

Lehrstuhl für Leichtbau
Technische Universität München

**Approximationen und paralleles Rechnen
bei der multidisziplinären
Strukturoptimierung**

Reiner Gleichmar

Vollständiger Abdruck der von der Fakultät für Maschinenwesen der Technischen Universität München zur Erlangung des akademischen Grades eines

Doktor-Ingenieurs (Dr.-Ing.)

genehmigten Dissertation.

Vorsitzender: Univ.-Prof. Dr.-Ing. Wolfgang A. Wall

Prüfer der Dissertation: 1. Univ.-Prof. Dr.-Ing. Horst Baier

2. Univ.-Prof. Dr.-Ing., Dr.-Ing. habil. Rudolf Schilling

Die Dissertation wurde am 22.01.2004 bei der Technischen Universität München eingereicht und durch die Fakultät für Maschinenwesen am 16.06.2004 angenommen.

Vorwort

Die vorliegende Arbeit entstand während meiner Tätigkeit als wissenschaftlicher Assistent am Lehrstuhl für Leichtbau der Technischen Universität München.

Mein herzlicher Dank gilt Herrn Prof. Dr.-Ing. Horst Baier, dem Leiter des Lehrstuhls. Die Anregungen in Fachgesprächen und die kritische Durchsicht der Dissertation waren eine wesentliche Stütze für das Gelingen der wissenschaftlichen Arbeit.

Herrn Prof. Dr.-Ing. Rudolf Schilling möchte ich für das entgegengebrachte Interesse an meiner Arbeit und die Übernahme des zweiten Gutachtens danken. Ferner gilt mein Dank Herrn Prof. Dr.-Ing. Wolfgang A. Wall für die Übernahme des Vorsitzes in der Prüfungskommission.

Allen Kollegen und Mitarbeitern des Lehrstuhls, die zum Gelingen dieser Arbeit beigetragen haben, möchte ich ebenfalls meinen Dank aussprechen. Zu nennen sind hier insbesondere Herr Dipl.-Ing. Harald Langer und Herr Dipl.-Ing. Tim Pühlhofer. Nicht zu vergessen ist auch der Anteil der durch Diplomanden, Semestranden und studentische Hilfskräfte erbracht wurde.

Zuletzt möchte ich meiner Familie und insbesondere meiner Frau Gunda für ihre liebevolle und kritische Unterstützung während der Entstehung dieser Arbeit danken in der Hoffnung, mich einmal entsprechend revanchieren zu dürfen.

Ingolstadt, im Juni 2005

Reiner Gleichmar

Inhaltsverzeichnis

1	Einleitung	1
1.1	Problemstellung	2
1.2	Stand der Technik	4
1.3	Zielsetzung und Gliederung der Arbeit	8
2	Multidisziplinäre Optimierung – Aufgaben und Lösungsverfahren	11
2.1	Multidisziplinäre Aufgaben der Strukturoptimierung	11
2.1.1	Thermisch-mechanische Kopplung im Entwurfsprozess	12
2.1.2	Strömungs-Struktur-Kopplung im Entwurfsprozess	15
2.1.3	Formulierung und Charakterisierung des allgemeinen multidisziplinären Optimierungsproblems	19
2.2	Das Entwurfsoptimierungssystem MOSES	20
2.3	Optimierungsalgorithmen und Lösungsverfahren	23
2.3.1	Behandlung mehrerer Ziele	23
2.3.2	Straffunktionsverfahren	24
2.3.3	Suchmethoden und gradientenfreie Verfahren	26
2.3.4	Gradientenverfahren und Newton-Verfahren	29
2.3.5	Direkte Verfahren zur Lösung restringierter Optimierungsaufgaben	30
2.3.6	Approximationsverfahren	31
2.3.7	Zusammenfassende Diskussion	35
3	Response Surface Methode und Design of Experiments	37
3.1	Grundlagen	37
3.1.1	Standard Ansatzfunktionen	38
3.1.2	Lineare Regression und Methode der kleinsten Fehlerquadrate .	39
3.1.3	Beispiel „Biegebalken“	41
3.2	Kriterien zur Beurteilung der Approximation	44
3.2.1	Globale Bewertung des Regressionsmodells	45
3.2.2	Analyse der Residuen	46
3.2.3	Einflussdiagnose	50
3.2.4	Zusammenfassende Betrachtung der Bewertungskriterien	51

3.3	Methoden zur Wahl der Stützstellen – Design of Experiments	52
3.3.1	Factorial Designs	53
3.3.2	Response Surface Designs für Regressionsmodelle zweiter Ordnung	58
3.3.3	Optimale Response Surface Designs	61
4	Anwendung der Response Surface Methode	63
4.1	Optimierungsprobleme mit konvexen Zielfunktionen und Restriktionen	63
4.1.1	Monotone Funktionsverläufe	63
4.1.2	Funktionsverläufe mit einem Extremum im Definitionsbereich .	70
4.2	Anmerkung zu nichtkonvexen Optimierungsaufgaben	77
4.3	Zusammenfassung und Ableitung von Verfahrensweisen	78
5	Erweiterung der Response Surface Methode	81
5.1	Transformation der Systemantwort	81
5.1.1	Anwendung bei Optimierungsaufgaben	82
5.1.2	Fazit	84
5.2	Transformation der Entwurfsvariablen und beliebige Ansatzterme . . .	84
5.2.1	Definition des erweiterten Approximationsmodells	85
5.2.2	Lineare Regression für das erweiterte Approximationsmodell . .	86
5.3	Bestimmung erweiterter Ansatzfunktionen durch ein diskret- nichtlineares Ausgleichsverfahren	89
5.3.1	Formulierung des Ausgleichsproblems	90
5.3.2	Lösungsverfahren für das diskrete Ausgleichsproblem	91
5.3.3	Lösungsverfahren für das nichtlineare Ausgleichsproblem	92
5.3.4	Zusammenspiel des diskreten und des nichtlinearen Lösungsver- fahrens – Empfehlung für die Anwendung	98
5.4	Geeignete Methoden des Design of Experiments	99
6	Grobparallelisierung auf einem PC-Cluster	101
6.1	Grobparallelisierung mit MOSES	101
6.1.1	Funktionsprinzip	101
6.1.2	Resultierende Leistungssteigerung	103
6.2	Auswahl eines Hardware- und Softwarekonzepts	104
6.2.1	Anforderungen	104
6.2.2	Verfügbare Systeme bzw. Konzepte	105
6.2.3	Auswahl eines Systems bzw. Konzepts	106
6.3	Beschreibung des realisierten PC-Clusters	107
6.3.1	Konfiguration	107
6.3.2	Job-Queueing-System	110

7	Anwendung bei multidisziplinären Optimierungsproblemen	115
7.1	Thermisch-mechanischer Entwurf einer Sandwichkonstruktion	115
7.1.1	Problemstellung	116
7.1.2	Beschreibung und Verifikation des Simulationsmodells	118
7.1.3	Referenzlösungen zur Beurteilung der Optimierungsergebnisse	122
7.1.4	RSA – lineare Terme und Interaktionsterme (22 Systemanalysen)	124
7.1.5	RSA – lineare Terme und Interaktionsterme (32 Systemanalysen)	124
7.1.6	RSA – vollständiger quadratischer Ansatz (45 Systemanalysen)	125
7.1.7	RSA – vollständiger quadratischer Ansatz und Transformation der Systemantworten	128
7.1.8	RSA – optimal gewählte erweiterte Ansatzterme (45 Systemana- lysen)	128
7.2	Strömungs-Struktur-Kopplung: Aeroelastic Tailoring eines Tragflügels	130
7.2.1	Problemstellung	130
7.2.2	Beschreibung des Simulationsmodells	133
7.2.3	Referenzlösungen zur Beurteilung der Optimierungsergebnisse	135
7.2.4	RSA für Lastfall 1 – quadratischer Ansatz (79 Systemanalysen)	138
7.2.5	RSAs für die Lastfälle 2 und 3 – quadratische Ansätze (je 79 Sys- temanalysen)	138
7.2.6	RSA – erweiterte Ansatzterme (89 Systemanalysen)	139
7.3	Behandlung von stark variierenden Systemantworten	141
7.3.1	Problemstellung	141
7.3.2	Lösung der Optimierungsaufgabe mit einem mathematischen Optimierungsverfahren	143
7.3.3	Glättung und Optimierung mit Hilfe einer Response Surface Ap- proximation	145
8	Zusammenfassung	147
	Literaturverzeichnis	151

Nomenklatur

Abkürzungen

BBD	Box-Behnken Design
CCC	Central Composite Circumscribed
CCD	Central Composite Design
CCF	Central Composite Faced
CCI	Central Composite Inscribed
DOE	Design of Experiments
FE	Finite Elemente
FEM	Finite Elemente Methode
MOSES	Multidisciplinary Optimization of Structures and Electromechanical Systems – Entwurfsoptimierungssystem
RSA	Response Surface Approximation
RSD	Response Surface Design
RSM	Response Surface Methode
UD	unidirektional – nur eine Faserrichtung

Symbole

Symbol	Beschreibung
\mathbf{A}	Matrix der fest vorgegebenen Ansatzterme
\mathbf{A}^*	Matrix der wählbaren Ansatzterme
\mathcal{E}	Einflussstärke der Variableneffekte
\mathbf{H}, h_{pp}	Hat-Matrix, Diagonalelemente der Hat-Matrix
MS_E	arithmetisches Mittel der Summe der Fehlerquadrate
$PRESS$	Prediction Error Sum of Squares
R^2	Bestimmtheitsmaß
R^2_{adj}	adjustiertes Bestimmtheitsmaß
R^2_{pred}	Maß für die Prognosegüte
SS_E	Summe der Fehlerquadrate
\mathbf{X}	Regressionsmatrix
\mathbf{Z}	Regressionsmatrix für ein erweitertes Approximationsmodell

Symbol	Beschreibung
\mathbf{a}	Parameter eines erweiterten Ansatzterms
b, \mathbf{b}	Koeffizienten des angepassten Regressionsmodells
$d_{(\)}$	standardisiertes Residuum
$e_{(\)}$	Residuum – Abweichung des prognostizierten Werts vom Stützwert
$e_{[1]}$	<i>PRESS</i> -Residuum
f, \mathbf{f}	Systemgleichungen
g, \mathbf{g}	Ungleichheitsrestriktionen
h, \mathbf{h}	Gleichheitsrestriktionen
l	Zahl der Levels eines Factorial Designs
NaN	Not a Number: Wird angegeben, wenn die Berechnung eines Wertes nicht möglich ist
n_C	Anzahl der Center Runs beim CCD und BBD
n_K	Anzahl der Koeffizienten bzw. Ansatzterme
n_P	Anzahl der Stützpunkte
n_S	Anzahl der wählbaren Ansatzterme
n_T	Anzahl der auszuwählenden Ansatzterme
n_V	Anzahl der Variablen
$r_{(\)}$	studentisiertes Residuum
t	Funktionsstyp eines erweiterten Ansatzterms
$t_{(\)}$	<i>R</i> -student Residuum
\mathbf{x}	Entwurfsvariablen
y, \mathbf{y}	Systemantworten
$\hat{y}, \hat{\mathbf{y}}$	approximierte Systemantworten
z, \mathbf{z}	Zielfunktion, Zielfunktionsvektor
α	Parameter für die Lage der sternförmig angeordneten Stützstellen eines CCDs
$\beta, \boldsymbol{\beta}$	Koeffizienten des Regressionsmodells
ε	Korrekturglied/Fehlerterm
λ	Parameter zur Transformation der Systemantworten
$\sigma^2, \hat{\sigma}^2$	Standardabweichung, Abschätzung für die Standardabweichung
$\xi, \boldsymbol{\xi}$	natürliche Variablen (nicht skaliert)

Indizes

Index	Beschreibung
$(\)_{\mathcal{A}, \mathcal{B}, \dots}$	zur physikalischen Disziplin $\mathcal{A}, \mathcal{B}, \dots$ gehörend
$(\)_{(p)}$	zum Stützpunkt p gehörend
$(\)_{[p]}$	einen nicht zur Anpassung des Approximationsmodells verwendeten Stützpunkt p betreffend
$(\)_{lb}, (\)_{ub}$	untere und obere Grenze
$\widehat{(\)}$	approximierte Größe

1 Einleitung

Leichtbaustrukturen sind bei nahezu allen technischen Systemen des Maschinenbaus anzutreffen. Zu den wohl bekanntesten Anwendungsgebieten gehört die Fahrzeugtechnik sowie der Einsatz in Luft- und Raumfahrtssystemen. Darüber hinaus gibt es eine Vielzahl von Sonderkonstruktionen, bei denen Leichtbaustrukturen eine wesentliche Funktion übernehmen. Betrachtet man die Komplexität der Gesamtsysteme, so kann deren Qualität und Leistungsfähigkeit meist nicht an strukturmechanischen Merkmalen allein gemessen werden.

Festigkeitsbezogene Kriterien können oft durch lokale Veränderungen an der Struktur erfüllt werden, die keinen oder nur einen sehr geringen Einfluss auf die globalen Struktureigenschaften haben. Ganz anders ist dies jedoch bei Anforderungen bezüglich Steifigkeiten und der Dynamik, die in der Regel die gesamte Struktur betreffen und damit auch Auswirkungen auf die Eigenschaften des Gesamtsystems haben. Somit ist es erforderlich, sowohl bei der Analyse als auch bei der Optimierung von Leichtbaustrukturen als Teil eines komplexen Systems, Anforderungen aus verschiedenen technisch-physikalischen Disziplinen gleichzeitig zu berücksichtigen (siehe Abbildung 1.1).

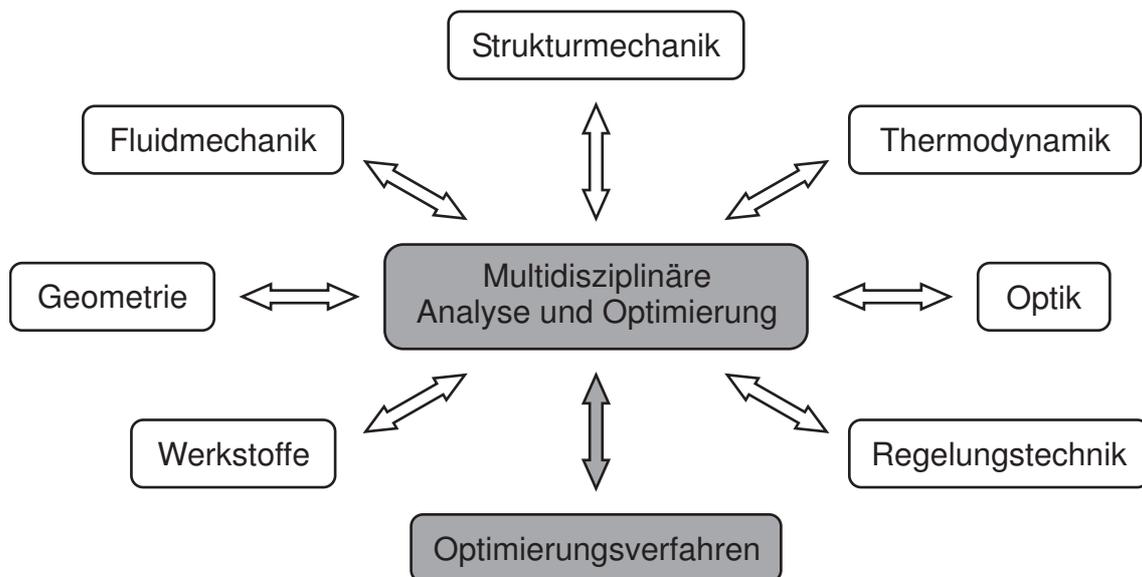


Abbildung 1.1: Multidisziplinäre Analyse und Optimierung

Fluidmechanische Anforderungen gelten beispielsweise für Tragflügel von Flugzeugen. So stellt sich bei der Aeroelastik die Aufgabe, die strukturmechanischen Eigenschaften eines Tragflügels so zu verändern, dass auch im Zusammenspiel mit den aerodynamischen Kräften eine möglichst optimale Funktionserfüllung gewährleistet ist. Dazu gehört zum Beispiel die Ruderwirksamkeit, bei der es darum geht, dass die durch einen Ruderausschlag hervorgerufene Änderung der aerodynamischen Kräfte auch tatsächlich eine Steuerung des Flugzeugs ermöglicht und nicht zu einer Verformung des Tragflügels führt, die mit den dann herrschenden Kräften die Ruderwirkung kompensiert oder gar umkehrt. Eine weitere Aufgabe aus der Aeroelastik ist das Verhindern von Flattern. Beim Flattern sind die strukturdynamischen Eigenschaften des Tragflügels so, dass dieser bei bestimmten Flugzuständen ins Schwingen kommt und dabei immer mehr Energie aus der umströmenden Luft aufnimmt. Dies kann zu derart hohen Schwingungsamplituden führen, dass die Tragflügelstruktur stark geschädigt wird. Ein ähnliches Verhalten kann auch bei falsch abgestimmten Flugregelungssystemen auftreten, so dass neben den strukturmechanischen und aerodynamischen Aspekten auch solche aus der Regelungstechnik berücksichtigt werden müssen.

Bei der Vibro-Akustik spielt ebenfalls die Fluidmechanik eine Rolle. Hier wird die Schallabstrahlung von schwingenden Strukturen betrachtet. Dabei kann es ebenfalls erforderlich sein, zusätzlich regelungstechnische Anforderungen mit einzubeziehen. Dies ist zum Beispiel dann gegeben, wenn aktive Maßnahmen zur Schwingungsdämpfung zum Einsatz kommen.

Weitere Interaktionen mit der Regelungstechnik finden sich bei der Regelung optomechanischer Systeme, wie zum Beispiel großer, hochgenauer Teleskope. Hierbei müssen Effekte wie Windlasten, seismische Störungen, thermisch bedingte Verformungen usw. berücksichtigt und kompensiert werden.

Die Thermodynamik in Leichtbaustrukturen kann ebenfalls eine Rolle spielen. Temperaturunterschiede bewirken Wärmedehnungen, welche die Funktion der Struktur zum Teil negativ beeinflussen. So kann zum Beispiel die optimale Funktion eines hochgenauen Reflektors nicht mehr gewährleistet werden, wenn die reflektierende Fläche von der vorgegebenen Form abweicht. Gleiches gilt auch bei hochgenauen Handhabungssystemen, die beispielsweise bei der Bestückung von Platinen oder der Fertigung mikroelektronischer Bauteile eingesetzt werden.

1.1 Problemstellung

Moderne computergestützte Analyse- und Simulationswerkzeuge ermöglichen heute detaillierte numerische Analysen komplexer Systeme. Obwohl die numerische Simulation ein etabliertes und handhabbares Werkzeug darstellt, entstehen Probleme bei der direkten Verwendung der detaillierten Analysen in einer Optimierungsumgebung.

Die größte Herausforderung stellt dabei die Problemgröße und der damit verbundene hohe Berechnungsaufwand für die einzelnen Systemanalysen dar. Die Ziel- und Restriktionsfunktionen besitzen bei den meisten Problemstellungen der multidisziplinären Optimierung einen nichtlinearen Verlauf, was selbst bei der Anwendung effizienter Optimierungsalgorithmen eine häufige Lösung der Systemgleichungen erforderlich macht. Anzumerken ist hierbei, dass der Anwender wie auch die Optimierungsalgorithmen meist nur das Systemverhalten in einem Designpunkt kennen und somit nicht das globale Verhalten im gesamten Designraum. Ohne eine globale Sicht kann es aber bezüglich der Effizienz der Optimierungsalgorithmen zu Problemen kommen. Aufgrund der Nichtlinearität sind die Optimierungsaufgaben häufig auch nichtkonvex. Das heißt, es existieren mehrere lokale Lösungen, von denen nur eine das globale Optimum darstellt. Hier muss ein zusätzlicher Aufwand betrieben werden, um neben den lokalen Optima auch das globale Optimum zu finden.

Insgesamt ist also festzuhalten, dass es schwierig und teilweise nicht möglich ist, den an der Rechenzeit gemessenen Optimierungsaufwand in einem vertretbaren Rahmen zu halten. Diese Tatsache wirkt der Forderung nach besseren Entwürfen in immer kürzer werdenden Entwicklungszyklen entgegen. Das vorrangige Ziel bei der Anwendung und Entwicklung von Verfahren für die Lösung multidisziplinärer wie auch anderer aufwendiger Optimierungsprobleme muss also sein, die Zahl der erforderlichen exakten Systemanalysen auf ein Minimum zu reduzieren.

Eine in den letzten Jahren verstärkt praktizierte Vorgehensweise ist der Einsatz globaler Approximationsverfahren wie die in dieser Arbeit behandelte Response Surface Methode (RSM). Diese liefert schnell auswertbare Approximationen für die Ziel- und Restriktionsfunktionen, die dann bei verschiedenen Optimierungsverfahren als Ersatz für die aufwendigen Systemanalysen verwendet werden. Dabei kann und muss es gerade bei extrem aufwendigen Problemen nicht Ziel sein, das Optimum exakt zu bestimmen. Die primäre Zielsetzung ist vielmehr, mit einem möglichst geringen Berechnungsaufwand möglichst nah an das exakte Optimum heranzukommen.

Um dies zu gewährleisten, müssen die Approximationsmodelle alle für eine Optimierung wesentlichen systeminhärenten Eigenschaften gut wiedergeben, ohne dabei auf eine große Zahl von Systemanalysen angewiesen zu sein. Dazu sind entsprechende Vorgehensweisen erforderlich, die im Rahmen der vorliegenden Arbeit untersucht und erarbeitet werden.

Ein ganz anderer Ansatz zur Reduktion des Berechnungsaufwands ist die Nutzung von Parallelrechnern. Fortschritte bei der Entwicklung parallelisierter Analysecodes können helfen, den für die einzelnen Systemanalysen erforderlichen Rechenzeitaufwand zu reduzieren. Dies gilt jedoch nur für eine begrenzte Zahl der vielfältigen multidisziplinären Problemstellungen. So lassen sich zum Beispiel Strömungsanalysen wegen der zugrunde liegenden Formulierung der Systemgleichungen recht effizient auf vielen Prozessoren durchführen, ohne dass die Berechnungsdauer weit über den theoretisch möglichen Wert steigt. Bei den meisten Disziplinen und insbesondere bei gekoppelten

multidisziplinären Analysen ist jedoch bei einer parallelisierten Berechnung ein wesentlich höherer Kommunikationsaufwand unter den verteilten Lösungsprozessen erforderlich. Dies führt dazu, dass die Lösungseffizienz bereits bei der Nutzung weniger Prozessoren erheblich unter dem theoretisch möglichen Wert liegt.

Ist das Ziel die Durchführung einer größeren Zahl von Systemanalysen in möglichst kurzer Zeit, so ist eine Grobparallelisierung der Analysen (eine pro Rechner/Prozessor) eine weitaus effizientere Vorgehensweise. Dabei gilt es, vorhandene, teilweise auch unterschiedliche oder nur temporär zur Verfügung stehende Ressourcen optimal auszunutzen. Solche Aufgaben müssen von einem so genannten Job-Management-System übernommen werden. Die Grobparallelisierung von Analysen auf einem aus vielen Einzelrechnern zusammengesetzten Cluster sowie dessen Nutzung über ein Entwurfs-optimierungssystem ist ein untergeordnetes, aber durchaus wichtiges Themengebiet der vorliegenden Arbeit. Bei den als zentrales Thema behandelten Verfahren zur Erstellung von RSAs und deren Anwendung bei Optimierungsaufgaben kann der zeitliche Aufwand durch eine Grobparallelisierung zusätzlich reduziert werden.

1.2 Stand der Technik

„Die Response Surface Methode ist eine Sammlung statistischer und mathematischer Methoden, die bei der Entwicklung, Verbesserung und Optimierung von Prozessen oder Produkten eingesetzt werden kann“ [Myers und Montgomery 1995]. Die Techniken und Methoden der RSM wurden ursprünglich entwickelt, um, basierend auf den Ergebnissen physikalischer Experimente, empirische Näherungsfunktionen aufzubauen und zu überprüfen. Die RSM umfasst Methoden zur Planung der Experimente (bekannt unter der Bezeichnung Design of Experiments, DOE), also der Festlegung, welche Datenpunkte zu generieren und untersuchen sind. Weiterhin vertreten sind Methoden zur Bestimmung der unbekannt Koeffizienten der Approximationsmodelle sowie zur Bewertung der Güte der resultierenden Response Surface Approximation (RSA). Aufgrund der Ähnlichkeiten zwischen detaillierten numerischen Simulationen und physikalischen Experimenten gibt es ein in letzter Zeit stetig wachsendes wissenschaftliches Interesse an der Nutzung von RSAs bei der Behandlung komplexer Optimierungsaufgaben.

Erste bedeutende Veröffentlichungen zur RSM finden sich bereits in den frühen 50er Jahren. Im Wesentlichen handelt es sich hierbei um Methoden des DOE. So stammt zum Beispiel das Central Composite Design (CCD) [Box und Wilson 1951] aus dieser Zeit, das heute noch zu den am meisten genutzten Response Surface Designs (RSDs) gehört. Als erster war es Myers [1971], der eine Fülle von Publikationen und einige bereits entstandene Monographien zu den genannten Teilgebieten der RSM in einem Werk zusammenfasste. Dieses wurde in Zusammenarbeit mit Montgomery wesentlich erweitert und aktualisiert [Myers und Montgomery 1995] und ist in dieser Form das meist zitierte Standardwerk heutiger Publikationen. Ein ähnlich umfassendes Werk,

das ebenfalls häufig in Literaturverzeichnissen zu finden ist, stammt von Box und Draper [1987]. Die beiden genannten Monographien unterscheiden sich abgesehen von der Aktualität nicht wesentlich in den dargestellten Methoden.

Auch in den letzten Jahren gab es eine Vielzahl von Veröffentlichungen zu Methoden und Vorgehensweisen der RSM.

Die Veröffentlichungen zu Methoden des DOE sind zwei Kategorien zuzuordnen. Zunächst sind die für regelmäßig geformte Definitionsbereiche geeigneten strukturierten RSDs zu nennen. Wie bereits erwähnt, sind diese seit langem bekannt und in [Myers und Montgomery 1995] oder [Box und Draper 1987] umfassend dargestellt. Bezüglich der strukturierten RSDs finden sich in der Literatur Untersuchungen zu geeigneten Vorgehensweisen bei der Anwendung und einige Verbesserungen zu bestehenden Methoden.

Für unregelmäßig geformte, also unterbrochene und/oder nichtlinear berandete Definitionsbereiche müssen optimale RSDs benutzt werden. Die formalen Anforderungen hierzu wurden größtenteils Anfang der 70er Jahre behandelt. So basiert das heute am häufigsten genutzte *D*-optimale Design auf Kriterien von Box und Draper [1971]. Einen umfassenden Überblick über die Theorie der optimalen DOEs gibt Pukelsheim [1993]. Die optimalen RSDs haben erst im Computerzeitalter große Popularität erlangt, da zu deren Bestimmung numerisch relativ aufwendige Algorithmen erforderlich sind. Viele aktuelle mathematische und statistische Programmpakete, darunter auch MATLAB [MATLAB Statistics Toolbox 2002], bieten Algorithmen zur Generierung *D*-optimaler Designs, die universell und effizient einsetzbar sind.

Die Verfügbarkeit der optimalen RSDs sollte nach Myers und Montgomery [1995] jedoch nicht dazu verleiten, diese immer und nur als „Black Box“ zu verwenden. Beide raten, wenn möglich die strukturierten Standard RSDs zu verwenden, da diese bezüglich vieler Kriterien optimal sind.

Bei der Anwendung der RSM für Optimierungsaufgaben können zwei Vorgehensweisen unterschieden werden. Zum einem handelt es sich um sequenzielle Verfahren, die in aktuellen Veröffentlichungen (z. B. [Rodríguez u. a. 2001]) unter dem Begriff „Sequential Approximate Optimization“ (SAO) zusammengefasst werden. Zum anderen zeigt eine Übersicht von Simpson u. a. [2001], dass es nach wie vor viele Autoren gibt, die die RSM in der Standardform nutzen.

Im Gegensatz zur Standard RSM werden bei den SAO-Verfahren die zum Aufstellen der RSAs benötigten Datenpunkte von einem übergeordneten Optimierungsalgorithmus bestimmt. Die Verfahren unterscheiden sich in der Vorgehensweise, wie der approximierte Bereich sukzessive eingeschränkt wird, um eine Konvergenz hin zu einem Optimum zu gewährleisten. So benutzten Tropov u. a. [1993] „Move Limits“, wie sie auch bei klassischen Approximationsverfahren (z. B. Sequenzielle Lineare Programmierung) Anwendung finden. In dem somit eingeschränkten Bereich um das Optimum der aktuellen RSA werden zusätzliche und bereits vorhandene Datenpunkte verwendet, um eine neue Approximation aufzubauen. Weitere „Move Limit“-Strategien finden

sich zum Beispiel in [Thomas u. a. 1992], [Chen 1993] und [Bloebaum u. a. 1994]. Ziel ist es dabei, den Bereich so einzugrenzen, dass man dort eine verlässliche Approximation für die Systemantwort erhält. So werden ähnliche, in letzter Zeit entwickelte Methoden unter dem Namen „Trust Region Model Management“ veröffentlicht (z. B. [Rodríguez u. a. 2000]). Die genannten Verfahren verwenden meist lineare und seltener quadratische Approximationsmodelle. Während die linearen Approximationen teilweise Konvergenzprobleme haben, liegt das Problem bei den quadratischen Ansätzen eher darin, dass im Laufe der iterativen Lösung viele exakte Systemanalysen benötigt werden.

Unger u. a. [1992] zählten zu den ersten, die diesem Nachteil begegneten, indem sie bei der aeroelastischen Optimierung eines Tragflügels Simulationsmodelle mit sehr unterschiedlicher Komplexität verwendeten. In einer Reihe von Veröffentlichungen wurde die Variable-Complexity-Methode am Beispiel des Tragflügels eines Hochgeschwindigkeitstransportflugzeugs (High Speed Civil Transport, HSCT) untersucht und weiterentwickelt (z. B. [Hutchinson u. a. 1994], [Kaufman u. a. 1996], [Giunta u. a. 1997] und [Balabanov u. a. 1999]). Die beschriebenen Verfahren nutzen zu Beginn sehr einfache Analysemodelle, um mit verhältnismäßig wenig Berechnungsaufwand viele Datenpunkte zu erzeugen. Mit diesen sind dann ungültige oder nicht realisierbare Entwürfe identifizierbar, die beim weiteren Vorgehen nicht mehr berücksichtigt werden. Aus dem verbleibenden Datensatz wird mittels des D -Optimalitätskriteriums eine definierte Menge von Datenpunkten ausgewählt. Die darin enthaltenen Systemantworten werden anschließend mit Daten ersetzt, die aus genaueren Analysemodellen stammen. Mit den neuen Daten kann nun eine RSA aufgebaut und deren Optimum bestimmt werden. Der Vorgang wiederholt sich mit veränderten Grenzen für den untersuchten Bereich solange, bis ein Abbruchkriterium erfüllt ist.

Sellar u. a. [1996] stellten mit der „Concurrent Subspace Optimization“ (CSSO) eine Methodik vor, die für spezielle multidisziplinäre Problemstellungen einsetzbar ist, bei denen sich das eigentliche Optimierungsproblem in eine Reihe von disziplinspezifischen Unterproblemen zerlegen lässt. Dabei bestimmen die Optima der Unterprobleme die Datenpunkte, die zum Aufstellen der disziplinübergreifenden Approximation verwendet werden. Sellar u. a. [1996] benutzen hierzu Neuronale Netze. Wenjie und Renaud [2001] und weitere Autoren nutzen die CSSO um geeignete Datenpunkte für die RSM zu generieren. Die CSSO-basierte Methodik hat jedoch das Problem, dass vorrangig Datenpunkte in den für die jeweilige Disziplin relevanten Bereichen gesammelt werden. Für erforderliche Zwischenlösungen stehen somit nur wenige Informationen zur Verfügung.

Die Variable-Complexity-Methode wie auch die CSSO können dazu beitragen, den Aufwand zum Erstellen eines Approximationsmodells deutlich zu reduzieren. Anzumerken ist jedoch, dass dabei häufig ein erhöhter Modellierungsaufwand betrieben werden muss. So ist es nicht als selbstverständlich anzusehen, dass neben einem aufwendig erstellten und gepflegten Simulationsmodell auch immer ein einfaches und stets an den aktuellen Stand im Entwicklungsprozess angepasstes Parallelmodell existiert. Zur

Anwendung der CSSO muss eine geeignete Dekomposition des Optimierungsproblems erfolgen. Sellar u. a. [1996] beschränken sich bei der Darstellung des Verfahrens auf ein sehr einfaches multidisziplinäres Simulationsmodell, bei dem die Disziplinen jeweils nur über eine Größe gekoppelt sind. Eine Anwendung der CSSO auf Probleme mit komplexeren Kopplungen erfolgt nicht. So verwenden auch Rodríguez u. a. [2001] stark vereinfachte Simulationsmodelle im Zusammenhang mit der CSSO. Dies führt zu der bereits beschriebenen Problematik, dass ein nicht vernachlässigbarer Aufwand zur Erstellung reduzierter Modelle betrieben werden muss. Insgesamt können die beiden Methoden also keinen Anspruch auf Allgemeingültigkeit erheben, da immer eine problemspezifische Anpassung erforderlich ist.

Nicht zuletzt aus diesem Grund gibt es auch in jüngerer Zeit eine Vielzahl von Veröffentlichungen, bei denen die Standard RSM eingesetzt wird. So nennen Simpson u. a. [2001] insgesamt 26 Publikationen, die sich mit der Anwendung der RSM bei allgemeinen und multidisziplinären Problemstellungen befassen. Die in den einzelnen Arbeiten dargestellten Vorgehensweisen unterscheiden sich vor allem in den verwendeten DOEs. Die Mehrzahl der Autoren verwendet CCDs und einige auch D -optimale Designs. Seltener sind zufällig gewählte RSDs und Fractional Factorial Designs. Mit den festgelegten Datenpunkten werden fast ausschließlich Approximationsmodelle zweiter Ordnung erstellt. Dabei wird bei etwa der Hälfte auf die Betrachtung von Interaktionstermen verzichtet.

Nur wenige Veröffentlichungen beschäftigen sich mit Verfahren zur Verbesserung der RSAs. So verwenden Venter u. a. [1996], Roux u. a. [1996] und Knill u. a. [1999] ein schrittweises Verfahren (Stepwise Regression, siehe [Myers und Montgomery 1995]), um wichtige Ansatzterme zu identifizieren. Dabei greifen Venter u. a. [1996] und Roux u. a. [1996] auch auf Terme vierter Ordnung zurück. Bei der Stepwise Regression können drei Vorgehensweisen unterschieden werden:

- Bei der Forward Selection erfolgt ein schrittweiser Aufbau des Approximationsmodells, indem zunächst alle zur Auswahl stehenden Ansatzterme einzeln in das Regressionsmodell aufgenommen werden. Der Term, der die größte Verbesserung bezüglich eines definierten Kriteriums bringt, wird im Approximationsmodell festgehalten und die Prozedur wiederholt sich, bis die vorgegebene Modellgröße erreicht ist. Der Nachteil dieser Vorgehensweise ist, dass ein in das Approximationsmodell integrierter Term nicht mehr entfernt werden kann. Es ist jedoch häufig so, dass zum Beispiel zwei Ansatzterme in der Kombination wesentlich besser geeignet sind als ein zuvor als wichtig eingestufte einzelner Term.
- Die Backward Elimination beginnt mit einem Approximationsmodell, das sich aus allen zur Auswahl stehenden Termen zusammensetzt. Ein Kriterium, das über die Wichtigkeit der einzelnen Ansatzterme Aufschluss gibt, bestimmt, welche der Terme schrittweise eliminiert werden. Das Problem dieser Verfahrensweise liegt darin, dass in der Regel nicht genug Datenpunkte zur Verfügung stehen, um die Regressionskoeffizienten zu bestimmen.

- So wird meist eine Kombination aus den beiden genannten Methoden verwendet: Mixed Stepwise Regression oder auch nur Stepwise Regression genannt. Auch mit dieser Methode ist nicht gewährleistet, dass am Ende das am besten geeignete Approximationsmodell resultiert [Myers und Montgomery 1995], da die Reihenfolge, in der die Terme in das Regressionsmodell integriert und eliminiert werden, eine wesentliche Rolle spielt.

Bei der Anwendung der drei Verfahren darf die Tatsache nicht unberücksichtigt bleiben, dass die Kriterien eigentlich nur für lineare Approximationsmodelle gelten [Myers und Montgomery 1995; Simpson u. a. 2001]. Dies zeigt auch eine Untersuchung in [Hausmann 2002]. Das Ziel dabei war der Aufbau eines Approximationsmodells, das nicht nur Polynomterme, sondern auch andere funktionale Zusammenhänge (z. B. $\ln(x)$, $\frac{1}{x}$, \sqrt{x} , ...) beinhaltet. Die Möglichkeit zur Verwendung solcher Ansatzterme und die daraus resultierenden Vorzüge bei der Beschreibung größerer Nichtlinearitäten werden von vielen Autoren lediglich erwähnt.

1.3 Zielsetzung und Gliederung der Arbeit

Am Lehrstuhl für Leichtbau wird derzeit das Entwurfsoptimierungssystem MOSES (Multidisciplinary Optimization of Structures and Electromechanical Systems) entwickelt, das verschiedene Verfahren und Ansätze verknüpft, um eine effiziente Lösung komplexer multidisziplinärer Optimierungsprobleme zu ermöglichen. Die vorliegende Arbeit soll hierzu in zweierlei Hinsicht einen Beitrag leisten:

- Das zentrale Thema ist die Untersuchung und Entwicklung von Verfahrensweisen, die auf der Basis der Response Surface Methode geeignete Approximationen für rechnerisch aufwendig zu ermittelnde technisch-physikalische Zusammenhänge liefern. Die Approximationen sollen bei der Anwendung verfügbarer Optimierungsverfahren als Ersatz für die Systemanalysen dienen.

Dabei ist es primäre Zielsetzung, mit möglichst wenig Berechnungsaufwand Approximationen zu erstellen, die möglichst nah an das exakte Optimum heranzuführen. Dazu müssen die Approximationsmodelle alle für eine Optimierung wesentlichen systeminhärenten Eigenschaften gut wiedergeben, ohne auf eine große Zahl von Systemanalysen angewiesen zu sein. Hierzu gilt es zunächst, die bestehenden Methoden aufzuarbeiten, um daraus geeignete Vorgehensweisen abzuleiten.

Da die Fähigkeiten der meist verwendeten quadratischen Polynomfunktionen bezüglich der Beschreibung nichtlinearer Zusammenhänge begrenzt sind, liegt ein großes Potential in der Verwendung erweiterter funktionaler Zusammenhänge in den einzelnen Ansatztermen. Die Aufarbeitung der Literatur zeigt, dass bei den zur Verfügung stehenden Methoden zur Zusammenstellung eines möglichst guten

Approximationsmodells noch Entwicklungsbedarf besteht. Deshalb soll ein Verfahren erarbeitet werden, das den Einsatz verschiedener mathematischer Funktionen ermöglicht und eine optimale Auswahl der damit zur Verfügung stehenden Ansatzterme gewährleistet. In diesem Zusammenhang sind auch geeignete Methoden zur Wahl der Stützstellen zu diskutieren.

- Ein untergeordneter Teil der Arbeit behandelt den Aufbau sowie die Konfiguration eines flexiblen, auf die Anforderungen des Entwurfsoptimierungssystems zugeschnittenen Parallelrechensystems auf der Basis vieler PCs. Durch die Grobparallelisierung von Prozessen soll das System dazu beitragen, den zeitlichen Aufwand für viele der implementierten Verfahren und Funktionalitäten drastisch zu reduzieren.

Bei der Grobparallelisierung gilt es, vorhandene, teilweise auch unterschiedliche oder nur temporär zur Verfügung stehende Ressourcen optimal auszunutzen. Die Koordination der Abarbeitung der vom Entwurfsoptimierungssystem angestoßenen Prozesse muss dabei ein Job-Management-System übernehmen. Hierzu ist ein geeignetes Rechner- und Softwarekonzept zu erarbeiten.

Anhand zweier Beispiele werden in Kapitel 2 zunächst die Problemstellungen der multidisziplinären Optimierung charakterisiert. Zur effizienten Lösung der Optimierungsaufgaben werden anhand des Entwurfsoptimierungssystems MOSES verschiedene Ansätze vorgestellt, zu denen auch die in dieser Arbeit behandelten Themen gehören. Weiterhin erfolgt eine Darstellung des zugrunde liegenden Funktionsprinzips, woraus später auch Anforderungen für die Grobparallelisierung abgeleitet werden. Da man bei der Optimierung mit RSAs auf Optimierungsverfahren angewiesen ist, die bestehender oder potentieller Bestandteil des Entwurfsoptimierungssystems sind, werden solche vorgestellt und unter anderem bezüglich ihrer Parallelisierbarkeit bewertet.

In Kapitel 3 erfolgt die Aufarbeitung bestehender Techniken der RSM. Dabei werden vor allem Kriterien zur Bewertung erstellter Approximationen sowie Methoden zur Wahl der Stützstellen dargestellt und diskutiert. Dies erfolgt zum einem aus dem genannten Ziel heraus, dass die Approximationen das Systemverhalten möglichst gut beschreiben sollen, zum anderen wegen der Anforderung, dies mit möglichst wenigen Stützpunkten zu erreichen. Somit wird eine Grundlage für die spätere Ableitung von Vorgehensweisen und Modifikationen geschaffen.

Basierend darauf werden in Kapitel 4 Optimierungsbeispiele mit nichtlinearen Ziel- und Restriktionsfunktionen behandelt. Für die RSAs werden dabei ausschließlich lineare und quadratische Ansätze verwendet. Mit den gewonnenen Erkenntnissen lässt sich eine Reihe von Vorgehensweisen ableiten, die teilweise zu einer Verbesserung des mit den Approximationen ermittelbaren Optimums führen, ohne auf eine Verkleinerung des approximierten Bereichs und damit auf zusätzliche Systemanalysen angewiesen zu sein. Es zeigt sich jedoch, dass dies bei der Anwendung quadratischer Ansatzfunktionen aufgrund größerer Nichtlinearitäten oftmals nicht vermeidbar sein wird.

Um dieser Problematik Rechnung zu tragen, befasst sich Kapitel 5 mit der Behandlung größerer Nichtlinearitäten. Um diese weiterhin auf der Basis quadratischer Approximationen beschreiben zu können, wird zunächst die Möglichkeit evaluiert, die Systemantworten zu transformieren. Da die Entwurfsvariablen gerade bei multidisziplinären Problemen teilweise sehr unterschiedliche, nichtlineare Einflüsse auf die Systemantworten erwarten lassen, wird als wesentlicher Teil ein Verfahren erarbeitet, mit dem verschiedene Ansatzterme in optimaler Weise kombiniert werden. Hierzu wird ein diskret-nichtlineares Ausgleichsproblem formuliert und dafür geeignete Lösungsverfahren vorgestellt. Abschließend steht die Vorstellung und Diskussion geeigneter Methoden zur Wahl der Stützpunkte.

Die Thematik der Grobparallelisierung auf einem im Rahmen dieser Arbeit aufgebauten PC-Cluster wird in Kapitel 6 behandelt. Da dieser als ein wichtiges Werkzeug für das Entwurfsoptimierungssystem zu sehen ist, wird zunächst das Funktionsprinzip für die Zusammenarbeit beider Systeme dargestellt. Aus den daraus resultierenden Aufgaben und weiteren Gesichtspunkten werden dann Anforderungen an das zu realisierende System formuliert. Danach erfolgt die Auswahl sowie die Umsetzung eines geeigneten Hard- und Softwarekonzepts, wobei auch detailliert auf das zur Steuerung erforderliche Job-Management-System eingegangen wird.

Um die Leistungsfähigkeit beim Einsatz von RSAs bei der Behandlung komplexer Optimierungsaufgaben und insbesondere der hierzu erarbeiteten Vorgehensweisen und Verfahren zu untersuchen, werden in Kapitel 7 abschließend drei Beispiele behandelt. Bei zwei der untersuchten Problemstellungen handelt es sich um die in Kapitel 2 vorgestellten multidisziplinären Optimierungsaufgaben, die hierzu detaillierter formuliert werden. Beim dritten Optimierungsproblem kommt die Eigenschaft der RSAs zum Tragen, stark variierende Systemantworten zu glätten. Ein wichtiger Punkt dieses Kapitels wird auch die Überprüfung der erarbeiteten Verfahren hinsichtlich ihrer Grenzen sein.

In Kapitel 8 werden die wesentlichen Inhalte und Erkenntnisse dieser Arbeit zusammengefasst.

2 Multidisziplinäre Optimierung – Aufgaben und Lösungsverfahren

In diesem Kapitel werden Problemstellungen der multidisziplinären Optimierung (MDO) vorgestellt und charakterisiert. Die zugrunde liegenden Aufgaben der multidisziplinären Analyse lassen sich in zwei Kategorien unterteilen, für die jeweils ein Beispiel angeführt wird. Anschließend wird das allgemeine multidisziplinäre Optimierungsproblem formuliert und dessen wesentliche Merkmale beschrieben.

Zur Lösung der komplexen Aufgaben der multidisziplinären Optimierung wird ein Entwurfsoptimierungssystem beschrieben, das verschiedene Ansätze zur Reduktion des Optimierungsaufwands verknüpft. Hierzu zählen Optimierungsverfahren, deren Eignung im Hinblick auf die Optimierung komplexer Problemstellungen diskutiert wird sowie die Grobparallelisierung der zugehörigen Prozesse. So werden die beschriebenen Optimierungsverfahren auch bezüglich ihrer Parallelisierbarkeit bewertet. Globale Approximationsverfahren, darunter die in dieser Arbeit behandelte Response Surface Methode (RSM), können ebenfalls dazu beitragen, den zeitlichen Aufwand für den Optimierungsprozess zu reduzieren. Auch deren wesentliche Eigenschaften werden verglichen und bewertet.

2.1 Multidisziplinäre Aufgaben der Strukturoptimierung

Beispiele für Entwurfsaufgaben, bei denen die Interaktion mehrerer technisch-physikalischer Disziplinen die Qualität des Gesamtentwurfs bestimmt, wurden bereits in Kapitel 1 aufgeführt. Allgemein lauten die Systemgleichungen der Disziplinen \mathcal{A} und \mathcal{B} :

$$\mathbf{f}_{\mathcal{A}}(\mathbf{x}_{\mathcal{A}}, \mathbf{x}_{\mathcal{AB}}, \mathbf{y}_{\mathcal{A}}, \mathbf{y}_{\mathcal{B}}) = 0 \quad \text{mit} \quad \mathbf{f}_{\mathcal{A}} = f_{\mathcal{A}(1)}, \dots, f_{\mathcal{A}(n_{\mathcal{A}})} \quad (2.1a)$$

$$\mathbf{f}_{\mathcal{B}}(\mathbf{x}_{\mathcal{B}}, \mathbf{x}_{\mathcal{AB}}, \mathbf{y}_{\mathcal{A}}, \mathbf{y}_{\mathcal{B}}) = 0 \quad \text{mit} \quad \mathbf{f}_{\mathcal{B}} = f_{\mathcal{B}(1)}, \dots, f_{\mathcal{B}(n_{\mathcal{B}})} \quad (2.1b)$$

Dabei sind $\mathbf{x}_{\mathcal{A}}$ und $\mathbf{x}_{\mathcal{B}}$ diejenigen Entwurfsvariablen, die nur in den Systemgleichungen der jeweiligen Disziplin auftreten. Die Entwurfsvariablen, die in den Systemgleichungen beider Disziplinen vorkommen und damit deren Systemantworten beeinflussen, sind in $\mathbf{x}_{\mathcal{AB}}$ zusammengefasst. Durch sie wird die Entwurfskopplung definiert. Das

zusätzliche Auftreten der Systemantworten \mathbf{y}_A und \mathbf{y}_B in den Systemgleichungen der jeweils anderen Disziplin definiert die physikalische Kopplung, die eine gekoppelte Lösung der Systemgleichungen erforderlich macht. Dabei können die disziplinbeschreibenden Systemgleichungen nicht immer gleichzeitig in einem umfassenden Gleichungssystem behandelt werden. In einigen Fällen gelingt dies, wobei dann zur Lösung der problemspezifisch formulierten Systemgleichungen nur spezialisierte Programmsysteme zur Verfügung stehen, die meist einen ebenso spezialisierten Anwender erfordern. In der Regel kann durch die Anwendung solcher Programmsysteme der numerische – und somit der zeitliche – Analyseaufwand reduziert werden.

Eine wesentlich allgemeinere Vorgehensweise stellt die Zerlegung oder Dekomposition der multidisziplinären Problemstellung in Teilaufgaben dar. Diese bietet den Vorteil, dass die einzelnen Disziplinen von fachlich spezialisierten Personen mit deren numerisch und rechnerisch spezialisierten Werkzeugen behandelt werden können. Dabei kann die Kopplung, welche die Abhängigkeiten der einzelnen Disziplinen definiert, sowohl hierarchisch als auch nichthierarchisch sein. Im ersten Fall besteht die Möglichkeit, die Lösung für das Gesamtproblem sequenziell (hierarchisch) zu ermitteln. Dies setzt voraus, dass die Systemantwort der Disziplin \mathcal{A} Einfluss auf das Systemverhalten der Disziplin \mathcal{B} hat, wobei ein umgekehrter Einfluss nicht oder nur in vernachlässigbarem Umfang existieren darf. Die Systemgleichungen für Problemstellungen mit hierarchischer Kopplung sehen somit wie folgt aus:

$$\mathbf{f}_A(\mathbf{x}_A, \mathbf{x}_{AB}, \mathbf{y}_A) = 0 \quad (2.2a)$$

$$\mathbf{f}_B(\mathbf{x}_B, \mathbf{x}_{AB}, \mathbf{y}_A, \mathbf{y}_B) = 0 \quad (2.2b)$$

Ein Vertreter der Kategorie (2.2) ist die *thermisch-mechanische Kopplung*. Als Beispiel für die Problemklasse (2.1) sind Aufgaben mit *Strömungs-Struktur-Kopplung* zu nennen, bei denen in der Regel eine nichthierarchische Kopplung vorliegt.

Beide Problemstellungen der multidisziplinären Analyse sowie deren Einordnung in einen Auslegungs- und Optimierungsprozess werden im Folgenden näher beschrieben.

2.1.1 Thermisch-mechanische Kopplung im Entwurfsprozess

Bei der thermisch-mechanischen Kopplung können die Disziplinen Thermodynamik und Strukturmechanik sequenziell gelöst werden. Die Berechnung der Temperaturverteilung wird von der Thermodynamik geliefert. Das somit bekannte Temperaturfeld wird wiederum in der Strukturmechanik benötigt, um zusätzlich zu den rein mechanisch bedingten Verformungen auch die durch Wärmedehnung hervorgerufenen Verformungen zu berechnen. Der Einfluss der Strukturverformung auf das Temperaturfeld ist in der Regel vernachlässigbar.

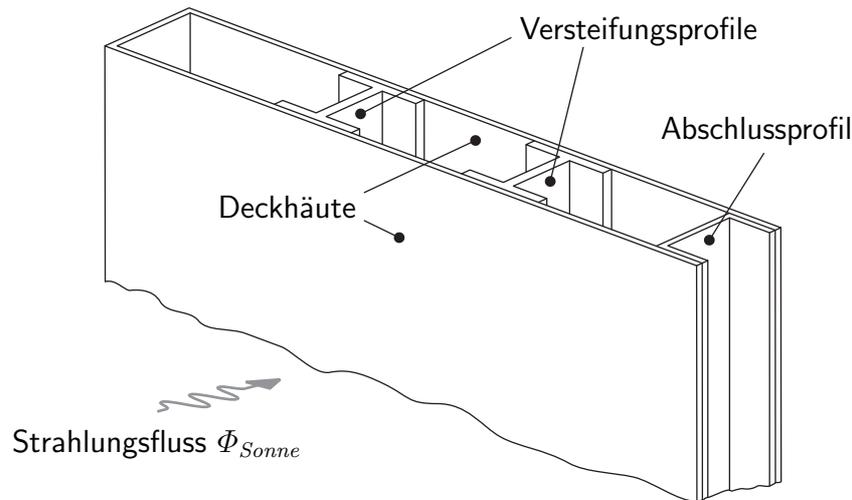


Abbildung 2.1: Thermisch-mechanische Problemstellung bei der sandwichartigen Konstruktion eines durch einseitige Sonneneinstrahlung belasteten Aufzugtürblatts

Einen häufigen Anwendungsfall findet man bei der Betrachtung der thermisch bedingten Verformung von Sandwichkonstruktionen. Abbildung 2.1 zeigt den sandwichartigen Querschnittsaufbau eines durch einseitige Sonneneinstrahlung belasteten Aufzugtürblatts, bei dem zwischen den Außenhäuten Versteifungsprofile integriert sind. Durch die absorbierte Strahlungsenergie der Sonne erwärmt sich die von der Schachtseite her gesehen außen liegende Deckhaut. Dadurch unterliegt die Außenseite einer größeren Wärmedehnung, wodurch sich das Türblatt stark krümmt und somit verklemmt.

Um das Verformungsverhalten positiv zu verändern, muss unter Berücksichtigung sowohl der thermodynamischen als auch der strukturmechanischen Effekte ein optimaler Entwurf gefunden werden. Ein möglicher Ansatz ist zum Beispiel die Verstärkung des Effekts, dass die absorbierte Strahlungsenergie durch eine Reihe von Wärmetransportmechanismen (Wärmeleitung, Konvektion und Strahlung) abgebaut bzw. abgeleitet wird. Zu den die Wärmeleitung beeinflussenden Entwurfsparametern gehören in erster Linie die Querschnittsabmessungen der Versteifungsprofile. Eine Variation dieser Entwurfsvariablen verändert jedoch nicht nur die Temperaturverteilung, sondern auch die mechanischen Eigenschaften des Türblatts, womit eine Entwurfskopplung gegeben ist. Der optimale Entwurf muss eine Reihe von Anforderungen erfüllen:

- Die Abmessungen des Querschnitts der Versteifungsprofile sollen so verändert werden, dass die Verformung des Türblatts so klein wie möglich wird.
- Da eine Variation der Querschnittsabmessungen nicht nur eine Veränderung der Temperaturverteilung, sondern auch veränderte mechanische Eigenschaften zur

Folge hat, müssen sowohl die thermodynamischen als auch die strukturmechanischen Systemgleichungen gelöst werden.

- Weiterhin soll der optimale Entwurf nicht schwerer sein als die ursprüngliche Konstruktion.
- Damit die Versteifungen (Aluminium-Strangpressprofile) herstellbar sind, dürfen deren Querschnittsabmessungen nur innerhalb gewisser Grenzen variiert werden.

Die genannten Anforderungen können auch in mathematischer Form angegeben werden:¹

$$\text{Minimiere} \quad u_{max} = u_{max}(\mathbf{x}_{QS}, \Delta\mathbf{T}) \quad (2.3a)$$

$$\text{mit} \quad \Delta\mathbf{T} = \Delta\mathbf{T}(\mathbf{x}_{QS}) \quad (2.3b)$$

$$\text{so, dass} \quad m = m(\mathbf{x}_{QS}) \leq m_{Ref} \quad (2.3c)$$

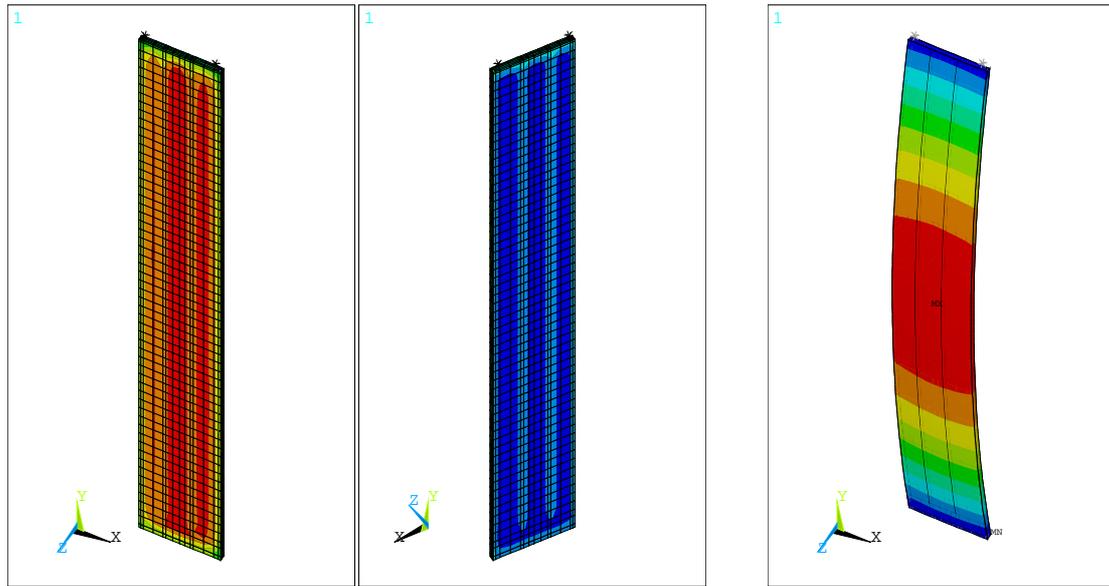
$$\text{und} \quad \mathbf{x}_{min} \leq \mathbf{x}_{QS} \leq \mathbf{x}_{max} \quad (2.3d)$$

Zur Lösung dieser Optimierungsaufgabe müssen die Systemgleichungen (2.3a) und (2.3b) gekoppelt gelöst werden. Hierzu zeigt Abbildung 2.2 den hierarchischen Ablauf zur Lösung der thermisch-mechanisch gekoppelten Systemgleichungen. Während der strukturmechanische Teil (2.3a) in vergleichsweise kurzer Zeit gelöst werden kann, gestaltet sich die Berechnung des Temperaturfelds (2.3b) hingegen sehr aufwendig:

- Die Sonneneinstrahlung wird mittels einer Fläche simuliert, deren Oberflächentemperatur so gewählt ist, dass der davon ausgehende Strahlungsfluss dem der Sonne entspricht. Dies macht die Berücksichtigung von Wärmestrahlungseffekten erforderlich, was zu nichtlinearen thermodynamischen Systemgleichungen führt.
- Der Wärmetransport innerhalb des Türblatts erfolgt primär durch Wärmeleitung. Zu einem gewissen Teil wird die Wärme durch Strahlung und über die Luft im Türblattinneren zur Schachtseite transportiert. Somit sind auch hier nichtlineare Wärmestrahlungseffekte zu berücksichtigen.
- Beim Wärmetransport über die Luft spielt die Konvektion eine wesentliche Rolle. Um diesen Vorgang ebenfalls korrekt zu simulieren, kann zusätzlich zu den Disziplinen Thermodynamik und Strukturmechanik zusätzlich noch eine Strömungssimulation erforderlich werden.

Die multidisziplinäre Analyse eines einzelnen Entwurfs ist somit mit einem hohen numerischen Aufwand verbunden. Dadurch wird der gesamte Optimierungsaufwand maßgeblich bestimmt.

¹Eine detaillierte Formulierung des Optimierungsproblems sowie dessen Lösung ist in Abschnitt 7.1 zu finden.



(a) Berechnete Temperaturverteilung für das gesamte Türblatt (hier: Vorder- und Rückseite) als Vorgabe für die Strukturmechanik (b) Verformung der Struktur aufgrund des Temperaturfelds

Abbildung 2.2: Hierarchischer Ablauf einer FE-Berechnung für eine thermisch-mechanisch gekoppelte Problemstellung

2.1.2 Strömungs-Struktur-Kopplung im Entwurfsprozess

Auch die Analyse von Problemen mit Strömungs-Struktur-Kopplung ist charakterisiert durch einen hohen Berechnungsaufwand bezüglich der meist nicht hierarchisierbaren Interaktion von Aerodynamik und Strukturmechanik. Ein anschauliches Beispiel hierzu ist die Berechnung des aeroelastischen Gleichgewichtszustands bei einem Tragflügel. Die Systemgleichungen hierfür sehen (vereinfacht) wie folgt aus:

$$\mathbf{f}_{strukt}(\mathbf{u}_{strukt}, \mathbf{p}_{aero}) = \mathbf{0} \quad (2.4a)$$

$$\mathbf{f}_{aero}(\mathbf{u}_{strukt}, \mathbf{p}_{aero}) = \mathbf{0} \quad (2.4b)$$

Hierbei rufen die über die Druckverteilung \mathbf{p}_{aero} gegebenen aerodynamischen Kräfte eine Verformung \mathbf{u}_{strukt} – insbesondere auch eine Verwindung – des Tragflügels nach (2.4a) hervor. Diese wiederum verändert die Umströmung des Tragflügels und damit die mittels Strömungssimulation (2.4b) berechnete Druckverteilung. Zur Lösung von (2.4) stehen zwei unterschiedliche Lösungsstrategien zur Auswahl, die weiter unten in diesem Abschnitt beschrieben und diskutiert werden.

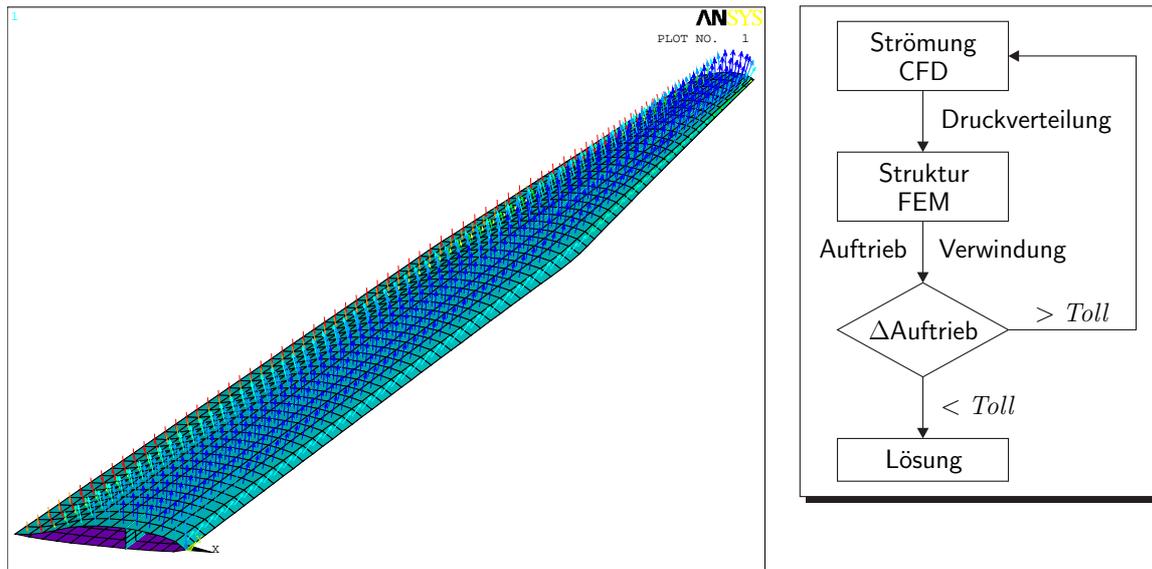


Abbildung 2.3: Berechnung des aeroelastischen Gleichgewichtszustands bei einem Tragflügel (links: Druckverteilung – rechts: Flussdiagramm des klassischen, iterativen Lösungsverfahrens)

Abbildung 2.3 zeigt die Druckverteilung auf den Tragflügel eines Hochleistungsseglflugzeugs sowie ein Flussdiagramm des klassischen, iterativen Lösungsverfahrens (siehe unten). Der Verlauf der Auftriebskraft während der einzelnen Iterationsschritte ist in Abbildung 2.4 dargestellt. Der Ablauf zur Berechnung eines aeroelastischen Gleichgewichtszustands mit vorgegebener Auftriebskraft ist folgendermaßen:

- Zu Beginn jedes Iterationsschritts wird der unverformte Tragflügel mit einem vorgegebenen Winkel angestellt.
 - Die sich aus der Strömungsanalyse ergebende Druckverteilung wird dann auf das Strukturmodell aufgebracht.
 - Mit der gegebenen Druckverteilung ergibt sich die Verformung der Struktur.
 - Für den verformten Tragflügel wird eine neue Druckverteilung ermittelt und wieder auf das Strukturmodell aufgebracht usw. ...
 - Ändert sich die Verformung nur noch geringfügig, so ist ein aeroelastischer Gleichgewichtszustand erreicht.
- Je nachdem, ob die erreichte Auftriebskraft unter oder über dem gewünschten Wert liegt, wird der Anstellwinkel an der Flügelwurzel vergrößert oder reduziert. Entspricht die Auftriebskraft im Gleichgewichtszustand dem geforderten Wert, so ist die Lösung konvergiert.

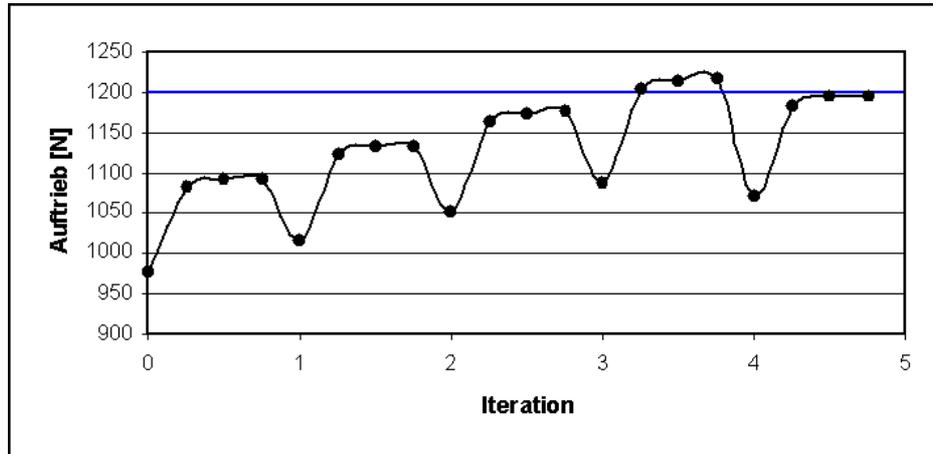


Abbildung 2.4: Iterativer Verlauf bei der Ermittlung eines aeroelastischen Gleichgewichtszustands mit vorgegebener Auftriebskraft

Bei einem Tragflügel im aeroelastischen Gleichgewichtszustand ist neben der Biegeverformung meist auch eine Verwindung der Tragflügelstruktur zu verzeichnen. Diese führt dazu, dass die aerodynamischen Profile in bestimmten Abschnitten des Flügels nicht innerhalb ihres optimalen Anstellwinkelbereichs angeströmt werden. Auch kann es vorkommen, dass Teile des Tragflügels durch die Torsionsverformung einen reduzierten Anstellwinkel besitzen. Um eine erforderliche Auftriebskraft zu erzeugen oder beizubehalten, muss das gesamte Flugzeug mit einem größeren Winkel gegen die Strömung angestellt werden. Die genannten Effekte und Konsequenzen einer Verwindung des Tragflügels führen fast ausschließlich zu einer Erhöhung des Strömungswiderstands.

Um den Widerstand zu minimieren, kann die Struktur des Tragflügels so verändert werden, dass die Verwindung im aeroelastischen Gleichgewichtszustand möglichst klein ist. Die zugehörige Optimierungsaufgabe lautet mathematisch formuliert:²

$$\text{Minimiere} \quad \Delta\alpha(\mathbf{u}_{\text{strukt}}) \quad (2.5a)$$

$$\text{mit } \mathbf{u}_{\text{strukt}} \text{ aus} \quad \mathbf{f}_{\text{strukt}}(\mathbf{x}_{\text{strukt}}, \mathbf{u}_{\text{strukt}}, \mathbf{p}_{\text{aero}}) = \mathbf{0} \quad (2.5b)$$

$$\text{und} \quad \mathbf{f}_{\text{aero}}(\mathbf{u}_{\text{strukt}}, \mathbf{p}_{\text{aero}}) = \mathbf{0} \quad (2.5c)$$

$$\text{so, dass} \quad \mathbf{x}_{\min} \leq \mathbf{x}_{\text{strukt}} \leq \mathbf{x}_{\max} \quad (2.5d)$$

Zur Lösung dieser Optimierungsaufgabe müssen die Strukturparameter $\mathbf{x}_{\text{strukt}}$ innerhalb konstruktionsbedingter Grenzen variiert und der sich damit einstellende aeroelastische Gleichgewichtszustand berechnet werden. Der Optimierungsaufwand wird dabei vorrangig durch das Lösungsverfahren für die Strömungs-Struktur-Interaktion und die Lösung der strömungsmechanischen Systemgleichungen bestimmt.

²Eine detaillierte Formulierung des Optimierungsproblems sowie dessen Lösung ist in Abschnitt 7.2 zu finden.

Lösungsverfahren für Problemstellungen mit Strömungs-Struktur-Kopplung

Bei der Strömungs-Struktur-Kopplung sind im Wesentlichen zwei unterschiedliche Lösungsstrategien zu nennen:

- Bei der ersten werden beide Teilprobleme iterativ abwechselnd gelöst, wobei die Ergebnisse der einzelnen Disziplinen jeweils mittels kinematischer und kinetischer Randbedingungen übertragen werden. Diese Vorgehensweise wird *klassische Methode* oder auch *schwache Kopplung* genannt. Sie hat den Vorteil, dass verschiedene, für die jeweilige Disziplin besonders geeignete Software-Tools genutzt werden können. Dies ist unter anderem deshalb sinnvoll, da es seitens der Strömungsmechanik eine Vielzahl von unterschiedlichen Modellen und Formulierungen gibt. Diese reichen von einfachen Euler- und Potentialgleichungen bis hin zu Navier-Stokes-Gleichungen und Turbulenzmodellen.

Allerdings ist die Lösung der strömungsmechanischen Systemgleichungen durch spezialisierte Programmcodes einer geschlossenen Lösung des Strömungs-Struktur-Kopplungsproblems wenig dienlich, da hier zum Beispiel instationäre Randbedingungen nicht immer in geeigneter Weise berücksichtigt werden können. Bei dieser Methode ist somit bei transienten Analysen – durch die erforderlichen strengeren Kopplungsbedingungen – mit Konvergenzproblemen und über die Zeitintegration hinweg wachsenden Fehlern zu rechnen.

- Bei der gemischten Euler-Lagrange'schen Betrachtungsweise ist eine *strenge Kopplung* formulierbar, so dass das Gesamtsystem einer geschlossenen Lösung zugänglich wird. Kinematische und kinetische Randbedingungen werden auf der Grenzfläche zwischen Struktur und Fluid erfüllt, indem genau dort von der Lagrange'schen zur Euler'schen Formulierung gewechselt wird. Während sich das numerische Netz bei der in der Strukturmechanik üblichen Finite-Elemente-Methode zusammen mit dem darin modellierten Material verformt (Lagrange'sches System), ist das Netz bei der Euler'schen Betrachtungsweise stationär und das Material (Fluid) strömt durch das Netz. Verformt oder bewegt sich nun die Struktur im Fluid, so ist dies mit einer Deformation des fluiddynamischen Netzes verbunden.

Dieser Problematik wird durch den so genannten ALE-Ansatz (Arbitrary Lagrangian-Eulerian-Ansatz) Rechnung getragen, mit dem definiert wird, dass die Verformung eines numerischen Netzes unabhängig vom damit analysierten Materialfluss ist. Die Lösung der das Gesamtsystem beschreibenden Gleichungen erfolgt letztendlich mittels eines mehrstufigen oder gestaffelten iterativen Lösungsverfahrens (z. B. Multistage Runge-Kutta). Das heißt, die Lösungen der Teilgebiete werden getrennt voneinander bestimmt und mittels einer übergeordneten Iterationsschleife zur Deckung gebracht. Eine analoge Verfahrensweise wurde von Ulrich [1999] in einem FE-Programm erfolgreich realisiert.

In [Gleichmar 1999] wurde ein Programmsystem zur streng gekoppelten Lösung von Problemen der Strömungs-Struktur-Interaktion untersucht und mit der in einem Finite-Elemente-Programm realisierten klassischen Methode verglichen. Daraus kann Folgendes festgehalten werden:

- Die Lösung der mit strenger Kopplung definierten Systemgleichungen ist gerade bei transienten Analysen numerisch effizienter und genauer als ein iteratives Verfahren mit schwach formulierter Kopplung.
- Bei stationären Aufgaben oder Problemstellungen, die kein Höchstmaß an Genauigkeit erfordern, kann beim iterativen Verfahren der Vorteil des nahezu beliebigen Einsatzes unterschiedlicher Strömungsanalyse-Tools zum Tragen kommen. Somit besteht die Möglichkeit, einfachere und damit schnellere 2D- und 3D-Strömungsanalyseverfahren einzusetzen.

Zusammenfassend ist anzumerken, dass bereits mit einfacheren CFD-Verfahren ein beträchtlicher Aufwand zur Lösung der gekoppelten strömungs- und strukturellen Systemgleichungen entsteht, der mit der Komplexität der strömungsmechanischen Systemgleichungen stark zunimmt. So wird auch bei Entwurfsaufgaben mit Strömungs-Struktur-Kopplung der Optimierungsaufwand maßgeblich durch die Lösung der zugrunde liegenden Systemgleichungen bestimmt.

2.1.3 Formulierung und Charakterisierung des allgemeinen multidisziplinären Optimierungsproblems

Das allgemeine multidisziplinäre Optimierungsproblem mit einem oder mehreren Zielen lautet mathematisch formuliert:

$$\text{Minimiere } \mathbf{z}(\mathbf{x}_A, \mathbf{x}_B, \mathbf{x}_{AB}, \mathbf{y}_A, \mathbf{y}_B) \quad \text{mit } \mathbf{z} = z_1, \dots, z_{n_Z} \quad (2.6a)$$

$$\text{so, dass } \mathbf{g}(\mathbf{x}_A, \mathbf{x}_B, \mathbf{x}_{AB}, \mathbf{y}_A, \mathbf{y}_B) \leq 0 \quad \text{mit } \mathbf{g} = g_1, \dots, g_{n_G} \quad (2.6b)$$

$$\text{und } \mathbf{h}(\mathbf{x}_A, \mathbf{x}_B, \mathbf{x}_{AB}, \mathbf{y}_A, \mathbf{y}_B) = 0 \quad \text{mit } \mathbf{h} = h_1, \dots, h_{n_H} \quad (2.6c)$$

$$\text{sowie } \mathbf{f}_A(\mathbf{x}_A, \mathbf{x}_{AB}, \mathbf{y}_A, \mathbf{y}_B) = 0 \quad \text{mit } \mathbf{f}_A = f_{A(1)}, \dots, f_{A(n_A)} \quad (2.6d)$$

$$\text{und } \mathbf{f}_B(\mathbf{x}_A, \mathbf{x}_{AB}, \mathbf{y}_A, \mathbf{y}_B) = 0 \quad \text{mit } \mathbf{f}_B = f_{B(1)}, \dots, f_{B(n_B)} \quad (2.6e)$$

Dabei ist \mathbf{z} der Zielfunktionsvektor, der für den Fall einer Optimierungsaufgabe mit mehreren Zielen auch mehrere Zielfunktionen beinhaltet. In Worten ausgedrückt bedeutet (2.6):

Minimiere die Zielfunktion(en) \mathbf{z} unter Einhaltung der in den Funktionsvektoren \mathbf{g} und \mathbf{h} zusammengefassten Ungleichheits- und Gleichheitsrestriktionen sowie unter Erfüllung der Systemgleichungen aller am Gesamtproblem beteiligten Disziplinen.

Mit $\mathbf{x}^T = [\mathbf{x}_A^T \mathbf{x}_B^T \mathbf{x}_{AB}^T \mathbf{y}_A^T \mathbf{y}_B^T]$ lautet (2.6) kurz geschrieben:

$$\min \left\{ \mathbf{z}(\mathbf{x}) \mid \mathbf{g}(\mathbf{x}) \leq \mathbf{0}, \mathbf{h}(\mathbf{x}) = \mathbf{0}, \mathbf{f}_A(\mathbf{x}) = \mathbf{0}, \mathbf{f}_B(\mathbf{x}) = \mathbf{0} \right\} \quad (2.7)$$

Wie in den beiden vorangegangenen Abschnitten dargestellt wurde, sind multidisziplinäre Optimierungsprobleme in erster Linie dadurch charakterisiert, dass die zugehörigen Systemgleichungen häufig nur mit einem hohen Berechnungsaufwand zu lösen sind. Die Zielfunktion(en) wie auch die Restriktionen besitzen dabei einen meist nichtlinearen Verlauf, was bei der Anwendung der in Abschnitt 2.3 beschriebenen Optimierungsverfahren eine relativ häufige Lösung der Systemgleichungen erforderlich macht. Dies ist vor allem auch darin begründet, dass es bei multidisziplinären Problemstellungen nur in relativ seltenen Fällen möglich ist, einem mathematischen Optimierungsalgorithmus analytisch bestimmte Sensitivitäten zur Verfügung zu stellen. Deshalb muss zur Ermittlung der Gradienten beispielsweise ein Finite-Differenzen-Verfahren eingesetzt werden, das – abhängig von der Zahl der Entwurfsvariablen – zusätzlich eine große Zahl exakter multidisziplinärer Analysen erforderlich macht.

Aufgrund der Nichtlinearität der Ziel- und/oder Restriktionsfunktionen muss häufig auch davon ausgegangen werden, dass die Optimierungsaufgabe nichtkonvex ist. Das heißt, es existieren mehrere lokale Lösungen, von denen nur eine das globale Optimum darstellt. Bei Aufgaben mit Nichtkonvexitäten muss ein zusätzlicher Aufwand betrieben werden, um neben den lokalen Optima möglichst auch das globale Optimum zu ermitteln.

2.2 Das Entwurfsoptimierungssystem MOSES

Zur Lösung der multidisziplinären Optimierungsaufgabe (2.6) gibt es eine Vielzahl unterschiedlicher Optimierungsverfahren. Allgemein ist anzumerken, dass dabei kein Verfahren zur Verfügung steht, das jedes Optimierungsproblem schnell und effizient löst. Je nach „Bauart“ der beteiligten Funktionen ist eine für das jeweilige Problem geeignete Methode anzuwenden. Aus diesem Grund wird am Lehrstuhl für Leichtbau das Entwurfsoptimierungssystem MOSES (Multidisciplinary Optimization of Structures and Electromechanical Systems) entwickelt, das verschiedene Verfahren und Ansätze verknüpft, um eine effiziente Lösung unterschiedlicher Aufgabenstellungen zu ermöglichen. Im Wesentlichen handelt es sich um folgende Ansätze, die in Abschnitt 2.3.7 noch einmal aufgegriffen werden:

- Um die Lösbarkeit der Optimierungsaufgabe (2.6) mit der vielfältigen Gestalt der darin vertretenen Funktionen zu gewährleisten, müssen die verwendeten Optimierungsverfahren sehr allgemeingültig und robust sein. In besonderem Maße gilt jedoch die Forderung nach einer möglichst **hohen Effizienz**, die hauptsächlich durch die Anzahl der zum Erreichen des Optimums erforderlichen exakten Systemanalysen charakterisiert ist.

- Bei den meisten Optimierungsverfahren lassen sich einige der exakten Systemanalysen – meist innerhalb eines Iterationsschritts – voneinander unabhängig und somit gleichzeitig durchführen. Durch diese so genannte **Grobparallelisierung** ist eine enorme Reduktion des Zeitaufwands für eine Optimierung erreichbar.
- **Globale Approximationsverfahren** wie die RSM liefern schnell auswertbare Approximationen, die dann bei den eigentlichen Optimierungsverfahren als Ersatz für die aufwendigen Systemanalysen angewendet werden können.

Im Prinzip ist MOSES ein Applikationsmanager. Kommerzielle Tools in diesem Umfeld sind beispielsweise iSIGHT der Firma Engineous Software Inc. [iSIGHT Homepage] und BOSS quattro der Firma Samtech [BOSS quattro Homepage]. Diese verfügen zwar über eine grafische Benutzeroberfläche, sind aber in ihren Funktionalitäten im Wesentlichen auf das beschränkt, was seitens der Softwarehäuser implementiert wurde.

Den Kern des Entwurfsoptimierungssystems MOSES bildet das mathematische Programmpaket MATLAB [Mathworks 2002]. Dieses stellt mit seinen Grundfunktionalitäten und besonders mit den als Toolbox bezeichneten Erweiterungspaketen (z. B. Optimization Toolbox, Statistics Toolbox, ...) eine sehr gute und flexible Basis für den Applikationsmanager dar. Die wesentlichen Elemente und Funktionalitäten von MOSES wie auch dessen Funktionsprinzip sind in Abbildung 2.5 dargestellt.

- Es werden Funktionalitäten wie Optimierungsalgorithmen (Evolutionäre Algorithmen, mathematische Optimierungsverfahren, ...) zur Verfügung gestellt.³
- Diese können durch implementierte Modellreduktions- und Approximationsverfahren wesentlich unterstützt werden.
- Zu den weiteren Funktionalitäten gehören Möglichkeiten zu Parameterstudien und statistischen Untersuchungen.
- Die Funktionalitäten sind mit der MATLAB eigenen Programmiersprache nahezu beliebig erweiterbar. Auf diese Weise wurde auch die in dieser Arbeit vorgestellte RSM⁴ inklusive der Verfahrensweise zur Zusammenstellung eines bestmöglichen Approximationsmodells⁵ implementiert.
- Eine interne Datenverwaltung speichert alle für die spätere Verarbeitung erforderlichen Daten.
- Die einzelnen Funktionalitäten stellen Daten für die benötigten Analysen zusammen. Der Applikationsmanager bereitet die zugehörigen Simulationsmodelle vor, indem er Eingabeskripte erzeugt oder über Schnittstellen manipuliert.

³Einen Überblick über Optimierungsverfahren gibt Abschnitt 2.3.

⁴Eine umfangreiche Beschreibung der RSM ist in Kapitel 3 zu finden.

⁵Die Verfahrensweise zur Zusammenstellung eines bestmöglichen Approximationsmodells ist in Kapitel 5 dargestellt.

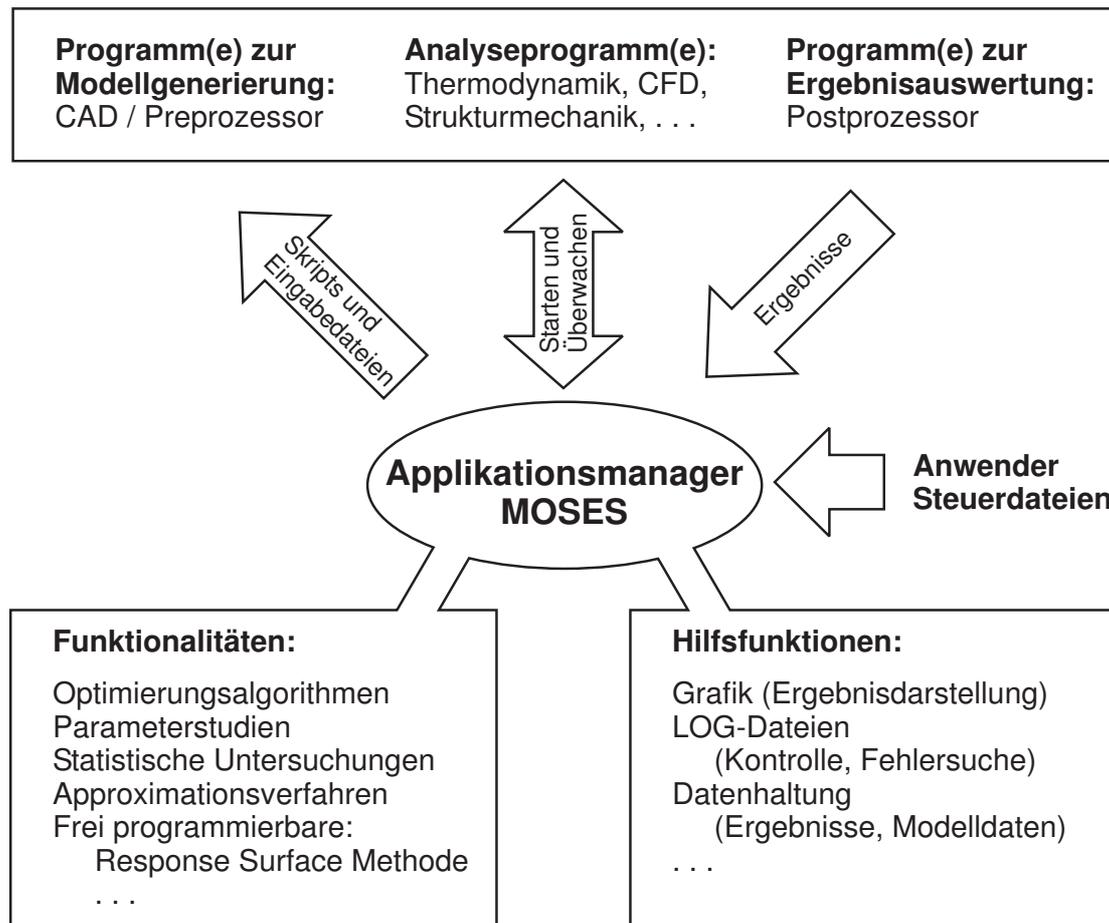


Abbildung 2.5: Funktionsprinzip des Applikationsmanagers MOSES

- Sind alle Daten für die Analysen zusammengestellt, werden diese parallel auf einem PC-Cluster⁶ durchgeführt und deren Fortschritt überwacht. Im Falle multidisziplinärer Analysen übernimmt der Applikationsmanager auch die Daten- und Prozesskoordination zu den disziplinspezifischen Simulations-Tools.
- Nach dem Abschluss der Analysen werden die relevanten Ergebnisse ausgelesen. Dies kann einerseits aus Dateien erfolgen, die speziell dafür vom Analyse-Tool erzeugt wurden und von vornherein nur die relevanten Daten enthalten. Andererseits kann auch über Schnittstellen auf die Ergebnisdateien der Analyse-Tools zugegriffen werden, falls deren Binärformat bekannt ist.
- Zur Visualisierung der Ergebnisse stehen Hilfsroutinen für die Erzeugung grafischer Ausgaben zur Verfügung.

⁶Der Aufbau und die Konfiguration einer solchen aus vielen Einzelrechnern zusammengesetzten Rechanlage wie auch deren Steuerung über MOSES ist in Kapitel 6 beschrieben.

2.3 Optimierungsalgorithmen und Lösungsverfahren

Die Aufgabe eines Optimierungsalgorithmus ist, eine möglichst zielgerichtete, effiziente und robuste Verbesserung der Entwurfsvariablen so zu generieren, dass dabei den Anforderungen eines Optimierungsproblems wie (2.6) entsprochen wird. Dabei gibt es Verfahren, die eine direkte Lösung des restringierten Optimierungsproblems ermöglichen und andere, die lediglich nichtrestringierte Probleme lösen. Besitzt das Optimierungsproblem mehrere Ziele, so muss dieses für die meisten Verfahren in geeigneter Weise aufbereitet werden.

2.3.1 Behandlung mehrerer Ziele

Wie bereits bei der Formulierung der allgemeinen multidisziplinären Optimierungsaufgabe erwähnt, können auch mehrere Zielfunktionen im Zielfunktionsvektor aufgelistet sein. Die Lösung dieser so genannten Vektoroptimierungsaufgabe wird meist dadurch erschwert, dass die einzelnen Werte der Zielfunktionen in gewissen Bereichen des Entwurfsraums gegenläufig sein können. Dies führt zu einem Entscheidungskonflikt, so dass es meist kein einzelnes Optimum gibt. Eine wesentliche Eigenschaft der Lösung der Vektoroptimierungsaufgabe muss dann die so genannte *Pareto-Optimalität* sein. Eine Lösung ist pareto-optimal, wenn durch eine Variation der Entwurfsvariablen keines der Zielkriterien besser erfüllt werden kann, ohne damit ein anderes schlechter zu erfüllen. Die Menge dieser optimalen Zielkompromisse wird *Pareto-Front* genannt. Dabei ist es auch möglich, dass diese mehrfach unterteilt ist und nicht nur einen einzigen Bereich des Entwurfsraums eingrenzt.

Um bei der Mehrzieloptimierung auch Verfahren zur Optimierung von Aufgaben mit nur einem Ziel verwenden zu können, wird die Vektoroptimierungsaufgabe in eine skalare Aufgabe transformiert:

- Eine Möglichkeit ist die Einführung von Wichtungsfaktoren, mit denen die einzelnen Zielfunktionen zu einer **Kostenfunktion** $u(\mathbf{x})$

$$u(\mathbf{x}) = \sum_{i=1}^{n_z} w_i z_i(\mathbf{x})$$

aufsummiert werden, die dann als Zielfunktion dient. Problematisch dabei ist, dass die Optimierung bei gewählten Wichtungsfaktoren nur einen Punkt auf der Pareto-Front zur Lösung haben kann. Um also eine globale Sicht auf den im Vorfeld meist nicht bekannten Verlauf der Pareto-Front zu gewinnen, müssen mehrere Optimierungsläufe mit unterschiedlichen Wichtungsfaktoren durchgeführt werden. Ist die Dauer eines Optimierungslaufs bereits sehr lang, so ist eine entsprechend feine „Abtastung“ der Pareto-Front aus Zeitgründen nahezu nicht mehr möglich.

- Bei der **Restriktionsformulierung** werden alle Zielfunktionen bis auf eine den Ungleichheitsrestriktionen $\mathbf{g}(\mathbf{x})$ zugeordnet, indem deren Werte nach oben beschränkt werden. Als verbleibende skalare Zielfunktion kann zum Beispiel jenes Optimierungsziel gewählt werden, dessen besondere Wichtigkeit von vornherein feststeht. Neben dieser Festlegung ist vor allem auch die Wahl der Schranken der anderen Ziele sehr schwierig. Darüber hinaus gestattet die Restriktionsformulierung eine weit weniger strukturierte Untersuchung der Pareto-Front als mit der Transformation auf eine Kostenfunktion.
- Die **Abstandsfunktionsformulierung** hat zum Ziel, den Abstand des Vektors der Zielfunktionswerte von einem angestrebten Zielfunktionsvektor zu minimieren. Dabei werden die einzelnen Abstände der Zielfunktionswerte von ihrem gewünschten Wert mit einem Wichtungsfaktor multipliziert und aufsummiert. Der angestrebte Zielfunktionsvektor kann dabei aus den Optima der einzelnen Ziele gebildet werden oder er wird mehr nach ingenieurmäßigen Gesichtspunkten vorgegeben.

Andere Verfahrensweisen zur Lösung von Optimierungsaufgaben mit mehreren Zielen sind interaktive Lösungsverfahren (siehe z. B. [Baier 1994]). Eine weitere Möglichkeit zur Behandlung von Optimierungsaufgaben mit mehreren Zielen ist der Einsatz von Evolutionären Algorithmen (siehe Abschnitt 2.3.3).

2.3.2 Straffunktionsverfahren

Um die Berücksichtigung von Restriktionen auch bei Optimierungsverfahren zur Lösung nichtrestringierter Problemstellungen zu ermöglichen, können Straffunktionsverfahren eingesetzt werden. Dabei wird das restringierte Problem (2.6) sukzessive durch Unterprobleme ersetzt, bei denen die Nebenbedingungen (2.6b) und (2.6c) als „Strafkosten“ in die Zielfunktion eingehen. Allgemein lässt sich dies für Optimierungsprobleme mit einer Zielfunktion wie folgt anschreiben:

$$\min_{\mathbf{x} \in \mathbb{D}} \left\{ z(\mathbf{x}) + \gamma(\mathbf{g}(\mathbf{x}), \mathbf{h}(\mathbf{x}), \boldsymbol{\alpha}^{(k)}); \quad k = 1, 2, \dots \right\} \quad (2.8)$$

Für die Straffunktion γ und die Wahl der Strafparameter $\boldsymbol{\alpha}^{(k)}$ gibt es prinzipiell drei Möglichkeiten:

- Bei der **Barrierefunktionsmethode (BFM)** soll die Barrierefunktion γ die Zielfunktion am Rand des zulässigen Bereichs derart erhöhen, dass ein Verlassen desselben praktisch unmöglich wird. Damit sind die Lösungen $\mathbf{x}^{(k)}$ stets zulässige Punkte des ursprünglichen Problems.

Mögliche Ansätze für γ bei Ungleichheitsrestriktionen sind:

$$\gamma(\mathbf{g}(\mathbf{x}), \alpha) = \alpha \cdot \sum_{j=1}^{n_G} \frac{1}{g_j(\mathbf{x})} \quad \text{oder} \quad \gamma(\mathbf{g}(\mathbf{x}), \alpha) = \alpha \cdot \sum_{j=1}^{n_G} (-\ln g_j(\mathbf{x}))$$

Der Startvektor $\mathbf{x}^{(1)}$ muss zulässig sein. Der Strafparameter $\alpha > 0$ wird bei jedem Iterationsschritt verkleinert.

Nachteilig kann sich auswirken, dass die Barrierefunktionsmethode keine Gleichheitsrestriktionen verarbeiten kann und die Lösung der Unterprobleme (2.8) gegebenenfalls numerisch schlecht konditioniert ist.

- Statt eine Barrierefunktion γ zu wählen, welche Restriktionsverletzungen erst gar nicht zulässt, wird bei der **Methode der äußeren Straffunktion** die Nichterfüllung einer Nebenbedingung „bestraft“ mit:

$$\gamma(\mathbf{g}(\mathbf{x}), \mathbf{h}(\mathbf{x}), \alpha) = \alpha \cdot \left(\sum_{j=1}^{n_G} \min(0, g_j(\mathbf{x}))^2 + \sum_{j=1}^{n_H} (0, h_j(\mathbf{x}))^2 \right)$$

In diesem Fall sorgt eine schrittweise Erhöhung des Strafparameters $\alpha > 0$ dafür, dass Restriktionsverletzungen im Laufe der Iterationsschritte immer stärker bestraft werden.

Diese Methode hat den Vorteil, dass keine zulässigen Startpunkte bekannt sein müssen und sowohl Ungleichheitsrestriktionen als auch Gleichheitsrestriktionen behandelt werden können. Nachteilig ist jedoch, dass die Lösungen der Unterprobleme (2.8) – da vom unzulässigen Bereich her angenähert – für (2.6) nicht zulässig sind.

- Die **Methode der erweiterten Lagrangefunktion** gleicht die schlechte Kondition der beiden bisher genannten Methoden durch Verwendung einfacherer Straffunktionen aus, die allerdings weitere Informationen benötigen. Dabei wird ausgehend von der Lagrangefunktion der Optimierungsaufgabe

$$L(\mathbf{x}, \boldsymbol{\lambda}_{(g)}, \boldsymbol{\lambda}_{(h)}) = z(\mathbf{x}) - \sum_{j=1}^{n_G} \lambda_{(g)j} g_j(\mathbf{x}) - \sum_{j=1}^{n_H} \lambda_{(h)j} h_j(\mathbf{x})$$

folgender Ansatz gemacht:

$$\begin{aligned} \gamma(\mathbf{h}^*(\mathbf{x}), \alpha) &= - \sum_{j=1}^{n_H+n_G} \lambda_{(h^*)j} h_j^*(\mathbf{x}) + \frac{1}{\alpha^{(j)}} \sum_{j=1}^{n_H+n_G} (h_j^*(\mathbf{x}))^2 \\ \text{mit } h_j^*(\mathbf{x}) &= h_j(\mathbf{x}) && (j = 1, \dots, n_H) \\ \text{und } h_{n_H+j}^*(\mathbf{x}) &= g_j(\mathbf{x}) - c_j^2 && (j = 1, \dots, n_G) \end{aligned}$$

Zu erkennen ist, dass die Ungleichheitsrestriktionen mit Hilfe der Schlupfvariablen c_j in Gleichheitsrestriktionen umgewandelt werden.

Alle dargestellten Straffunktionsverfahren haben die negative Eigenschaft, dass sie sehr viele Systemanalysen und Gradientenberechnungen benötigen. Sind die Systemanalysen sehr aufwendig und zudem keine daraus ableitbaren analytischen oder semi-analytischen Gradienten verfügbar, wie es bei multidisziplinären Aufgabenstellungen

häufig der Fall ist, so ist schnell ein Punkt erreicht, bei dem ein auf dem Straffunktionsverfahren basierender Algorithmus nicht mehr praktikabel oder zumindest äußerst ineffizient ist.

2.3.3 Suchmethoden und gradientenfreie Verfahren

Unter dem Begriff Suchmethoden versteht man Algorithmen, die das Optimum nur mittels des wiederholt berechneten Zielfunktionswerts $z(\boldsymbol{x})$ bestimmen, ohne dabei auf die Berechnung der Gradienten der Zielfunktion angewiesen zu sein. Aus dem Verzicht auf die Nutzung von Ableitungsinformationen ergibt sich zwangsläufig der Nachteil, dass Suchverfahren nicht an die Effizienz gradientenbasierter Verfahren heranreichen. Je nach Wahl der Steuerparameter können sie jedoch als robust eingestuft werden. Bleibt der Suchraum auf den zulässigen Bereich beschränkt, ist auch die Lösung restringierter Aufgaben möglich.

Vollständige Enumeration

Bei der vollständigen Enumeration wird der n_V -dimensionale Suchraum (Definitionsbereich der Entwurfsvariablen) mit einem Suchgitter überzogen. Bei dieser Methode ist klar, dass bereits eine sehr moderate Zahl von Entwurfsvariablen zu einem extremen Rechenaufwand führt. Dieser wächst exponentiell mit der Zahl der Gitterpunkte pro Entwurfsvariable. Da der gesamte Definitionsbereich untersucht wird, kann die Methode dazu dienen, einen oder mehrere erfolgversprechende Startvektoren für mathematische Optimierungsverfahren zu ermitteln.

Sequenzielle Suchverfahren

Die hierzu gehörenden Suchverfahren unterscheiden sich in der Methodik, mit der die Suchrichtung in den einzelnen Iterationsschritten festgelegt ist bzw. bestimmt wird. Während die *Sukzessive Variation*, auch Gauß-Seidel-Strategie genannt, einen verbesserten Punkt jeweils nur in einer Dimension durch achsenparallele Schritte sucht, bietet das *Hooke-Jeeves-Verfahren* eine vergleichsweise effiziente Suche, indem die Suchrichtung nicht an die Achsen des Entwurfsvariablenraums gekoppelt ist, sondern dem Verlauf der Zielfunktion angepasst wird. Beide Verfahren sind jedoch für restringierte Optimierungsprobleme ohne den zusätzlichen Einsatz eines Straffunktionsverfahrens nicht anwendbar, was die ohnehin mäßige Effizienz weiter vermindert.

Monte-Carlo-Verfahren

Beim Monte-Carlo-Verfahren werden die untersuchten Punkte nicht nach einem strengen Muster im Definitionsbereich der Entwurfsvariablen verteilt, sondern gewissermaßen zufällig gesetzt. Dabei werden die Punkte um einen Ausgangspunkt gestreut, wobei die einzelnen Koordinaten dieses Punkts die Mittelwerte für eine Wahrscheinlichkeitsverteilung (z. B. Gauß'sche Normalverteilung, Gleichverteilung, ...) bestimmen. Unter Einhaltung der gewählten Verteilungsform und Mittelwerte wird dann ein Satz von Entwurfsvarianten erzeugt, die unabhängig voneinander analysiert werden können.

Im Anschluss an die Systemanalysen wird der beste gültige Entwurf selektiert und um diesen ein weiterer Satz von zu untersuchenden Designs gebildet. Dabei kann auch die Art oder Form der zu berücksichtigenden Zufallsverteilung variiert werden, indem beispielsweise deren Streubreite sukzessive verkleinert wird. Der beschriebene Suchvorgang wird solange wiederholt, bis ein vorgegebenes Abbruchkriterium erreicht ist. Dies kann zum Beispiel die Vorgabe der Anzahl an Iterationsschritten, die Einhaltung einer vorgegebenen Rechendauer oder das Erreichen bestimmter Zielvorgaben sein.

Wie alle Suchverfahren ist auch das Monte-Carlo-Verfahren in Bezug auf den Rechenaufwand nicht sehr effizient, wenn die Lösung nur auf einem Rechner ermittelt wird. Nutzt man jedoch die Eigenschaft, dass die Systemanalysen aller Entwürfe eines Berechnungssatzes unabhängig voneinander auf vielen Rechnern parallel durchgeführt werden können, kann der Nachteil der erforderlichen großen Zahl von Systemanalysen gut ausgeglichen werden. Die Anzahl der in einem Berechnungssatz enthaltenen Entwürfe ist zudem frei wählbar. Somit kann die Zahl der Analysen als ein ganzzahliges Vielfaches der zur Verfügung stehenden Prozessoren gewählt werden und eine Anpassung an die verfügbaren Kapazitäten der Rechenanlagen erfolgen.

Die Methode ist sehr robust, da fehlende Berechnungsergebnisse, die zum Beispiel durch Ausfall einzelner Rechenknoten oder nicht-konvergierten Lösungen bedingt sind, den Suchprozess nicht blockieren. Je nachdem, ob es sich bei der „verlorenen“ Analyse um einen bezüglich des Optimierungsziels guten – und damit wichtigen – Entwurf handelt, wird die für das Optimierungsproblem gefundene Lösung mehr oder weniger stark beeinflusst.

Evolutionäre Algorithmen

Die eng miteinander verwandten, oft unter dem Begriff „Evolutionäre Algorithmen“ zusammengefassten Evolutionsstrategien und Genetische Algorithmen sind höher entwickelte Suchverfahren, die sich am Vorbild der Evolution und Genetik orientieren. Die Entwurfsparameter werden dabei – ähnlich wie Gene auf den Chromosomen – als String oder anderweitig kodiert zusammengefasst. Die Verfahren arbeiten mit einer Menge (*Population*) von Punkten (*Individuen*) im Suchraum, die dann entsprechend dem Vorbild der Natur einer Rekombination, Mutation und Selektion (Survival of

the fittest) unterworfen werden. Die am besten geeigneten Individuen einer Population werden als „*Eltern*“ einer neuen Generation selektiert. Die Entwurfsvariablen der als „*Kinder*“ bezeichneten Individuen der Folgegeneration werden aus den Genen der ausgewählten Eltern durch *Rekombination* gebildet. Dies erfolgt bei den Evolutionsstrategien durch eine Vermischung der Eigenschaften und bei den Genetischen Algorithmen durch Rekombination, wodurch auch die Behandlung diskreter Variablen möglich wird. Um nicht vollständig auf die mit der zufällig verteilten Ausgangspopulation festgelegten Eigenschaften beschränkt zu bleiben, wirkt zudem noch das Element der *Mutation*. Diese ermöglicht die vor allem in frühen Generationen wichtige Erschließung neuer Ausprägungen der Entwurfsvariablen und damit bisher nicht untersuchter Bereiche des Suchraums. Die Varianten der Evolutionären Algorithmen unterscheiden sich im Wesentlichen durch die Art der Selektion der Eltern, der Rekombination (intermediär oder diskret, mehr als nur zwei Eltern, ...) sowie der Variation der Mutationsmechanismen im Laufe der Generationen.

Zusammenfassend betrachtet stellen Evolutionsstrategien und Genetischen Algorithmen eine Klasse einfacher, universeller und robuster Optimierungsverfahren dar. Sie eignen sich insbesondere für Probleme mit diskreten Variablen und unterbrochenen Suchräumen sowie für kombinatorische Probleme. Bezüglich des Optimierungsaufwands und der damit verbundenen Effizienz sind zu den Monte-Carlo-Verfahren analoge Eigenschaften zu nennen. Mit der Größe einer Population ist die Anzahl der erforderlichen, parallel zueinander durchführbaren Systemanalysen frei wählbar. Somit können die zur Verfügung stehenden Rechenressourcen bei der Festlegung der betreffenden Steuerparameter des Algorithmus einbezogen werden.

Mit geeigneter Wahl der für die Mutation verantwortlichen Steuerparameter lässt sich vermeiden, dass der Algorithmus die Population schnell auf ein vermeintliches Optimum eingrenzt. Somit ist das Verfahren auch sehr gut für Probleme mit zahlreichen lokalen Optima geeignet, da suboptimale Bereiche wieder verlassen werden können.

Bei der Mehrzieloptimierung besitzen Evolutionäre Algorithmen Vorteile gegenüber der bei anderen Optimierungsverfahren eher üblichen, starren Zusammenfassung der Zielfunktionen über Wichtungsfaktoren. Bei entsprechender Wahl der Steuerparameter nähern sich die Individuen von Generation zu Generation der Front der pareto-optimalen Lösungen an. Hierzu wird am Lehrstuhl für Leichtbau ein spezieller Algorithmus mit dem Namen GAME (Genetic Algorithm for Multidisciplinary Engineering) entwickelt (siehe z. B. [Langer u. a. 2002] und [Pühlhofer u. a. 2003]).

2.3.4 Gradientenverfahren und Newton-Verfahren

Eine Steigerung der Effizienz lässt sich erreichen, wenn zur Bestimmung der Suchrichtung und der zugehörigen Schrittweite neben dem reinen Funktionswert noch zusätzliche Informationen über die Zielfunktion genutzt werden. Bei den Gradientenverfahren ist dies die erste Ableitung der Zielfunktion, wohingegen bei den Newton-Verfahren noch die zweite Ableitung hinzugenommen wird.

Gradientenverfahren

Die **Methode des steilsten Abstiegs** (Method of Steepest Descent) macht sich die Eigenschaft des Gradientenvektors $\nabla z(\mathbf{x}) = \partial z / \partial \mathbf{x}$ zunutze, in die Richtung maximaler Zunahme des Zielfunktionswerts zu zeigen. Als Suchrichtung wird die entgegengesetzte Richtung verwendet, so dass diese in Richtung des steilsten Abstiegs der Zielfunktion weist. In dieser Richtung wird mit veränderlicher Schrittweite solange gesucht, bis die eindimensionale Liniensuche ein Minimum gefunden hat. Dort wird erneut die Richtung des steilsten Abstiegs bestimmt und in dieser Richtung weiter gesucht. Diese Iteration wird solange durchlaufen, bis im erreichten Punkt kein Abstieg mehr möglich ist.

Die Methode neigt dazu, das Ziel mit wachsender Nichtlinearität der Zielfunktion auf immer größer werdenden Umwegen zu erreichen. Durch das Einbeziehen von Informationen aus dem vorangegangenen Iterationsschritt kann die **Methode der konjugierten Richtungen** (Method of Conjugate Gradients) die Effizienz steigern, indem eine neue Suchrichtung berechnet wird mit:

$$\mathbf{s}^{(k+1)} = -\nabla z^{(k)} + \frac{|\nabla z^{(k)}|^2}{|\nabla z^{(k-1)}|^2} \cdot \mathbf{s}^{(k)}$$

Newton-Verfahren

Im Gegensatz zu den Gradientenverfahren werden bei den Newton-Verfahren neben den Ableitungen erster Ordnung auch die in der Hesse-Matrix $\mathbf{H}^{(k)} = \mathbf{H}(\mathbf{x}^{(k)})$ enthaltenen Ableitungen zweiter Ordnung verwendet. Dabei wird die Zielfunktion an der Stelle $\mathbf{x}^{(k)}$ über eine Taylor-Reihe bis zu Termen zweiter Ordnung approximiert. Im analytisch bestimmbar Minimum dieser quadratischen Näherung wird erneut eine Taylor-Reihen-Entwicklung mit Termen bis zur zweiten Ordnung durchgeführt und deren Minimum bestimmt. Dieser Vorgang wird solange wiederholt, bis eine zufriedenstellende Lösung gefunden ist. Wie auch bei den Gradientenverfahren kann eine zwischengeschaltete eindimensionale Liniensuche die Methode ergänzen und somit das Konvergenzverhalten verbessern.

Die Bestimmung der Hesse-Matrix bedeutet einen deutlich höheren Aufwand pro Iterationsschritt als bei den Gradientenverfahren. Allerdings können mit den Newton-Verfahren rein quadratische Probleme in einem Schritt gelöst werden und auch sonst werden weitaus weniger Iterationen benötigt. Bei multidisziplinären Problemen der Strukturoptimierung mit ihren aufwendigen Systemanalysen ist eine exakte Berechnung der Hesse-Matrix häufig nicht von Vorteil, da der erforderliche Berechnungsaufwand nicht immer in einem günstigen Verhältnis zum erreichten Nutzen steht.

Aus diesem Grund arbeiten die **Quasi-Newton-Verfahren** mit der approximierten Inversen der Hesse-Matrix, die bei jedem Iterationsschritt korrigiert wird. Die am häufigsten verwendete Vorschrift zur Korrektur der Hesse-Matrix ist die Broydon-Fletcher-Goldfarb-Shanno (BFGS)-Formel, die auch bei anderen Algorithmen und in vielen Softwarepaketen zum Einsatz kommt (siehe z. B. [MATLAB Optimization Toolbox 2002]).

2.3.5 Direkte Verfahren zur Lösung restringierter Optimierungsaufgaben

Die im Folgenden beschriebenen Verfahren optimieren keine Unterprobleme, deren Lösungen dann gegen das Optimum des eigentlichen Problems konvergieren, sondern behandeln direkt das restringierte Problem. Hierbei geht es neben der Minimierung der Zielfunktion vor allem darum, die Einhaltung der Restriktionen zu gewährleisten. Wichtigster Bestandteil dieser Verfahren ist die Ermittlung geeigneter Suchrichtungen.

- Einer der bekanntesten Algorithmen ist die **Methode der zulässigen Richtungen** (Method of Feasible Directions, MFD). Hierbei wird die Suchrichtung im Sektor des brauchbaren und zulässigen Bereichs des Designraums ermittelt, also aus der Menge aller Richtungen \mathbf{s} , für die gilt:

$$\nabla z^T \mathbf{s} \leq 0 \quad \text{und} \quad \nabla g_j^T \mathbf{s} \geq 0, \quad j \in I_{akt}$$

Die erste Bedingung sorgt für die Minimierung der Zielfunktion, während die übrigen Bedingungen eine Verletzung der aktiven Restriktionen verhindern sollen.

Die **Modifizierte Methode der zulässigen Richtungen** (Modified Method of Feasible Directions, MMFD) macht sich die Eigenschaften der nachfolgend beschriebenen Methode der generalisierten Gradienten zunutze, um zu vermeiden, dass das Optimum nur auf größeren Umwegen erreicht wird. Der durch dieses Verfahren entstehende Lösungspfad verlässt den Rand des zulässigen Gebiets nur, wenn das Optimum innerhalb des gültigen Bereichs liegt. Die MMFD stellt eine sehr effiziente Methode zur direkten Lösung restringierter Optimierungsaufgaben dar.

- Bei der erweiterten bzw. **generalisierten Methode der reduzierten Gradienten** (Generalized Reduced Gradient Method, GRGM) werden die Ungleichheitsrestriktionen durch Einführung von Schlupfvariablen in Gleichheitsrestriktionen umgewandelt. Die Entwurfsvariablen werden dann in gebundene und freie Variablen unterteilt, wobei bezüglich der freien Variablen der so genannte generalisierte Gradient gebildet wird. Während der Liniensuche entlang dieses Gradienten erfolgt jeweils ein „Korrekturschritt“, welcher gegebenenfalls entstandene Verletzungen von Restriktionen kompensiert.
- Das wohl bekannteste und bislang effizienteste direkte Verfahren ist die **Methode der sequenziellen quadratischen Programmierung** (Sequential Quadratic Programming, SQP). Die Suchrichtung wird in diesem Fall aus einem Unterproblem mit quadratischer Zielfunktion und linearen Restriktionen ermittelt. Dies erfolgt nicht über eine Approximation der Zielfunktion $z(\mathbf{x})$, sondern mittels der in einer Taylor-Reihe entwickelten Lagrange-Funktion $L(\mathbf{x})$, die für das Unterproblem nach dem quadratischen und für die Restriktionen hinter dem linearen Term abgebrochen wird. Ein Iterationsschritt lautet also $\mathbf{x}^{(k+1)} = \mathbf{x}^{(k)} + \rho^{(k)} \mathbf{s}^{(k)}$, mit $\mathbf{s}^{(k)}$ aus:

$$\min_{\mathbf{s}} \left\{ z(\mathbf{x}^{(k)}) + \mathbf{s}^T \nabla z(\mathbf{x}^{(k)}) + \frac{1}{2} \mathbf{s}^T \mathbf{H}^{(k)} \mathbf{s} \right\}$$

so, dass

$$\begin{cases} g_j(\mathbf{x}^{(k)}) + \mathbf{s}^T \nabla g_j(\mathbf{x}^{(k)}) \geq 0, & j = 1, \dots, n_G \\ h_j(\mathbf{x}^{(k)}) + \mathbf{s}^T \nabla h_j(\mathbf{x}^{(k)}) = 0, & j = 1, \dots, n_H \end{cases}$$

Die Matrix $\mathbf{H}^{(k)}$ ist dabei die Approximation der Hesse-Matrix der Lagrange-Funktion im Punkt $\mathbf{x}^{(k)}$. Auch hier erfolgt ein Update der Hesse-Matrix meist nach der BFGS-Formel.

2.3.6 Approximationsverfahren

Bei den Approximationsverfahren sind prinzipiell zwei Varianten zu unterscheiden. Während bei den *lokalen* Approximationsverfahren die Zielfunktion und Restriktionen durch lokale Näherungen ersetzt werden, liegt der Grundgedanke bei den *globalen* Approximationsverfahren darin, zu einer umfassenderen Sicht bezüglich der Zielfunktion(en) und gegebener Restriktionen zu gelangen. Dazu wird die Approximation bei den globalen Verfahren nicht lokal – in einem Punkt des Entwurfsvariablenraums –, sondern mittels im Definitionsbereich der Entwurfsvariablen verteilter Entwurfsvarianten gebildet.

Lokale Approximationsverfahren

Die Vorgehensweise bei den klassischen lokalen Approximationsverfahren liegt, ähnlich wie bei den Straffunktionsverfahren, in der sukzessiven Lösung von Unterproble-

men. Im Gegensatz zu den Straffunktionsverfahren werden dabei nicht die exakten Funktionen $z(\mathbf{x})$, $\mathbf{g}(\mathbf{x})$, $\mathbf{h}(\mathbf{x})$ benutzt, sondern lediglich lokale Näherungen derselben. Dies führt zu einer beträchtlichen Steigerung der Effizienz, da sich somit die Zahl der benötigten exakten Systemanalysen auf ein geringes Maß reduzieren lässt. Die Teilprobleme werden dann mit Hilfe sehr effizienter Verfahren gelöst. Die lokalen Approximationsverfahren unterscheiden sich in der Art und Weise, wie Zielfunktion und Restriktionen approximiert werden:

- Eine einfache Form der Approximation bietet die **sequenzielle lineare Programmierung** (SLP). Hierzu werden Zielfunktion und Restriktionen an der Stelle $\mathbf{x}^{(k)}$ in Taylor-Reihen entwickelt, die nach den linearen Termen abgebrochen werden. Die somit entstandene lineare Ersatzaufgabe kann dann effizient mit dem Simplex-Verfahren (siehe [Bronstein 1991]) gelöst werden. Eine Konvergenz ist bei der SLP nur dann gewährleistet, wenn das Optimierungsproblem mehr Restriktionen als Optimierungsvariablen besitzt. Ist dies nicht der Fall, kann eine Lösung ermöglicht werden, indem der Betrachtungsbereich durch Einführung zusätzlicher Restriktionen, der so genannten „Move Limits“, eingeschränkt wird. Die zugrunde liegende lineare Repräsentation führt jedoch bei stärkerer Nichtlinearität der eigentlichen Optimierungsaufgabe zu Konvergenzproblemen.
- Die **sequenzielle invers-lineare Programmierung** (SILP) kann die Form der Approximationen gegebenenfalls besser an die Gegebenheiten spezieller Restriktionstypen anpassen. Dies ist gegeben, wenn eine Entwurfsvariable einen umgekehrt proportionalen Einfluss auf den Wert einer Restriktionsfunktion hat. Ein strukturmechanisches Beispiel hierfür ist die Spannung in einem Stab ($\sigma = \frac{N}{A}$), die durch die Normalkraft N und die reziprok eingehende Querschnittsfläche A bestimmt wird. In diesem – nicht verallgemeinerbaren – Idealfall ist zur Lösung nur eine einzige Näherung erforderlich, da diese bereits eine globale Approximation darstellt. Anstelle des invers-linearen Näherungsansatzes sind auch andere, zum Beispiel invers-quadratische Ansätze, denkbar. Die Entscheidung für eine andere als lineare Approximation sollte generell aber nur dann getroffen werden, wenn bekannt ist, in welchem funktionalen Zusammenhang die Optimierungsvariablen in die Zielfunktion oder Restriktionen eingehen. Dies ist im Allgemeinen und besonders bei Aufgaben der multidisziplinären Optimierung eher nicht gegeben ist. In diesen Fällen stellt die SLP das konservativere und damit vorzuziehende Verfahren dar.
- Weitere lokale Approximationsverfahren sind die *Konvexe Linearisierung* (CONLIN) und deren Weiterentwicklung die *Method of Moving Asymptotes* (MMA). Der übergeordnete Begriff dieser und ähnlicher Verfahren ist die **sequenzielle konvexe Approximation**, deren Ziel es ist, konvexe und damit gut lösbare Ersatzprobleme zu generieren. So entscheidet bei der CONLIN das Vorzeichen der Ableitung darüber, ob bezüglich einer Entwurfsvariable direkt oder invers-linear approximiert wird, so dass eine gemischt direkt-inverse Linearisierung erfolgt.

Globale Approximationsverfahren

Der wesentlichste Unterschied der globalen Approximationsverfahren zu den lokalen ist, dass letztere immer nur einen eng begrenzten Bereich um einem Punkt (Startvektor $\mathbf{x}^{(0)}$ oder Zwischenlösungen $\mathbf{x}^{(k)}$) approximieren und sich auf einem Lösungspfad in Richtung des Optimums vorarbeiten. Die globalen Näherungsverfahren basieren hingegen auf einer möglichst umfassenden Sicht auf des Optimierungsproblem. Das Ziel der globalen Approximationsverfahren ist es, eine möglichst im gesamten Definitionsbereich anwendbare, gute und numerisch schnell auswertbare Näherung für die Zielfunktion(en) und Restriktionen zu generieren. Zum Auffinden des Optimums wird einer der vorgestellten Algorithmen verwendet, wobei die Effizienz gegenüber der Forderung nach Allgemeingültigkeit und Robustheit des verwendeten Verfahrens in den Hintergrund rückt.

Es gibt eine Reihe teilweise sehr unterschiedlicher Approximationsverfahren, die im Folgenden erläutert werden.

- Ein beschreitbarer Weg ist die Verwendung **Neuronaler Netze**. Diese „erlernen“ funktionale Zusammenhänge, indem die Parameter der so genannten Transferfunktion der netzartig miteinander verbundenen Knotenpunkte (*Neuronen*) so bestimmt werden, dass aus einem vorgegebenen Entwurfsvariablenvektor \mathbf{x} eine Antwort resultiert, die möglichst nah am gewünschten Wert liegt.

Der Nachteil der Neuronalen Netze ist, dass diese meist mit einer großen Zahl exakter Systemanalysen trainiert werden müssen, um eine gute Näherung zu erreichen. Mit zu wenigen Datenpunkten passiert es nicht selten, dass die approximierte Systemantwort deutlich vom erwarteten Wert abweicht. Zur Sicherstellung einer entsprechenden Prognosegüte wird ein Teil der Datenpunkte nicht zum Trainieren des Neuronalen Netzes, sondern zum Testen der prognostizierten Systemantworten verwendet. Deshalb ist es gängige Praxis, die Testdaten in einer Art Rotationsprinzip mit den Trainingsdaten auszutauschen, was jedoch bei komplexen Neuronalen Netzen und vielen Datenpunkten ein entsprechendes Vielfaches des sonst nur einmal zu betreibenden Trainingsaufwands zur Folge hat.

- Bei den **Interpolationsverfahren** wird ein Satz von Datenpunkten durch eine Linearkombination elementarer Basisfunktionen interpoliert. Der prognostizierte Funktionswert für einen neuen Datenpunkt \mathbf{x}^* wird dabei gebildet aus

$$\hat{y}(\mathbf{x}^*) = \sum_{k=1}^{n_D} a_k \pi_k(\mathbf{x}^*) + \sum_{l=1}^{n_P} b_l \varphi(\mathbf{x}^* - \mathbf{x}_l) \quad (2.9)$$

Hierbei ist $\varphi(\mathbf{x}^* - \mathbf{x}_l)$ die in jedem Stützpunkt zentrierte interpolierende Basisfunktion. Übliche Repräsentationen der *spline-basierten* Interpolationsfunktionen sind lineare, quadratische und kubische Formen sowie deren Abwandlungen.

Die π_k sind Polynomterme maximal G -ten Grades, also der Form $x_1^{d_1} x_2^{d_2} \cdots x_{n_V}^{d_{n_V}}$ (mit $\sum d_i \leq G$). Die Interpolation wird gebildet, indem das Gleichungssystem gelöst wird, das durch Einsetzen aller n_P Stützpunkte $\mathbf{x}_{(p)}$ und deren zugehöriger Stützwerte $y_{(p)}$ in (2.9) entsteht. Die fehlenden n_D Gleichungen werden über Nebenbedingungen für die Koeffizienten b_l formuliert:

$$\sum_{l=1}^{n_P} b_l \pi_k(\mathbf{x}_l) \quad \text{mit } k = 1, \dots, n_D$$

Eine Sonderform der Interpolationsverfahren ist das so genannte *Kriging*, bei dem, ähnlich wie bei der RSM, zusätzlich Elemente der Statistik Anwendung finden. So wird beispielsweise der Korrelation zweier Systemantworten y_a und y_b sowie dem Abstand der erzeugenden Stützpunkte $\|\mathbf{x}_a - \mathbf{x}_b\|$ Rechnung getragen. Für eine detailliertere Beschreibung sei auf [Stein 1999] verwiesen.

Allen Interpolationsverfahren gemein ist die Eigenschaft, dass die damit erzeugte Fläche „stur“ durch die jeweiligen Datenpunkte verläuft. Zwar wird mit der beschriebenen Methodik ein Überspringen zwischen den Datenpunkten vermieden, jedoch ist die mit wachsender Nichtlinearität steigende Anzahl erforderlicher Datenpunkte recht groß. Dies gilt insbesondere dann, wenn auch die Zahl der Entwurfsvariablen groß ist. Ein weiterer Nachteil tritt auf, wenn die Datenpunkte durch Fehler stark verrauscht sind. Hier wird eine Welligkeit der Antwortfläche generiert, die es eigentlich nicht gibt, was das Auffinden eines globalen Optimums durch Suggestion vieler lokaler Optima verhindern kann.

- Bei der **Response Surface Methode** wird der Verlauf einer durch Linearkombination meist linearer und quadratischer Terme erzeugbaren Antwortfläche durch Bestimmung der Linearkoeffizienten so verändert, dass diese möglichst nah an den Datenpunkten vorbei verläuft. Auf diese Weise wird beispielsweise ein „Rauschen“ bei streuenden Datenpunkten oder auch stark variierende Systemantworten geglättet.⁷

Das Bestimmen der Parameter des Approximationsmodells ist mit einem sehr geringen Rechenaufwand verbunden. Die RSM stellt weiterhin eine Vielzahl von Beurteilungskriterien bereit, die ebenfalls eine Auswertung in kürzester Zeit erlauben. Diese dienen der Bewertung eines Approximationsansatzes im Hinblick auf dessen globale und lokale Approximationsfehler, dessen Prognosegüte und die Wichtigkeit einzelner Ansatzterme. Einige der Beurteilungskriterien erlauben eine Bewertung der Datenpunkte, so dass eventuell vorhandene, stark fehlerbehaftete Datenpunkte ausgegrenzt werden können.

Der Nachteil der RSM ist, dass größere Nichtlinearitäten mit den meist verwendeten quadratischen Approximationsmodellen nicht abgebildet werden können. Hierzu wird in Kapitel 5 ein Verfahren vorgestellt, das den Einsatz erweiterter Ansatzterme ermöglicht. Diese werden durch die Lösung eines diskret-nichtlinearen Regressionsproblems bestimmt.

⁷Ein Beispiel hierfür ist in Abschnitt 7.3 zu finden.

2.3.7 Zusammenfassende Diskussion

Effizienz der Optimierungsverfahren

Die Effizienz eines Optimierungsverfahrens ist hauptsächlich durch die Anzahl der zum Erreichen des Optimums erforderlichen exakten Systemanalysen charakterisiert. Den Eigenschaften der in Abschnitt 2.3 beschriebenen Verfahren zufolge sind unter den „klassischen“ Optimierungsverfahren die sequenzielle quadratische Programmierung sowie die sequenzielle konvexe Approximation (speziell die MMA) die geeignetsten. Suchverfahren bieten, rein an der Zahl der erforderlichen Systemanalysen gemessen, die geringste Effizienz.

Grobparallelisierung der Systemanalysen

Prinzipiell sind alle genannten Optimierungsverfahren parallelisierbar. Eine solche Vorgehensweise bietet sich beispielsweise bei der wiederholt erforderlichen Bestimmung der Gradienten für einen Optimierungsalgorithmus an. Durch die Parallelisierbarkeit gewinnen auch Suchverfahren enorm an Bedeutung. So lassen sich beispielsweise bei einem Evolutionären Algorithmus die Systemanalysen der Individuen einer Population in einem Bruchteil der für sequenzielle Analysen benötigten Zeit durchführen (siehe [Baier, Gleichmar u. a. 2001] und [Pühlhofer, Gleichmar u. a. 2002]). Dies gilt in vollem Umfang auch für die globalen Approximationsverfahren.

Generell ist es für die Parallelisierung besser, wenn die erforderlichen Systemanalysen in möglichst wenige Blöcke unterteilt werden müssen, da so die zur Verfügung stehenden Rechenressourcen optimal genutzt werden können. Bei den mathematischen Optimierungsverfahren sind hier Grenzen gesetzt: Zum einen ist eine iterative Vorgehensweise unumgänglich und nur durch die Effizienz des Algorithmus bestimmt. Zum anderen werden zur Bestimmung eines Gradienten mittels finiter Differenzen für n_V Entwurfsvariablen genau $n_V + 1$ Systemanalysen benötigt. Stehen nur n_V Rechner bzw. Prozessoren zur Verfügung, verdoppelt sich die Optimierungsdauer, da die verbleibende Systemanalyse erst im Anschluss an die ersten n_V Analysen durchgeführt werden kann, wofür jedoch die selbe Zeit benötigt wird.

Im Hinblick auf eine Parallelisierung ist es demnach vorteilhaft, wenn ein verwendetes Verfahren nur ein oder wenige Male eine bestimmte Anzahl von Analysen benötigt oder eine Anpassung an die vorhandenen Rechenressourcen erlaubt. Ersteres ist vor allem bei den globalen Approximationsmethoden gegeben. Unter den anderen Verfahren sind es speziell die Suchmethoden „Monte-Carlo-Verfahren“ und die „Evolutionären Algorithmen“, die eine relativ große Flexibilität bezüglich der Zahl der pro Iteration benötigten Systemanalysen besitzen. Dieser Vorteil wird aufgrund der im Vergleich zu anderen Verfahren doch größeren Zahl an Systemanalysen zumindest geschmälert.

Globale Approximationsverfahren

Globale Approximationsverfahren wie die RSM bieten zusammen mit einem der beschriebenen Optimierungsverfahren ein großes Potential zur Reduktion des zeitlichen Optimierungsaufwands. Zwar muss auch hier eine bestimmte Anzahl exakter Systemanalysen durchgeführt werden, die manchmal ähnlich groß oder gar größer als bei einem effizienten mathematischen Optimierungsverfahren sein kann. Die Approximationsverfahren besitzen jedoch eine Reihe vorteilhafter Eigenschaften, die einen gegebenenfalls höheren Aufwand zur Erstellung der Approximationen mehr als ausgleichen:

- **Wiederverwendbarkeit und Erweiterung von Daten aus exakten Systemanalysen**

Die globalen Approximationsverfahren greifen jeweils auf eine Datenbasis zurück, um freie Funktionsparameter eines Approximationsmodells zu bestimmen. Kommen im Laufe eines Entwicklungszyklus weitere Informationen hinzu, so können die zugrunde liegenden Approximationsmodelle angepasst und/oder neu aufgebaut werden. Somit ist eine schrittweise Verbesserung der Approximation leicht möglich, wobei auch Daten aus der Datenbasis entfernt werden können, wenn diese eher in eine falsche Richtung „führen“.

- **Schnelle Auswertbarkeit und wiederholte Optimierungsläufe**

Sowohl bei den rein gradientenbasierten Verfahren als auch bei den Newton-Verfahren kommt der Wahl des Startpunkts eine große Bedeutung zu. Gerade bei Existenz mehrerer lokaler Minima entscheidet dieser darüber, ob der Algorithmus nur ein lokales oder das globale Optimum erreicht. Um sicher zu gehen, muss der Algorithmus an verschiedenen Punkten des Definitionsbereichs der Entwurfsvariablen gestartet werden, was einen enormen Berechnungsaufwand zur Folge hat. Wird nun statt der exakten Systemanalysen ein globales Approximationsverfahren eingesetzt, dessen Auswertung nur Bruchteile einer Sekunde in Anspruch nimmt, können sehr viele Optimierungsläufe mit unterschiedlichen Startpunkten in sehr kurzer Zeit durchgeführt werden.

Gleiches gilt auch bei der Mehrzieloptimierung (siehe Abschnitt 2.3.1), bei der zur Ermittlung der Pareto-Front viele Optimierungsläufe – mit zum Beispiel variierenden Wichtungsfaktoren der Kostenfunktion – durchgeführt werden müssen. Die Aufstellung eines Approximationsmodells erfordert dabei in der Regel nur einmal den Aufwand zur Ermittlung der Datenbasis.

3 Response Surface Methode und Design of Experiments

Der Aufwand zur Behandlung multidisziplinärer Optimierungsprobleme wird sehr stark durch die Lösung der zugrunde liegenden Systemgleichungen bestimmt. Zur Reduktion des Optimierungsaufwands können mit der Response Surface Methode (RSM) erstellte Approximationen als Ersatz für die rechnerisch aufwendig zu ermittelnden Zusammenhänge zwischen den Entwurfsvariablen und einer daraus resultierenden Systemantwort verwendet werden. Zur Erstellung sowie zur Bewertung geeigneter Approximationen bietet die RSM eine Vielzahl von Techniken, die in diesem Kapitel vorgestellt und diskutiert werden. Dadurch soll eine Grundlage für die spätere Ableitung von Verfahrensweisen und Modifikationen geschaffen werden.

3.1 Grundlagen

Die für eine Optimierung relevante Systemantwort y ist abhängig von den Entwurfsvariablen $\xi_1, \xi_2, \dots, \xi_{n_V}$ und steht mit diesen im Zusammenhang:

$$y = y(\xi_1, \xi_2, \dots, \xi_{n_V}) = y(\boldsymbol{\xi}) \quad (3.1)$$

Wenn eine Response Surface Approximation (RSA) \hat{y} die reale Systemantwort y annähern soll, gilt:

$$y = \hat{y}(\xi_1, \xi_2, \dots, \xi_{n_V}) + \varepsilon = \hat{y}(\boldsymbol{\xi}) + \varepsilon \quad (3.2)$$

Das Korrekturglied ε kompensiert dabei in erster Linie Fehler der Approximation ($y(\boldsymbol{\xi}) - \hat{y}(\boldsymbol{\xi})$) aber auch Fehler, die bei der experimentellen Messung oder durch systeminhärente (z. B. numerische oder modellbedingte) Fehlerquellen entstehen. Um die Approximation \hat{y} so zu bestimmen, dass deren Gesamtfehler minimal wird, stützt sich die RSM auf einen Datensatz, der vorher zu generieren ist. Dieser Datensatz besteht aus n_P Stützpunkten.

Aus numerischen Gründen ist es bei der RSM üblich, die in ihren gängigen Einheiten verwendeten natürlichen Variablen $\xi_1, \xi_2, \dots, \xi_{n_V}$ in kodierte, dimensionslose Variablen x_1, x_2, \dots, x_{n_V} zu transformieren. Dies erfolgt für jede Stützstelle $\boldsymbol{\xi}_{(p)}$ und jede Entwurfsvariable ξ_i , meist unter Anwendung folgender Normierung auf das Intervall $[-1; +1]$:

$$x_{i(p)} = \frac{\xi_{i(p)} - \frac{\max(\xi_{i(j)}) + \min(\xi_{i(j)})}{2}}{\frac{\max(\xi_{i(j)}) - \min(\xi_{i(j)})}{2}} \quad \text{mit} \begin{cases} i = 1, 2, \dots, n_V \\ j, p = 1, 2, \dots, n_P \end{cases} \quad (3.3)$$

Bezüglich der normierten Variablen lässt sich der Approximationsansatz (3.2) wie folgt anschreiben:

$$y = \hat{y}(x_1, x_2, \dots, x_{n_V}) + \varepsilon = \hat{y}(\boldsymbol{x}) + \varepsilon \quad (3.4)$$

3.1.1 Standard Ansatzfunktionen

In vielen Fällen kann die Systemantwort mit einem Polynomansatz geringer Ordnung ausreichend gut approximiert werden. Einige Autoren (z. B. Myers und Montgomery [1995]) bezeichnen die RSM sogar nur dann als solche, wenn nur Polynomansätze mit maximal quadratischen Termen eingesetzt werden. Die in der Literatur am häufigsten verwendeten Ansatzfunktionen sind lineare Modelle

$$\hat{y} = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_{n_V} x_{n_V} = \beta_0 + \sum_{i=1}^{n_V} \beta_i x_i \quad (3.5)$$

und quadratische Modelle

$$\hat{y} = \beta_0 + \sum_{i=1}^{n_V} \beta_i x_i + \sum_{i=1}^{n_V} \sum_{j=i}^{n_V} \beta_{ij} x_i x_j \quad (3.6)$$

Mit einem Ansatz nach (3.5) lassen sich die Haupteffekte für die einzelnen Entwurfsvariablen identifizieren. Um Interaktionen zwischen zwei Entwurfsvariablen zu erfassen, müssen dem Ansatz die Terme $\beta_{ij} x_i x_j$ ($i \neq j$, $j > i$) hinzugefügt werden. Durch eine Erweiterung mit den quadratischen Termen $\beta_{ij} x_i x_j$ ($i = j$) erhält man den vollständigen quadratischen Ansatz (3.6). Mit diesem kann bereits ein breites Spektrum an Funktionsverläufen beschrieben werden.

Die unbekanntenen Koeffizienten β in (3.5) und (3.6) werden mittels einer linearen Regression bestimmt.

3.1.2 Lineare Regression und Methode der kleinsten Fehlerquadrate

Der Ansatz (3.5) stellt bereits ein standardmäßiges lineares Regressionsmodell dar. Um komplexere Ansätze wie (3.6) einer linearen Regression zugänglich zu machen, müssen die Terme höherer Ordnung substituiert werden. Beispielsweise kann der vollständige quadratische Ansatz für zwei Entwurfsvariablen

$$\hat{y} = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_{11} x_1^2 + \beta_{22} x_2^2 + \beta_{12} x_1 x_2 \quad (3.7)$$

mit den Substitutionen $x_3 = x_1^2$, $x_4 = x_2^2$, $x_5 = x_1 x_2$, $\beta_3 = \beta_{11}$, $\beta_4 = \beta_{22}$ und $\beta_5 = \beta_{12}$ in das lineare Regressionsmodell

$$\hat{y} = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \beta_5 x_5 \quad (3.8)$$

umgewandelt werden. Dies gilt auch für andere Ansatzfunktionen, die linear in den Parametern β sind.

Die gängigste Methode zur Bestimmung der unbekanntenen Regressionskoeffizienten β ist die Methode der kleinsten Fehlerquadrate, auch Gauß'sche Fehlerquadratminimierung genannt. Die Zahl der Stützpunkte im dabei verwendeten Datensatz¹ muss mindestens der Zahl der unbekanntenen Regressionskoeffizienten entsprechen ($n_P \geq n_V$). Zu jeder Stützstelle $\mathbf{x}_{(p)}$ gibt es eine Systemantwort (Stützwert $y_{(p)}$), so dass sich die Daten der Stützpunkte folgendermaßen darstellen lassen:

y	x_1	x_2	\dots	x_{n_V}
$y_{(1)}$	$x_{1(1)}$	$x_{2(1)}$	\dots	$x_{n_V(1)}$
$y_{(2)}$	$x_{1(2)}$	$x_{2(2)}$	\dots	$x_{n_V(2)}$
\vdots	\vdots	\vdots	\ddots	\vdots
$y_{(n_P)}$	$x_{1(n_P)}$	$x_{2(n_P)}$	\dots	$x_{n_V(n_P)}$

Mit den Stützpunkten und (3.4) lässt sich das lineare Regressionsmodell wie folgt anschreiben:

$$\begin{aligned} y_{(p)} &= \beta_0 + \beta_1 x_{1(p)} + \beta_2 x_{2(p)} + \dots + \beta_{n_V} x_{n_V(p)} + \varepsilon_{(p)} \\ &= \beta_0 + \sum_{i=1}^{n_V} \beta_i x_{i(p)} + \varepsilon_{(p)} \quad \text{mit } p = 1, 2, \dots, n_P \end{aligned} \quad (3.9)$$

In Matrixschreibweise lautet (3.9):

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \quad (3.10)$$

¹Methoden zur Planung und Festlegung der für einen solchen Datensatz zu generierenden Stützpunkte werden in Abschnitt 3.3 beschrieben.

Dabei gelten folgende Definitionen:

$$\mathbf{y} = \begin{bmatrix} y(1) \\ y(2) \\ \vdots \\ y(n_P) \end{bmatrix}, \quad \boldsymbol{\varepsilon} = \begin{bmatrix} \varepsilon(1) \\ \varepsilon(2) \\ \vdots \\ \varepsilon(n_P) \end{bmatrix}, \quad \boldsymbol{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_{n_V} \end{bmatrix},$$

$$\mathbf{X} = \begin{bmatrix} 1 & x_{1(1)} & x_{2(1)} & \cdots & x_{n_V(1)} \\ 1 & x_{1(2)} & x_{2(2)} & \cdots & x_{n_V(2)} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{1(n_P)} & x_{2(n_P)} & \cdots & x_{n_V(n_P)} \end{bmatrix}$$

Die Methode der kleinsten Fehlerquadrate beruht darauf, die Koeffizienten β so zu wählen, dass sich für die Summe der Quadrate der Fehler $\varepsilon_{(p)}$ der kleinstmögliche Wert ergibt.

$$\begin{aligned} L &= \sum_{p=1}^{n_P} \varepsilon_{(p)}^2 \\ &= \sum_{p=1}^{n_P} \left(y_{(p)} - \beta_0 - \sum_{i=1}^{n_V} \beta_i \xi_{i(p)} \right)^2 \end{aligned} \quad (3.11)$$

Gleichung (3.11) lautet in Matrizenschreibweise:

$$\begin{aligned} L &= \boldsymbol{\varepsilon}^T \boldsymbol{\varepsilon} = (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^T (\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) \\ &= \mathbf{y}^T \mathbf{y} - \boldsymbol{\beta}^T \mathbf{X}^T \mathbf{y} - \mathbf{y}^T \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\beta}^T \mathbf{X}^T \mathbf{X}\boldsymbol{\beta} \\ &= \mathbf{y}^T \mathbf{y} - 2\boldsymbol{\beta}^T \mathbf{X}^T \mathbf{y} + \boldsymbol{\beta}^T \mathbf{X}^T \mathbf{X}\boldsymbol{\beta} \end{aligned} \quad (3.12)$$

Zur Minimierung der Summe aller Fehlerquadrate L muss folgende notwendige Bedingung erfüllt sein:

$$\left. \frac{\partial L}{\partial \boldsymbol{\beta}} \right|_{\boldsymbol{\beta}=[b_0, b_1, \dots, b_{n_V}]^T} = -2\mathbf{X}^T \mathbf{y} + 2\mathbf{X}^T \mathbf{X}\mathbf{b} = \mathbf{0} \quad (3.13)$$

Somit ergeben sich die Koeffizienten des angepassten Regressionsmodells zu

$$\mathbf{b} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{y} \quad (3.14)$$

und folglich die an die Stützpunkte angepasste Approximationsfunktion:

$$\hat{\mathbf{y}} = \mathbf{X}\mathbf{b} \quad (3.15)$$

Die Differenz zwischen dem Stützwert $y_{(p)}$ und der approximierten Systemantwort $\hat{y}_{(p)}$ ist das Residuum $e_{(p)} = y_{(p)} - \hat{y}_{(p)}$. Der Residuenvektor ist bestimmt durch:

$$\mathbf{e} = \mathbf{y} - \hat{\mathbf{y}} \quad (3.16)$$

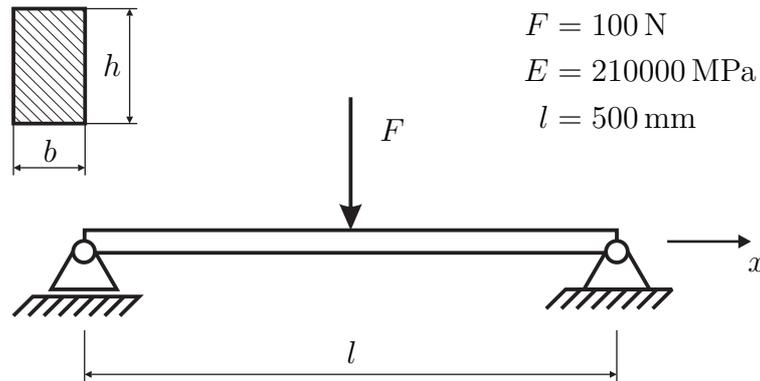


Abbildung 3.1: Zweifach gelagerter Balken mit mittig aufgebrachtener Querkraft

3.1.3 Beispiel „Biegebalken“

Im Folgenden wird ein einfaches Beispiel vorgestellt, an dem die RSM veranschaulicht werden soll. Es handelt sich um einen zweifach gelagerten Balken, der durch eine mittig aufgebrachte Kraft belastet wird (siehe Abbildung 3.1). Der Balken besitzt einen Rechteckquerschnitt, dessen Höhe h und Breite b die Entwurfsvariablen sind. Die zu approximierende Systemantwort ist die maximale Durchsenkung w_{max} in der Balkenmitte. Der bekannte Zusammenhang hierfür lautet:

$$w_{max} = w(x = l/2) = \frac{Fl^3}{4E} \frac{1}{bh^3} \quad (3.17)$$

Als Approximationsfunktion soll ein vollständiger quadratischer Ansatz nach (3.6) dienen.

$$w_{max} = \beta_0 + \beta_1 \bar{h} + \beta_2 \bar{b} + \beta_{11} \bar{h}^2 + \beta_{22} \bar{b}^2 + \beta_{12} \bar{h}\bar{b} + \varepsilon \quad (3.18)$$

Die Größen \bar{h} und \bar{b} sind die nach (3.3) normierten Querschnittsabmessungen h und b . Mit den Substitutionen $\bar{h} = x_1$, $\bar{b} = x_2$, $\bar{h}^2 = x_3$, $\bar{b}^2 = x_4$, $\bar{h}\bar{b} = x_5$ und $w_{max} = y$ folgt das lineare Regressionsmodell:

$$\hat{w}_{max} = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \beta_5 x_5 \quad (3.19)$$

Die für die Bestimmung der Koeffizienten β benötigten Stützstellen werden im Definitionsbereich der Querschnittsabmessungen ($b, h \in [5\text{mm}, 15\text{mm}]$) durch ein Full Factorial Design² dritter Ordnung gewählt. Dies bedeutet, dass b und h jeweils einen Wert aus $\{5\text{mm}, 10\text{mm}, 15\text{mm}\}$ annehmen. Daraus ergeben sich neun Kombinationen. Diese sind zusammen mit der sich jeweils aus (3.17) ergebenden maximalen Durchsenkung w_{max} in Tabelle 3.1 aufgeführt. Hier sind weiterhin die nach (3.3) normierten Größen \bar{h} und \bar{b} sowie die daraus resultierenden Werte für die substituierten Regressionsvariablen x_1, x_2, \dots, x_5 zu finden.

²Siehe Abschnitt 3.3

Stützpunkt	w_{max} [mm]	h [mm]	b [mm]	\bar{h} [-]	\bar{b} [-]	x_1 \bar{h}	x_2 \bar{b}	x_3 \bar{h}^2	x_4 \bar{b}^2	x_5 $\bar{h}\bar{b}$
1	23.81	5	5	-1	-1	-1	-1	1	1	1
2	11.91	5	10	-1	0	-1	0	1	0	0
3	7.937	5	15	-1	1	-1	1	1	1	-1
4	2.976	10	5	0	-1	0	-1	0	1	0
5	1.488	10	10	0	0	0	0	0	0	0
6	0.992	10	15	0	1	0	1	0	1	0
7	0.882	15	5	1	-1	1	-1	1	1	-1
8	0.441	15	10	1	0	1	0	1	0	0
9	0.294	15	15	1	1	1	1	1	1	1

Tabelle 3.1: Stützpunkte für Beispiel Biegebalken

Mit den Daten aus Tabelle 3.1 ergeben sich der Antwortvektor \mathbf{y} und die Regressionsmatrix \mathbf{X} zu:

$$\mathbf{y} = \begin{bmatrix} 23.81 \\ 11.91 \\ 7.937 \\ 2.976 \\ 1.488 \\ 0.992 \\ 0.882 \\ 0.441 \\ 0.294 \end{bmatrix} \quad \mathbf{X} = \begin{bmatrix} 1 & -1 & -1 & 1 & 1 & 1 \\ 1 & -1 & 0 & 1 & 0 & 0 \\ 1 & -1 & 1 & 1 & 1 & -1 \\ 1 & 0 & -1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 \\ 1 & 1 & -1 & 1 & 1 & -1 \\ 1 & 1 & 0 & 1 & 0 & 0 \\ 1 & 1 & 1 & 1 & 1 & 1 \end{bmatrix}$$

Die Koeffizienten der Approximationsfunktion (3.18) erhält man mit (3.14).

$$\begin{aligned} \mathbf{b} &= (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{y} \\ &= [0.794 \quad -7.006 \quad -3.074 \quad 5.726 \quad 1.537 \quad 3.821]^T \end{aligned}$$

Somit ergibt sich die Approximation $\hat{w}_{max}(\bar{h}, \bar{b})$ in den normierten Variablen zu:

$$\hat{w}_{max}(\bar{h}, \bar{b}) = 0.794 - 7.006 \bar{h} - 3.074 \bar{b} + 5.726 \bar{h}^2 + 1.537 \bar{b}^2 + 3.821 \bar{h} \bar{b} \quad (3.20)$$

Meist interessiert jedoch eine Approximation in den natürlichen Variablen ($\hat{w}_{max}(h, b)$). Hierzu ist die Normierung nach (3.3) wieder rückgängig zu machen, indem

$$\begin{aligned} x_i &= \frac{\xi_i - \frac{\max(\xi_{i(j)}) + \min(\xi_{i(j)})}{2}}{\frac{\max(\xi_{i(j)}) - \min(\xi_{i(j)})}{2}} \\ &= \frac{\xi_i - 10}{5} = 0.2 \xi_i - 2 \end{aligned} \quad \text{mit } \begin{cases} x_i = \bar{h}, \bar{b}, & \xi_i = h, b \\ j = 1, 2, \dots, 9 \end{cases}$$

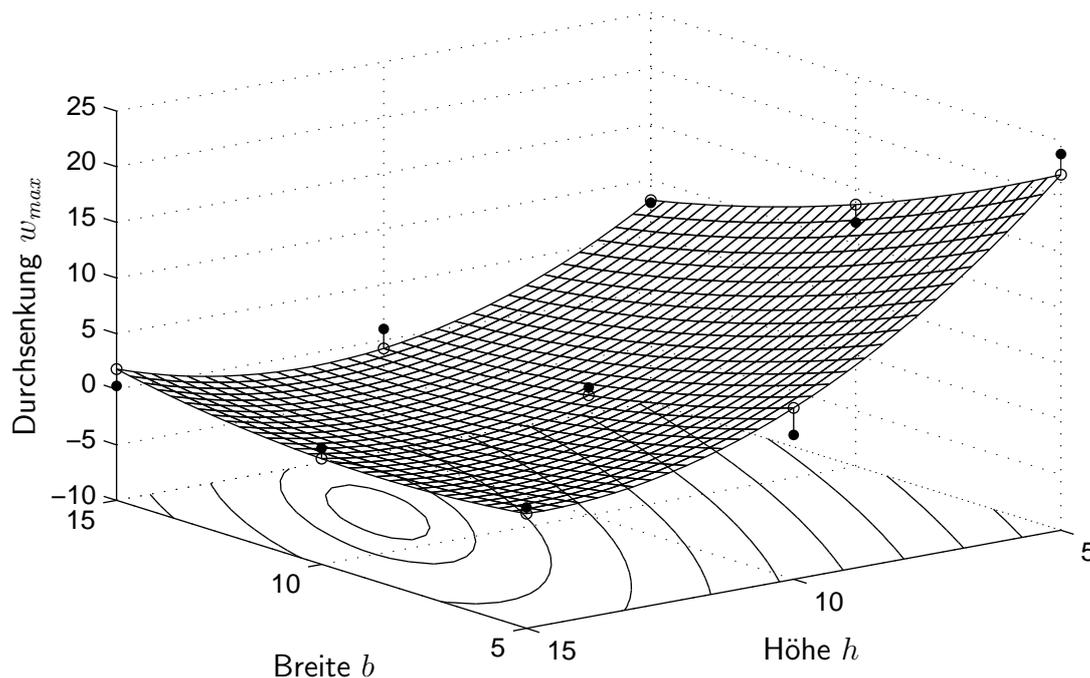


Abbildung 3.2: Maximale Durchsenkung w_{max} des Biegebalken, approximiert durch einen quadratischen Ansatz

in (3.20) eingesetzt wird.

$$\begin{aligned}
 \hat{w}_{max}(h, b) &= 0.794 - 7.006(0.2h - 2) - 3.074(0.2b - 2) \\
 &\quad + 5.726(0.2h - 2)^2 + 1.537(0.2b - 2)^2 \\
 &\quad + 3.821(0.2h - 2)(0.2b - 2) \\
 \hat{w}_{max}(h, b) &= 65.29 - 7.510h - 3.373b + 0.229h^2 + 0.0615b^2 + 0.153hb \quad (3.21)
 \end{aligned}$$

In Abbildung 3.2 ist die aufgestellte Approximationsfunktion grafisch dargestellt. Als Punkte sind die über den Stützstellen aufgetragenen Stützwerte gekennzeichnet. Die zugehörigen Werte der Approximationsfunktion sind mit Kreisen markiert.

Der tatsächliche Verlauf der maximalen Durchsenkung w_{max} wird durch (3.21) relativ gut wiedergegeben. Um genauere Aussagen über die Qualität der Approximation machen zu können, gibt es eine Reihe von Beurteilungskriterien, die im Folgenden beschrieben und diskutiert werden.

3.2 Kriterien zur Beurteilung der Approximation

Das mit einer RSA ermittelbare Optimum kann nur so gut wie die Approximation selbst sein. Die Qualität einer Approximation ist also im Prinzip daran zu messen, wie gut die für eine Optimierung relevanten Systemeigenschaften wiedergegeben werden. Angesichts des hohen Berechnungsaufwands für die Systemanalysen stehen zur Bewertung lediglich die Stützpunkte zur Verfügung, die auch zur Anpassung des Approximationsmodells verwendet werden. Da diese wiederum fehlerbehaftet sein können, muss deren Einfluss auf die Approximation bestimmt werden.

Die RSM umfasst verschiedene statische Methoden die eine Bewertung der Approximation erlauben. Ausgangspunkt dazu bilden folgende Definitionen, deren detaillierte Herleitung in [Myers und Montgomery 1995] zu finden ist.

Die Kovarianzmatrix des Koeffizientenvektors \mathbf{b} ist

$$\text{Cov}(\mathbf{b}) = \sigma^2(\mathbf{X}^T \mathbf{X})^{-1} \quad (3.22)$$

Dabei sind die Diagonalelemente $[\text{Cov}(\mathbf{b})]_{ii}$ die Varianzen der Koeffizienten b_i . Die Kovarianz zwischen den Koeffizienten b_i und b_j ($i \neq j$) ist gegeben durch die Elemente $[\text{Cov}(\mathbf{b})]_{ij}$. Eine üblicherweise verwendete Abschätzung für σ^2 ist

$$\hat{\sigma}^2 = MS_E = \frac{SS_E}{n_P - n_K} \quad (3.23)$$

wobei SS_E die Summe der Fehlerquadrate (Sum of Squares of Errors)

$$SS_E = \sum_{p=1}^{n_P} e_{(p)}^2 = \mathbf{e}^T \mathbf{e} \quad (3.24)$$

und MS_E das arithmetische Mittel der Summe der Fehlerquadrate (Mean of Squares of Errors) bezogen auf die $(n_P - n_K)$ Freiheitsgrade der Regression ist.

Für die weitere Berechnung statistischer Bewertungsgrößen benötigte Werte sind die Summe der Quadrate der Abweichungen der Stützwerte $y_{(p)}$ von deren Mittelwert \bar{y} (Sum of Squares Total)

$$SS_T = \sum_{p=1}^{n_P} (y_{(p)} - \bar{y})^2 \quad \text{mit} \quad \bar{y} = \frac{\sum_{p=1}^{n_P} y_{(p)}}{n_P} \quad (3.25)$$

oder
$$SS_T = \mathbf{y}^T \mathbf{y} - \frac{\left(\sum_{p=1}^{n_P} y_{(p)} \right)^2}{n_P}$$

sowie die Summe der Quadrate der Abweichungen der approximierten Werte des Regressionsmodells $\hat{y}_{(p)}$ vom Mittelwert der Stützwerte (Sum of Squares of Regression).

$$SS_R = \sum_{p=1}^{n_P} (\hat{y}_{(p)} - \bar{y})^2$$

$$\text{oder } SS_R = \mathbf{b}^T \mathbf{X}^T \mathbf{y} - \frac{\left(\sum_{p=1}^{n_P} y_{(p)} \right)^2}{n_P} \quad (3.26)$$

Mit diesen Werten können nun verschiedene Bewertungskriterien definiert werden.

3.2.1 Globale Bewertung des Regressionsmodells

Eine nahe liegende Größe zur Bewertung eines verwendeten Regressionsmodells ist die **Summe der Fehlerquadrate** SS_E . Da mit SS_E jedoch kein Bezug zu den Eigenschaften (Anzahl, Verteilung usw.) der verwendeten Stützpunkte gegeben ist, kann ein Vergleich verschiedener Regressionsmodelle nur dann sinnvoll erfolgen, wenn für diese jeweils die gleichen Stützpunkte verwendet werden.

Ein Wert, der die Eigenschaften der verwendeten Stützwerte mit einbezieht, ist das **Bestimmtheitsmaß** R^2 :

$$R^2 = \frac{SS_R}{SS_T} = 1 - \frac{SS_E}{SS_T} \quad (3.27)$$

Es ist ein Maß für die Minderung der Varianz der durch das Regressionsmodell approximierten Werte, bezogen auf die Gesamtvarianz der Stützwerte. Somit gibt es an, wie gut die verschiedenen Systemantworten approximiert werden. Der mögliche Wertebereich ist $0 \leq R^2 \leq 1$. Ein Bestimmtheitsmaß $R^2 \ll 1$ bedeutet ein sehr unzureichendes Approximationsvermögen des mathematischen Modells. Ein großer Wert hingegen besagt nicht notwendigerweise, dass das Regressionsmodell auch wirklich sehr gut ist. Dies ist darin begründet, dass eine größere Anzahl von Termen bzw. Regressionskoeffizienten in der Regel zu einer besseren Anpassung der Approximationsfunktion an die Stützwerte führt. Dies wiederum führt zu einer Reduktion der Summe der Fehlerquadrate und somit zu einem größeren Bestimmtheitsmaß R^2 .

Um verschiedene Regressionsmodelle mit variierender Anzahl an Stützpunkten vergleichen zu können, wird beim **adjustierten Bestimmtheitsmaß** R_{adj}^2 sowohl die Zahl der Regressionskoeffizienten als auch die Anzahl der Stützpunkte berücksichtigt.

$$R_{adj}^2 = 1 - \frac{SS_E / (n_P - n_K)}{SS_T / (n_P - 1)} = 1 - \left(\frac{n_P - 1}{n_P - n_K} \right) (1 - R^2) \quad (3.28)$$

Im Allgemeinen wächst R_{adj}^2 nicht notwendigerweise mit einer größeren Zahl von Regressionskoeffizienten. Oftmals sinkt der Wert, wenn dem Regressionsmodell unnötige Terme hinzugefügt werden.

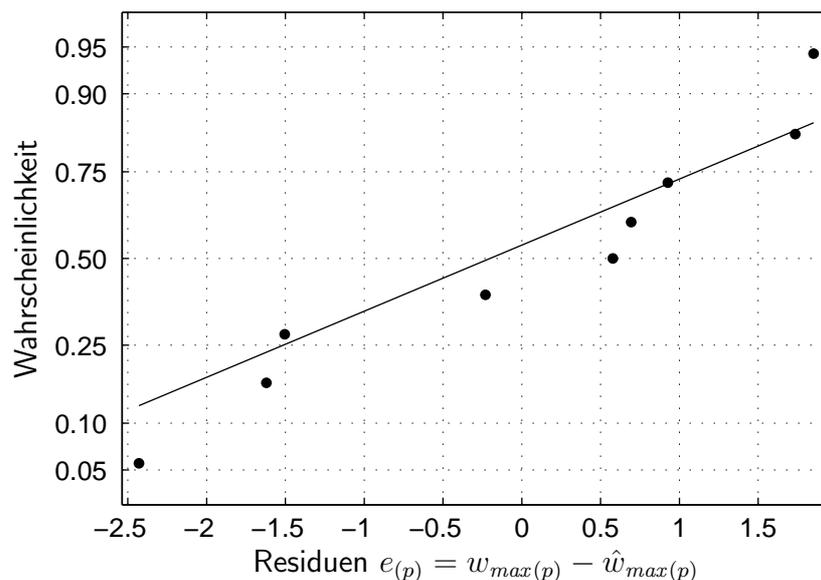


Abbildung 3.3: Wahrscheinlichkeit der normalverteilten Residuen des Beispiels „Biegebalken“

3.2.2 Analyse der Residuen

Den Residuen der Methode der kleinsten Fehlerquadrate kommt bei der Bewertung des Approximationsmodells ebenfalls eine große Bedeutung zu. Hierzu gibt es eine Reihe an grafischen Auswertungs- und Beurteilungsmethoden, die im Folgenden anhand der Residuen des Beispiels „Biegebalken“ vorgestellt werden. Diese ergeben sich mit (3.21) und Tabelle 3.1 zu:

$$\begin{aligned} \mathbf{e} &= \mathbf{w}_{max} - \hat{\mathbf{w}}_{max} \\ &= [1.85 \quad -1.62 \quad -0.23 \quad -2.43 \quad 0.69 \quad 1.74 \quad 0.58 \quad 0.93 \quad -1.50]^T \end{aligned} \quad (3.29)$$

Bei der RSM wird davon ausgegangen, dass die Residuen eine normalverteilte Menge mit dem Erwartungswert 0 bilden. Die Überprüfung dieser Annahme erfolgt, indem man über die aufsteigend sortierten Residuen $e_{(p)}$ die Werte der kumulierten Dichtefunktion einer normierten und zentrierten Normalverteilung aufträgt. Diese Werte entsprechen der Wahrscheinlichkeit, dass ein Residuum e eines beliebigen Punkts im Definitionsbereich der Entwurfsvariablen kleiner ist als $e_{(p)}$ und werden allgemein vorgegeben mit:

$$F(p) = F(e_{(p)}) = \frac{p - \frac{1}{2}}{n_P} \quad (3.30)$$

Abbildung 3.3 zeigt eine grafische Darstellung dieses **normalverteilten Residuenplots**. Liegen die Residuen näherungsweise auf einer geraden Linie, so ist die Annahme der Normalverteilung erfüllt. Ist der Verlauf der Residuen zwar geordnet, jedoch nicht

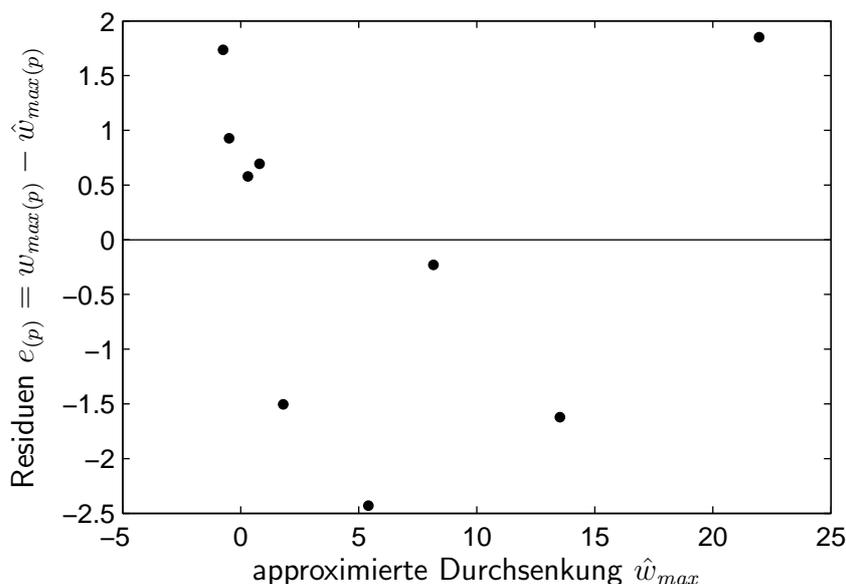


Abbildung 3.4: Residuen $e_{(p)}$ aufgetragen über die Werte der approximierten Durchsenkung $\hat{w}_{max(p)}$ aus dem Beispiel „Biegebalken“

auf einer geraden Linie, so ist dies ein Zeichen dafür, dass andere funktionale Zusammenhänge zugrunde liegen.

Eine weitere grafische Bewertungsmethode ist das Auftragen der Residuen über die approximierten Systemantwort. Diese Form der Darstellung wird als **Residuenplot** bezeichnet und ist für das Beispiel „Biegebalken“ in Abbildung 3.4 zu sehen. Die Verteilung der Punkte soll dabei möglichst zufällig sein. So weist zum Beispiel eine trichterförmige Anordnung der Punkte auf eine Korrelation und somit wiederum auf andere funktionale Zusammenhänge hin. Dies kann jedoch oft erst mit einer größeren Anzahl von Stützpunkten erkennbar werden.

Durch Skalierung der Residuen können in vielen Fällen weitere Informationen gewonnen werden. So werden die Residuen oft auf ihre durchschnittliche Standardabweichung normiert (**standardisierte Residuen**).

$$d_{(p)} = \frac{e_{(p)}}{\hat{\sigma}} \quad \text{mit} \quad \hat{\sigma} = \sqrt{\frac{\sum e_{(p)}^2}{n_P - n_K}} \quad (3.31)$$

Für das Beispiel „Biegebalken“ ergibt sich damit die in Abbildung 3.5(a) dargestellte Verteilung. Die standardisierten Residuen haben einen Durchschnittswert von 0 und eine Standardabweichung von ungefähr 1. Bei Werten, die im Vergleich zu den anderen unverhältnismäßig weit vom Erwartungswert 0 entfernt liegen oder außerhalb des Intervalls $-3 \leq d_{(p)} \leq 3$ zu finden sind, sollten die zugehörigen Stützpunkte genauer untersucht werden. Hier kann es sich einerseits um Fehler bei der Messung bzw. Be-

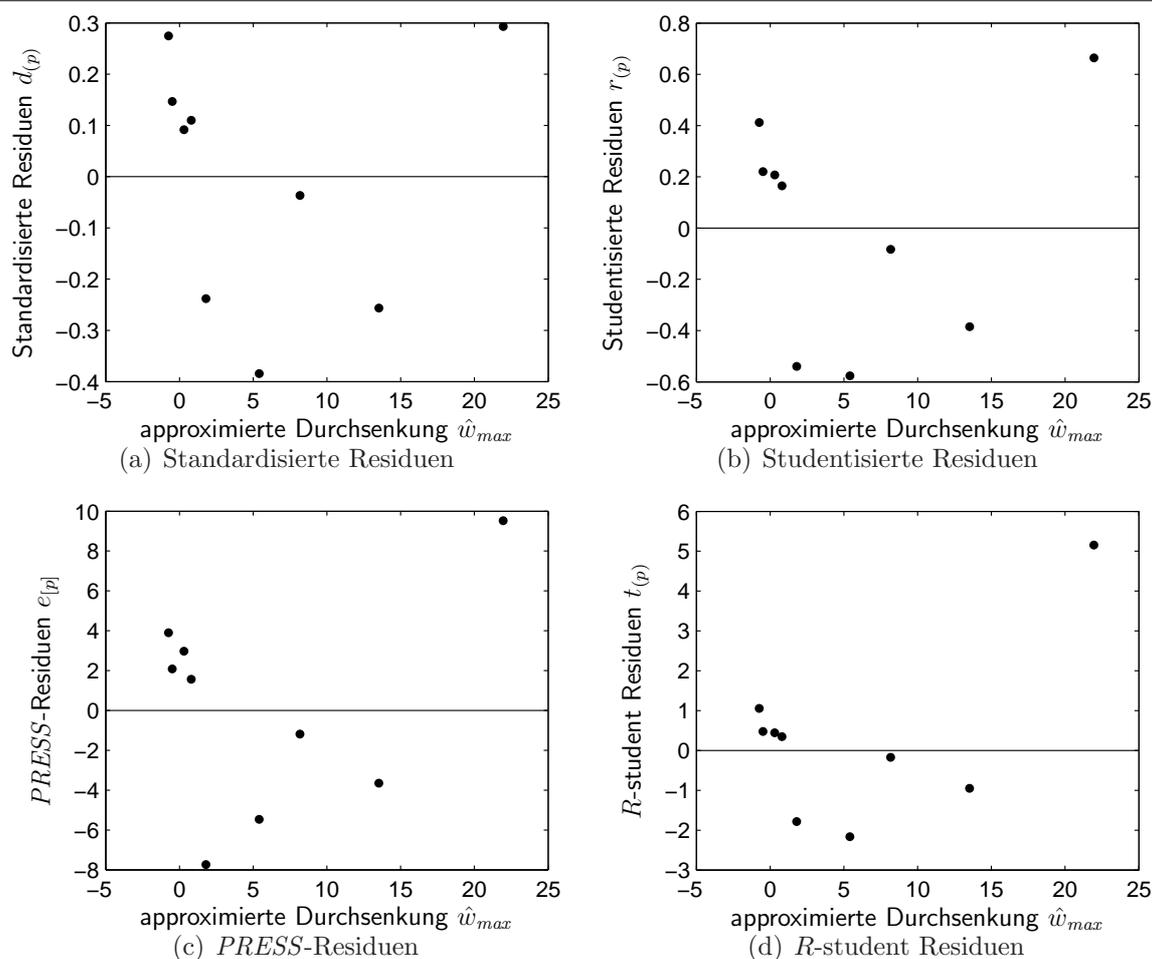


Abbildung 3.5: Verschiedene Skalierungen der Residuen (Daten aus dem Beispiel „Biegebalken“)

rechnung der Systemantwort handeln. Andererseits kann das Approximationsmodell grobe Unzulänglichkeiten in diesem Bereich aufweisen.

Je nach Stützpunkt-Datensatz können die Residuen teilweise sehr unterschiedliche Standardabweichungen besitzen. Ausgehend vom Vektor der approximierten Systemantworten $\hat{\mathbf{y}}$ (3.15) lässt sich mit (3.14) die Hat-Matrix ($\hat{\mathbf{y}} \rightarrow$ engl. „hat“) bestimmen:

$$\mathbf{H} = \mathbf{X}(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \quad (3.32)$$

Diese bildet den Stützwertvektor \mathbf{y} auf den Vektor der approximierten Systemantworten $\hat{\mathbf{y}}$ ab.

$$\hat{\mathbf{y}} = \mathbf{H} \mathbf{y} \quad (3.33)$$

Schreibt man nun die Kovarianzmatrix der Residuen mit (3.33), erhält man

$$\begin{aligned}\text{Var}(\mathbf{e}) &= \text{Var}((\mathbf{I} - \mathbf{H}) \mathbf{y}) \\ &= (\mathbf{I} - \mathbf{H}) \text{Var}(\mathbf{y})(\mathbf{I} - \mathbf{H})^T \\ &= \sigma^2(\mathbf{I} - \mathbf{H})\end{aligned}\quad (3.34)$$

Damit ergibt sich die Varianz des p -ten Residuums zu

$$\text{Var}(e_{(p)}) = \sigma^2(1 - h_{pp}) \quad (3.35)$$

Hierbei ist h_{pp} das p -te Diagonalelement von \mathbf{H} und liegt im Intervall $[0; 1]$. Da die h_{pp} durch die Lage der Stützstellen im Designraum bestimmt werden, gilt dies auch für die Varianzen der zugehörigen Residuen. Meist sind die $\text{Var}(e_{(p)})$ nahe dem Zentrum des Designraums größer als an dessen Rändern. Verletzungen der Modellannahmen sind an den Rändern des Designraums zwar wahrscheinlicher, jedoch mit Beobachtung der Residuen $e_{(p)}$ oder standardisierten Residuen $d_{(p)}$ schwerer zu entdecken, da diese dort in der Regel kleiner sind. Bei den **studentisierten Residuen**

$$r_{(p)} = \frac{e_{(p)}}{\sqrt{\hat{\sigma}^2(1 - h_{pp})}} \quad \text{mit} \quad \hat{\sigma}^2 = \frac{\sum e_{(p)}^2}{n_P - n_K} \quad (3.36)$$

werden diese Unterschiede berücksichtigt. Für das Beispiel „Biegebalken“ sind diese in Abbildung 3.5(b) dargestellt.

Eine weitere nützliche Skalierung wird über die **PRESS-Residuen** (Prediction Error Sum of Squares) definiert. Hierzu wird jeweils ein Stützpunkt $\mathbf{x}_{(p)}$ aus dem Datensatz entfernt und der Approximationsansatz an die verbleibenden $(n_P - 1)$ Stützpunkte angepasst. Mit der so aufgestellten Approximation kann dann eine prognostizierte Antwort für die p -te Stützstelle $\hat{y}_{[p]}$ errechnet werden. Der Prognosefehler für den p -ten Stützwert ist dann $e_{[p]} = y_{(p)} - \hat{y}_{[p]}$. Führt man dies für alle n_P Stützpunkte durch und summiert die Quadrate der *PRESS*-Residuen auf, erhält man:

$$PRESS = \sum_{p=1}^{n_P} e_{[p]}^2 = \sum_{p=1}^{n_P} (y_{(p)} - \hat{y}_{[p]})^2 \quad (3.37)$$

Um die *PRESS*-Residuen $e_{[p]}$ und somit den *PRESS*-Wert zu bestimmen ist es jedoch nicht nötig n_P Regressionsmodelle anzupassen. Unter Verwendung der Hat-Matrix (3.32) kann dieser errechnet werden mit:

$$PRESS = \sum_{p=1}^{n_P} \left(\frac{e_{(p)}}{1 - h_{pp}} \right)^2 \quad (3.38)$$

Aus (3.38) ist ersichtlich, dass die *PRESS*-Residuen die mittels der Diagonalelemente der Hat-Matrix gewichteten gewöhnlichen Residuen $e_{(p)}$ sind. Ein Stützpunkt, dessen h_{pp} groß ist, besitzt ein großes *PRESS*-Residuum und kann im Allgemeinen als

sehr einflussreich eingestuft werden. Unter der Voraussetzung, dass der zugehörige Stützwert keinem Fehler unterliegt, wird ein ohne diesen Stützpunkt angepasstes Approximationsmodell die Systemantwort in der Regel schlechter approximieren als ein mit diesem Punkt angepasstes Modell. Abbildung 3.5(c) zeigt die *PRESS*-Residuen für das Beispiel „Biegebalken“.

Mit dem *PRESS*-Wert kann zudem das **Maß für die Prognosegüte** R_{pred}^2 bestimmt werden.

$$R_{pred}^2 = 1 - \frac{PRESS}{S_{yy}} \quad \text{mit} \quad S_{yy} = \sum_{p=1}^{n_P} y_{(p)}^2 - \frac{\left(\sum_{p=1}^{n_P} y_{(p)}\right)^2}{n_P} \quad (3.39)$$

Für das Beispiel „Biegebalken“ ergibt sich der Wert $R_{pred}^2 = 0.996$. Demzufolge beschreibt der verwendete quadratische Ansatz 99.6% der Variabilität beim Prognostizieren von Systemantworten.

Die studentisierten Residuen $r_{(p)}$ (3.36) werden oft benutzt, um Ausreißer unter den Stützpunkten zu identifizieren. Hierbei wird $\hat{\sigma}^2$ durch Einbeziehen aller Stützpunkte abgeschätzt. Eine andere Möglichkeit, σ^2 aus (3.35) abzuschätzen ist, dies jeweils ohne den p -ten Stützpunkt zu tun. Das so studentisierte **R-student Residuum** ergibt sich zu:³

$$t_{(p)} = \frac{e_{(p)}}{\sqrt{S_{[p]}^2(1 - h_{pp})}} \quad \text{mit} \quad S_{[p]}^2 = \frac{PRESS - \frac{e_{(p)}^2}{(1-h_{pp})}}{n_P - n_K - 1} \quad (3.40)$$

Meist unterscheiden sich die *R*-student Residuen $t_{(p)}$ nur wenig von den studentisierten Residuen $r_{(p)}$. Hat der p -te Stützpunkt jedoch großen Einfluss, so können sich $S_{[p]}^2$ und $\hat{\sigma}^2$ signifikant unterscheiden, so dass die *R*-student Residuen hier sensitiver sind. Diese sind – wiederum für das Beispiel „Biegebalken“ – in Abbildung 3.5(d) dargestellt.

Eine zusammenfassende Diskussion der Residuenanalyse wird in Abschnitt 3.2.4 geführt.

3.2.3 Einflussdiagnose

In einigen Fällen kann es vorkommen, dass eine kleine Gruppe von Stützpunkten überproportionalen Einfluss auf die Bestimmung der Koeffizienten des Regressionsmodells ausübt. Diese Punkte zu finden und zu untersuchen, kann durchaus wichtig sein. Stellt sich heraus, dass die Systemantworten der Punkte fehlerbehaftet sind, müssen

³An dieser Stelle sei auf einen Fehler in [Myers und Montgomery 1995] hingewiesen. Die dort angegebene Formel für $S_{[p]}^2$ führt teilweise zu negativen Werten und somit zu imaginären *R*-student Residuen.

diese aus dem Datensatz entfernt werden. Falls es sich um verlässliche Werte handelt, können die Informationen dazu dienen, eine bessere Einsicht in die physikalischen bzw. mathematischen Zusammenhänge zu gewinnen.

Zur Beurteilung des Einflusses kann die Hat-Matrix verwendet werden. Als **einfache Faustregel** gilt: Der Einfluss eines Stützpunkts $[y_{(p)} \mathbf{x}_{(p)}^T]^T$ auf die approximierte Antwort $\hat{y}_{(p)}$ wird als überproportional bezeichnet, wenn das Diagonalelement h_{pp} größer ist als $2n_K/n_P$.

Das **Cook'sche Einflussmaß** bewertet den Einfluss eines Stützpunkts auf die Koeffizienten, indem der Abstand des Koeffizientenvektors \mathbf{b} zum Koeffizientenvektor $\mathbf{b}_{[p]}$ zur Berechnung herangezogen wird. Dabei wird $\mathbf{b}_{[p]}$ unter Weglassen des p -ten Stützpunkts ermittelt. Man erhält:

$$\begin{aligned} D_{(p)} &= \frac{(\mathbf{b}_{[p]} - \mathbf{b})^T \mathbf{X}^T \mathbf{X} (\mathbf{b}_{[p]} - \mathbf{b})}{n_K MS_E} \\ &= \frac{r_{(p)}^2 \text{Var}(\hat{y}(\mathbf{x}_{(p)}))}{n_K \text{Var}(e_{(p)})} = \frac{r_{(p)}^2}{n_K} \frac{h_{pp}}{(1 - h_{pp})} \end{aligned} \quad (3.41)$$

Punkte mit $D_{(p)} > 1$ können in der Regel als einflussreich betrachtet werden. Das Verhältnis $h_{pp}/(1 - h_{pp})$ entspricht dem Abstand des Vektors $\mathbf{x}_{(p)}$ zum „Schwerpunkt“ der restlichen Daten. Somit spiegelt eine Komponente von $D_{(p)}$ wider, wie gut das Modell den p -ten Stützwert approximiert. Die andere Komponente misst die Distanz der p -ten Stützstelle zu den übrigen Stützstellen und nimmt dadurch eine Wichtung vor.

3.2.4 Zusammenfassende Betrachtung der Bewertungskriterien

Betrachtet man für das Beispiel „Biegebalken“ zunächst den normalverteilten Residuenplot in Abbildung 3.3, so liegen die Residuen annähernd auf einer Linie. Daraus kann geschlossen werden, dass die Fehler der Approximation nahezu eine Gauß'sche Normalverteilung haben. Demzufolge ist eine geordnete Verteilung und damit wiederum gravierende Mängel in den durch das Regressionsmodell beschreibbaren funktionalen Zusammenhängen eher auszuschließen. Auch der in Abbildung 3.4 dargestellte Residuenplot deutet nur wenig auf die Notwendigkeit anderer funktionaler Zusammenhänge in den Ansatztermen hin. Erst durch die Hinzunahme weiterer Stützpunkte (z. B. durch Verwendung eines 5^2 -Designs) wird eine Korrelation der Residuen erkennbar. Im normalverteilten Residuenplot entfernen sich dann die Punkte im Bereich kleiner und großer approximierter Durchsenkungen deutlich vom linearen Verlauf.

Mit einem einfachen Residuenplot sind Ausreißer unter den Stützwerten identifizierbar. Für diese Aufgabe jedoch besser geeignete Werte sind die in den Abbildungen 3.5(a) bzw. 3.5(b) dargestellten, auf ihre durchschnittliche Standardabweichung bezogenen

standardisierten Residuen und die auf ihre Varianzen bezogenen studentisierten Residuen. Letztere bieten eine individuellere – auch von der Lage der jeweiligen Stützstelle beeinflusste – Skalierung und sind den standardisierten Residuen vorzuziehen. Im Beispiel „Biegebalken“ sind mit diesen beiden skalierten Residuen keine Ausreißer identifizierbar, da aufgrund des verwendeten, klar definierten mathematischen Zusammenhangs keine Fehler bei den Stützwerten auftreten.

Die beiden anderen Varianten der Skalierung sind in den Abbildungen 3.5(c) und 3.5(d) dargestellt. Sie berücksichtigen den Einfluss jedes Stützpunkts auf die Varianz der Koeffizienten des Regressionsmodells und somit den Einfluss jedes Stützpunkts auf die Approximation. Insbesondere bei den R -student Residuen erkennt man den im Vergleich zu den restlichen Stützpunkten, überproportionalen Einfluss des Stützpunkts bei $[h = 5 \text{ mm}, b = 5 \text{ mm}]$. Dies ist vorrangig darin begründet, dass die Höhe h in (3.17) in der dritten Potenz eingeht. Eine sichere Identifikation besonders einflussreicher Punkte ist ebenso über die im Abschnitt 3.2.3 beschriebene Faustregel und das Cook'sche Einflussmaß gewährleistet.

Wird der so identifizierte Stützpunkt bei der Anpassung des Regressionsmodells ausgegrenzt, so ändert sich die Approximation sehr stark. In Bezug auf eine eventuell angestrebte Optimierung muss dies jedoch nicht zwangsläufig bedeuten, dass die Approximation dadurch schlechter wird. Ohne den genannten Stützpunkt sinkt im Beispiel „Biegebalken“ der maximale Betrag der Residuen von 2.4 mm auf 0.67 mm. Die Vermutung einer demzufolge besseren Approximation der verbleibenden Stützpunkte lässt sich anhand des adjustierten Bestimmtheitsmaßes R_{adj}^2 bestätigen. Dieses steigt mit dem Weglassen des Stützpunkts von $R_{adj}^2 = 0.935$ auf $R_{adj}^{2*} = 0.978$.

3.3 Methoden zur Wahl der Stützstellen – Design of Experiments

Das wesentliche Ziel bei der Wahl der Stützstellen ist eine möglichst gute Festlegung bzw. Gestaltung der (numerischen) Experimente (Design of Experiments, DOE). Die daraus gewonnenen Stützpunkte sollen eine möglichst gute Anpassung des Regressionsmodells an die realen Zusammenhänge gewährleisten und zugleich Informationen für Tests zur Beurteilung des Approximationsmodells liefern. Gerade bei sehr zeitaufwendigen Systemanalysen ist es zudem fast unabdingbar, diesen Anforderungen mit möglichst wenigen Stützpunkten gerecht zu werden.

In der Literatur wie auch in vielen Softwarepaketen (z. B. [MATLAB Statistics Toolbox 2002]) wird eine Vielzahl verschiedener Klassen von DOE-Methoden beschrieben bzw. angeboten. Diese reichen von den Standard Response Surface Designs (RSDs) bis hin zum computergestützten (Computer Aided) Design of Experiments (CADOE). Letzteres bietet sich vor allem dann an, wenn die Standard RSDs nicht angewendet werden

können, zum Beispiel wenn der Definitionsbereich der Entwurfsvariablen unregelmäßig geformt oder unterbrochen ist.

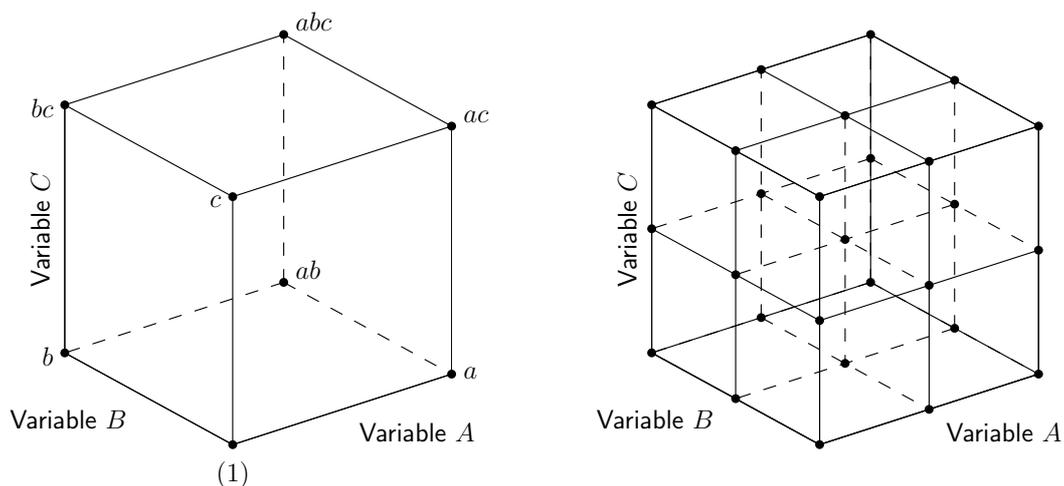
Im Folgenden werden die am häufigsten verwendeten RSDs einschließlich des CADOE beschrieben und diskutiert.

3.3.1 Factorial Designs

Bei den Factorial Designs wird der Definitionsbereich jeder Entwurfsvariablen gleichmäßig unterteilt, so dass jede Entwurfsvariable l Werte (*Levels*) annimmt. Die Gesamtheit der Stützstellen $\mathbf{x}_{(p)}$ wird aus allen möglichen Kombinationen gebildet. Deshalb wird dieses RSD häufig auch als vollständiges (*Full*) Factorial Design bezeichnet. Allgemein spricht man vom **l -Level Factorial Design, Factorial Design l -ter Ordnung** oder kurz **l^{n_V} -Design**. Für n_V Variablen mit jeweils l Levels ergeben sich l^{n_V} Stützstellen.

Am meisten verbreitet sind Factorial Designs zweiter und dritter Ordnung. Diese sind für den Fall $n_V = 3$ in den Abbildungen 3.6(a) bzw. 3.6(b) veranschaulicht. Man erkennt, dass für das 2-Level Factorial Design $n_p = 2^3 = 8$ Stützpunkte ermittelt werden müssen. Beim 3-Level Factorial Design sind es bereits $n_p = 3^3 = 27$ Stützpunkte.

Bei den Factorial Designs bietet sich auch die Möglichkeit, unterschiedliche Levels für die einzelnen Entwurfsvariablen vorzugeben, je nachdem welcher Grad an nichtlinearem Einfluss zu erwarten ist. Die Anzahl der dann zu bestimmenden Stützpunkte ergibt sich zu $n_p = \prod_{i=1}^{n_V} l_i$, wobei l_i die Anzahl der Levels der i -ten Entwurfsvariablen ist.



(a) **2³-Design** – Factorial Design zweiter Ordnung für drei Designvariablen (b) **3³-Design** – Factorial Design dritter Ordnung für drei Designvariablen

Abbildung 3.6: Full Factorial Designs für drei Entwurfsvariablen

Stützstelle	Variable/Haupteffekt			Kombination	Interaktionseffekt			
	A	B	C		AB	BC	AC	ABC
1	-	-	-	(1)	+	+	+	-
2	+	-	-	a	-	-	+	+
3	-	+	-	b	-	+	-	+
4	-	-	+	c	+	-	-	+
5	+	+	-	ab	+	-	-	-
6	+	-	+	ac	-	+	-	-
7	-	+	+	bc	-	-	+	-
8	+	+	+	abc	+	+	+	+

Tabelle 3.2: Stützstellen und untersuchbare Einflüsse eines 2^3 -Designs

Obwohl beim **Factorial Design zweiter Ordnung** jede Variable nur zwei Werte annimmt, hat dieses für die RSM eine nicht zu unterschätzende Bedeutung. Mit den Informationen der so generierten Stützpunkte können sowohl Haupteigenschaften als auch Interaktionen der Entwurfsvariablen bestimmt werden. In Tabelle 3.2 sind die bei einem 2^3 -Design möglichen Kombinationen dreier Eingangsgrößen A , B und C aufgelistet. Hierbei steht die Notation „+“ für die Obergrenze einer Eingangsgröße und „-“ für deren Untergrenze. Die Kombination (1) bedeutet, dass alle Variablen den Wert der Untergrenze ihres jeweiligen Definitionsbereichs annehmen. Die Präsenz eines a und/oder b und/oder c bedeutet, dass die zugehörige Eingangsgröße mit der Obergrenze ihres Definitionsbereichs verwendet wird. In Abbildung 3.6(a) sind die auf diese Weise kodierten Stützstellen gekennzeichnet.

Ausgehend vom Stützpunkt (1) besitzt ein Factorial Design zweiter Ordnung allgemein $n_{DOF} = 2^{n_V} - 1$ Freiheitsgrade. Mit diesen können n_V Haupteffekte, $n_V! / ((n_V - 2)! 2!)$ Zwei-Faktor-Interaktionen, $n_V! / ((n_V - 3)! 3!)$ Drei-Faktor-Interaktionen, ... und eine n_V -Faktor-Interaktion untersucht werden (vgl. Tabelle 3.2).

Zur Abschätzung des Einflusses einer (oder mehrerer) Eingangsgröße(n) wird die durch die Änderung des Levels dieser Eingangsgröße(n) hervorgerufene, durchschnittliche Änderung der Systemantwort berechnet. Der Einfluss der Eingangsgröße A ergibt sich für das 2^3 -Design aus Tabelle 3.2 zu:

$$\begin{aligned}
 \mathcal{E}_A &= \bar{y}_{A+} - \bar{y}_{A-} \\
 &= \frac{y_a + y_{ab} + y_{ac} + y_{abc}}{4} - \frac{y_{(1)} + y_b + y_c + y_{bc}}{4} \\
 \mathcal{E}_A &= \frac{1}{4} [y_a + y_{ab} + y_{ac} + y_{abc} - y_{(1)} - y_b - y_c - y_{bc}] \tag{3.42}
 \end{aligned}$$

Die Vorzeichen in (3.42) werden den jeweiligen Spalten der Tabelle 3.2 entnommen. Die Vorzeichen für die Interaktionseffekte ergeben sich aus denen der Haupteffekte durch Anwendung der mathematischen Vorzeichenregeln. Der Einfluss der Interaktion BC ist demnach:

$$\mathcal{E}_{BC} = \frac{1}{4} [y_{(1)} + y_a + y_{bc} + y_{abc} - y_b - y_c - y_{ab} - y_{ac}] \quad (3.43)$$

Aufgrund der – im Vergleich zu anderen RSDs – geringen Zahl an Stützstellen wird das Factorial Design zweiter Ordnung oft zu Beginn einer Response Surface Studie (beim sog. Screening) eingesetzt, wo es primär darum geht, die wesentlichen Eigenschaften der Eingangsgrößen zu identifizieren. Weiterhin bildet es meist einen Basisdatensatz für andere RSDs. Die bereits vorhandenen Stützwerte müssen dann nicht erneut ermittelt werden.

Fractional Factorial Designs

Die Anzahl der Stützstellen, und damit der erforderlichen Systemanalysen, wächst exponentiell mit der Zahl der Entwurfsvariablen. Ein Factorial Design zweiter Ordnung für sechs Entwurfsvariablen benötigt zum Beispiel $2^6 = 64$ (numerische) Experimente. Gerade einmal 6 der 63 Freiheitsgrade dienen hierbei zur Identifikation von Haupteffekten der Entwurfsvariablen. Weitere 15 Freiheitsgrade dienen zur Identifikation von Zwei-Faktor-Interaktionen. Mit den restlichen 42 Freiheitsgraden können Interaktionen dritter und höherer Ordnung identifiziert werden.

Unter der häufig zutreffenden Annahme, dass die Systemantwort maßgeblich nur durch Haupteffekte und Interaktionen niedriger Ordnung beeinflusst wird, benötigt man zur Bestimmung der Koeffizienten des Regressionsmodells lediglich einen Bruchteil (*Fractional*) eines Full Factorial Designs. Es besteht die Möglichkeit die Hälfte, ein Viertel oder allgemein den (2^{n_F}) -ten Bruchteil eines Factorial Designs auszuwählen. Dieses wird kurz als **$2^{n_V - n_F}$ -Design** bezeichnet.

Zwei wesentliche Eigenschaften des Fractional Factorial Designs sind:

- Ein Fractional Factorial Design lässt sich in einen Unterraum der Entwurfsvariablen projizieren und stellt für diesen dann ein vollständigeres bzw. vollständiges Factorial Design dar (siehe Abbildung 3.7).
- Mehrere Fractional Factorial Designs können sequenziell kombiniert werden, bis sie ein vollständiges Factorial Design bilden.

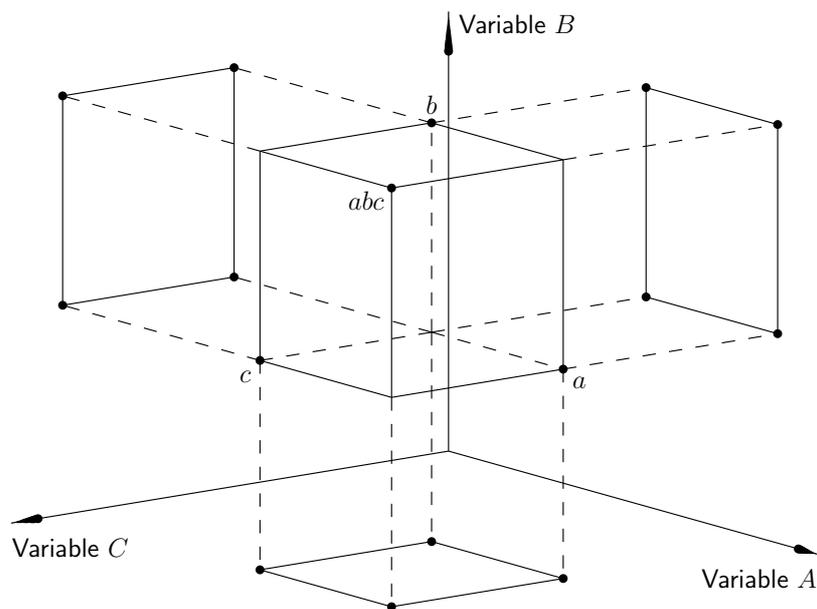


Abbildung 3.7: Projektion eines 2^{3-1}_{III} -Designs auf drei 2^2 -Designs

Die Konstruktion eines Fractional Factorial Designs erfolgt nach folgender Vorgehensweise:

- Zunächst wird als Basisdesign ein vollständiges Factorial Design für $n_V - n_F$ Designvariablen aufgestellt. Dies ist für ein 2^{3-1} -Design in Tabelle 3.3 dargestellt.
- Zur methodischen Bestimmung der Vorzeichenkodierung der n_F nicht im Basisdesign enthaltenen Variablen müssen n_F Erzeuger festgelegt werden. Diese werden ausgehend vom Interaktionseffekt höchster Ordnung gewählt (vgl. Tabelle 3.2) und dürfen jeweils nur eine der n_F nicht im Basisdesign enthaltenen Variablen enthalten.
- Die Anzahl der am erzeugenden Effekt beteiligten Variablen bestimmt die so genannte Länge der (des) Erzeuger(s). Durch den kürzesten Erzeuger wird die Auflösung eines Factorial Designs festgelegt, die in der Kurzbezeichnung als In-

Stützstelle	2^2 Basisdesign		$2^{3-1}_{\text{III}}, I = ABC$			$2^{3-1}_{\text{III}}, I = -ABC$		
	A	B	A	B	C = AB	A	B	C = -AB
1	-	-	-	-	+	-	-	-
2	+	-	+	-	-	+	-	+
3	-	+	-	+	-	-	+	+
4	+	+	+	+	+	+	+	-

Tabelle 3.3: Konstruktion eines 2^{3-1} -Designs (Hauptteil und komplementärer Teil)

design mit römischen Zahlen genannt wird. Im Fall des 2_{III}^{3-1} -Designs in Tabelle 3.3 ist der Erzeuger $\pm ABC$ und hat die Länge 3. Mit diesem Erzeuger kann nun mittels $C = +AB$ der Hauptteil und mit $C = -AB$ der komplementäre Teil des 2_{III}^{3-1} -Designs konstruiert werden. Die somit festgelegten Stützstellen entsprechen den Zeilen in Tabelle 3.2, bei denen der Interaktionseffekt ABC die Kodierung „+“ beziehungsweise „-“ hat.

- Werden zwei oder mehr Erzeuger gewählt, so bilden diese zusammen mit ihren Produkten die *generalisierten Erzeuger*. Sind zum Beispiel für ein 2_{IV}^{6-2} -Design die Erzeuger $ABCE$ und $BCDF$ gewählt, so bestimmt sich deren generalisierte Interaktion zu:

$$ABCE \cdot BCDF = AB^2 C^2 DEF = ADEF$$

Die Abschätzung der Einflüsse der Effekte erfolgt analog zum Factorial Design. Allerdings steht nur ein Bruchteil der Stützpunkte zur Verfügung, was dazu führt, dass einige Effekte „gekoppelt“ sind. Vergleicht man beispielsweise beim 2_{III}^{3-1} -Design in Tabelle 3.2 die Einflussstärke des Haupteffekts A mit der des Interaktionseffekts BC , so erkennt man anhand der Vorzeichen dieser Effekte (nur Zeilen mit $ABC = „+“$ oder „-“), dass diese identisch sind und somit keine differenzierbare Aussage liefern. Effekte mit identischer Vorzeichenkombination bilden ein *Alias*. Formal ergibt sich die Alias-Struktur eines Fractional Factorial Designs aus der Multiplikation aller Variableneffekte mit den das Design definierenden generalisierten Erzeugern. Im Fall des 2_{III}^{3-1} -Designs gibt es nur den Erzeuger $\pm ABC$, so dass sich das Alias für B wie folgt berechnen lässt:

$$B \cdot (\pm ABC) = \pm AB^2 C = \pm AC$$

Jeder Variableneffekt hat $(2^{n_F} - 1)$ Aliase. Da man bei einer relativ großen Zahl von Entwurfsvariablen Interaktionseffekte sehr hoher Ordnung oft ausschließen kann, vereinfacht sich die Aliasstruktur entsprechend, weil die berechenbare Einflussstärke jeweils nur noch einer reduzierten Menge an Interaktionseffekten zugeordnet werden muss. Da die Wahl der Erzeuger die Aliasstruktur der Interaktionseffekte stark beeinflusst, ist diesem Schritt eine besondere Bedeutung beizumessen. Generell sollen die Erzeuger so gewählt werden, dass daraus ein Fractional Factorial Design mit der höchstmöglichen Auflösung resultiert.

Mit den bisher beschriebenen RSDs können nur lineare Approximationsmodelle angepasst werden. Im Folgenden werden RSDs vorgestellt, die auch die Bestimmung quadratischer und somit nichtlinearer Effekte ermöglichen.

3.3.2 Response Surface Designs für Regressionsmodelle zweiter Ordnung

RSDs für Ansatzfunktionen, die durch ein Polynom zweiten Grades definiert werden, müssen neben den zu Beginn des Abschnitts 3.3 erwähnten Anforderungen folgende zwei Voraussetzungen erfüllen:

- Jede Entwurfsvariable muss mit mindestens drei Levels vertreten sein.
- Das RSD muss mindestens $1 + 2n_V + n_V(n_V - 1)/2$ unterschiedliche Stützstellen vorgeben.

Während bei RSDs zur Untersuchung von Haupt- und Interaktionseffekten die Orthogonalität⁴ die dominierende Eigenschaftsanforderung darstellt, rückt bei RSDs für Approximationsmodelle zweiter oder höherer Ordnung die damit erreichbare Prognosegüte deutlich in den Vordergrund.

Central Composite Design

Das zweifellos bekannteste und am weitesten verbreitete RSD zweiter Ordnung ist das Central Composite Design (CCD). Das CCD wird zunächst aus einem Factorial Design zweiter Ordnung oder auch einem Fractional Factorial Design mit Auflösung V aufgebaut. Bei Verwendung eines Fractional Factorial Designs der Auflösung III spricht man von einem *Small Composite Design* (SCD). Hinzu kommen die folgenden $2n_V$ auf den Achsen des kodierten Entwurfsvariablenraums liegenden Stützstellen:

$$\begin{array}{cccc} x_1 & x_2 & \dots & x_{n_V} \\ \hline -\alpha & 0 & \dots & 0 \\ +\alpha & 0 & \dots & 0 \\ 0 & -\alpha & \dots & 0 \\ 0 & +\alpha & \dots & 0 \\ 0 & 0 & \dots & -\alpha \\ 0 & 0 & \dots & +\alpha \end{array}$$

Der Parameter α gibt dabei die Distanz der axialen Stützstellen zum Ursprung des kodierten Entwurfsvariablenraums an. Weiterhin folgen noch n_C Stützstellen im Ursprung des kodierten Entwurfsvariablenraums (sog. *Center Runs*). Diesen drei Komponenten des CCD sind jeweils unterschiedliche Aufgaben zugeordnet.

- Das (Fractional) Factorial Design trägt einzig dazu bei, die Koeffizienten der linearen Terme sowie der Interaktionsterme zu bestimmen.

⁴Ein kodiertes RSD ist orthogonal, wenn $\mathbf{X}^T \mathbf{X}$ eine Diagonalmatrix ist.

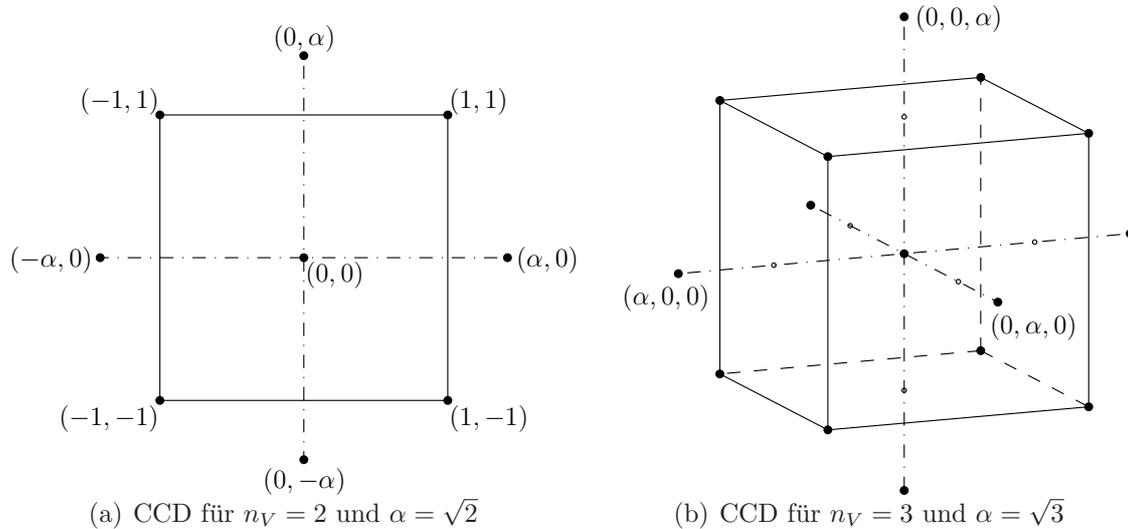


Abbildung 3.8: CCDs für zwei und drei Entwurfsvariablen

- Die auf den Achsen des Entwurfsvariablenraums lokalisierten Stützpunkte dienen zur Bestimmung der Koeffizienten der rein quadratischen Terme, nicht aber der Bestimmung der Interaktionsterme. Ohne sie kann lediglich ein Wert für die Summe der Koeffizienten der quadratischen Terme $\sum_{i=1}^{n_V} \beta_{ii}$ bestimmt werden.
- Mit den im Ursprung des Entwurfsvariablenraums lokalisierten Stützpunkten können Hinweise auf die Existenz nichtlinearer Effekte gewonnen werden. Sie tragen somit zur Bestimmung der Koeffizienten der quadratischen Terme bei.

Durch die Wahl der beiden Parameter α und n_C kann das CCD und dessen Einfluss auf die Approximation zum Teil deutlich beeinflusst werden. Die Wahl von α ist dabei primär geprägt vom Definitionsbereich der Entwurfsvariablen. Die Festlegung von n_C hat häufig Einfluss auf die Verteilung der Varianz der approximierten Systemantwort. Als Richtwert wird in [Myers und Montgomery 1995] $n_C = 1 \dots 2$ für $\alpha = 1.0$ und $n_C = 3 \dots 5$ für $\alpha = \sqrt{n_V}$ angegeben.

Die Abbildungen 3.8(a) und 3.8(b) zeigen CCDs für zwei und drei Entwurfsvariablen. Beim Fall $n_V = 2$ ist für α der Wert $\sqrt{2}$ gewählt; für $n_V = 3$ ist α gleich $\sqrt{3}$. Somit liegen beim CCD für zwei Entwurfsvariablen acht Stützstellen äquidistant auf einem Kreis. Dazu kommen noch die Stützstellen im Ursprung des Entwurfsvariablenraums. Für drei Entwurfsvariablen liegen 14 Stützstellen auf einer Kugel. Auch hier kommen noch die Center Runs hinzu.

Die Werte für α variieren im Allgemeinen zwischen 1.0 und $\sqrt{n_V}$. Wird für α der Wert $\sqrt{n_V}$ gewählt, liegen alle Stützstellen – außer die Center Runs – auf einer Kugel im n_V -dimensionalen Entwurfsvariablenraum. In diesem Fall werden zwei Varianten des CCD unterschieden. Umschreibt die Kugel den Definitionsbereich der Entwurfsvariablen, so trägt die Variante den Namen **Central Composite Circumscribed (CCC)**. Ist

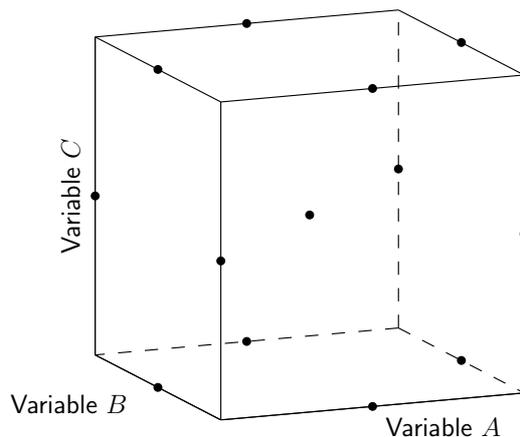


Abbildung 3.9: Box-Behnken Design für drei Entwurfsvariablen

die Kugel dem Definitionsbereich eingeschrieben, so nennt man die Variante **Central Composite Inscribed (CCI)**. Für $\alpha = 1.0$ liegen alle axialen Stützstellen im Zentrum der Seitenflächen des Würfels oder Hyperwürfels ($n_V > 3$). Dieses Variante des CCD trägt deshalb auch den Namen **Central Composite Faced (CCF)**.

Das zum CCD „konkurrierende“ RSD ist das Factorial Design dritter Ordnung. Für zwei Entwurfsvariablen erzeugen beide die gleiche Anzahl von Stützstellen. Wird beim CCD $\alpha = 1$ gesetzt, ist auch die Lage der Stützstellen identisch. Die Zahl der Stützstellen steigt jedoch beim Factorial Design deutlich stärker mit der Zahl der Entwurfsvariablen als dies beim CCD der Fall ist.

Box-Behnken Design

Das von Box und Behnken entwickelte gleichnamige Box-Behnken Design (BBD) besitzt im Gegensatz zum CCD nur drei Levels für jede Entwurfsvariable. Die Konstruktion des BBD ist abhängig von der Zahl der Entwurfsvariablen.

Für $n_V = 3 \dots 5$ wird für jedes Variablenpaar x_i und x_j ($j > i$ und $i, j \in [1 \dots n_V]$) ein Zwei-Level Factorial Design gebildet. Die verbleibenden Variablen x_k ($k \neq i, j$ und $k \in [1 \dots n_V]$) werden auf den Mittelwert ihres Definitionsbereichs gesetzt. Analog zum CCD werden die so bestimmten Stützstellen durch n_C Stützstellen im Ursprung ergänzt. Somit hat ein BBD für $n_V = 3 \dots 5$ allgemein $n_P = 4 \left[\frac{n_V!}{((n_V - 2)! 2!)} \right] + n_C$ Stützstellen. Wie beim CCD wird auch beim BBD vorgeschlagen, etwa $n_C = 3 \dots 5$ Center Runs zu benutzen. Abbildung 3.9 zeigt ein BBD für drei Entwurfsvariablen.

Im Fall $n_V \geq 6$ wird ein Factorial Design zweiter Ordnung mit drei statt nur zwei Variablen verwendet. Allerdings würde sich bei Verwendung aller Variablen-tripel eine sehr große Stützstellenzahl ergeben. Aus diesem Grund geben Box und Behnken bis zu

einer Variablenzahl $n_V \leq 16$ an, welche der Variablen-triplet verwendet werden sollen. Somit entspricht die Anzahl der generierten Stützstellen in etwa der des CCDs.

Wie auch beim CCD mit $\alpha = \sqrt{n_V}$, haben alle Stützstellen eines BBDs den gleichen Abstand vom Ursprung des kodierten Entwurfsvariablenraums. Vergleicht man die Abbildungen 3.8(b) und 3.9, so ist unschwer zu erkennen, dass das CCD mit den durch das Factorial Design definierten Stützstellen den gesamten Definitionsbereich der Designvariablen abdeckt. Dies ist beim BBD nicht der Fall, weshalb es in erster Linie nur dann zum Einsatz kommen sollte, wenn man mehr an einer guten Approximation in den inneren Bereichen des Designraums interessiert ist.

3.3.3 Optimale Response Surface Designs

Die bisher beschriebenen RSDs stellen in Bezug auf Lage und Anzahl der Stützstellen zum großen Teil bereits optimale Lösungen dar. Ein Factorial Design zweiter Ordnung ist beispielsweise das optimale RSD, wenn das Regressionsmodell nur lineare Terme und Interaktionsterme besitzt. Kommen noch quadratische Terme hinzu, so stellen CCD und Box-Behnken Design einen Satz von Stützstellen bereit, mit dem eine äußerst effiziente Bestimmung der Regressionskoeffizienten möglich ist. Was aber, wenn

- (a) bekannt ist, dass einige Terme absolut unnötig sind?
- (b) der Definitionsbereich der Entwurfsvariablen unregelmäßig geformt ist und möglicherweise Abhängigkeiten zwischen zulässigen Wertebereichen einer Entwurfsvariablen und den eingesetzten Werten anderer Entwurfsvariablen bestehen?
- (c) anstelle der Produkte der Entwurfsvariablen auch andere (z. B. trigonometrische, logarithmische, exponentielle, ...) Ansatzterme verwendet werden sollen?

In diesen Fällen führen starre RSDs dazu, dass

- (a) einige Stützstellen nur einen sehr unwesentlichen Beitrag zur Bestimmung der Regressionskoeffizienten liefern.
- (b) viele Stützwerte nicht bestimmbar sind oder der zulässige Designraum nur zu einem Bruchteil untersuchbar ist.
- (c) die Stützstellen keine optimale Verteilung in Hinblick auf eine robuste Bestimmung der Koeffizienten des Regressionsmodells haben.

Für diese Fälle stellt die RSM Kriterien zur Verfügung, die aus einem Pool von n_Q zulässigen Stützstellen (Designs) die n_P am besten geeigneten Stützstellen auswählen. Dies erfolgt in einem iterativen Prozess unter Berücksichtigung des verwendeten Regressionsmodells und weiterer Randbedingungen bzw. Ziele. Das bekannteste dieser Kriterien ist die D -Optimalität.

Die **D-Optimalität** basiert auf der Forderung nach einer möglichst guten und vor allem robusten Bestimmung der Regressionskoeffizienten. Hierzu werden aus n_Q Stützstellen diejenigen n_P ausgewählt, bei denen die Determinante der Informationsmatrix $\mathbf{X}^T \mathbf{X}$ maximal wird.

$$\max_{\zeta} |\mathbf{X}^T \mathbf{X}| \quad (3.44)$$

Der Parameter ζ kennzeichnet dabei alle möglichen Stützstellenkombinationen. Die Informationsmatrix ist proportional zur Inversen der Kovarianzmatrix $\text{Cov}(\mathbf{b}) = \sigma^2(\mathbf{X}^T \mathbf{X})^{-1}$, wodurch eine Maximierung von $|\mathbf{X}^T \mathbf{X}|$ gleichbedeutend ist mit der Minimierung der Determinante der Kovarianzmatrix der Regressionskoeffizienten. Ergebnis eines mittels der *D-Optimalität* generierten RSDs sind also Stützstellen, die kleine Varianzen und Kovarianzen für die Koeffizienten b erzeugen. Ein *D-optimales* Design wird meist über ein Zeilenaustauschverfahren bestimmt, da eine vollständige Enumeration aller möglichen Stützstellenkombinationen häufig nicht praktikabel ist.

Bei der **A-Optimalität** werden lediglich die Varianzen der Koeffizienten als Bemessungsgrundlage verwendet. Ziel ist somit ein Satz von Stützstellen für den die Spur der Kovarianzmatrix minimal wird.

$$\min_{\zeta} \sum_{i=1}^{n_P} \left[[\mathbf{X}^T \mathbf{X}]^{-1} \right]_{ii} \quad (3.45)$$

Aus dieser Definition ist leicht ersichtlich, dass der numerische Aufwand zum Generieren eines *A-optimales* Designs deutlich höher ist als bei der *D-Optimalität*. In (3.45) ist für jede, aus dem Pool der n_Q Stützstellen generierbare Kombination von n_P Stützstellen eine Matrix zu invertieren.

Ein weiteres Optimalitätskriterium ist die **G-Optimalität**. Sie hat zum Ziel, die maximale Varianz aller im Definitionsbereich \mathbb{D} der Entwurfsvariablen prognostizierten Systemantworten zu minimieren.

$$\min_{\zeta} \left[\max_{\mathbf{x} \in \mathbb{D}} \left[n_P \mathbf{x}^{(m)T} [\mathbf{X}^T \mathbf{X}]^{-1} \mathbf{x}^{(m)} \right] \right] \quad (3.46)$$

Hierbei reflektiert $\mathbf{x}^{(m)}$ sowohl die Lage eines Designs \mathbf{x} im Designraum \mathbb{D} , als auch die Eigenschaften des Regressionsmodells, indem $\mathbf{x}^{(m)}$ analog zu den Zeilen der Regressionsmatrix \mathbf{X} mittels der Ansatzterme gebildet wird. Auch bei der *G-Optimalität* ist der numerische Aufwand beträchtlich. Grund hierfür ist zum einen die erforderliche Invertierung der Informationsmatrix. Zum anderen kann die Varianz der prognostizierten Systemantworten jeweils nur in diskreten Punkten des Entwurfsvariablenraums bestimmt werden.

4 Anwendung der Response Surface Methode

Basierend auf den in Kapitel 3 dargestellten und diskutierten Methoden werden in diesem Kapitel Optimierungsaufgaben mit nichtlinearen Ziel- und Restriktionsfunktionen behandelt. Bei den durchzuführenden Untersuchungen soll damit dem nichtlinearen Charakter vieler multidisziplinärer Problemstellungen Rechnung getragen werden. Mit den gewonnenen Erkenntnissen erfolgt die Ableitung von Vorgehensweisen, die zu einer Verbesserung des mittels der Response Surface Approximationen (RSAs) ermittelbaren Optimums führen sollen. Da dies ohne zusätzliche, im Realfall aufwendige Systemanalysen erfolgen soll, bleiben die verwendeten Approximationsmodelle auf lineare und quadratische Ansatzterme beschränkt.

4.1 Optimierungsprobleme mit konvexen Zielfunktionen und Restriktionen

Sind die Zielfunktion(en) wie auch die Restriktionen konvex, so liegt die Vermutung nahe, dass hier Ansatzfunktionen mit linearen Termen, Interaktionstermen bis hin zum vollständigen quadratischen Ansatz mit Erfolg eingesetzt werden können. Auch wenn das reale Systemverhalten in größeren Bereichen teilweise einen nichtkonvexen Verlauf zeigt, so gilt dies zumindest für mehr oder weniger große Abschnitte. Somit ist bereits bei der Festlegung des zu approximierenden Bereichs besonderer Wert darauf zu legen, diesen so klein wie möglich, jedoch so groß wie nötig zu wählen.

Haben die realen Systemantworten tatsächlich einen quadratischen Verlauf, dann entspricht das Optimum der zugehörigen quadratischen RSA exakt dem Optimum des physikalischen Systems. Da dies in der praktischen Anwendung jedoch eher selten der Fall ist, soll die im Folgenden geführte Untersuchung mehr auf allgemeine konvexe Optimierungsaufgaben ausgerichtet sein.

4.1.1 Monotone Funktionsverläufe

Zunächst soll noch einmal das Beispiel „Biegebalken“ (siehe Abschnitt 3.1.3) betrachtet werden. Das Ziel der Optimierung soll sein, die maximale Durchsenkung in der

Balkenmitte zu minimieren. Zum Erreichen dieses Ziels kann sowohl die Höhe h als auch die Breite b des Rechteckquerschnitts in gewissen Grenzen verändert werden. Mit $h, b \in [5\text{mm}, 15\text{mm}]$ ist der Entwurfsvariablenraum festgelegt. Da Optimierungsprobleme in der Praxis fast immer durch Nebenbedingungen beschränkt sind, soll zur Formulierung der Optimierungsaufgabe eine obere Schranke für die Querschnittsfläche des Balken eingeführt werden ($A(h, b) \leq A_{max}$). Die Optimierungsaufgabe lautet somit:

$$\begin{aligned} \text{Minimiere} \quad & w_{max}(h, b) = \frac{Fl^3}{4E} \frac{1}{bh^3} \\ \text{so, dass} \quad & \frac{A(h, b)}{A_{max}} - 1 \leq 0 \quad \text{mit} \quad A(h, b) = hb \end{aligned} \quad (4.1)$$

Mit der Normierung der Restriktion(en) wird eine numerisch günstigere Situation erreicht, gerade wenn mehrere Restriktionen mit sehr unterschiedlicher Größenordnung zu berücksichtigen sind. Die aus der Beschränkung des Entwurfsraums resultierenden Nebenbedingungen sind in (4.1) nicht mit aufgeführt. Solche linearen Restriktionen werden bei der RSM – bei der ja letztendlich ein mathematischer Optimierungsalgorithmus das Optimum der Approximation bestimmt – auf die gleiche Weise berücksichtigt wie bei einer Optimierung bezüglich der realen Zielfunktion und Restriktionen. Der Verlauf der Ziel- und Restriktionsfunktion aus (4.1) ist zusammen mit den dabei verwendeten Werten in Abbildung 4.1 dargestellt.

Zum Vergleich der Effizienz der Optimierung mit Hilfe einer RSA wird (4.1) zunächst mit einem mathematischen Optimierungsalgorithmus gelöst. Somit steht auch das exakte Optimum des realen Problems zu Verfügung. Hierzu dient die MATLAB-Funktion `fmincon` (siehe [MATLAB Optimization Toolbox 2002]), die eine Lösung restringierter Probleme mit nichtlinearen Ziel- und Restriktionsfunktionen erlaubt. Der verwendete

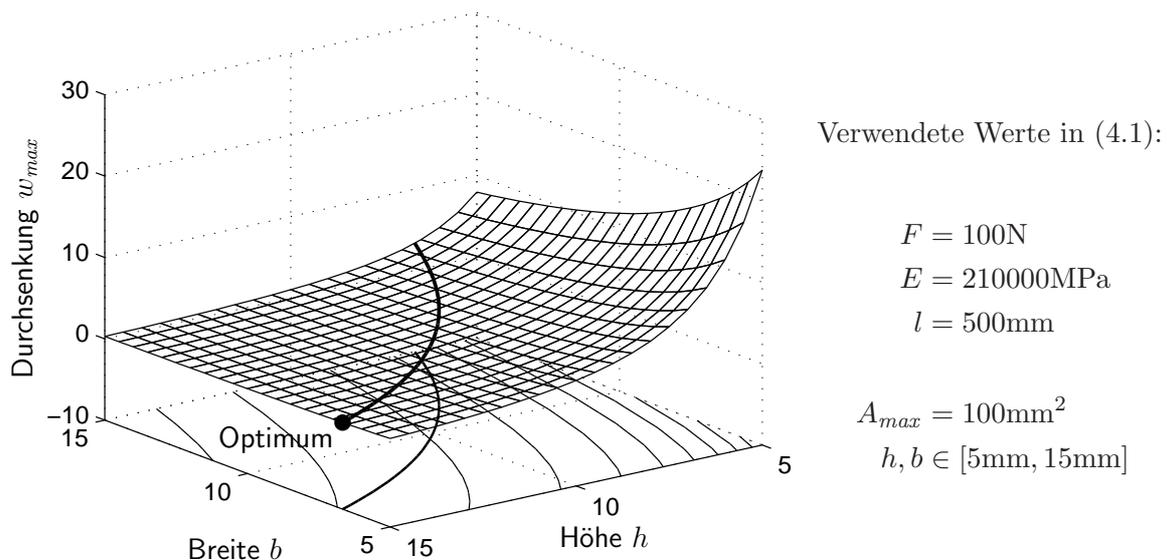


Abbildung 4.1: Optimierungsbeispiel „Biegebalken“

SQP	Startvektor		System- analysen	Optimum		
	h [mm]	b [mm]		w_{opt} [mm]	h [mm]	b [mm]
	[10 10]		21	15.0	6.67	0.66
	[5 5]		37	15.0	6.67	0.66
	[15 15]		24	15.0	6.67	0.66

RSM	RS-Design	Ansatz	System- analysen	Optimum		
	h [mm]	b [mm]		\hat{w}_{opt} [mm]	h [mm]	b [mm]
	2^2	linear	4	15.0	5.00	4.70
	2^2	linear mit Interakt.	4	15.0	6.67	0.78
	$3^2 \equiv$ CCF	quadratisch	9	13.4	7.45	-0.69
	CCC	quadratisch	9	13.7	5.00	-3.74
	CCI	quadratisch	9	13.2	7.57	0.28

Tabelle 4.1: Optimierungsbeispiel „Biegebalken“: Vergleich der benötigten Systemanalysen und des ermittelten Optimums bei einem SQP-Verfahren mit den realen Systemgleichungen und verschiedenen RSA, basierend auf unterschiedlichen DOE

te Algorithmus ist dabei ein SQP-Verfahren, wie es in Abschnitt 2.3.5 beschrieben wird.

Für verschiedene Startvektoren $\boldsymbol{x} = [h; b]$ sind in Tabelle 4.1 die Zahl der bis zum Erreichen des Optimums erforderlichen Funktionsauswertungen sowie das jeweils gefundene Optimum zusammengefasst. Das SQP-Verfahren findet von jedem Startvektor aus das exakte Optimum, wofür unterschiedlich viele Systemanalysen benötigt werden. Dabei werden Ziel- und Restriktionsfunktion gleich oft ausgewertet und als nur eine Systemanalyse gewertet. Die Funktionsauswertungen zur Behandlung der unteren und oberen Schranke für h und b werden dabei nicht gezählt. Startet man im Zentrum des Entwurfsraums, so müssen 21 Systemanalysen durchgeführt werden.

Bei den mit der RSM aufgestellten Approximationen variiert die Zahl der erforderlichen Systemanalysen weniger stark. Mit dem Response Surface Design (RSD) steht die Zahl der Analysen fest. Jetzt entscheidet die Qualität der Approximation über die Lage des Optimums. Im hier angeführten Beispiel wird das Optimum bereits mit einer einfachen RSA, die ausschließlich lineare Terme besitzt, relativ exakt ermittelt.

Nutzt man alle Informationen des 2^3 -Designs aus, so lässt sich noch der Interaktionsterm „ bh “ hinzufügen. In diesem Beispiel ist somit die Restriktionsfunktion exakt beschreibbar, wenn alle anderen Koeffizienten der Ansatzfunktion zu null gesetzt werden. Da sich mit dem Interaktionsterm zwar der Verlauf der Approximation der Zielfunktion ändert, nicht aber die globale Tendenz zum Rand des Entwurfsraums hin, wird hier sogar das exakte Optimum erreicht. Dies bedeutet jedoch nicht, dass das verwendete Approximationsmodell die Werte der Zielfunktion besonders gut prognostiziert.

Für einen vollständigen quadratischen Ansatz muss durch ein entsprechend geeignetes RSD die Zahl der Stützpunkte erhöht werden. Möglich sind hierbei ein 3-Level Factorial Design oder eine Variante der Central Composite Designs (CCDs) (ein Box-Behnken Design (BBD) ist für zwei Variablen nicht gegeben). Für zwei Entwurfsvariablen ist das Central Composite Faced (CCF) Design ($\alpha = 1$) identisch mit dem 3-Level Factorial Design. Auch wenn die Zahl der Stützpunkte bei den letztgenannten RSDs gleich ist, gibt es dennoch Unterschiede. Die RSDs für den vollständigen quadratischen Ansatz sind in Tabelle 4.2 mit allen zur Diskussion erforderlichen Informationen zusammengefasst.

Auch die quadratischen Ansätze, deren Koeffizienten jeweils über die Stützpunkte eines der genannten RSDs zweiter Ordnung angepasst wurden, dienen bei Optimierungsläufen als Ersatz für die Ziel- und Restriktionsfunktion. Während die Nebenbedingung weiterhin exakt approximiert wird, zeigt die approximierte Zielfunktion einen für die Optimumsuche ungünstigen Verlauf, was sich auch an den gefundenen Optima erkennen lässt (siehe Tabelle 4.1). Dies liegt vorrangig darin begründet, dass die Minima der quadratischen Approximationen immer innerhalb des Definitionsbereichs der Entwurfsvariablen liegen und somit der monoton fallende Charakter der realen Zielfunktion nicht korrekt wiedergegeben wird. Bezüglich der Qualität des Optimums kann für diesen Fall eine Tendenz festgestellt werden: Je größer der durch die Stützstellen abgedeckte Bereich des Designraums, desto besser ist das mit dem zugehörigen Approximationsmodell gefundene Optimum.

In der Praxis ist der reale Verlauf der Ziel- und Restriktionsfunktionen in der Regel nicht bekannt, so dass eine Beurteilung der Qualität der gefundenen Optima schwierig sein wird. Eine Möglichkeit ist, das gefundene optimale Design mit einer exakten Simulation zu analysieren und die Werte der Ziel- und Restriktionsfunktionen mit den approximierten Werten zu vergleichen. In diesem Beispiel genügt bereits eine Betrachtung der approximierten Werte der Zielfunktion. Bei allen quadratischen Approximationen werden negative Werte für die Durchsenkung der Balkenmitte prognostiziert, was physikalisch unmöglich und damit in Frage zu stellen ist.

Um einerseits Informationen über die Güte der Näherung zu erhalten und andererseits zu einer im Hinblick auf den Einsatz bei der Optimierung besseren Approximation zu gelangen, können die in Abschnitt 3.2 beschriebenen Kriterien eingesetzt werden.

Zunächst werden die globalen Beurteilungskriterien für den quadratischen Ansatz und die zugehörigen RSDs betrachtet (siehe Tabelle 4.2). Aufgrund der unterschiedlich großen Variationsbreite der Stützstellen ist allerdings kein direkter Vergleich möglich. Der *PRESS*-Wert ist jeweils ein Vielfaches der Summe der Fehlerquadrate SS_E , was ein deutliches Zeichen dafür ist, dass das Approximationsmodell den tatsächlichen Verlauf der Zielfunktion eher ungenügend annähert. Dies spiegelt sich auch in den Bestimmtheitsmaßen R^2 und R_{adj}^2 wieder. Diese sind beim 3-Level Factorial Design und besonders beim CCC Design relativ weit vom Idealwert 1 entfernt. Beim CCI Design können generell bessere Werte erzielt werden, da hier das Approximationsmodell innerhalb eines kleineren Bereichs angepasst wird. Betrachtet man das Maß für die

**3-Level Factorial Design bzw.
Central Composite Faced, 1 Center Run**

			SS_E	$PRESS$	R^2	R^2_{adj}	R^2_{pred}			
			19.0	226.0	0.9620	0.8987	0.5470			
p	Var_1	Var_2	$y_{(p)}$	$\hat{y}_{(p)}$	$e_{(p)}$	h_{pp}	$d_{(p)}$	$r_{(p)}$	$e_{[p]}$	$t_{(p)}$
1	5.00	5.00	23.81	21.96	1.851	0.806	0.737	1.670	9.52	5.153
2	5.00	10.00	11.90	13.53	-1.621	0.556	-0.645	-0.967	-3.65	-0.952
3	5.00	15.00	7.94	8.17	-0.231	0.806	-0.092	-0.208	-1.19	-0.171
4	10.00	5.00	2.98	5.41	-2.429	0.556	-0.966	-1.450	-5.47	-2.162
5	10.00	10.00	1.49	0.79	0.694	0.556	0.276	0.414	1.56	0.348
6	10.00	15.00	0.99	-0.74	1.735	0.556	0.690	1.035	3.90	1.055
7	15.00	5.00	0.88	0.30	0.578	0.806	0.230	0.521	2.97	0.446
8	15.00	10.00	0.44	-0.49	0.927	0.556	0.369	0.553	2.09	0.476
9	15.00	15.00	0.29	1.80	-1.504	0.806	-0.598	-1.357	-7.74	-1.784

Central Composite Circumscribed, 1 Center Run, ($\alpha = \sqrt{2}$)

			SS_E	$PRESS$	R^2	R^2_{adj}	R^2_{pred}			
			494.3	3514.8	0.8384	0.5691	-0.1492			
p	Var_1	Var_2	$y_{(p)}$	$\hat{y}_{(p)}$	$e_{(p)}$	h_{pp}	$d_{(p)}$	$r_{(p)}$	$e_{[p]}$	$t_{(p)}$
1	2.93	10.00	59.22	45.83	13.40	0.625	1.044	1.704	35.72	7.811
2	5.00	5.00	23.81	33.16	-9.35	0.625	-0.729	-1.190	-24.94	-1.337
3	5.00	15.00	7.94	19.91	-11.98	0.625	-0.933	-1.524	-31.94	-2.618
4	10.00	2.93	5.08	2.87	2.21	0.625	0.172	0.281	5.90	0.233
5	10.00	10.00	1.49	1.49	0.00	1.000	0.000	0.000	0.00	0.000
6	10.00	17.07	0.87	-5.06	5.93	0.625	0.462	0.754	15.81	0.684
7	15.00	5.00	0.88	-2.96	3.84	0.625	0.299	0.488	10.24	0.416
8	15.00	15.00	0.29	-0.92	1.21	0.625	0.094	0.154	3.23	0.126
9	17.07	10.00	0.30	5.56	-5.26	0.625	-0.410	-0.669	-14.02	-0.592

Central Composite Inscribed, 1 Center Run, ($\alpha = \sqrt{2}$)

			SS_E	$PRESS$	R^2	R^2_{adj}	R^2_{pred}			
			4.4	31.1	0.9669	0.9117	0.7645			
p	Var_1	Var_2	$y_{(p)}$	$\hat{y}_{(p)}$	$e_{(p)}$	h_{pp}	$d_{(p)}$	$r_{(p)}$	$e_{[p]}$	$t_{(p)}$
1	5.00	10.00	11.90	10.73	1.177	0.625	0.974	1.591	3.138	3.287
2	6.46	6.46	8.52	9.17	-0.652	0.625	-0.540	-0.881	-1.738	-0.836
3	6.46	13.54	4.07	5.25	-1.184	0.625	-0.981	-1.602	-3.158	-3.435
4	10.00	5.00	2.98	3.06	-0.083	0.625	-0.068	-0.112	-0.220	-0.091
5	10.00	10.00	1.49	1.49	0.000	1.000	0.000	0.000	0.000	0.000
6	10.00	15.00	0.99	0.32	0.671	0.625	0.555	0.907	1.788	0.869
7	13.54	6.46	0.93	0.33	0.597	0.625	0.494	0.807	1.591	0.744
8	13.54	13.54	0.44	0.38	0.064	0.625	0.053	0.086	0.170	0.071
9	15.00	10.00	0.44	1.03	-0.589	0.625	-0.488	-0.796	-1.570	-0.732

Tabelle 4.2: Informationen zum 3^2 -Design \equiv CCF, CCC und CCI für die Zielfunktion des Beispiels „Biegebalken“

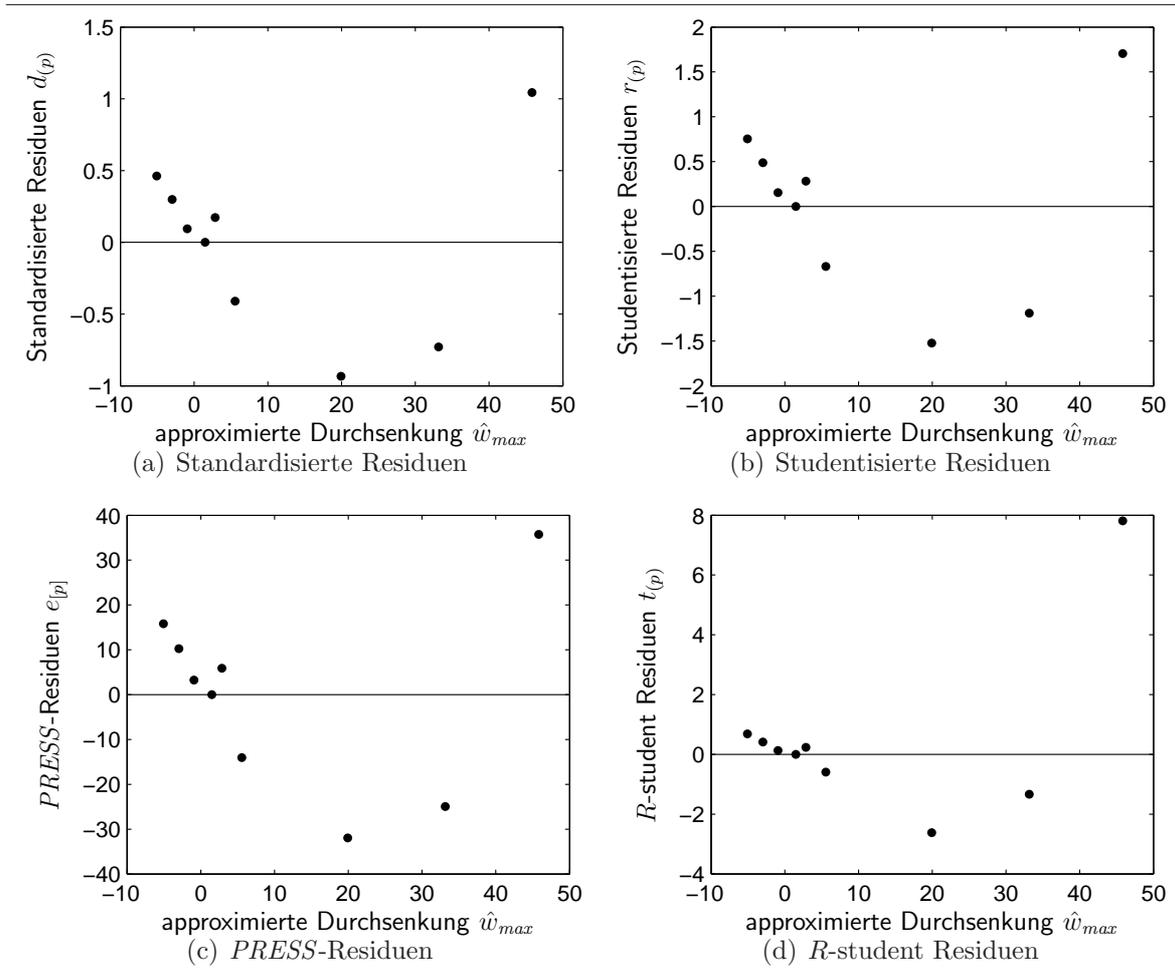


Abbildung 4.2: Skalierte Residuen für das Beispiel „Biegebalken“ und CCC-Design

Prognosegüte R^2_{pred} , so wird auch dies mit wachsender Variationsbreite der Stützstellen immer schlechter.

Um nicht gleich eine neue Approximation mit verkleinertem Definitionsbereich um das bis dahin ermittelte Optimum erstellen zu müssen, können zusätzlich die auf die einzelnen Stützstellen bezogenen Bewertungskriterien herangezogen werden (siehe Tabelle 4.2). Das Ziel dabei ist, Stützpunkte zu identifizieren, die einen großen Einfluss auf die Bestimmung der Koeffizienten des Approximationsmodells haben. Liegen diese Stützpunkte weit vom bisher ermittelten Optimum entfernt, so wirkt sich deren großer Einfluss eher negativ auf das Optimierungsergebnis aus. Zum Auffinden besagter Stützpunkte ist besonders die Analyse der Residuen von Interesse. Zur besseren Übersicht ist eine Auswahl von Residuenplots für das CCC Design in Abbildung 4.2 dargestellt. Auf der Ordinate sind die approximierten Werte für die Durchsenkung \hat{w}_{max} aufgetragen. Anhand der Abbildungen 4.2(a), 4.2(b) und 4.2(c) erkennt man eine häufige Eigenschaft der CCDs mit $\alpha = \sqrt{n_V}$. Die unterschiedlichen Skalierungen führen, abgesehen von den Wertebereichen, zu nahezu gleichen Verteilungen der

Punkte. Nur bei den R -student Residuen $t_{(p)}$ in Abbildung 4.2(d) fallen drei Punkte wesentlich deutlicher auf als bei den anderen Residuenplots. Dabei handelt es sich, beginnend mit dem stärksten Einfluss auf die Prognosegüte, um die Stützpunkte 1, 3 und 2. Da alle drei Stützstellen genügend weit vom bisherigen optimalen Design entfernt sind und es sich auch nicht um den wichtigen Center Run handelt, können diese aus dem RSD gestrichen werden.

Mit den verbleibenden sechs Stützpunkten kann der vollständige quadratische Ansatz erneut angepasst werden. So lässt sich mit dieser Vorgehensweise eine Näherung erstellen, welche die Funktionswerte im Bereich des vermutlichen Optimums wesentlich besser approximiert, ohne dabei auf weitere exakte Systemanalysen angewiesen zu sein. Die Werte der globalen Beurteilungskriterien sehen dann wie folgt aus:

SS_E	$PRESS$	R^2	R_{adj}^2	R_{pred}^2
0.0	3.5	1.0000	NaN	0.7864

Da das Approximationsmodell $n_K = 6$ frei bestimmbare Koeffizienten besitzt und auch nur ebenso viele Stützpunkte verwendet werden ($n_P = 6$), verläuft die RSA exakt durch die Stützpunkte. Deshalb ist die Summe der Fehlerquadrate $SS_E = 0.0$ und das Bestimmtheitsmaß R^2 erreicht den Maximalwert 1.0. Dass die approximierten Werte weiterhin einen Prognosefehler aufweisen, erkennt man an der Summe der Quadrate der $PRESS$ -Residuen sowie am Wert für R_{pred}^2 . Eine Bestimmung des adjustierten Bestimmtheitsmaßes R_{adj}^2 ist nicht möglich, da für den Fall $n_P = n_K$ in (3.28) der Term „ $\infty \cdot 0$ “ auszuwerten ist.

Führt man nun eine Optimierung mit der diesbezüglich verbesserten RSA durch, so erhält man als optimales Design $h = 14.9$ mm und $b = 6.73$ mm. Das prognostizierte Optimum ist dabei $\hat{w}_{opt} = 0.37$ mm. Somit wird das optimale Design mit einer Genauigkeit ≤ 0.1 mm bestimmt. Der Fehler bei der Prognose der minimalen Durchsenkung liegt bei 0.29 mm.

Fazit und Zusammenfassung

Für die Optimierung von Problemstellungen mit monoton fallenden oder steigenden Ziel- und Restriktionsfunktionen liefert die RSM relativ gute Approximationen. Dies gilt auch dann, wenn die Systemantworten einen nichtlinearen Verlauf aufweisen. Wie im obigen Beispiel gezeigt, benötigt auch ein als effizient zu bezeichnender Optimierungsalgorithmus wie die SQP mehr exakte Systemanalysen als ein vollständiger quadratischer Ansatz. Der Grund hierfür ist, dass mit der RSA globale Tendenzen der Ziel- und Restriktionsfunktionen relativ gut abgebildet werden können. Die mit lokalen Informationen arbeitenden mathematischen Optimierungsalgorithmen stoßen hierbei gegebenenfalls wiederholt in ungünstige Richtungen vor.

Im folgenden Abschnitt wird untersucht, wie gut sich RSAs bei Problemstellungen einsetzen lassen, deren Ziel- und Restriktionsfunktionen jeweils ein Extremum im Definitionsbereich der Entwurfsvariablen haben.

4.1.2 Funktionsverläufe mit einem Extremum im Definitionsbereich

Zur Definition der Problemstellung sei die in Abbildung 4.3 dargestellte Faserverbundplatte betrachtet. Diese besitzt einen symmetrischen Laminataufbau, was bedeutet, dass die Ausrichtung der Fasern sowie die Dicke der einzelnen Schichten ausgehend von der Symmetrieebene identisch sind. Die beiden äußeren Schichten sind dabei mit dem Winkel α_1 ausgerichtet, die beiden inneren Schichten sind um α_2 gegenüber der x -Achse gedreht. Dadurch ergibt sich bezüglich der Steifigkeiten der Platte ein anisotropes (richtungsabhängiges) Verhalten.

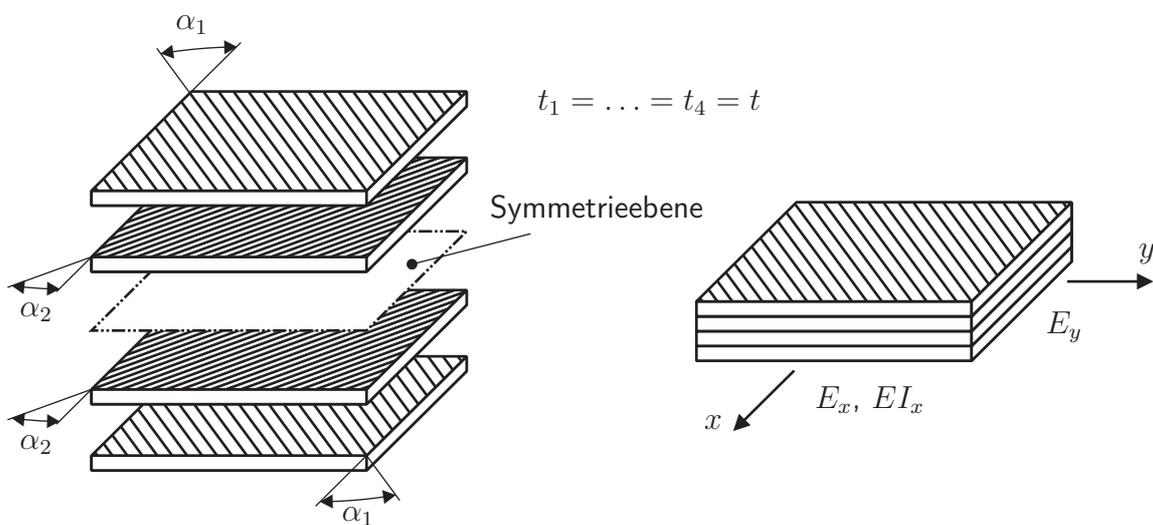


Abbildung 4.3: Optimierungsbeispiel „Faserverbundplatte“

Ziel der Optimierung ist es, die bezogene Biegesteifigkeit des Laminats um die y -Achse (EI_x) zu maximieren.¹ Ohne Restriktionen liegt das Optimum bei $[\alpha_1, \alpha_2] = [0^\circ, 0^\circ]$. Dies führt jedoch dazu, dass die bezogene Längssteifigkeit in y -Richtung E_y sehr klein gegenüber E_x wird. Deshalb wird als Nebenbedingung eingeführt, dass die Längssteifigkeit in y -Richtung nicht weniger als die Hälfte der Längssteifigkeit in x -Richtung werden darf. Die Optimierungsaufgabe lautet somit wie folgt:

$$\begin{aligned} &\text{Maximiere} && EI_x(\alpha_1, \alpha_2) \\ &\text{so, dass} && 1 - \frac{2 E_y(\alpha_1, \alpha_2)}{E_x(\alpha_1, \alpha_2)} \leq 0 \\ &\text{mit} && \alpha_1, \alpha_2 \in [-45^\circ, 45^\circ] \end{aligned} \quad (4.2)$$

Der Verlauf der Ziel- sowie der Restriktionsfunktion ist in Abbildung 4.4 dargestellt. Zur Berechnung der bezogenen Laminatsteifigkeiten wird eine Implementierung der

¹Der Index x bei EI_x bezieht sich auf die Berechnung der Krümmung um die y -Achse aus der Durchsenkung w : $\kappa_x = \partial^2 w / \partial x^2$

SQP	Startvektor		System- analysen	Optimum		
	α_1 [°]	α_2 [°]		EI_x [N]		
	[0 0]		6	<i>NaN</i>	<i>NaN</i>	<i>NaN</i>
	[45 45]		23	-2.3	42.2	3498
	[-45 45]		23	-2.3	42.2	3498

RSM	RS-Design	Ansatz	System- analysen	Optimum		
	α_1 [°]	α_2 [°]		\widehat{EI}_x [N]		
	2 ²	linear	4	<i>NaN</i>	<i>NaN</i>	<i>NaN</i>
	2 ²	linear mit Interakt.	4	<i>NaN</i>	<i>NaN</i>	<i>NaN</i>
	3 ² \equiv CCF	quadratisch	9	-0.40	39.1	3646
	CCC	quadratisch	9	-0.47	33.3	3835
	CCI	quadratisch	9	-4.06	40.6	3160

Tabelle 4.3: Optimierungsbeispiel „Faserverbundplatte“: Vergleich der benötigten Systemanalysen und des ermittelten Optimums bei einem SQP-Verfahren mit den realen Systemgleichungen und verschiedenen RSAs basierend auf unterschiedlichen DOEs

klassischen Laminattheorie [Baier 2002a] in MATLAB benutzt. Die beiden gekennzeichneten Optima sind dabei absolut gleichwertig. Die optimalen Entwürfe lassen sich durch einfaches Wenden der Faserverbundplatte erzeugen.

Auch diese Optimierungsaufgabe wurde mit dem in MATLAB integrierten SQP-Verfahren gelöst. Zum Vergleich sind die Ergebnisse in Tabelle 4.3 dargestellt. Mit dem Startvektor im Zentrum des Entwurfsraums $([\alpha_1, \alpha_2]) = [0^\circ, 0^\circ]$ kommt die Funktion `fmincon` nicht weiter. In diesem Punkt liegt das Maximum der bezogenen Biegesteifigkeit EI_x . Zusätzlich ist der Verlauf der Konturlinien zu diesem Punkt symmetrisch. Auch die Restriktionsfunktion liefert in diesem Fall keine Suchrichtung, da diese ebenfalls bei $[0^\circ, 0^\circ]$ ihren Extremwert annimmt und die Höhenlinien symmetrisch zur α_1 -Achse und α_2 -Achse verlaufen. Um dies festzustellen, werden sechs Systemanalysen durchgeführt, deren Ergebnisse nicht weiter benutzt werden können. Die beiden anderen Startvektoren ermöglichen eine exakte Bestimmung des Optimums. Dabei werden jedes Mal 23 Systemanalysen benötigt.

Bei der RSM kann angesichts des Verlaufs der Ziel- und Restriktionsfunktion auf die Anwendung eines linearen Approximationsansatzes verzichtet werden. Verwendet man das hierfür übliche 2-Level Factorial Design, liefert die Ziel- bzw. Restriktionsfunktion identische Werte in gegenüberliegenden bzw. in allen Ecken des Definitionsbereichs. Aufgrund des punktsymmetrischen Verlaufs wird hier auch kein anderes strukturiertes RSD Abhilfe schaffen. Die RSA ist somit für die Zielfunktion wie auch für die Nebenbedingung immer eine ebene Fläche, die zusätzlich noch parallel zur α_1 - α_2 -Ebene verläuft. Deshalb ist ein linearer Ansatz für dieses Problem absolut unbrauchbar. Auch die Einführung eines Interaktionsterms bringt keinen Erfolg. Beim approximierten Ver-

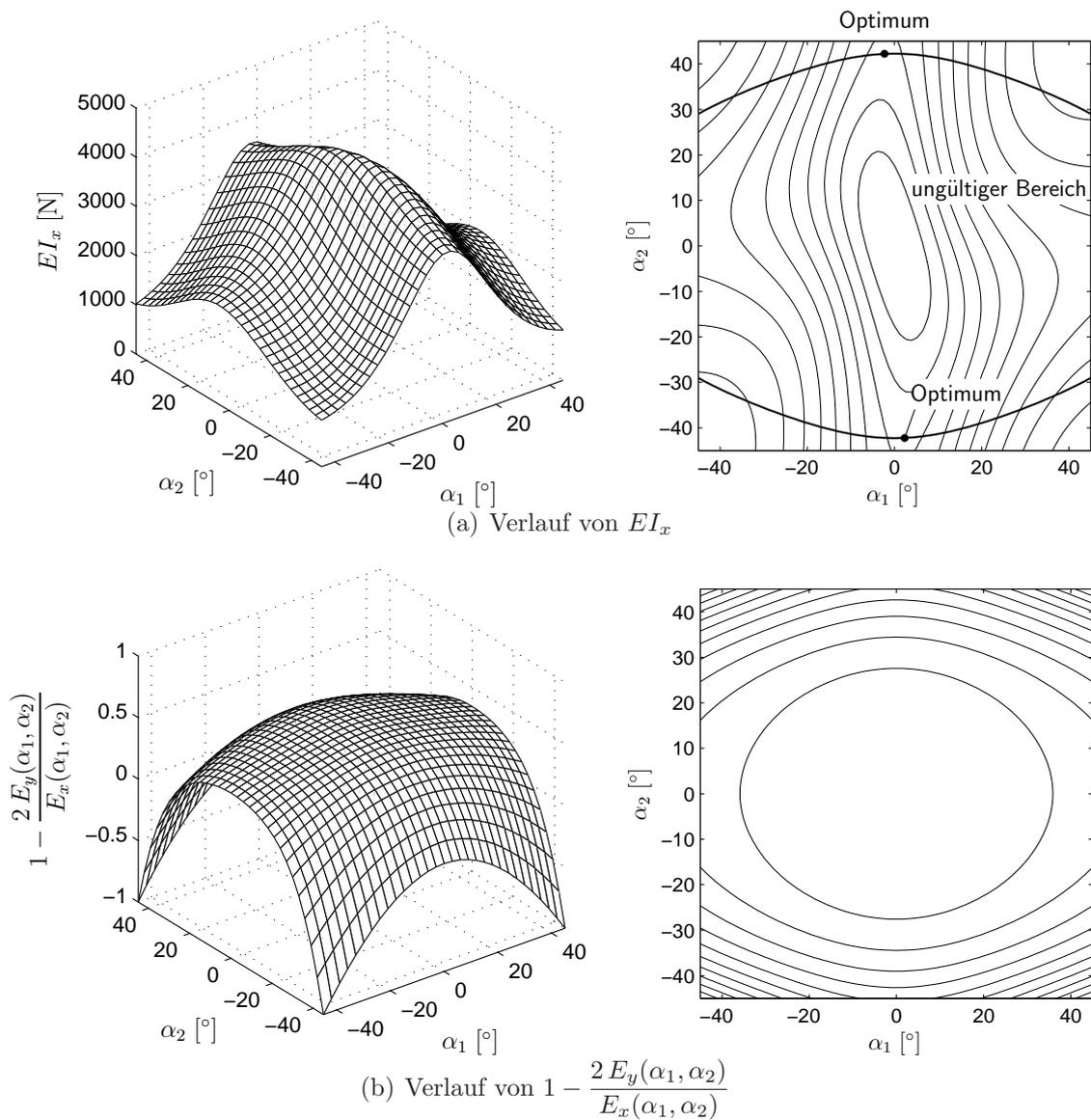


Abbildung 4.4: Verlauf der Zielfunktion und der Restriktionsfunktion für das Optimierungsbeispiel „Faserverbundplatte“

lauf \widehat{EI}_x bildet sich ein Rücken aus, der unter 45° zur α_1 -Achse und weiterhin parallel zur α_1 - α_2 -Ebene verläuft. Aufgrund der Symmetrie zur α_1 -Achse und α_2 -Achse bleibt die Approximation der Nebenbedingung weiterhin eben. Mit einem linearen Ansatz inklusive eines Interaktionsterms lässt sich somit kein Optimum bestimmen.

Erst ein quadratischer Ansatz bringt Erfolg. Bei der Optimierung mit den Approximationen wird das exakte Optimum mehr oder weniger gut getroffen. Dies hängt wieder vom jeweils zugrunde liegenden RSD ab. Dabei wird die Lage des Optimums mit der über die Stützpunkte des CCI Designs erstellten Approximation am besten bestimmt. Ähnlich gut gelingt dies mit den Stützstellen des 3-Level Factorial Designs (\equiv CCF Design). Zuletzt folgt die über das CCC Design aufgestellte Approximation. Zum Vergleich der RSDs sind die Werte der Bewertungskriterien in Tabelle 4.4 dargestellt.

Das schlechtere Abschneiden des CCC Designs liegt an dessen Eigenschaft, dass einige Stützstellen deutlich außerhalb des Definitionsbereichs der Entwurfsvariablen liegen. Wenn sich – so wie in diesem Fall – gerade dort der Verlauf der Zielfunktion stark verändert, wird die mit diesen Stützpunkten angepasste Approximation negativ beeinflusst.

Betrachtet man zunächst die Werte der globalen Beurteilungskriterien für die Approximation der Zielfunktion (siehe Tabelle 4.4), so bestätigt sich auch hier, dass der quadratische Ansatz die Stützpunkte des CCC Designs am schlechtesten approximieren kann. Dies gilt ebenfalls – wenn auch nicht so stark – für das CCI Design. Die Stützpunkte des 3-Level Factorial Designs können am besten approximiert werden, was sich an den überaus guten Werten R^2 und R_{adj}^2 zeigt. Auch die Prognosegüte scheint in diesem Fall fast nicht mehr zu verbessern zu sein. Dies ist allerdings mit den Erkenntnissen aus den beiden CCDs (mit $\alpha = \sqrt{2}$) als fehlerhafte Diagnose einzustufen.

Betrachtet man den Verlauf der Restriktionsfunktion in Abbildung 4.4(b), so kann dieser mit einem quadratischen Ansatz in seinen globalen Tendenzen gut approximiert werden. Für eine genauere Bestimmung der Lage des Optimums sollte jedoch auch der Funktionswert möglichst genau prognostiziert werden. Für die Nebenbedingung hängt dies entscheidend vom verwendeten RSD ab (siehe Tabelle 4.5). Betrachtet man die globalen Beurteilungswerte, so kann der verwendete quadratische Ansatz die Stützpunkte des CCI Designs hervorragend gut annähern. Dies gilt insbesondere für die prognostizierten Funktionswerte, was sich durch den sehr guten Wert für R_{pred}^2 zeigt. Gerade diesbezüglich hat der quadratische Ansatz mit den Stützpunkten der beiden (bzw. drei) anderen RSDs größere Schwierigkeiten.

Auch an diesem Beispiel wurde untersucht, inwieweit eine Reduktion der Stützpunkte eine Verbesserung des ermittelbaren Optimums ermöglicht. Dabei ist jedoch eine direkte Auswahl über die Werte eines oder mehrerer Beurteilungskriterien nicht möglich. Diese sind bei den Varianten des CCDs mit $\alpha = \sqrt{2}$ identisch (mit Ausnahme des

**3-Level Factorial Design bzw.
Central Composite Faced, 1 Center Run**

		SS_E	$PRESS$	R^2	R^2_{adj}	R^2_{pred}				
		4583.0	34091.9	0.9997	0.9991	0.9975				
p	Var_1	Var_2	$y_{(p)}$	$\hat{y}_{(p)}$	$e_{(p)}$	h_{pp}	$d_{(p)}$	$r_{(p)}$	$e_{[p]}$	$t_{(p)}$
1	-45.00	-45.00	946	957	-11.28	0.806	-0.289	-0.655	-58.0	-1.000
2	-45.00	0.00	1988	1965	22.57	0.556	0.577	0.866	50.8	1.000
3	-45.00	45.00	1022	1033	-11.28	0.806	-0.289	-0.655	-58.0	-1.000
4	0.00	-45.00	3432	3410	22.57	0.556	0.577	0.866	50.8	1.000
5	0.00	0.00	4334	4380	-45.13	0.556	-1.155	-1.732	-101.5	∞
6	0.00	45.00	3432	3410	22.57	0.556	0.577	0.866	50.8	1.000
7	45.00	-45.00	1022	1033	-11.28	0.806	-0.289	-0.655	-58.0	-1.000
8	45.00	0.00	1988	1965	22.57	0.556	0.577	0.866	50.8	1.000
9	45.00	45.00	946	957	-11.28	0.806	-0.289	-0.655	-58.0	-1.000

Central Composite Circumscribed, 1 Center Run, ($\alpha = \sqrt{2}$)

		SS_E	$PRESS$	R^2	R^2_{adj}	R^2_{pred}				
		5337378.1	37954689.0	0.5912	-0.0901	-1.9070				
p	Var_1	Var_2	$y_{(p)}$	$\hat{y}_{(p)}$	$e_{(p)}$	h_{pp}	$d_{(p)}$	$r_{(p)}$	$e_{[p]}$	$t_{(p)}$
1	-63.64	0.00	1909	1092	817	0.625	0.612	1.000	2178	1.000
2	-45.00	-45.00	946	1763	-817	0.625	-0.612	-1.000	-2178	-1.000
3	-45.00	45.00	1022	1838	-817	0.625	-0.612	-1.000	-2178	-1.000
4	0.00	-63.64	3326	2509	817	0.625	0.612	1.000	2178	1.000
5	0.00	0.00	4334	4334	0	1.000	0.000	0.000	0	0.000
6	0.00	63.64	3326	2509	817	0.625	0.612	1.000	2178	1.000
7	45.00	-45.00	1022	1838	-817	0.625	-0.612	-1.000	-2178	-1.000
8	45.00	45.00	946	1763	-817	0.625	-0.612	-1.000	-2178	-1.000
9	63.64	0.00	1909	1092	817	0.625	0.612	1.000	2178	1.000

Central Composite Inscribed, 1 Center Run, ($\alpha = \sqrt{2}$)

		SS_E	$PRESS$	R^2	R^2_{adj}	R^2_{pred}				
		2500545.8	17781658.8	0.7243	0.2648	-0.9605				
p	Var_1	Var_2	$y_{(p)}$	$\hat{y}_{(p)}$	$e_{(p)}$	h_{pp}	$d_{(p)}$	$r_{(p)}$	$e_{[p]}$	$t_{(p)}$
1	-45.00	0.00	1988	1429	559	0.625	0.612	1.000	1491	1.000
2	-31.82	-31.82	1344	1903	-559	0.625	-0.612	-1.000	-1491	-1.000
3	-31.82	31.82	1839	2398	-559	0.625	-0.612	-1.000	-1491	-1.000
4	0.00	-45.00	3432	2873	559	0.625	0.612	1.000	1491	1.000
5	0.00	0.00	4334	4334	0	1.000	0.000	0.000	0	0.000
6	0.00	45.00	3432	2873	559	0.625	0.612	1.000	1491	1.000
7	31.82	-31.82	1839	2398	-559	0.625	-0.612	-1.000	-1491	-1.000
8	31.82	31.82	1344	1903	-559	0.625	-0.612	-1.000	-1491	-1.000
9	45.00	0.00	1988	1429	559	0.625	0.612	1.000	1491	1.000

Tabelle 4.4: Bewertungskriterien zum 3^2 -Design \equiv CCF, CCC und CCI für die Zielfunktion des Beispiels „Faserverbundplatte“

**3-Level Factorial Design bzw.
Central Composite Faced, 1 Center Run**

		SS_E	$PRESS$	R^2	R^2_{adj}	R^2_{pred}				
		0.8	6.0	0.7108	0.2287	-1.1514				
p	Var_1	Var_2	$y_{(p)}$	$\hat{y}_{(p)}$	$e_{(p)}$	h_{pp}	$d_{(p)}$	$r_{(p)}$	$e_{[p]}$	$t_{(p)}$
1	-45.00	-45.00	-1.000	-0.851	-0.149	0.806	-0.289	-0.655	-0.769	-1.00
2	-45.00	0.00	0.345	0.046	0.299	0.556	0.577	0.866	0.673	1.00
3	-45.00	45.00	-1.000	-0.851	-0.149	0.806	-0.289	-0.655	-0.769	-1.00
4	0.00	-45.00	-0.123	-0.421	0.299	0.556	0.577	0.866	0.673	1.00
5	0.00	0.00	-0.123	0.475	-0.598	0.556	-1.155	-1.732	-1.345	$-\infty$
6	0.00	45.00	-0.123	-0.421	0.299	0.556	0.577	0.866	0.673	1.00
7	45.00	-45.00	-1.000	-0.851	-0.149	0.806	-0.289	-0.655	-0.769	-1.00
8	45.00	0.00	0.345	0.046	0.299	0.556	0.577	0.866	0.673	1.00
9	45.00	45.00	-1.000	-0.851	-0.149	0.806	-0.289	-0.655	-0.769	-1.00

Central Composite Circumscribed, 1 Center Run, ($\alpha = \sqrt{2}$)

		SS_E	$PRESS$	R^2	R^2_{adj}	R^2_{pred}				
		0.4	3.0	0.8424	0.5796	-0.1210				
p	Var_1	Var_2	$y_{(p)}$	$\hat{y}_{(p)}$	$e_{(p)}$	h_{pp}	$d_{(p)}$	$r_{(p)}$	$e_{[p]}$	$t_{(p)}$
1	-63.64	0.00	0.135	-0.093	0.228	0.625	0.612	1.000	0.608	1.000
2	-45.00	-45.00	-1.000	-0.772	-0.228	0.625	-0.612	-1.000	-0.608	-1.000
3	-45.00	45.00	-1.000	-0.772	-0.228	0.625	-0.612	-1.000	-0.608	-1.000
4	0.00	-63.64	-1.223	-1.451	0.228	0.625	0.612	1.000	0.608	1.000
5	0.00	0.00	-0.123	-0.123	0.000	1.000	0.000	0.000	0.000	0.000
6	0.00	63.64	-1.223	-1.451	0.228	0.625	0.612	1.000	0.608	1.000
7	45.00	-45.00	-1.000	-0.772	-0.228	0.625	-0.612	-1.000	-0.608	-1.000
8	45.00	45.00	-1.000	-0.772	-0.228	0.625	-0.612	-1.000	-0.608	-1.000
9	63.64	0.00	0.135	-0.093	0.228	0.625	0.612	1.000	0.608	1.000

Central Composite Inscribed, 1 Center Run, ($\alpha = \sqrt{2}$)

		SS_E	$PRESS$	R^2	R^2_{adj}	R^2_{pred}				
		0.0000002	0.0000014	0.9999995	0.9999987	0.9999965				
p	Var_1	Var_2	$y_{(p)}$	$\hat{y}_{(p)}$	$e_{(p)}$	h_{pp}	$d_{(p)}$	$r_{(p)}$	$e_{[p]}$	$t_{(p)}$
1	-45.00	0.00	0.345	0.346	-1.6e-4	0.625	-0.612	-1.000	-4.1e-4	-1.000
2	-31.82	-31.82	0.112	0.112	1.6e-4	0.625	0.612	1.000	4.1e-4	1.000
3	-31.82	31.82	0.112	0.112	1.6e-4	0.625	0.612	1.000	4.1e-4	1.000
4	0.00	-45.00	-0.123	-0.122	-1.6e-4	0.625	-0.612	-1.000	-4.1e-4	-1.000
5	0.00	0.00	0.547	0.547	0.00000	1.000	0.000	0.000	0.00000	0.000
6	0.00	45.00	-0.123	-0.122	-1.6e-4	0.625	-0.612	-1.000	-4.1e-4	-1.000
7	31.82	-31.82	0.112	0.112	1.6e-4	0.625	0.612	1.000	4.1e-4	1.000
8	31.82	31.82	0.112	0.112	1.6e-4	0.625	0.612	1.000	4.1e-4	1.000
9	45.00	0.00	0.345	0.346	-1.6e-4	0.625	-0.612	-1.000	-4.1e-4	-1.000

Tabelle 4.5: Bewertungskriterien zum 3^2 -Design \equiv CCF, CCC und CCI für die Restriktionsfunktion des Beispiels „Faserverbundplatte“

Center Runs). Auch beim 3-Level Factorial Design gibt es kein deutliches Auswahlkriterium. Deshalb wird folgende Vorgehensweise angewendet:

- Zunächst wird der Stützpunkt an der Stelle entfernt, die den größten Abstand zum bisher ermittelten Optimum besitzt.
- Weiterhin werden auch die Stützpunkte ausgegrenzt, die zur bereits entfernten Stützstelle um 90° versetzt liegen.

Für das jeweilige RSD sind dies die in nachfolgender Tabelle dargestellten Punkte. Hier sind auch die mit den neu angepassten Approximationsmodellen ermittelten Optima aufgelistet.

RS-Design	entfernte Stützpunkte	Optimum		
		α_1 [°]	α_2 [°]	\widehat{EI}_x [N]
$3^2 \equiv$ CCF	4,2,8	-0.4	40.7	3597
CCC	1,6,9	-0.2	33.3	3903
CCI	1,6,9	-2.3	40.7	3609

Durch das Entfernen der Stützpunkte aus den RSDs wird die Lage des Optimums beim 3-Level Factorial Design und beim CCI Design besser bestimmt. Dies gilt auch für das jeweils prognostizierte Optimum. Beim CCC Design wird sowohl die Lage des Optimums als auch der hierfür prognostizierte Wert geringfügig schlechter bestimmt.

Fazit und Zusammenfassung

Bei der untersuchten Optimierungsaufgabe wird deutlich, wie wichtig die in Kapitel 3 beschriebene Möglichkeit ist, eine RSA sequenziell aufzubauen. Durch die Informationen des 2-Level Factorial Designs wird klar, dass zusätzliche Daten benötigt werden. Mit der Kenntnis oder Vermutung zugrunde liegender physikalischer Zusammenhänge könnte in diesem Fall das CCC Design bereits im Vorfeld ausgeschlossen werden. Dies bestätigt sich auch anhand der ermittelten Ergebnisse bezüglich der Lage und des prognostizierten Werts des Optimums.

Mit den RSAs, die mittels der Stützpunkte der verbleibenden RSDs angepasst werden, wird die Lage des Optimums relativ gut bestimmt ($\Delta\alpha \leq 2.3^\circ$). Dies entspricht weniger als 2.6% des Definitionsbereichs. Der prognostizierte Wert $\widehat{EI}_{x_{opt}}$ verfehlt das exakte Optimum um maximal 338 N und somit um weniger als 10%.

Durch die Reduktion des 3-Level Factorial Designs bzw. des CCI Designs um Stützstellen, die weit vom bis dahin ermittelten Optimum entfernt liegen, lässt sich eine Verbesserung bezüglich des dann ermittelbaren Optimums erreichen. Es wird die Stützstelle entfernt, die punktsymmetrisch zu der dem Optimum nächstgelegenen Stützstelle liegt. Weiterhin werden die zwei um $\pm 90^\circ$ versetzt liegenden Stützstellen nicht verwendet. Damit sinkt der Fehler bei der Bestimmung der Lage des Optimums auf weniger als

1.7° (\cong 1.9%). Der Fehler beim prognostizierten Wert $\widehat{EI}_{x_{opt}}$ liegt nur noch bei 111 N bzw. bei 3.2%.

4.2 Anmerkung zu nichtkonvexen Optimierungsaufgaben

Die Nichtkonvexität bei Optimierungsaufgaben hat zur Folge, dass mehrere Lösungen existieren können. Dies kann durch eine nichtkonvexe Zielfunktion und/oder einen nichtkonvexen zulässigen Bereich bedingt sein, wie in Abbildung 4.5 deutlich wird. Dort ist ebenfalls zu erkennen, dass die Funktionsverläufe im dargestellten Bereich nicht durch eine quadratische Approximation wiedergegeben werden können. Dies ist nur innerhalb kleinerer Abschnitte zu gewährleisten. Da diese nicht immer von vornherein eingegrenzt werden können, wird in solchen Fällen ein iteratives Vorgehen nicht vermeidbar sein.

In der Literatur werden dazu verschiedene Verfahren beschrieben, die im Wesentlichen die Aufgabe haben, den approximierten Bereich um ein Optimum herum einzugrenzen (z. B. [Rodríguez u. a. 2000]). Vorgehensweisen hierzu wurden in [Thalhauser 2000] realisiert und untersucht. Eine schrittweise Unterteilung des Definitionsbereichs führte dabei zu sehr vielen Systemanalysen, obwohl Stützpunkte in benachbarten Bereichen wiederverwendet wurden. Bei der schrittweisen Verschiebung und Verkleinerung des Definitionsbereichs kam es vor, dass nur eine lokale Lösung ermittelt werden konnte.

So entstehen bei der Optimierung nichtkonvexer Problemstellungen mit Hilfe von RSAs ähnliche Probleme wie bei den mathematischen Optimierungsalgorithmen, auch

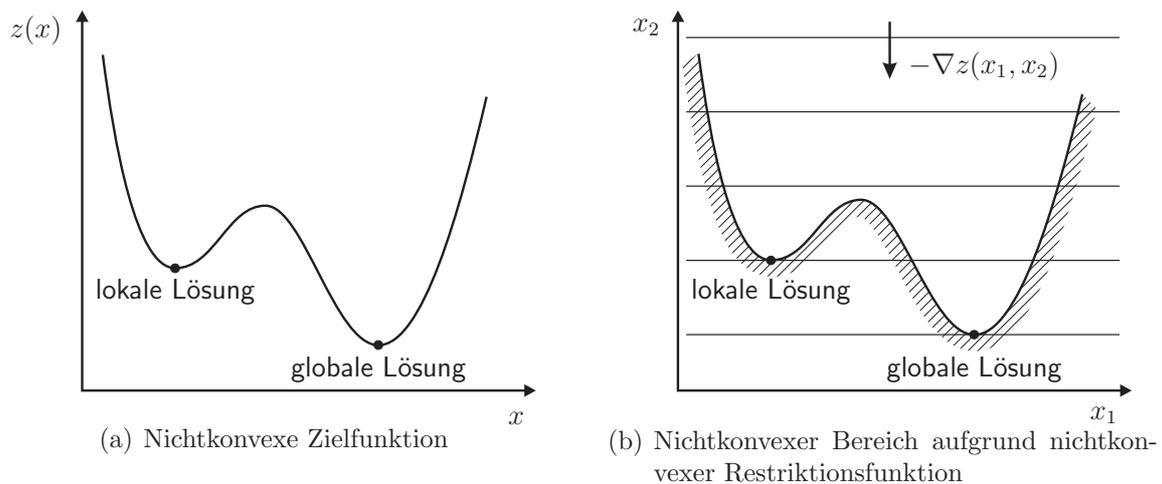


Abbildung 4.5: Lokale Lösungen bei nichtkonvexen Funktionsverläufen

wenn der Aufwand dabei geringer ist, da ein Teil der bereits durchgeführten Systemanalysen wiederholt verwendet werden kann. Aus diesem Grund beschäftigt sich Kapitel 5 mit einem Verfahren zur Behandlung größerer Nichtlinearitäten, das auch in der Lage ist, Approximationen für nichtkonvexe Funktionsverläufe zu erstellen. Hierzu sind allerdings mehr Stützpunkte und damit mehr Systemanalysen erforderlich als für die Approximation konvexer Ziel- und Restriktionsfunktionen.

Bezüglich der Nichtkonvexität des zulässigen Bereichs sei abschließend noch angemerkt, dass diese auch durch mehrere konvexe Restriktionsfunktionen bedingt sein kann. Dies wird in Abbildung 4.6 verdeutlicht.

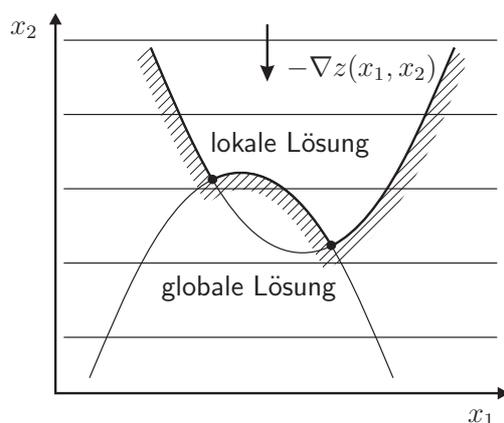


Abbildung 4.6: Durch konvexe Restriktionsfunktionen erzeugter nichtkonvexer Bereich

Streng genommen wird die Nichtkonvexität des zulässigen Bereichs dabei durch die Kombination eines konvexen und eines konkaven Funktionsverlaufs erzeugt. Beide lassen sich mit den in den vorangegangenen Abschnitten erarbeiteten und im Folgenden zusammengefassten Verfahrensweisen sehr gut behandeln.

4.3 Zusammenfassung und Ableitung von Verfahrensweisen

Bei den in Abschnitt 4.1 untersuchten Optimierungsaufgaben wird deutlich, dass der in Kapitel 3 beschriebenen Möglichkeit, eine RSA sequenziell aufzubauen, gerade im Hinblick auf die Verwendung möglichst weniger Stützpunkte (Systemanalysen) eine Bedeutung zukommt. Die zunächst mit relativ geringem Analyseaufwand zu gewinnenden Grundinformationen können dabei gut als Entscheidungsgrundlage für weitere Schritte eingesetzt werden. Bis dahin verwendete Stützpunkte können bei kompatiblen RSDs weiterhin genutzt werden.

Zusammenfassend kann folgende Vorgehensweise vorgeschlagen werden:

- Zunächst wird ein lineares Approximationsmodell (ggf. mit Interaktionstermen) unter Verwendung eines 2-Level Factorial Designs angepasst. Bei vielen Entwurfsvariablen sind 2-Level Fractional Factorial Designs mit einem wesentlich geringeren Rechenaufwand verbunden. Sind die Werte für die Einflüsse \mathcal{E} der ermittelbaren Variableneffekte sehr klein oder gar null, kann vermutet werden, dass eine Symmetrie beim zu approximierenden Funktionsverlauf vorliegt und weitere Systeminformationen erforderlich sind.
- Werden einflussreiche Effekte ermittelt, so können und sollen diese mittels eines quadratischen Ansatzes und einem passenden RSD näher untersucht werden. Sind die Stützwerte dabei für eine oder mehrere Entwurfsvariablen monoton fallend, so ist ein CCC Design die richtige Wahl, da hierbei globale Tendenzen auch außerhalb des Definitionsbereichs erfasst werden. Dabei ist es prinzipiell möglich, dass ein Approximationsmodell mit linearen Termen und Interaktionstermen die Lage des Optimums besser wiedergibt.
- Vermutet man – gestützt durch physikalische Einsicht – das Extremum der zu approximierenden Systemantwort eher innerhalb des Definitionsbereichs der Entwurfsvariablen, so ist das CCI Design die geeignetste Methode zum Generieren der Stützstellen für einen quadratischen Ansatz. Umgekehrt kann das CCC Design verwendet werden, wenn das Optimum des zu approximierenden Funktionsverlaufs relativ sicher am Rand des Entwurfsvariablenraums erwartet wird. Bestehen diesbezüglich Unsicherheiten, stellt das 3-Level Factorial Design bzw. das CCF Design eine konservative Wahl dar.
- Kann davon ausgegangen werden, dass sich der Verlauf der zu approximierenden Systemantwort außerhalb des Definitionsbereichs der Entwurfsvariablen stark verändert, so ist vom CCC Design abzuraten. Dies sollte auch dann gelten, wenn man sich dessen nicht sicher ist.
- Lässt man bei der Anpassung der Approximationsmodelle ungeeignete Stützpunkte weg, so kann meist ein genaueres Optimum bestimmt werden. Als Entscheidungshilfe hierfür dienen die in Abschnitt 3.2 beschriebenen Bewertungskriterien. Vorrangig handelt es sich dabei um die skalierten Residuen und um Größen, die den Einfluss jedes Stützpunkts auf die Approximation bewerten. Liegen die auf diese Weise identifizierten Stützpunkte ausreichend weit vom Optimum entfernt, können diese aus der Regressionsmatrix gestrichen werden.
- Die Untersuchungen zeigen, dass es nicht „das“ RSD gibt. Will man jedoch ein für möglichst viele Problemstellungen geeignetes Stützstellen-Design für quadratische Ansatzfunktionen, so kann als Empfehlung das CCF Design sowie das 3-Level Factorial Design angegeben werden. Bei einer größeren Zahl von Entwurfsvariablen ist das CCF Design vorzuziehen.
- Benötigt man ein genaueres Optimum, so kann der Definitionsbereich im Bereich des bisher gefundenen Optimums deutlich verkleinert werden.

Bezüglich der Genauigkeit von Lage und Wert des ermittelbaren Optimums sind noch folgende Bemerkungen angebracht:

Bei Optimierungsproblemen mit sehr zeitaufwendigen Systemanalysen, wie sie gerade auch bei multidisziplinären Problemstellungen anzutreffen sind, ist die primäre Aufgabe nicht unbedingt die Bestimmung eines exakten Optimums. Hier geht es vielmehr darum, mit möglichst wenigen Systemanalysen so nah wie möglich an das Optimum heranzukommen. Bei extrem komplexen Problemstellungen genügt es teilweise, lediglich einen verbesserten Entwurf zu erhalten.

Dies kann mit mathematischen Optimierungsverfahren nicht gewährleistet werden. Bei diesem ist nur ein iteratives Herantasten an das Optimum möglich. Werden dabei zu große Schrittweiten gewählt, so verläuft der Lösungspfad alles andere als optimal. Zwar können die Toleranzen für Abbruchkriterien stark gelockert werden, was in Grenzen zu einer Reduktion der erforderlichen Systemanalysen führt. Dann besteht jedoch die Gefahr, dass das Optimierungsverfahren an Punkten „stehen bleibt“, bei denen zwar die Abbruchkriterien erfüllt werden, diese aber noch keinerlei Optimum darstellen.

Generell ist beim Aufstellen einer RSA auf eine geeignete Wahl der zu approximierenden Größe zu achten. Anstatt zum Beispiel die von einer Normalkraft $N(\mathbf{x})$ und den Entwurfsvariablen \mathbf{x} abhängige Spannung direkt zu approximieren, kann dies auch für die Schnittkraft selbst erfolgen. Lässt sich die Schnittkraft besser approximieren und sind die Spannungen daraus analytisch bestimmbar, führt dies zu einer wesentlich besseren Approximation. Analoges gilt auch für Aufgaben aus der Dynamik. Hier können zum Beispiel die approximierten Eigenwerte/Eigenfrequenzen verwendet werden, um die Verschiebungsamplituden bei der Berechnung einer harmonischen Antwort explizit zu bestimmen. Diese Überlegungen gelten auch für Entwurfsvariablen, deren Einfluss auf eine Systemantwort mit einer RSA beschrieben werden soll. So lassen sich zum Beispiel Querschnittabmessungen mit der Querschnittsfläche oder einem Flächenträgheitsmoment zusammenfassen. Diese können dann anstelle der einzelnen Querschnittsabmessungen verwendet werden.

5 Erweiterung der Response Surface Methode

Für Ziel- und Restriktionsfunktionen mit nur einem Extremum innerhalb des approximierten Bereichs stellt die Response Surface Methode (RSM) ein effizientes Werkzeug zur Approximation der zugrunde liegenden technischen und physikalischen Zusammenhänge dar. Gerade wenn diese numerisch nur sehr aufwendig ermittelbar sind, können die aufgestellten Approximationen bei einer Optimierung gut als Ersatz für die Ziel- und Restriktionsfunktionen eingesetzt werden.

In den vorangegangenen Untersuchungen hat sich jedoch gezeigt, dass es bei funktionalen Zusammenhängen, die mit linearen und quadratischen Termen nicht ausreichend beschrieben werden können, zu Problemen kommen kann. Obwohl die Lage des Optimums hierbei noch relativ gut ermittelt werden kann, gilt dies meist weniger für den Funktionswert des Optimierungsziels oder der Nebenbedingungen. Ist der Verlauf der zu approximierenden Systemantwort innerhalb des Definitionsbereichs nichtkonvex und besitzt somit zwei oder mehr Extrema, so ist eine Optimierung mittels quadratischer Approximation selten auf Anhieb erfolgreich. Hier müssen andere funktionale Zusammenhänge als Terme des Approximationsmodells eingesetzt werden.

Soll eine Response Surface Approximation (RSA) bei multidisziplinären Analysen als Ersatz für die numerische Analyse einer Teildisziplin eingesetzt werden, müssen die Funktionswerte ebenfalls möglichst exakt wiedergegeben werden. Auch dies ist mit einem quadratischen Ansatz nicht zu bewerkstelligen. Im Folgenden werden Möglichkeiten vorgestellt, die diesbezüglich eine Verbesserung bringen können.

5.1 Transformation der Systemantwort

Myers und Montgomery [1995] wie auch Box und Draper [1987] schlagen vor, die zu approximierende Systemantwort y zu transformieren. Beide gehen dabei den Weg über die so genannte „Power Transformation“. Hierzu wird die Systemantwort mittels der Vorschrift

$$w = \frac{y^\lambda - 1}{\lambda} \tag{5.1}$$

transformiert. Mit dem üblicherweise aus dem Intervall $[-2, 2]$ gewählten Parameter λ lässt sich eine Reihe von Transformationsfunktionen definieren, beispielsweise:

$$\begin{array}{cccccc} \lambda = 1 & \lambda = \frac{1}{2} & \lambda \rightarrow 0 & \lambda = -\frac{1}{2} & \lambda = -1 & \dots \\ w \sim y & w \sim \sqrt{y} & w \sim \ln y & w \sim \frac{1}{\sqrt{y}} & w \sim \frac{1}{y} & \dots \end{array}$$

Mit den so transformierten Stützwerten wird das Regressionsmodell angepasst. Die approximierten Systemantwort errechnet sich durch Rücktransformation von (5.1) zu:

$$\hat{y}(\mathbf{x}) = (\lambda \hat{w}(\mathbf{x}) + 1)^{\frac{1}{\lambda}} \quad (5.2)$$

Der Parameter λ muss so bestimmt werden, dass die Summe der Fehlerquadrate bei dem jeweils neu anzupassenden Regressionsmodell minimal wird. Dies ließe sich leicht mit einem der in Abschnitt 2.3.4 Gradientenverfahren bewerkstelligen. Aufgrund der Möglichkeit, dass das Optimierungsverfahren in ein lokales Optimum laufen kann, müsste man jedoch mit verschiedenen Werten für λ eine Optimierung starten. Da die einzelne Regressionsanalyse nur eine äußerst kurze Berechnungsdauer hat, ist es auch möglich, mittels einer vollständigen Enumeration mit kleiner Schrittweite ($\Delta\lambda \leq 0.1$) das Optimum zu suchen.

Im folgenden Abschnitt wird das Verbesserungspotential der „Power Transformation“ anhand der in den Abschnitten 4.1.1 und 4.1.2 vorgestellten Optimierungsbeispiele untersucht.

5.1.1 Anwendung bei Optimierungsaufgaben

Beispiel „Biegebalken“ (Abschnitt 4.1.1)

Wendet man die Transformationsvorschrift (5.1) auf die Restriktionsfunktion des Beispiels „Biegebalken“ an, so erhält man für alle Response Surface Designs (RSDs) $\lambda_{opt} = 1$. Dies war auch zu erwarten, da bereits ein linearer Ansatz mit Interaktionsterm den Funktionswert exakt wiedergibt.

Anders sieht es bei der Approximation der Zielfunktion aus. Hier konnte die globale Tendenz mit einem vollständigen quadratischen Ansatz eher nur unzureichend abgebildet werden. Eine Transformation kann in diesem Fall eine enorme Verbesserung der Näherung für die Zielfunktion und somit auch für das ermittelbare Optimum bewirken. Die sehr gute Qualität der Approximation erkennt man auch anhand der im Folgenden dargestellten Werte der globalen Beurteilungskriterien.

RS-Design	λ	SS_E	$PRESS$	R^2	R_{adj}^2	R_{pred}^2
$3^2 \equiv CCF$	0.01	0.00198	0.0428	0.9999960	0.9999894	0.9999141
CCC	-0.45	47.5875567	338.4002449	0.9844414	0.9585104	0.8893612
CCI	-0.47	0.0645947	0.4593403	0.9995110	0.9986960	0.9965227

Mit den zurücktransformierten quadratischen Ansätzen, die mittels der transformierten Stützpunkte der verschiedenen RSDs angepasst wurden, kann nun jeweils das exakte Optimum ermittelt werden. Dabei wird der Wert der realen Zielfunktion im Optimum beim 3-Level Factorial Design um 0.02 mm, beim Central Composite Circumscribed (CCC) Design um 0.01 mm und beim Central Composite Inscribed (CCI) Design um 0.01 mm verfehlt. Dies ist auf die enorme Verbesserung der Prognosegüte der RSAs zurückzuführen.

Beispiel „Faserverbundplatte“ (Abschnitt 4.1.2)

In diesem Beispiel kann die Approximation durch Transformation des Werts der Restriktionsfunktion nicht verbessert werden. Bei allen verwendeten RSDs ergibt sich als optimaler Wert $\lambda = 1$.

Auf die Zielfunktion angewendet, sehen die Werte der globalen Beurteilungskriterien sowie die optimalen Transformationsparameter wie folgt aus:

RS-Design	λ	SS_E	$PRESS$	R^2	R^2_{adj}	R^2_{pred}
$3^2 \equiv CCF$	1.1	31.0	248.1	0.9999977	0.9999939	0.9999817
CCD	0.6	5155929.1	36664384.4	0.6051041	-0.053055	-1.808148
CCI	0.45	2369000.1	16846223.6	0.7388107	0.3034952	-0.857346

Wie auch bei der Anwendung der RSM ohne Transformation der Systemantwort zeigt das Approximationsmodell mit den Stützpunkten des 3-Level Factorial Designs eine hohe Approximationsgüte. Dies wird durch die Transformation sogar noch etwas verbessert. Bei den beiden anderen RSDs schneidet die jeweils zugehörige Approximation trotz Transformation der Werte der Zielfunktion schlecht ab. Dies ist nach wie vor ein deutliches Zeichen dafür, dass andere funktionale Zusammenhänge im Regressionsmodell beschreibbar sein müssten. Das zeigt sich auch anhand der Optima, die mit den zurücktransformierten quadratischen Ansätzen ermittelt werden.

RS-Design	λ	Optimum		
		α_1 [°]	α_2 [°]	\widehat{EI}_x [N]
$3^2 \equiv CCF$	1.10	-0.4	39.1	3653
CCC	0.60	-0.7	33.3	3744
CCI	0.45	-5.5	40.5	3063

Sowohl bei der Lage des Optimums als auch beim hierfür prognostizierten Wert der Zielfunktion ist – wenn überhaupt – keine nennenswerte Verbesserung zu erkennen. Mit den Stützpunkten des 3-Level Factorial Designs erhält man identische optimale Werte für α_1 und α_2 . Lediglich die prognostizierte bezogene Biegesteifigkeit \widehat{EI}_x wird um einen geringen Betrag besser. Beim über das CCC Design angepassten Approximationsmodell wird der optimale Wert für α_1 und der optimale Wert für \widehat{EI}_x etwas besser bestimmt. Verwendet man das CCI Design, so führt eine Optimierung mit der RSA bei allen drei Werten zu geringfügig schlechteren Werten.

5.1.2 Fazit

Mit der beschriebenen Transformation der Systemantwort lassen sich in vielen Fällen deutlich bessere Approximationen erstellen. Dies gilt jedoch nur dann, wenn sich dadurch der Verlauf der Systemantwort in eine Form bringen lässt, die mit einem quadratischen Ansatz wesentlich besser approximiert werden kann. Wird der zu approximierende Funktionsverlauf an unterschiedlichen Stellen des Definitionsbereichs ungleich stark beeinflusst, so haben Transformationen selten eine deutlich positive Wirkung. Dies zeigt sich am zweiten Beispiel, der Faserverbundplatte. Aus der klassischen Laminattheorie weiß man, dass die Schichtwinkel über Cosinus- und Sinus-Terme – teilweise sogar in der 4. Potenz – Einfluss nehmen. Ein quadratischer Ansatz auch mit Unterstützung durch eine Transformation der zu approximierenden Systemantwort ist hier eher ein unzureichendes Werkzeug zum Generieren einer guten RSA.

5.2 Transformation der Entwurfsvariablen und beliebige Ansatzterme

Aus den bisherigen Untersuchungen wird deutlich, dass für eine gute Approximation häufig erweiterte funktionale Zusammenhänge im Regressionsmodell vorhanden sein müssten. Um individuelle Einflüsse ausgleichen zu können, führen Box und Draper [1987] auch für die Entwurfsvariablen Transformationen ein. Die Entwurfsvariablen werden dabei durch logarithmische Funktionen, Wurzelfunktionen sowie durch reziproke Zusammenhänge transformiert, die – wie bereits bei der Transformation der Systemantwort – mit einer Potenzfunktion dargestellt werden können. Die transformierten Werte dienen als Stützstellen für einen vollständigen quadratischen Ansatz. Die optimalen Exponenten der Transformationsvorschriften für die einzelnen Entwurfsvariablen müssen in einem iterativen Prozess ermittelt werden.

Auch mit den von Box und Draper [1987] vorgeschlagenen Transformationen lassen sich gute Erfolge bei der Verbesserung der Approximation erzielen. Das Problem dabei ist, dass zum einen die erzeugbaren Transformationsvorschriften nur einen Teil der denkbaren funktionalen Zusammenhänge abdecken. Zum anderen gibt es auch Fälle, bei denen es von großem Vorteil ist, bestimmte Eingangsgrößen mit unterschiedlichen Transformationen zu verändern oder nur in bestimmten Ansatztermen zu transformieren.¹

Aus diesem Grund wird in dieser Arbeit ein anderer Weg beschritten. Dieser ermöglicht eine wesentlich flexiblere Gestaltung der Ansatzterme.

¹Ein Beispiel hierfür ist in Abschnitt 7.2 zu finden.

5.2.1 Definition des erweiterten Approximationsmodells

Die RSM bietet in dem in Kapitel 3 beschriebenen Umfang eine Vielzahl von Verfahrensweisen für die Behandlung linearer Regressionsprobleme. Besonders hervorzuheben sind hierbei die Beurteilungskriterien. Um diese weiterhin nutzen zu können, wird das erweiterte Approximationsmodell ebenfalls über eine Linearkombination mathematischer Ansatzfunktionen $h_i(\mathbf{x})$ definiert.

$$\hat{y}(\mathbf{x}) = \sum_{i=1}^{n_K} b_i h_i(\mathbf{x}) = \mathbf{b}^T \mathbf{h}(\mathbf{x}) \quad \text{mit} \quad \mathbf{h}(\mathbf{x}) = [h_1(\mathbf{x}), h_2(\mathbf{x}), \dots, h_{n_K}(\mathbf{x})]^T \quad (5.3)$$

Für eine möglichst flexible Definition der einzelnen Ansatzfunktionen $h_i(\mathbf{x})$ und besonders auch im Hinblick auf die erforderliche Optimierung der Ansatzterme, werden diese in einer parametrisierten Form definiert.

$$h_i(\mathbf{x}) = h(\mathbf{x}, \mathbf{a}_{(i)}, t_i) = \begin{cases} x_1^{a_{1(i)}} x_2^{a_{2(i)}} \dots x_{n_V}^{a_{n_V(i)}} = \prod_{j=1}^{n_V} x_j^{a_{j(i)}} & \text{für } t = 1 \\ e^{\sum_{j=1}^{n_V} a_{j(i)} x_j} = e^{\mathbf{a}_{(i)}^T \mathbf{x}} & \text{für } t = 2 \\ \ln \left(\sum_{j=1}^{n_V} a_{j(i)} x_j \right) = \ln(\mathbf{a}_{(i)}^T \mathbf{x}) & \text{für } t = 3 \\ \sin \left(\sum_{j=1}^{n_V} a_{j(i)} x_j \right) = \sin(\mathbf{a}_{(i)}^T \mathbf{x}) & \text{für } t = 4 \\ \cos \left(\sum_{j=1}^{n_V} a_{j(i)} x_j \right) = \cos(\mathbf{a}_{(i)}^T \mathbf{x}) & \text{für } t = 5 \\ \vdots & \end{cases} \quad (5.4)$$

mit $\mathbf{a}_{(i)} \in \mathbb{R}^{n_V}$ und $t_i \in \mathbb{N}$

Mit der Wahl des Parameters t_i kann der Typ der Ansatzfunktion für jeden Ansatzterm gewählt werden. Um die Variationsbreite der möglichen Ansatzterme zu erhöhen, kann mit den Parametern $a_{j(i)}$ auf die Gestalt der einzelnen Ansatzfunktionstypen Einfluss genommen werden. Mit dem Typ $t = 1$ ist somit eine ähnlich große Vielfalt von funktionalen Zusammenhängen beschreibbar wie mit der von Box und Draper [1987] vorgeschlagenen Verfahrensweise. Der Unterschied liegt jedoch darin, dass der Parametervektor $\mathbf{a}_{(i)}$ wie auch der Typ t_i für jeden Ansatzterm $h_i(\mathbf{x})$ verschieden gewählt werden kann. Anders als bei der „Power Transformation“ wird ein Parameter $a_{j(i)}$ in (5.4) zu null gesetzt, wenn die j -te Entwurfsvariable im i -ten Ansatzterm nicht vorkommen soll.

Die in (5.4) angegebenen parametrisierten Ansatzterme stellen sicherlich nur eine Auswahl möglicher funktionaler Zusammenhänge dar. Sie sind in der softwaretechnischen

Realisierung der hier beschriebenen Verfahrensweise implementiert und können leicht erweitert oder verändert werden. Betrachtet man die Sinus- und Cosinus-Terme ($t = 4$ und $t = 5$), so ist hier durch gezielte Anwendung der Beziehungen zwischen den trigonometrischen Funktionen (siehe z. B. [Bronstein 1991]) ebenfalls eine große Variationsbreite gegeben. Um dennoch auch andere – in der parametrisierten Form nicht beschreibbare – Ansatzterme zu verwenden, können diese explizit als String vorgegeben werden. Ein solcher Anwendungsfall ist zum Beispiel gegeben, wenn man die funktionalen Zusammenhänge einer oder mehrerer Entwurfsvariablen bereits mit physikalischer Einsicht ableiten kann.

Durch die Vorgabe einer Ansatzmatrix \mathbf{A} ($\dim(\mathbf{A}) = n_K \times (n_V + 1)$) können nahezu beliebige Approximationsmodelle definiert werden.

$$\mathbf{A} = \begin{bmatrix} t_1 & \mathbf{a}_{(1)}^T \\ t_2 & \mathbf{a}_{(2)}^T \\ \vdots & \vdots \\ t_{n_K} & \mathbf{a}_{(n_K)}^T \end{bmatrix} \quad (5.5)$$

Diese sieht für einen vollständigen quadratischen Ansatz mit zwei Entwurfsvariablen wie folgt aus:

$$\mathbf{A} = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & 1 & 1 \\ 1 & 2 & 0 \\ 1 & 0 & 2 \end{bmatrix}$$

Mit (5.4) und (5.3) folgt daraus:

$$\hat{y} = b_1 + b_2 x_1 + b_3 x_2 + b_4 x_1 x_2 + b_5 x_1^2 + b_6 x_2^2$$

5.2.2 Lineare Regression für das erweiterte Approximationsmodell

Die Regressionsanalyse zur Bestimmung der Linearkoeffizienten wurde bereits in Abschnitt 3.1.2 ausführlich beschrieben. Die Regressionsmatrix für das erweiterte Approximationsmodell lautet:

$$\mathbf{Z} = \begin{bmatrix} h_1(\mathbf{x}_1) & h_2(\mathbf{x}_1) & \dots & h_{n_K}(\mathbf{x}_1) \\ h_1(\mathbf{x}_2) & h_2(\mathbf{x}_2) & & \vdots \\ \vdots & & \ddots & \vdots \\ h_1(\mathbf{x}_{n_P}) & \dots & \dots & h_{n_K}(\mathbf{x}_{n_P}) \end{bmatrix} \quad (5.6)$$

Somit kann der Vektor der Linearkoeffizienten mit 3.14 bestimmt werden.

$$\mathbf{b} = (\mathbf{Z}^T \mathbf{Z})^{-1} \mathbf{Z}^T \mathbf{y} \quad (5.7)$$

Normierung der Regressionsmatrix

Bei der Standard RSM ist es üblich, die Entwurfsvariablen jeweils auf das Intervall $[-1, 1]$ zu normieren (siehe Gleichung (3.3)). Bei Verwendung von Polynomansätzen führt dies zu einer normierten und somit numerisch günstigeren Regressionsmatrix. Die Koeffizienten des eigentlichen Approximationsmodells müssen durch „Rücknormierung“ ermittelt werden. Dies ist bei den erweiterten Ansätzen aus zweierlei Gründen nicht praktikabel bzw. sinnvoll:

- Die Normierung der Entwurfsvariablen beeinflusst den „natürlichen“ Verlauf der Funktionen der erweiterten Ansatzterme und müsste deshalb teilweise wieder rückgängig gemacht werden. Dies gelingt jedoch nicht mit den Parametern $a_{j(i)}$ und führt bei den Ansatztermen vom Typ 2 bis 5 zu unerwünschten Effekten.
- Wie bereits erwähnt müssen die Koeffizienten des eigentlichen Approximationsmodells durch Rückgängigmachung der Normierung ermittelt werden. Würde man ungeachtet des bereits genannten Arguments dennoch eine Normierung nach (3.3) durchführen, ließe sich auch dies bei den Ansatztermen vom Typ 2 bis 5 nicht sauber bewerkstelligen.

Um dennoch eine numerisch günstigere Regressionsmatrix zu erhalten, werden zwei Normierungsvorschriften vorgeschlagen und angewendet:

- **Normierung durch einen konstanten Faktor**

Diese Form der Normierung kann ohne Einschränkung bezüglich der Ansatzterme auf die Regressionsmatrix angewendet werden. Die Regressionsmatrix wird dabei spaltenweise – also bezüglich der Funktionswerte der Ansatzterme – normiert. Für die normierte Regressionsmatrix gilt dann:

$$\bar{\mathbf{Z}} = \mathbf{Z} \operatorname{diag}(\mathbf{c})^{-1} \quad \text{mit} \quad c_j = \max_{i=1..n_K} (|Z_{ij}|) \quad (j = 1 \dots n_p) \quad (5.8)$$

Die Koeffizienten der normierten Regressionsmatrix erfüllen somit die Bedingung

$$-1 \leq \bar{Z}_{ij} \leq 1 \quad \text{mit} \quad i = 1 \dots n_P \quad \text{und} \quad j = 1 \dots n_K$$

wobei das Intervall meist nicht vollständig ausgefüllt ist. Wird mit der Regressionsanalyse mit $\bar{\mathbf{Z}}$ erhält man den „normierten“ Koeffizientenvektor $\bar{\mathbf{b}}$. Der reale Koeffizientenvektor ergibt sich aus dem Koeffizientenvergleich von

$$\hat{\mathbf{y}} = \bar{\mathbf{Z}} \bar{\mathbf{b}} = \mathbf{Z} \mathbf{b} = \bar{\mathbf{Z}} \operatorname{diag}(\mathbf{c}) \mathbf{b}$$

zu

$$\mathbf{b} = \operatorname{diag}(\mathbf{c})^{-1} \bar{\mathbf{b}} \quad (5.9)$$

• **Lineare Normierung**

Bei einer linearen Normierung ist Voraussetzung, dass der Ansatz einen konstanten Term $h(\mathbf{x}) = 1$ besitzt, das heißt

$$\mathbf{Z} = [\mathbf{1}, \mathbf{Z}^*] \quad \dim(\mathbf{Z}^*) = n_P \times (n_K - 1)$$

$$\text{mit } \mathbf{1} = \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix} \quad \dim(\mathbf{1}) = n_P \times 1$$

\mathbf{Z}^* stellt dabei die Regressionsmatrix der frei wählbaren Ansätze dar.

Bei spaltenweiser linearer Normierung der Regressionsmatrix \mathbf{Z} auf das Intervall $[-1; 1]$ ergibt sich folgender Zusammenhang:

$$\mathbf{Z}^* = \mathbf{1}\mathbf{c}^T + \bar{\mathbf{Z}}^* \text{diag}(\mathbf{d}) \quad (5.10)$$

$$\left. \begin{array}{l} \text{mit } d_j = \frac{\max_i(Z_{ij}) - \min_i(Z_{ij})}{2} \\ \text{und } c_j = \min_i(Z_{ij}) + d_j \end{array} \right\} \begin{array}{l} i = 1 \dots n_K \\ j = 1 \dots n_P \end{array}$$

$$\mathbf{Z} = [\mathbf{1}, \mathbf{1}\mathbf{c}^T + \bar{\mathbf{Z}}^* \text{diag}(\mathbf{d})] \quad (5.11)$$

Hierbei stellt $\bar{\mathbf{Z}}^*$ die normierte Regressionsmatrix ohne den konstanten Term dar. Aus (5.10) bzw. (5.11) ergibt sich die normierte Regressionsmatrix $\bar{\mathbf{Z}}^*$ bzw. $\bar{\mathbf{Z}}$ zu

$$\bar{\mathbf{Z}}^* = (\mathbf{Z}^* - \mathbf{1}\mathbf{c}^T) \text{diag}(\mathbf{d})^{-1} \quad (5.12)$$

$$\bar{\mathbf{Z}} = [\mathbf{1}, \bar{\mathbf{Z}}^*] \quad (5.13)$$

Setzt man nun die approximative Antwort $\hat{\mathbf{y}}$ in normierter Form

$$\hat{\mathbf{y}} = \bar{\mathbf{Z}}\bar{\mathbf{b}} \quad (5.14)$$

an, so erhält man den normierten Koeffizientenvektor $\bar{\mathbf{b}}$. Der Zusammenhang zwischen diesem und dem realen Koeffizientenvektor ergibt sich folgendermaßen:

Spaltet man $\bar{\mathbf{b}}$ in die Anteile

$$\bar{\mathbf{b}} = \begin{bmatrix} \bar{b}_0 \\ \bar{\mathbf{b}}^* \end{bmatrix} \quad (5.15)$$

auf, und setzt diesen und (5.13) in (5.14) ein, so ergibt sich

$$\hat{\mathbf{y}} = \bar{\mathbf{Z}}\bar{\mathbf{b}} = [\mathbf{1}, \bar{\mathbf{Z}}^*] \begin{bmatrix} \bar{b}_0 \\ \bar{\mathbf{b}}^* \end{bmatrix} = \mathbf{1}\bar{b}_0 + \bar{\mathbf{Z}}^*\bar{\mathbf{b}}^* \quad (5.16)$$

Des Weiteren lässt sich der Antwortvektor auch in nicht normierter Form darstellen

$$\hat{\mathbf{y}} = \mathbf{Z}\mathbf{b} \quad (5.17)$$

spaltet man hier wiederum \mathbf{b} in

$$\mathbf{b} = \begin{bmatrix} b_0 \\ \mathbf{b}^* \end{bmatrix} \quad (5.18)$$

auf und setzt diesen und (5.11) in (5.17) ein, ergibt sich

$$\begin{aligned} \hat{\mathbf{y}} = \mathbf{Z}\mathbf{b} &= [\mathbf{1}, \mathbf{1}\mathbf{c}^T + \bar{\mathbf{Z}}^* \text{diag}(\mathbf{d})] \begin{bmatrix} b_0 \\ \mathbf{b}^* \end{bmatrix} \\ \Leftrightarrow \hat{\mathbf{y}} &= \mathbf{1}b_0 + \mathbf{1}\mathbf{c}^T\mathbf{b}^* + \bar{\mathbf{Z}}^* \text{diag}(\mathbf{d})\mathbf{b}^* \end{aligned} \quad (5.19)$$

Aus dem Koeffizientenvergleich von (5.19) und (5.16)

$$\begin{aligned} \mathbf{1}(b_0 + \mathbf{c}^T\mathbf{b}^*) &= \bar{\mathbf{1}}\bar{b}_0 \\ \Leftrightarrow b_0 + \mathbf{c}^T\mathbf{b}^* &= \bar{b}_0 \end{aligned} \quad (5.20)$$

$$\text{und } \text{diag}(\mathbf{d})\mathbf{b}^* = \bar{\mathbf{b}}^* \quad (5.21)$$

folgt die Lösung des Koeffizientenvektors \mathbf{b} zu

$$\mathbf{b}^* = \text{diag}(\mathbf{d})^{-1}\bar{\mathbf{b}}^* \quad (5.22)$$

$$b_0 = \bar{b}_0 - \mathbf{c}^T\mathbf{b}^* \quad (5.23)$$

$$\mathbf{b} = \begin{bmatrix} b_0 \\ \mathbf{b}^* \end{bmatrix} \quad (5.24)$$

Die beschriebenen Normierung haben keinerlei Einfluss auf die Werte der Bewertungskriterien nach Abschnitt 3.2. Lediglich die nicht auf der Hauptdiagonalen liegenden Elemente der Hat-Matrix \mathbf{H} (3.32) nehmen andere Werte an. Da diese jedoch bei keinem in dieser Arbeit verwendeten Bewertungskriterium benötigt werden, spricht nichts gegen die Anwendung der Normierungsvorschriften.

5.3 Bestimmung erweiterter Ansatzfunktionen durch ein diskret-nichtlineares Ausgleichsverfahren

Bedingt durch die Vielfalt der möglichen Ansatzterme nach (5.4) ist es sehr schwierig, ein gutes Approximationsmodell „von Hand“ aufzustellen. Dabei ist es sicherlich möglich und außerordentlich wichtig, für einige der Entwurfsvariablen mit Kenntnis physikalischer Zusammenhänge den Typ t_i und möglichst auch die zugehörigen Parameter $a_{i(j)}$ festzulegen. Für die Fälle, in denen dies nicht gelingt, wird im Folgenden ein Verfahren vorgestellt, mit dem die Ansatzterme automatisiert und optimal in Bezug auf die Qualität der Approximation bestimmt werden. Hierzu werden verschiedene der in Abschnitt 2.3 beschriebenen Optimierungsverfahren eingesetzt.

5.3.1 Formulierung des Ausgleichsproblems

Mit der in Abschnitt 5.2.1 definierten parametrisierten Form der Ansatzterme (5.4) stehen die zu bestimmenden Parameter fest. Jeder Ansatzterm ist dabei durch den Typ t_i sowie den kontinuierlichen Parametervektor $\mathbf{a}_{(i)}$ bestimmt. Während der Typ t_i nur die diskreten Werte 1 bis 5 annehmen kann, müssen für den Parametervektor eine untere Grenze $\mathbf{a}_{\text{lb}(i)}$ sowie eine obere Grenze $\mathbf{a}_{\text{ub}(i)}$ vorgegeben werden.

Da der Wertebereich des Parametervektors für jeden Ansatzfunktionstyp anders sein kann bzw. muss, können t_i und $\mathbf{a}_{(i)}$ nicht willkürlich variiert werden. Deshalb wird ein zur Auswahl stehender Ansatzterm i durch Vorgabe des Typs t_i und durch Angabe der dazu passenden Grenzen $\mathbf{a}_{\text{lb}(i)}$ und $\mathbf{a}_{\text{ub}(i)}$ definiert.

$$\mathbf{a}_{(i)}^* = [t_i, \mathbf{a}_{\text{lb}(i)}^{\text{T}}, \mathbf{a}_{\text{ub}(i)}^{\text{T}}]^{\text{T}} \quad (5.25)$$

Der Vektor $\mathbf{a}_{(i)}^*$ enthält alle für die optimale Bestimmung des Ansatzterms benötigten Informationen. Auf diese Weise werden alle gewünschten Ansatzterme definiert und in der Matrix der zu bestimmenden Ansatzterme zusammengefasst.

$$\mathbf{A}^* = \begin{bmatrix} \mathbf{a}_{(1)}^{*\text{T}} \\ \mathbf{a}_{(2)}^{*\text{T}} \\ \vdots \\ \mathbf{a}_{(n_S)}^{*\text{T}} \end{bmatrix} = \begin{bmatrix} t_1 & \mathbf{a}_{\text{lb}(1)}^{\text{T}} & \mathbf{a}_{\text{ub}(1)}^{\text{T}} \\ t_2 & \mathbf{a}_{\text{lb}(2)}^{\text{T}} & \mathbf{a}_{\text{ub}(2)}^{\text{T}} \\ \vdots & \vdots & \vdots \\ t_{n_S} & \mathbf{a}_{\text{lb}(n_S)}^{\text{T}} & \mathbf{a}_{\text{ub}(n_S)}^{\text{T}} \end{bmatrix} \quad (5.26)$$

Sollen nun n_T Terme ausgewählt und deren Parameter $a_{(i)j}$ optimiert werden, so ergibt sich für die insgesamt zu bestimmenden Parameter der Vektor

$$\mathbf{x} = [p_1, \dots, p_{n_T}, \mathbf{a}_{(1)}^{\text{T}}, \dots, \mathbf{a}_{(n_T)}^{\text{T}}]^{\text{T}} \quad (5.27)$$

Die Parameter p_i stehen für die Zeilennummer der Matrix \mathbf{A}^* .

Somit kann ein diskret-nichtlineares Ausgleichsproblem formuliert werden:

$$\text{Minimiere} \quad SS_E(\mathbf{x}) = SS_E(p_1, \dots, p_{n_T}, \mathbf{a}_{(1)}^{\text{T}}, \dots, \mathbf{a}_{(n_T)}^{\text{T}}) \quad (5.28a)$$

$$\text{so, dass} \quad 1 \leq p_1 < p_2 < \dots < p_{n_T} \leq n_S \quad (5.28b)$$

$$\text{und} \quad \mathbf{a}_{\text{lb}(p_i)} \leq \mathbf{a}_{(p_i)} \leq \mathbf{a}_{\text{ub}(p_i)} \quad (i = 1 \dots n_T) \quad (5.28c)$$

In Worten bedeutet dies:

(5.28a) Minimiere die Summe der Fehlerquadrate durch die optimale Auswahl von n_T Ansatztermen aus einem Pool von n_S möglichen Ansatztermen sowie durch die optimale Bestimmung der Parameter jedes Ansatzterms.

(5.28b) Dabei darf kein Ansatzterm mehrfach gewählt werden.

(5.28c) Die Parameter \mathbf{a} dürfen bei keinem Ansatzterm die zulässigen Grenzen überschreiten.

Das Ausgleichsproblem (5.28) ist mit kaum einem der in Abschnitt 2.3 beschriebenen Optimierungsverfahren zu geschlossen lösen. Zum einen kann ein Teil des Parametervektors \mathbf{x} nur diskrete Werte annehmen. Darüber hinaus sind die Grenzen der Parameter $a_{(i)j}$ an die Wahl des Ansatzterms gekoppelt. Zusätzlich kann sich auch noch die Zahl der zu bestimmenden Parameter ändern. Dies ist der Fall, wenn statt eines Ansatzterms mit einem freien Parameter ein Ansatzterm gewählt wird, bei dem die Parameter für zwei oder mehr Entwurfsvariablen bestimmt werden sollen.

Aus diesem Grund ist es für die Lösung des Ausgleichsproblems nicht nur sinnvoll, sondern auch notwendig, verschiedene Lösungsverfahren zu kombinieren/verschachteln. Dabei wählt ein äußeres Lösungsverfahren die diskreten Parameter p_i . Für das innere Lösungsverfahren sind somit die Zahl der kontinuierlichen Parameter und deren Grenzen bekannt.

Eine solche Aufteilung ist außerdem auch deshalb sinnvoll, da man zunächst mit dem äußeren Lösungsverfahren eine geeignete Auswahl von Ansatzfunktionstypen gewinnen kann. Hierzu werden die unteren und oberen Grenzen der Parameter $\mathbf{a}_{(i)}$ mit identischen Werten belegt, so dass das innere Lösungsverfahren hinfällig wird. Für diese Vorauswahl können dann mit dem inneren Lösungsverfahren die optimalen Werte für die Parameter der Ansatzterme bestimmt werden.

5.3.2 Lösungsverfahren für das diskrete Ausgleichsproblem

Wie bereits erwähnt ist die Aufgabe des äußeren Lösungsverfahrens das Zusammenstellen der n_T besten Ansatzterme.

$$\begin{aligned} \text{Minimiere } SS_E(\mathbf{x}) &= SS_E(p_1, \dots, p_{n_T}) \\ \text{so, dass } 1 &\leq p_1 < p_2 < \dots < p_{n_T} \leq n_S \end{aligned} \quad (5.29)$$

Als ein mögliches Verfahren kommt die vollständige Enumeration aller Kombinationen der n_S zur Auswahl stehenden Ansatzterme in Frage. Das sind $n_S! / ((n_S - n_T)! n_T!)$ Kombinationen. Auch wenn je nach Anzahl der Terme im Approximationsmodell und Anzahl der Stützpunkte etwa 500 bis 1000 Kombinationen pro Sekunde ausgewertet werden können, ist hier jedoch schnell eine Grenze erreicht. Sollen zum Beispiel aus 25 möglichen Ansatztermen die zehn besten ausgewählt werden, so ist die Aufgabe mit Hilfe der vollständigen Enumeration noch innerhalb etwa eines halben Tages lösbar. Die besten 25 aus 50 Ansatztermen auszuwählen, würde etwas mehr als 4000 Jahre in Anspruch nehmen. Somit ist die vollständige Enumeration – wenn überhaupt – nur für eine eher kleine Zahl von möglichen Ansatztermen geeignet.

Auch andere, eher suchrichtungsorientierte Optimierungsverfahren, die mit diskreten Variablen umgehen können, sind für die Aufgabe des äußeren Lösungsverfahrens nicht

geeignet. Der Grund hierfür ist, dass es keinen Zusammenhang zwischen dem Approximationsfehler zweier benachbarter Kombinationen gibt. Das heißt, wenn ein Ansatzterm p_i durch den Term $p_{\pm 1}$ ersetzt wird, ändert sich der Approximationsfehler meist sprunghaft. Somit können keinerlei Suchrichtungsinformationen ermittelt werden.

Das für diese Aufgabe am besten geeignete Optimierungsverfahren ist ein einfacher Genetischer Algorithmus, dessen Arbeitsweise bereits in Abschnitt 2.3.3 erläutert wurde. Eine weitere Beschreibung auch der softwaretechnischen Umsetzung ist zum Beispiel in [Baier 2002b] zu finden.

Die Variation der Approximationsmodelle (Kombination verschiedener Ansatzterme) erfolgt beim Genetischen Algorithmus durch Mutation und Rekombination. Durch Mutation kommen neue Ansatzterme in die untersuchten Approximationsmodelle. Die Rekombination sorgt dafür, dass Approximationsmodelle mit guten Ansatztermen diese weiter vererben. Deshalb sollte die Mutationsrate zu Beginn der Ansatzoptimierung höher sein und zum Ende hin stark reduziert werden. Dann sind die wichtigen Ansatzterme in der Regel in den Approximationsmodellen der letzten Generationen vertreten und können mit Hilfe der Rekombination genauer identifiziert werden.

Sind die oberen und unteren Grenzen der zu den Ansatzfunktionen gehörenden Parametervektoren gleichgesetzt, so wird das innere Lösungsverfahren nicht benötigt. Die Zielfunktion ist allein die Summe der Fehlerquadrate des Approximationsmodells. Können einige der kontinuierlichen Parameter innerhalb vorgegebener Grenzen variiert werden, so kommt das innere Lösungsverfahren zur Anwendung. Dieses bestimmt die kontinuierlichen Parameter so, dass sich ein kleinstmöglicher Wert für die Summe der Fehlerquadrate ergibt. Der Wert dient dann dem Genetischen Algorithmus zur Bewertung der gewählten Kombination von Ansatztermen.

5.3.3 Lösungsverfahren für das nichtlineare Ausgleichsproblem

Das innere Lösungsverfahren soll die optimalen Parameter $\mathbf{a}_{(i)}$ der vom Genetischen Algorithmus zusammengestellten Ansatzterme bestimmen. Dabei ist das Ziel die Minimierung der Fehlerquadrate.

$$\begin{aligned} \text{Minimiere } SS_E(\mathbf{x}) &= SS_E(\mathbf{a}_{(1)}^T, \dots, \mathbf{a}_{(n_T)}^T) \\ \text{so, dass } \mathbf{a}_{\text{lb}(p_i)} &\leq \mathbf{a}_{(p_i)} \leq \mathbf{a}_{\text{ub}(p_i)} \quad (i = 1 \dots n_T) \end{aligned} \tag{5.30}$$

Da es sich bei den freien Parametern um kontinuierliche Variablen handelt, bieten sich zur Lösung sowohl Gradientenverfahren als auch Evolutionäre Algorithmen an. Bevor auf die Lösungsverfahren für das nichtlineare Ausgleichsproblem eingegangen wird, soll zunächst an einem Beispiel der Verlauf einer hierbei möglichen Zielfunktion dargestellt werden.

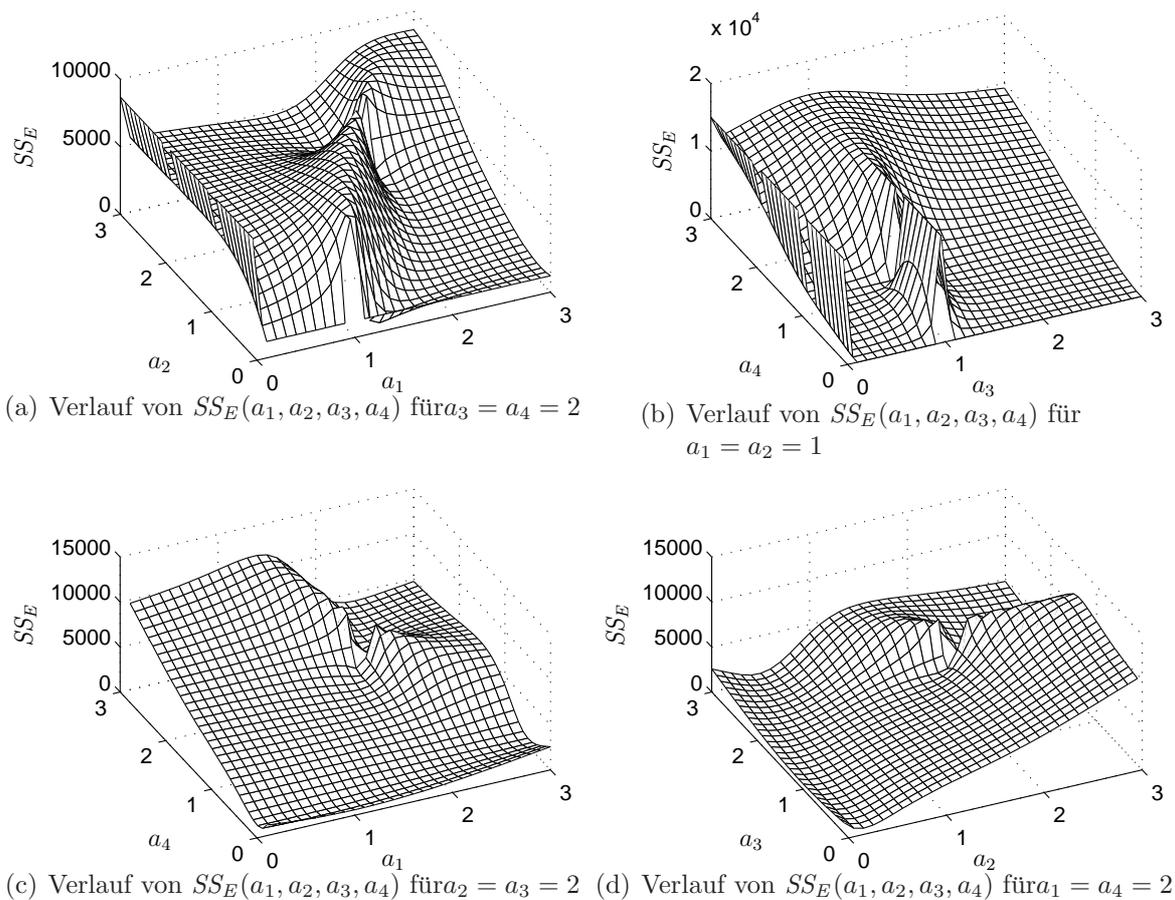


Abbildung 5.1: Verlauf der Zielfunktion des inneren Lösungsverfahrens in verschiedenen Punkten des Entwurfsraums

Beispiel für eine innere Zielfunktion

Als zu approximierende Funktion sei gegeben:

$$y(x_1, x_2, x_3) = 5 + 3x_1 - 2x_2 + 3x_3 + 2x_1^2 + 3x_2^2 + 2x_1x_2 \quad (5.31)$$

Das dabei zu verwendende Approximationsmodell mit wählbaren Ansatztermen sieht wie folgt aus:

$$\mathbf{A} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 2 & 0 \end{bmatrix} \quad \mathbf{A}^* = \begin{bmatrix} 1 & a_{1lb} & a_{2lb} & 0 & a_{3ub} & a_{4ub} & 0 \\ 1 & a_{1lb} & a_{2lb} & 0 & a_{3ub} & a_{4ub} & 0 \\ 1 & 0 & 0 & 0 & 3 & 3 & 0 \\ 1 & 0 & 0 & 0 & 3 & 3 & 0 \end{bmatrix}$$

$$\hat{y}(x_1, x_2, x_3) = b_1 + b_3x_2 + b_2x_3 + b_4x_2^2 + b_5x_1 + b_6x_1^{a_1}x_2^{a_2} + b_7x_1^{a_3}x_2^{a_4} \quad (5.32)$$

Mit dem inneren Lösungsverfahren müssen nun die Parameter a_1 , a_2 , a_3 und a_4 bestimmt werden. Diese können aus dem Intervall $[0, 3]$ gewählt werden. Aus dem Vergleich von (5.31) und (5.32) folgt, dass (5.31) durch die Approximation für den Fall $a_1 = 1$, $a_2 = 1$, $a_3 = 2$, $a_4 = 0$ bzw. $a_1 = 2$, $a_2 = 0$, $a_3 = 1$, $a_4 = 1$ exakt wiedergegeben werden kann. Das Residuum ist in beiden Fällen gleich 0.

Das hier dargestellte nichtlineare Ausgleichsproblem stellt einen eher schwierigen Fall dar, da die beschreibbaren Funktionen identisch sind und somit zwei gleichrangige Lösungen existieren. Ein weiteres Problem ist, dass die Zielfunktion des inneren Lösungsverfahrens viele lokale Optima hat, was anhand der in Abbildung 5.1 dargestellten Funktionsverläufe zu erkennen ist. In allen dargestellten Flächen findet man jeweils Lücken an den Punkten, bei denen beide Ansatzterme identisch sind und somit die Regressionsmatrix singular wird. Für diese Punkte existiert also keine Lösung für die Koeffizienten des Approximationsmodells. Beide Eigenschaften erschweren eine Lösung.

Im Folgenden werden zwei Lösungsverfahren beschrieben, mit denen eine Lösung des nichtlinearen Ausgleichsproblems möglich ist.

Einsatz eines Evolutionären Algorithmus

Als Evolutionärer Algorithmus dient ein einfacher Standardalgorithmus, wie er in [Bai-er 2002b] beschrieben ist. Der Wertebereich für die Parameter $\mathbf{a}_{(i)}$ ist mit der vom äußeren Lösungsverfahren festgelegten Auswahl an Ansatztermen gegeben. Da sich die Grenzen jedoch dynamisch ändern, ist es sinnvoll, die Standardabweichung der Mutation in Abhängigkeit vom Wertebereich der Parameter festzulegen.

Der Ablauf ist dann wie folgt:

1. Zunächst wird eine Ausgangspopulation erzeugt. Dies erfolgt durch eine zufällig gewählte Verteilung der Parametervektoren innerhalb ihrer Grenzen. Die Anzahl der erzeugten Individuen soll dabei größer sein als die Anzahl der Individuen, die in den nachfolgenden Generationen vertreten sind.
2. Als nächstes wird die Fitness jedes der Individuen bewertet. Hierzu muss das lineare Regressionsmodell jedes Individuums angepasst und die Summe der Fehlerquadrate oder der Wert eines anderen globalen Beurteilungskriteriums (R^2 , R^2_{pred} , $PRESS$) bestimmt werden.

3. Danach startet die Iterationsschleife für die vorgegebene Zahl von Generationen mit folgenden Ablauf:
 - a) Beginnend mit dem am besten bewerteten Individuum wird eine bestimmte Anzahl von Individuen als Eltern selektiert.
 - b) Die Parameter der selektierten Individuen werden anschließend mutiert. Dies erfolgt durch zufällige Streuung um den aktuellen Wert mit einer wählbaren Standardabweichung.
 - c) Die so erhaltenen Individuen werden wieder in ihrer Fitness bewertet und die n schlechtesten Individuen der Elterngeneration durch die n besten Kinder ersetzt.
4. Abbruchkriterium ist das Erreichen der vorgegebenen Anzahl von Generationen. Das insgesamt am besten bewertete Individuum liefert letztendlich den optimalen Parametervektor sowie den Fitness-Wert an den äußeren Teil der Ansatzoptimierung zurück.

Wird die Populationsgröße sehr groß gewählt, kann mit hoher Wahrscheinlichkeit das globale Optimum bestimmt werden. Allerdings ist damit der Berechnungsaufwand für das innere Lösungsverfahren ebenfalls sehr groß. Der Vorteil eines Evolutionären Algorithmus ist, dass es keine Rolle spielt, wenn die Fitness-Bewertung für einige Individuen fehlschlägt.

Einsatz eines Gradientenverfahrens

Ein Gradientenverfahren kann die optimalen Parameter im Gegensatz zu einem Evolutionären Algorithmus wesentlich zielgerichteter bestimmen. Somit werden deutlich weniger Berechnungen benötigt. Wie jedoch anhand des Beispiels für eine Zielfunktion gezeigt wurde, besteht eine nicht unerhebliche Gefahr, dass lediglich ein lokales Optimum gefunden wird.

Um dies zu verdeutlichen, wird das gleiche Beispiel noch einmal herangezogen. Um das Verhalten grafisch darstellen zu können, wird die Anzahl der zu bestimmenden Parameter a_i im Approximationsmodell (5.32) auf 3 reduziert, indem ein Parameter auf seinen optimalen Wert gesetzt wird ($a_4 = 1$). Startet man nun an verschiedenen Stellen des Definitionsbereichs ein Gradientenverfahren, so erhält man eine Reihe unterschiedlicher Optima. Dieses Verhalten ist in Abbildung 5.2 dargestellt. Die Optimierung wurde an insgesamt 64 Stellen gestartet, wobei 14 verschiedene Optima (gekennzeichnet durch Kreuze) ermittelt wurden. Die Linien verbinden jeden Startpunkt mit dem zugehörigen Optimum. Eine große Zahl der Optimierungen endet dabei im globalen Optimum ($a_1 = 2$, $a_2 = 0$ und $a_3 = 1$). Bei einer nicht unerheblichen Menge von Startvektoren läuft das Gradientenverfahren jedoch in eines der lokalen Optima.

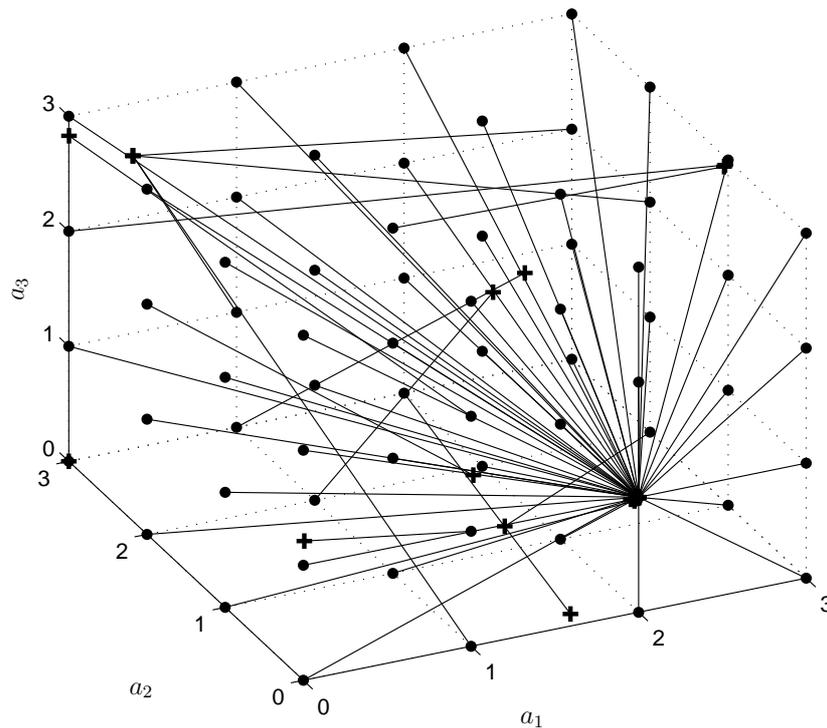


Abbildung 5.2: Mit einem Gradientenverfahren ermittelte Optima der Zielfunktion des Beispiels für eine innere Zielfunktion

Aus diesem Grund kann ein Gradientenverfahren nicht ohne weiteres eingesetzt werden. Nur durch Vorschalten bestimmter Verfahren, welche die Aufgabe haben, einen guten Startvektor zu ermitteln, wird die Effizienz eines gradientenbasierten Optimierungsverfahrens nutzbar. Somit kann eine schnelle und stabile Suche nach dem globalen Optimum gewährleistet werden.

- **Vorschalten eines Scans**

Ein anwendbares Verfahren ist, den Definitionsbereich der Parameter mit Hilfe eines Scans zu untersuchen. Darunter ist die vollständige Enumeration einer gut verteilten Menge von Punkten zu verstehen. Diese Menge kann zum Beispiel mit einem l -Level Factorial Design erzeugt werden. Für jeden der Punkte werden die Linearkoeffizienten des Approximationsmodells angepasst und der (oder die) Wert(e) zur Beurteilung der Approximation berechnet.

Startet man nun das Gradientenverfahren mit dem am besten beurteilten Parametervektor, so ist die Wahrscheinlichkeit relativ hoch, dass bei der Optimierung das globale Optimum ermittelt wird. Da jedoch nicht sichergestellt ist, dass der

Punkt mit der besten Beurteilung auch tatsächlich in der Nähe des globalen Optimums liegt, ist es sinnvoll und notwendig, mehrere Optimierungen an den paar besten Punkten zu starten.

Mit der Dichte der Scanpunkte erhöht sich die Wahrscheinlichkeit, das globale Optimum zu finden. Allerdings steigt damit der Berechnungsaufwand exponentiell an.

• Vorschalten eines Evolutionären Algorithmus

Eine wesentlich effizientere Methode ist das Vorschalten eines Evolutionären Algorithmus. Dabei wird das gleiche Verfahren benutzt, wie es in Abschnitt 5.3.3 beschrieben wurde. Die Population durchläuft jedoch deutlich weniger Generationen als zur genauen Ermittlung des globalen Optimums erforderlich wäre. Bis dahin haben sich die Individuen bereits in ausreichender Form im Bereich möglicher Optima gruppiert.

Verwendet man – wie auch beim Scan – ein paar der besten Individuen als Startvektor, kann mit hoher Wahrscheinlichkeit das globale Optimum ermittelt werden. Da die Suche der Startvektoren mit einem Evolutionären Algorithmus „intelligenter“ verläuft, ist der Aufwand deutlich geringer als bei der zuvor beschriebenen Methode.

5.3.4 Zusammenspiel des diskreten und des nichtlinearen Lösungsverfahrens – Empfehlung für die Anwendung

Bei der Beschreibung des diskreten Lösungsverfahrens wurde bereits erörtert, wie schnell die Komplexität des Ausgleichsproblems mit der Zahl der zur Auswahl stehenden Ansatzterme steigt. Dies gilt bereits dann, wenn der Zielfunktionswert nur durch eine einzige Regressionsanalyse und die Bestimmung der Summe der Fehlerquadrate SS_E ermittelt wird. Beim Lösungsverfahren für das nichtlineare Ausgleichsproblem müssen – abhängig von der Zahl der freien Parameter – relativ viele Regressionsanalysen durchgeführt werden, so dass sich der Gesamtaufwand entsprechend vervielfacht. Aus diesem Grund ist bei der Verwendung des inneren Lösungsverfahrens Vorsicht geboten.

Als Empfehlung kann aus einer Vielzahl von Untersuchungen der in Kapitel 7 behandelten Optimierungsprobleme folgende Vorgehensweise abgeleitet werden:

- Wie auch bei der Standard RSM sollten beim Aufbau eines Approximationsmodells mit erweiterten Ansatztermen zu Beginn nur die Haupteffekte der Variablen untersucht werden. Dies führt zu einer drastischen Reduktion der möglichen Ansatzterme.

- Hierbei erweist es sich im Hinblick auf den Lösungsaufwand als sehr sinnvoll, nur die wirklich notwendigen Parameter $\mathbf{a}_{(i)}$ der Ansatzterme $h_i(\mathbf{x})$ für die innere Optimierung freizugeben. Dabei handelt es sich insbesondere um die Trigonometrischen Terme (Typ 4 und 5), bei denen eine Skalierung der Entwurfsvariable(n) sehr wichtig sein kann.
- Besonders bei den am häufigsten eingesetzten Potenzfunktionen als Ansatzterm erweist es sich als nützlich, für jede interessierende Variable x_j mehrere Ansatzterme zur Auswahl zu stellen. Dabei werden die Parametervektoren so gesetzt, dass zunächst nur üblicherweise verwendete Potenzfunktionen erzeugt werden. Konkret bedeutet dies: $a_{j(i)} = -2$ für $1/x_j^2$, $a_{j(i)} = -1$ für $1/x_j$, $a_{j(i)} = -1/2$ für $1/\sqrt{x_j}$ und $a_{j(i)} = 1/2$ für $\sqrt{x_j}$. Auf die Einbeziehung linearer und quadratischer Terme kann meist verzichtet werden, da diese in aller Regel bereits im Vorfeld untersucht wurden.
- Da das Approximationsmodell aus der Ansatzmatrix der fixen Terme \mathbf{A} (5.5) und der Matrix der wählbaren Ansätze \mathbf{A}^* (5.26) gebildet wird, sollte nur für möglichst wenige Variablen ein erweiterter Ansatzterm in Betracht gezogen werden. Als fest vorgegebener Teil des Approximationsmodells erweisen sich weiterhin lineare und quadratische Ansätze als geeignet. Sie können helfen, die mit den wählbaren Ansatztermen nicht erfassten Variableneffekte zu kompensieren.
- Zuletzt, wenn die am besten geeigneten Ansatzterme zuverlässig eingegrenzt sind, können mehr Parameter für das innere Lösungsverfahren freigegeben werden. Stehen alle Ansatzterme fest, kann das diskrete Lösungsverfahren umgangen werden, so dass das Lösungsverfahren für das nichtlineare Ausgleichsproblem nur einmal durchlaufen wird. Dies ist der Fall, wenn mit $n_T = n_S$ nur eine mögliche Kombination von wählbaren Ansatztermen existiert und die Zahl der Individuen wie auch die Zahl der Generationen auf 1 gesetzt wird.

5.4 Geeignete Methoden des Design of Experiments

Für den Einsatz einiger der erweiterten Ansatzterme liefern die Standard RSDs für quadratische Ansatzfunktionen (CCD, BBD) nicht genug Informationen. Für eine diesbezügliche Erweiterung eignet sich das CCD am besten:

- Zur Bestimmung der Haupteffekte kann das CCD durch sternförmig angeordnete Stützstellen erweitert werden, deren Parameter α^* die Hälfte des Parameters α der bereits vorhandenen Stützstellen beträgt: $\alpha^* = \frac{\alpha}{2}$
- Sollen nichtlineare Interaktionseffekte bestimmt werden, so können die Stützstellen, die zum Fractional Factorial Design des CCDs gehören mit dem selben Faktor skaliert werden. Hierfür soll der negative Wert verwendet werden, um somit den komplementären Teil des Fractional Factorial Designs zu erzeugen.

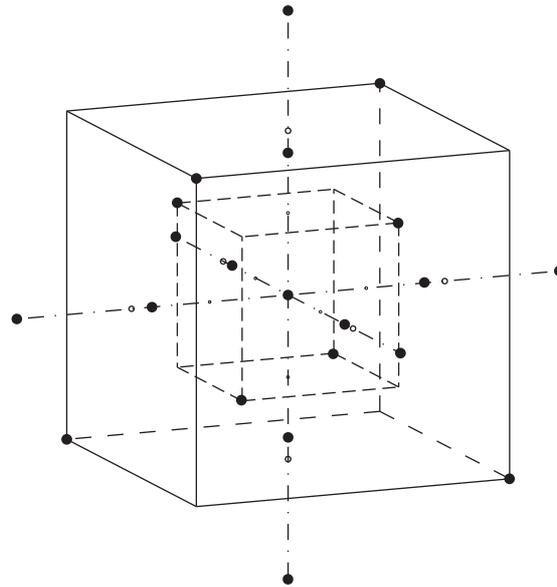


Abbildung 5.3: CCD und skaliertes komplementäres CCD zur Bestimmung nichtlinearer Haupt- und Interaktionseffekte

Damit stehen auch mehr Informationen in gegenüberliegenden Bereichen des Entwurfsraums zur Verfügung.

Das daraus entstehende RSD ist in Abbildung 5.3 dargestellt.

Für unregelmäßig geformte Entwurfsvariablenräume ist folgende Vorgehensweise geeignet:

- Zunächst werden Stützstellen mit einem Full Factorial Design hoher Ordnung (≥ 5) erstellt.
- Aus diesen sind dann die unzulässigen Stützstellen zu streichen.
- Anschließend können mit dem D -Optimalitätskriterium die für einen gewählten Ansatz geeignetsten ausgewählt werden.

6 Grobparallelisierung auf einem PC-Cluster

Anhand des Entwurfsoptimierungssystems MOSES wurden in Abschnitt 2.2 verschiedene Ansätze für die Reduktion des zeitlichen Aufwands bei Optimierungsprozessen vorgestellt. Hierzu erfolgte bereits eine Diskussion bezüglich der Effizienz verschiedener Optimierungsverfahren. Als zentrales Thema dieser Arbeit wurden Verfahrensweisen zur Erstellung von Response Surface Approximationen (RSAs) zur Anwendung bei Optimierungsaufgaben behandelt.

Für diese und andere in MOSES implementierten Funktionalitäten gilt, dass deren Anwendung die Durchführung einer mehr oder weniger großen Zahl von Systemanalysen erfordert. Da diese unabhängig voneinander durchgeführt werden können, bietet sich eine Grobparallelisierung der Prozesse an, was eine enorme Leistungssteigerung erwarten lässt. Im Rahmen der vorliegenden Arbeit wurde ein für diese Form der Parallelisierung geeignetes Rechnersystem zusammengestellt, das somit als wichtiges Werkzeug für ein integriertes Entwurfsoptimierungssystem zu sehen ist. Die hardware- und softwaretechnische Konfiguration soll in diesem Kapitel erarbeitet und beschrieben werden.

Zunächst wird das Zusammenwirken zwischen MOSES und dem PC-Cluster dargestellt.

6.1 Grobparallelisierung mit MOSES

6.1.1 Funktionsprinzip

Obwohl jedes System für sich durchaus komplex ist, gestaltet sich deren Zusammenwirken relativ einfach. Abbildung 6.1 zeigt das dabei zugrunde liegende Funktionsprinzip, das im Folgenden erläutert wird:

- Die einzelnen Funktionalitäten (z. B. die RSM) stellen Daten für die benötigten Analysen zusammen. Das Entwurfsoptimierungssystem generiert die zugehörigen Simulationsmodelle und erzeugt gegebenenfalls Steuerskripte. Dies kann auch außerhalb der Clusterumgebung auf einem Arbeitsplatzrechner erfolgen. Ein solcher Fall ist zum Beispiel gegeben, wenn ein CAD-Tool oder andere Programme zur

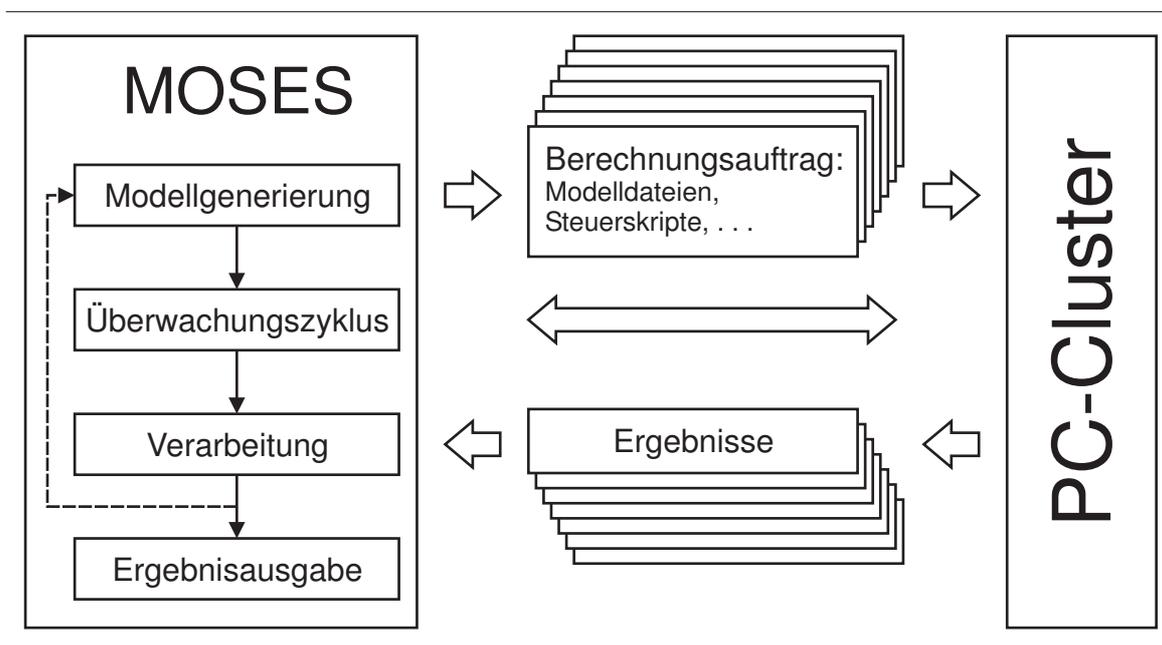


Abbildung 6.1: Funktionsprinzip zwischen dem Entwurfsoptimierungssystem MOSES und dem PC-Cluster

Modellgenerierung nur auf Windows-PCs oder Workstations lauffähig sind. Die Modelldaten müssen dann in einem Zwischenschritt in ein von außen nutzbares Dateisystem kopiert werden.

- Sind alle Daten für die Analysen zusammengestellt, werden diese beim Cluster in Auftrag gegeben. Dieser arbeitet alle Berechnungsaufträge in der später beschriebenen Weise ab. Dabei überwacht das Entwurfsoptimierungssystem den Fortschritt der Analysen und wartet, bis alle Ergebnisdateien vorliegen.
- Nach dem erfolgreichen Abschluss der Analysen werden die relevanten Ergebnisse ausgelesen. Wie auch bei der Modellgenerierung ist es dabei nicht zwingend erforderlich, dass die auswertenden Programme (Postprozessoren, Laminatanalyse) auf dem Cluster ausgeführt werden.
- Die so gewonnenen Simulationsergebnisse können dann verarbeitet werden. Im Falle der RSM bedeutet dies die Erstellung von Approximationen für die Ziel- und Restriktionsfunktionen und den Aufruf eines Optimierungsverfahrens. Bei iterativen Verfahren, wie zum Beispiel einem Evolutionären Algorithmus, kann nach einer Zwischenauswertung ein neuer Satz von Berechnungsaufträgen erstellt und abgearbeitet werden.
- Das Endergebnis wird abschließend in geeigneter Form ausgegeben.

Die Grobparallelisierung von Prozessen hat den Vorteil, dass lediglich zu Beginn und zum Abschluss eines Berechnungsauftrags ein gewisser Kommunikationsaufwand zwi-

schen Steuer- und Rechenprozessen erforderlich ist. Damit ist eine mit der Zahl der verwendeten Prozessoren nahezu linear steigende Leistung zu erwarten.

6.1.2 Resultierende Leistungssteigerung

Die Leistungssteigerung bei der Anwendung des PC-Clusters hängt unter anderem vom Aufwand ab, der vor und nach der Durchführung eines Berechnungsauftrags erforderlich ist. Der Aufwand wird dabei im Wesentlichen durch das Kopieren von Modell- und Ergebnisdateien bestimmt. Dafür kann bei besonders großen Datenmengen relativ viel Zeit benötigt werden. Je nachdem, wie lange die zugehörige Analyse dauert, ergibt sich ein unterschiedlicher Leistungssteigerungsfaktor (Speedup-Faktor).

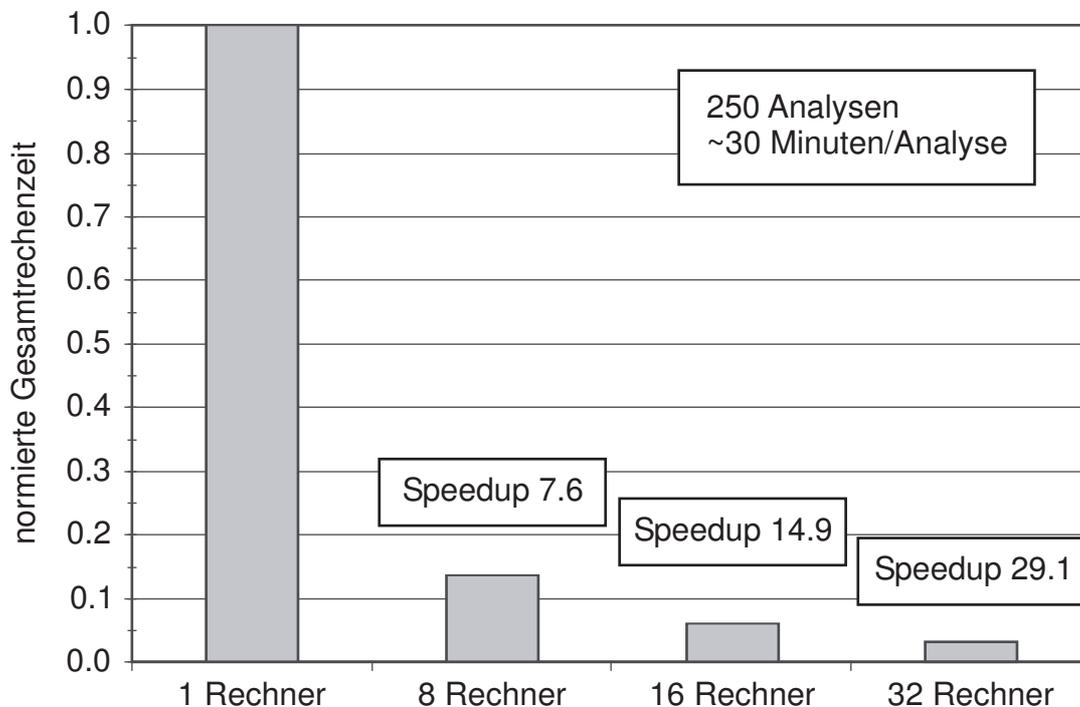


Abbildung 6.2: Leistungssteigerung bei der Bearbeitung eines durchschnittlichen Problems auf dem PC-Cluster

Für die Bearbeitung eines durchschnittlichen Problems ist in Abbildung 6.2 die durch die Grobparallelisierung auf dem Cluster erzielbare Leistungssteigerung dargestellt. Dazu wurden 250 Analysen mit einer Berechnungsdauer von jeweils 30 Minuten auf einem Rechner sowie auf 8, 16 und 32 Rechnern durchgeführt und die dabei benötigte Zeit gemessen. Dabei betrug die sequenzielle Bearbeitungszeit mehr als 5 Tage, während die Analysen auf 32 Rechnern etwas mehr als 4 Stunden durchgeführt waren. Der Speedup-Faktor ergibt sich aus der mit einem Rechner entstehenden Bearbeitungszeit bezogen auf die mit mehreren Rechnern benötigte Zeit. Der theoretisch mögliche Speedup-Faktor entspricht jeweils der Anzahl der verwendeten Rechner.

Die Untersuchung und der tägliche Betrieb zeigen, dass durch die Grobparallelisierung mit dem realisierten PC-Cluster eine sehr gute Leistungssteigerung erzielt werden kann. Der Aufbau des PC-Cluster, beginnend mit der Auswahl bis hin zur Umsetzung eines geeigneten Hardware- und Softwarekonzepts, wird im Folgenden beschrieben.

6.2 Auswahl eines Hardware- und Softwarekonzepts

6.2.1 Anforderungen

Zur Zusammenstellung eines geeigneten Hardware- und Softwarekonzepts sind zunächst die Anforderungen an das zu realisierende System zu klären. Diese ergeben sich aus den zu bewältigenden Aufgaben und anderen Gesichtspunkten folgendermaßen:

- Viele der beschriebenen Aufgaben, darunter auch die Response Surface Methode (RSM), erfordern die Durchführung vieler, voneinander unabhängiger Systemanalysen. Da deren Berechnung in möglichst kurzer Zeit erfolgen soll, ist es umso besser, je mehr Analysen gleichzeitig durchführbar sind. Das System muss also über möglichst viele Prozessoren/Rechner verfügen, die jeweils den Anforderungen der einzelnen Analysen bezüglich Arbeitsspeicher, Festplattenplatz, usw. genügen.
- Weiterhin ist es wünschenswert, vorhandene Ressourcen (Arbeitsplatzrechner, Workstations, ...) immer dann zusätzlich zum Grundsystem nutzen zu können, wenn diese nicht für andere Zwecke benötigt werden. Dies ist zum Beispiel in der Nacht, am Wochenende oder generell bei Abwesenheit der jeweiligen Benutzer gegeben.
- Daraus ergibt sich wiederum die Forderung, dass heterogene Rechnerarchitekturen mit unterschiedlichen Betriebssysteme in das System eingebunden werden können.

- Betrachtet man die rasche Entwicklung auf dem Computersektor, so ist es sinnvoll und notwendig, dass ein bestehendes System mit wenig Aufwand und kostengünstig erweitert und verbessert werden kann. Gerade bei voll ausgelasteten Systemen soll der Betrieb durch Erweiterungsmaßnahmen nicht unterbrochen und möglichst nur wenig gestört werden.
- Ähnliche Eigenschaften müssen auch bezüglich der Ausfallsicherheit gewährleistet sein. Falls einzelne Komponenten wie Prozessoren, Speichermodule und Festplatten durch einen Defekt ausfallen, so darf davon nicht das gesamte System in Mitleidenschaft gezogen werden.
- Bei der Erfüllung der Anforderungsliste darf der Kostenfaktor nicht außer Acht gelassen werden. Hier gilt es, mit einem verfügbaren Budget möglichst alle genannten Anforderungen optimal zu erfüllen.

6.2.2 Verfügbare Systeme bzw. Konzepte

Im Bereich der skalierbaren Parallelrechner stehen mehrere Grundkonzepte zur Auswahl. Die wesentlichen Eigenschaften dieser Systeme werden im Folgenden diskutiert:

- Zur Verfügung stehende kommerzielle Systeme sind oft für die parallelisierte Berechnung einer einzelnen Analyse optimiert, bei der ein hoher Kommunikationsaufwand zwischen den Prozessen erforderlich ist. Diese Symmetric Multiprocessor (SMP) Systeme teilen sich die zentralen Einheiten wie Arbeitsspeicher, Systembus, usw. und verfügen in der Regel über 2^n Prozessoren, die in ihrer Leistungsfähigkeit den PC-Prozessoren überlegen sind¹. Das Gesamtsystem wird von einer einzigen Instanz des Betriebssystems gesteuert.
- Andere kommerzielle Systeme, die über eine variablere und größere Zahl an Prozessoren verfügen, stehen mit den Massively Parallel Processor (MPP) Systemen zur Verfügung. Jeder der Knoten verfügt über einen Prozessor sowie die wesentlichen zentralen Einheiten (meist keine Festplatten) und kann somit als unabhängig betrachtet werden. Auf jedem Knoten läuft eine Instanz des Betriebssystems. Die Verbindung erfolgt über ein Hochgeschwindigkeitsnetzwerk. Auch auf diesen Systemen ist die Bearbeitung parallelisierter Einzelanalysen möglich.
- Von den Eigenschaften her betrachtet, sind Cluster den MPP-Systemen relativ ähnlich. Ein Cluster ist ein Parallelrechner, der aus mehreren eigenständigen Rechnern (PCs, Workstations oder auch SMPs) zusammengesetzt ist. Die Knoten sind mit mehr oder weniger leistungsfähigen Standard Netzwerkkomponenten verbunden. Meist besitzen die Knoten eine einheitliche Hardware- und Softwarekonfiguration. Die Einzelrechner können auf ein gemeinsames Dateisystem zugreifen, das meist von einem Server verwaltet wird. Je nach Betriebssystem

¹Stand 2000

und Simulationssoftware kann auch auf einem Cluster eine einzelne Analyse auf mehreren Prozessoren durchgeführt werden.

- Wie die Bezeichnung „verteilte Rechner“ bereits verrät, sind die Einzelrechner bei einem solchen System nicht an den gleichen Ort gebunden. Dabei ist es häufig der Fall, dass die in das System integrierten Rechner eine andere Hardwarekonfiguration besitzen. Teilweise liegt sogar eine andere Rechnerarchitektur vor, wenn die Knoten zum Beispiel aus PCs und Workstations bestehen.

6.2.3 Auswahl eines Systems bzw. Konzepts

Gerade SMP-Systeme sind aufgrund ihrer Spezialisierung auf die parallelisierte Berechnung einer einzelnen Analyse und den damit verbundenen hohen Anforderungen an die zentralen Einheiten und Prozessoren sehr teuer. Die pro Zeiteinheit durchführbare Zahl an Einzelanalysen steht hier in einem schlechten Verhältnis zu den Kosten eines solchen Systems. Weiterhin sind SMP-Systeme anfällig gegenüber dem Ausfall einzelner Komponenten. Hier ist verhältnismäßig schnell mit einem Totalausfall zu rechnen. Die Erweiterung und/oder Verbesserung des Systems ist mit einem hohen Aufwand und vor allem mit hohen Kosten verbunden. Dies ist im Wesentlichen dadurch begründet, dass einzelne Komponenten nicht einzeln und unabhängig voneinander ausgetauscht werden können.

Bei MPP-Systemen ist die Leistung – gemessen an der pro Zeiteinheit durchführbare Zahl an Einzelanalysen – bezogen auf die Kosten schon deutlich höher. Auch bezüglich der Erweiterbarkeit und der Betriebssicherheit bei Ausfall einzelner Komponenten der Knoten schneiden diese Systeme gegenüber SMP-Systemen wesentlich besser ab. Betrachtet man den Leistung/Kosten -Faktor, so können kommerzielle SMP- und auch MPP-Systeme ausgegrenzt werden.

Cluster und verteilte Rechner können die in Abschnitt 6.2.1 aufgeführten Anforderungen am besten erfüllen. Als Knotenrechner sind PCs am besten geeignet. Diese sind in den letzten Jahren äußerst leistungsfähig geworden, wobei die Kosten sogar gesunken sind. Da die primäre Aufgabe des aufzubauenden Systems in der möglichst raschen Abarbeitung vieler voneinander unabhängiger Analysen liegt, werden bezüglich der Vernetzung der Einzelrechner keine hohen Anforderungen gestellt. Die einzelnen Berechnungsaufträge werden, wie auch bei den zuvor genannten Systemen, über ein Queueing-System auf die einzelnen Rechner verteilt. Dabei können unterschiedlich leistungsstarke Rechner, auch verschiedener Architektur, eingesetzt werden. Schnellere Rechner bekommen in diesem Fall mehr Berechnungsaufträge zugewiesen als langsamere.

6.3 Beschreibung des realisierten PC-Clusters

Der realisierte Cluster ist in Abbildung 6.3 zu sehen. Abbildung 6.4 zeigt schematisch die hardwaretechnische Konfiguration des System. Diese sowie die einzelnen Komponenten werden im Folgenden näher beschrieben.



Abbildung 6.3: Foto des Clusters

6.3.1 Konfiguration

Rechner und Prozessoren

In der ursprünglichen Zusammenstellung hatten die Knoten-PCs eine identische Konfiguration. Um den Anforderungen komplexerer Simulationsmodelle zu genügen und

teilweise wegen Hardwaredefekten, wurden 16 Rechner gegen leistungsfähigere ausgetauscht.

Um sehr große Datenmengen verarbeiten zu können, verfügt das Serversystem über zwei Prozessoren und kann die Daten über eine schnelle SCSI-Verbindung mit bis zu 160 MB/s auf ein RAID-System schreiben. Dieses bietet neben dem Vorteil, dass die Daten parallel auf mehrere Festplatten verteilt und somit schneller geschrieben werden, auch eine entsprechende Datensicherheit. Durch den Verteilungsalgorithmus werden die Daten in einer speziellen Form gespeichert, so dass bei Ausfall einer Festplatte alle Daten ohne große Verzögerung rekonstruiert werden können.

Netzwerkkomponenten

Die Vernetzung der Rechner ist in Abbildung 6.4 schematisch dargestellt. Alle Knoten verfügen über eine 100 Mbit-Netzwerkkarte. Zur Aufteilung der Netzwerklast sind jeweils 16 der 32 Rechner mit 100 Mbit/s mit je einem Netzwerkschicht verbunden. Damit die Verbindungen vom Switch zum Server keinen „Flaschenhals“ darstellen, ist hier jeweils eine maximale Übertragungsrate von 1 Gbit/s möglich. Der Server verfügt hierzu über insgesamt drei 1 Gbit-Netzwerkkarten, von denen die dritte für die Verbindung zu den Arbeitsplatzrechnern und Workstations konfiguriert ist.

Wie in Abbildung 6.4 zu sehen, sind die einzelnen Knoten nur über den Server erreichbar. Dies erleichtert unter anderem auch den Schutz des Clusters vor unbefugten, möglicherweise schädlichen Zugriffen, indem der Server als Firewall konfiguriert wird.

Betriebssystem

Das für den Betrieb eines PC-Clusters geeignetste Betriebssystem ist Linux, für das die meisten Hersteller gängiger Analyse-Software (ANSYS, NASTRAN, LS-Dyna, MATLAB, ...) inzwischen Versionen anbieten. Die meist frei verfügbaren Distributionen des Unix für PCs stellen im Prinzip die gleichen Funktionalitäten zur Verfügung, die auch die Unix-Varianten für große Parallelrechner bieten.

Hierzu gehören unter anderem folgende Funktionalitäten:

- Eine zentrale Benutzerverwaltung erfolgt über NIS (Network Information System).
- Das Netzwerkdateisystem NFS (Network File System) ermöglicht den gemeinsamen Zugriff auf Daten, auch von außen.
- Mehrere Benutzer können gleichzeitig am System (Server) arbeiten und dort, wie auch auf den Knoten, mehrere Prozesse starten.

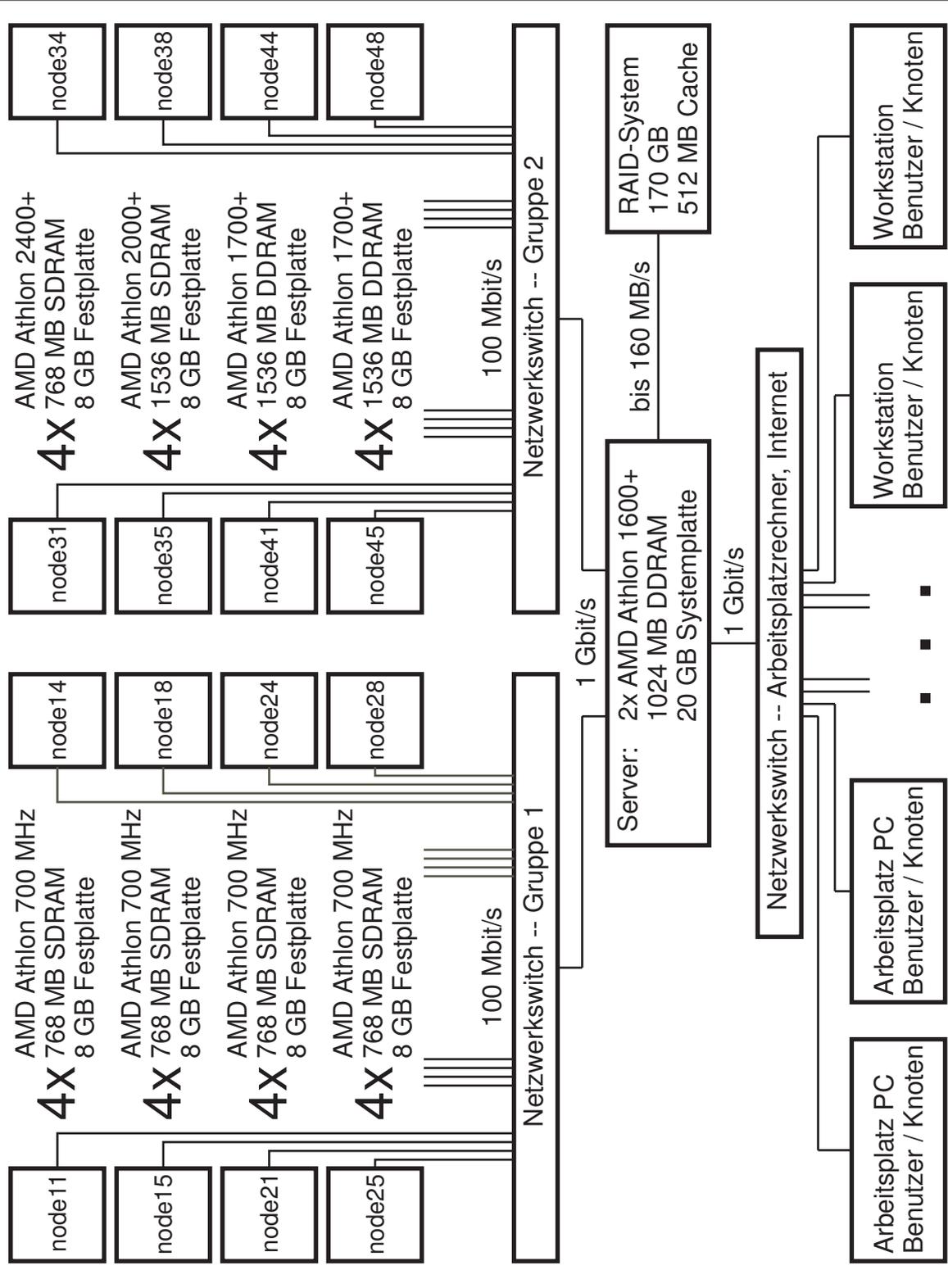


Abbildung 6.4: Schematischer Aufbau des Clusters

Darüber hinaus gibt es für Linux eine Vielzahl nützlicher, kostenfreier Softwarepakete, die ihren kommerziellen Vorbildern oft nur in wenigen Punkten unterlegen sind. Hierzu gehört auch das für den Betrieb des Clusters wichtige Job-Queueing-System OpenPBS (Portable Batch System).

6.3.2 Job-Queueing-System

Für den Betrieb des Clusters stellt sich die Frage, wie die Analysen auf die einzelnen Knoten verteilt werden können. Einfache Programme zur Lastverteilung ermitteln regelmäßig einen Wert für die Auslastung jedes Knotens. Der Knoten mit der geringsten Auslastung bekommt einen Prozess zugewiesen. Dies erfolgt solange, bis die Auslastung aller Knoten einen Grenzwert erreicht hat.

Bei vielen Benutzern mit sehr unterschiedlichen Berechnungsaufträgen (bzgl. Rechen-dauer, Anforderungen an Ressourcen wie Arbeitsspeicher, Festplattenplatz, ...) ist mit diesen reinen Load-Balancing-Systemen schnell ein unbefriedigender Zustand erreicht. Mit ihnen lässt sich in der Regel lediglich eine Abarbeitung nach dem FCFS-Prinzip (First Come – First Served) bewerkstelligen. Hat nun ein Benutzer 20000 Analysen für eine statistische Untersuchung in Auftrag gegeben (Gesamtdauer beispielsweise 48 Stunden), so muss ein anderer Benutzer, der zum Beispiel 10 Analysen mit je 1/2 Stunde Berechnungsdauer durchführen möchte, zu lange warten.

Anforderungen

Aus der beispielhaft dargestellten Problematik lassen sich, zusammen mit anderen Überlegungen, folgende Anforderungen an ein System zur Verteilung der Berechnungsaufträge zusammenstellen:

- Das System soll anstelle eines reinen FCFS-Prinzips eine prioritätsgesteuerte Abarbeitung der Berechnungsaufträge erlauben.
- Da auch dies mit dem Nachteil verbunden ist, dass Aufträge mit niedriger Priorität sehr spät oder gar nicht zum Zuge kommen, müssen Mechanismen zur Verfügung stehen, die solchen Berechnungsaufträgen nach entsprechender Zeit, ungeachtet deren Priorität, den Vorzug geben.
- Im Falle vieler, von verschiedenen Benutzern stammender Aufträge mit gleicher Priorität ist es sinnvoll, diese nach einem Rotationsprinzip abzuarbeiten. Dies gilt allerdings nur, wenn diese eine ähnliche Berechnungsdauer haben.
- Die Aufträge müssen abhängig von den Anforderungen, die sie an die Systemressourcen (Arbeitsspeicher, Festplattenplatz, ...) stellen, auf geeigneten Knoten bearbeitet werden.

- Knoten sollen leicht und ohne Betriebsstörung hinzugefügt oder entfernt werden können. Damit können Erweiterungen und Verbesserungen der Knoten problemlos vorgenommen werden.
- Bietet das Job-Management-System die Möglichkeit, Knoten zeitgesteuert in den Clusterbetrieb aufzunehmen und ebenso wieder herauszunehmen, kann der Cluster temporär durch verteilte Rechner erweitert werden. Dies bietet sich zum Beispiel für Arbeitsplatzrechner an, wenn diese bei Abwesenheit der jeweiligen Benutzer nicht genutzt werden.
- Um hier möglichst alle in Frage kommenden Rechner einbeziehen zu können, stellt sich die Forderung, dass auch heterogene Rechnerarchitekturen und Betriebssysteme unterstützt werden.
- Schlägt ein Berechnungsauftrag fehl, so muss dieser problemlos neu gestartet werden können. Auf diese Weise lässt sich eine Fehlerbehandlung realisieren.
- Als letzter, nicht unwesentlicher Punkt ist noch die Möglichkeit zum interaktiven Arbeiten an einzelnen Knoten zu nennen. Dies ist unter anderem dann sinnvoll und erforderlich, wenn die Analysen in ihrem Fortgang anhand grafischer Ausgaben vom Benutzer überwacht werden müssen.

Auswahl eines Systems

Zum Zeitpunkt des Aufbaus des hier beschriebenen Clusters standen verschiedene kommerzielle und frei verfügbare Job-Management-Systeme zur Auswahl. Als kommerzieller Vertreter sei hier das Produkt LSF (Load Sharing Facility) genannt. Was die Funktionalität und auch die Verbreitung angeht, ist LSF nach wie vor Marktführer.² Bei unveränderten Kosten von etwa 1200 Dollar pro Knoten, wird jedoch ein zur Verfügung stehendes Budget so stark reduziert, dass nahezu nur noch die Hälfte an Rechenleistung beschafft werden kann.

Da in der Regel bei weitem nicht alle Funktionalitäten der kommerziellen Vertreter benötigt werden, um die genannten Anforderungen zu erfüllen, wurde ein frei verfügbares Job-Queueing-System gewählt. Als solche standen das DQS (Distributed Queueing System) und das bereits genannte OpenPBS (Open Portable Batch System)³ zur Auswahl. Da DQS jedoch in seinem Entwicklungsstand „eingefroren“ wurde, um unter dem Namen CODINE auf die kommerzielle Schiene zu wechseln, wurde dies nur in einer frühen Phase des Clusters eingesetzt. Somit fiel die Entscheidung auf das OpenPBS, das die beschriebenen Anforderungen ohne Einschränkungen erfüllt. Als Vorteil ist dabei zu sehen, dass die Weiterentwicklung von OpenPBS durch eine verfügbare kommerzielle Version (PBS Pro) vorangetrieben wird.

²Stand 2003

³Siehe [OpenPBS Homepage]

Bezüglich CODINE sei der Vollständigkeit halber noch angemerkt, dass dies in einer Weiterentwicklung mittlerweile unter dem Namen SGE (Sun Grid Engine)⁴ wieder frei zur Verfügung steht. Die freie Version von SGE bietet – ähnlich wie OpenPBS – eingeschränkte Funktionalitäten der kommerziellen Version.

Realisiertes Job-Queueing-System mit temporären Knoten

Das zentrale Konzept von Job-Management-Systemen besteht darin, alle Berechnungsaufträge in einer Queue („Warteschlange“) zu sammeln und diese unter Berücksichtigung verschiedener Kriterien zur Bearbeitung auszuwählen. Das einfachste Kriterium ist dabei das FCFS-Prinzip, dessen entscheidender Nachteil bereits genannt wurde. Zur Erfüllung der zu Beginn dieses Abschnitts aufgeführten Anforderungen wurde mit OpenPBS folgendes Konzept realisiert (vgl. Abbildung 6.5):

- Jeder Benutzer bekommt eine eigene Queue zugewiesen. Innerhalb dieser kann der Benutzer die Prioritäten seiner Berechnungsaufträge frei festlegen. Darüber hinaus besteht auch die Möglichkeit, eigene Aufträge, die bereits in „Warteposition“ stehen, für eine Bearbeitung zu sperren. Somit ergeben sich für den einzelnen Benutzer vielfältige Möglichkeiten, um auf die Reihenfolge der Abarbeitung von Berechnungsaufträgen Einfluss zu nehmen.
- Den Berechnungsaufträgen aller Benutzerqueues können nur Prioritäten bis zu einer festgelegten Größe zugewiesen werden. Damit sind diese Queues gleichberechtigt.
- Um besonders dringend benötigten Analysen den Vorzug zu geben, steht eine Queue mit hoher Priorität zur Verfügung (`important_queue`), deren Berechnungsaufträge immer zuerst ausgeführt werden.
- Die Berechnungsaufträge werden vom Scheduler zur Bearbeitung ausgewählt. Dieser ist folgendermaßen konfiguriert:
 - Sind alle Aufträge der Benutzerqueues bezüglich ihrer Priorität – und in gewissen Grenzen auch bezüglich ihrer angegebenen Rechendauer – gleich, so werden diese im Rotationsprinzip abgearbeitet.
 - Haben die Berechnungsaufträge spezielle Ressourcenanforderungen, so werden diese so lange zurückgestellt, bis ein geeigneter Knoten frei ist. Währenddessen werden weiter hinten eingereihte Aufträge bearbeitet („Backfill“).
 - Der Scheduler ermöglicht es auch, dass Prozesse nur zu bestimmten Zeiten (z. B. nachts, am Wochenende, ...) gestartet werden, wenn der Benutzer dies vorgibt. Somit kann eine Entlastung für den Tagesbetrieb erreicht werden.

⁴Siehe [SGE Homepage]

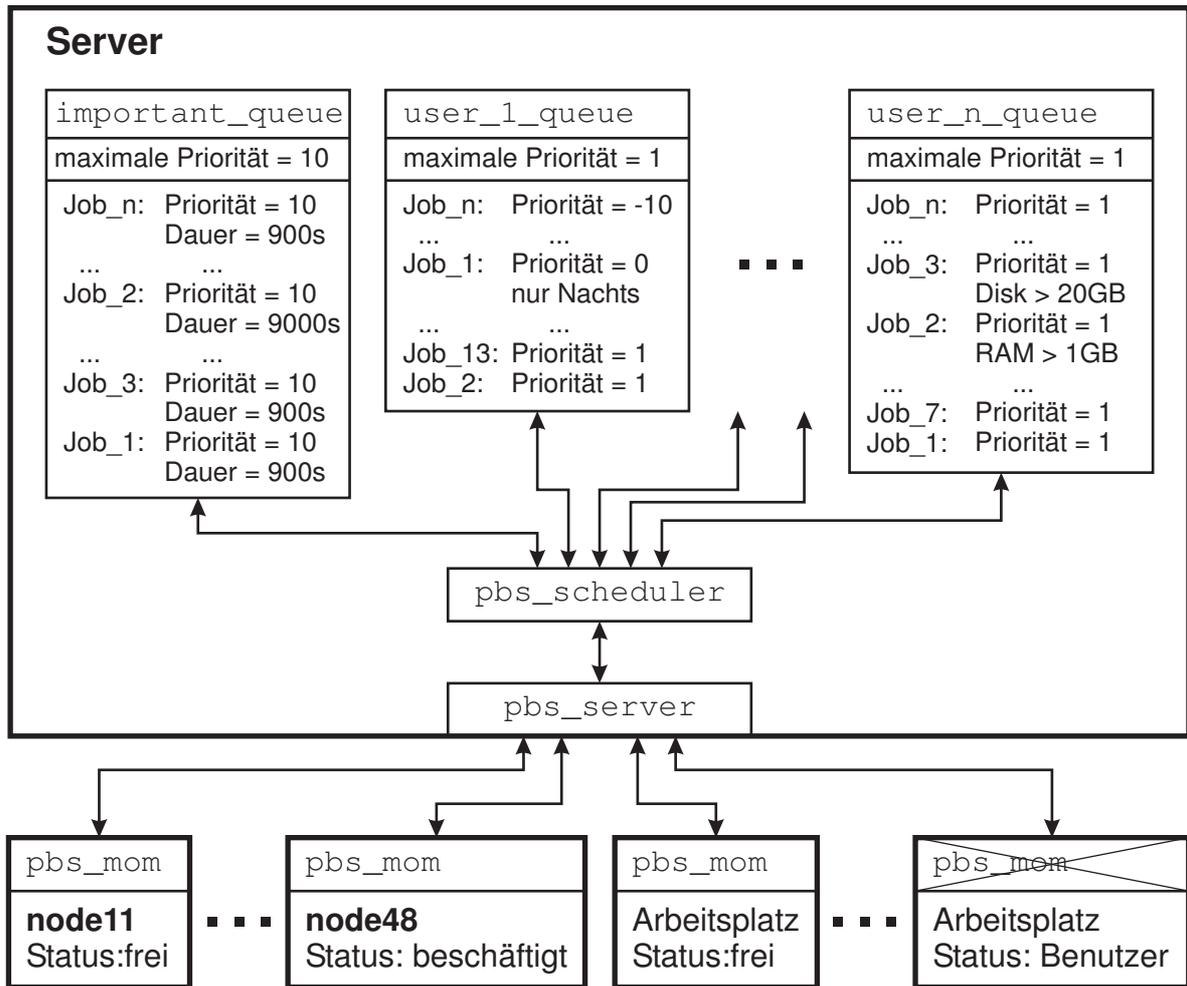


Abbildung 6.5: Realisiertes Job-Queueing-System basierend auf OpenPBS

- Um interaktives Arbeiten zu ermöglichen, muss der Benutzer einen speziell gekennzeichneten Auftrag erteilen. Dieser beinhaltet lediglich die Zuweisung einer Shell (Kommandozeileninterpreter) auf dem nächsten freien Knoten, der für die Zeit des interaktiven Betriebs keine weiteren Berechnungsaufträge zugewiesen bekommt.

Mit dem realisierten Konzept sind alle zu Beginn dieses Abschnitts beschriebenen Anforderungen erfüllt. Die bis jetzt noch nicht beschriebene Möglichkeit, den Cluster temporär durch nicht benötigte Arbeitsplatzrechner zu erweitern, wird im Folgenden erläutert:

- Wie bereits beschrieben, bestimmt der Scheduler (`pbs_scheduler`) welche Aufträge wann und in welcher Reihenfolge bearbeitet werden. Hierzu wird auf Informationen des Job-Servers (`pbs_server`) zurückgegriffen. Dieses Programm

überwacht den Status der Knoten und koordiniert den Datentransfer zu und von den Knoten, indem es mit den dort laufenden Programmen (`pbs_mom`) kommuniziert. Somit ist jederzeit bekannt, ob ein Knoten zur Bearbeitung eines Auftrags zur Verfügung steht oder nicht.

- Da OpenPBS auch für andere Betriebssysteme als Linux zur Verfügung steht, können andere Rechner, die auf unterstützten Plattformen laufen, problemlos genutzt werden. Diese werden hierzu als Knoten konfiguriert, indem die PBS-Client-Programme (insbesondere `pbs_mom`) installiert und gestartet werden. Weiterhin werden die zusätzlichen Knoten in die Datenbank des Job-Servers eingetragen, die unter anderem auch Informationen über die verfügbaren Analyse-Tools enthält.
- Da die meisten Arbeitsplatzrechner mit Microsoft Windows betrieben werden, dies jedoch von OpenPBS nicht unterstützt wird, muss hier ein höherer Aufwand betrieben werden:
 - Auf den Arbeitsplatz-PCs muss Linux als zweites Betriebssystem installiert sein. Die Konfiguration ist dabei ähnlich wie bei den eigentlichen Knoten des Clusters. Da die Festplatten der meisten Arbeitsplatzrechner meist vollständig vom vorhandenen Betriebssystem genutzt werden, würde dies eine vollständige Neuinstallation beider Betriebssysteme erfordern, was vor allem auch die Zustimmung der Benutzer erfordert und gegebenenfalls zum Verlust lokal gespeicherter Daten führt, wenn diese nicht gesichert wurden. Die Einbindung von Arbeitsplatz-PCs konnte deshalb nur an zwei Rechnern kurzzeitig getestet werden.
 - Bei einem Arbeitsplatzrechner stellt sich das Problem, dass dieser zu einem bestimmten Zeitpunkt dem Benutzer wieder zur Verfügung stehen muss. Um zu vermeiden, dass dann noch Berechnungsaufträge bearbeitet werden, werden den temporären Knoten nur Aufträge zugewiesen, die bis zu einem vorgegebenen Zeitpunkt durchführbar sind. Die individuellen Zeitfenster können dazu in die Datenbank des Schedulers eingetragen werden. Sollten Aufträge nach der vom Eigentümer angegebenen, maximalen Berechnungsdauer nicht abgeschlossen sein, so werden diese abgebrochen. Der temporäre Knoten wird nach Ablauf des Zeitfensters automatisch heruntergefahren und steht somit dem Benutzer wieder zur Verfügung.

7 Anwendung bei multidisziplinären Optimierungsproblemen

In diesem Kapitel werden drei Optimierungsbeispiele behandelt. Das Ziel dabei ist die Verdeutlichung sowie die Beurteilung der Effizienz, die bei der Lösung komplexer Optimierungsaufgaben durch den Einsatz von Response Surface Approximationen (RSAs) und insbesondere der dafür erarbeiteten Vorgehensweisen und Verfahren erreichbar ist. Dabei soll auch eine Überprüfung der erarbeiteten Verfahren hinsichtlich ihrer Grenzen erfolgen. Als Vergleich dienen jeweils die mit einem mathematischen Optimierungsalgorithmus ermittelten Optima sowie der dabei erforderliche Berechnungsaufwand.

Bei zwei der untersuchten Problemstellungen handelt es sich um die in Kapitel 2 vorgestellten multidisziplinären Optimierungsaufgaben:

- *Thermisch-mechanischer Entwurf einer Sandwichkonstruktion*
- *Aeroelastic Tailoring eines Tragflügels (Strömungs-Struktur-Kopplung)*

Da RSAs die Verläufe stark variierender Systemantworten glätten, wird als drittes Optimierungsproblem die *Gestaltoptimierung einer Chopperscheibe aus kohlefaserverstärktem Kunststoff (CFK)* beschrieben. Hier führen kleine Änderungen in den Entwurfsvariablen teilweise zu sprunghaften Veränderungen in den Systemantworten.

7.1 Thermisch-mechanischer Entwurf einer Sandwichkonstruktion

Wie in Abschnitt 2.1.1 beschrieben, liegt bei der Optimierung thermisch-mechanisch gekoppelter Systeme fast ausschließlich eine reine Entwurfskopplung vor. Ein hierbei häufig anzutreffender Anwendungsfall ist die Betrachtung thermisch und mechanisch bedingter Verformungen bei Sandwichkonstruktionen. Ein Beispiel hierfür wird im Folgenden vorgestellt und untersucht.

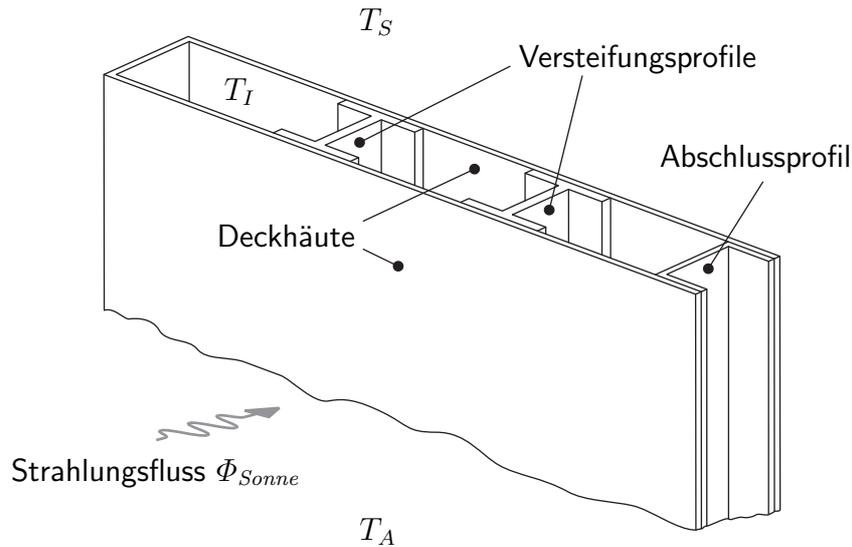


Abbildung 7.1: Thermisch-mechanische Problemstellung bei der sandwichartigen Konstruktion eines durch einseitige Sonneneinstrahlung belasteten Aufzugtürblatts

7.1.1 Problemstellung

Abbildung 7.1 zeigt den sandwichartigen Querschnittsaufbau eines durch einseitige Sonneneinstrahlung belasteten Aufzugtürblatts, bei dem zwischen den Außenhäuten Versteifungsprofile integriert sind. Durch die absorbierte Wärmestrahlung der Sonne erwärmt sich die – von der Schachtseite her gesehen – außen liegende Deckhaut. Dadurch unterliegt die Außenseite bei symmetrischem Aufbau des Türblatts einer größeren Wärmedehnung, wodurch sich das Türblatt stark krümmt und somit auch verklemmt. Gerade bei hohen Aufzugtüren (bis 4 m) ist dieser Effekt besonders stark ausgeprägt.

Um das Verformungsverhalten positiv zu verändern, wurden im Rahmen eines Projekts [Gleichmar 2002] verschiedene Maßnahmen diskutiert:

- Mit aktiven Maßnahmen, wie
 - schachtseitig angebrachten Heizelementen im Türblatt
 - Lüfter im Türblatt (für erzwungene Konvektion)

wären gute Erfolge zu erzielen. Diese wurden jedoch aufgrund des höheren Aufwands bei der Fertigung der Türblätter wie auch für die Realisierung einer Stromversorgung und Regelung verworfen. Zudem wäre eine Wartung und gegebenenfalls Instandsetzung der aktiven Elemente nach dem Zusammenbau kaum möglich, da es sich um eine Schweißkonstruktion handelt.

- Als passive Maßnahmen wurden folgende Varianten untersucht:
 - Verwendung eines dickeren Blechs für die Versteifungsprofile¹ ⇒ Trotz deutlich erhöhtem Gewicht nur eine geringe Reduktion der Verformung
 - Kohlefaserlaminat (sehr geringe Temperaturdehnung) auf der Innenseite der Türblattfront ⇒ Hohe Kosten im Vergleich zur erzielten Reduktion der Verformung
 - Beschichtung der Innenseiten der Außenhäute mit Emaillack (hoher Emissionsgrad) zur Wärmeübertragung durch Strahlung ⇒ Geringe Wirkung im Vergleich zum erhöhten Fertigungsaufwand
 - Verkleidung des Türblatts, so dass eine Isolierwirkung erzielt wird ⇒ Die erforderliche Isolierstärke würde eine Umkonstruktion des Führungssystems erfordern
 - Aluminium-Strangpressprofile (höhere Wärmeleitfähigkeit) anstelle der Versteifungsprofile aus Stahlblech ⇒ Wurde als erfolgversprechende Lösung ausgewählt

Mit den Querschnittsabmessungen der Aluminium-Strangpressprofile und deren Positionierung stehen die Entwurfsvariablen fest (siehe Abbildung 7.2). Hier muss nun unter Berücksichtigung sowohl der thermodynamischen als auch der strukturmechanischen Effekte ein optimaler Entwurf gefunden werden. Das Ziel der Optimierung ist die Minimierung der maximalen Verformung des Türblatts. Als Nebenbedingung darf die Masse des Türblatts nicht über der des Referenz-Türblatts liegen. Die Optimierungsaufgabe kann somit formuliert werden mit:

$$\begin{aligned} \text{Minimiere} \quad u_{max} &= u_{max}(\mathbf{x}, \Delta\mathbf{T}) \\ &= u_{max}(t_{G,\text{front}}, t_{G,\text{rück}}, t_S, b_G, p_{S1}, p_{S2}, \Delta\mathbf{T}) \end{aligned} \quad (7.1a)$$

$$\begin{aligned} \text{mit} \quad \Delta\mathbf{T} &= \Delta\mathbf{T}(\mathbf{x}) \\ &= \Delta\mathbf{T}(t_{G,\text{front}}, t_{G,\text{rück}}, t_S, b_G, p_{S1}, p_{S2}) \end{aligned} \quad (7.1b)$$

$$\begin{aligned} \text{so, dass} \quad & \left\{ \begin{array}{ll} m & \leq m_{Ref} \simeq 8 \text{ kg} \\ 1 \text{ mm} \leq t_{G,\text{front}}, t_{G,\text{rück}}, t_S & \leq 8 \text{ mm} \\ 10 \text{ mm} \leq b_G & \leq 40 \text{ mm} \\ 150 \text{ mm} \leq p_{S1} & \leq 210 \text{ mm} \\ 335 \text{ mm} \leq p_{S2} & \leq 395 \text{ mm} \end{array} \right. \end{aligned} \quad (7.1c)$$

Der Definitionsbereich der Entwurfsvariablen ist dabei im Wesentlichen durch Fertigungsrestriktionen bestimmt.

¹In der ursprünglichen Konstruktion handelt es sich um durch Abkanten eines Stahlblechs hergestellte C-Profile.

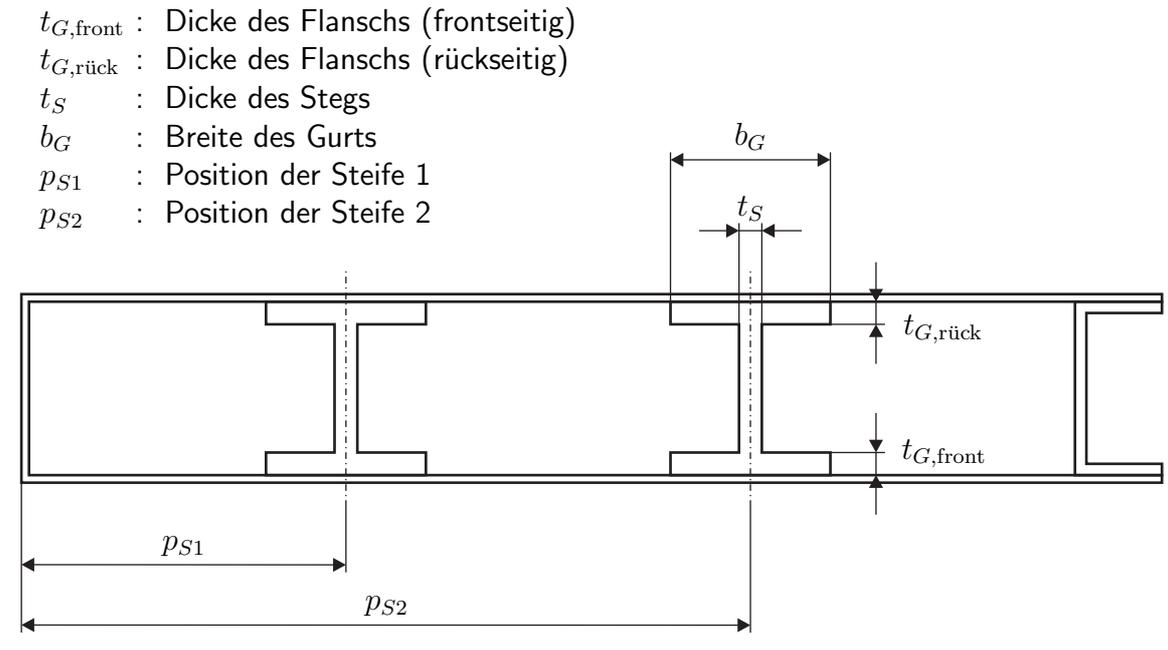


Abbildung 7.2: Entwurfsvariablen für die Positionierung und Dimensionierung der Versteifungen des Aufzugtürblatts

7.1.2 Beschreibung und Verifikation des Simulationsmodells

Beschreibung des Simulationsmodells

Das strukturmechanische Finite-Elemente-Modell wurde nach vorliegenden Konstruktionsplänen erstellt. Das Modell kommt in unveränderter Form auch bei der thermodynamischen Analyse zum Einsatz. Dabei werden lediglich andere Randbedingungen und ein geeigneter Elementtyp verwendet.

Zur Simulation der Sonneneinstrahlung wird eine Fläche verwendet, deren Oberflächentemperatur so gewählt ist, dass der davon ausgehende Strahlungsfluss dem der Sonne entspricht. Hierzu wurde das Strahlungsgesetz von Stefan und Boltzmann verwendet (siehe z. B. [Kuchling 1991]). Die von der Türblattfront absorbierte Strahlungswärme wird durch eine Reihe von Wärmetransportmechanismen abgebaut bzw. verteilt. Diese sind ebenfalls berücksichtigt:

- Die Wärmeleitung ist durch die Elemente des thermodynamischen Modells erfasst.
- Die vom Türblatt absorbierte Wärmemenge wird über Strahlung und freie Konvektion² an die Umgebung abgegeben. Hierbei sind die Außentemperatur T_A und die Temperatur im Schacht T_S maßgeblich.

²Freie Konvektion stellt in diesem Fall eine konservative Annahme dar.

- Zu einem gewissen Teil wird die Wärme auch durch Strahlung und durch die Luft im Türblattinneren zur Schachtseite transportiert. Beim letztgenannten Wärmetransportmechanismus spielt die Konvektion eine wesentliche Rolle. Um nicht für alle Analysen – zusätzlich zu den Teildisziplinen Thermodynamik und Strukturmechanik – auch noch eine Strömungssimulation durchführen zu müssen, werden die einfacher zu lösenden Konvektionsrandbedingungen bei der thermodynamischen Simulation benutzt. Hierzu wird die Temperatur der Luft im Türblattinneren aus dem Mittelwert zwischen der maximalen Temperatur an der Türblatfront und der minimalen Temperatur an der schachtseitigen Außenhaut gebildet.

Experimentelle Verifikation des Simulationsmodells

Um die Gültigkeit der im Simulationsmodell getroffenen Annahmen experimentell zu überprüfen, wurde ein Mustertürblatt gefertigt und getestet. Das Mustertürblatt besitzt eine Gesamthöhe von 1 m. Die zwei Aluminium-Profile sind im Verhältnis $1/3 : 1/3 : 1/3$ über die Türblattbreite ausgerichtet. Im Querschnitt, auf halber Türblatthöhe, sind insgesamt 6 Temperatursensoren und 4 Dehnmessstreifen angebracht. Abbildung 7.3 zeigt die Positionierung und die Bezeichnung der einzelnen Sensoren und Messpunkte.

Der Testaufbau ist in Abbildung 7.4 zu sehen. Das Mustertürblatt ist an den üblichen Lagerpunkten in einen steifen Rahmen gehängt. Die Wärmezufuhr erfolgt über einen gasbetriebenen Heizstrahler, der in einem Abstand von ca. 30 cm auf die Mitte der Türblatfront strahlt. Der Abstand wurde so gewählt, dass sich am Temperatursensor in der Mitte der Türblatfront eine Temperatur von ca. 65 °C einstellt. Abbildung 7.5 zeigt die Verläufe der Temperaturen an den sechs Messstellen während der Aufheiz- und auch der Abkühlphase. Der Aufheizvorgang wurde dabei solange durchgeführt bis sich ein stationärer Zustand eingestellt hat. Die Abkühlphase sollte weitere Daten zur Beurteilung des Messsystems liefern.

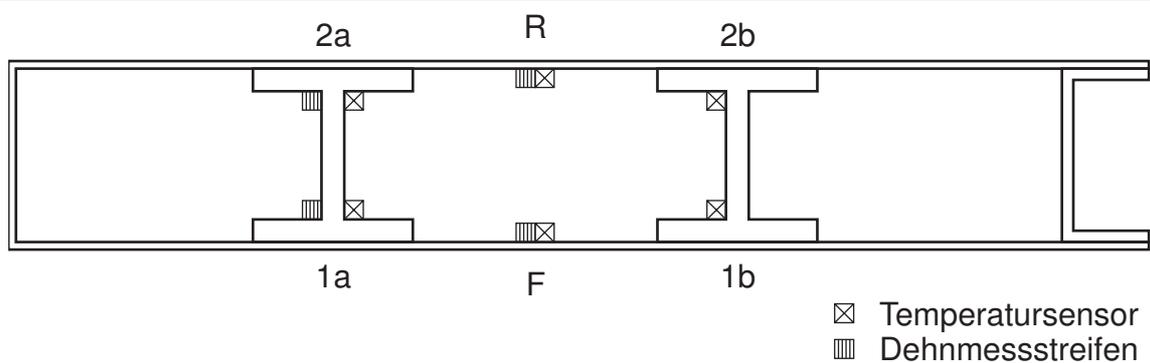


Abbildung 7.3: Position der Sensoren im Querschnitt des Mustertürblatts



Abbildung 7.4: Versuchsaufbau für den Test des Mustertürblatts

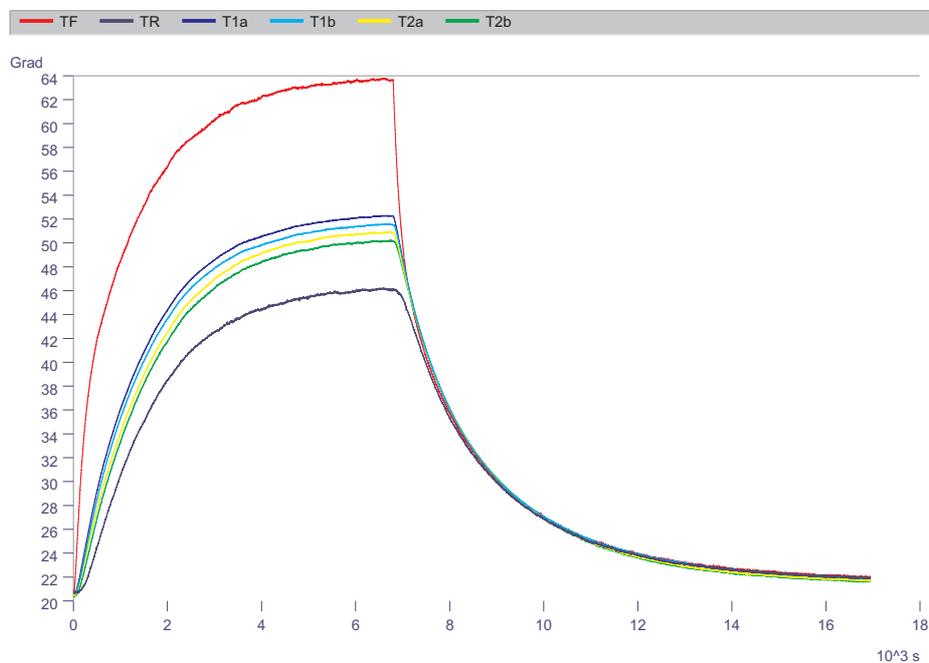


Abbildung 7.5: Temperaturverlauf an den sechs Messstellen beim Aufheizen und Abkühlen

Betrachtet man die Temperaturverläufe, so steigt die Temperatur in der Mitte der Türblattfront am stärksten. Mit zeitlicher Verzögerung folgen die Temperaturen der Messstellen an den Aluminium-Strangpressprofilen und mit weiterer Verzögerung die Temperatur in der Mitte der Türblattrückwand. Wie aufgrund der hohen Wärmeleitfähigkeit von Aluminium zu erwarten, liegen die Temperaturen der Messstellen 1a, 2a, 1b und 2b dicht beieinander. Der Unterschied erklärt sich einerseits durch den unsymmetrischen Querschnitt des Türblatts und andererseits durch die Tatsache, dass die Strahlungsintensität des Heizstrahlers nicht ideal verteilt ist. Beim Abkühlen liegen die Temperaturen binnen sehr kurzer Zeit beisammen, wobei wie erwartet die Temperaturen im Inneren des Türblatts etwas höher sind als außen. Das Verhalten spricht somit für einen zuverlässigen Testaufbau.

Vergleichsrechnung zur Beurteilung der Annahmen im Simulationsmodell anhand der experimentellen Daten

Zur Beurteilung der Modellannahmen wurde ein dem Mustertürblatt entsprechendes Simulationsmodell erstellt und die damit berechneten Werte mit den experimentellen Daten verglichen. Zur Simulation des Heizstrahlers wurde die sonst zur Repräsentation der Sonneneinstrahlung verwendete Fläche verkleinert. Dieser wurde die beim Versuch mit einem Temperaturfühler gemessene Temperatur am Heizstrahler zugewiesen. Die so ermittelte Temperaturverteilung ist in Abbildung 7.6 dargestellt.

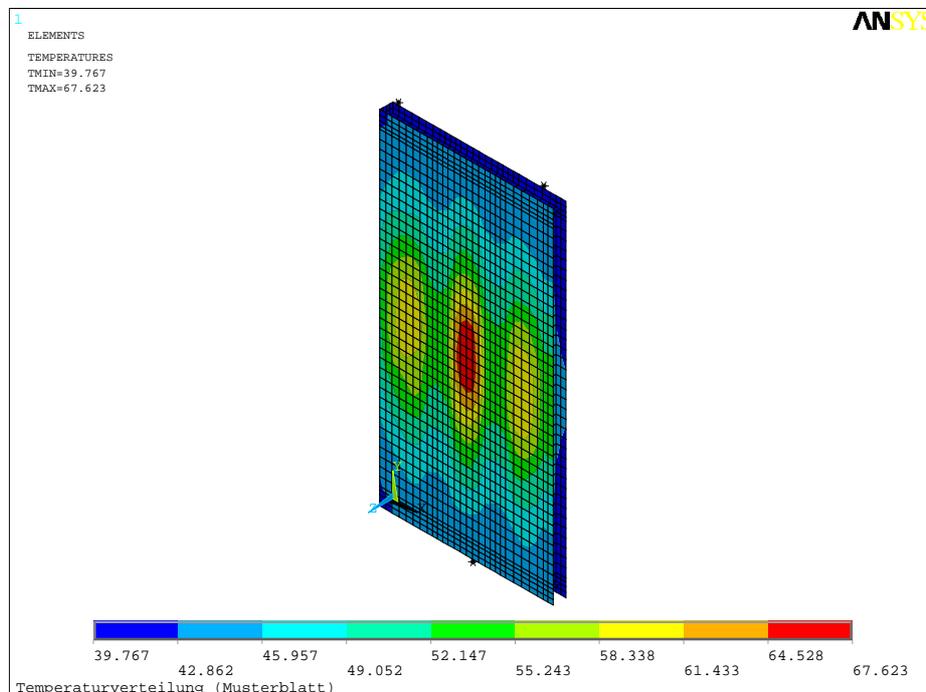


Abbildung 7.6: Temperaturverteilung im FE-Modell des Mustertürblatts

Positiv fällt beim Vergleich mit den gemessenen Werten auf, dass der rechnerisch ermittelte Temperaturgradient größer ist als der experimentell gemessene. Die rechnerisch ermittelte Verformung liegt somit erwartungsgemäß ebenfalls über dem experimentell gemessenen Wert. Dies lässt den Schluss zu, dass die Idealisierungen bezüglich der Wärmeübertragung als konservativ eingestuft werden können. Der Unterschied kommt von der fehlenden Modellierung des Luftvolumens im Inneren des Türblatts. Der hierdurch zusätzlich wirksame Wärmeaustausch durch Wärmeleitung und Konvektion verringert den Temperaturgradienten weiter, was in einem einmalig erstellten, sehr rechenaufwendigen Simulationsmodell bestätigt werden konnte. Hier liegt der Unterschied zwischen Messung und Rechnung nur noch bei ca. 2%.

7.1.3 Referenzlösungen zur Beurteilung der Optimierungsergebnisse

Verformungsverhalten der Ausgangskonstruktion

Um die bei der Optimierung erzielte Verbesserung quantifizieren zu können, werden die Daten und Simulationsergebnisse der Ausgangskonstruktion kurz dargestellt. In der ursprünglichen Ausführung des Türblatts werden als Versteifungen zwei C-Profile verwendet. Diese sind aus 1.5 mm starkem Stahlblech durch Abkanten hergestellt und besitzen zusammen eine Masse von 7.93 kg.

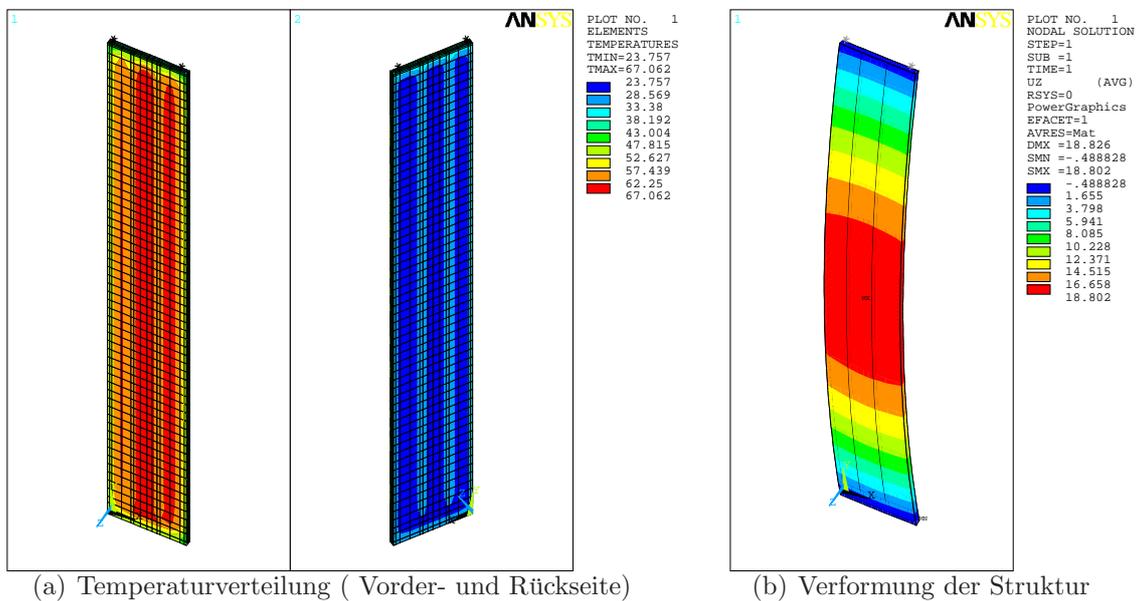


Abbildung 7.7: Temperaturverteilung und Verformung bei der Ausgangskonstruktion des Aufzugtürblatts

Mit der thermodynamischen Simulation ergibt sich die in Abbildung 7.7(a) dargestellte Temperaturverteilung. Der maximale Temperaturunterschied zwischen Türblattfront und -rückwand beträgt 43.3°C . Mit dieser Temperaturverteilung ergibt sich bei der strukturmechanischen Analyse eine maximale Verformung von 18.8 mm.

Optimierung mit einem SQP-Verfahren unter Verwendung der exakten Lösung der Systemgleichungen

Zur Beurteilung des mit einer RSA für (7.1a) ermittelten Optimums, wurde die Optimierungsaufgabe (7.1) mit dem in MATLAB verfügbaren SQP-Verfahren (`fmincon`) gelöst. Dabei waren, je nach Startvektor, zwischen 141 und 172 Systemanalysen erforderlich. Bei sequenzieller Durchführung der Analysen auf einem Rechner ergab sich eine Gesamtoptimierungsdauer von bis zu 34 Stunden. Die Daten der einzelnen Optimierungsläufe sind in nachfolgender Tabelle zusammengestellt.

Startwerte [mm]						System- analysen	Optimaler Entwurf [mm]						u_{max} [mm]
$t_{F,front}$	$t_{F,rück}$	t_S	b_F	p_{S1}	p_{S2}		$t_{F,front}$	$t_{F,rück}$	t_S	b_F	p_{S1}	p_{S2}	
4	4	4	20	180	365	172	1.0	1.0	6.8	40.0	180.2	360.5	3.48
1	1	1	10	150	395	153	1.0	1.0	6.8	40.0	167.7	372.5	3.42
8	8	8	40	210	335	141	1.0	1.0	6.8	40.0	200.2	343.7	3.55

Von allen Startpunkten aus wird das gleiche Optimum für die Querschnittsabmessungen der Versteifungen ermittelt. Für die Position der Steifen ergeben sich jedoch jeweils andere Werte. Dies liegt unter anderem daran, dass die Toleranz für die Konvergenz des verwendeten Optimierungsalgorithmus nicht klein genug gewählt wurde. Bei einem kleineren Wert wären jedoch noch mehr Systemanalysen erforderlich, um zum Optimum zu gelangen.

Die Ergebnisse zeigen, dass der Einfluss der Position der Versteifungen unerwartet gering ausfällt. Durch Einführen einer Restriktion bezüglich der mechanischen Steifigkeit (eventuell auch als zusätzliches Optimierungsziel), könnte die Positionierung der Versteifungen verbessert werden. Diesbezüglich stellt die Ausgangskonstruktion bereits eine gute Lösung dar. Im Hinblick auf die Biegesteifigkeit fällt bei den Querschnittsabmessungen in erster Linie die Stegdicke ins Gewicht. Über die Stege muss die Biegebelastung durch Schub auf die Deckhäute übertragen werden. Da der Steg beim Optimum bezüglich der thermisch bedingten Verformung eine Dicke besitzt, die gegenüber den Versteifungen aus Stahlblech sogar zu einer höheren Schubsteifigkeit führt, kann auf eine weitere Untersuchung verzichtet werden.

7.1.4 RSA – lineare Terme und Interaktionsterme (22 Systemanalysen)

Analog der in Abschnitt 4.3 beschriebenen Verfahrensweise wird in einem ersten Schritt ein einfaches Approximationsmodell erstellt.

$$\hat{u}_{max} = b_0 + \sum_{i=1}^6 b_i x_i + \sum_{i=1}^5 \sum_{j=i+1}^6 b_{(6i+j)} x_i x_j \quad (7.2)$$

$$\text{mit } \mathbf{x} = [t_{G,\text{front}} \quad t_{G,\text{rück}} \quad t_S \quad b_G \quad p_{S1} \quad p_{S2}]^T$$

Dieses beinhaltet einen konstanten Faktor, sechs lineare Terme und 15 Interaktionsterme, in denen jeweils nur zwei Entwurfsvariablen vertreten sind. Somit sind insgesamt 22 Koeffizienten zu bestimmen.

Als Response Surface Design (RSD) käme ein 2^{6-1} Fractional Factorial Design in Frage, das die Berechnung von 32 Stützwerten erfordert. Um jedoch ausschließlich die tatsächlich benötigten Stützstellen zu verwenden, wird ein anderer Weg beschritten. Dazu wird mit den Stützstellen des 2^{6-1} -Designs (oder auch des 2^6 -Designs) und den 22 Ansatztermen eine Regressionsmatrix aufgestellt. Aus dieser können dann mit dem D -Optimalitätskriterium die 22 wichtigsten Stützstellen ermittelt werden. Nach Berechnung der zugehörigen Stützwerte können die Koeffizienten des Regressionsmodells bestimmt werden.

Eine Bewertung des Approximationsmodells ist in diesem Fall wenig sinnvoll, da die Response Surface Approximation (RSA) exakt durch alle Stützpunkte verläuft. Somit sind alle Residuen gleich null und das Bestimmtheitsmaß ergibt sich zu $R^2 = 1.0$. Die Einflüsse \mathcal{E} der Haupteffekte der Parameter $t_{F,\text{front}}$, $t_{F,\text{rück}}$, t_S und b_F sind signifikant. Bei den Entwurfsvariablen p_{S1} und p_{S2} sind diese Effekte eher schwach ausgeprägt.

Die Optimierung mit der auf diese Weise approximierten Verformung \hat{u}_{max} führt zu folgenden Ergebnissen:

Optimaler Entwurf [mm]						\hat{u}_{max}	u_{max}
$t_{F,\text{front}}$	$t_{F,\text{rück}}$	t_S	b_F	p_{S1}	p_{S2}	[mm]	[mm]
1.0	1.0	8.0	30.2	150.0	395.0	9.2	5.3

Wenn auch das exakte Optimum hierbei nicht getroffen wird, so ist dennoch eine deutliche Verbesserung des Entwurfs erreicht.

7.1.5 RSA – lineare Terme und Interaktionsterme (32 Systemanalysen)

Mit 10 weiteren Stützpunkten kann das bisher verwendete D -optimale RSD zum 2^{6-1} -Design vervollständigt werden. Mit den Stützpunkten kann mit (7.2) das selbe Appro-

ximationsmodell wie im vorigen Abschnitt angepasst werden, wobei dann auch eine Bewertung möglich ist:

SS_E	$PRESS$	R^2	R_{adj}^2	R_{pred}^2
32.076	328.46	0.9625396	0.8838728	0.6164056

Anhand der Werte ist klar zu erkennen, dass die Approximation und insbesondere deren Prognosegüte nicht besonders gut ist. Das ermittelte Optimum ist:

Optimaler Entwurf [mm]						\hat{u}_{max}	u_{max}
$t_{F,front}$	$t_{F,rück}$	t_S	b_F	p_{S1}	p_{S2}	[mm]	[mm]
1.0	1.0	8.0	30.2	210.0	335	11.3	5.1

Dabei fällt auf, dass sich die Position der Versteifungen drastisch verändert. Diese liegen in den beiden bisher ermittelten optimalen Entwürfen jeweils an entgegengesetzten Rändern ihres Definitionsbereichs. Dieses Verhalten lässt sich mit den schwachen Einflüssen \mathcal{E} der beiden Parameter p_{S1} und p_{S2} erklären. So kann es vorkommen, dass die Koeffizienten der beiden zugehörigen linearen Ansatzterme ihr Vorzeichen ändern, wenn andere (oder mehr) Stützpunkte verwendet werden. Da die Position der Versteifungen keinen Einfluss auf die Masse hat, kann die Massenrestriktion nicht greifen.

Weitere Interaktionsterme, wie zum Beispiel die 20 möglichen Drei-Faktor-Interaktionen

$$\sum_{i=1}^4 \sum_{j=i+1}^5 \sum_{k=j+1}^6 x_i x_j x_k ,$$

können nicht ohne weiteres in das Regressionsmodell aufgenommen werden, da nicht genügend Stützpunkte vorhanden sind. Hier kann die in Abschnitt 5.3 vorgestellte Methode zur Optimierung der Ansatzterme benutzt werden, um aus den 20 möglichen die 5 bis 10 wichtigsten Drei-Faktor-Interaktionen auszuwählen (siehe Abschnitt 7.1.8).

7.1.6 RSA – vollständiger quadratischer Ansatz (45 Systemanalysen)

Mit einem Central Composite Design (CCD) kommen 12 sternförmig angeordnete Stützstellen sowie ein Center Run hinzu.³ Das Approximationsmodell kann nun auch quadratische Terme enthalten. Um die bisherigen Stützpunkte verwenden zu können, kann nur das CCF verwendet werden. Das CCI würde die Berechnung von 45 neuen Stützwerten erfordern. Mit dem CCC ergäben sich physikalisch unsinnige Stützstellen wie beispielsweise negative Wandstärken.

³Auch wenn mehrere Center Runs als Stützpunkte benutzt werden, muss der zugehörige Stützwert nur einmal berechnet werden.

Mit dem vollständigen quadratischen Ansatz

$$\hat{u}_{max} = b_0 + \sum_{i=1}^6 b_i x_i + \sum_{i=1}^6 \sum_{j=i}^6 b_{(6i+j)} x_i x_j \quad (7.3)$$

$$\text{mit } \mathbf{x} = [t_{G,\text{front}} \quad t_{G,\text{rück}} \quad t_S \quad b_G \quad p_{S1} \quad p_{S2}]^T$$

ergibt sich folgendes Optimum:

Optimaler Entwurf [mm]						\hat{u}_{max}	u_{max}
$t_{F,\text{front}}$	$t_{F,\text{rück}}$	t_S	b_F	p_{S1}	p_{S2}	[mm]	[mm]
1.0	5.73	6.43	15.1	210	335	8.7	7.9

Betrachtet man den optimalen Entwurf, so liegt dieser weit vom exakten Optimum entfernt. Der nichtlinear monoton fallende oder steigende Einfluss der Entwurfsvariablen kann mit dem quadratischen Ansatz nicht wiedergegeben werden. Die verwendeten Stützpunkte „zwingen“ der Approximation eine Krümmung auf, die zu falschen Gradienten für den verwendeten Optimierungsalgorithmus führt. Abbildung 7.8 zeigt den approximierten Verlauf der maximalen Verformung u_{max} bei Variation der Entwurfsvariablen $t_{F,\text{rück}}$ und t_S . Die verbleibenden Entwurfsvariablen sind dabei auf ihre „optimalen“ Werte gesetzt.

Die Bewertung der Approximation liefert die in der nachfolgenden Tabelle zusammengefassten Werte:

SS_E	$PRESS$	R^2	R^2_{adj}	R^2_{pred}
36.45	323.5	0.9640803	0.9327955	0.6812182

Insbesondere der Wert R^2_{pred} weist hierbei auf Unzulänglichkeiten bezüglich der Prognosegüte des Approximationsmodells hin. Aus diesem Grund ist es sinnvoll, Stützstellen aus dem RSD zu entfernen. Dazu wird die in Abschnitt 4.1.1 erarbeitete Vorgehensweise benutzt. Es werden also Stützstellen entfernt, zu deren Stützwerten ein großes Residuum entsteht, wenn diese nicht zur Anpassung des Regressionsmodells verwendet werden. Als Entscheidungsgrundlage dienen demnach die in Abbildung 7.9 dargestellten $PRESS$ -Residuen beziehungsweise deren studentisierte Form, die R -student Residuen. Eine Stützstelle darf dabei jedoch nur dann entfernt werden, wenn diese einen vorgegebenen Mindestabstand zum vorläufigen Optimum nicht unterschreitet.

Mit dem reduzierten RSD und dem damit angepassten Regressionsmodell (7.3) kann folgender optimaler Entwurf ermittelt werden:

Optimaler Entwurf [mm]						\hat{u}_{max}	u_{max}
$t_{F,\text{front}}$	$t_{F,\text{rück}}$	t_S	b_F	p_{S1}	p_{S2}	[mm]	[mm]
1.0	1.2	7.6	30.8	210	335	5.3	5.1

Dieser liegt dem exakten Optimum wesentlich näher. Der Nachteil dieser Vorgehensweise ist, dass die RSA zwar im Bereich des Optimums besser ist, die maximale Verformung u_{max} jedoch in den anderen Bereichen wesentlich schlechter prognostiziert wird. Bei Änderung oder Hinzufügen von Restriktionen kann dies zu Problemen führen.

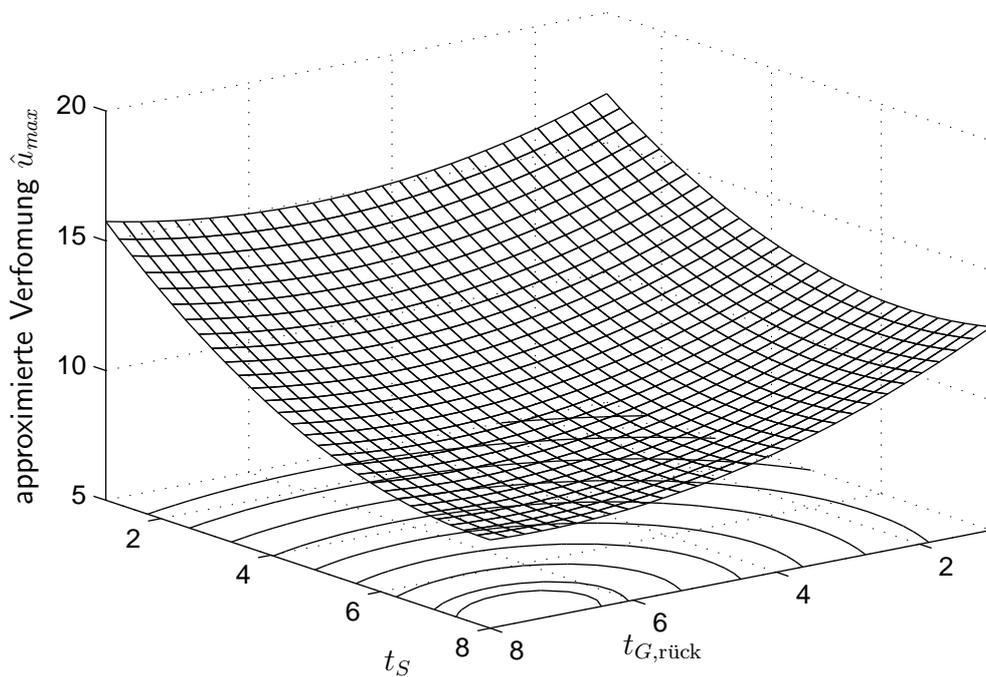


Abbildung 7.8: Approximierter Verlauf der maximalen Verformung u_{max} bei Variation der Entwurfsvariablen $t_{F,rück}$ und t_s – ($t_{F,front} = 1.0$ mm, $b_F = 15.1$ mm, $p_{S1} = 210$ mm und $p_{S2} = 335$ mm)

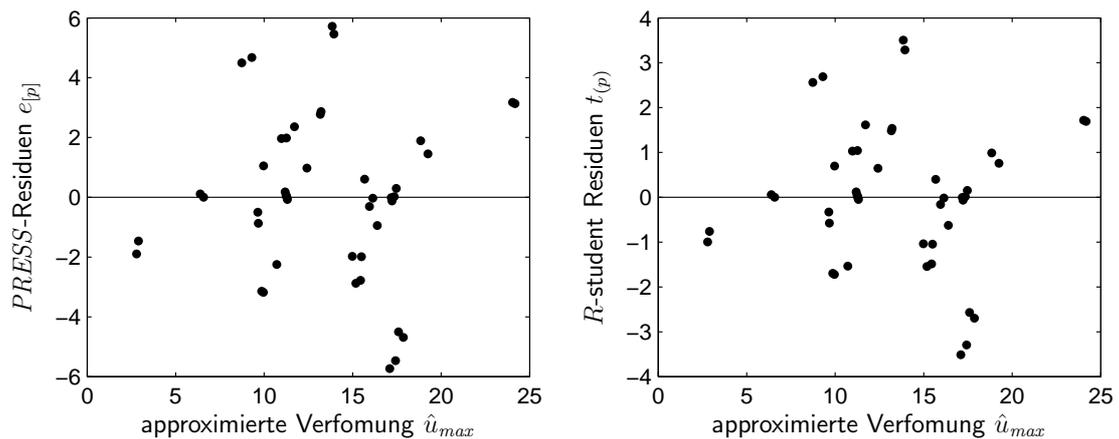


Abbildung 7.9: *PRESS*-Residuen und *R*-student Residuen für das thermisch-mechanische Optimierungsproblem „Aufzugtürblatt“ – Approximationsmodell nach (7.3)

7.1.7 RSA – vollständiger quadratischer Ansatz und Transformation der Systemantworten

Mit der in Abschnitt 5.1 beschriebenen „Power Transformation“ ist keine Verbesserung der Approximation zu erzielen. Als optimaler Wert für den Transformationsparameter ergibt sich $\lambda = 1.03$, was an der Approximation nahezu nichts ändert. Der Grund hierfür ist, dass die Entwurfsvariablen $t_{G,\text{front}}$, $t_{G,\text{rück}}$, t_S , b_G , p_{S1} und p_{S2} jeweils einen deutlich unterschiedlichen Einfluss auf die Systemantwort besitzen.

Dieser Problematik wird im folgenden Abschnitt Rechnung getragen.

7.1.8 RSA – optimal gewählte erweiterte Ansatzterme (45 Systemanalysen)

Wie bereits erwähnt, reicht der vollständige quadratische Ansatz (7.3) nicht aus, um die Zusammenhänge zwischen Entwurfsvariablen und Systemantwort ausreichend gut zu approximieren. Aus diesem Grund werden zunächst erweiterte Ansatzterme bezüglich der Haupteffekte der Entwurfsvariablen eingeführt. Dazu wird für jede Variable der quadratische Term durch eine Potenzfunktion

$$x_i^{\alpha_i} \quad \text{mit } i = 1, \dots, 6$$

$$\text{und } \mathbf{x} = [t_{G,\text{front}} \quad t_{G,\text{rück}} \quad b_G \quad p_{S1} \quad p_{S2}]^T$$

ersetzt. Somit entfällt der in Abschnitt 5.3 beschriebene äußere Teil der Ansatzoptimierung. Als Wertebereich für α_i wird das Intervall $[-2, 2]$ verwendet. Die Ansatzoptimierung führt lediglich für die zwei Entwurfsvariablen t_S und $t_{G,\text{rück}}$ zu einer Veränderung. Die Parameter α_{t_S} und $\alpha_{t_{G,\text{rück}}}$ erhalten beide einen „optimalen“ Wert nahe -1 , so dass die zugehörigen Entwurfsvariablen einen umgekehrt proportionalen Einfluss im Approximationsmodell haben. Für die verbleibenden Entwurfsvariablen bleibt mit $\alpha_i = 2$ der quadratische Ansatzterm unverändert. Insgesamt ist nur eine mäßige Verbesserung der Approximationsqualität zu verzeichnen, was die Werte in der folgenden Tabelle zeigen:

SS_E	$PRESS$	R^2	R_{adj}^2	R_{pred}^2
30.32	276.8	0.9757021	0.9443905	0.7971640

Am ermittelten Optimum ändert sich demzufolge auch nur sehr wenig. Auch exponentielle und logarithmische Ansatzterme bringen lediglich einen ähnlich mäßigen Erfolg. Für trigonometrische Terme sind nicht genügend Stützpunkte vorhanden. Da solche Zusammenhänge jedoch sicher auszuschließen sind, soll auf die Berechnung zusätzlicher Stützwerte verzichtet werden.

Die Ergebnisse der Ansatzoptimierung mit ausschließlich Haupteffekten legt die Vermutung nahe, dass die Systemantwort auch durch Interaktionsterme höherer Ordnung

maßgeblich beeinflusst wird. Da die verfügbaren Stützpunkte eine Bestimmung aller möglichen Interaktionsterme höherer Ordnung nicht gestatten (die Regressionsmatrix wird dann singulär), werden im nächsten Schritt die geeignetsten ausgewählt. Dabei werden alle möglichen Interaktionsterme dem äußeren Teil der Ansatzoptimierung zur Auswahl übergeben. Der innere Teil der Ansatzoptimierung wird hierzu umgangen, indem die Parameter $a_{(i)j}$ der zur Auswahl stehenden Ansatzterme einen festen Wert zugewiesen bekommen. Hierbei ist zu beachten, dass das CCD zur Bestimmung der Koeffizienten von Interaktionstermen lediglich ein 2^{6-1} Fractional Factorial Design verwendet. Dies hat zur Folge, dass die einzelnen Entwurfsvariablen in den Interaktionstermen nur linear Eingang finden dürfen. Deshalb werden alle Parameter $a_{(i)j}$ zu 1 gesetzt.

Zur Auswahl stehen somit 20 Drei-Faktor-Interaktionen, 15 Vier-Faktor-Interaktionen und 6 Fünf-Faktor-Interaktionen. Bei Präsenz einiger Fünf-Faktor-Interaktionen im Approximationsmodell wird die Regressionsmatrix singulär, da die benötigten Stützpunkte nicht vorhanden sind. Die eine Sechs-Faktor-Interaktion kann ohnehin nicht bestimmt werden.

Mit der Ansatzoptimierung werden nun – zusätzlich zum bisher bestimmten Approximationsmodell – weitere 12 Ansatzterme aus den genannten Interaktionstermen ausgewählt. Als Resultat erhält man ausschließlich Terme, in denen die Querschnittsabmessungen $t_{G,front}$, $t_{G,rück}$, t_S und b_G vorkommen. Die Bewertung der Approximation sieht wie folgt aus:

SS_E	$PRESS$	R^2	R_{adj}^2	R_{pred}^2
1.378	3.744	0.9999750	0.9998405	0.9987554

Als Ergebnis der Optimierung mit der ansatzoptimierten Approximation erhält man:

Optimaler Entwurf [mm]						\hat{u}_{max}	u_{max}
$t_{F,front}$	$t_{F,rück}$	t_S	b_F	p_{S1}	p_{S2}	[mm]	[mm]
1.0	1.0	6.9	39.3	210	335	4.3	3.5

Während die optimalen Querschnittsabmessungen nun sehr gut bestimmt werden, kann die optimale Position der Versteifungen immer noch nicht bestimmt werden. Der Grund hierfür ist, dass deren Einfluss gegenüber dem Einfluss der Querschnittsabmessungen zu gering ist. Diese Problematik wurde bereits bei der Optimierung mit dem SQP-Verfahren beschrieben und erklärt auch das Nicht-Vorhandensein der Positionierungsvariablen in den optimierten Ansatztermen.

7.2 Strömungs-Struktur-Kopplung: Aeroelastic Tailoring eines Tragflügels

Unter „Aeroelastic Tailoring“ versteht man das Bestreben, die Struktur eines angeströmten Tragflügels so zu gestalten, dass unter aerodynamischen Lasten eine gezielte elastische Verformung eintritt beziehungsweise unerwünschte Verformungen reduziert werden. Ziele sind dabei unter anderem das Verhindern der so genannten Divergenz⁴, die Reduktion des Widerstands (durch Beeinflussung der Auftriebsverteilung) und die Flutterprävention⁵.

Die Aeroelastik spielt insbesondere dann eine Rolle, wenn Tragflügel aufgrund ihrer Bauform und -größe nur relativ geringe Steifigkeiten aufweisen. Dies ist zum Beispiel bei großen Passagier- oder Transportflugzeugen mit Spannweiten bis zu 75 m der Fall. Auch bei Hochleistungssegelflugzeugen ist die Untersuchung der aeroelastischen Effekte wichtig. Hier kommen Flügel mit sehr hoher Streckung und gleichzeitig immer geringeren Profildicken zum Einsatz, wodurch die Steifigkeit der Tragflügelstruktur eingeschränkt wird.

Von den genannten Zielen des „Aeroelastic Tailoring“ steht bei Segelflugzeugen die Reduktion des Widerstands im Vordergrund. Neben der optimalen aerodynamischen Gestaltung der Flugzeugzelle, der Tragflügelgeometrie und -profile spielt dabei auch die Verformung im aeroelastischen Gleichgewichtszustand eine große Rolle.

In [Plengorth 2000] wurde das Prinzip des „Aeroelastic Tailoring“ an einem einfachen Balkenmodell sowie an einem FE-Modell des Tragflügels eines Hochleistungssegelflugzeugs untersucht. Die dort detailliert beschriebene Konstruktion des Tragflügels und in Grundzügen auch das dabei verwendete FE-Modell bilden die Basis für die im Folgenden dargestellten Untersuchungen. Abbildung 7.10 zeigt eine Prinzipskizze des Tragflügels.

7.2.1 Problemstellung

Um den Widerstand des Tragflügels zu minimieren soll die Struktur des Tragflügels so verändert werden, dass die durch die aerodynamischen Lasten bedingte Verwindung möglichst klein wird.

⁴Unter Divergenz versteht man das statische Auskippen des elastischen Tragflügels.

⁵Mit Flattern ist der Resonanzfall gemeint, bei dem sich die Schwingungsamplitude eines Tragflügels immer weiter erhöht, indem die schwingende Tragflügelstruktur immer mehr Energie von der umströmenden Luft aufnimmt.

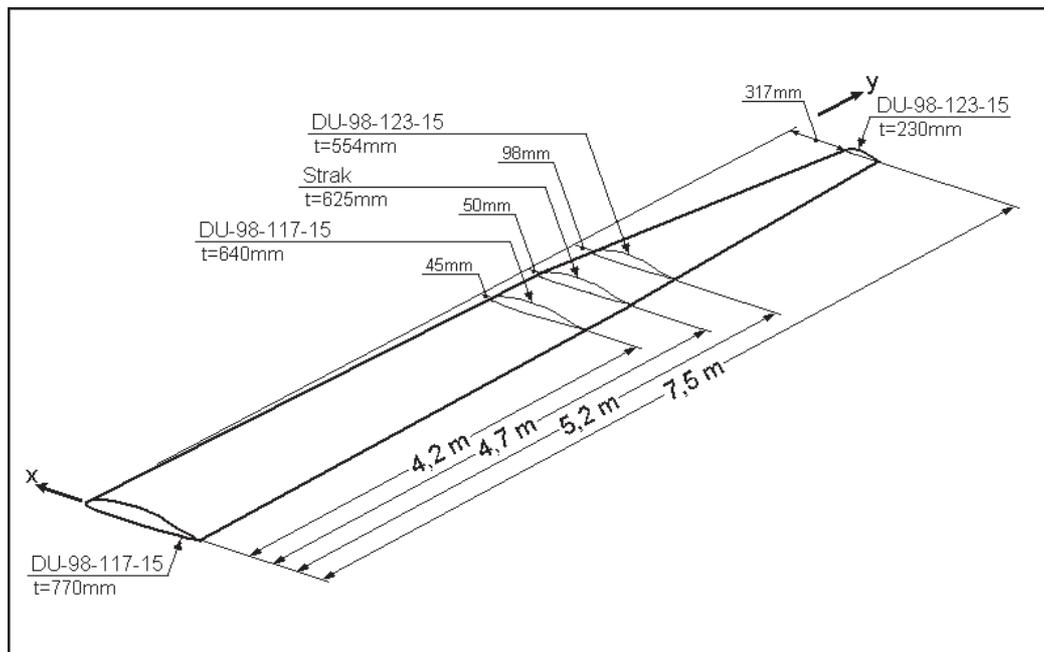


Abbildung 7.10: Prinzipskizze des untersuchten Segelflugzeugtragflügels

Zur Beeinflussung des Verformungsverhaltens werden folgende Maßnahmen gewählt:

- Ein übliches Werkzeug des „Aeroelastic Tailoring“ ist die Ausnutzung gerichteter Steifigkeiten. Dies kann bei großen Tragflügelstrukturen durch das Strukturkonzept selbst [Piening 1998] oder allgemein durch die anisotropen Eigenschaften von Faserverbundwerkstoffen erfolgen [Kaiser 1998]. Da der Tragflügel ohnehin in Faserverbundbauweise gefertigt wird, soll über die Zug-Schub-Kopplung einer Schicht unidirektional (UD) angeordneter Fasern eine Biege-Torsions-Kopplung beim Tragflügel erzeugt werden. In Abbildung 7.11 ist die Wirkungsweise der Zug-Schub-Kopplung bei einer UD-Schicht veranschaulicht, deren Fasern in einem Winkel zur Lastrichtung verlaufen. Bei einer durch ein positives Lastvielfaches hervorgerufenen Biegeverformung erfährt die obere UD-Schicht eine Druckbelastung, die untere wird auf Zug beansprucht. Durch die Zug-Schub-Kopplung auf der Tragflügeloberseite und -unterseite ist somit die Biege-Torsions-Kopplung gegeben.

Um die Biege-Torsions-Kopplung in unterschiedlichen Bereichen des Tragflügels individuell beeinflussen zu können, sollen die UD-Schichten in 5 Bereichen einzeln, unter einem definierten Winkel zur Flügellängsachse ausgerichtet werden (siehe Abbildung 7.12).

- Als weitere Maßnahme wird der normalerweise gerade, senkrecht zur Flugzeuglängsachse verlaufende Holm in Lage und Verlauf verändert. Dadurch wird

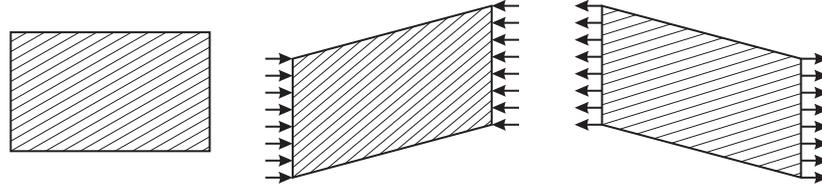


Abbildung 7.11: Zug-Schub-Kopplung in einer UD-Schicht

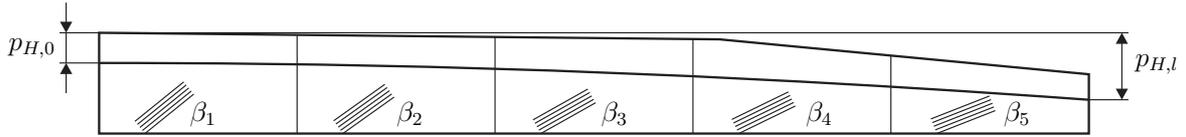


Abbildung 7.12: Entwurfsvariablen für das „Aeroelastic Tailoring“ des Tragflügels

die Lage des Schubmittelpunkts beeinflusst. Da die Angriffspunkte der resultierenden aerodynamischen Kräfte in etwa gleich bleiben, ändert sich der „Hebelarm“ zwischen diesen und dem Schubmittelpunkt. Das Biege-Torsions-Verhalten kann somit durch Effekte der schiefen Biegung beeinflusst werden.

Der veränderbare Holm besitzt einen parabelförmigen Verlauf. Die Parameter dieser Parabel werden so bestimmt, dass diese an der Flügelwurzel senkrecht in den Rumpf mündet und dort, wie auch an der Flügelspitze, durch einen definierten Punkt läuft (siehe Abbildung 7.12). Der Definitionsbereich dieser Punkte wird so gewählt, dass die Holmhöhe nicht unter $\sim 90\%$ gegenüber der bereits ausgelegten Tragflügelstruktur sinkt.

Somit kann die Optimierungsaufgabe formuliert werden:

$$\text{Minimiere } \Delta\alpha_{ges}(\mathbf{u}_{strukt}) = \sum_{i=1}^{n_E} (\Delta\alpha_i(\mathbf{u}_{strukt}))^2 \quad (7.4a)$$

$$\text{mit } \mathbf{u}_{strukt} \text{ aus } \mathbf{f}_{strukt}(\mathbf{x}_{strukt}, \mathbf{u}_{strukt}, \mathbf{p}_{aero}) = \mathbf{0}$$

$$\text{bzw. } \mathbf{f}_{strukt}(p_{H,0}, p_{H,l}, \beta_1, \beta_2, \beta_3, \beta_4, \beta_5, \mathbf{u}_{strukt}, \mathbf{p}_{aero}) = \mathbf{0} \quad (7.4b)$$

$$\text{und } \mathbf{f}_{aero}(\mathbf{u}_{strukt}, \mathbf{p}_{aero}) = \mathbf{0} \quad (7.4c)$$

$$\text{so, dass } \begin{cases} 200 \text{ mm} \leq p_{H,0} \leq 500 \text{ mm} \\ 350 \text{ mm} \leq p_{H,l} \leq 470 \text{ mm} \\ -45^\circ \leq \beta_i \leq 45^\circ \end{cases} \quad \text{mit } i = 1, \dots, 5 \quad (7.4d)$$

In (7.4a) entspricht n_E der Anzahl der Elemente entlang der Flügellängsachse. Über die Verschiebungen der Knoten an der Flügelvorderkante und -hinterkante können die einzelnen Torsionswinkel $\Delta\alpha_i$ bestimmt werden.

Da die Forderung nach einer möglichst kleinen Verwindung des Tragflügels auch für verschiedene, typische Flugzustände gelten soll, gibt es für (7.4) genau genommen mehrere Ziele. Als betrachtete Flugzustände sind dabei die zwei Auslegungslastfälle „Kreisflug“ (LF1) und „schneller Geradeausflug“ (LF3) zu nennen. Als weiterer, gerade im Hinblick auf den Widerstand wichtiger Flugzustand wird noch der „Geradeausflug mit der Geschwindigkeit des besten Gleitens“ (LF2) betrachtet. Bei allen genannten Flugzuständen hat die Wölbklappe jeweils einen anderen Ausschlag. Dies verändert die Auftriebsverteilung und somit die aerodynamischen Lasten. Es stellt sich also immer ein anderer aeroelastischer Gleichgewichtszustand ein.

Im folgenden Abschnitt wird das verwendete Simulationsmodell und die dafür erforderliche Realisierung der Strömungs-Struktur-Kopplung kurz vorgestellt.

7.2.2 Beschreibung des Simulationsmodells

Abbildung 7.13 zeigt das FE-Modell des Tragflügels, an dessen Knoten bereits die aus der Druckverteilung resultierenden Kräfte aufgebracht sind. Die Druckverteilung wird dabei über die Profiltheorie bestimmt. Das heißt, bei der Berechnung der Druckverteilung wird von einem Flügel mit unendlicher Streckung ausgegangen. Veränderungen der Strömung im Randbogenbereich und am Flügel-Rumpf-Übergang werden nicht berücksichtigt.

Zur Berechnung des Druckbeiwerts C_p wird das Profilberechnungsprogramm XFOIL [Drela 2001] verwendet. Dieses liefert die Werte für $C_p(x_{rel}, \alpha)$ in Abhängigkeit von der relativen Profiltiefe x_{rel} und dem Anstellwinkel α . Da XFOIL keine Schnittstelle besitzt, kann der Verlauf der Druckbeiwerte nicht direkt aus dem übergeordneten Strukturberechnungsprogramm angefordert werden. Deshalb sind die Druckbeiwerte für diskrete relative Profiltiefen und Anstellwinkel in einer Datenbasis abgelegt. Abbildung 7.14 zeigt den Verlauf des C_p -Wertes (Flügeloberseite) über die Profiltiefe und einem Anstellwinkelbereich von -2° bis 5° . Eine solche Datenbasis existiert für alle verwendeten Profilgeometrien (siehe Abbildung 7.10) und alle zu den Flugzuständen gehörenden Wölbklappenausschlägen: $\delta_{LF1} = +20^\circ$, $\delta_{LF2} = 0^\circ$ und $\delta_{LF3} = -3^\circ$.

Der iterative Ablauf zur Berechnung des aeroelastischen Gleichgewichtszustands ist in Abbildung 7.15 dargestellt und wird im folgenden beschrieben:

- Zu Beginn jedes Iterationsschritts wird der unverformte Tragflügel mit einem vorgegebenen Winkel angestellt.
 - Die sich dabei ergebende Druckverteilung wird dann auf die Elemente des Strukturmodells verteilt. Dies erfolgt durch Interpolation der Werte aus den Datenbasen und unter Berücksichtigung der Größe der Elemente.
 - Damit ein Überschwingen vermieden wird, erfolgt das Aufbringen der Druckverteilung im ersten Schritt nicht vollständig.

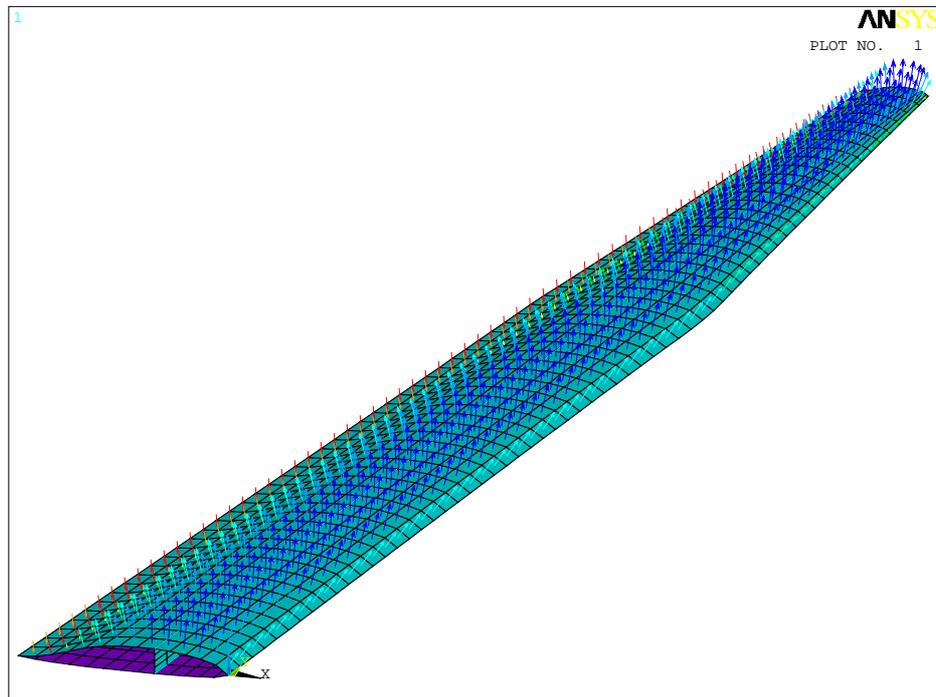


Abbildung 7.13: Druckverteilung auf das FE-Modell des Tragflügels

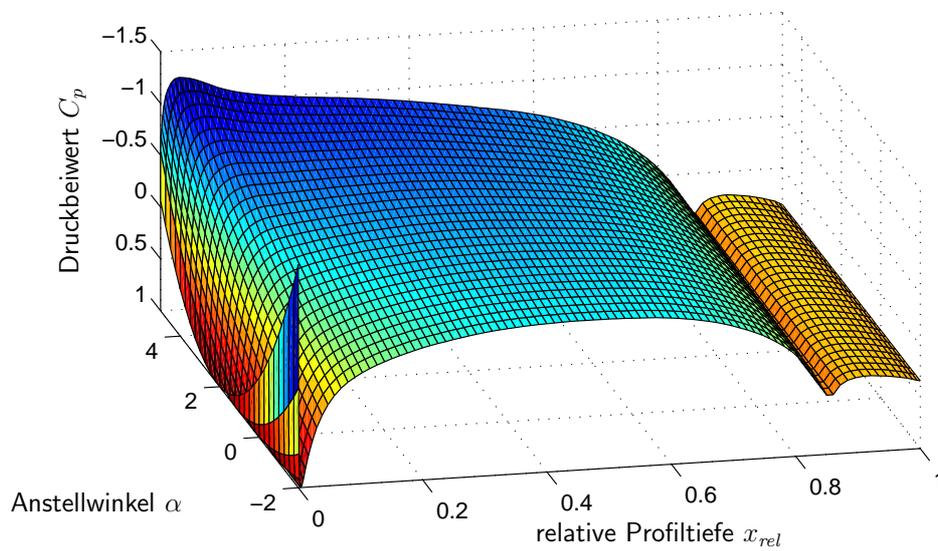


Abbildung 7.14: Verlauf des Druckbeiwerts C_p auf der Flügeloberseite bei einem Wölbklappenausschlag von 0° (LF2) und einem Anstellwinkelbereich von -2° bis 5°

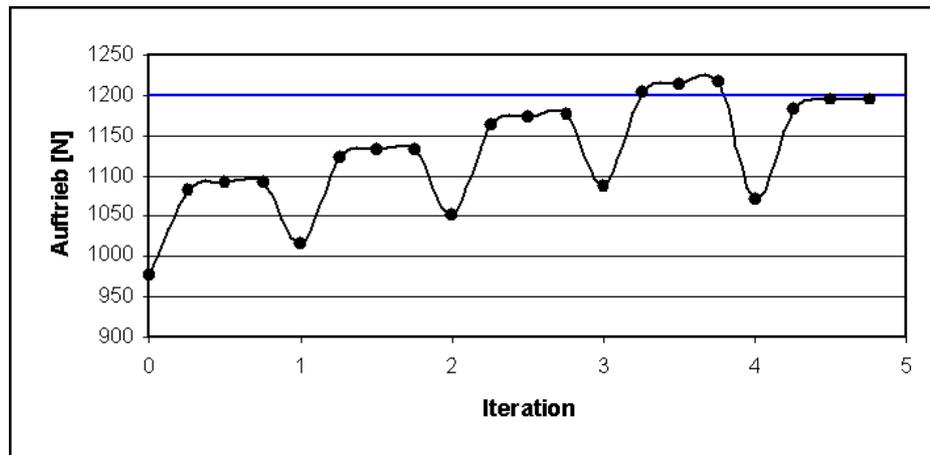


Abbildung 7.15: Ermittlung eines aeroelastischen Gleichgewichtszustands mit vorgegebener Auftriebskraft

- Mit der gegebenen Druckverteilung ergibt sich die Verformung der Struktur. Damit kann in den Abständen der Knoten (entlang der Flügellängsachse) der Torsionswinkel bestimmt werden.
- Für den verformten Tragflügel wird eine neue Druckverteilung ermittelt und wieder auf das Strukturmodell aufgebracht ...
- Ändern sich die Torsionswinkel nur noch geringfügig, so ist ein aeroelastischer Gleichgewichtszustand erreicht.
- Je nachdem, ob die erreichte Auftriebskraft unter oder über dem gewünschten Wert liegt, wird der Anstellwinkel an der Flügelwurzel vergrößert oder reduziert. Entspricht die Auftriebskraft im Gleichgewichtszustand dem geforderten Wert, so ist die Lösung konvergiert.

7.2.3 Referenzlösungen zur Beurteilung der Optimierungsergebnisse

Flügelverwindung der Ausgangskonstruktion

Bei der Ausgangskonstruktion wird ein gerader Holm verwendet. Die Position des Holms an der Flügelwurzel $p_{H,0}$ wie auch an der Flügelspitze $p_{H,l}$ beträgt 370 mm. Die UD-Schichten sind unter 0° zur Flügellängsachse ausgerichtet. Damit ergeben sich die in Abbildung 7.16 dargestellten Flügelverwindungen für die Lastfälle 1 bis 3.

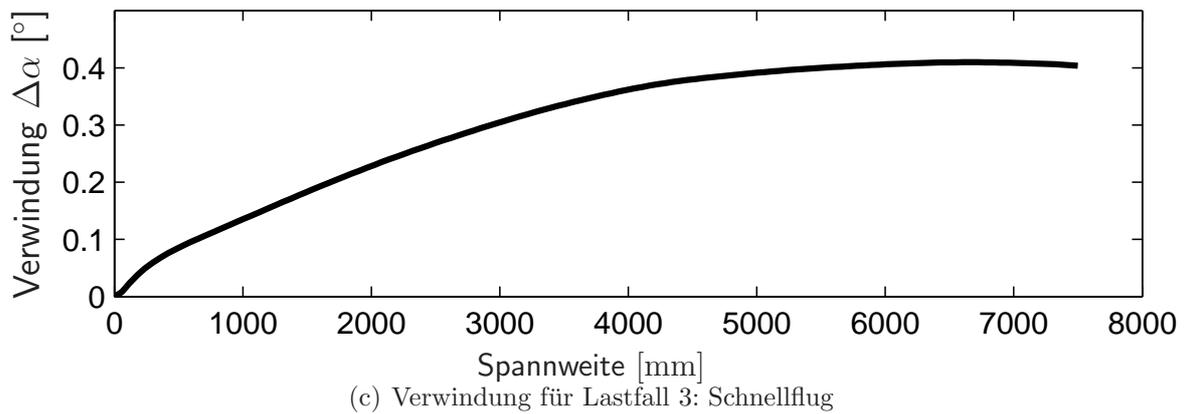
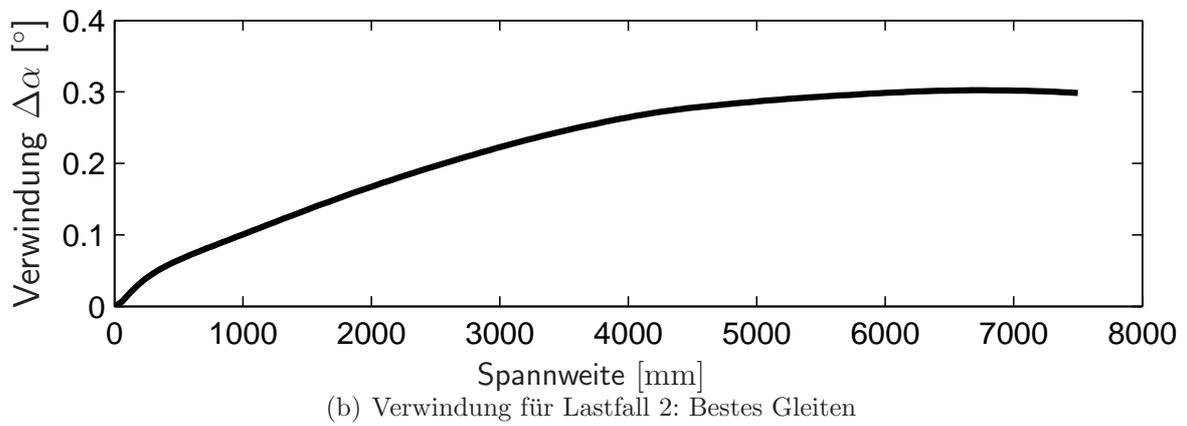
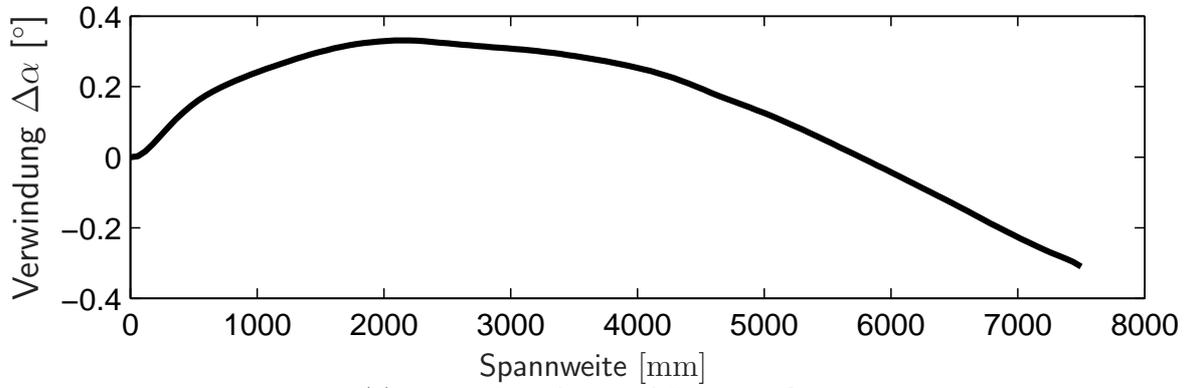


Abbildung 7.16: Flügelverwindung der Ausgangskonstruktion für die Lastfälle 1 bis 3

Die daraus resultierenden Gesamtverwindungen $\Delta\alpha_{ges}$ sind in nachfolgender Tabelle zusammengefasst:

	Kreisflug (LF1)	Bestes Gleiten (LF2)	Schnellflug (LF3)
$\Delta\alpha_{ges}$	6.7	7.1	13.2

Optimierung mit einem SQP-Verfahren unter Verwendung der exakten Lösung der Systemgleichungen

Um die mit RSAs ermittelten optimalen Entwürfe beurteilen zu können, werden zunächst die mit einem SQP-Verfahren berechneten Optima dargestellt. Die Lösungen der Optimierungsaufgabe (7.4) für die drei Lastfälle sind:

Lastfall		$p_{H,0}$	$p_{H,l}$	β_1	β_2	β_3	β_4	β_5	$\Delta\alpha_{ges}$	Analysen
LF1	Startwerte	350	410	0	0	0	0	0		
	Optimum	344	417	13	7	2	1	0	0.20	168
LF2	Startwerte	350	410	0	0	0	0	0		
	Optimum	200	350	34	29	29	27	3	0.66	256
LF3	Startwerte	350	410	0	0	0	0	0		
	Optimum	200	350	36	30	29	28	4	2.4	224

Mit bis zu 256 Lösungen für die exakten Systemgleichungen ergibt sich bei einer durchschnittlichen Berechnungsdauer von knapp 30 Minuten eine sequenzielle Bearbeitungszeit von 5 Tagen. Diese kann durch eine Grobparallelisierung der für die Gradientenberechnung erforderlichen Systemanalysen auf 16 Stunden reduziert werden.

Die ermittelten Optima lassen erkennen, dass die Lösungen für die Lastfälle 2 und 3 sehr ähnlich sind. In beiden Fällen wird der Holm sehr weit vorne platziert. Die durch den negativen Wölbklappenausschlag beim Lastfall 3 veränderten aerodynamischen Kräfte erfordern gegenüber dem Lastfall 2 lediglich eine geringfügig größere Kompensation durch die Schichtwinkel.

Für den Lastfall 1 ergibt sich ein optimales „Aeroelastic Tailoring“ mit einer gänzlich anderen Form und Lage des Holms. Bei der Ausrichtung der UD-Schichten würde man aufgrund der erforderlichen Biege-Torsions-Kopplung negative Winkel erwarten. Die erzielbare Verbesserung spricht jedoch gegen diese Annahme. Der Grund dafür ist, dass Form and Lage des Holms einen dominanten Einfluss auf das Biege-Torsions-Verhalten des Tragflügels besitzen. Die Biege-Torsions-Kopplung über die Ausrichtung der UD-Schichten dient hier der bereichsweisen Verbesserung des durch den Holm bestimmten Verformungsverhaltens.

7.2.4 RSA für Lastfall 1 – quadratischer Ansatz (79 Systemanalysen)

Die Kenntnis physikalischer Zusammenhänge aus der Laminattheorie legt die Vermutung nahe, dass der Einfluss der Ausrichtung der UD-Schichten stark nichtlinear ist. Aus diesem Grund wird von vornherein ein quadratischer Ansatz verwendet:

$$\widehat{\Delta\alpha}_{ges} = b_0 + \sum_{i=1}^7 b_i x_i + \sum_{i=1}^7 \sum_{j=i}^7 b_{(6i+j)} x_i x_j \quad (7.5)$$

$$\text{mit } \mathbf{x} = [p_{H,0} \ p_{H,l} \ \beta_1 \ \beta_2 \ \beta_3 \ \beta_4 \ \beta_5]^T$$

Als RSD wird das CCC benutzt. Somit müssen 79 Stützwerke berechnet werden. Mit diesen lassen sich die Koeffizienten des Ansatzes (7.5) bestimmen und man erhält folgende Bewertung für die Approximation:

SS_E	$PRESS$	R^2	R_{adj}^2	R_{pred}^2
923.8524367	1322.4879879	0.8858747	0.8525882	0.8366305

Die Werte zeugen von einer eher mäßigen Approximation, was sich auch am ermittelten Optimum zeigt:

	$p_{H,0}$	$p_{H,l}$	β_1	β_2	β_3	β_4	β_5	$\widehat{\Delta\alpha}_{ges}$	$\Delta\alpha_{ges}$
Optimum	217	470	45	-45	45	45	45	8.8	3.4

Die Position der Holmendpunkte und damit auch die Form des Holms werden im Vergleich zum exakten Optimum relativ gut bestimmt. Anders ist dies bei den Winkeln der UD-Schichten. Hier besitzt einer (β_2) eine „falsche“ Tendenz.

7.2.5 RSAs für die Lastfälle 2 und 3 – quadratische Ansätze (je 79 Systemanalysen)

Auch für den Lastfall 2 wird der Ansatz (7.5) und das selbe RSD verwendet. Die Werte der globalen Beurteilungskriterien sprechen für eine gute Approximation:

SS_E	$PRESS$	R^2	R_{adj}^2	R_{pred}^2
312.9238311	467.3347867	0.9708304	0.9624621	0.9564368

Das ermittelte Optimum

	$p_{H,0}$	$p_{H,l}$	β_1	β_2	β_3	β_4	β_5	$\widehat{\Delta\alpha}_{ges}$	$\Delta\alpha_{ges}$
Optimum	200	350	45	45	45	45	15.9	7.6	4.3

zeugt ebenfalls dafür, dass die globalen Tendenzen/Einflüsse ausreichend gut durch die Approximation beschrieben werden.

Beim ähnlich gelagerten Lastfall 3 wird bis auf den Schichtwinkel β_5 der gleiche optimale Entwurf ermittelt. Der Winkel liegt bei der oberen Grenze seines Definitionsbereichs.

7.2.6 RSA – erweiterte Ansatzterme (89 Systemanalysen)

Die bisherigen Approximationsmodelle – insbesondere das für den Lastfall 1 (Kreisflug) – weisen noch Schwächen beim Beschreiben des Verlaufs der Zielfunktion auf. Bessere Approximationen für die einzelnen Lastfälle könnten nun für einen kleineren Definitionsbereich (z. B. $\beta_i \in [0, 45]$) erstellt werden.

Hier soll jedoch der Weg über die erweiterten Ansatzterme und der optimalen Zusammensetzung zu einem guten Approximationsmodell gegangen werden. Da sich die Zug-Schub-Kopplung und damit die Biege-Torsions-Kopplung mit der Ausrichtung der UD-Schichten ändert, liegt die Verwendung trigonometrischer Terme nahe. Für die Ansatzoptimierung werden zur Beschreibung der nichtlinearen Haupteffekte der Schichtwinkel Sinus- und Cosinus-Terme zur Auswahl gestellt. Hierbei müssen gleichzeitig die zugehörigen Parameter $a_{(i)j}$ variiert werden, was bereits im Vorfeld einen hohen Aufwand für die Lösung des diskret-nichtlinearen Regressionsproblems erwarten lässt. Hier muss für jede vom Genetischen Algorithmus gewählte Kombination von Ansatztermen, der Parametervektor $\mathbf{a}_{(i)}$ jedes Sinus- oder Cosinus-Terms durch das innere Lösungsverfahren bestimmt werden.

Durch den Einsatz der trigonometrischen Terme werden zusätzliche Stützpunkte benötigt. Dazu werden die sternförmig angeordneten Stützstellen des CCC Designs mit dem Faktor $1/2$ skaliert. Da dies nur für die 5 Schichtwinkel erfolgen muss, sind 10 weitere Stützwerte und somit Systemantworten zu berechnen.

Als Ansatzterme stehen

- 5 Sinus-Terme
- 5 Cosinus-Terme
- 7 lineare Terme
- 7 quadratische Terme
- 21 Zwei-Faktor-Interaktionen
- 35 Drei-Faktor-Interaktionen
- 35 Vier-Faktor-Interaktionen
- 21 Fünf-Faktor-Interaktionen
- 7 Sechs-Faktor-Interaktionen

zur Auswahl. Davon werden für die Schichtwinkel ausschließlich Sinus-Terme in das resultierende Approximationsmodell aufgenommen. Lineare Terme sind für alle Entwurfsvariablen enthalten. Quadratische Terme finden sich nur noch für die Positionen der Holmendenpunkte ($p_{H,0}$ und $p_{H,l}$). Diese sind sehr stark in den Zwei- und Mehrfaktor-Interaktionen vertreten. Von den Schichtwinkeln tritt β_1 am häufigsten auf, die weiter außen liegenden immer seltener.

Die Bewertungen der Approximationen für die einzelnen Lastfälle sehen wie folgt aus:

	SS_E	$PRESS$	R^2	R_{adj}^2	R_{pred}^2
RSA – LF1	321.3568197	510.4936994	0.9845781	0.9721478	0.9668695
RSA – LF2	201.9624621	297.2246771	0.9942475	0.9882676	0.9795014
RSA – LF3	219.3472819	306.9335670	0.9901493	0.9834587	0.9729874

Die Werte zeigen, dass mit den zur Auswahl gestellten erweiterten Ansatztermen noch keine perfekten Approximationen erstellt werden können. Anhand der ermittelbaren Optima ist jedoch zu erkennen, dass der Verlauf der Zielfunktionen besser beschrieben wird.

	Optimum							$\widehat{\Delta\alpha}_{ges}$	$\Delta\alpha_{ges}$
	$p_{H,0}$	$p_{H,l}$	β_1	β_2	β_3	β_4	β_5		
RSA – LF1	312	443	23	18	15	12	11	1.2	1.4
RSA – LF2	200	350	38	33	34	30	10	2.3	2.4
RSA – LF3	200	350	39	34	34	31	14	3.5	4.3

Für die einflussreicheren Entwurfsvariablen können demzufolge bessere optimale Werte ermittelt werden, auch wenn diese immer noch nicht dem exakten Optimum entsprechen. Bei Entwurfsvariablen mit einem, im Verhältnis gesehen schwachen Einfluss auf den Zielfunktionswert liegen die optimalen Werte weit vom exakten Optimum entfernt. Dies führt jedoch nicht dazu, dass insgesamt eine beachtliche Verbesserung bezüglich der Flügelverwindung erzielt wird.

Der Aufwand für die Erstellung dieser Approximationen ist allerdings beträchtlich. Hierzu wird jeweils mehr als ein Tag benötigt, was im Vergleich zu den 16 Stunden, die ein Gradientenverfahren mit grobparallelisierten Systemanalysen benötigt, als nicht gerechtfertigt anzusehen ist.

Dies relativiert sich jedoch, wenn zum Beispiel die Lösung der genannten Mehrziel-aufgabe angegangen werden soll. Mit guten Approximationen kann eine Pareto-Front sehr schnell erzeugt werden. Diese ist zwar nicht exakt, kann aber als wesentliche Hilfe bei der Auswahl eines optimalen Kompromisses dienen.

7.3 Behandlung von stark variierenden Systemantworten

In diesem Abschnitt wird abschließend ein Optimierungsproblem behandelt, dessen Zielfunktionswert stark variiert. Solche Probleme treten zum Beispiel bei der Gestaltoptimierung auf, wenn sich numerische Netze ändern, wie es auch bei der Optimierung der Querschnittsform einer schnell rotierenden Chopperscheibe aus kohlefaserverstärktem Kunststoff (CFK) der Fall war (siehe [Friedrich 2003]).

7.3.1 Problemstellung

Chopperscheiben dienen zum Filtern eines Neutronenstrahls, so dass lediglich Neutronen mit einer bestimmten Energie im Strahl verbleiben. Hierzu sind mehrere rotierende Scheiben mit Schlitz (Fenstern) hintereinander angeordnet, deren Phasenverschiebung so abgestimmt ist, dass nur Neutronen mit einer bestimmten Geschwindigkeit passieren können. Um ein möglichst scharf begrenztes Energiespektrum zu erhalten werden Drehzahlen von bis zu 30000 1/min gefordert. Dies führt zu enormen Fliehkräften, die gerade im Fenstergrund hohe Beanspruchungen hervorrufen.

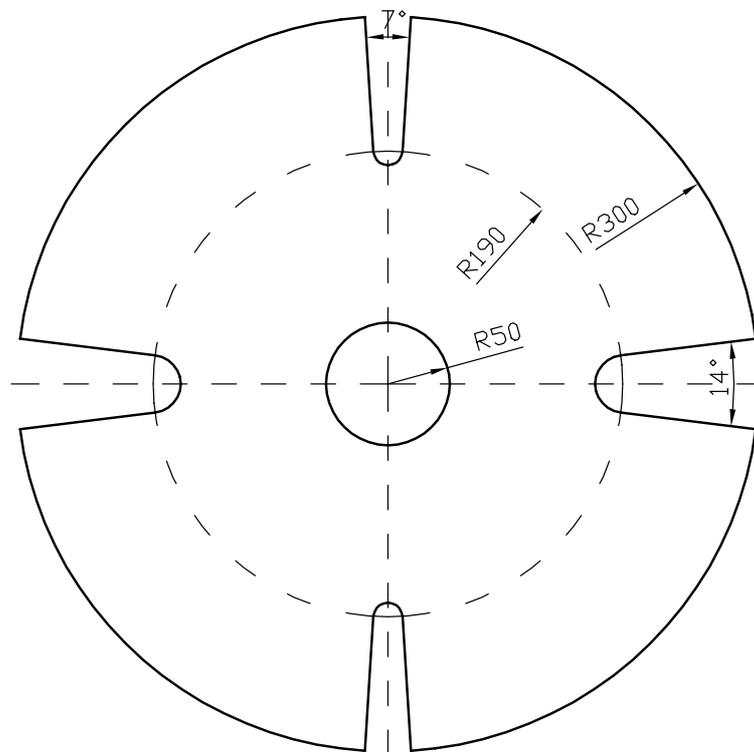


Abbildung 7.17: Abmessungen der Chopperscheibe

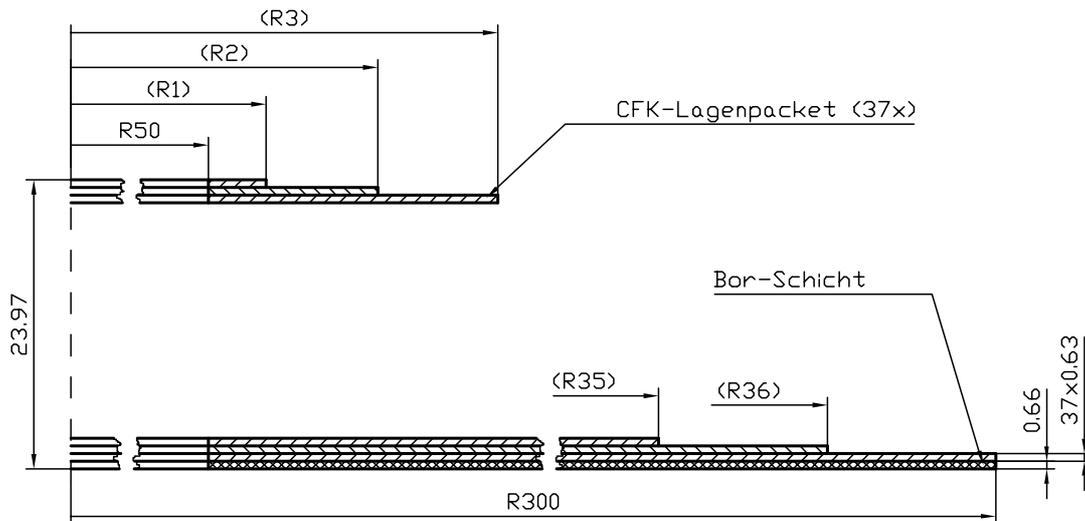


Abbildung 7.18: Querschnittsaufbau der Chopperscheibe über die Radien der einzelnen CFK-Lagen

Abbildung 7.17 zeigt die Abmessungen einer solchen Scheibe. Der Aufbau des Querschnitts über die einzelnen CFK-Lagen ist in Abbildung 7.18 dargestellt. Durch die Abstufung der Radien dieser Lagen soll die Form des Querschnitts so variiert werden, dass die mit einem Versagenskriterium quantifizierbare Beanspruchung minimiert wird.

Eine Verwendung der Radien zur Beschreibung der Querschnittsform hätte ein Optimierungsproblem mit 36 Entwurfsvariablen zur Folge. Zudem müssten im Hinblick auf kontinuierliche Abstufung der Lagen zahlreiche Restriktionen eingeführt werden, die dafür Sorge tragen, dass ein Radius R_i immer kleiner ist als der Radius R_{i+1} . Zur Reduktion der Entwurfsvariablen wird die Querschnittsform durch einen B-Spline beschrieben (siehe Abbildung 7.19). Die Describer des B-Splines können in der dargestellten Weise verschoben werden und bestimmen somit die Abstufung der einzelnen Lagen. Die Optimierungsaufgabe lautet:

$$\text{Minimiere} \quad \max(\mathcal{F}_{\text{Tasai/Wu}}(\mathbf{x})) = \max(\mathcal{F}_{\text{Tasai/Wu}}(R_1, R_2, z_2, R_3, z_3, R_4)) \quad (7.6a)$$

$$\text{so, dass} \quad \left\{ \begin{array}{l} 60 \text{ mm} \leq R_1 \leq 100 \text{ mm} \\ R_1 + 5 \text{ mm} \leq R_2 \leq 165 \text{ mm} \\ 5 \text{ mm} \leq z_2 \leq 18 \text{ mm} \\ 190 \text{ mm} \leq R_3 \leq 230 \text{ mm} \\ 3 \text{ mm} \leq z_3 \leq z_2 \\ z_3 \leq 12 \text{ mm} \\ R_3 + 5 \text{ mm} \leq R_4 \leq 250 \text{ mm} \end{array} \right. \quad (7.6b)$$

Hierbei sind die R_i und z_i die veränderbaren Koordinaten der Describer. Die Restriktionen sorgen für einen monotonen Verlauf, der aus vernetzungstechnischen Gründen auch nicht zu steil sein darf.

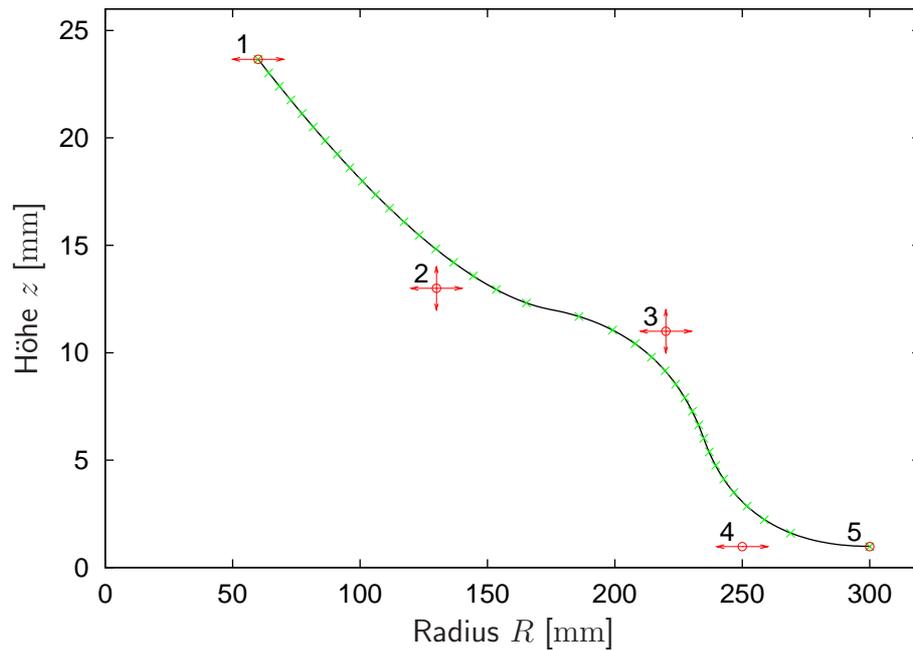


Abbildung 7.19: Parametrisierte Beschreibung der Querschnittsform der Chopperscheibe über einen B-Spline

7.3.2 Lösung der Optimierungsaufgabe mit einem mathematischen Optimierungsverfahren

Bei der Lösung der Optimierungsaufgabe (7.6) mit einem gradientenbasierten Optimierungsverfahren zeigen sich enorme Konvergenzprobleme. Der Grund hierfür liegt in der teilweise sprunghaften Veränderung des Zielfunktionswerts $\mathcal{F}_{\text{Tsaï/Wu}}$. Dies wiederum liegt an der gestuften Realisierung des Querschnittsverlaufs im FE-Modell. So entscheidet in einigen Fällen eine kleine Änderung einer Entwurfsvariable darüber, ob sich ein Radius und damit eine CFK-Lage in den Fenstergrund erstreckt oder nicht. Dies führt zu einem deutlich veränderten FE-Netz und in diesem Fall auch zu stark variierenden Modellantworten. Werden die Gradienten durch Finite Differenzen bestimmt, so sind diese stark verfälscht und erlauben keine zielgerichtete Suche nach einem Optimum.

Das beschriebene Verhalten zeigt sich anhand einer Parameterstudie mit 10000 Entwürfen. In Abbildung 7.20 ist der Verlauf der Zielfunktion (7.6a) bei einer Variation der Describerkoordinaten z_2 und z_3 dargestellt. Der Verlauf ist sehr unruhig und weist viele lokale Minima auf. Eine globale Tendenz, die das Optimierungsverfahren allerdings nicht kennt, ist jedoch deutlich zu erkennen.

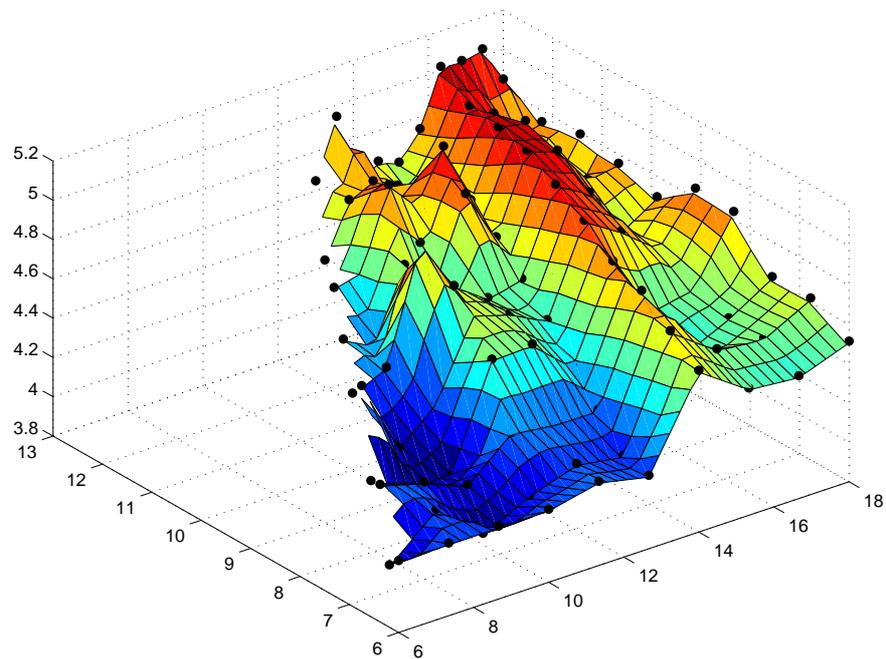


Abbildung 7.20: Verlauf des Zielfunktionswerts $\mathcal{F}_{\text{Tsai/Wu}}$ bei Variation der Deskriberkoordinaten z_2 und z_3

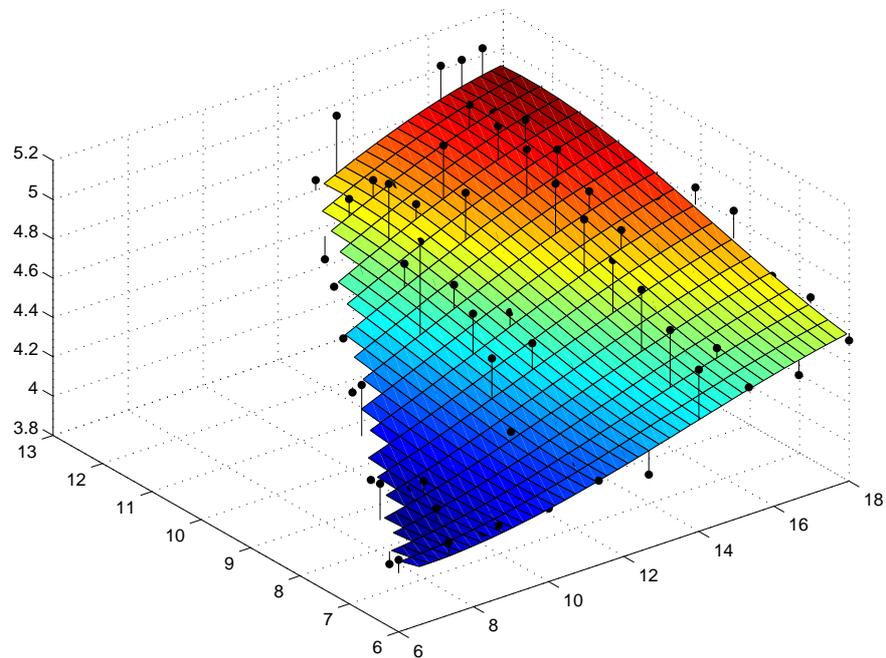


Abbildung 7.21: Verlauf des approximierten Zielfunktionswerts $\hat{\mathcal{F}}_{\text{Tsai/Wu}}$ bei Variation der Deskriberkoordinaten z_2 und z_3

7.3.3 Glättung und Optimierung mit Hilfe einer Response Surface Approximation

Werden die bei der Parameterstudie gewonnenen Daten zur Anpassung eines Approximationsmodells benutzt, so werden die stark variierenden Systemantworten ausgemittelt. Damit entsteht ein glatter Verlauf für den approximierten Beanspruchungswert $\widehat{\mathcal{F}}_{\text{Tsai/Wu}}$, für den eine Optimierung problemlos durchführbar ist. Als Optimum ergibt sich:

R_1	R_2	z_2	R_3	z_3	R_4	$\widehat{\mathcal{F}}_{\text{Tsai/Wu}}$
100 mm	130 mm	7 mm	190 mm	7 mm	195 mm	4.0

Eine ähnlich gute Approximation lässt sich auch mit deutlich weniger Stützpunkten erstellen. Den Ausgangspunkt dazu bildet ein CCF Design (45 Stützpunkte), das dann parallel zu den Achsen des Entwurfsvariablenraums um einen bestimmten Betrag verschoben und kopiert wird. Mit den 6 Entwurfsvariablen ergeben sich somit 315 Stützpunkte. Die damit erstellte RSA führt zu einem fast identischen Optimum.

8 Zusammenfassung

In vielen technischen Bereichen übernehmen Leichtbaustrukturen wichtige Funktionen als Teil eines komplexen Systems. Beispiele solcher Systeme sind in der Fahrzeugtechnik, in Luft- und Raumfahrtssystemen und zahlreichen Sonderkonstruktionen zu finden. Hier sind es neben der Strukturmechanik auch andere technisch-physikalische Disziplinen, die über die Qualität und Leistungsfähigkeit des Gesamtsystems entscheiden. Diese zu verbessern, ist Aufgabe der multidisziplinären Strukturoptimierung. Zur Reduktion des dabei erforderlichen Optimierungsaufwands konzentriert sich die Arbeit auf zwei Ansätze:

- Anwendung von auf der Response Surface Methode (RSM) basierenden Approximationen als Ersatz für aufwendige Systemanalysen
- Grobparallelisierung von erforderlichen exakten Systemanalysen auf einem PC-Cluster

Der Aufwand zur Lösung multidisziplinärer Optimierungsprobleme wird primär durch den hohen Berechnungsaufwand zur Lösung der zugrunde liegenden, gekoppelten Systemgleichungen bestimmt. Hierzu werden Problemstellungen mit hierarchischer und nichthierarchischer Kopplung vorgestellt. Während der Lösungsaufwand bei hierarchischer Kopplung durch die einzelnen, sequenziell durchführbaren Analysen bestimmt wird, vervielfältigt sich der Aufwand bei Aufgaben mit nichthierarchischer Kopplung aufgrund der erforderlichen iterativen Lösung der Systemgleichungen. Weiterhin sind multidisziplinäre Optimierungsprobleme meist nichtlinear und häufig auch nichtkonvex.

Zur effizienten Lösung komplexer Optimierungsprobleme wird ein Entwurfsoptimierungssystem vorgestellt, das im Wesentlichen drei Ansätze miteinander verknüpft: Hohe Effizienz der Optimierungsalgorithmen, Verwendung von Approximationen und Grobparallelisierung der Prozesse. Somit lassen sich dort die in dieser Arbeit behandelten Themen einordnen. Als bestehender oder potentieller Bestandteil des Entwurfsoptimierungssystems werden Optimierungsalgorithmen vorgestellt und bezüglich ihrer Effizienz und Parallelisierbarkeit bewertet. In Bezug auf die Effizienz werden unter den „klassischen“ Optimierungsverfahren die Sequenzielle Quadratische Programmierung (SQP) sowie die Sequenzielle Konvexe Approximation als die geeignetsten angesehen. Suchverfahren bieten, rein an der Zahl der erforderlichen Systemanalysen gemessen, die geringste Effizienz. Sie gewinnen jedoch durch die Grobparallelisierung an Bedeutung. Von der Grobparallelisierung profitieren in erster Linie die Verfahren, die nur

ein oder wenige Male eine bestimmte Anzahl von Analysen benötigen und/oder eine Anpassung an die vorhandenen Rechenressourcen erlauben. Ersteres ist vor allem bei den globalen Approximationsmethoden, so auch bei der RSM gegeben. Unter den anderen Verfahren sind es speziell die Suchmethoden Monte-Carlo-Verfahren und die Evolutionären Algorithmen, die eine relativ große Flexibilität bezüglich der Zahl der pro Iteration benötigten Systemanalysen besitzen. Diese soll als ganzzahliges Vielfaches der zur Verfügung stehenden Rechner/Prozessoren gewählt werden. Bei den mathematischen Optimierungsverfahren sind hier Grenzen gesetzt, da die Zahl der benötigten Systemanalysen mit der Zahl der Entwurfsvariablen festgelegt ist.

Im Hinblick auf die Erarbeitung geeigneter Vorgehensweisen erfolgt eine Aufarbeitung bestehender Techniken der RSM. Als wichtiger Bestandteil werden dabei die Kriterien zur Beurteilung einer erstellten Approximation erachtet. Diese werden später in einem Verfahren eingesetzt, das eine bestmögliche Wahl erweiterter Ansatzterme für das Approximationsmodell gewährleistet. Durch die Analyse der Residuen gelingt eine Bewertung der einzelnen Stützpunkte, die für die Erstellung der Response Surface Approximationen (RSAs) benötigt werden. Eventuell fehlerbehaftete Stützpunkte können somit identifiziert und genauer untersucht werden. Im Hinblick auf die Ermittlung eines Optimums ist es ebenfalls erforderlich, den Einfluss jedes Stützpunkts auf den Verlauf der RSA zu bestimmen und gegebenenfalls ungeeignete Stützpunkte nicht zur Anpassung der Approximationsmodelle zu verwenden. Für diese Aufgabe erweisen sich skalierte Residuen als am besten geeignet. Den Methoden zur Wahl der Stützpunkte (Design of Experiments, DOE) kommt insofern eine Bedeutung zu, indem sie mit möglichst wenigen Systemanalysen genügend Informationen bereitstellen, um eine gute Approximation zu erhalten. Aufgrund des hohen Berechnungsaufwands für die einzelnen Systemanalysen erscheinen insbesondere solche als geeignet, die ein sequenzielles Vorgehen ermöglichen, bei dem sukzessive ein immer höherwertigeres Approximationsmodell angepasst werden kann.

Die ausgewählten Methoden werden anhand verschiedener Optimierungsbeispiele mit nichtlinearen Ziel- und Restriktionsfunktionen untersucht. Für monoton fallende oder steigende Funktionsverläufe erweisen sich lineare Approximationsmodelle als überraschend gut geeignet. Mit ihnen können die globalen Tendenzen gut erfasst werden, so dass sie bereits relativ nah an das Optimum heranzuführen. Die dafür prognostizierte Systemantwort liegt jedoch meist deutlich vom exakten Wert entfernt. Das mit einem quadratischen Approximationsmodell ermittelbare Optimum kann bei monotonen Funktionsverläufen schlechter sein, als das mit einer linearen Approximation ermittelte. Dabei zeigt sich, dass der Verlauf der quadratischen Approximation durch das verwendete DOE stark beeinflusst wird. Bei Funktionsverläufen mit einem Extremum im Definitionsbereich können mit quadratischen Approximationsmodellen erwartungsgemäß gute Erfolge erzielt werden. Auch hier hat das DOE teilweise einen deutlichen Einfluss auf die Approximation und somit auf das ermittelbare Optimum. Aus den Untersuchungen folgt die Angabe eines für die meisten Problemstellungen geeigneten DOEs. Dieses ermöglicht das sequenzielle Vorgehen, indem zunächst mit einer kleineren Zahl von Stützpunkten wichtige Variableneinflüsse identifiziert werden,

während weitere Systemanalysen zur Anpassung einer höherwertigen Approximation durchgeführt werden.

Zur Beschreibung von Nichtlinearitäten, die mit einem quadratischen Approximationsmodell nicht erfassbar sind, wird das durch eine Transformation der Systemantworten erzielbare Verbesserungspotential untersucht. Dabei zeigt sich, dass nichtlineare, monotone Funktionsverläufe mit einem geringem Aufwand wesentlich besser approximierbar sind, obwohl weiterhin ein quadratisches Approximationsmodell zugrunde liegt und deshalb keine zusätzlichen Informationen in Form von Systemanalysen erforderlich sind. Dies gilt jedoch bei Funktionsverläufen mit einem Extremum im approximierten Bereich nur noch eingeschränkt. Hier lassen sich zwar Transformationen identifizieren, die zu einer besseren Approximation der Stützpunkte führen, das hiermit ermittelbare Optimum ändert sich jedoch nur unwesentlich. Gerade bei multidisziplinären Problemstellungen liegen meist sehr unterschiedliche Variableneinflüsse vor. Diese können mit Transformationen der Systemantworten nicht individuell erfasst und behandelt werden.

Aus diesem Grund werden erweiterte funktionale Zusammenhänge in den einzelnen Ansatztermen eingeführt. Für eine möglichst flexible Definition der einzelnen Ansatzfunktionen werden diese in einer parametrisierten Form definiert. Mit der Wahl eines diskreten Parameters kann der Funktionstyp für jeden Ansatzterm gewählt werden. Als Funktionstypen stehen Potenzfunktionen, exponentielle, logarithmische und trigonometrische Funktionen zur Verfügung. Um die Variationsbreite der möglichen Ansatzterme zu erhöhen, kann der Einfluss jeder einzelnen Variablen durch einen kontinuierlichen Parameter bestimmt werden. Die somit beschreibbaren Ansatzterme werden als Linearkombination zusammengesetzt. Auf diese Weise können die Bewertungskriterien der RSM weiterhin genutzt werden. Zudem sind die Linearkoeffizienten sehr einfach bestimmbar. Anders ist dies bei der Auswahl einer geeigneten Kombination verschiedener Ansatzterme bei denen jeweils noch die zugehörigen Parameter so zu bestimmen sind, dass die resultierende Approximation die Stützpunkte möglichst gut wiedergibt. Diese Aufgabe wird mit einem diskret-nichtlinearen Ausgleichsproblem formuliert, für das geeignete Lösungsverfahren vorgestellt werden. Der diskret-kombinatorische Teil wird von einem Genetischen Algorithmus übernommen. Für die dort eingebettete Lösung des nichtlinearen Regressionsproblems eignen sich Gradientenverfahren und Evolutionsstrategien. Da hier teilweise sehr viele lokale Lösungen existieren, erweist sich eine Kombination der Verfahren als am besten geeignet. Der auf Evolutionsstrategien basierende Algorithmus dient zum Lokalisieren von Bereichen, die eine globale Lösung beinhalten können. Das Gradientenverfahren kann anschließend schnell eine exakte Lösung bestimmen. Das beschriebene Verfahren zur Zusammenstellung eines Approximationsmodells mit erweiterten Ansatztermen stellt ein komplexes Werkzeug dar. Deshalb wird eine strukturierte Vorgehensweise vorgeschlagen, mit der ein unverhältnismäßig großer Aufwand vermeidbar ist. Da für einige der verwendbaren Ansatzterme mehr Stützpunkte benötigt werden, erfolgt die Angabe geeigneter Methoden zur Wahl weiterer Stützstellen.

Die zur Bestimmung der Stützpunkte erforderlichen Systemanalysen können unabhängig voneinander durchgeführt werden. Durch die Grobparallelisierung der Analysen auf einem PC-Cluster lässt sich der hierfür benötigte zeitliche Aufwand drastisch reduzieren. Hierzu wird zunächst das Zusammenwirken zwischen dem Entwurfs-optimierungssystem und dem PC-Cluster beschrieben und die daraus resultierende Leistungssteigerung quantifiziert. Dabei ist ein mit der Zahl der verwendeten Rechner/Prozessoren nahezu linear steigender Leistungszuwachs festzuhalten. Aus dem zugrunde liegenden Funktionsprinzip und weiteren Gesichtspunkten werden Anforderungen an eine zu realisierende Hardware- und Softwarekonfiguration abgeleitet, deren Umsetzung beschrieben wird. Das System ermöglicht eine prioritätsgesteuerte Abarbeitung vieler, von unterschiedlichen Benutzern stammender Berechnungsaufträge. Dabei können Arbeitsplatzrechner temporär (z. B. in der Nacht, am Wochenende) in das System eingebunden werden, um die Rechenkapazität weiter zu erhöhen.

Die erarbeiteten Vorgehensweisen und Verfahren werden anhand von komplexen Optimierungsproblemen hinsichtlich ihrer Effizienz und Grenzen bewertet. Bei der Behandlung der ersten Optimierungsaufgabe können die erarbeiteten Methoden mit Erfolg eingesetzt werden. Hierbei zeigt sich erneut, dass bereits mit einer linearen Approximation und wenigen Systemanalysen ein vergleichsweise gutes Optimum ermittelt werden kann, wenn die Ziel- und Restriktionsfunktionen einen monoton steigenden oder fallenden Verlauf haben. Eine deutliche Verbesserung wird mit erweiterten Ansatztermen in Verbindung mit der optimalen Auswahl einflussreicher Interaktionsterme höherer Ordnung erreicht. Gegenüber einem mathematischen Optimierungsverfahren wird nur ein Drittel an exakten System benötigt. Bei der Lösung des zweiten Optimierungsproblems zeigen quadratische Ansatzfunktionen ein geringes Approximationsvermögen. Erst mit einem optimal zusammengestellten Approximationsmodell mit erweiterten Ansatztermen gelangt man in die Nähe des exakten Optimums. Hierbei kann festgestellt werden, dass dies für Entwurfsvariablen mit großem Einfluss auf den Wert der Zielfunktion besser gelingt, als für weniger einflussreiche Entwurfsvariablen.

Eine Schwäche des verwendeten Verfahrens ist im Aufwand zu sehen, der für die Zusammenstellung eines Approximationsmodells erforderlich ist. Während der diskrete oder der nichtlineare Teil der Lösung des zugrunde liegenden Regressionsproblems durchaus in akzeptablen Zeiten zu bewerkstelligen ist, steigt der Aufwand extrem stark, wenn eine kombinierte Anwendung erforderlich ist.

Literaturverzeichnis

- [**Baier 1994**] BAIER, H. ; SEESSELBERG, C. ; SPECHT, B.: *Optimierung in der Strukturmechanik*. Braunschweig/Wiesbaden : Friedrich Vieweg & Sohn Verlagsgesellschaft mbH, 1994
- [**Baier 2002a**] BAIER, H.: *Faserverbundwerkstoffe*. Skriptum zur Vorlesung und Übung, Lehrstuhl für Leichtbau, Technische Universität München. 2002
- [**Baier 2002b**] BAIER, H.: *Multidisciplinary Design Optimization*. Skriptum zur Vorlesung und Übung, Lehrstuhl für Leichtbau, Technische Universität München. 2002
- [**Baier, Gleichmar u. a. 2001**] BAIER, H. ; GLEICHMAR, R. ; LANGER, H. ; PÜHLHOFER, T.: A Grain Parallelized Structural MDO Engine which includes Shape and Topology Variables. In: *CEAS Conference on Multidisciplinary Aircraft Design and Optimization, Köln*, 2001
- [**Balabanov u. a. 1999**] BALABANOV, V. O. ; GIUNTA, A. A. ; GOLOVIDOV, O. ; GROSSMAN, B. ; MASON, W. H. ; WATSON, L. T. ; HAFTKA, R. T.: Reasonable Design Space Approach to Response Surface Approximation. In: *Journal of Aircraft* 36 (1999), Nr. 1, S. 308–315
- [**Bloebaum u. a. 1994**] BLOEBAUM, C. L. ; HONG, W. ; PECK, A.: Improved Moved Limit Strategies for Approximate Optimization. In: *Proceedings of the 5th AIAA/USAF/NASA/ISSMO Symposium on Multidisciplinary Analysis & Optimization*. Panama City, 1994, S. 843–850
- [**Blumhardt 2001**] BLUMHARDT, R.: *Numerische Optimierung des Crashverhaltens von Fahrzeugstrukturen und -komponenten*, Fakultät für Maschinenwesen der Technischen Universität München, Dissertation, 2001
- [**BOSS quattro Homepage**] N.N.: *BOSS quattro der Firma Samtech*. www.samcef.com → Products → BOSS quattro
- [**Box und Draper 1971**] BOX, G. E. P. ; DRAPER, N. R.: Factorial Designs, the $X^T X$ Criterion and Some Related Matter. In: *Technometrics* 13 (1971), S. 731–742

- [**Box und Draper 1987**] BOX, G. E. P. ; DRAPER, N. R.: *Empirical Modell-Building and Response Surfaces*. New York : John Wiley & Sons, Inc., 1987 (Wiley series in probability and mathematical statistics)
- [**Box und Wilson 1951**] BOX, G. E. P. ; WILSON, K. P.: On the Experimental Attainment of Optimum Conditions. In: *Journal of the Royal Statistical Society Series B* (1951), S. 1–45
- [**Bronstein 1991**] BRONSTEIN, I. N. ; SEMENDJAJEW, K. A.: *Taschenbuch der Mathematik*. 25. Stuttgart – Leipzig : B. G. Teubner Verlagsgesellschaft, 1991
- [**Chen 1993**] CHEN, T. Y.: Calculation of Move Limits for the Sequential Linear Programming. In: *International Journal for Numerical Methods in Engineering* 36 (1993), S. 2661–2679
- [**Drela 2001**] DRELA, M.: *XFOIL User Guide*. MIT (Veranst.), 2001
- [**Friedrich 2003**] FRIEDRICH, M.: *Anwendung der Response Surface Methode beim Entwurfs von schnell rotierenden Chopperscheiben aus CFK*, Lehrstuhl für Leichtbau, Technische Universität München, Semesterarbeit, TUM–MW 65/0330–SA, 2003
- [**Giunta u. a. 1997**] GIUNTA, A. A. ; BALABANOV, V. ; HAIM, D. ; GROSSMAN, B. ; MASON, W. H. ; WATSON, L. T.: Multidisciplinary Optimization of a Supersonic Transport Using Design of Experiments Theory and Response Surface Modelling. In: *The Aeronautical Journal* (1997), S. 347–356
- [**Gleichmar 1999**] GLEICHMAR, R.: Vergleiche zur Strömungs-Struktur-Kopplung mit iterativen und direkten Kopplungsmethoden auf der Basis von ANSYS/FLOTRAN und Spectrum. In: *Proceedings zum 17. CAD-FEM Users' Meeting*, 1999
- [**Gleichmar 2002**] GLEICHMAR, R.: Thermisch-mechanischer Entwurf eines in Sandwichbauweise ausgeführten Aufzugtürblatts unter Sonneneinstrahlung / Lehrstuhl für Leichtbau – Meiller Aufzugtüren GmbH. 2002. – Abschlussbericht
- [**Golovidov 1997**] GOLOVIDOV, O. B.: *Variable-Complexity Response Surface Approximations for Aerodynamic Parameters in HSCT Optimization*. Blacksburg, Virginia, Virginia Polytechnic Institute and State University, Diplomarbeit, 1997
- [**Hausmann 2002**] HAUSMANN, M.: *Anwendung der Response Surface Methode beim mechanisch-thermischen Entwurf von Sandwichkonstruktionen*, Lehrstuhl für Leichtbau, Technische Universität München, Diplomarbeit, TUM–MW 65/0205–DA, 2002
- [**Hutchinson u. a. 1994**] HUTCHINSON, M. G. ; UNGER, E. R. ; MASON, W. H. ; GROSSMAN, B. ; HAFTKA, R. T.: Variable Complexity Aerodynamic Optimization of an HSCT Wing Using Structural Wing-Weight Equations. In: *Journal of Aircraft* 31 (1994), Nr. 1, S. 110–116

-
- [**iSIGHT Homepage**] N.N.: *iSIGHT der Firma Engineous Software Inc.*
www.engineous.com → iSIGHT
- [**Jones 2001**] JONES, D. R.: A Taxonomy of Global Optimization Methods Based on Response Surfaces. In: *Journal of Global Optimization* 21 (2001), S. 345–383
- [**Kaiser 1998**] KAISER, C. ; RÖSSLER, N.: Aeroelastic Tailoring am CFK-Tragflügel eines Verkehrsflugzeugs. In: *Proceedings zum 16. CAD-FEM Users' Meeting*. Bad Neuenahr, 1998
- [**Kaufman u. a. 1996**] KAUFMAN, M. ; BALABANOV, V. ; BURGEE, S. L. ; GIUNTA, A. A. ; GROSSMAN, B. ; HAFTKA, R. T. ; MASON, W. H. ; WATSON, L. T.: Variable-Complexity Response Surface Approximations for Wing Structural Weight in HSCT Design. In: *Computational Mechanics* 18 (1996), S. 112–126
- [**Knill u. a. 1999**] KNILL, D. L. ; GIUNTA, A. A. ; BAKER, C. A. ; GROSSMAN, B. ; MASON, W. H. ; HAFTKA, R. T. ; WATSON, L. T.: Response Surface Models Combining Linear and Euler Aerodynamics for Supersonic Transport Design. In: *Journal of Aircraft* 36 (1999), Nr. 1, S. 75–86
- [**Kuchling 1991**] KUCHLING, H.: *Taschenbuch der Physik*. 13. Thun und Frankfurt/Main : Verlag Harri Deutsch, 1991
- [**Langer u. a. 2002**] LANGER, H. ; PÜHLHOFER, T. ; BAIER, H.: An Approach for Shape and Topology Optimization Integrating CAD Parametrization and Evolutionary Algorithms. In: *9th AIAA/ISSMO Symposium on Multidisciplinary Analysis and Optimization*. Atlanta – Georgia, USA, 2002
- [**Mathworks 2002**] THE MATHWORKS: *MATLAB Users's Guide*, 2002
- [**MATLAB Optimization Toolbox 2002**] THE MATHWORKS: *MATLAB Optimization Toolbox Users's Guide*, 2002
- [**MATLAB Statistics Toolbox 2002**] THE MATHWORKS: *MATLAB Statistics Toolbox Users's Guide*, 2002
- [**OpenPBS Homepage**] N.N.: *OpenPBS Homepage*. www.openpbs.org
- [**Myers 1971**] MYERS, R. H.: *Response Surface Methodology*. Boston : Allyn and Bacon, 1971
- [**Myers und Montgomery 1995**] MYERS, R. H. ; MONTGOMERY, D. C.: *Response Surface Methodology: Process and Product Optimization Using Designed Experiments*. New York : John Wiley & Sons, Inc., 1995 (Wiley series in probability and statistics)
- [**Pfister 1998**] PFISTER, G.: *In Search of Clusters*. 2. Prentice Hall PTR, NJ, 1998
- [**Pühlhofer u. a. 2003**] PÜHLHOFER, T. ; LANGER, H. ; BAIER, H.: Towards a Configuration Design Optimization Method. In: *5th World Congress on Structural and Multidisciplinary Optimization (ISSMO)*. Lido di Jesolo – Venedig, Italien, 2003

- [**Pühlhofer, Gleichmar u. a. 2002**] PÜHLHOFER, T. ; GLEICHMAR, R. ; LANGER, H. ; MÜLLER, M. ; MÜLLER, U. ; BAIER, H.: Multidisciplinary Optimization and Statistical Evaluations of Structures and Electro-Mechanical Components with ANSYS on a PC-Cluster. In: *Proceedings zum 20. CAD-FEM Users' Meeting*, 2002
- [**Piening 1998**] PIENING, M.: Ein Strukturkonzept für Tragflügel in Faserverbundbauweise mit Verformungskopplungen. In: *Proceedings zur Aeroelastik-Tagung der DGLR*. Göttingen, 1998
- [**Plengorth 2000**] PLENGORTH, A.: *Konstruktive Auslegung des Tragflügelkonzepts eines Hochleistungssegelflugzeugs nach dem Prinzip des Aeroelastic Tailoring*, Lehrstuhl für Leichtbau, Technische Universität München, Semesterarbeit, 2000
- [**Pukelsheim 1993**] PUKELSHEIM, F.: *Optimal Design of Experiments*. New York : Chapman & Hall, 1993
- [**Rodríguez u. a. 2001**] RODRÍGUEZ, J. F. ; PÉREZ, V. M. ; PADMANABHAN, D. ; RENAUD, J. E.: Sequential Approximate Optimization Using Variable Fidelity Response Surface Approximations. In: *Structural and Multidisciplinary Optimization* 22 (2001), S. 24–34
- [**Rodríguez u. a. 2000**] RODRÍGUEZ, J. F. ; RENAUD, J. E. ; WUJEK, B. A. ; TAPPETA, R. V.: Trust Region Model Management in Multidisciplinary Design Optimization. In: *Journal of Computational and Applied Mathematics* 124 (2000), S. 139–154
- [**Roux u. a. 1996**] ROUX, W. J. ; STANDER, N. ; HAFKTA, R. T.: Response Surface Approximations for Structural Optimization. In: *6th AIAA/USAF/NASA/ISSMO Symposium on Multidisciplinary Analysis and Optimization, AIAA 96-4050-CP*. Bellevue, WA, 1996, S. 642–654
- [**Sellar u. a. 1996**] SELLAR, R. S. ; BATILL, S. M. ; RENAUD, J. E.: Response Surface Based, Concurrent Subspace Optimization for Multidisciplinary System Design. In: *34th AIAA Aerospace Science Meeting and Exhibit, AIAA 96-0714*. Reno, Nevada, 1996
- [**SGE Homepage**] N.N.: *Sun Grid Engine*. www.sun.com/software/gridware
- [**Simpson u. a. 2001**] SIMPSON, T. W. ; PEPLINSKI, J. D. ; KOCH, P. N. ; ALLEN, J. K.: Metamodels for Computer-based Engineering Design: Survey and Recommendations. In: *Engineering with Computers* 17 (2001), S. 129–150
- [**Sobieski und Haftka 1996**] SOBIESZCZANSKI-SOBIESKI, J. ; HAFTKA, R. T.: Multidisciplinary Aerospace Optimization: Survey of Recent Developments. In: *34th AIAA Aerospace Sciences Meeting and Exhibit, AIAA 96-0711* (1996)
- [**Stein 1999**] STEIN, M. L.: *Interpolation of Spatial Data – Some Theory for Kriging*. New York : Springer Verlag, 1999

- [**Thalhauser 2000**] THALHAUSER, J.: *Anwendung der Response Surface Methode zur Berechnung strukturmechanischer Größen für ein Laminat unter Temperaturbelastung*, Lehrstuhl für Leichtbau, Technische Universität München, Semesterarbeit, TUM-MW 65/0018-SA, 2000
- [**Thomas u. a. 1992**] THOMAS, H. L. ; VANDERPLAATS, G. N. ; SHYY, Y. K.: A Study of Move Limit Adjustment strategies in the Approximation Concept Approach to Structural Synthesis. In: *Proceedings of the 4th AIAA/USAF/NASA/ISSMO Symposium on Multidisciplinary Analysis & Optimization*. Cleveland, OH, 1992, S. 507–512
- [**Tropov u. a. 1993**] TROPOV, V. V. ; FILATOV, A. A. ; POLYNKIN, A. A.: Multiparameter Structural Optimization Using FEM and Multipoint Explicit Approximations. In: *Structural Optimization 6* (1993), S. 7–14
- [**Ulrich 1999**] ULRICH, J.: *Ein Modell zur Simulation von Mikropumpen auf Basis der Finite-Elemente-Methode*, Fakultät für Maschinenwesen der Technischen Universität München, Dissertation, 1999
- [**Unal u. a. 1998**] UNAL, R. ; LEPSCH, R. A. ; MCMILLIN, M. L.: *Response Surface Model Building and Multidisciplinary Optimization Using D-optimal Designs*. AIAA Paper No. 98-4759. 1998
- [**Unger u. a. 1992**] UNGER, E. A. ; HUTCHISON, M. G. ; RAIS-ROHANI, M. ; HAFTKA, R. T. ; GROSSMAN, B.: Variable-Complexity Design of a Transport Wing. In: *International Journal of Systems Automation: Research and Applications 2* (1992), S. 87–113
- [**Venter 1998**] VENTER, G.: *Non-Dimensional Response Surfaces for structural Optimization with Uncertainty*, Graduate School of the University of Florida, Dissertation, 1998
- [**Venter u. a. 1996**] VENTER, G. ; HAFTKA, R. T. ; STARNES, J. H.: Construction of Response Surfaces for Design Optimization Applications. In: *6th AIAA/USAF/NASA/ISSMO Symposium on Multidisciplinary Analysis and Optimization, AIAA 96-4040-CP*. Bellevue, WA, 1996, S. 548–564
- [**Wenjie und Renaud 2001**] WENJIE, L. ; RENAUD, J. E.: A Comparative Study of Trust Region Managed Approximate Optimization. In: *4^{2nd} AIAA/ASME/ASCE/AHS/ASC Structures, Structural Dynamics and Materials Conference and Exhibit, AIAA 2001-1499*. Seattle, WA, 2001