

**Eine adaptive sequentielle Prozedur zur effizienten
Optimierung des CNC-gesteuerten Drückprozesses**

Dissertation

zur Erlangung des Grades
eines Doktors der Naturwissenschaften
der Universität Dortmund

Dem Fachbereich Statistik
der Universität Dortmund

vorgelegt von

Nadine Henkenjohann

Dortmund, Dezember 2006

Inhaltsverzeichnis

1	Einleitung	1
2	Der CNC-gesteuerte Drückprozess	3
2.1	Beschreibung des Drückprozesses	3
2.2	Anforderungen an die Optimierungsprozedur	10
2.3	Eine adaptive sequentielle Optimierungsprozedur	12
3	Fallbasiertes Schließen	14
3.1	Prinzip des fallbasierten Schließens	14
3.2	Erweiterung des fallbasierten Schließens	17
4	Raumfüllende Versuchspläne bei unbekanntem Restriktionen	19
4.1	Raumfüllende Versuchspläne	19
4.1.1	Distanzbasierte Versuchspläne	21
4.1.2	Latin Hypercube Designs	23
4.1.3	Uniform Coverage Designs	25
4.1.4	Kombinierte raumfüllende Versuchspläne	26
4.2	Unbekannte konvexe Restriktionen	28
4.2.1	Schrittweise Identifikation des Versagensbereichs	28
4.2.2	Schrittweise Anpassung des Versuchsbereichs	30

4.2.3	Wahl des Kriteriums	30
4.2.4	Beschreibung des Algorithmus	32
4.2.5	Robustheit des Algorithmus	34
4.3	Simulationsstudie	36
4.3.1	Simulationsaufbau	36
4.3.2	Ergebnisse	39
4.3.3	Zusammenfassung der Simulationsergebnisse	47
5	Gauß-Prozess-Modell	49
5.1	Stochastische Prozesse	49
5.2	Motivation des Gauß-Prozess-Modells	52
5.3	Definition des Gauß-Prozess-Modells	53
5.4	Korrelationsfunktionen	54
5.5	Parameterschätzung	56
5.5.1	Bekannte Korrelationsfunktion	59
5.5.2	Unbekannte Korrelationsfunktion	60
5.6	Modellvorhersage	61
5.6.1	Bekannte Korrelationsfunktion	62
5.6.2	Unbekannte Korrelationsfunktion	63
5.6.3	Interpretation des BLUP	64
5.7	Modellvalidierung	65
6	Multivariate Expected Improvement	67
6.1	Das Expected-Improvement-Kriterium	67
6.2	Univariates Simulationsbeispiel	74
6.3	Wünschbarkeitsfunktionen	76

6.3.1	Wünschbarkeitsfunktionen von Harrington	76
6.3.2	Wünschbarkeitsfunktionen von Derringer und Suich	78
6.3.3	Wünschbarkeitsindex	80
6.3.4	Verteilungsfunktionen der Wünschbarkeiten und des Wünschbarkeitsindexes	81
6.4	Das multivariate Expected-Improvement-Kriterium	86
6.4.1	Ansatz 1 - Improvement des Wünschbarkeitsindexes	87
6.4.2	Ansatz 2 - Improvement des Wünschbarkeitsvektors	90
6.5	Multivariates Anwendungsbeispiel	94
7	Optimierung des Drückprozesses	99
7.1	Vorhersage des stabilen Bereichs	101
7.2	Sequentielle Generierung eines raumfüllenden Versuchsplans	104
7.3	Modellselektion	108
7.4	Sequentielle Modellverfeinerung	114
8	Zusammenfassung	118
9	Ausblick	123
A	Appendix	125
A.1	Bedingte multivariate Normalverteilung	125
A.2	Tabellen	126
A.3	Abbildungen	131
A.4	Programmcode	134

Kapitel 1

Einleitung

Diese Arbeit entstand im Rahmen des Teilprojekts C6 „Methoden der statistischen Versuchsplanung zur Analyse und Optimierung von komplexen Umformprozessen am Beispiel des CNC-gesteuerten Drückens“ des Sonderforschungsbereichs SFB 475 „Komplexitätsreduktion in multivariaten Datenstrukturen“. Das globale Ziel des Teilprojekts ist die Weiterentwicklung des CNC-gesteuerten Drückverfahrens durch den Einsatz effizienter statistischer Methoden zur Prozessauslegung und Optimierung.

Obwohl das Drückverfahren es ermöglicht komplexe rotationssymmetrische Geometrieformen mit hoher Form- und Maßgenauigkeit herzustellen, werden in der Industrie andere Fertigungsverfahren wie das Tiefziehen bevorzugt. Grund dafür ist die hohe Komplexität des Drückprozesses, die bisher eine effiziente Optimierung des Drückprozesses verhindert hat. Die Effizienz des Optimierungsverfahrens spielt beim Drückprozess aber eine besonders große Rolle, da Bauteile oft nur in geringer Stückzahl produziert werden.

Zur Analyse und Optimierung des Drückprozesses wird standardmäßig das Expertenwissen mit der One-Factor-at-a-Time-Methode kombiniert. Zwar existieren Finite-Elemente-Methoden, die den Drückprozess simulieren, diese sind aber im Rahmen der Prozessoptimierung nicht hilfreich, da schon die Simulation einer einzelnen Bahnbewegung der Drückrolle mehrere Stunden benötigt. In der Literatur konnten keine Hinweise gefunden werden, dass statistische Methoden

zur Analyse und Optimierung des Drückprozesses eingesetzt werden.

Zu Beginn der Forschungsarbeiten im Rahmen des SFB-Projekts wurden klassische Methoden der statistischen Versuchsplanung eingesetzt. Obwohl mit diesen Methoden erste Erfolge erzielt werden konnten, erwiesen sich diese Verfahren als sehr ineffizient. Zudem erlauben es diese Verfahren nicht, das Optimierungspotenzial auszunutzen. Um eine effiziente Optimierung des Drückprozesses zu ermöglichen, wurde eine neue Optimierungsprozedur entwickelt, die speziell auf die Eigenschaften des Drückprozesses zugeschnitten ist. Diese Optimierungsprozedur wird in dieser Arbeit vorgestellt.

In Kapitel 2.1 und 2.2 wird zunächst der Drückprozess beschrieben und aus den speziellen Eigenschaften des Drückprozesses die Anforderungen an eine effiziente Optimierungsprozedur hergeleitet. Basierend auf diesen Anforderungen wurde eine adaptive sequentielle Optimierungsprozedur (ASOP) entwickelt, die aus einer Kombination von aufeinander abgestimmten Modellen und Methoden besteht. In Kapitel 2.3 wird zunächst das Grundkonzept der ASOP erläutert. Der Ablauf der Prozedur lässt sich schematisch als geschlossener Kreislauf darstellen, der im Wesentlichen vier Bearbeitungsschritte enthält. In jedem der vier Bearbeitungsschritte werden Methoden und Modelle eingesetzt, die eine Weiterentwicklung bereits existierender Methoden darstellen oder die neu erarbeitet wurden. Diese vier Bausteine werden ausführlich in den Kapiteln 3 bis 6 vorgestellt. Die Verifikation der ASOP für den Drückprozess erfolgt anschließend anhand der Optimierung einer speziellen Bauteilgeometrie. Die Ergebnisse der einzelnen Optimierungsschritte werden in Kapitel 7 präsentiert. In Kapitel 8 wird eine Zusammenfassung dieser Arbeit gegeben. Kapitel 8 behandelt noch offene Arbeitspunkte und gibt abschließend einen Überblick über geplante Einsatzbereiche der ASOP.

Kapitel 2

Der CNC-gesteuerte Drückprozess

2.1 Beschreibung des Drückprozesses

Das Umformverfahren Drücken stellt ein leistungsfähiges, sehr flexibles, aber zugleich hochkomplexes Fertigungsverfahren dar. Das Verfahren wird zur Herstellung rotationssymmetrischer Hohlkörper mit nahezu beliebiger Mantellinie eingesetzt. Der Drückprozess besitzt eine lange Historie, wobei die Ursprünge des Drückens in der Zeit des Mittelalters vermutet werden, vgl. Abbildung 2.1. Beim Drücken wurden Bauteile ursprünglich in Handarbeit gefertigt, wobei das Erfahrungswissen und das Geschick des Handwerkers von wesentlicher Bedeutung waren. Heutzutage ist das Drückverfahren CNC-gesteuert (computerized numerical control). Auf diese Weise kann eine automatische Produktion mit deutlich höheren Bauteilqualitäten und schnellerer Produktionszeit erfolgen.

Ausgangsmaterial beim CNC-gesteuerten Drückprozess sind Metallronden. Die Anforderungen an das Bauteil legen dabei das Material, die Blechdicke und den Durchmesser der Ronde fest. Im ersten Produktionsschritt wird die Metallronde mit Hilfe eines Hydraulikreitstocks an das Drückfutter gepresst. Die gesamte Einheit, bestehend aus dem Drückfutter, der Ronde und dem Hydraulikreitstock, wird in Rotation versetzt. Durch die Bewegung des Umformwerkzeugs, der so

genannten Drückrolle, wird die Ronde schrittweise umgeformt, bis sie an die Form des Drückfutters angepasst ist, siehe Abbildung 2.2.



Abbildung 2.1: Holzschnitt des Drückhandwerks (Quelle: Runge (1993))

Die Geometrie des Drückfutters kann konkaver oder konvexer Natur sein, wobei komplexere Bauteile durch eine Kombination beider Krümmungsarten erzeugt werden. Während der Umformung kommt in der Regel ein Rondengegenhalter zum Einsatz, um den Prozess zu stabilisieren. Eine detaillierte Beschreibung des Prozesses wird in Ewers (2005) gegeben.

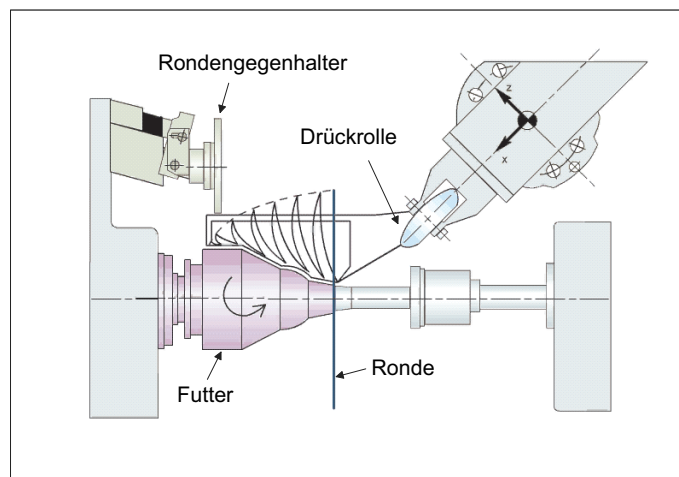


Abbildung 2.2: Schematische Darstellung des Drückprozesses

Bauteile, die mit Hilfe des Drückprozesses hergestellt werden, finden ihre Anwendung hauptsächlich in der Chemie-, Lampen-, Automobil- und Luftfahrtindustrie. In Abbildung 2.3 sind typische Bauteile des Drückverfahrens abgebildet.



Abbildung 2.3: Typische Bauteile

Gefertigt wird in erster Linie in kleinen und mittleren Stückzahlen. Mit dem Drückprozess können eine Vielzahl von Materialien wie Stahl, Edelstahl und Aluminium verarbeitet werden. Die Verfahrensbewegung der Drückrolle wird durch den so genannten Stadienplan beschrieben, vgl. Abbildung 2.4. Der Stadienplan lässt sich durch den Verlauf der Randbegrenzungskurve (RBK), der Anzahl der Stufen, der Form der Bahnen und zwei weiteren Parametern e_{AQU} und V_{AQU} charakterisieren.

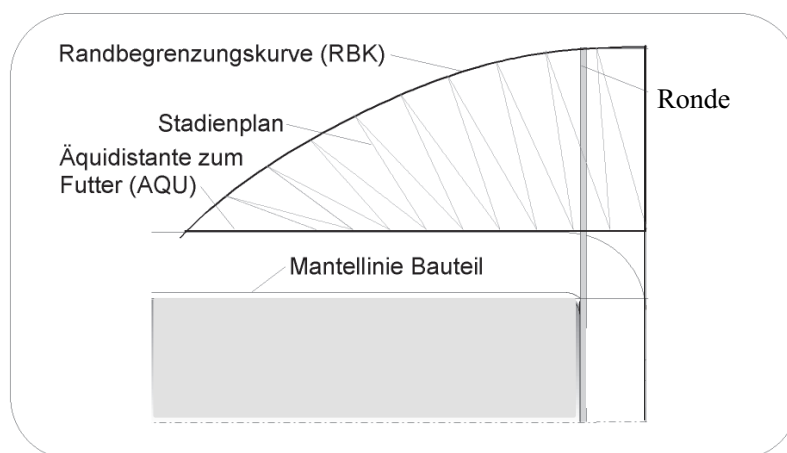


Abbildung 2.4: Stadienplan mit zehn Stufen und geradlinigen Bahnen

Der Parameter e_{AQU} gibt den Endpunkt der Rückbewegung der Drückrolle der letzten Bahn auf der Äquidistanten zum Futter (AQU) an. Die Angabe erfolgt in Prozent der Gesamtlänge der Äquidistanten. In Abbildung 2.5 ist dieser Sachverhalt für zwei unterschiedliche Werte von e_{AQU} graphisch dargestellt. Die Verteilung der Bahnen auf der Äquidistanten wird durch eine geometrische Folge beschrieben, die durch den Parameter V_{AQU} definiert ist. Für $V_{AQU} = 1$ besitzen die Bahnen den gleichen Abstand auf der Äquidistanten. Während für $V_{AQU} \leq 1$ Abstände der Bahnen mit steigender Stufenzahl sinken, steigen diese für $V_{AQU} > 1$, vgl. Abbildung 2.5.

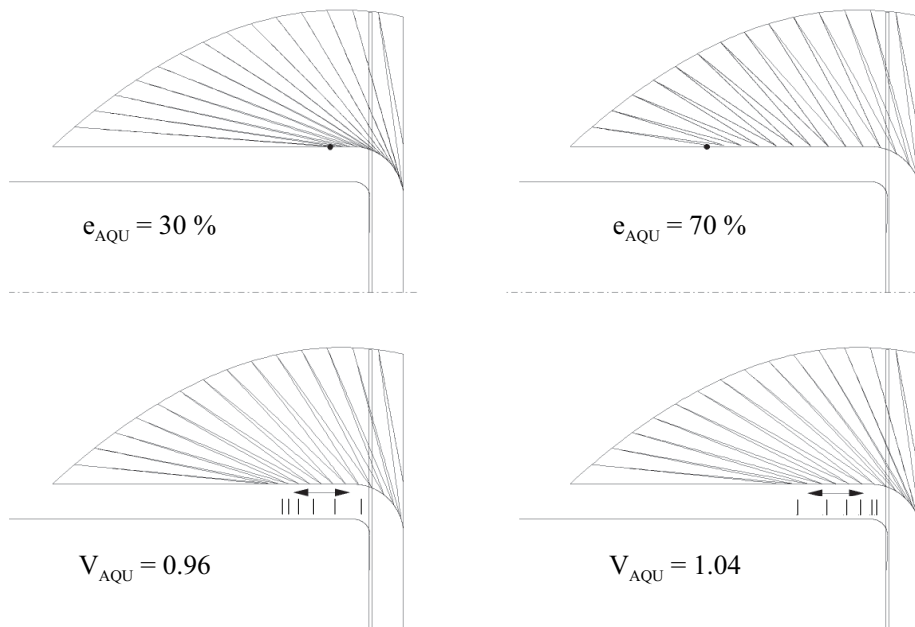


Abbildung 2.5: Einfluss von e_{AQU} und V_{AQU} auf die Stadienplangeometrie

Das Drückverfahren wird hauptsächlich eingesetzt, wenn andere Umformverfahren wie das Tiefziehen und ähnliche Prozesse an ihre Grenzen stoßen. Das Drückverfahren erlaubt z. B. die Produktion weitaus komplexerer Geometrieformen, als sie mit dem Tiefziehen realisierbar sind. Des Weiteren zeichnet sich das Formdrücken durch eine hohe Form- und Maßgenauigkeit aus, die die spanende Nachbearbeitung der Werkstücke häufig überflüssig macht, vgl. Runge (1993). Weitere Vorteile sind sehr gute Oberflächenqualitäten und die gute Umformbarkeit ansonsten schwer zu verarbeitender Werkstoffe.

Trotz dieser Vorteile des Drückprozesses ist das Tiefziehen deutlich populärer. Ein Grund dafür ist das begrenzte Anwendungsfeld durch die Einschränkung auf rotationssymmetrische Bauteile. Der wesentliche Grund ist aber die Komplexität des Prozesses. In Abbildung 2.6 sind einige Eigenschaften des Drückprozesses angegeben, die zur Komplexität des Prozesses beitragen.

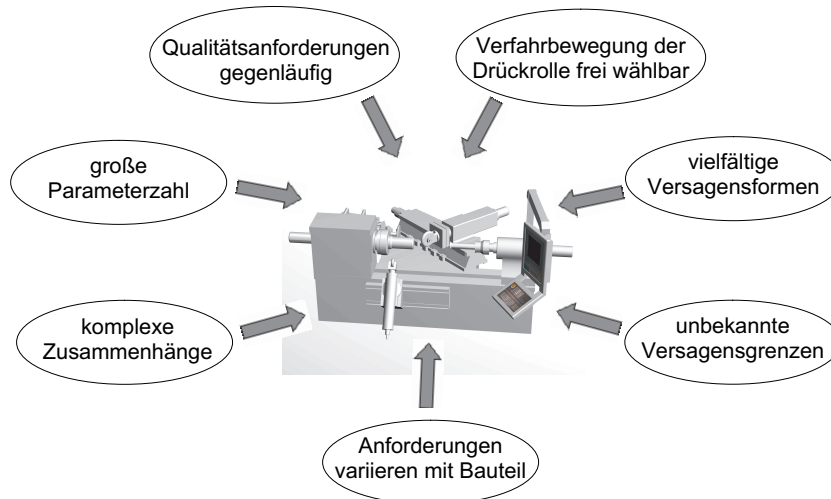


Abbildung 2.6: Eigenschaften des Drückprozesses

Um die Gesamtqualität eines Bauteils beurteilen zu können, sind Abweichungen von der Sollgeometrie, die Oberflächenqualität und wirtschaftliche Aspekte wie die Produktionszeit zu erheben. Da für diese Zielgrößen in der Regel unterschiedliche Maßeinheiten verwendet werden, ist es schwierig, verschiedene Zielgrößen miteinander zu vergleichen. Zudem tendieren diese Zielgrößen dazu, sich gegenläufig zu verhalten. Kann eine Zielgröße durch Veränderung der Parametereinstellungen verbessert werden, führt dies in der Regel zu einer Verschlechterung einer anderen Zielgröße. Dies erfordert einen guten Kompromiss zwischen den Zielgrößen. Mit dieser Problematik beschäftigt sich die Arbeit von Auer et al. (2004), in der verschiedene multivariate Methoden bezüglich ihrer Anwendbarkeit auf den Drückprozess und ihrer Ergebnisse verglichen werden.

Des Weiteren zeichnet sich der Drückprozess durch eine große Anzahl von möglichen Einflussparametern aus, wobei zum Teil komplexe Zusammenhänge zwischen den Faktoren und den Zielgrößen existieren.

Ein weiterer Grund für die Komplexität des Prozesses ist die im Prinzip frei wählbare Verfahrensbewegung der Drückrolle. Von besonderer Schwierigkeit ist es dabei den Verlauf der RBK zu bestimmen, der von Bauteil zu Bauteil variiert. Wird die RBK nicht korrekt berechnet, so kann die Drückrolle über das Werkstück hinausfahren. Durch die elastische Rückfederung der Ronde würde auf der Rückbewegung der Drückrolle zum Futter eine Kollision mit dem Werkstück erfolgen, was zum sofortigen Versagen des Bauteils führt. Eine zu flache RBK hingegen erhöht die Wahrscheinlichkeit, dass Faltenbildung auftritt. In Ewers (2005) wird ein Ansatz beschrieben, wie die RBK abhängig von der Bauteilgeometrie bestimmt werden kann. Zudem wurden dort ausführliche Untersuchungen zur Wahl der Stufenanzahl und der Form der Bahnen durchgeführt.

Ein weiteres Problem beim Drückprozess sind die für diesen Prozess typischen Versagensformen Riss- und Faltenbildung, vgl. Abbildung 2.7. Das Auftreten von Rissen wird durch zu starkes Ausdünnen des Materials hervorgerufen. Eine detaillierte Beschreibung der Entstehung von Falten ist in Ewers (2005) zu finden. Ein bestimmter Grad der Faltenbildung in den ersten Stufen des Prozesses ist noch tolerabel, da eine geringe Faltenbildung im weiteren Verlauf der Umformung geglättet wird. Erst das Überschreiten einer kritischen Faltenstärke führt zum Versagen des Bauteils während der Fertigung. Beim Drückprozess hängt der Anteil des Versagensbereichs am Parameterraum von der Wahl der Bauteilgeometrie ab. Je nach Komplexität der Geometrie kann der Versagensbereich einen Großteil des Parameterraums ausmachen.

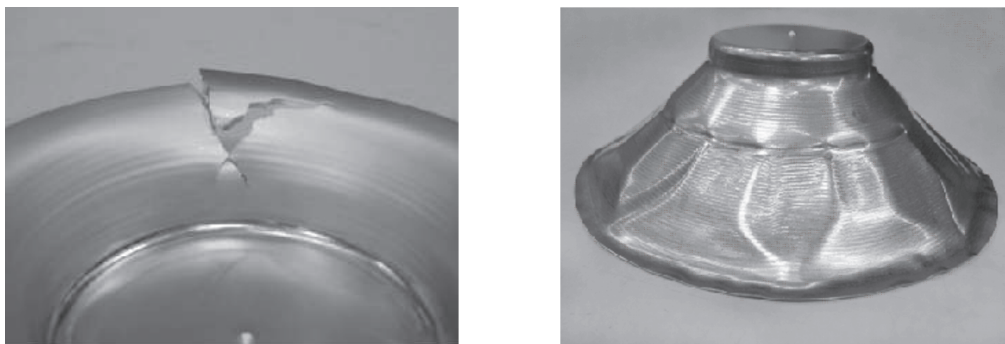


Abbildung 2.7: a) Bauteil mit radialem Riss b) Bauteil mit starker Faltenbildung

Da die Faltenbildung eine wesentliche Einschränkung der Wirtschaftlichkeit des Formdruckprozesses darstellt, sind mathematische Modelle gefragt, die es erlauben, das Auftreten der Faltenbildung und die Stärke der Falten zu modellieren. In Erdbrügge et al. (2001) wurden dazu so genannte Proportional-Odds-Modelle erfolgreich eingesetzt. Die Problematik lag hier in der Nichtmessbarkeit der Stärke der Faltenbildung. Stattdessen wird eine qualitative Skala verwendet, mit deren Hilfe Experten die Stärke der Faltenbildung eines Bauteils beurteilen. Die Einteilung in die verschiedenen Kategorien erfolgt dabei, ohne dass die Experten Hintergrundinformationen über die Fertigung des jeweiligen Bauteils besitzen. Diese Proportional-Odds-Modelle lassen sich auch zur Modellierung der Rissbildung verwenden. Die eigentliche Problematik beim Auftreten von Versagensteilen ist, dass die Versagensgrenzen nicht bekannt sind und gute Bauteile sich zudem am Rand des Versagensbereichs befinden. Durch experimentelle Untersuchungen (Ewers, 2005) konnte zudem gezeigt werden, dass Bauteile mit sehr guter Qualität in der Regel sogar von mehreren Seiten von Bauteilversagen umgeben sind, vgl. Abbildung 2.8, in der ein Schnitt durch den Parameterraum von e_{AQU} und dem Vorschub F gegeben ist. Beim Auftreten von Rissen oder zu starker Faltenbildung muss der Produktionsvorgang sofort gestoppt werden, um Schäden an der Maschine zu verhindern. Dies führt zu fehlenden Werten für die Zielgrößen.

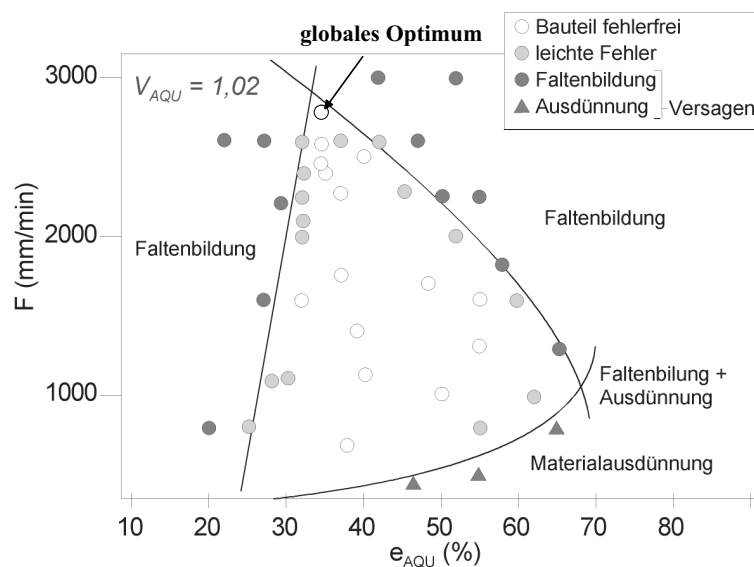


Abbildung 2.8: Schnitt durch den Parameterraum

2.2 Anforderungen an die Optimierungsprozedur

Der Drückprozess wird hauptsächlich in der Kleinserienfertigung angewandt. Aus diesem Grund kommt der effizienten Optimierung des Prozesses eine besondere wirtschaftliche Bedeutung zu. Eine *effiziente* Optimierung liegt vor, falls das globale Optimum mit einem möglichst geringen Versuchsaufwand identifiziert werden konnte.

In der industriellen Praxis wird zur Optimierung des Drückprozesses hauptsächlich auf die Erfahrung des Drückexperten zurückgegriffen. Dies hat den Nachteil, dass das gesamte Prozesswissen personengebunden ist. Bei systematischem Vorgehen hingegen werden empirische Ansätze, wie die One-Factor-at-a-Time-Methode, zur Optimierung des Drückprozesses eingesetzt. Diese Methode ist in der Regel jedoch sehr ineffektiv und erlaubt zudem nicht, Wechselwirkungen zwischen den Einflussfaktoren zu berücksichtigen. Aus diesem Grund kann das globale Optimum des Prozesses in der Regel nicht identifiziert werden. Eine Analyse und Optimierung des Drückprozesses mit Hilfe statistischer Versuchsplanung ist in der Literatur nicht zu finden.

Da in dieser Arbeit ein problemorientierter Ansatz zur Entwicklung einer effizienten statistischen Optimierungsprozedur gewählt wurde, sind zunächst die Anforderungen an die Optimierungsprozedur zu bestimmen.

- Aufgrund der großen Anzahl an zu berücksichtigenden Zielgrößen sind multivariate Methoden in die Optimierungsprozedur zu integrieren.
- Die komplexen Zusammenhänge zwischen Einflussfaktoren und Zielgrößen erfordern außerdem den Einsatz flexibler Modelle.
- Solche flexiblen Modelle wiederum stellen spezielle Ansprüche an die statistischen Versuchspläne. Zu Beginn einer Optimierung ist in der Regel nicht bekannt, wie komplex die Zusammenhänge sind. Aus diesem Grund sind Versuchspläne sinnvoll, die nicht auf modellbasierten Informationen beruhen, sondern den Versuchsraum möglichst gleichmäßig abdecken.

- Zudem stellen die unbekanntenen Versagensgrenzen zusätzliche Anforderungen an die Versuchspläne. Da Bauteile mit guter Gesamtqualität an der Grenze des Versagensbereichs liegen, ist bei der Optimierung mit einer gewissen Anzahl von fehlenden Beobachtungen zu rechnen. Die verwendeten Versuchspläne müssen also auch mit einer großen Anzahl von fehlenden Werten umgehen können.
- Beim Drückprozess werden unterschiedlichste Bauteilgeometrien oft in geringer Stückzahl produziert. Aus diesem Grund muss die Optimierungsprozedur zudem in der Lage sein, schnell auf die neue Aufgabe zu reagieren. Dies könnte ermöglicht werden, indem Informationen über schon produzierte Bauteile für die Optimierung genutzt werden.
- Da aufgrund der Kleinserienfertigung der Aufwand der Optimierung einen Großteil der Gesamtproduktionskosten ausmacht, ist der Effizienzaspekt von besonders großer Bedeutung. Um die Effizienz der Optimierungsprozedur verbessern zu können, wäre z. B. ein sequentielles Vorgehen zu empfehlen.

Insgesamt muss die Optimierungsprozedur somit eine große Anzahl von wesentlichen Anforderungen erfüllen. In der Literatur gibt es keine Ansätze, die diese Summe von Anforderungen erfüllen. Bei der klassischen statistischen Versuchsplanung z. B. kann schon eine geringe Anzahl von fehlenden Beobachtungen dazu führen, dass der komplette Versuchsplan nicht mehr auswertbar ist. Zum anderen sind Polynome zweiter Ordnung, die zur Modellierung eingesetzt werden, nicht ausreichend, um die komplexen Zusammenhänge zu beschreiben. Des Weiteren ermöglicht die klassische statistische Versuchsplanung es nicht, Informationen über schon produzierte Bauteile direkt für die Optimierung zu nutzen. Diese Probleme, die sich beim Einsatz der Methoden der klassischen Versuchsplanung ergeben, sind ausführlich anhand eines konkreten Optimierungsproblems in Göbel et al. (2004) beschrieben.

2.3 Eine adaptive sequentielle Optimierungsprozedur

Die adaptive sequentielle Optimierungsprozedur (ASOP) unterteilt die Optimierungsaufgabe in vier Schritte, vgl. Abbildung 2.9. Ausgehend von der Fertigungsaufgabe wird im ersten Schritt mittels einer wissensbasierten Vorhersage ein geeigneter Startbereich im Parameterraum bestimmt. Zu diesem Zweck wird die Methode des fallbasierten Schließens auf den Drückprozess adaptiert, vgl. Kapitel 3. Diese Methode greift auf eine Falldatenbank zu, in der die Ergebnisse schon produzierter Bauteile gespeichert sind. Beim Drückprozess lässt sich eine Datenbank mit unterschiedlichen Bauteilgeometrien relativ schnell aufbauen, da aufgrund der Kleinserienfertigung die Bauteilgeometrie häufig wechselt. Für neue Optimierungsaufgaben werden ähnliche Bauteile aus der Datenbank gefiltert und das dort gewonnene Wissen auf die aktuelle Fertigungsaufgabe übertragen. Für die Bauteilreihenfertigung wurde die Methode des fallbasierten Schließens zudem erweitert, so dass sie es ermöglicht, die Prozesseigenschaften zu interpolieren. Auf diese Weise kann z. B. der komplette Versagensbereich für neue Bauteilgeometrien vorhergesagt werden.

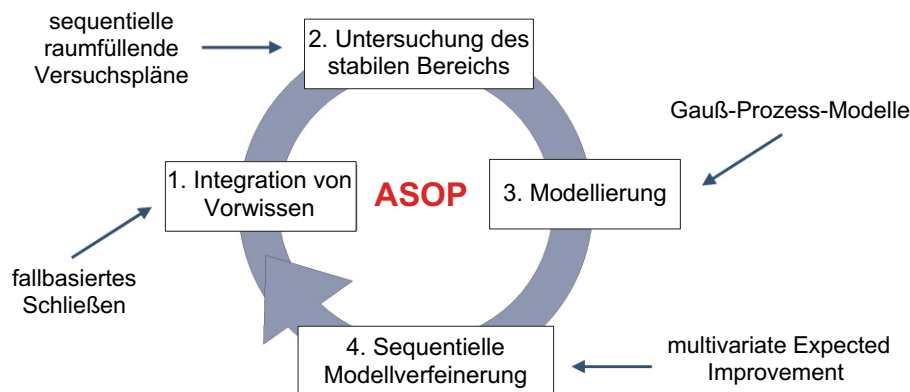


Abbildung 2.9: Prinzip der adaptiven sequentiellen Optimierungsprozedur

Im zweiten Schritt ist ein geeigneter Versuchsplan zu generieren. Da die Form und das Ausmaß der Komplexität der Zusammenhänge zwischen den Einflussfaktoren und den Zielgrößen zu Beginn der Untersuchungen nicht bekannt sind, wur-

den raumfüllende Versuchspläne verwendet. Ein Kriterium zur Generierung eines raumfüllenden Versuchsplans wurde in einem neuentwickelten Algorithmus implementiert, der sequentiell vorgeht. Dabei werden die Informationen über die Lage von Versuchspunkten, die zum Bauteilversagen führten, genutzt, um Teile des Versagensbereichs zu identifizieren. Durch das sequentielle Vorgehen wird verhindert, dass neue Versuchspunkte in schon bekannte Bereiche des Versagensraums platziert werden. Dieses Vorgehen bewirkt eine Konzentration der Versuchspunkte auf den unbekannt stabilen Bereich, also den Bereich, in dem kein Bauteilversagen auftritt, vgl. Kapitel 4.

Um im dritten Schritt der ASOP die Zusammenhänge zwischen Einflussfaktoren und Zielgrößen beschreiben zu können, werden Gauß-Prozess-Modelle an die Daten angepasst, vgl. Kapitel 5. Diese Modelle zeichnen sich durch eine große Flexibilität aus und sie erlauben komplexe Zusammenhänge mit relativ geringem Versuchsaufwand verlässlich zu approximieren.

Im vierten Schritt wird ein sequentielles Optimierungsverfahren eingesetzt. Dazu wurde ein univariater Ansatz basierend auf der Expected Improvement mit Hilfe von Wünschbarkeitsfunktionen auf den multivariaten Fall verallgemeinert, vgl. Kapitel 6. Im Gegensatz zu anderen sequentiellen Optimierungsverfahren wird bei der Wahl neuer Versuchspunkte nicht nur die Vorhersage, sondern auch die Unsicherheit des Modells berücksichtigt. Dabei kann der Anwender die Gewichtung zwischen der lokalen und globalen Suche steuern.

Die bei der Optimierung des Bauteils gewonnenen Informationen werden abschließend in die Datenbank eingespeist. Auf diese Weise können diese Informationen für weitere Optimierungsaufgaben genutzt werden. Durch dieses Vorgehen findet eine ständige Vergrößerung der Datenbank statt, was die Effizienz der ASOP wiederum steigert. Die adaptive sequentielle Optimierungsprozedur lässt sich somit schematisch als ein geschlossener Kreislauf darstellen, vgl. Abbildung 2.9. Zu erwähnen bleibt, dass die Anwendung sequentieller Methoden im Rahmen der ASOP hier unproblematisch ist, da die Reproduzierbarkeit der Ergebnisse des Drückprozesses sehr gut ist. In den folgenden Kapiteln werden die einzelnen Bausteine der ASOP im Detail vorgestellt.

Kapitel 3

Fallbasiertes Schließen

3.1 Prinzip des fallbasierten Schließens

Das Fallbasierte Vorgehen ist eine Methode aus dem Bereich der künstlichen Intelligenz, vgl. Lenz et al. (1991) und Bergmann et al. (1999). Das Vorgehen des fallbasierten Schließens ähnelt stark dem menschlichem Vorgehen beim Lösen von Problemen. Ähnliche Erfahrungen werden herangezogen und auf das neue Problem übertragen. Ausgangsbasis ist eine Datenbank, in der schon produzierte Bauteile gespeichert sind. Das fallbasierte Schließen extrahiert Bauteile aus dieser Datenbank, deren Geometrie zu der neuen Fertigungsaufgabe möglichst ähnlich sind. Die zugehörige Einstellung der Maschinenparameter wird dann auf das neue Problem adaptiert und angewandt.

Diese Methode beruht auf zwei grundsätzlichen Annahmen. Zum einen wird gefordert, dass ähnliche Probleme ähnliche Lösungen haben. Somit wird vorausgesetzt, dass bei vergleichbaren Problemstellungen nicht grundsätzlich andere Lösungsansätze benötigt werden. Um auch in komplexen Situationen schnell Problemlösungen präsentieren zu können, sollte sich zum anderen die Fertigungsaufgabe systematisieren lassen. Für den Druckprozess lässt sich die Fertigungsaufgabe beschreiben durch das Tripel Geometrie, Werkstoff und Qualitätsanforderungen. Ein allgemeiner Ansatz zur Systematisierung von Bauteilgeometrien wird z. B. in Kawai et al. (1989) gegeben. Eine vergleichbare Strukturierung kann eben-

falls für den Bauteilwerkstoff und die Qualitätsmerkmale vorgenommen werden. Ein Fall in der Datenbank beinhaltet neben der Beschreibung der Fertigungsaufgabe und der gewählten Parametereinstellungen auch die beobachteten Zielgrößen sowie die Beurteilung der Gesamtqualität des Bauteils. Um die Effizienz des Verfahrens zu steigern, sollten dabei nicht nur Bauteile mit guter Gesamtqualität berücksichtigt werden. Werden z. B. Versagensbauteile integriert, kann eine Einschränkung des Parameterraums erfolgen.

Der Grundzyklus des fallbasierten Schließens lässt sich in vier Schritte unterteilen und ist in Abbildung 3.1 dargestellt. Im ersten Schritt findet die Fallselektion statt. Aus der Falldatenbank werden zunächst Bauteile extrahiert, deren Fertigungsaufgabe ähnlich zur neuen Fertigungsaufgabe ist. Zentraler Aspekt ist dabei die Definition eines geeigneten Ähnlichkeitsmaßes zur quantitativen Beurteilung der Vergleichbarkeit zweier Fertigungsaufgaben. Die Berechnung der Ähnlichkeit erfolgt in der Regel in zwei Schritten. Zuerst wird für jeden Faktor, der die Bauteilgeometrie beschreibt, eine lokale Ähnlichkeit bestimmt, die im Intervall $[0,1]$ liegt. Je größer dieser Wert ist, desto größer ist die lokale Ähnlichkeit. Diese lokalen Ähnlichkeiten werden anschließend zu einer globalen Ähnlichkeit zusammengefasst. Die Beschreibung der lokalen Ähnlichkeit hängt vom vorliegenden Datentyp ab. Für stetige Faktoren aus \mathbb{R}^+ haben sich Ähnlichkeitsmaße etabliert, die über den relativen Abstand im Wertebereich des Faktors definiert sind. Ein Beispiel für solch ein Ähnlichkeitsmaß ist gegeben durch

$$\text{sim}(x_1, x_2) = 1 - \frac{|x_1 - x_2|}{\max(x_1, x_2)}, \quad (3.1)$$

wobei x_1 und x_2 zwei Einstellungen eines Faktors sind.

Für den Drückprozess wurden lokale Ähnlichkeitsmaße verwendet, die analog zur Wünschbarkeitsfunktion definiert wurden, vgl. Ewers (2005). Bei qualitativen Einflussfaktoren erfolgt die Definition eines Ähnlichkeitsmaßes über einen Zuordnungsvektor oder eine Zuordnungsmatrix. Für jede Ausprägungskombination lässt sich die lokale Ähnlichkeit dann direkt ablesen. Die Erstellung solch einer Matrix erfordert jedoch ein fundiertes Hintergrundwissen über den Prozess. Für die Berechnung der globalen Ähnlichkeit wird häufig die gewichtete Summe der lokalen Ähnlichkeiten verwendet. Alternativ lässt sich aber auch ein gewichtetes

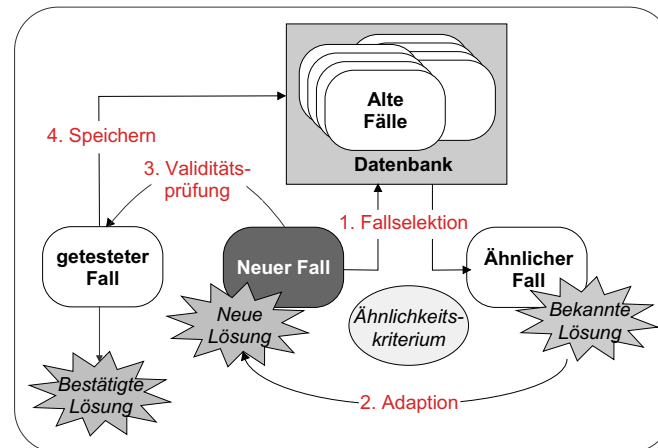


Abbildung 3.1: Schema des fallbasierten Schließens

geometrisches Mittel einsetzen. Im Gegensatz zur gewichteten Summe wird das geometrische Mittel Null, sobald eine der lokalen Ähnlichkeiten Null ist. Je nach Prozesseigenschaften und Einflussstärke des Faktors kann dies eine realistischere Annahme sein.

In der zweiten Phase des fallbasierten Schließens findet die Adaption der ausgewählten Fälle auf die aktuelle Fertigungsaufgabe statt. Neben der Definition der Ähnlichkeitsmaße ist die Erstellung eines leistungsfähigen Adaptionsmodells die Hauptschwierigkeit bei der Anwendung des fallbasierten Schließens. In Beierle und Kern-Isberner (2000) werden verschiedene Methoden zur Falladaption vorgestellt. Bei der Substitutionsmethode werden z. B. Teile der alten Lösung so angepasst, dass sie besser zur neuen Fertigungsaufgabe passen. Für industrielle Produktionsprozesse bedeutet dies in der Regel eine Anpassung der Maschinenparameter.

Bevor der adaptierte Lösungsvorschlag durchgeführt wird, erfolgt in der dritten Phase des fallbasierten Schließens zunächst eine Validitätsprüfung. Dabei wird überprüft, ob die Lösung in einer ähnlichen Situation bereits fehlgeschlagen ist. Anhand des adaptierten Lösungsvorschlags wird anschließend ein Bauteil gefertigt. Die Beurteilung der Gesamtqualität erfolgt mit Hilfe des Wünschbarkeitsindex.

Der komplette Fall wird im vierten Schritt der Datenbank hinzugefügt.

Die Leistungsfähigkeit der Methode des fallbasierten Schließens hängt von der Zahl der in der Datenbank gespeicherten Fälle ab. Eine große Fallbasis erfordert jedoch den Einsatz effizienter Algorithmen und eine gut strukturierte Organisation der Datenbank. Neben der Fallzahl spielt aber auch der Informationsgehalt der einzelnen Fälle eine Rolle. Für den Drückprozess haben diejenigen Bauteile einen besonders hohen Informationsgehalt, die an der Grenze zum Bauteilversagen produziert wurden. Neben der Gesamtqualität geben diese Bauteile zusätzlich noch Auskunft über den Verlauf der unbekanntens Versagensgrenze.

Das fallbasierte Schließen konnte bereits erfolgreich zur Optimierung des Drückprozesses genutzt werden, vgl. Ewers (2005).

3.2 Erweiterung des fallbasierten Schließens

Mit Hilfe des Drückprozesses werden häufig Bauteilreihen gefertigt. Eine Bauteilreihenfertigung zeichnet sich dadurch aus, dass sich die Geometrie der Bauteile nicht stark verändert. Oft werden sogar nur einzelne Geometrieaspekte geändert, wie z. B. die Höhe des Bauteils. In diesem Fall kann das Wissen aus der Falldatenbank dazu genutzt werden, das Prozessverhalten zu interpolieren und es damit für neue Bauteile mit ähnlicher Geometrie vorherzusagen. Zu diesem Zweck wird im Folgenden ein geeigneter Ansatz vorgestellt. Wesentliche Voraussetzung ist auch hier wieder die Forderung, dass ähnliche Bauteile ein ähnliches Prozessverhalten zeigen.

Ein erster Ansatz zur Interpolation des stabilen Bereichs wurde in Langewiesche (2005) vorgestellt. Dabei wurde der stabile Bereich durch n -dimensionale Ellipsen beschrieben und die Veränderung der Ellipsen in Abhängigkeit der interessierenden Geometriecharakteristik modelliert. Eine Anpassung der n -dimensionalen Ellipsen erfolgte dabei durch die Methode der Hauptkomponentenanalyse. Obwohl mit diesem Ansatz erste Erfolge erzielt werden konnten, traten bei diesem Vorgehen jedoch einige schwerwiegende Probleme auf.

Stattdessen ist ein alternativer Ansatz zu empfehlen, der auf Hyper-Modellen beruht. Die Idee dieser Hyper-Modelle ist, die Geometriecharakteristiken, durch

die sich die Bauteile aus der Falldatenbank von der zu fertigenden Geometrie unterscheiden, als zusätzliche Faktoren mit in das Modell aufzunehmen. Auf diese Weise lässt sich die Veränderung der Zielgrößen abhängig von der Bauteilgeometrie approximieren.

Da beim Drückprozess ein großer Anteil des Versuchsumfangs für die Bestimmung der unbekanntes Versagensgrenzen benötigt wird, konzentrieren wir uns im Folgenden auf Hyper-Modelle, die es ermöglichen, den stabilen Bereich für neue Bauteilgeometrien vorherzusagen. Dazu wird ausgenutzt, dass jedes Bauteil aus der Datenbank einer der drei Kategorien "kein Bauteilversagen", "Grenzfall" und "Bauteilversagen" zugeordnet wurde. Bei Bauteilen der mittleren Kategorie wurde im Fertigungsprozess eine leichte Faltenbildung beobachtet, die aber im Verlauf der Fertigung wieder ausgeglichen werden konnte. Mit Hilfe dieser Information wird der Zusammenhang zwischen den Faktoren und dem Bauteilversagen nun durch ein Hyper-Proportional-Odds-Modell beschrieben. Das Proportional-Odds-Modell stellt eine Erweiterung der logistischen Regression auf mehr als zwei Kategorien dar, vgl. Agresti (1990). Im Rahmen des SFB Projekts konnte mit Hilfe von Experimenten bestätigt werden, dass das Hyper-Proportional-Odds-Modell die Veränderung der Lage und der Form des stabilen Bereichs verlässlich interpoliert. Neben der Vorhersage des stabilen Bereichs kann dieser Ansatz aber auch genutzt werden, um das Prozessverhalten bezüglich anderer relevanter Zielgrößen zu interpolieren. Zur Interpolation des Prozessverhaltens stetiger Zielgrößen, die z. B. die Abweichung von der Sollgeometrie beschreiben, ist jedoch das Hyper-Gauß-Prozess-Modell vorzuziehen, vgl. Kapitel 5.

Kapitel 4

Raumfüllende Versuchspläne bei unbekannten Restriktionen

4.1 Raumfüllende Versuchspläne

Raumfüllende Versuchspläne wurden für die Analyse von Computerexperimenten entwickelt, deren Einsatz in den vergangenen Jahren stark zugenommen hat. Grund dafür ist die schnelle Entwicklung im Bereich der Computertechnologie, welche es z. B. ermöglicht, selbst komplexe physikalische Prozesse sehr gut durch Computerprogramme zu simulieren. Innerhalb dieser Programme werden häufig Finite-Elemente-Methoden verwendet. Da die Durchführung von realen Experimenten sehr kostenintensiv sein kann, werden bevorzugt Computersimulationen der Prozesse eingesetzt. Anwendung finden Computersimulationen realer Prozesse z. B. in der Automobilindustrie, Luft- und Raumfahrttechnik, Maschinenbau, Geophysik und Medizintechnik. Konkrete Beispiele solcher Computersimulationen werden z. B. in Booker et al. (1999), Berk et al. (2002) und Santner et al. (2003) gegeben. Jedoch sind Computersimulationen aufgrund der Komplexität der abgebildeten Prozesse sehr rechenintensiv. Die Ermittlung der Zielgrößen eines Versuches kann dabei mehrere Stunden bis Tage in Anspruch nehmen. Dies ist insbesondere für die Optimierung eines Prozesses nicht tolerierbar. Als Ausweg haben Sacks et al. (1989) die statistische Modellierung von Computerexperimenten

ten vorgeschlagen. Da in diesem Fall ein statistisches Modell eines Computermodells erstellt wird, spricht man auch von Meta-Modellierung, vgl. Kleijnen (2005). Das Ziel bei der statistischen Modellierung ist es, mit möglichst wenigen Versuchen die Zusammenhänge möglichst gut abzubilden, um dann an unbeobachteten Versuchspunkten gute Vorhersagen tätigen zu können. Im Rahmen dieser statistischen Modellierung hat sich der Einsatz von Gauß-Prozess-Modellen bewährt, da diese Modelle flexibel sind und auch komplexe Zusammenhänge approximieren können. Von besonderer Bedeutung bei der Planung der Versuche ist die Tatsache, dass Computereperimente oft deterministisch sind. Aus diesem Grund greifen die klassischen Prinzipien der Versuchsplanung - Randomisation, Blockbildung und Wiederholungsmessungen - hier nicht. Stattdessen sollte beachtet werden, dass für die Gauß-Prozess-Modelle der Vorhersagefehler für Punkte im Versuchsraum vom Abstand zu den Versuchspunkten abhängt. Aus diesem Grund ist es sinnvoll, Versuchspunkte raumfüllend in den Versuchsraum zu legen.

In der Literatur wurden verschiedene Kriterien zur Beurteilung der raumfüllenden Eigenschaft von Versuchsplänen vorgeschlagen. Dabei haben sich die folgenden drei Klassen von Kriterien etabliert:

- Distanzbasierte Kriterien
- Kriterien, welche auf Sampling Methoden beruhen
- Kriterien, welche die Gleichverteilung der Versuchspunkte beurteilen

Im Folgenden werden die raumfüllenden Versuchspläne, die auf den hier genannten Kriterien beruhen, und deren Eigenschaften vorgestellt.

4.1.1 Distanzbasierte Versuchspläne

Distanzbasierte Versuchspläne verwenden Abstandsmaße, um die raumfüllende Eigenschaft der Versuchspunkte beurteilen zu können. Dabei wird in der Regel der euklidische Abstand δ verwendet. Sei \mathcal{X} eine kompakte Teilmenge des \mathbb{R}^d und $\mathbf{x}_1, \mathbf{x}_2 \in \mathcal{X}$ zwei Punkte im Versuchsraum. Der euklidische Abstand ist definiert durch

$$\delta(\mathbf{x}_1, \mathbf{x}_2) := \sqrt{\sum_{k=1}^d (x_{1k} - x_{2k})^2}.$$

Die bekanntesten Vertreter distanzbasierter Versuchspläne sind Minimax und Maximin Designs, vgl. Johnson et al. (1990). Sei $\mathcal{D} = \{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n\} \subset \mathcal{X}$ ein beliebiger Versuchsplan vom Umfang n . Ein Maximin Design \mathcal{D}_{Mm} maximiert den minimalen Abstand für je zwei Versuchspunkte unter allen Versuchsplänen $\mathcal{D} \subset \mathcal{X}$ vom Umfang n . Auf diese Weise wird sichergestellt, dass die Versuchspunkte nicht zu nah beieinander liegen. Formal lässt sich dies wie folgt ausdrücken.

Definition 4.1.1 *Ein Versuchsplan $\mathcal{D}_{Mm} \subset \mathcal{X}$ ist ein Maximin Design, falls*

$$\delta_{Mm} := \min_{\substack{\mathbf{x}_1, \mathbf{x}_2 \in \mathcal{D}_{Mm} \\ \mathbf{x}_1 \neq \mathbf{x}_2}} \delta(\mathbf{x}_1, \mathbf{x}_2) = \max_{\mathcal{D} \subset \mathcal{X}} \min_{\substack{\mathbf{x}_1, \mathbf{x}_2 \in \mathcal{D} \\ \mathbf{x}_1 \neq \mathbf{x}_2}} \delta(\mathbf{x}_1, \mathbf{x}_2). \quad (4.1)$$

Das Konstruktionsprinzip von Maximin Designs führt dazu, dass sich Kreise mit Radius $r = \delta_{Mm}/2$, die um die Versuchspunkte gelegt werden, nicht schneiden.

Zu beachten ist, dass Maximin Designs in der Regel nicht eindeutig sind. Das klassische Vorgehen in der Praxis ist es, sich zufällig für eines der Maximin Designs zu entscheiden. Alternativ kann ein zweites Entscheidungskriterium eingeführt werden. Beispielsweise kann für jedes Maximin Design der Index I_{Mm} bestimmt werden, der die Anzahl der Versuchspunktpaare angibt, deren Abstand zueinander δ_{Mm} ist. Unter der Menge der Maximin Designs wird dann das Design mit kleinstem Index I_{Mm} gewählt, vgl. Johnson et al. (1990). Dieses zweite Entscheidungskriterium garantiert aber immer noch keine Eindeutigkeit des Versuchsplans.

Während beim Maximin Design nur die Anordnung der Versuchspunkte untereinander betrachtet wird, ist beim Minimax Design die Lage der Versuchspunkte

zu beliebigen Punkten aus dem Versuchsraum \mathcal{X} von Interesse. Das Minimax Design platziert die Versuchspunkte so, dass jeder Punkt $\mathbf{x} \in \mathcal{X}$ möglichst nahe an einem Versuchspunkt liegt. Unter der realistischen Annahme, dass der absolute Vorhersagefehler der Zielgröße im Versuchsraum proportional vom Abstand zum nächsten beobachteten Versuchspunkt abhängt, minimiert ein Minimax Design intuitiv den absoluten Vorhersagefehler.

Definition 4.1.2 Sei $\delta(\mathbf{x}, \mathcal{D}) = \min_{\mathbf{x}_i \in \mathcal{D}} \delta(\mathbf{x}, \mathbf{x}_i)$. Ein Versuchsplan $\mathcal{D}_{mM} \subset \mathcal{X}$ ist ein Minimax Design, falls

$$\delta_{mM} = \max_{\mathbf{x} \in \mathcal{X}} \delta(\mathbf{x}, \mathcal{D}_{mM}) = \min_{\mathcal{D} \subset \mathcal{X}} \max_{\mathbf{x} \in \mathcal{X}} \delta(\mathbf{x}, \mathcal{D}). \quad (4.2)$$

Auch Minimax Designs sind in der Regel nicht eindeutig. Als zusätzliches Entscheidungskriterium bietet sich hier an, für jedes Minimax Design den Index I_{mM} zu bestimmen, der die Anzahl der Versuchspunkte zählt, für die ein $\mathbf{x} \in \mathcal{X}$ mit maximalen Abstand δ_{mM} zum Versuchsplan existiert. Auch hier liefert die zusätzliche Berücksichtigung des zweiten Entscheidungskriteriums noch keine eindeutigen Versuchspläne.

Johnson et al. (1990) konnten nachweisen, dass bei Verwendung von Gauß-Prozess-Modellen mit schnell fallender Kovarianzfunktion Minimax und Maximin Designs asymptotisch optimal sind. Das Minimax Design minimiert die maximale Vorhersagevarianz des Prozesses, während das Maximin Design die Kovarianzmatrix des Gauß-Prozesses maximiert. Diese Eigenschaften sind vergleichbar mit der G-Optimalität bzw. D-Optimalität aus der klassischen Versuchsplanung.

Ein Vergleich der beiden Versuchspläne zeigt, dass Minimax Designs insbesondere für kleine Versuchsumfänge eine bessere raumfüllende Eigenschaft besitzen als Maximin Designs. Da beim Maximin-Kriterium der minimale Abstand der Versuchspunkte untereinander maximiert wird, tendieren diese Versuchspläne dazu, bei kleinem n die Versuchspunkte an den Rand des Versuchsbereichs zu legen. Dies führt zu einer schlechten Modellanpassung innerhalb des Versuchsraums. Um diesen Sachverhalt zu verdeutlichen, ist in Abbildung 4.1 a) ein Maximin Design auf $[0, 1]^2$ mit $n = 7$ Punkten eingezeichnet. Insgesamt 6 der 7 Versuchspunkte liegen am Rand des Versuchsbereichs. Im Gegensatz dazu liegen beim Minimax Design nur 2 Versuchspunkte direkt am Versuchsrand, vgl. Abbildung 4.1 b).

Trotz der besseren raumfüllenden Eigenschaft von Minimax Designs bei kleinem Versuchsumfang werden Maximin Designs in der Praxis weitaus häufiger eingesetzt. Der Grund dafür ist, dass diese Versuchspläne deutlich einfacher zu generieren sind.

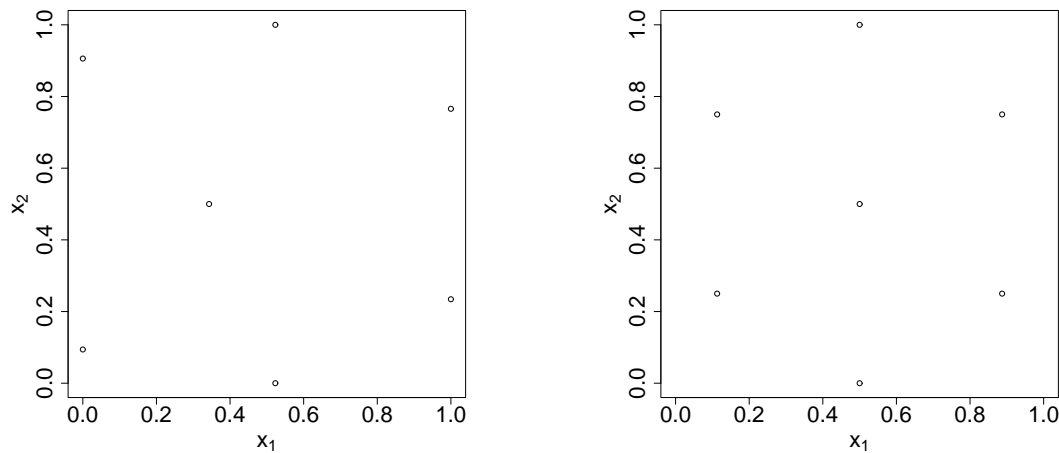


Abbildung 4.1: a) Maximin Design auf $[0, 1]^2$ mit $n = 7$, b) Minimax Design auf $[0, 1]^2$ mit $n = 7$

4.1.2 Latin Hypercube Designs

Latin Hypercube Designs (LHDs) wurden von McKay et al. (1979) eingeführt und zählen zu den populärsten raumfüllenden Versuchsplänen. Gründe hierfür sind ihre einfache Erzeugung, auch bei einer hohen Anzahl von Faktoren, sowie ihre gute Projektionseigenschaft. Diese Eigenschaft garantiert, dass eine beliebige Projektion der Versuchspunkte eines LHD auf eine der Koordinatenachsen sich durch eine äquidistante Verteilung der Punkte auszeichnet.

Zur Generierung eines Latin Hypercube Designs (LHD) mit n Versuchspunkten für d Faktoren wird der Wertebereich für jeden der d Faktoren in n gleich große Intervalle geteilt. Die Menge aller möglichen kartesischen Produkte dieser Intervalle unterteilt den Versuchsraum in n^d d -dimensionale Zellen, auch Hypercubes genannt. Die Bestimmung der n Versuchspunkte des LHD kann nun als Urnenexperiment aufgefasst werden. Dabei werden d Urnen, die alle die Zahlen $1, \dots, n$ enthalten, verwendet. Aus jeder dieser d Urnen wird nun n -mal ohne Zurück-

legen gezogen. Die i -te gezogene Zahlenkombination definiert die i -te Zelle des Versuchsplans. Die Versuchspunkte sind durch die Mittelpunkte der ausgewählten Zellen festgelegt. Alternativ können die Versuchspunkte auch zufällig innerhalb einer Zelle ausgewählt werden. Ein zweidimensionales Beispiel solch eines Latin Hypercube Designs mit $n = 7$ Versuchspunkten ist in Abbildung 4.2 gegeben. Zu beachten ist, dass LHDs existieren, denen intuitiv keine gute raumfüllende Eigenschaft zugeordnet wird. Beispielsweise kann ein LHD aus Punkten bestehen, die ausschließlich auf einer der Diagonalachsen des Versuchsraums liegen.

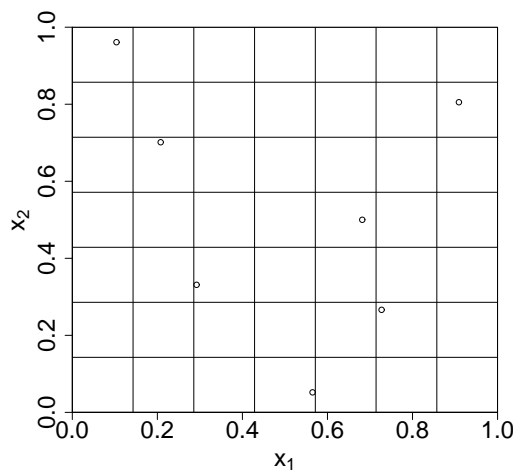


Abbildung 4.2: Latin Hypercube Design auf $[0, 1]^2$ mit $n = 7$

Eine Erweiterung der LHDs, die so genannten Randomised Orthogonal Arrays (Wu und Hamada, 2000), erlauben es, die Projektionseigenschaft auf höhere Dimensionen zu erweitern. Dies ist von großem Interesse, da zu Beginn eines Experimentes oft nicht bekannt ist, welche der Faktoren einen Einfluss auf die Zielgröße besitzen. Da aufgrund der Factorsparsity in der Regel nur wenige Faktoren tatsächlich aktiv sind, kann mit Hilfe der Projektionseigenschaft garantiert werden, dass die Versuchspunkte im Unterraum der aktiven Faktoren ebenfalls gute raumfüllende Eigenschaften aufweisen. Nachteil dieser Versuchspläne ist jedoch, dass sie nur für bestimmte Versuchsumfänge n existieren. Eine andere Erweiterung von LHDs sind die cascading LHDs. Sie wurden von Handcock (1991) vorgestellt und werden wie folgt erzeugt: Ausgehend von einem LHD vom Umfang n_1 mit Versuchspunkten im Mittelpunkt der Zellen wird um jeden dieser Versuchspunkte ein kleiner Bereich betrachtet, in dem ein zusätzliches LHD vom Umfang

n_2 generiert wird. Dieser Bereich kann z. B. gleich der jeweiligen Zelle gewählt werden. In jeder ausgewählten Zelle würde dann mit Hilfe des Latin Hypercube Samplings noch zusätzlich n_2 Versuchspunkte platziert werden. Auf diese Weise werden raumfüllend Cluster von LHDs erzeugt. Vorteil dieser cascading LHDs ist, dass sie es ermöglichen, sowohl die lokale als auch globale Korrelationsstruktur für das Gauß-Prozess-Modell zu untersuchen. Nachteil ist jedoch der benötigte Gesamtversuchsumfang, der schnell auf nicht mehr vertretbare Größen wächst.

4.1.3 Uniform Coverage Designs

Ähnlich wie bei der Generierung der LHDs wird bei Uniform Coverage Designs unterstellt, dass der Versuchsbereich einem d -dimensionalen Hyper-Rechteck $\mathcal{X} = \times_{i=1}^d [a_i, b_i]$ entspricht. Sei $\mathcal{D} = \{\mathbf{x}_1, \dots, \mathbf{x}_n\}$ ein beliebiger Versuchsplan. Um zu verdeutlichen, dass $\mathbf{x} \in \mathcal{X}$ nun als Zufallsvariable betrachtet wird, wird die Notation \mathbf{X} verwendet. Es gelte, dass $\mathbf{X} \sim F(\cdot)$, wobei

$$F(\mathbf{x}) = \prod_{i=1}^d \left(\frac{x_i - a_i}{b_i - a_i} \right)$$

die Verteilungsfunktion der Gleichverteilung auf \mathcal{X} ist. Betrachte zudem die empirische Verteilungsfunktion der Punkte in \mathcal{D} . Sei $\mathbf{x}_0 = (x_{01}, \dots, x_{0d})$ ein fester Punkt im Versuchsraum, dann gilt

$$F_n(\mathbf{x}_0) = \frac{1}{n} \sum_{i=1}^n I \{X_{i1} \leq x_{01}, \dots, X_{id} \leq x_{0d}\}, \quad (4.3)$$

wobei $I\{\cdot\}$ die Indikatorfunktion bezeichnet.

Der Abstand $|F_n(\mathbf{x}) - F(\mathbf{x})|$ wird nun als Diskrepanz bezeichnet. Um beurteilen zu können, wie stark die Versuchspunkte aus \mathcal{D} von einer Gleichverteilung der Punkte abweichen, schlagen Fang et al. (2000) das folgende Diskrepanzmaß vor

$$D(\mathcal{D}) = \sup_{\mathbf{x} \in \mathcal{X}} |F_n(\mathbf{x}) - F(\mathbf{x})|. \quad (4.4)$$

Dieses Diskrepanzmaß entspricht der Kolmogorov-Smirnov-Statistik des Tests auf Gleichverteilung. Die Versuchspunkte eines Uniform Coverage Designs minimieren dieses Diskrepanzmaß.

Ähnlich zum Latin Hypercube Design können mit dem obigen Diskrepanzkriterium Uniform Coverage Designs erzeugt werden, die intuitiv nicht als raumfüllend angesehen werden würden, vgl. Santner et al. (2003). Ein weiterer Nachteil dieser Versuchspläne ist ihre aufwändige Generierung. Ein oft praktizierter Ausweg ist es, den Wertebereich des Versuchsraums zu diskretisieren. Als Konsequenz können dann in der Regel nur Versuchspläne erzeugt werden, die annähernd gleichverteilt sind. Solche Versuchspläne werden als Near Uniform Coverage Designs bezeichnet.

4.1.4 Kombinierte raumfüllende Versuchspläne

In den vorherigen Kapiteln wurden die bekanntesten raumfüllenden Versuchspläne vorgestellt. Dabei zeigte sich, dass keiner dieser Versuchspläne an sich überzeugt. Maximin Designs und Latin Hypercube Designs werden aufgrund ihrer einfachen Generierung in der Praxis am häufigsten verwendet. Bei beiden Versuchsplänen können jedoch Realisierungen auftreten, denen intuitiv eine schlechte raumfüllende Eigenschaft zugeordnet wird. Auch Uniform Coverage Designs garantieren keine gute raumfüllende Eigenschaft. Zudem sind diese Versuchspläne schwer zu generieren. Minimax Designs zeichnen sich stets durch eine gute raumfüllende Verteilung der Versuchspunkte aus, werden in der Praxis allerdings kaum eingesetzt, da auch sie schwer zu erzeugen sind.

Da jeder dieser Versuchspläne gewisse Nachteile besitzt, wurden in der Literatur kombinierte Kriterien vorgeschlagen. Dabei wird ein erstes raumfüllendes Kriterium verwendet, um die Menge der möglichen raumfüllenden Versuchspläne auf eine endliche Menge zu begrenzen. Im zweiten Schritt wird dann ein zweites Kriterium herangezogen, um ein Design aus dieser beschränkten Menge zu bestimmen. Morris and Mitchell (1995) haben einen kombinierten raumfüllenden Versuchsplan vorgeschlagen, der leicht zu erzeugen ist. Dabei wird zunächst für das gewünschte n die Menge aller LHDs mit Versuchspunkten im Mittelpunkt der ausgewählten Zellen bestimmt und anschließend das LHD gewählt, das den minimalen Abstand zweier Versuchspunkte maximiert. Solch ein Maximin LHD besitzt sehr gute raumfüllende Eigenschaften, da sich durch Kombination der

Kriterien die negativen Eigenschaften beider Versuchspläne eliminieren. Um dies zu verdeutlichen, ist in Abbildung 4.3 ein Maximin LHD vom Umfang $n = 7$ dargestellt.

Im Gegensatz zum einfachen Maximin Design desselben Umfangs wird nun das Innere des Versuchsraums besser untersucht, vgl. Abbildung 4.1. Morris und Mitchell (1995) haben sich ausführlich mit den Eigenschaften der Maximin LHDs beschäftigt. Beispielsweise liegen die Versuchspunkte immer äquidistant auf parallelen Linien, vgl. Abbildung 4.3.

Zusammenfassend lässt sich sagen, dass bei einem Vergleich der hier vorgestellten raumfüllenden Versuchspläne das Maximin LHD am besten abschneidet. Neben der einfachen Generierung garantiert dieses Design sehr gute raumfüllende Eigenschaften.

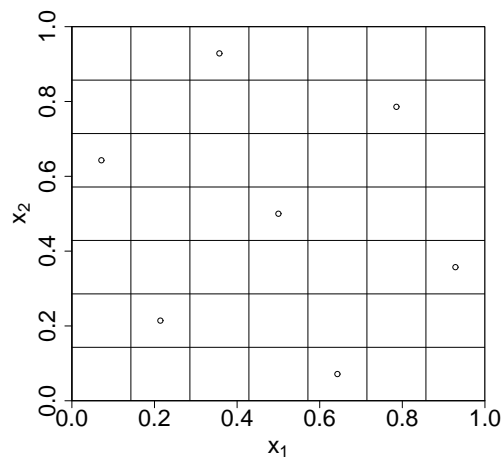


Abbildung 4.3: Maximin Latin Hypercube Design auf $[0, 1]^2$ mit $n = 7$

4.2 Unbekannte konvexe Restriktionen

In der Literatur werden unterschiedliche Ansätze zur Generierung von raumfüllenden Versuchsplänen bei eingeschränktem Versuchsbereich vorgestellt, vgl. z. B. Stinstra et al. (2003) und Trosset (1999). Alle diese Ansätze basieren aber auf der Annahme, dass die Restriktionsgleichungen bekannt sind. Motiviert durch die Eigenschaften des Drückprozesses wird im Folgenden ein Algorithmus zur Generierung von raumfüllenden Versuchsplänen bei unbekanntem Restriktionen des Versuchsraums entwickelt. Angelehnt an den Drückprozess wird der Versuchsbereich außerhalb der Restriktionsgrenzen im Folgenden als Versagensbereich und innerhalb dieser Grenzen als stabiler Bereich bezeichnet. Ziel des Algorithmus ist es, den kompletten unbekanntem stabilen Bereich zu untersuchen und dabei möglichst wenig Versuche zu fahren, die zu Bauteilversagen führen. Um dieses Ziel zu erreichen, generiert der Algorithmus die raumfüllenden Versuchspläne sequentiell. Auf diese Weise können die bereits gewonnenen Informationen berücksichtigt werden.

Der Algorithmus basiert auf zwei Ansätzen, einer schrittweisen Identifikation des Versagensbereichs kombiniert mit einer schrittweisen Anpassung des Versuchsplans in Richtung des stabilen Bereichs.

4.2.1 Schrittweise Identifikation des Versagensbereichs

Das Verfahren zur schrittweisen Identifikation des Versagensbereichs basiert auf einer wesentlichen Annahme. Beim Drückprozess hat sich gezeigt, dass der stabile Bereich konvex ist, vgl. Abbildung 2.8. Diese Eigenschaft wird im Algorithmus ausgenutzt, um Teile des Versagensbereichs zu identifizieren.

Gegeben sei eine Menge von Punkten im d -dimensionalen Versuchsraum mit n_1 Punkten im Versagensbereich und n_2 Punkten im stabilen Bereich. Das Ziel dieses Verfahrens ist nun, basierend auf den $n_1 + n_2$ Versuchspunkten ein Maximum an Information über die Lage des Versagensbereichs zu gewinnen. Die unbekannte Grenze zum Versagensbereich liegt zwischen der Menge der Punkte im stabilen Bereich und der Menge der Punkte im Versagensbereich. Aufgrund der

Konvexität des stabilen Bereichs ist der mögliche Verlauf der Versagensgrenzen eingeschränkt. Dieser Sachverhalt ist in Abbildung 4.4 a) dargestellt.

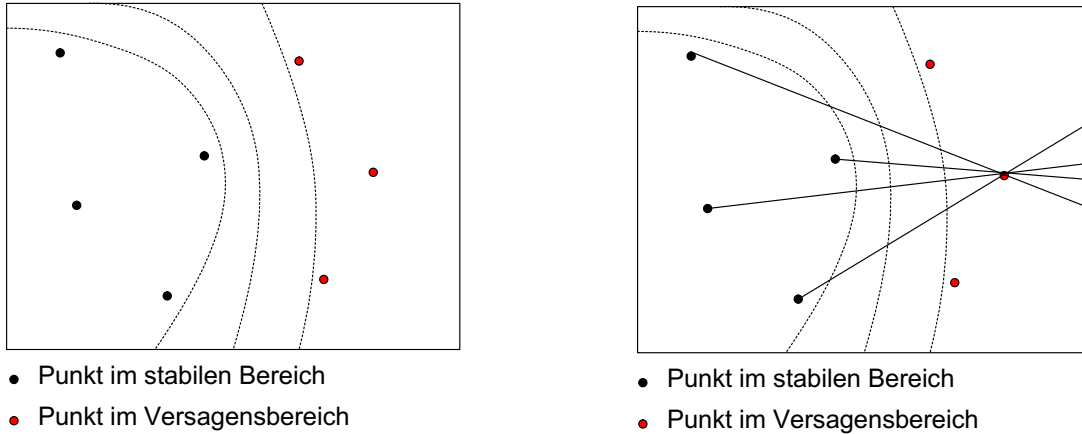


Abbildung 4.4: a) Mögliche Verläufe der unbekanntten konvexen Versagensgrenze, b) Erstellung von Verbindungsvektoren für einen Punkt im Versagensbereich

Ausgehend von einem beliebigen Punkt im Versagensbereich werden nun Verbindungsvektoren zu allen Punkten im stabilen Bereich aufgespannt, vgl. Abbildung 4.4 b). Falls $n_2 \geq d$ kann durch diese Verbindungsvektoren jenseits vom Versagenspunkt eine konvexe Hülle aufgespannt werden. Tritt der Fall ein, dass mehrere Verbindungsvektoren sich überlagern, so liegt eine lineare Abhängigkeit vor. In diesem Fall erhöht sich die benötigte Anzahl an Punkten im stabilen Bereich. Unter der obigen Annahme ist die konvexe Hülle der Strahlen jenseits des Versagenspunktes Teil des Versagensbereichs, vgl. Abbildung 4.5 a). Durch Erzeugen solcher konvexen Hüllen für jeden Versagenspunkt wird ein Teil des Versagensbereichs sicher identifiziert, vgl. Abbildung 4.5 b). Werden nun sequentiell neue Versuchspunkte hinzugefügt, wird verhindert, dass diese Punkte in schon identifizierte Bereiche des Versagensraums platziert werden.

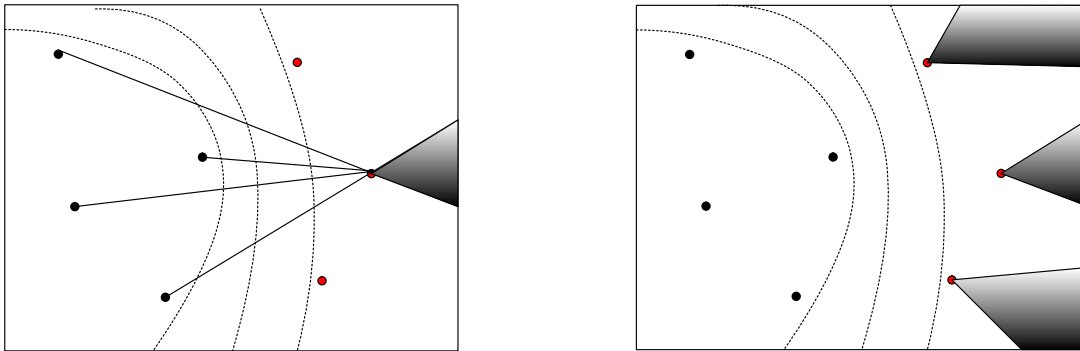


Abbildung 4.5: a) Bestimmung der konvexen Hülle, die durch die Verbindungsvektoren aufgespannt wird, b) Identifikation von Teilen des Versagensbereichs mit Hilfe der konvexen Hüllen

4.2.2 Schrittweise Anpassung des Versuchsbereichs

Zu Beginn des Algorithmus wird ein relativ kleiner Anteil des interessierenden Versuchsraums untersucht. Dieser Versuchsraum wird abhängig von der Lage der Versagenspunkte schrittweise angepasst. In jedem Iterationsschritt erfolgt dazu zunächst eine Verschiebung des Versuchsplans in Richtung des stabilen Bereichs. Der Mittelpunkt des Versuchsplans wird dazu gleichgesetzt mit dem arithmetischen Mittel der Versuchspunkte im stabilen Bereich. Anschließend wird der Versuchsplan vergrößert. Dabei entscheidet der Anwender abhängig von der Lage der Versagenspunkte in welche Richtung und wie stark der Versuchsplan erweitert wird.

4.2.3 Wahl des Kriteriums

Bei der Wahl eines geeigneten Kriteriums ist neben einer guten raumfüllenden Eigenschaft zu beachten, dass dieses Kriterium eine sequentielle Generierung des Versuchsplans ermöglichen muss. Versuchspläne, die in Kapitel 4.1 vorgestellt wurden und eine gute raumfüllende Eigenschaft besitzen, sind Minimax Designs und Maximin LHDs. Für Minimax Designs lassen sich vorgegebene Punkte problemlos in die Generierung eines Versuchsplans integrieren. Eine Methode zur sequentiellen Generierung von Maximin LHDs wird in Wang (2003) vorgestellt.

Da der Versuchsbereich durch Restriktionen eingeschränkt ist, sollte das gewählte Kriterium zusätzlich in der Lage sein, raumfüllende Versuchspläne auch bei nicht kubischen Versuchsräumen zu erstellen. Diese Anforderung ist jedoch ein Ausschlusskriterium für das Maximin LHD, da das Prinzip des Latin Hypercube Designs auf der Unabhängigkeit der Faktoren beruht. Für den Algorithmus wird daher das Minimax-Kriterium verwendet, wobei der Versuchsraum auf eine endliche Punktmenge reduziert wird. Dieses Vorgehen ist in der Regel problemlos, da in der Praxis viele Faktoren oft nicht beliebig genau eingestellt werden können.

Wird der Versuchsplan sequentiell erzeugt, so kann oft schon nach dem ersten Schritt ein räumliches Regressionsmodell angepasst werden. Obwohl solch ein Modell sehr grob ist und nicht zur Prognose verwendet werden soll, können durch die Schätzer der Modellparameter erste Hinweise auf die Relevanz der einzelnen Faktoren gewonnen werden. Dazu sind die Einflussfaktoren auf z. B. $[0, 1]$ zu standardisieren.

Diese Information wird in den weiteren Iterationsschritten bei der Generierung des raumfüllenden Versuchsplans durch ein zweites Entscheidungskriterium berücksichtigt. Dieses zweite Entscheidungskriterium misst mit Hilfe des Minimax-Kriteriums die raumfüllende Eigenschaft auf dem Unterraum der aktiven Faktoren. Durch zusätzlicher Berücksichtigung des zweiten Entscheidungskriteriums wird aus der Menge der Minimax Designs das Design bestimmt, dessen Projektion auf den Unterraum der aktiven Faktoren am besten ist. Im Folgenden bezeichne ein Minimax Design, das das obige zweite Entscheidungskriterium berücksichtigt als erweitertes Minimax Design \mathcal{D}_{mMe} . Um den Unterschied zum einfachen Minimax Design zu verdeutlichen, betrachte als Beispiel den dreidimensionalen Versuchsraum $[0, 1]^3$, bei dem die ersten beiden Faktoren einen deutlich größeren Einfluss auf die Zielgröße besitzen als der dritte Einflussfaktor. In Abbildung 4.6 a) und b) ist die Projektion eines einfachen und eines erweiterten Minimax Designs auf die ersten beiden Dimensionen abgebildet. Offensichtlich ist die raumfüllende Eigenschaft in den ersten beiden Dimensionen des Versuchsraums für das erweiterte Minimax Design deutlich besser.

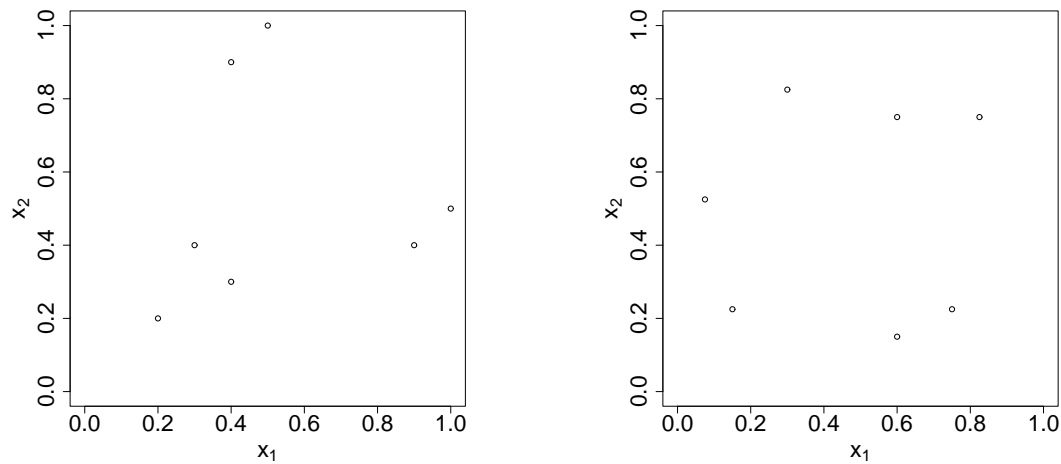


Abbildung 4.6: a) Projektion eines einfachen Minimax Designs auf $[0, 1]^2$, b) Projektion eines erweiterten Minimax Designs auf $[0, 1]^2$

4.2.4 Beschreibung des Algorithmus

Der Algorithmus zur sequentiellen Generierung eines raumfüllenden Versuchsplans bei unbekanntem stabilen Bereich geht nun wie folgt vor:

1. Generierung eines raumfüllenden Versuchsplans
2. Einschränkung des Versuchsraums
3. Verschiebung und Erweiterung des Versuchsplans
4. Hinzufügen von neuen Versuchspunkten
5. Wiederholung von Schritt 2 bis 4 bis Stoppkriterium erfüllt ist.

Im ersten Schritt wird ein kubischer raumfüllender Versuchsplan generiert. Dazu wird das einfache Minimax-Kriterium verwendet. Der Anteil der Fläche am Parameterraum des Versuchsplans sollte relativ klein gehalten werden. Im zweiten Schritt werden mit Hilfe konvexer Hüllen Teile des Versagensbereichs identifiziert und ausgeschlossen, vgl. Abbildung 4.7 a). Anschließend erfolgt im dritten Schritt eine Verschiebung des Versuchsplans in Richtung der stabilen Punkte. Neuer Mittelpunkt des zu untersuchenden Versuchsraums wird der Mittelpunkt aller Punkte im stabilen Bereich. Zudem erfolgt eine Vergrößerung des zu untersuchenden Versuchsbereichs, wie in Kapitel 4.2.2 beschrieben. Im vierten Schritt

werden in den so erweiterten und verschobenen Versuchsraum neue Versuchspunkte hinzugefügt, vgl. Abbildung 4.7 b).

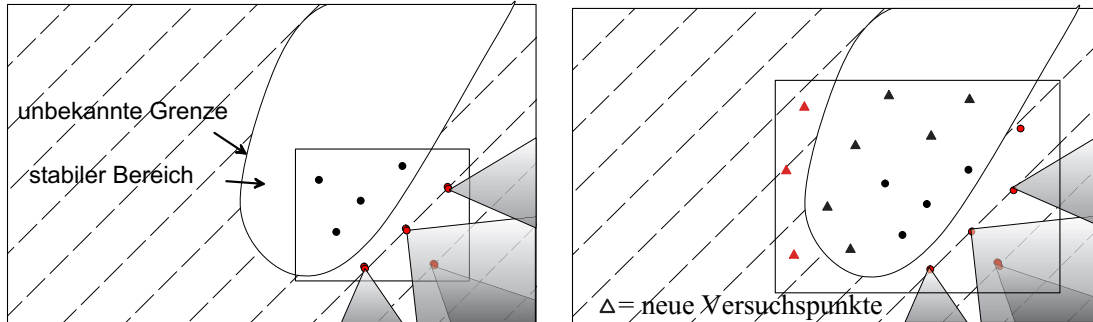


Abbildung 4.7: a) Schritt 1 und 2, b) Schritt 3 und 4 des Algorithmus

Dabei wird nun das erweiterte Minimax-Kriterium verwendet, vgl. Kapitel 4.2.3. Schritt 2 bis 4 werden nun iterativ wiederholt, bis ein geeignetes Stoppkriterium erfüllt ist, vgl. Abbildungen 4.8 a) und 4.8 b). Dies kann z. B. eine bestimmte Gesamtzahl von Versuchen sein, die zur Verfügung steht. Um bei komplexen Zusammenhängen eine ausreichend gute Approximation der Zusammenhänge zu erlauben, kann alternativ eine vorher festgelegte Gesamtzahl von Versuchspunkten innerhalb des stabilen Bereichs verlangt werden.

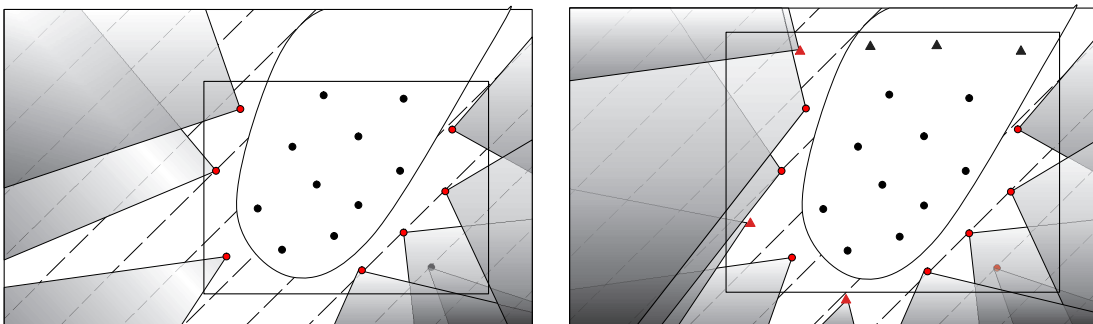


Abbildung 4.8: a) Schritt 2 des zweiten Durchlaufs des Algorithmus, b) Schritt 2 des dritten Durchlaufs des Algorithmus

Durch das oben beschriebene sequentielle Vorgehen zur Generierung von raumfüllenden Versuchsplänen bei unbekanntem Restriktionen werden fehlende Beobachtungen in der Regel nur am Rand des Versagensbereichs beobachtet.

Auf diese Weise wird der unbekannte konvexe stabile Bereich vollständig untersucht, wobei die Anzahl der fehlenden Beobachtungen im Vergleich zum einfachen raumfüllenden Versuchsplan deutlich reduziert werden.

4.2.5 Robustheit des Algorithmus

Eine wesentliche Annahme des Algorithmus ist die Konvexität des stabilen Bereichs. Ist diese Annahme verletzt, so ist es möglich, dass die konvexen Hüllen, die durch den Algorithmus erzeugt werden, fälschlicherweise auch Teile des stabilen Bereichs einschließen. Als Konsequenz würden diese Teile des stabilen Bereichs nicht untersucht werden. Um beurteilen zu können, wie empfindlich der Algorithmus auf Verletzung der Konvexitätsannahme reagiert, wurden Worst-Case-Szenarien für verschiedene nichtkonvexe stabile Bereiche betrachtet. Es wurden also für diese nichtkonvexen stabilen Bereiche Versuchspunkt-Konstellationen betrachtet, die zu einem maximalen Anteil des stabilen Bereichs führen, der durch die konvexen Hüllen ausgeschlossen wird. Wie zu erwarten, ist bei einer starken Abweichung von der Konvexitätsannahme der Einsatz des Algorithmus nicht zu empfehlen. Interessanterweise zeigte sich aber, dass ein gewisser Grad an konkaver Krümmung selbst bei den Worst-Case-Szenarien noch sehr gut toleriert werden kann. Dieser Sachverhalt ist in Abbildung 4.9 dargestellt, in der zwei unterschiedliche nichtkonvexe stabile Bereiche betrachtet wurden. Offensichtlich ist selbst für ein Worst-Case-Szenario der fälschlicherweise ausgeschlossene stabile Bereich nur sehr gering.

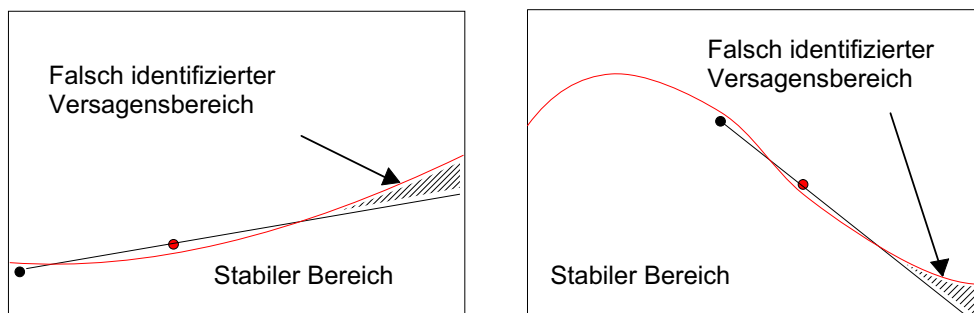


Abbildung 4.9: Worst-Case-Szenarien für zwei nichtkonvexe stabile Bereiche

In Abbildung 4.10 ist ein Fall dargestellt, bei dem ein Großteil des stabilen Bereichs ausgeschlossen werden würde. Solch ein Fall ist aber sehr unrealistisch, da aufgrund des Erfahrungswissens über den Drückprozess bekannt ist, dass die Versagensgrenzen relativ glatt verlaufen. Solch eine extreme lokale Einschnürung ist somit beim Drückprozess nicht zu erwarten.

Somit scheint der Algorithmus relativ robust auf Verletzungen der Konvexitätsannahme des stabilen Bereichs zu reagieren. Diese Eigenschaft könnte die Übertragung des für den Drückprozess entwickelten Algorithmus auf andere Prozesse, bei denen die Versagensgrenzen ebenfalls unbekannt sind, erleichtern.

Selbst wenn das wahre globale Optimum in dem ausgeschlossenen Bereich liegen sollte, ist dies unproblematisch. Grund dafür ist die im ASOP implementierte sequentielle Modellverfeinerung. Bei der Suche nach dem globalen Optimum spielt es dabei keine Rolle mehr, dass im Rahmen der Generierung eines raumfüllenden Versuchsplans bestimmte Bereiche ausgeschlossen wurden.

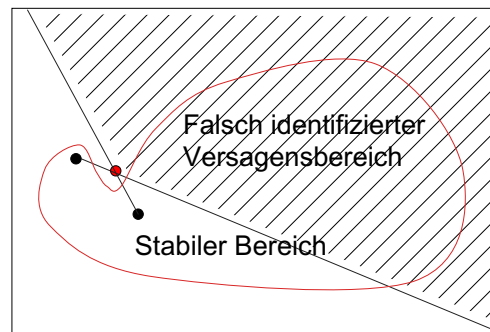


Abbildung 4.10: Worst-Case-Szenarium für einen nichtkonvexen stabilen Bereich mit lokaler Einschnürung

4.3 Simulationsstudie

In diesem Kapitel werden die Ergebnisse einer Simulationsstudie zur Untersuchung der Güte des in Kapitel 4.2 beschriebenen Algorithmus präsentiert. Diese Studie gibt Hinweise, wie die Parameter des Algorithmus sinnvoll einzustellen sind. Zudem erlaubt sie den Effizienzgewinn bezüglich der Anzahl der Versuchspunkte im stabilen Bereich im Vergleich zum einfachen raumfüllenden Versuchsplan unter verschiedenen Randbedingungen zu beurteilen.

4.3.1 Simulationsaufbau

Der Gesamtversuchsbereich wurde auf $[0, 1]^2$ festgelegt. Variiert wurden die folgenden Ausgangsbedingungen:

- Form des konvexen stabilen Versuchsbereichs
- Anteil des stabilen Versuchsbereichs am Gesamtversuchsbereich
- Anzahl der Iterationen des Algorithmus
- Vorgehen bei der schrittweisen Vergrößerung der Versuchspläne

Insgesamt wurden sechs verschiedene Formen des konvexen stabilen Versuchsbereichs untersucht. Dabei wurden zwei unterschiedliche Gruppen betrachtet. Die erste Gruppe zeichnet sich durch stabile Bereiche aus, die am Rand des Versuchsraums liegen. Die unterschiedlichen Formen innerhalb dieser Gruppe lassen sich durch den funktionalen Verlauf der Versagensgrenze eindeutig beschreiben. Betrachtet wurde eine lineare Funktion, ein Polynom und ein Viertelkreis mit Mittelpunkt $(1,1)$. Der stabile Bereich ist jeweils durch die Fläche, die unterhalb dieser Funktionen liegt, definiert. In der zweiten Gruppe befinden sich stabile Bereiche, die komplett oder fast komplett vom Versagensbereich umschlossen sind. Berücksichtigt wurde eine Parabel, eine Ellipse mit Mittelpunkt $(0.5,0.5)$ und ein Kreis mit Mittelpunkt $(0.5,0.5)$. Die stabilen Bereiche wurden so angepasst, dass der Anteil A des stabilen Bereichs für jede der sechs verschiedenen Formen vier verschiedene Werte annehmen kann (25, 35, 45 und 55%). Insgesamt wurden

somit 24 verschiedene stabile Bereiche betrachtet. In Abbildung 4.11 a) und b) sind die unterschiedlichen stabilen Bereiche für $A = 25\%$ abgebildet.

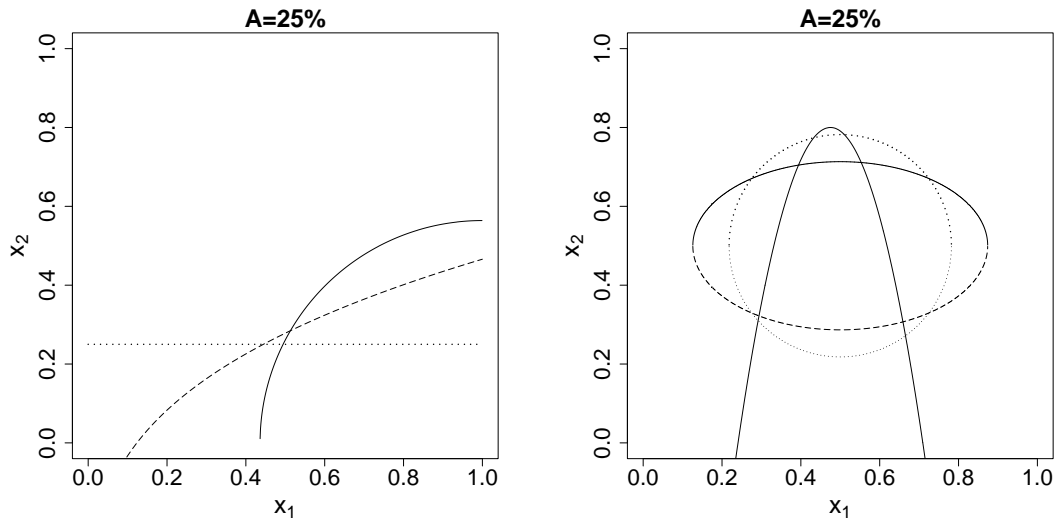


Abbildung 4.11: a) Stabile Bereiche, die am Rand des Versuchsraums liegen, b) Stabile Bereiche, die (fast) komplett vom Versagensbereich umschlossen sind

In der Literatur hat sich für die benötigten Versuchspunkte zur Anpassung von Gauß-Prozess-Modellen die Faustformel von $\sim 10d$ etabliert, wobei d die Anzahl der Faktoren angibt, vgl. Jones et al. (1998). Für den zweidimensionalen Fall würden daher ca. 20 Versuchspunkte im stabilen Bereich benötigt. Da aber mit einer gewissen Anzahl an fehlenden Beobachtungen zu rechnen ist, wurde ein Gesamtversuchsumfang von $N = 30$ gewählt. Für diese Simulationsstudie wurde das einfache Minimax-Kriterium verwendet.

Die Anzahl der Iterationsschritte IS ergibt sich aus der Anzahl der Versuchspunkte n_s , die in jedem Schritt hinzugefügt werden. Um die Vergleichbarkeit der Ergebnisse zu garantieren, wurde nur Werte von n_s verwendet, deren ganzzahliges Vielfaches den Gesamtversuchsumfang $N = 30$ ergibt. Dies sind die folgenden Einstellungen: $n_s = \{1, 2, 3, 5, 6, 10, 15, 30\}$. Die Wahl von n_s wird dabei über die einzelnen Iterationsschritte festgehalten. Für $n_s = 1$ werden 30 Iterationen des Algorithmus benötigt um einen Gesamtversuchsumfang von $N = 30$ zu erreichen. Die Wahl von $n_s = 30$ präsentiert das einfache raumfüllende Design. Für jede der 24 verschiedenen konvexen stabilen Bereiche wurde jede der 8 Möglichkeiten n_s

einzustellen betrachtet. Insgesamt wurden für jede Kombinationsmöglichkeit 100 Versuchspläne mit einem Gesamtumfang von $N = 30$ erzeugt.

In der Praxis erfolgt die Anpassung des Versuchsplans in jedem Iterationsschritt des Algorithmus durch den Drückexperten. Dabei berücksichtigt er die Lage der Versagenspunkte. Dieses adaptive Vorgehen lässt sich aber für die Simulationsstudie schlecht umsetzen. Aus diesem Grund wurde ein automatisiertes Verfahren zur Anpassung des Versuchsplans in den Algorithmus implementiert. Die Verschiebung des Versuchsplans erfolgt dabei wie bisher. Als Mittelpunkt des Versuchsplans wird das arithmetische Mittel der Versuchspunkte, die im stabilen Bereich liegen, gewählt. Die prozentuale Vergrößerung des Versuchsplans wird durch eine geometrische Folge $p_i = A_1 \cdot q^{i-1}$ mit $i = 2, \dots, IS$ festgelegt. Der Parameter A_1 bestimmt den Flächenanteil des ersten Versuchsplans am interessierenden Gesamtversuchsplan. Dieser Versuchsplan wird nun in den IS Iterationsschritten des Algorithmus vergrößert, bis er im letzten Schritt den kompletten Versuchsplan $[0, 1]^2$ untersucht. Der zweite Parameter q regelt die schrittweise Vergrößerung des Versuchsplans. Der Flächenanteil des i -ten Versuchsplans lässt sich aus der geometrischen Folge wie folgt bestimmen. Aus der Gleichung $c \sum_{i=2}^{IS} p_i = (1 - A_1)$ lässt sich zunächst die Konstante c berechnen. Damit gilt für den Flächeninhalt A_i des i -ten Versuchsplans $A_i = A_1 + cp_2 + \dots + cp_i$ mit $i = 2, \dots, IS$. Im i -ten Iterationsschritt erhöht sich der Flächenanteil somit additiv um den Wert cp_i . Während dieser Wert für $q > 1$ mit steigender Iterationszahl zunimmt, sinkt dieser für $q < 1$.

Für die Simulationsstudie wurden für die beiden Parameter die folgenden Einstellungsmöglichkeiten betrachtet: $A_1 = \{10\%, 20\%, 30\%\}$ und $q = \{1, 2, 3\}$. Werte von q kleiner 1 wurden nicht berücksichtigt, da es bei dem sequentiellen Vorgehen des Algorithmus wenig sinnvoll erscheint, zu Beginn sofort einen Großteil des Versuchsraums zu untersuchen. Der Einfluss von q lässt sich für ein festes A_1 verdeutlichen. Sei $IS = 3$ und $A_1 = 10\%$, dann wird für $q = 1$ im zweiten Schritt ein Versuchsplan generiert mit 55% Flächenanteil am Gesamtversuchsraum $[0, 1]^2$. Für $q = 2$ sinkt der Anteil des zweiten Versuchsplans auf 40% und für $q = 3$ auf 33%.

4.3.2 Ergebnisse

4.3.2.1 Einfluss der Form und Lage des stabilen Bereichs

In Abbildung 4.12 a) sind die Simulationsergebnisse für die verschiedenen Formen, die (fast) komplett vom Versagensbereich umschlossen sind für $A = 25\%$ anhand von Box Plots zusammengefasst. Dabei wurde für die einzelnen Formen über alle möglichen Iterationsschritte größer eins gemittelt. Da ein Iterationsschritt dem einfachen raumfüllenden Versuchsplan entspricht, werden die zugehörigen Ergebnisse als Referenzwerte zur Bewertung der Effizienz des Algorithmus verwendet. Nimmt der stabile Bereich z. B. 25% der Gesamtfläche ein, so werden mit dem einfachen raumfüllenden Versuchsplan bei $N = 30$ im Schnitt 7.5 Versuchspunkte in den unbekanntem stabilen Bereich gelegt.

Offensichtlich besitzt die Form des stabilen Bereichs einen deutlichen Einfluss auf die Simulationsergebnisse. Existiert ein kreisförmiger stabiler Bereich in der Mitte des Versuchsraums, so liegt die mittlere Anzahl von Versuchspunkten im stabilen Bereich mit 8.5 nur geringfügig über dem Mittel von 7.5, das beim einfachen raumfüllenden Versuchsplan beobachtet wurde. Die Effizienzsteigerung beträgt für den Kreis im Mittel somit nur 13.5 Prozent, vgl. Abbildung 4.12 b).

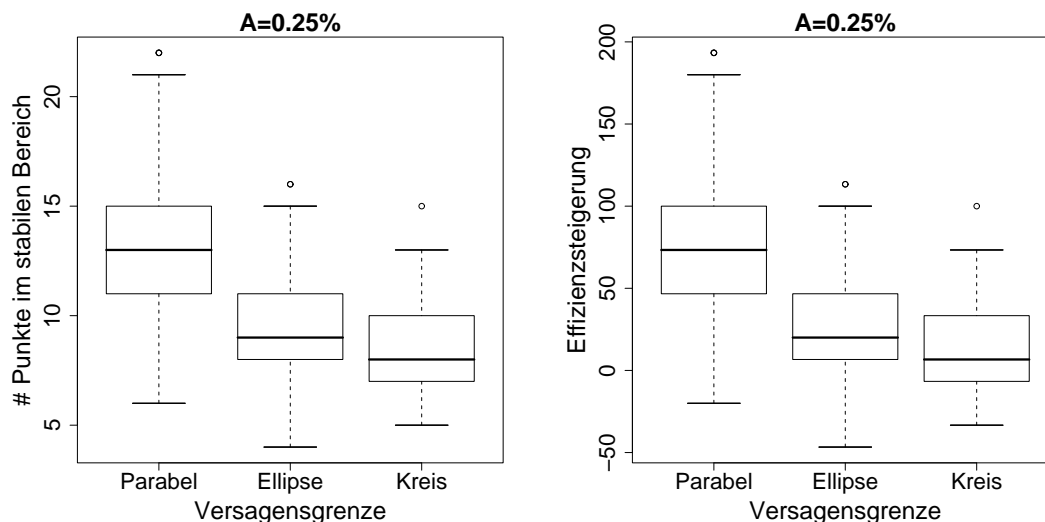


Abbildung 4.12: a) Anzahl der Versuchspunkte im stabilen Bereich b) Effizienzsteigerung im Vergleich zum einfachen raumfüllenden Versuchsplan

Für den ellipsen- und insbesondere den parabelförmigen stabilen Bereich sind die Ergebnisse mit im Mittel 9.5 bzw. 13.2 Versuchspunkten im stabilen Bereich schon deutlich besser. Die mittlere Effizienzsteigerung liegt hier bei 26.6 bzw. 75.5 %. Die Streuung in den Simulationsergebnissen ist für alle drei Formen groß. Während für den Kreis die Spannweite bei zehn Versuchspunkten liegt, wächst sie für die Parabel auf 16 Versuchspunkte. Die relativ große Spannweite ist dadurch zu erklären, dass die Ergebnisse über alle möglichen Einstellungen für die Anzahl der Iterationsschritte gemittelt wurden.

Liegt der stabile Bereich am Rand des Versuchsbereichs, werden deutlich mehr Versuchspunkte in diesem Bereich platziert, vgl. Abbildung 4.13 a) und b). Die besten Ergebnisse wurden bei der linearen Versagensgrenze beobachtet. In diesem Fall konnten im Schnitt 22.2 Versuchspunkte im unbekanntem stabilen Bereich beobachtet werden, vgl. Abbildung 4.13 a). Dies entspricht einer mittleren Effizienzsteigerung von 196 %, vgl. Abbildung 4.13 b). Die Anzahl der Versuchspunkte im unbekanntem stabilen Bereich konnte durch den Algorithmus damit fast verdreifacht werden. Die maximale Effizienzsteigerung liegt bei 273 %. Bei diesen Durchläufen konnten 28 von 30 Versuchspunkten im stabilen Bereich platziert werden. Es wurden nur zwei Versuchspunkte benötigt, um den Versagensbereich zu identifizieren. In welchen Situationen dieses sehr gute Ergebnis eintreten kann, wird im nächsten Abschnitt erklärt. Für das Polynom und den Viertelkreis wurden im Mittel 19.83 bzw. 17.76 Versuchspunkte im stabilen Bereich beobachtet. Die Effizienzsteigerung liegt im Mittel bei 137 bzw. 164 %. Die Streuung der Anzahl der Punkte im stabilen Bereich scheint für die drei Funktionen gleich zu sein.

Aus diesen Erkenntnissen lässt sich schließen, dass die Anzahl der Versuchspunkte im stabilen Bereich sowohl stark von der Lage als auch der Form des stabilen Bereichs abhängt. Von großem Interesse wäre nun ein Maß zu identifizieren, das erlaubt, Aussagen über die Größe des Effizienzgewinns abhängig von der Form und Lage des stabilen Bereichs zu tätigen. Offensichtlich hängt der Flächenanteil der konvexen Hüllen am Gesamtversuchsbereich von der Anordnung der Versuchspunkte im stabilen Bereich zu den Versuchspunkten im Versagensbereich ab. Um dies zu verdeutlichen, betrachte einen Versagenspunkt im d -dimensionalen Ver-

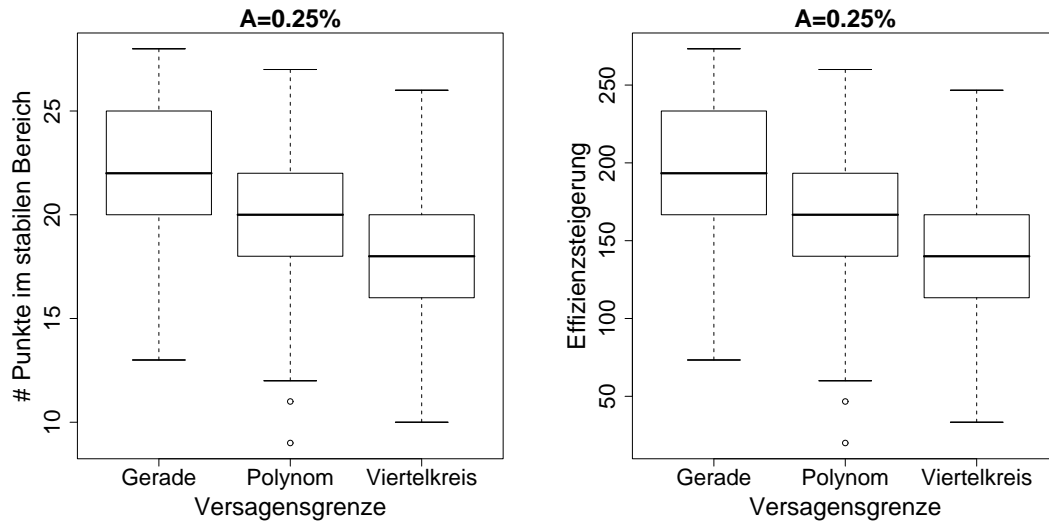


Abbildung 4.13: a) Anzahl der Versuchspunkte im stabilen Bereich b) Effizienzsteigerung im Vergleich zum einfachen raumfüllenden Versuchsplan

suchsraum und d Punkte im stabilen Bereich. Definiere zudem mit E die Ebene, die durch diese d Punkte aufgespannt wird. Je weiter nun diese d Punkte auseinander liegen und je kleiner der orthogonale Abstand des Versagenspunktes zu der Ebene E ist, desto größer wird das Volumen der zugehörigen konvexen Hülle. Dieser Sachverhalt ist für den zweidimensionalen Fall in Abbildung 4.14 dargestellt.

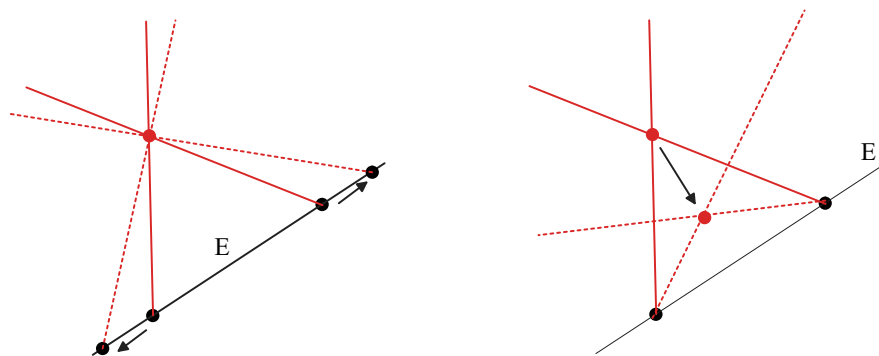


Abbildung 4.14: Vergrößerung der konvexen Hülle durch a) Vergrößerung des Abstandes der Punkte im stabilen Bereich b) Verringerung des orthogonalen Abstandes des Versagenspunktes zur Gerade E

Daraus lässt sich schließen, dass der Anteil der konvexen Hüllen am Gesamtversuchsbereich abhängig von der Krümmung der Versagensgrenze ist. Je weniger stark die Versagensgrenze gekrümmt ist, desto weiträumiger können die Punkte im stabilen Bereich verteilt sein und desto näher können die Versagenspunkte an der Ebene E liegen. Der Einfluss der Krümmung ist für die beiden Versagensgrenzen, die sich durch eine Gerade und einen Viertelkreis beschreiben lassen, in Abbildung 4.15 für $A = 25\%$ dargestellt. Dabei wurde die Anordnung der Versuchspunkte gewählt, die die Fläche der konvexen Hülle maximiert.

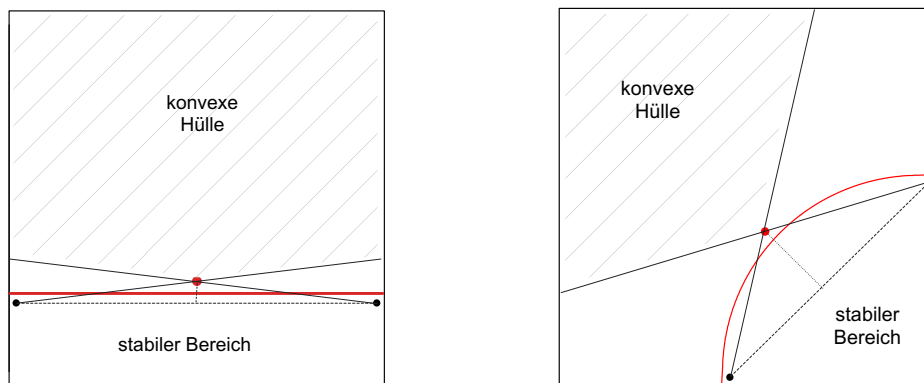


Abbildung 4.15: Abhängigkeit des Flächeninhalts der konvexen Hülle von der Krümmung der Versagensgrenze

Um zu überprüfen, ob die Krümmung der Versagensgrenze in der Tat ein aussagekräftiges Maß für die Effizienz des Algorithmus ist, soll im Folgenden zunächst ein Maß für die mittlere Krümmung der Versagensgrenze entwickelt werden. In der Analysis existiert ein Maß, das die Krümmung einer Funktion $f(x)$ an einer Stelle x definiert, vgl. Kaballo (1996) Seite 198. Dieses Maß ist gegeben durch

$$k(x) := \frac{f''(x)}{(1 + f'(x)^2)^{\frac{3}{2}}}. \quad (4.5)$$

Die mittlere Krümmung sei nun wie folgt definiert. Bestimme eine Folge von Punkten x_1, \dots, x_n , so dass die zugehörigen Werte $f(x_1), \dots, f(x_n)$ äquidistant auf der Funktionskurve platziert sind. Berechne nun $k(x_i)$ für $i = 1, \dots, n$. Die mittlere Krümmung sei dann für $n \rightarrow \infty$ definiert als $k_M = \frac{1}{n} \sum_{i=1}^n k(x_i)$. In Tabelle 4.1 ist die mittlere Krümmung für die sechs stabilen Bereiche angegeben.

Fläche	Gerade	Polynom	Viertelkreis	Parabel	Ellipse	Kreis
0.25 %	0	0.62	1.77	1.68	3.10	3.54
0.35 %	0	0.75	1.52	1.56	2.71	2.99
0.45 %	0	0.85	1.33	1.37	2.44	2.64
0.55 %	0	0.89	1.20	1.26	2.24	2.39

Tabelle 4.1: Mittlere Krümmung abhängig von der Form des stabilen Bereichs

Ein Vergleich mit den Simulationsergebnissen ergibt, dass in der Tat mit steigender mittlerer Krümmung der Effizienzgewinn abnimmt, vgl. Abbildung 4.12 b) und 4.13 b). Nur für $A = 25\%$ ist der Effizienzgewinn der Parabel trotz etwas kleinerer Krümmung geringer als für den Viertelkreis. Dies liegt daran, dass neben der Krümmung auch die Länge der Grenze im Versuchsraum eine Rolle spielt. Der Algorithmus erreicht somit seinen maximalen Effizienzgewinn wenn die Versagensgrenze linear ist und deren Länge möglichst klein ist. Liegt jedoch der stabile Bereich in der Mitte des Versuchsbereichs und ist er zudem kreisförmig, so ist der Effizienzgewinn am niedrigsten.

4.3.2.2 Einfluss der Größe des stabilen Bereichs

Der mittlere prozentuale Effizienzgewinn in Abhängigkeit vom Anteil A des stabilen Bereichs ist in Abbildung 4.16 dargestellt. Offensichtlich sinkt mit steigendem Anteil des stabilen Bereichs der Effizienzgewinn. Dies ist darin zu begründen, dass bei einem großen stabilen Bereich der Anteil von Punkten in diesem Bereich selbst beim einfachen raumfüllenden Versuchsbereich schon hoch ist. Somit ist das Potential zur Effizienzsteigerung geringer. Der funktionale Verlauf dieses Effizienzverlusts ist dabei für alle drei Formen des stabilen Bereichs ähnlich. Es kann aber beobachtet werden, dass mit Abnahme der mittleren Krümmung die absolute Differenz des mittleren Effizienzgewinns zwischen $A = 25\%$ und $A = 55\%$ steigt.

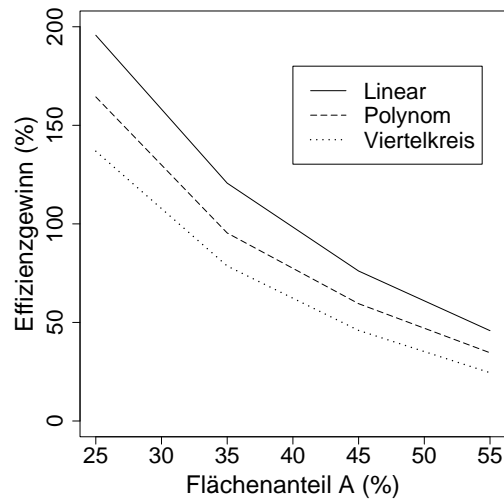


Abbildung 4.16: Mittlerer Effizienzgewinn in Abhängigkeit des Flächenanteils A für verschiedene Versagensgrenzen

4.3.2.3 Einfluss der Anzahl der Iterationsschritte

Die weitere Auswertung der Simulationsergebnisse zeigt, dass wie erwartet ein streng monoton steigender Zusammenhang zwischen der Anzahl der Iterationen und der Effizienz des Algorithmus besteht. Für die funktionale Versagensgrenze $x^{0.3} - c$ sind die zugehörigen Ergebnisse in Abbildung 4.17 zusammengefasst. Wird der raumfüllende Versuchsplan in nur zwei Iterationen mit je 15 Versuchspunkten generiert, führt dies insbesondere bei einem kleinen stabilen Bereich schon zu einer deutlichen Effizienzsteigerung. Der maximale Effizienzgewinn wird erreicht, falls pro Iteration nur ein Versuchspunkt hinzugefügt wird. Interessant ist, dass der Effizienzgewinn mit steigender Iterationsanzahl relativ schnell konvergiert. Die Effizienzsteigerung zwischen 10 und 30 Iterationen ist nur noch gering. Das Sättigungsverhalten des Effizienzgewinns ist bei der Wahl der Iterationszahl des Algorithmus in der Praxis von sehr großer Bedeutung. Da der Zeitaufwand bei der Erhebung eines sequentiellen Versuchsplans von der Iterationszahl abhängt, wird man versuchen den Algorithmus so einzustellen, dass ein gutes Kosten-Nutzen-Verhältnis erreicht werden kann. Dieses ergibt sich an der Bruchstelle zwischen der in den ersten Iterationsschritten rasch ansteigenden Effizienzkurve und dem Übergang in die fast stationäre Phase. Für den zweidimensionalen Fall aus der

Simulationsstudie scheint der Bruchpunkt bei sechs Iterationen zu liegen, vgl. Abbildung 4.17.

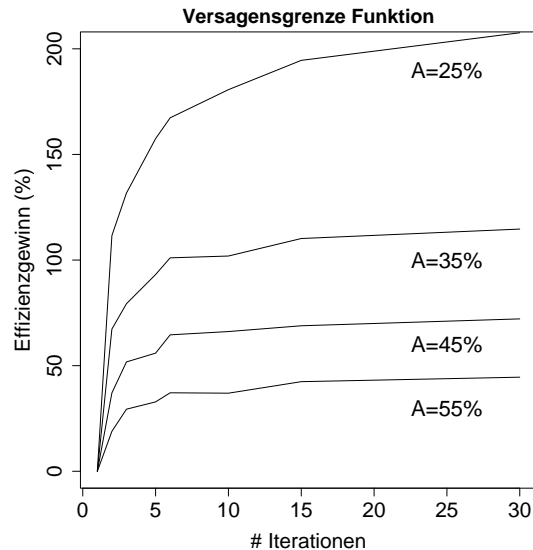


Abbildung 4.17: Mittlerer Effizienzgewinn in Abhängigkeit der Iterationszahl für die funktionale Versagensgrenze $x^{0.3} - c$

Um eine allgemeine Einstellungsempfehlung der Iterationszahl geben zu können, wurde der Effizienzgewinn für unterschiedliches N sowohl im zwei- als auch dreidimensionalen Versuchsraum untersucht. Dabei zeigte sich, dass in beiden Fällen der Bruchpunkt bei ca. $N/5$ Iterationen liegt. Ein sehr gutes Kosten-Nutzen-Verhältnis scheint bei niedrigen Dimensionen somit bei $N/5$ Iterationen vorzuliegen.

4.3.2.4 Einfluss der schrittweisen Vergrößerung der Versuchspläne

Wie in Kapitel 4.3.1 beschrieben, lässt sich mit Hilfe einer geometrischen Reihe ein automatisiertes Vorgehen zur schrittweisen Vergrößerung der Versuchspläne umsetzen. Die schrittweise Vergrößerung ist dabei durch die Wahl der beiden Parameter A_1 und q der geometrischen Reihe steuerbar. In Tabelle 4.2 ist die mittlere Anzahl der Versuchspunkte im stabilen Bereich beispielhaft für den stabilen Bereich mit linearer Versagensgrenze und drei Iterationsschritten gegeben.

Eine erste wesentliche Erkenntnis der Simulationsergebnisse ist, dass die mittlere Anzahl der Versuchspunkte im stabilen Bereich zwar abhängig von den Einstellungen dieser beiden Parameter ist, diese aber nicht stark variiert. Im Vergleich dazu besitzt die Wahl der Anzahl der Iterationsschritte des Algorithmus einen deutlich stärkeren Einfluss. Die Simulationsergebnisse zeigen, dass eine optimale Einstellungsempfehlung der beiden Parameter von der Größe des stabilen Bereichs abhängt. Für $A = 25\%$ wurde die maximale Versuchsanzahl im stabilen Bereich für $A_1 = 10\%$ und $q = 2$ beobachtet. Wie zu erwarten ist es in dieser Situation sinnvoll, mit einem kleinen Startdesign zu beginnen und die Grenzen langsam zu erweitern. Für $A = 55\%$ wurden drei der vier besten Ergebnisse für $q = 2$ erzielt. Ein größeres Initialdesign führt somit auch zu guten Ergebnissen. Insgesamt haben sich die Einstellungen als optimal erwiesen, die dazu führten, dass bei einer Erweiterung des Versuchsplans dieser nur knapp die Versagensgrenzen überschreitet. Auf diese Weise werden Versagenspunkte direkt hinter der Versagensgrenze beobachtet, wodurch ein Großteil des Versagensraums durch die konvexen Hüllen ausgeschlossen werden kann. Für die in der Praxis verwendete adaptive sequentielle Vergrößerung kann basierend auf diesen Ergebnissen somit die folgende Vorgehensweise empfohlen werden: Stößt man bei der Untersuchung des Versuchsraums an die Versagensgrenze, so ist diese erst genauer zu erkunden und der Versuchsbereich in dieser Phase nicht oder nur geringfügig zu vergrößern.

	A=55 %			A=25 %		
	$A_1=10\%$	$A_1=20\%$	$A_1=30\%$	$A_1=10\%$	$A_1=20\%$	$A_1=30\%$
q=1	23.84	23.25	22.69	20.84	20.43	19.13
q=2	24.33	23.94	23.85	21.49	19.85	19.45
q=4	23.68	23.78	23.96	21.2	20.37	19.55

Tabelle 4.2: Mittlere Anzahl der Versuchspunkte im stabilen Bereich mit linearer Versagensgrenze und drei Iterationsschritten

4.3.2.5 Effizienzverlust mit steigender Dimension der Daten

Da mit steigender Dimension der Daten der Anteil der konvexen Hüllen am Gesamtversuchsbereich sinkt, lässt auch die Effizienz des Algorithmus nach. Dieser Sachverhalt ist in Abbildung 4.18 für den Übergang vom zwei- zum dreidimensionalen Versuchsraum verdeutlicht. Der Anwendung des Algorithmus sollte somit ein Screening der wichtigsten Faktoren vorangegangen sein.

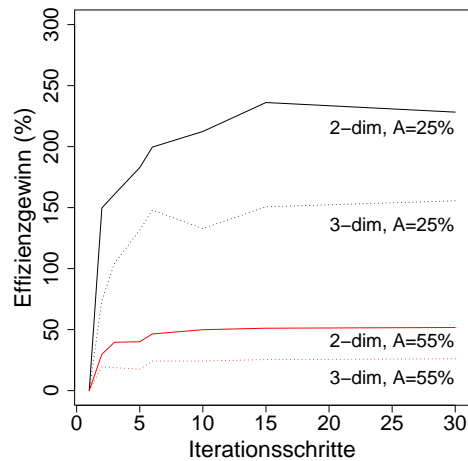


Abbildung 4.18: Vergleich des mittleren Effizienzgewinns für zwei- und dreidimensionale Versuchsräume bei linearer Versagensgrenze

4.3.3 Zusammenfassung der Simulationsergebnisse

Mit Hilfe der Simulationsstudie konnten ausführliche Erkenntnisse über die Stärken und Schwächen des Algorithmus gewonnen werden. Es zeigte sich, dass die Effizienz des Algorithmus stark von der Größe, Form und Lage des stabilen Bereichs abhängt. Je kleiner der stabile Bereich, desto besser war die Performance des Algorithmus. Um die Form des stabilen Bereichs charakterisieren zu können, wurde der Verlauf der Versagensgrenze betrachtet. Bei festem Anteil A des stabilen Bereichs stieg die Effizienz des Algorithmus mit abnehmender Krümmung und Länge der Versagensgrenze. Eine maximale Effizienzsteigerung wurde somit bei linearen Versagensgrenzen, die möglichst kurz sind, beobachtet. Bezüglich der Lage des stabilen Bereichs konnte gezeigt werden, dass stabile

Bereiche am Rand des Versuchsbereichs einen deutlich größeren Effizienzgewinn ermöglichen als stabile Bereiche im Zentrum des Versuchsbereichs, die komplett vom Versagensbereich umschlossen sind. Eine Untersuchung des Einflusses der Einstellparameter des Algorithmus ergab, dass die Anzahl der Iterationsschritte einen wesentlich stärkeren Einfluss auf die Effizienz besitzt als die Parameter, die die schrittweise Anpassung des Versuchsraums steuern. Bezüglich der Anzahl der Iterationsschritte konnte gezeigt werden, dass bei niedriger Dimension der Daten unabhängig von dieser $N/5$ Iterationsschritte ein sehr gutes Kosten-Nutzen-Verhältnis liefern. Durch die Untersuchung des Einflusses der Parameter, die die schrittweise Anpassung des Versuchsplans steuern, konnte eine Vorgehensstrategie für die adaptive Anpassung des Versuchsplans in der Praxis entwickelt werden. Sobald bei der Untersuchung des Versuchsraums die Grenzen des Versagensbereichs erreicht werden, ist zu empfehlen, diese Grenzen zunächst genauer zu untersuchen, bevor der Versuchsplan weiter vergrößert wird.

Insgesamt lässt sich sagen, dass der Algorithmus zur Generierung von raumfüllenden Versuchsplänen bei niedrig-dimensionalen Daten mit unbekanntem konvexen Versagensgrenzen sehr zu empfehlen ist. Insbesondere bei stabilen Bereichen, die am Rand des Versuchsbereichs liegen, konnte ein deutlicher Effizienzgewinn beobachtet werden. Bei linearen Versagensgrenzen betrug die mittlere Effizienzsteigerung im zweidimensionalen Fall bis zu 200 %.

Kapitel 5

Gauß-Prozess-Modell

5.1 Stochastische Prozesse

Im Folgenden werden wesentliche Begriffe aus der Theorie stochastischer Prozesse vorgestellt, die zur Definition des Gauß-Prozess-Modells benötigt werden. Eine gut strukturierte Einführung in stochastische Prozesse wird in Christakos (1992) gegeben.

Definition 5.1.1 *Ein stochastischer Prozess ist definiert als Menge von Zufallsvariablen $\{Z(\mathbf{x}) : \mathbf{x} \in \mathcal{X}\}$ indiziert durch eine unendliche Indexmenge $\mathcal{X} \subseteq \mathbb{R}^d$.*

Somit ist $Z(\mathbf{x})$ für jedes $\mathbf{x} \in \mathcal{X}$ eine Zufallsvariable. Andere synonym verwendete Begriffe für den stochastischen Prozess sind *Zufallsprozess*, *Zufallsfunktion* oder *Zufallsfeld*. Eine spezielle Klasse der stochastischen Prozesse bilden die Gauß-Prozesse.

Definition 5.1.2 *Ein stochastischer Prozess $\{Z(\mathbf{x}) : \mathbf{x} \in \mathcal{X}\}$ heißt Gauß-Prozess, falls für beliebiges $n \in \mathbb{N}$ und beliebige $\mathbf{x}_1, \dots, \mathbf{x}_n \in \mathcal{X}$ der Vektor $(Z(\mathbf{x}_1), \dots, Z(\mathbf{x}_n))'$ eine multivariate Normalverteilung besitzt.*

Ein Gauß-Prozess ist vollständig bestimmt durch seine Erwartungswertfunktion $\mu(\mathbf{x}) := E[Z(\mathbf{x})]$ mit $\mathbf{x} \in \mathcal{X}$ und seine Kovarianzfunktion $C(\mathbf{x}_i, \mathbf{x}_j) := \text{Cov}(Z(\mathbf{x}_i), Z(\mathbf{x}_j))$ mit $\mathbf{x}_i, \mathbf{x}_j \in \mathcal{X}$.

Um eine Realisation eines stochastischen Prozesses an einem neuen Versuchspunkt \mathbf{x}_0 in \mathcal{X} vorhersagen zu können, sind Annahmen über die räumliche Homogenität zu treffen. Als ein Maß für die räumliche Homogenität kann die Stationaritätsannahme angesehen werden.

Definition 5.1.3 Sei ein stochastischer Prozess $\{Z(\mathbf{x}) : \mathbf{x} \in \mathcal{X}\}$ gegeben mit $\mathcal{X} \subset \mathbb{R}^d$. Für beliebige $\mathbf{x}_1, \dots, \mathbf{x}_n \in \mathcal{X}$ mit $n \geq 1$ und $\mathbf{h} \in \mathbb{R}^d$ sei $\mathbf{x}_1 + \mathbf{h}, \dots, \mathbf{x}_n + \mathbf{h} \in \mathcal{X}$. Dann heißt der stochastische Prozess $\{Z(\mathbf{x}) : \mathbf{x} \in \mathcal{X}\}$ streng stationär, falls $(Z(\mathbf{x}_1), \dots, Z(\mathbf{x}_n))'$ und $(Z(\mathbf{x}_1 + \mathbf{h}), \dots, Z(\mathbf{x}_n + \mathbf{h}))'$ für alle solche \mathbf{h} dieselbe Verteilung besitzen.

Da dies die Kenntnis der gemeinsamen Verteilung von $\mathbf{Z} = (Z(\mathbf{x}_1), \dots, Z(\mathbf{x}_n))$ voraussetzt, wird stattdessen häufig eine schwächere Form der räumlichen Homogenität unterstellt, die nur über die ersten beiden Momente der Verteilung definiert wird.

Definition 5.1.4 Ein stochastischer Prozess $\{Z(\mathbf{x}) : \mathbf{x} \in \mathcal{X}\}$ heißt schwach stationär, falls

- $E[Z(\mathbf{x}_i)] = \mu$ für alle $\mathbf{x}_i \in \mathcal{X}$
- $Cov[Z(\mathbf{x}_i), Z(\mathbf{x}_j)] = C(\mathbf{h})$ für alle $\mathbf{x}_i, \mathbf{x}_j \in \mathcal{X}$ mit $\mathbf{h} = \mathbf{x}_i - \mathbf{x}_j$.

Dies bedeutet, dass für schwach stationäre stochastische Prozesse der Erwartungswert der Zufallsvariable Z in \mathcal{X} konstant ist und dass die Kovarianz zwischen zwei beliebigen Punkten in \mathcal{X} nur vom Differenzvektor \mathbf{h} abhängt. Für schwach stationäre Prozesse gilt

$$var[Z(\mathbf{x})] = \sigma_Z^2 \quad \text{für alle } \mathbf{x} \in \mathcal{X}. \quad (5.1)$$

Da ein Gauß-Prozess vollständig durch die ersten beiden Momente bestimmt ist, sind für solch einen stochastischen Prozess die starke und schwache Stationarität äquivalent.

Bemerkung Für beliebige $\mathbf{x}_1, \dots, \mathbf{x}_n \in \mathcal{X}$ mit $n \geq 1$ und $\mathbf{h} \in \mathbb{R}^d$ sei $\mathbf{x}_1 + \mathbf{h}, \dots, \mathbf{x}_n + \mathbf{h} \in \mathcal{X}$. Ein Gauß-Prozess ist stark stationär, falls $(Z(\mathbf{x}_1), \dots, Z(\mathbf{x}_n))'$ und $(Z(\mathbf{x}_1 + \mathbf{h}), \dots, Z(\mathbf{x}_n + \mathbf{h}))'$ denselben Erwartungswert und dieselbe Kovarianzmatrix besitzen.

Aufbauend auf der schwachen Stationarität lassen sich nun stärkere räumliche Homogenitätsbegriffe definieren. Zwei wichtige Konzepte dabei sind die *Isotropie* und die *Trennbarkeit*, vgl. Zimmermann (1989).

Definition 5.1.5 *Ein schwach stationärer stochastischer Prozess heißt isotrop, falls für alle $\mathbf{x}_i, \mathbf{x}_j \in \mathcal{X}$ mit $\mathbf{h} = \mathbf{x}_i - \mathbf{x}_j$ sich die Kovarianz $C(\mathbf{h})$ als Funktion darstellen lässt, die nur von dem euklidischen Abstand $\|\mathbf{h}\|$ abhängt.*

Die Isotropie bei schwach stationären Prozessen führt dazu, dass jedem Paar von Punkten in \mathcal{X} mit demselben Abstand dieselbe Kovarianz zugeordnet wird. In der Praxis ist diese Annahme jedoch häufig zu einschränkend, da ein Prozess sich in verschiedene Richtungen des Versuchsraums unterschiedlich verändern kann. Kovarianzfunktionen stationärer Gauß-Prozesse, die nicht isotrop sind, hängen somit für alle Paare \mathbf{x}_i und \mathbf{x}_j aus \mathcal{X} sowohl vom Abstand als auch der Richtung zwischen \mathbf{x}_i und \mathbf{x}_j ab. Nicht isotrope Kovarianzfunktionen werden im Folgenden als anisotrope Kovarianzfunktionen bezeichnet. In einigen Fällen lässt sich die Anisotropie durch geometrische Transformationen wie Rotation und Reskalierung der Einflussfaktoren aufheben. Solche Kovarianzfunktionen $C(\mathbf{h})$ werden dann als *geometrisch anisotrop* bezeichnet.

Mit Hilfe der Trennbarkeit lässt sich die Anzahl benötigter Kovarianzparameter weiter einschränken. Dabei wird ausgenutzt, dass das Produkt von Kovarianzfunktionen wieder eine Kovarianzfunktion ergibt. Gleiches gilt im Übrigen auch für die Korrelationsfunktion.

Definition 5.1.6 *Ein schwach stationärer stochastischer Prozess heißt trennbar, falls für alle $\mathbf{x}_i, \mathbf{x}_j \in \mathcal{X}$ mit $\mathbf{h} = \mathbf{x}_i - \mathbf{x}_j$ für die Kovarianz gilt*

$$C(\mathbf{h}) = C_1(h_1) \cdot \dots \cdot C_d(h_d), \quad \text{wobei } \mathbf{h} = (h_1, \dots, h_d).$$

Eine spezielle Klasse von schwach stationären stochastischen Prozessen sind die so genannten White-Noise-Prozesse.

Definition 5.1.7 Sei ein stochastischer Prozess $\{Z(\mathbf{x}) : \mathbf{x} \in \mathcal{X}\}$ gegeben mit $\mathcal{X} \subset \mathbb{R}^d$. Dann ist $\{Z(\mathbf{x}) : \mathbf{x} \in \mathcal{X}\}$ ein White-Noise-Prozess, falls für $\mathbf{h} = \mathbf{x}_i - \mathbf{x}_j$ mit $\mathbf{x}_i, \mathbf{x}_j \in \mathcal{X}$ gilt:

$$E[Z(\mathbf{x})] = 0 \text{ für alle } \mathbf{x} \in \mathcal{X} \text{ und } C(\mathbf{h}) = \begin{cases} \sigma_Z^2 & , \text{für } \mathbf{h} = 0 \\ 0 & , \text{für } \mathbf{h} \neq 0 \end{cases}. \quad (5.2)$$

Die Kovarianzfunktion lässt sich häufig durch einen Korrelationsparametervektor $\boldsymbol{\theta}$ parametrisieren. Um diesen Sachverhalt zu verdeutlichen, wird im Folgenden die Notation $C(\mathbf{h}|\boldsymbol{\theta})$ verwendet.

5.2 Motivation des Gauß-Prozess-Modells

Gauß-Prozess-Modelle wurden ursprünglich für Computer Experimente entwickelt, deren Zielgrößen deterministisch sind. Die Erweiterung dieser Modelle für physikalische Experimente ist aber unproblematisch. Um den Einsatz des Gauß-Prozess-Modells zu motivieren, betrachte die folgende zweidimensionale Funktion

$$z = -\frac{\sin\left(\sqrt{x_1^2 + x_2^2}\right)}{\sqrt{x_1 + x_2^2}} \quad \text{mit } -12 \leq x_1, x_2 \leq 0, \quad (5.3)$$

deren Response-Oberfläche in Abbildung 5.1 a) dargestellt ist. Basierend auf 20 Versuchspunkten, die sich gleichmäßig im Versuchsraum verteilen, wurde ein Gauß-Prozess-Modell angepasst.

Die Approximation der Responsefläche durch dieses Gauß-Prozess-Modell ist in Abbildung 5.1 b) dargestellt. Wie anhand dieses Beispiels zu sehen ist, erlaubt das Gauß-Prozess-Modell komplexe Zusammenhänge mit relativ geringem Versuchsumfang gut zu approximieren.

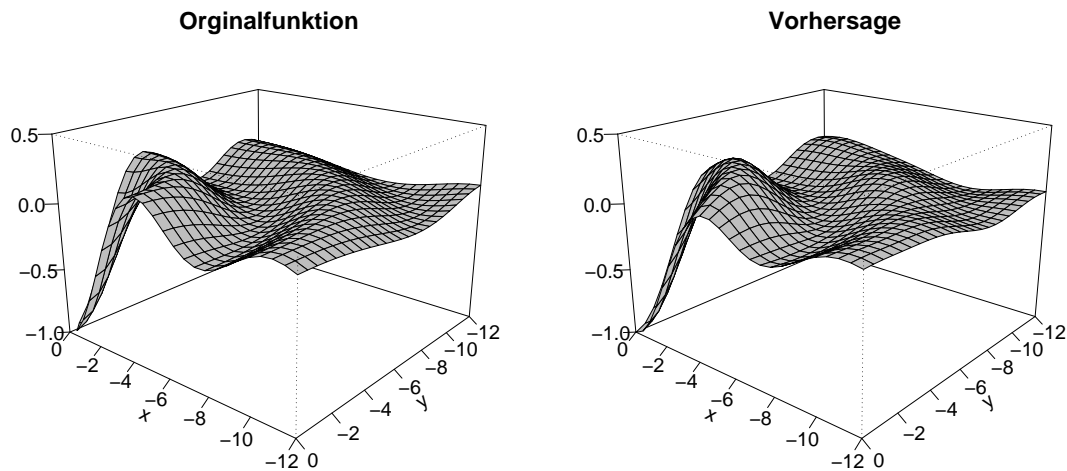


Abbildung 5.1: a) Komplexe zweidimensionale Funktion b) Anpassung durch das Gauß-Prozess-Modell

5.3 Definition des Gauß-Prozess-Modells

Ein Großteil der statistischen Theorie wird durch die Annahme vereinfacht, dass die Zufallsvariablen unabhängig sind. In der Praxis sind aber oft Modelle realistischer, die statistische Abhängigkeit berücksichtigen. Genauso wie z. B. Zeitreihendaten können auch räumliche Daten korreliert sein. Ziel bei räumlichen Daten ist es somit, diese Abhängigkeiten zu modellieren. Diesen Ansatz verfolgt das Gauß-Prozess-Modell (Santner et al., 2003), indem es Versuche mit d Einflussfaktoren als Datenpunkte im d -dimensionalen Versuchsraum betrachtet und die räumliche Abhängigkeit der Versuchspunkte beschreibt. Wie der Name des Modells schon andeutet, werden dazu stationäre Gauß-Prozesse verwendet.

In dem Versuchsraum $\mathcal{X} \subseteq \mathbb{R}^d$ seien n Versuchspunkte $\mathbf{x}_1, \dots, \mathbf{x}_n$ beobachtet worden. Das Gauß-Prozess-Modell behandelt die beobachteten Zielgrößen $y(\mathbf{x}_i)$ mit $i = 1, \dots, n$ als Realisationen eines stochastischen Prozesses $\{Y(\mathbf{x}) : \mathbf{x} \in \mathcal{X}\}$. Das Modell beschreibt die Veränderung des Mittelwertvektors des zugrundeliegenden stochastischen Prozesses $\{Y(\mathbf{x}) : \mathbf{x} \in \mathcal{X}\}$ mit klassischen Regressionsmodellen und die Residualvariation durch einen stationären Gauß-Prozess. Sei die Kurznotation eines stochastischen Prozess $\{Z(\mathbf{x}) : \mathbf{x} \in \mathcal{X}\}$ gegeben durch $Z(\mathbf{x})$.

Definition 5.3.1 Sei $\mathcal{X} \subseteq \mathbb{R}^d$. Das Gauß-Prozess-Modell für deterministische Zielgrößen ist definiert durch

$$Y(\mathbf{x}) = \mu(\mathbf{x}) + Z(\mathbf{x}), \quad (5.4)$$

wobei $E[Y(\mathbf{x})] = \mu(\mathbf{x}) = \sum_{j=1}^p f_j(\mathbf{x})\beta_j$ mit $f_1(\cdot), \dots, f_p(\cdot)$ bekannte Regressionsfunktionen, $\boldsymbol{\beta} = (\beta_1, \dots, \beta_p)'$ der Vektor der unbekanntenen Regressionskoeffizienten und $Z(\mathbf{x})$ ein stationärer Gauß-Prozess mit Erwartungswertvektor $\mathbf{0}$ und Kovarianzfunktion $C(\mathbf{h})$.

Das Gauß-Prozess-Modell unterstellt somit, dass die räumliche Heterogenität durch zwei Arten von räumlicher Abhängigkeit beschrieben werden kann. Die Makro-Abhängigkeit wird dabei durch die deterministische Trendkomponente modelliert und die Mikro-Abhängigkeit durch die Korrelation des stationären Gauß-Prozesses beschrieben. Aus der Definition des Gauß-Prozess-Modells ergibt sich, dass $Y(\mathbf{x})$ nicht stationär ist. Für nicht deterministische Zielgrößen mit Wiederholungsmessungen wird das Modell additiv um eine Komponente $\epsilon(\mathbf{x})$ erweitert, die einen White-Noise-Prozess darstellt. In der Literatur ist dieses Drei-Komponenten-Modell auch bekannt als *räumliches Regressionsmodell*, vgl. O'Connell und Wolfinger (1997).

Das Gauß-Prozess-Modell ist vollständig durch die Wahl der Regressionsfunktionen und der Kovarianzfunktion $C(\mathbf{h})$ spezifiziert.

5.4 Korrelationsfunktionen

Da der stationäre Gauß-Prozess $Z(\mathbf{x})$ einen Erwartungswertvektor von $\mathbf{0}$ besitzt, ist dieser stochastische Prozess damit vollständig durch die Kovarianzfunktion $C(\cdot)$ bestimmt. Eine Kovarianzfunktion $C(\cdot)$ eines stationären Prozesses $Z(\mathbf{x})$ muss die folgenden Eigenschaften besitzen:

- $C(\mathbf{0}) = \sigma_Z^2 > 0$
- $C(\mathbf{h}) = C(-\mathbf{h})$

- $\sum_{i=1}^n \sum_{j=1}^n w_i w_j C(\mathbf{x}_i - \mathbf{x}_j) \geq 0$ für $n \geq 1$, beliebige Konstanten $w_1, \dots, w_n \in \mathbb{R}$ und beliebige $\mathbf{x}_1, \dots, \mathbf{x}_n \in \mathcal{X}$.

Während die zweite Eigenschaft die Symmetrie der Kovarianzfunktion verlangt, besagt die dritte Eigenschaft, dass $C(\cdot)$ positiv semidefinit sein muss. Die oben aufgelisteten Anforderungen an eine Kovarianzfunktion sind jedoch nicht konstruktiv. Eine geeignete Methode mit der gültige Kovarianzfunktionen erzeugt werden können wird in Santner et al. (2003) vorgestellt.

Eine häufig verwendete Klasse von gültigen Korrelationsfunktionen sind die Power-Exponential-Korrelationsfunktionen, da diese sehr flexible Gauß-Prozess-Modelle ermöglichen. Für isotrope Prozesse ist die Power-Exponential-Korrelationsfunktion wie folgt definiert

$$R(\mathbf{h}) = \exp \left\{ - \left(\frac{\|\mathbf{h}\|}{\theta} \right)^s \right\}, \quad (5.5)$$

wobei $\theta > 0$ und $0 < s \leq 2$.

Die Eigenschaften des Gauß Prozesses werden dabei durch die Parameter θ und s gesteuert. Um den Einfluss dieser Parameter beurteilen zu können, betrachte einen Gauß-Prozess auf $\mathcal{X} = [0, 1]$ mit Erwartungswert $\mu = 0$ und Varianz $\sigma_Z^2 = 1$. In Abbildung 5.2 ist der Einfluss des Parameters s bei konstantem Parameter θ dargestellt. Offensichtlich bestimmt der Parameter s die Glattheit der Realisierungen. Je kleiner s gewählt wird, desto schneller fällt die Korrelation bei kleinem Abstand h ab. Für $s = 2$ sind die Realisierungen eines Gauß-Prozesses sehr glatt und fast sicher unendlich oft differenzierbar. Der Parameter θ wird als *Range*-Parameter bezeichnet. Der Range gibt den Abstand $\|\mathbf{h}\|$ an, ab dem $Z(\mathbf{x})$ und $Z(\mathbf{x} + \mathbf{h})$ mit $\mathbf{x} \in \mathcal{X}$ als zumindest approximativ unabhängig betrachtet werden können. Für $s = 2$ ist dieser Range gegeben durch $\sqrt{3}\theta$, vgl. Santner et al. (2003).

In Abbildung 5.3 ist der Einfluss des Parameters θ auf die Form der Realisierungen des Gauß-Prozesses dargestellt. Falls $s = 2$, kontrolliert dieser Parameter die Anzahl der lokalen Minima und Maxima. Mit kleiner werdendem θ fällt die Korrelationsfunktion schneller gegen Null, was die Sprunghaftigkeit der Realisierungen des Gauß-Prozesses erhöht.

Eine anisotrope und trennbare Erweiterung der Power-Exponential-Korrelationsfunktion auf \mathbb{R}^d ist gegeben durch

$$R(\mathbf{h}) = \exp \left\{ - \sum_{j=1}^d \left(\frac{|h_j|}{\theta_j} \right)^{s_j} \right\}, \quad (5.6)$$

wobei $\theta_j > 0$ und $0 < s_j \leq 2$ für $j = 1, \dots, d$.

Wie anhand der obigen Abbildungen zu sehen ist, können Gauß-Prozesse mit Power-Exponential-Korrelationsfunktion sehr unterschiedliche Formen annehmen. Realisierungen von Gauß-Prozessen mit anderen Korrelationsfunktionen sind in Santner et al. (2003) abgebildet.

In Kombination mit der anisotropen Power-Exponential-Korrelationsfunktion ist es oft ausreichend, die deterministische Trendkomponente auf einen Intercept β_0 festzulegen. Solche Modelle können dennoch sehr komplexe Zusammenhänge modellieren, vgl. Jones et al. (1998). Dieses Zusammenspiel hat auch den Vorteil, dass bei kleinem Stichprobenumfang ein Overfitting vermieden wird. Wird z. B. im dreidimensionalen Versuchsraum die Trendkomponente im Gauß-Prozess-Modell durch ein Polynom zweiter Ordnung beschrieben (10 Parameter) und zudem die anisotrope Power-Exponential-Korrelationsfunktion mit festem s gewählt (4 Parameter), so sind 14 Parameter zu schätzen. Die Problematik des Overfittings beim Gauß-Prozess-Modell wurde in Schaller (2004) näher untersucht.

5.5 Parameterschätzung

In diesem Abschnitt werden Schätzer der unbekannt Parameter des Gauß-Prozess-Modells hergeleitet. Diese unbekannt Parameter sind der Vektor der Regressionskoeffizienten $\boldsymbol{\beta}$, der Vektor der Korrelationsparameter $\boldsymbol{\theta}$ und die Varianz σ_Z^2 des Gauß Prozesses $Z(\mathbf{x})$. Werden nicht deterministische Zielgrößen betrachtet, so ist zudem die Varianz σ_ϵ^2 des White-Noise-Prozesses $\epsilon(\mathbf{x})$ zu berücksichtigen. Im Folgenden beschränken wir uns zunächst auf Gauß-Prozess-Modelle deterministischer Zielgrößen. Die unbekannt Parameter des Gauß-Prozess-Modells werden mit Hilfe der Maximum-Likelihood-Methode (ML) oder der Restricted Maximum Likelihood (REML) bestimmt.

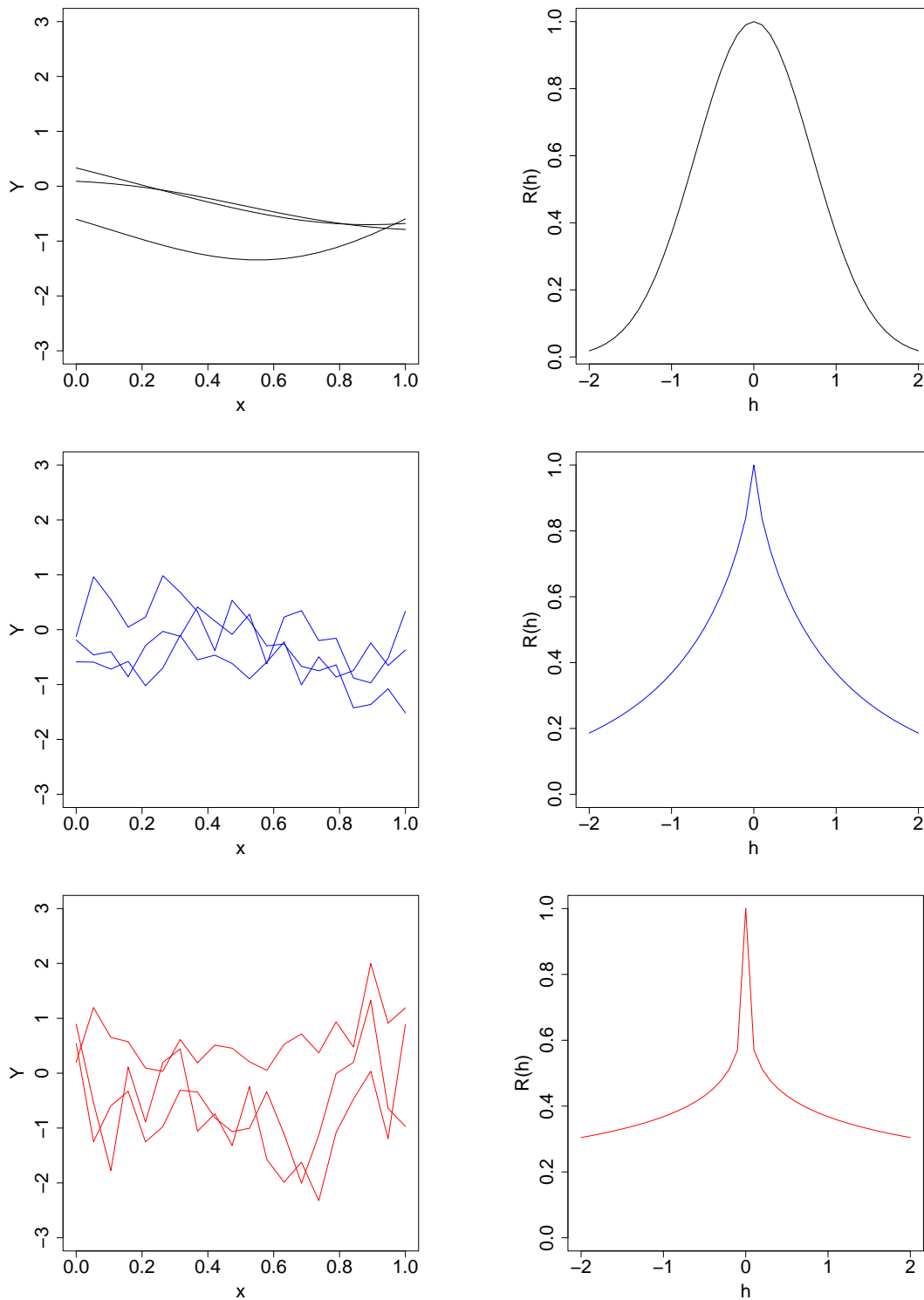


Abbildung 5.2: Drei Realisationen eines Gauß-Prozesses mit Erwartungswert $\mu = 0$ und Varianz $\sigma_Z^2 = 1$ mit einer Power-Exponential-Korrelationsfunktion mit festem $\theta = 1$ und $s = 2$ (schwarz), $s = 0.75$ (blau), $s = 0.25$ (rot)

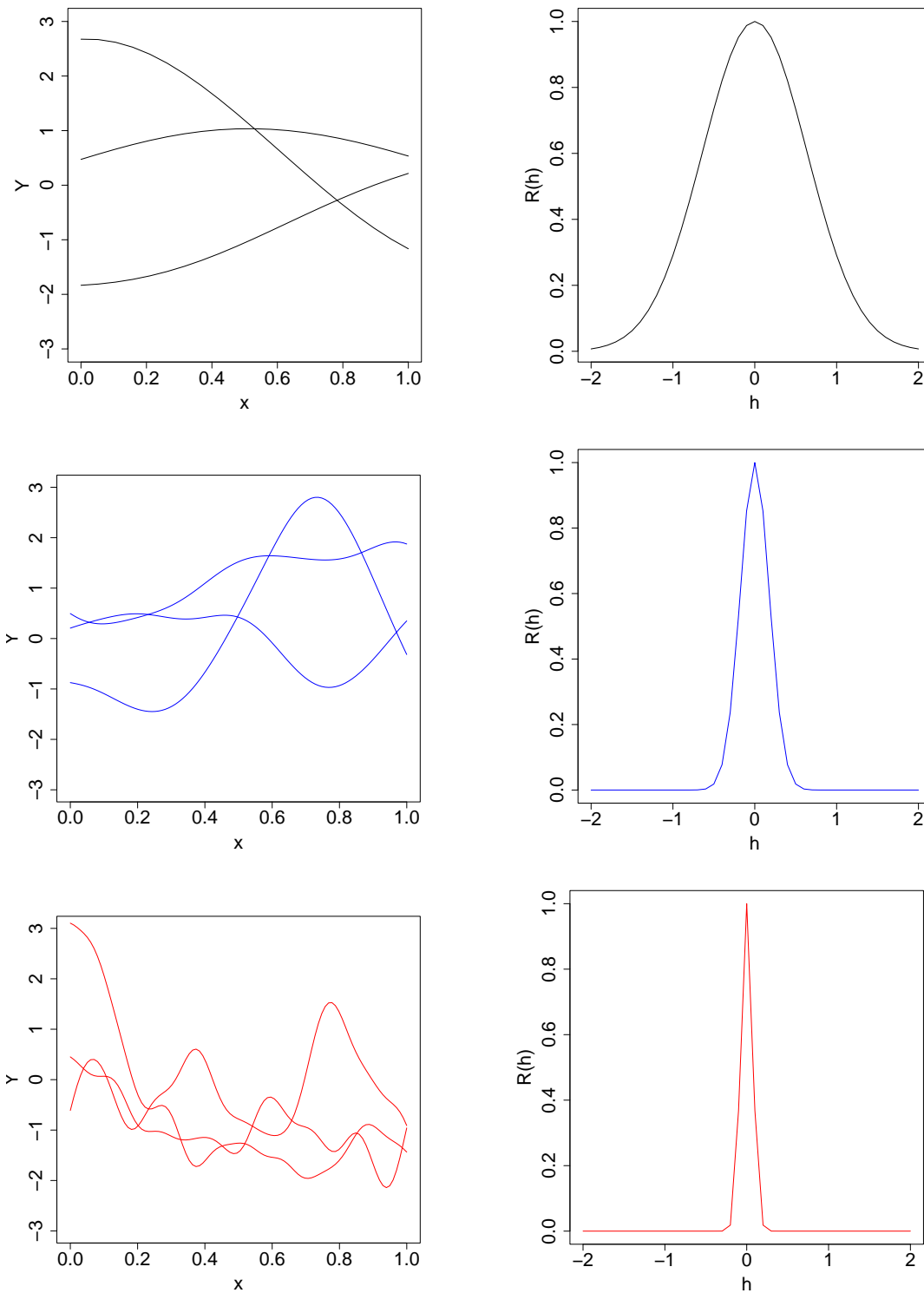


Abbildung 5.3: Drei Realisationen eines Gauß-Prozesses mit Erwartungswert $\mu = 0$ und Varianz $\sigma_Z^2 = 1$ mit einer Power-Exponential-Korrelationsfunktion mit festem $s = 2$ und $\theta = 0.9$ (schwarz), $\theta = 0.25$ (blau), $\theta = 0.1$ (rot)

5.5.1 Bekannte Korrelationsfunktion

Sei eine endliche Anzahl von Versuchen $\mathbf{x}_1, \dots, \mathbf{x}_n \in \mathcal{X}$ durchgeführt worden. Unter dem Gauß-Prozess-Modell ist der Zufallsvektor $\mathbf{Y}^{(n)} = (Y(\mathbf{x}_1), \dots, Y(\mathbf{x}_n))'$ n -dimensional multivariat normalverteilt

$$\mathbf{Y}^{(n)} \sim N_n(\mathbf{F}\boldsymbol{\beta}, \sigma_Z^2 \mathbf{R}), \quad (5.7)$$

wobei $\mathbf{F} = (\mathbf{f}(\mathbf{x}_1), \dots, \mathbf{f}(\mathbf{x}_n))'$ und \mathbf{R} die Korrelationsmatrix des Zufallsvektors $\mathbf{Y}^{(n)}$ ist. Die Log-Likelihood für die Realisationen $\mathbf{y}^{(n)}$ ist dann gegeben durch

$$l(\boldsymbol{\beta}, \boldsymbol{\theta}, \sigma_Z^2) = -\frac{1}{2} \left[n \log(2\pi) + n \log(\sigma_Z^2) + \log(|\mathbf{R}|) + \frac{(\mathbf{y}^{(n)} - \mathbf{F}\boldsymbol{\beta})' \mathbf{R}^{-1} (\mathbf{y}^{(n)} - \mathbf{F}\boldsymbol{\beta})}{\sigma_Z^2} \right],$$

wobei $|\mathbf{R}|$ die Determinante von \mathbf{R} bezeichnet. Durch partielles Ableiten der Log-Likelihood nach $\boldsymbol{\beta}$ und Nullsetzen des Ausdrucks ergibt sich der ML-Schätzer von $\boldsymbol{\beta}$ zu

$$\hat{\boldsymbol{\beta}} = (\mathbf{F}' \mathbf{R}^{-1} \mathbf{F})^{-1} \mathbf{F}' \mathbf{R}^{-1} \mathbf{y}^{(n)}. \quad (5.8)$$

Der ML-Schätzer von $\boldsymbol{\beta}$ ist also gleich dem verallgemeinerten Gauß-Markov Schätzer. Dieser Schätzer ist zudem der beste linear unverzerzte Schätzer (BLUE) für $\boldsymbol{\beta}$. Den ML-Schätzer für σ_Z^2 erhält man, indem man $\hat{\boldsymbol{\beta}}$ in die Log-Likelihood des Gauß-Prozess-Modells einsetzt und diesen Ausdruck bezüglich σ_Z^2 maximiert. Der resultierende ML-Schätzer lautet

$$\hat{\sigma}_Z^2 = \frac{1}{n} (\mathbf{y}^{(n)} - \mathbf{F}\hat{\boldsymbol{\beta}})' \mathbf{R}^{-1} (\mathbf{y}^{(n)} - \mathbf{F}\hat{\boldsymbol{\beta}}). \quad (5.9)$$

Für nicht deterministische Zielgrößen wird das Modell additiv um einen White-Noise-Prozess ergänzt. Der Zufallsvektor $\mathbf{Y}^{(n)} = (Y(\mathbf{x}_1), \dots, Y(\mathbf{x}_n))'$ ist dann weiterhin n -dimensional multivariat normalverteilt mit Erwartungswertvektor $\mathbf{F}\boldsymbol{\beta}$. Die Kovarianzmatrix verändert sich nun zu $\mathbf{V} = \sigma_Z^2 \mathbf{R} + \sigma_\epsilon^2 \mathbf{I} = \sigma_\epsilon^2 \left(\frac{\sigma_Z^2}{\sigma_\epsilon^2} \mathbf{R} + \mathbf{I} \right)$. Die ML-Schätzer für den Vektor der Regressionskoeffizienten $\boldsymbol{\beta}$ und für σ_ϵ^2 lassen sich äquivalent zum Modell mit deterministischen Zielgrößen über die Log-Likelihood-Funktion herleiten. Mit $\mathbf{W} = \frac{\sigma_Z^2}{\sigma_\epsilon^2} \mathbf{R} + \mathbf{I}$ ergeben sich die ML-Schätzer aus Formel (5.8) und (5.9), wobei \mathbf{R} durch \mathbf{W} ersetzt wird.

5.5.2 Unbekannte Korrelationsfunktion

Der ML-Schätzer für $\boldsymbol{\beta}$ sowie der ML-Schätzer für σ_Z^2 hängen von den unbekanntem Modellparametern nur von den Korrelationsparametern ab. Um diesen Sachverhalt zu verdeutlichen, wird die Schreibweise $\hat{\boldsymbol{\beta}}(\boldsymbol{\theta})$ und $\hat{\sigma}_Z^2(\boldsymbol{\theta})$ verwendet. Um den ML-Schätzer von $\boldsymbol{\theta}$ zu bestimmen, wird nun $\hat{\boldsymbol{\beta}}(\boldsymbol{\theta})$ und $\hat{\sigma}_Z^2(\boldsymbol{\theta})$ aus Formel (5.8) und (5.9) in die Log-Likelihood des Gauß-Prozess-Modells eingesetzt. Die Log-Likelihood lässt sich dann auf den folgenden Ausdruck reduzieren

$$l(\hat{\boldsymbol{\beta}}, \hat{\sigma}_Z^2, \boldsymbol{\theta}) = -\frac{1}{2} \left[n \log(2\pi) + n \log(\hat{\sigma}_Z^2(\boldsymbol{\theta})) + \log(|\mathbf{R}(\boldsymbol{\theta})| + n) \right]. \quad (5.10)$$

Der ML-Schätzer für $\boldsymbol{\theta}$ ergibt sich durch Maximierung dieses Ausdrucks. Ein geschlossener Ausdruck für den ML-Schätzer existiert nicht. Das Maximum des Ausdrucks in (5.10) wird mit Hilfe von numerischen Verfahren bestimmt. Wolfinger et al. (1994) schlagen vor, den *ridge-stabilized* Newton Raphson Algorithmus zu verwenden. Im Gegensatz zum klassischen Newton Raphson Algorithmus wird zu der Korrelationsmatrix \mathbf{R} ein bestimmtes Vielfaches der Einheitsmatrix addiert, um Unstabilitäten bei der Invertierung zu verhindern. Die ML-Schätzer für $\boldsymbol{\beta}$ sowie σ_Z^2 erhält man indem man in Formel 5.8 und 5.9 die Korrelationsmatrix \mathbf{R} durch ihren Schätzer $\hat{\mathbf{R}}$ ersetzt.

Die ML-Schätzer von σ_Z^2 und $\boldsymbol{\theta}$ können jedoch insbesondere bei kleinen Stichproben verzerrt sein, vgl. Mardia und Marshall (1984). Der Grund dafür ist, dass der Verlust an Freiheitsgraden durch das Schätzen von $\boldsymbol{\beta}$ nicht berücksichtigt wird. Je höher die Dimension von $\boldsymbol{\beta}$ ist, desto stärker ist die Verzerrung von σ_Z^2 und $\boldsymbol{\theta}$. Durch die Verwendung der Restricted (Residual) Maximum Likelihood (REML) kann dieses Problem gelöst werden, vgl. Patterson und Thompson (1974). Sei der Spaltenrang von \mathbf{F} voll, also $rg(\mathbf{F}) = p$. Die Idee dieser Methode ist, eine Matrix \mathbf{C} von vollem Zeilenrang $n - p$ zu bestimmen mit $\mathbf{CF} = \mathbf{0}$. Der REML-Schätzer von $\boldsymbol{\theta}$ maximiert dann die Likelihood der transformierten Daten $\mathbf{CY}^{(n)}$. Da $\mathbf{CY}^{(n)} \sim N(\mathbf{CF}\boldsymbol{\beta} = \mathbf{0}, \sigma_Z^2 \mathbf{CR}(\boldsymbol{\theta})\mathbf{C}')$, hängt die zugehörige Likelihood nicht von $\boldsymbol{\beta}$ ab. Die Matrix $\mathbf{CY}^{(n)}$ kann als Menge von Linearkombinationen der $\mathbf{Y}^{(n)}$ aufgefasst werden, wobei jede Linearkombination orthogonal zu dem Erwartungswertvektor $\mathbf{F}\boldsymbol{\beta}$ von $\mathbf{Y}^{(n)}$ ist. Wie bei der ML-Methode wird der Schätzer von $\boldsymbol{\theta}$

nun numerisch bestimmt. Der REML-Schätzer für σ_Z^2 ergibt sich zu

$$\tilde{\sigma}_Z^2 = \frac{1}{n-p} (\mathbf{y}^{(n)} - \mathbf{F}\hat{\boldsymbol{\beta}})' \mathbf{R}^{-1} (\mathbf{y}^{(n)} - \mathbf{F}\hat{\boldsymbol{\beta}}). \quad (5.11)$$

5.6 Modellvorhersage

Ziel der Modellierung des Gauß-Prozess-Modells ist die Zielgröße an unbeobachteten Stellen des Versuchsraums \mathcal{X} zu prognostizieren. In diesem Kapitel wird aus der Klasse der linearen und unverzerrten Prädiktoren der beste Prädiktor hergeleitet.

Definition 5.6.1 Ein Prädiktor $\hat{Y}(\mathbf{x}_0)$ heißt *linear*, falls

$$\hat{Y}(\mathbf{x}_0) = \sum_{i=1}^n l_i Y(\mathbf{x}_i) + l_0, \quad (5.12)$$

wobei l_i mit $i = 0, 1, \dots, n$ beliebige feste Zahlen sind.

Definition 5.6.2 Ein Prädiktor $\hat{Y}(\mathbf{x}_0)$ heißt *lokal unverzerrt* in \mathbf{x}_0 , falls

$$E[\hat{Y}(\mathbf{x}_0)] = E[Y(\mathbf{x}_0)]. \quad (5.13)$$

Ein Prädiktor ist unverzerrt, falls er für alle $\mathbf{x}_0 \in \mathcal{X}$ lokal unverzerrt ist. Um den besten linear und unverzerrten Prädiktor bestimmen zu können, ist zunächst ein Kriterium zu definieren, mit dem die Vorhersagegüte verschiedener Prädiktoren verglichen werden kann. Ein häufig verwendetes Kriterium ist der mittlere quadratische Vorhersagefehler (Mean Squared Prediction Error). Gegeben sei der Beobachtungsvektor $\mathbf{Y}^{(n)} = (Y(\mathbf{x}_1), \dots, Y(\mathbf{x}_n))'$ mit $\mathbf{x}_1, \dots, \mathbf{x}_n \in \mathcal{X}$. Bezeichne mit $\hat{Y}(\mathbf{x}_0)|\mathbf{Y}^{(n)}$ einen Prädiktor der Zielgröße in einem neuen Punkt $\mathbf{x}_0 \in \mathcal{X}$ gegeben $\mathbf{Y}^{(n)}$. Der mittlere quadratische Vorhersagefehler in \mathbf{x}_0 ist dann definiert als

$$\text{MSPE}[\hat{Y}(\mathbf{x}_0)|\mathbf{Y}^{(n)}] = E\left[\left(Y(\mathbf{x}_0) - \hat{Y}(\mathbf{x}_0)|\mathbf{Y}^{(n)}\right)^2\right]. \quad (5.14)$$

Ein Prädiktor aus einer Klasse zulässiger Prädiktoren wird als *bester* dieser Klasse bezeichnet, falls der den MSPE minimiert.

5.6.1 Bekannte Korrelationsfunktion

Satz 5.6.1 *Gegeben sei das Modell aus Definition 5.3.1, wobei Z nun ein beliebiger schwach stationärer stochastischer Prozess ist. Falls die Korrelationsfunktion bekannt ist, ist der beste linear unverzerrte Schätzer (BLUP) für $y(\mathbf{x}_0)$ gegeben durch*

$$\hat{Y}(\mathbf{x}_0)|\mathbf{Y}^{(n)} = \mathbf{f}(\mathbf{x}_0)' \hat{\boldsymbol{\beta}} + \mathbf{r}(\mathbf{x}_0)' \mathbf{R}^{-1} (\mathbf{Y}^{(n)} - \mathbf{F} \hat{\boldsymbol{\beta}}). \quad (5.15)$$

Der Beweis dieses Satzes ist in Schonlau (1997) gegeben. Für nicht deterministische Zielgrößen ergibt sich der BLUP aus Formel (5.15), wobei \mathbf{R} durch $\mathbf{U} = \mathbf{R} + \sigma_\epsilon^2 / \sigma_Z^2 \mathbf{I}$ ersetzt wird.

Es ist bekannt, dass für beliebige stochastische Prozesse der beste Prädiktor an der Stelle \mathbf{x}_0 gleich dem bedingten Erwartungswert von $Y(\mathbf{x}_0)$ gegeben $\mathbf{Y}^{(n)}$ ist, vgl. Cressie (1993). Dies ist ein intuitiv einleuchtender Schätzer, da $Y(\mathbf{x}_0)$ und $\mathbf{Y}^{(n)}$ korreliert sind, und somit $\mathbf{Y}^{(n)}$ Informationen über $Y(\mathbf{x}_0)$ enthält. Um den bedingten Erwartungswert berechnen zu können, ist zunächst die gemeinsame Verteilung von $\mathbf{Y}^{(n)}$ und $Y(\mathbf{x}_0)$ zu definieren. Diese ist im Gauß-Prozess-Modell gegeben durch die multivariate Normalverteilung

$$\begin{pmatrix} Y(\mathbf{x}_0) \\ \mathbf{Y}^{(n)} \end{pmatrix} \sim N_{1+n} \left[\begin{pmatrix} \mathbf{f}(\mathbf{x}_0)' \\ \mathbf{F} \end{pmatrix} \boldsymbol{\beta}, \sigma_Z^2 \begin{pmatrix} 1 & \mathbf{r}(\mathbf{x}_0)' \\ \mathbf{r}(\mathbf{x}_0) & \mathbf{R} \end{pmatrix} \right], \quad (5.16)$$

wobei $\mathbf{r}(\mathbf{x}_0)' = (R(\mathbf{x}_0, \mathbf{x}_1), \dots, R(\mathbf{x}_0, \mathbf{x}_n))$ dem Vektor der Korrelationen zwischen $Y(\mathbf{x}_0)$ und den Beobachtungen $\mathbf{Y}^{(n)}$ entspricht. Mit Hilfe des Satzes A.1 lässt sich nun zeigen, dass der bedingte Erwartungswert die folgende Form annimmt

$$E [Y(\mathbf{x}_0)|\mathbf{Y}^{(n)}] = \mathbf{f}(\mathbf{x}_0)' \boldsymbol{\beta} + \mathbf{r}(\mathbf{x}_0)' \mathbf{R}^{-1} (\mathbf{Y}^{(n)} - \mathbf{F} \boldsymbol{\beta}). \quad (5.17)$$

Somit ist für Gauß-Prozesse der BLUP aus Formel (5.15) erwartungstreu für den besten Prädiktor, der Kenntnis von $\boldsymbol{\beta}$ voraussetzt.

Für deterministische Zielgrößen ist der MSPE des BLUP im Punkt \mathbf{x}_0 gegeben durch

$$MSPE(\hat{Y}(\mathbf{x}_0)|\mathbf{Y}^{(n)}) = \sigma_Z^2 \left(1 - \mathbf{r}(\mathbf{x}_0)' \mathbf{R}^{-1} \mathbf{r}(\mathbf{x}_0) + \mathbf{a}' (\mathbf{F}' \mathbf{R}^{-1} \mathbf{F})^{-1} \mathbf{a} \right), \quad (5.18)$$

wobei $\mathbf{a} = \mathbf{f}(\mathbf{x}_0) - \mathbf{F}'\mathbf{R}^{-1}\mathbf{r}(\mathbf{x}_0)$, vgl. Santner (2003), Seite 93. Für nicht deterministische Zielgrößen ist der MSPE

$$MSPE(\hat{Y}(\mathbf{x}_0)|\mathbf{Y}^{(n)}) = \sigma_Z^2 \left(1 + \frac{\sigma_\epsilon^2}{\sigma_Z^2} - \mathbf{r}(\mathbf{x}_0)'\mathbf{U}^{-1}\mathbf{r}(\mathbf{x}_0) + \mathbf{a}'(\mathbf{F}'\mathbf{U}^{-1}\mathbf{F})^{-1}\mathbf{a} \right), (5.19)$$

wobei $\mathbf{a} = \mathbf{f}(\mathbf{x}_0) - \mathbf{F}'\mathbf{U}^{-1}\mathbf{r}(\mathbf{x}_0)$, vgl. O'Connell und Wolfinger (1997). Während sich der MSPE an beobachteten Versuchspunkten bei deterministischen Zielgrößen auf Null reduziert, ist dieser für nicht deterministische Zielgrößen σ_ϵ^2 .

5.6.2 Unbekannte Korrelationsfunktion

Da in der Praxis die Parameter der Korrelationsfunktion $R(\cdot)$ unbekannt sind, werden zur Berechnung von $\hat{Y}(\mathbf{x}_0)|\mathbf{Y}^{(n)}$ in Formel (5.15) die Schätzer $\hat{\mathbf{r}}(\mathbf{x}_0)$ und $\hat{\mathbf{R}}$ bzw. $\hat{\mathbf{U}}$ eingesetzt. Der daraus resultierende Schätzer $\hat{Y}(\mathbf{x}_0)|\mathbf{Y}^{(n)}$ wird als der empirisch beste linear unverzerrte Schätzer (EBLUP) bezeichnet. Zu beachten ist, dass solch ein EBLUP in der Regel jedoch nicht mehr linear bzw. unverzerrt ist.

Bei unbekannter Korrelation ergibt sich ein Schätzer der Varianz des EBLUP, indem in Formel (5.18) die Schätzer $\hat{\sigma}_Z^2$, $\hat{\mathbf{r}}(\mathbf{x}_0)$ und $\hat{\mathbf{R}}$, bzw. in Formel (5.19) die Schätzer $\hat{\sigma}_Z^2$, $\hat{\sigma}_\epsilon^2$, $\hat{\mathbf{r}}(\mathbf{x}_0)$ und $\hat{\mathbf{U}}$ eingesetzt werden.

5.6.3 Interpretation des BLUP

Der BLUP ergibt sich additiv aus der Prognose des Erwartungswertvektors von $Y(\mathbf{x})$ und dem Korrekturterm $\mathbf{r}(\mathbf{x}_0)' \mathbf{R}^{-1} (\mathbf{Y}^{(n)} - \mathbf{F} \hat{\boldsymbol{\beta}})$, vgl. Formel (5.15). Für deterministische Zielgrößen lässt sich zeigen, dass dieser Korrekturterm eine Interpolation der Daten $\mathbf{Y}^{(n)}$ erzwingt. Sei $\mathbf{x}_0 = \mathbf{x}_i$ für ein beliebiges i mit $i = 1, \dots, n$. Dann gilt, dass $\mathbf{f}(\mathbf{x}_0)' = \mathbf{f}(\mathbf{x}_i)'$ und $\mathbf{r}(\mathbf{x}_0)' = (R(\mathbf{x}_i - \mathbf{x}_1), \dots, R(\mathbf{x}_i - \mathbf{x}_n))$. Da $\mathbf{r}(\mathbf{x}_0)'$ dann der i -ten Zeile von \mathbf{R} entspricht ist $\mathbf{R}^{-1} \mathbf{r}(\mathbf{x}_0)$ gleich dem i -ten Einheitsvektor. Damit gilt

$$\begin{aligned} \hat{Y}(\mathbf{x}_0) &= \mathbf{f}(\mathbf{x}_0)' \hat{\boldsymbol{\beta}} + \mathbf{r}(\mathbf{x}_0)' \mathbf{R}^{-1} (\mathbf{Y}^{(n)} - \mathbf{F} \hat{\boldsymbol{\beta}}) \\ &= \mathbf{f}(\mathbf{x}_i)' \hat{\boldsymbol{\beta}} + (Y(\mathbf{x}_i) - \mathbf{f}(\mathbf{x}_i)' \hat{\boldsymbol{\beta}}) = Y(\mathbf{x}_i). \end{aligned}$$

Die Vorhersage in einem schon beobachteten Versuchspunkt ist somit gleich der dort erhobenen Zielgröße. Der Korrekturterm für einen noch nicht beobachteten Punkt lässt sich als Linearkombination der Residuen $(\mathbf{Y}^{(n)} - \mathbf{F} \hat{\boldsymbol{\beta}})$ interpretieren

$$\mathbf{r}(\mathbf{x}_0)' \mathbf{R}^{-1} (\mathbf{Y}^{(n)} - \mathbf{F} \hat{\boldsymbol{\beta}}) = \sum_{i=1}^n c_i(\mathbf{x}_0) (\mathbf{Y}^{(n)} - \mathbf{F} \hat{\boldsymbol{\beta}})_i, \quad (5.20)$$

wobei das Gewicht $c_i(\mathbf{x}_0)$ dem i -ten Element von $\mathbf{r}(\mathbf{x}_0)' \mathbf{R}^{-1}$ entspricht und $(\mathbf{Y}^{(n)} - \mathbf{F} \hat{\boldsymbol{\beta}})_i$ das i -te Residuum des Gauß-Prozess-Modells definiert. Damit ist der BLUP definiert durch die Vorhersage des Regressionsterms $\mathbf{f}(\mathbf{x}_i)' \hat{\boldsymbol{\beta}}$ plus einer gewichteten Summe der Residuen des Modells. Die Gewichtung hängt dabei von der Lage des Punkts \mathbf{x}_0 ab. In Abbildung 5.4 a) ist die Vorhersage eines Gauß-Prozess-Modells basierend auf sechs Versuchspunkten abgebildet. Der Regressionsterm wurde in diesem Modell auf einen Intercept β_0 festgelegt und die Power Exponential Korrelationsfunktion mit $s = 2$ gewählt. Als Schätzer für β_0 ergab sich der Wert 0.3 und für den Korrelationsparameter θ der Wert 0.422. Die Residuen des Modells sind durch gestrichelte Linien dargestellt.

In Abbildung 5.4 b) sind die Gewichte der Residuen für den Punkt $x_0 = 0.5$ eingezeichnet. Die Prognose der Zielgröße in diesem Punkt beruht somit hauptsächlich auf den beiden Residuen der umliegenden Versuchspunkte $x = 0.4$ und $x = 0.55$.

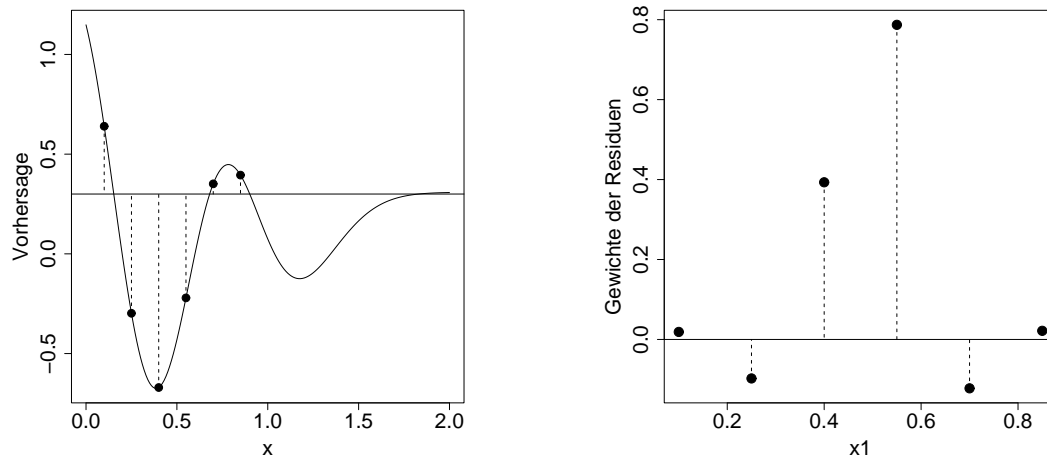


Abbildung 5.4: a) Residuen des Gauß-Prozess-Modells, b) Gewichte der Residuen für $\mathbf{x}_0 = 0.5$

Wie anhand dieses Beispiels deutlich wird, können die Gewichte der Residuen auch negativ sein. Da die Gewichte von der Korrelation des Punktes \mathbf{x}_0 zu den Versuchspunkten $\mathbf{x}_1, \dots, \mathbf{x}_n$ abhängen, wird die Vorhersage umso stärker in Richtung der Vorhersage des Regressionsterms gezogen, je größer der Abstand des Punktes \mathbf{x}_0 zu den Versuchspunkten ist. Wie in Kapitel 5.4 beschrieben, ist der Range für die Power Exponential Korrelationsfunktion mit $s = 2$ gleich $\sqrt{3}\theta$. Für das in Abbildung 5.4 betrachtete Beispiel werden Punkte, deren Abstand größer als $\sqrt{3} \cdot 0.422 = 0.73$ ist, somit als approximativ unabhängig betrachtet. Da der größte Versuchspunkt bei $x = 0.85$ lag, gilt dies hier für Werte von $x_0 > 1.6$. Für diese x_0 -Werte reduziert sich der Korrekturterm fast auf 0 und die Prognose der Zielgröße ist damit ungefähr identisch mit dem Regressionsschätzer $\hat{\beta}_0 = 0.3$, vgl. Abbildung 5.4 a).

5.7 Modellvalidierung

Bevor ein Gauß-Prozess-Modell zur Prognose verwendet wird, ist die Modellgültigkeit zu überprüfen. Zu diesem Zweck lassen sich verschiedene diagnostische Plots einsetzen, die auf der Methode der Kreuzvalidierung beruhen, vgl. Jones et al. (1998). Diese Methode wurde ursprünglich von Stone (1974) und Geisser (1975) entwickelt, um die Vorhersagegüte eines Modells beurteilen zu

können. Bei dieser Methode wird jeweils eine Beobachtung eliminiert und aus den restlichen Daten das Modell angepasst. Die Vorhersagegüte lässt sich grafisch darstellen, indem die Vorhersage gegen die wahren Werte der Zielgröße geplottet wird. Bei einem gutem Modell liegen die Punkte entlang der Diagonalen. Sei $\hat{y}_{(-1)}(\mathbf{x}_i)$ die Vorhersage an der Stelle \mathbf{x}_i ohne Berücksichtigung von $y(\mathbf{x}_i)$ und $\sigma_{\hat{y}_{(-1)}}^2(\mathbf{x}_i)$ die geschätzte Varianz dieser Vorhersage. Falls das Modell gültig ist, gilt approximativ

$$\frac{y(\mathbf{x}_i) - \hat{y}_{(-1)}(\mathbf{x}_i)}{\sqrt{\sigma_{\hat{y}_{(-1)}}^2(\mathbf{x}_i)}} \sim N(0, 1) \quad \text{mit } 1 \leq i \leq n.$$

Für jedes i mit $i = 1, \dots, n$ sollte diese Größe somit ungefähr im Intervall $[-3, 3]$ liegen. Ist die Vorhersagegüte nicht gut, kann eine Transformation der Zielgröße oft hilfreich sein, vgl. Jones et al. (1998).

Da der Anwender sehr flexibel ist bei der Spezifikation des Gauß-Prozess-Modells, sind Kriterien zu verwenden, die einen Vergleich verschiedener Modelle ermöglichen. Um die Vorhersagegüte unterschiedlicher Modelle beurteilen zu können, wird der Mean Squared Crossvalidation Error (MSCVE) betrachtet

$$\frac{1}{n} \sum_{i=1}^n \left(y(\mathbf{x}_i) - \hat{y}_{(-i)}(\mathbf{x}_i) \right)^2.$$

Neben der Vorhersagegüte sollten auch Maße zur Beurteilung des Fits berücksichtigt werden. Klassische Maße wie das Akaike-Kriterium oder Schwarz-Kriterium beruhen dabei auf dem maximierten Likelihood-Wert l^* des Modells. Bezeichne die Deviance D mit $D = -2\log(l^*)$. Das Akaike-Kriterium (AIC) ist definiert als $AIC = D + 2u$, wobei u die Anzahl der zu schätzenden Modellparameter ist, vgl. Akaike (1974). Dabei ist der Fit des Modells um so besser, je kleiner das AIC -Kriterium ist. Der zweite Summand des AIC -Kriteriums kann als Strafterm interpretiert werden, der von der Anzahl der zu schätzenden Parameter u abhängt. Für das Schwarz-Kriterium (BIC) gilt $BIC = D + \log(n)u$, wobei n die Anzahl der Versuchspunkte definiert, vgl. Schwarz (1975). Ein Vergleich der beiden Kriterien zeigt, dass das BIC -Kriterium für $n > 7$ die Anzahl der zu schätzenden Parameter stärker bestraft.

Kapitel 6

Multivariate Expected Improvement

6.1 Das Expected-Improvement-Kriterium

Bedingt durch den wissenschaftlichen und industriellen Fortschritt ist es möglich, neue Prozesse zu entwickeln bzw. das Potenzial bestehender Prozesse zu verbessern. Als Konsequenz ist aber eine steigende Komplexität der Prozesse zu beobachten. Aus diesem Grund sind die Methoden zur Prozessoptimierung an die Anforderungen dieser Prozesse anzupassen. Ein Ansatz, der sich in der Praxis etabliert hat, wurde ursprünglich für Computerexperimente entwickelt, lässt sich aber auch auf reale Prozesse anwenden. Dabei werden raumfüllende Versuchspläne in Kombination mit Gauß-Prozess-Modellen verwendet. Nachteil dieses Ansatzes ist jedoch, dass der benötigte Stichprobenumfang sehr groß sein kann. Grund dafür ist, dass die gleiche Stichprobendichte im gesamten Versuchsraum verwendet wird. Sinnvoller wäre es, sich bei der Prozessoptimierung auf viel versprechende Bereiche im Versuchsraum zu konzentrieren. Bei der in Abbildung 5.1 dargestellten zweidimensionalen komplexen Funktion ist der Funktionswert in der rechten Hälfte des abgebildeten Versuchsraums nahezu konstant, während in der linken Hälfte der Funktionswert stark variiert. Um die Zusammenhänge effizient zu erfassen, sollte in der rechten Hälfte des abgebildeten

Versuchsraums eine geringere Stichprobendichte als in der linken Hälfte verwendet werden.

Eine alternative Möglichkeit ist ein sequentielles Vorgehen. Auf diese Weise kann die aus den Vorschriften gegebene Information genutzt werden, um sich auf interessierende Bereiche des Versuchsraums zu konzentrieren. Es gibt eine Vielzahl von Methoden, die mit Hilfe von unterschiedlichen Ansätzen den Versuchsraum schrittweise einschränken, vgl. Wang (2003), Osio und Amon (1996) und Wang und Simpson (2004). Der Nachteil dieser Methoden ist jedoch, dass die Einschränkung des Versuchsraums irreversibel ist und der Aspekt der Unsicherheit der Modelle nicht berücksichtigt wird. Jones et al. (1998) stellen einen sequentiellen Ansatz vor, der diese beiden Nachteile nicht besitzt.

Die Idee dieses Ansatzes ist es, mit Hilfe eines ersten raumfüllenden Versuchsplans ein Modell zu erstellen, das die Zusammenhänge im Wesentlichen beschreibt, aber keine exakte Approximation liefert. Jones et al. (1998) empfehlen einen Versuchsumfang von $10 \cdot d$ für das Initialdesign, wobei d die Anzahl der Faktoren angibt. Dieses Initialmodell wird anschließend verfeinert, indem sequentiell Versuchspunkte hinzugefügt werden. Nach jedem Iterationsschritt wird das Modell neu angepasst. Das Kriterium zur Wahl neuer Versuchspunkte basiert auf der Expected Improvement und platziert neue Versuchspunkte in viel versprechenden Bereichen. Im Gegensatz zu anderen Kriterien berücksichtigt die Expected Improvement nicht nur Bereiche mit sehr guten Vorhersagen, sondern auch Bereiche mit großer Unsicherheit. Dem Anwender ist es dabei möglich, die Gewichtung zwischen der so genannten lokalen und globalen Suche zu steuern. Im Folgenden wird das Expected-Improvement-Kriterium vorgestellt und deren Anwendung an einem Simulationsbeispiel verdeutlicht.

Ohne Beschränkung der Allgemeinheit sei ein Maximierungsproblem gegeben, für das im d -dimensionalen Versuchsraum ein Versuchsplan vom Umfang n durchgeführt worden ist. Basierend auf den beobachteten Zielgrößen $\mathbf{y}^{(n)} = (y(\mathbf{x}_1), \dots, y(\mathbf{x}_n))'$ wird ein Gauß-Prozess-Modell angepasst, mit dem in jedem Punkt \mathbf{x} im Versuchsraum die Schätzer der Zielgröße und deren Varianz bestimmt werden. Bezeichne dazu die Prognose der Zielgröße im Punkt \mathbf{x} mit $\hat{y}(\mathbf{x})$ und den Schätzer für die Varianz der Prognose mit $s_{\hat{y}(\mathbf{x})}^2$.

In Abbildung 6.1 a) wurde beispielhaft ein Gauß-Prozess-Modell basierend auf vier Versuchspunkten angepasst. Zudem ist in dieser Abbildung die Standardabweichung der Prognose $s_{\hat{y}(\mathbf{x})}$ angegeben.

Die Unsicherheit der Vorhersage an einer Stelle \mathbf{x} wird nun berücksichtigt, indem unterstellt wird, dass die Zielgröße $y(\mathbf{x})$ gegeben $\mathbf{y}^{(n)}$ normalverteilt sei mit der Modellvorhersage als Mittelwert und der Vorhersagevarianz als Varianz, also $(Y(\mathbf{x})|\mathbf{y}^{(n)}) \sim N(\hat{y}(\mathbf{x}), s_{\hat{y}(\mathbf{x})}^2)$. Die resultierende Unsicherheitsdichte ist in Abbildung 6.1 b) für die beiden Versuchspunkte $x = 4.5$ und $x = 8$ angegeben. Im Versuchspunkt $x = 4.5$ ist die Vorhersage des Modells z. B. $\hat{y}(\mathbf{x}) = 6.55$ mit $s_{\hat{y}(\mathbf{x})} = 0.5$.

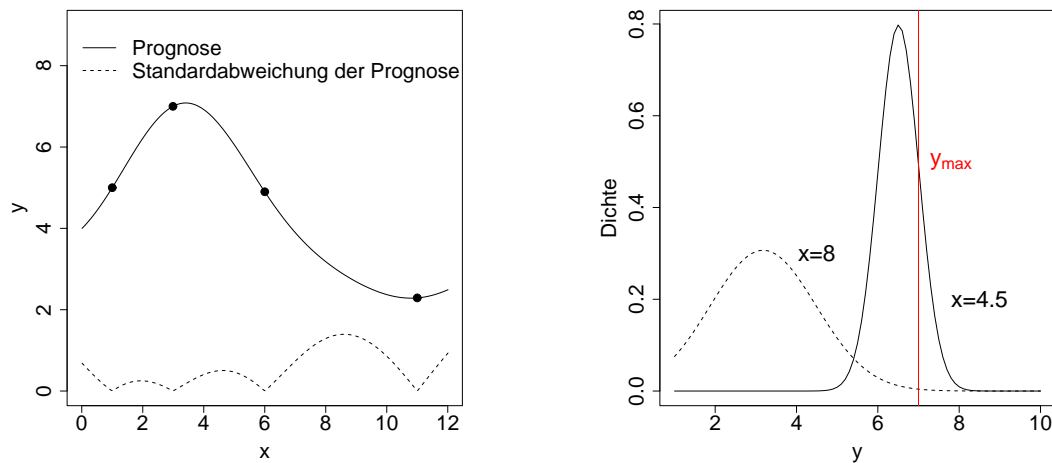


Abbildung 6.1: a) Anpassung ein Gauß-Prozess-Modells, b) resultierende Unsicherheitsverteilung für verschiedene x

Sei $y_{max} = \max(y(\mathbf{x}_1), \dots, y(\mathbf{x}_n))$ das aktuell beobachtete Maximum der Zielgröße. Für das Beispiel aus Abbildung 6.1 a) ist $y_{max} = 7$. In Abbildung 6.1 b) ist diese Größe durch eine rote Linie symbolisiert. Um das Optimierungspotential der Zielgröße bezüglich des aktuell beobachteten Maximums y_{max} für jedes $x \in \mathcal{X}$ bestimmen zu können, wird der Begriff der *Improvement* eingeführt. Die Improvement sei für y -Werte größer als y_{max} definiert als der Abstand von y zu y_{max} . Für y -Werte kleiner gleich y_{max} ist die Improvement 0. Formal lässt sich

die Improvement in einem Punkt \mathbf{x} definieren als

$$I(\mathbf{x}) = \begin{cases} Y(\mathbf{x}) - y_{max} & , \text{ falls } Y(\mathbf{x}) > y_{max} \\ 0 & , \text{ falls } Y(\mathbf{x}) \leq y_{max} \end{cases} . \quad (6.1)$$

Da $I(\mathbf{x})$ eine Zufallsvariable ist, lässt sich nun die erwartete Verbesserung berechnen. Formal entspricht dies dem bedingten Erwartungswertes von $I(\mathbf{x})$ gegeben die Beobachtungen $\mathbf{y}^{(n)}$:

$$E[I(\mathbf{x})|\mathbf{y}^{(n)}] = \int_{y_{max}}^{\infty} (z - y_{max}) f_{Y(\mathbf{x})|\mathbf{y}^{(n)}}(z) dz. \quad (6.2)$$

Eine hohe Expected Improvement ist in Bereichen zu erwarten, in denen die Vorhersage sehr gut ist und/oder die Modellunsicherheit groß ist. Für das Beispiel aus Abbildung 6.1 ist das Optimierungspotenzial für $x = 4.5$ größer als für $x = 8$. Da für deterministische Computereperimente in beobachteten Versuchspunkten die Vorhersagevarianz Null ist, ist die Expected Improvement dort ebenfalls Null.

Satz 6.1.1 *Die Expected Improvement lässt sich in geschlossener Form darstellen als*

$$E[I(\mathbf{x})|\mathbf{y}^{(n)}] = s_{\hat{y}(\mathbf{x})} \phi\left(\frac{y_{max} - \hat{y}(\mathbf{x})}{s_{\hat{y}(\mathbf{x})}}\right) + (y_{max} - \hat{y}(\mathbf{x})) \left(1 - \Phi\left(\frac{y_{max} - \hat{y}(\mathbf{x})}{s_{\hat{y}(\mathbf{x})}}\right)\right),$$

wobei Φ und ϕ die Verteilungsfunktion und die Dichtefunktion der Standardnormalverteilung definieren.

Beweis Mit Hilfe der Substitutionsregel lässt sich der bedingte Erwartungswert aus (6.2) umformen. Definiere dazu $z := \frac{y - \hat{y}(\mathbf{x})}{s_{\hat{y}(\mathbf{x})}}$, dann ist

$$\frac{dz}{dy} = \frac{1}{s_{\hat{y}(\mathbf{x})}} \Leftrightarrow dz = \frac{1}{s_{\hat{y}(\mathbf{x})}} dy.$$

Mit $q(\mathbf{x}) := \frac{y_{max} - \hat{y}(\mathbf{x})}{s_{\hat{y}(\mathbf{x})}}$ lässt sich der Ausdruck $y(\mathbf{x}) - y_{max}$ umformulieren zu

$$\begin{aligned}
y - y_{max} &= y - \hat{y}(\mathbf{x}) + \hat{y}(\mathbf{x}) - y_{max} \\
&= s_{\hat{y}(\mathbf{x})}z - (y_{max} - \hat{y}(\mathbf{x})) \\
&= s_{\hat{y}(\mathbf{x})}(z - q(\mathbf{x}))
\end{aligned}$$

Für den bedingten Erwartungswert gilt somit

$$\begin{aligned}
E[I(\mathbf{x})|\mathbf{y}^{(n)}] &= \int_{y_{max}}^{\infty} (y - y_{max})f_{Y(\mathbf{x})|\mathbf{y}^{(n)}}(y)dy \\
&= \int_{q(\mathbf{x})}^{\infty} s_{\hat{y}(\mathbf{x})}(z - q(\mathbf{x}))\phi(z)dz \\
&= s_{\hat{y}(\mathbf{x})} \underbrace{\int_{q(\mathbf{x})}^{\infty} z\phi(z)dz}_{\phi(q(\mathbf{x}))} - s_{\hat{y}(\mathbf{x})}q(\mathbf{x}) \int_{q(\mathbf{x})}^{\infty} \phi(z)dz \\
&= s_{\hat{y}(\mathbf{x})}\phi(q(\mathbf{x})) - s_{\hat{y}(\mathbf{x})}q(\mathbf{x})(1 - \Phi(q(\mathbf{x}))) \\
&= s_{\hat{y}(\mathbf{x})}\phi\left(\frac{y_{max} - \hat{y}(\mathbf{x})}{s_{\hat{y}(\mathbf{x})}}\right) + (y_{max} - \hat{y}(\mathbf{x}))\left(1 - \Phi\left(\frac{y_{max} - \hat{y}(\mathbf{x})}{s_{\hat{y}(\mathbf{x})}}\right)\right).
\end{aligned}$$

Die sequentielle Modellverfeinerung mit Hilfe des Expected-Improvement-Kriteriums wird so lange fortgesetzt, bis ein Abbruchkriterium erfüllt ist. Schonlau (1997) schlägt vor, eine feste Toleranzgrenze von z. B. 0.001 zu verwenden. Die sequentielle Modellverfeinerung wird gestoppt, falls das Maximum der Expected Improvement kleiner als diese Toleranzgrenze ist. Da die Expected Improvement aber vom Wertebereich der Zielgröße abhängt, ist dieses Vorgehen nicht zu empfehlen. Sinnvoller ist es, die Toleranzgrenze in Abhängigkeit von $\mathbf{y}^{(n)}$ zu wählen. Jones et al. (1998) stoppen die sequentielle Modellverfeinerung, sobald der maximale Wert der Expected Improvement weniger als 1 % des aktuell beobachteten Maximums y_{max} ist. Dieses Vorgehen besitzt jedoch drei Schwachstellen. Zum einen sollte als Schwellenwert 1 Prozent des Betrages des aktuell beobachteten Maximums verwendet werden, um auch Zielgrößen mit negativen Werten berücksichtigen zu können. Zum anderen ergibt sich bei diesem

Stoppkriterium das Problem, wie zu verfahren ist, wenn y_{max} gleich Null ist. In diesem Fall ist dieses Verfahren nicht anwendbar. Das schwerwiegendste Problem ist jedoch, dass dieses Vorgehen nicht den Range der beobachteten Zielgröße berücksichtigt. Um die sich daraus ergebende Problematik zu verdeutlichen, betrachte Abbildung 6.2 a), in der die Anpassung eines Gauß-Prozess-Modells an vier Datenpunkte dargestellt ist. Nach dem Vorschlag von Jones et al. (1998) würde in dieser Situation die sequentielle Modellverfeinerung gestoppt, falls das Maximum der Expected Improvement kleiner als 0.07 ist. Bei einer additiven Vergrößerung dieser Zielgrößen um eine Konstante von $c=100$ vergrößert sich im Gauß-Prozess-Modell nur der Achsenabschnitt der deterministischen Trendkomponente um diese Konstante c . Die Expected Improvement in Abhängigkeit von x ändert sich durch diese Transformation aber nicht. Für diese Daten würde die sequentielle Modellverfeinerung somit schon gestoppt, falls das Maximum der Expected Improvement kleiner als 1.07 ist. Aus diesen Gründen wird ein alternatives Abbruchkriterium vorgeschlagen, das die sequentielle Modellverfeinerung stoppt, sobald das Maximum der Expected Improvement kleiner als 1 Prozent von $|y_{min} - y_{max}|$, der aktuell beobachteten maximalen Differenz der Zielwerte, ist. Da sich die Varianz der Expected Improvement im Versuchsraum ändert, sollte dies bei der Wahl des Abbruchkriteriums berücksichtigt werden. Dazu könnte eine obere Grenze definiert werden, bestehend aus der Expected Improvement plus deren Varianz. Die Varianz lässt sich im Übrigen leicht numerisch berechnen. Mit der so modifizierten Version des obigen Abbruchkriterium würde die sequentielle Modellverfeinerung gestoppt, falls die obere Grenze an der Stelle mit maximaler Expected Improvement kleiner als 1 Prozent von $|y_{min} - y_{max}|$ ist. Alternativ zu der Toleranzgrenze, die über $|y_{min} - y_{max}|$ definiert ist, kann vom Anwender auch die untere Grenze spezifiziert werden, ab der eine erwartete Verbesserung der Zielgröße nicht mehr relevant ist. Da die Werte der Expected Improvement über die Iterationsschritte nicht monoton fallen, kann es sinnvoll sein, erst abzubrechen, falls mehrere aufeinanderfolgende Werte der Expected Improvement die Toleranzgrenze unterschreiten. Zudem können zusätzliche Anforderungen gestellt werden, wie z. B. dass bei der Wahl eines neuen Versuchspunkts mit Hilfe der Expected Improvement nur Punkte des Versuchsraums in Frage kommen, die einen bestimmten Mindestabstand zu den bestehenden Versuchspunkten aufwei-

sen. Um den Versuchsaufwand zu beschränken, ist eine obere Schranke für den Gesamtversuchsumfang festzulegen. Kann das Abbruchkriterium bis dahin nicht erreicht werden, wird die sequentielle Modellverfeinerung gestoppt. Bei der Wahl der Toleranzgrenze ist zudem zu berücksichtigen, dass der verwendete Standardfehler der Vorhersage des Gauß-Prozess-Modells nicht berücksichtigt, dass die Parameter der Korrelationsfunktion geschätzt werden. Als Folge wird damit der Standardfehler der Vorhersage und damit ebenfalls die Expected Improvement unterschätzt.

Eine von Schonlau et al. (1998) vorgeschlagene Erweiterung des Expected-Improvement-Kriteriums erlaubt es, die Gewichtung zwischen lokaler und globaler Suche zu steuern. Diese erweiterte Expected Improvement unterscheidet sich von der einfachen Expected Improvement durch einen Exponenten g .

$$E [I^g(\mathbf{x}) | \mathbf{y}^{(n)}] = \int_{y_{max}}^{\infty} (z - y_{max})^g f_{Y(\mathbf{x})|\mathbf{y}^{(n)}}(z) dz. \quad (6.3)$$

Für $g = 1$ ergibt sich das einfache Expected-Improvement-Kriterium. Je größer der Exponent g gewählt wird, desto stärker werden große Werte der Improvement gewichtet. Da große Werte der Improvement bei Unsicherheitsverteilungen von $y(\mathbf{x})$ mit breiten Tails mit größerer Wahrscheinlichkeit auftreten, wird für größere Werte von g die Suche in Bereichen mit hoher Unsicherheit stärker gewichtet. Eine geschlossene Form der verallgemeinerten Expected Improvement wird in Schonlau et al. (1998) für Minimierungsprobleme gegeben. Schonlau et al. (1998) schlagen vor, für die verallgemeinerte Expected Improvement eine Abbruchregel zu verwenden, die auf $(E [I^g(\mathbf{x}) | \mathbf{y}^{(n)}])^{1/g}$ statt $E [I(\mathbf{x}) | \mathbf{y}^{(n)}]$ beruht. Für $g \geq 1$ gilt bei gleicher Toleranzgrenze, dass Abbruchregeln, die auf $(E [I^g(\mathbf{x}) | \mathbf{y}^{(n)}])^{1/g}$ statt $E [I(\mathbf{x}) | \mathbf{y}^{(n)}]$ beruhen, konservativer sind und somit mehr Punkte sequentiell hinzufügen. Dies lässt sich mit Hilfe der Jensenschen Ungleichung zeigen. Da I nichtnegativ und I^g eine konvexe Funktion von I ist, gilt $(E [I^g(\mathbf{x}) | \mathbf{y}^{(n)}])^{1/g} \geq E [I(\mathbf{x}) | \mathbf{y}^{(n)}]$.

Aufgrund der Effizienz hat sich die Prozessoptimierung mit Hilfe der Expected Improvement etabliert. Dies zeigt sich auch an der hohen Zahl von Autoren, die die Veröffentlichung von Jones et al. (1998) zitieren (77 Zitate im ISI Web of Science, Stand 1.11.2006). Dieser Ansatz zur effizienten Prozessoptimierung kann

ohne Probleme auch auf nicht deterministische Prozesse angewandt werden. Zu beachten ist in diesem Fall lediglich, dass aufgrund der Fehlervarianz in den Daten die Expected Improvement sich in schon beobachteten Versuchspunkten nicht auf 0 reduziert.

6.2 Univariates Simulationsbeispiel

Um das Vorgehen der sequentiellen Modellverfeinerung mit dem Expected-Improvement-Kriterium zu verdeutlichen, wird das Beispiel aus Abbildung 6.1 a) wieder aufgegriffen. In Abbildung 6.2 a) ist neben der Modellvorhersage der wahre funktionale Zusammenhang abgebildet. Offensichtlich ist die Modellanpassung im Großteil des Versuchsraums schlecht. Das globale Maximum befindet sich an der Stelle $x = 8.35$. Um bei der sequentiellen Modellverfeinerung auch Bereiche mit höherer Unsicherheit berücksichtigen zu können, wurde der Parameter g auf 3 gesetzt. Die Expected Improvement lässt sich nun wie oben beschrieben für jedes x berechnen, vgl. Abbildung 6.2 b).

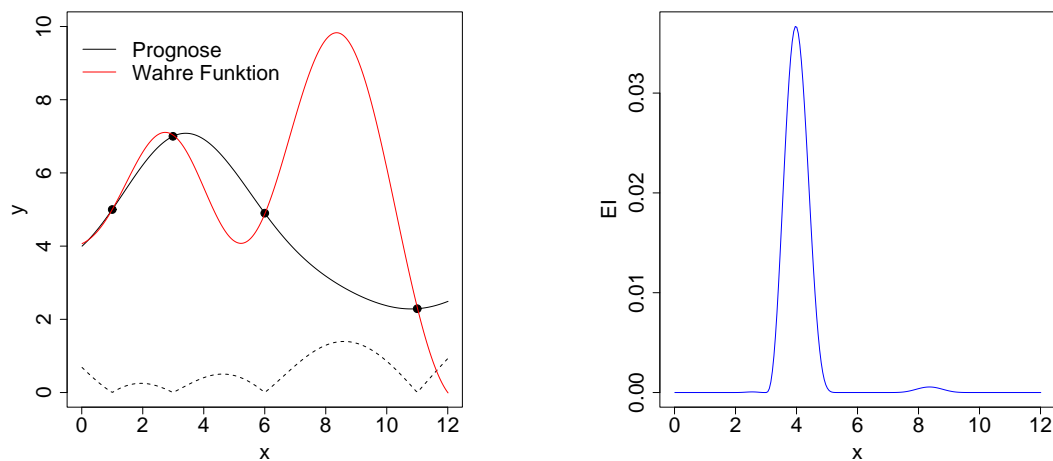


Abbildung 6.2: a) Initial Gauß-Prozess-Modell, b) Expected Improvement

Das Maximum der Expected Improvement wird an dem vorhergesagten Optimum des Initial-Modells bei $x = 3.96$ beobachtet. Ein erster Versuchspunkt wird somit an dieser Stelle hinzugefügt. Nachdem dort die Zielgröße erhoben wurde, kann das Gauß-Prozess-Modell aktualisiert werden. Die neue Modellvorhersage

ist in Abbildung 6.4 oben links geplottet. Die erneute Berechnung der Expected Improvement ergibt zwei Peaks. Das Maximum der Expected Improvement liegt dabei mit $x = 8.1$ im Bereich mit großer Unsicherheit, vgl. Abbildung 6.4 oben rechts. Wird die Zielgröße in diesem Versuchspunkt berücksichtigt, erlaubt das Gauß-Prozess-Modell schon eine sehr gute Approximation der wahren Zusammenhänge, vgl. Abbildung 6.4 Mitte links. Da im Versuchspunkt $x = 8.1$ das aktuell beobachtete Maximum mit 9.73 deutlich vergrößert werden konnte, fokussiert sich die sequentielle Modellverfeinerung nun auf den Bereich um diesen Versuchspunkt. Dabei wird ein neuer Versuchspunkt bei $x = 8.91$ platziert, vgl. Abbildung 6.4 Mitte rechts. Das aktualisierte Modell liefert nun eine sehr genaue Anpassung an die wahre Funktion, vgl. Abbildung 6.4 unten links. Die Expected Improvement wird nun maximal bei $x = 8.35$, vgl. Abbildung 6.4 unten rechts. Wird dieser Versuchspunkt bei der Modellanpassung berücksichtigt und die Expected Improvement erneut berechnet, so ist diese im gesamten Versuchsbereich gleich 0. Dies führt zum Abbrechen der sequentiellen Modellverfeinerung. Mit vier sequentiellen Versuchspunkten konnte das globale Optimum somit identifiziert werden. Mit steigender Komplexität der Zusammenhänge steigt die Effizienz dieser Methode im Vergleich zum Ansatz, der die gleiche Stichprobendichte verwendet.

Der Parameter g , der die Gewichtung zwischen lokaler und globaler Suche steuert, ist bei der Optimierung vom Anwender zu spezifizieren. Um den Einfluss dieses Parameters zu untersuchen, wurde das obige Optimierungsproblem mit $g \in \{1, 2, \dots, 11\}$ wiederholt. In allen elf Fällen konnte das globale Optimum identifiziert werden, wobei zwischen vier und fünf sequentielle Versuchspunkte benötigt wurden. Für $g < 10$ wurde jeweils im ersten neuen Versuchspunkt das lokale Optimum untersucht. Erst für $g = 11$, also einer sehr starken Gewichtung der globalen Suche, wurde der erste sequentielle Versuchspunkt in den Bereich mit großer Unsicherheit gelegt. Der Bereich des Versuchsraum, der bei diesen elf Fällen untersucht wurde ist in Abbildung 6.3 dargestellt. Die Suche nach dem globalen Optimum beschränkt sich somit nur auf zwei Bereiche. In dem restlichen Versuchsraum wurde in keinem der elf Fälle ein Versuchspunkt platziert.

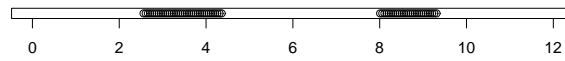


Abbildung 6.3: Untersucher Versuchsbereich aller elf Optimierungsdurchläufe

Diese Untersuchung lässt vermuten, dass die Identifizierung des globalen Optimums nicht von der Wahl des Parameters g abhängt. Es verändert sich lediglich der benötigte Stichprobenumfang. Die Variation scheint aber gering zu sein, auch wenn hier nur ein eindimensionales Optimierungsproblem betrachtet wurde.

6.3 Wünschbarkeitsfunktionen

Da der univariate Ansatz von Jones zur Prozessoptimierung im Sinne von Exaktheit und Effizienz sehr gute Ergebnisse liefert, soll er auf den multivariaten Fall verallgemeinert werden. In einem ersten Schritt ist dazu aber zunächst ein geeignetes Maß zu definieren, mit dem die Gesamtqualität mehrerer Zielgrößen verglichen werden kann. Dabei ist zu berücksichtigen, dass die Zielgrößen in der Regel unterschiedlich skaliert sind und unterschiedliche Qualitätsanforderungen besitzen. In Harrington (1965) wurde ein Ansatz entwickelt, der es ermöglicht die Gesamtqualität verschiedener Produkte trotz dieser Probleme zu vergleichen. Dabei werden Wünschbarkeitsfunktionen verwendet, welche als Indikator für den Grad der Übereinstimmung mit den Prozessspezifikationen der einzelnen Zielgrößen dienen. Jede Zielgröße Y_j mit $j = 1, \dots, I$ wird dazu in das skalenfreie Intervall $[0, 1]$ transformiert. Je größer die Wünschbarkeit, desto besser ist die Qualität der Zielgröße.

6.3.1 Wünschbarkeitsfunktionen von Harrington

Harrington (1965) schlägt zwei Arten von Wünschbarkeitsfunktionen vor, um eine Transformation der Zielgrößen in das skalenfreie Intervall $[0, 1]$ zu erreichen. Ist eine Zielgröße Y zu minimieren bzw. zu maximieren, werden einseitige Wünschbarkeitsfunktionen eingesetzt.

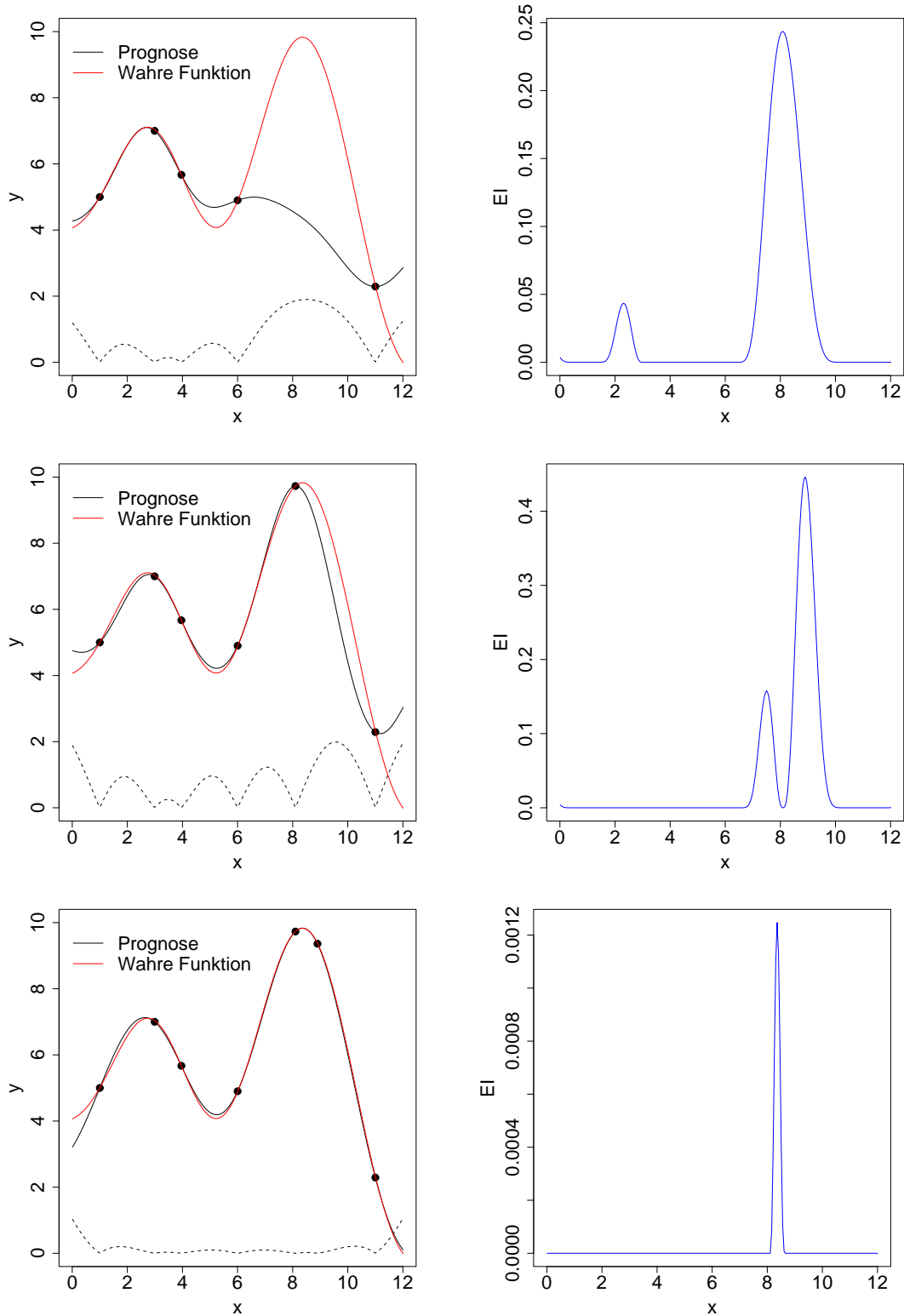


Abbildung 6.4: Gauß-Prozess-Modell (2. Iteration: oben links, 3. Iteration: Mitte links 4. Iteration: unten links), Expected Improvement (2. Iteration: oben rechts, 3. Iteration: Mitte rechts 4. Iteration: unten rechts)

Harrington definiert die einseitige Wünschbarkeitsfunktion über eine spezielle Form der Gompertz-Kurve

$$w(Y) = \exp(-\exp(-(b_0 + b_1 Y))), \text{ mit } b_0, b_1 \in \mathbb{R}. \quad (6.4)$$

Die Form der Wünschbarkeitsfunktion wird eindeutig festgelegt, indem an zwei Stellen von y die Wünschbarkeiten bestimmt werden. Die zugehörigen Parameter b_0 und b_1 lassen sich somit mit Hilfe eines linearen Gleichungssystems bestehend aus zwei Gleichungen aus Formel (6.4) berechnen. In Abbildung 6.5 a) sind zwei unterschiedliche Wünschbarkeitsfunktionen für Zielgrößen, die zu maximieren sind, angegeben.

Für Zielgrößen, bei denen Abweichungen vom Zielwert in beide Richtungen von Interesse sind, schlägt Harrington eine zweiseitige Wünschbarkeitsfunktion vor. Für die Charakterisierung solch einer Wünschbarkeitsfunktion ist zunächst die Bestimmung der unteren und oberen Spezifikationsgrenze USG und OSG für die Zielgröße Y nötig. Die Spezifikationsgrenzen sind dabei symmetrisch um den Zielwert zu legen. Die zweiseitige Wünschbarkeitsfunktion nach Harrington ist nun definiert durch

$$w(Y) = \exp\left(-\left|\frac{Y - \frac{OSG+USG}{2}}{\frac{OSG-USG}{2}}\right|^\nu\right) \text{ für } 0 < \nu < \infty. \quad (6.5)$$

Der Parameter ν bestimmt die Kurtosis der Funktion, vgl. Abbildung 6.5b). Je kleiner dieser Parameter gewählt wird, desto stärker werden kleine Abweichungen vom Sollwert bestraft. Die Wünschbarkeit an den Spezifikationsgrenzen ist $1/e$.

6.3.2 Wünschbarkeitsfunktionen von Derringer und Suich

In Derringer und Suich (1980) wurde ein alternativer Ansatz zur Definition von Wünschbarkeitsfunktionen vorgestellt. Diese lassen sich durch stückweise definierte Funktionen beschreiben, wodurch sich eine größere Flexibilität in der Form der Wünschbarkeitsfunktion ergibt. Für Zielgrößen, bei denen Werte kleiner gleich dem Sollwert T zu einem Qualitätsverlust führen, ist die einseitige Wünschbar-

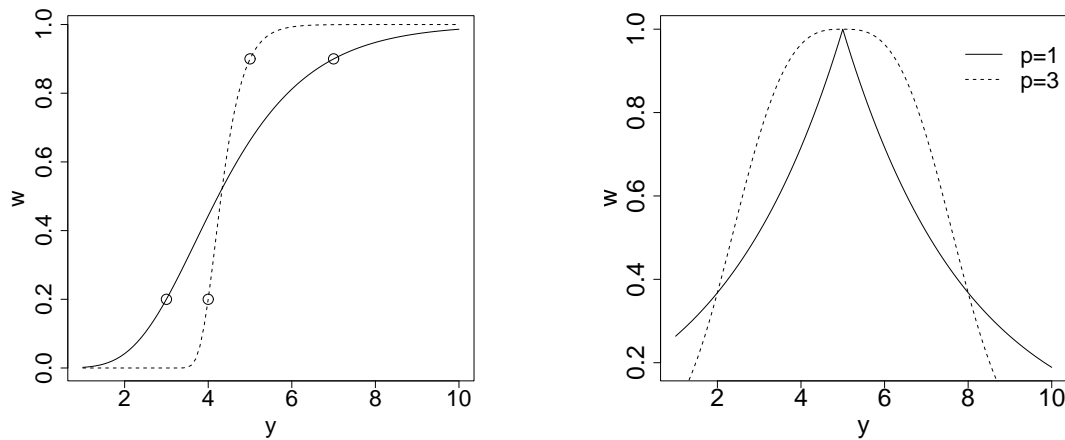


Abbildung 6.5: a) Einseitige Wünschbarkeitsfunktionen nach Harrington, b) Zweiseitige Wünschbarkeitsfunktionen nach Harrington

keitsfunktion von Derringer und Suich gegeben durch

$$w(Y) = \begin{cases} 0 & , \text{für } Y < USG \\ \left(\frac{Y-USG}{T-USG}\right)^l & , \text{für } USG \leq Y \leq T \\ 1 & , \text{für } T < Y \end{cases} \quad (6.6)$$

wobei USG die untere Spezifikationsgrenze ist und $l \in \mathbb{R}^+$. Unterhalb der Spezifikationsgrenze sind Abweichungen vom Sollwert nicht mehr tolerabel und führen damit zu einer Wünschbarkeit von Null. Mit Hilfe des Parameters l lässt sich einstellen, wie stark Abweichungen vom Sollwert bestraft werden. Je größer der Parameter l gewählt wird, desto schwerwiegender sind kleine Abweichungen vom Sollwert T zu beurteilen. In Abbildung 6.6 a) sind typische Verläufe der einseitigen Wünschbarkeitsfunktion dargestellt. Die Definition der einseitigen Wünschbarkeitsfunktion für Zielgrößen, bei denen Werte größer gleich dem Sollwert zu einem Qualitätsverlust führen, erfolgt entsprechend.

Die zweiseitige Wünschbarkeitsfunktion für Zielgrößen, bei denen Abweichungen vom Sollwert T in beide Richtungen zu einem Qualitätsverlust führen, ist definiert

als

$$w(Y) = \begin{cases} 0 & , \text{für } Y \leq USG \\ \left(\frac{Y-USG}{T-USG}\right)^l & , \text{für } USG < Y \leq T \\ \left(\frac{OSG-Y}{OSG-T}\right)^r & , \text{für } T < Y \leq OSG \\ 0 & , \text{für } OSG < Y \end{cases} \quad (6.7)$$

wobei USG, OSG die untere und obere Spezifikationsgrenze angeben und $l, r \in \mathbb{R}^+$. Ähnlich wie im einseitigen Fall bestimmen die Parameter l und r , wie stark Abweichungen vom Sollwert nach oben oder unten bestraft werden. Im Gegensatz zu Harringtons Ansatz können auch asymmetrische Wünschbarkeitsfunktionen um den Sollwert betrachtet werden. In Abbildung 6.6 b) sind typische Verläufe der zweiseitigen symmetrischen Wünschbarkeitsfunktion dargestellt.

Im Gegensatz zu Harringtons Ansatz besitzen die Wünschbarkeitsfunktionen von Derringer und Suich die intuitiv ansprechende Eigenschaft den Wert 0 anzunehmen, wenn die Zielgröße außerhalb der Spezifikationsgrenzen liegt. Eine Wünschbarkeit von 1 wird erreicht, wenn die Zielgröße dem Sollwert entspricht. Die Wünschbarkeit von 1 kann bei Harringtons Wünschbarkeitsfunktionen per Definition nicht erreicht werden. Die Wünschbarkeitsfunktionen von Derringer und Suich besitzen aber neben den zahlreichen Fallunterscheidungen, die zu berücksichtigen sind, einen weiteren Nachteil. Dieser ist die Nichtdifferenzierbarkeit der Wünschbarkeitsfunktion in den Knickstellen, was sich als problematisch beim Einsatz gradientenbasierter Optimierungsverfahren erweist.

6.3.3 Wünschbarkeitsindex

Um die Gesamtqualität gegenläufiger Zielgrößen beurteilen zu können, schlägt Harrington vor, den *Wünschbarkeitsindex* WI zu verwenden. Dieser ist definiert als das geometrische Mittel der Einzelwünschbarkeiten $WI = \prod_{j=1}^I w_j^{1/I}$. Auf diese Weise wird das multivariate Problem in ein univariates Problem transformiert. Ein Wünschbarkeitsindex nahe 1 spricht für eine sehr gute Gesamtqualität des Produkts. Ist eine zusätzliche Gewichtung der Zielgrößen zu berücksichtigen, kann statt des einfachen geometrischen Mittels eine gewichtete Version $WI = \prod_{j=1}^I w_j^{g_j}$ verwendet werden, wobei $\sum_{j=1}^I g_j = 1$.

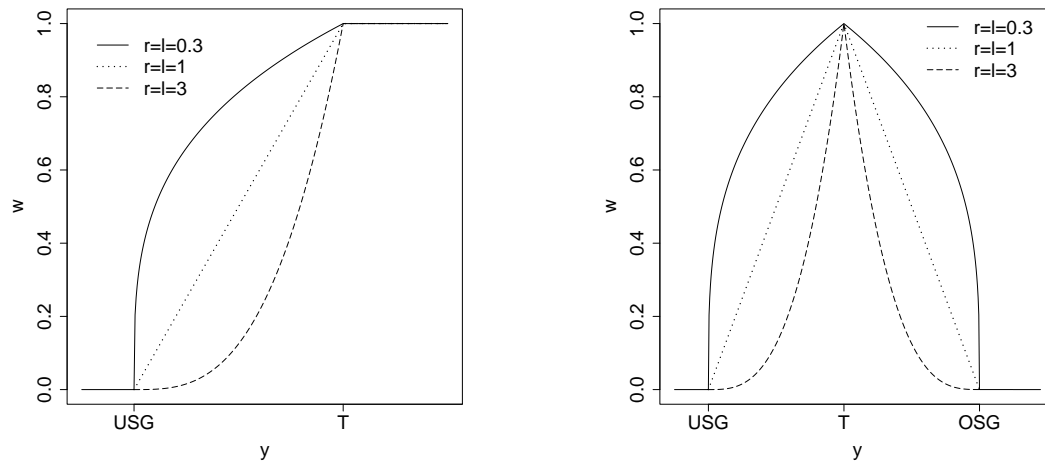


Abbildung 6.6: a) Einseitige Wünschbarkeitsfunktionen nach Derringer und Suich, b) Zweiseitige Wünschbarkeitsfunktionen nach Derringer und Suich

Mit Hilfe der Wünschbarkeitsfunktionen wird der I -dimensionale Wertebereich der Zielgrößen in den I -dimensionalen Einheitswürfel $[0, 1]^I$ transformiert. Für jeden Punkt im transformierten Wertebereich kann der Wünschbarkeitsindex WI berechnet werden. Im zweidimensionalen Fall lässt sich die Veränderung von WI in Abhängigkeit der Einzelwünschbarkeiten plotten. In Abbildung 6.7 sind Konturplots der Veränderung für unterschiedliche Gewichte von WI dargestellt. Dabei besitzen z. B. Punkte im Wünschbarkeitsraum, die auf der untersten Konturkurve liegen, einen Wünschbarkeitsindex von 0.1. Diese graphische Darstellungsform wird im folgenden Kapitel für die Definition der multivariaten Expected Improvement benötigt.

6.3.4 Verteilungsfunktionen der Wünschbarkeiten und des Wünschbarkeitsindexes

In der Praxis werden Wünschbarkeitsfunktionen zur multivariaten Optimierung von Prozessen häufig in Kombination mit statistischer Versuchsplanung eingesetzt. Basierend auf den beobachteten Versuchspunkten wird für jede Zielgröße ein lineares Regressionsmodell angepasst. Mit Hilfe dieser Modelle kann für jeden Punkt im Versuchsraum der Wünschbarkeitsindex geschätzt werden.

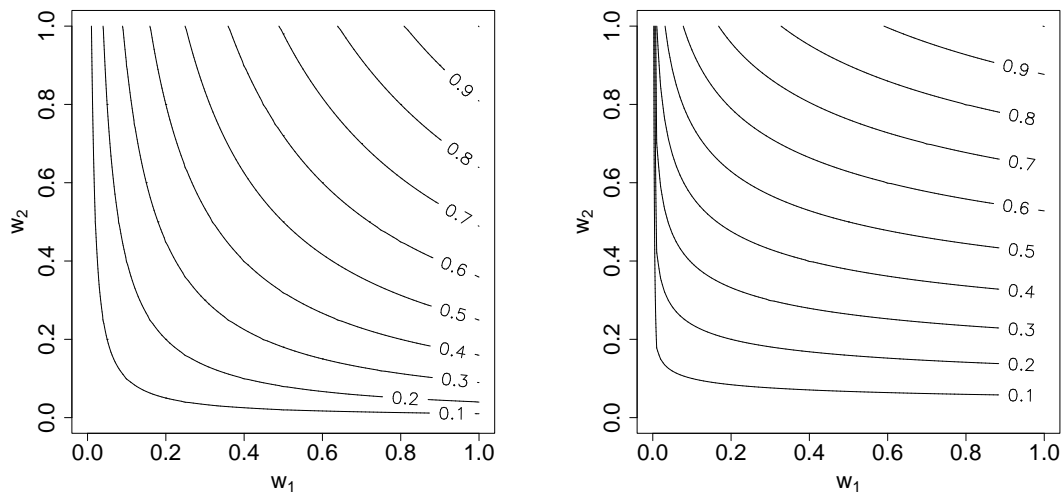


Abbildung 6.7: Konturplot des Wünschbarkeitsindexes mit Gewichtung des geometrischen Mittels von a) $g_1 = 0.5, g_2 = 0.5$ und b) $g_1 = 0.3, g_2 = 0.7$.

Die Optimierung des Prozesses erfolgt dann mittels eines Gradientenverfahrens über den Wünschbarkeitsindex. Dieses Vorgehen besitzt jedoch eine wesentliche methodische Schwachstelle. Bei der Transformation der modellierten Zielgrößen in Wünschbarkeiten wird der Fehlerterm der linearen Regressionsmodelle vernachlässigt. In Steuer (2005) werden die Folgen dieses Vorgehens anhand einer Simulationsstudie verdeutlicht. Das identifizierte Optimum lag teilweise deutlich vom wahren Optimum entfernt. In 5% der Fälle besaß das geschätzte Optimum sogar eine wahre Wünschbarkeit von 0. Grund dafür ist, dass für einen Punkt \mathbf{x} im Versuchsraum die Prognose der Wünschbarkeit mit Hilfe der linearen Regressionsmodelle nicht gleich dem Erwartungswert der Verteilung des Wünschbarkeitsindexes ist. Diese Ergebnisse zeigen, dass es sinnvoller ist, die Werte der Wünschbarkeitsfunktionen und des Wünschbarkeitsindexes als Zufallsvariablen zu betrachten und die Optimierung des Prozesses über z. B. den geschätzten Erwartungswert dieser Zufallsvariablen durchzuführen.

Unter der Annahme, dass $Y \sim N(\mu, \sigma^2)$ lässt sich durch wiederholtes Anwenden des Transformationssatzes die Verteilung der ein- und zweiseitigen Wünschbarkeitsfunktion von Harrington herleiten, vgl. Trautmann und Weihs (2006).

Satz 6.3.1 Für eine Zielgröße $Y \sim N(\mu, \sigma^2)$ besitzt die einseitige Wünschbarkeitsfunktion nach Harrington eine Doppel-Lognormal-Verteilung mit der Dichte

$$f(w) = \frac{-1}{\sqrt{2\pi}\tilde{\sigma} \log(w)w} \exp\left[-\frac{1}{2\tilde{\sigma}^2} (\log(-\log(w)) - \tilde{\mu})^2\right],$$

wobei $\tilde{\mu} = -(b_0 + b_1\mu)$ und $\tilde{\sigma}^2 = b_1^2\sigma^2$.

In Abbildung 6.8 ist der unterschiedliche Einfluss der Parameter $\tilde{\mu}$ und $\tilde{\sigma}$ dargestellt. Offensichtlich ist die Form der Dichtefunktion sehr variabel und kann sich durch ihre mögliche Asymmetrie deutlich von der Dichte der Normalverteilung unterscheiden. Eine ausführliche Untersuchung des Einflusses der Parameter $\tilde{\mu}$ und $\tilde{\sigma}$ auf die Form der Dichtefunktion und deren Eigenschaften ist in Holland und Ahsanullah (1989) gegeben.

Satz 6.3.2 Für eine Zielgröße $Y \sim N(\mu, \sigma^2)$ ist die Dichte der zweiseitigen Wünschbarkeitsfunktion nach Harrington gegeben durch

$$f(w) = c \left[\exp\left(\frac{-\left((-\log(w))^{\frac{1}{\nu}} - \tilde{\mu}\right)^2}{2\tilde{\sigma}^2}\right) + \exp\left(\frac{-\left((-\log(w))^{\frac{1}{\nu}} + \tilde{\mu}\right)^2}{2\tilde{\sigma}^2}\right) \right],$$

$$\text{mit } c = \frac{(-\log(w))^{\frac{1}{\nu}-1}}{\sqrt{2\pi}\tilde{\sigma}w\nu}, \quad \tilde{\mu} = \frac{2}{OSG-USG}\mu + \frac{OSG+USG}{OSG-USG}, \quad \text{und}$$

$$\tilde{\sigma}^2 = \left(\frac{2}{OSG-USG}\right)^2 \sigma^2.$$

Typische Dichten der zweiseitigen Wünschbarkeitsfunktion nach Harrington sind in Abbildung A.1 dargestellt. Ähnlich wie im einseitigen Fall ist die Form der Dichtefunktion sehr vielfältig. Auch hier ist in der Regel ein asymmetrisches Verhalten zu beobachten.

Für den Spezialfall mit $r, l = 1$ wurden in Steuer (2005) die Dichtefunktionen der ein- und zweiseitigen Wünschbarkeitsfunktionen von Derringer und Suich hergeleitet. Für den allgemeinen Fall mit $r, l \in \mathbb{R}^+$ erweist sich die Bestimmung der Dichtefunktion als deutlich schwieriger, da sich diese nicht mehr über die Normalverteilung herleiten lässt.

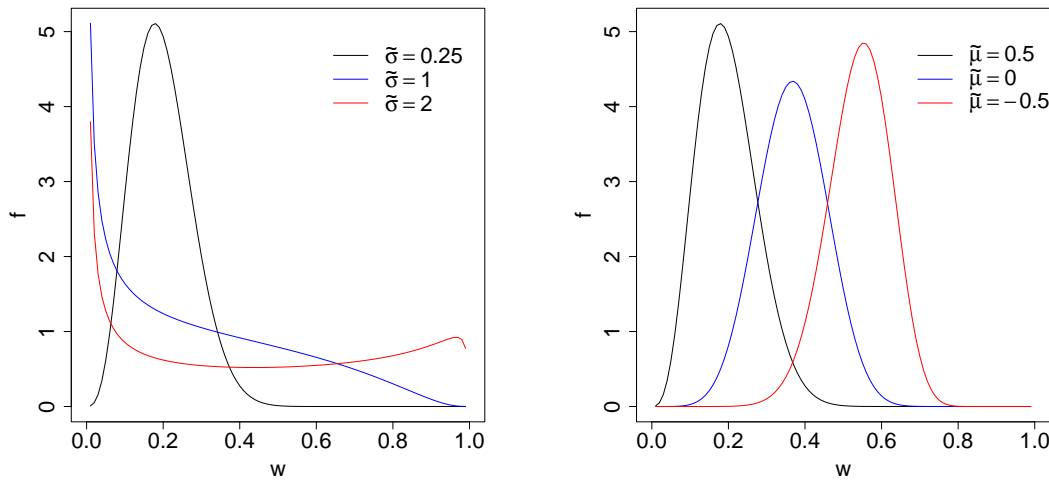


Abbildung 6.8: Graphen der Dichten einseitiger Wünschbarkeitsfunktionen nach Harrington mit a) $\tilde{\mu} = 0.5$ und b) $\tilde{\sigma} = 0.25$

Satz 6.3.3 Für eine Smaller-the-better-Zielgröße $Y \sim N(\mu, \sigma^2)$ besitzt die einseitige Wünschbarkeitsfunktion von Derringer und Suich mit $r, l = 1$ die folgende Dichtefunktion

$$f(w) = \begin{cases} \Phi\left(\frac{USG - \mu}{\sigma}\right) & , \text{für } w = 0 \\ \frac{T - USG}{\sigma} \phi\left(\frac{USG + w(T - USG) - \mu}{\sigma}\right) & , \text{für } w \in (0, 1) \\ 1 - \Phi\left(\frac{T - \mu}{\sigma}\right) & , \text{für } w = 1 \end{cases} ,$$

wobei Φ und ϕ die Verteilungsfunktion und die Dichtefunktion der Standardnormalverteilung definieren.

Diese Dichtefunktion lässt sich als Mischdichte von Normalverteilungsdichten mit einem Punktgewicht auf der 0 und 1 darstellen. In Abbildung 6.9 a) sind typische Verläufe der Dichtefunktion angegeben. Die Dichten zeigen eine ausgeprägte Rechts- oder Linksschiefe. Wie für Harringtons Wünschbarkeitsfunktionen weichen die Dichtefunktionen auch hier deutlich von der Gestalt einer symmetrischen Normalverteilungsdichte ab.

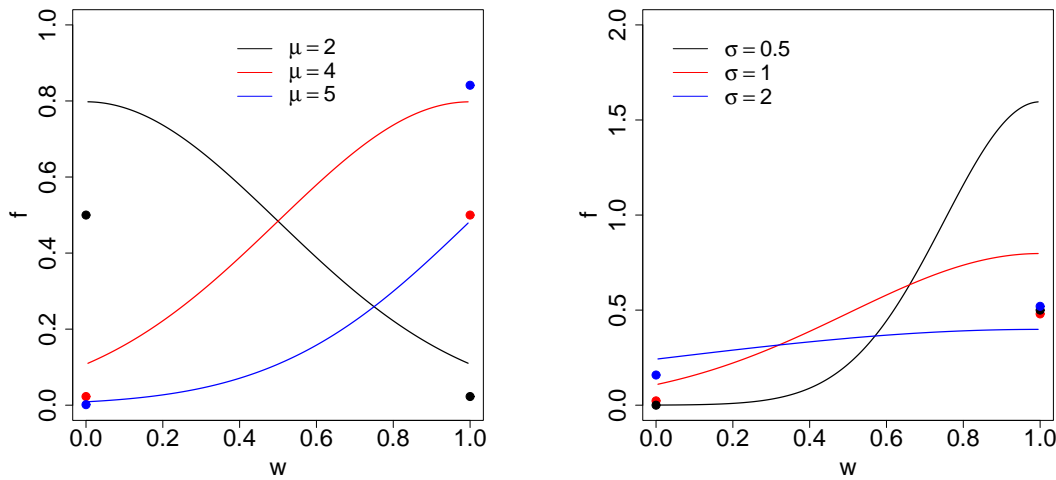


Abbildung 6.9: Graphen der Dichten einseitiger Wünschbarkeitsfunktionen nach Derringer und Suich mit $USG = 2, T = 4$ und a) $\sigma = 1$ b) $\mu = 4$

Satz 6.3.4 Für eine Zielgröße $Y \sim N(\mu, \sigma^2)$ besitzt die zweiseitige Wünschbarkeitsfunktion von Derringer und Suich mit $r, l = 1$ die folgende Dichtefunktion

$$f(w) = \begin{cases} \Phi\left(\frac{USG-\mu}{\sigma}\right) + 1 - \Phi\left(\frac{OSG-\mu}{\sigma}\right) & , \text{für } w = 0 \\ \frac{T-USG}{\sigma}\phi(q_1) + \frac{OSG-T}{\sigma}\phi(q_2) & , \text{für } w \in (0, 1) \\ 0 & , \text{für } w = 1 \end{cases} ,$$

$$\text{mit } q_1 = \frac{USG+w(T-USG)-\mu}{\sigma}$$

$$\text{und } q_2 = \frac{OSG-w(OSG-T)-\mu}{\sigma}.$$

Der Einfluss der Dichteparameter ist in Abbildung A.2 verdeutlicht. Auch hier ergibt sich die Dichtefunktion als Mischdichte von Normalverteilungsdichten mit einem Punktgewicht auf der Null. Die Form der Dichte ist sehr flexibel. Dies reicht von extrem rechts- bzw. linksschiefen Verteilungen bis hin zu symmetrischen Verteilungen.

Da der Wünschbarkeitsindex durch das geometrische Mittel definiert wird, ist die Bestimmung der exakten Verteilung des Wünschbarkeitsindexes sehr schwierig. Aus diesem Grund wurden in Steuer (2005) Simulationsstudien durchgeführt, um Aussagen über die Verteilung des Wünschbarkeitsindexes zu erlangen. In

Trautmann und Weihs (2006) konnte für den bivariaten Spezialfall zweiseitiger Wünschbarkeitsfunktionen nach Harrington mit dem Exponent $\nu = 1$ eine exakte Verteilung des Wünschbarkeitsindex hergeleitet werden. Bei Beschränkung auf einseitige Wünschbarkeitsfunktionen nach Harrington konnte dort zudem eine approximative Verteilung hergeleitet werden. Die Form der Dichten des Wünschbarkeitsindex sind ähnlich wie die Dichten der Wünschbarkeitsfunktionen sehr flexibel und teilweise stark asymmetrisch.

6.4 Das multivariate Expected-Improvement-Kriterium

Schonlau et al. (1998) präsentieren eine Modifikation des in Kapitel 6.1 vorgestellten Verfahrens, das eine Optimierung einer Zielgröße ermöglicht wobei weitere Zielgrößen in vorgegebenen Intervallen liegen müssen. Dieser Ansatz erlaubt aber nicht die gleichzeitige Optimierung mehrerer Zielgrößen. Aus diesem Grund wird innerhalb dieses Kapitels der auf der Expected Improvement beruhende Ansatz auf den multivariaten Fall verallgemeinert.

Eine Möglichkeit, das sequentielle Verfahren auch bei Berücksichtigung mehrerer Zielgrößen anwenden zu können, wäre für jede Zielgröße ein Gauß-Prozess-Modell anzupassen und für jede dieser Zielgrößen die univariate Expected Improvement zu berechnen. Da die Zielgrößen häufig unterschiedlich skaliert sind und unterschiedliche Spezifikationsanforderungen besitzen, fehlt jedoch die Vergleichbarkeit der einzelnen Expected Improvements. Um dieses Problem zu lösen, bietet sich der Einsatz der Wünschbarkeitsfunktionen an. Jedoch sollte keine direkte Modellierung der Wünschbarkeiten erfolgen. Zum einen erhöht die Transformation der Zielgrößen in Wünschbarkeiten oft die Komplexität der Zusammenhänge. Zum anderen gehen insbesondere beim Einsatz zweiseitiger Wünschbarkeitsfunktionen wertvolle Informationen verloren. Da nicht mehr erkennbar ist, in welche Richtung eine Abweichung vom Sollwert beobachtet wurde, können die Prognosen der Modelle sehr schlecht sein. Um dies zu verdeutlichen, betrachte den Fall, bei dem im Punkt $x = a$ eine Abweichung vom Sollwert von -1 und im Punkt $x = b$ eine Abweichung vom Sollwert von $+1$ beobachtet wurde. Bei linearem

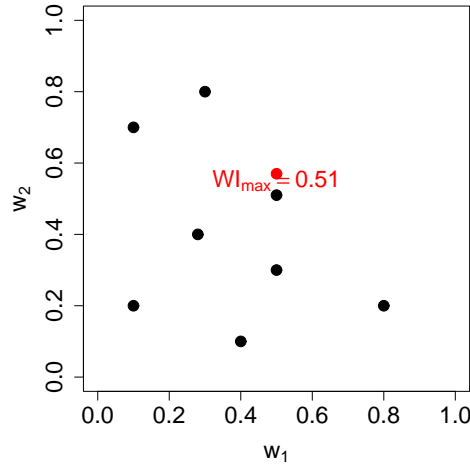
Zusammenhang ergäbe sich somit im Punkt $x = (a + b)/2$ der Sollwert. Bei symmetrischer zweiseitiger Wünschbarkeitsfunktion würde jedoch an den beiden Punkten $x = a$ und $x = b$ der gleiche Wert beobachtet. Ein lineares Modell der Wünschbarkeiten würde im Punkt $x = (a + b)/2$ fälschlicherweise die gleiche Wünschbarkeit wie an den beiden Punkten $x = a$ und $x = b$ vorhersagen. Diese Problematik tritt in noch stärker ausgeprägter Form bei der direkten Modellierung des Wünschbarkeitsindex auf. Aus diesem Grund ist es vorzuziehen, statt direkter Modellierung der Wünschbarkeiten oder des Wünschbarkeitsindex die Zielgrößen selbst zu modellieren. Für jeden Punkt \mathbf{x} im Versuchsraum lassen sich dann anhand der Prognosen der Modelle die Einzelwünschbarkeiten und damit auch der Wünschbarkeitsindex schätzen.

Im Folgenden werden zwei Ansätze zur Verallgemeinerung der in Kapitel 6.1 beschriebenen sequentiellen Modellverfeinerung vorgestellt. Basierend auf n Versuchspunkten sei dazu für jede der Zielgrößen Y_j mit $j = 1, \dots, I$ ein separates Gauß-Prozess-Modell angepasst worden. Um die Unsicherheit der Prognosen der I Zielgrößen an einer Stelle \mathbf{x} des Versuchsraums zu berücksichtigen, sei äquivalent zum univariaten Fall $\mathbf{Y}(\mathbf{x}) | \mathbf{y}_1^{(n)}, \dots, \mathbf{y}_I^{(n)} \sim N(\hat{\mathbf{y}}(\mathbf{x}), \hat{\Sigma}(\mathbf{x}))$. Für $\hat{\Sigma}(\mathbf{x})$ gilt dabei, dass das j -te Diagonalelement mit $j = 1, \dots, I$ gleich $s_{\hat{\mathbf{y}}_j(\mathbf{x})}^2$ ist und die Off-Diagonalelemente die Kovarianz der Zielgrößen widerspiegeln.

6.4.1 Ansatz 1 - Improvement des Wünschbarkeitsindex

Dieser Ansatz reduziert das multivariate auf ein univariates Optimierungsproblem, indem die Expected Improvement über den Wünschbarkeitsindex definiert wird. Seien $\mathbf{w}(\mathbf{x}_1), \dots, \mathbf{w}(\mathbf{x}_n)$ die Wünschbarkeitsvektoren der beobachteten Versuchspunkte und $WI(\mathbf{x}_1), \dots, WI(\mathbf{x}_n)$ die zugehörigen Wünschbarkeitsindizes. Bezeichne mit $WI_{max} = \max(WI(\mathbf{x}_1), \dots, WI(\mathbf{x}_n))$ das aktuell beobachtete Maximum der Wünschbarkeitsindizes der n Versuchspunkte. Für das Beispiel aus Abbildung 6.10 ist der Wünschbarkeitsvektor mit maximalem Wünschbarkeitsindex WI_{max} rot markiert.

Da $\mathbf{Y}(\mathbf{x})$ als Zufallsvariable betrachtet wird, ist auch $WI(\mathbf{x})$ eine Zufallsvariable. Die Improvement bezüglich des Wünschbarkeitsindex an der Stelle \mathbf{x} wird nun

Abbildung 6.10: WI der Wünschbarkeitsvektoren \mathbf{w}

definiert als

$$I^g(\mathbf{x}) = \begin{cases} (WI(\mathbf{x}) - WI_{max})^g & , \text{ falls } WI(\mathbf{x}) > WI_{max} \\ 0 & , \text{ falls } WI(\mathbf{x}) \leq WI_{max} \end{cases} . \quad (6.8)$$

Für die Expected Improvement gilt somit

$$E[I^g(\mathbf{x}) | \mathbf{y}_1^{(n)}, \dots, \mathbf{y}_I^{(n)}] = \int_{WI_{max}}^1 (z - WI_{max})^g f_{WI(\mathbf{x}) | \mathbf{y}_1^{(n)}, \dots, \mathbf{y}_I^{(n)}}(z) dz. \quad (6.9)$$

Zur Berechnung der Expected Improvement wird die bedingte Dichte $f_{WI(\mathbf{x}) | \mathbf{y}_1^{(n)}, \dots, \mathbf{y}_I^{(n)}}$ des Wünschbarkeitsindex WI benötigt. Diese Dichte wäre nun theoretisch aus der multivariaten Unsicherheitsverteilung des Zielgrößenvektors $\mathbf{Y}(\mathbf{x})$ mit $\mathbf{Y}(\mathbf{x}) \sim N(\hat{\mathbf{y}}(\mathbf{x}), \hat{\Sigma}(\mathbf{x}))$ zu bestimmen. Wie in Kapitel 6.3.4 schon argumentiert wurde, gestaltet sich die Herleitung der Verteilung des Wünschbarkeitsindex jedoch als sehr schwierig. Selbst unter der Annahme unkorrelierter Zielgrößen konnte die exakte Verteilung des Wünschbarkeitsindex nur für einige wenige Spezialfälle hergeleitet werden, vgl. Trautmann und Weihs (2006). Ein Ausweg bei der Bestimmung der Verteilung des Wünschbarkeitsindex ist, diese mit Hilfe von Monte Carlo Simulationen zu erzeugen. Dazu werden für einen festen Versuchspunkt \mathbf{x} N Realisationen des Zufallsvektors $\mathbf{Y}(\mathbf{x})$ mit $\mathbf{Y}(\mathbf{x}) \sim N(\hat{\mathbf{y}}(\mathbf{x}), \hat{\Sigma}(\mathbf{x}))$ gezogen und für jede dieser Realisationen der Wünschbarkeitsindex berechnet. Mit steigendem Stichprobenumfang N nähert

sich die empirische Dichte der wahren Dichte des Wünschbarkeitsindexes immer weiter an.

Um das Konvergenzverhalten zu verdeutlichen, wurde ein zweidimensionales Beispiel betrachtet. Dazu wurde unterstellt, dass $\mathbf{Y}(\mathbf{x}) \sim N((3, 6), \Sigma)$, wobei $\Sigma_{jj} = 0.5$ mit $j=1,2$. Die Korrelation zwischen den Zielgrößen wurde auf 0.7 gesetzt. Zudem wurden einseitige Wünschbarkeitsfunktion von Derringer und Suich verwendet mit $USG_1 = 1$, $T_1 = 5$ und $USG_2 = 4$, $T_2 = 8$. Zur Berechnung des Wünschbarkeitsindexes wurden gleiche Gewichte für die beiden Zielgrößen verwendet. In Abbildung 6.11 sind für zwei unterschiedliche Stichprobenumfänge N je drei Realisationen der Monte-Carlo-Simulation der Dichte des Wünschbarkeitsindexes angegeben.

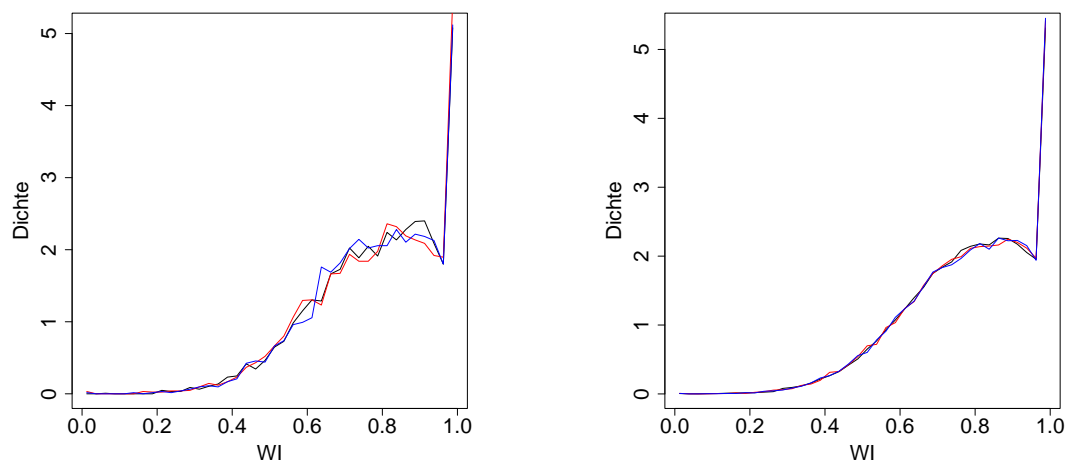


Abbildung 6.11: a) Empirische Dichten von WI mit a) $N = 5.000$, b) $N = 50.000$

Für eine gute Anpassung ist offensichtlich ein sehr großer Stichprobenumfang nötig. Da die Vorhersagen $\hat{\mathbf{y}}(\mathbf{x})$ und $\hat{\Sigma}(\mathbf{x})$ vom Versuchspunkt \mathbf{x} abhängen, müssten diese Monte Carlo Simulationen für jeden Versuchspunkt $\mathbf{x} \in \mathcal{X}$ erneut durchgeführt werden. Der Zeitaufwand dieses Vorgehens würde jeglichen Zeitrahmen sprengen.

Statt die Expected Improvement über die Dichtefunktion des Wünschbarkeitsindexes zu bestimmen, kann die Expected Improvement auch direkt mit Hilfe des schwachen Gesetzes der großen Zahlen approximiert werden. Dabei werden wie bei der Monte-Carlo-Simulation zunächst Realisationen des Zufallsvektors

$\mathbf{Y}(\mathbf{x})$ bestimmt und für jede dieser Realisationen die Improvement $I^g(\mathbf{Y}(\mathbf{x}))$ berechnet. Das arithmetische Mittel der Improvements $\overline{I^g(\mathbf{Y}(\mathbf{x}))}$ konvergiert dann fast sicher gegen die Expected Improvement $E[I^g(\mathbf{x})]$. Durch zusätzliche Berechnung von $(I^g(\mathbf{Y}(\mathbf{x})))^2$ kann zudem ein Schätzer für die Varianz der Improvement gewonnen werden, da gilt $\text{var}[I^g(\mathbf{Y}(\mathbf{x}))] = \overline{(I^g(\mathbf{Y}(\mathbf{x})))^2} - (\overline{I^g(\mathbf{Y}(\mathbf{x}))})^2$. Um eine gute Approximation für die Expected Improvement zu bekommen, reicht im Vergleich zur Monte-Carlo-Simulation ein kleinerer Stichprobenumfang, der aber auch von der Vorhersagevarianz der Zielgrößen im Versuchspunkt \mathbf{x} abhängt. Jedoch erweist sich auch dieser Ansatz selbst bei relativ kleinem Stichprobenumfang als zu rechenintensiv. Als Beispiel betrachte einen dreidimensionalen Wünschbarkeitsraum, bei dem jede Achse in 100 äquidistante Punkte unterteilt wird. Schon in diesem niedrig-dimensionalen Versuchsraum dauert die Berechnung der Expected Improvement zu lange. Die Berechnung der Expected Improvement für alle Gitterpunkte bei einem Stichprobenumfang von nur 5000 pro Versuchspunkt benötigt bei einem AMD Athlon 2600+ mehr als 10 Stunden. Um dieses Problem zu lösen, kann die Expected Improvement für jeden Gitterpunkt aus dem Versuchsraum zunächst mit einem kleinen Stichprobenumfang von z. B. 100 approximiert werden. Der Versuchsraum lässt sich dann einschränken, indem alle Punkte des Gitters eliminiert werden, für die eine obere Schranke der approximierten Expected Improvement einen bestimmten Grenzwert nicht überschreitet. Die Bestimmung einer oberen Schranke erfolgt mit Hilfe von Konfidenzintervallen für $\overline{I^g(\mathbf{Y}(\mathbf{x}))}$. Für die restlichen Punkte des Gitters wird anschließend eine genauere Approximation der Expected Improvement durch Erhöhung des Stichprobenumfangs durchgeführt.

6.4.2 Ansatz 2 - Improvement des Wünschbarkeitsvektors

Da die Berechnung der Expected Improvement von WI zu rechenintensiv ist, wird im Folgenden ein Ansatz präsentiert, der die Improvement über den Wünschbarkeitsvektor definiert. Sei F_{max} die Punktmenge im Wünschbarkeitsraum $[0, 1]^I$, deren Wünschbarkeitsindex gleich WI_{max} ist, vgl. Abbildung 6.12. Per Konstruktion liegt mindestens einer der n beobachteten Wünschbarkeitsvektoren auf F_{max} . Da für einen neuen Versuchspunkt \mathbf{x} die Zielgröße $\mathbf{Y}(\mathbf{x})$ als Zufallsvariable be-

trachtet wird, ist auch der Wünschbarkeitsvektor $\mathbf{W}(\mathbf{x})$ eine Zufallsvariable.

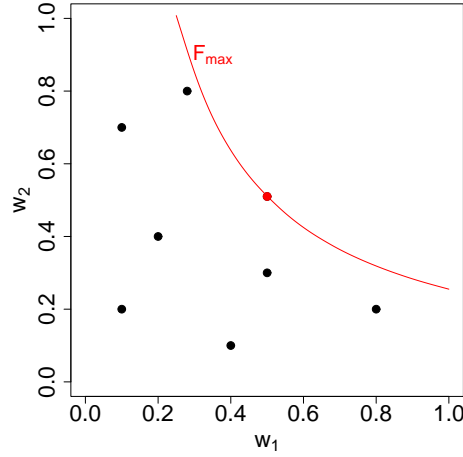


Abbildung 6.12: Punktmenge F_{max} mit $WI(\mathbf{w}) = WI_{max}$

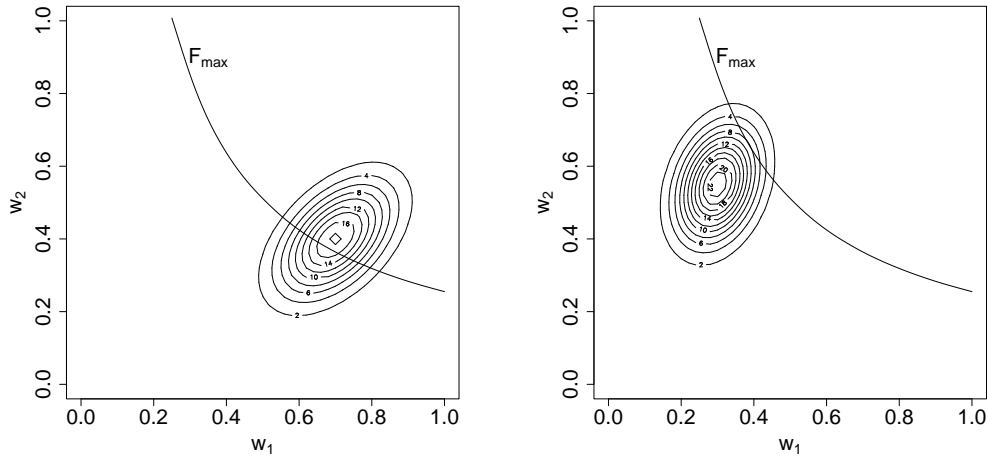
Die Improvement an einer Stelle \mathbf{x} wird nun als minimaler Abstand von $\mathbf{W}(\mathbf{x})$ zur Punktmenge F_{max} definiert. Formal lässt sich die Improvement unter Berücksichtigung des Exponenten g darstellen als

$$I^g(\mathbf{x}) = \begin{cases} \min_{\mathbf{f} \in F_{max}} \|\mathbf{f} - \mathbf{W}(\mathbf{x})\|^g & , \text{ falls } WI(\mathbf{W}(\mathbf{x})) > WI_{max} \\ 0 & , \text{ falls } WI(\mathbf{W}(\mathbf{x})) \leq WI_{max} \end{cases} \quad (6.10)$$

Liegt der Wünschbarkeitsvektor unterhalb der Punktmenge F_{max} ist die Improvement gleich 0. Die multivariate Expected Improvement an der Stelle \mathbf{x} ist nun definiert als

$$E [I^g(\mathbf{x}) | \mathbf{y}_1^{(n)}, \dots, \mathbf{y}_I^{(n)}] = \int_{\substack{\mathbf{z} \in [0,1]^I: \\ WI(\mathbf{z}) > WI_{max}}} \min_{\mathbf{f} \in F_{max}} \|\mathbf{f} - \mathbf{z}\|^g f_{\mathbf{W}(\mathbf{x}) | \mathbf{y}_1^{(n)}, \dots, \mathbf{y}_I^{(n)}}(\mathbf{z}) d\mathbf{z}. \quad (6.11)$$

In Abbildung 6.13 sind für den zweidimensionalen Fall neben F_{max} die Konturplots der bedingten Dichtefunktion von $\mathbf{W}(\mathbf{x})$ für zwei unterschiedliche Werte von \mathbf{x} eingezeichnet. Offensichtlich ist das Optimierungspotenzial in der linken Abbildung in \mathbf{x} deutlich höher.


 Abbildung 6.13: Dichte des Wünschbarkeitsvektors für unterschiedliche \mathbf{x}

Zur Berechnung der multivariaten Expected Improvement wird die bedingte Dichte $f_{\mathbf{W}(\mathbf{x})|y_1^{(n)}, \dots, y_I^{(n)}}$ des Wünschbarkeitsvektors benötigt. Für unkorrelierte Zielgrößen ergibt sich diese Dichte aus dem Produkt der Einzeldichten $f_{W_j(\mathbf{x})|y_j^{(n)}}$ für $j = 1, \dots, I$. Diese wurden in Kapitel 6.3 sowohl für Harringtons als auch Derringer und Suichs Wünschbarkeitsfunktionen gegeben. Bei Derringer und Suichs Wünschbarkeitsfunktionen beschränkt sich die Angabe der Dichten jedoch auf stückweise lineare Wünschbarkeitsfunktionen. Für korrelierte Zielgrößen ist die Verteilung des Wünschbarkeitsvektors direkt über die multivariate Normalverteilung des Zielgrößenvektors mit $\mathbf{Y}(\mathbf{x}) \sim N(\hat{\mathbf{y}}(\mathbf{x}), \hat{\Sigma}(\mathbf{x}))$ herzuleiten. Dies erweist sich jedoch als sehr schwierig. Nur für den Spezialfall einseitiger Wünschbarkeitsfunktionen nach Harrington konnte die Unsicherheitsverteilung von $\mathbf{W}(\mathbf{x})$ bestimmt werden. Dabei ist die Dichte von $e^{-e^{-\mathbf{Y}^*}}$ zu bestimmen, mit $\mathbf{Y}^* = (Y_1^*, \dots, Y_I^*)'$ und $Y_j^* = b_{0j} + b_{1j}Y_j$ für $j = 1, \dots, I$, vgl. Kapitel 6.3.1. Die Herleitung dieser Dichte erfolgt in drei Schritten, wobei zunächst vom allgemeinen Fall mit $\mathbf{Y} \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ mit $\mathbf{Y} = (Y_1, \dots, Y_I)$ ausgegangen wird.

Schritt 1. Sei $\mathbf{Y} \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. Dann gilt, dass $-\mathbf{Y}^* \sim N(\boldsymbol{\mu}^*, \boldsymbol{\Sigma}^*)$ mit $\boldsymbol{\mu}^* = -(\mathbf{b}_0 + \mathbf{b}_1 \mathbf{Y})$ und $\boldsymbol{\Sigma}^* = \mathbf{b}_1 \boldsymbol{\Sigma} \mathbf{b}_1'$.

Schritt 2. Mit Hilfe des Transformationssatzes für multivariate Dichten (Rohatgi, 1974, Seite 135) lässt sich zeigen, dass die Dichte $g(\mathbf{z})$ mit $\mathbf{Z} = e^{-\mathbf{Y}^*}$

gegeben ist durch

$$g(\mathbf{z}) = \frac{1}{\sqrt{(2\pi)^I |\boldsymbol{\Sigma}^*| \prod_{j=1}^I z_j}} e^{-\frac{1}{2} (\log(\mathbf{z}) - \boldsymbol{\mu}^*)' \boldsymbol{\Sigma}^{*-1} (\log(\mathbf{z}) - \boldsymbol{\mu}^*)}$$

Schritt 3. Abschließend bleibt die Dichte von $\mathbf{W} = e^{-\mathbf{Z}}$ zu bestimmen. Auch hier lässt sich durch erneutes Anwenden des Transformationssatzes zeigen, dass gilt

$$g(\mathbf{w}) = \frac{(-1)^I}{\sqrt{(2\pi)^I |\boldsymbol{\Sigma}^*| \prod_{j=1}^I \log(w_j) w_j}} e^{-\frac{1}{2} \mathbf{a}' \boldsymbol{\Sigma}^{*-1} \mathbf{a}},$$

wobei $\mathbf{a} = (\log(-\log(\mathbf{w})) - \boldsymbol{\mu}^*)$.

Satz 6.4.1 Sei $\mathbf{Y}(\mathbf{x}) \sim N(\hat{\mathbf{y}}(\mathbf{x}), \hat{\boldsymbol{\Sigma}}(\mathbf{x}))$ und $w_j(\mathbf{x})$ für $j = 1, \dots, I$ einseitige Wünschbarkeitsfunktionen nach Harrington. Dann ist die Dichtefunktion von $\mathbf{w}(\mathbf{x}) = (w_1(\mathbf{x}), \dots, w_I(\mathbf{x}))'$ gegeben durch

$$f(\mathbf{w}(\mathbf{x})) = \frac{(-1)^I}{\sqrt{(2\pi)^I |\tilde{\boldsymbol{\Sigma}}(\mathbf{x})| \prod_{j=1}^I \log(w_j(\mathbf{x})) w_j(\mathbf{x})}} e^{-\frac{1}{2} \mathbf{a}' \tilde{\boldsymbol{\Sigma}}(\mathbf{x})^{-1} \mathbf{a}}, \quad (6.12)$$

wobei $\mathbf{a} = (\log(-\log(\mathbf{w}(\mathbf{x}))) - \tilde{\boldsymbol{\mu}}(\mathbf{x}))$ mit $\tilde{\boldsymbol{\mu}}(\mathbf{x}) = \mathbf{b}_0 + \mathbf{b}_1 \hat{\mathbf{y}}(\mathbf{x})$. Die Matrix $\tilde{\boldsymbol{\Sigma}}(\mathbf{x})$ ist definiert durch $\tilde{\Sigma}_{jj}(\mathbf{x}) = b_{1j}^2 s_{y_j(\mathbf{x})}^2$ für $j = 1, \dots, I$ und $\tilde{\Sigma}_{ij}(\mathbf{x}) = b_{1i} b_{1j} \Sigma_{ij}(\mathbf{x})$ mit $i \neq j$ für $i, j = 1, \dots, I$.

Die Berechnung der multivariaten Expected Improvement aus Formel (6.11) kann nun numerisch erfolgen. Da die Konturlinien des Wünschbarkeitsindex ab Werten von $WI = 0.6$ ungefähr parallel zueinander liegen, kann erst für $WI_{max} > 0.6$ von einem relativ konstanten Verhältnis der Expected Improvement zur Verbesserung des aktuell beobachteten Maximums WI_{max} ausgegangen werden, vgl. Abbildungen 6.7 a) und b). Diese Tatsache ist bei der Wahl des Abbruchkriteriums zu berücksichtigen. Für Werte von $WI_{max} > 0.6$ wäre somit ein mögliches Abbruchkriterium die Modellverfeinerung zu stoppen, sobald ein bestimmtes

Vielfaches der Expected Improvement kleiner als 1 Prozent des aktuell beobachteten Maximums WI_{max} ist. Für Werte von $WI_{max} < 0.6$, sollte die Entscheidung Abzubrechen von der Lage der Verteilung des Wünschbarkeitsvektors im Wünschbarkeitsraum abhängen.

6.5 Multivariates Anwendungsbeispiel

Im Folgenden wird die multivariate Expected Improvement, wie sie in Kapitel 6.4.2 definiert wurde, auf ein Optimierungsproblem mit zwei nicht deterministischen Zielgrößen Y_1 und Y_2 angewandt. Aus Visualisationsgründen wird nur eine Einflussgröße x mit $x \in [1, 12]$ betrachtet. Der Wertebereich von Y_1 sei das Intervall der reellen Zahlen $[8, \infty)$ und für Y_2 sei der Wertebereich durch \mathbb{R}^+ definiert. Der funktionale erwartete Zusammenhang zwischen den Zielgrößen und dem Einflussfaktor ist in Abbildung 6.14 a) gegeben. Die Korrelation zwischen den beiden Zielgrößen beträgt -0.23 . Beide Zielgrößen sind zu minimieren. Die Form der Wünschbarkeitsfunktion wird durch Harringtons einseitige Wünschbarkeitsfunktion beschrieben, vgl. Abbildung 6.14 b). Mit Hilfe dieser Wünschbarkeitsfunktionen lassen sich nun für jeden Versuchspunkt x die erwarteten Zielgrößen zu Wünschbarkeiten transformieren, vgl. Abbildung 6.15 a).

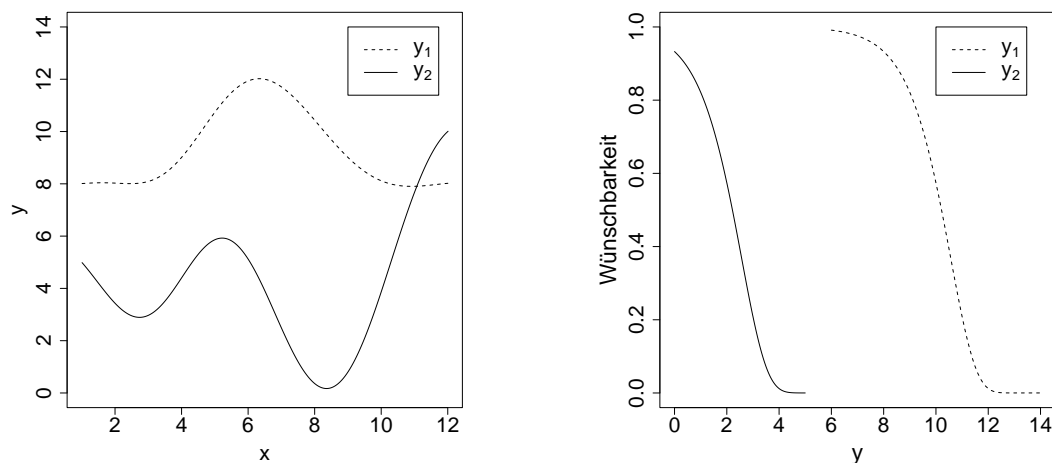


Abbildung 6.14: a) Erwarteter funktionaler Zusammenhang b) Harringtons einseitige Wünschbarkeitsfunktionen

Für die Zielgröße Y_1 sind Werte des Einflussfaktors x an den beiden Rändern des Versuchsbereichs optimal. Für die Zielgröße Y_2 ist der Verlauf der Wünschbarkeiten bimodal. Die maximale Wünschbarkeit wird bei $x = 8.2$ erreicht. Diese Einzelwünschbarkeiten lassen sich nun zum Wünschbarkeitsindex kombinieren. Beide Zielgrößen seien gleich wichtig. Daher wurde zur Berechnung des Wünschbarkeitsindex das einfache geometrische Mittel verwendet. Der resultierende Wünschbarkeitsindex in Abhängigkeit von x ist in Abbildung 6.15 b) zu sehen. Das Maximum des Wünschbarkeitsindex stellt unter den gegebenen Spezifikationsanforderungen den besten Kompromiss zwischen den beiden Zielgrößen dar. Das globale Optimum des Wünschbarkeitsindex befindet sich an der Schnittstelle der beiden Wünschbarkeiten bei $x = 8.9$, vgl. Abbildung 6.15 a) und b). Zudem wird ein lokales Optimum bei $x = 2.7$ beobachtet. Im Folgenden wird die multivariate Expected Improvement auf dieses Beispiel angewandt.

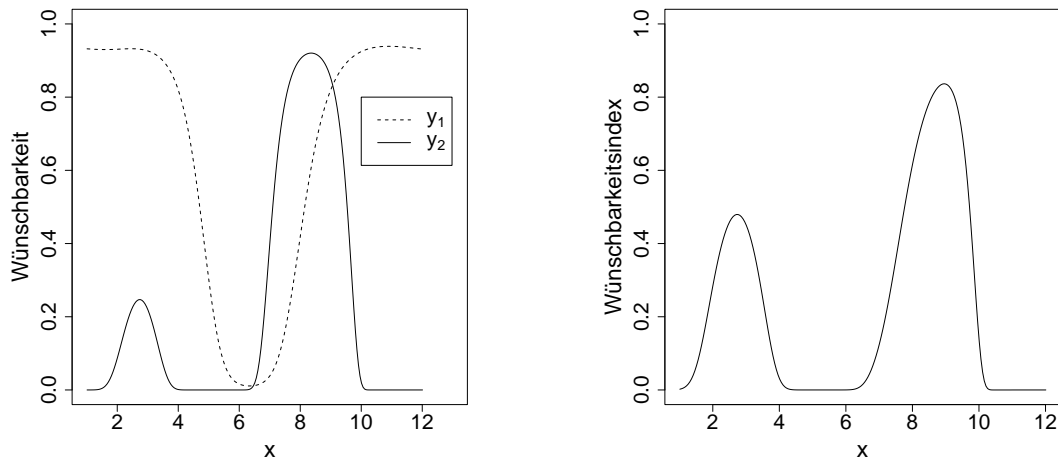


Abbildung 6.15: a) Erwartete Wünschbarkeiten abhängig von x , b) Erwarteter Wünschbarkeitsindex abhängig von x

Gegeben seien Beobachtungen der beiden Zielgrößen an den Stellen $x = 1, 2, 3, 4, 5$ und 12 . Zur Simulation nicht deterministischer Zielgrößen wurde zu den Zielgrößen additiv ein Zufallsfehler e hinzugefügt mit $e \sim N(0, 0.001)$. Um die Optimierungsaufgabe schwieriger zu gestalten, wurden im Bereich des globalen Optimums bewusst keine Versuchspunkte platziert. Basierend auf diesen Versuchspunkten wurde für jede der beiden Zielgrößen ein Gauß-Prozess-Modell angepasst. Die deterministische Trendkomponente wurde dabei durch einen Inter-

cept beschrieben. Zudem wurde die Power-Exponential-Korrelationsfunktion aus (5.5) mit $s = 2$ verwendet. Die Vorhersage der Initialmodelle und die Wurzel der Vorhersagevarianz sind in Abbildung 6.16 dargestellt. Beide Initialmodelle bilden den wahren Zusammenhang im Bereich $[5,12]$ nicht gut ab, da dort keine Versuchspunkte beobachtet wurden. Dies spiegelt sich auch an der hohen Standardabweichung des Vorhersagefehlers in diesem Bereich wider.

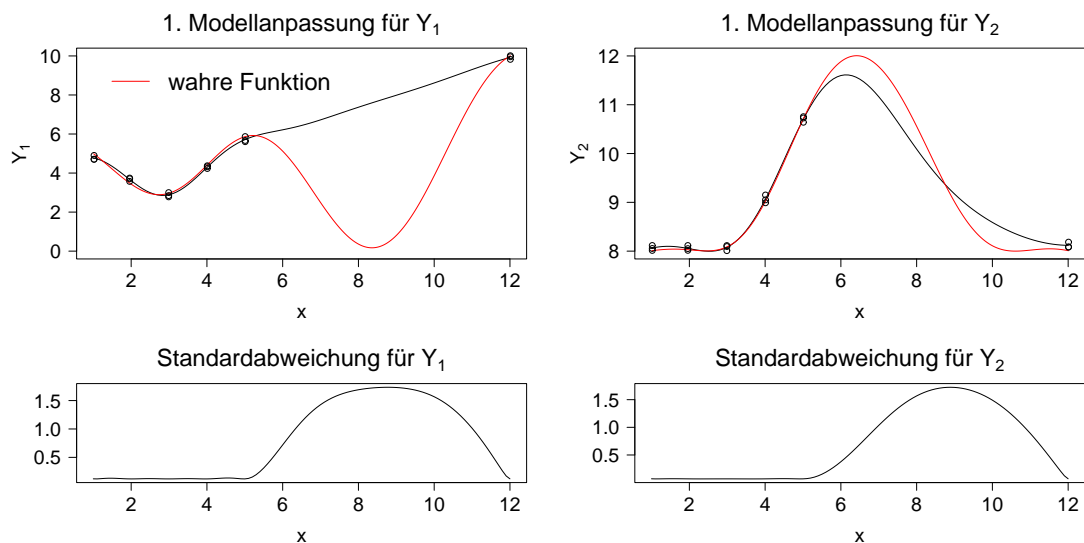


Abbildung 6.16: 1. Modellanpassung der beiden Zielgrößen

Aufbauend auf diesen Modellen wird nun die multivariate Expected Improvement aus Kapitel 6.4.2 berechnet. Der Parameter g , der die Balancierung zwischen lokaler und globaler Suche steuert, wurde auf 3 gesetzt. Die resultierende multivariate Expected Improvement besitzt zwei Peaks, vgl. Abbildung 6.17. Während der eine Peak im Bereich liegt wo die Kombination der Vorhersagen der beiden Zielgrößen optimal ist, liegt der zweite Peak im Bereich mit großer Unsicherheit. Das Maximum der Expected Improvement liegt im ersten Peak bei $x = 2.7$. Die multivariate Expected Improvement identifiziert somit im ersten Schritt das lokale Optimum des Prozesses, vgl. Abbildung 6.15 b). In diesem Versuchspunkt werden drei neue Beobachtungen erhoben und die Modelle angepasst. Der Fit beider Modelle ändert sich nur minimal, da zum einen der Abstand des neu hinzugefügten Versuchspunktes nur 0.3 zum nächsten Versuchspunkt des Initialdesigns beträgt und zum anderen die Vorhersage an dieser Stelle sehr gut ist.

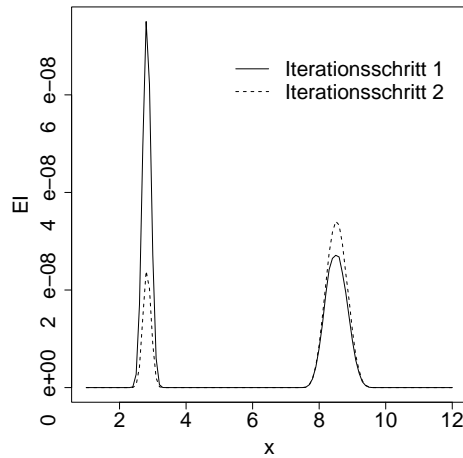


Abbildung 6.17: Multivariate Expected Improvement (Schritt 1 und 2)

Das Maximum der aktualisierten multivariaten Expected Improvement verschiebt sich nun zum zweiten Peak und liegt bei $x = 8.5$, vgl. Abbildung 6.17. Somit werden im zweiten Schritt drei Wiederholungsmessungen im Bereich mit großer Modellunsicherheit erhoben. Dies führt zu einer deutlichen Verbesserung der Modellanpassung, vgl. Abbildung 6.18 a) und b).

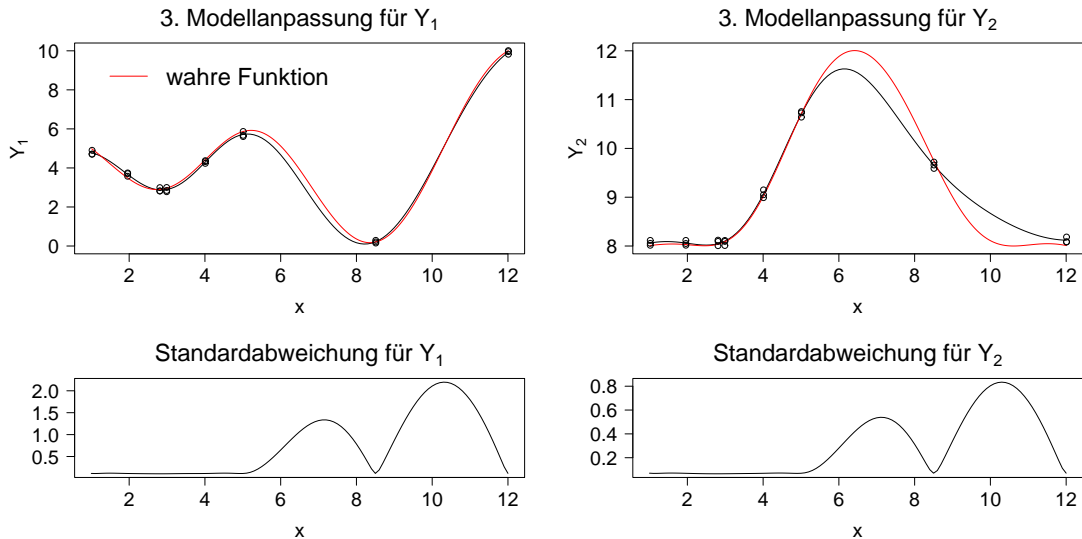


Abbildung 6.18: Modelle der Zielgrößen nach dem 2. Iterationsschritt

Da in dem aktuell hinzugefügten Versuchspunkt der resultierende Wünschbarkeitsindex deutlich größer ist als an den bisher beobachteten Versuchspunkten, wird nun auf der Suche nach dem globalen Optimum der Bereich um $x = 8.5$

näher untersucht. Die multivariate Expected Improvement identifiziert im dritten Schritt die Stelle $x = 9.6$, vgl. Abbildung 6.19 a).

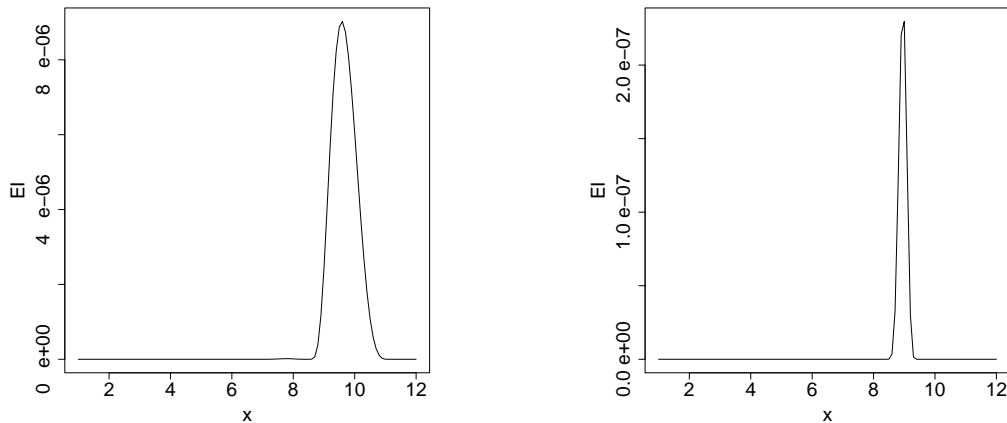


Abbildung 6.19: a) Multivariate Expected Improvement (Iterationsschritt 3), b) Multivariate Expected Improvement (Iterationsschritt 4)

Werden die dort erhobenen Wiederholungsmessungen bei der Modellanpassung berücksichtigt, ergibt sich ein sehr guter Fit für beide Zielgrößen, vgl. Abbildung 6.20. Mit Hilfe der multivariaten Expected Improvement wird anschließend im vierten Iterationsschritt das globale Optimum bei $x = 8.9$ identifiziert, vgl. Abbildung 6.19 b).

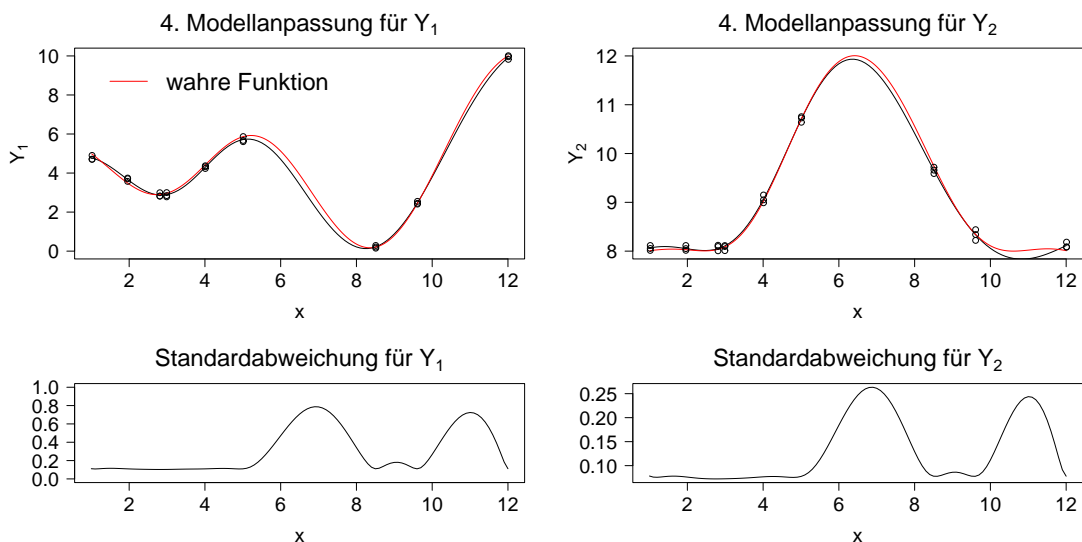


Abbildung 6.20: Modelle der Zielgrößen nach dem 3. Iterationsschritt

Kapitel 7

Optimierung des Drückprozesses

Im Folgenden wird die in den vorherigen Kapiteln vorgestellte adaptive sequentielle Optimierungsprozedur auf den Drückprozess angewandt. Als zu fertigende Bauteilgeometrie wird dabei ein Napf ausgewählt, vgl. Abbildung 7.1 a). Die Sollgeometrie des Napfes entspricht einer Blechdicke BD von 1 mm, einem Durchmesser D von 71 mm und einer Napftiefe NT von 81 mm, vgl. Abbildung 7.1 b). Diese Eigenschaften des zu fertigenden Napfes legen den Durchmesser (165 mm) und die Blechdicke (1 mm) der Metallronde fest, aus der das Bauteil geformt wird. Als Material des Napfes wurde halbhartes Aluminium Al99,5 gewählt.

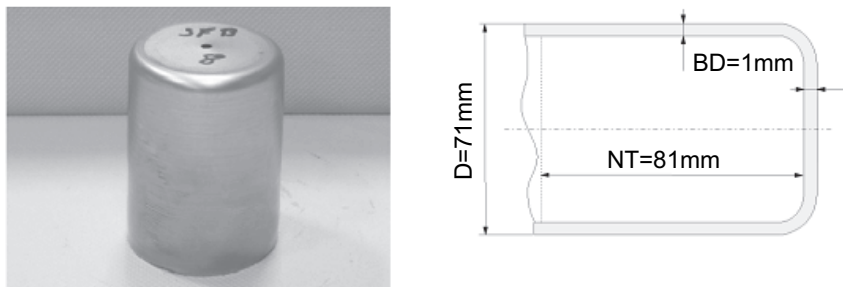


Abbildung 7.1: a) Musterbauteil eines Napfes b) Sollgeometrie

Aufgrund der so genannten Zipfelbildung am Rand des gefertigten Napfes ist ein Nachbearbeitungsschritt nötig, um den Rand durch Abschneiden zu begradigen, vgl. Abbildung 7.1 b). Dazu wird an verschiedenen Messpositionen die Napftiefe des Bauteils gemessen und die minimale Napftiefe bestimmt. Das überstehende

Material wird anschließend bei dieser Napftiefe abgeschnitten. Diese minimale Napftiefe NT wird im Folgenden als Zielgröße erhoben. Da die Bauteile nach der Fertigung unterschiedlich eng am Futter anliegen und die Bauteile in der Regel nicht exakt rund verlaufen, wird zudem der maximale Durchmesser D_{max} des Napfes gemessen. Außerdem wird die binäre Zielgröße Versagen beobachtet, wobei nicht zwischen den Versagensformen Riss und Faltenbildung unterschieden wird. Grund dafür ist das eher seltene Auftreten von Rissbildung. Des Weiteren wird als wirtschaftlicher Faktor die Produktionszeit t berücksichtigt.

Anhand ausführlicher Untersuchungen im SFB 475 Projekt C6 „Methoden der statistischen Versuchsplanung zur Analyse und Optimierung von komplexen Umformprozessen am Beispiel des CNC-gesteuerten Drückens“ konnte gezeigt werden, dass die Betrachtung von drei wesentlichen Maschinenparametern ausreicht, um sehr gute Optimierungsergebnisse erzielen zu können, vgl. Ewers (2005). Dies sind neben dem Vorschub f die beiden Parameter e_{AQU} und V_{AQU} , die neben anderen Parametern die Verfahrbewegung der Drückrolle beschreiben, vgl. Kapitel 2.1. Die Grenzen des zu untersuchenden Versuchsraums dieser drei Parameter wurden in Zusammenarbeit mit Drückexperten bestimmt, vgl. Tabelle 7.1.

	e_{AQU}	V_{AQU}	f
untere Grenze	20	0.95	650
obere Grenze	60	1.05	3000

Tabelle 7.1: Versuchsraum

Die restlichen Maschinenfaktoren wurden konstant gehalten, wobei bei der Wahl der Einstellungen die bisherigen Untersuchungsergebnisse des SFB-Projekts berücksichtigt wurden. Die Spindeldrehzahl wurde auf 1000 Umdrehungen pro Minute eingestellt, ein Durchmesser der Drückrolle von 10 mm gewählt und die Anzahl der Stufen im Stadienplan auf 20 gesetzt. Zudem wurde eine leicht konvex gekrümmte Form der Bahnen gewählt.

7.1 Vorhersage des stabilen Bereichs

Im Rahmen der Untersuchungen des SFB Projekts C6 wurden bei der Optimierung kompletter Bauteile bisher hauptsächlich Näpfe betrachtet. Da in der Falldatenbank mehrere Optimierungsergebnisse unterschiedlicher Näpfe gespeichert sind, wird im Folgenden mit Hilfe des fallbasierten Schließens der stabile Bereich für den zu fertigenden Napf vorhergesagt. Eine Durchsichtung der Falldatenbank ergab, dass zwei Optimierungsfälle herangezogen werden können. Bei beiden Optimierungsfällen wurden Näpfe mit dem gleichen Futter hergestellt. Der Sollwert des maximalen Durchmessers D_{max} für diese Bauteile beträgt damit ebenfalls 71 mm. Zudem wurde das gleiche Material verwendet. Diese Bauteilreihen unterscheiden sich von der zu fertigenden Bauteilgeometrie somit nur in der Napftiefe (69 und 86 mm). Für das Bauteil mit Napftiefe $NT = 69$ mm sind 30 Versuchspunkte in der Datenbank gespeichert und für das Bauteil mit Napftiefe $NT = 86$ mm 22 Versuche, vgl. Tabelle A.1 und A.2. In Abbildung 7.2 sind diese Versuchspunkte bezüglich des Kriteriums Bauteilversagen dargestellt. Offensichtlich ist der stabile Bereich für das Bauteil mit kleinerer Napftiefe deutlich größer.

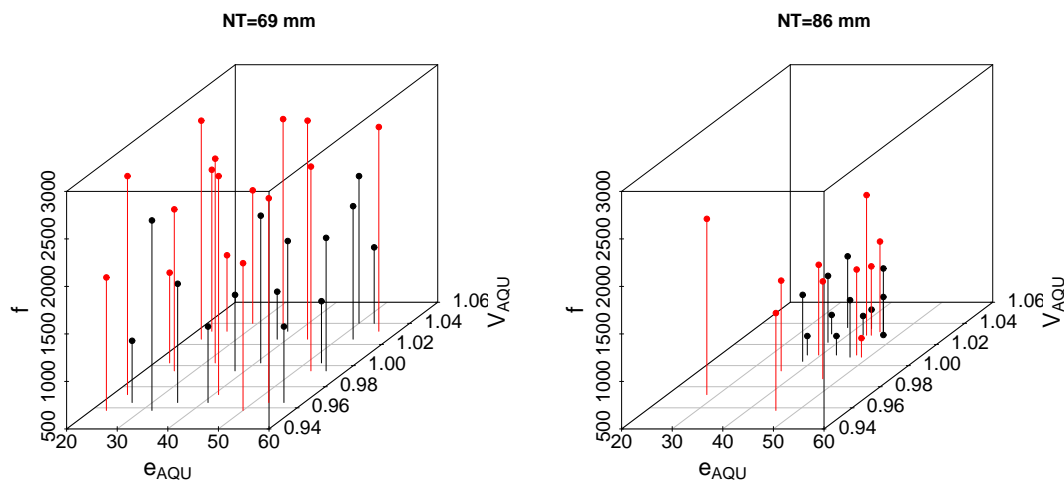


Abbildung 7.2: a) Versuchspunkte aus der Falldatenbank (rot: Bauteilversagen) für a) $NT = 69$ mm und b) $NT = 86$ mm

Betrachtet man für das Bauteil mit Napftiefe $NT = 69$ mm die Lage der Punkte im stabilen Bereich, so lässt sich vermuten, dass sich der stabile Bereich paral-

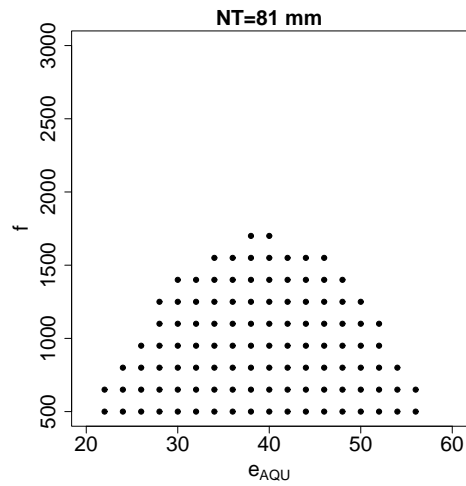
lel zur Achse V_{AQU} als Schlauch durch den Versuchsraum zieht, vgl. Abbildung 7.2 a).

Um nun den stabilen Bereich für den Napf mit einer Tiefe von $NT = 81$ mm vorherzusagen, wird das Hyper-Modell der logistischen Regression verwendet. Dazu werden beide Datensätze kombiniert und der zusätzliche Faktor Napftiefe eingeführt. Neben dem vollen quadratischen Modell der drei Maschinenparameter e_{AQU} , V_{AQU} und f wird zudem der Faktor Napftiefe NT und dessen 2-fach-Interaktionen mit den drei Maschinenparametern berücksichtigt. Um die signifikanten Faktoren dieses Modells zu bestimmen, wurde die Vorwärtsselektion zum Niveau 0.05 durchgeführt. Die Parameter des reduzierten Modells, die Standardabweichungen der Parameterschätzer und die p-Werte des Tests von Wald sind in Tabelle 7.2 aufgelistet. Dieses Modell setzt sich aus drei Haupteffekten und dem quadratischen Faktor von e_{AQU} zusammen. Nur der Faktor V_{AQU} besitzt keinen signifikanten Einfluss. Ein negatives Vorzeichen bedeutet, dass mit steigenden Werten des Faktors die Wahrscheinlichkeit für ein Bauteilversagen steigt.

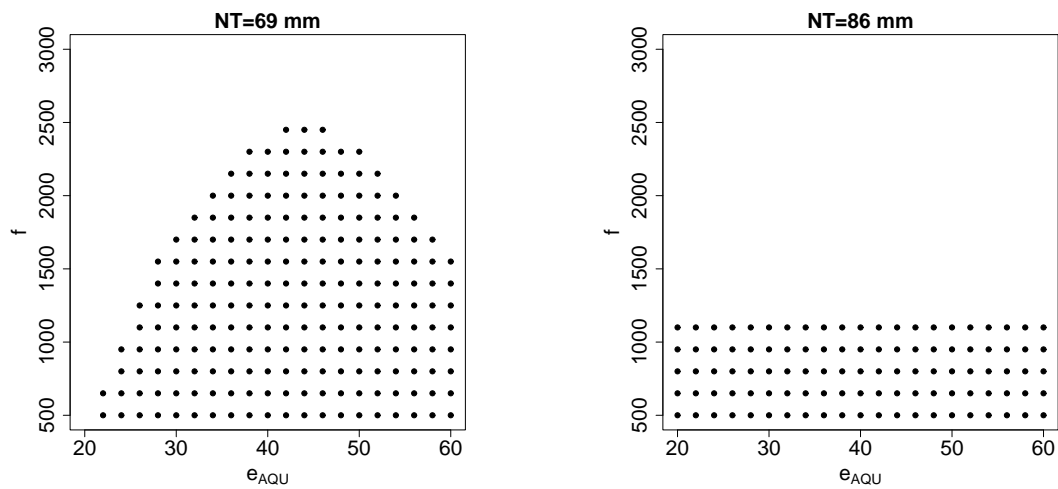
Parameter	Schätzer	Standard- abweichung	p-Wert (Wald)
Intercept	42.0013	15.9276	0.0084
e_{AQU}	1.8742	0.8276	0.0235
f	-0.0054	0.0015	0.0005
NT	-0.4410	0.1530	0.0039
e_{AQU}^2	-0.0220	0.0100	0.0281

Tabelle 7.2: Reduziertes Hyper-Modell der logistischen Regression

Anhand dieses Modells wird nun der stabile Bereich für den zu fertigenden Napf mit $NT = 81$ mm vorhergesagt. Da der Parameter V_{AQU} keinen Einfluss besitzt, lässt sich die Vorhersage durch einen zweidimensionalen Schnitt im Parameterraum darstellen, vgl. Abbildung 7.3. Die Punktwolke gibt dabei den Bereich im Versuchsraum an, bei dem die prognostizierte Wahrscheinlichkeit, dass kein Bauteilversagen eintritt, größer als 50 % ist.

Abbildung 7.3: Vorhersage des stabilen Bereichs für $NT = 81$ mm

Um die Veränderung des stabilen Bereichs mit steigender Bauteilhöhe zu verdeutlichen, sind in den Abbildungen 7.4 a) und b) die Vorhersagen des stabilen Bereichs für die beobachteten Napftiefen von $NT = 69$ mm und $NT = 86$ mm geplottet.

Abbildung 7.4: Vorhersage des stabilen Bereichs für a) $NT = 69$ mm und b) $NT = 86$ mm

7.2 Sequentielle Generierung eines raumfüllenden Versuchsplans

Um den gesamten unbekannt stabilen Bereich mit möglichst wenig Versagenspunkten untersuchen zu können, wird nun der in Kapitel 4.2 vorgestellte Algorithmus zur sequentiellen Generierung eines raumfüllenden Versuchsplans eingesetzt. Die Verwendung des erweiterten Minimax-Kriteriums aus Kapitel 4.2.3 war zu diesem Zeitpunkt wegen fehlender Effizienz bezüglich der Rechenzeit leider noch nicht möglich. Da die Berechnung der konvexen Hüllen mit der Software R erfolgt, wurde zur Generierung von raumfüllenden Versuchsplänen ein Kriterium verwendet, das in dieser Software bereits implementiert ist. Dieses Kriterium ist ähnlich zum Minimax-Kriterium, da es ebenfalls für jeden Gridpunkt die Abstände zu den Punkten des Versuchsplans berechnet. Während das Minimax Design mit dem minimalen Abstand eines jeden Gridpunktes zum Versuchsplan arbeitet, minimiert dieses Kriterium die Summe der summierten Abstände eines jeden Gridpunktes zu jedem Punkt des Versuchsplans. Dieses Kriterium besitzt eine gute raumfüllende Eigenschaft auch bei kleinem Versuchsumfang. Zudem können diese Versuchspläne sequentiell generiert werden und sie benötigen keinen kubischen Versuchsraum.

In einem ersten Schritt wurden mit diesem Kriterium nun zehn Versuchspunkte raumfüllend in den vorhergesagten stabilen Bereich gelegt. Die Produktion dieser Bauteile führte in vier Fällen zu einem Bauteilversagen, vgl. Abbildung 7.5 oben links.

Bei der Wahl der Anzahl der Iterationsschritte wird auf die Ergebnisse der Simulationsstudie aus Kapitel 4.3 zurückgegriffen. Demnach ergibt sich ein gutes Kosten-Nutzen-Verhältnis bei $N/5$ Iterationen, wobei N der Gesamtstichprobenumfang ist. Da für die komplette Optimierung 40 Metallrunden zur Verfügung stehen und einige Metallrunden für die sequentielle Modellverfeinerung benötigt werden, sollte der Umfang des raumfüllenden Versuchsplans ca. $N = 35$ sein. Damit ergibt sich ein gutes Kosten-Nutzen-Verhältnis bei sieben Iterationsschritten, wobei pro Iterationsschritt fünf Versuchspunkte hinzugefügt werden. Da im ersten Iterationsschritt schon zehn Versuchspunkte erhoben wurden, werden hier

insgesamt nur sechs Iterationsschritte durchgeführt.

Für die nächsten fünf sequentiellen Versuchspunkte wird als Versuchsraum ein Kubus gewählt, der den mit Hilfe des Hyper-Modells der logistischen Regression vorhergesagten stabilen Bereich zum Großteil umschließt. Die Grenzen der Parameter wurden dabei wie folgt festgelegt: $e_{AQU} = [28, 50]$, $V_{AQU} = [0.95, 1.05]$ und $f = [500, 1500]$. Bevor neue Versuchspunkte hinzugefügt werden, sind zunächst die konvexen Hüllen der ersten zehn Versuchspunkte zu berechnen. Da der Vorschub f dieser Versuchspunkte nicht stark variiert, können zunächst nur kleine Bereiche als Versagensbereich identifiziert und ausgeschlossen werden. Die Lage der fünf Versuchspunkte, die sequentiell hinzugefügt werden, ist in Abbildung 7.5 oben rechts zu sehen. Bei der Herstellung dieser Bauteile trat nur für eine Parametereinstellung ein Bauteilversagen auf.

Im dritten Iterationsschritt wird zunächst der Versuchsraum in Richtung größerer Vorschubwerte erweitert, um die Produktionszeit reduzieren zu können. Dabei wurde in Absprache mit dem Drückexperten die obere Grenze des Vorschubs von 1500 auf 1750 erhöht. Nach Aktualisierung der konvexen Hüllen der 15 beobachteten Versuchspunkte werden nun fünf weitere Versuchspunkte hinzugefügt. Für diese Versuchspunkte liegt der Vorschub zwischen 1500 und 1750. Da für drei dieser Versuchspunkte ein Bauteilversagen eintrat, wurde die Grenze zum Bauteilversagen in Richtung des Vorschubs überschritten, vgl. Abbildung 7.5 Mitte links. Wie in der Simulationsstudie aus Kapitel 4.2 empfohlen, wird diese Grenze in den folgenden Iterationsschritten genauer untersucht. Im vierten Iterationsschritt wird somit der Versuchsbereich nicht weiter vergrößert. Die konvexen Hüllen, die sich aus den 20 Versuchspunkten ergeben, sind in der Abbildung 7.6 a) und b) als zweidimensionale Schnitte durch den Parameterraum dargestellt. Statt die exakten Grenzen anzugeben sind die konvexen Hüllen hier durch Punktwolken symbolisiert. Mit Hilfe dieser konvexen Hüllen konnte im vierten Iterationsschritt bereits 30% des gesamten Versuchsraums als Versagensbereich identifiziert werden.

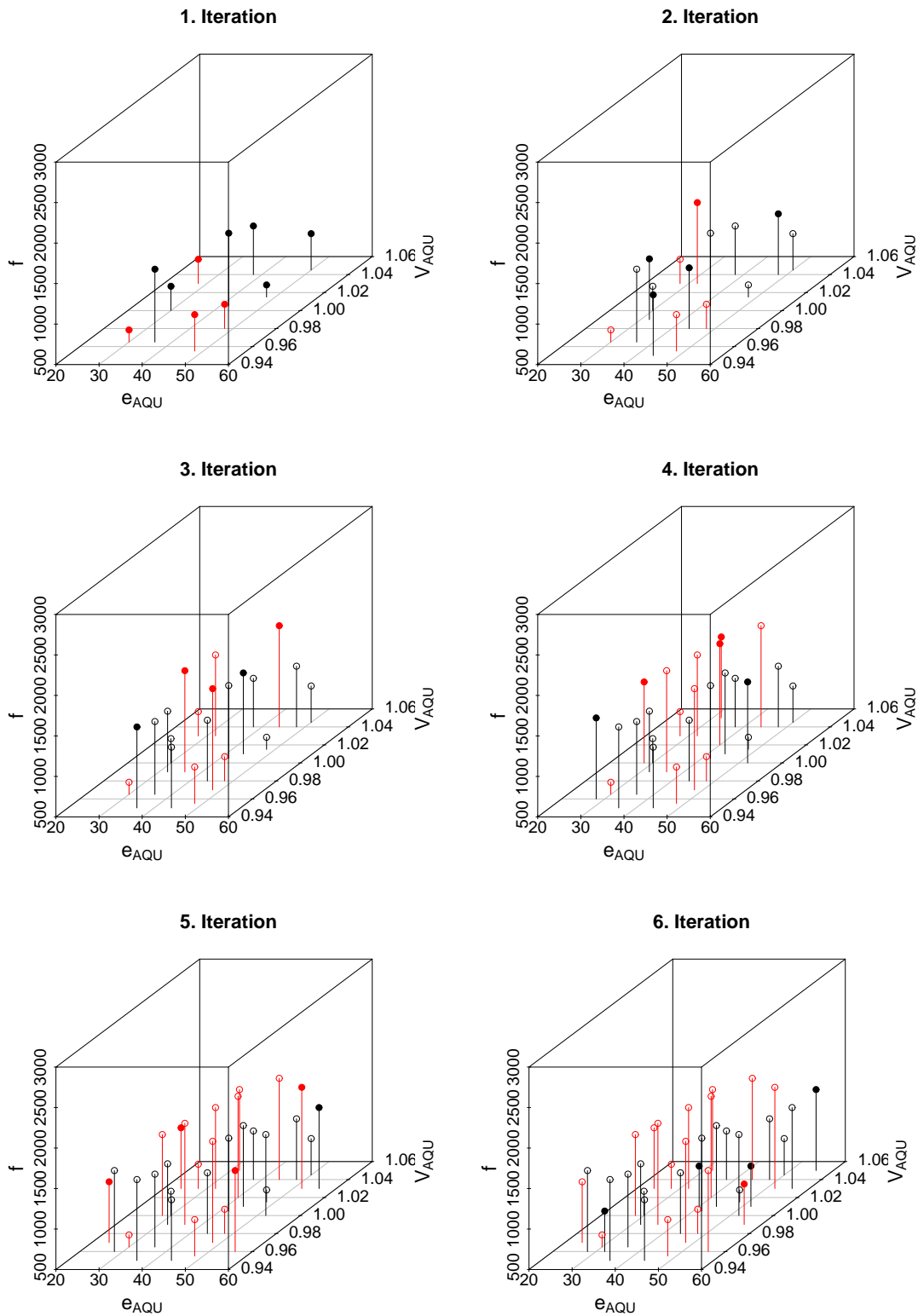


Abbildung 7.5: Sequentielle Generierung eines raumfüllenden Versuchsplans mit $N = 35$ in 6 Iterationsschritten (ausgefüllter Punkt: neuer Versuchspunkt, rot: Versagen)

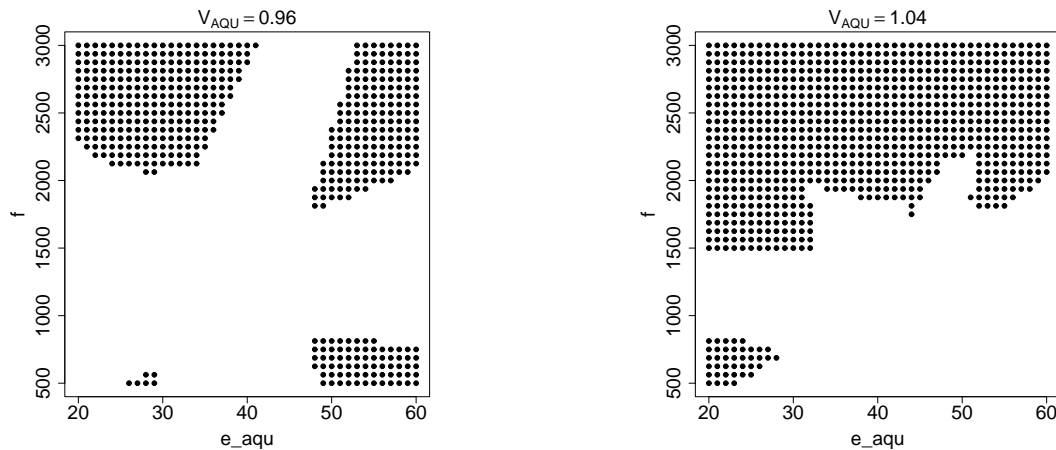


Abbildung 7.6: Konvexe Hüllen im vierten Iterationsschritt

Von den fünf neuen Versuchspunkten, die im vierten Iterationsschritt hinzugefügt wurden, trat in drei Fällen ein Bauteilversagen auf, vgl. Abbildung 7.5 Mitte rechts. Im fünften Iterationsschritt wurde der Versuchsbereich in Richtung der e_{AQU} -Achse von $e_{AQU} = [28, 50]$ auf $e_{AQU} = [24, 56]$ vergrößert, vgl. Abbildung 7.5 unten links. Die fünf neuen Versuchspunkte werden an den unteren und oberen Versuchsrand bezüglich der e_{AQU} -Ebene platziert. Bei vier von diesen fünf neuen Versuchspunkten wurde ein Bauteilversagen beobachtet. Aus diesem Grund wurde der Versuchsbereich im letzten Iterationsschritt nicht weiter vergrößert. In diesem Iterationsschritt wurde nur in einem Fall ein Bauteilversagen beobachtet, vgl. Abbildung 7.5 unten rechts. Der komplette raumfüllende Versuchsplan und die Ergebnisse sind in Tabelle A.3 gegeben. Anhand der Anordnung der Versagensbauteile zu den Nichtversagensbauteilen entsteht der Eindruck, dass der Schlauch des stabilen Bereichs im Parameterraum nicht exakt parallel zur Achse von V_{AQU} verläuft. Bei steigenden Werten von V_{AQU} bewegt sich der Schlauch in Bereiche mit größeren e_{AQU} -Werten. Zudem scheint der Schlauch mit kleiner werdendem V_{AQU} in Richtung größerem Vorschubs f anzusteigen. In Abbildung 7.7 ist der durch die konvexen Hüllen identifizierte Anteil des Versagensbereichs am gesamten Versuchsraum über die Iterationsschritte dargestellt. Nach dem sechsten Iterationsschritt konnte 51% des gesamten Versuchsraums als Versagensbereich identifiziert werden. Von den insgesamt 35 Versuchspunkten wurden 19 Versuchspunkte raumfüllend im stabilen Bereich platziert. Aufgrund des speziellen Vorgehens des Algorithmus wurden Versagenspunkte nur am Rand des

Versagensbereichs beobachtet, vgl. Abbildung 7.5 unten rechts. Die Ergebnisse aus Abbildung 7.5 lassen vermuten, dass der Anteil des stabilen Bereichs kleiner als 25 % ist. Unter der Annahme, dass der Anteil des stabilen Bereichs 25 % ist, wären bei einem einfachen raumfüllenden Versuchsplan im Mittel $35/4=8.75$ Versuchspunkte im stabilen Bereich beobachtet worden. Damit konnte durch die Anwendung des Algorithmus die Anzahl der Versuchspunkte im stabilen Bereich mehr als verdoppelt werden.

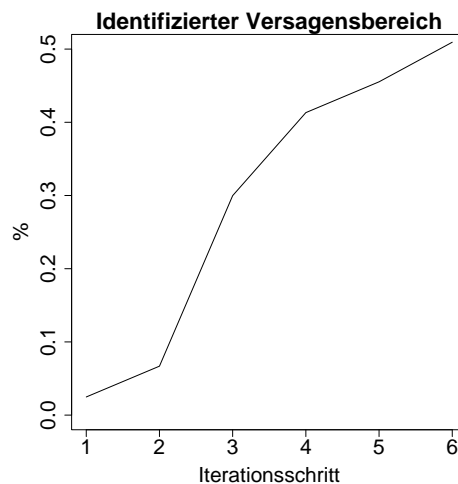


Abbildung 7.7: Prozentualer Anteil des identifizierten Versagensbereichs am gesamten Versuchsraum

7.3 Modellselektion

Basierend auf den Versuchspunkten im stabilen Bereich wird nun separat für jede Zielgröße ein Gauß-Prozess-Modell angepasst. Zunächst ist der Versuchsraum aus Tabelle 7.1 auf $[0, 1]^3$ zu standardisieren, um Konvergenzprobleme bei der Anpassung des Gauß-Prozess-Modells zu verhindern. Da der Anwender bei der Spezifikation des Gauß-Prozess-Modells sowohl für die deterministische Trendkomponente als auch für die Kovarianzfunktion verschiedene Auswahlmöglichkeiten besitzt, erfolgt zunächst eine Modellselektion für jede Zielgröße. Bei der Bestimmung der deterministischen Trendkomponente kann zwischen einem reinem Intercept, einem linearen Term oder einem linearen Term mit 2-fach Wech-

selwirkungen gewählt werden. Das volle quadratische Modell wurde nicht berücksichtigt, da in diesem Fall nur alleine für die deterministische Trendkomponente schon zehn Parameter zu schätzen sind. Addiert man die unbekanntes Kovarianzparameter hinzu, kann es sehr schnell zu einem Overfitting der Daten kommen. Aufgrund von Vorwissen über den Drückprozess ist bekannt, dass die Zusammenhänge zwischen den hier berücksichtigten drei Einflussfaktoren nicht sehr sprunghaft sind. Deswegen wurde die Wahl des Parameters s der Power-Exponential-Korrelationsfunktion auf 1 oder 2 reduziert. Werte von s kleiner als 1 führen zu sehr sprunghaften Oberflächenplots, vgl. Abbildung 5.2. Da sowohl die isotrope als auch die anisotrope Variante der Power Korrelationsfunktion gewählt werden kann, gibt es vier Spezifikationsmöglichkeiten für die Korrelationsfunktion der Gauß-Prozess-Modelle. Jede dieser Korrelationsfunktionen kann dabei mit den drei möglichen Trendkomponenten kombiniert werden. Ausgeschlossen wird lediglich die Kombination einer anisotropen Kovarianzfunktion mit einem linearen Trend unter Berücksichtigung der 2-fach Wechselwirkungen, da in diesem Modell bei festen s_i mit $i = 1, 2, 3$ sieben Parameter der Trendfunktion und vier Kovarianzparameter, also insgesamt 11 Parameter, zu schätzen wären. Insgesamt werden pro Zielgröße somit zehn Gauß-Prozess-Modelle angepasst. Zur Auswahl eines Gauß-Prozess-Modells wurde sowohl die Vorhersagegüte als auch der Fit betrachtet. Die Vorhersagegüte wurde mit Hilfe des Mean Squared Crossvalidation Error (*MSCVE*) bewertet. Der Fit der Gauß-Prozess-Modelle wurde anhand des *BIC*-Kriteriums beurteilt, das eine große Anzahl von Modellparametern stärker bestraft als das *AIC*-Kriterium. Dies ist hier sinnvoll, da die Modelle basierend auf nur 19 Versuchspunkten angepasst werden. Die Ergebnisse bezüglich dieser Maße sind in den Tabellen A.4, A.5 und A.6 aufgelistet. Für die Zielgrößen D_{max} und t existiert ein Gauß-Prozess-Modell, bei dem sowohl der *MSCVE* als auch der Fit optimal sind. Beide Modelle verwenden die anisotrope Kovarianzfunktion mit $s_i = 2$ für $i = 1, 2, 3$. Während für die Zielgröße D_{max} die Trendkomponente auf einen Intercept reduziert ist, wird für die Zielgröße Produktionszeit t ein linearer Trend verwendet. Die geschätzten Modellparameter für diese beiden Modelle sind in den Tabellen 7.3 und 7.4 gegeben.

Modellkomponente	Parameter	Schätzer
Deterministische Trendkomponente	Intercept	74.78
Anisotrope Power-Exponential Korrelationsfunktion ($s_i = 2$) für $i = 1, 2, 3$	θ_e	0.342
	θ_v	0.718
	θ_f	0.256
	σ_Z^2	0.148

Tabelle 7.3: Gauß-Prozess-Modell für den maximalen Durchmesser D_{max}

Modellkomponente	Parameter	Schätzer
Deterministische Trendkomponente	Intercept	182.98
	e_{AQU}	-28.589
	V_{AQU}	9.559
	f	-198.440
Anisotrope Power-Exponential Korrelationsfunktion ($s_i = 2$) für $i = 1, 2, 3$	θ_e	2.933
	θ_v	2.628
	θ_f	0.188
	σ_Z^2	229.200

Tabelle 7.4: Gauß-Prozess-Modell für die Produktionszeit t

Bevor diese Modelle verwendet werden, ist deren Modellgültigkeit zu überprüfen. Dazu wurden die in Kapitel 5.7 vorgestellten diagnostischen Plots verwendet, vgl. Abbildungen A.3 und A.4. Der Plot der Kreuzvalidierungsvorhersagen gegen die Beobachtungen zeigt für die Zielgröße D_{max} kein auffälliges Verhalten. Die Punkte streuen im akzeptablen Rahmen um die Diagonale, wobei ein Ausreißer existiert. Für die Zielgröße t wird eine sehr gute Anpassung beobachtet. Diese sehr gute Anpassung kommt nicht durch ein Overfitting der Daten zustande. Wird z. B. zufällig ein Drittel der Beobachtungen gelöscht, ist die Anpassung immer noch vergleichbar gut. Der Grund für die sehr gute Anpassung ist, dass der Zusammenhang zwischen der Fertigungszeit und den drei Einflussfaktoren sehr einfach ist. Wahrscheinlich hätte für diese Zielgröße auch ein klassisches quadra-

tisches Regressionsmodell gereicht. Für die sequentielle Modellverfeinerung sind aber Gauß-Prozess-Modelle zu verwenden. Falls die beiden Modelle gültig sind, sollten zudem die standardisierten Beobachtungen aus Formel (5.7) im Intervall $[-3, 3]$ liegen. Da auch diese Forderung erfüllt ist, werden diese Modelle im Folgenden zur Prognose der beiden Zielgrößen verwendet. Die Oberflächenplots für einen Schnitt durch die e_{AQU} Ebene sind in den Abbildungen 7.8 a) und b) dargestellt.

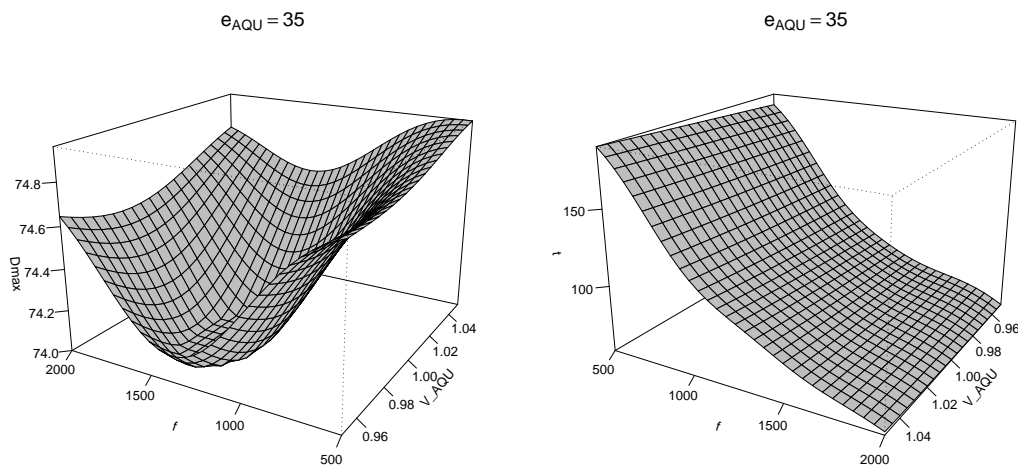


Abbildung 7.8: Oberflächenplot des Gauß-Prozess-Modells für a) D_{max} , b) t

Der Vorschub f besitzt den größten Einfluss auf die Produktionszeit. Zum einen ist der lineare Parameter des deterministischen Trends deutlich größer als für die anderen beiden Parameter. Zum anderen ist aber auch der Korrelationsparameter für f deutlich kleiner. Je größer der Korrelationsparameter ist, desto glatter ist die Modelloberfläche. Auch für den maximalen Durchmesser scheint der Vorschub den größten Einfluss zu besitzen.

Für die Zielgröße NT ist das beste Modell je nachdem, ob die Vorhersagegüte oder der Fit betrachtet wird, unterschiedlich, vgl. Tabelle A.6. Aus diesem Grund ist bei der Wahl eines geeigneten Modells ein guter Kompromiss zwischen Vorhersagegüte und Fit zu finden. Unter den zehn Gauß-Prozess-Modellen sind drei Modelle, die sowohl eine gute Vorhersagegüte als auch einen guten Fit aufweisen. Dies sind die beiden Modelle mit isotroper Kovarianzfunktion mit linearem Trend und linearem Trend mit 2-fach-Wechselwirkungen. Das dritte Modell ver-

wendet eine anisotrope Kovarianzfunktion mit ebenfalls linearem Trend. Für alle drei Modelle sind die Parameter $s_i = 1$ für alle $i = 1, 2, 3$. Um zu entscheiden, welches dieser drei Modelle schließlich zur Prognose verwendet wird, werden die diagnostischen Plots herangezogen, vgl. Abbildung A.5. Beim Vergleich der Plots, die die Kreuzvalidierungsvorhersagen gegen die Beobachtungen abtragen, scheint das Modell mit anisotroper Kovarianzfunktion leicht besser als die anderen Modelle zu sein. Während für das Modell mit isotroper Kovarianzfunktion und linearem Trend eine leichte Krümmung in den Punkten existiert, besitzt das Modell, das zusätzlich die 2-fach-Wechselwirkungen im Trend berücksichtigt, eine größere Streuung als die anderen beiden Modelle. Auch beim Vergleich der Plots der mittels Kreuzvalidierung standardisierten Beobachtungen schneidet das anisotrope Gauß-Prozess-Modell am besten ab. Insbesondere beim Modell mit isotroper Kovarianzfunktion und linearem Trend fallen zwei Ausreißer auf, die außerhalb des Intervalls $[-3, 3]$ liegen. Somit wird das Gauß-Prozess-Modell mit anisotroper Kovarianzfunktion mit $s_i = 1$ für $i = 1, 2, 3$ und linearem Trend zur Prognose der Zielgröße NT verwendet. Die geschätzten Modellparameter sind in Tabelle 7.5 aufgelistet. Der Faktor e_{AQU} besitzt nur einen linearen Einfluss auf den deterministischen Trend. Wie für die anderen beiden Zielgrößen scheint der Faktor f wieder den größten Einfluss zu besitzen. Der Oberflächenplot für $e_{AQU} = 35$ ist in Abbildung 7.9 gegeben.

Modellkomponente	Parameter	Schätzer
Deterministische Trendkomponente	Intercept	84.54
	e_{AQU}	15.22
	V_{AQU}	-8.40
	f	-12.04
Anisotrope Exponential Korrelationsfunktion ($s_i = 1$) für $i = 1, 2, 3$	θ_e	10.010
	θ_v	0.667
	θ_f	0.026
	σ_Z^2	0.747

Tabelle 7.5: Gauß-Prozess-Modell für die Napftiefe NT

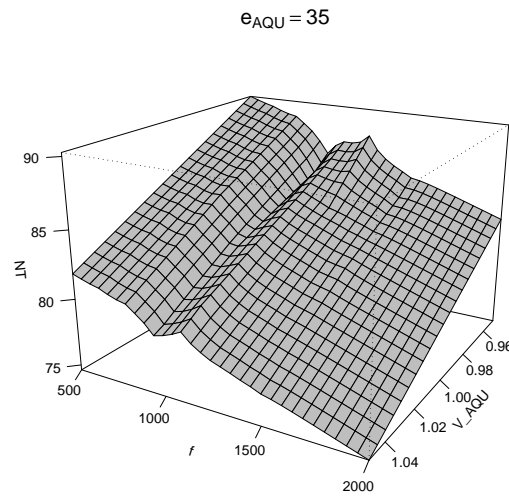


Abbildung 7.9: Oberflächenplot des Gauß-Prozess-Modells für die Napftiefe NT

Abschließend ist der Zusammenhang zwischen den Maschinenparametern und dem Auftreten von Bauteilversagen zu modellieren. Da die Zielgröße Bauteilversagen binär ist, wurde dazu ein logistisches Regressionsmodell verwendet. Während zur Modellierung der Zielgrößen NT , D_{max} und t nur die Versuchspunkte im stabilen Bereich berücksichtigt wurden, basiert das logistische Regressionsmodell auf dem kompletten Versuchsumfang von $N=35$ Bauteilen. In Tabelle A.7 sind die geschätzten Modellparameter des reduzierten Modells angegeben. Wie beim Hyper-Modell der logistischen Regression, das im ersten Optimierungsschritt zur Vorhersage des stabilen Bereichs für $NT = 81$ mm verwendet wurde, besitzt auch in diesem Modell der Parameter V_{AQU} keinen Einfluss auf das Auftreten von Bauteilversagen. Dies ermöglicht den vorhergesagten stabilen Bereich im zweidimensionalen Raum von e_{AQU} und f darzustellen, vgl. Abbildung 7.10. Anscheinend kann der stabile Bereich als Schlauch im dreidimensionalen Versuchsraum betrachtet werden, der parallel zur V_{AQU} -Ebene liegt. Mit diesem logistischen Regressionsmodell lässt sich nun im Nachhinein die Vorhersagegüte des Hyper-Proportional-Odds-Modells beurteilen. Ein Vergleich der beiden Vorhersagen des stabilen Bereichs aus Abbildung 7.3 und 7.10 ergibt, dass die Formen des vorhergesagten stabilen Bereichs sich zwar unterscheiden, die Lage des stabilen Bereichs aber gut vorhergesagt werden konnte.

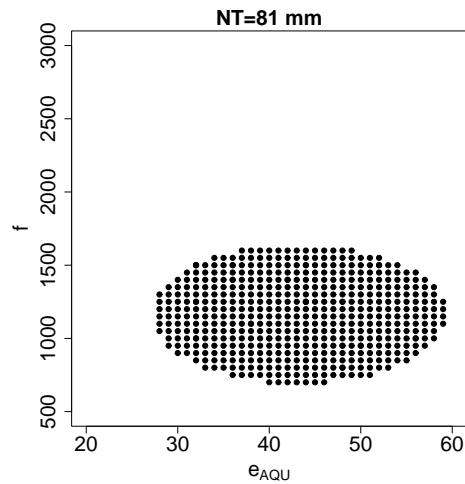


Abbildung 7.10: Vorhersage des stabilen Bereichs der logistischen Regression

7.4 Sequentielle Modellverfeinerung

Bevor die sequentielle Prozessoptimierung mit Hilfe der multivariaten Expected Improvement erfolgen kann, sind zunächst die Qualitätsanforderungen der Zielgrößen durch Wünschbarkeitsfunktionen zu spezifizieren. Dazu werden Wünschbarkeitsfunktionen von Harrington verwendet, die in Zusammenarbeit mit Drückexperten definiert wurden. Da für die Zielgrößen D_{max} und NT Werte kleiner gleich dem Sollwert von 71 mm bzw. 81 mm nicht realisiert werden können, wurden einseitige Wünschbarkeitsfunktionen verwendet. Zu beachten ist, dass der optimale Wert für den maximalen Durchmesser D_{max} von 71 mm aufgrund des Unrundlaufs des Drückfutters nicht erreicht werden kann. Messungen haben ergeben, dass der Unrundlauf dieses Futters den maximalen Durchmesser eines jeden Bauteils um 3 mm erhöht. Aus diesem Grund ergibt sich für den zu fertigenden Napf ein Sollwert für D_{max} von 74 mm. Die Parameter, die die einseitige Wünschbarkeitsfunktion nach Harrington spezifizieren, sind für die drei Zielgrößen in Tabelle 7.6 gegeben. Die so definierten Wünschbarkeitsfunktionen sind in Abbildung 7.11 eingezeichnet.

	D_{max}	NT	t
b_0	458.09	65.7	7.64
b_1	64.7	-0.771	-0.077

Tabelle 7.6: Parameter der einseitigen Wünschbarkeitsfunktion von Harrington

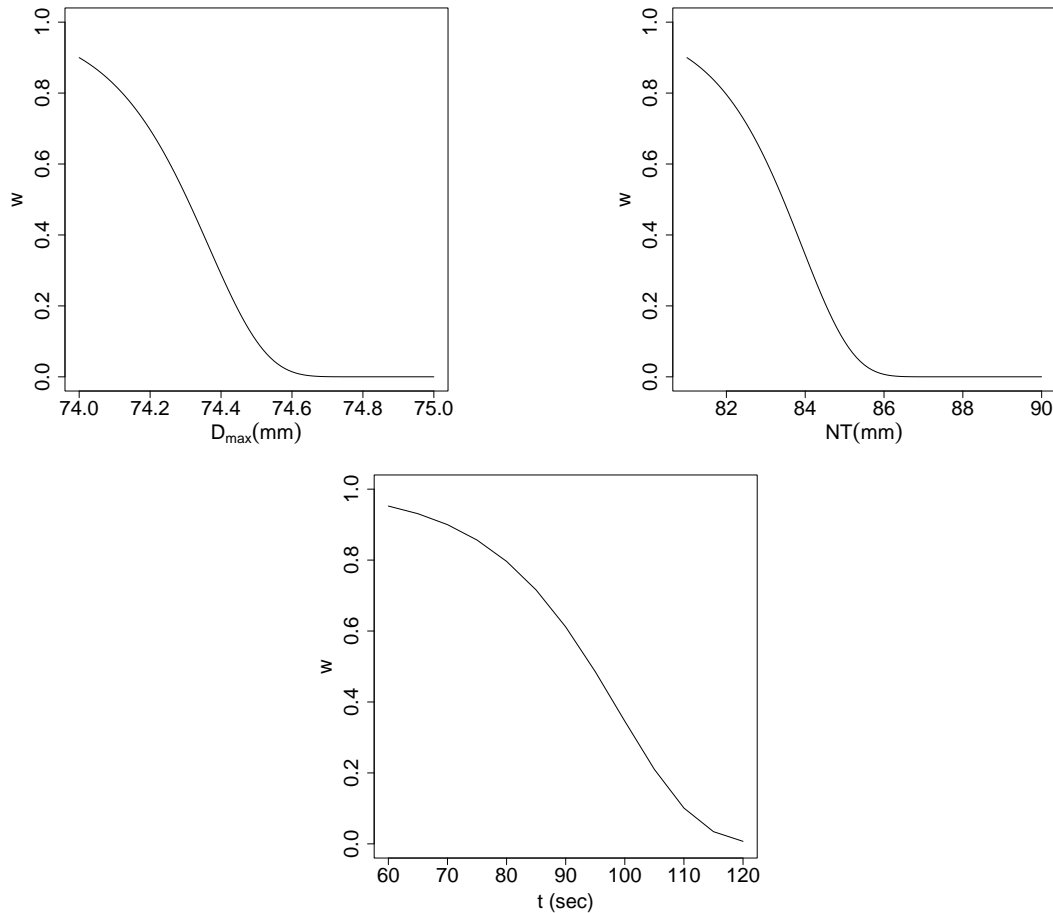


Abbildung 7.11: Einseitige Wünschbarkeitsfunktionen nach Harrington

Neben der Spezifikation der Wünschbarkeitsfunktionen ist zur Berechnung des Wünschbarkeitsindex die Gewichtung der Zielgrößen zu bestimmen. Dazu wurde eine qualitative Skala von 1 bis 10 verwendet, wobei Zielgrößen, die in die Kategorie 10 eingeordnet werden, als sehr wichtig einzustufen sind. Die vom Drückexperten gewählte Gewichtung ist $D_{max} = 9$, $NT = 9$ und $t = 7$.

Mit Hilfe der umfangreichen Falldatenbank des Drückprozesses können gute

Schätzer für die Korrelation der Zielgrößen gewonnen werden. Die Korrelationschätzer der Zielgrößen sind in Tabelle 7.7 aufgelistet. Während die beiden Zielgrößen Produktionszeit t und Napftiefe NT unkorreliert sind, ist die Zielgröße D_{max} sowohl mit NT als auch mit t ungefähr im gleichen Maße positiv korreliert.

	D_{max}	NT	t
D_{max}	1	0.589	0.657
NT	0.589	1	<0.001
t	0.657	<0.001	1

Tabelle 7.7: Geschätzte Korrelation der Zielgrößen

Die sequentielle Modellverfeinerung erfolgt nun mit dem multivariaten Expected-Improvement-Kriterium, das in Kapitel 6.4.2 vorgestellt wurde. Dabei wurde der Parameter g , der die Gewichtung zwischen lokaler und globaler Suche steuert auf 3 gesetzt. Um die wichtige binäre Zielgröße Versagen zu berücksichtigen, könnte die Expected Improvement an der Stelle \mathbf{x} mit der Wahrscheinlichkeit $P_B(\mathbf{x})$ für ein Bauteilversagen multipliziert werden. Damit wird aber verhindert das neue Versuchspunkte am Rand des stabilen Bereichs beobachtet werden. Da Bauteile mit guter Gesamtqualität aber an der Grenze zum Bauteilversagen liegen, ist dieser Ansatz somit nicht sinnvoll. Um den abrupten Übergang zum Bauteilversagen zu simulieren, wird die Expected Improvement mit der Indikatorfunktion $I_{(P_B(\mathbf{x}) < c)}$ multipliziert. Ist die Wahrscheinlichkeit $P_B(\mathbf{x})$ größer als c wird die Expected Improvement Null. Für $P_B(\mathbf{x}) < c$ bleibt die Expected Improvement unverändert. Für dieses Optimierungsbeispiel wurde der Grenzwert c auf 0.6 gesetzt.

Um die Ergebnisse der nun folgenden sequentiellen Optimierung besser einschätzen zu können, sollte berücksichtigt werden, dass das beste Bauteil aus dem raumfüllenden Versuchsplan (Lauf 24) schon eine sehr gute Gesamtqualität besitzt. Der maximale Durchmesser weicht mit 74.1 mm nur 0.1 mm von Sollwert ab. Bei der Bauteilhöhe wurde mit 81.5 eine Abweichung von 0.5 mm beobachtet. Da auch die Produktionszeit mit 79 Sekunden relativ kurz war, ergibt sich für dieses Bauteil ein Wünschbarkeitsindex von 0.83. Selbst bei exakten Bauteilgeo-

metrie und einer sehr schnellen Produktionszeit von 60 Sekunden kann durch die oben definierten Wünschbarkeitsfunktionen ein Wünschbarkeitsindex von maximal 0.88 beobachtet werden. Somit ist das Optimierungspotenzial begrenzt.

Insgesamt wurden fünf Versuchspunkte sequentiell hinzugefügt. Die Einstellungen der Maschinenparameter dieser Versuche und die beobachteten Zielgrößen sowie der Wünschbarkeitsindex sind in Tabelle 7.8 gegeben. Im vierten Iterationsschritt trat ein Bauteilversagen auf. Im darauf folgenden Iterationsschritt wurde diese Information berücksichtigt, indem der nächste Versuchspunkt etwas mehr zur Mitte des stabilen Bereichs platziert wurde. In diesem fünften Versuchspunkt konnte ein Bauteil mit nahezu optimaler Geometrie produziert werden. Der Durchmesser entsprach dem Zielwert von 81 mm und die Napftiefe wich nur um 0.1 mm von der Spezifikation ab. Zudem wurde zur Herstellung dieses Bauteils eine relativ kurze Produktionszeit benötigt. Die Optimierung wurde hier abgebrochen, da der zur Verfügung stehende Gesamtversuchsumfang von $N = 40$ erreicht wurde. Offensichtlich konnte mit Hilfe der sequentiellen Modellverfeinerung die Gesamtqualität des besten Bauteils noch um 2 von maximal 5 Prozentpunkten gesteigert werden. Zusammenfassend lässt sich sagen, dass mit Hilfe der ASOP erstmalig eine erfolgreiche und effiziente Prozessoptimierung vorgenommen werden konnte.

Iterations- schritt	e_{AQU}	V_{AQU}	f	Bauteil- versagen	t	D_{max}	NT	WI
1	30.5	0.968	1400	nein	83	74	82	0.82
2	38.0	0.980	1315	nein	83	74.1	84.2	0.53
3	33.5	0.968	1360	nein	84	74.3	83.6	0.54
4	25.2	0.958	1560	ja
5	26.5	0.950	1520	nein	79	74.1	81	0.85

Tabelle 7.8: Ergebnisse der sequentiellen Modellverfeinerung

Kapitel 8

Zusammenfassung

Das Ziel der Arbeit war die Entwicklung einer effizienten Optimierungsprozedur für den CNC-gesteuerten Drückprozess. Dazu wurde ein problemorientierter Ansatz gewählt. Basierend auf den Eigenschaften des Drückprozesses erfolgte zunächst die Bestimmung der wesentlichen Anforderungen an die Optimierungsprozedur. Unter Berücksichtigung dieser Anforderungen wurde anschließend eine adaptive sequentielle Optimierungsprozedur (ASOP) entwickelt, die sich im wesentlichen in vier Bearbeitungsschritte untergliedern lässt. Die dazu verwendeten Methoden und Modelle wurden aufeinander abgestimmt.

Im ersten Schritt der ASOP wird mit Hilfe des fallbasierten Schließens das Vorwissen über den Prozess in die Optimierung integriert. Dazu wird auf eine Falldatenbank zugegriffen, in der Informationen über gefertigte Bauteile gespeichert sind. Mit Hilfe des fallbasierten Schließens können dann Bauteile aus dieser Datenbank extrahiert werden, deren Geometrie möglichst ähnlich zu der neuen Fertigungsaufgabe ist. Die gespeicherten Informationen dieser Bauteile werden anschließend auf das neue Problem übertragen. Bei der Adaption des fallbasierten Schließens auf den Drückprozess war die Bestimmung geeigneter Ähnlichkeitsmaße und die Systematisierung der Bauteilgeometrien von wesentlicher Bedeutung. Um ein Maximum an Informationen aus der Falldatenbank extrahieren zu können, wurden zudem Hyper Modelle eingesetzt. Da beim Drückprozess oft Bauteilreihen gefertigt werden, ermöglichen es diese Modelle, das Prozessverhalten, wie z. B die Lage und Form des stabilen Bereichs, zu interpolieren.

Im zweiten Schritt der ASOP wird ein Versuchsplan im stabilen Bereich generiert. Da die Struktur der Zusammenhänge unbekannt und komplex ist, sollte die Generierung des Versuchsplans nicht auf Modellinformationen beruhen. Stattdessen ist zu empfehlen, die Versuchspunkte möglichst gleichmäßig im Versuchsraum zu verteilen. In dieser Arbeit wurde ein ausführlicher Überblick über mögliche Kriterien zur Generierung solcher raumfüllender Versuchspläne gegeben. Ein Vergleich dieser Versuchspläne ergab, dass während bei kubischen Versuchsräumen Maximin Latin Hypercube Designs zu empfehlen sind, bei nicht kubischen Versuchsräumen der Einsatz von Minimax Designs zu bevorzugen ist. Der Drückprozess stellt aber nun spezielle Anforderungen an die raumfüllenden Versuchspläne, da die Grenzen zum Versagensbereich unbekannt sind. Die klassischen raumfüllenden Versuchspläne erweisen sich in dieser Situation als sehr ineffizient, da abhängig von der Größe des stabilen Bereichs mit einer mehr oder minder großen Anzahl an fehlenden Beobachtungen zu rechnen ist. Motiviert durch diese Problematik wurde ein Algorithmus zur sequentiellen Generierung von raumfüllenden Versuchsplänen entwickelt. Dieser Algorithmus beruht auf der Annahme, dass der stabile Bereich konvex ist. Durch diese Annahme können anhand der Lage der Versuchspunkte zueinander komplette Bereiche des Versagensraums identifiziert und damit ausgeschlossen werden. Mit der sequentiellen Generierung des Versuchsplans wird somit verhindert, dass neue Versuchspunkte in die schon identifizierten Versagensbereiche platziert werden. Da der Algorithmus sequentiell vorgeht, wurde eine erweiterte Version des Minimax-Kriteriums in den Algorithmus implementiert. Dieses Kriterium garantiert eine gute raumfüllende Eigenschaft auch auf den aktiven Unterräumen des Parameterraums. Anhand einer ausführlichen Simulationsstudie konnte gezeigt werden, dass dieser Algorithmus eine deutliche Reduzierung der Anzahl der Versagensfälle im Vergleich zum einfachen raumfüllenden Versuchsplan ermöglicht. Dabei hängt die Effizienz des Algorithmus bezogen auf die Reduktion der Versagensfälle von der Lage, Form und Größe des stabilen Bereichs im Parameterraum ab. Bei festem Gesamtversuchsumfang N steigt die Effizienz des Verfahrens je kleiner der Anteil A des stabilen Bereichs am Parameterraum ist und je weniger die Versagensgrenze gekrümmt ist. Eine maximale Reduzierung der Anzahl der Versagensfälle erfolgt bei einem kleinen stabilen Bereich, der am Rand des Parameterraums liegt und dessen Grenze zum Versagensbereich linear

verläuft. Im zweidimensionalen Versuchsraum konnte in diesen Fällen der Anteil der Versuchspunkte im stabilen Bereich im Schnitt verdreifacht werden. Zudem konnte mit Hilfe der Simulationsstudie ein günstiges Kosten-Nutzen-Verhältnis für die Anzahl der Iterationsschritte bestimmt werden. Dieses scheint bei niedriger Dimension der Daten bei $N/5$ zu liegen.

Basierend auf diesem Versuchsplan erfolgt im dritten Schritt der ASOP die Modellierung der Zielgrößen. Dazu werden Gauß-Prozess-Modelle verwendet, da sie es ermöglichen, komplexe Zusammenhänge mit relativ geringem Versuchsaufwand zu approximieren. Eine detaillierte Beschreibung dieser Modelle wurde in dieser Arbeit gegeben. Da der Anwender bei der Spezifikation des Gauß-Prozess-Modells flexibel ist, sollte im ersten Schritt eine Modellwahl erfolgen, die sowohl die Vorhersagegüte als auch den Fit des Modells berücksichtigt. Dazu wurde die Vorhersagegüte mit Hilfe des mittleren Kreuzvalidierungsfehlers und der Fit der Modelle mit Hilfe des klassischen BIC-Kriterium bewertet. Bevor das nach diesen Kriterien ausgewählte Modell zur Vorhersage verwendet wird, sollte die Modellgültigkeit durch diagnostische Plots überprüft werden.

Im vierten Schritt der ASOP wird zur Bestimmung des globalen Optimums eine sequentielle Modellverfeinerung durchgeführt. Dazu wurde ein bestehender univariater Ansatz von Jones et al. (1998) basierend auf dem Expected-Improvement-Kriterium auf den multivariaten Fall verallgemeinert. Im Gegensatz zu anderen sequentiellen Optimierungsansätzen berücksichtigt dieser univariate Ansatz bei der Wahl neuer Versuchspunkte nicht nur die Vorhersage des Modells, sondern auch dessen Unsicherheit. Die Gewichtung zwischen der lokalen und globalen Suche kann dabei vom Anwender bestimmt werden. Durch dieses kombinierte Vorgehen wird eine effiziente Identifizierung des globalen Optimums ermöglicht. In dieser Arbeit wurden zwei verschiedene Ansätze vorgestellt, die es ermöglichen, die Idee der Expected Improvement auf den multivariaten Fall zu übertragen. Dazu wird das Konzept der Wünschbarkeiten verwendet. Der erste beschriebene Ansatz zur sequentiellen Modellverfeinerung reduziert das multivariate Problem auf ein univariates Problem, indem die Expected Improvement des Wünschbarkeitsindex betrachtet wird. Nachteil dieses Ansatzes ist jedoch, dass die zur Berechnung der Expected Improvement benötigte Verteilung des Wünschbarkeits-

indexes nicht bekannt ist. Um dieses Problem zu lösen, könnten Monte-Carlo-Simulationen zur Generierung der Verteilungsfunktion eingesetzt werden. Dieses Vorgehen ist jedoch aufgrund des hohen Zeitaufwands nicht praktikabel. Selbst bei direkter Approximation der Expected Improvement mit Hilfe des Gesetzes der großen Zahlen ist der Rechenaufwand zu groß.

Aus diesem Grund wurden ein alternativer Ansatz vorgestellt, bei dem die Expected Improvement über den Wünschbarkeitsvektor definiert ist. Zur Berechnung dieses Expected-Improvement-Kriteriums wird die Dichte des Wünschbarkeitsvektors benötigt. Für unkorrelierte Zielgrößen ergibt sich diese Dichte aus dem Produkt der Dichten der Einzelwünschbarkeiten. Zur Herleitung dieser Dichten wurde auf die Arbeiten von Trautmann und Weihs (2006) sowie Steuer (2005) zurückgegriffen. Für korrelierte Zielgrößen konnte diese bisher nur für den Spezialfall einseitiger Wünschbarkeitsfunktionen nach Harrington hergeleitet werden.

Zu Beginn der Forschungsarbeiten wurde eine erste Version der ASOP mit der klassischen Response Surface Methodology verglichen, vgl. Göbel et al. (2004). Obwohl die erste Version der ASOP bei Weitem nicht so ausgereift war wie die jetzige Endversion, sprechen die Optimierungsergebnisse eindeutig für die ASOP. Bei gleichem Gesamtversuchsumfang von $N = 55$ konnte neben der Reduktion der Produktionszeit um 30% auch die technologische Bauteilqualität gesteigert werden. Um die in dieser Arbeit beschriebene Endversion der ASOP zu verifizieren, wurde diese Prozedur zur Optimierung einer neuen Bauteilgeometrie eingesetzt. Das so optimierte Bauteil besitzt eine sehr gute Gesamtqualität und lässt sich in kurzer Fertigungszeit herstellen. Somit konnte eine erfolgreiche Optimierung des Prozesses basierend auf nur 40 Versuchspunkten durchgeführt werden. Zudem ergeben sich durch die ASOP allgemeingültige Vorteile. Da es mit Hilfe des implementierten Algorithmus nun möglich ist, den Verlauf der Versagensgrenzen auf effiziente Weise zu bestimmen, konnten nun erstmalig genauere Untersuchungen zur Veränderung des stabilen Bereichs durchgeführt werden, vgl. Ewers (2005). Dabei zeigte sich, dass der stabile Bereich mit ansteigender Komplexität des Bauteils kleiner wird und gegen einen so genannten stabilen Punkt konvergiert. Durch dieses Wissen konnte der Schwierigkeitsgrad bei der Bauteilproduktion deutlich gesteigert werden. Beispielsweise ließ sich auf diese Weise

das in der Literatur bekannte maximale Druckverhältnis von Durchmesser zu Höhe eines Bauteils um 65 % steigern. Auch die Interpolation des Prozesswissens mittels Hyper-Modellen führte zu einem tieferen Prozessverständnis. Insgesamt lässt sich sagen, dass mit der ASOP erstmalig eine effiziente Optimierungsprozedur für den CNC-gesteuerten Druckprozess bereitgestellt werden konnte, vgl. Henkenjohann et al. (2005) und Göbel et al. (2005).

Für die Herstellung rotationssymmetrischer Bauteile war die extrem aufwändige Optimierung des Druckprozesses bisher der wesentliche Grund, warum andere Verfahren wie das Tiefziehen in der Industrie bevorzugt wurden. Die Existenz der ASOP könnte nun dazu beitragen, dass sich der CNC-gesteuerte Druckprozess als Alternative zum Tiefziehen etabliert, da sich das Drücken durch eine hohe Form- und Maßgenauigkeit auszeichnet und der Prozess komplexere Geometrieformen herzustellen kann als dies mit dem Tiefziehen möglich ist.

Kapitel 9

Ausblick

Um die ASOP auch auf Daten höherer Dimension anwenden zu können, sind die implementierten R-Programme zur Berechnung der konvexen Hüllen zunächst in Bezug auf die Rechenzeit zu optimieren. Anschließend könnten die Simulationsergebnisse zur Effizienz des Algorithmus auch auf mehr als drei Dimensionen ausgedehnt werden.

Im Rahmen des SFB-Projekts wurden bisher relativ einfache Geometrieformen untersucht. Durch die Erkenntnisse, die durch diese Untersuchungen gewonnen werden konnten, ist es aber nun möglich auf die Produktion komplexer Geometrieformen überzugehen. Dies erfordert eine konsequente Weiterentwicklung der ASOP.

Für das fallbasierte Schließen ist eine Erweiterung der Ähnlichkeitsmaße auf komplexe Geometrieformen vorzunehmen. Zudem ist der Ansatz der Hypermodellierung zu erweitern. Dabei sind geeignete Modellansätze zu finden, mit deren Hilfe die Veränderung verschiedener wesentliche Aspekte des Prozessverhaltens approximiert werden kann.

Mit steigender Komplexität der Geometrieformen könnte auch die Komplexität der Form der Versagensgrenzen ansteigen. Somit ist der Algorithmus zur sequentiellen Generierung raumfüllender Versuchspläne anzupassen. Da die Entwicklung bzw. Weiterentwicklung des ASOP aber problemorientiert erfolgt, sind zunächst ausführliche Untersuchungen zum Verlauf der Versagensgrenzen bei zusammen-

gesetzten Bauteilen durchzuführen. Dazu könnte die Entwicklung von Versuchsplänen hilfreich sein, die speziell den Verlauf der Versagensgrenze untersuchen.

Die Analyse des Drückprozesses hat gezeigt, dass Bauteile mit guter Gesamtqualität durchgehend an der Grenze zum Versagen zu finden sind. Das Ziel ist es somit, die Produktion von Bauteilen so nah wie möglich an diese physikalischen Grenzen zu bringen. Um dies zu erreichen, sollte der Aspekt der Robustheit gegen unkontrollierbare Faktoren berücksichtigt werden. Erste Ansätze zur Berechnung der Expected Improvement unter Berücksichtigung von Störgrößen existieren bereits, vgl. Lehmann (2004) und Williams et al. (2000).

Aufgrund der guten Ergebnisse und der Effizienz der ASOP ist dieses Verfahren auf ein großes Interesse seitens der Industrie und Forschung gestoßen. Daher findet seit Juli 2006 ein Transferprojekt im Rahmen des SFB 475 statt, bei dem eine Adaption der Methodik auf das Einziehen durchgeführt wird. Mit dem Einziehen werden konturierte rohrförmige Bauteile wie Achsen und Druckgasbehälter hergestellt. Konkrete industrierelevante Einziehprobleme im Transferprojekt werden durch die Kooperationspartner VW und Benteler bereitgestellt. Im Rahmen der Adaption der ASOP auf das Einziehen sind z. B. die Ähnlichkeitsmaße, die im Rahmen des fallbasierten Schließens eingesetzt werden, auf das Einziehen anzupassen. Außerdem sind wesentliche Annahmen wie die Konvexität des stabilen Bereichs zu überprüfen. Leichte Abweichung von der Konvexitätsannahme sind jedoch tolerabel, wie die Untersuchungen in dieser Arbeit gezeigt haben.

Zudem startet im Januar 2007 der SFB 708 „3D-Surface Engineering für Werkzeugsysteme der Blechformteilefertigung – Erzeugung, Modellierung, Bearbeitung“. Das Teilprojekt C3 „Strategien zur Kompensation rückfederungsbedingter Formabweichungen“ beschäftigt sich mit der Entwicklung und Erforschung einer auf statistischen Ansätzen beruhenden Methode zur Auslegung der Aktivflächen von Umformwerkzeugen. Das Ziel des Projekts ist die Entwicklung einer Methode, die rückfederungsbedingte Formabweichungen kompensiert und damit die Herstellung von komplexen freigeformten Blechformteilen mit der geforderten Geometrie bei minimalem Optimierungsaufwand ermöglicht. Als ein Hilfsmittel dazu soll die ASOP verwendet werden.

Anhang A

Appendix

A.1 Bedingte multivariate Normalverteilung

$$\text{Sei } \begin{pmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \end{pmatrix} \sim N_{m+n} \left[\begin{pmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{pmatrix}, \begin{pmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{pmatrix} \right],$$

wobei $\boldsymbol{\mu}_1$ $m \times 1$, $\boldsymbol{\mu}_2$ $n \times 1$, $\boldsymbol{\Sigma}_{11}$ $m \times m$, $\boldsymbol{\Sigma}_{12} = \boldsymbol{\Sigma}'_{21}$ $m \times n$ und $\boldsymbol{\Sigma}_{22}$ $n \times n$. Die bedingte Verteilung von $\mathbf{X}_1|\mathbf{X}_2$ ist dann gegeben durch (vgl. Santner et al. 2003)

$$N_m \sim \left(\boldsymbol{\mu}_1 + \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}(\mathbf{X}_2 - \boldsymbol{\mu}_2), \boldsymbol{\Sigma}_{11} - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21} \right).$$

A.2 Tabellen

Nummer	e_{AQU}	V_{AQU}	f	Versagen
1	46.56	0.989	1525	ja
2	43.75	1.029	1975	ja
3	27.81	0.973	2350	ja
4	45.62	0.958	1525	ja
5	44.69	1.029	1225	ja
6	36.25	0.995	1450	ja
7	47.00	1.029	1200	nein
8	47.00	1.029	900	nein
9	47.00	1.029	500	nein
10	43.00	1.029	700	nein
11	43.00	1.035	700	nein
12	43.00	1.010	700	nein
13	38.00	1.036	1250	nein
14	48.50	1.008	700	ja
15	38.00	1.004	1200	nein
16	38.00	1.022	1200	nein
17	46.25	1.008	1100	nein
18	36.50	1.030	700	nein
19	37.25	1.010	700	nein
20	45.50	1.032	1450	ja
21	39.50	1.010	1450	ja
22	47.00	1.010	1400	ja

Tabelle A.1: Versuchspunkte aus der Falldatenbank für $NT = 86$ mm

Nummer	e_{AQU}	V_{AQU}	f	Versagen
1	35	0.965	1750	nein
2	29	1.040	1900	ja
3	23	1.033	2200	ja
4	38	1.033	1450	nein
5	23	0.958	1900	ja
6	26	1.033	1300	ja
7	56	1.033	2650	ja
8	23	0.987	2800	ja
9	53	0.965	2650	ja
10	41	0.965	1300	nein
11	56	0.995	1900	nein
12	50	0.958	2050	ja
13	53	1.003	1150	nein
14	41	0.973	2800	ja
15	53	1.025	1900	nein
16	56	0.965	1300	nein
17	32	0.958	2500	nein
18	35	1.04	2650	ja
19	23	1.003	1450	ja
20	53	0.995	2650	ja
21	38	1.025	1000	nein
22	44	1.025	2800	ja
23	53	1.040	1300	nein
24	38	0.995	1300	nein
25	26	0.995	2200	ja
26	26	0.965	1150	nein
27	50	1.040	2050	nein
28	41	1.003	2050	nein
29	23	1.025	2800	ja
30	32	1.003	2650	ja

Tabelle A.2: Versuchspunkte aus der Falldatenbank für $NT = 69$ mm

Versuch	e_{AQU}	V_{AQU}	f	Versagen	t	D_{max}	NT
1	38	1.040	1100	nein	104	74.5	81.8
2	30	0.965	650	ja			
3	28	1.030	800	ja			
4	48	1.015	650	nein	156	75.2	89.3
5	42	1.005	1400	nein	78	74.1	83.9
6	50	1.045	950	nein	110	74.6	84.9
7	48	0.955	950	ja			
8	48	0.980	800	ja			
9	36	0.965	1400	nein	79	74.0	85.0
10	30	1.000	800	nein	146	74.6	83.3
11	32	0.990	1250	nein	93	74.0	82.2
12	48	1.040	1250	nein	86	74.2	84.0
13	44	0.980	1250	nein	83	74.4	87.4
14	32	1.030	1500	ja			
15	44	0.950	1250	nein	81	74.7	91.6
16	44	1.040	1750	ja			
17	48	0.970	1750	ja			
18	36	0.990	1750	ja			
19	36	0.950	1500	nein	74	74.2	85.6
20	44	1.010	1500	nein	72	74.2	83.9
21	40	1.020	1750	ja			
22	28	1.000	1500	ja			
23	52	1.000	1500	nein	67	74.3	87.9
24	28	0.960	1500	nein	79	74.1	81.5
25	32	1.050	1500	ja			
26	52	1.030	1750	ja			
27	56	0.960	1500	ja			
28	56	1.030	1500	nein	67	74.3	86.7
29	24	0.970	1250	ja			
30	24	1.030	1250	ja			
31	56	0.990	1000	ja			
32	52	1.010	1000	nein	99	74.7	88.8
33	56	1.050	1500	nein	69	74.5	84.9
34	40	1.010	1000	nein	109	74.4	84.3
35	32	0.960	1000	nein	113	74.3	84.0

Tabelle A.3: Versuchsplan und Ergebnisse für $NT = 81$ mm

Trendkomponente	Power Exponential Korrelationsfunktion	# Modell- parameter	MSCVE	BIC
Intercept	Isotrop (p=1)	3	0.0364	10.43
Linear	Isotrop (p=1)	6	0.0354	5.67
Linear+WW	Isotrop (p=1)	9	0.0644	6.09
Intercept	Isotrop (p=2)	3	0.0943	11.03
Linear	Isotrop (p=2)	6	0.0710	5.66
Linear+WW	Isotrop (p=2)	9	0.1046	8.40
Intercept	Anisotrop (p=1)	5	0.0558	12.52
Linear	Anisotrop (p=1)	8	0.0388	7.66
Intercept	Anisotrop (p=2)	5	0.0212	5.32
Linear	Anisotrop (p=2)	8	0.0264	9.75

Tabelle A.4: Gauß-Prozess-Modelle für den maximalen Durchmesser D_{max}

Trendkomponente	Power Exponential Korrelationsfunktion	# Modell- parameter	MSCVE	BIC
Intercept	Isotrop (p=1)	3	169.53	164.83
Linear	Isotrop (p=1)	6	82.60	118.36
Linear+WW	Isotrop (p=1)	9	106.35	95.10
Intercept	Isotrop (p=2)	3	199.00	164.03
Linear	Isotrop (p=2)	6	81.60	118.36
Linear+WW	Isotrop (p=2)	9	117.98	95.11
Intercept	Anisotrop (p=1)	5	26.26	126.02
Linear	Anisotrop (p=1)	8	8.08	83.35
Intercept	Anisotrop (p=2)	5	4.13	92.62
Linear	Anisotrop (p=2)	8	2.03	64.86

Tabelle A.5: Gauß-Prozess-Modelle für die Fertigungszeit t

Trendkomponente	Power Exponential Korrelationsfunktion	# Modell- parameter	MSCVE	BIC
Intercept	Isotrop (p=1)	3	1.48	81.033
Linear	Isotrop (p=1)	6	0.81	50.466
Linear+WW	Isotrop (p=1)	9	0.98	35.999
Intercept	Isotrop (p=2)	3	1.76	74.933
Linear	Isotrop (p=2)	6	1.19	52.766
Linear+WW	Isotrop (p=2)	9	1.41	38.099
Intercept	Anisotrop (p=1)	5	2.13	90.122
Linear	Anisotrop (p=1)	8	0.76	53.250
Intercept	Anisotrop (p=2)	5	2.54	78.122
Linear	Anisotrop (p=2)	8	1.49	55.355

Tabelle A.6: Gauß-Prozess-Modelle für die Napftiefe NT

Parameter	Schätzer	Standard- abweichung	p-Wert (Wald)
Intercept	-37.34920	13.02068	0.0047
e_{AQU}	1.01840	0.45440	0.0250
f	0.03170	0.01230	0.0104
e_{AQU}^2	-0.01180	0.00547	0.0311
f^2	-0.00001	<0.00001	0.0076

Tabelle A.7: Logistisches Regressionsmodell der Zielgröße $Bauteilversagen$ nach Vorwärtsselektion zum Niveau 0.05

A.3 Abbildungen

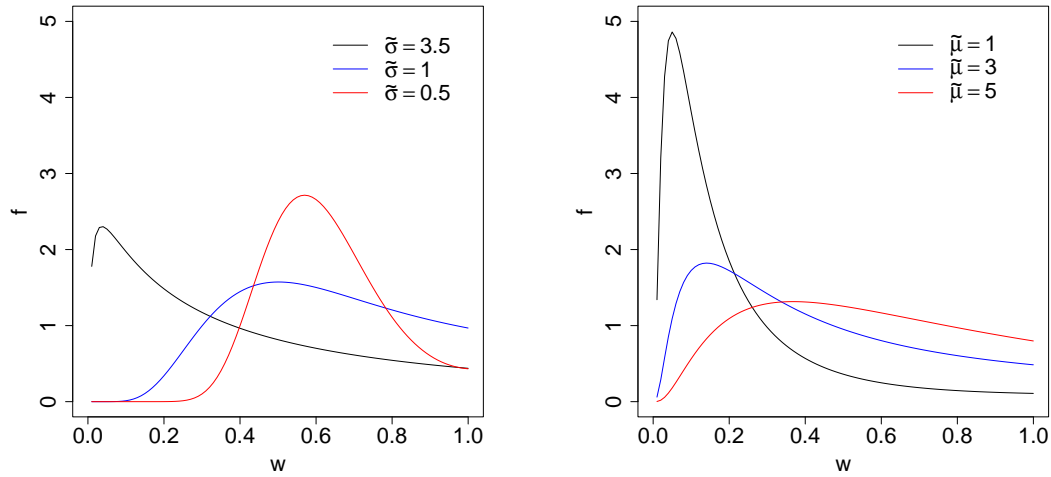


Abbildung A.1: Graphen der Dichten zweiseitiger Wünschbarkeitsfunktionen nach Harrington mit a) $\tilde{\mu} = 4$ und b) $\tilde{\sigma} = 2$

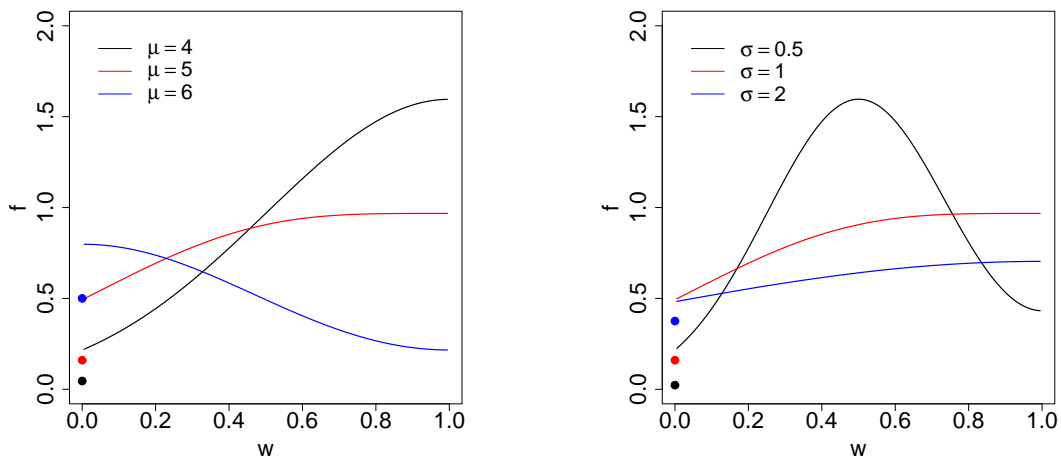


Abbildung A.2: Graph der Dichten zweiseitiger Wünschbarkeitsfunktionen nach Derringer und Suich mit $(USG, T, OSG) = (2, 4, 6)$ und a) $\sigma = 1$ b) $\mu = 5$

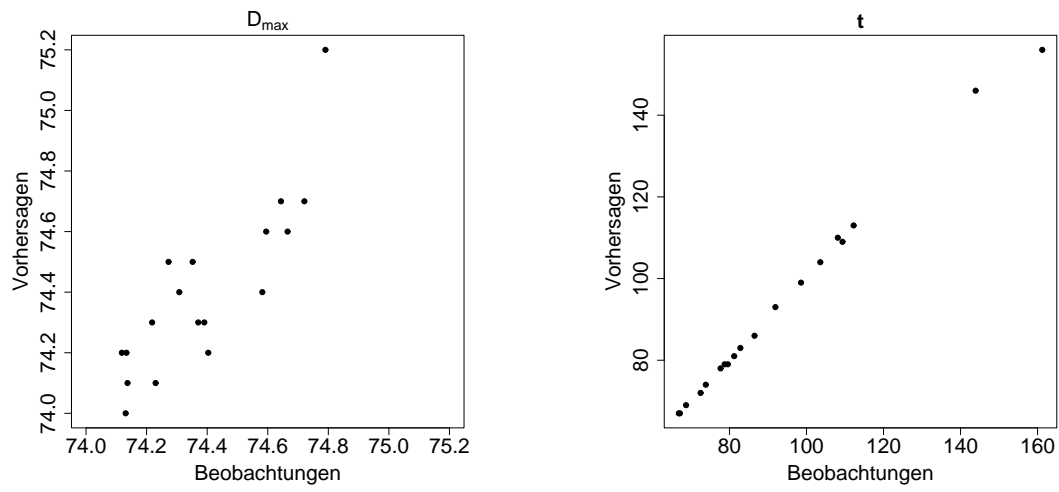


Abbildung A.3: Vergleich der Kreuzvalidierungsvorhersagen mit den Beobachtungen für a) D_{max} , b) t

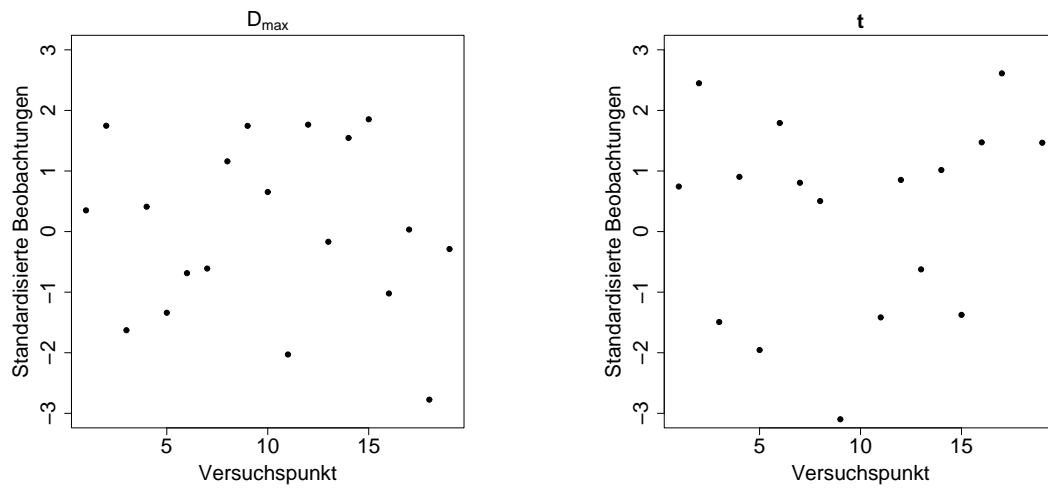


Abbildung A.4: Standardisierte Beobachtungen mittels Kreuzvalidierung für a) D_{max} , b) t

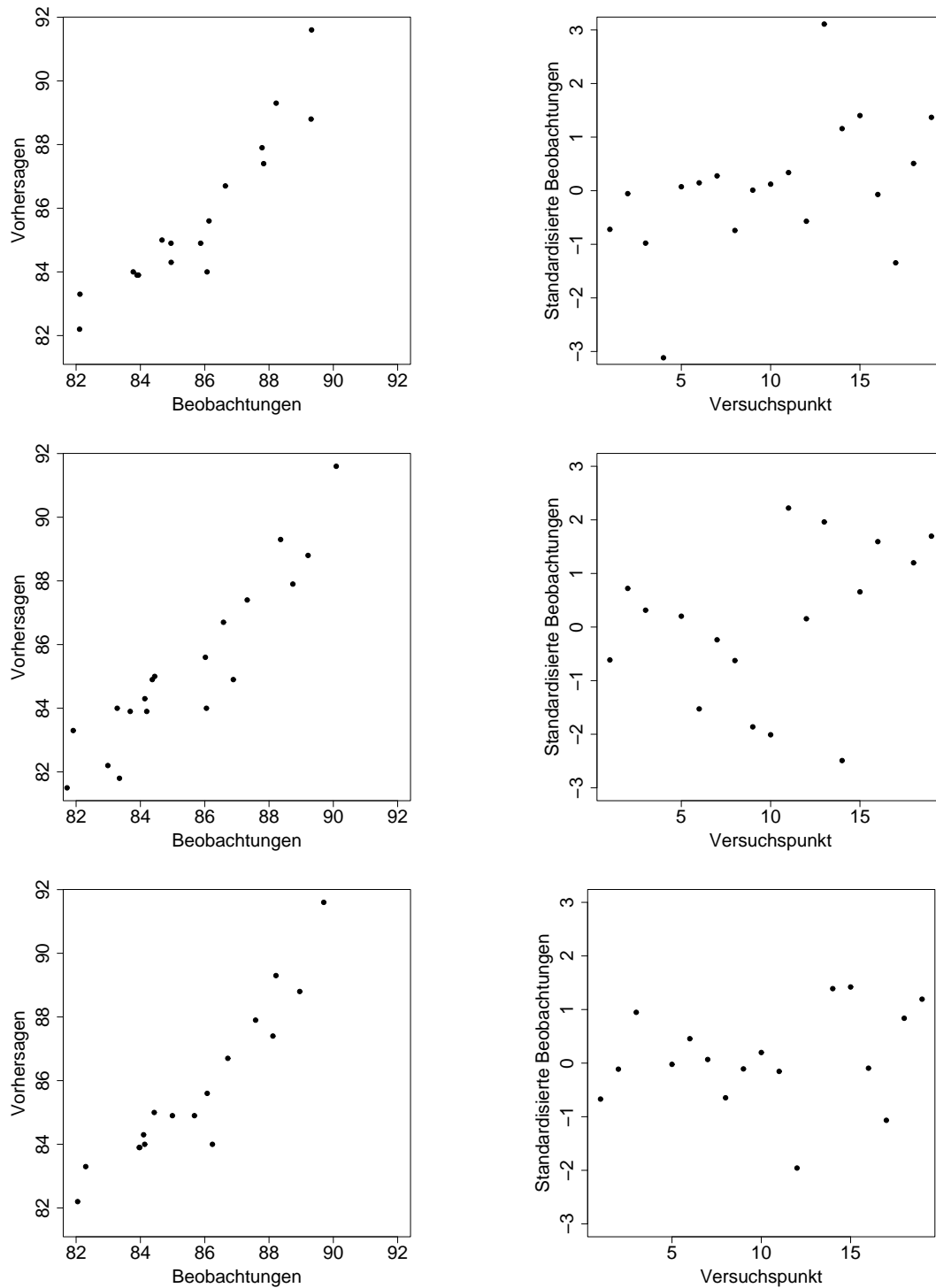


Abbildung A.5: Diagnostische Plots der Gauß-Prozess-Modelle für NT . Erste Reihe: linearer Trend und isotrope Korrelationsfunktion ($p=1$). Zweite Reihe: linearer Trend plus Wechselwirkungen und isotrope Korrelationsfunktion ($p=1$). Dritte Reihe: linearer Trend und anisotrope Korrelationsfunktion ($p=1$).

A.4 Programmcode

Maximin.R (*Generierung von Maximin Designs*)

```
library(gtools)
dimension <- 2      #Dimension der Daten
n <- 5             #Anzahl der Versuchspunkte
m <- 10000        #Wiederholungen
schrittweite <- 0.01 #Feinheit des Grids

a <- seq(0,1,by=schrittweite)
t <- permutations(length(a),dimension,a,repeats.allowed=TRUE)
index <- seq(1,dim(t)[1],by=1)
dMn <- 0

for (l in 1:m) {
  index2 <- sample(index,n,replace=FALSE)
  punkte <- t[index2,]
  minimum <- min(dist(punkte))
  if (minimum>dMn) {dMn <- minimum; Mn <- punkte}
}
```

Miminax.R (*Generierung von Minimax Designs*)

```
dimension <- 3      #Dimension der Daten
n <- 5             #Anzahl der Versuchspunkte
m <- 10000        #Wiederholungen
schrittweite <- 0.1 #Feinheit des Grids

a <- seq(0,1,by=schrittweite)
t <- permutations(length(a),dimension,a,repeats.allowed=TRUE)
index <- seq(1,dim(t)[1],by=1)
dmM <- 100
```

```

for (l in 1:m) {
  index2 <- sample(index,n,replace=FALSE)
  punkte <- t[index2,]
  abstand.vp <- rep(NA,n)
  abstand <- rep(NA,dim(t)[1])
  for (i in 1:dim(t)[1]){
    for (j in 1:n){
      abstand.vp[j] <- sqrt(sum((punkte[j,]-t[i,])^2))
      abstand[i] <- min(abstand.vp)
    }
  }
  if (max(abstand)<dmM) {dmM <- max(abstand); minimax <- punkte}
  print(max(abstand))
}
minimax

```

MnLHD.R (*Generierung von Maximin LHDs*)

```

dimension <- 2      #Dimension der Daten
n <- 10            #Anzahl der Versuchspunkte
x <- seq(1,n)
test <- 0
int <- seq(-10,10,1)
MnLHD <- matrix(rep(0,2*n),n,2)
for (i in 1:10000){
  p <- matrix(c(rep(NA,n*dimension)),n,dimension)
  for (j in 1:dimension){
    p[,j] <- sample(x1,n,relace=FALSE)/n-1/(2*n)
  }
  min <- min(dist(p))
  if (min>test) {test <- min; MnLHD <- p }
}

```

unterraum.R (*Generierung von raumfüllenden Versuchsplänen, die auch auf einem gewissen Unterraum eine gute raumfüllende Eigenschaft besitzen*)

```
library(fields)
dimension <- 3           #Dimension der Daten
unterdimension <- 2     #Unterdimension der aktiven Faktoren
schrittweite <- 0.05    #Feinheit des Grids
anzahl <- 7            #Anzahl der Versuchspunkte
wiederholung <- 100

seq <- seq(0,1,by=schrittweite)
t <- permutations(length(seq),dimension,seq,repeats.allowed=TRUE)
t2 <- permutations(length(seq),unterdimension,seq,repeats.allowed=TRUE)
index <- 100

for (m in 1:wiederholung) {
  design <- cover.design(t,anzahl)
  test <- matrix(rep(NA,dim(t2)[1]*length(design[,1])),
                dim(t2)[1],length(design[,1]))
  a <- rep(NA,dim(t2)[1])
  for (i in 1:dim(t2)[1]){
    for (j in 1:length(design[,1])){
      test[i,j] <- dist(matrix(c(t2[i,],
                                design[j,1:unterdimension]),2,unterdimension,byrow=TRUE))
      a[i] <- min(test[i,])
    }
  }

  if (max(a)$<$index) {index <- max(a); design.opt <- design}
  print(m)
}
```

konvex.R (*Dieses Programm berechnet die konvexen Hüllen für einen Versuchsplan*)

```

library(gtools)

# k Vektor der Zielgröße Versagen (0= kein versagen, 1= versagen)
# X Versuchsplan
#gitter = Grid

gridpunkte <- function(X,gitter,k) {

m <- dim(X)[2]                #Anzahl der Versuchspunkte
n <-dim(X)[1]                 #Dimension
zp <-matrix(X[,which(k==0)],nrow=n) #Anzahl zulässige Punkte
nzp <-matrix(X[,which(k!=0)],nrow=n) #Anzahl nichtzulässiger Punkte
anzahl_nzp <- dim(nzp)[2]
anzahl_zp <- dim(zp)[2]

if (anzahl_zp <- (n-1)) {return(gitter)} else
if (anzahl_nzp==0) {return(gitter)}
else {komb <- combinations(anzahl_zp, n)

#Funktion zul_n identifiziert Punkt, der in konvexer Hülle liegt
zul_n <- function(x){
  for(i in 1:anzahl_nzp){
    for (j in 1:length(komb[,1])) {

      #vermeiden von singularitäten
      punkt <- zp[,komb[j,]]
      #2 zp und 1 nzp auf einer Geraden mit Steigung 0
      if (punkt[1,1]==punkt[1,2] && punkt[1,1]==nzp[1,i]) {j=j+1}
      #2 zp und 1 nzp auf einer Geraden mit Steigung unendlich
      if (punkt[2,1]==punkt[2,2] && punkt[2,1]==nzp[2,i]) {j=j+1}
      test1 <- (punkt[1,1]-nzp[1,i])/(punkt[2,1]-nzp[2,i])
      test2 <- (punkt[1,2]-nzp[1,i])/(punkt[2,2]-nzp[2,i])
      #2 zp und 1 nzp auf einer Geraden mit Steigung ungleich
    }
  }
}

```

```
#0 und unendlich
if (test1==test2) {j=j+1}

s <- matrix(c(rep(nzp[,i], n)), nrow=n)-zp[,komb[j,]]
p <- x-nzp[,i]
loesung <- solve(s, p)
#mit i=anzahl_nzp+1 wird erreicht, dass wenn der Punkt
#in einem der Räume ist, die anderen Räume nicht mehr
#untersucht werden
if (all(loesung$>=$0)) {return(x) ; i=anzahl_nzp+1 }
}
}

hilfe <- 0
#Für jeden Gridpunkt wird überprüft, ob er in einer
#konvexen Hülle liegt
for (o in 1:dim(gitter)[1]) {erg <- zul_n(gitter[o,])
  hilfe <- append(hilfe,erg)}
#Matrix m der Punkte, die in einer der konvexen Hüllen liegen
m <- matrix(hilfe[-1],ncol=n,byrow=TRUE)
#Mit Hilfe der Matrix m werden alle Punkte bestimmt, die
#nicht in konvexer Hülle liegen
total <- rbind(gitter,m)
hilfe2 <- apply(total, 1, paste, collapse="")
total1 <- total[table(hilfe2)[hilfe2] == 1,]
return(total1)
}
}
```

algorithmus.R (*Dieses Programm führt den Algorithmus zur sequentiellen Generierung eines raumfüllenden Versuchsplans aus*)

```
# erst Programm convex laufen lassen
# vp = aktueller Versuchsplan, Spalte 1 bis 3: Faktoren,
#     Spalte 4: Versagen (0=nein, 1=ja)

#Transformation der Faktoren in [0,1]-Intervall
#für die Funktion gridpunkte
e_aqu_t <- (vp[,1]-20)/40
v_aqu_t <- (vp[,2]-0.95)/0.1
f_t <- (vp[,3]-500)/2500
out <- cbind(e_aqu_t,v_aqu_t,f_t)

out2 <- out
out3 <- t(out2)
#addiere kleinen Zufallsfehler um Singularitäten bei
#der Funktion gridpunkte zu vermeiden
e <- abs(rnorm(length(out3[2,])*3)/100)
e.matrix <- matrix(e,3,length(out3[2,]),byrow=T)
out4 <- out3-e.matrix

#Schritt 2: erzeuge konvexe hüllen
schrittweite <- 0.1
a <- seq(0,1,by=schrittweite)
t <- permutations(length(a),3,a,repeats.allowed=TRUE)
k <- vp[,4]
grid <- gridpunkte(out4,t,k)

e_aqu <- grid[,1]*40+20
v_aqu <- grid[,2]*0.1+0.95
f <- grid[,3]*2500+500
plot3d(e_aqu,v_aqu,f,xlim=c(20,60),
       ylim=c(0.95,1.05),zlim=c(500,3000))
```

```
#Schritt 3: Bestimmung des erweiterten Versuchsbereichs
e_aqu_min <- 30
e_aqu_max <-50
v_aqu_min <-0.95
v_aqu_max <-1.05
f_min <- 500
f_max <- 1500
e_aqu_min_c <- (e_aqu_min-20)/40
e_aqu_max_c <- (e_aqu_max-20)/40
v_aqu_min_c <- (v_aqu_min-0.95)/0.1
v_aqu_max_c <- (v_aqu_max-0.95)/0.1
f_min_c <-(f_min-500)/2500
f_max_c <-(f_max-500)/2500
e_aqu_c <- seq(e_aqu_min_c,e_aqu_max_c,by=0.05)
v_aqu_c <- seq(v_aqu_min_c,v_aqu_max_c,by=0.05)
f_c <- seq(f_min_c,f_max_c,by=0.05)

index <- 0
for (i in 1:length(e_aqu_c)){
  for (j in 1:length(v_aqu_c)){
    for (k in 1:length(f_c)){
      index <- append(index,c(e_aqu_c[i],v_aqu_c[j],f_c[k]))
    }
  }
}
t.xyz <- matrix(index[-1],length(index[-1])/3,3,byrow=TRUE)

all <- rbind(grid,t.xyz)
#Gridpunkte im erweiterten Vp ohne konvexe Hüllen
all1 <- all[which(duplicated(all)==TRUE),]

#Schritt 4: Hinzufügen von neuen Versuchspunkten
anzahl <- 5
zusammen <- rbind(out2,all1)
```



```

#in row werden die Zeilen vom Grid bestimmt die als
# feste Versuchspunkte übernommen werden sollen
row <- seq(1,length(out2[,1]),1)
zusammen1 <- zusammen[which(duplicated(zusammen)==FALSE),]
out <- cover.design(zusammen1,anzahl,nruns=1,fixed=row,nn=FALSE)

e_aqu <- out[,1]*40+20
v_aqu <- out[,2]*0.1+0.95
f <- out[,3]*2500+500
farbe <- c(rep(1,5),rep(2,10))
scatterplot3d(e_aqu,v_aqu,f,type='h',pch=19,xlab=expression(e[AQU]),
ylab=expression(V[AQU]),zlab=expression(f),cex.axis=1.9,cex.lab=1.6,
zlim=c(500,3000),xlim=c(20,60),ylim=c(0.95,1.05),cex.sub=1.7)
s3d\$\points3d(e_aqu[1:5],v_aqu[1:5],f[1:5],type='h',col='blue',pch=19)

```

ei1.SAS (*Dieses Programm passt die Modelle der Zielgrößen an und berechnet die Vorhersagen in jedem Gridpunkt*)

```

data grid;
do e_aqu=20 to 60 by 1;
do v_aqu=0.95 to 1.05 by 0.005;
do f=700 to 2000 by 50;
output;
end;end;end;
run;

data zusammen;
set grid daten165;
e_aqu_t=(e_aqu-20)/40;
v_aqu_t=(v_aqu-0.95)/0.1;
f_t=(f-500)/2500;
run;

proc logistic data=zusammen;
model falten=e_aqu f f*f e_aqu*e_aqu;

```

```
output out=predict predprobs=i;
run;quit;

data t_falten;
set predict;
if falten=.;
wk_falten=IP_1;
run;

data daten1;
set daten165;
b0=64.7;
b1=-0.771;
des_nt=exp(-exp(-(b0+b1*nt)));
b0=458.09;
b1=-6.16;
des_dmax=exp(-exp(-(b0+b1*dmax)));
b0=7.64;
b1=-0.077;
des_zeit=exp(-exp(-(b0+b1*zeit)));
odes=(des_nt**(9/25))*(des_dmax**(9/25))*(des_zeit**(7/25));
run;

PROC EXPORT DATA= WORK.daten1
OUTFILE='C:\user\nadine\mei_w.txt'
DBMS=TAB REPLACE;
RUN;

proc mixed data=zusammen method=reml scoring=50 convh=1e-8
alpha=0.05 asycov noinfo noitprint;
model zeit=e_aqu_t v_aqu_t f_t /alphap=0.01 solution
outpred=predicted;
parms (0 to 10 by 1)(0 to 1 by 0.1)(0 to 50 by 1)
(2) (2) (2) (0.1) /hold=4,5,6;
repeated /type=SP(expa)(e_aqu_t v_aqu_t f_t) subject=intercept;
```

```
run;  
quit;
```

```
data t_zeit;  
set predicted;  
if falten NE . then delete;  
if pred<60 then pred=60;  
pred_zeit=pred;  
s_zeit=stderrpred;  
run;
```

```
proc mixed data=zusammen method=reml scoring=50 convh=1e-8  
alpha=0.05 asycov noinfo noitprint;  
model dmax= /alphap=0.01 solution outpred=predicted;  
parms (0 to 10 by 1)(0 to 1 by 0.1)(0 to 50 by 1)  
(2) (2) (2) (0.1) /hold=4,5,6;  
repeated /type=SP(expa)(e_aqu_t v_aqu_t f_t) subject=intercept;  
run;  
quit;
```

```
data t_dmax;  
set predicted;  
if falten NE . then delete;  
if pred<74 then pred=74;  
pred_dmax=pred;  
s_dmax=stderrpred;  
run;
```

```
proc mixed data=zusammen method=reml scoring=50 convh=1e-8  
alpha=0.05 asycov noinfo noitprint;  
model nt=e_aqu_t v_aqu_t f_t /alphap=0.01 solution  
outpred=predicted;  
parms (0.1)(0 to 1 by 0.1)(0 to 50 by 1)  
(1) (1) (1) (0.1) /hold=1,4,5,6;  
repeated /type=SP(expa)(e_aqu_t v_aqu_t f_t) subject=intercept;
```

```
run;
quit;

data t_nt;
set predicted;
if falten NE . then delete;
if pred<81 then pred=81;
pred_nt=pred;
s_nt=stderrpred;
run;

data zusammen1;
merge t_zeit t_dmax t_nt t_falten;
keep e_aqu_t v_aqu_t f_t pred_zeit s_zeit
pred_nt s_nt pred_dmax s_dmax wk_falten;
run;

PROC EXPORT DATA= WORK.ZUSAMMEN1
OUTFILE='C:\user\nadine\mei.txt'
DBMS=TAB REPLACE;
RUN;
```

ei2.R (*Dieses Programm berechnet die Expected Improvement nach Ansatz 3*)

```
library(djmr1)
library(MASS)

#Grid zur Vorhersage
daten <- read.table('c:/user/nadine/mei.txt',header=TRUE,sep='\t')

#Versuchsplan mit Wünschbarkeitsindex für jeden Versuchspunkt
daten_w <- read.table('c:/user/nadine/mei_w.txt',
header=TRUE,sep='\t')

#maximaler Wünschbarkeitsindex
```

```
wopt <- max(daten_w[,14],na.rm=TRUE)

#Vorhersagen der Zielgrößen für jeden Gridpunkt
ydach <- cbind(daten[,4],daten[,6],daten[,8],daten[,10])
wkfalten <- daten[,10]

#Standardabweichung der Vorhersage für jeden Gridpunkt
s1 <- daten[,5]
s2 <- daten[,7]
s3 <- daten[,9]

#Konstante b0 und b1 der Wünschbarkeitsfunktion von Harrington
b0 <- c(7.64,458.09,64.7)
b1 <- c(-0.077,-6.16,-0.771)
aopt<-c(0,0,0)
interval<-0.02

#Gewichtung zwischen lokaler und globaler Suche
g<-3

ei <-rep(NA,length(ydach[,1]))

#Berechnung der Punktmenge Fmax
nt <- seq(0,1,0.02)
dmax <- seq(0,1,0.02)
zeit <- seq(0,1,0.02)
a <- c(0,0,0)
for (i in 1:length(zeit)){
  for (j in 1:length(dmax)){
    for (k in 1:length(nt)){
      w <- (zeit[i]^(9/25))*(dmax[j]^(9/25))*(nt[k]^(7/25))

      if (w>=wopt & w<wopt+0.013)
        {a <- cbind(a,c(nt[i],dmax[j],zeit[k]))}
    }
  }
}
```

```

    }
  }
}

aopt <- a[,-1]
plot3d(a[1,2:dim(a)[2]],a[2,2:dim(a)[2]],a[3,2:dim(a)[2]],
xlim=c(0,1),ylim=c(0,1),zlim=c(0,1))

#Berechnung des Grids mit WI$>$Fmax
wopt.raum <- rep(0,3)
interval <- 0.015
x1 <- seq(0.001,0.999,interval) #zeit
x2 <- seq(0.001, 0.999,interval) #dmax
x3 <- seq(0.001, 0.999,interval) #nt
for (i in 1:length(x1)){
  for (j in 1:length(x2)){
    for (k in 1:length(x3)){
      x <- c(x1[i],x2[j],x3[k])
      w <- (x1[i]^(9/25))*(x2[j]^(9/25))*(x3[k]^(7/25))
      if (w>=wopt)
        {wopt.raum <- cbind(wopt.raum,x)}
    }
  }
}

wopt.raum <- wopt.raum[,-1]
plot3d(wopt.raum[1,],wopt.raum[2,],wopt.raum[3,],
xlim=c(0,1),ylim=c(0,1),zlim=c(0,1))

#Distanzmatrix des Grids zu Fmax
dist <- rep(NA,dim(wopt.raum)[2])
dist1 <- rep(NA,dim(aopt)[2])
for (i in 1:dim(wopt.raum)[2]) {
  for (j in 1:dim(aopt)[2]){

```

```

        dist1[j] <- sqrt(sum((wopt.raum[,i]-aopt[,j])^2))
        dist[i] <- (min(dist1))^g
    }
}

#Berechnung der Expected Improvement
for (m in 1:length(s1)){
  if (wkfalten[m]> 0.3) {ei[m] <- 0} else
  {mu <- -(b0+b1*ydach[m,1:3])
  var1 <- (b1[1]^2*s1[m]^2)
  var2 <- (b1[2]^2*s2[m]^2)
  var3 <- (b1[3]^2*s3[m]^2)
  Sigma <- matrix(c(var1,0.656*sqrt(var1)*sqrt(var2),0,
  0.656*sqrt(var1)*sqrt(var2),var2,0.588*sqrt(var2)*sqrt(var3),
  0,0.588*sqrt(var2)*sqrt(var3),var3),3,3)
  gSigma <- ginv(Sigma)

  punkte <- c(0,0)
  for (z in 1:dim(wopt.raum)[2]){
    x <- wopt.raum[,z]
    f <- (-1/(((2*pi)^3)*sqrt(det(Sigma))*
    x[1]*log(x[1])*x[2]*log(x[2])*x[3]*log(x[3]))) *
    exp(-1/2*t(log(-log(x))-mu)\%*\%gSigma\%*\%
    (log(-log(x))-mu))
    if (f>0.01) {punkte <- rbind(punkte,c(f,dist[z]))}
    if (length(punkte)>2)
    {ei[m] <- sum(punkte[,1]*punkte[,2]*interval^3)*
    (1-wkfalten[m])} else {ei[m] <- 0}
  }
  print(m)
}

daten[which(ei==max(ei)),]

```

Literaturverzeichnis

- [1] Agresti, A. (1990): *Categorical Data Analysis*, Wiley, New York.
- [2] Akaike, H. (1974): Information theory and extension of the maximum likelihood principle, in *Proceedings of the 2nd International Symposium on Information*, edited by B. N. Petrov and F. Czaki, Akademic Kiado, Budapest.
- [3] Auer, C., Erdbrügge, M., Göbel, R. (2004): Comparison of Multivariate Methods for Robust Parameter Design in Sheet Metal Spinning, *Applied Stochastic Models in Business and Industry* ASMBI, **20**, 201-218.
- [4] Beierle, C., Kern-Isberner, G. (2000): *Methoden wissensbasierter Systeme*, 2. Auflage, Vieweg 2000, ISBN 3-528-15723-2.
- [5] Bergmann, R., Althoff, K. D., Breenm, S., Gökerm, M., Managom, M., Traphönerm, R., Wessm, S. (1999): *Developing Industrial Case-Based Reasoning Applications*, *Lecture Notes in Computer Science*, Springer, Berlin.
- [6] Berk, R., Bickel, P., Campbell, K., Fovell, R., Keller-McNulty, S., Kelly, E., Linn, R., Park, B., Perelson, A., Roupail, N., Sacks, J., Schoenberg, F. (2002): Workshop on statistical approaches for the evaluation of complex computer models, *Statistical Science*, **17**(2), 173-192.
- [7] Booker, A. J., Dennis, J. E., Frank, P. D., Serafini, D. B., Torczon, V., Trosset, M. W. (1999): A rigorous framework for optimization of expensive functions by surrogates, *Structural Optimization*, **17**, 1-13.

-
- [8] Christakos, G. (1992): Random Field Models in Earth Sciences, Academic Press, San Diego.
- [9] Cressie, N. A. (1993): Statistics for Spatial Data, J. Wiley, New York.
- [10] Derringer, G., Suich, R. (1980): Simultaneous Optimization of Several Response Variables, *Journal of Quality Technology*, **12**, 214-219.
- [11] Erdbrügge, M., Göbel, R., Kleiner M., Kunert, J. (2001): Optimales Drücken. Statistische Versuchplanung bei qualitativen Zielgrößen, *Qualität und Zuverlässigkeit*, **46**, 1180-1183
- [12] Ewers, R. (2005): Prozessauslegung und Optimierung des CNC-gesteuerten Drückens, Dissertation, Fachbereich Maschinenbau, Universität Dortmund.
- [13] Fang, K. T., Lin, D. K., Winker, P., Zhang, Y. (2000): Uniform Design: Theory and Application, *Technometrics*, **42**, 237-248.
- [14] Geisser, S. (1975): The predictive sample reuse method with applications, *Journal of the American Statistical Association*, **70**, 320-328.
- [15] Göbel, R., Henkenjohann, N., Kleiner, M. (2004): Process Optimization in Sheet Metal Spinning by Adaptive Sequential Design of Experiments, ENBIS conference, CD-ROM, 20-22. September, Kopenhagen, Dänemark.
- [16] Göbel, R., Kleiner, M., Henkenjohann, N. (2005): New Approach for Process Planning and Optimization in Sheet Metal Spinning, *Advanced Materials Research*, **6-8**, 493-500.
- [17] Handcock, M. S. (1991): On cascading latin hypercube designs and additive models for experiments, *Communications Statistics-Theory Methods*, **20**, 417-439.
- [18] Harrington, J. (1965): The desirability function, *Industrial Quality Control*, **21**, 494-498.
- [19] Henkenjohann, N., Göbel, R., Kleiner, M., Kunert, J. (2005): An Adaptive Sequential Procedure for Efficient Optimization of the Sheet Metal Spinning

- Process, *Quality and Reliability Engineering International* (Special Issue), **21**, 439-455.
- [20] Holland, B., Ahsanullah, M. (1989): Further Results on a Distribution of Meinhold and Singpurwalla, *The American Statistician*, **43**, 216-219.
- [21] Johnson, M. E., Moore, L. M., Ylvisaker, D. (1990): Minimax and maximin distance designs, *Journal of statistical planning and inference*, **26**, 131-148.
- [22] Jones, D. R., Schonlau, M., Welch, W. J. (1998): Efficient Global Optimization of Expensive Black-Box Functions, *Journal of Global Optimization*, **13**, 455-492.
- [23] Kaballo, W. (1996): Einführung in die Analysis I, Spektrum Akademischer Verlag, Heidelberg.
- [24] Kawai, K., Sawano, S., Itho, H. (1989): A Trail Approach to Spinning Data Base, *Journal of JSTP*, **Vol 30**, No. 345, 1411-1415.
- [25] Kleijnen, J. P. C., Sanchez, S. M., Lucas, T. W., Cloppa, T. M. (2005): A user's guide to the brave new world of simulation experiments, *INFORMS Journal on Computing*, **17**, 263-289.
- [26] Langewiesche, N. (2005): Interpolation von Stochastic Process Models für unterschiedliche Werkstückgeometrien beim CNC-gesteuerten Drücken, Diplomarbeit, Fachbereich Statistik, Universität Dortmund.
- [27] Lehman, J. S., Santner, T. J., Notz, W. I. (2004): Designing computer experiments to determine robust control variables, *Statistica Sinica*, **14**, 517-590.
- [28] Lenz, M., Bartsch-Spörl, B., Burkhard, H.-D., Wess, S. (1991): Case-Based Reasoning Technology - From Foundations to Applications, LNAI 1400, Springer, London.
- [29] Mardia, K. V., Marshall, R. J. (1984): Maximum likelihood estimation of models for residual covariance in spatial regression, *Biometrika*, **71**, 135-146.

-
- [30] McKay, M. D., Beckmann, R. J., Conover, W. J. (1979): A comparison of three methods for selecting values of input variables in the analysis of output from a computer code, *Technometrics*, **21**, 239-245.
- [31] Morris, M. D., Mitchell, T. J. (1995): Exploratory designs for computational experiments, *Journal of Statistical Planning and Inference*, **43**, 381-402.
- [32] O'Connell, M. A., Wolfinger, R. D. (1997): Spatial Regression Models, Response Surfaces, and Process Optimization, *Journal of Computational and Graphical Statistics*, **6**, 224-241.
- [33] Osio, I. G., Amon, C. H. (1996): An engineering design methodology with multistage Bayesian surrogates and optimal sampling, *Research in Engineering Design*, **8**, 189-206.
- [34] Patterson, H. D., Thompson, R. (1974): Maximum likelihood estimation of variance components, *Proceedings of the 8th International Biometric Conference*, 197-207.
- [35] Rohatgi, V. K. (1974): An introduction to probability theory and mathematical statistics, Wiley, New York.
- [36] Runge, M. (1993): Drücken und Drückwalzen, Verlag Moderne Industrie, Landsberg/Lech.
- [37] Sacks, J., Welch, W. J., Mitchell, T. J., Wynn, H. P. (1989): Design and Analysis of Computer Experiments, *Statistical Science*, **4**, 409-423.
- [38] Santner, T. J., Williams, B. J., Notz, W. I. (2003): The Design and Analysis of Computer Experiments, Springer, New York.
- [39] Schaller, M. (2004): Comparison of Models for the Analysis of a Data Set in Surface Grinding, Diplomarbeit, Fachbereich Statistik, Universität Dortmund.
- [40] Schonlau, M. (1997): Computer experiments and global optimization, PhD thesis, Department of Statistics and Actuarial Science, University of Waterloo, Waterloo, CA.

-
- [41] Schonlau, M., Welch, W. J., Jones, D. R. (1998): Global versus Local Search in Constrained Optimization of Computer Models, In: New Developments and Applications in Experimental Design, N. Flournoy, W.F. Rosenberger, and W.K. Wong, Institute of Mathematical Statistics, Hayward, California, *34*, 11-25.
- [42] Schwarz, G. (1975): Estimating the dimension of a model, *Annals of Statistics*, **16**, 461-464.
- [43] Steuer, D. (2005): Statistische Eigenschaften der multikriteriellen Optimierung mittels Wünschbarkeiten, Dissertation, Universität Dortmund.
- [44] Stinstra E., den Hertog D., Stehouwer P., Vestjens A. (2003): Constrained Maximin Designs for Computer Experiments, *Technometrics*, **45**, No. 4, 340-346.
- [45] Stone, M. (1974): Cross-validatory choice and assessment of statistical predictions, *Journal of the Royal Statistical Association B*, **36**, 111-147.
- [46] Trautmann, H., Weihs, C. (2006): On the distribution of the desirability index using Harrington's desirability function, *Metrika*, **63**, 207-213.
- [47] Trosset, M. W. (1999): Approximate Maximin Distance Designs, *ASA Proceedings of the Section on Physical and engineering Sciences*, 223-227, American Statistical Association (Alexandria, VA).
- [48] Wang, G. G. (2003): Adaptive response surface method using inherited Latin hypercube design points, *ASME Journal of Mechanical Design*, **125**, 210-220.
- [49] Wang, G. G., Simpson, T. W. (2004): Fuzzy clustering based hierarchical metamodeling for space reduction and design optimization, *Journal of Engineering Optimization*, **36**, 313-335.
- [50] Williams, B. J., Santner, T. J., Notz, W. I. (2000): Sequential design of computer experiments to minimize integrated response functions, *Statistica Sinica*, **10**, 1133-1152.

-
- [51] Wolfinger, R., Tobias, R., Sall, J. (1994): Computing Gaussian Likelihood and their Derivatives for General Linear Mixed Models, *SIAM Journal on Scientific Computing*, **15**, 1294-1310.
- [52] Wu, C. F. J., Hamada, M. (2000): Experiments: Planning, Analysis, and Parameter Design Optimization, J. Wiley, New York.
- [53] Zimmermann, D. L. (1989): Computational exploitable structure of covariance matrices and generalized covariance matrices in spatial models, *Journal of Statistical Computation and Simulation*, **32**, 1-15.