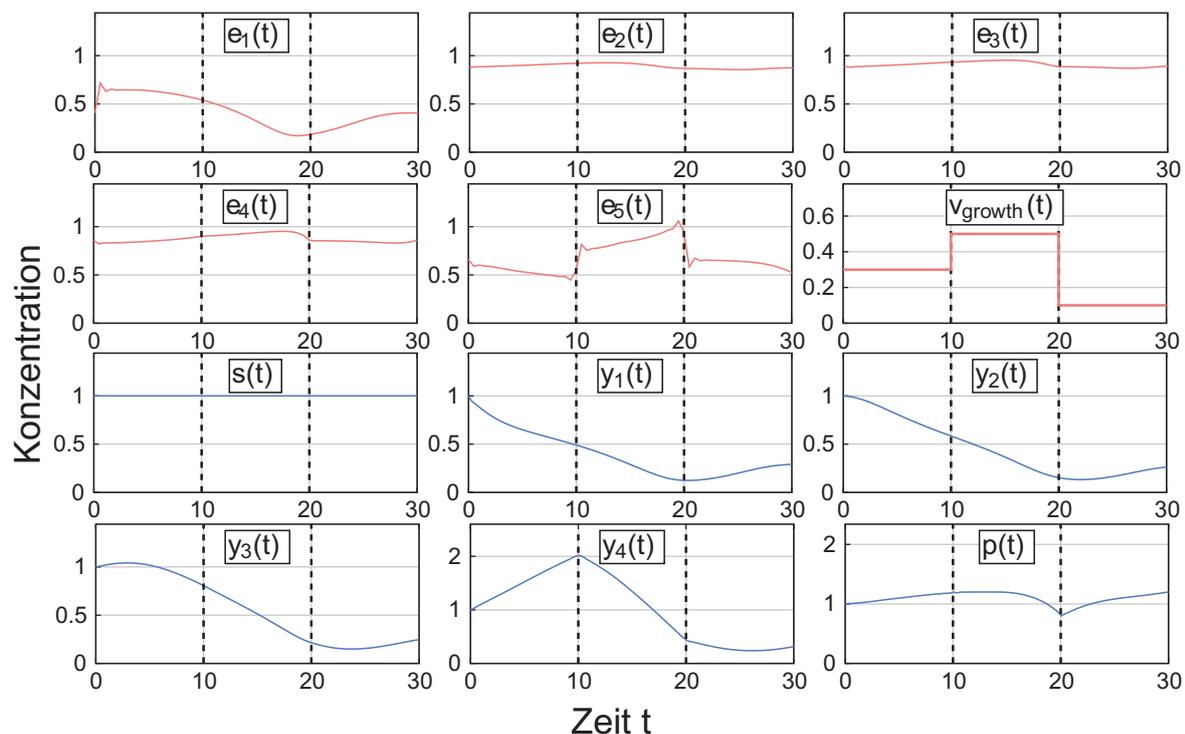


Abbildung 5.5: Optimales regulatorisches Programm zur minimalen transkriptionellen Regulation

Diese Art der Netzwerkregulation entspricht einer geringen Koordination des Netzwerkes, welches ähnlich *in vivo* beobachtet wurde (vergleiche Kapitel 5.1) und wird in der vorliegenden Arbeit als dünn verteilte, transkriptionelle Regulation bezeichnet. Dies stellt die Kernaussage dieses Kapitels dar und wird im Folgendem auf ihre Aussagekraft in den jeweiligen Unterabschnitten geprüft.

### Untersuchung der Qualität der Lösung

Da die Verwendung eines numerischen Lösungsverfahrens zu lokalen Optima führen kann, wurde für alle Analysen der Suchraum zufällig initialisiert und die Problemstellung 100mal gelöst. Für alle gezeigten Regulationsstrategien wurde/wird jeweils die beste Lösung (hier entsprechend Gleichung (5.2) die kleinste) ausgewählt und die dazugehörigen Ergebnisse dargestellt.

Um nun die Optimalität der hier getroffenen Kernaussage und dementsprechend den Vorteil dieser dünn verteilten, transkriptionellen Regulation sicherzustellen, empfiehlt es sich die eventuell verschiedenen Zielfunktionswerte der einzelnen Lösungen und deren dynamische Eigenschaften näher anzuschauen. Dazu lassen sich die Dynamiken der einzelnen Enzymprofile durch eine separate Berechnung des Zielfunktionsanteils

$$\int_{t=0}^{t=30} (e_i(t) - e_i(0))^2 dt$$

leicht ermitteln. Kleine Werte stellen das Maß für eine geringe Dynamik dar, wohingegen große Werte das Maß für eine hohe Dynamik darstellen (siehe hierzu Abbildung 5.6).

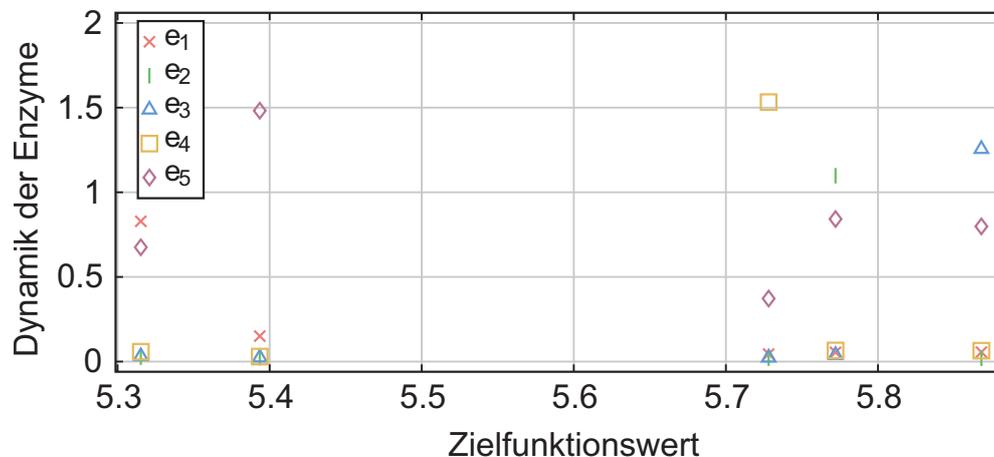


Abbildung 5.6: Variation in den Dynamiken der Enzymprofile in Abhängigkeit von verschiedenen Lösungen (global bzw. lokale)

Eine Analyse von Abbildung 5.6 ergibt fünf Lösungen des Optimierungsproblems mit verschiedenen Optima und verschiedensten Dynamiken der einzelnen Enzymprofile. Hierbei ist leicht zu erkennen, dass sich für kleiner werdende Zielfunktionswerte ein regulatorisches Programm ergibt, welches mit dynamischen Anpassungen des ersten und letzten Enzyms die Kontrolle des Netzwerkes realisiert. Im Rückschluss stellt somit die oben erwähnte dünn verteilte, transkriptionelle Regulation eine klare Verbesserung der Zielfunktionswerte und einen hohen Vorteil während der Netzwerkkontrolle dar.

### Einfluss der Umgebungsbedingungen

Variationen in den Umgebungsbedingungen können verschiedene Wachstumsraten hervorrufen und somit zu verändertem  $v_{growth}$  führen. Dies wird durch

$$v_{growth}(t) = \begin{cases} 0,1 & \text{für } 0 \leq t < 10 \\ 0,5 & \text{für } 10 \leq t < 20 \\ 0,3 & \text{für } 20 \leq t \leq t_f = 30 \end{cases}$$

exemplarisch untersucht. Die hierfür ermittelte optimale Regulationsstrategie ist in Abbildung 5.7 dargestellt. Wie anhand der Enzymverläufe zu erkennen ist, überwiegt die dynamische Anpassung des ersten und letzten Enzyms wiederum sehr stark. Somit kann auch unter anderen Umgebungsbedingungen eine dünn verteilte, transkriptionelle Regulation von Vorteil sein.

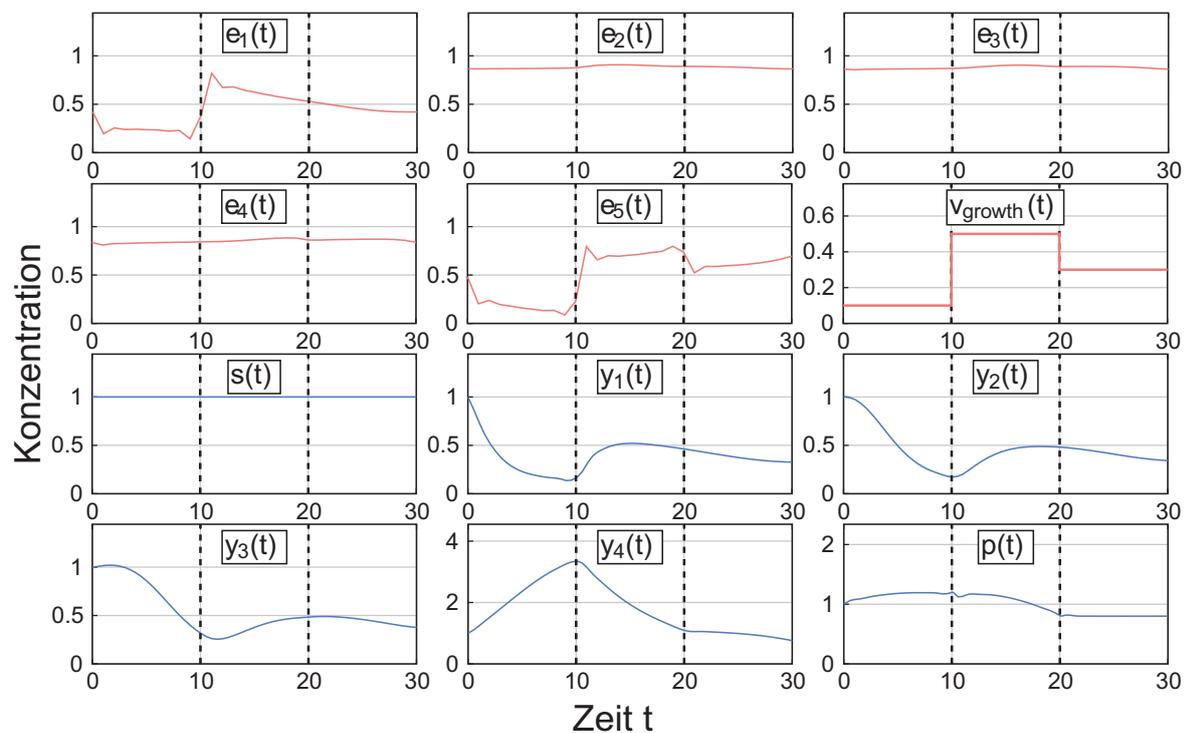


Abbildung 5.7: Optimales, regulatorisches Programm zur minimalen Flusskontrolle für veränderte Ausflussraten

Für konstante Umgebungsbedingungen besteht ein weiteres wichtiges Szenarium in der Anpassung von Produktkonzentrationen einzelner Stoffwechselnetzwerke. Hierfür muss jedoch das Standardproblem aus Kapitel 5.2, wie im Folgendem erklärt, modifiziert werden. Um den Einfluss dieser Anforderung untersuchen zu können, werden die Umgebungsbedingungen als konstant angenommen und es wird entsprechend  $v_{growth}$  aus Gleichung (5.1) für alle

Zeitpunkte zu

$$v_{growth} = 0,3$$

gewählt. Weiterhin wird das Ziel der Produktkonzentrationsanpassung durch eine Modifikation von Gleichung (5.3) realisiert. Hierfür wird ein Anstieg der Produktkonzentration und eine nachfolgende Reduzierung der Konzentration untersucht nach:

$$\text{für } 0 \leq t < 10 \text{ gilt } 0,8 \leq p(t) \leq 1,2$$

$$\text{für } 10 \leq t < 15 \text{ gilt } 0,8 \leq p(t) \leq 3,6$$

$$\text{für } 15 \leq t < 25 \text{ gilt } 2,4 \leq p(t) \leq 3,6$$

$$\text{für } 25 \leq t < 30 \text{ gilt } 0,8 \leq p(t) \leq 3,6$$

$$\text{für } 30 \leq t < 40 \text{ gilt } 0,8 \leq p(t) \leq 1,2.$$

Die berechnete, optimale Regulationsstrategie ist in Abbildung 5.8 dargestellt, wobei die Anforderungen an die Produktkonzentrationen zusätzlich illustriert werden. In diesem Fall ist es wieder ausreichend das erste und letzte Enzym des Netzwerkes einer dynamischen Kontrolle zu unterziehen. Somit ist auch bei dem Ziel einer Produktkonzentrationsanpassung eine dünn verteilte, transkriptionelle Regulation von Vorteil.

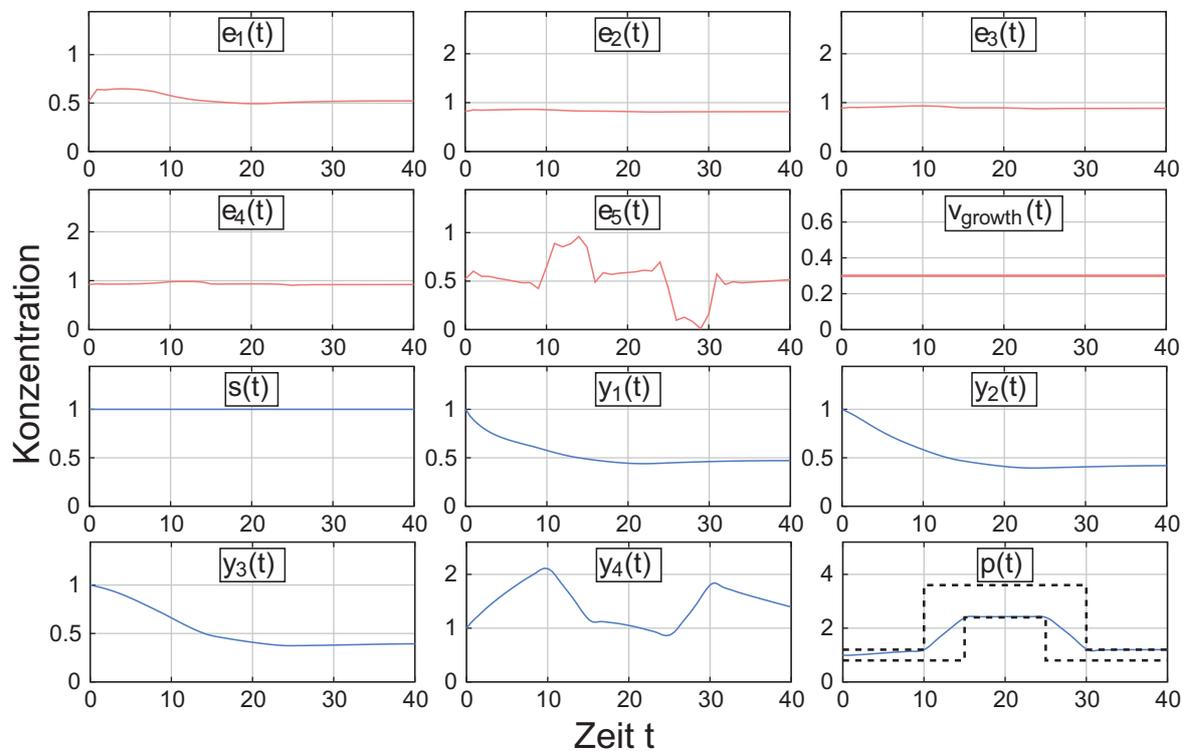


Abbildung 5.8: Optimales, regulatorisches Programm zur minimalen Flusskontrolle für Änderungen der Produktkonzentration

### Untersuchung verschiedener Proteinkosten

Da die letztendlich investierten Proteinmengen für die Katalyse von Stoffwechselprodukten *in vivo* stark variieren, ist es notwendig auch deren Einfluss auf die Regulationsstrategien zu untersuchen. Dabei werden die verschiedenen Proteinkosten über den Einfluss des Wichtungsfaktors  $\sigma$  der Zielfunktion (5.2) in der Formulierung des optimalen Steuerungsproblems behandelt. Es gilt somit für jede Bewertung (verschiedener Wichtungsfaktor) das optimale Steuerungsproblem aus Kapitel 5.2 wiederholt zu lösen. Die Auswertung der berechneten optimalen Regulationsstrategien geschieht (ähnlich wie Abbildung 5.6) über eine Analyse der Dynamik der Enzymprofile für die verschiedenen Gewichtungen in Abbildung 5.9. Hierbei entsteht ein interessantes Verhalten der Regulationsstrategien:

- Für fallende Kosten ergibt sich eine Regulation des Netzwerkanfangs und Netzwerkes, wobei die inneren Enzyme nahe ihres Anfangswertes bleiben. Somit ergibt sich für kleine Proteinkosten die bisher beobachtete dünn verteilte, transkriptionelle Regulation.
- Für steigende Kosten unterliegen nun auch die inneren Enzymprofile einer Regulation. Diese Art der Regulation aller Enzyme entspricht einer starken Koordination der Netzwerke und wird in der vorliegenden Arbeit als umfassende, transkriptionelle Regulation bezeichnet.

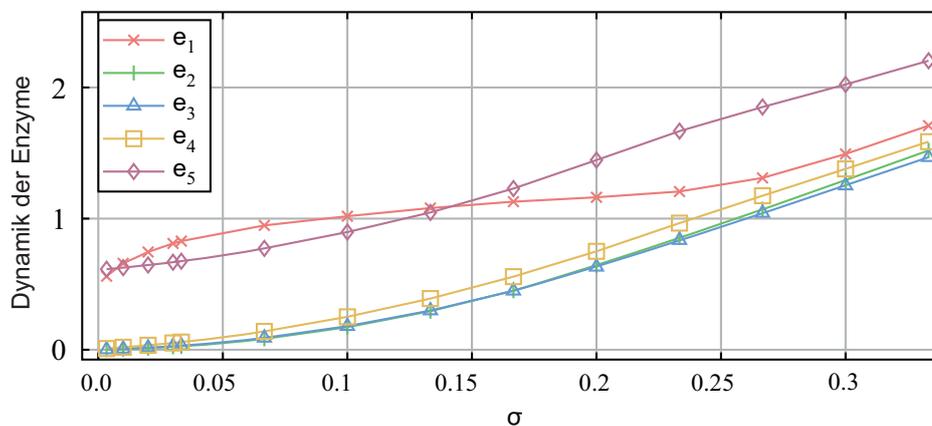


Abbildung 5.9: Variation in den Dynamiken der Enzymprofile in Abhängigkeit verschiedener Proteinkosten (Wichtungsfaktor  $\sigma$ )

Zusammenfassend zeigt dies, dass mit steigenden Proteinkosten eine Verhaltensänderung in der Regulationsstrategie von Vorteil ist. Diese Veränderung im Regulationsverhalten ist gekennzeichnet durch einen Übergang der transkriptionellen Regulation von Anfangs- und Endenzymen (dünn verteilte, transkriptionelle Regulation) zu der transkriptionellen Regulation aller Enzyme (umfassende, transkriptionelle Regulation).

### Einfluss von zufälligen kinetischen Parametern

In diesem Abschnitt werden die Abhängigkeiten der dünn verteilten, transkriptionellen Regulation von den kinetischen Parametern des Stoffwechselnetzwerkes untersucht. Diese Sensitivität zu den Modellparametern wird realisiert durch eine zufällige Erzeugung von 100 Parametersets (gleichverteilt auf dem Intervall  $[0,2]$ ), für welche jeweils die optimale Steuerung aus Kapitel 5.2 gelöst wird. Hierbei wurde aus Gründen der Rechenzeit die Lösung der einzelnen Optimierungsprobleme auf eine 50malige Konvergenz mit zufälligen Startwerten reduziert. Die Lösungen mit dem kleinsten Zielfunktionswert sind in Abbildung 5.10 dargestellt. Es ist zu erkennen, dass die dünn verteilte, transkriptionelle Regulation für initiale Enzyme und am Netzwerkende befindlichen Enzyme nicht für alle Parametersets gilt.

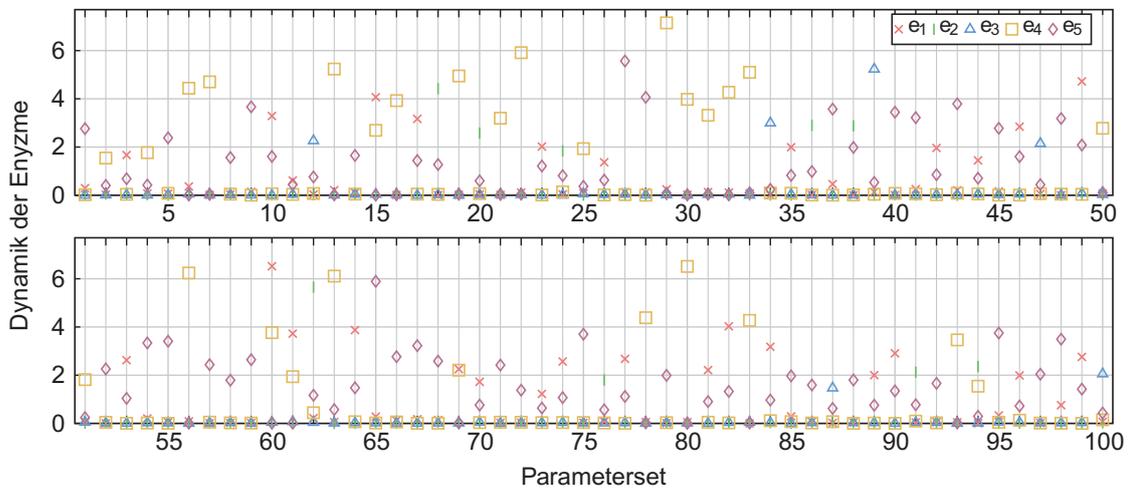


Abbildung 5.10: Variation in den Enzymprofilen in Abhängigkeit zufälliger Parametersets

Werden nun zusätzlich diese Regulationsstrategien positionsabhängig analysiert, ergibt sich die in Abbildung 5.11 gezeigte Häufigkeitsverteilung. Dabei wurde ein Enzym als reguliert identifiziert, wenn die Werte aus Abbildung 5.10 über den Grenzwert 0,1 liegt. Es kann nun geschlussfolgert werden, dass die oben definierte, dünn verteilte, transkriptionelle Regulation auch für Variationen in den kinetischen Parametern dominierend erhalten bleibt.

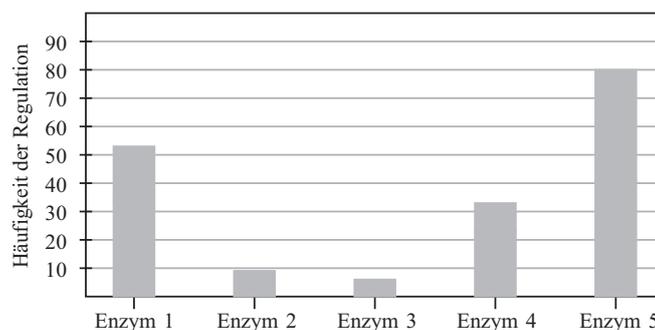


Abbildung 5.11: Positionsabhängige Häufigkeit der Regulation für zufällige Parametersets

### 5.3.2 Rolle der Enzyme am Netzwerkanfang und -ende

Hier gilt es die transkriptionelle Regulation des letzten und ersten Enzyms aus Kapitel 5.3.1 auf ihre Ursache zu untersuchen. Dafür werden verschiedene Optimalsteuerungsprobleme gelöst um die Rolle der einzelnen Enzyme während der Flusskontrolle aufzuzeigen.

Zu Beginn wird untersucht, welches Enzym hauptsächlich die Akkumulation der Stoffwechselzwischenprodukte kontrolliert verhindert. Hierfür wird nun wieder das Optimalsteuerungsproblem aus Kapitel 5.2 betrachtet. Dabei wird jedoch die Modifikation vorgenommen, dass die Nebenbedingung für die Akkumulation der Intermediate aus Gleichung (5.4) entfernt wird (bzw.  $\Omega = \infty$ ). Dies ergibt die optimale Regulationsstrategie aus Abbildung 5.12. Dabei wird überwiegend das letzte Enzym reguliert, wobei die anderen Enzyme nur einer sehr geringen Anpassungen unterliegen. Auffällig sind hierbei die höheren Konzentrationen der Intermediate  $y_{1,\dots,3}$  über den untersuchten Zeitraum und dem starken Anstieg von  $y_4$  im Vergleich zur Lösung des Originalproblems (siehe Abbildung 5.5). Eine derartige Regulationsstrategie kann somit zu einer toxischen Anhäufung von Intermediaten führen und wird *in vivo* mit hoher Wahrscheinlichkeit vermieden.

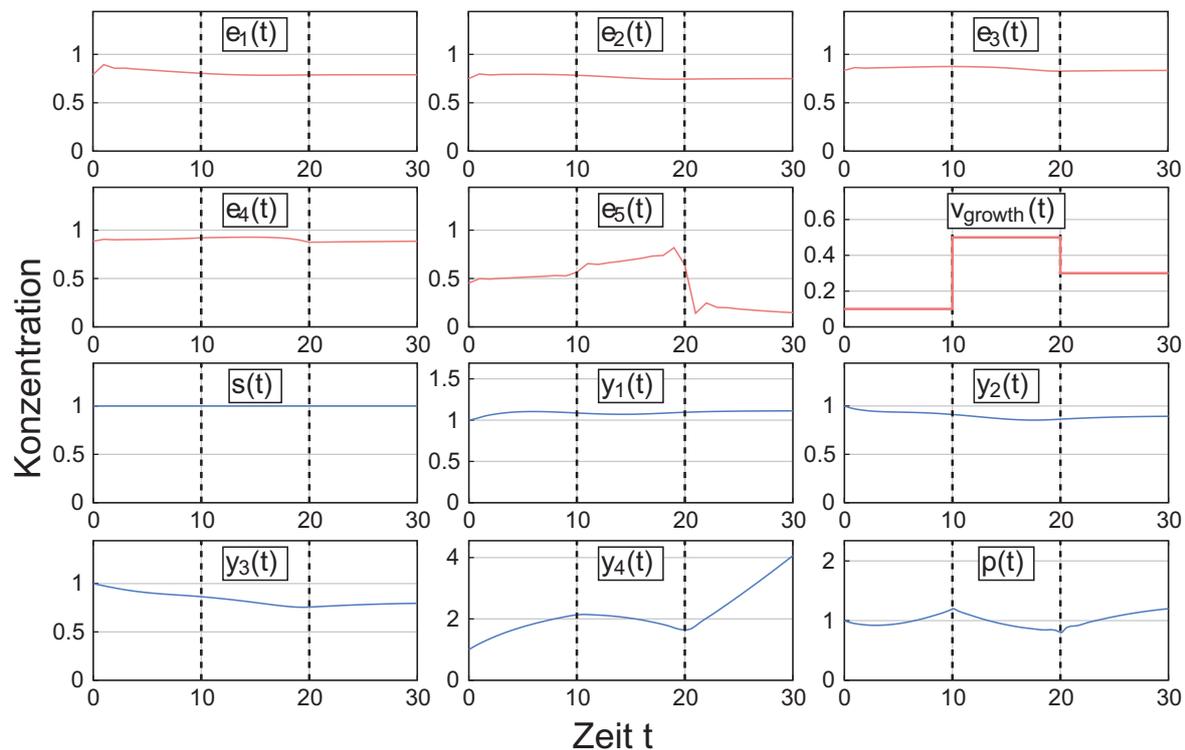


Abbildung 5.12: Optimales, regulatorisches Programm zur minimalen Flusskontrolle ohne Kontrolle der Intermediatanhäufung

Das *in vivo* realisierte Verhalten der einzelnen Enzyme könnte innerhalb des Übergangs von beiden Regulationsstrategien aus Abbildung 5.5 zu 5.12 liegen. Hierfür wird entsprechend ein

Parameter  $\alpha$  in Gleichung (5.4) nach

$$\sum_{i=2}^5 z_i(t) \leq \alpha \cdot \Omega$$

eingeführt. Hierbei werden die Analysen für längere Abschnitte mit konstanten Wachstum durchgeführt, da dies verstärkt zu einer Anhäufung von Intermediaten über einen längeren Zeitraum führen kann. Es wird  $t_f = 90$  beispielhaft gewählt und Gleichung (5.1) entsprechend angepasst:

$$v_{growth}(t) = \begin{cases} 0,3 & \text{für } 0 \leq t < 30 \\ 0,5 & \text{für } 30 \leq t < 60 \\ 0,1 & \text{für } 60 \leq t \leq 90 \end{cases}$$

Zur Analyse der Ergebnisse sind die dynamischen Eigenschaften der einzelnen Enzyme für verschiedene Werte des Wichtungsfaktors  $\alpha$  in Abbildung 5.13 dargestellt. Auch in dieser Vielzahl von optimalen, regulatorischen Programmen, gilt für das letzte Enzym eine Dominanz der Dynamik im Vergleich zu den anderen Enzymen. Die Komplexität der Regulation des ersten Enzyms nimmt mit steigendem  $\alpha$  ab und ist konform zu den obigen Untersuchungen ohne Intermediatskontrolle aus Gleichung (5.1).

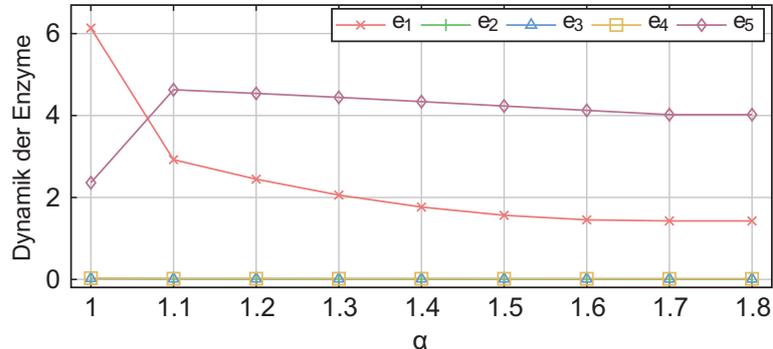


Abbildung 5.13: Variation in den Dynamiken der Enzymprofile in Abhängigkeit verschiedener Intermediatanhäufung (Wichtungsfaktor  $\alpha$ )

In einem nächsten Schritt werden die Ergebnisse aus Abbildung 5.5, 5.7 und 5.12 nochmals herangezogen, jedoch nun bezüglich der Realisierung des Produktflusses durch die geforderte Ausflussraten  $v_{growth}$  analysiert. Dabei ergibt sich die Vermutung, dass der Fluss in das Produkt durch das letzte Enzym überwiegend kontrolliert wird, da dessen Verlauf der geforderten Ausflussrate prinzipiell ähnlich ist. Jedoch können auch ähnliche Aussagen für das erste Enzym aus Abbildung 5.5 und deutlicher aus Abbildung 5.7 getroffen werden. Um die Rolle dieser Enzyme nun besser untersuchen zu können und so eine eindeutigere Aussage abzuleiten, wird das Optimierungsproblem aus Kapitel 5.2 wiederum modifiziert.

Die Modifikationen stellen sich durch eine Fixierung der Konzentration von betreffenden Enzymprofilen (hier auf 1,0) und einer Veränderung der Aufgabenstellung in eine optimale Umsteuerung dar, um so den Einfluss des ersten und letzten Enzymes auf die Produktflussanpassung näher zu untersuchen. Hierfür werden alle Anfangskonzentrationen (Enzyme, Intermediate, Produkt) zu 1,0 angenommen und es ergibt sich  $v_{growth} = 0,5$  für den stationären Betrieb des Netzwerkes. Die optimale Umsteuerung wird realisiert durch eine Minimierung der quadratischen Abweichung des Produktkonzentrationsverlaufes zur initialen Produktkonzentration durch die veränderte Zielfunktion

$$\min_{e(t)} \int_{t=0}^{t=30} (p(t) - p(0))^2 dt.$$

Dabei entfällt entsprechend die Nebenbedingung des Produktverlaufes aus Gleichung (5.3). Die optimale Umsteuerung wird nun auf einen anderen Netzwerkzustand von  $v_{growth} = 0,5$  zu  $v_{growth} = 0,3$  berechnet.

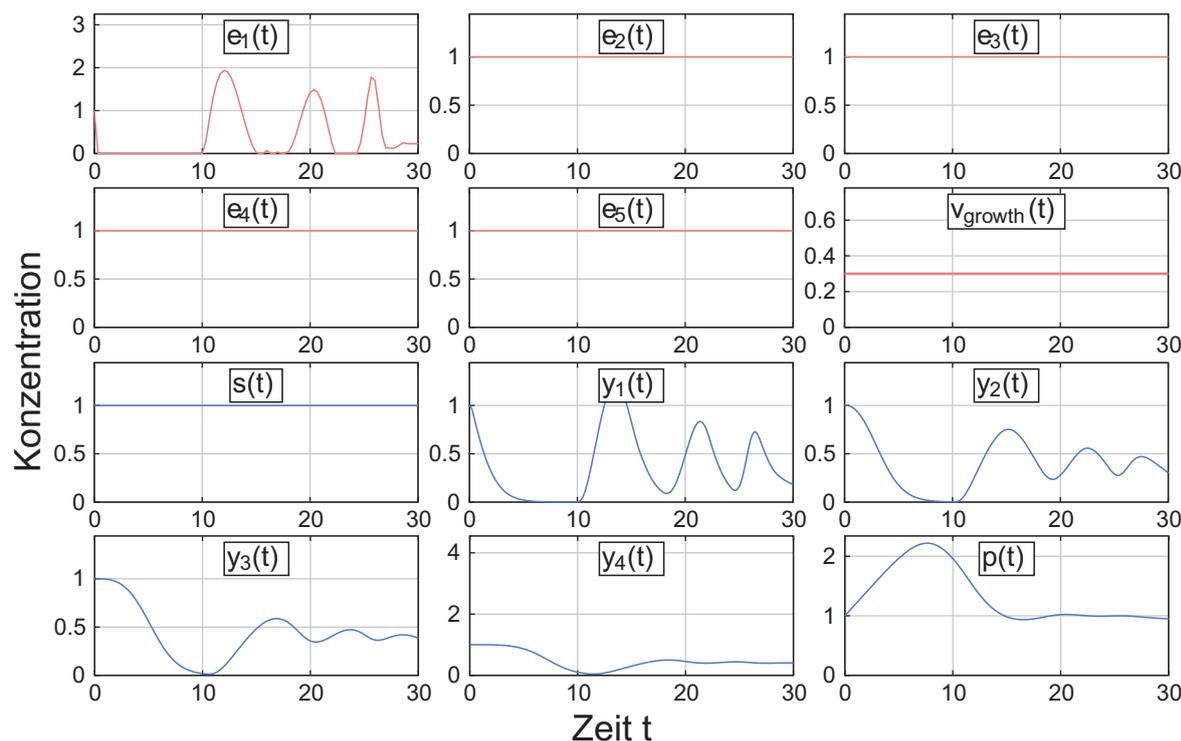


Abbildung 5.14: Optimales regulatorisches Programm zum Umsteuern des Netzwerkbetriebspunktes durch das erste Enzym

In einer ersten Studie ist es nur erlaubt die Umsteuerung des Netzwerkes durch das erste Enzym zu realisieren. Somit wurden entsprechend die anderen Enzymverläufe auf  $e_2(t) = e_3(t) = e_4(t) = e_5(t) = 1,0$  fixiert. Das dazugehörige Ergebnis ist in Abbildung 5.14 dargestellt. Die optimale Umsteuerung unter ausschließlicher Kontrollmöglichkeit des ersten En-

zyms resultiert, nach der Veränderung der Ausflussrate, in einem starken Überschwingen der Produktkonzentration, bis diese sich bei  $t \approx 20$  dem neuen stationären Betriebszustand hinsichtlich des Netzwerkproduktes annähert. In einer zweiten Studie wird nun erlaubt, die Umsteuerung des Netzwerkes über das erste und letzte Enzym umzusetzen. Damit ergibt sich für die restlichen Enzymverläufe eine Fixierung auf  $e_2(t) = e_3(t) = e_4(t) = 1, 0$ . Hierbei wurde die optimale Umsteuerung aus Abbildung 5.15 ermittelt, in welcher sich der neue stationäre Betriebszustand hinsichtlich des Netzwerkproduktes bereits nach  $t \approx 3$  einstellt. Eine derartige Netzwerkkontrolle mit den geringen Dynamiken des ersten und letzten Enzymes kann einfach mittels einer gleichzeitigen Reduktion in der Expression beider Enzyme *in vivo* umgesetzt werden.

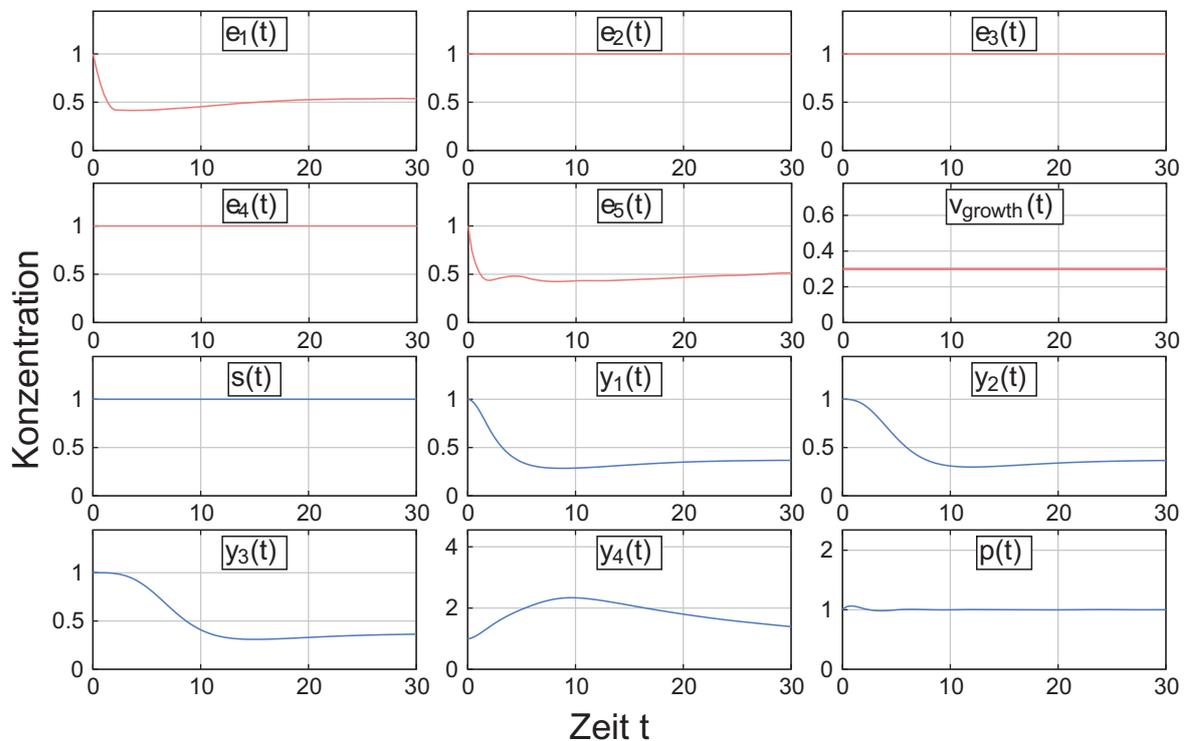


Abbildung 5.15: Optimales, regulatorisches Programm zum Umsteuern des Netzwerkbetriebspunktes durch das erste und letzte Enzym

**Zusammenfassend** gilt:

- Es lässt sich die Rolle des ersten Enzyms durch die Aktivierung des Netzwerkes, der Kontrolle des Stoffflusses in das Netzwerk und der Gesamtkonzentrationen der Intermediate beschreiben.
- Im Gegensatz hierzu, spielt die Regulation des letzten Enzymes bei der direkten Kontrolle des Produktflusses und somit auch bei der gezielten Beeinflussung der Produktkonzentration eine entscheidende Rolle.

### 5.3.3 Rolle der inneren Enzyme

In diesem Abschnitt wird die Rolle der inneren Enzyme näher untersucht. Dabei lässt sich aus Abbildung 5.5 vermuten, dass die inneren Enzyme in ausreichender Konzentration (hier entsprechend minimale Konzentration) zur Verfügung stehen sollten, um den Fluss zum Netzwerkprodukt aufrecht erhalten zu können. Diese Vermutung ist zusätzlich konform mit allen weiteren bisherigen Ergebnissen (siehe dementsprechend auch Abbildung 5.6 bis 5.15).

Um diese Aussage nun näher zu untersuchen, werden wieder verschiedene Szenarien der optimalen Regulation des Netzwerkes betrachtet, indem die Anfangskonzentration der inneren Enzyme vorgegeben wird (sie sind somit keine freien Variablen mehr). Hierbei wird das originale Optimalsteuerungsproblem aus Kapitel 5.2 entsprechend betrachtet bzw. modifiziert. Beobachtungen aus *in vivo* Experimenten [83] zeigten hierfür eine Änderung der ent-

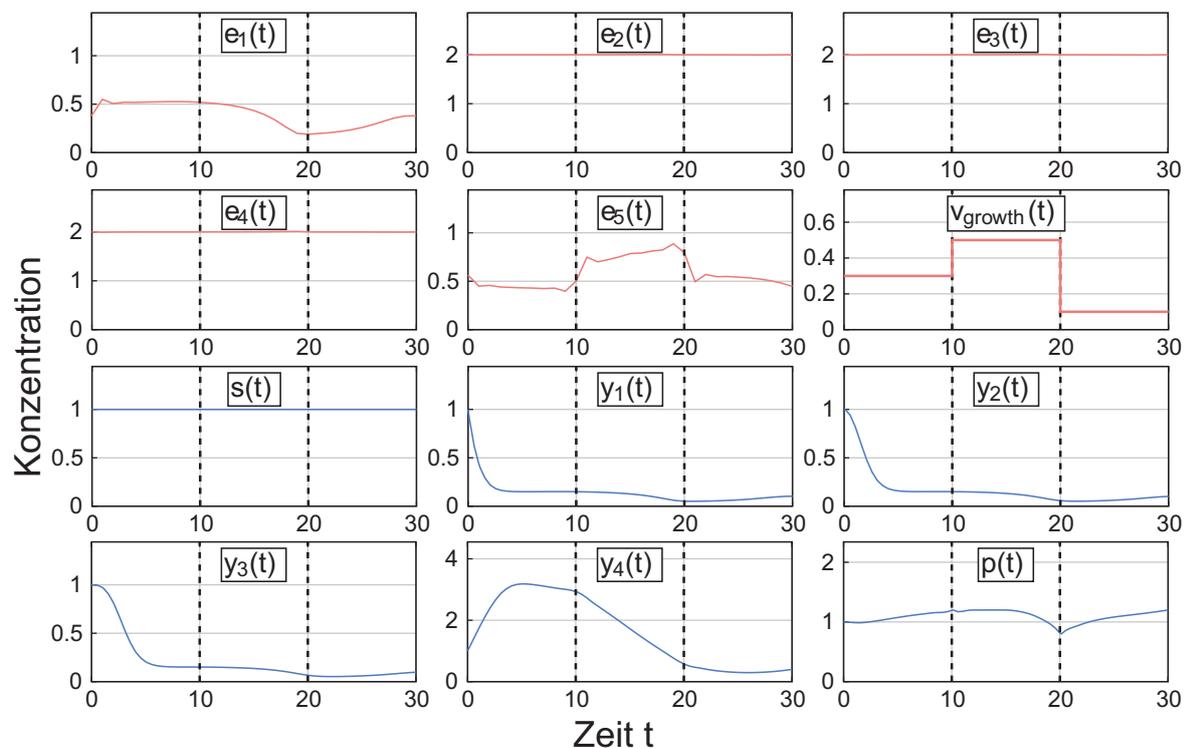


Abbildung 5.16: Optimales, regulatorisches Programm zur minimalen Flusskontrolle mit Anfangswerten der inneren Enzyme auf 2, 0

sprechenden Enzymkonzentrationen des ungefähr vierfachen Wertes für verschiedene Wachstumsraten. In einem solchen Fall wäre es zudem vorteilhaft, wenn sich für die optimalen Regulationsstrategien keine prinzipiellen Änderungen für verschiedene Wachstumsraten ergeben würden. Diese Regulationen könnten somit als robuste Strategie auch bei verschiedenen Wachstumsraten *in vivo* Anwendung finden ohne diese in Abhängigkeit der Wachstumsrate verschiedenartig zu implementieren. In Abbildung 5.16 ist die optimale Regulationsstrate-

gie für zweifache  $e_2(0) = e_3(0) = e_4(0) = 2,0$  und in Abbildung 5.17 für das vierfache  $e_2(0) = e_3(0) = e_4(0) = 4,0$  der Anfangskonzentrationen zu finden. Wie aus Abbildung 5.16 und 5.17 zuerkennen ist, ergeben sich ausschließlich geringe Änderungen in den Dynamiken der optimalen Regulationsstrategien. Dabei werden die Anfangskonzentrationen der inneren Enzyme auf ihrem Anfangswert belassen, obwohl diese Enzymprofile prinzipiell in ihrem Zeitverlauf für die Optimierung anpassbar sind. Als **Schlussfolgerung** lässt sich hierbei ableiten, dass

- die inneren Enzyme in ausreichender Menge vorliegen müssen, um den Netzwerkfluss zu ermöglichen und hierbei keinen großen Einfluss auf die Art der optimalen Regulation haben.

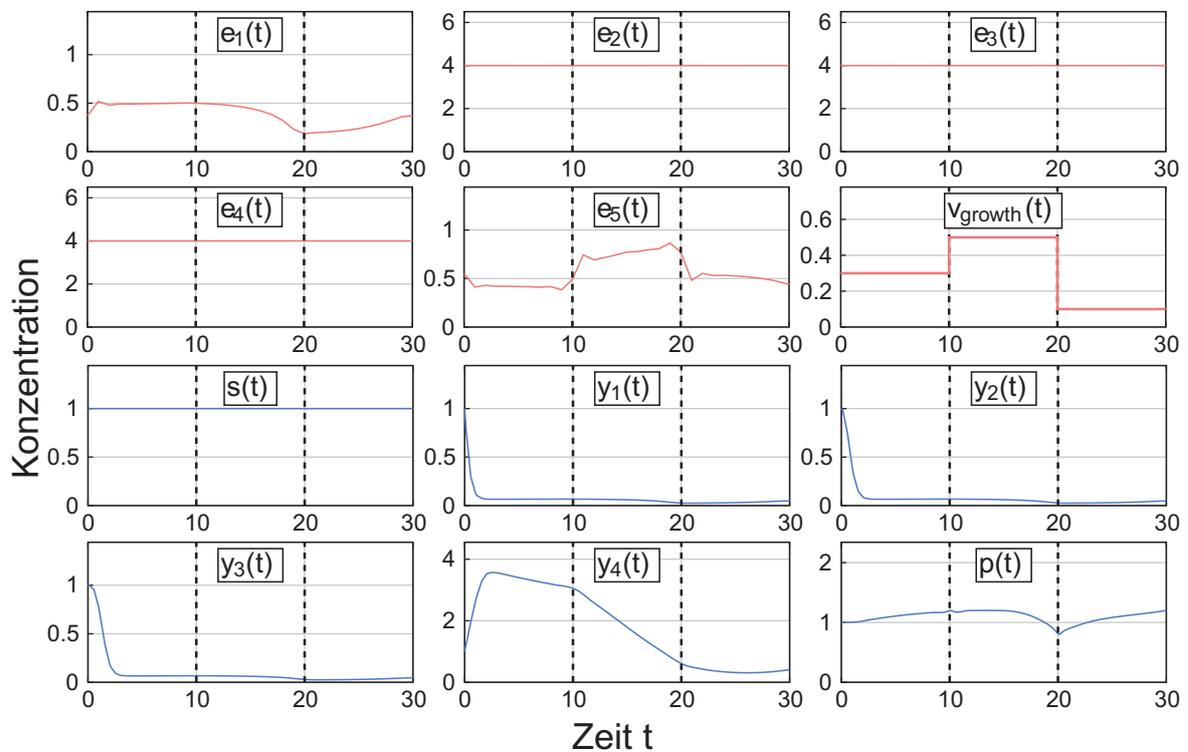


Abbildung 5.17: Optimales, regulatorisches Programm zur minimalen Flusskontrolle mit Anfangswerten der inneren Enzyme auf 4,0

## 5.4 Konsequenzen für die Regulation in *Escherichia coli*

In diesem Kapitel soll untersucht werden, welche Konsequenzen sich aus den in Kapitel 5.3 aufgezeigten optimalen Regulationsstrategien für die Regulation in *Escherichia coli* ergeben und ob bzw. wie sich diese in den *in vivo* implementierten Regulationsstrategien entsprechend zeigen. Dazu werden die prinzipiellen Eigenschaften der optimalen Regulationsstrategie zusammengefasst und deren mögliche Realisierung *in vivo* besprochen.

### 5.4.1 Ableitung von Vorhersagen

Aus Gründen der Übersichtlichkeit wird vorerst basierend auf den Aussagen aus Kapitel 5.3 ein Überblick für die verschiedenen Aufgaben der einzelnen Enzyme gegeben. Die darin angesprochene dünn verteilte, transkriptionelle Regulation basiert auf einer ausschließlichen transkriptionellen Regulation des ersten und letzten Enzyms, wobei die inneren Enzymprofile relativ konstante Verläufe aufweisen (siehe Kapitel 5.3.1). Dabei dient das erste Enzym dazu, das Netzwerk zu aktivieren, den Stofffluss in das Netzwerk zu regulieren und so auch die Anhäufungen von Intermediat zu vermeiden. Das letzte Enzym ermöglicht die direkte Regulation des Stoffflusses aus dem Netzwerk in das Produkt (siehe Kapitel 5.3.2). Dies kann besonders für lange Stoffwechselwege in *Escherichia coli* von Vorteil sein, da eine derartige Kontrolle besser auf variierende Produkthanforderung reagieren kann. Hierbei ist hervorzuheben, dass die inneren Enzyme in ausreichender Menge vorhanden sein müssen, um den Fluss durch das Stoffwechselnetzwerk zu garantieren (siehe Kapitel 5.3.3). Dies wird für hohe Kosten der inneren Enzyme problematisch (z.B. verbrauchte Proteinmenge), was jedoch durch eine höhere Komplexität in der Regulation der inneren Enzyme kompensiert werden kann. Diese Sachverhalte werden zusammenfassend als grafischer Überblick in Abbildung 5.18 veranschaulicht.

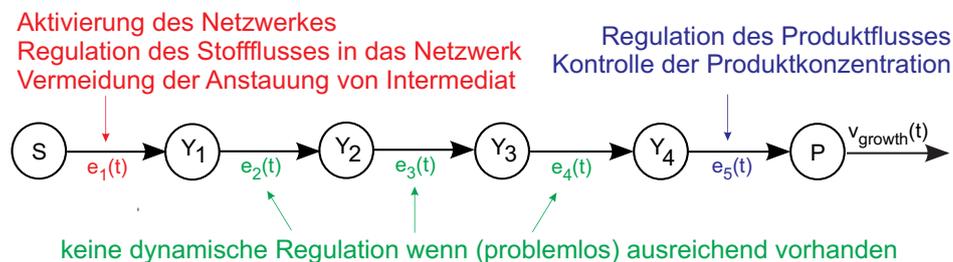


Abbildung 5.18: Vorhersage über die Aufgabe der einzelnen Enzyme für die dünn verteilte, transkriptionelle Regulation

Die bisher gezeigten zeitlichen Verläufe der Enzymkonzentrationen sind nur schwer und nicht in ausreichender Menge für alle Stoffwechselnetzwerke als Messdaten verfügbar. Um die Aussagen der bisherigen Ergebnisse jedoch zu verifizieren, müssen andere Zusammenhänge zwischen den Enzymprofilen und der zur Verfügung stehenden Daten gefunden werden. Anhand dieser können dann die ermittelten Vorhersagen aus der optimalen Regulation für die Regulation in *Escherichia coli* analysiert werden. Eine prinzipiell mögliche Implementierung für die Regulation der Enzymkonzentration kann *in vivo* durch eine transkriptionelle Regulation zur Synthese von benötigten Enzymen (siehe Kapitel 2.2) vorliegen. Bei der gleichzeitigen Regulation der Enzyme, kann die dynamische Anpassung der Enzymkonzentration

durch eine Koexpression der Enzyme oder auch einer unabhängigen Regulation der einzelnen Enzyme ermöglicht werden. Die Realisierung für konstante innere Enzyme kann *in vivo* durch eine konstitutive Expression der betreffenden Enzyme geschehen, um so eine kontinuierliche Nachführung mit ausreichender Enzymmenge zu garantieren. Dies führt **zusammenfassend** zu den folgenden Vorhersagen:

- Unterliegen die Enzyme eines Stoffwechselfades keiner Koexpression, so müssen mindestens die Anfangsreaktion und die Endreaktionen eine Regulation aufweisen.
- Bei einer ausschließlich transkriptionellen Regulation am Netzwerkanfang und -ende, müssen die inneren Enzyme konstitutiv exprimiert werden.
- Bei steigenden Kosten der Enzymsynthese, findet ein Übergang für dünn verteilte, transkriptionelle Regulation zu einer umfassenden, transkriptionellen Regulation statt.

#### 5.4.2 Verhalten in *Escherichia coli*

Die vorhergehenden Ergebnisse ermöglichen nun eine zielgerichtete Datenanalyse für *Escherichia coli* basierend auf den abgeleiteten Vorhersagen aus Kapitel 5.4.1. Dadurch kann nun untersucht werden, ob das identifizierte optimale (minimale) Regulationsverhalten für die Kontrolle eines Stoffwechselnetzwerkes die in Kapitel 5.1 aufgezeigte geringe Kontrolle von Stoffwechselwegen erklärt. Es gilt entsprechend, die Regulationsstrategie und welchen Einfluss die Kosten der Enzymsynthese auf die Art der Regulationsstrategie haben, zu untersuchen.

Die hier aufgezeigten Zusammenhänge sind Ergebnisse einer Datenanalyse durchgeführt von den Kooperationspartnern der Friedrich Schiller Universität (Christoph Kaleta, Frank Wessely, Stefan Schuster) und dem Hans Knöll Institut Jena (Reinhard Guthke). Durch die interdisziplinäre Ausrichtung der vorliegenden Arbeit sind diese Ergebnisse jedoch zum Verständnis zwingend notwendig und werden somit in Kurzform dargestellt.

#### Analyse der Regulation in Teilsystemen mit geringer Koordination

Die Analysen basieren auf einer positionsabhängigen Betrachtung der zugrundeliegenden Stoffwechselfade in *Escherichia coli*. Dabei wurde die Position der katalysierenden Enzyme anhand der Aussagen aus Abbildung 5.18 in die erste Reaktion, die inneren Reaktionen und die letzte Reaktion unterteilt (Anfang, Mitte und Ende der jeweiligen Netzwerke). Bezüglich der Regulation wird hier in transkriptionelle (inklusive der translationalen) Regulation und post-translationalen Regulation unterschieden. Dabei wurden die Teilsysteme „Cofactor and Prosthetic Group Biosynthesis“, „Glycerophospholipid Metabolism“, „Murein

Recycling“, „Nucleotide Salvage Pathway“ und „Pentose Phosphate Pathway“ als gering koordiniert identifiziert (definiert durch eine Koexpression von weniger als 25% der Reaktionswege in einem Teilsystem des Metabolismus).

Die Ergebnisse für die dünn verteilte, transkriptionelle Regulation der zugrundeliegenden Stoffwechselfade sind in Abbildung 5.19 durch eine Violinengrafik dargestellt. Dabei illustrieren Punkte den Anteil der jeweiligen Regulationsart positionsabhängig für jedes Teilsystem und jede der darin enthaltenen zugrundeliegenden Stoffwechselfade. Mehrere horizontale Punkte verdeutlichen eine mehrfach anteilige Regulationsart von Stoffwechselfaden. Die dargestellten Hüllkurven entsprechen der Dichteverteilungsfunktion der Regulationsarten mit den dazugehörigen Mittelwerten als schwarze Balken. Werden die Ergebnisse der positionsabhängigen, transkriptionellen Regulation (oberer Teil aus Abbildung 5.19) analysiert, ergibt sich bevorzugt eine erhöhte Kontrolle in Form einer transkriptionellen Regulation der Anfangsreaktion ( $P = 8,7 \cdot 10^{-4}$ ) und Endreaktion ( $P = 4,7 \cdot 10^{-6}$ ) im Vergleich zu den inneren Reaktionen der Stoffwechselnetzwerke. Hierfür lässt sich auch eine statistische Signifikanz nachweisen (P-Werte des Mann-Whitney-Wilcoxon Tests in Klammern).

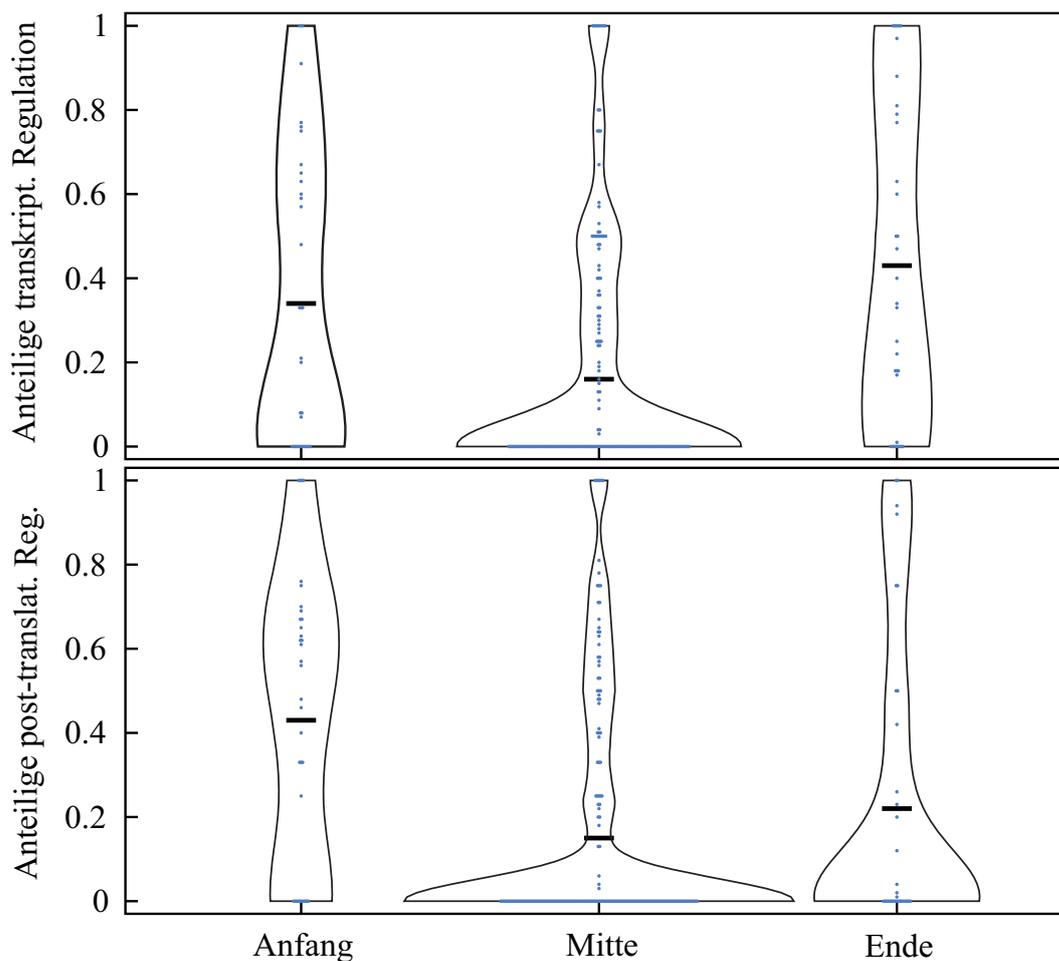


Abbildung 5.19: Positionsabhängige Kontrolle in Teilsystemen mit geringer Koordination

Der geringe Anteil einer transkriptionellen Regulation der inneren Reaktionen in Abbildung 5.19 spiegelt zudem wieder, dass die zugehörigen Enzyme konstitutiv exprimiert werden. Diese Aussage kann zusätzlich durch eine Analyse der Varianz in der Genexpression unterstützt werden, wobei die Varianzen für Teilsysteme mit geringer Koordination entsprechend klein ausfallen, im Vergleich zu den Teilsystemen mit komplexer Koordination. Siehe hierzu weitere Erläuterungen im Anhang A.2 und in Abbildung A.2.

Eine weiterhin häufig vorkommende Kontrollmöglichkeit von Stoffwechselwegen besteht in einer post-translationalen Regulation und wird somit für die Teilsysteme mit geringer Koordination ebenfalls untersucht. Die Ergebnisse hierfür sind im unteren Teil von Abbildung 5.19 zu sehen. Hier ergibt sich ein ähnliches Bild, wie bei der transkriptionellen Regulation, mit einer bevorzugten Kontrolle in Form einer post-translationalen Regulation der Anfangsreaktion ( $P = 7,5 \cdot 10^{-7}$ ) und Endreaktion (jedoch nur mit geringem P-Wert von  $P = 8,5 \cdot 10^{-2}$ ) im Vergleich zu den inneren Reaktionen der Stoffwechselnetzwerke.

Die stark signifikante, post-transkriptionelle Regulation der ersten Reaktion unterstützt ebenfalls die allgemein bekannte negative Rückkoppelregulation von Netzwerkprodukten auf die Substrataufnahme der entsprechenden Netzwerke (z.B. „negative feedback“). Dies entspricht zudem der in Kapitel 5.3.2 dargestellten Aufgabe des ersten Enzyms, welches dazu dient den Fluss in das Netzwerk zu regulieren und somit die Anhäufung von Intermediat zu vermeiden. Das sich hier ergebende **Gesamtbild** für die Teilsysteme mit einer **geringen Koordination** kann wie folgt zusammengefasst werden zu:

- Es ergibt sich eine erhöhte Kontrolle der transkriptionellen Regulation der Anfangsreaktion und Endreaktion im Vergleich zu den inneren Reaktionen des Netzwerkes.
- Es ergibt sich eine erhöhte Kontrolle in Form einer post-translationalen Regulation der Anfangsreaktionen der Stoffwechselnetzwerke.

#### **Analyse der Regulation in Teilsystemen mit komplexer Koordination**

Die hier gezeigten Analysen basieren wiederum auf einer positionsabhängigen Betrachtung der zugrundeliegenden Stoffwechselfade von *Escherichia coli*, mit der Unterteilung dieser Stoffwechselfade in erste Reaktion, innere Reaktionen und letzte Reaktion (Anfang, Mitte, Ende), welche bezüglich der Regulation in transkriptionelle Regulation (inklusive der translationalen) und post-translationalen Regulation unterschieden werden. Dabei wurden die Teilsysteme als komplex koordiniert angenommen, in welchen mehr als 25% der beteiligten Enzyme reguliert werden (diese entsprechen den nicht fettgedruckten Teilsystemen aus Abbildung A.2 im Anhang).

Die Ergebnisse für umfassende, transkriptionelle Regulation der zugrundeliegenden Stoffwechselfade sind in Abbildung 5.20 wie im vorherigen Abschnitt durch eine Violinengrafik dargestellt. Hierbei ergibt sich eine positionsabhängige, transkriptionelle Regulation (oberer Teil aus Abbildung 5.20), **ohne** bevorzugte Kontrolle der Anfangsreaktion ( $P = 1,5 \cdot 10^{-1}$ ) und Endreaktion ( $P = 1,8 \cdot 10^{-2}$ ) im Vergleich zu den inneren Reaktionen der Stoffwechselnetzwerke (entsprechender Signifikanztest mittels Mann-Whitney-Wilcoxon Tests in Klammern). Dies liegt darin begründet, dass die Enzyme zu einem großen Teil unabhängig von ihrer Position im Stoffwechselfad transkriptionell reguliert werden. Somit liegt hier eine konstitutive Expression der inneren Enzyme größtenteils nicht vor (siehe auch Abbildung A.2 im Anhang).

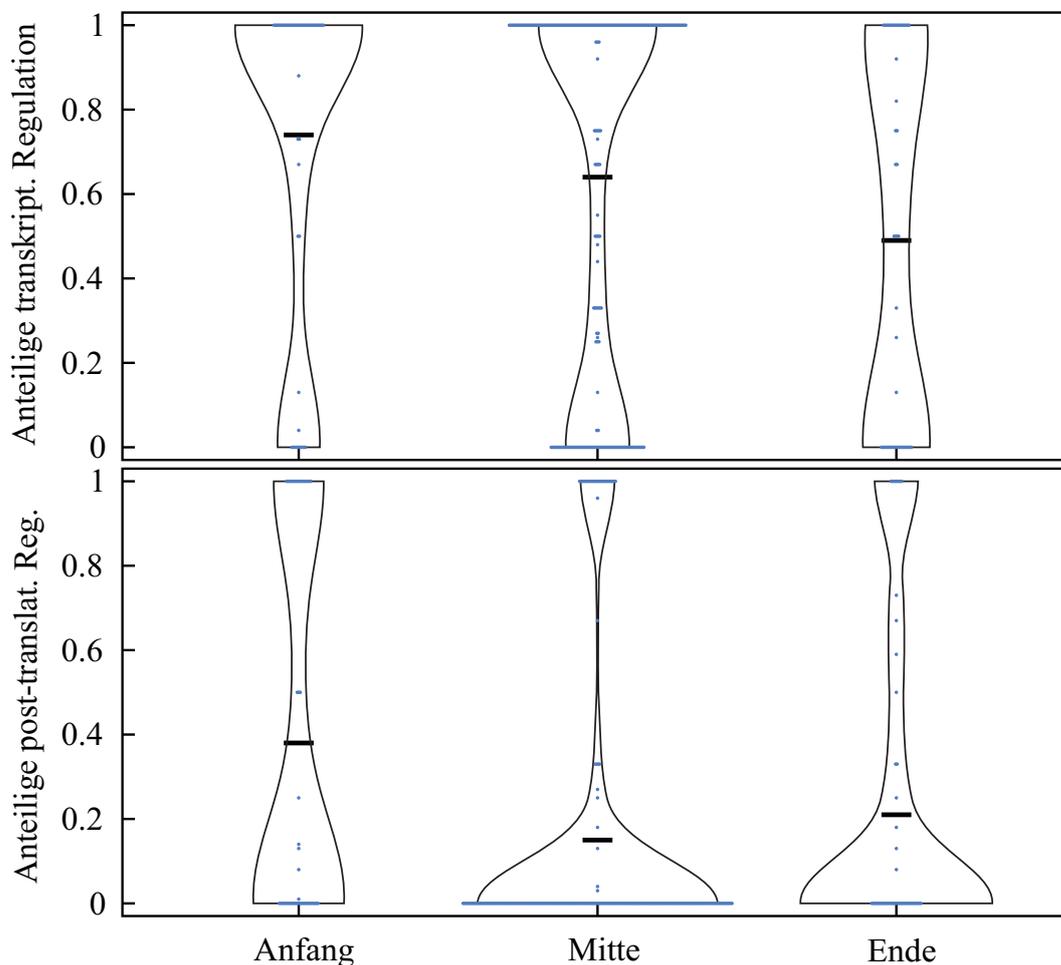


Abbildung 5.20: Positionsabhängige Kontrolle in Teilsystemen mit komplexer Koordination

Für die post-translationale Regulation (siehe unterer Teil in Abbildung 5.20) ergeben sich ähnliche Zusammenhänge, wie bei der Regulation von gering koordinierten Teilsystemen (Vergleich Abbildung 5.19 unten). Entsprechend gibt es auch hier eine bevorzugte Kontrolle der Anfangsreaktion durch eine post-translationale Regulation ( $P = 2 \cdot 10^{-6}$ ), welche ebenfalls

die allgemein bekannte negative Rückkoppelregulation von Netzwerkprodukten auf die Substrataufnahme der entsprechenden Netzwerke unterstützt. Hierbei kommt hinzu, dass diese Regulation nun auch unabhängig von den beteiligten Teilsystemen signifikant auftritt. Es ergeben sich nun die **Zusammenhänge** für die Teilsysteme mit einer **komplexen Koordination** zu:

- Es ergibt sich keine erhöhte Kontrolle der transkriptionellen Regulation der Anfangsreaktion und Endreaktion im Vergleich zu den inneren Reaktionen der Netzwerke.
- Die transkriptionelle Regulation der Enzyme tritt sehr häufig und unabhängig von der Netzwerkposition auf.
- Es ergibt sich eine erhöhte Kontrolle in Form einer post-translationalen Regulation der Anfangsreaktionen aller Stoffwechselnetzwerke.

#### **Analyse des Einflusses der Proteinmasse auf die Regulation**

In diesem Abschnitt wird untersucht, ob sich Komplexität der Regulationsstrategien in Abhängigkeit der Enzymkosten von einer ausschließlichen Kontrolle der Anfangs- und Endreaktionen (dünn verteilte, transkriptionelle Regulation) zu einer Kontrolle aller Reaktionen (umfassende, transkriptionelle Regulation) mit steigenden Kosten vergrößert. Die hier gezeigten Analysen basieren auf den erweiterten Daten für Proteinmassen aus [95], welche für *in vivo* typisch in Dalton (atomare Masseneinheit) ermittelt werden. Diese Proteinmassen stehen hier stellvertretend für die Kosten der Enzyme. Dabei wurden absolute Proteinabundanzdaten verwendet und die Proteinmassen bestimmt (Anzahl der Proteine pro Zelle mal Masse des Einzelproteins).

Um den Einfluss der Proteinkosten auf die Koordinationskomplexität der Teilsysteme mit der dünn verteilten, transkriptionellen Regulation und Teilsysteme mit der umfassenden, transkriptionellen Regulation zu untersuchen, werden die Proteinmassen für die zugrunde liegenden Stoffwechselwege teilsystem-abhängig analysiert und deren Mittelwerte berechnet. Die Ergebnisse für die einzelnen Teilsysteme sind in Abbildung 5.21 sortiert nach den Proteinkosten dargestellt. Hierbei befinden sich, abgesehen des „Pentose Phosphate Pathway“, die Teilsysteme mit der dünn verteilten, transkriptionellen Regulation „Cofactor and Prosthetic Group Biosynthesis“, „Glycerophospholipid Metabolism“, „Murein Recycling“ und „Nucleotide Salvage Pathway“ in der unteren Hälfte von Abbildung 5.21.

Anhand dieser Ergebnisse, lässt sich eine Umschaltung zwischen der dünn verteilten, transkriptionellen Regulation zu der umfassenden, transkriptionellen Regulation in Abhängigkeit

der Proteinmassen vermuten. Um dies zu untersuchen, werden die relativen Anteile der transkriptionellen Regulation der Teilsysteme in Abhängigkeit der mittleren Proteinmassen in Abbildung 5.22 dargestellt. Hierbei entsprechen die grünen Punkte den Teilsystemen mit der dünn verteilten, transkriptionellen Regulation. Abbildung 5.22 veranschaulicht, dass Teilsysteme mit kleinen mittleren Proteinmassen verschiedenartig, transkriptionell reguliert werden und somit in ihrer Regulation stark variieren. Jedoch unterliegen Teilsysteme mit hohen Proteinmassen tendenziell keiner geringen transkriptionellen Regulation. Diese Variation der transkriptionellen Regulation für kleine mittlere Proteinmassen, kann durch die Betrachtung von Teilsystemen hervorgerufen werden. Dies motiviert die Untersuchung der einzelnen Proteine unabhängig von einer Unterteilung in Teilsysteme. Hierfür ist in Abbildung 5.23 oben ein Histogramm der Proteinmassen (graue Balken) und unten der dazugehörige Anteil für transkriptionell regulierte Proteine (schwarze Balken) und entsprechend nicht regulierte Proteine (grüne Balken) dargestellt.

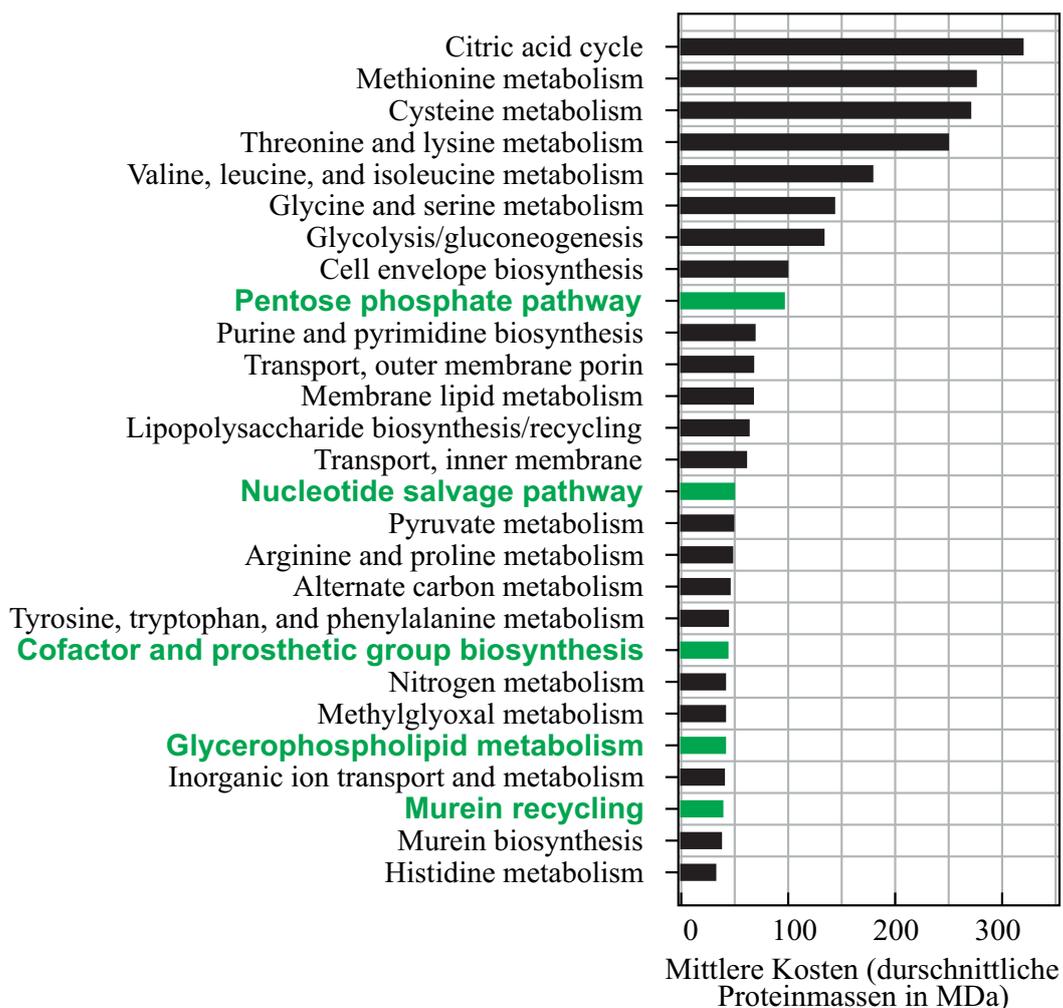


Abbildung 5.21: Teilsystem-bezogene, mittlere Proteinmassen

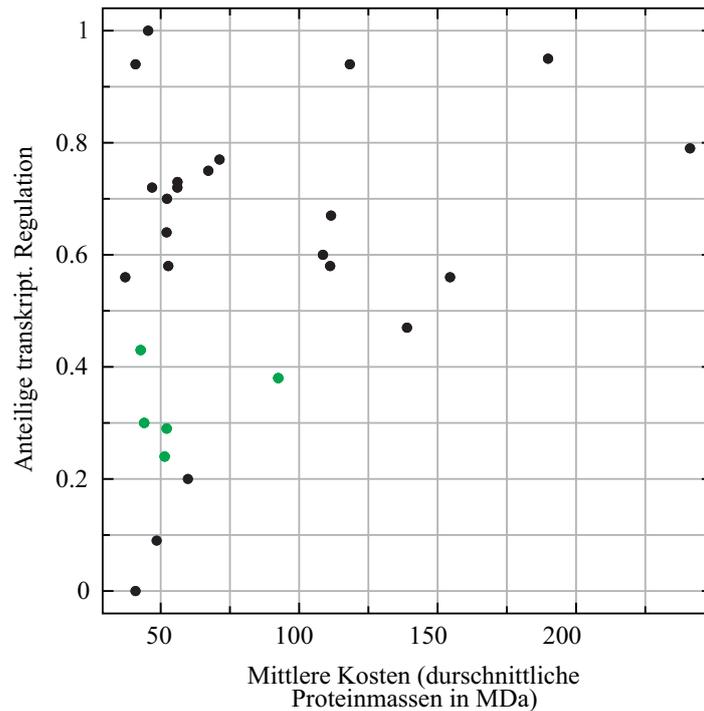


Abbildung 5.22: Regulation der Teilsysteme in Abhängigkeit der Proteinmassen

Die augenscheinlich verschiedenen Verteilungen aus Abbildung 5.23 unten (linksschief für nicht-regulierte und rechtsschief für regulierte) sind zudem statistisch signifikant unterschiedlich: Whitney-Wilconxon-Test mit  $P$ -Wert  $2,3 \cdot 10^{-5}$ . Es ergibt sich somit eine bevorzugte transkriptionelle Regulation von Enzymen mit hohen Proteinmassen (umfassende, transkriptionelle Regulation) und eine Reduzierung der transkriptionellen Regulation von Enzymen mit niedrigen Proteinmassen (dünn verteilte, transkriptionelle Regulation). Eine ähnliche Analyse für post-translationale Regulation zeigt, dass Proteinmassen keinen Einfluss auf die Bevorzugung dieser Regulation haben (nicht signifikant unterschiedlich  $P = 0,45$ ).

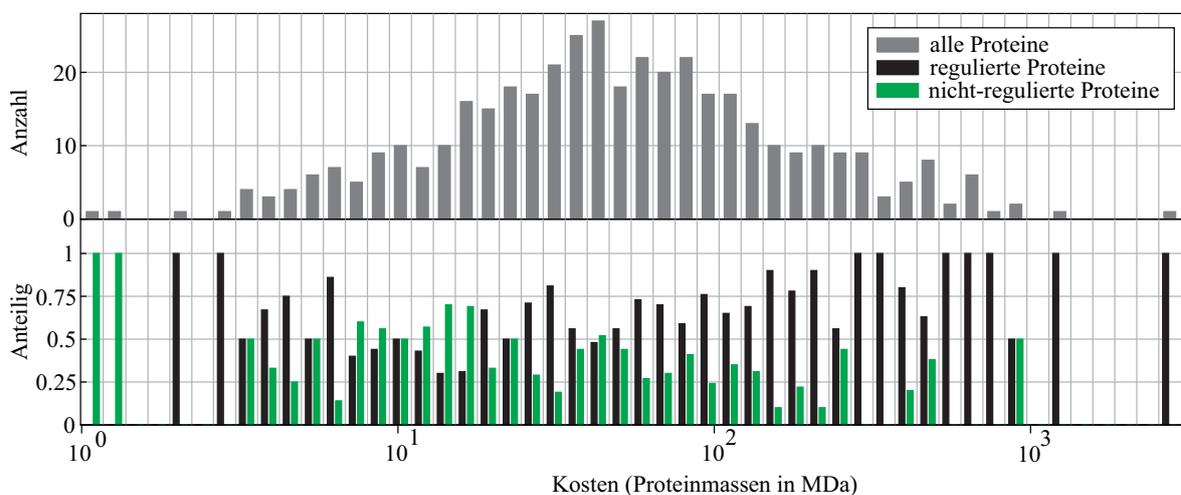


Abbildung 5.23: Regulation der Proteine in Abhängigkeit ihrer Massen

Die in diesem Abschnitt aufgezeigten Zusammenhänge in *Escherichia coli* entsprechen den Vorhersagen des Kapitels 5.4.1 aus den Optimierungsergebnissen und werden zur Veranschaulichung **zusammenfassend** dargestellt:

- Eine dünn verteilte, transkriptionelle Regulation tritt bevorzugt in Teilsystemen mit geringen Proteinkosten auf.
- Es existiert eine Umschaltung von einer dünn verteilten, transkriptionellen Regulation zu einer umfassenden, transkriptionellen Regulation mit steigenden Proteinmassen.
- Enzyme mit hohen Proteinmassen werden bevorzugt transkriptionell reguliert.
- Es gibt keinen Zusammenhang zwischen der post-translationalen Regulation von Enzymen und der jeweiligen Proteinmassen.

### 5.5 Diskussion

Die erzielten Ergebnisse zeigen eine zu den klassischen Vorstellungen der transkriptionellen Regulation aller Enzyme (umfassende, transkriptionelle Regulation genannt) verschiedenartige Regulation in einigen Teilsystemen von *Escherichia coli*. Dies wird hier als dünn verteilte, transkriptionelle Regulation bezeichnet, wobei eine Flusskontrolle des Netzwerkes über eine transkriptionelle Regulation von Anfangs- und Endreaktionen erfolgt. Die Regulation aller Enzyme oder nur der Anfangs- und Endreaktionen erlaubt jedoch eine Kontrolle des Stoffflusses in das Netzwerk und aus diesem heraus.

Die bevorzugte transkriptionelle Regulation von Enzymen mit hohen Proteinmassen weist darauf hin, dass *in vivo* Mechanismen bestehen, welche eine derartige Regulation der Enzyme hervorrufen. Dieses zugrunde liegende Prinzip, kann durch eine Balance zwischen der Kostenreduktion für die Proteinsynthese und einer schnellen Reaktion auf veränderte Umweltbedingungen erklärt werden. Somit werden Enzyme mit hohen Proteinmassen erst synthetisiert, wenn diese auch von der Zelle benötigt werden (veranschaulicht in Abbildung 5.24). Problematisch kann sich dabei die Bereitstellungszeit für die benötigten Enzyme auswirken, welche sich bei einer transkriptionellen Regulation in *Escherichia coli* im Bereich von mehreren Minuten befindet [158].

Die Reaktion auf Umweltveränderungen in Form von Stress stellt eine zeitkritische Aufgabenstellung der Zelle dar und kann durch einen „Standby“ billiger Stoffwechselnetzwerke reduziert werden (dünn verteilte, transkriptionelle Regulation). Dieser „Standby“ wird durch eine konstitutive Expression der inneren Enzyme realisiert, sodass diese unabhängig von den

Umweltbedingungen in der Zelle kontinuierlich vorliegen. Die Untersuchungen der vorliegenden Arbeit ergeben eine konstitutive Expression für Netzwerke mit tendenziell günstigen Enzymen und entsprechenden geringen Proteinmassen.

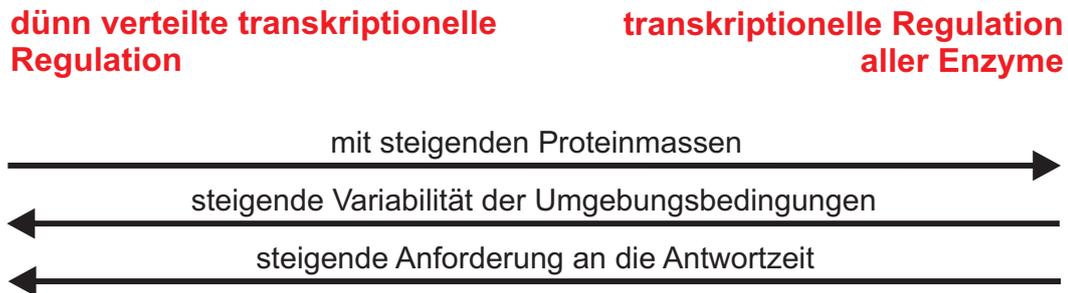


Abbildung 5.24: Regulationsstrategien und ihre Abhängigkeiten

Die Kontrolle des Netzwerkes durch eine umfassende, transkriptionelle Regulation kann eine ökonomische Ausnutzung von vorhandene Ressourcen bezogen auf den Proteinverbrauch in der Zelle realisieren. Wohingegen die Kontrolle des Netzwerkes durch die hier ermittelte dünn verteilte, transkriptionelle Regulation mit den vorhandenen Ressourcen bezogen auf den Proteinverbrauch verschwenderisch umgeht. Wird dies in Verbindung mit z.B. vorliegenden Umweltbedingungen betrachtet, so kann für spärliche Ressourcen der Umgebungsbedingungen die umfassende, transkriptionelle Regulation eine dominante Regulationsstrategie und für nährstoffreiche Umgebungsbedingungen eine dünn verteilte, transkriptionelle Regulation vermehrt auftreten. Befindet sich die Zelle jedoch öfters in veränderlichen Umweltbedingungen, z.B. regelmäßige Übergänge zwischen verschiedenen Nährstoffen, so stellt die dünn verteilte, transkriptionelle Regulation durch die schnelle Anpassungsfähigkeit einen hohen selektiven Vorteil im Gegensatz zur umfassenden, transkriptionellen Regulation dar. Dies gilt auch wenn Proteine durch die konstitutive Expression in einer ersten Einschätzung „verschwendet“ werden.

*Abschließend wird darauf hingewiesen, dass diese interdisziplinäre Arbeit in einem gemeinsamen Artikel in „Molecular Systems Biology“*

*F. Wessely, M. Bartl, R. Guthke, P. Li, S. Schuster, und C. Kaleta. „Optimal regulatory strategies for metabolic pathways in escherichia coli depending on protein costs“ (7:515,2011)*

*publiziert wurde, welcher im nachfolgendem Text gekennzeichnet wird mit [154]. Die detaillierte Vorgehensweise der Datenanalyse von Escherichia coli aus Kapitel 5.4.2 kann zudem hier nachgelesen werden.*

# 6 Signaturen optimaler Aktivierung von Stoffwechselnetzwerken in Prokaryoten

Damit Mikroorganismen in einer veränderlichen Umgebung überleben, müssen diese sich neuen Umweltbedingungen sehr schnell anpassen können und gegen konkurrierende Organismen effektiv bestehen [5, 15, 118, 121, 154]. In den vorhergehenden Kapiteln wurde bereits gezeigt, dass sich Optimalitätsprinzipien für die Erklärung von einer Vielzahl wichtiger biologischer Phänomene als geeignet herausstellten [siehe auch 35, 64, 65, 74]. Hierbei unterliegen jedoch alle Mikroorganismen einer beschränkten Kapazität in der Proteinsynthese um letztendlich die benötigten Enzyme bereitzustellen. In diesem Kapitel wird untersucht, wie sich diese Synthesekapazität von Protein auf die Anpassungsfähigkeit von Mikroorganismen auswirkt und wie diese optimal eingesetzt werden kann. Hierfür wird im Folgenden die zu untersuchende (bzw. zu lösende) Problemstellung motiviert.

## 6.1 Motivation

Im vorhergehenden Kapitel 5 wurde untersucht, ob sich die Kosten in der Proteinsynthese (dargestellt durch die entsprechenden Enzymmassen) auf die optimale Regulationsstrategie auswirken. Dabei wurde festgestellt, dass sich für geringe Proteinmassen eine dünn verteilte, transkriptionelle Regulation als optimal ergibt und die betreffenden Antwortzeiten der Stoffwechselnetzwerke, durch eine kontinuierliche Verfügbarkeit von inneren Enzymen gering ausfällt (auch als „Standby“ bezeichnet). Werden nun jedoch Stoffwechselnetzwerke mit hohen Enzymmassen (hohe Kosten in der Proteinsynthese) betrachtet, wird dieser „Standby“ *in vivo* nicht umgesetzt und die betreffenden Stoffwechselnetzwerke müssen aktiviert werden.

Die schnelle Anpassung von Stoffwechselflüssen spielt z.B. eine wichtige Rolle bei der Reduzierung von Verzögerungszeiten in der Versorgung mit Biomassebestandteilen und während starken Zustandsänderungen im Wachstumsverhalten [55, 54, 121]. Die schnelle Anpassung auf sich verändernde Umwelteinflüsse durch eine optimale Aktivierung von Stoffwechselnetzwerken wurde bereits in früheren Arbeiten [10, 80, 103, 104, 122, 157, 158] umfangreich

untersucht. Diese Arbeiten erzielten eine sequentielle Synthese der Enzyme entlang der Netzwerke als optimal, unter Annahme einer zu verteilenden Proteinmenge, sowie der Minimierung von investiertem Protein. Die sequentielle Aktivierungsstrategie ist in der Literatur auch als „just-in-time-activation“ bekannt und wurde in *Escherichia coli* für einige Aminosäurebiosynthesenetze experimentell beobachtet [158] und für Stoffwechselnetzwerke in *Saccharomyces Cerevisiae* nachgewiesen [29].

Die Kernaussage dieser Arbeiten stellt jedoch die Bevorzugung einer sequentiellen Aktivierungsstrategie vor anderen Aktivierungsstrategien dar. Dies ist insoweit problematisch, da eine operon-organisierte Regulation von Stoffwechselnetzwerken in Prokaryoten vorliegt (siehe Kapitel 2.2). Exprimierte Proteine eines gemeinsamen Operons werden nahezu gleichzeitig synthetisiert [4], weshalb sequentielle Aktivierungsstrategien für Prokaryoten *in vivo* nicht immer umgesetzt werden. Erklärungen hierfür finden sich in der Literatur bezüglich der Balance zwischen dem Vorteil einer sequentiellen Aktivierung und dem Vorteil der operon-basierten Regulation zur Reduzierung des Einflusses von stochastischen Eigenschaften der biochemischen Reaktionen [85, 112] oder zur einfachen Reduzierung von Promotersequenzen [157]. Weiterhin kann hier als problematisch angesehen werden, dass die Synthese von Proteinen eine ständige Belastung der Zelle darstellt [42, 132], welche durch vorhergehende Arbeiten nur unvollständig beachtet wurden. Dies betrifft z.B., dass Proteinkosten in den Prozessen der Proteinsynthese liegen [132] oder gar die Annahme, dass Proteine mit jeglicher notwendiger Syntheserate produziert werden können [10, 80, 103, 104, 158].

Während der Aktivierung von Stoffwechselnetzwerken stellen hohe Enzymmassen in Verbindung mit einer dynamischen Proteinsynthesekapazität (bzw. Syntheserate) der Zelle einen limitierenden Faktor dar. In diesem Kapitel wird durch Lösungen von optimalen Steuerungsproblemen untersucht, wie sich limitierende Faktoren in der Kapazität der Proteinsynthese auf die Aktivierung von Stoffwechselnetzwerken auswirken.

## 6.2 Aktivierung von Stoffwechselnetzwerken

In der vorliegenden Arbeit wird die Aktivierung von Stoffwechselnetzwerken durch die Formulierung eines dynamischen Optimierungsproblems berechnet. Die berechneten Ergebnisse stellen somit eine zielgerichtete Beeinflussung (Steuerung) des dynamischen Systems dar, welche der Lösung des Optimierungsproblems entspricht. Die Eigenschaften der berechneten Lösung des Optimierungsproblems geben wiederum Anhaltspunkte für eine zielgerichtete Datenanalyse und ermöglichen das Verhalten der *in vivo* umgesetzten Aktivierungsstrategien besser zu verstehen.

### 6.2.1 Das untersuchte Stoffwechselnetzwerk

Wie schon in Kapitel 5 erwähnt, sind die angewendeten Methoden modell-basiert und benötigen eine vollständige Beschreibung des dynamischen Systems. Auch hier stellt sich die begrenzte Verfügbarkeit von kinetischen Parametern als problematisch heraus und die angestrebten Analysen werden anhand eines theoretischen Stoffwechselnetzwerkes ohne Verzweigungen untersucht (siehe Abbildung 6.1). Dabei wird ein Stoffwechselnetzwerk für die

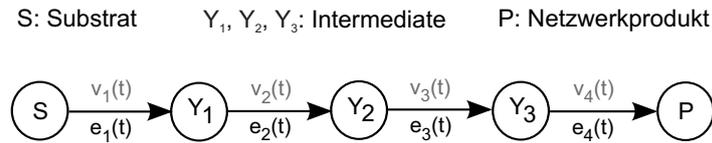


Abbildung 6.1: Das zu aktivierende Stoffwechselnetzwerk

Umsetzung eines Substrates in ein Produkt über vier enzymatisch, katalysierte Reaktionen untersucht. Hierbei findet wieder eine einheiten-lose Betrachtung für Konzentrationen, Flüsse, Modellparameter und Simulationszeit statt, welche jedoch eine prinzipielle Analyse der Aktivierungsstrategien in Abhängigkeit der Proteinsynthesekapazität ermöglicht. Für das zu untersuchende Stoffwechselnetzwerk können durch entsprechende Synthese der Enzyme

$$\mathbf{e}(t) = [e_1(t), \dots, e_4(t)]^T ; \mathbf{e}(t=0) = \mathbf{e}_0$$

die einzelnen Stoffwechselreaktionen aktiviert werden. Zudem soll hierbei der Einfluss des Wachstums durch entsprechende Verdünnungseffekte modelliert werden. Dies resultiert in einer Beschreibungsform für zeitlich veränderliche Enzyme des stöchiometrischen Modells und einer Verdünnung durch eine Wachstumsrate  $\mu(t)$  nach

$$\dot{\mathbf{z}}(t) = \mathbf{N} \cdot \mathbf{v} - \mu(t) \cdot \mathbf{z}(t) ; \mathbf{z}(t=0) = \mathbf{z}_0,$$

$$v_i(z_i(t), e_i(t)) = \frac{e_i(t) \cdot k_{cat,i} \cdot z_i(t)}{K_{M,i} + z_i(t)}, ; i = 1, \dots, 4,$$

$$k_{cat,i} = K_{M,i} = 1 ; i = 1, \dots, 4,$$

$$\mathbf{z}(t) = [s(t), y_1(t), \dots, y_3(t), p(t)]^T$$

$$\mathbf{N} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 1 & -1 & 0 & 0 & 0 \\ 0 & 1 & -1 & 0 & 0 \\ 0 & 0 & 1 & -1 & 0 \\ 0 & 0 & 0 & 1 & 0 \end{bmatrix} .$$

Für die Untersuchungen wird ein vollständig deaktiviertes Netzwerk nach  $\mathbf{z}_0 = [1, 0, 0, 0, 0, 0]^T$  und  $\mathbf{e}_0 = [0, 0, 0, 0]^T$  angenommen.

### 6.2.2 Zielstellungen und Anforderungen an die Aktivierung

Das in Kapitel 6.2.1 dargestellte Stoffwechselnetzwerk beinhaltet die Schwierigkeit die Aktivierung derart umzusetzen, dass ein gewünschtes Produkt schnellstmöglich und in ausreichender Menge zur Verfügung steht. Dies entspricht einer Kontrolle des Stoffwechselnetzwerkes durch ein regulatorisches Programm, welches letztendlich über die Anpassung der Enzyme  $e_1(t), \dots, e_4(t)$  realisiert wird. In dieser Arbeit unterteilen sich die Untersuchungen vorerst in zwei Betrachtungen:

1. Die Hauptaufgabe des Stoffwechselnetzwerkes kann darin bestehen, das Produkt so schnell wie möglich zu erzeugen und somit die Verzögerungszeit für biologischen Anforderungen gering zu halten. Um das Produkt schnellstmöglich bereitzustellen, muss das Substrat über die Intermediate zum Produkt in minimaler Zeit umgesetzt werden. Hierfür wurden bisher die Übergangszeit [93] z.B. in [80] oder die reale Produktionszeit betrachtet. Aufbauend auf eigenen vorhergehenden Arbeiten [10] soll hier entsprechend die zeitoptimale Aktivierung von Stoffwechselnetzwerken untersucht werden, welche die Produktionszeit einer bestimmten Produktmenge minimiert.
2. Es gilt hier eine Formulierung abzuleiten, welche eine auf das Zellwachstum bezogene Aktivierung darstellt. In vielen Fällen wird jedoch die Aktivierung eines Netzwerkes benötigt um ein Zellwachstum wieder aufzunehmen. Dies trifft z.B. zu wenn eine Aminosäure im Wachstumsmedium nicht mehr vorhanden ist oder entfernt wurde. Da hierbei das Produkt einen festen Bestandteil der Biomasse darstellt, kann ein Wachstum nicht wieder einsetzen bis das Produkt in ausreichender Menge vorhanden ist. Eine derartige Kontrolle des Wachstums wird z.B. in *Escherichia coli* durch eine sogenannte „stringente Antwort“ umgesetzt [46], durch welche das Zellwachstum reduziert werden, wenn keine Aminosäuren vorhanden sind.

Ein weiterer Schwerpunkt besteht darin den Einfluss der Proteinsynthesekapazität der Zelle auf die implementierten Aktivierungsstrategien zu untersuchen. Hierbei kann die Synthesekapazität der einzelnen Enzyme und die (momentan) freie Proteinsynthesekapazität der gesamten Zelle von Bedeutung sein. Die Synthesekapazität einzelner Enzyme wird durch die Kopienanzahl der entsprechenden „messenger RNA“ (mRNA) und die Effektivität in der Translation bestimmt [137]. In Abhängigkeit der benötigten Konzentration eines Enzyms zur

Aktivierung des Netzwerkes (dies entspricht der Enzymmasse) können die Vervielfältigung der mRNA und der Effektivität der Translation durch evolutionäre Entwicklungen angepasst werden, um entsprechend die Synthesekapazität der einzelnen Enzyme zu erhöhen oder zu reduzieren. Die freie Proteinsynthesekapazität einer Zelle wird durch die freien Ribosomen bestimmt, deren Zahl z.B. durch die Synthese von Enzymen anderer Netzwerke beeinflusst werden [125]. Während der Aktivierung eines Stoffwechselnetzwerkes kann es durch die Untersuchungen der Proteinsynthesekapazitäten zu starken Anhäufungen von Intermediat kommen (siehe z.B. [122]). Entsprechend gilt es zusätzlich diese Akkumulation auf toxische Level zu verhindern.

### 6.2.3 Formulierung des Optimierungsproblems

Das Ziel dieser optimalen Steuerung ist es Aktivierungsstrategien zu identifizieren, indem die Zeitverläufe  $e_1(t), \dots, e_5(t)$  so angepasst werden, dass zwei prinzipielle Zielfunktionen minimiert werden. Es ergibt sich durch die in Kapitel 6.2.2 aufgezeigte 1. Zielstellung

$$\min_{\mathbf{e}(t), t_f} t_f \tag{6.1}$$

mit

$$p(t_f) = p_f.$$

Hierbei stellt  $t_f$  die zu berechnende minimale Produktionszeit und  $p_f$  den zu erreichenden aktiven Zustand des Stoffwechselnetzwerkes dar. Der aktive Zustand definiert sich beispielhaft durch eine 90-prozentige Umsetzung des Substrates in das Produkt  $p_f = 0,9 \cdot s(t = 0)$ .

Die aufgezeigte 2. Zielstellung bedingt eine Erweiterung des Optimierungsproblems durch eine weitere freie Steuerung der Wachstumsrate  $\mu(t)$ . Dabei soll eine Zielfunktion abgeleitet werden, die für geringe Produktkonzentrationen auch nur einen geringen Vorteil formuliert. Dies wird hier durch eine Zielfunktion erreicht, welche die Biomasseerzeugung limitierend durch die Produktsynthese maximiert nach

$$\max_{\mathbf{e}(t), \mu(t)} \int_0^{t_f} \mu(t) \cdot p(t) dt. \tag{6.2}$$

Hierbei stellt  $t_f$  den betrachteten Zeithorizont dar, welcher für eine Ermittlung eines aktiven Netzwerkzustandes groß gewählt werden sollte ( $t_f = 1000$  angenommen). Zudem ermöglicht die formulierte Zielfunktion (6.2) eine Aktivierung des Netzwerkes ohne zusätzliche Angabe eines zu erreichenden aktiven Zustandes, wie dies für die zeitoptimale Aktivierung aus

Zielfunktion (6.1) notwendig ist. Die explizite Behandlung der Wachstumsrate während der Netzwerkaktivierung ermöglicht eine verbesserte Modellierung des Einflusses von Zellwachstum bei der Untersuchung der Aktivierungsstrategien.

Die Anforderung einer beschränkten Akkumulation der Intermediate und des Produktes wird beschrieben durch

$$\sum_{j=2}^5 z_j(t) \leq \Omega,$$

ohne das Substrat einzubeziehen ( $\Omega = 4$  angenommen). Durch die angestrebte numerische Lösung muss die Bedingung

$$\mathbf{e}(t) \geq \mathbf{0}$$

erfüllt sein, damit physikalisch richtige und somit biologisch bedeutsame Ergebnisse erzielt werden können.

Die Einhaltung der Gleichungsnebenbedingungen aus Kapitel 6.2.1 wird durch die Methoden aus Kapitel 4 realisiert. Wie in den Betrachtungen aus Kapitel 5 beschrieben, ergibt sich unter Anwendung der quasi-sequentiellen Methode die Ermittlung von kontinuierlichen Enzymverläufen, wie sie *in vivo* auftreten, als problematisch. Dies betrifft konkret die stückweise konstante Approximation und die Kontinuität der Enzymverläufe und kann durch eine lineare Approximation der kontinuierlichen Enzymverläufe mit den freien Variablen  $d_i(t)$  geschickt gelöst werden. Der Integrationsansatz der quasi-sequentiellen Methode, garantiert dabei auf einfachem Wege die Einhaltung der Kontinuität zwischen den Intervallen. Es ergibt sich somit unter Beachtung des Verdünnungseffektes der Wachstumsrate eine Approximation der Enzymverläufe durch die Differentialgleichungen

$$\frac{de_i(t)}{dt} = d_i(t) - \mu(t) \cdot e_i(t) ; e_i(t) \geq 0, i = 1, \dots, 4.$$

Die geforderten Analysen der Proteinsynthesekapazität (bezogen auf die Proteinsynthese einzelner Enzyme) lassen sich hierbei wiederum sehr effektiv umsetzen. Dies kann über die Syntheseraten der einzelne Enzyme geschehen, für welche ein expliziter Zugriff über  $d_i(t)$  in jedem finiten Element nach

$$|d_i(t)| \leq d_{i,\max} \quad (6.3)$$

besteht (in den Untersuchungen wird  $d_{i,\max}$  variabel sein). Entsprechend kann nun die Formulierung der freien Proteinsynthesekapazität abgeleitet werden zu

$$\sum_i |d_i(t)| \leq d_{\max}, \quad (6.4)$$

welche beispielhaft zu  $d_{\max} = 0,01$  angenommen wird. Für die Anpassung der Wachstumsrate gelten ähnliche Anforderungen wie bei der Approximation der *in vivo* vorliegenden Enzymkonzentrationen, welche wieder durch eine Differentialgleichung modelliert wird nach

$$\dot{\mu}(t) = d_{\mu}(t); \mu(0) = 0.$$

Dies ermöglicht eine dynamische Anpassung der Wachstumsrate an sich z.B. veränderte Umgebungsbedingungen. Hier wird wiederum eine maximale Rate

$$|d_{\mu}(t)| \leq d_{\mu,\max}$$

angenommen. Dabei soll ein Zeitbereich betrachtet werden, welcher kleiner ist als die Anpassungsmöglichkeiten der Enzymsynthese (hier zu  $d_{\mu,\max} = 0,001$  gewählt).

### 6.3 Optimale Aktivierungsstrategien

Die optimalen Aktivierungsstrategien werden durch die in Kapitel 6.2.3 formulierten optimalen Steuerungsprobleme unter der entsprechenden Zielfunktion mit dem in Kapitel 4.3 erweiterten Verfahren berechnet. Die Verwendung eines numerischen Lösungsverfahrens kann zu lokalen Optima führen. Um diesen Einfluss zu reduzieren, wurde wiederum für alle Analysen der Suchraum zufällig initialisiert und die Problemstellung 100mal gelöst. Die gezeigten Aktivierungsstrategien stellen die jeweilige beste Lösung des Optimierungsproblems dar.

#### 6.3.1 Zeitoptimale Aktivierung

In diesem Kapitel werden Aktivierungsstrategien untersucht, welche die Produktionszeit eines Produktes minimiert und Kapazitäten in der Proteinbiosynthese beachtet. Für die Untersuchungen der zeitoptimalen Aktivierung ist eine Behandlung der Wachstumsrate nicht vorgesehen und wird entsprechend zu  $\mu(t) = 0$  gewählt.

#### Einfluss einzelner Faktoren in der Proteinsynthesekapazität

Es soll vorerst der Einfluss einzelner Faktoren in der Proteinsynthesekapazität, d.h. der Enzymsyntheserate aus Gleichung (6.3) und der freien Proteinsynthesekapazität aus Gleichung (6.4) untersucht werden. Es ergeben sich die in Abbildung 6.2 dargestellten optimalen Enzymverläufe und optimalen metabolischen Konzentrationsverläufe. Hierbei wird links eine ausschließliche Beschränkung einzelner Enzymsyntheseraten und rechts eine ausschließlich zu

verteilende (freie) Proteinsyntheserate auf alle Enzyme betrachtet. Das optimale Verhalten kann wie folgt zusammengefasst werden:

- Es ergibt sich für die Beschränkung der einzelnen Enzymsyntheseraten eine gleichzeitige Aktivierung aller Enzyme entlang des untersuchten Stoffwechselnetzwerkes.
- Für die Beschränkung einer auf die Enzyme zu verteilenden Proteinsyntheserate werden die Enzyme sukzessiv entlang des untersuchten Stoffwechselnetzwerkes aktiviert.

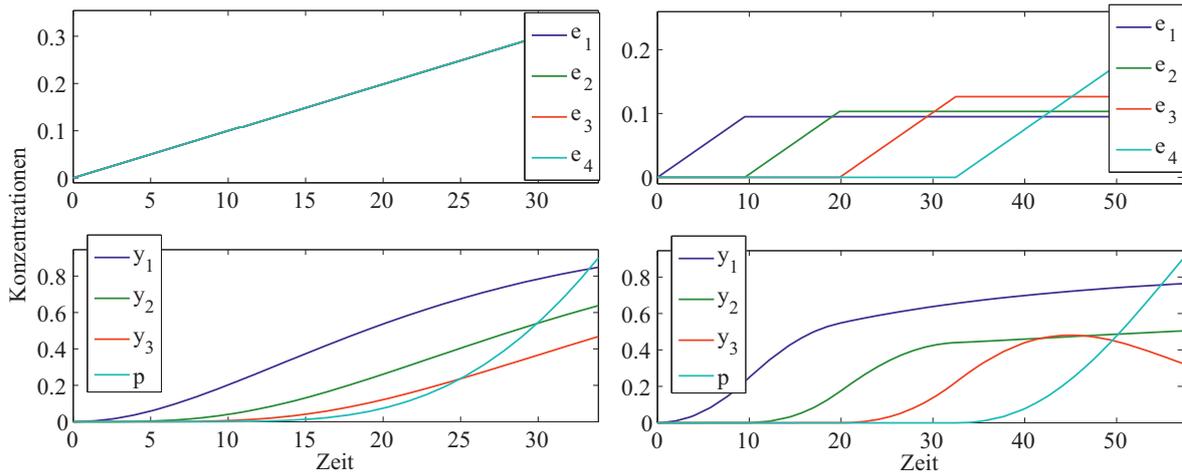


Abbildung 6.2: Zeitoptimale Aktivierung für verschiedene Proteinsynthesebeschränkungen (links  $d_{i,max} = 0,01$  und rechts  $d_{max} = 0,01$ )

Für den Fall der zu verteilenden Proteinsyntheserate wird dementsprechend diese Kapazität auf die einzelnen Enzyme sequentiell und nicht gleichzeitig verteilt. Diese Art der Aktivierungsstrategie entspricht der experimentell beobachteten „just-in-time-activation“ [158], welche ausgiebig in der Literatur untersucht wurde. Jedoch ist hier anzumerken, dass eine derartige Aktivierungsstrategie nicht immer experimentell beobachtet wurde und mit einer häufig auftretenden Regulation mittels Operons in Prokaryoten nur schlecht vereinbar ist [31, 51].

### Kombination der Faktoren in der Proteinsynthesekapazität

Unter dem Gesichtspunkt, dass eine „just-in-time-activation“ nicht immer vorliegt, werden in weiteren Untersuchungen das Zusammenspiel beider Beschränkungen analysiert. Es kann zudem davon ausgegangen werden, dass die beiden Beschränkungen *in vivo* gleichzeitig vorliegen. Die entsprechenden Ergebnisse sind in Abbildung 6.3 dargestellt. Dabei wird die freie Proteinsynthesekapazität fixiert und die Enzymsyntheserate variiert. Ist die Beschränkung der einzelnen Enzymsyntheserate der limitierende Faktor (links oben), wird die gleichzeitige Synthese aller Enzyme zur optimalen Aktivierungsstrategie. Wirkt jedoch die Beschränkung

## 6 Signaturen optimaler Aktivierung von Stoffwechselnetzwerken in Prokaryoten

der freien Proteinsynthesekapazität limitierend, ergibt sich eine sequentielle Aktivierung entlang des Netzwerkes (rechts unten). Diese Aktivierungsstrategien sind entsprechend ähnlich zu den Betrachtungen aus Abbildung 6.2 und bleiben vom prinzipiellen Verhalten auch für kleinere bzw. höhere Enzymsyntheseraten gleich (siehe Anhang A.3). Wird das Zusammenspiel beider Beschränkungen betrachtet (Abbildung 6.3 rechts oben und links unten) ergibt sich

- ein Mischverhalten aus gleichzeitiger und sequentieller Aktivierung mit der Aktivierung einzelner Netzwerkteile.

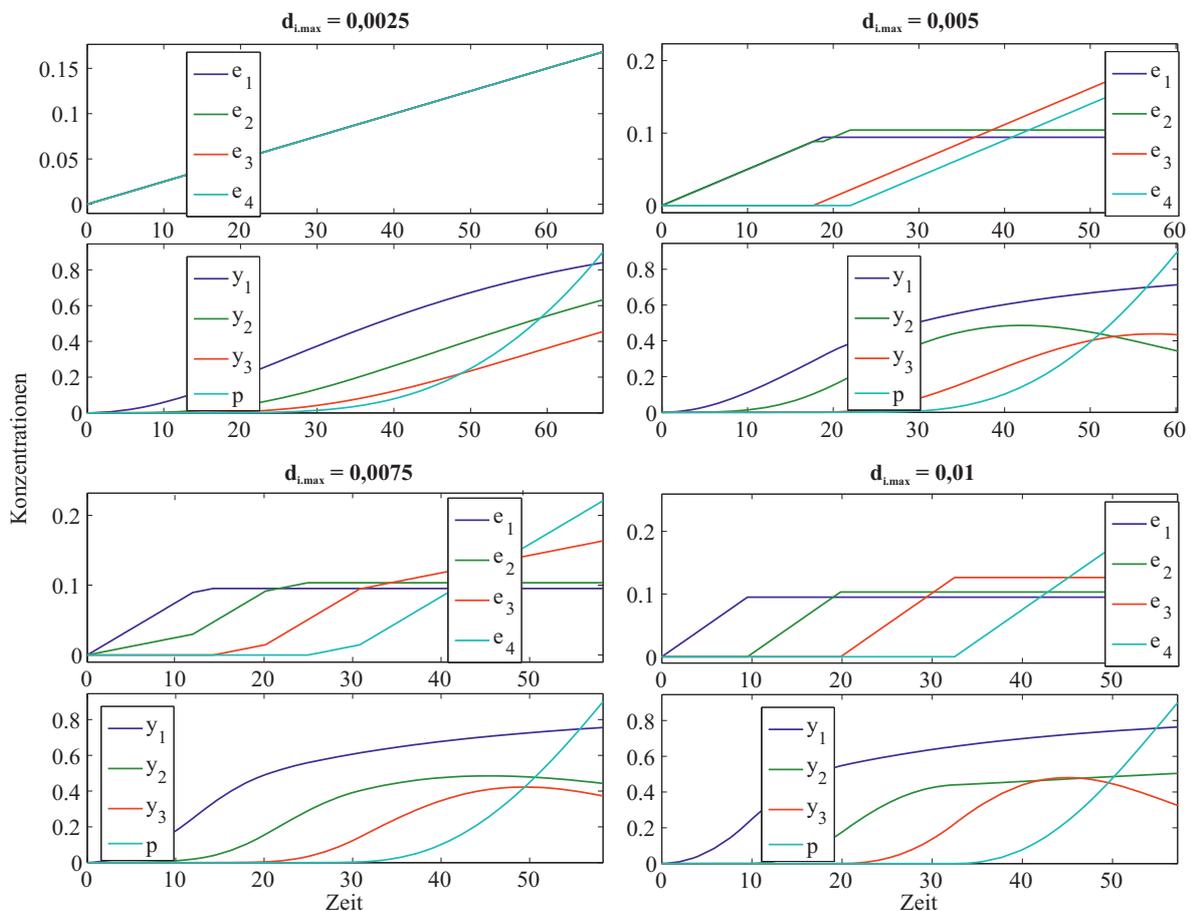


Abbildung 6.3: Zeitoptimale Aktivierung für das Zusammenspiel beider Beschränkungen ( $d_{i,max}$  variabel mit  $d_{max} = 0,01$ )

Diese Aktivierungsstrategien können *in vivo* durch eine Organisation in Operons realisiert werden. Dies kann z.B. für Abbildung 6.2 von einer Ansammlung aller Enzyme in ein Operon (links oben), die Enzyme  $e_{1,2}$  in ein Operon und  $e_{3,4}$  separat (rechts oben), bis hin zur einzelnen Regulation der Enzyme umgesetzt werden (rechts unten). Wird nun die Anzahl von Genen in einem Operon betrachtet, dann fällt der Wert für die steigenden Enzymsyntheseraten von 4 Genen pro Operon auf 1 Gen pro Operon. Für eine konsistente Darstellung der

Ergebnisse, wird in dieser Arbeit der Begriff „1 Gen pro Operon“ verwendet, auch wenn laut exakter Definition ein Operon mindestens zwei Gene beinhaltet.

### Einfluss von zufälligen kinetischen Parametern

In einer nächsten Untersuchung soll aufgezeigt werden, wie sensitiv die operon-basiert Kontrolle der Stoffwechselnetzwerke zu der Wahl der kinetischen Parameter ist. Dazu werden zufällig 100 Parametersets im Intervall  $[0,2]$  erzeugt und die zeitoptimale Steuerung mit variierenden Enzymsyntheseraten für alle Parametersets jeweils berechnet.

Die Ergebnisse werden in Abbildung 6.4 zusammengefasst und Gene pro Operon mittels Fehlerbalken dargestellt (siehe Anhang A.3 für die Ermittlung der Gene pro Operon). Für kleine Werte der einzelnen Enzymsyntheseraten ergibt sich die gleichzeitige Synthese aller Enzyme zur optimalen Aktivierungsstrategie (4 Gene pro Operon). Steigen die Werte der Enzymsyntheseraten ergibt sich wieder eine Mischform mit tendenziell gleicher Operonorganisation, wie für die Wahl von Einheitsparametern (Gene pro Operon zwischen 4 und 1). Für große Werte der einzelnen Enzymsyntheseraten wird letztendlich die sequentielle Aktivierung entlang des Netzwerkes dominierend umgesetzt (meist 1 Gen pro Operon). Dies zeigt, dass eine operon-basierte Regulation auch für eine Vielzahl von Realisierungen der Stoffwechselparameter optimal ist.

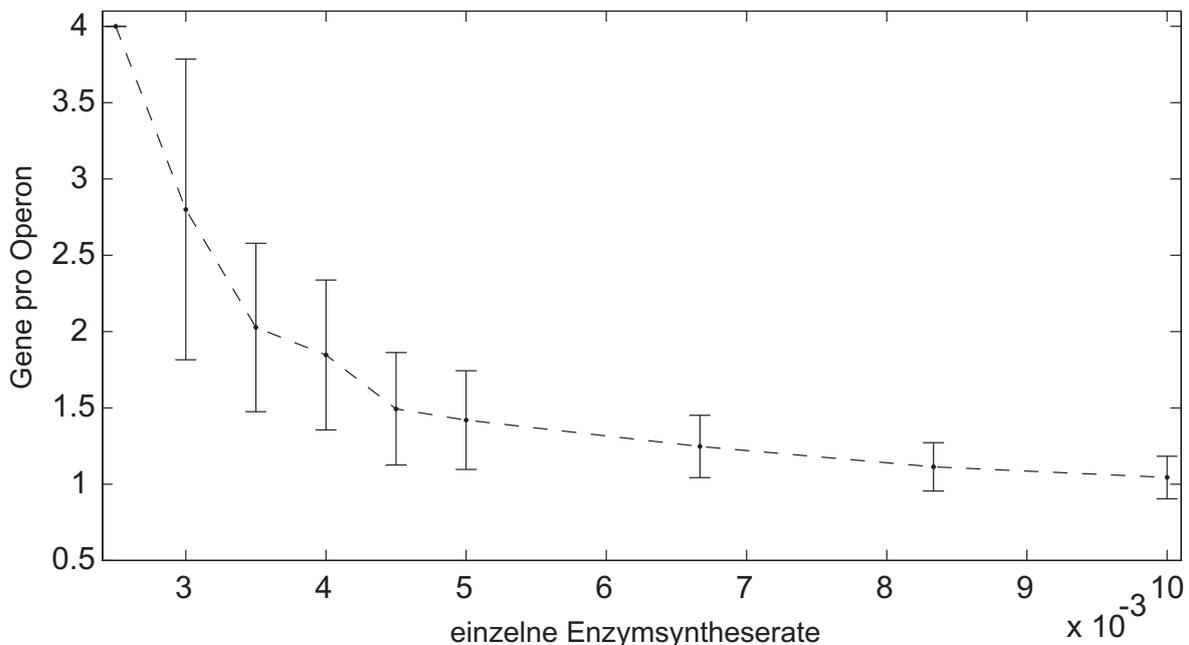


Abbildung 6.4: Operon-basierte, zeitoptimale Aktivierung für zufällige Parametersets ( $d_{i,max}$  variabel mit  $d_{max} = 0,01$ )

### 6.3.2 Wachstumsoptimale Aktivierung

Im vorhergehenden Abschnitt wurden Aktivierungsstrategien untersucht, die die Produktionszeit eines Produktes minimieren. Der Ansatz der Zeitoptimalität hat jedoch Nachteile, wie die ungenügende Beachtung des Wachstums und der Definition einer „künstlichen“ Nebenbedingung zur Aktivierung des Netzwerkes ( $p(t_f) = p_f$ ). Da selbst unter diesen Nachteilen der Einfluss bzw. die Notwendigkeit der Beschränkungen in der Proteinsynthese zu erkennen ist, werden in einem nächsten Schritt die Untersuchungen der wachstumsoptimalen Aktivierung unter den limitierenden Faktoren der Proteinsynthesekapazität gleichzeitig durchgeführt.

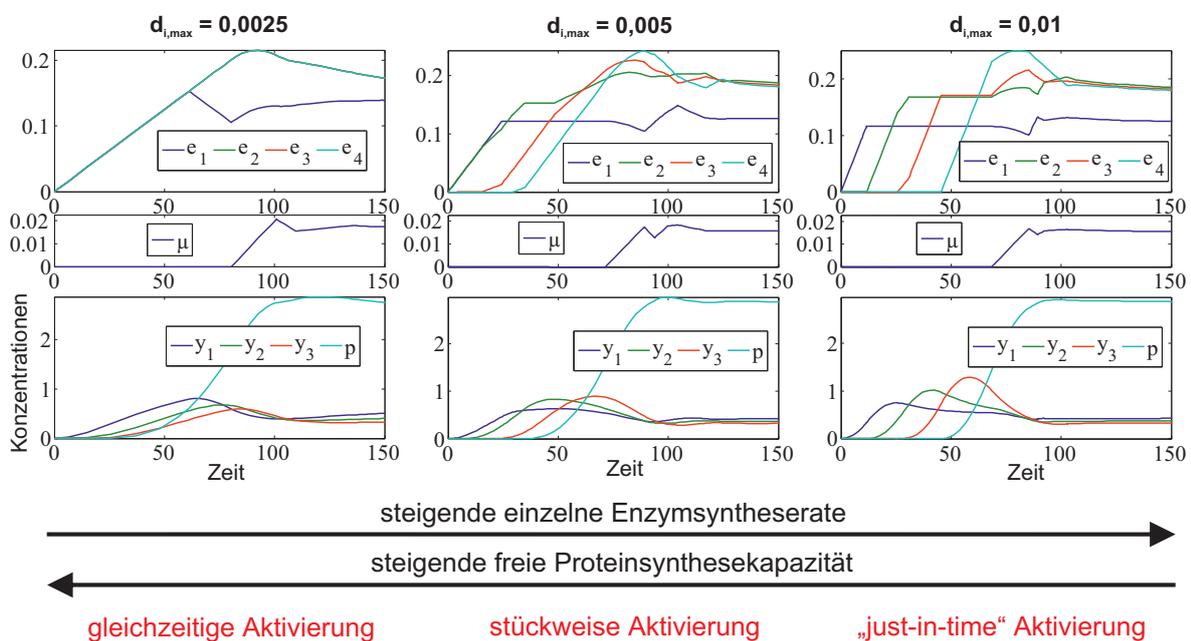


Abbildung 6.5: Wachstumsoptimale Aktivierung für das Zusammenspiel beider Beschränkungen ( $d_{i,max}$  variabel mit  $d_{max} = 0,01$ )

Die entsprechenden Ergebnisse sind in Abbildung 6.5 dargestellt. Dabei wird die freie Proteinsynthesekapazität fixiert und die Enzymsyntheserate variiert (wie bei den zeitoptimalen Untersuchungen). Zur Veranschaulichung wird ausschließlich der Zeitraum gezeigt bis das Netzwerk den aktiven Zustand erreicht hat (die vollständigen Verläufe bis  $t_f = 1000$  sind im Anhang A.3 dargestellt). Dieser stellt dann entsprechend einen optimalen Betriebspunkt des Netzwerkes dar. Für den Ansatz der Zeitoptimalität ist dieser Zustand ungeklärt und muss zusätzlich definiert werden (in dieser Arbeit über  $p_f$ ). Aus Abbildung 6.5 wird zudem ersichtlich, dass für den optimalen Betriebspunkt hohe Produktkonzentrationen erreicht werden. Dieser Zusammenhang kann auch in experimentellen Daten für *Escherichia coli* [28] und *Saccharomyces cerevisiae* [115] gefunden werden.

Es ergeben sich, im Vergleich zum zeitoptimalen Ansatz, für die Aktivierungsphasen ähnliche optimale Strategien der Netzwerkaktivierung in Abhängigkeit der Proteinsynthesekapazität. Ist die Summe der einzelnen Enzymsyntheseraten gleich (oder kleiner) als die freie Proteinsynthesekapazität werden alle Enzyme von Beginn an gleichzeitig produziert (Abbildung 6.5 links,  $e_1, e_2, e_3, e_4$  liegen übereinander). Es ergibt sich somit eine gleichzeitige optimale Aktivierung des Netzwerkes. In Abbildung 6.5 rechts wird die optimale Aktivierungsstrategie für Enzymsyntheseraten dargestellt, welche gleich (oder größer) der freien Proteinsynthesekapazität sind. Hierfür ergibt sich eine sequentielle Synthese der Enzyme entlang der Position im Stoffwechselnetzwerk, was der schon angesprochenen „just-in-time“-Aktivierung entspricht. Für den Fall, dass die einzelnen Enzymsyntheseraten kleiner sind als die freie Proteinsynthesekapazität und deren Summe größer als die freie Proteinsynthesekapazität ergibt sich wieder ein Mischverhalten durch eine stückweise Aktivierung als optimal (Abbildung 6.5 Mitte).

Diese Ergebnisse werden zur Veranschaulichung in Abbildung 6.5 unten zusammenfassend dargestellt. Mit steigender Enzymsyntheserate verändert sich die Aktivierungsstrategie von einer gleichzeitigen Aktivierung, auf eine stückweise Aktivierung und zu einer sequentiellen („just-in-time“) Aktivierung. Da diese Strategien sich in Abhängigkeit von den Verhältnissen der Beschränkungen (einzelnen Enzymsyntheserate und freie Proteinsynthesekapazität) ergeben, kann geschlussfolgert werden, dass sich mit steigender freier Proteinsynthesekapazität die Aktivierungsstrategie von einer sequentiellen Aktivierung, auf eine stückweise Aktivierung und zu einer gleichzeitigen Aktivierung verändert.

#### **Optimale Aktivierungsstrategien für zufällige kinetische Parameter**

Hier gilt es zu untersuchen, wie sensitiv eine operon-basierte wachstumsoptimale Aktivierung der Stoffwechselnetzwerke zu der Wahl der kinetischen Parameter ist. Dazu werden die 100 zufälligen Parametersets aus 6.3.1 verwendet und die optimalen Aktivierungen mit variierenden Enzymsyntheseraten für diese Parametersets berechnet. Die Ergebnisse der wachstumsoptimalen Aktivierung sind in Abbildung 6.6 durch ihrer Merkmale in der Operonrealisierung dargestellt (siehe wiederum Anhang A.3 für die Ermittlung der Gene pro Operon). Das prinzipielle Verhalten ist ähnlich zur zeitoptimalen Aktivierung von Abbildung 6.4. Zusammenfassend wird aus Abbildung 6.6 ersichtlich, dass

- die Größe der Operons (entspricht der Anzahl von Genen im Operon) sich mit steigender Enzymsyntheserate verkleinert.
- Das heißt, wenn sich die einzelnen Enzymsyntheseraten relativ zur freien Proteinsynthesekapazität vergrößern, dann reduziert sich auch die optimale Größe der Operons.

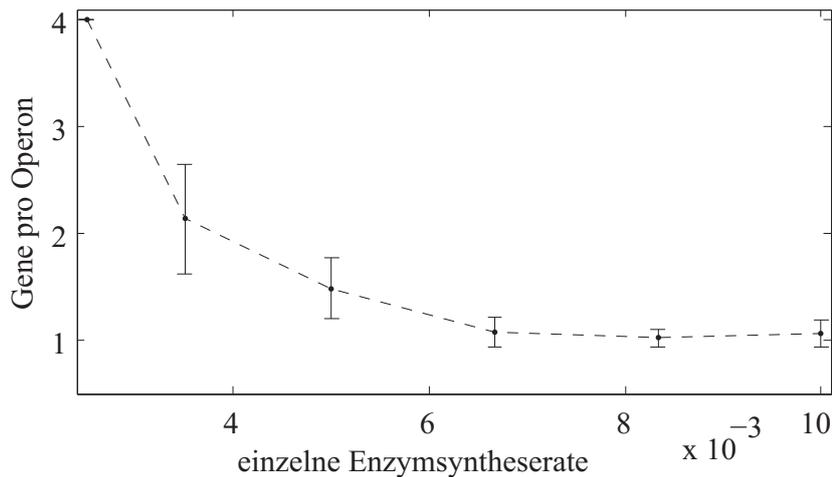


Abbildung 6.6: Operon-basierte, wachstumsoptimale Aktivierung für zufällige Parametersets ( $d_{i,max}$  variabel mit  $d_{max} = 0,01$ )

## 6.4 Konsequenzen für die Aktivierungsstrategien in Prokaryoten

### 6.4.1 Ableitung von Vorhersagen

Die im vorhergehenden Kapitel dargestellten Ergebnisse verdeutlichen den Einfluss des Zusammenspiels der einzelnen Enzymsyntheseraten und der freien Proteinsynthesekapazität für die Annahmen einer zeitoptimalen Aktivierung 6.3.1 und der Annahme einer wachstumsoptimalen Aktivierung 6.3.2. Hierbei stellt sich eine operon-basierte Aktivierung heraus, welche mit steigenden einzelnen Enzymsyntheseraten relativ zur freien Proteinsynthesekapazität eine fallende Größe der Operons als optimal vorhersagt (siehe Abbildung 6.3 und 6.5). Diese Aussage behält auch für zufällige Parametersets und für beide Annahmen der Aktivierungsstrategien ihre Gültigkeit (vergl. Abbildung 6.4 und 6.6).

Für weitere Ableitungen werden ausschließlich die wachstumsoptimalen Aktivierungen weiter untersucht. Dies bewahrt nicht nur die Übersichtlichkeit der vorliegenden Arbeit, sondern ist auch darin begründet, dass diese die Zielstellung des Netzwerkes besser widerspiegelt (explizite Behandlung der Wachstumsrate und das Erreichen eines optimalen Stoffwechselzustandes). Für die Einheitsparameter ergibt sich eine sequentielle oder stückweise sequentielle bis hin zur „just-in-time“-Aktivierung von Enzymen als optimal. Werden die optimalen Aktivierungszeitpunkte für zufällige Parametersets näher untersucht, ergibt sich dass diese Reihenfolge häufig durch eine vorgezogene Aktivierung von Enzymen aus dem hinteren Netzwerkteil nicht eingehalten wird. Diese Umstrukturierung in den Aktivierungssequenzen wird durch verschiedene zu erreichende Abundanzen der Enzyme relativ zu den anderen Enzymen im Netzwerk bestimmt, welche durch die zufälligen Parameter und somit durch die verschiedenen katalytischen Effektivitäten verursacht werden. Zur Veranschaulichung dient

Abbildung 6.7, welchen den Einfluss der relativen Enzymabundanz auf die Reihenfolge in der Aktivierungssequenz für die zufälligen 100 Parametersets mit  $d_{max,i} = 0,005$  darstellt (hierfür wäre eine stückweise sequentielle Aktivierung anhand der vorhergehenden Untersuchungen zu erwarten). Im Anhang A.3 wird die Ermittlung von Enzymabundanz und der Positionen in den Aktivierungssequenzen dargestellt.

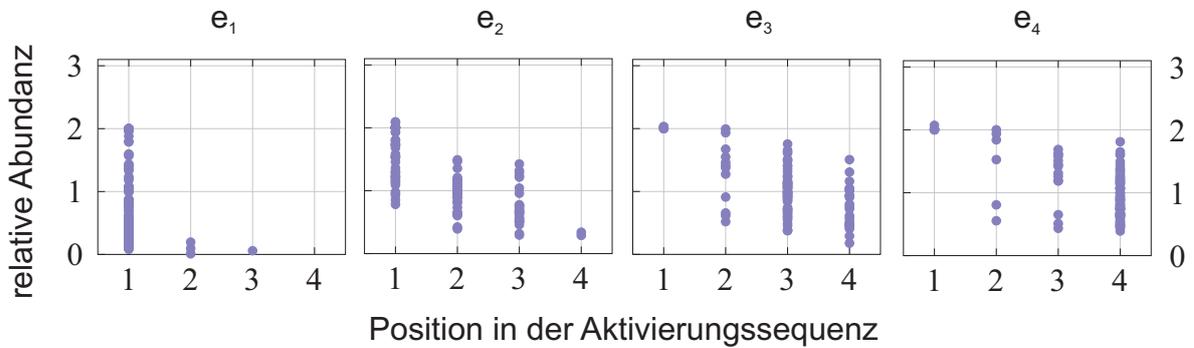


Abbildung 6.7: Einfluss der Enzymabundanz auf die Position in der Aktivierungssequenz

Es wird aus der Abbildung 6.7 ersichtlich, dass sich eine vorgezogene Aktivierung von Enzymen mit einer hohen relativen Abundanz im Vergleich zu Enzymen mit durchschnittlicher Abundanz in einem Netzwerk als optimal ergibt. Außerdem kann hier ein weiterer Zusammenhang abgeleitet werden, dass Enzyme mit einer niedrigen relativen Abundanz tendenziell später aktiviert werden als die umgebenen Enzyme. Im Anhang A.3 wird gezeigt, dass sich auch ein prinzipiell ähnliches Verhalten für verschiedene einzelnen Enzymsyntheseraten ergibt.

Diese Ergebnisse werden nun auf eine operon-basierte Aktivierung von Netzwerken übertragen, also auf eine entsprechende Koexpression der Gene. Es kann angenommen werden, dass Gene für hohe Enzymabundanz relativ zu anderen Enzymen eines Netzwerkes mit vorhergehenden Enzymen des gleichen Netzwerkes tendenziell koexpressiert werden. Für die spätere Aktivierung von gering abundanten Enzymen ergibt sich analog eine Koexpression von relativ betrachteten niedrigen Enzymabundanz mit darauf folgenden Enzymen eines Netzwerkes. In einem nächsten Schritt wird analysiert, ob diese Art der operon-basierten Aktivierung für die optimalen Steuerungen verifiziert werden können. Dazu wird für die berechneten Ergebnisse aller zufälligen Parametersets eine positionsabhängige Verschiebung der Koexpression mit anderen Enzymen durch Operons für gering (siehe links in Abbildung 6.8) und abundante Enzyme (siehe rechts in Abbildung 6.8) ermittelt. Das nötige Vorgehen für die Erstellung dieser Abbildung wird im Anhang A.3 beschrieben. Diese berechnete Verteilung verdeutlicht, wie oft Enzyme in einem Operon mit früheren (negative positionsabhängige Verschiebung) und späteren Enzymen (positive positionsabhängige Verschiebung) des selben

Netzwerkes koexpressiert werden. Für Enzyme mit einer geringen Abundanz, ergibt sich eine positive positionsabhängige Verschiebung und für Enzyme mit einer hohen Abundanz eine negative positionsabhängige Verschiebung.

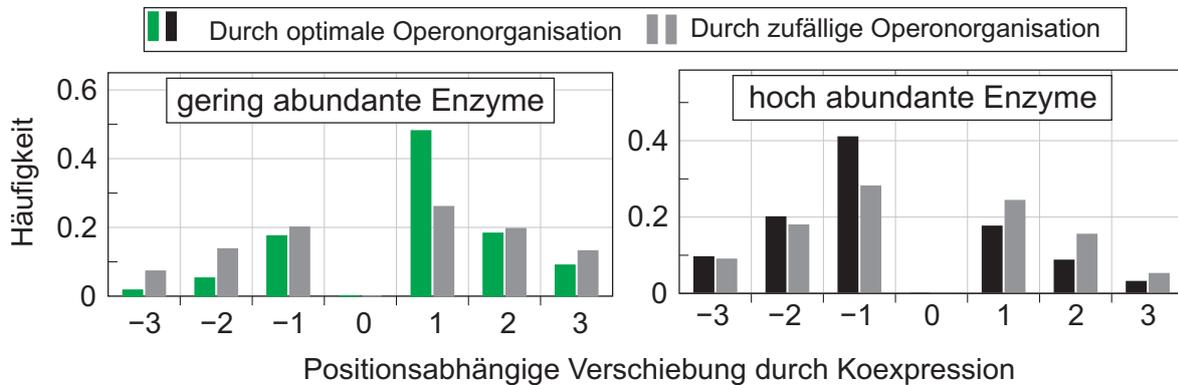


Abbildung 6.8: Positionsabhängige Verschiebung durch Koexpression in Abhängigkeit der Enzymabundanz

Zusätzlich wird an dieser Stelle überprüft, ob sich eine derartige operon-basierte Aktivierung zufällig ergeben kann (graue Balken in Abbildung 6.8) und somit die Enzymabundanz diese Art Realisierung nicht bedingen. Es ergibt sich für gering abundante Enzyme eine signifikant erhöhte Koexpression mit späteren Enzymen eines Netzwerkes (Wilcoxon-Test  $P = 6,64 \cdot 10^{-4}$ ) und für hoch abundante Enzyme eine signifikant erhöhte Koexpression mit früheren Enzymen eines Netzwerkes (Wilcoxon-Test  $P = 3,38 \cdot 10^{-7}$ ) im Vergleich zur zufälligen Operonaktivierung. Dies verdeutlicht, dass sich in der optimalen Operonorganisation eine von den Enzymabundanz beeinflusste Aktivierung widerspiegelt. Die hier durchgeführten Ableitungen führen zusammenfassend zu den folgenden Vorhersagen:

- Im Zusammenspiel der einzelnen Enzymsyntheseraten und der freien Proteinsynthesekapazität ergibt sich mit steigender Enzymsyntheserate eine Reduzierung der Gene pro Operon und mit steigender Proteinsynthesekapazität ergibt sich eine Erhöhung der Gene pro Operon.
- Im Zusammenspiel der einzelnen Enzymsyntheseraten und der freien Proteinsynthesekapazität ergibt sich mit fallender Enzymsyntheserate eine Erhöhung der Gene pro Operon und mit fallender Proteinsynthesekapazität ergibt sich eine Reduzierung der Gene pro Operon.
- Für abundante Enzyme, ergibt sich eine Aktivierung mit vorhergehenden Enzymen und für gering abundante Enzyme, gilt entsprechend eine Aktivierung mit darauffolgenden Enzymen eines Netzwerkes.

### 6.4.2 Verhalten in Prokaryoten

In diesem Kapitel wird basierend auf den Vorhersagen des vorhergehenden Kapitels 6.4.1 eine zielgerichtete Datenanalyse für die Operonorganisation von Stoffwechselnetzwerken in Prokaryoten durchgeführt. Dies geschieht für alle Netzwerke von 550 Prokaryoten aus der Datenbank MicorCyc [139] für welche Informationen zur Operonorganisation in MicrobesOnline [41] vorliegen. Dadurch kann untersucht werden, ob die identifizierten optimalen Aktivierungsstrategien der Stoffwechselnetzwerke unter den eingeführten Kapazitäten in der Proteinsynthese die Operonorganisation für Prokaryoten erklärt. Es gilt hierbei entsprechend das Zusammenspiel der einzelnen Enzymsyntheseraten und der freien Proteinsynthesekapazität auf die Realisierung der Operongrößen und den Einfluss der Abundanz von Enzymen auf die Aktivierungssequenz zu untersuchen.

Die hier aufgezeigten Zusammenhänge sind Ergebnisse einer Datenanalyse, durchgeführt von Kooperationspartnern der Friedrich Schiller Universität (Christoph Kaleta, Stefan Schuster). Durch die interdisziplinäre Ausrichtung sind diese Ergebnisse jedoch zum Verständnis zwingend notwendig und werden somit in Kurzform dargestellt.

#### **Einfluss der einzelnen Enzymsyntheseraten und der freien Proteinsynthesekapazität auf die Größe der Operons**

Da Messdaten für die Syntheseraten der Enzyme und der freien Proteinsynthesekapazität für die untersuchten Prokaryoten nicht verfügbar sind, werden hier korrelierende Eigenschaften untersucht.

Ein wichtiger Faktor stellt die Anzahl von protein-erzeugenden Genen im Genom dar, welcher sich negativ auf die freie Proteinsynthesekapazität auswirkt. Diese Annahme kann damit begründet werden, dass mit einer steigenden Anzahl von protein-erzeugenden Genen der verfügbare Anteil der Proteinsynthesekapazität des betrachteten Stoffwechselnetzwerkes fällt, da entsprechend mehr Proteine zur selben Zeit synthetisiert werden müssen, um andere zelluläre Funktionen aufrecht erhalten zu können [125].

Prinzipiell wirkt sich eine steigende Konzentration von Ribosomen in der Zelle positiv auf die freie Proteinsynthesekapazität aus, da mehr Proteine durch mehr Ribosomen synthetisiert werden können. Dies wird durch positive Korrelationsbeziehungen der Wachstumsrate zu Ribosomenkonzentration [21] und zu der Kopienanzahl der ribosomalen RNA (rRNA) [144] untersucht. Dementsprechend können verfügbare Daten für die Kopienanzahl der rRNA Operons zur Analyse hinzugezogen werden.

In einer nächsten Überlegung wird der Einfluss von Enzymabundanzen dargestellt, deren relative Veränderung durch den „codon adaptionindex“ für vorliegende Gensequenzen abgeschätzt werden kann [126]. Befinden sich durchschnittlich hohe abundante Enzyme in einem Stoffwechselnetz, dann reduziert sich der Abstand von einzelnen Enzymsyntheseraten und freier Proteinsynthesekapazität, im Vergleich zu Netzwerken mit gering abundanten Enzymen.

Entsprechend dieser Zusammenhänge (abgeleitet aus den optimalen Aktivierungsstrategien) ist es anzunehmen, dass die Anzahl von protein-erzeugende Gene, die Anzahl der rRNA Operons und die relative Abundanz der Enzyme („codon adaption index“) einen Einfluss auf die Größe der Operons von Stoffwechselwegen haben. Dies resultiert in folgende Hypothesen:

- **Hypothese 1:** Die Größe der Operons entlang eines Stoffwechselnetzwerkes reduziert sich mit einer steigenden Anzahl von Protein erzeugenden Genen im Genom, da sich entsprechend die freie Proteinsynthesekapazität verringert.
- **Hypothese 2:** Die Größe der Operons entlang eines Stoffwechselnetzwerkes vergrößert sich mit einer steigenden Anzahl von rRNA Operons im Genom, da sich entsprechend die freie Proteinsynthesekapazität vergrößert.
- **Hypothese 3:** Die Größe der Operons entlang eines Stoffwechselnetzwerkes reduziert sich mit höheren Abundanzen der Enzyme des Stoffwechselnetzwerkes, da sich in Summe die einzelnen Enzymsyntheseraten der freien Proteinsynthesekapazität entsprechend annähert.

Diese Hypothesen werden nun für metabolische Netzwerke der Prokaryoten getestet, welche in mindestens 100 Prokaryoten vorkommen. Dies ergibt eine Datenanalyse von 99 Netzwerken in 550 Prokaryoten aus MicorCyc [139]. Anschließend wird die Spearman Korrelation zwischen der Größe der Operons zu den Faktoren aus Hypothese 1, 2 und 3 berechnet (jeweils kontrolliert auf die Beeinflussung der anderen Faktoren). Die Ergebnisse werden netzwerkbezogen in Abbildung 6.9 dargestellt, wobei zusätzlich die Anzahl der Netzwerke in den 550 Prokaryoten angegeben wird. In Abbildung 6.9 entspricht Grün einer „false discovery rate“ FDR [17] von bis zu 5% und Hellgrün bis 20%, wenn diese mit den Hypothesen übereinstimmen und Rot mit einer FDR von bis 5% und Hellrot bis 20%, wenn diese die Hypothesen ablehnen. Gezeigt werden nur Netzwerke mit mindestens einer FDR von unter 5% (für Annahme und Ablehnung) was zu einer Reduzierung auf 69 Netzwerke führt. Eine Auswertung von Abbildung 6.9 ergibt, dass

## 6.4 Konsequenzen für die Aktivierungsstrategien in Prokaryoten

Netzwerkname	#	Hyp. 1	Hyp. 2	Hyp. 3
galactose degradation I (Leloir pathway)	170	-0,21	0,23	-0,48
UDP-N-acetyl-D-glucosamine biosynthesis I	462	0,04	-0,17	-0,11
uridine-5'-phosphate biosynthesis	462	-0,19	0,33	0,05
mixed acid fermentation	364	0,04	0,30	0,09
N-acetylglucosamine degradation II	104	0,28	0,23	-0,30
7-keto-8-aminopelargonate biosynthesis I	313	-0,12	0,09	-0,32
molybdenum cofactor biosynthesis	160	-0,24	0,09	0,03
leucine biosynthesis	402	-0,16	0,27	-0,23
tetrahydrofolate biosynthesis	199	-0,20	-0,11	-0,16
glycogen degradation I	169	-0,23	0,42	-0,27
ubiquinol-8 biosynthesis (prokaryotic)	270	-0,03	0,19	-0,11
urate biosynthesis/inosine 5'-phosphate degradation	220	0,27	-0,15	-0,14
inosine-5'-phosphate biosynthesis I	262	0,00	0,23	-0,04
5-aminoimidazole ribonucleotide biosynthesis II	208	-0,42	-0,30	-0,06
valine biosynthesis	409	0,04	0,20	-0,18
5-aminoimidazole ribonucleotide biosynthesis I	431	-0,27	0,36	-0,02
adenosine nucleotides de novo biosynthesis	514	-0,18	0,25	0,04
guanosine nucleotides de novo biosynthesis	452	0,02	0,32	-0,14
cholesterol degradation to androstenedione I (cholesterol oxidase)	200	-0,16	0,15	-0,23
cholesterol degradation to androstenedione II (cholesterol dehydrogenase)	164	-0,13	0,16	-0,27
heme biosynthesis from uroporphyrinogen-III I	296	-0,19	0,00	-0,01
isoleucine biosynthesis I (from threonine)	393	-0,05	0,23	-0,27
TCA cycle VI (obligate autotrophs)	291	0,13	0,34	0,06
pentose phosphate pathway (oxidative branch)	289	0,06	-0,25	-0,05
arginine biosynthesis II (acetyl cycle)	354	-0,12	0,19	0,05
glycolysis III (glucokinase)	307	-0,22	0,14	-0,04
lysine biosynthesis VI	235	-0,34	0,13	0,24
pyrimidine ribonucleosides degradation I	109	-0,23	-0,25	0,29
pyrimidine deoxyribonucleosides degradation	104	0,15	0,18	0,32
proline biosynthesis I	376	-0,19	0,29	-0,27
dTDP-L-rhamnose biosynthesis I	305	-0,16	0,19	-0,26
heme biosynthesis from uroporphyrinogen-III II	336	-0,27	0,18	0,02
aerobic respiration (cytochrome c)	147	-0,03	0,07	0,39
leucine degradation I	113	-0,09	0,03	-0,36
cis-dodecenoyl biosynthesis	323	0,19	-0,11	-0,21
pyrimidine deoxyribonucleotides de novo biosynthesis I	343	-0,04	0,14	-0,24
ADP-L-glycero-beta;-D-manno-heptose biosynthesis	105	-0,19	-0,44	0,12
tyrosine biosynthesis I	137	0,03	-0,16	0,26
respiration (anaerobic)	412	0,02	0,25	0,05
UDP-N-acetylmuramoyl-pentapeptide biosynthesis III (meso-DAP-containing)	467	0,39	-0,28	-0,19
palmitoleate biosynthesis I	436	0,06	0,00	-0,21
fatty acid beta;-oxidation I	255	-0,14	0,04	-0,32
CMP-KDO biosynthesis I	200	0,11	0,30	0,28
lysine biosynthesis I	286	0,00	0,09	0,34
lipid IVA biosynthesis	291	0,15	0,09	-0,24
adenosylcobalamin biosynthesis I (early cobalt insertion)	236	-0,22	0,26	-0,06
adenosylcobalamin biosynthesis from cobyrinate a,c-diamide II	249	0,09	0,12	-0,29
adenosylcobalamin biosynthesis from cobyrinate a,c-diamide I	265	0,09	0,11	-0,28
adenosylcobalamin biosynthesis II (late cobalt incorporation)	236	-0,03	0,22	-0,11
arginine biosynthesis I	401	-0,16	0,13	0,03
peptidoglycan biosynthesis I (meso-diaminopimelate containing)	465	0,13	-0,17	-0,20
tryptophan biosynthesis	436	-0,17	0,20	-0,28
tetrapyrrole biosynthesis I	384	0,04	0,25	0,01
xylose degradation IV	216	0,01	-0,28	-0,12
gluconeogenesis II (Methanobacterium thermoautotrophicum)	212	-0,52	-0,10	0,02
6-hydroxymethyl-dihydropterin diphosphate biosynthesis I	354	-0,17	0,01	-0,11
1,4-dihydroxy-2-naphthoate biosynthesis I	191	-0,25	0,37	-0,35
serine biosynthesis	250	0,02	-0,26	-0,08
threonine degradation I	101	0,24	0,05	-0,45
coenzyme A biosynthesis	408	-0,19	0,04	-0,31
siroheme biosynthesis	269	-0,16	0,28	0,25
flavin biosynthesis I (bacteria and plants)	403	-0,16	0,17	0,06
glycogen biosynthesis I (from ADP-D-Glucose)	217	0,09	0,39	-0,35
methylerythritol phosphate pathway	391	0,08	0,22	-0,06
histidine biosynthesis	416	0,00	0,16	-0,37
chorismate biosynthesis from 3-dehydroquinate	465	-0,18	0,12	-0,06
GDP-mannose biosynthesis	200	0,07	-0,04	-0,58
phosphopantothenate biosynthesis I	342	-0,23	0,10	-0,17
histidine degradation I	128	0,07	0,00	-0,32
palmitate biosynthesis II (bacteria and plants)	355	0,23	-0,16	-0,17

Abbildung 6.9: Einfluss der Proteinsynthesekapazität auf die Größe der Operons *in vivo*

- **Hypothese 1** für 22 der 99 untersuchten Netzwerke verifiziert werden konnte und für nur 4 Netzwerke abgelehnt wurde,
- **Hypothese 2** für 29 der untersuchten Netzwerke verifiziert werden konnte und für 9 Netzwerke abgelehnt wurde und
- **Hypothese 3** für 32 der untersuchten Netzwerke verifiziert werden konnte und für 8 Netzwerke abgelehnt wurde.

Für insgesamt 5 Stoffwechselnetzwerke wurden alle Hypothesen angenommen, wobei keine derartige Konstellation für die Ablehnung der Hypothesen ermittelt wurde. Hier ergibt sich, bei den angenommenen Hypothesen ein Anteil von 3 Stoffwechselnetzwerken aus der Aminosäurebiosynthese (Lysin, Prolin, Tryptophan).

Dies zeigt, dass die individuellen Enzymsyntheseraten und freien Proteinsynthesekapazitäten starke Einflüsse auf die Größe der Operons in Stoffwechselwegen besitzen. Auch wenn die Hypothesen nicht in allen Fällen angenommen wurden, ergibt sich doch ein überwiegendes Verhältnis von 83 Annahmen zu 21 Ablehnungen über alle 99 Stoffwechselnetzwerke. Die Annahme aller Hypothesen für 3 Stoffwechselnetzwerke aus der Aminosäuresynthese, spiegelt letztendlich wieder, dass die Synthese von benötigten Aminosäuren notwendig ist, um ein Wachstum wieder aufnehmen zu können und ein dementsprechend hoher evolutionärer Druck für die effektiven Aktivierungen dieser Stoffwechselwege vorliegt.

### **Einfluss der Enzymabundanz auf den Aktivierungszeitpunkt eines Enzyms**

Die zweite Vorhersage aus Kapitel 6.4.1 ergibt für hoch abundante Enzyme, eine Aktivierung mit vorhergehenden Enzymen und für gering abundante Enzyme eine Aktivierung mit darauffolgenden Enzymen des Netzwerkes. Wie bereits beschrieben, werden die Enzymabundanz durch den „codon adaption index“ abgeschätzt und in hoch bzw. gering abundante Enzyme unterteilt.

Die Untersuchungen basieren auf einer Bestimmung der positionsabhängigen Verschiebung durch Operons in Abhängigkeit von hoch bzw. gering abundanten Enzymen der Organismen aus MicroCyc. Die positionsabhängige Verschiebung beschreibt, wie stark Enzyme in einem Operon mit früheren (negative positionsabhängige Verschiebung) und späteren Enzymen (positive positionsabhängige Verschiebung) des selben Netzwerkes koexprimiert werden. Die Ergebnisse sind in Abbildung 6.10 dargestellt, dabei entspricht jeder Punkt der durchschnittlichen positionsabhängigen Verschiebung eines Organismus. Dargestellt werden weiterhin die Verteilung der positionsabhängigen Verschiebung über die Organismen (Hüllkurven) und der

absolute Mittelwert (schwarze Balken). Es ergibt sich ein Organismus übergreifender Trend einer positiven positionsabhängigen Verschiebung für Enzyme mit einer geringen Abundanz und einer negativen positionsabhängigen Verschiebung für Enzyme mit einer hohen Abundanz. Ein Vergleich mit einer zufälligen Realisierung der Operons über die Organismen, ergibt für gering abundante Enzyme eine signifikant erhöhte Koexpression mit späteren Enzymen eines Netzwerkes (Wilconxon-Test  $P = 1,1 \cdot 10^{-8}$ ) und für hoch abundante Enzyme eine signifikant erhöhte Koexpression mit früheren Enzymen eines Netzwerkes (Wilconxon-Test  $P = 2,1 \cdot 10^{-4}$ ) im Vergleich zur zufälligen Operonorganisation. Entsprechend ergibt sich ein Trend für hoch abundante Enzyme mit einer vorgezogenen Aktivierung und ein Trend für gering abundante Enzyme mit einer späteren Aktivierung.

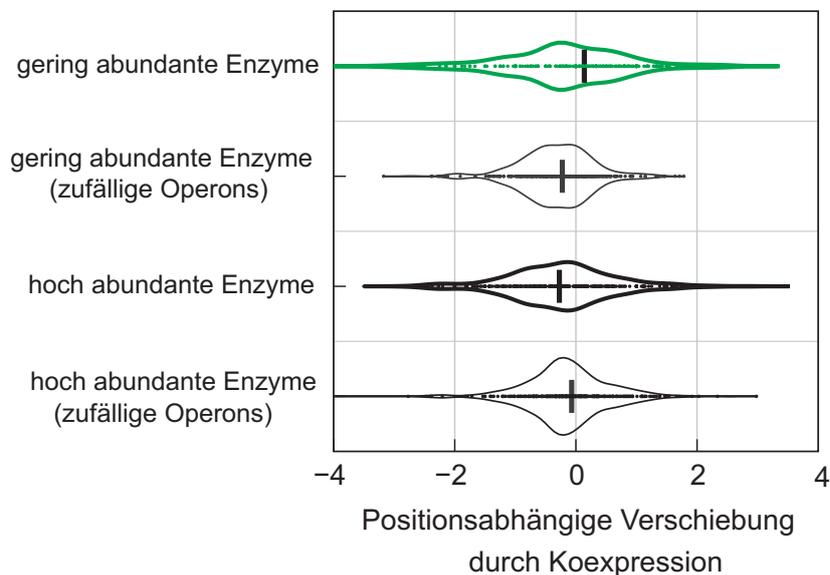


Abbildung 6.10: Einfluss der Enzymabundanz auf die Operonorganisation *in vivo*

Dies motiviert eine konkrete Analyse der Aktivierung in der Argininbiosynthese von *Escherichia coli* [158], worin eine vorgezogene Aktivierung des Enzyms ArgG (7. Postion im Netzwerk) experimentell beobachtet wurde. Eine Analyse von Abundanzdaten aus *Escherichia coli* [154] zeigt das ArgG das höchst abundante Enzym des Netzwerkes ist (zur Veranschaulichung dient Abbildung 6.11). Dies ist konform mit den obigen Ergebnissen einer vorgezogenen

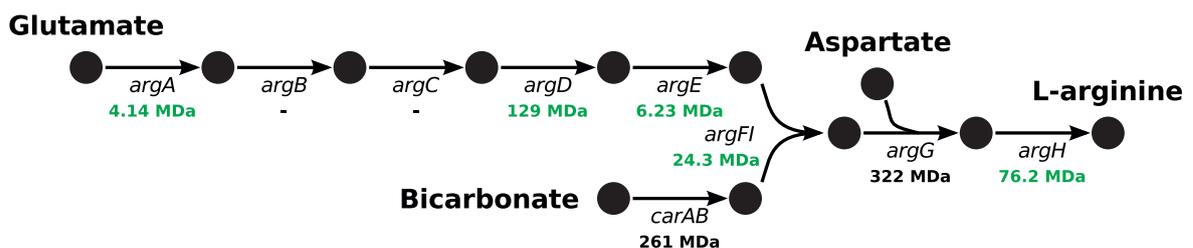


Abbildung 6.11: Enzymabundanz in der Argininbiosynthese von *Escherichia coli*

Aktivierung von hoch abundanten Enzymen. Weiterhin ist die Aktivierung von *carAB* (auch eine hoch abundantes Enzym) der Aktivierung von *argE* ebenfalls vorgezogen, obwohl deren Produkte für den nächsten Stoffwechselschritt gleichzeitig notwendig sind.

### 6.5 Diskussion

Dieses Kapitel beschreibt eine weitere Anwendung der dynamischen Optimierung (im speziellen der optimalen Steuerung) zur Identifikation und anschließender Verifikation von Regulationsmechanismen in Stoffwechselnetzwerken. Im Vergleich zu Kapitel 5 wurden hier die identifizierten Signaturen, diese entsprechen den prinzipiellen Eigenschaften der optimalen Aktivierung, für eine Vielzahl von Stoffwechselwegen in einigen hundert Prokaryoten nachgewiesen. Zudem wurde eine auf die Optimalität des Wachstums (Wachstumsmaximierung) formulierte Problemstellung abgeleitet, um die *in vivo* vorliegende Situation besser abbilden zu können.

Die durchgeführten Untersuchungen basierten auf den plausiblen Annahmen einer limitierten Enzymsyntheserate und Kapazitätsbegrenzungen in der freien Proteinbiosynthese zur Aktivierung von Stoffwechselnetzwerken. Die Ergebnisse der dynamischen Optimierung und der Validierung durch eine Datenanalyse zeigen, dass die Abundanz von Enzymen im Stoffwechselnetzwerk einen sehr großen Einfluss auf die *in vivo* umgesetzten Regulationsstrategien zur Kontrolle der Stoffwechselnetzwerke haben. Hierbei ergibt sich für gering abundante Proteine eine gleichzeitige Aktivierung der Enzyme als optimal, wohingegen sich eine sequentielle Aktivierung der Enzyme für hoch abundante Proteine als optimal ergibt. Dies ist zudem konform zu den Untersuchungen aus Kapitel 5 und [154], in welchen eine fein ausgeprägte Regulation von Stoffwechselnetzwerken in *Escherichia coli* für steigende Proteinmassen gezeigt werden konnte. Ebenfalls konnte gezeigt werden, dass in Abhängigkeit der relativen Enzymmassen (bezogen auf die Enzyme im untersuchten Stoffwechselnetzwerk) eine vorgezogene Aktivierung von hoch abundanten Proteinen und eine verzögerte Aktivierung von gering abundanten Proteinen der Stoffwechselnetzwerke auftritt. Die Ergebnisse werden in Abbildung 6.12 veranschaulicht.

In einer weiteren Untersuchung wurde die Operonorganisation zur Kontrolle der Stoffwechselwege analysiert. Hierbei wurde der Einfluss der individuellen Enzymsyntheserate und der freien Proteinsynthesekapazität auf die Organisation von Enzymen in Operons aufgezeigt. Es ergibt sich ein einfacher Zusammenhang mit einer prinzipiellen Veränderung der Größe der Operons (Anzahl von Enzymen in einem Operon) fallend mit steigenden Proteinmassen und steigend mit steigender Proteinsynthesekapazität. Dementsprechend stellt das Zusam-

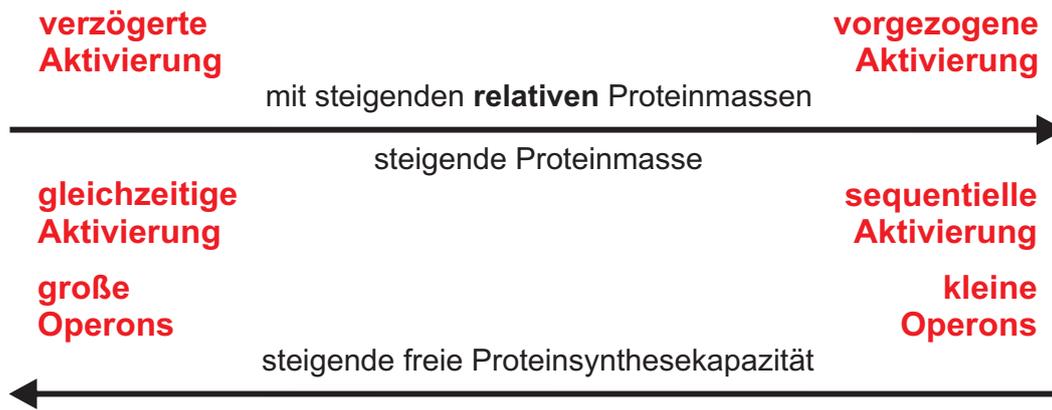


Abbildung 6.12: Aktivierungsstrategien und ihre Abhängigkeiten

Beispiel von zu synthetisierenden Proteinabundanzen und verfügbare Kapazitäten für diese Proteinsynthese einen Einflussfaktor für die Vergrößerung und der Verkleinerung von Operon zur Regulation des Stoffwechsels dar. Da diese Faktoren während einer evolutionären Entwicklung starken Veränderungen unterliegen, könnte dieser Zusammenhang eine mögliche Erklärung für die starken Variationen in der evolutionär betrachteten Operonorganisation [72, 109] darstellen.

Die hier aufgezeigten Ergebnisse widersprechen der gegenwärtigen Auffassung, geprägt durch vorhergehende Arbeiten (siehe z.B. [158]) insoweit, dass eine „just-in-time“-Aktivierung in sequentieller Form nur ein Spezialfall für hoch abundante Proteine darstellt. Weiterhin kann geschlossen werden, dass unter veränderten Umgebungsbedingungen auch ein Übergang zwischen den verschiedenen optimalen regulatorischen Programmen auftritt. So steigen z.B. die Menge der Ribosomen mit der Wachstumsrate (siehe Kapitel 6.4.2), was dazu führt, dass für Umgebungsbedingungen mit hohen Wachstumsraten eine gleichzeitige Aktivierung optimal ist und für Umgebungsbedingungen mit niedrigen Wachstumsraten eine sequentielle Aktivierung umgesetzt wird. Dieser Zusammenhang kann eine Erklärung für aktuelle Beobachtungen sein, wo andere Aktivierungen in der Argininbiosynthese bei hohen Wachstumsraten [57], im Vergleich zu Beobachtungen bei niedrigen Wachstumsraten [158] in *Escherichia coli* auftraten.

Teile dieser interdisziplinären Arbeit wurden in „Nature Communications“

*Bartl, M. and Kötzing, M. and Schuster, S. and Li, P. and Kaleta, C. „Dynamic optimization identifies optimal programs for pathway regulation in prokaryotes“ (2013)*

veröffentlicht (im nachfolgendem Text gekennzeichnet mit [8]). Die detaillierte Vorgehensweise der Datenanalyse dieses Kapitels kann hier nachgelesen werden.

## 7 Schlussbemerkungen und Ausblick

Die vorliegende Arbeit beinhaltet Beiträge in verschiedenen Fachgebieten. Dabei wurden Methodenerweiterungen der dynamischen Optimierung in Kapitel 4 erreicht und Beiträge auf Seiten der Systembiologie zur Untersuchung minimaler Regulationsstrategien von Stoffwechselnetzwerken in Kapitel 5 und zur Untersuchung der Aktivierungsstrategien von Stoffwechselnetzwerken in Kapitel 6 geleistet. Durch die interdisziplinäre Ausrichtung, wurde aus Gründen der Übersichtlichkeit, bereits in den jeweiligen Kapiteln eine detaillierte Diskussion der Ergebnisse durchgeführt und es soll an dieser Stellen nur darauf verwiesen werden. Im Folgendem werden die erzielten Ergebnisse zusammenfassend dargestellt und weitere mögliche Arbeiten aufgezeigt.

### **Erweiterung der quasi-sequentiellen Methode**

Es ergeben sich durch die Erweiterung der quasi-sequentiellen Methode aus Kapitel 4 verbesserte Konvergenzeigenschaften mittels einer Approximationskontrolle der Zustandsgrößen und die Möglichkeit einer adaptiven Diskretisierung (qMFE). Diese adaptive Diskretisierung und eine dazugehörige Approximationskontrolle (vergl. Abbildung 4.12) ermöglichen den Einfluss einer vorher fest definierten Elementplatzierung auf die Verläufe der optimalen Steuerung zu reduzieren und harte Umschaltungen zu detektieren (z.B. Bang-Bang-Steuerung). Diese Erweiterungen sind essentiell und führen zusammenfassend zu einer Aktualisierung des Vergleiches der quasi-sequentiellen Methode mit anderen direkten Methoden (vorher beschrieben in Kapitel 3) in Abbildung 7.1. Zudem wurde eine C++ Bibliothek für die Umsetzung der Funktionalitäten implementiert und entsprechend die Rechenzeiten im Vergleich zu [9] erheblich verkürzt.

In zukünftigen Arbeiten können zusätzlich Information aus der optimalen Steuerungstheorie (diskretisierte Darstellung der Hamilton-Funktion) bei der Lösung des diskretisierten Optimierungsproblems einbezogen werden. In den Arbeiten [134, 135, 88] werden Ansatzpunkte aufgezeigt, welche vielversprechende Verbesserungen der erzielten, optimalen Ergebnisse zur Folge haben. Zudem könnte eine Verbesserung des Rechenaufwandes in der Simulationsschicht der quasi-sequentiellen Methode (vergl. rot hervorgehoben in Abbildung 7.1) durch

eine Vorsimulation zur Startwertermittlung des Newton-Verfahrens erzielt werden, welche entsprechend die nötigen Iterationen des Lösungsverfahrens stark reduzieren kann. Eine weitere Möglichkeit den Rechenaufwand zu reduzieren, besteht in der Kombination des direkten Mehrfachschießverfahrens unter Anwendung der quasi-sequentiellen Methode [133].

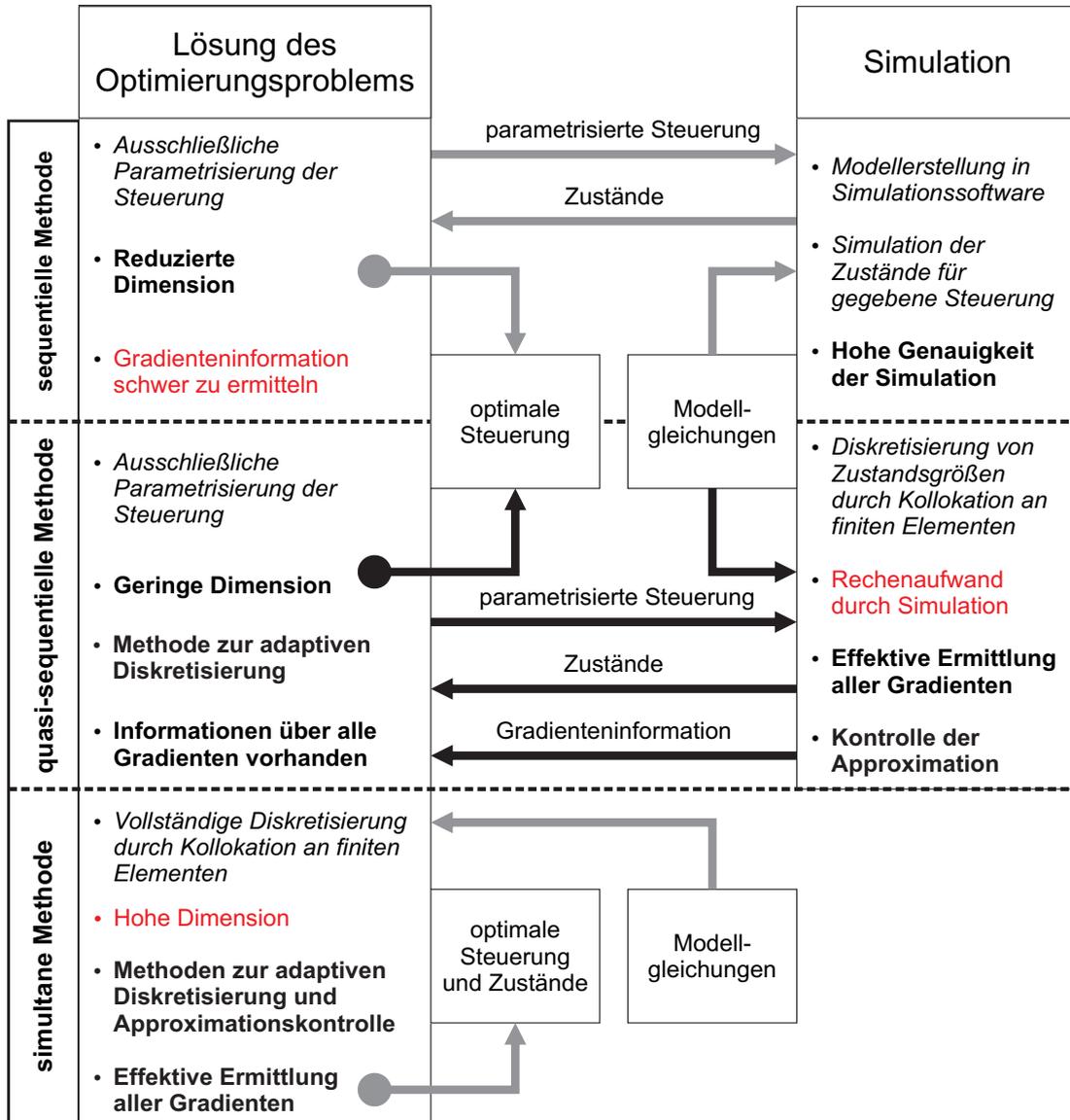


Abbildung 7.1: Direkte Methoden nach den Erweiterungen aus Kapitel 4 im Vergleich

Weiterhin bietet sich die Analyse von „Mathematical Programs with Equilibrium Constraints“ durch die erweiterten Funktionalitäten der quasi-sequentiellen Methode an, da deren Lösung eine adaptive Diskretisierung bedingt [19]. Anschließend kann der Vergleich von verschiedenen Problemstellungen mit dem Lösungsverhalten der adaptiven Diskretisierung der simultanen Methode durchgeführt werden. Dabei können sich Hinweise ergeben, für welche Problemstellungen, sich die MFE der simultanen oder die qMFE der quasi-sequentiellen Methode eignen.

## Optimale Regulation von Stoffwechselnetzwerken

Die gezeigten Analysen haben ergeben, dass sich regulatorische Strategien optimal an verändernde Umgebungsbedingungen anpassen. Dabei wurde gezeigt, welches Potential die Kombination von Werkzeugen der Ingenieurwissenschaften (die optimale Steuerung mit entsprechender Modellierung und Simulation) und Werkzeuge der Lebenswissenschaften (Netzwerkanalyse, Bioinformatik und entsprechende Datenanalysen) haben, um die Prinzipien der regulatorischen Mechanismen von Stoffwechselnetzwerken aufzuzeigen und zu erklären.

Die bevorzugte transkriptionelle Regulation von Enzymen mit hohen Proteinmassen aus Kapitel 5 zeigt, dass sich eine Balance zwischen der Kostenreduktion für die Proteinsynthese und einer schnellen Reaktion auf veränderte Umweltbedingungen in den regulatorischen Mechanismen widerspiegeln [154]. Die Enzyme mit hohen Proteinmassen werden erst synthetisiert, wenn diese von der Zelle benötigt werden. Die durchgeführten Untersuchungen ergaben eine konstitutive Expression für tendenziell günstige Enzyme mit entsprechend geringen Proteinmassen.

Unterliegen Stoffwechselnetzwerke nicht einer dünn verteilten, transkriptionellen Regulation, dann kann sich die Bereitstellungszeit für die benötigten Enzyme problematisch auswirken, da sich eine transkriptionelle Regulation im Bereich von mehreren Minuten befinden kann (siehe z.B. für *Escherichia coli* in [158]). Dieser Einfluss kann durch ein sinnvolles Einsetzen von zur Verfügung stehenden Ressourcen reduziert werden und war Gegenstand der optimalen Aktivierung von Stoffwechselnetzwerken aus Kapitel 6. Dabei wurde der Einfluss einer limitierenden Enzymsyntheserate und Kapazitätsbegrenzungen in der freien Proteinbiosynthese auf die *in vivo* umgesetzten Regulationsstrategien zur Aktivierung der Stoffwechselnetzwerke gezeigt. Es ergab sich ein nachweisbarer Einfluss dieser limitierenden Faktoren auf die Art der *in vivo* umgesetzten Regulationsmechanismen für eine Vielzahl von Stoffwechselwegen in einigen hundert Prokaryoten [siehe auch 8].

Werden nun die in Kapitel 5 und 6 dargestellten Ergebnisse in einem gemeinsamen Kontext gebracht, ergibt sich das Schema aus Abbildung 7.2 für den Übergang der dünn verteilten, transkriptionellen Regulation zu einer optimalen Aktivierung von Stoffwechselnetzwerken mit steigenden bzw. fallenden Proteinmassen. Dieses Gesamtschema veranschaulicht zusammenfassend die einzelnen Einflussfaktoren auf die verschiedenen Arten der Regulationsstrategien.

Die hier durchgeführten Analysen können auf Untersuchungen von groben Ganzzellmodellen (siehe z.B. [98]) in weiteren Arbeiten übertragen werden. Dabei können z.B. zellglobale Strategien (optimale Umsteuerung) zur Anpassung auf verschiedene Zellzyklen oder Stresssituationen untersucht werden. Weiterhin können in der Zelle Stoffwechselnetzwerke

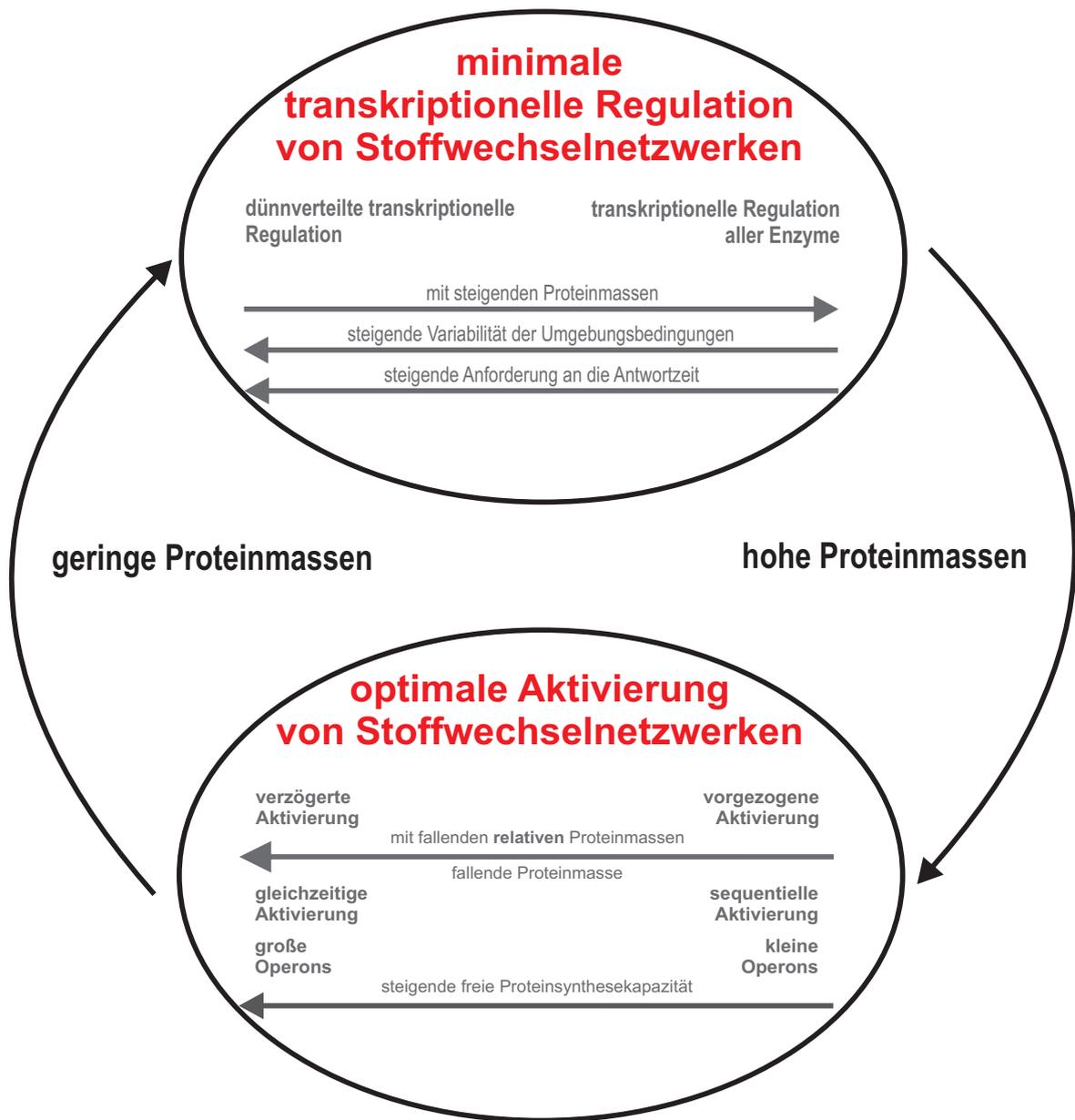


Abbildung 7.2: Regulationsstrategien aus Kapitel 5 und 6 im Kontext

mit hohen Proteinmassen vorliegen, welche eine dünn verteilte, transkriptionelle Regulation aufweisen und erst durch eine Betrachtung von zusammenhängenden, zellulären Prozessen erklärt werden können. Hierbei wirkt der Pentosephosphatweg motivierend (siehe Abbildung 5.21), welcher eine dünn verteilte, transkriptionelle Regulation aufweist und doch relativ hohe Proteinmassen benötigt. Dies entspricht auch früheren experimentellen Beobachtungen [50] und kann damit erklärt werden, dass der Pentosephosphatweg benötigte Stoffe für viele Biosynthesenetzwerke bereitstellt. Da die Datenanalyse aus Kapitel 5 auf *Escherichia coli* basierten, könnten außerdem in weiteren Arbeiten vergleichbare Analysen spezies-übergreifend durchgeführt werden.

Weiterhin besteht Entwicklungsbedarf bei der Reduzierung von zu optimierenden Enzymverläufen (z.B. Unterscheidung in dynamische oder zeitlich gering veränderliche Enzymverläufe oder gar A-Priori-Untersuchungen) und bei einer weicheren Formulierung von Nebenbedingungen (wie z.B. einer kurzfristigen, osmotischen Verletzung). Zudem wurden in Kapitel 2.1 die stochastischen Eigenschaften von Stoffwechselreaktion angesprochen [59, 60, 61, 111]. Diese können dazu führen, dass deterministisch berechnete Regulationsstrategien keine robuste Kontrolle der Stoffwechselnetzwerke ergeben. Die Behandlung von Unsicherheiten kann in weiteren Arbeiten z.B. durch Wahrscheinlichkeitsrestriktionen betrachtet werden, um eine robuste und optimale Steuerung bzw. Regelung zu ermitteln [48, 56, 151].

Während der Bearbeitung der vorliegenden Arbeit wurden zudem Ansätze entwickelt, welche die Optimalität in der Zonierung des Stoffwechselverhaltens im Lebergewebe untersuchen. Basierend auf einem 2-Kompartimentmodell aus [159] konnten detaillierte Zonierungen im Stickstoffstoffwechsel der Rattenleber durch Optimalitätsansätze untersucht werden. Dies betrifft bisher unerklärte, experimentelle Expressionen der Glutaminase [100], örtliche Harnstoffbildungen [33] und kleine Zonen der Glutaminsynthese [52]. Die Beschreibung des Ansatzes und erste erzielte Ergebnisse wurden in [13] publiziert und weiterführende Arbeiten auf Tagungen präsentiert [14, 11, 12]. Die bisherigen Ergebnisse lassen ein hohes Potential für weitere Arbeiten erkennen, welche sich z.B. in 3D-Simulationen der Leber einordnen könnten.

# Anhang

## A.1 Anhang zum Kapitel 4

### Parameter zur Fehlerabschätzung

$i$	$\tau_{nonc=i}$	$c_{nonc=i}$
1	0,056350832689630	0,0939585033519458
2	0,306350832689630	0,0304818132989253
3	0,693649167310370	-0,0304818132989254

Tabelle A.1: Parameter zur Fehlerabschätzung ( $i$ 'ten Nicht-Kollokationspunkt)

$j$	$dl_{j,nonc=1}$	$dl_{j,nonc=2}$	$dl_{j,nonc=3}$
0	-8,80947501821704	0,749999999558192	0,749999999558193
1	8,77906529081166	-3,69111181539446	-1,10912291764631
2	0,0423388843878621	3,08198889758751	-2,08198889790881
3	0,0119291569824848	-0,140877081751235	2,44111181599692

$j$	$l_{j,nonc=1}$	$l_{j,nonc=2}$	$l_{j,nonc=3}$
0	0,415473751041354	-0,435710626316106	0,435710626316105
1	0,614415278788021	1,01934218859213	-0,769342188565354
2	-0,0351813904067050	0,459526249018040	1,04047375092841
3	0,00529236057732955	-0,0431578112940637	0,293157811320839

Tabelle A.2: Parameter zur Interpolation ( $j$ 'ten Kollokationspunkt)

### Modell des Fermentierungsprozesses

Die mathematische Beschreibung der Zustandsvariablen sind in Gleichung (4.20) dargestellt, hierbei gelten die Anfangsbedingungen  $g(0) = 70\text{mol}/\text{m}^3$ ,  $m(0) = 220\text{mol}/\text{m}^3$ ;  $n(0) = 40\text{mol}/\text{m}^3$  und  $temp(0) = 8^\circ\text{C}$  und die folgende Beschreibung der Symbole:

- Zustände  $g(t)$ : Glukosekonzentration,  $m(t)$ : Maltosekonzentration,  $n(t)$ : Maltotriosekonzentration,  $xb(t)$ : Biomassekonzentration,  $e(t)$ : Ethanolkonzentration in  $\text{mol}/\text{m}^3$  und  $temp(t)$ : Temperatur in  $^\circ\text{C}$ ,

- Parameter  $\mu_i$ : spezifische Reaktionsrate in  $h^{-1}$ ,  $C_p$ : Wärmekapazität in  $kJ/kg/^\circ C$ ,  $\rho$ : Dichte in  $kg/m^3$ ,  $\Delta H_{Fi}$ : Reaktionsenthalpie in  $kJ/mol$ ,  $TEMP_C$ : Temperatur des Kühlmittels (hier = 0  $^\circ C$ ) und  $R_{Xi}$ ,  $R_{Ei}$ : stöchiometrische Koeffizienten.

Die Reaktionsraten werden durch Michaelis-Menten Kinetiken beschrieben nach

$$\mu_1(g(t)) = \frac{V_g g(t)}{K_g + g(t)}$$

$$\mu_2(m(t), g(t)) = \frac{V_m m(t)}{K_m + m(t)} \frac{K'_g}{K'_g + g(t)}$$

$$\mu_3(n(t), m(t), g(t)) = \frac{V_n n(t)}{K_n + n(t)} \frac{K'_g}{K'_g + g(t)} \frac{K'_m}{K'_m + m(t)}$$

dimensionslos		$\Delta H$ [kJ/mol]	verschieden
$R_{X_G} = 0,134$	$R_{E_G} = 1,92$	$\Delta H_{F_G} = -91,2$	$\rho = 1040 \frac{kg}{m^3}$
$R_{X_M} = 0,268$	$R_{E_M} = 3,84$	$\Delta H_{F_M} = -226,3$	$C_p = 4016 \frac{J}{kg \cdot K}$
$R_{X_N} = 0,402$	$R_{E_N} = 5,76$	$\Delta H_{F_N} = -361,3$	$R = 8,314 \frac{J}{mol \cdot K}$
$E$ [kJ/mol]			
$E_{V_G} = 94,6$	$E_{V_N} = 31,8$	$E_{E_N} = -83,3$	$E_{K'_G} = 42,7$
$E_{V_M} = 47,3$	$E_{K_M} = -43,2$	$E_{E_{K_G}} = -284,5$	$E_{K'_M} = 110$

Tabelle A.3: Modellparameter des Fermentierungsprozesses

Des Weiteren werden die Temperaturabhängigkeiten der maximalen Reaktionsrate  $V$ , der Halbsättigungskonstante  $K$  und die Abschwächung der Reaktionsraten  $K'$  durch die Arrhenius-Beziehung beschrieben:

$$\left. \begin{aligned} V_i(t) &= V_{i0} \exp[-E_{V_i} / (R(\text{temp}(t) + 273.15))] \\ K_i(t) &= k_{i0} \exp[-E_{K_i} / (R(\text{temp}(t) + 273.15))] \\ K'_i(t) &= k'_{i0} \exp[-E_{K'_i} / (R(\text{temp}(t) + 273.15))] \end{aligned} \right\} i = g, m, n$$

mit

- $V_{i0}$ : Arrhenius-Faktoren der Maximalen Reaktionsraten in  $h^{-1}$ ,
- $k_{i0}, k'_{i0}$ : Arrhenius Faktoren der Michaelis-Menten-Konstanten in  $mol/m^3$ ,
- $E_{V_i}$ : Arrhenius-Aktivierungsenergie der maximalen Reaktionsraten in  $kJ/mol$ ,
- $E_{K_i}, E_{K'_i}$ : Arrhenius-Aktivierungsenergie der Michaelis-Menten-Konstanten in  $kJ/mol$  und

- $R$ : Gaskonstante in  $J/mol/K$ .

Die konkreten Werte der Modellparameter sind in Tabelle A.3 zu finden. Für Details der Modellanpassung sei hier auf [53] verwiesen.

### Optimale Zustände

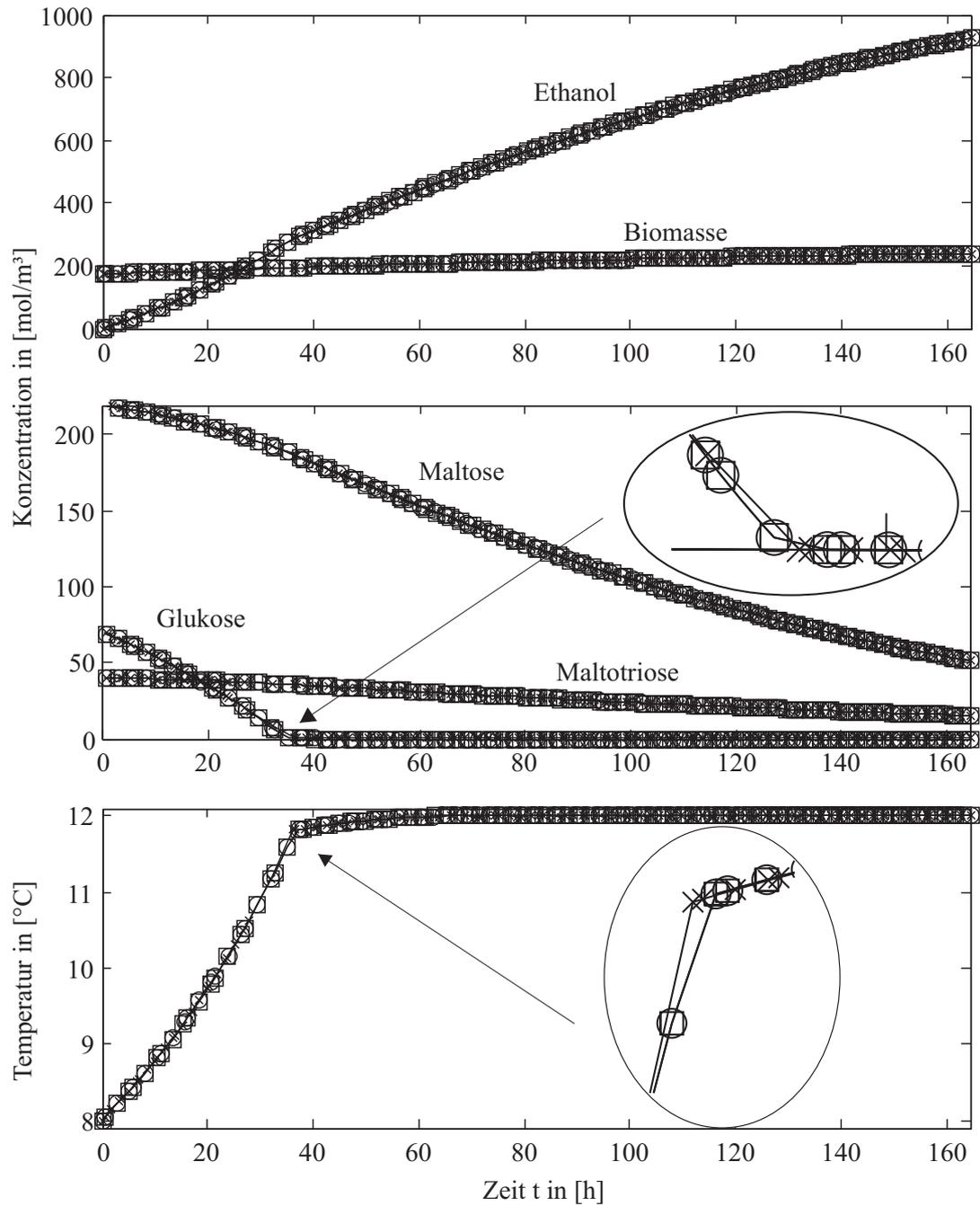


Abbildung A.1: Optimale Zustände (ohne Approximationskontrolle „x“, Variante 1 „o“, 2 „□“)

## A.2 Anhang zum Kapitel 5

### Varianz in der Genexpression

Hier wird eine Auswertung gezeigt, welche die Veränderung der Genexpression für die Teilsysteme aus Kapitel 5.1 untersucht. Dabei wird für jedes Gen die Varianz in den Expressionswerten der „microarray experiments“ aus  $M^{3D}$  [43] berechnet. Die Ergebnisse werden in Abbildung A.2 auf die Teilsysteme bezogen dargestellt (Durchschnittswerte der beteiligten Gene) und die Teilsysteme mit geringer Koordination (dünn verteilte, transkriptionelle Regulation) durch Fettschrift hervorgehoben. Die Teilsysteme mit der dünn verteilten, transkriptionellen Regulation haben hierbei geringe Varianzen in den Genexpressionen. Dies weist auf eine konstitutive Expression vieler beteiligter Gene hin.

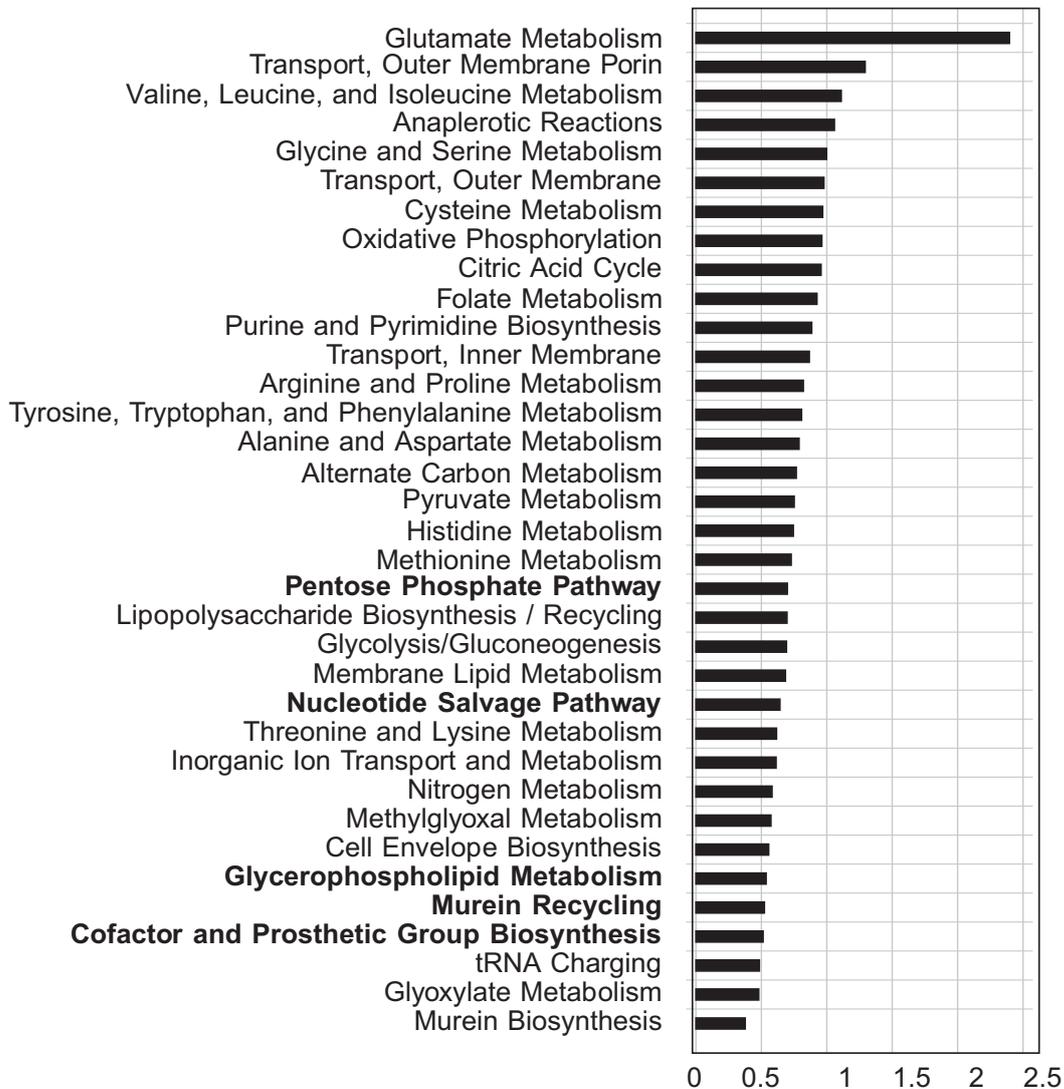


Abbildung A.2: Teilsystembezogene Varianz in der Genexpression

## A.3 Anhang zum Kapitel 6

### Zeitoptimale Aktivierung für kleine und große Enzymsyntheseraten

In Abbildung A.3 werden die zeitoptimalen Aktivierungsstrategien für die Verhältnisse von  $d_{i,max} \ll d_{max}$  und  $d_{i,max} \gg d_{max}$  der einzelnen Enzymsyntheseraten und der freien Proteinsynthesekapazität dargestellt. Es ergibt sich für  $d_{i,max} \ll d_{max}$  eine gleichzeitige Enzymsynthese und für  $d_{i,max} \gg d_{max}$  eine entsprechende sequentielle Synthese der Enzyme. Es kann davon ausgegangen werden, dass dieses prinzipielle Verhalten auch für andere Wertepaare mit den Bedingungen  $d_{i,max} \ll d_{max}$  und  $d_{i,max} \gg d_{max}$  bestehen bleibt.

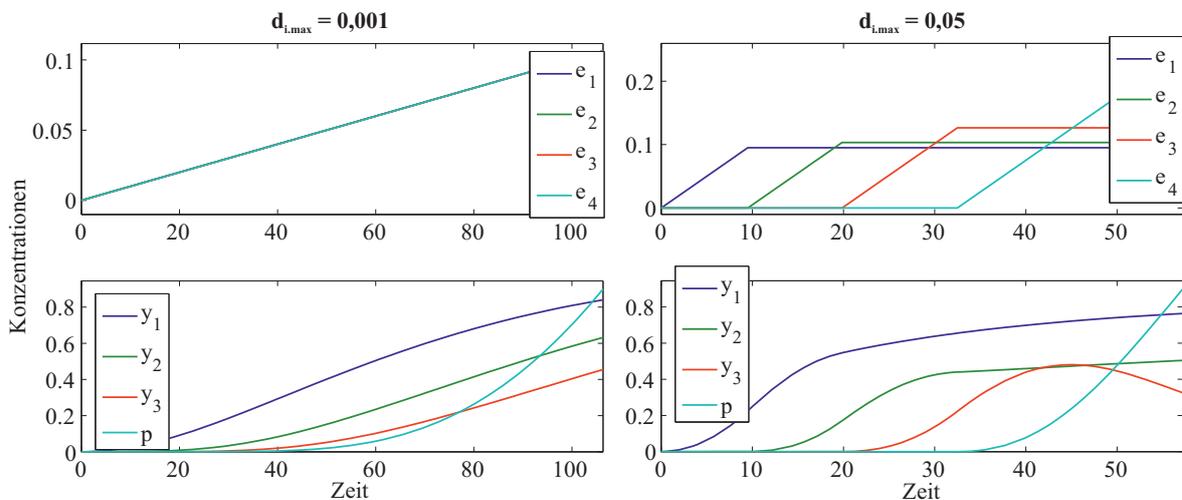


Abbildung A.3: Zeitoptimale Aktivierung für das Zusammenspiel beider Beschränkungen ( $d_{i,max} = 0,001$  und  $0,05$  mit  $d_{max} = 0,01$ )

### Wachstumsoptimale Aktivierung im vollständigen Zeitverlauf

Die vollständigen Zeitverläufe der wachstumsoptimalen Aktivierungsstrategien sind in Abbildung A.4 für verschiedene Enzymsyntheseraten dargestellt. Diese entsprechen den optimalen Steuerungen für das Modell aus Kapitel 6.2 mit Einheitsparametern. Die berechneten Aktivierungsstrategien können in Phasen eingeteilt werden, beginnend mit einer Aktivierungsphase, gefolgt von konstantem Wachstum. Die Phase des konstanten Wachstums entspricht einem optimalen Betriebspunkt für den aktiven Zustand des Netzwerkes. Die Betrachtung einer finiten Zeitspanne (wenn auch einer langen  $t_f = 1000$ ) resultiert in einer für diese Art optimalsteuerungsproblem-typischen Ausnutzung der Ressourcen im letzten Zeitabschnitt (eine Art Artefakt der technischen Realisierung). Somit wird in der vorliegenden Arbeit dieser Zeitabschnitt keiner biologischen Deutung unterzogen und entsprechend in den angestrebten Analysen nicht mit einbezogen.

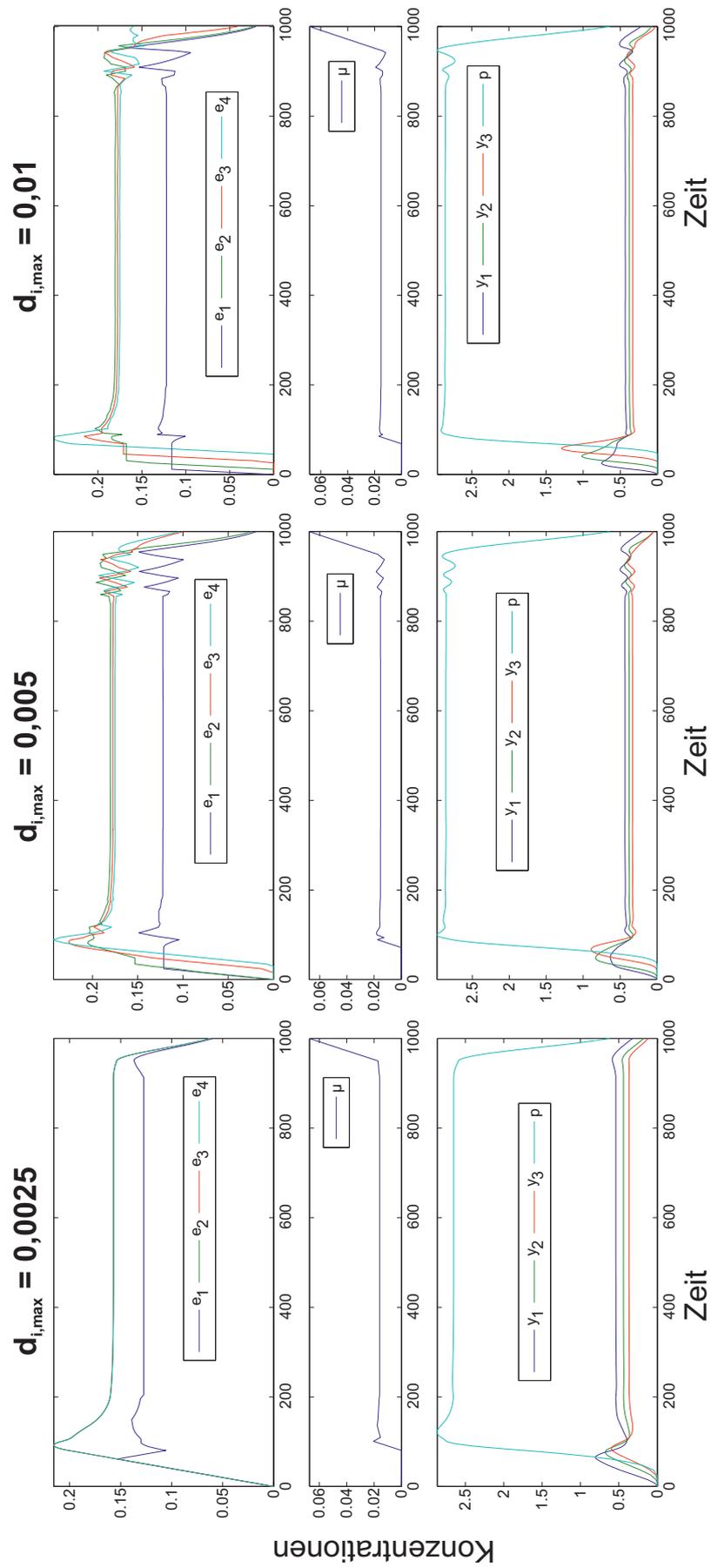


Abbildung A.4: Wachstumsoptimale Aktivierung im vollständigem Zeitverlauf

### **Ermittlung der Aktivierungszeitpunkte und der Gene pro Operon**

Der Aktivierungszeitpunkt  $t_{aktive,i}$  wird durch die Bedingung  $e_i(t_{aktive,i}) \geq d_{max,i}$  ermittelt, wobei dies der Konzentration des Enzyms  $i$  nach einer Investition der maximalen Enzymsyntheserate für mindestens eine Zeiteinheit entspricht. Die Ermittlung der Gene pro Operon basiert auf der Gruppierung von Enzymen in einem Operon, wenn deren Aktivierungszeitpunkt nicht mehr als eine Zeiteinheit auseinander liegt. Die Aktivierungszeitpunkte  $t_{aktive,i}$  der Enzyme definieren sich wie oben beschrieben.

### **Ermittlung der relativen Enzymabundanzen**

Die relative Abundanzen entsprechen hier den Konzentrationen nach der Netzwerkaktivierung der wachstumsoptimalen Strategien und anschließender Normierung durch den Mittelwert der Enzyme der einzelnen Netzwerke für die jeweils zufälligen Parametersets. Die bestimmten relativen Abundanzen sind entsprechend auf das jeweilige Netzwerk bezogen und ergeben eine Normierung auf den Mittelwert 1. Es ergibt sich somit für ein hoch abundantes Enzym (im Vergleich zu den anderen Enzymen des selben Netzwerkes) eine relative Abundanz von größer 1 und für ein gering abundantes Enzym (im Vergleich zu den anderen Enzymen des selben Netzwerkes) eine relative Abundanz von kleiner 1.

### **Bestimmung der Position in der Aktivierungssequenz**

Die Position in der Aktivierungssequenz wird bestimmt durch eine Sortierung der Aktivierungszeitpunkte der Enzyme in steigender Reihenfolge und einer anschließenden Ermittlung der entsprechenden Position eines Enzyms in dieser sortierten Liste. Die Aktivierungszeitpunkte  $t_{aktive,i}$  der Enzyme definiert sich wie oben beschrieben.

### **Einfluss der Enzymabundanzen auf die Position in der Aktivierungssequenz**

Entsprechend zur Ergänzung von Abbildung 6.7 wird der Einfluss der relativen Enzymabundanzen auf die Reihenfolge in der Aktivierungssequenz für die zufälligen 100 Parametersets der verschiedenen einzelnen Enzymsyntheseraten dargestellt (siehe Abbildung A.5). Die relativen Enzymabundanzen werden wie oben beschrieben bestimmt. Daraus wird ersichtlich, dass sich prinzipiell ähnliche Aussagen, wie für Abbildung 6.7 ableiten lassen. Es ergibt sich eine vorgezogene Aktivierung von Enzymen mit einer hohen relativen Abundanz im Vergleich zu Enzymen mit durchschnittlicher Abundanz und Enzyme mit einer niedrigen relativen Abundanz werden tendenziell später aktiviert als die umgebenen Enzyme.

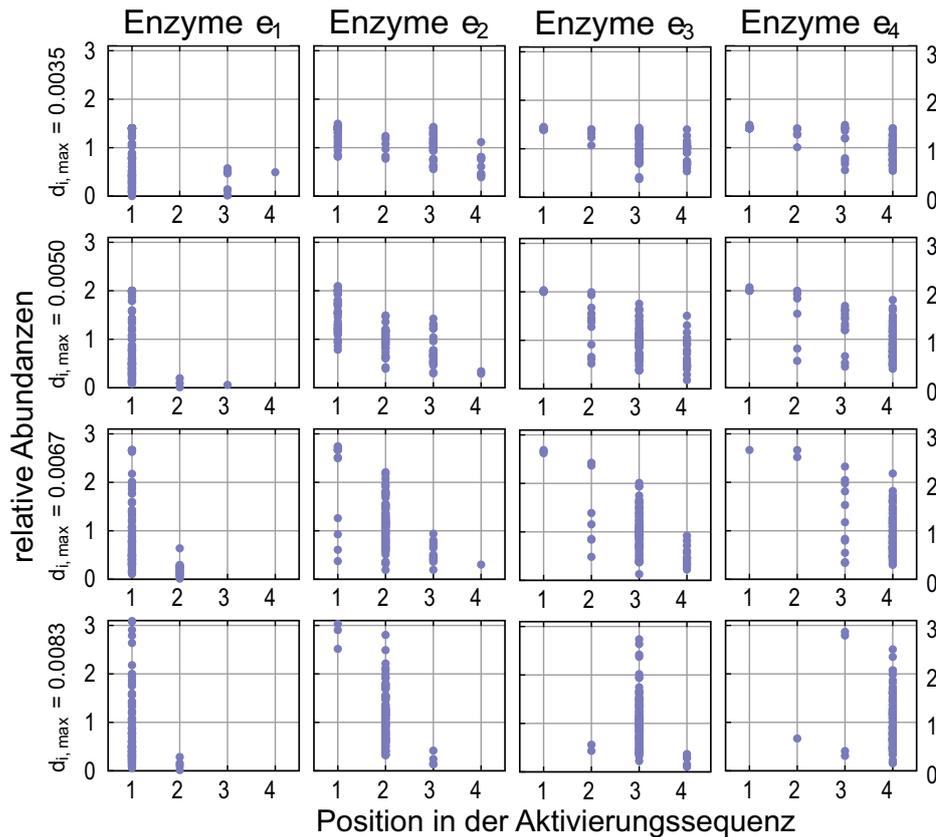


Abbildung A.5: Einfluss der Enzymabundanz auf die Position in der Aktivierungssequenz für verschiedene einzelne Enzymsyntheseraten

### Positionsabhängige Verschiebung durch Koexpression

Die relativen Enzymabundanz werden wie oben beschrieben bestimmt und in hoch abundante Enzyme mit einer relativen Abundanz von größer als  $1 * 1.1$  und in gering abundante Enzyme mit einer relativen Abundanz von kleiner als  $1 : 1.1$  unterschieden. Die positionsabhängige Verschiebung durch Koexpression in einem Operon entspricht der Distanz von der Position in der Aktivierungssequenz zur Position im Stoffwechselnetzwerk (siehe auch oben für die Position in der Aktivierungssequenz). Befindet sich z.B. ein hoch abundantes Enzym e<sub>4</sub> (dies katalysiert entsprechend die 4. Position im Stoffwechselnetzwerk) in einem Operon mit Enzym e<sub>1</sub> (dies katalysiert entsprechend die 1. Position im Stoffwechselnetzwerk) ergibt sich eine positionsabhängige Verschiebung durch Koexpression von -3. Befindet sich z.B. ein gering abundantes Enzym e<sub>1</sub> in einem Operon mit Enzym e<sub>4</sub> ergibt sich +3.

Um die positionsabhängige Verschiebung durch Koexpression mit einer zufälligen Operonorganisation zu vergleichen, werden zufällige Zugehörigkeiten von Enzymen zu Operons erzeugt. Die dargestellten Häufigkeitsverteilungen entsprechen den Operonorganisationen aus jeweils 100 unabhängigen Zufallsexperimenten.

# Literaturverzeichnis

- [1] B. Alberts, A. Johnson, J. Lewis, M. Raff, K. Roberts, and P. Walter. *Molecular Biology of the Cell*. Wiley-VCH, 2002.
- [2] B. Alberts, A. Johnson, J. Lewis, M. Raff, K. Roberts, and P. Walter. *Molekularbiologie der Zelle*. Wiley-VCH, Übersetzt von Jaenicke, L. und Arndt, O. et. al - Weinheim, 2004.
- [3] U. Alon. *An Introduction to Systems Biology: Design Principles of Biological Circuits*. Chapman & Hall/CRC, Boca Raton, FL, 2007.
- [4] D. H. Alpers and G. M. Tomkins. The order of induction and deinduction of the enzymes of the lactose operon in *e. coli*. *Proc Natl Acad Sci USA*, 53:797–802, 1965.
- [5] R. Alves and M. A. Savageau. Evidence of selection for low cognate amino acid bias in amino acid biosynthetic enzymes. *Mol Microbiol*, 56(4):1017–34, 2005.
- [6] U. Ascher. Collocation for two-point boundary value problems revisited. *SIAM J Numer Anal*, 23(3):596–609, 1986.
- [7] M. M. Babu and L. Aravind. Adaptive evolution by optimizing expression levels in different environments. *Trends Microbiol*, 14(1):11–4, 2006.
- [8] M. Bartl, M. Kötzing, S. Schuster, P. Li, and C. Kaleta. Dynamic optimization identifies optimal programs for pathway regulation in prokaryotes. *Nat Commun*, 4, 2013.
- [9] M. Bartl, P. Li, and L. Biegler. Improvement of state profile accuracy in nonlinear dynamic optimization with the quasi-sequential approach. *AIChE J*, 57(8):2185–2197, 2011.
- [10] M. Bartl, P. Li, and S. Schuster. Modelling the optimal timing in metabolic pathway activation—use of pontryagin’s maximum principle and role of the golden section. *Bio-systems*, 101(1):67–77, 2010.

- [11] M. Bartl, M. Pfaff, D. Driesch, S. Zellmer, S. Schuster, R. Gebhardt, and P. Li. Using metabolic modelling and optimization methods in organ-oriented systems biology: Prediction of adaptive liver zonation during regeneration. In *German Conference on Bioinformatics (GCB)*, Jena, Germany, 2012.
- [12] M. Bartl, M. Pfaff, D. Driesch, S. Zellmer, S. Schuster, R. Gebhardt, and P. Li. Using metabolic modelling and optimization methods in organ-oriented systems biology: Prediction of liver zonation under different metabolic conditions. In *Workshop Organ-oriented Systems Biology 2012*, Jena, Germany, 2012.
- [13] M. Bartl, M. Pfaff, S. Toepfer, S. Zellmer, R. Gebhardt, S. Schuster, and P. Li. Model-based optimization to explain liver zonation. In P. Scharff, editor, *Proc. 55th International Scientific Colloquium*, Ilmenau, Germany, 2010.
- [14] M. Bartl, M. Pfaff, S. Töpfer, S. Zellmer, S. Schuster, R. Gebhardt, and P. Li. A model-based optimization approach to describe and predict liver zonation. In *The 12th International Conference on Systems Biology (ICSB)*, Heidelberg, Germany, 2011.
- [15] H. J. Beaumont, J. Gallie, C. Kost, G. C. Ferguson, and P. B. Rainey. Experimental evolution of bet hedging. *Nature*, 462(7269):90–3, 2009.
- [16] R. Bellman. *Dynamic Programming*. Princeton University Press, New Jersey, 1957.
- [17] Y. Benjamini and D. Yekutieli. The control of the false discovery rate in multiple testing under dependency. *Annals of statistics*, pages 1165–1188, 2001.
- [18] K. Bettenbrock, S. Fischer, A. Kremling, K. Jahreis, T. Sauter, and E. D. Gilles. A quantitative approach to catabolite repression in escherichia coli. *J Biol Chem*, 281(5):2578–84, 2006.
- [19] L. T. Biegler. An overview of simultaneous strategies for dynamic optimization. *Chem Eng Process*, 46(11):1043–1053, 2007.
- [20] L. T. Biegler, A. M. Cervantes, and A. Wachter. Advances in simultaneous strategies for dynamic process optimization. *Chem Eng Sci*, 57(4):575–593, 2002.
- [21] H. Bremer and P. P. Dennis. Modulation of chemical composition and other parameters of the cell by growth rate. In F. Neidhardt, editor, *Escherichia coli and Salmonella*, pages 1553–1569. American Society for Microbiology, Washington, DC, Georgetown, 2nd edition, 1996.

- [22] G. E. Briggs and J. B. S. Haldane. A note on the kinetics of enzyme action. *Biochem. J.*, 19:338–339, 1925.
- [23] M. Caldara, G. Dupont, F. Leroy, A. Goldbeter, L. De Vuyst, and R. Cunin. Arginine biosynthesis in escherichia coli - experimental perturbation and mathematical modeling. *J Biol Chem*, 283(10):6347–6358, 2008.
- [24] S. L. Campbell and C. W. Gear. The index of general nonlinear daes. *Numer. Math.*, 72(2):173–196, 1995.
- [25] C. Canuto, M. Y. Hussaini, A. Quarteroni, and T. A. Zang. *Spectral Methods in Fluid Dynamics*. Springer-Verlag, Heidelberg, 1988.
- [26] S. Ceric and Z. Kurtanjek. Model identification, parameter estimation, and dynamic flux analysis of e-coli central metabolism. *Chem Biochem Eng Quarterly*, 20(3):243–253, 2006.
- [27] C. Chassagnole, D. A. Fell, B. Rais, B. Kudla, and J. P. Mazat. Control of the threonine-synthesis pathway in escherichia coli: a theoretical and experimental approach. *Biochem J*, 356:433–444, 2001.
- [28] C. Chassagnole, N. Noisommit-Rizzi, J. W. Schmid, K. Mauch, and M. Reuss. Dynamic modeling of the central carbon metabolism of escherichia coli. *Biotechnol Bioeng*, 79(1):53–73, 2002.
- [29] G. Chechik, E. Oh, O. Rando, J. Weissman, A. Regev, and D. Koller. Activity motifs reveal principles of timing in transcriptional control of the yeast metabolic network. *Nat Biotechnol*, 26(11):1251–1259, 2008.
- [30] T. Chen, H. L. He, and G. M. Church. Modeling gene expression with differential equations. *Pac Symp Biocomput*, pages 29–40, 1999.
- [31] C. S. Chin, V. Chubukov, E. R. Jolly, J. DeRisi, and H. Li. Dynamics and design principles of a basic regulatory architecture controlling metabolic pathways. *Plos Biology*, 6(6):1343–1356, 2008.
- [32] J. Christiansen and R. D. Russell. Error analysis for spline collocation methods with application to knot selection. *Math Comput*, 32(142):415–419, 1978.
- [33] J. F. Comar, F. Suzuki-Kemmelmeier, J. Constantin, and A. Bracht. Hepatic zonation of carbon and nitrogen fluxes derived from glutamine and ammonia transformations.

- J. Biomed. Sci.*, 17(doi: 10.1186/1423-0127-17-1), 2010. Comar, Jurandir F. Suzuki-Kemmelmeier, Fumie Constantin, Jorgete Bracht, Adelar.
- [34] B. A. Conway. A survey of methods available for the numerical optimization of continuous dynamic systems. *J Optim Theory Appl*, 152(2):271–306, 2012.
- [35] A. Cornish-Bowden. *The Pursuit of Perfection: Aspects of Biochemical Evolution*. Oxford University Press, Oxford, 2004.
- [36] J. E. Cuthrell and L. T. Biegler. On the optimization of differential-algebraic process systems. *AIChE J*, 33(8):1257–1270, 1987.
- [37] C. De Boor. Good approximation by splines with variable knots. ii. *Lect Notes Math*, 363:12–20, 1974.
- [38] C. De Boor. A practical guide to splines. *Appl Math Sci*, 27, 1978.
- [39] C. De Boor and B. Swartz. Collocation at gaussian points. *SIAM J Numer Anal*, 10(4):582–606, 1973.
- [40] D. Degenring. *Erstellung und Validierung mechanistischer Modelle für den mikrobiellen Stoffwechsel ur Auswertung von Substrat-Puls-Experimenten*. PhD thesis, University of Rostock, 2004.
- [41] P. S. Dehal, M. P. Joachimiak, M. N. Price, J. T. Bates, J. K. Baumohl, D. Chivian, G. D. Friedland, K. H. Huang, K. Keller, P. S. Novichkov, I. L. Dubchak, E. J. Alm, and A. P. Arkin. Microbesonline: an integrated portal for comparative and functional genomics. *Nucleic Acids Res*, 38(Database issue):D396–400, 2010.
- [42] E. Dekel and U. Alon. Optimality and evolutionary tuning of the expression level of a protein. *Nature*, 436(7050):588–92, 2005.
- [43] J. J. Faith, M. E. Driscoll, V. A. Fusaro, E. J. Cosgrove, B. Hayete, F. S. Juhn, S. J. Schneider, and T. S. Gardner. Many microbe microarrays database: uniformly normalized affymetrix compendia with structured experimental metadata. *Nucleic Acids Res*, 36(Database issue):D866–70, 2008.
- [44] A. M. Feist, C. S. Henry, J. L. Reed, M. Krummenacker, A. R. Joyce, P. D. Karp, L. J. Broadbelt, V. Hatzimanikatis, and B. O. Palsson. A genome-scale metabolic reconstruction for escherichia coli k-12 mg1655 that accounts for 1260 orfs and thermodynamic information. *Mol Syst Biol*, 3:121, 2007.

- [45] F. Felgner, L. Liu, and G. Frey. Vergleich numerischer löser zur simulation steifer und hybrider systeme. In *Proceedings of the Kongress Automation 2011, VDI-Berichte 2143*, pages 303–306, (extended 12-pages paper on CD), Baden-Baden, Germany, 2011.
- [46] D. J. Ferullo and S. T. Lovett. The stringent response and cell cycle arrest in escherichia coli. *PLoS Genet*, 4(12):e1000300, 2008.
- [47] B. Finlayson. *Nonlinear analysis in chemical engineering*. McGraw-Hill, New York, 1980.
- [48] T. Flemming, M. Bartl, and P. Li. Set-point optimization for closed-loop control systems under uncertainty. *Ind Eng Chem Res*, 46(14):4930–4942, 2007.
- [49] S. S. Fong, A. P. Burgard, C. D. Herring, E. M. Knight, F. R. Blattner, C. D. Maranas, and B. O. Palsson. In silico design and adaptive evolution of escherichia coli for production of lactic acid. *Biotechnol Bioeng*, 91(5):643–8, 2005.
- [50] S. S. Fong, A. Nanchen, B. . Palsson, and U. Sauer. Latent pathway activation and increased pathway capacity enable escherichia coli adaptation to loss of key metabolic enzymes. *J Biol Chem*, 281:8024–8033, 2006.
- [51] S. Gama-Castro, H. Salgado, M. Peralta-Gil, A. Santos-Zavaleta, L. Muniz-Rascado, H. Solano-Lira, V. Jimenez-Jacinto, V. Weiss, J. S. Garcia-Sotelo, A. Lopez-Fuentes, L. Porron-Sotelo, S. Alquicira-Hernandez, A. Medina-Rivera, I. Martinez-Flores, K. Alquicira-Hernandez, R. Martinez-Adame, C. Bonavides-Martinez, J. Miranda-Rios, A. M. Huerta, A. Mendoza-Vargas, L. Collado-Torres, B. Taboada, L. Vega-Alvarado, M. Olvera, L. Olvera, R. Grande, E. Morett, and J. Collado-Vides. Regulondb version 7.0: transcriptional regulation of escherichia coli k-12 integrated within genetic sensory response units (gensor units). *Nucleic Acids Res*, 39:D98–D105, 2011.
- [52] R. Gebhardt and D. Mecke. Heterogeneous distribution of glutamine synthetase among rat liver parenchymal cells in situ and in primary culture. *EMBO J.*, 2(4):567–570, 1983. Embo J. Article ISI Document Delivery No.: QL752 Times Cited: 317 Cited Reference Count: 24 OXFORD UNIV PRESS OXFORD.
- [53] D. A. Gee and W. F. Ramirez. Optimal temperature control for batch beer fermentation. *Biotechnol Bioeng*, 31(3):224–234, 1988.
- [54] N. Geisel. Constitutive versus responsive gene expression strategies for growth in changing environments. *PLoS One*, 6(11):e27033, 2011.

- [55] N. Geisel, J. M. Vilar, and J. M. Rubi. Optimal resting-growth strategies of microbial populations in fluctuating environments. *PLoS One*, 6(4):e18622, 2011.
- [56] A. Geletu, M. Klöppel, H. Zhang, and P. Li. Advances and applications of chance-constrained approaches to systems optimisation under uncertainty. *Int J Syst Sci*, 44(7):1209–1232, 2013.
- [57] L. Gerosa, K. Kochanowski, M. Heinemann, and U. Sauer. Dissecting specific and global transcriptional regulation of bacterial gene expression. *Mol. Sys. Biol.*, 9:658, 2013.
- [58] P. E. Gill, W. Murray, and M. A. Saunders. Snopt: An sqp algorithm for large-scale constrained optimization. *SIAM J Optim*, 12(4):979–1006, 2002.
- [59] D. T. Gillespie. A general method for numerically simulating the stochastic time evolution of coupled chemical reactions. *J Comp Phys*, 22(4):403–434, 1976.
- [60] D. T. Gillespie. Exact stochastic simulation of coupled chemical reactions. *J Phys Chem*, 173(MAR20):128–128, 1977.
- [61] D. T. Gillespie. Approximate accelerated stochastic simulation of chemically reacting systems. *J Chem Phys*, 115(4):1716–1733, 2001.
- [62] F. Gnad, S. Ren, J. Cox, J. Olsen, B. Macek, M. Oroshi, and M. Mann. Phosida (phosphorylation site database): management, structural and evolutionary investigation, and prediction of phosphosites. *Genome Biology*, 8(11):R250, 2007.
- [63] M. Heinemann and U. Sauer. Systems biology of microbial metabolism. *Current Opinion in Microbiology*, 13(3):337 – 343, 2010. Ecology and industrial microbiology, Special section: Systems biology.
- [64] R. Heinrich and S. Schuster. Is metabolic channelling the complicated solution to the easy problem of reducing transient times? *J Theor Biol*, 152(1):57–61, 1991.
- [65] R. Heinrich and S. Schuster. *The Regulation of Cellular Systems*. Chapman & Hall, New York, 1996.
- [66] R. Heinrich, S. Schuster, and H. G. Holzhütter. Mathematical analysis of enzymic reaction systems using optimization principles. *Eur J Biochem*, 201(1):1–21, 1991.
- [67] W. Hong, S. Wang, P. Li, G. Wozny, and L. T. Biegler. A quasi-sequential approach to large-scale dynamic optimization problems. *AIChE J*, 52(1):255–268, 2006.

- [68] G. Huntington. *Advancement and analysis of a gauss pseudospectral transcription for optimal control*. Ph.D. dissertation. Department of Aeronautics and Astronautics, Massachusetts Institute of Technology, Cambridge, Massachusetts, 2007.
- [69] G. T. Huntington and A. V. Rao. Comparison of global and local collocation methods for optimal control. *J Guid Contr Dynam*, 31(2):432–436, 2008.
- [70] R. U. Ibarra, J. S. Edwards, and B. O. Palsson. Escherichia coli k-12 undergoes adaptive evolution to achieve in silico predicted optimal growth. *Nature*, 420(6912):186–9, 2002.
- [71] J. Ihmels, R. Levy, and N. Barkai. Principles of transcriptional control in the metabolic network of saccharomyces cerevisiae. *Nat Biotechnol*, 22(1):86–92, 2004.
- [72] T. Itoh, K. Takemoto, H. Mori, and T. Gojobori. Evolutionary instability of operon structures disclosed by sequence comparisons of complete microbial genomes. *Mol Biol Evol*, 16(3):332–46, 1999.
- [73] F. Jacob and J. Monod. Genetic regulatory mechanisms in the synthesis of proteins. *J Mol Biol*, 3(3):318–356, 1961.
- [74] H. Kacser and R. Beeby. Evolution of catalytic proteins. on the origin of enzyme species by means of natural selection. *J Mol Evol*, 20(1):38–51, 1984.
- [75] C. Kaleta, L. F. de Figueiredo, and S. Schuster. Can the whole be less than the sum of its parts? pathway analysis in genome-scale metabolic networks using elementary flux patterns. *Genome Res*, 19(10):1872–83, 2009.
- [76] M. Kanehisa and S. Goto. Kegg: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Res*, 28(1):27–30, 2000.
- [77] I. M. Keseler, J. Collado-Vides, A. Santos-Zavaleta, M. Peralta-Gil, S. Gama-Castro, L. Muñiz-Rascado, C. Bonavides-Martinez, S. Paley, M. Krummenacker, T. Altman, P. Kaipa, A. Spaulding, J. Pacheco, M. Latendresse, C. Fulcher, M. Sarker, A. G. Shearer, A. Mackie, I. Paulsen, R. P. Gunsalus, and P. D. Karp. Ecocyc: a comprehensive database of escherichia coli biology. *Nucleic Acids Res*, 39(suppl 1):D583–D590, 2011.
- [78] H. Kitano. *Foundations of Systems Biology*. MIT Press, Cambridge, MA, 2001.
- [79] H. Kitano. Systems biology: a brief overview. *Science*, 295(5560):1662–1664, 2002.

- [80] E. Klipp, R. Heinrich, and H. G. Holzhütter. Prediction of temporal gene expression - metabolic optimization by re-distribution of enzyme activities. *Eur J Biochem*, 269(22):5406–5413, 2002.
- [81] E. Klipp, W. Liebermeister, C. Wierling, A. Kowald, H. Lehrach, and R. Herwig. *Systems Biology. A Textbook*. Wiley-Blackwell, Weinheim, 2009.
- [82] E. Klipp, B. Nordlander, R. Kruger, P. Gennemark, and S. Hohmann. Integrative model of the response of yeast to osmotic shock. *Nat Biotechnol*, 23(8):975–982, 2005.
- [83] S. Klumpp, Z. Zhang, and T. Hwa. Growth rate-dependent global effects on gene expression in bacteria. *Cell*, 139(7):1366–1375, 2009.
- [84] O. Kotte, J. B. Zaugg, and M. Heinemann. Bacterial adaptation through distributed sensing of metabolic fluxes. *Mol Syst Biol*, 6:355, 2010.
- [85] K. Kovacs, L. D. Hurst, and B. Papp. Stochasticity in protein levels drives colinearity of gene order in metabolic operons of escherichia coli. *PLoS Biol*, 7(5):e1000115, 2009.
- [86] A. Kremling. *Kompendium Systembiologie - Mathematische Modellierung und Modellanalyse*. Springer-Vieweg, München, 2012.
- [87] M. Kötzing. *Diplomarbeit: Untersuchung von Optimalitätskriterien in transkriptionellen Regulationsvorgängen des zentralen Stoffwechsels von Escherichia coli*. Technische Universität Ilmenau, Ilmenau, Germany, 2010.
- [88] Y. D. Lang and L. T. Biegler. A software environment for simultaneous dynamic optimization. *Comput Chem Eng*, 31(8):931–942, 2007.
- [89] N. Lewis, H. Nagarajan, and B. Palsson. Constraining the metabolic genotype-phenotype relationship using a phylogeny of in silico methods. *Nat Rev Micro*, 10(4):291–305, Apr. 2012.
- [90] P. Li. *Entwicklung optimaler Führungsstrategien für Batch-Destillationsprozesse*. Fortschritt-Berichte, Reihe 3, Nr. 560, VDI-Verlag, Berlin, 1998.
- [91] P. Li, H. A. Garcia, G. Wozny, and E. Reuter. Optimization of a semibatch distillation process with model validation on the industrial site. *Ind Eng Chem Res*, 37(4):1341–1350, 1998.

- [92] P. Li and G. Wozny. Application of a decomposition approach to optimal temperature control of a batch beer fermentation reactor. *European Control Conference, ECC 99*, 1999.
- [93] M. Llorens, J. C. Nuno, Y. Rodriguez, E. Melendez-Hevia, and F. Montero. Generalization of the theory of transition times in metabolic pathways: A geometrical approach. *Biophys J*, 77(1):23–36, 1999.
- [94] W. Lorenz. *Diplomarbeit: Effiziente Implementierung des quasisequentiellen Ansatzes für die Untersuchung optimaler Regulationsvorgänge in Stoffwechselwegen*. Technische Universität Ilmenau, Ilmenau, Germany, 2011.
- [95] P. Lu, C. Vogel, R. Wang, X. Yao, and E. M. Marcotte. Absolute protein expression profiling estimates the relative contributions of transcriptional and translational regulation. *Nat Biotechnol*, 25(1):117–24, 2007.
- [96] R. Mahadevan, J. S. Edwards, and F. J. Doyle. Dynamic flux balance analysis of diauxic growth in escherichia coli. *Biophys J*, 83(3):1331–1340, 2002.
- [97] L. Michaelis and M. L. Menten. Die kinetik der invertinwirkung. *Biochem Z*, 49:333–369, 1913.
- [98] D. Molenaar, R. van Berlo, D. de Ridder, and B. Teusink. Shifts in growth strategies reflect tradeoffs in cellular economics. *Mol Syst Biol*, 5:323, 2009.
- [99] J. Monod and G. Cohen-Bazire. The effect of specific inhibition in biosynthesis of tryptophan-desmase by aerobacter aerogenes. *CR Acad Sci, Paris*, 5:323, 1953.
- [100] A. F. M. Moorman, P. A. J. Deboer, M. Watford, M. A. Dingemanse, and W. H. Lamers. Hepatic glutaminase mrna is confined to part of the urea cycle domain in the adult rodent liver lobule. *FEBS Lett.*, 356(1):76–80, 1994.
- [101] I. E. Nikerel, W. A. van Winden, P. J. T. Verheijen, and J. J. Heijnen. Model reduction and a priori kinetic parameter identifiability analysis using metabolome time series for metabolic reaction networks with linlog kinetics. *Metab Eng*, 11(1):20–30, 2009.
- [102] R. A. Notebaart, B. Teusink, R. J. Siezen, and B. Papp. Co-regulation of metabolic genes is better explained by flux coupling than by network distance. *PLoS Comput Biol*, 4(1):e26, 2008.

- [103] D. Oyarzún, B. Ingalls, R. Middleton, and D. Kalamatianos. Sequential activation of metabolic pathways: a dynamic optimization approach. *Bull Math Biol*, 71(8):1851–1872, 2009.
- [104] D. A. Oyarzun. Optimal control of metabolic networks with saturable enzyme kinetics. *IET Syst Bio*, 5(2):110–119, 2011.
- [105] B. O. Palsson. *Systems Biology. Properties of Reconstructed Networks*. Cambridge Univ. Press, Cambridge [u.a.], 2006.
- [106] M. Papageorgiou. *Optimierung: Statische, dynamische, stochastische Verfahren für die Anwendung*. Oldenbourg Wissenschaftsverlag, 1996.
- [107] B. Papp, B. Teusink, and R. A. Notebaart. A critical view of metabolic network adaptations. *HFSP J*, 3(1):24–35, 2009.
- [108] L. S. Pontryagin, V. G. Boltyanskii, R. V. Gamkrelidze, and E. Mishchenko. *The Mathematical Theory of Optimal Processes*. International Series of Monographs in Pure and Applied Mathematics. Pergamon Press, Oxford London New York, 1964.
- [109] M. N. Price, A. P. Arkin, and E. J. Alm. The life-cycle of operons. *PLoS Genet*, 2(6):e96, 2006.
- [110] A. V. Rao. A survey of numerical methods for optimal control. *Astrodynamics Specialist Conference, AAS Paper*, pages 9–334, 2009.
- [111] M. Rathinam, L. R. Petzold, Y. Cao, and D. T. Gillespie. Stiffness in stochastic chemically reacting systems: The implicit tau-leaping method. *J Chem Phys*, 119(24):12784–12794, 2003.
- [112] J. C. Ray and O. A. Igoshin. Interplay of gene expression noise and ultrasensitive dynamics affects bacterial operon organization. *PLoS Comput Biol*, 8(8):e1002672, 2012.
- [113] J. G. Renfro, A. M. Morshedi, and O. A. Asbjornsen. Simultaneous optimization and solution of systems described by differential algebraic equations. *Comput Chem Eng*, 11(5):503–517, 1987.
- [114] C. A. M. Riascos and J. M. Pinto. Optimal control of bioreactors: a simultaneous approach for complex systems. *Chem Eng J*, 99(1):23–34, 2004.

- [115] M. Rizzi, M. Baltes, U. Theobald, and M. Reuss. In vivo analysis of metabolic dynamics in *saccharomyces cerevisiae*: Ii. mathematical model. *Biotechnol Bioeng*, 55(4):592–608, 1997.
- [116] S. Rump. Intlab - interval laboratory. In T. Csendes, editor, *Developments in Reliable Computing*, pages 224–234. Kluwer Academic Publishers, Dordrecht, 1999.
- [117] R. D. Russell and J. Christiansen. Adaptive mesh selection strategies for solving boundary value problems. *SIAM J Numer Anal*, 15(1):59–80, 1978.
- [118] D. Satory, A. J. Gordon, J. A. Halliday, and C. Herman. Epigenetic switches: can infidelity govern fate in microbes? *Curr Opin Microbiol*, 14(2):212–7, 2011.
- [119] T. Schlitt and A. Brazma. Current approaches to gene regulatory network modelling. *BMC Bioinformatics*, 8 Suppl 6:S9, 2007.
- [120] R. Schuetz, L. Kuepfer, and U. Sauer. Systematic evaluation of objective functions for predicting intracellular fluxes in *escherichia coli*. *Mol Syst Biol*, 3:119, 2007.
- [121] R. Schuetz, N. Zamboni, M. Zampieri, M. Heinemann, and U. Sauer. Multidimensional optimality of microbial metabolism. *Science*, 336(6081):601–4, 2012.
- [122] S. Schuster and R. Heinrich. Time hierarchy in enzymatic reaction chains resulting from optimality principles. *J Theor Biol*, 129(2):189–209, 1987.
- [123] S. Schuster and R. Heinrich. Minimization of intermediate concentrations as a suggested optimality principle for biochemical networks. 1. theoretical-analysis. *J Math Biol*, 29(5):425–442, 1991.
- [124] S. Schuster, T. Pfeiffer, and D. A. Fell. Is maximization of molar yield in metabolic networks favoured by evolution? *J Theor Biol*, 252(3):497–504, 2008.
- [125] M. Scott, C. W. Gunderson, E. M. Mateescu, Z. Zhang, and T. Hwa. Interdependence of cell growth and gene expression: origins and consequences. *Science*, 330(6007):1099–1102, 2010.
- [126] P. M. Sharp and W.-H. Li. The codon adaptation index—a measure of directional synonymous codon usage bias, and its potential applications. *Nucleic Acids Res*, 15(3):1281–1295, 1987.

- [127] P. Smolen, D. A. Baxter, and J. H. Byrne. Modeling transcriptional control in gene networks—methods, recent results, and future directions. *Bull Math Biol*, 62(2):247–92, 2000.
- [128] Software. Adol-c. <https://projects.coin-or.org/ADOL-C/> (Stand: 01. Januar 2013).
- [129] Software. Cmake. <http://www.cmake.org/> (Stand: 01. Januar 2013).
- [130] Software. doxygen. <http://www.doxygen.org/> (Stand: 01. Januar 2013).
- [131] J. Stelling, S. Klamt, K. Bettenbrock, S. Schuster, and E. D. Gilles. Metabolic network structure determines key aspects of functionality and regulation. *Nature*, 420(6912):190–193, 2002.
- [132] D. M. Stoebel, A. M. Dean, and D. E. Dykhuizen. The cost of expression of escherichia coli lac operon proteins is in the process, not in the products. *Genetics*, 178(3):1653–60, 2008.
- [133] J. Tamimi and P. Li. A combined approach to nonlinear model predictive control of fast systems. *Journal of Process Control*, 20(9):1092 – 1102, 2010. <ce:title>ADCHEM 2009 Special Issue</ce:title>.
- [134] P. Tanartkit and L. T. Biegler. Reformulating ill-conditioned differential-algebraic equation optimization problems. *Ind Eng Chem Res*, 35(6):1853–1865, 1996.
- [135] P. Tanartkit and L. T. Biegler. A nested, simultaneous approach for dynamic optimization problems - ii: the outer problem. *Comput Chem Eng*, 21(12):1365–1388, 1997.
- [136] D. Tieu, W. R. Cluett, and A. Penlidis. A comparison of collocation methods for solving dynamic optimization problems. *Comput Chem Eng*, 19(4):375–381, 1995.
- [137] T. Tuller, I. Veksler-Lublinsky, N. Gazit, M. Kupiec, E. Ruppin, and M. Ziv-Ukelson. Composite effects of gene determinants on the translation speed and density of ribosomes. *Genome Biology*, 12(11):R110, 2011.
- [138] J. J. Tyson, A. Csikasz-Nagy, and B. Novak. The dynamics of cell cycle regulation. *Bioessays*, 24(12):1095–109, 2002.
- [139] D. Vallenet, E. Belda, A. Calteau, S. Cruveiller, S. Engelen, A. Lajus, F. Le Fevre, C. Longin, D. Mornico, D. Roche, Z. Rouy, G. Salvignol, C. Scarpelli, A. A. Thil Smith, M. Weiman, and C. Medigue. Microscope - an integrated microbial resource for the

- curation and comparative analysis of genomic and metabolic data. *Nucleic Acids Res*, 41(D1):D636–47, 2013.
- [140] J. van den Brink, A. B. Canelas, W. M. van Gulik, J. T. Pronk, J. J. Heijnen, J. H. de Winde, and P. Daran-Lapujade. Dynamics of glycolytic regulation during adaptation of *saccharomyces cerevisiae* to fermentative metabolism. *Appl Environ Microb*, 74(18):5710–5723, 2008.
- [141] J. van Schijndel and E. Pistikopoulos. Towards the integration of process design, process control and process operability - current status and future trends. *Foundations of Computer Aided Process Design 99, AIChE Symp Ser*, 323:99–112, 2000.
- [142] S. Vasantharajan and L. T. Biegler. Simultaneous strategies for optimization of differential-algebraic systems with enforcement of error criteria. *Comput Chem Eng*, 14(10):1083–1100, 1990.
- [143] S. Vaseghi, A. Baumeister, M. Rizzi, and M. Reuss. In vivo dynamics of the pentose phosphate pathway in *saccharomyces cerevisiae*. *Metab Eng*, 1(2):128–140, 1999.
- [144] S. Vieira-Silva and E. P. Rocha. The systemic imprint of growth and its uses in ecological (meta)genomics. *PLoS Genet*, 6(1):e1000808, 2010.
- [145] J. Villadsen and W. E. Stewart. Solution of boundary-value problems by orthogonal collocation. *Chem Eng Sci*, 22(11):1483–1502, 1967.
- [146] D. Visser, J. W. Schmid, K. Mauch, M. Reuss, and J. J. Heijnen. Optimal re-design of primary metabolism in *escherichia coli* using linlog kinetics. *Metab Eng*, 6(4):378–390, 2004.
- [147] T. T. Vu and J. Vohradsky. Nonlinear differential equation model for quantification of transcriptional regulation applied to microarray data of *saccharomyces cerevisiae*. *Nucleic Acids Res*, 35(1):279–287, 2007.
- [148] T. T. Vu and J. Vohradsky. Inference of active transcriptional networks by integration of gene expression kinetics modeling and multisource data. *Genomics*, 93(5):426–433, 2009.
- [149] S. A. Wahl, M. D. Haunschild, M. Oldiges, and W. Wiechert. Unravelling the regulatory structure of biochemical networks using stimulus response experiments and large-scale model selection. *IEE Proc Syst Biol*, 153(4):275–285, 2006.

- [150] A. Wächter and L. T. Bielger. On the implementation of a primal-dual interior point filter line search algorithm for large-scale nonlinear programming. *Mathe Prog*, 106(1):25–57, 2006.
- [151] M. Wendt, P. Li, and G. Wozny. Nonlinear chance-constrained process optimization under uncertainty. *Ind Eng Chem Res*, 41(15):3621–3629, 2002.
- [152] S. Werner, G. Diekert, and S. Schuster. Revisiting the thermodynamic theory of optimal atp stoichiometries by analysis of various atp-producing metabolic pathways. *J Mol Evol*, 71(5-6):346–355, 2010.
- [153] J. Wernstedt. *Experimentelle Prozessanalyse*. Verl. Technik, Berlin, 1989.
- [154] F. Wessely, M. Bartl, R. Guthke, P. Li, S. Schuster, and C. Kaleta. Optimal regulatory strategies for metabolic pathways in escherichia coli depending on protein costs. *Mol Sys Biol*, 7:515, 2011.
- [155] N. Wiener. *Cybernetics: Or Control and Communication in the Animal and the Machine*. MIT Press, Cambridge, MA, 1948.
- [156] K. Xie. *Diplomarbeit: Entwicklung eines Programmpakets für die nichtlineare dynamische Optimierung unter Anwendung des quasisequentiellen Ansatzes*. Technische Universität Ilmenau, Ilmenau, Germany, 2008.
- [157] A. Zaslaver, A. Bren, M. Ronen, S. Itzkovitz, I. Kikoin, S. Shavit, W. Liebermeister, M. G. Surette, and U. Alon. A comprehensive library of fluorescent transcriptional reporters for escherichia coli. *Nature Methods*, 3(8):623–628, 2006.
- [158] A. Zaslaver, A. E. Mayo, R. Rosenberg, P. Bashkin, H. Sberro, M. Tsalyuk, M. G. Surette, and U. Alon. Just-in-time transcription program in metabolic pathways. *Nat Genet*, 36(5):486–491, 2004.
- [159] S. Zellmer, S. Toepfer, D. Driesch, D. Woetzel, R. Guthke, M. Pfaff, and R. Gebhardt. Modelling of hepatic glutamine metabolism. In F. Allgöwer and M. Reuss, editors, *Proc. 2nd Conference Foundations of Systems Biology in Engineering*, pages 183–187, Stuttgart, Germany, 2007.
- [160] A. Zomorodi and C. Maranas. Optcom: A multi-level optimization framework for the metabolic modeling and analysis of microbial communities. *PLoS Comput Biol*, 8(2):e1002363, 02 2012.