

Effektive Beobachtung von zufälligen Funktionen
unter besonderer Berücksichtigung von Ableitungen

von der Fakultät für Mathematik und Informatik

der Technischen Universität Bergakademie Freiberg
genehmigte

DISSERTATION

zur Erlangung des akademischen Grades
doctor rerum naturalium
Dr. rer. nat.,

vorgelegt
von Diplom-Mathematiker Markus Holtmann

geboren am 31. Mai 1971 in Georgsmarienhütte

Gutachter: Prof. Dr. rer. nat. Wolfgang Näther, Freiberg
Prof. Dr. ing. Joachim Menz, Freiberg
Prof. Dr. rer. nat. Jürgen Pilz, Klagenfurt (Österreich)

Tag der Verleihung: 15. Juni 2001

Inhaltsverzeichnis

1. Einleitung und Zusammenfassung	1
2. Definitionen und wesentliche Begriffe	5
2.1. Zufällige Funktionen	5
2.1.1. Definition	5
2.1.2. Momente von zufälligen Funktionen	6
2.1.3. Stationarität und Isotropie	7
2.2. Kovarianzfunktionen	8
2.2.1. Wesentliche Eigenschaften	9
2.2.2. Hilberträume	9
2.2.3. Stetigkeit, Differenzierbarkeit und Integrierbarkeit	10
2.2.4. Spektraldarstellungen	13
3. Lineare Rekonstruktion und Vorhersage	17
3.1. Lineare Approximationsprobleme	17
3.1.1. Definition	17
3.1.2. Kernreproduzierende Hilberträume	20
3.2. Krigingmethoden	22
3.2.1. Simple Kriging und BLP	22
3.2.2. Universal Kriging und BLUP	24
3.3. Asymptotische Versuchsplanung	26
3.3.1. Versuchsplanung	27
3.3.2. Asymptotische Verfahren nach MÜLLER-GRONBACH und RITTER	28
3.3.3. Versuchsplanung mit nichthermiteschen Daten	34
3.4. Splinemethoden	40
3.4.1. Allgemeine Äquivalenz zwischen dem BLP und Splinemethoden	42
3.4.2. Ein splinebasiertes Vorhersageverfahren bei Brownschen Bewegungen	48

4. Versuchsplanung für Regression	53
4.1. Einordnung in den Kontext	53
4.2. Relativer Einfluß von Ableitungen auf die Varianz des Regressionsschätzers	55
4.3. Versuchsplanung bei lokaler Regression	58
5. Versuchsplanung mit endlichen Designs	61
5.1. Versuchsplanung für den BLP mittels gestörter Kovarianzmatrizen	61
5.1.1. Verallgemeinerung und Störung des \square -Funktionals	63
5.1.2. Berechnung der Richtungsableitung von \square^γ	64
5.1.3. Ein Algorithmus und Beispiele	68
5.1.4. Zusammenfassung und weitere Verwendungsmöglichkeiten	70
Literaturverzeichnis	73
A. Anhang	77
B. Anhang	78
B.1. Konvergenzverhalten von alternativen Varianztermen	78
C. Anhang	80
C.1. Zerlegung der Schätzvarianz bei lokaler Regression	80

1. Einleitung und Zusammenfassung

Es liegt in der Natur der Dinge, daß eine Vielzahl von Phänomenen, welche den Menschen umgeben, oft genug nicht in ihrer Gesamtheit erfaßt werden können oder deren vollständige Erfassung sehr große Anstrengungen erfordern würde. Geologische Erscheinungen, Probleme in der Geodäsie, Wettervorhersagen, Klimaentwicklungen, Börsenkurse und die Verifizierung komplexer Computerprogramme sind Beispiel hierfür. In anderen Fällen treten Phänomene auf, für die (noch) keine physikalischen Gesetzmäßigkeiten in Form einer funktionalen Abhängigkeit bekannt sind, so z.B. bei vielen Wetterereignissen, wie Wirbelstürmen oder Hochwasserfluten.

Man versucht sich dieser Unzulänglichkeit dadurch zu entziehen, daß Merkmale dieser Erscheinungen in definierter Art und Weise beobachtet werden. D.h. es werden sogenannte Samples genommen. Diese Samples geben allerdings nur einen kleinen Ausschnitt dessen wieder, was dem Betrachter ansonsten verborgen bleibt. Daher können nun vielfältige mathematische Approximationsverfahren auf diese Beobachtungen angewendet werden, welche es dem Betrachter ermöglichen sollen, einen Eindruck von der Qualität des beobachteten Merkmals an Stellen zu bekommen, die (noch) nicht betrachtet wurden. Ein wichtiges Unterscheidungsmerkmal der verschiedenen Approximationsverfahren ist, ob sie davon ausgehen, ob das zu beobachtende Merkmal eines Phänomens zufällig ist oder nicht. Die sogenannte Geostatistik kann in diesem Zusammenhang als ein Teilbereich der Räumlichen Statistik angesehen werden, welche die beobachteten Merkmale eines bestimmten Phänomens stochastisch-deterministisch modelliert (siehe in [12], [23],[33]). Dabei werden makroskopischen Strukturen, deren Interpretation oftmals durch bekannte geologische Zusammenhänge gestützt werden, deterministisch modelliert, die Mikrostruktur des betreffenden Merkmals dagegen als Realisierung einer zufälligen Funktion. Dabei spiegelt die Zufälligkeit hier das Unwissen über die genauen kausalen Zusammenhänge wider. Die Approximationsverfahren, welche in diesem Zusammenhang meist angewendet werden, sind die sogenannten Kriging-Verfahren. Dabei funktionieren diese Approximationsverfahren gerade so, daß der mittlere quadratische Fehler minimiert wird (siehe auch in [12], [31]). Die deterministischen Verfahren gehen dagegen davon aus, daß sich das betreffende Merkmal des interessierenden Phänomens durch eine deterministische Funktion beschreiben läßt, die jedoch einer gewissen Klasse von Funktionen angehören soll. Die sogenannten Spline-Verfahren sind die wichtigsten Vertreter diese Approximationsart. Sie haben ihre Wurzeln u.a. in der Computergraphik, wo mit ihrer Hilfe eine einfache und effektive Codierung gekrümmter Flächen im Raum möglich ist, wie es in vielen CAD-Anwendungen erforderlich ist. Insbesondere WAHBA beschreibt in [60] die vielfältigen Möglichkeiten von Spline-Verfahren für die Approximation. Schließlich gibt es die Regressionsverfahren, die zwar annehmen, daß sich das zu untersuchende Merkmal des Phänomens durch eine deterministische Funktion beschrei-

ben läßt, die Samples jedoch durch zufällige Meßfehler gestört werden. Zur Konstruktion der Approximation werden die Samples gewichtet gemittelt, wobei die Gewichtung im Falle der lokalen Regression lokal abhängig ist vom Vorhersageort. Dies führt zu sehr einfach realisierbaren Approximationen, die in vielen Bereichen der Natur-, Umwelt- und Geowissenschaften häufig angewendet werden (siehe z.B. in [11] für eine Analyse des Zusammenhangs zwischen Ozon-Konzentration und Windgeschwindigkeit in New York).

Die Wahl einer geeigneten Approximationsmethode ist von großer Wichtigkeit für die Qualität, worunter auch z.B. die Interpretierbarkeit der Approximation zu zählen ist. Damit jedoch eine gewählte Approximationsmethode effektiv nutzbar ist, muß zuvor eine geeignete Planung der Sample-Auswahl erfolgen, die sogenannte Versuchsplanung. Während Erkundungskampagnen in der Praxis in aller Regel von einer endlichen Sample-Anzahl ausgehen, beschäftigt sich der Großteil der mathematisch analytisch motivierten Arbeiten mit dem Fall, daß die Anzahl der Samples im Grenzfall gegen unendlich geht, der sogenannten asymptotischen Versuchsplanung. Die fundamentalen Arbeiten von SACKS/YLVISAKER in [49, 50, 51] sind als Grundstein dieses Ansatzes zu sehen. Die Wichtigkeit einer sinnvollen Versuchsplanung mit endlich vielen Samples wird am Beispiel der Exploration von Offshore-Ölfeldern deutlich. Dort können, aufgrund der damit verbundenen hohen Kosten, oft nur sehr wenige Bohrungen vorgenommen werden, die obendrein nach der Erkundungsphase oftmals als Förderquellen dienen sollen. Versuchsplanungen dieser Art werden zu meist direkt mittels numerischer Optimierungsverfahren behandelt, was aus mathematisch-analytischer Sicht oft unbefriedigend erscheint.

Eine wichtige Alternative zur direkten Beobachtung des interessierenden Merkmals des Phänomens selbst, stellt das Betrachten von relativen Daten dar. Im Rahmen des sogenannten Gradientenkrigings für die Erkundung von Lagerstättenparametern wurden in [32] erfolgreich einfach zu beschaffende Ableitungs-Samples verwendet. Prinzipiell ist dieses Verfahren ein Spezialfall des Co-Kriging, wo die Approximation eines bestimmten Merkmals unter Zuhilfenahme gewisser, mit dem interessierenden Merkmal korrelierter Samples anderer Merkmale erfolgt. In dem Bereich der analytischen Versuchsplanung für Approximationen mit derartigen Ableitungs-Samples mit insgesamt endlich vielen Samples, existieren jedoch praktisch keine Arbeiten. Ebenso ist der Stellenwert von Ableitungs- im Vergleich zu Funktionswert-Samples weitestgehend unbekannt.

Die vorliegende Arbeit stellt sich der genannten Herausforderung, die Rolle von Ableitungs-Samples in der Approximationstheorie sowie in der Versuchsplanung für Approximationen mit Einbeziehung von Ableitungen zu untersuchen. Dabei werden sowohl Kriging-, Spline- als auch Regressionsverfahren betrachtet. Ein Schwerpunkt der Arbeit liegt auf einer möglichst vielseitigen Darstellung von möglichen Einordnungen dieses Problems in verschiedene Kontexte. Um so wichtiger ist damit das einleitende Kapitel 2, in dem der Begriff der zufälligen Funktion sowie deren wesentlichen Eigenschaften definiert werden. Die Auswertung von Krigingverfahren basiert auf dem mittleren quadratischen Fehler. Dieser hängt ausschließlich von Größen maximal zweiter Ordnung ab, also insbesondere der Kovarianzfunktion. Diese wird im Verlauf der Arbeit, soweit relevant, stets als bekannt vorausgesetzt. Sie hat allerdings einen wichtigen Einfluß auf die Qualität von zugehörigen Krigingverfahren, wie von STEIN in [56] hervorgehoben wird. Daher sollen im Abschnitt 2.2 wesentliche Eigenschaften von Kovarianzfunktionen dargestellt werden. Desweiteren werden Bedingungen an die Kovarianzfunktion für die Stetigkeit, Differenzierbarkeit und

Integrierbarkeit von zufälligen Funktionen erläutert. Abschließend wird die sehr vielseitige Klasse der Matérnschen Kovarianzfunktionen vorgestellt, bei der insbesondere die Differenzierbarkeit durch einen stetigen Parameter gesteuert werden kann (siehe in [30], [56]).

In Kapitel 3 wird dann zunächst der Begriff der Approximation, also insbesondere der Vorhersage und der Rekonstruktion, allgemein eingeführt und anschließend eine kurze Einführung in die Theorie der Kernreproduzierenden Hilberträume gegeben. Diese stellen ein sehr hilfreiches Mittel zur Darstellung und Lösung der nachfolgend beschriebenen Kriging- und Spline-Verfahren dar, was zunächst anhand des Simple und Universal Krigings gezeigt wird. Der Schwerpunkt liegt dabei auf einer allgemein gehaltenen Darstellung der Verfahren, zum einen, um die diversen Sonderfälle des Kriging einheitlich wiederzugeben und zum anderen, um die nachfolgend bewiesene Äquivalenz zu gewissen Spline-Verfahren zu unterstreichen. Der Abschnitt 3.4 behandelt erstmals die Versuchsplanung, die zunächst auch allgemein definiert wird. Schwerpunktmäßig wird dort die asymptotische Versuchsplanung vorgestellt, deren mathematisch analytische Behandlung im wesentlichen auf drei Regularitätsbedingungen für die Kovarianzfunktion fußt, welche auf SACKS/YLVISAKER zurückgehen (siehe in [49, 50, 51]). In [44] und [46] werden diese Bedingungen für zufällige Funktionen verallgemeinert, die bis zu einem Grad r differenzierbar sind. Für eine Teilmenge der Klasse der Matérnschen Kovarianzfunktionen kann vom Verfasser gezeigt werden, daß die ersten beiden dieser modifizierten Regularitätsbedingungen erfüllt werden. Alsdann werden einige Ergebnisse zur asymptotischen Versuchsplanung mit hermiteschen Daten vorgestellt, welche von RITTER in [44] bewiesen worden sind. Insbesondere ein adaptives Verfahren zur Rekonstruktion zufälliger Funktionen mit lokal variabler Regularität wird in einer durch den Verfasser der vorliegenden Arbeit verallgemeinerten Form dargestellt. Zum Abschluß des Abschnitts 3.3 wird ein Verfahren zur Versuchsplanung entwickelt, das auf der Basis endlich vieler, nichthermitescher Daten für den Fall asymptotisch verschwindender Korrelation funktioniert. Ein wichtiges Merkmal dieser Versuchsplanung ist insbesondere, daß es im Unterschied zu den zuvor dargestellten asymptotischen Verfahren keinerlei Beschränkung bezüglich der Dimensionalität der Definitionsmenge der betrachteten zufälligen Funktion gibt. Wie dort gezeigt wird, führt das Kriterium auf eine rein geometrische Versuchsplanungsdirektive. Der Themenkomplex Splinemethode wird im letzten Abschnitt von Kapitel 3 behandelt. Nach einer Einführung und Definition beweist der Autor eine allgemeine Äquivalenz zwischen Spline- und den zuvor eingeführten Kriging-Methoden. Zwar wurde die Optimalität von Spline-Methoden bereits von RITTER in [44] behandelt, doch fehlt eine Darstellung des mathematischen Gerüsts, welches notwendig ist, um insbesondere die besagte Äquivalenz zu beweisen. Abschließend wird als Beispiel für die Verwendung von Spline-Verfahren bei zufälligen Funktionen ein einfaches Approximationsverfahren vorgestellt, daß einfach integrierte Brownsche Bewegungen rekonstruiert, wobei praktisch ausschließlich Ableitungs-Samples verwendet werden.

In Kapitel 4 werden die relativen Auswirkungen auf die Varianz eines Regressionsschätzers bei Verwendung eines Ableitungs-Samples im Vergleich zur Verwendung einer weiteren Funktionsmessung untersucht. Zunächst wird auch hier eine kurze Einführung in die Funktionsweise von Regressionsschätzern gegeben. Die oben genannten Untersuchungen basieren auf der Möglichkeit, den Varianzterm bei der klassischen Regression in zwei Teile additiv zu zerlegen, wobei ein Teil exakt der Varianz des Schätzers ohne die Verwendung des neu auszuwählenden Samples ist. Diese Zerlegung ist auch für die sogenannten lokalen Regres-

sionsverfahren möglich, welche im letzten Abschnitt des Kapitels dargestellt werden. Im Unterschied zu den üblichen Regressionsverfahren gehen sie von einer lokal vom Schätzzort abhängigen Gewichtung der Samples aus. Insbesondere die lokale polynomiale Regression (auch local response genannt) erfreut sich starker Beliebtheit in vielen Wissenschaftsbereichen (siehe dazu in [11], [29]). Wichtigstes Ergebnis der Untersuchung ist die Feststellung, daß die Verwendung von Ableitungs-Samples im Rahmen der Regression durchaus vorteilhafter in bezug auf die Varianz des resultierenden Schätzers sein kann als die Verwendung von Funktionswertmessungen.

Im letzten Kapitel stellt der Autor ein Verfahren vor, daß den Titel Versuchsplanung für die beste lineare Vorhersage (engl.: Best Linear Predictor, BLP) mittels gestörter Kovarianzfunktionen trägt. Die zugrundeliegende Idee ist, daß mittels einer kontrollierten Störung der Kovarianzmatrix durch das Hinzufügen einer Diagonalmatrix, der ursprüngliche mittlere quadratische Fehler des BLP differenzierbar modifiziert wird, wobei die Differenzierbarkeit bezüglich eines sogenannten Designmaßes, also einer Gewichtsfunktion für die Samples zu verstehen ist. Diese Art der Modifizierung wird von MÜLLER in [37, 41] vorgestellt. Sie erlaubt ein ähnliches Vorgehen in der Versuchsplanung wie dies bei der klassischen Versuchsplanung mit unkorrelierten Samples der Fall ist. Es wird exemplarisch ein Bewertungsalgorithmus vorgestellt, der eine Bewertung von Samples bei Eintreten infinitesimaler Meßfehler vornimmt.

Trotz oder vielleicht gerade wegen des Versuchs, in der hier vorgestellten Arbeit, einen möglichst vielseitigen Einblick in die Thematik zu geben, bleiben einige Fragen offen. In Kapitel 4 wird die relative Auswirkung eines Ableitungs-Samples auf die Varianz eines Schätzers untersucht. Ein analoges Vorgehen für die Vorhersage ist jedoch ein offenes Problem, das insbesondere an der Komplexität des mittleren quadratischen Fehlers in diesem Fall liegt. Die Anwendung des in Kapitel 5 vorgestellten Regularisierungsverfahrens auf den mittleren quadratischen Fehler der besten linearen erwartungstreuen Vorhersage (engl.: best linear unbiased predictor, BLUP) könnte ein erster Schritt in diese Richtung sein.

2. Definitionen und wesentliche Begriffe

2.1. Zufällige Funktionen

Zufällige Experimente, deren Ergebnisse Funktionen sind, werden aus der Sicht der Modellierung aus zweierlei Gründen verwendet. Die gleichsam kanonische Motivation ist in den Fällen gegeben, wo das beobachtete Phänomen tatsächlich einem Zufall unterworfen ist, der Zufall also systembedingt ist. Zeitreihen, z.B. Börsenkurse, sind ein Beispiel hierfür. Dort ist der Wert eines Finanzguts zu einem zukünftigen Zeitpunkt de facto unbekannt, und man kann nur versuchen, durch geeignete Modelle dessen Entwicklung in der Zukunft zu beschreiben. Derartige Modelle versuchen im allgemeinen, funktionale Abhängigkeiten von Ausgangsgrößen auf die beobachtete Größe abstrakt widerzuspiegeln (siehe z.B. in [9], [42]).

Eine andere Situation liegt vor, wenn die betrachtete Funktion prinzipiell nicht-zufällig ist, jedoch nur relativ wenig Daten im Rahmen eines Versuches für eine weitere Analyse zur Verfügung stehen. In diesem Zusammenhang verwendet man das Hilfsmittel der zufälligen Funktionen, um die versuchsbedingte Unwissenheit über das System zu modellieren. Dieser Ansatz wird häufig in geostatistischen Anwendungen und neuerdings auch bei der Modellierung von komplexen Computersystemen und -programmen benutzt. (siehe z.B. [12], [19], [48], [13])

2.1.1. Definition

Unter einer zufälligen Funktion Z versteht man eine meßbare Abbildung

$$Z : \Omega \times D \rightarrow \mathbb{C}, \quad D \subset \mathbb{R}^n \quad (2.1)$$

wobei für alle $t \in D$ die Funktionen $Z_\omega(t) := Z(\omega, t)$ jeweils Zufallsvariablen in ω auf einem abstrakten Wahrscheinlichkeitsraum (Ω, \mathcal{A}, P) seien. Für feste $\omega \in \Omega$ bezeichnet man die Funktion $z(\cdot) := Z_\omega(\cdot)$ auch als Realisation oder Pfad dieser zufälligen Funktion. Die zu einer zufälligen Funktion gehörigen Realisationen sollen fortan stets mit Kleinbuchstaben bezeichnet werden. Schließlich sei P_Z das Bildmaß von P unter der Abbildung $\omega \rightarrow Z_\omega(\cdot)$.

Vielfach ist die Zerlegung von Z in Real- und Imaginärteil sinnvoll,

$$Z_\omega(t) = U_\omega(t) + iV_\omega(t), \quad \text{für alle } t \in D, \omega \in \Omega.$$

Dabei seien U und V reellwertige zufällige Funktionen. In sehr vielen praktischen Anwendungsgebieten kann man von einer reellwertigen zufälligen Funktion ausgehen, es gilt dann $V \equiv 0$.

Für den Fall $n = 1$ spricht man bei Z auch von einem stochastischen Prozeß, insbesondere Zeitreihen sind Beispiele hierfür. In den Fällen $n = 2, 3$ findet man die meisten Anwendungen aus dem Bereich der Umweltwissenschaften wieder (siehe z.B. in [18], [17]). Hier werden z.B. Dichten von Erdschichten in einer bestimmten Tiefe, über der Ebene verteilt beobachtet oder Temperaturen im Raum. In den Bereichen Computerprogramm- und Computersystemsimulation findet man dagegen Beispiele, wo der Parameterraum wesentlich höhere Dimensionen hat ($n \gg 3$) (siehe z.B. in [13], [48]).

Die Beschränkung des Definitionsbereichs auf eine Teilmenge D des \mathbb{R}^n ist in der Regel durch die realen Vorgaben der zu modellierenden Variable gegeben. So wird eine reale Größe oftmals nur auf einem Teilbereich des \mathbb{R}^n beobachtbar bzw. überhaupt definiert sein.

Formal ist eine zufällige Funktion eine indizierte Menge von Zufallsvariablen. Die Indexmenge D wurde hier als Teilmenge des \mathbb{R}^n gewählt, kann jedoch auch eine beliebige Menge sein (siehe z.B. in [14] für eine allgemeine Definition und in [13] für ein Beispiel aus der Simulation von Computerprogrammen). Nach dem Satz von Kolmogorov ist die Verteilung einer zufälligen Funktion vollständig durch deren endlichdimensionale Verteilungen beschrieben, d.h. es reicht aus, Wahrscheinlichkeiten von Mengen aus Ω der Form

$$\{(Z(t_1), \dots, Z(t_l)) \in A\} \text{ mit } (t_1, \dots, t_l) \in D^l, A \in \mathcal{B}^l \quad (2.2)$$

für beliebige $l \in \mathbb{N}$ zu kennen (siehe z.B. in [8]). \mathcal{B}^l sei dabei die Borel- σ -Algebra des \mathbb{C}^l .

2.1.2. Momente von zufälligen Funktionen

In der Regel wird man die Verteilung einer zufälligen Funktion jedoch nicht kennen, denn dies würde voraussetzen, daß man die Wahrscheinlichkeiten von allen Mengen der Form in (2.2) bestimmen könnte. Wie auch in der klassischen Statistik beschränkt man sich daher auf die Schätzung weniger Kenngrößen der zufälligen Funktion. Die einfachste ist die Erwartungswertfunktion:

$$m(t) := E Z(t) = \int_{\Omega} Z(\omega, t) P(d\omega), \text{ mit } t \in D \quad (2.3)$$

Der Mittelwert beschreibt eine systematische Veränderung der Größe $Z(t)$, den sogenannten Trend.

Das Hauptmerkmal von zufälligen Funktion, sprich die zufälligen Veränderungen, werden mit der Erwartungswertfunktion nicht erfaßt. Die einfachste Kenngröße, um diese Variation zu beschreiben, ist das zweite zentrierte Moment, die sogenannte Kovarianzfunktion,

$$K(t, s) := E (Z(t) - E Z(t)) \overline{(Z(s) - E Z(s))}, \text{ mit } t, s \in D. \quad (2.4)$$

Ein Überstrich über einer Größe bedeute hier die Bildung des komplex Konjugierten. Mit $\text{Var}(Z(t)) = K(t, t)$ erhält man die Varianzfunktion der zufälligen Funktion Z . Es sei betont, daß im Verlauf der Arbeit stets davon ausgegangen wird, daß die Kovarianzfunktion der zufälligen Funktion bekannt, sowie endlich auf dem Definitionsgebiet ist. In einem praktischen Umfeld ist natürlich eine der ersten Aufgaben im Zusammenhang mit der Analyse

einer zufälligen Funktion die Schätzung der zugrundeliegenden Kovarianzfunktion. Hierauf beruht das weitere Vorgehen, wie die Vorhersage oder Regression und schließlich auch die Versuchsplanung. Der Teil der Theorie der zufälligen Funktionen, welcher nur auf die Größen maximal zweiter Ordnung zurückgreift, wird auch Theorie der Prozesse zweiter Ordnung oder Korrelationstheorie genannt (s. [61]).

Die sogenannten Gaußschen Prozesse bilden eine wichtige Prozeßklasse, sie sind durch die ersten beiden Momente bereits vollständig beschrieben. Ein Gaußscher Prozeß liegt dann vor, wenn alle endlichdimensionalen Verteilungen multivariat normalverteilt sind. Mit Hilfe einer beliebigen Kovarianzfunktion K und einer Erwartungswertfunktion m läßt sich stets ein Gaußscher Prozeß konstruieren, der gerade diese beiden Funktionen als Kovarianz- bzw. Erwartungswertfunktion hat (siehe in [14]). D.h. wenn die Diskussion über eine zufällige Funktion nur die ersten beiden Momente einschließt, so stellt es keine Einschränkung dar, allein Gaußsche Prozesse zu betrachten.

Ein wichtiges Beispiel für einen Gaußschen Prozeß ist die sogenannte Brownsche Bewegung B auf dem Einheitsintervall $[0, 1]$ (siehe z.B. in [52]). B hat eine Erwartungswertfunktion konstant Null, fängt P -fast-sicher in 0 an und hat die Kovarianzfunktion $K(t, s) = \min(t, s)$.

2.1.3. Stationarität und Isotropie

In vielen Anwendungen wird man nur eine Realisierung einer zufälligen Funktion Z zur Verfügung haben, z.B. bei Probebohrungen im Bergbaubereich. Aus diesen Daten läßt sich jedoch, wenn keine Einschränkungen an die Verteilung der zufälligen Funktion gemacht werden, im allgemeinen die Kovarianzfunktion nicht schätzen. Eine hinreichende Forderung, um dies zu ermöglichen, ist, daß die Verteilung translationsinvariant ist. Dies setzt notwendig voraus, daß der Definitionsbereich D der ganze \mathbb{R}^n ist. Sei $\{T_\tau, \tau \in \mathbb{R}^n\}$ die Gruppe der Translationen in \mathbb{R}^n , d.h. es gilt $T_\tau t = t + \tau$ für beliebige t und τ .

Definition 2.1. *Eine zufällige Funktion Z heißt stationär, falls deren Verteilung invariant gegenüber Translationen von Z ist, d.h. es gilt*

$$Z(T_\tau(\cdot)) \stackrel{d}{=} Z(\cdot).$$

Dabei bedeutet das Zeichen $\stackrel{d}{=}$, daß die Verteilungen der Größen auf den beiden Seiten der Relation gleich sind. Insbesondere ist in diesem Fall die Erwartungswertfunktion eine konstante Funktion und wird im folgenden soweit nicht anders vermerkt o.B.d.A. als die Nullfunktion angenommen. Die Kovarianzfunktion hängt von ihren Argumenten nur durch deren Differenz ab, d.h. es gilt

$$K(t, s) = K(t - s), \text{ mit } t, s \in \mathbb{R}^n.$$

Im eindimensionalen Fall gilt für stationäre Prozesse ein starkes Gesetz der großen Zahl (siehe [14]). Falls Z ein meßbarer stochastischer Prozeß ist, bei dem $m := E |Z(t_0)| < \infty$ für ein beliebiges $t_0 \in \mathbb{R}$ vorliegt, so gilt:

$$\lim_{s \rightarrow \infty} \frac{1}{s} \int_0^s Z(t) dt = E(Z(t_0) | \mathcal{J}) \quad P\text{-f.s.},$$

wobei \mathcal{J} die σ -Algebra der invarianten Mengen ist. Mittelwertbildung von Funktionswerten führt also zu einem konsistenten Schätzer für den Mittelwert von Z , falls die Indizes (sprich z.B. die Orte) der Daten in der Grenze den ganzen Definitionsbereich überdecken. Ein stochastischer Prozeß ist dann meßbar, falls der Definitionsbereich D Lebesgue-meßbar ist sowie Z meßbare Funktion auf dem Produktraum $\Omega \times D$ ist (siehe [14]).

Werden nur Prozesse zweiter Ordnung betrachtet, so reichen Homogenitätsforderungen an die ersten beiden Momente bereits aus.

Definition 2.2. *Eine zufällige Funktion Z heißt schwach stationär, falls deren beiden ersten Momente translationsinvariant sind.*

Jede stationäre zufällige Funktion ist somit sofort schwach stationär, und jeder Gaußsche Prozeß, der schwach stationär ist, ist auch stationär. Letzteres folgt daraus, daß Gaußsche Prozesse bereits vollständig durch ihre ersten beiden Momente bestimmt sind.

Beschreibt der Begriff der Stationarität die Invarianz der Verteilung einer zufälligen Funktion gegenüber Translationen, ist es zusätzlich möglich, eine Invarianz gegenüber beliebigen Drehungen zu fordern. Sei \mathbf{H} eine orthogonale $n \times n$ Matrix, sprich eine Drehmatrix.

Definition 2.3. *1. Eine zufällige stationäre Funktion Z heißt isotrop, falls deren Verteilung invariant gegenüber Drehungen von Z ist, d.h. es gilt:*

$$Z(\mathbf{T}_\tau \circ \mathbf{H}(\cdot)) \stackrel{d}{=} Z(\cdot)$$

für beliebige $\tau \in \mathbb{R}^n$ und beliebige Drehmatrizen \mathbf{H} .

2. Eine zufällige stationäre Funktion Z heißt schwach isotrop, falls deren beiden ersten Momente invariant gegenüber Drehungen sind.

Die Kovarianzfunktion hängt im Fall von (schwacher) Isotropie nur vom Abstand ihrer Argumente ab, so daß

$$K(\mathbf{t}, \mathbf{s}) = K(r) \text{ mit } r = \|\mathbf{s} - \mathbf{t}\|$$

für alle $\mathbf{s}, \mathbf{t} \in \mathbb{R}^n$ gilt. Die Schreibweise $\|\cdot\|$ bezeichne dabei den euklidischen Abstand im \mathbb{R}^n .

2.2. Kovarianzfunktionen

In der Theorie der Prozesse zweiter Ordnung nimmt die Kovarianzfunktion eine zentrale Rolle ein. Wie bereits im ersten Abschnitt betont wurde, soll in der vorliegenden Arbeit die Kovarianzstruktur als bekannt vorausgesetzt werden. Dies ist eine notwendige Voraussetzung, um L_2 -Vorhersagen bzw. darauf aufbauend, Versuchsplanung für Vorhersageexperimente durchführen zu können. In Anwendungen muß also zunächst die Kovarianzfunktion ermittelt werden, bzw. sie muß aus den zur Verfügung stehenden Daten geschätzt werden. Im Falle von parametrischen Schätzverfahren setzt dies voraus, daß auf Klassen von Kovarianzfunktionen zurückgegriffen werden kann, welche wichtige analytische Eigenschaften

des zugrundeliegenden Feldes wiedergeben können. Insbesondere die Differenzierbarkeit der Kovarianzfunktion hat einen sehr starken Einfluß auf das Vorhersageergebnis. In [56] findet man zahlreiche Beispiele dafür, daß eine Schätzung unter Annahme der sogenannten Gaußschen Kovarianzfunktion, welche unendlich oft differenzierbar ist, stets auf zu optimistischen Vorhersagen hinausläuft. Dort wird die sogenannte Matérnsche Kovarianzklasse propagiert, bei der die Differenzierbarkeit mittels eines kontinuierlichen Parameters gesteuert werden kann.

2.2.1. Wesentliche Eigenschaften

Beschreibt der Trend einer zufälligen Funktion im wesentlichen deren Verhalten in globaleren Maßstäben, so ist die Kovarianzfunktion eine Kenngröße, welche deren lokales Verhalten kennzeichnet.

Sei K die Kovarianzfunktion einer zufälligen Funktion, so ist K positiv semidefinit. D.h. es gilt für beliebige $l \in \mathbb{N}$, sowie beliebige reelle Faktoren c_1, \dots, c_l und Orte $t_1, \dots, t_l \in D^l$

$$\sum_{i,j=1}^l c_i c_j K(t_i, t_j) \geq 0. \quad (2.5)$$

Dies folgt daraus, daß $\sum_{i,j=1}^l c_i c_j K(t_i, t_j) = \text{Var} \sum_{j=1}^l c_j Z(t_j)$ ist. Allgemein läßt sich zu einer beliebigen positiv semidefiniten Funktion K auf $\mathbb{R}^n \times \mathbb{R}^n$ und einer ebenso beliebigen Funktion m auf \mathbb{R}^n ein Gaußscher Prozeß konstruieren, welcher diese Funktionen als Kovarianz- bzw. Erwartungswertfunktion hat (siehe z.B. in [56]). Desweiteren ist K definitionsbedingt symmetrisch in ihren Argumenten, und es gilt $K(t, t) \geq 0$ für alle $t \in D$.

Für den Fall, daß Z eine schwach stationäre Funktion mit Kovarianzfunktion K ist, erhalten die letztgenannten Eigenschaften die folgende Form. Es sei dabei $D^* = \{l \in \mathbb{R}^n \mid l = s - t \text{ mit } s, t \in D\}$.

$$K(0) \geq 0 \quad (2.6)$$

$$K(t) = \overline{K(-t)}, \quad t \in D^* \quad (2.7)$$

$$|K(t)| \leq K(0), \quad t \in D^* \quad (2.8)$$

Die Ungleichung in (2.8) ist dabei ein Spezialfall der allgemeineren Ungleichung $|K(t, s)| \leq (K(t, t)K(s, s))^{1/2}$, welche aus der Cauchy-Schwarzschen Ungleichung folgt und für allgemeine Kovarianzfunktionen gilt. Die oben genannte Forderung nach Endlichkeit der Kovarianzfunktion auf ihrem Definitionsgebiet ist also notwendig erfüllt, falls für alle $t \in D$ gilt, daß $\text{Var} Z(t) < \infty$.

2.2.2. Hilberträume

Aus formal mathematischer Sicht läßt sich das Konzept von zufälligen Funktionen mit zugehörigen Kovarianzfunktionen wesentlich allgemeiner darstellen. Es wird im folgenden die in [56] eingeführte Notation verwendet.

Sei Z eine zufällige Funktion auf $D \subset \mathbb{R}^n$ mit Kovarianzfunktion K und Erwartungswertfunktion m . Definiere den komplexen Vektorraum $\mathcal{H}_D^0 = \text{span}\{Z(t), t \in D\}$, der durch Z implizierten Zufallsvariablen im Raum aller komplexwertigen Zufallsvariablen auf dem gegebenen Wahrscheinlichkeitsraum (Ω, \mathcal{A}, P) .

Die Vorhersage mit kleinsten mittleren Fehlerquadraten setzt voraus, daß ein Abstand zwischen zwei zufälligen Größen in diesem Raum bestimmt werden kann. Da häufig nur wenige oder gar eine Realisierung einer zufälligen Funktion zur Verfügung steht, wird ein sinnvoller Abstandsbegriff von der Verteilung der betrachteten zufälligen Funktion abhängen. Im Kontext der Prozesse zweiter Ordnung ist die Kovarianzfunktion die stärkste Beschreibung der zugrundeliegenden Verteilung, und so erhält man mit dem sogenannten L_2 -Produkt, welches für beliebige $X, Y \in \mathcal{H}_D^0$ wie $\langle X, Y \rangle := E(\overline{XY})$ definiert ist, die Norm (Abstand) $\|X\| := \sqrt{\langle X, X \rangle}$. Für ein beliebiges $t \in D$ ist die Norm von $Z(t)$, also der Abstand zu der Zufallsvariable, welche P-f.s. konstant Null ist, gerade $\sqrt{K(t, t) + |m(t)|^2}$. Falls $m \equiv 0$ ist, dann ist der Abstand von $Z(t)$ zur Nullfunktion um so größer, je stärker deren Varianz ist. $\langle \cdot, \cdot \rangle$ erfüllt die folgenden drei Eigenschaften und ist somit ein inneres Produkt.

1. $\langle X, X \rangle \geq 0$ wobei die Gleichheit genau dann gilt, wenn $X \equiv 0$ ist.
2. $\langle X, Y \rangle = \overline{\langle Y, X \rangle}$ für alle $X, Y \in \mathcal{H}_D^0$
3. $\langle aX + bY, Z \rangle = a\langle X, Z \rangle + b\langle Y, Z \rangle$ für alle $X, Y, Z \in \mathcal{H}_D^0$ und $a, b \in \mathbb{C}$

Das Tupel $(\mathcal{H}_D^0, \langle \cdot, \cdot \rangle)$ ist ein Prähilbertraum. Schließt man diesen Raum bezüglich der oben definierten Norm ab, so erhält man einen Hilbertraum, der mit $\mathcal{H}_D(K, m)$ bezeichnet werden soll. Die wichtigste Eigenschaft eines Hilbertraums ist, daß jede Cauchy-Folge konvergiert. Sei $\{X_n\}$ eine solche Folge in $\mathcal{H}_D(K, m)$, d.h. es gilt $\lim_{n, m \rightarrow \infty} \|X_n - X_m\| = 0$. Dann existiert ein $X \in \mathcal{H}_D(K)$ derart, daß $\lim_{n \rightarrow \infty} \|X_n - X\| = 0$.

2.2.3. Stetigkeit, Differenzierbarkeit und Integrierbarkeit im quadratischen Mittel

Die vermutlich anschaulichste Ausprägung des Begriffes Stetigkeit, die P-fast sichere Stetigkeit von Realisierungen allgemeiner zufälliger Funktionen, ist im Kontext der Prozesse zweiter Ordnung nicht praktikabel, da diese Eigenschaft Forderungen an die Verteilung als solche stellt, welche aber i.a. nicht bekannt sein dürfte. Im Fall von zufälligen Prozessen ($n = 1$) kann z.B. gezeigt werden, daß die Brownsche Bewegung P-fast-sicher stetige Realisierungen hat (s. [52]).

Sind allerdings nur die Größen maximal zweiter Ordnung einer zufälligen Funktion bekannt, so ist die Konvergenz im quadratischen Mittel die in diesem Kontext stärkste Formulierung des Konvergenzbegriffs. Sei X_n eine Folge von Zufallsvariablen auf dem Wahrscheinlichkeitsraum (Ω, \mathcal{A}, P) und X eine weitere Zufallsvariable auf demselben Raum, so konvergiert X_n im quadratischen Mittel gegen X , in Zeichen $X_n \xrightarrow{2} X$, falls

$$E(X_n - X)^2 \rightarrow 0, \text{ für } n \rightarrow \infty.$$

Eine zufällige Funktion Z heißt nun stetig im quadratischen Mittel in $t \in D$, falls für beliebige Punktfolgen $\{t_n\} \subset D$, welche in D gegen t konvergieren, gilt:

$$\lim_{n \rightarrow \infty} E(Z(t_n) - Z(t))^2 = 0.$$

Sei Z nun eine schwach stationäre zufällige Funktion mit Kovarianzfunktion K , so gilt $E(Z(t_n) - Z(t))^2 = 2(K(0) - K(t_n - t))$. Z ist also genau dann in t im quadratischen Mittel stetig, falls dessen Kovarianzfunktion in 0 stetig ist. Die Stetigkeit von K im Ursprung ist demnach notwendiges und hinreichendes Kriterium für die Stetigkeit im quadratischen Mittel von Z für alle $t \in D$.

Insbesondere ist die Stetigkeit von K im Ursprung ein hinreichendes Kriterium für dessen Stetigkeit auf ganz D . Sei $t \in D$ beliebig aber fest gewählt und $\{t_n\} \subset D$ eine beliebige Folge, welche gegen t konvergiert, so gilt wegen der Cauchy-Schwarzschen Ungleichung:

$$\begin{aligned} |K(t) - K(t_n)| &= |E(Z(t)\overline{Z(0)}) - E(Z(t_n)\overline{Z(0)})| \\ &= |E(\overline{Z(0)}\{Z(t_n) - Z(t)\})| \\ &\leq (E|Z(0)|^2 E|Z(t_n) - Z(t)|^2)^{\frac{1}{2}} \\ &= (K(0)\{2K(0) - K(t_n - t) - \overline{K(t_n - t)}\})^{\frac{1}{2}} \\ &\rightarrow 0, \text{ für } n \rightarrow \infty. \end{aligned}$$

Differenzierbarkeit

Die Definition der Differenzierbarkeit im quadratischen Mittel orientiert sich an der Definition für deterministische Funktionen, verwendet jedoch wiederum die Konvergenz im quadratischen Mittel. Sei Z zunächst ein stochastischer Prozeß, d.h es ist $D \subset \mathbb{R}$. So heißt Z in $t \in D$ differenzierbar im quadratischen Mittel, falls für beliebige Folgen $\{t_n\} \subset D$, welche gegen t konvergieren,

$$\frac{Z(t_n) - Z(t)}{t_n - t} \xrightarrow{2} Z^{(1)}(t) \tag{2.9}$$

gilt. Die Größe $Z^{(1)}(t)$ wird erste Ableitung im quadratischen Mittel von Z an der Stelle t genannt.

Unter Ausnutzung der in Abschnitt 2.2.2 genannten Hilbertraumeigenschaften läßt sich in bekannter Art und Weise ein hinreichendes Kriterium für die Differenzierbarkeit formulieren (siehe z.B. in [61]). Z sei dazu nun zusätzlich schwach stationär. $t \in D$ sei ein beliebiger aber fester Punkt und $\{h_n\} \subset D$ sei eine positive Nullfolge. Betrachte die Doppelfolge in n

und m

$$\left\| \frac{Z(t + h_n) - Z(t)}{h_n} - \frac{Z(t + h_m) - Z(t)}{h_m} \right\|^2 \quad (2.10)$$

$$= E \left| \frac{Z(t + h_n) - Z(t)}{h_n} - \frac{Z(t + h_m) - Z(t)}{h_m} \right|^2$$

$$= E \left| \frac{Z(t + h_n) - Z(t)}{h_n} \right|^2 + E \left| \frac{Z(t + h_m) - Z(t)}{h_m} \right|^2 \quad (2.11)$$

$$- E \left(\frac{Z(t + h_n) - Z(t)}{h_n} \frac{Z(t + h_m) - Z(t)}{h_m} \right) \quad (2.12)$$

$$- E \left(\frac{Z(t + h_m) - Z(t)}{h_m} \frac{Z(t + h_n) - Z(t)}{h_n} \right). \quad (2.13)$$

Der erste Summand in (2.11) kann umgeschrieben werden in $\Delta^{(2)}(K)_n := (2K(0) - K(-h_n) - K(h_n))/h_n^2$. Unter der Bedingung, daß $K^{(2)}(0)$ existiert, gilt für den Grenzübergang $n \rightarrow \infty$:

$$\Delta^{(2)}(K)_n \rightarrow -K^{(2)}(0).$$

Der zweite Summand verhält sich unter der genannten Voraussetzung analog. Der Ausdruck in (2.12) ergibt $\Delta_*^{(2)}(K)_{n,m} := (K(h_n - h_m) + K(0) - K(h_n) - K(-h_m))/h_n h_m$, und unter der Bedingung, daß $K^{(2)}(0)$ existiert, gilt auch hier für den Grenzübergang $m, n \rightarrow \infty$:

$$\Delta_*^{(2)}(K)_{n,m} \rightarrow -K^{(2)}(0).$$

Zusammengenommen konvergiert die Doppelfolge in (2.10) genau dann gegen Null, wenn $-K^{(2)}(0)$ existiert. Demnach ist $(Z(t + h_n) - Z(t))/h_n$ für beliebige $t \in D$ unter dieser Voraussetzung eine Cauchy-Folge, und somit existiert ein eindeutiger Grenzwert, welcher gerade die erste Ableitung im quadratischen Mittel von $Z(t)$ ist (siehe Abschnitt 2.2.2).

Die Existenz von $K^{(2)}(0)$ ist ein hinreichendes Kriterium für die Existenz der zweiten Ableitung von K an beliebigen Stellen in D^* . Aus dieser Bedingung folgt, wie gerade gezeigt wurde, die Differenzierbarkeit von Z , und die Kovarianzfunktion dieser ersten Ableitung im quadratischen Mittel an einer beliebigen Stelle $s \in D^*$ ist gerade ($t \in D$ beliebig)

$$\lim_{n \rightarrow \infty} E \left(\frac{Z(t + s + h_n) - Z(t + s)}{h_n} \frac{Z(t + h_n) - Z(t)}{h_n} \right)$$

$$= \lim_{n \rightarrow \infty} \left(\frac{2K(s) - K(s - h_n) - K(s + h_n)}{h_n^2} \right) = -K^{(2)}(s).$$

Wie in [14] erwähnt wird, ist dieses Kriterium ebenfalls hinreichend für die P-f.s. pfadweise Stetigkeit des betreffenden Prozesses. Ist Z zusätzlich separabel, so stimmen zusätzlich die pfadweisen Ableitungen von Realisierungen mit den Realisierungen der Ableitung im quadratischen Mittel P-f.s. überein. Die Eigenschaft der Separabilität stellt keine Einschränkung an die endlichdimensionalen Verteilungen einer zufälligen Funktion dar, sondern bezieht sich auf den Übergang hin zu unendlichdimensionalen Verteilungen (siehe [14]). Gaußsche Prozesse z.B. können somit stets als separabel angenommen werden.

Integrierbarkeit

Das Konzept für die Definition von Integralen von zufälligen Funktionen verläuft nach demselben Schema wie bei den Ableitungen von zufälligen Funktionen. Die Konstruktion verwendet als Ansatz die Riemanschen Summen und benutzt die Konvergenz im quadratischen Mittel für die Grenzwertbildung (siehe in [61]).

Der Einfachheit halber sei Z ein schwach stationärer Prozeß mit Kovarianzfunktion K . Weiter sei f eine komplexe Funktion auf D und $a < b$ die Integrationsgrenzen, welche ebenfalls in D liegen. Für die Konstruktion der Riemanschen Summen zerlege das Intervall $[a, b]$ derart in n Teilintervalle $\delta_j := t_j - t_{j-1}$ mit $a = t_1 < t_2 < \dots < t_n = b$, daß $\delta_{\max}(n) := \max\{\delta_j : j = 1, \dots, n\}$ für $n \rightarrow \infty$ gegen Null konvergiert. Damit definiere für den Fall, daß der Grenzwert existiert,

$$\sum_{j=1}^n f(t'_j)Z(t'_j)\delta_j \xrightarrow{2} \int_a^b f(t)Z(t) dt \text{ für } n \rightarrow \infty \quad (2.14)$$

mit $t_{j-1} \leq t'_j \leq t_j$.

Auf analoge Weise wie im Falle der Differenzierbarkeit, zeigt man, daß ein hinreichendes Kriterium zur Existenz des Grenzwertes in (2.14) gerade die Bedingung

$$\int_a^b \int_a^b f(t)f(s)K(t-s) ds dt < \infty \quad (2.15)$$

ist. Aufgrund der Definition des Integrals ist u.a. die Linearität in Z sichergestellt. Damit ist das Doppelintegral in (2.15) die Varianz von $\int_a^b f(t)Z(t) dt$ für den Fall, daß $m \equiv 0$ ist.

Eine wichtige Eigenschaft von meßbaren, zufälligen Funktionen, die im quadratischen Mittel integrierbar sind, ist, daß deren Realisierungen P -f.s. Lebesgue-integrierbar sind und daß die Integrale von Realisierungen der Form $\int_a^b f(t)z(t) dt$ Realisierungen des Integrals $\int_a^b f(t)Z(t) dt$ sind (siehe [14],[61]).

2.2.4. Spektraldarstellungen

Die Fourierentwicklung von deterministischen Funktionen ist eine gängige Methode in der deterministischen Analysis. Allerdings ist ihre Anwendbarkeit auf Funktionen auf \mathbb{R} sehr eingeschränkt, was den Definitionsbereich anbelangt. So müssen Funktionen, die sich Fourier entwickeln lassen, absolut integrierbar sein und ausserdem auf allen endlichen Intervallen des \mathbb{R} die sogenannten Dirichletbedingungen erfüllen. D.h. in je einem endlichen Intervall darf die Funktion nur endlich viele Teilintervalle haben, auf denen sie jeweils stetig und monoton ist und an deren Rändern sie endliche Sprünge hat.

Die Dirichletschen Forderungen sind i.a. von Realisierungen zufälliger Funktionen nicht erfüllt. Z.B. werden bei schwach stationären Prozessen die Realisierungen nicht absolut integrierbar sein, da sie dazu P -fast-sicher. stark gegen Null fallen müßten, was aber der

Stationarität widersprechen würde. Dennoch gibt es eine auf Kolmogorov zurückgehende Theorie über die Spektraldarstellung von schwach stationären zufälligen Funktionen (siehe die Referenzen in [61, 62]).

Sei Z eine beliebige schwach stationäre zufällige Funktion mit Kovarianzfunktion K . So existiert ein zufälliges Maß \mathcal{M} derart, daß folgende Darstellung P -fast-sicher definiert ist:

$$Z(\mathbf{t}) = \int_{\mathbb{R}^n} e^{i\langle \mathbf{v}, \mathbf{t} \rangle} \mathcal{M}(d\mathbf{v}). \quad (2.16)$$

Das Integral ist als Grenzwert im quadratischen Mittel von Riemannschen Summen zu verstehen, wobei im Unterschied zum vorhergehenden Abschnitt hier das Integrations-Maß und nicht der Integrand zufällig ist (siehe auch in [61] und [14]). Das verwendete innere Produkt sei das euklidische. \mathcal{M} ist ein komplexwertiges, orthogonales Maß auf \mathcal{B}^n , der Borel- σ -Algebra des \mathbb{R}^n , und es gelten folgende Eigenschaften:

$$\begin{aligned} E \mathcal{M}(\Delta) &= 0, \text{ für alle } \Delta \in \mathcal{B}^n \\ \mathcal{M}(\Delta) &= \overline{\mathcal{M}(-\Delta)}, \text{ für alle } \Delta \in \mathcal{B}^n \\ \mathcal{M}\left(\bigcup_{i=1}^n \Delta_i\right) &= \sum_{i=1}^n \mathcal{M}(\Delta_i), \text{ für alle } \Delta_i \in \mathcal{B}^n \text{ paarw. disjunkt.} \end{aligned}$$

Eine pfadweise Interpretation des Integrals in (2.16) ist, im Gegensatz zu der Situation, wie sie im vorherigen Abschnitt vorlag, nicht möglich (s. [61]). Allerdings gewinnt die Theorie an Stärke, wenn nicht das zufällige Maß \mathcal{M} direkt sondern dessen quadratisches Mittel betrachtet wird.

Satz 2.4 (von Bochner). *Eine komplexwertige Funktion K auf \mathbb{R}^n ist genau dann Kovarianzfunktion eines schwach stationären zufälligen Feldes auf \mathbb{R}^n , falls es ein positives endliches Maß \mathcal{F} auf $(\mathbb{R}^n, \mathcal{B}^n)$ gibt mit*

$$K(\mathbf{t}) = \int_{\mathbb{R}^n} e^{i\langle \mathbf{t}, \mathbf{v} \rangle} \mathcal{F}(d\mathbf{v}).$$

\mathcal{F} heißt *Spektralmaß* von Z .

\mathcal{F} erfüllt die folgenden Eigenschaften:

$$\mathcal{F}(\Delta) = E |\mathcal{M}(\Delta)|^2, \text{ für alle } \Delta \in \mathcal{B}^n \quad (2.17)$$

$$E \mathcal{M}(\Delta_1) \overline{\mathcal{M}(\Delta_2)} = \mathcal{F}(\Delta_1 \cap \Delta_2) \text{ Orthogonalität.} \quad (2.18)$$

Ist \mathcal{F} absolutstetig bezüglich des Lebesguemaßes, so existiert eine eindeutige Spektraldichtefunktion f , so daß $K(\mathbf{t}) = \int_{\mathbb{R}^n} e^{i\langle \mathbf{t}, \mathbf{v} \rangle} f(\mathbf{v}) d\mathbf{v}$ erfüllt ist. Für isotrope zufällige Funktionen gilt, sofern die Dichte f existiert, die folgende Umkehrformel (s. [61]):

$$f(\mathbf{v}) = \frac{1}{(2\pi)^{n/2}} \int_0^\infty \frac{J_{(n-2)/2}(\mathbf{v}r)}{(\mathbf{v}r)^{(n-2)/2}} r^{n-1} K(r) dr.$$

Dabei ist J_l die Besselfunktion der Ordnung l (siehe in [1]). Insbesondere ist jede positive, integrierbare Funktion isotrope Spektraldichte einer zufälligen, isotropen Funktion. Es folgen zwei Beispiele für isotrope Kovarianzfunktionen bzw. Spektraldichten.

Beispiel 2.5. 1. Gaußsche Kovarianzfunktion:

$$\begin{aligned} f(\omega) &= \frac{c}{2^n (\pi\alpha)^{n/2}} e^{-\frac{\omega^2}{4\alpha}} \\ K(r) &= ce^{-\alpha r^2}, \quad r \in \mathbb{R}, c > 0, \alpha > 0. \end{aligned}$$

Diese Kovarianzfunktion ist unendlich oft differenzierbar, was auch aus der exponentiell abfallenden Spektraldichte gefolgert werden kann. Dieses Verhalten ist der Grund für eine kritische Bewertung in STEIN (1999). Dort wird u.a. bemängelt, daß bei Verwendung dieser Kovarianzfunktion, etwaige Vorhersagen zu optimistisch bewertet werden.

2. Matérnsche Kovarianzklasse (siehe in [20], [30], [56]): $\eta = (\sigma^2, \nu, \rho) \in \mathbb{R}_+^3 \setminus \{0\}$

$$\begin{aligned} f_\eta(\omega) &= \frac{\sigma^2 c(\nu, \rho)}{\left(\frac{4\nu}{\rho} + \omega^2\right)^{(\nu+1/2)}} \\ K_\eta(r) &= \frac{\sigma^2}{2^{\nu-1} \Gamma(\nu)} \left(\frac{2\sqrt{\nu}r}{\rho}\right)^\nu \mathcal{K}_\nu\left(\frac{2\sqrt{\nu}r}{\rho}\right). \end{aligned}$$

\mathcal{K}_ν ist die modifizierte Besselfunktion der Ordnung ν (siehe in [1]). Wichtigster Parameter ist ν , das Maß der Differenzierbarkeit, Z ist genau m -mal differenzierbar, falls $m < \nu$ und $m+1 > \nu$ ist. Desweiteren steuert ρ die Reichweite der Kovarianzfunktion, und σ^2 ist der Wert von K_η an der Stelle Null.

3. Lineare Rekonstruktion und Vorhersage

Der Ausgangspunkt des linearen Rekonstruktions- bzw. Vorhersageproblems ist ein unbekannter beschränkter Operator S auf einer Menge reellwertiger Funktionen, der mittels einer Linearkombination einfacher linearer Operatoren $\lambda_1, \dots, \lambda_n$, der sogenannten Samplefunktionale, approximiert werden soll. Prinzipiell sind wenigstens zwei grundsätzlich verschiedene Verfahren möglich. Die deterministischen Verfahren, die annehmen, daß S auf einer Menge von reellwertigen Funktionen operiert, und stochastische Verfahren, die auf der Annahmen beruhen, daß eben diese Menge einer gewissen Wahrscheinlichkeitsverteilung unterworfen ist. Zufällige Funktionen seien demnach fortan stets als reellwertig angenommen.

Im folgenden Abschnitt wird zunächst der Begriff der linearen Probleme eingeführt, dabei wird die in [44] eingeführte Notation verwendet. Insbesondere Approximation und Vorhersage fallen in diesen Themenbereich. Als ein sehr effektives Hilfsmittel im Hinblick auf eine zusammenfassende Theorie der sogenannten Kriging- und Splinemethoden werden dann die Kernreproduzierenden Hilberträume definiert (siehe in [44],[60]). Alsdann folgt ein Abschnitt über Krigingverfahren, die vom Typ her stochastisch sind. Dabei wird vor allem auf das Versuchsplanungproblem, also die optimale Auswahl der Samplefunktionale, eingegangen. Schließlich wird die Splineinterpolation als Vertreter einer deterministischen Approximationsmethode vorgestellt. Dabei liegt der Fokus auf der Darstellung einer bisher nicht erwähnten, allgemeinen Äquivalenz zwischen Kriging- und Splinemethoden. In dem folgenden Kapitel wird stets die kanonische Sicht einer zufälligen Funktion gewählt, d.h. Wahrscheinlichkeitsmaße sind stets auf Funktionenmengen definiert, Elementarereignisse sind somit stets Funktionen. Die besagte Äquivalenz gewinnt dadurch erheblich an Transparenz.

3.1. Lineare Approximationsprobleme

3.1.1. Definition

Es sei G ein normierter Raum, so ist die Ausgangsposition eines linearen Problems ein linearer, beschränkter Operator

$$S : C^r(D) \rightarrow G. \quad (3.1)$$

$C^r(D)$ ist die Menge der r -fach stetig differenzierbaren Funktionen auf $D \subset \mathbb{R}^n$. Führe nun einen weiteren Operator

$$\begin{aligned} S_k : F &\rightarrow G, \text{ mit } F \subset C^r(D) \\ f &\mapsto \sum_{j=1}^k a_j \lambda_j(f) \end{aligned} \quad (3.2)$$

ein. Die sogenannten Samplefunktionale λ_j seien aus einer gewissen Menge Λ linearer Funktionale auf $C^r(D)$, und S_k wird in diesem Zusammenhang auch lineare Methode genannt. S_k heißt lineare Approximation von S auf F , wenn die Koeffizienten $a_j \in G$ so gewählt werden, daß der Fehler

$$e(S_k, S) := \|S - S_k\| \quad (3.3)$$

minimal wird. Die Norm sei dabei auf dem Raum $L(C^r(D), G)$ aller linearen beschränkten Operatoren von $C^r(D)$ nach G definiert.

In der Mehrzahl der praktisch relevanten Fälle wird eine der nachfolgenden Normen gewählt:

- *Mittlere Fehler*

Sei Q ein Maß auf F , so ist der p -mittlere Fehler einer Methode S_k , die S approximiert, gerade

$$e(S_k, S) = e_p(S_k, S, Q) := \left(\int_F \|S(f) - S_k(f)\|_G^p Q(df) \right)^{1/p}, \quad (3.4)$$

mit $1 \leq p < \infty$. $\|\cdot\|_G$ sei die Norm im Raum G .

Im Kontext von stochastischen Approximationsverfahren stellt man sich $Q = P_Z$ als die Verteilung einer zufälligen Funktion Z vor. Falls $p = 2$ ist, spricht man vom mittleren quadratischen Fehler. Dieser hängt nur von Größen maximal zweiter Ordnung ab, also von der Erwartungswert- und Kovarianzfunktion. Beste lineare (erwartungstreue) Vorhersagen, also insbesondere Krigingverfahren, arbeiten ebenfalls mit diesem Fehlerbegriff. Wenn nicht besonders darauf hingewiesen wird, sei im folgenden stets der mittlere quadratische Fehler als verwendete Fehlerevaluation angenommen.

- *Maximale Fehler*

Der mittlere quadratische Fehler interpretiert den in (3.3) angegebenen allgemeinen Fehler als die mittlere quadratische Abweichung von Auswertungen der Funktionale S und S_k im Raum G . Der maximale Fehler ist dagegen die größt mögliche Abweichung dieser Art, d.h.

$$e(S_k, S) = e_{\max}(S_k, S, F) := \max_{f \in F} \|S(f) - S_k(f)\|_G. \quad (3.5)$$

Es werden fortan nur zwei Typen linearer Approximationsprobleme behandelt, die lineare Vorhersage und die lineare Rekonstruktion. Das Attribut linear wird fortan weggelassen, sofern nicht ausdrücklich nichtlineare Methoden gemeint sind.

1. Vorhersage:

Sei o.B.d.A. $G = \mathbb{R}$, ausgestattet mit der euklidischen Norm. Die Formulierung des Problems für komplexwertige Funktionen erfolgt analog. Für einen festen Punkt $t_0 \in D$, den sogenannten Vorhersageort, ist hier $S := \text{Pre}(t_0)$, das Auswertungsfunktional an der Stelle t_0 , gewählt. Für beliebige Funktionen $f \in C^r(D)$ gilt demnach $S(f) := \delta_{t_0}(f) = f(t_0)$. Dabei ist mit $\delta_t(\cdot)$ die gewöhnliche δ - oder Diracdistribution gemeint. Mit $S_k^{t_0}$ bezeichne in diesem Fall die lineare Approximation von S von der Form wie in (3.2) auf der Basis gewisser Samplefunktionale λ_j .

Sei Λ^{std} die Menge aller Diracdistributionen über $C^r(D)$, so bezeichne mit Λ^Q deren Abschluß bezüglich des L_2 -Produkts

$$\langle \lambda_1, \lambda_2 \rangle = \int_{\mathbb{F}} \lambda_1(f) \lambda_2(f) Q(df)$$

für beliebige Funktionale $\lambda_1, \lambda_2 \in \Lambda^Q$. Dieser Raum ist ein Hilbertraum (siehe auch [44]), womit eine besonders einfache Darstellung der besten linearen Approximation möglich ist, wie im Abschnitt 3.2 gezeigt wird. Es gilt $\text{Pre}(t_0) \in \Lambda^Q$ und mit $\lambda_1, \dots, \lambda_k \in \Lambda^Q$ auch $S_k^{t_0} \in \Lambda^Q$. Wie das obige Skalarprodukt vorgibt, ist der kanonische Fehler in diesem Fall gerade der mittlere quadratische, kurz $e(S_k^{t_0}, \text{Pre}(t_0)) = e_2(S_k^{t_0}, \text{Pre}(t_0), Q)$. Dieses lineare Vorhersageproblem zusammen mit der Evaluierung durch den mittleren Fehler ist grundlegend für die Betrachtungen in dieser Arbeit.

Bei der Vorhersage werden in vielen Fällen allein Beobachtungen der Funktionswerte der betrachteten Funktion verwendet. Prinzipiell können jedoch auch Beobachtungen von Ableitungen dieser Funktion für die Konstruktion einer Approximation benutzt werden. Dieses Verfahren wird erstmals von MENZ in [32] im Rahmen der statistischen Vorhersage, also unter Verwendung des mittleren quadratischen Fehlers als Fehlerkriterium, vorgeschlagen und findet z.B. bei der Erkundung von Kohlevorkommen Verwendung (siehe in [21]). In Abschnitt 3.3 wird insbesondere das Problem der optimalen Wahl der Beobachtungsorte von Funktions- und Ableitungswerten behandelt, was unter dem Begriff Versuchsplanung einzuordnen ist.

2. Rekonstruktion:

Sei $G = L_q(D)$ der Raum der q -fach integrierbaren, reellwertigen Funktionen auf D mit der gewöhnlichen L_q -Norm für $1 \leq q < \infty$. Auch hier erfolgt die Formulierung analog für den Fall komplexwertiger Funktionen. $S := \text{App}_q$ sei die Einbettung des $C^r(D)$ in diesen L_q -Raum. Bezeichne hier mit $S_{k,q}^D$ die zu S gehörige lineare Approximation von der Form wie in (3.2) auf der Basis gewisser gegebener Samplefunktionale. Falls $q = 2$ ist, soll der entsprechende Index fortan weggelassen werden.

Die Rekonstruktion $S_{k,q}^D$ eines Operators App_q kann bei Verwendung des mittleren Fehlers mit $p = q$ auf eine Familie von Vorhersageproblemen $\text{Pre}(t_0)$ mit $t_0 \in D$

zurückgeführt werden. Mit Fubini erhält man:

$$\begin{aligned}
e_p(S_{k,p}^D, \text{App}_p, Q)^p &= \int_F \|S(f) - S_{k,p}^D(f)\|_p^p Q(df) \\
&= \int_F \int_D \left| (S(f))(t_0) - (S_{k,p}^D(f))(t_0) \right|^p dt_0 Q(df) \\
&= \int_D e_p(\delta_{t_0} \circ S_{k,p}^D, \text{Pre}(t_0), Q)^p dt_0
\end{aligned}$$

Die gesuchte Rekonstruktion läßt sich demnach punktweise aus Vorhersagen ermitteln.

Für den Fall, daß Funktionen $f \in C^r(D)$ auf einer endlichen Teilmenge $D^* \subset D$ approximiert werden sollen, wählt man $G = \mathbb{R}^N$ mit der euklidischen Norm, wobei N gerade die Anzahl der Elemente in D^* ist. Diese Art der Approximation soll fortan mit App_{D^*} bezeichnet werden. Bei Verwendung des mittleren quadratischen Fehlers $e_2(S_k^{D^*}, \text{App}_{D^*}, Q)$ ist dieses Rekonstruktionsproblem wiederum gleichwertig zu einer Familie von N Vorhersageproblemen $\text{Pre}(t_0)$ mit $t_0 \in D^*$.

3.1.2. Kernreproduzierende Hilberträume

Sei Λ ein Raum reellwertiger Funktionale auf einem normierten Funktionenraum F . Ein Element $\lambda \in \Lambda$ heißt beschränkt in Λ , falls es eine reelle Konstante M gibt, so daß

$$|\lambda(f)| \leq M \|f\|$$

für alle $f \in F$ erfüllt ist. Aus der Beschränktheit von λ folgt insbesondere dessen Stetigkeit.

Definition 3.1. Sei H ein Hilbertraum reellwertiger Funktionen auf $D \subset \mathbb{R}^n$. Falls für alle $t \in D$ das zugehörige Auswertungsfunktional δ_t auf H beschränkt ist, so heißt H Kernreproduzierender Hilbertraum (engl.: *Reproducing Kernel Hilbert Space, RKHS*)

Es gelten die folgenden Eigenschaften:

1. Nach dem Darstellungssatz von Riesz existiert zu jedem Auswertungsfunktional δ_t mit $t \in D$ ein eindeutiger Repräsentant $K(\cdot, t) \in H$ (siehe z.B. in [43]). Wenn $\langle \cdot, \cdot \rangle$ das innere Produkt des Hilbertraumes H ist, so gilt demnach für alle $t \in D$ und $f \in H$ die Reproduzierenden-Eigenschaft

$$\delta_t(f) = \langle K(\cdot, t), f \rangle = f(t).$$

Dieser Repräsentant wird reproduzierender Kern genannt. Führe außerdem die Notation $(H(K), \langle \cdot, \cdot \rangle_K) := (H, \langle \cdot, \cdot \rangle)$ ein.

2. Sei $K : D^2 \rightarrow \mathbb{R}$ eine positiv semidefinite Funktion, so kann ein eindeutiger RKHS reellwertiger Funktionen auf D konstruiert werden, dessen reproduzierender Kern gerade K ist. Insbesondere zu Kovarianzfunktionen zufälliger Funktionen lassen sich solche RKHS konstruieren.

Zum Beweis von Punkt 2 siehe z.B. in [44]. Dort wird ein Raum $H_0 = \text{span}\{K(\cdot, t) : t \in D\}$ zusammen mit einem inneren Produkt $\langle \cdot, \cdot \rangle_K$ eingeführt. Dieses funktioniert für Funktionen $f = \sum_{j=1}^n a_j K(\cdot, t_j)$ und $g = \sum_{i=1}^m b_i K(\cdot, s_i)$ mit $t_j, s_i \in D$ und $a_j, b_i \in \mathbb{R}$ folgendermaßen:

$$\langle f, g \rangle_K := \sum_{j=1}^n \sum_{i=1}^m a_j b_i K(t_j, s_i).$$

Der Abschluß von H_0 bzgl. der aus diesem Skalarprodukt resultierenden Norm ist ein RKHS, wie in der genannten Quelle gezeigt wird.

Beispiel 3.2. Gegeben sei die Brownsche Bewegung auf dem Einheitsintervall in \mathbb{R} . D.h. es gilt $D = [0, 1]$ und $K(s, t) = \min(s, t)$, und der zugehörige RKHS ist gerade

$$H(K) = \{f \in W_2^1([0, 1]) : f(0) = 0\} \quad (\text{siehe in [44]})$$

mit dem Sobolevraum

$$W_2^1([0, 1]) = \{h \in C([0, 1]) : h \text{ absolutstetig, } h^{(1)} \in L_2([0, 1])\} \quad (\text{siehe in [43]}).$$

Die Schreibweise $h^{(r)}$ bezeichne hierbei die r -te Ableitung einer Funktion in \mathbb{R} .

Sei Q ein Wahrscheinlichkeitsmaß auf F mit Kovarianzfunktion K und verschwindender Erwartungswertfunktion. Die unter Punkt 1 angedeutete Zuordnung von Funktionen aus $H(K)$ zu Funktionalen aus Λ^Q kann zu einem linearen Hilbertraumisomorphismus

$$\mathcal{J} : \Lambda^Q \rightarrow H(K)$$

ergänzt werden, wobei $\mathcal{J}(\delta_t) = K(\cdot, t)$ für alle $t \in D$ (siehe in [27]) gilt. \mathcal{J} ist insbesondere normerhaltend. Das im vorherigen Unterabschnitt eingeführte spezielle lineare Vorhersageproblem kann somit isomorph in einen RKHS übertragen werden. Die Konstruktion der Lösung gewinnt bei dieser Vorgehensweise an Einfachheit, wie im folgenden Abschnitt gezeigt wird.

Bemerkung 3.3. 1. Sei Q ein Wahrscheinlichkeitsmaß auf einer Menge F , reellwertiger Funktionen auf $D \subset \mathbb{R}^n$ mit verschwindender Erwartungswertfunktion. Ferner sei die zugehörige Kovarianzfunktion K genau r -mal stetig differenzierbar auf D^2 . In [44] wird gezeigt, daß dann

$$(C^r(D))' \subset \Lambda^Q \tag{3.6}$$

sowie $\lambda(h) = \langle \mathcal{J}\lambda, h \rangle_K$ für alle $\lambda \in (C^r(D))'$ und $h \in H(K)$ gelten. $(C^r(D))'$ bezeichne hier die Menge aller stetigen linearen Funktionalen auf $C^r(D)$, dessen sogenannter Dualraum. Insbesondere liefert der Isomorphismus \mathcal{J} für beliebige lineare stetige Funktionale auf $C^r(D)$ deren Repräsentant in $H(K)$ wieder. Wie in [44] gezeigt wird, ist $(C^r(D))'$ sogar dicht in Λ^Q . Bezüglich der Genauigkeit von Approximationen spielt es daher keine Rolle, aus welcher der zwei Mengen die Samplefunktionalen gewählt werden.

2. Mit Hilfe der Reproduzierenden-Eigenschaft ist es sehr leicht möglich, die gemäß der Abbildung \mathcal{J} korrespondierenden Paare von Funktion und Funktional zu bestimmen. Sei z.B. Q ein Wahrscheinlichkeitsmaß auf $F \subset C^1(D)$ mit Definitionsbereich $D \subset \mathbb{R}$, verschwindender Erwartungswertfunktion und Kovarianzfunktion K . Bezeichne mit D_t das Funktional, welches die Ableitung einer Funktion an einer Stelle t zurückliefert. Es gilt:

$$\begin{aligned}
K^{(0,1)}(s, t) &= \int_F f^{(1)}(t) f(s) Q(df) \\
&= \int_F D_t(f) \delta_s(f) Q(df) \\
&= \langle D_t, \delta_s \rangle_{\Lambda^Q} = \langle \mathcal{J}D_t, \mathcal{J}\delta_s \rangle_K \\
&= \langle \mathcal{J}D_t, K(\cdot, s) \rangle_K = (\mathcal{J}D_t)(s),
\end{aligned}$$

also $\mathcal{J}D_t = K^{(0,1)}(\cdot, t)$. Diese Konstruktion funktioniert analog bei höherdimensionalen Definitionsbereichen wie in [44] gezeigt wird.

Beispiel 3.4. Die sogenannten glättenden Polynomsplines (siehe in [60]), sind Lösungen linearer Rekonstruktionsprobleme in Sobolevräumen $W_2^{r+1} = \{h \in C^r(D) : h^{(r)} \text{ abs. stetig, } h^{(r+1)} \in L_2(D)\}$, wobei $r \in \mathbb{N}_0$ geeignet gewählt wird. Dieser Raum ist ein RKHS wie in [44, 46] gezeigt wird und ist durch die Abbildung \mathcal{J} isomorph zu Λ^Q , wobei $Q := P_{Z_r}$ die Verteilung einer zufälligen Funktion der Art $Z_r(t) = \sum_{k=0}^{r-1} \frac{\theta_k}{k!} t^k + \int_D B(u) \frac{(t-u)^{r-1}}{(r-1)!} du$ ist. Dabei sind die zufälligen Größen $\theta_1, \dots, \theta_{r-1}, \{B(u) \mid u \in D\}$ unkorreliert, die θ_k haben Erwartungswert Null und endliche Varianz $\sigma^2 > 0$, und B ist eine Brownsche Bewegung auf dem Definitionsbereich D . Insbesondere der zweite Summand von Z_r ist eine r -fach integrierte Brownsche Bewegung (siehe [46], [60]).

3.2. Krigingmethoden

Die sogenannten Krigingmethoden sind Vertreter der stochastischen Vorhersageverfahren und werden insbesondere bevorzugt in den Umweltwissenschaften eingesetzt. Im nun folgenden Abschnitt werden das sogenannte Simple und das Universal Kriging als die eigentlich wichtigsten Vertreter dieser Verfahrensklasse vorgestellt. Die Darstellung der Verfahren sowie die Herleitung der Lösung gehört zum Standardrepertoire vieler Bücher aus dem Bereich der Geostatistik (siehe z.B. in [9], [12], [19], [23]). Der Schwerpunkt hier soll deshalb auf einer allgemeinen und abstrakten Darstellung der Probleme liegen. Zum einen ermöglicht dies, die verschiedenen Sonderfälle des Kriging (Block-Kriging, Gradientenkriging, Indikatorkriging) zusammenfassend zu behandeln und zum anderen die bereits angedeutete allgemeine Äquivalenz zwischen Kriging- und Splineverfahren wesentlich offensichtlicher zu gestalten.

3.2.1. Simple Kriging und BLP

Unter Kriging im allgemeinen versteht man eine erwartungstreue, lineare Vorhersage mit Fehlerevaluierung nach mittleren Quadraten. Diese Umbenennung eines hinlänglich bekannt-

ten Problems erfolgte durch MATHERON zu Ehren des südafrikanischen Bergbau-Ingenieurs D. G. Krige (siehe in [31]).

Es seien hierzu die folgenden Bedingungen erfüllt:

$$\lambda_1, \dots, \lambda_k, \text{Pre}(\mathbf{t}_0) \in (C^r(D))', \quad \mathbf{t}_0 \in D, \quad (3.7)$$

$$E \lambda_j = \lambda_j(\mathbf{m}), \quad j = 1, \dots, k. \quad (3.8)$$

Schließlich sei die Vorhersage $S_k^{\mathbf{t}_0}$ von der Form wie in (3.2). Die geforderte Erwartungstreue der Vorhersage bedeutet dabei, daß die Beziehung

$$E \text{Pre}(\mathbf{t}_0) = \mathbf{m}(\mathbf{t}_0) \stackrel{!}{=} S_k^{\mathbf{t}_0}(\mathbf{m}) \quad (3.9)$$

erfüllt ist. Die Erwartungswertbildung wie auch die Fehlerevaluierung geschieht dabei durch eine Wahrscheinlichkeitsverteilung Q auf einer Menge reellwertiger Funktionen $F \subset C^r(D)$ mit Erwartungswertfunktion \mathbf{m} . Die zugehörige Kovarianzfunktion $K \in C^{(r,r)}(D^2)$ sei als bekannt vorausgesetzt. Für ein geeignet gewähltes $r \in \mathbb{N}$ sei K also genau r -mal stetig differenzierbar. Der deterministische Anteil \mathbf{m} wird auch Trend genannt und beschreibt das Makroverhalten von Z .

Q ist Verteilung einer zufälligen Funktion $Z(\mathbf{t}) = \mathbf{m}(\mathbf{t}) + \varepsilon(\mathbf{t})$ für $\mathbf{t} \in D$, wobei $\text{Cov}(Z(\mathbf{t}), Z(\mathbf{s})) = \text{Cov}(\varepsilon(\mathbf{t}), \varepsilon(\mathbf{s})) = K(\mathbf{t}, \mathbf{s})$ gilt. Simple Kriging geht insbesondere davon aus, daß auch \mathbf{m} vollständig bekannt ist. Die Lösung dieses Vorhersageproblems wird auch BLP (engl.: best linear predictor) genannt.

$\mathbf{m} \equiv 0$

Nach Bemerkung 3.3 (1) gilt $\text{Pre}(\mathbf{t}_0), \lambda_1, \dots, \lambda_k \in \Lambda^Q$. Es liegt also ein Approximationsproblem in einem Hilbertraum vor, und damit ist die eindeutige Lösung die orthogonale Projektion von $\text{Pre}(\mathbf{t}_0)$ auf $\text{span}\{\lambda_1, \dots, \lambda_k\}$. Hierzu isomorph ist es, die orthogonale Projektion von $\mathcal{J}\text{Pre}(\mathbf{t}_0) = K(\cdot, \mathbf{t}_0)$ auf $\text{span}\{\mathcal{J}\lambda_1, \dots, \mathcal{J}\lambda_k\}$ zu bestimmen. Sei $\mathcal{J}S_k^{\mathbf{t}_0}$ diese Projektion, so gilt wegen der Linearität von \mathcal{J} und der genannten Orthogonalität:

$$\begin{aligned} & \langle \mathcal{J}S_k^{\mathbf{t}_0} - \mathcal{J}\text{Pre}(\mathbf{t}_0), \mathcal{J}S_k^{\mathbf{t}_0} \rangle_K = 0 \\ \Leftrightarrow & \sum_{i,j=1}^k a_i a_j \langle \mathcal{J}\lambda_i, \mathcal{J}\lambda_j \rangle_K = \sum_{j=1}^k a_j \langle \mathcal{J}\lambda_j, K(\cdot, \mathbf{t}_0) \rangle_K \\ \Leftrightarrow & \mathbf{a}^T \boldsymbol{\Sigma}_\lambda \mathbf{a} = \mathbf{a}^T \boldsymbol{\Sigma}_{\mathbf{t}_0} \\ \Rightarrow & \boldsymbol{\Sigma}_\lambda \mathbf{a} = \boldsymbol{\Sigma}_{\mathbf{t}_0} \end{aligned}$$

mit

$$\begin{aligned} \mathbf{a}^T &= (a_1, \dots, a_k) \\ \boldsymbol{\Sigma}_\lambda &= (\langle \mathcal{J}\lambda_i, \mathcal{J}\lambda_j \rangle_K)_{i,j=1}^k \\ \boldsymbol{\Sigma}_{\mathbf{t}_0} &= (\langle \mathcal{J}\lambda_j, K(\cdot, \mathbf{t}_0) \rangle_K)_{j=1}^k = ((\mathcal{J}\lambda_j)(\mathbf{t}_0))_{j=1}^k. \end{aligned}$$

Die Invertierbarkeit von $\boldsymbol{\Sigma}_\lambda$ vorausgesetzt, erhält man die bekannte Lösung

$$S_k^{\mathbf{t}_0} = \boldsymbol{\Sigma}_{\mathbf{t}_0}^T \boldsymbol{\Sigma}_\lambda^{-1} \boldsymbol{\lambda}, \quad (3.10)$$

wobei $\lambda^T = (\lambda_1, \dots, \lambda_k)$ ist. Wegen Eigenschaft (3.8) ist sichergestellt, daß diese Lösung erwartungstreu ist.

Bemerkung 3.5. Die Invertierbarkeit ist beispielsweise dann gewährt, wenn die Funktionen $\mathcal{J}\lambda_1, \dots, \mathcal{J}\lambda_k$ linear unabhängig sind in $H(K)$. Lineare Abhängigkeit bedeutet, daß redundante Sampleinformationen vorliegen. Es existiert dann also mindestens ein Funktional λ_{j_0} mit $j_0 \in \{1, \dots, k\}$, das als Linearkombination der verbleibenden Samplefunktionale dargestellt werden kann. Dieses Funktional kann für die Vorhersage weggelassen werden, da damit keine zusätzliche Information geliefert wird.

$\mathbf{m} \neq 0$

Wegen der geforderten Erwartungstreue ist $S_k^{t_0}$ notwendigerweise affin (siehe [44]). Wähle also den Ansatz $S_k^{t_0} = \mathbf{m}(t_0) + \sum_{j=1}^k \alpha_j (\lambda_j - \lambda_j(\mathbf{m})) =: \mathbf{m}(t_0) + \tilde{S}_k^{t_0}$ mit $\tilde{S}_k^{t_0} := \sum_{j=1}^k \alpha_j \tilde{\lambda}_j$. Dabei erfüllen die Funktionale λ_j die Bedingungen (3.7) und (3.8).

Die Funktionale $\tilde{\lambda}_j$ haben einen verschwindenden Erwartungswert, und es gilt $\Sigma_{\tilde{\lambda}} = \Sigma_{\lambda}$. Wählt man $\tilde{S}_k^{t_0}$ als die Approximation von $\widetilde{\text{Pre}}(t_0) := \text{Pre}(t_0) - \mathbf{m}(t_0)$ mit Fehlerevaluierung im quadratischen Mittel, so liegt bezüglich dieses Teilproblems die gleiche Situation wie im vorhergehenden Fall vor, und man erhält die Lösung des Gesamtproblems:

$$S_k^{t_0} = \mathbf{m}(t_0) + \Sigma_{t_0}^T \Sigma_{\lambda}^{-1} (\lambda - \lambda(\mathbf{m})). \quad (3.11)$$

Der Vorhersagefehler ist unabhängig von der gewählten Trendfunktion (also auch bei $\mathbf{m} \equiv 0$):

$$e_2(S_k^{t_0}, S, Q)^2 = K(t_0, t_0) - \Sigma_{t_0}^T \Sigma_{\lambda}^{-1} \Sigma_{t_0}. \quad (3.12)$$

Bemerkung 3.6. 1. Zur Konstruktion des BLP sowie des zugehörigen Vorhersagefehlers müssen somit wenigstens die Werte $\mathbf{m}(t_0), \lambda_1(\mathbf{m}), \dots, \lambda_k(\mathbf{m})$ sowie die Kovarianzmatrizen Σ_{t_0} und Σ_{λ} bekannt sein.

2. Die Eigenschaft (3.8) ist Voraussetzung für die Erwartungstreue und wird z.B. von Diracfunktionalen und Differentiationsoperatoren (Differenziation im quadratischen Mittel) erfüllt. Funktionale von der Form $\lambda_i = 1/|B_i| \int_{B_i} \delta_t dt$, mit $B_i \subset \mathbb{R}^n$, erfüllen diese Bedingung ebenfalls, sie führen zum sogenannten Blockkriging (siehe in [9], [23]).

3.2.2. Universal Kriging und BLUP

Im Unterschied zum Simple Kriging wird beim Universal Kriging die Forderung fallen gelassen, daß die Trendfunktion \mathbf{m} vollständig bekannt ist. Vielmehr geht man bei diesem Ansatz davon aus, daß der Trend eine deterministische Linearkombination bekannter Ansatzfunktionen ist, die Koeffizienten jedoch unbekannt sind.

Seien nun die Bedingungen (3.7) sowie

$$E \lambda_j = \lambda_j(\mathbf{f})^T \boldsymbol{\beta}, \quad j = 1, \dots, k \quad (3.13)$$

erfüllt. Analog zum Fall beim Simple Kriging bedeutet die Erwartungstreue der Vorhersage, daß die Bedingung

$$E\text{Pre}(t_0) = \mathbf{f}(t_0)^\top \boldsymbol{\beta} \stackrel{!}{=} S_k^{t_0}(\mathbf{m}) \quad (3.14)$$

erfüllt wird. Dabei ist $\boldsymbol{\beta} \in \mathbb{R}^r$ der unbekannte aber feste Parametervektor und $\mathbf{f}^\top = (f_1, \dots, f_r)$ sind die bekannten reellwertigen Ansatzfunktionale des Trends auf D . Hinsichtlich der Voraussetzungen an das für die Mittelung und Fehlerevaluierung verwendete Wahrscheinlichkeitsmaß Q , und deren zugehörige Kovarianzfunktion K seien dieselben Voraussetzungen wie beim Simple Kriging erfüllt.

Ist der Parametervektor $\boldsymbol{\beta}$ bekannt, so hat die Lösung $S_k^{t_0, \boldsymbol{\beta}} = \sum_{j=1}^k a_j^\beta \lambda_j$ des linearen Vorhersageproblems mit $\text{Pre}(t_0)$ die gleiche Form wie beim Simple Kriging,

$$S_k^{t_0, \boldsymbol{\beta}} = \mathbf{f}(t_0)^\top \boldsymbol{\beta} + \boldsymbol{\Sigma}_{t_0}^\top \boldsymbol{\Sigma}_\lambda^{-1} (\boldsymbol{\lambda} - \mathbf{F}\boldsymbol{\beta}) \quad (3.15)$$

mit $\mathbf{F} = (\lambda_i(f_j)) \in \mathbb{R}^{k \times r}$, $i = 1, \dots, k$ und $j = 1, \dots, r$. Da der Parametervektor $\boldsymbol{\beta}$ aber unbekannt ist, muß er durch $\hat{\boldsymbol{\beta}}$ linear und erwartungstreu geschätzt werden. D.h., daß $\hat{\boldsymbol{\beta}}$ ein lineares Funktional auf F mit der zusätzlichen Eigenschaft $\hat{\boldsymbol{\beta}}(\mathbf{m}) = \boldsymbol{\beta}$ ist. Ein klassisches Resultat besagt, daß in diesem Fall der beste, lineare, erwartungstreue Schätzer (engl.: best linear unbiased estimator, BLUE) in (3.15) eingesetzt werden kann, um die Lösung des anfangs genannten Vorhersageproblems zu erhalten (siehe z.B. in [10],[44],[56]):

$$S_k^{t_0} = \mathbf{f}(t_0)^\top \hat{\boldsymbol{\beta}} + \boldsymbol{\Sigma}_{t_0}^\top \boldsymbol{\Sigma}_\lambda^{-1} (\boldsymbol{\lambda} - \mathbf{F}\hat{\boldsymbol{\beta}}) \quad (3.16)$$

mit dem BLUE

$$\hat{\boldsymbol{\beta}} = (\mathbf{F}^\top \boldsymbol{\Sigma}_\lambda^{-1} \mathbf{F})^{-1} \mathbf{F}^\top \boldsymbol{\Sigma}_\lambda^{-1} \boldsymbol{\lambda} \quad (3.17)$$

für $\boldsymbol{\beta}$. Die Lösung $S_k^{t_0}$ wird auch BLUP (engl.: best linear unbiased predictor) genannt.

Bemerkung 3.7. 1. Sei $\boldsymbol{\beta}^* = \mathbf{C}\boldsymbol{\lambda}$ ein beliebiger linearer erwartungstreuer Schätzer für $\boldsymbol{\beta}$. In (3.15) eingesetzt erhält man $S_k^{t_0, \boldsymbol{\beta}^*} = \mathbf{f}(t_0)^\top \boldsymbol{\beta}^* + \boldsymbol{\Sigma}_{t_0}^\top \boldsymbol{\Sigma}_\lambda^{-1} (\boldsymbol{\lambda} - \mathbf{F}\boldsymbol{\beta}^*) =: A + B$ mit $A = \mathbf{f}(t_0)^\top \boldsymbol{\beta}^*$, dem Schätzer für den Trend an der Stelle t_0 . B dagegen ist die Vorhersage der trendbereinigten zufälligen Funktion an der Stelle t_0 . Man zeigt leicht, daß die beiden Summanden A und B genau dann unkorreliert sind, wenn $\boldsymbol{\beta}^*$ der BLUE ist.

2. Der Vorhersagefehler bei Verwendung der Methode in (3.16) ist:

$$e_2(S_k^{t_0}, S, Q) = K(t_0, t_0) - \boldsymbol{\Sigma}_{t_0}^\top \boldsymbol{\Sigma}_\lambda^{-1} \boldsymbol{\Sigma}_{t_0} + \boldsymbol{\gamma}^\top (\mathbf{F}^\top \boldsymbol{\Sigma}_\lambda^{-1} \mathbf{F})^{-1} \boldsymbol{\gamma} \quad (3.18)$$

mit $\boldsymbol{\gamma} = \mathbf{f}(t_0) - \mathbf{F}^\top \boldsymbol{\Sigma}_\lambda^{-1} \boldsymbol{\Sigma}_{t_0}$.

3. Die allgemeine Formulierung des Universal Kriging schließt insbesondere die in der Geostatistik bekannten Fälle mit ein. Beim sogenannten Gradientenkriging (siehe in [21], [32]) verwendet man zusätzlich zu Messungen von Funktionswerten der betrachteten Funktion (entsprechend Deltafunktionalen) Gradienten derselben Funktion, was hier den Differentiationsfunktionalen entspricht. Die Matrizen $\boldsymbol{\Sigma}$ sind dann die dort hinlänglich bekannten Kovarianzmatrizen.

3.3. Asymptotische Versuchsplanung

Nachdem im vorhergehenden Abschnitt die Konstruktion linearer Vorhersagen bei einer vorgegebenen Menge von Samplefunktionalen, einem sogenannten *Sampledesign*, erläutert wurde, steht nun eine effiziente Wahl der Samplefunktionale im Mittelpunkt. Zunächst wird der Begriff der Versuchsplanung eingeführt, wobei besonders die Unterscheidung zwischen der sogenannten asymptotischen Versuchsplanung, bei der die Anzahl der verwendeten Samplefunktionale gegen Unendlich strebt, und der Versuchsplanung mit endlichen Anzahlen von Samplefunktionalen verdeutlicht wird. Vom mathematisch analytischen Standpunkt ist die Behandlung des zweiten Problems wesentlich aufwendiger, wenngleich dieser Fall der eigentlich für die Praxis wichtige ist. Probleme dieser Art werden in der Regel durch globale numerische Optimierungsverfahren wie *Simulated Annealing* gelöst (siehe in [47] zu einer Variante des sogenannten *DETMAX*-Algorithmus, in [4, 5] für Beispiele aus dem Bereich der Meßnetzplanung in Zusammenhang mit *Kriging* unter Verwendung von Gradienteninformationen). In [7] wird zusätzlich der Suchraum durch die Verwendung sogenannter *Latin-Hypercube-Designs* eingeschränkt. Diese Verfahren ermöglichen eine einfache und schnelle Generierung von raumausfüllenden Samplepunkten im Beobachtungsraum. In [16] wird ein sehr interessanter Ansatz verfolgt, bei dem die Kovarianzstruktur des zufälligen Vorhersagefehlers bei einem bereits benutzten Design geschätzt wird. Mit Hilfe dieser Kovarianzfunktion läßt sich die Verkleinerung des Vorhersagefehlers bei Hinzunahme weiterer Designpunkte unabhängig von dem bereits vorhandenem Design ausdrücken. Dadurch lassen sich numerisch sehr schnell diejenigen Punkte feststellen, um die das Design vergrößert werden müßte, um den mittleren Vorhersagefehler zu minimieren. Dabei geschieht hier die Mittelung über die Vorhersagefehler, die an allen möglichen Vorhersageorten auftreten.

Zur asymptotischen Versuchsplanung dagegen existieren viele neuere Ansätze, wobei hier insbesondere jene von MÜLLER-GRONBACH und RITTER [44, 46, 45] vorgestellt werden. Schwerpunkt soll dabei insbesondere die Verwendbarkeit von Ableitungsinformationen sein. Allerdings ist eine wesentliche Einschränkung der genannten Arbeiten die Beschränkung der Anwendbarkeit auf zufällige Funktionen in \mathbb{R}^1 und die Beschränkung auf Hermitesche Daten. Letzteres heißt, daß an jedem Beobachtungsort jeweils stets der Funktionswert sowie alle Ableitungen bis zu einer gewissen Ordnung $r \in \mathbb{N}_0$ gemessen werden. In [57] wird zwar versucht, ein ähnliches Verfahren für die Versuchsplanung von zufälligen Funktionen in der Ebene zu übertragen, doch wird dort durch die Verwendung von stückweise konstanten Vorhersagefunktionen die Regularität der zugrundeliegenden zufälligen Funktion nicht ausgenutzt. Wenngleich das eigentliche Anliegen einer asymptotisch optimalen Versuchsplanung das Erzielen der bestmöglichen Rate ist, mit der das Designkriterium fällt, so ist manchen Verfahren die Vermutung immanent, bereits für endliche, genügend große Designs ein nahezu optimales gefunden zu haben.

Der Autor stellt in Abschnitt 3.3.3 eine Erweiterung eines von SACKS et. al. in [48] vorgestellten asymptotischen Verfahrens dar. Im Sinne des soeben erläuterten Prinzips einer asymptotischen Versuchsplanung liegt hier im engeren Sinne keine asymptotische Versuchsplanung vor, bei der die Anzahl der Beobachtungen gegen Unendlich strebt. Vielmehr verschwindet im Grenzfall die Korrelation zwischen den — endlich vielen — Beobachtungen. Im Unterschied zu den oben zitierten Arbeiten zur asymptotischen Versuchsplanung gibt es bei dem hier vorgestellten Verfahren keine Einschränkung bzgl. der Dimension des

Definitionsgebietes der betrachteten zufälligen Funktionen, und die Verwendung von nicht-hermiteschen Daten wird ermöglicht.

3.3.1. Versuchsplanung

Betrachte das lineare Vorhersageproblem mit $S = \text{Pre}(t_0)$, $t_0 \in D$ oder das Approximationsproblem $S = \text{App}_2$. Für fest vorgegebene Samplefunktionale $\lambda_1, \dots, \lambda_k$ sei $S_k = \sum_{j=1}^k \alpha_j \lambda_j$ die gemäß $e_2(S_k, S, Q)$ beste Methode. Dabei sei Q wiederum ein Wahrscheinlichkeitsmaß auf einer Menge $F \subset C^r(D)$. Das Tupel $T_k = (\lambda_1, \dots, \lambda_k)$ linearer Samplefunktionale auf $C^r(D)$ wird Ensemble oder auch Design genannt. In Fällen, wo alle Samplefunktionale von gleicher Art sind, z.B. nur Beobachtungen der Funktion oder nur Beobachtungen der r -ten Ableitung, so soll auch das Tupel der Beobachtungsorte selbst als Design bezeichnet werden.

Man kann im wesentlichen zwei Arten der Versuchsplanung unterscheiden. Diejenige, bei der die Anzahl der Samplefunktionale konstant bleibt und diejenige, bei der Folgen von Designs wachsender Größe betrachtet werden.

Ziel der erstgenannten Art der Versuchsplanung ist es, dasjenige Design $T_k^* = (\lambda_1^*, \dots, \lambda_k^*) \subset (\Lambda^{\text{poss}})^k$ mit $\Lambda^{\text{poss}} \subset (C^r(D))'$ zu bestimmen, so daß ein gewisses Fehlerfunktional Δ_S minimiert wird. Definiere dieses gemäß

$$\Delta_S : (\Lambda^{\text{poss}})^k \longrightarrow \mathbb{R}. \quad (3.19)$$

Das Ziel der letztgenannten Art der Versuchsplanung ist dagegen, eine Folge von Designs $(T_k^*)_{k \in \mathbb{N}}$ zu finden, so daß die Rate, mit der $\Delta_S(T_k^*)$ für größer werdende k fällt, maximal wird. Dieses Vorgehen heißt asymptotische Versuchsplanung (siehe in [49, 50, 51]). Es folgen zwei Beispiele für Versuchsplanungsprobleme und die dort verwendeten Fehlerfunktionale.

Beispiel 3.8. 1. Sei Q ein Gaußsches Maß auf $C^1(D)$, $D \subset \mathbb{R}^n$, mit verschwindender Erwartungswertfunktion und Kovarianzfunktion K . Betrachte dazu das lineare Approximationsproblem $S = \text{App}_{D^*}$ auf einer endlichen Menge $D^* \subset D$ mit Fehlerevaluierung im quadratischen Mittel. Zur Approximation wird ein Design $T_k = (\lambda_1, \dots, \lambda_k) \subset (\Lambda^{\text{poss}})^k$ fester Größe k verwendet. Dabei sollen den Funktionalen aus Λ^{poss} ausschließlich Beobachtungen von Funktionswerten und Ableitungen jeweils an Orten aus der Menge D^* entsprechen.

Sei \tilde{Q} das bedingte Maß Q gegeben $\lambda_1, \dots, \lambda_k$, so ist \tilde{Q} wieder ein Gaußsches Maß (siehe z.B. in [58]). Die im quadratischen Mittel beste Vorhersage von $\text{Pre}(t_0)$ für ein $t_0 \in D^*$ unter Verwendung des Designs T_n ist gerade die bedingte Erwartung $\tilde{m}(t_0) := \tilde{E} \text{Pre}(t_0)$, wobei die Erwartungswertbildung gerade bezüglich des Wahrscheinlichkeitsmaßes \tilde{Q} erfolgt. Da Q ein Gaußsches Maß ist, ist diese bedingte Erwartung bekanntermaßen linear in den Designelementen aus T . Die gesuchte Approximation $S_k^{D^*}$ ist also gerade \tilde{m}_{D^*} , die Einschränkung von \tilde{m} auf die Menge D^* .

\tilde{Q} hat die Kovarianzfunktion

$$\tilde{K}(s, t) = K(s, t) - \Sigma_s^T \Sigma_\lambda^{-1} \Sigma_t$$

mit $\Sigma_\lambda = (\langle \mathcal{J}\lambda_i, \mathcal{J}\lambda_j \rangle_{\mathcal{K}})_{i,j=1}^k$ und $\Sigma_x = (\langle \mathcal{J}\lambda_j, \mathcal{K}(x, \cdot) \rangle_{\mathcal{K}})_{j=1}^k = ((\mathcal{J}\lambda_j)(x))_{j=1}^k$ für ein $x \in D^*$. Das sogenannte D-Kriterium verwendet als Fehlerfunktional für die Versuchsplanung

$$\begin{aligned} \Delta_S(\lambda_1, \dots, \lambda_n) &= \det \left(\text{Cov}(\lambda_i, \lambda_j)_{\lambda_i, \lambda_j \in \Lambda^{\text{poss}} \setminus \{\lambda_1, \dots, \lambda_k\}} \right) \\ &= \det \left((\langle \mathcal{J}\lambda_i, \mathcal{J}\lambda_j \rangle_{\mathcal{K}})_{\lambda_i, \lambda_j \in \Lambda^{\text{poss}} \setminus \{\lambda_1, \dots, \lambda_k\}} \right). \end{aligned}$$

Nach SHEWRY und WYNN [55] ist dieses Problem äquivalent zum Problem

$$\max_{(\lambda_1, \dots, \lambda_k) \in (\Lambda^{\text{poss}})^k} \det \left((\langle \mathcal{J}\lambda_i, \mathcal{J}\lambda_j \rangle_{\mathcal{K}})_{\lambda_i, \lambda_j \in \{\lambda_1, \dots, \lambda_k\}} \right). \quad (3.20)$$

2. Betrachte ein Approximationsproblem App_2 auf einer Menge $F \subset C(D)^r$ mit Verteilung Q und $(\lambda_1, \dots, \lambda_k) \subset (\Lambda^{\text{poss}})^k$. Als Fehlerfunktional für die Versuchsplanung wähle $\Delta = e_2(S_k^D, \text{App}_2, Q)$.

Der Fall, daß endlich viele Beobachtungen vorliegen, ist in der Literatur bisher nur numerisch behandelt worden. Dort wird mittels Optimierungsalgorithmen das Funktional Δ minimiert (siehe z.B. in [48]). Für den Fall, daß man Folgen $((\lambda_1^*, \dots, \lambda_k^*))_{k \in \mathbb{N}}$ von Designs mit wachsendem Umfang betrachtet, stellen MÜLLER-GRONBACH und RITTER (1997) in [45] ein adaptives Verfahren der Versuchsplanung vor, das asymptotisch optimal ist, d.h. es gilt:

$$e_2(S_k^{D,*}, S, Q) = O \left(\inf_{(\lambda_1, \dots, \lambda_k) \subset (\Lambda^{\text{poss}})^k} e_2(S_k^D, S, Q) \right).$$

Dabei ist $S_k^{D,*}$ eine in $\lambda_1^*, \dots, \lambda_k^*$ lineare Approximation von S .

3.3.2. Asymptotische Verfahren nach Müller-Gronbach und Ritter

In den fundamentalen Arbeiten von SACKS/YLVISAKER [49, 50, 51] werden Regularitätsbedingungen für Kovarianzfunktionen definiert, die im wesentlichen unverändert bis heute als Voraussetzung für eine analytische Betrachtung der asymptotischen Versuchsplanung gelten. In der ursprünglichen Version werden nur Kovarianzfunktionen zufälliger Funktionen auf dem Einheitsintervall $[0, 1]$ betrachtet, die nicht differenzierbar im quadratischen Mittel sind. In [44] und [46] werden diese Bedingungen auf den Fall, der r -fachen Differenzierbarkeit erweitert, und in [44] werden verschiedene Approximationsprobleme unter diesen modifizierten Bedingungen behandelt, auf die im folgenden eingegangen wird. In Form eines Beispiels wird vom Autor gezeigt, daß eine gewisse Teilmenge der sehr vielseitigen Matérn-Kovarianzklasse (siehe Beispiel 2.5 (2)) zwei wichtige Teile der Sacks-Ylvisaker-Bedingungen erfüllen. Im letzten Teil des Abschnitts wird ein adaptives Verfahren nach MÜLLER-GRONBACH und RITTER zur Rekonstruktion zufälliger Funktionen mit lokal veränderlicher Variabilität erläutert. Dabei wird das Verfahren jedoch in einer vom Autor modifizierten Version für die Rekonstruktion der r -ten Ableitung einer derartigen Funktion dargestellt.

Im folgenden sei stets $D = [0, 1]$ angenommen. Wenngleich die Definitionen und Verfahren auf andere Definitionsgebiete in \mathbb{R} übertragen werden können, so ist doch die Forderung nach der Eindimensionalität von D grundlegend.

Sacks/Ylvisaker-Bedingungen der Ordnung r

Zunächst einige Notationen und Definitionen. Das offene Einheitsquadrat in \mathbb{R}^2 sei disjunkt in die Diagonale und die Mengen $\Omega_- = \{(s, t) \in (0, 1)^2 : s < t\}$ und $\Omega_+ = \{(s, t) \in (0, 1)^2 : s > t\}$ zerlegt. Sei L eine auf $\Omega_+ \cup \Omega_-$ stetige Funktion, wobei für $j \in \{+, -\}$ die Einschränkung L_{Ω_j} stetig auf $\bar{\Omega}_j$ ergänzbar ist. Mit L_j bezeichne dann diejenige Ergänzung von L auf $[0, 1]^2$, die stetig auf $\bar{\Omega}_j$ sowie $[0, 1]^2 \setminus \bar{\Omega}_j$ ist.

Falls möglich, definiere für ein $r \in \mathbb{N}_0$ die Funktion $L(s, t) := K^{(r,r)}(s, t)$ für $(s, t) \in [0, 1]^2$. Eine Kovarianzfunktion K erfüllt dann die sogenannten Sacks/Ylvisaker-Bedingungen der Ordnung $r \in \mathbb{N}_0$ (kurz SY(r)), falls die folgenden Bedingungen erfüllt sind:

- (A) L ist stetig auf $[0, 1]^2$. Für Multiindizes $\kappa = (\kappa[1], \kappa[2])$ mit $\kappa[i] \in \{0, 1\}$ und $\|\kappa\| \leq 2$ sei $L^\kappa \in C(\Omega_+ \cup \Omega_-)$ und L^κ sei stetig erweiterbar auf $\bar{\Omega}_+$ und $\bar{\Omega}_-$.
- (B) $L_-^{(1,0)}(t, t) - L_+^{(1,0)}(t, t) = \text{const.}$, für alle $t \in [0, 1]$
- (C) $L_+^{(2,0)}(t, \cdot) \in H(L)$ für alle $t \in [0, 1]$ und $\sup_{t \in [0, 1]} \|L_+^{(2,0)}(t, \cdot)\|_L < \infty$

Beispiel 3.9. 1. Für $s, t \in [0, 1]$, definiere die Kovarianzfunktionen

$$\begin{aligned} K_1(s, t) &= (1 - |s - t|)/2 \\ K_2(s, t) &= \exp(-|s - t|)/2. \end{aligned}$$

Beide erfüllen SY(0) (siehe [44]), insbesondere K_2 ist die Kovarianzfunktion des Ornstein-Uhlenbeck-Prozesses (siehe in [8]).

In [26] und [34] wird gezeigt, daß sich mittels iterierter Integration stationäre, differenzierbare Kovarianzfunktionen erzeugen lassen, die wiederum zu gewissen integrierten zufälligen gaußschen Funktionen gehören. Dazu sei X ein Gaußscher Prozeß mit verschwindender Erwartungswertfunktion und Kovarianzfunktion K_0 , welche eine der oben genannten Kovarianzfunktionen K_1 oder K_2 ist. Es lassen sich dann für beliebige aber feste $r \in \mathbb{N}_0$ positive Konstanten $\sigma_1^2, \dots, \sigma_r^2$ finden, so daß stationäre gaußsche Prozesse $Z^{(r-i)}$, für $i = 1, \dots, r$ mit Kovarianzfunktion

$$K_{(i)}(\mathbf{h}) := \sigma_i^2 - \int_0^h \int_0^u K_{(i-1)}(s) ds du \quad i = 1, \dots, r \quad (3.21)$$

existieren. Insbesondere existiert also ein $Z^{(0)}$, dessen r -te Ableitung gerade X und dessen zugehörige Kovarianzfunktion $K_{(r)}$ ist. Die Prozesse $Z^{(r-i)}$ werden dabei durch iterierte Integration gebildet:

$$Z^{(r-i)}(t) = Z_0^{(r-i)} + \int_0^t Z^{(r-i+1)}(s) ds \quad i = 1 \dots, r. \quad (3.22)$$

Dabei sind die Größen $Z_0^{(r-i)}$ gaußsche Zufallsvariable mit Erwartungswert 0 und endlicher Varianz, deren nichttriviale Konstruktion in den zitierten Arbeiten beschrieben

wird. Würde man z.B. paarweise unkorrelierte Größen wählen, so wären die so konstruierten Prozesse i.a. nicht stationär.

Die integrierten Kovarianzfunktionen $K_{(\mathbf{r})}$ erfüllen definitionsgemäß die Bedingungen SY(\mathbf{r}).

2. Betrachte die Matérnsche Kovarianzklasse mit einer anderen Parametrisierung als in Beispiel 2.5 angegeben (siehe in [56]). Sei $\eta = (\alpha, \phi, \nu) \in \mathbb{R}_+^3 \setminus \{\mathbf{0}\}$ ein Parametervektor, so definiere die Kovarianzfamilie

$$K_{\eta}(\mathbf{r}) = \frac{\pi^{1/2}\phi}{s^{\nu-1}\Gamma(\nu+1/2)\alpha^{2\nu}}(\alpha\mathbf{r})^{\nu}\mathcal{K}_{\nu}(\alpha\mathbf{r}). \quad (3.23)$$

Sei nun $\nu = m + 1/2$, wobei $m \in \mathbb{N}_0$ ist, so ist K_{η} genau $2m$ -mal stetig differenzierbar und hat die folgende Form

$$K_{\eta}(\mathbf{r}) = f(\mathbf{r})g(\mathbf{r})$$

mit

$$\begin{aligned} f(\mathbf{r}) &= e^{-\alpha\mathbf{r}} \\ g(\mathbf{r}) &= \text{Polynom in } \mathbf{r} \text{ vom Grad } m \quad (\text{siehe in [56]}). \end{aligned}$$

Insbesondere gilt $K_{(\alpha,\phi,1/2)}(\mathbf{r}) = \pi\phi\alpha^{-1}e^{-\alpha\mathbf{r}}$ bzw. $K_{(\alpha,\phi,3/2)}(\mathbf{r}) = (1/2)\pi\phi\alpha^{-3}e^{-\alpha\mathbf{r}} \times (1 + \alpha\mathbf{r})$, wobei $K_{(\alpha,\phi,1/2)}$ die Bedingung SY(0) erfüllt (siehe K_2 im obigen Beispiel).

Behauptung 3.10. Für $m \in \mathbb{N}_0$ erfüllt die Kovarianzfunktion $K_{(\alpha,\phi,m+1/2)}$ die Bedingungen SY(m), (A) und (B).

Beweis. Sei $m \in \mathbb{N}$ (für $m = 0$ siehe oben) und schreibe $L(\mathbf{t}, s) = K_{(\alpha,\phi,m+1/2)}(|\mathbf{t} - s|)^{(m,m)}$.

Teil (A): Mit wiederholter Anwendung der Produktregel und weil $g^{(i)} \equiv 0$ für $i > m$ gilt, erhält man

$$L(\mathbf{t}, s) = f(|\mathbf{t} - s|) \sum_{j=0}^m \binom{2m}{j} g^{(j)}(|\mathbf{t} - s|) (-\alpha)^{2m-j}.$$

L ist als Summe stetiger Funktionen stetig auf $[0, 1]^2$. Betrachte nun die Fälle L^{κ} mit $\kappa \in \{(1, 0), (1, 1)\}$.

$\kappa = (1, 0)$ und $(\mathbf{t}, s) \in \Omega_+$: Analog zu oben erhält man

$$L^{\kappa}(\mathbf{t}, s) = -\alpha f(\mathbf{t} - s) \sum_{j=0}^m \binom{2m+1}{j} g^{(j)}(\mathbf{t} - s) (-\alpha)^{2m-j}, \quad (3.24)$$

und L^{κ} ist stetig auf Ω_+ und stetig erweiterbar auf $\bar{\Omega}_+$.

$\kappa = (1, 0)$ und $(\mathbf{t}, s) \in \Omega_-$:

$$L^{\kappa}(\mathbf{t}, s) = \alpha f(s - \mathbf{t}) \sum_{j=0}^m \binom{2m+1}{j} g^{(j)}(s - \mathbf{t}) (-\alpha)^{2m-j} \quad (3.25)$$

ist stetig auf Ω_- und stetig erweiterbar auf $\bar{\Omega}_-$.

$\kappa = (1, 1)$ und $(t, s) \in [0, 1]^2$:

$$L^\kappa(t, s) = -\alpha^2 f(|t-s|) \sum_{j=0}^m \binom{2m+2}{j} g^{(j)}(|t-s|) (-\alpha)^{2m-j}$$

ist stetig auf $[0, 1]^2$. Der Fall $\kappa = (0, 1)$ folgt aus Symmetriegründen.

Teil (B): Aus den Teilen (3.24) und (3.25) erhält man $L_-^{(1,0)}(t, t) - L_+^{(1,0)}(t, t) = 2\alpha \sum_{j=0}^m \binom{2m+1}{j} g^{(j)}(0) (-\alpha)^{2m-j} < \infty$. \square

Bemerkung 3.11. Für $m = 1$ ist die Bedingung (C) nicht erfüllt, wie leicht nachzuweisen ist. Sei dazu $t \in [0, 1]$ beliebig aber fest gewählt und $s \in [0, 1]$, so gilt:

$$L_+^{(2,0)} = \alpha^3 f(|t-s|) \left\{ \alpha g^{(0)}(|t-s|) - 4g^{(1)}(|t-s|) \right\}.$$

Nehme nun an, daß die Bedingung (C) gilt, so existiert eine Linearkombination $\sum_{j=1}^l \alpha_j L(s_j, s)$ mit $\alpha_j \in \mathbb{R}$ und $s_j \in [0, 1]$, die identisch $L_+^{(2,0)}(t, s)$ für alle $s \in [0, 1]$ ist bzw. deren Grenzwert für $l \rightarrow \infty$ identisch mit dieser Funktion ist. Ein Koeffizientenvergleich führt jedoch zu $l = 1$, und somit muß ein Faktor $\alpha_1 \in \mathbb{R}$ und ein Punkt $s_1 \in [0, 1]$ existieren, so daß

$$\begin{aligned} \alpha_1 \left(\alpha^2 f(|s_1-s|) g(|s_1-s|) \right) - \alpha_1 \left(2\alpha f(|s_1-s|) g^{(1)}(|s_1-s|) \right) \\ = \left(\alpha^4 f(|t-s|) g(|t-s|) \right) - \left(4\alpha^3 f(|t-s|) g^{(1)}(|t-s|) \right) \end{aligned}$$

gilt. Führe wieder einen Koeffizientenvergleich durch und betrachte nur die jeweils ersten Summanden der obigen Gleichung, so erhält man $s_1 = t$ und $\alpha_1 = \alpha^2$. Dies führt allerdings zum Widerspruch bei den jeweils zweiten Summanden. Damit ist die Bedingung (C) jedoch nicht erfüllt.

Asymptotisch optimale Versuchsplanung mit Hermiteschen Daten

Sei Q ein Wahrscheinlichkeitsmaß auf einer Menge $F \subset C^r([0, 1])$ mit verschwindender Erwartungswertfunktion und Kovarianzfunktion K , welche die Bedingungen SY(r) mit $r \in \mathbb{N}_0$ erfüllt. Es sollen nun lineare Rekonstruktionsprobleme App_2 mit Methoden $S_k^{[0,1]}$ betrachtet werden, die ausschließlich sogenannte Hermite-Daten der Ordnung r benutzen, d.h. es werden Samplevektoren der Form $(D_0(t_j), \dots, D_r(t_j))$ verwendet. Dabei ist $D_0(t_j) = \delta(t_j)$ und $D_i(t_j)(\square) = \frac{\partial^i}{\partial t^i} \square(t)|_{t=t_j}$. Die Rekonstruktion ist demnach von der Form

$$S_k^{[0,1],\psi} = \sum_{j=1}^l \sum_{k=0}^r \alpha_{j,k} D_k(t_j), \quad \text{mit } k = l(r+1). \quad (3.26)$$

Die Orte t_j sind gerade die Quantile der Funktion ψ , d.h. es gilt für alle $j = 1, \dots, l$, $j/l = \int_0^{t_j} \psi(s) ds$. Dabei sei ψ eine stetige Wahrscheinlichkeitsdichtefunktion auf $[0, 1]$. Derartige Designs werden als regulär und durch ψ erzeugt bezeichnet.

In [44] wird der folgende Satz bewiesen. Dabei sind zwei reellwertige Folgen a_k, b_k stark äquivalent (in Zeichen $a_k \approx b_k$), falls $a_k/b_k \rightarrow 1$ für $k \rightarrow \infty$ gilt.

Satz 3.12. *Unter den genannten Bedingungen gilt:*

$$e_2(S_k^{[0,1],\psi}, App_2, Q) \approx \left(\frac{2r+1}{(4r+3)!} \right)^{1/2} \frac{(2r)!}{r!} I_r(\psi) \iota^{-r+1/2} \quad (3.27)$$

mit

$$I_r(\psi) = \left(\int_0^1 \psi(s)^{-(2r+1)} ds \right)^{1/2} \left(\int_0^1 \psi(s) ds \right)^{r+1/2}. \quad (3.28)$$

Der minimale Fehler in der Klasse aller linearen Methoden, die Hermitesche Daten der Ordnung r verwenden, ist stark äquivalent zu dem Fehler, der bei Verwendung von $S_k^{[0,1],1}$ entsteht, also bei äquidistanten Beobachtungsorten.

Bemerkung 3.13. Der Satz gilt in einer allgemeineren Fassung für Rekonstruktion, die durch ρ -gewichtete mittlere quadratische Fehler konstruiert werden (siehe in [44]).

Adaptive, asymptotisch optimale Versuchsplanung mit Funktionswerten

In [45] wird ein Verfahren zur Rekonstruktion zufälliger Funktionen auf $[0, 1]$ vorgestellt, die lokal Hölder stetig mit Hölder-Exponent $1/2$ sind und $SY(0)$ erfüllen. In einem ersten Schritt wird dabei die lokale Hölderkonstante geschätzt, und in einem zweiten Schritt erfolgt eine Erweiterung des Designs entsprechend dieser geschätzten Funktion. Die Versuchsplanung ist somit adaptiv in dem Sinne, daß nicht alle Beobachtungspunkte unabhängig voneinander ausgewählt werden (siehe auch in [39]). Durch das Konzept der lokalen Hölder Stetigkeit wird es insbesondere möglich, zufällige Funktionen mit einer lokal veränderlichen Variabilität zu betrachten.

Dieser Ansatz läßt sich, wie vom Autor gezeigt wird, leicht erweitern auf die Rekonstruktion von Ableitungen der Ordnung r gewisser Prozesse, die zu den in [45] genannten, analoge Bedingungen erfüllen. Die Vorgehensweise wird nur kurz vorgestellt, wobei insbesondere bezüglich der Details zum adaptiven Versuchsplanungsteil wie auch zur Schätzung der lokalen Hölderkonstante auf die Ausarbeitung in [45] verwiesen wird.

Betrachte eine zufällige Funktion

$$Y(t) = m(t) + g(t)Z(f(t)) \text{ für } t \in [0, 1]. \quad (3.29)$$

Dabei sei m die Erwartungswertfunktion von Y und R dessen Kovarianzfunktion. Die zufällige Funktion Z mit Verteilungsfunktion Q sei gaußsch, und ihre Kovarianzfunktion K erfülle die Bedingungen $SY(r)$, wobei die Konstante in der Teilbedingung (B) gerade 1 sei. Die Erwartungswertfunktion m sowie die Skalierung g und die Transformation f erfüllen außerdem die folgenden Bedingungen:

$$(D) \quad m \in C^{r+1}([0, 1])$$

(E) $g \in C^{r+1}([0, 1])$ und positiv

(F) $f([0, 1]) \subset [0, 1]$, f ist differenzierbar, und $f^{(1)}$ ist positiv und Lipschitz stetig.

Lipschitz Stetigkeit von f bedeutet, daß eine positive Konstante L existiert, so daß für alle $s, t \in [0, 1]$ die Bedingung $|f(s) - f(t)| < L|s - t|$ erfüllt ist. Die lokale Hölderfunktion $\alpha(t)$ von $Y^{(r)}$ ist schließlich definiert als der Limes

$$\alpha(t) = \lim_{s \rightarrow t} \frac{(\mathbb{E} (Y^{(r)}(s) - Y^{(r)}(t))^2)^{1/2}}{|s - t|^{1/2}}, \quad t \in [0, 1]. \quad (3.30)$$

Behauptung 3.14. *Erfüllt die zufällige Funktion Y aus (3.29) die Bedingungen (A) bis (F), so gilt $\alpha(t) = g(t) (f^{(1)}(t))^{r+1/2}$.*

Beweis. Mit Hilfe wiederholter Anwendung der Produktregel erhalte

$$Y^{(r)}(t) = m^{(r)}(t) + \Phi(Z, \dots, Z^{(r-1)}, f, g) + g(t) \left(f^{(1)}(t) \right)^r Z^{(r)}(f(t)). \quad (3.31)$$

Durch einfache Fallunterscheidung ($s > t$ oder $t > s$) und weil $R^{(r,r)} \in C([0, 1]^2)$ ist, gilt:

$$\alpha(t) = \lim_{s \rightarrow t} \frac{(\mathbb{E} (Y^{(r)}(s) - Y^{(r)}(t))^2)^{1/2}}{|s - t|^{1/2}} \quad (3.32)$$

$$= \left(R_{-/+}^{(r+1,r)}(t, t) - R_{+}^{(r+1,r)}(t, t) \right)^{1/2}. \quad (3.33)$$

Mit der Beziehung in (3.31) und erneuter Anwendung der Produktregel kann der Ausdruck in (3.33) in Termen der Kovarianzfunktion K ausgedrückt werden:

$$\begin{aligned} R_{-/+}^{(r+1,r)}(s, t) &= g(s)g(t) K_{-/+}^{(r+1,r)}(f(s), f(t)) \left(f^{(1)}(s) \right)^{r+1} \left(f^{(1)}(t) \right)^r \\ &\quad + \tilde{\Phi}(K^{(0,0)}, \dots, K^{(r,0)}, \dots, K^{(r,r)}, f, g). \end{aligned} \quad (3.34)$$

Bildet man nun die Differenz aus (3.33), so fallen die Terme mit $\tilde{\Phi}$ weg, da die Kovarianzfunktion $K^{(r,r)} \in C([0, 1]^2)$ ist. Zusammen mit der Bedingung SY(r),(B) für K erhält man:

$$\alpha(t) = \left(g(t)^2 f^{(1)}(t)^{2r+1} \left[K_{-}^{(r+1,r)}(t, t) - K_{+}^{(r+1,r)}(t, t) \right] \right)^{1/2} = g(t) \left(f^{(1)}(t) \right)^{r+1/2}.$$

□

Die Rekonstruktion der r -ten Ableitung von Y soll asymptotisch optimal in der Klasse der linearen Methoden sein, die ausschließlich Beobachtungen der r -ten Ableitung verwenden. Seien $S_k^{[0,1]}$ genau diese linearen Methoden. Die Konvergenzordnung des zugehörigen minimalen Fehlers findet man in [45]:

$$\inf_{S_k^{[0,1]}} e_2(S_k^{[0,1]}, \text{App}_2, Q) \approx c \|\alpha\|_1 k^{-1/2}, \quad c = \text{const.} (\in \mathbb{R}). \quad (3.35)$$

Insbesondere ist diejenige lineare Methode $S_k^{r,\psi}$ asymptotisch optimal, welche die Beobachtungen an den durch die Funktion

$$\psi(\mathbf{t}) = \alpha(\mathbf{t}) \|\alpha\|_1 \quad (3.36)$$

induzierten Designpunkten linear interpoliert (siehe ebenfalls in [45]).

Deshalb wird im ersten (nichtadaptiven) Schritt des Verfahrens die lokale Hölderkonstante mit $\hat{\alpha}$ geschätzt. Alsdann werden zu dem bestehenden Design Beobachtungspunkte gemäß der Dichte $\hat{\psi} := \psi(\hat{\alpha})$ (siehe (3.36)) zugefügt. In der asymptotischen Betrachtung wird der Anteil der so zugefügten Beobachtungspunkte stetig erhöht. Die Anzahl dieser Punkte wird in der zitierten Arbeit mit N_k bezeichnet und ist stark äquivalent zu k .

Die Rekonstruktion $S_{N_k}^{[0,1],\hat{\psi}}$ erfolgt entsprechend durch lineare Interpolation zwischen den Beobachtungen aller Beobachtungsstellen. Da $Y^{(r)}$ wiederum ein Gaußscher Prozeß ist, gilt der in [45] bewiesene Satz:

Satz 3.15. *Erfüllt der Prozeß Y die oben genannten Bedingungen, so ist die Rekonstruktion von $Y^{(r)}$ mittels $S_{N_k}^{[0,1],\hat{\psi}}$ asymptotisch optimal, d.h. es gilt*

$$e_2(S_{N_k}^{[0,1],\hat{\psi}}, \text{App}_2, Q) \approx c \|\alpha\|_1 k^{-1/2}.$$

3.3.3. Versuchsplanung mit nichthermiteschen Daten

$D \subset \mathbb{R}$

Das nachfolgend dargestellte Verfahren basiert auf einer Idee von MORRIS, MITCHELL und YLVISAKER, die in [36] vorgestellt wird. Dort wird die Rekonstruktion $S = \text{App}_{D^*}$ für eine endliche Menge $D^* \subset D$ behandelt. Die Versuchsplanung ist in dem Sinne asymptotisch, daß ein optimales Design für den Fall einer asymptotisch verschwindenden Korrelation bestimmt werden soll. Die Korrelation wird in diesem Fall durch einen Parameter Θ gesteuert. Im Unterschied zu der genannten Arbeit sollen im vorliegenden Fall jedoch nichthermitesche Daten verwendet werden.

Sei Q ein Gaußsches Wahrscheinlichkeitsmaß auf $C^1(D)$ mit $D \subset \mathbb{R}$ und der Kovarianzfunktion $K(\mathbf{t}, \mathbf{s}) = K(|\mathbf{t} - \mathbf{s}|)^\Theta = (e^{-\alpha|\mathbf{t} - \mathbf{s}|^2})^\Theta$, $\alpha > 0$. Der Exponent Θ bewirkt bei $\Theta \rightarrow \infty$ das asymptotische Verschwinden der Kovarianz bei dem zu betrachtenden Grenzwertprozeß. Die Wahl der Kovarianzfunktion entspricht derjenigen in [36], wo sie im Rahmen der Analyse von Computerprogrammen zu guten Ergebnissen führt. Dennoch sei betont, daß das folgende Verfahren für beliebige isotrope Kovarianzfunktionen funktioniert, die mindestens zweimal stetig differenzierbar auf dem Definitionsbereich sind.

Die Approximation S_n basiere stets auf einem Design mit genau k Funktionswertmessungen und m Ableitungsmessungen. Es soll also von der Art $T = (\delta_{t_1}, \dots, \delta_{t_k}, D_{s_1}, \dots, D_{s_m})$ mit $t_1, \dots, t_k, s_1, \dots, s_m \in D^*$ sein. Damit liegt genau die im Beispiel 3.8(1) skizzierte Situation vor, und es soll nun ein Design T^* gefunden werden, so daß für $\Theta \rightarrow \infty$ die Determinante

$$D(T, \Theta) := \det \Sigma = \det \left((\langle \mathcal{J}\lambda_i, \mathcal{J}\lambda_j \rangle_{K^\Theta})_{\lambda_i, \lambda_j \in T} \right) \quad (3.37)$$

maximal wird (siehe Beispiel 3.8(1)). Die Matrixelemente von Σ lassen sich dabei entsprechend der Bemerkung 3.3 berechnen.

Die Varianzen von δ_{t_i} sowie D_{t_j} hängen nicht von der Wahl der t_i bzw. t_j ab. Es gilt $\text{Var } \delta_{t_i} = 1$ und $\text{Var } D_{s_j} = -\Theta g''(0)$, wobei $g(t) = \ln K(t)$ bedeuten soll. Bezüglich der Minimierung der Determinante in (3.37) ist es demnach äquivalent, statt der dort verwendeten Kovarianzmatrix Σ zur Korrelationsmatrix Ψ überzugehen. Schreibe Ψ als die folgende Blockmatrix:

$$\Psi = \begin{pmatrix} \Psi^{00} & \Psi^{10} \\ (\Psi^{10})^\top & \Psi^{11} \end{pmatrix}.$$

Dabei enthält die Teilmatrix Ψ^{00} alle Korrelationsterme zwischen den in T enthaltenen Deltafunktionalen $\delta_{t_i}, \delta_{t_j}$, die Teilmatrix Ψ^{11} dagegen alle Korrelationsterme zwischen den in T enthaltenen Differentiationsfunktionalen D_{s_i} und D_{s_j} . Die Teilmatrix Ψ^{10} schließlich enthält die Korrelationsterme zwischen je einem Delta- und einem Differentiationsfunktional. Die Elemente in Ψ^{00} haben demnach die folgende Form

$$\frac{\text{Cov}(\delta_{t_i}, \delta_{t_j})}{1} = K^\Theta(t_i - t_j).$$

Die Elemente aus Ψ^{10} sind von der Form:

$$\frac{\text{Cov}(\delta_{t_i}, D_{s_j})}{\sqrt{-\Theta g''(0)}} = -\sqrt{\frac{\Theta}{-g''(0)}} K^\Theta(t_i - s_j) g'(t_i - s_j),$$

und die Elemente aus Ψ^{11} schreiben sich wie

$$\frac{\text{Cov}(D_{s_i}, D_{s_j})}{-\Theta g''(0)} = \frac{\Theta}{g''(0)} K^\Theta(s_i - s_j) g'(s_i - s_j)^2 + \frac{1}{g''(0)} K^\Theta(s_i - s_j) g''(s_i - s_j).$$

Zur Berechnung von $\det \Psi$ wird die Determinante nach Diagonalelementen von Ψ entwickelt (siehe [3]). Zwecks einfacherer Darstellung seien die $(k+m)^2$ Elemente von Ψ durch das in T vorgegebene Schema indiziert.

$$\begin{aligned} \det \Psi &= \prod_l \Psi_{ll} - \sum_i \sum_{j < i} \Psi_{ij} \Psi_{ji} \prod_{l \neq i, j} \Psi_{ll} + \dots + \det(\Psi - \mathbf{I}) \\ &= 1 - \sum_{i=2}^k \sum_{j < i} K^{2\Theta}(t_i - t_j) - \sum_{i=1}^k \sum_{j=1}^m K^{2\Theta}(t_i - s_j) \frac{\Theta}{-g''(0)} (g'(t_i - s_j))^2 \\ &\quad - \sum_{i=2}^m \sum_{j < i} \left[\frac{\Theta}{g''(0)} K^\Theta(s_i - s_j) g'(s_i - s_j)^2 \right. \\ &\quad \left. + \frac{1}{g''(0)} K^\Theta(s_i - s_j) g''(s_i - s_j) \right]^2 \\ &\quad + o(\Theta^2 \bar{K}^{2\Theta}). \end{aligned}$$

Dabei sei $\bar{K} = \max\{\bar{K}_{00}, \bar{K}_{01}, \bar{K}_{11}\}$ definiert, mit $\bar{K}_{00} = \max_{l \neq k} K(t_l - t_k)$, $\bar{K}_{01} = \max_{l, k} K(t_l - s_k)$ und $\bar{K}_{11} = \max_{l \neq k} K(s_l - s_k)$. Führe außerdem die folgenden Mengen $P_{00} = \{(t_l, t_k) : K(t_l - t_k) = \bar{K}_{00}\}$ mit der Mächtigkeit (Index) J_{00} ein, P_{01}, P_{11} bzw. J_{01}, J_{11} analog.

Die Kovarianzfunktion ist eine Funktion des euklidischen Abstands. Dadurch wird bei kleiner werdendem Abstand der beiden Elemente eines Tupels aus z.B. P_{00} die zugehörige Größe \bar{K}_{00} vergrößert. Diese minimalen Abstände bezeichne mit \bar{d}_{00} , \bar{d}_{01} und \bar{d}_{11} . Die Mengen P_{ij} enthalten also gerade die Punktpaare mit dem jeweiligen minimalen Abstand \bar{d}_{ij} .

Mit diesen Bezeichnungen erhält man für $\Theta \rightarrow \infty$:

$$\begin{aligned} \det \Psi &= 1 - \sum_{i=2}^k \sum_{j<i}^k K^{2\Theta}(t_i - t_j) - \sum_{i=1}^k \sum_{j=1}^m K^{2\Theta}(t_i - s_j) \frac{\Theta}{-g''(0)} g'(t_i - s_j)^2 \\ &\quad - \sum_{i=2}^m \sum_{j<i}^m \frac{\Theta^2}{g''(0)^2} K^{2\Theta}(s_i - s_j) g'(s_i - s_j)^4 + o(\Theta^2 \bar{K}^{2\Theta}) \\ &= 1 - \sum_{(t_i, t_j) \in P_{00}} \bar{K}_{00}^{2\Theta} - \sum_{(t_i, s_j) \in P_{01}} \bar{K}_{01}^{2\Theta} \frac{4\alpha^2 \Theta}{-g''(0)} (t_i - s_j)^4 \\ &\quad - \sum_{(s_i, s_j) \in P_{11}} \bar{K}_{11}^{2\Theta} \frac{16\alpha^4 \Theta^2}{g''(0)^2} (s_i - s_j)^4 + o(\Theta^2 \bar{K}^{2\Theta}). \end{aligned}$$

Unter Beachtung der Definition der Größen J_{00} , J_{01} und J_{11} erhält man schließlich den folgenden Ausdruck:

$$\det \Psi = 1 - J_{00} \bar{K}_{00}^{2\Theta} - J_{01} \bar{K}_{01}^{2\Theta} \frac{4\alpha^2 \Theta}{-g''(0)} \bar{d}_{01}^2 - J_{11} \bar{K}_{11}^{2\Theta} \frac{16\alpha^4 \Theta^2}{g''(0)^2} \bar{d}_{11}^4 + o(\Theta^2 \bar{K}^{2\Theta}). \quad (3.38)$$

Um $\det \Psi$ für den asymptotischen Fall $\Theta \rightarrow \infty$ zu maximieren, ist demnach die folgende Strategie anzuwenden:

- (i) Minimiere \bar{K} , also \bar{K}_{00} , \bar{K}_{01} und \bar{K}_{11} . D.h., daß die Abstände von Punktpaaren aus den Mengen P_{ij} maximiert werden müssen.
- (ii) Minimiere die Indizes J_{00} , J_{01} sowie J_{11} . D.h. die jeweilige Anzahl der Punktpaare mit minimalem Abstand ist zu minimieren.

Aus (i) folgt insbesondere, daß die minimalen Abstände \bar{d}_{00} , \bar{d}_{01} und \bar{d}_{11} maximiert werden müssen, wobei jedoch, wie aus (ii) ersichtlich ist, die Anzahl der Paare mit diesem minimalen Abstand so klein wie möglich zu halten ist. Diese Art einer rein geometrische Konstruktion heißt Maximin-Design mit minimalem Index (siehe in [48]).

Übertragen auf diesen Fall ist ein Maximin-Design zu konstruieren, welches drei Klassen von Meßpunkt-Abständen berücksichtigt: Solche zwischen je zwei Funktionsmeßorten, solche zwischen je einem Funktionsmeß- und einem Ableitungsmeßort und schließlich solche zwischen je zwei Ableitungsmeßorten. In jeder dieser Klassen ist eine Maximin-Konstruktion anzuwenden, wobei der jeweilige Beitrag zur Minimierung von $\det \Psi$ über den Ausdruck (3.38) gesteuert wird. Die einzelnen Summanden, die aus den einzelnen Abstandsklassen resultieren, werden dort unterschiedlich gewichtet, zum einen durch Potenzen von den zugehörigen Terme \bar{d}_{ij} und zum anderen durch Potenzen von Θ . Diese gewichtete Kombination

von Maximin-Design-Direktiven ist ein wichtiger Unterschied zu dem in [48] entwickelten Verfahren.

$\mathbf{D} \subset \mathbb{R}^n$ mit $n > 1$

Das soeben vorgestellte Verfahren läßt sich auf zufällige Funktionen im mehrdimensionalen Raum übertragen. Im Unterschied zum eindimensionalen Fall betrachte nun ein Gaußsches Wahrscheinlichkeitsmaß Q auf der Menge $C^1(D)$ der einmal stetig differenzierbaren Funktionen auf $D \subset \mathbb{R}^n$. Es sei $K(s, t) = K^\Theta(\|s - t\|) = \prod_{j=1}^n K_j^\Theta(t^j - s^j)$ mit $K_j(h) = e^{-\alpha_j h^2}$ ($\alpha_j > 0$) die zugehörige Kovarianzfunktion. Dabei bezeichne mit t^j die j -te Koordinate eines Vektors $t \in \mathbb{R}^n$. Auch hier gilt für die Wahl der Kovarianzfunktion das für den Fall $n = 1$ gesagte. Für das Funktionieren des nun beschriebenen Verfahrens ist es jedoch notwendig, daß die Kovarianzfunktion dieselbe Produktstruktur aufweist, wie die hier verwendete.

Als Daten über die betrachtete zufällige Funktion stehen sowohl Funktionswertmessungen $\delta_{t_1}, \dots, \delta_{t_k}$ sowie Messungen der ersten Ableitungen in Koordinatenrichtung der Funktion, $D_{\kappa_1}(s_1), \dots, D_{\kappa_m}(s_m)$, zur Verfügung. Dabei stammen sämtliche Multiindizes κ_j aus einer a-priori festgelegten Konfigurationsmenge I , und alle Meßorte seien aus der ebenfalls a-priori definierten, endlichen Menge D^* .

Die Elemente aus I sind also Multiindizes, d.h. für ein $\kappa_j \in I$ gilt $\kappa_j = (\kappa_j[1], \dots, \kappa_j[n])$, wobei hier jedoch insbesondere noch $\kappa_j[i] \in \{0, 1\}$ gilt und der zugehörige Betrag $\|\kappa\|$ genau Eins ist^s. Für ein $\kappa_j := \kappa_j(p) = (0, \dots, 0, 1, 0, \dots, 0)$ mit genau einer 1 an der p -ten Stelle bedeutet z.B. das zugehörige Funktional D_{κ} die Ableitung nach der p -ten Koordinate.

Das zugehörige Design T besteht also aus genau k Funktionswertmessungen und m Ableitungsmessungen gemäß der oben definierten Menge I . Analog zum eindimensionalen Fall besteht nun die Aufgabe der Versuchsplanung darin, ein Design T^* mit der soeben definierten Zusammensetzung zu finden, so daß die Determinante in (3.37) maximiert wird. Mit der gleichen Begründung wie im obigen Fall soll auch hier zur zugehörigen Korrelationsmatrix Ψ übergegangen werden. Die in der Matrix Σ enthaltenen Varianzterme sind hier $\text{Var } \delta_t = 1$ und $\text{Var } D_{\kappa(p)}(s) = -\Theta g_p''(0)$. Dabei sei in Verallgemeinerung zum eindimensionalen Fall die Funktion $g_p(t) := \ln K_p(t)$ definiert. Auch hier wird die Matrix Ψ in vier Blöcke zerlegt. Die Elemente aus Ψ^{00} sind dabei von der Form

$$\frac{\text{Cov}(\delta_{t_i}, \delta_{t_j})}{1} = K^\Theta(\|t_i - t_j\|).$$

Die Funktion $\iota: I \rightarrow \{1, \dots, n\}$ liefere zu einem Multiindex κ_j genau die Stelle zurück, an der die Eins steht, z.B. $\iota(\kappa(p)) = p$. Die Elemente aus dem Block Ψ^{10} erhalten damit die Gestalt

$$\frac{\text{Cov}(\delta_{t_i}, D_{\kappa_j}(s_j))}{\sqrt{-\Theta g_{\iota(\kappa_j)}''(0)}} = -\sqrt{\frac{\Theta}{-g_{\iota(\kappa_j)}''(0)}} g'_{\iota(\kappa_j)}(t_i^{\iota(\kappa_j)} - s_j^{\iota(\kappa_j)}) K^\Theta(\|t_i - s_j\|),$$

^sStrenggenommen ist I_1 keine Menge im eigentlichen Sinne, sondern eine (nicht-injektive) Abbildung von $\{1, \dots, m\}$ in die Menge der soeben beschriebenen Multiindizes. Zwecks einer Vereinfachung der Schreibweise behalte jedoch die Mengenschreibweise bei.

und diejenigen aus Ψ^{11} sehen schließlich wie folgt aus:

$$\frac{\text{Cov}(D_{\kappa_i}(s_i), D_{\kappa_j}(s_j))}{\Theta \sqrt{g''_{\iota(\kappa_i)}(0) g''_{\iota(\kappa_j)}(0)}} = \begin{cases} -\frac{\Theta}{\sqrt{g''_{\iota(\kappa_i)}(0) g''_{\iota(\kappa_j)}(0)}} g'_{\iota(\kappa_i)}(s_i^{\iota(\kappa_i)} - s_j^{\iota(\kappa_i)}) \\ \quad \times g'_{\iota(\kappa_j)}(s_i^{\iota(\kappa_j)} - s_j^{\iota(\kappa_j)}) K^\Theta(\|s_i - s_j\|) & \text{falls } \iota(\kappa_i) \neq \iota(\kappa_j) \\ \frac{\Theta}{g''_{\iota(\kappa_i)}(0)} K^\Theta(\|s_i - s_j\|) g'_{\iota(\kappa_i)}(s_i^{\iota(\kappa_i)} - s_j^{\iota(\kappa_i)})^2 \\ \quad + \frac{1}{g''_{\iota(\kappa_i)}(0)} K^\Theta(\|s_i - s_j\|) g''_{\iota(\kappa_i)}(s_i^{\iota(\kappa_i)} - s_j^{\iota(\kappa_i)}) & \text{falls } \iota(\kappa_i) = \iota(\kappa_j). \end{cases}$$

Analog zu der im eindimensionalen Fall beschriebenen Technik soll auch hier die Determinante $\det \Psi$ nach den Diagonalelementen von Ψ entwickelt werden. Man erhält in diesem Fall den folgenden Ausdruck:

$$\begin{aligned} \det \Psi &= 1 - \sum_{i=2}^k \sum_{j<i}^k K^{2\Theta}(\|t_i - t_j\|) \\ &\quad - \sum_{i=1}^k \sum_{j=1}^m -\frac{\Theta}{g''_{\iota(\kappa_j)}(0)} K^{2\Theta}(\|t_i - s_j\|) g'_{\iota(\kappa_j)}(t_i^{\iota(\kappa_j)} - t_j^{\iota(\kappa_j)})^2 \\ &\quad - \sum_{i=2}^m \sum_{\substack{j<i \\ \iota(\kappa_i) \neq \iota(\kappa_j)}}^m \frac{\Theta^2}{g''_{\iota(\kappa_i)}(0) g''_{\iota(\kappa_j)}(0)} g'_{\iota(\kappa_i)}(t_i^{\iota(\kappa_i)} - t_j^{\iota(\kappa_i)})^2 \times \\ &\quad \quad \times g'_{\iota(\kappa_j)}(t_i^{\iota(\kappa_j)} - t_j^{\iota(\kappa_j)})^2 \\ &\quad - \sum_{i=2}^m \sum_{\substack{j<i \\ \iota(\kappa_i) = \iota(\kappa_j)}}^m \left[\frac{\Theta}{g''_{\iota(\kappa_i)}(0)} K^\Theta(\|t_i - t_j\|) g'_{\iota(\kappa_i)}(t_i^{\iota(\kappa_i)} - t_j^{\iota(\kappa_i)})^2 + \right. \\ &\quad \quad \left. + \frac{1}{g''_{\iota(\kappa_i)}(0)} K^\Theta(\|t_i - t_j\|) g''_{\iota(\kappa_i)}(t_i^{\iota(\kappa_i)} - t_j^{\iota(\kappa_i)}) \right]^2 + o(\Theta^2 \bar{R}^{2\Theta}). \end{aligned}$$

Für große Θ ist der eingekästelte Teil vernachlässigbar gegenüber dem zweiten, darüberstehenden Summandenterm. Mit der für den eindimensionalen Fall eingeführten Notation vereinfacht sich die Determinante dann zu:

$$\begin{aligned} \det \Psi &= 1 - \sum_{(t_i, t_j) \in P_{00}} \bar{K}_{00}^{2\Theta} \\ &\quad - \sum_{(t_i, s_j) \in P_{01}} \frac{\Theta}{-g''_{\iota(\kappa_j)}(0)} \bar{K}_{01}^{2\Theta} g'_{\iota(\kappa_j)}(t_i^{\iota(\kappa_j)} - s_j^{\iota(\kappa_j)})^2 \\ &\quad - \sum_{(t_i, t_j) \in P_{11}} \frac{\Theta^2}{g''_{\iota(\kappa_i)}(0) g''_{\iota(\kappa_j)}(0)} \bar{K}_{11}^{2\Theta} g'_{\iota(\kappa_i)}(t_i^{\iota(\kappa_i)} - t_j^{\iota(\kappa_i)})^2 g'_{\iota(\kappa_j)}(t_i^{\iota(\kappa_j)} - t_j^{\iota(\kappa_j)})^2 \\ &\quad + o(\Theta^2 \bar{K}^{2\Theta}). \end{aligned}$$

Definiere dabei $P_{00} = \{(t_i, t_j) : K(t_i, t_j) = \bar{K}_{00}\}$ mit der Mächtigkeit (Index) J_{00} und die Mengen P_{01} und P_{11} mit den zugehörigen Indizes J_{01} bzw. J_{11} analog. Damit kann man für $\Theta \rightarrow \infty$ das Funktional wie folgt schreiben:

$$\begin{aligned} \det \Psi &= 1 - J_{00} \bar{K}_{00}^{2\Theta} - \bar{K}_{01}^{2\Theta} \sum_{(t_i, s_j) \in P_{01}} \frac{4\Theta \alpha_{t(\kappa_j)}^2}{-g_{t(\kappa_j)}''(0)} \left(t_i^{t(\kappa_j)} - s_j^{t(\kappa_j)} \right)^2 \\ &- \bar{K}_{11}^{2\Theta} \sum_{(s_i, s_j) \in P_{11}} \frac{16\Theta^2 \alpha_{t(\kappa_i)}^2 \alpha_{t(\kappa_j)}^2}{g_{t(\kappa_i)}''(0) g_{t(\kappa_j)}''(0)} \left(s_i^{t(\kappa_i)} - s_j^{t(\kappa_i)} \right)^2 \left(s_i^{t(\kappa_j)} - s_j^{t(\kappa_j)} \right)^2 + o(\Theta^2 \bar{K}^{2\Theta}). \end{aligned} \quad (3.39)$$

Die Minimierung von $\det \Psi$ folgt hier nach sehr ähnlichen Prinzipien wie im eindimensionalen Fall.

- (i) Minimiere \bar{K} , was einer Minimierung von \bar{K}_{00} , \bar{K}_{01} und \bar{K}_{11} gleichkommt. Übersetzt heißt dies, daß die Abstände von Punktepaaren in den Mengen P_{ij} maximiert werden müssen.
- (ii) Die auftretenden Summen in (3.39) müssen minimiert werden, d.h. es sind
 - a) die Indizes J_{ij} klein zu halten und
 - b) die Abstände $(t_i^{t(\kappa_j)} - s_j^{t(\kappa_j)})$ bzw. $(s_i^{t(\kappa_i)} - s_j^{t(\kappa_i)})$ klein zu halten.

Wenn man also nur die Punkte (i) sowie (ii)a.) beachtet, ist die Designstrategie wie im eindimensionalen Fall ein Maximin-Design, wobei auch hier die drei, bereits oben genannten Klassen von Punktabständen berücksichtigt werden müssen. Der Punkt (ii).b) scheint den anderen Direktiven auf den ersten Blick zu widersprechen. Es wird hier jedoch nur gefordert, daß gewisse Koordinaten eines Punktepaars minimalen Abstand haben sollen, nicht jedoch etwa alle Koordinaten. Diese Koordinatenrichtungen sind genau diejenigen, in die Ableitungen gemessen, und als Daten für eine Vorhersage ausgewertet werden sollen.

Analog zum eindimensionalen Fall ist auch hier eine Gewichtung der Maximin-Konstruktion innerhalb der verschiedenen Abstandsklassen zu beobachten.

Beispiel 3.16. Betrachte eine zufällige Funktion in $D \subset \mathbb{R}^2$. Wie im obigem Fall sollen genau n Funktionswertmessungen und m Ableitungsmessungen für eine Vorhersage innerhalb einer endlichen Menge $D^* \subset D$ verwendet werden. Die Ableitungen sollen jedoch alle nach einer ausgezeichneten Koordinate erfolgen. Nenne diese e_1 . Für die oben definierten Multiindizes gelte also $t(\kappa_j) = 1$ für alle $j \in \{1, \dots, m\}$. Die Determinante in (3.39) erhält dann die folgende Gestalt:

$$\begin{aligned} \det \Psi &= 1 - J_{00} \bar{K}_{00}^{2\Theta} - \frac{4\Theta \alpha_1^2}{-g_1''(0)} \bar{K}_{01}^{2\Theta} \sum_{(t_i, s_j) \in P_{01}} \left(t_i^1 - s_j^1 \right)^2 \\ &- \frac{16\Theta^2 \alpha_1^4}{g_1''(0)^2} \bar{K}_{11}^{2\Theta} \sum_{(s_i, s_j) \in P_{11}} \left(s_i^1 - s_j^1 \right)^4 + o(\Theta^2 \bar{K}^{2\Theta}). \end{aligned} \quad (3.40)$$

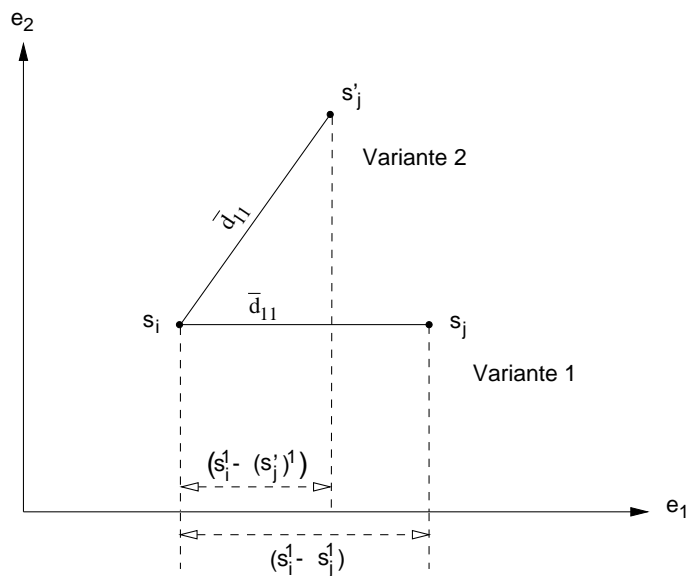


Abbildung 3.1.: Zwei mögliche Konstellationen von Ableitungsmessungen an den Punkten s_i, s_j bzw. s_j' . e_1 ist hier die Koordinate, nach der die Ableitungen gebildet werden.

In Abbildung 3.1 sind zwei mögliche Varianten der Plazierung von je zwei Ableitungspunkten dargestellt. Dabei sei der Abstand der minimal auftretende zwischen je zwei Ableitungspunkten innerhalb eines gegebenen Designs, \bar{d}_{11} , in der oben eingeführten Bezeichnungswiese. Die zweite Variante ist jedoch gegenüber der ersten die bessere, da der Abstand der e_1 -Koordinaten dort kleiner ist als bei der ersten Variante und somit der Beitrag in der entsprechenden Summe (über Paare aus P_{11}) in 3.40 kleiner ist.

3.4. Splinemethoden

Wie bereits in der Einleitung zu diesem Kapitel erwähnt wurde, sind die Splinemethoden in die Reihe der deterministischen Rekonstruktionsverfahren einzuordnen. Dieses Approximationsverfahren ist linear in den verwendeten Daten und beschränkt sich auf die Suche nach einer ausgezeichneten Funktion innerhalb eines a-priori festgelegten Funktionenraumes. Genauer gesagt, werden RKHS verwendet, und die ausgezeichnete Funktion ist der Interpolator bezüglich der gegebenen Daten, der die minimale Norm in diesem RKHS unter allen dort zu diesem Problem existierenden Interpolatoren hat. Wenngleich also die Ausgangssituation wie auch die konkrete Konstruktion von Kriging- bzw. Splinemethoden unterschiedlich sind, so liefern die Verfahren für den Fall, daß auf der Krigingseite zufällige Funktionen mit Gaußscher Verteilung benutzt werden, die gleichen Ergebnisse.

Die allgemeine Äquivalenz der beiden Verfahren wird im folgenden noch genauer untersucht. Doch zuvor soll, quasi als Einführung in die Funktionsweise von Splinemethoden, gezeigt werden, daß der BLP mit verschwindendem Mittelwert sowohl als das Ergebnis von (punktweise angewendeten) Splinemethoden wie auch von Kriging-Methoden angesehen werden kann.

Äquivalenz bei Verwendung des BLP

Das Ergebnis verwendet die gleichen Voraussetzungen, wie die in Abschnitt 3.2.1 gemachten. Dort war von einer Verteilung Q auf der Menge $F \subset C^r(D)$ ausgegangen worden. D war als Teilmenge des \mathbb{R}^n gewählt worden, K bezeichnete die zugehörige Kovarianzfunktion, und die Erwartungswertfunktion war konstant Null. Für die Konstruktion des BLP $S_k^{t_0}$ als Approximation für $\text{Pre}(t_0)$ mit $t_0 \in D$ sollten k lineare Samplefunktionale $\lambda_1, \dots, \lambda_k \in (C^r(D))'$ verwendet werden.

Wird die in (3.10) dargestellte lineare Vorhersage $S_k^{t_0}$ auf eine Funktion $f \in F$ angewendet, so erhält man:

$$\tilde{f}(t_0) := S_k^{t_0}(f) = ((\mathcal{J}\lambda_1)(t_0), \dots, (\mathcal{J}\lambda_k)(t_0)) \boldsymbol{\Sigma}_\lambda^{-1}(\lambda_1 f, \dots, \lambda_k f)^T. \quad (3.41)$$

Diese Konstruktion kann für beliebige $t_0 \in D$ durchgeführt werden, und man erhält auf diese Weise eine lineare Rekonstruktion $S_k^D : f \mapsto \tilde{f}$ auf der Basis der bereits genannten Samplefunktionale (siehe in Abschnitt 3.1.1)).

Satz 3.17. *Die Funktion $\tilde{f} : D \rightarrow \mathbb{R}$ erfüllt die folgenden drei Bedingungen:*

- (i) $\tilde{f} \in V_\lambda := \text{span}\{\mathcal{J}\lambda_1, \dots, \mathcal{J}\lambda_k\}$
- (ii) $\langle \mathcal{J}\lambda_j, \tilde{f} \rangle_K = \lambda_j f$ für alle $j = 1, \dots, k$ (Interpolationseigenschaft)
- (iii) $\|\tilde{f}\|_K = \min\{\|\mathbf{h}\|_K : \mathbf{h} \in H(K), \langle \lambda_j, \mathbf{h} \rangle = \lambda_j f \text{ für alle } j = 1, \dots, k\}$.

Damit ist \tilde{f} ein sogenannter abstrakter Interpolationsspline für die gegebenen Sampledaten $\lambda_1 f, \dots, \lambda_k f$ (siehe in [6] für eine allgemeine Definition).

Beweis. Wegen der Definition von \tilde{f} und des Raumes V_λ ist Punkt (i) klar. Betreffs Punkt (ii) fixiere ein beliebiges $i_0 \in \{1, \dots, k\}$. Erhalte damit

$$\langle \mathcal{J}\lambda_{i_0}, \tilde{f} \rangle_K = (\langle \mathcal{J}\lambda_{i_0}, \mathcal{J}\lambda_1 \rangle_K, \dots, \langle \mathcal{J}\lambda_{i_0}, \mathcal{J}\lambda_k \rangle_K) \boldsymbol{\Sigma}_\lambda^{-1}(\lambda_1 f, \dots, \lambda_k f)^T.$$

Dabei ist $(\langle \mathcal{J}\lambda_{i_0}, \mathcal{J}\lambda_1 \rangle_K, \dots, \langle \mathcal{J}\lambda_{i_0}, \mathcal{J}\lambda_k \rangle_K)$ gerade die i_0 -te Zeile von $\boldsymbol{\Sigma}_\lambda$, und somit ist das Produkt zwischen diesem Vektor und der Inversen $\boldsymbol{\Sigma}_\lambda^{-1}$ gerade ein Vektor mit lauter Nullen und nur einer Eins an der i_0 -ten Stelle. Damit ist jedoch Punkt (ii) erfüllt.

Sei $\mathbf{h} \in H(K)$ derart, daß $\langle \mathcal{J}\lambda_j, \mathbf{h} \rangle_K = \lambda_j f$ für alle $j = 1 \dots, k$ erfüllt ist. Zerlege \mathbf{h} orthogonal, d.h.:

$$\mathbf{h} = \mathbf{h}_\parallel + \mathbf{h}_\perp,$$

wobei $\mathbf{h}_\parallel \in V_\lambda$ und $\mathbf{h}_\perp \in V_\lambda^\perp$, dem orthogonalen Komplement von V_λ , ist. Für die Norm von \mathbf{h} in $H(K)$ gilt dementsprechend $\|\mathbf{h}\|_K^2 = \|\mathbf{h}_\parallel\|_K^2 + \|\mathbf{h}_\perp\|_K^2$ und für beliebige $j = 1, \dots, k$ außerdem:

$$\begin{aligned} \langle \mathcal{J}\lambda_j, \mathbf{h} \rangle_K &= \langle \mathcal{J}\lambda_j, \mathbf{h}_\parallel \rangle_K + \langle \mathcal{J}\lambda_j, \mathbf{h}_\perp \rangle_K \\ &= \langle \mathcal{J}\lambda_j, \mathbf{h}_\parallel \rangle_K \\ &= \lambda_j f \end{aligned}$$

Insbesondere ist $\langle \mathcal{J}\lambda_j, \mathbf{h}_{||} \rangle_{\mathcal{K}} = \lambda_j f$ für alle $j = 1, \dots, k$ erfüllt. $\mathbf{h}_{||}$ und \tilde{f} haben damit die gleiche Darstellung in V_λ ,

$$\mathbf{h}_{||} = \tilde{f} = \sum_{j=1}^k (\lambda_j f) \mathcal{J}\lambda_j$$

und somit die gleiche Norm. Es folgt $\|\tilde{f}\|_{\mathcal{K}} \leq \|\mathbf{h}_{||}\|_{\mathcal{K}}$, und somit ist Punkt (iii) erfüllt. \square

Die lineare Rekonstruktion S_k^D wird entsprechend auch (abstrakte) Splinemethode genannt und liefert zu jeder Funktion $f \in F$ die zugehörige (abstrakte) Splinefunktion \tilde{f} zurück. Dabei erfüllt \tilde{f} genau die oben genannten drei Punkte. Zusätzlich entspricht \tilde{f} punktweise einer Realisierung des BLP mit den Daten $\lambda_1 f, \dots, \lambda_k f$.

Bemerkung 3.18. Bei Approximationsproblemen App_q mit $q > 2$, bedarf es Zusatzannahmen, damit auch hier die Splinemethode die beste (bezüglich des Fehlers $e_q(S_n, \text{App}_q, Q)$), in den Daten lineare Approximation ergibt. Verantwortlich ist hierfür im wesentlichen, daß dann kein Approximationsproblem in einem Hilbertraum vorliegt. Ist Q z.B. ein Gaußsches Maß, so bleibt die Optimalität von Splinemethoden erhalten (siehe in [44]).

3.4.1. Allgemeine Äquivalenz zwischen dem BLP und Splinemethoden

Wie bereits in der Einleitung zum übergeordneten Abschnitt angekündigt wurde, soll nun die allgemeine Äquivalenz zwischen dem BLP und Splinemethoden für den Fall gezeigt werden, daß fehlerbehaftete Sampledaten zur Verfügung stehen. In der Literatur sind bisher in den meisten Fällen sogenannte polynomiale Glättungssplines behandelt worden (siehe z.B. in in [24], [28], [35], [60] und auch in Abschnitt 3.1.2). Diese Analogie ist jedoch auch in einem allgemeineren Kontext, wie er im folgenden dargestellt werden soll, gültig. In [44] wird die Optimalität des allgemeinen Glättungsspline-Verfahrens bereits genannt. Dort liegt der Schwerpunkt jedoch auf dem Beweis eben dieser Optimalität und nicht darauf, die Analogie zum Kriging mit Meßfehlern aufzuzeigen. Diese Analogie soll dagegen hier hergeleitet werden, wobei ein Schwerpunkt auf der Entwicklung der benötigten mathematischen Hilfsmittel liegt.

Aus der Sicht von Kriging ist der wesentliche Unterschied zu der oben dargestellten Situation derjenige, daß nun Sampledaten mit Meßfehlern zur Verfügung stehen. Jede Messung durch ein Samplefunktional λ_j wird durch einen zufälligen Fehler „verrauscht“. Dabei seien Meßfehler verschiedener Messungen unkorreliert, jedoch von gleicher Struktur. So sei er im Mittel Null und habe die Varianz σ^2 . Splines sollen stets so, wie bereits oben dargestellt wurde, in einer sehr allgemeinen Definition nach ATTEIA verstanden werden (siehe in [6], [59]).

Im Gegensatz zu einer „fehlerfreien“ Approximation, die von unverfälschten Daten ausgeht, liegen bei der Approximation mit fehlerbehafteten Daten die eigentlichen Nutzdaten in der Messung nicht direkt vor, sondern sind durch einen zufälligen Fehler verrauscht. Ist die Fehlervarianz σ^2 nicht bekannt, so muß, anschaulich gesprochen, zunächst abgewogen werden, inwieweit man den Daten „traut“, um eine Approximation nach kleinsten Quadraten durchzuführen.

Betrachte, wie im vorherigen Abschnitt, eine Verteilung Q auf einer Menge $F \subset C^r(D)$, wobei auch hier der Grad der Differenzierbarkeit r geeignet gewählt wurde und D eine Teilmenge des \mathbb{R}^n ist. Q habe eine verschwindende Erwartungswertfunktion, und die Kovarianzfunktion sei mit K bezeichnet. Weiter sei P eine Verteilung auf \mathbb{R}^k , ebenfalls mit verschwindendem Erwartungswert und mit der Kovarianzfunktion $R_{ij} = \delta_{i,j}\sigma^2$ ($\sigma^2 < \infty$). Dabei ist $\delta_{ij} = 1$, falls $i = j$ und Null sonst. Die Verteilungen Q und P seien stochastisch unabhängig.

Für eine lineare Vorhersage sollen k Samplefunktionale verwendet werden, wobei diese nun auf dem Produktraum $F \times \mathbb{R}^k$ operieren. Nenne diese

$$c_j = \lambda_j + \pi_j \quad , j = 1, \dots, k. \quad (3.42)$$

π_j bezeichne die Projektion eines Vektors aus dem \mathbb{R}^k auf seine j -te Komponente.

Skalarprodukt auf $F \times \mathbb{R}^k$

Für beliebige $(f_1, \varepsilon_1), (f_2, \varepsilon_2) \in F \times \mathbb{R}^k$ definiere das Skalarprodukt

$$\langle (f_1, \varepsilon_1), (f_2, \varepsilon_2) \rangle_{F \times \mathbb{R}^k} := \langle f_1, f_2 \rangle_2 + \langle \varepsilon_1, \varepsilon_2 \rangle_2, \quad (3.43)$$

wobei die Skalarprodukte $\langle \cdot, \cdot \rangle_2$ jeweils die kanonischen L_2 -Skalarprodukte auf den Mengen F bzw. \mathbb{R}^k sind. Es gilt also speziell $\langle \varepsilon_1, \varepsilon_2 \rangle_2 = \sum_{j=1}^k \varepsilon_1^j \varepsilon_2^j$. Der Raum $F \times \mathbb{R}^k$ ist somit wiederum ein Hilbertraum (siehe in [22]), demnach kann bei der Konstruktion der besten linearen Approximation, wie unten gezeigt wird, auf die bereits oben angewendete Technik der orthogonalen Projektion zurückgegriffen werden.

Lineare Funktionale auf $F \times \mathbb{R}^k$

Sei ϕ ein lineares Funktional auf $F \times \mathbb{R}^k$, so gibt es nach dem Darstellungssatz von Riesz ein eindeutiges Tupel $(f_\phi, \varepsilon_\phi) \in F \times \mathbb{R}^k$, so daß für alle $(f, \varepsilon) \in F \times \mathbb{R}^k$ gilt:

$$\phi[(f, \varepsilon)] = \langle (f_\phi, \varepsilon_\phi), (f, \varepsilon) \rangle_{F \times \mathbb{R}^k}.$$

Nach Definition des Skalarprodukts auf $F \times \mathbb{R}^k$ gilt also insbesondere

$$\phi[(f, \varepsilon)] = \lambda_\phi(f) + \pi_\phi(\varepsilon) \quad (3.44)$$

mit $\lambda_\phi(f) := \langle f_\phi, f \rangle_2$ und $\pi_\phi(\varepsilon) := \langle \varepsilon_\phi, \varepsilon \rangle_2$.

Bemerkung 3.19. Nach dem Satz von Riesz sind alle linearen Funktionale auf $F \times \mathbb{R}^k$ von dieser Form.

Identifiziere also fortan alle linearen Funktionale ϕ auf dem Raum $F \times \mathbb{R}^k$ mit Tupeln (λ_ϕ, π_ϕ) , wobei jeweils $\lambda_\phi \in F'$ und $\pi_\phi \in (\mathbb{R}^k)'$, den zugehörigen Dualräumen, sind. Die Auswertung für ein beliebiges Tupel (f, ε) funktioniert dann wie unter (3.44) dargestellt.

Definiere schließlich die Menge der sogenannten Standardfunktionale,

$$\Lambda^{\text{std}} := \text{span}\{ \{(\delta_t, 0) : t \in D\} \cup \{(0, \pi_i) : i \in \{1, \dots, k\}\} \}. \quad (3.45)$$

Skalarprodukt auf Λ^{std}

Für zwei Funktionale $\phi_1 = (\delta_{t_1}, \pi_{i_1}), \phi_2 = (\delta_{t_2}, \pi_{i_2}) \in \Lambda^{std}$ sei deren Skalarprodukt entsprechend dem gewöhnlichen L_2 -Produkt definiert, d.h. es gilt:

$$\begin{aligned}
 \langle \phi_1, \phi_2 \rangle_\Lambda &:= \int_{F \times \mathbb{R}^k} \phi_1[(f, \varepsilon)] \phi_2[(f, \varepsilon)] d(P \otimes Q)((f, \varepsilon)) \\
 &= \int_{F \times \mathbb{R}^k} \left\{ f(t_1) + \varepsilon^{i_1} \right\} \left\{ f(t_2) + \varepsilon^{i_2} \right\} d(P \otimes Q)((f, \varepsilon)) \\
 &= \int_{F \times \mathbb{R}^k} f(t_1) f(t_2) + f(t_1) \varepsilon^{i_2} + f(t_2) \varepsilon^{i_1} + \varepsilon^{i_1} \varepsilon^{i_2} d(P \otimes Q)((f, \varepsilon)) \\
 &= \int_F f(t_1) f(t_2) dP(f) + \int_{\mathbb{R}^k} \varepsilon^{i_1} \varepsilon^{i_2} dQ(\varepsilon) \\
 &=: \langle \delta_{t_1}, \delta_{t_2} \rangle_{\Lambda^1} + \langle \pi_{i_1}, \pi_{i_2} \rangle_{\Lambda^2}.
 \end{aligned}$$

Dabei wurde in der vorletzten Zeile ausgenutzt, daß die Verteilungen P und Q unabhängig sind, sowie einen verschwindenden Mittelwert haben. Für die gewählten Funktionale lauten die in der letzten Zeile definierten Skalarprodukte einfach $\langle \delta_{t_1}, \delta_{t_2} \rangle_{\Lambda^1} = K(t_1, t_2)$ bzw. $\langle \pi_{i_1}, \pi_{i_2} \rangle_{\Lambda^2} = R_{i_1 i_2}$.

Schließe nun den Raum Λ^{std} bezüglich dieses Skalarprodukts ab, und bezeichne den so entstehenden Raum mit Λ^{all} .

Bemerkung 3.20. Aufgrund der Struktur des Skalarprodukts $\langle \cdot, \cdot \rangle_\Lambda$ werden solche Funktionale in Λ^{all} identifiziert, die $(P \otimes Q)$ -fast sicher auf $F \times \mathbb{R}^n$ übereinstimmen.

Konstruktion eines RKHS

Im Sinne einer Übertragung des Approximationsproblems auf Splinemethoden ist es erforderlich, einen zum Raum Λ^{all} zufälliger Funktionale isomorphen Kernreproduzierenden Hilbertraum (RKHS) zu konstruieren.

Sei $H = H(K) \times H(R)$. So definiere wie folgt ein Skalarprodukt auf H :

$$\langle (f_1, \varepsilon_1), (f_2, \varepsilon_2) \rangle_H := \langle f_1, f_2 \rangle_{H(K)} + \langle \varepsilon_1, \varepsilon_2 \rangle_{H(R)}, \quad (f_1, \varepsilon_1), (f_2, \varepsilon_2) \in H.$$

Bezeichne mit (K, R) den reproduzierenden Kern auf H . Um dieser Schreibweise jedoch Sinn zu verleihen, muß zunächst der Auswertungsmechanismus von Elementen aus H definiert werden. Man nutzt dazu die Reproduzierenden-Eigenschaft des Kerns aus. Für ein $(f, \varepsilon) \in H$ funktioniert der Mechanismus an der Stelle $(s, i) \in \mathbb{R}^n \times \{1, \dots, k\}$ wie folgt:

$$\begin{aligned}
 (f, \varepsilon)[(s, i)] &= \langle (K(s, \cdot), R_i), (f, \varepsilon) \rangle_H \\
 &= \langle K(s, \cdot), f \rangle_{H(K)} + \langle R_i, \varepsilon \rangle_{H(R)} \\
 &= f(s) + \varepsilon^i.
 \end{aligned}$$

Dabei ist R_i die i -te Zeile der Kovarianzmatrix, die zur Verteilung Q gehört. Es gilt also $(f, \varepsilon)[(s, i)] := f(s) + \varepsilon^i$.

Bemerkung 3.21. Es gilt

$$\text{Cov} \left((f(\mathbf{t}) + \varepsilon^i)(f(\mathbf{s}) + \varepsilon^j) \right) = (\mathbf{K}(\mathbf{t}, \square), \mathbf{R}_{i\Delta})_{|(\square=\mathbf{s}, \Delta=j)}.$$

In dieser Schreibweise ist also das Tupel (\mathbf{K}, \mathbf{R}) tatsächlich ein Kovarianzkern.

Isomorphismus zwischen \mathbf{H} und Λ^{all}

Sei $\phi \in \Lambda^{\text{std}}$ mit $\phi = \sum_{j=1}^m (a_j \delta_{t_j}, b_j \pi_{i_j})$, wobei für alle $j = 1, \dots, m$ die Koeffizienten a_j, b_j reellwertig sind, die Punkte t_j aus D stammen und die Indizes i_j aus der Menge $\{1, \dots, k\}$.

Definiere die Abbildung \mathcal{J} auf Λ^{std} mit

$$\begin{aligned} \mathcal{J}\phi &= \mathcal{J} \left(\sum_{j=1}^m (a_j \delta_{t_j}, b_j \pi_{i_j}) \right) \\ &\stackrel{!}{=} \sum_{j=1}^m (a_j \mathbf{K}(t_j, \cdot), b_j \mathbf{R}_{i_j}). \end{aligned}$$

Ergänzt man \mathcal{J} auf ganz Λ^{all} , so erhält man den gewünschten Isomorphismus. \mathcal{J} ist isometrisch, wie folgende Rechnung zeigt.

Das Normquadrat des oben definierten ϕ ist

$$\begin{aligned} \|\phi\|_{\Lambda^{\text{all}}}^2 &= \langle \phi, \phi \rangle_{\Lambda^{\text{all}}} \\ &= \sum_{j=1}^m \sum_{l=1}^m \langle (a_j \delta_{t_j}, b_j \pi_{i_j}), (a_l \delta_{t_l}, b_l \pi_{i_l}) \rangle_{\Lambda^{\text{all}}} \\ &= \sum_{j=1}^m \sum_{l=1}^m \{ a_j a_l \langle \delta_{t_j}, \delta_{t_l} \rangle_{\Lambda^1} + b_j b_l \langle \pi_{i_j}, \pi_{i_l} \rangle_{\Lambda^2} \} \\ &= \sum_{j=1}^m \sum_{l=1}^m \left\{ a_j a_l \int_{\mathbb{F}} f(t_j) f(t_l) dP(f) + b_j b_l \int_{\mathbb{R}^n} \varepsilon_{i_j} \varepsilon_{i_l} dQ(\varepsilon) \right\} \\ &= \sum_{j=1}^m \sum_{l=1}^m \{ a_j a_l \mathbf{K}(t_j, t_l) + b_j b_l \mathbf{R}_{i_j i_l} \}. \end{aligned}$$

Das Normquadrat des Bildes von ϕ unter \mathcal{J} ist

$$\begin{aligned} \|\mathcal{J}\phi\|_{\mathbf{H}}^2 &= \langle \mathcal{J}\phi, \mathcal{J}\phi \rangle_{\mathbf{H}} \\ &= \sum_{j=1}^m \sum_{l=1}^m \langle (a_j \mathbf{K}(t_j, \cdot), b_j \mathbf{R}_{i_j}), (a_l \mathbf{K}(t_l, \cdot), b_l \mathbf{R}_{i_l}) \rangle_{\mathbf{H}} \\ &= \sum_{j=1}^m \sum_{l=1}^m \left\{ a_j a_l \langle \mathbf{K}(t_j, \cdot), \mathbf{K}(t_l, \cdot) \rangle_{\mathbf{H}(\mathbf{K})} + b_j b_l \langle \mathbf{R}_{i_j}, \mathbf{R}_{i_l} \rangle_{\mathbf{H}(\mathbf{R})} \right\} \\ &= \sum_{j=1}^m \sum_{l=1}^m \{ a_j a_l \mathbf{K}(t_j, t_l) + b_j b_l \mathbf{R}_{i_j i_l} \} = \|\phi\|_{\Lambda^{\text{all}}}^2. \end{aligned}$$

Skalarprodukt in $H(\mathbb{R})$

Das Skalarprodukt in $H(\mathbb{R})$ kann konkret angegeben werden, da der Raum isomorph zum \mathbb{R}^k mit dem gewöhnlichen L_2 -Produkt ist.

Nehme einen beliebigen Vektor $x \in \mathbb{R}^k$, und schreibe ihn in der Basisdarstellung $s = \sum_{j=1}^k x^j e_j$. Die Vektoren e_j , $j = 1, \dots, k$ sind dabei die Einheitsvektoren mit einer Eins an der j -ten Position. Wegen der Reproduzierenden-Eigenschaft erhält man

$$\begin{aligned} \langle x, R_l \rangle_{H(\mathbb{R})} &= \sum_{j=1}^k x^j \langle e_j, R_l \rangle_{H(\mathbb{R})} \\ &= \sum_{j=1}^k x^j \langle e_j, \sigma^2 e_l \rangle_{H(\mathbb{R})} \\ &\stackrel{!}{=} x^l. \end{aligned}$$

Daraus folgt $\langle e_j, e_l \rangle_{H(\mathbb{R})} = 1/\sigma^2 \delta_{jl}$. Für allgemeine $x, y \in \mathbb{R}^k$ erhält man die Umrechnungsformel

$$\langle x, y \rangle_{H(\mathbb{R})} = \frac{1}{\sigma^2} \langle x, y \rangle_2. \quad (3.46)$$

Lösung des Approximationsproblems

Betrachte das Vorhersageproblem $\text{Pre}(t_0) := (\delta_{t_0}, 0)$ mit $t_0 \in D$, das durch die lineare Methode $S_k^{t_0} = \sum_{j=1}^k a_j(t_0)(\lambda_j, \pi_j)$ bezüglich des Fehlers $e_2(S_k^{t_0}, \text{Pre}(t_0), Q \times P) = \|S_k^{t_0} - \text{Pre}(t_0)\|_\Lambda$ approximiert werden soll. Es gilt:

$$\begin{aligned} \|S_k^{t_0} - \text{Pre}(t_0)\|_\Lambda &= \|(\delta_{t_0}, 0) - \sum_{j=1}^k a_j(t_0)(\lambda_j, \pi_j)\|_\Lambda \\ &= \|\mathcal{J}(\delta_{t_0}, 0) - \mathcal{J} \sum_{j=1}^k a_j(t_0)(\lambda_j, \pi_j)\|_H \\ &= \|(\mathcal{K}(t_0, \cdot) - \sum_{j=1}^k a_j(t_0)(\mathcal{J}\lambda_j, R_j))\|_H \\ &\stackrel{\text{s. Anhang}}{=} \sup_{(h, p) \in B} \left| \langle (\mathcal{K}(t_0, \cdot), 0), (h, p) \rangle_H - \sum_{j=1}^k a_j(t_0) \langle (\mathcal{J}\lambda_j, R_j), (h, p) \rangle_H \right| \\ &= \sup_{(h, p) \in B} \left| h(t_0) - \sum_{j=1}^k a_j(t_0) \{ \lambda_j(h) + p_j \} \right|. \end{aligned}$$

Dabei sei B die Einheitskugel in H . Die rechte Seite entspricht dabei genau einem linearen Problem mit den exakten Daten $c_j = \lambda_j(f) + \varepsilon^j$ für $j = 1, \dots, k$. Die Lösung von

diesem Problem ist gerade der interpolierenden Spline $(\tilde{\mathbf{h}}, \tilde{\mathbf{p}}) \in \mathbf{H}$ (siehe in [59]). Dies ist definitionsgemäß aber gerade die Lösung des Optimierungsproblems

$$\|(\mathbf{h}, \mathbf{p})\|_{\mathbf{H}} = \|\mathbf{h}\|_{\mathbf{H}(\mathbf{K})} + \frac{1}{\sigma^2} \sum_{j=1}^k (\mathbf{p}^j)^2 \rightarrow \min$$

auf $\{(\mathbf{h}, \mathbf{p}) \in \mathbf{H} : \lambda_j(\mathbf{h}) + \mathbf{p}^j = \mathbf{c}_j, j = 1, \dots, n\}$. (3.47)

Andererseits handelt es sich bei dem gestellten Problem, formal gesehen, um ein lineares Approximationsproblem auf dem Raum $\mathbf{F} \times \mathbb{R}^k$. Die Lösung $\mathbf{S}_k^{\mathbf{t}_0}$ ist also die Projektion von $\text{Pre}(\mathbf{t}_0)$ auf den Raum $\text{span}\{(\lambda_1, 0) : j = 1, \dots, k\} \cup \{(0, \pi_j) : j = 1, \dots, k\}$, die zusätzlich die Daten $\mathbf{c}_1, \dots, \mathbf{c}_n$ interpoliert. Es gilt also:

$$\begin{aligned} & \left\langle \mathcal{J}\mathbf{S}_k^{\mathbf{t}_0} - \mathcal{J}\text{Pre}(\mathbf{t}_0), \mathcal{J}\mathbf{S}_k^{\mathbf{t}_0} \right\rangle_{\mathbf{H}} \stackrel{!}{=} 0 \\ \Leftrightarrow & \left\langle \mathcal{J}\mathbf{S}_k^{\mathbf{t}_0}, \mathcal{J}\mathbf{S}_k^{\mathbf{t}_0} \right\rangle_{\mathbf{H}} = \left\langle \mathcal{J}\text{Pre}(\mathbf{t}_0), \mathcal{J}\mathbf{S}_k^{\mathbf{t}_0} \right\rangle_{\mathbf{H}} \\ \Leftrightarrow & \sum_{i,j=1}^k \mathbf{a}_i(\mathbf{t}_0) \mathbf{a}_j(\mathbf{t}_0) \langle (\mathcal{J}\lambda_i, \mathbf{R}_i), (\mathcal{J}\lambda_j, \mathbf{R}_j) \rangle_{\mathbf{H}} = \sum_{j=1}^k \mathbf{a}_j(\mathbf{t}_0) \langle (\mathcal{J}\lambda_j, \mathbf{R}_j), (\mathbf{K}(\mathbf{t}_0, \cdot), 0) \rangle_{\mathbf{H}} \\ \Leftrightarrow & \mathbf{a}(\mathbf{t}_0)^{\mathbf{T}} (\boldsymbol{\Sigma}_{\lambda} + \mathbf{R}) \mathbf{a}(\mathbf{t}_0) = \mathbf{a}(\mathbf{t}_0)^{\mathbf{T}} \boldsymbol{\Sigma}_{\mathbf{t}_0} \\ \Rightarrow & \mathbf{a}(\mathbf{t}_0) = (\mathbf{a}_1(\mathbf{t}_0), \dots, \mathbf{a}_k(\mathbf{t}_0))^{\mathbf{T}} = (\boldsymbol{\Sigma}_{\lambda} + \mathbf{R})^{-1} \boldsymbol{\Sigma}_{\mathbf{t}_0}. \end{aligned}$$

Mit $\boldsymbol{\Sigma}_{\mathbf{t}_0} = ((\mathcal{J}\lambda_1, 0), \dots, (\mathcal{J}\lambda_k, 0))(\mathbf{t}_0)^{\mathbf{T}} = ((\mathcal{J}\lambda_1)(\mathbf{t}_0), \dots, (\mathcal{J}\lambda_k)(\mathbf{t}_0))^{\mathbf{T}}$ erhält man die erste Komponente der Approximation,

$$\tilde{\mathbf{h}}(\mathbf{t}) := (\mathbf{c}_1, \dots, \mathbf{c}_k)(\boldsymbol{\Sigma}_{\lambda} + \mathbf{R})^{-1} \boldsymbol{\Sigma}_{\mathbf{t}}. \quad (3.48)$$

Die Approximation soll außerdem die Daten interpolieren. Mit einem Ansatz (\mathbf{h}, \mathbf{p}) , wobei $\mathbf{p} \in \mathbb{R}^k$ ist, muß also für alle $i \in \{1, \dots, k\}$ gelten:

$$(\lambda_i, \pi_i) \left[(\tilde{\mathbf{h}}, \tilde{\mathbf{p}}) \right] = \mathbf{c}_i.$$

Man erhält mit dieser Bedingung:

$$\tilde{\mathbf{p}} = (\mathbf{c}_1, \dots, \mathbf{c}_k)(\boldsymbol{\Sigma}_{\lambda} + \mathbf{R})^{-1} \mathbf{R}. \quad (3.49)$$

Fügt man die beiden Komponenten $\tilde{\mathbf{h}}$ und $\tilde{\mathbf{p}}$ zusammen, so erhält man die gesuchte Approximation

$$(\tilde{\mathbf{h}}, \tilde{\mathbf{p}})[(\mathbf{t}_0, \mathbf{i})] = (\mathbf{c}_1, \dots, \mathbf{c}_k)(\boldsymbol{\Sigma}_{\lambda} + \mathbf{R})^{-1} ((\mathcal{J}\lambda_1, \mathbf{R}_1), \dots, (\mathcal{J}\lambda_k, \mathbf{R}_k))[(\mathbf{t}_0, \mathbf{i})]^{\mathbf{T}} \quad (3.50)$$

für $(\mathbf{t}_0, \mathbf{i}) \in \mathbf{D} \times \{1, \dots, k\}$.

- Die Lösung (3.50) ist genau der interpolierende Spline, der das Optimierungsproblem (3.47) löst. Es liegt also eine formale Übereinstimmung der Lösung des Approximationsproblems, welche Splinemethoden wie auch Kriging-Methoden liefern, vor.

- Der optimale „Algorithmus“, d.h. die beste lineare Methode zur oben gestellten Approximationsaufgabe, ist also

$$S_k(t_0)[(f, \varepsilon)] = S(t_0)[(\tilde{h}, \tilde{p})] = \tilde{h}(t_0). \quad (3.51)$$

Die Schätzung der Meßfehler ist ein „Nebenprodukt“, das bei der Berechnung des Splines anfällt.

- Die Varianz des Meßfehlers beeinflusst das qualitative Verhalten des Optimierungsproblems (3.47) ganz entscheidend. Schreibe die Aufgabe dazu um:

$$\|h\|_{H(K)} + \frac{1}{\sigma^2} \sum_{j=1}^k (c_j - \lambda_j(h))^2 \rightarrow \min. \quad (3.52)$$

Dabei sei h aus dem RKHS $H(K)$.

Eine größere Meßgenauigkeit, sprich ein größeres σ^2 , führt dazu, daß der Summenterm in (3.52) sehr klein skaliert wird und somit die Lösung \tilde{h} nicht so sehr die Daten interpolieren wird. Im umgekehrten Fall, wird die Lösung in erster Linie die Daten nach kleinsten Quadraten interpolieren. Dieser Mechanismus erfüllt also in geeigneter Weise das im ersten Abschnitt geforderte Verhalten der Approximation in bezug auf die Qualität der Meßdaten.

Es gibt jedoch, so sei abschließend bemerkt, keine allgemeingültigen Gründe, Spline-Verfahren den Kriging-Verfahren vorzuziehen. Das Problem der Krigingverfahren, eine geeignete Kovarianzfunktion für das zu modellierende Merkmal zu finden, entspricht bei den Splines die geeignete Wahl eines RKHS. Hier wird in der Regel auf Standardverfahren wie die bereits erwähnten polynomialen Splines ausgewichen. Dies gibt den tatsächlichen Sachverhalt aber u.U. nur ungenügend wieder. Hier scheinen die Krigingverfahren überlegen zu sein, da es stets möglich ist, empirische Kovarianzfunktionen zu bestimmen. In [56] wird jedoch ausdrücklich davon abgeraten, so vorzugehen und stattdessen eine Matérnsche Kovarianzfunktion an die Sampledaten anzupassen. Damit wäre man auch hier wieder in parametrisierten Klassen von Kovarianzfunktionen. Schließlich gibt es bei beiden Verfahren zu bedenken, daß die Varianz des Meßfehlers bzw. der sogenannte Smoothingfaktor σ^2 geeignet bestimmt werden muß. In [60] wird hierfür Cross-Validation verwendet, was in gleicher Weise auch bei Verwendung von Kriging-Verfahren angewendet werden kann.

3.4.2. Ein splinebasiertes Vorhersageverfahren bei Brownschen Bewegungen

Zum Abschluß des Kapitels wird ein einfaches Approximationsverfahren vorgestellt, das praktisch ausschließlich auf der Benutzung von Ableitungsinformationen beruht. Zunächst soll das Verfahren auf deterministische Funktionen angewendet werden. Sei $f \in C^1([0, 1])$ die zu approximierende Funktion. Als Design verwende $T = \{\delta_0, D_{t_0}, \dots, D_{t_{k+1}}\}$. Es sei also die Funktion f an der Stelle Null, sowie die Werte ihrer ersten Ableitung $f^{(1)}$ an Stellen $0 = t_0 < t_1 < \dots < t_{k+1} = 1$ bekannt.

Das Ziel des Verfahrens ist eine Approximation von f durch eine in diesen Daten lineare Interpolation der Daten.

Konstruktion der Approximation

In einem ersten Schritt wird die Ableitung $f^{(1)}$ auf $[0, 1]$ stückweise linear durch einen Schätzer $\widehat{f}_k^{(1)}$ approximiert, wobei die Daten entsprechend dem Design $T \setminus \{\delta_0\}$ als Stützstellen verwendet werden. Für $t \in [t_j, t_{j+1}]$ und $j \in \{0, \dots, k\}$ gilt demnach:

$$\widehat{f}_k^{(1)}(t) = \frac{f^{(1)}(t_{j+1}) - f^{(1)}(t_j)}{t_{j+1} - t_j}(t - t_j) + f^{(1)}(t_j).$$

Im zweiten Schritt wird die Funktion f durch Integration von $\widehat{f}_k^{(1)}$ approximiert. Definiere also für $t \in [0, 1]$

$$\widehat{f}_k(t) = f(0) + \int_0^t \widehat{f}_k^{(1)}(s) ds.$$

Führt man außerdem die Ausdrücke

$$\begin{aligned} l_j &= \frac{f^{(1)}(t_{j+1}) - f^{(1)}(t_j)}{t_{j+1} - t_j}, \quad j \in \{0, \dots, k\}, \\ j_k(t) &= \max \{j \in \{0, \dots, k\} : t \geq t_j\} \end{aligned}$$

ein, so läßt sich \widehat{f}_k auch schreiben als

$$\begin{aligned} \widehat{f}_k(t) = f(0) + \sum_{j=0}^{j_k(t)-1} & \left[\frac{l_j}{2} (t_{j+1} - t_j)^2 + f^{(1)}(t_j)(t_{j+1} - t_j) \right] \\ & + \left[\frac{l_{j_k(t)}}{2} (t - t_{j_k(t)})^2 + f^{(1)}(t_{j_k(t)})(t - t_{j_k(t)}) \right]. \end{aligned} \quad (3.53)$$

Setzt man $t_{-1} = 0$, so läßt sich die Summe in (3.53) nach Faktoren von Ableitungen von f ordnen:

$$\begin{aligned} \widehat{f}_k(t) &= f(0) + \sum_{j=0}^{j_k(t)-1} f^{(1)}(t_j) \left[\frac{t_{j+1} - t_{j-1}}{2} \right] \\ &+ f^{(1)}(t_{j_k(t)}) \left[\frac{t_{j_k(t)} - t_{j_k(t)-1}}{2} + (t - t_{j_k(t)}) - \frac{(t - t_{j_k(t)})^2}{2(t_{j_k(t)+1} - t_{j_k(t)})} \right] \\ &+ f^{(1)}(t_{j_k(t)+1}) \left[\frac{(t - t_{j_k(t)})^2}{2(t_{j_k(t)+1} - t_{j_k(t)})} \right]. \end{aligned} \quad (3.54)$$

Bemerkung 3.22. \widehat{f}_k ist offensichtlich eine datenlineare Approximation für f . Konstruktionsbedingt ist es außerdem interpolierend, d.h. es gilt $\delta_0 \widehat{f}_k = \delta_0 f$ und $D_{t_j} \widehat{f}_k = D_{t_j} f$ für alle $j = 1, \dots, k$.

Anwendung auf zufällige Felder

Sei Q ein Wahrscheinlichkeitsmaß auf $F \subset C^1([0, 1])$ mit Erwartungswertfunktion m und Kovarianzfunktion K . Damit ist also insbesondere $m \in C^1([0, 1])$.

Bezeichne mit d_k den maximalen Abstand zweier aufeinanderfolgender Designpunkte t_j , t_{j+1} , also $d_k = \max\{t_{j+1} - t_j : j \in \{0, \dots, k\}\}$. Wähle ein beliebiges aber festes $t \in [0, 1]$ und definiere dann die lineare Methode $S_k(f) := \hat{f}_k$ für alle $f \in F$, wobei $\hat{f}_k(t)$ die in (3.54) definierte Funktion ist.

Ist nun $d_k = o(1/k)$, so ist die Methode S_k erwartungstreu für $t = 0$ und asymptotisch erwartungstreu, falls $t > 0$. D.h. es gilt dann $E S_k \rightarrow m(t)$, für $k \rightarrow \infty$.

Beweis. Für $t = 0$ ist S_k erwartungstreu gemäß Definition. Sei also $t > 0$. Aus Gleichung (3.54) erhält man:

$$\begin{aligned} E S_k &= m(0) + \sum_{j=0}^{j_k(t)-1} m^{(1)}(t_j) \left[\frac{t_{j+1} - t_{j-1}}{2} \right] + m^{(1)}(t_{j_k(t)})(t - t_{j_k(t)}) \\ &\quad + m^{(1)}(t_{j_k(t)}) \left[\frac{t_{j_k(t)} - t_{j_k(t)-1}}{2} \right] \\ &\quad + \left(m^{(1)}(t_{j_k(t)+1}) - m^{(1)}(t_{j_k(t)}) \right) \left[\frac{(t - t_{j_k(t)})^2}{2(t_{j_k(t)+1} - t_{j_k(t)})} \right]. \end{aligned} \quad (3.55)$$

Es gilt $(t - t_{j_k(t)}) = o(1/k)$, und somit verschwinden die Summanden in der zweiten und dritten Zeile in (3.55) für $k \rightarrow \infty$. Mit den Voraussetzungen folgt

$$A_k := \sum_{j=0}^k m^{(1)}(t_j)(t_{j+1} - t_j) + m^{(1)}(t_{j_k(t)})(t - t_{j_k(t)}) \xrightarrow{k \rightarrow \infty} \int_0^t m^{(1)}(s) ds = m(t) - m(0).$$

Bildet man die Differenz

$$E S_k(t) - A_k = \sum_{j=0}^k m^{(1)}(t_j) [t_j - \tilde{t}_j], \quad \text{mit } \tilde{t}_j := \frac{t_{j+1} + t_{j-1}}{2}, \quad (3.56)$$

so erhält man $(t_j - \tilde{t}_j) = o(t_j - t_{j-1})$. Es existiert also eine Nullfolge α_k , so daß $(t_j - \tilde{t}_j) = \alpha_k(t_j - t_{j-1})$. Damit konvergiert die Folge in (3.56) gegen Null, und es gilt die Behauptung. \square

Als Beispiel für eine Anwendung des Verfahrens betrachte den Fall, daß Q die Verteilung einer einfach integrierten Brownschen Bewegung ist. Die Menge F sei entsprechend der Träger dieses Maßes, also die kleinste abgeschlossene Menge $M \subset C^1([0, 1])$, so daß $Q(M) = 1$ gilt (siehe in [44]). Für ein $f \in F$ gilt demnach

$$f(t) = \int_0^t b(s) ds, \quad t \in [0, 1], \quad (3.57)$$

wobei b die Realisierung einer Brownsche Bewegung B ist. f erfüllt ferner die Bedingung $f(0) = f^{(1)}(0) = 0$. Die Kovarianzfunktion von B , die Minimumfunktion, sei mit K_0 bezeichnet. Unter Beachtung der Definition in (3.57) erhält man damit die Kovarianzfunktion $K_1(s, t) = \int_0^1 (s-u)_+(t-u)_+ du$ von Q .

Wende nun das oben beschriebene Verfahren auf ein $f \in F$ an. Dazu wird zunächst die stückweise lineare Funktion $\widehat{f}_n^{(1)}$ aus Ableitungsmessungen an den Orten $0 < t_1 < \dots < t_{n+1} = 1$ konstruiert (da $f^{(1)}(t) = 0$ für $t = 0$ als bekannt vorausgesetzt wurde, ist dort keine Messung erforderlich).

$f^{(1)}$ ist definitionsgemäß eine Realisierung einer Brownschen Bewegung, also ist $\widehat{f}_k^{(1)}$ bereits die beste Approximation aus den gegebenen Daten, d.h. es gilt

$$\widetilde{f^{(1)}} \equiv \widehat{f}_k^{(1)}.$$

Dabei ist $\widetilde{f^{(1)}}$, wie unter (3.41) angegeben, der zu diesem Problem gehörige abstrakte Spline, also

$$\widetilde{f^{(1)}}(s) = (\delta_{t_1}, \dots, \delta_{t_{k+1}}) f^{(1)} (\Sigma_\lambda^{(0)})^{-1} (K_0(t_1, s), \dots, K_0(t_{k+1}, s))^T$$

mit $\Sigma_\lambda^{(0)} := ((\mathcal{J}\delta_{t_i}, \mathcal{J}\delta_{t_j})_{K_0})_{i,j=1}^k = (K_0(t_i, t_j))_{i,j=1}^k$.

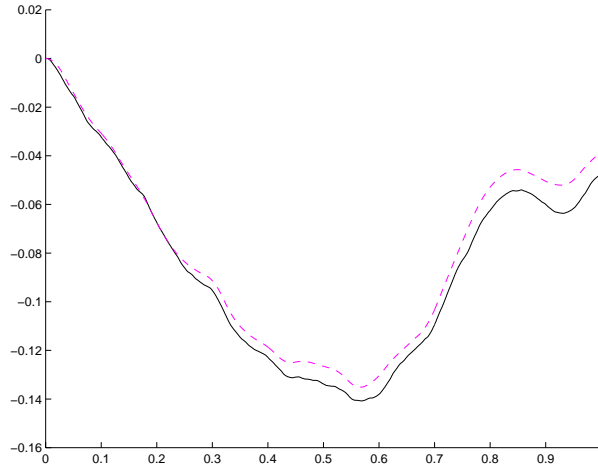


Abbildung 3.2.: Eine Realisierung von f sowie eine Approximation (gestrichelte Kurve) basierend auf 17 Ableitungsmessungen

In der zweiten Stufe des Verfahrens ist die Funktion $\widehat{f}_k^{(1)}$, wie oben allgemein gezeigt wurde, zu integrieren. Man erhält

$$\begin{aligned} \widehat{f}_k(s) &= (\delta_{t_1} f^{(1)}, \dots, \delta_{t_{k+1}} f^{(1)}) (\Sigma_\lambda^{(0)})^{-1} \int_0^s (K_0(t_1, u), \dots, K_0(t_{k+1}, u))^T du \\ &= (D_{t_1} f, \dots, D_{t_{k+1}} f) (\Sigma_\lambda^{(1)})^{-1} \left(K_1^{(1,0)}(t_1, s), \dots, K_1^{(1,0)}(t_{k+1}, s) \right)^k, \quad (3.58) \end{aligned}$$

wobei

$$\Sigma_{\lambda}^{(1)} := \left(\langle \mathcal{J}D_{t_i}, \mathcal{J}D_{t_j} \rangle_{K_1} \right)_{i,j=1}^{k+1} = \Sigma_{\lambda}^{(0)}$$

gesetzt wurde. Dabei ist bei der Berechnung des Integrals in (3.58) zu beachten, daß $K_0(s, t) = K_1^{(1,1)}(s, t)$ und $K_1^{(1,0)}(t, 0) \equiv 0$ gilt. \hat{f}_n ist somit ein Spline, der die Daten $f^{(1)}(t_1), \dots, f^{(1)}(t_{k+1})$ interpoliert. Insbesondere liefert dieses Verfahren also die beste Approximationsmethode (gemittelt über alle einfach integrierten Brownschen Bewegungen), welches diese Daten interpoliert.

In Abbildung 3.2 ist der Vergleich einer von Herrn Jaroslav Šimák erzeugten Realisierung einer simulierten, einfach integrierten Brownschen Bewegung mit dem Ergebnis des vorgestellten Approximationsverfahrens dargestellt.[¶]

[¶]Herr Šimák ist zum Zeitpunkt der Erstellung der vorliegenden Arbeit Doktorand am Graduiertenkolleg für Räumliche Statistik unter der Betreuung von Prof. Dr. W. Näther

4. Versuchsplanung für Regression

4.1. Einordnung in den Kontext

Im Unterschied zu Krigingverfahren, die von einer unbekanntem zufälligen Funktion ausgehen, welche durch diese Verfahren an einer gewissen Stelle vorhergesagt wurde, ist das Objekt der Regression eine unbekannte deterministische Funktion. Sei diese fortan mit g bezeichnet. Soweit liegt also, gemäß der Einordnung der Verfahren aus der Einleitung zu Kapitel 3, ein deterministisches Verfahren vor. Ähnlich wie bei den Smoothing-Spline-Verfahren liegen Informationen über die Funktion g in Form von Realisierungen linearer Funktionale vor, die durch Meßfehler gestört werden. Bei der gewöhnlichen Regression wird der Ansatz für die Rekonstruktion mittels einer Linearkombination bekannter sogenannter Basisfunktionen gemacht. Danach bestimmt man die gewichtete Summe der Fehlerquadrate, die durch diesen Ansatz entstehen und bestimmt die Faktoren des Ansatzes der Linearkombination so, daß eben dieses Fehlermaß minimiert wird. Der Hauptunterschied zu den genannten Smoothing-Splines besteht bei den unten beschriebenen lokalen Regressionsverfahren darin, daß eine lokale Gewichtung der Samples stattfindet. Ein wichtiger Unterschied zu den oben beschriebenen Krigingmethoden ist, daß die Samples bei der Regression als unkorreliert angesehen werden. Im allgemeinen war bei den Krigingverfahren von korrelierten Samplefunktionalen ausgegangen worden, deren Korrelationsstruktur als bekannt vorausgesetzt worden war.

Der letztgenannte Punkt mag die Krigingmethoden auf den ersten Blick als mächtigere Rekonstruktionsmethode im Vergleich zu den einfach strukturierten Regressionsmethoden erscheinen lassen, da dort eine spezifizierte Korrelationsstruktur bei der Vorhersage miteinbezogen werden kann. Allerdings ist dieser Punkt auch eine wichtige Fehlerquelle, die in den meisten Fällen nicht in die Qualität von Vorhersagen miteinbezogen wird. Denn die Bestimmung der Korrelationsstruktur unterliegt ebenfalls einem Fehler, der das Vorhersageergebnis u.U. stark verschlechtern kann (siehe in [56] zur Argumentation für die Verwendung der Matérnschen Kovarianzfunktion). Dagegen machen Regressionsverfahren keine Annahmen über spezielle Korrelationsstrukturen sondern gehen davon aus, daß die einzige Zufälligkeit, die bei den Samples auftritt, durch unkorrelierte Meßfehler bedingt ist. Erscheinen diese Verfahren vielleicht aus stochastischer Sicht einfach, so sind sie eine sinnvolle Methode bei einer ersten Trendschätzung im Rahmen einer Erkundungskampagne für ein bestimmtes zu untersuchendes Merkmal. Dies ist insbesondere vor dem Hintergrund einer nachfolgenden Schätzung der Kovarianzstruktur wichtig, da die Verfahren hierfür in der Regel von trendbereinigten zufälligen Funktionen ausgehen (siehe in [12], [37]).

Betrachte fortan also eine Funktion $g \in C^r(D)$, wobei $D \subset \mathbb{R}^n$ ist. Im folgenden wird der Übersichtlichkeit halber jedoch stets auf den eindimensionalen, $n = 1$, Fall zurückgegrif-

fen. Es seien ferner Informationen über die Funktion g in Form von meßfehlerbehafteten Realisierungen linearer Samplefunktionale vorhanden. D.h. es steht für die Regression ein Vektor $Z^T = (Z_1, \dots, Z_k)$ von zufälligen Größen zur Verfügung, mit $Z_j = \lambda_j g + \varepsilon_j$, wobei die zufälligen Größen ε_j und deren Realisierungen gleich bezeichnet werden. Die Samplefunktionale λ_j seien linear und beschränkt, d.h. sie stammen aus dem Dualraum $(C^r(D))'$. Für den Samplevektor gelte weiter:

$$\begin{aligned} E Z^T &= (\lambda_1 g, \dots, \lambda_k g) \\ \text{Cov } Z &= \text{Cov } \varepsilon = \text{diag}(\text{Var } \varepsilon_1, \dots, \text{Var } \varepsilon_k). \end{aligned}$$

Betrachte nun die Schätzung von g an einem Punkt $t_0 \in D$ mittels des Ansatzes $\hat{g}(t_0) = \sum_{j=0}^p \beta_j f_j(t_0)$, wobei die sogenannten Basisfunktionen f_j als bekannt und linear unabhängig vorausgesetzt sind. Vielfach werden Polynome bis zu einer gewissen Ordnung verwendet, insbesondere die unten betrachtete lokale polynomiale Regression verwendet diese Ansatzfunktionen. Es sind jedoch z.B. auch trigonometrische Funktionen verwendbar (siehe in [25], [37]). Die unbekanntenen Koeffizienten β_j werden so gewählt, daß

$$e := \sum_{i=1}^k (Z_i - \lambda_i \hat{g})^2 \stackrel{!}{=} \min, \quad (4.1)$$

gilt, d.h. die Summe der Fehlerquadrate bei Verwendung des genannten Ansatzes minimal wird. Standardverfahren führen auf die sogenannten Normalgleichungen. Dabei sei $p < k$ und $l \in \{0, \dots, p\}$ vorausgesetzt:

$$\begin{aligned} \frac{\partial e}{\partial \beta_l} &= 0 \\ \Leftrightarrow -2 \sum_{i=1}^k (Z_i - \lambda_i \hat{g}) \frac{\partial}{\partial \beta_l} \sum_{j=0}^p \beta_j (\lambda_i f_j) &= 0 \\ \Leftrightarrow \sum_{i=1}^k \sum_{j=0}^p \beta_j (\lambda_i f_j) (\lambda_i f_l) &= \sum_{i=1}^k (\lambda_i f_l) Z_i \\ \Leftrightarrow \mathbf{F}^T \mathbf{F} \boldsymbol{\beta} &= \mathbf{F}^T \mathbf{Z} \end{aligned}$$

mit

$$\begin{aligned} \mathbf{F} &= (\lambda_i f_j)_{i=1, \dots, k; j=0, \dots, p} \in \mathbb{R}^{k \times (p+1)} \\ \boldsymbol{\beta}^T &= (\beta_0, \dots, \beta_p). \end{aligned} \quad (4.2)$$

Unter der Voraussetzung, daß die Matrix $\mathbf{F}^T \mathbf{F}$ invertierbar ist, definiere den Vektor $\mathbf{L}^T(t_0) = \mathbf{f}^T(t_0)(\mathbf{F}^T \mathbf{F})^{-1} \mathbf{F}^T$ mit $\mathbf{f}^T(t_0) = (f_0(t_0), \dots, f_p(t_0))$. Damit erhält man die bekannte Lösung:

$$\hat{g}(t_0) = \mathbf{L}^T(t_0) \mathbf{Z}. \quad (4.3)$$

Die Varianz von $\hat{g}(t_0)$ ist demnach:

$$\text{Var } \hat{g}(t_0) = \mathbf{L}^T \text{Cov}(Z) \mathbf{L}. \quad (4.4)$$

4.2. Relativer Einfluß von Ableitungen auf die Varianz des Regressionsschätzers

Im folgenden soll der relative Einfluß einer Ableitungsmessung im Vergleich zu einer Funktionswertmessung auf die Varianz eines Regressionsschätzers für den Fall untersucht werden, daß bereits k Samplefunktionale zur Verfügung stehen. Insbesondere bei lokaler polynomialer Regression ist die Varianz des Schätzers im allgemeinen nicht nach oben beschränkt (siehe in [54]), weshalb eine sorgfältige Versuchsplanung zur Minimierung der Varianz wichtig ist. In der Literatur gibt es außerdem Ansätze, die Qualität eines Schätzers nur durch dessen Bias zu definieren (siehe in [37] und die Kritik dort zu diesem Vorgehen), bzw. den mittleren quadratischen Fehler zu betrachten. Diese Vorgehensweise wird hier nicht verfolgt. Die hier vorgestellten Ergebnisse sollen vielmehr als ein Schritt bei der Wahl eines Versuchsplans für eine Regressionsschätzung verstanden werden.

Sei $\mathbf{t}_0 \in D$, so betrachte zwei verschiedene Regressionsschätzer für die unbekannte Funktion g auf D :

$$\hat{g}_{(l)}(\mathbf{t}_0) = L_{(l)}^T(\mathbf{t}_0)Z_{(l)} \quad , l = 1, 2. \quad (4.5)$$

Für $l = 1$ bzw. 2 seien die Vektoren $L_{(l)}^T(\mathbf{t}_0) = f^T(\mathbf{t}_0)(\mathbf{F}_{(l)}^T \mathbf{F}_{(l)})^{-1} \mathbf{F}_{(l)}^T$ definiert. Die Basisfunktionen seien außerdem differenzierbar auf ihrem Definitionsgebiet. Die ersten k Zeilen der Matrizen $\mathbf{F}_{(1)}$ und $\mathbf{F}_{(2)}$, sowie die ersten k Elemente von Z seien dabei identisch. Nenne die Teilmatrix, die aus diesen ersten k Zeilen besteht, $\tilde{\mathbf{F}}$. Das zu $\mathbf{F}_{(1)}$ gehörige Funktional der $(k+1)$ -ten Zeile sei ein Deltafunktional $\delta_{\mathbf{t}_{k+1}}$ mit $\mathbf{t}_{k+1} \in D$, und das analoge zu $\mathbf{F}_{(2)}$ gehörige sei eine Ableitungsmessung $D_{s_{k+1}}$, ebenfalls mit $s_{k+1} \in D$. Die $(k+1)$ -ten Elemente der Vektoren $Z_{(1)}$ bzw. $Z_{(2)}$ enthalten die entsprechenden Samplefunktionale.

Die Matrizen $(\mathbf{F}_{(l)}^T \mathbf{F}_{(l)})$ lassen sich wie folgt zerlegen:

$$(\mathbf{F}_{(l)}^T \mathbf{F}_{(l)}) = (\tilde{\mathbf{F}}^T \tilde{\mathbf{F}} + \mathbf{a}_{(l)} \mathbf{a}_{(l)}^T). \quad (4.6)$$

Dabei seien die Vektoren $\mathbf{a}_{(1)}^T = (f_1(\mathbf{t}_{k+1}), \dots, f_p(\mathbf{t}_{k+1}))$ und $\mathbf{a}_{(2)}^T = (D_{s_{k+1}} f_1, \dots, D_{s_{k+1}} f_p)$ definiert. Sei nun o.B.d.A. $\text{Cov } Z_{(l)} = \sigma^2 \mathbf{I}_{k+1}$, angenommen, die Varianzen der Meßfehler seien also alle identisch σ^2 . Damit erhalten die Varianzen der oben definierten Schätzer die einfache, bekannte Form $\text{Var } \hat{g}_{(l)}(\mathbf{t}_0) = \sigma^2 f^T(\mathbf{t}_0)(\mathbf{F}_{(l)}^T \mathbf{F}_{(l)})^{-1} f(\mathbf{t}_0)$. Mit der Zerlegung aus (4.6) und einer bekannten Beziehung für inverse Matrizen aus [42] erhält man:

$$\text{Var } \hat{g}_{(l)}(\mathbf{t}_0) = \sigma^2 f^T(\mathbf{t}_0) \left((\tilde{\mathbf{F}}^T \tilde{\mathbf{F}})^{-1} - \frac{(\tilde{\mathbf{F}}^T \tilde{\mathbf{F}})^{-1} \mathbf{a}_{(l)} \mathbf{a}_{(l)}^T (\tilde{\mathbf{F}}^T \tilde{\mathbf{F}})^{-1}}{1 + \mathbf{a}_{(l)}^T (\tilde{\mathbf{F}}^T \tilde{\mathbf{F}})^{-1} \mathbf{a}_{(l)}} \right) f(\mathbf{t}_0). \quad (4.7)$$

Die Varianz der beiden unterschiedlichen Schätzer läßt sich also additiv in zwei Teile zerlegen. Dabei ist der erste Summand die Varianz eines Regressionsschätzers unter Verwendung der ersten k Samples aus den Vektoren $Z_{(l)}$. Diese stimmen bei beiden Schätzern überein. Der zweite Summand dagegen ist der Teil der Varianz, um welchen sich die soeben genannte ändert, wenn eine weitere Funktionswertmessung bzw. eine weitere Ableitungswertmessung zur Schätzung hinzugezogen wird. Mittels dieser Darstellung ist es nun also möglich, den

unterschiedlichen Einfluß von Ableitungs- bzw. Funktionswertmessung auf die Varianz eines Regressionsschätzers zu quantifizieren. Dies soll im folgenden exemplarisch anhand einer linearen Regression für den eindimensionalen Fall gezeigt werden.

Beispiel 4.1. Für den linearen Regressionsansatz für den eindimensionalen Fall werden die beiden Basisfunktionen $f_0(t) \equiv 1$ und $f_1(t) = t$ verwendet. Desweiteren soll der Fall $k = 2$ betrachtet werden, wobei die verwendeten Samplefunktionale Funktionswertmessungen an den Stellen t_1 bzw. t_2 sind. Schließlich sei $\sigma^2 = 1$ gewählt.

Berechne die Matrix $(\tilde{\mathbf{F}}^T \tilde{\mathbf{F}})^{-1}$ aus (4.7):

$$\mathbf{S} := (\tilde{\mathbf{F}}^T \tilde{\mathbf{F}})^{-1} = \frac{1}{(t_1 - t_2)^2} \begin{pmatrix} t_1^2 + t_2^2 & -t_1 - t_2 \\ -t_1 - t_2 & 2 \end{pmatrix}.$$

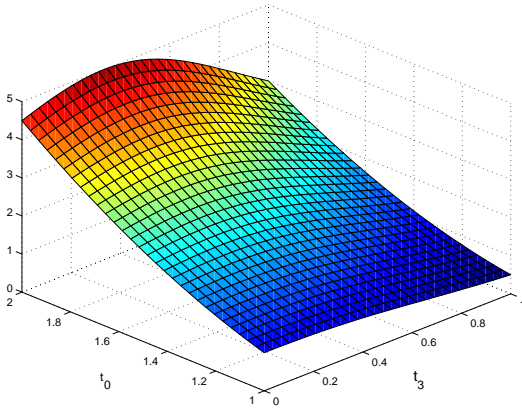
Mit der Beziehung (4.7) lassen sich die Varianzen der beiden Schätzer $\hat{g}_{(1)}$ und $\hat{g}_{(2)}$, jeweils additiv in zwei Teile zerlegt, darstellen. Der jeweils erste Summand entspricht dabei wiederum der Varianz des Regressionsschätzers, welcher nur auf den beiden Funktionsmessungen an den Stellen t_1 und t_2 beruht. Die zweiten Summanden spiegeln dagegen den unterschiedlichen Beitrag zur Veränderung der Varianz der Regressionsschätzer bei Hinzunahme einer weiteren Funktionswertmessung an einer Stelle t_3 bzw. einer Ableitungsmessung an einer beliebigen Stelle wider. Dabei hat der Ort der Ableitungsmessung bei einem linearen Ansatz keinen Einfluß auf die Varianz, da die Steigung des verwendeten Schätzeransatzes an jedem Ort gleich ist. Für die Differenz der beiden Schätzvarianzen erhält man mit der obigen Bezeichnungsweise:

$$\begin{aligned} & \text{Var } \hat{g}_{(1)} - \text{Var } \hat{g}_{(2)} & (4.8) \\ &= \frac{1}{c_{(2)}} \mathbf{S}_{22}^2 \left(t_0 + \frac{\mathbf{S}_{12}}{\mathbf{S}_{22}} \right)^2 - \frac{1}{c_{(1)}} \left(\mathbf{S}_{12}^2 \left(t_0 + \frac{\mathbf{S}_{11}}{\mathbf{S}_{12}} \right)^2 \right. \\ & \quad \left. + 2t_3(\mathbf{S}_{11}\mathbf{S}_{12} - t_0(\mathbf{S}_{12}^2 + \mathbf{S}_{11}\mathbf{S}_{22}) - t_0^2\mathbf{S}_{12}\mathbf{S}_{22}) + t_3^2 \left(t_0 + \frac{\mathbf{S}_{12}}{\mathbf{S}_{22}} \right)^2 \right), \end{aligned}$$

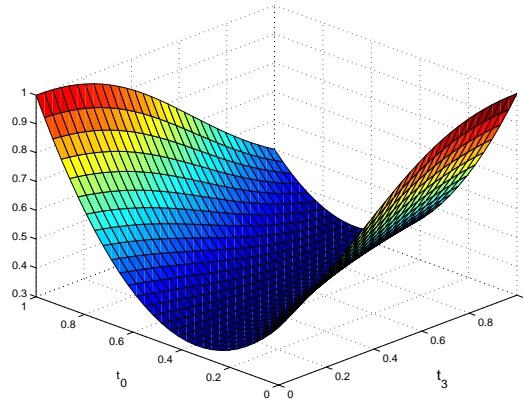
mit

$$\begin{aligned} c_{(1)} &= 1 + \mathbf{S}_{11} + 2t_3\mathbf{S}_{12} + t_3^2\mathbf{S}_{22} \\ c_{(2)} &= 1 + \mathbf{S}_{22}. \end{aligned}$$

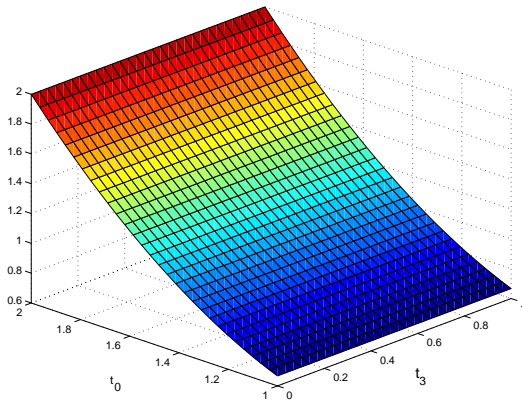
In Abbildung 4.1 sind für zwei verschiedene Konstellationen für die Wahl des Intervalls für den Beobachtungsort t_0 jeweils die Varianzen der Schätzer $\hat{g}_{(1)}$ und $\hat{g}_{(2)}$ sowie die soeben berechnete Differenz aus beiden. Für diesen einfachen Fall, daß nur zwei Funktionsmessungen t_1 bei 0 und t_2 bei 1 verwendet werden und dieses Design nun um eine weitere Funktions- bzw. Ableitungswertmessung vergrößert werden soll, erscheint der Aufwand, die Beziehung (4.8) anzugeben, übertrieben. Doch bleibt diese Beziehung auch für große Designs gültig und liefert so eine schnell zu realisierende Bewertungsmöglichkeit. In den Teilbildern (a),(c) und (e) stammt der Beobachtungsort t_0 aus dem Intervall $[1, 2]$, in den übrigen Teilbildern dagegen aus dem Intervall $[0, 1]$. In allen Fällen ist jeweils der Ort der neuwählenden Funktionsmessung, t_3 aus dem Intervall $[0, 1]$. Aus der Beziehung (4.8) wird deutlich, daß, falls t_3 so gewählt wird, daß die Varianzendifferenz positiv ist, dann eine Ableitungsmessung besser hinsichtlich der Schätzvarianz ist.



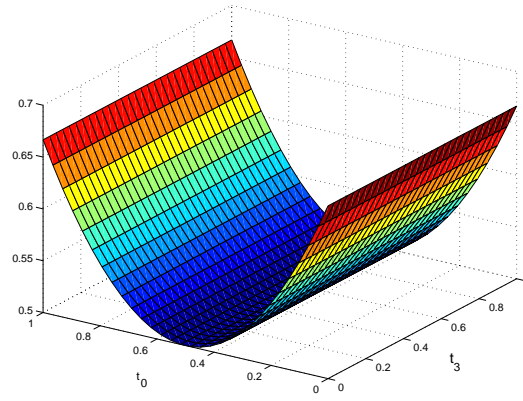
(a) $\text{Var } \hat{g}_{(1)}$ mit $t_1 = 0$ und $t_2 = 1$



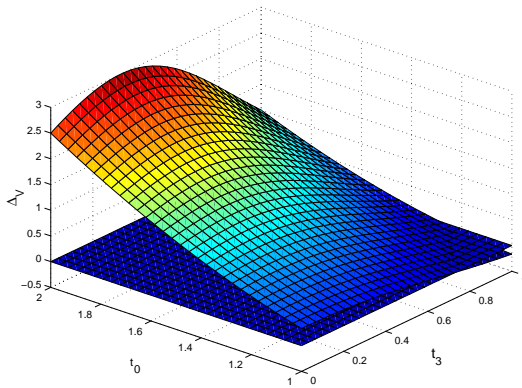
(b) $\text{Var } \hat{g}_{(1)}$ mit $t_1 = 0$ und $t_2 = 1$



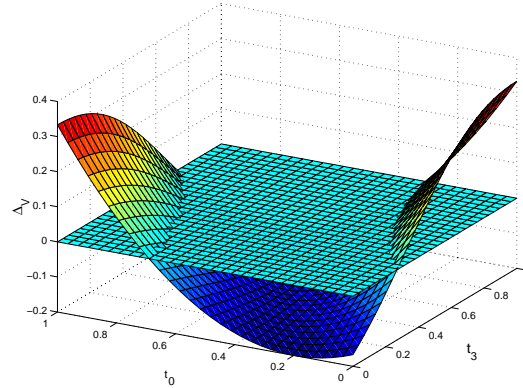
(c) $\text{Var } \hat{g}_{(2)}$ mit $t_1 = 0$ und $t_2 = 1$



(d) $\text{Var } \hat{g}_{(2)}$ mit $t_1 = 0$ und $t_2 = 1$



(e) $\text{Var } \hat{g}_{(1)} - \text{Var } \hat{g}_{(2)}$ mit $t_1 = 0$ und $t_2 = 1$



(f) $\text{Var } \hat{g}_{(1)} - \text{Var } \hat{g}_{(2)}$ mit $t_1 = 0$ und $t_2 = 1$

Abbildung 4.1.: Die Varianzen der linearen Regressionsschätzer $\hat{g}_{(1)}$ und $\hat{g}_{(2)}$ sowie deren Differenz. t_3 ist der Ort der zusätzlichen Funktionswertmessung.

In Teilbild (e) wird dies besonders deutlich, wenn der Beobachtungsort $t_0 = 2$ gewählt wird. Dann nämlich gibt es in dem Intervall $[0, 1]$ keinen Punkt, an dem eine Funktionswert- einer Ableitungswertmessung überlegen wäre. Erst mit zunehmender Annäherung des Vorhersageortes an den Bereich, der bereits mit Funktionsmessungen erkundet wurde (t_1 und 2), gibt es stets auch Möglichkeiten, daß eine Funktionswertmessung besser ist als eine Ableitungswertmessung. Festzuhalten bleibt jedoch, daß für den Fall, daß eine Schätzung außerhalb eines erkundbaren Bereiches gemacht werden soll, dann scheinbar Ableitungsmessungen den Funktionswertmessungen überlegen zu sein scheinen. Dies ändert sich für den Fall, daß die Schätzung in dem Bereich vorgenommen wird, wo auch Erkundungen möglich sind, hier also im Intervall $[0, 1]$. Doch auch dann kann eine unglückliche Wahl des Ortes einer weiteren Funktionswertmessung zu einem schlechteren Ergebnis führen, als bei Verwendung einer Ableitungswertmessung.

Bemerkung 4.2. 1. Die Varianz σ^2 der zufälligen Meßfehler wird in der Regel nicht bekannt sein. Er ist daher zunächst aus den Daten und unter Beachtung des Schätzansatzes zu schätzen. In [37] wird

$$\hat{\sigma}^2 = \frac{\hat{\varepsilon}^T \hat{\varepsilon}}{\text{tr}(\mathbf{L} - \mathbf{I}_k)(\mathbf{L} - \mathbf{I}_k)^T}$$

vorgeschlagen, wobei $\hat{\varepsilon}^T = (\lambda_1 \hat{g} - Z_1, \dots, \lambda_k \hat{g} - Z_k)$ der Vektor der empirischen Meßfehler bei Anwendung des Schätzers \hat{g} ist. Außerdem sei $\mathbf{L}^T = (\lambda_1 L(\cdot), \dots, \lambda_k L(\cdot))$ definiert. Voraussetzung für diesen Schätzer ist allerdings, daß die empirischen Schätzer $\hat{\varepsilon}_j$ unabhängig identisch verteilt mit Erwartungswert 0 sind. Letzteres bedeutet wiederum, daß der Regressionsschätzer im Mittel meßwerttreu sein muß.

2. Die Verwendung der Summe der Fehlerquadrate setzt voraus, daß mit normalverteilten Meßfehlern gerechnet wird. In diesem Fall nämlich entspricht der so berechnete Regressionsschätzer dem Maximum-Likelihood-Schätzer. In [11] wird auf diesen Umstand hingewiesen und mögliche andere Anpassungskriterien für den Fall, daß diese Voraussetzung nicht erfüllt ist, genannt. Insbesondere falls die Meßfehler gemäß einer heavy-tail Verteilung verteilt sind, dürfte die Berechnung der Parameter auf der Basis von Fehlerquadraten problematisch sein.

Prinzipiell ist jedoch die Möglichkeit, auf verschiedene Verteilungsvoraussetzungen der Meßfehler reagieren zu können, als ein weiterer Vorteil für die Verwendung von Regressionsverfahren zu werten. Kriging, als Beispiel, stellt nur bei angenommenen gaußschen zufälligen Funktionen das beste Voraussageverfahren im Sinne der Methode der kleinsten Quadrate dar. Die gesamte Theorie beruht jedoch sehr stark auf der Verwendung dieses Fehlerbegriffs, so daß von dieser Seite nicht auf veränderte Verteilungen der zufälligen Funktionen reagiert werden kann.

4.3. Versuchsplanung bei lokaler Regression

Im Unterschied zum Ansatz der klassischen Regression werden bei der lokalen Regression die Samples lokal gewichtet. Die zugrundeliegende Idee ist, daß Samples, die sich in der Nähe des Ortes t_0 befinden, für den die unbekannte Funktion g geschätzt werden soll, stärker gewertet werden, als weiter entfernt liegende. In [11] und [29] skizzieren CLEVELAND/LOADER

bzw. MARRON die historische Entwicklung dieser Verfahren, wobei sie besonders auf die vielfältigen Formen und Anwendungen dieser Verfahren verweisen. Insbesondere werden die leichte Interpretierbarkeit und die einfache Berechenbarkeit der lokalen Regressionschätzer hervorgehoben.

Sei fortan vorausgesetzt, daß der Samplevektor Z nur Funktions- und Ableitungsmessungen enthält, die mit Meßfehlern verfälscht vorliegen. Der wesentliche Unterschied zum klassischen Regressionsansatz, wie er oben dargestellt wurde, ist, daß nun das Fehlermaß

$$e(\mathbf{t}_0) = \sum_{i=1}^k (Z_i - \lambda_i \hat{g})^2 w_i(\mathbf{t}_0) \quad (4.9)$$

zur Bestimmung des Schätzers verwendet wird. Dabei gelte für die neu eingeführten Koeffizienten $w_i(\mathbf{t}_0) = w(\|\mathbf{t}_i - \mathbf{t}_0\|/h)$, wobei w eine sogenannte Gewichtsfunktion mit der Bandweite h ist. Bezüglich der Wahl der Gewichtsfunktion werden in [11] und [37] einige Vorschläge gemacht. Allen gemein ist, daß sie die Regularität des zu modellierenden Merkmals widerspiegeln sollen, d.h. die Stetigkeit und Differenzierbarkeit. Ferner wird gefordert, daß sie positiv sind und entweder einen beschränkten Träger haben oder wenigstens $w(\mathbf{u}) \rightarrow 0$ für $\|\mathbf{u}\| \rightarrow \infty$ gelten soll. Beispiele sind:

$$w(\mathbf{u}) = \begin{cases} (1 - \|\mathbf{u}\|^3)^3 & \|\mathbf{u}\| < 1 \\ 0 & \|\mathbf{u}\| > 1 \end{cases},$$

$$w(\mathbf{u}) = \frac{3}{4}(1 - u^2)_+,$$

dabei ist der letztere der sogenannte Epanechnikov-Kern (siehe in [54]). In [25] wird ein lokales Regressionsverfahren vorgestellt, welches interpolierend ist. Dies wird dadurch erreicht, daß die Gewichtsfunktion bei 0 einen Pol mindestens der Ordnung 2 erhält und ansonsten das soeben geforderte Verhalten hat. In der zitierten Arbeit wird eine Gewichtung $w(\mathbf{u}) = e^{-u^2}/u^2$ vorgeschlagen.

Die Bandweite hat ebenfalls einen starken Einfluß auf das Verhalten des Regressionschätzers. Allgemein gilt, daß eine Verkleinerung der Bandweite zu einem Schätzer mit geringerem Bias aber mit höherer Varianz führt. Des weiteren hat die Wahl einen Einfluß auf die Lösbarkeit des Regressionsproblems. In [54] werden Möglichkeiten der Bandweitenwahl diskutiert, der Schwerpunkt liegt dabei auf einer adaptiven Wahl.

Wird die Summe der Fehlerquadrate als Anpassungskriterium verwendet, so geht die Herleitung der Regressionslösung analog zum obigen Fall. Bei einem zum klassischen Fall analogen Ansatz $\hat{g}(\mathbf{t}_0) = \sum_{j=0}^p \beta_{0j} f_j(\mathbf{t}_0)$ für die Schätzfunktion, wobei der Koeffizientenvektor $\beta_0^T = (\beta_{00}, \dots, \beta_{0p})$ nun von \mathbf{t}_0 abhängt, erhält man die Lösung:

$$\hat{g}(\mathbf{t}_0) = L_{\text{lok}}^T(\mathbf{t}_0)Z$$

mit

$$L_{\text{lok}}^T(\mathbf{t}_0) = \mathbf{f}^T(\mathbf{t}_0) \left(\mathbf{F}^T \mathbf{W}(\mathbf{t}_0) \mathbf{F} \right)^{-1} \mathbf{F}^T \mathbf{W}(\mathbf{t}_0).$$

Dabei ist $\mathbf{W}(\mathbf{t}_0) = \text{diag}(w_1(\mathbf{t}_0), \dots, w_p(\mathbf{t}_0))$ die Matrix der Gewichte.

Die Schätzvarianz stellt sich wie folgt dar:

$$\text{Var } \hat{g}(t_0) = f^T(t_0) \left(\mathbf{F}^T \mathbf{W}(t_0) \mathbf{F} \right)^{-1} \mathbf{F}^T \mathbf{W}(t_0) \text{Cov}(Z) \mathbf{W}(t_0) \mathbf{F} \left(\mathbf{F}^T \mathbf{W}(t_0) \mathbf{F} \right)^{-1} f(t_0).$$

Auch hier wird in der Regel die Kovarianzfunktion der Samples nicht bekannt sein. Eine Möglichkeit deren Schätzung ist $\hat{\sigma}^2 \mathbf{I}_k$ unter Beachtung von Bemerkung 4.2(1). In [37] wird dagegen $\widehat{\text{Cov}}(Z) = \hat{\sigma}^2 \mathbf{W}^{-1}(t_0)$ vorgeschlagen, was zu einem stark vereinfachten Ausdruck für die obige Varianz führt.

Da die Matrix \mathbf{W} diagonal ist, kann die Schätzvarianz in zum klassischen Fall analoger Weise additiv zerlegt werden. Legt man den Fall zugrunde, daß die Varianz des Meßfehlers σ^2 bekannt ist, so ist jedoch der Summand, welcher durch das Hinzufügen eines neuen Samples zur Berechnung des Schätzers entsteht, weitaus komplexer als dies bei der klassischen Regression der Fall war (siehe Anhang C). Dennoch ist auch in diesem Fall eine differenzierte Bewertung von Funktions- im Vergleich zu Ableitungswertmessungen im Hinblick auf deren Einfluß auf die Schätzvarianz möglich.

5. Versuchsplanung mit endlichen Designs

5.1. Versuchsplanung für den BLP mittels gestörter Kovarianzmatrizen

Sei Q eine zufällige Funktion auf einer geeigneten Menge $F \subset C^r(D)$ mit Kovarianzfunktion K . Betrachte in diesem Kontext das Approximationsproblem $\text{Pre}(t_0)$ mit einem $t_0 \in \mathbb{R}^n$, wobei die Vorhersage mittels des BLP auf der Basis einer Menge $T = \{\lambda_1, \dots, \lambda_k\}$ von linearen Samplefunktionalen geschieht. T sei die maximale mögliche Menge von Samplefunktionalen in einer gegebenen Situation. Eine Möglichkeit ist z.B., daß nur Deltafunktionalen mit Beobachtungsorten innerhalb einer gewissen endlichen Menge $D^* \subset D$ für die Vorhersage verwendet werden können. Es liegt also eine analoge Situation wie in Abschnitt 3.2.1 vor, und der BLP sei wie dort mit $S_k^{t_0}$ bezeichnet.

Das Hauptziel der Betrachtung wird es sein, ein Designkriterium zu definieren, das differenzierbar nach einem sogenannten Designmaß, also einer Gewichtungsfunktion für die beteiligten Samplefunktionalen, ist und einen unmittelbaren Bezug zum mittleren quadratischen Vorhersagefehler hat. Zu diesem Zweck sollen die Beobachtungen mit Meßfehler künstlich gestört werden. Diese Meßfehler werden kontrolliert hinzugefügt. D.h. zum Zwecke dieses Verfahrens werden die oben genannten Samplefunktionalen modifiziert, und man verwendet nun eine Menge T^{err} von Samplefunktionalen auf $F \times \mathbb{R}^k$ von der Form

$$\lambda_j^{\text{err}} := \lambda_j + \pi_j(\gamma, \xi), \quad j = 1, \dots, k. \quad (5.1)$$

Die Funktionale π_j sind wie im Abschnitt 3.4.1 die Projektionen von Vektoren $\varepsilon \in \mathbb{R}^k$ auf die j -te Komponente. Die Vektoren ε sind dabei wiederum mit einer Verteilung P ausgestattet, die stochastisch unabhängig von Q ist, verschwindenden Erwartungswert hat und deren Kovarianzfunktion gerade $R_{ij} = \delta_{ij} \sigma^2(i, \gamma, \xi)$ ist. Dabei sei $\gamma \in \mathbb{R}_+$ der Parameter, der die Stärke der Störung durch einen Meßfehler steuert. Der Parameter ξ ist ein Designmaß auf T , d.h. es gilt:

1. $\xi(\lambda_j) \geq 0$, für alle $j=1, \dots, k$
2. $\sum_{j=1}^k \xi(\lambda_j) = 1$.

Man spricht von einem uniformen Designmaß, falls für alle $\lambda \in T$ gerade $\xi(\lambda) = \xi_{\max}$ gilt, wobei ξ_{\max} der Maximalwert ist, den ξ über T annimmt. In diesem Fall gibt das Designmaß nur die Zugehörigkeit zur Menge der für eine Vorhersage verwendeten Samplefunktionalen

wieder. In Übereinstimmung mit der Schreibweise in [37] sollen derartige Designmaße mit ξ_T bezeichnet werden. Die sogenannten atomaren Designmaße schließlich sind uniforme Designmaße, deren Support einelementig ist. Der Support ist gerade die Menge aller Funktionale aus der Menge T , die ein positives Gewicht bekommen.

In der klassischen Versuchsplanung mit unkorrelierten Beobachtungen gibt ein Designmaß die relative Häufigkeit an, mit der an einer bestimmten Stelle Beobachtungen gemacht werden sollen. Im Falle der Vorhersage zufälliger Funktionen mittels Beobachtungen ohne Meßfehler ist dies keine mögliche Interpretation mehr. Einzig unitäre Designmaße haben eine unmittelbare Interpretation als Indikatorfunktion der verwendeten Samplefunktionalen. Im allgemeinen Fall allerdings soll ein Designmaß in diesem Fall angeben, wie wichtig ein bestimmtes Samplefunktional für eine genaue Vorhersage bezüglich eines bestimmten Fehlerkriteriums ist (siehe auch in [40]). Am Ende des Abschnitts wird ein Algorithmus vorgestellt, der iterativ ein solches Designmaß bestimmt, wobei die einzelnen Samplefunktionalen gewichtet werden für den Fall, daß ein Meßfehler eintritt. Vom Standpunkt einer praktischen Anwendung ist dies ein wichtiger Aspekt, da es in der Regel kein ideales, d.h. fehlerfreies Meßgerät gibt. Versuchsplanung wird jedoch in aller Regel für die Verwendung fehlerfreier Messungen gemacht. Das nun vorgestellte Verfahren untersucht genau den Fall, daß vom fehlerfreien in den fehlerbehafteten Meßfall übergegangen wird.

Der Ausgangspunkt der weiteren Betrachtung ist der mittlere quadratische Fehler, der bei der Verwendung der Methode $S_k^{t_0}$ entsteht, wobei die Funktionale λ_j mit $j = 1, \dots, k$ verwendet werden. Aus 3.12 erhält man

$$\begin{aligned} e_2(S_k^{t_0}, \text{Pre}(t_0), Q) &= K(t_0, t_0) - \Sigma_{t_0}^T \Sigma_\lambda^{-1} \Sigma_{t_0} \\ &=: K(t_0, t_0) - \square(t_0, T). \end{aligned} \tag{5.2}$$

Bemerkung 5.1. 1. Entspricht die Menge $\{\lambda_1, \dots, \lambda_k\}$ der potentiell größtmöglichen Menge an Samplefunktionalen im Beobachtungsraum, so ist der mittlere quadratische Fehler in (5.2) nicht weiter zu minimieren. Allerdings werden in einem praktischen Umfeld oftmals nicht an allen möglichen Orten Beobachtungen durchführbar sein, da dies z.B. zu hohen Erkundungskosten führt. Damit ist also abzuwägen, welche Samplefunktionalen aus dieser Maximalmenge auszuwählen sind, um im Rahmen der Möglichkeiten die beste Vorhersage zu bekommen.

2. Die Schreibweise in (5.2) deutet an, daß das \square -Funktional insbesondere nur davon abhängt, ob ein bestimmtes Samplefunktional in der Menge der für die Vorhersage verwendeten Funktionale enthalten ist oder nicht.
3. Die verwendete Methode, Matrizen in einer, wie unten gezeigt wird, differenzierbaren Weise zu stören, wird erstmals im Zusammenhang sogenannter approximierter Informationsmatrizen bei der Schätzung von Trendparametern zufälliger räumlicher Funktionen von MÜLLER in [37] angewendet (siehe auch in [41] für einen verfeinerten Algorithmus für die iterative Suche nach einem optimalen Design mittels gestörter Informationsmatrizen).

5.1.1. Verallgemeinerung und Störung des \square -Funktionals

Sei ξ ein beliebiges Designmaß auf \mathbb{T} , so definiere:

$$\square(\mathbf{t}_0, \xi) := \square(\mathbf{t}_0, S_\xi). \quad (5.3)$$

Dabei sei mit S_ξ der Support des Maßes ξ bezeichnet.

Diese Verallgemeinerung ist zunächst nur eine andere Schreibweise des in (5.2) definierten Funktionals, da das Designmaß in das \square -Funktional nur durch dessen Support, nicht jedoch durch die mit diesem Maß erzielte Gewichtung von Elementen aus \mathbb{T} eingeht. Diese Eigenschaft wird durch die folgende Definition verwirklicht:

$$\square^\gamma(\mathbf{t}_0, \xi) := \boldsymbol{\Sigma}_{\mathbf{t}_0}(S_\xi)^\top (\boldsymbol{\Sigma}_\lambda(S_\xi) + \mathbf{W}^\gamma(\xi))^{-1} \boldsymbol{\Sigma}_{\mathbf{t}_0}(S_\xi) \quad (5.4)$$

mit

$$\mathbf{W}^\gamma(\xi) = \text{diag}(\sigma^2(j, \gamma, \xi) : \lambda_j \in S_\xi). \quad (5.5)$$

Dabei ist $\mathbf{W}^\gamma(\xi)$ gerade die Kovarianzmatrix der künstlich hinzugefügten Meßfehler (siehe 5.1)), wobei allerdings nur diejenigen Messungen berücksichtigt werden, deren zugehörige Samplesfunktionale im Support des Designmaßes ξ enthalten sind.

In [37] werden die folgenden Forderungen an die Terme $\sigma^2(j, \gamma, \xi)$ gestellt:

$$\sigma^2(j, \gamma, \xi) = \begin{cases} 0 & \xi(\lambda_j) = \xi_{\max} \\ \infty & \xi(\lambda_j) = 0 \\ \rightarrow 0 & \gamma \rightarrow 0 \text{ and } \xi(\lambda_j) > 0. \end{cases} \quad (5.6)$$

Dabei sei eine Menge $\mathbb{T} = \{\lambda_1, \dots, \lambda_k\}$ von Samplefunktionalen zugrundegelegt. Diese Eigenschaften lassen sich recht anschaulich interpretieren. Wird ein Samplefunktional $\lambda_j \in \mathbb{T}$ mit der größtmöglichen Gewichtung, ξ_{\max} versehen, so wird kein künstlicher Meßfehler hinzuaddiert. Genau umgekehrt wird im entgegengesetzten Fall verfahren, falls also $\lambda_j \in \mathbb{T}$ nicht im Support von ξ ist. Dann wird ein künstlicher Meßfehler mit unendlicher Varianz hinzuaddiert, womit diese Messung als unbrauchbar deklariert wird. Die letzte Zeile sichert schließlich, daß für einen gegen Null verschwindenden Steuerparameter γ die Wirkung der Störung der Kovarianzmatrix $\boldsymbol{\Sigma}_\lambda$ linear verschwindet.

Insbesondere ist aufgrund dieser Eigenschaften sichergestellt, daß bei Verwendung unitärer Designmaße $\xi_{\mathbb{T}}$ die Matrix $\mathbf{W}^\gamma(\xi_{\mathbb{T}})$ identisch Null ist, es gilt dann also $\square^\gamma(\mathbf{t}_0, \xi_{\mathbb{T}}) \equiv \square(\mathbf{t}_0, \mathbb{T})$. Mögliche Varianten für die Wahl der Varianzen $\sigma^2(j, \gamma, \xi)$ sind:

1. $\sigma^2(j, \gamma, \xi) = \gamma \ln\left(\frac{\xi^{(\gamma)}}{\xi(j)}\right)$. Dabei ist $\xi^{(\gamma)} = \left(\sum_{j: \lambda_j \in S_\xi} \xi(j)^{\frac{1}{\gamma}}\right)^\gamma$ eine differenzierbare $L_{1/\gamma}$ -Näherung von ξ_{\max} , d.h. es gilt $\xi^{(\gamma)} \rightarrow \xi_{\max}$, falls γ gegen Null strebt. In [37] und [41] wird diese Variante verwendet.
2. $\sigma^2(j, \gamma, \xi) = \left(\left(\frac{1}{\xi(j)}\right)^\gamma - \left(\frac{1}{\xi^{(\gamma)}}\right)^\gamma\right)$. Bei gleichem Wert für γ werden bei dieser Variante „schlechte“ Designpunkte λ_j , also solche, bei denen das Gewicht $\xi(j)$ relativ klein ist, mit einem Meßfehler relativ höherer Varianz als bei der ersten Variante versehen (siehe Abbildung 5.1).

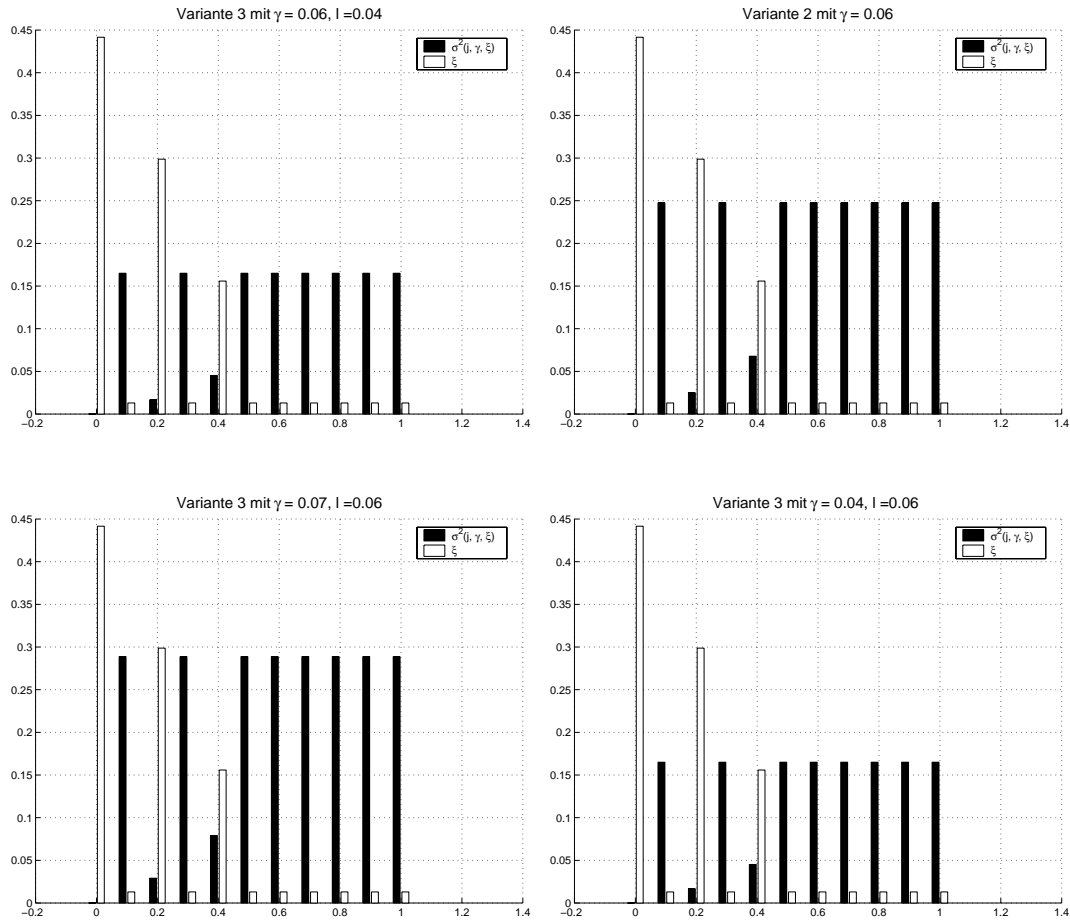


Abbildung 5.1.: 3 mögliche Varianten für die Wahl des künstlich zugefügten Meßfehlers. ξ ist ein Designmaß mit 11 äquidistanten Punkten im Intervall $[0, 1]$.

3. $\sigma^2(j, \gamma, \xi) = \frac{\gamma}{l} \left(\left(\frac{1}{\xi(j)} \right)^l - \left(\frac{1}{\xi(j\gamma)} \right)^l \right)$, mit einem $l \in \mathbb{R}_+ \setminus \{0\}$. Der Parameter l übernimmt teilweise die Funktion, die γ in den anderen Varianten hatte. Verändert man den Parameter γ bei dieser Variante, so kann erreicht werden, daß „schlechte“ Designpunkte entweder stärker oder schwächer gewichtet werden als dies bei der Variante 1 der Fall ist (siehe ebenfalls Abbildung 5.1, untere Zeile).

5.1.2. Berechnung der Richtungsableitung von \square^γ

Das \square -Funktional in der ursprünglichen Definition in (5.2) ist bereits maximal, da für die Konstruktion der Vorhersage sämtliche potentiell zur Verfügung stehenden Informationen in Form von Samplefunktionalen verwendet werden. Andererseits ist das unmodifizierte sowie das modifizierte (siehe 5.3) \square -Funktional nicht differenzierbar nach dem Designmaß ξ . Dies ist bedingt dadurch, daß die beiden Funktionale von einem Designmaß ξ nur durch dessen Support abhängen und somit keine sinnvolle Definition einer Richtungs-

ableitung möglich ist. Es ist mit diesen Ansätzen also nicht möglich, Richtungsableitungen in Abhängigkeit des Designmaßes ξ zu bestimmen. Somit kann kein Algorithmus konstruiert werden, der mittels Richtungsableitungen des \square -Funktionales analytisch ein Designmaß ermittelt oder annähert, dessen Support z.B. kleiner ist als die potentiell mögliche Maximalgröße.

Das Funktional \square^γ dagegen ist differenzierbar, der Beweis aus [37] zur Differenzierbarkeit der dort eingeführten approximierten Informationsmatrizen kann hierfür verwendet werden. Seien also ξ und ξ' zwei Designmaße, so bestimme die Richtungsableitung von $\square^\gamma(t_0, \xi)$ in Richtung ξ' :

$$\begin{aligned} \phi_{t_0}^\gamma(\xi, \xi') &= \lim_{\alpha \rightarrow 0} \frac{\square^\gamma(t_0, \xi_\alpha) - \square^\gamma(t_0, \xi)}{\alpha} \\ &\stackrel{\text{l'Hospital}}{=} \lim_{\alpha \rightarrow 0} \frac{\partial}{\partial \alpha} \square^\gamma(t_0, \xi_\alpha). \end{aligned} \quad (5.7)$$

Dabei sei $\xi_\alpha = (1 - \alpha)\xi + \alpha\xi'$ definiert. Berechne weiter die partielle Ableitung aus dem obigen Ausdruck:

$$\begin{aligned} \frac{\partial}{\partial \alpha} \square^\gamma(t_0, \xi_\alpha) &= \frac{\partial}{\partial \alpha} \sum_{i,j:\lambda_i,\lambda_j \in S_{\xi_\alpha}} \boldsymbol{\Sigma}_{t_0}(S_\xi)_i^\top (\boldsymbol{\Sigma}_\lambda(S_{\xi_\alpha}) + \mathbf{W}^\gamma(\xi_\alpha))_{ij}^{-1} \boldsymbol{\Sigma}_{t_0}(S_\xi)_j^\top \\ &\stackrel{(*)}{=} \sum_{i,j:\lambda_i,\lambda_j \in S_{\xi_\alpha}} \boldsymbol{\Sigma}_{t_0}(S_\xi)_i^\top \frac{\partial}{\partial \alpha} (\boldsymbol{\Sigma}_\lambda(S_\xi) + \mathbf{W}^\gamma(\xi_\alpha))_{ij}^{-1} \boldsymbol{\Sigma}_{t_0}(S_\xi)_j^\top \\ &\stackrel{(**)}{=} - \sum_{i,j:\lambda_i,\lambda_j \in S_{\xi_\alpha}} \boldsymbol{\Sigma}_{t_0}(S_\xi)_i^\top \left[(\boldsymbol{\Sigma}_\lambda(S_\xi) + \mathbf{W}^\gamma(\xi_\alpha))^{-1} \frac{\partial}{\partial \alpha} \mathbf{W}^\gamma(\xi_\alpha) (\boldsymbol{\Sigma}_\lambda(S_\xi) \right. \\ &\quad \left. + \mathbf{W}^\gamma(\xi_\alpha))^{-1} \right]_{ij} \boldsymbol{\Sigma}_{t_0}(S_\xi)_j^\top \\ &= - \sum_{i,j,l:\lambda_i,\lambda_j,\lambda_l \in S_{\xi_\alpha}} \boldsymbol{\Sigma}_{t_0}(S_\xi)_i^\top (\boldsymbol{\Sigma}_\lambda(S_\xi) + \mathbf{W}^\gamma(\xi_\alpha))_{il}^{-1} \left(\frac{\partial}{\partial \alpha} \mathbf{W}^\gamma(\xi_\alpha) \right)_{ll} \\ &\quad \times (\boldsymbol{\Sigma}_\lambda(S_\xi) + \mathbf{W}^\gamma(\xi_\alpha))_{ij}^{-1} \boldsymbol{\Sigma}_{t_0}(S_\xi)_j^\top. \end{aligned} \quad (5.8)$$

Dabei ist für die Umformung bei (*) zu beachten, daß $S_{\xi_\alpha} = S_\xi \cup S_{\xi'}$ für alle $\alpha \in (0, 1)$ gilt und somit gar nicht von α abhängt. Für $\alpha = 0$ gilt außerdem $S_{\xi_0} = S_\xi$. Bei (**) wurde eine bekannte Beziehung zur Differentiation parametrisierter, inverser Matrizen verwendet (siehe [53]).

Der verbleibende Differentiationsterm im obigen Ausdruck soll zunächst für den Fall, daß $\sigma^2(j, \gamma, \xi) = \gamma \ln\left(\frac{\xi(j)}{\xi(j)}\right)$ ist, berechnet werden. Sei also ein beliebiger Index l gewählt, so daß

$\lambda_l \in S_\xi$ ist. Dann gilt:

$$\begin{aligned}
\left(\frac{\partial}{\partial \alpha} \mathbf{W}^\gamma(\xi_\alpha) \right)_u &= \frac{\partial}{\partial \alpha} \gamma (\ln \xi_\alpha^{(\gamma)} - \ln \xi_\alpha(\lambda_l)) \\
&= \gamma \left(\frac{1}{\xi_\alpha^{(\gamma)}} \frac{\partial}{\partial \alpha} \xi_\alpha^{(\gamma)} - \frac{1}{\xi_\alpha(\lambda_l)} \frac{\partial}{\partial \alpha} \xi_\alpha(\lambda_l) \right) \\
&= \gamma \left(\frac{\gamma}{\xi_\alpha^{(\gamma)}} \left(\sum_{\lambda_j \in S_\xi} \xi_\alpha^{\frac{1}{\gamma}}(\lambda_j) \right)^{\gamma-1} \sum_{\lambda_j \in S_\xi} \frac{\xi_\alpha^{\frac{1}{\gamma}-1}(\lambda_j)}{\gamma} (\xi'(\lambda_j) - \xi(\lambda_j)) \right. \\
&\quad \left. - \frac{\xi'(\lambda_l) - \xi(\lambda_l)}{\xi_\alpha(\lambda_l)} \right) \\
&= \gamma \left(\frac{\sum_{\lambda_j \in S_\xi} \xi_\alpha^{1/\gamma}(\lambda_j) \frac{\xi'(\lambda_j) - \xi(\lambda_j)}{\xi_\alpha(\lambda_j)}}{\sum_{\lambda_j \in S_\xi} \xi_\alpha^{1/\gamma}(\lambda_j)} - \frac{\xi'(\lambda_l) - \xi(\lambda_l)}{\xi_\alpha(\lambda_l)} \right).
\end{aligned}$$

Da $\mathbf{W}^\gamma(\xi_\alpha)$ für $\alpha \rightarrow 0$ gegen $\mathbf{W}^\gamma(\xi)$ konvergiert, läßt sich die Richtungsableitung in (5.7) wie folgt darstellen:

$$\phi_{t_0}^\gamma(\xi, \xi') = -\gamma \boldsymbol{\Sigma}_{t_0}(S_\xi)^\top (\boldsymbol{\Sigma}_\lambda(S_\xi) + \mathbf{W}^\gamma(\xi))^{-1} \Xi^\gamma(\xi, \xi') (\boldsymbol{\Sigma}_\lambda(S_\xi) + \mathbf{W}^\gamma(\xi))^{-1} \boldsymbol{\Sigma}_{t_0}(S_\xi) \quad (5.9)$$

mit

$$\Xi^\gamma(\xi, \xi') = \text{diag} \left(\frac{\sum_{\lambda_j \in S_\xi} \xi_\alpha^{\frac{1}{\gamma}}(\lambda_j) \frac{\xi'(\lambda_j)}{\xi(\lambda_j)}}{\sum_{\lambda_j \in S_\xi} \xi_\alpha^{\frac{1}{\gamma}}(\lambda_j)} - \frac{\xi'(\lambda_i)}{\xi(\lambda_i)} : \lambda_i \in S_\xi \right). \quad (5.10)$$

Die Richtungsableitung in (5.9) soll, wie oben schon erwähnt wurde, für infinitesimal kleine Werte für γ betrachtet werden.

Satz 5.2. *Mit den obigen Voraussetzungen gilt $\phi_{t_0}^\gamma(\xi, \xi') = o(\gamma)$.*

Beweis. Gemäß Definition gilt $\mathbf{W}^\gamma(\xi) \rightarrow 0$ für $\gamma \rightarrow 0$. Damit genügt es, die Diagonalmatrix $\Xi^\gamma(\xi, \xi')$ im Grenzfall $\gamma \rightarrow 0$ zu betrachten.

Definiere die Menge $B_\xi = \{\lambda_j \in S_\xi : \xi(\lambda_j) = \xi_{\max}\}$, deren Kardinalität mit N_{B_ξ} bezeichnet

sei. Damit gilt:

$$\begin{aligned}
\frac{\sum_{\lambda_j \in S_\xi} \xi^{\frac{1}{\gamma}}(\lambda_j) \left(\frac{\xi'(\lambda_j) - \xi(\lambda_j)}{\xi(\lambda_j)} \right)}{\sum_{\lambda_j \in S_\xi} \xi^{\frac{1}{\gamma}}} &= \frac{\sum_{\lambda_j \in B_\xi} \xi_{\max}^{\frac{1}{\gamma}} \left(\frac{\xi'(\lambda_j) - \xi_{\max}}{\xi_{\max}} \right)}{\sum_{\lambda_j \in S_\xi} \xi^{\frac{1}{\gamma}}(\lambda_j)} + \frac{\sum_{\lambda_j \in S_\xi \setminus B_\xi} \xi_{\max}^{\frac{1}{\gamma}} \left(\frac{\xi'(\lambda_j) - \xi(\lambda_j)}{\xi(\lambda_j)} \right)}{\sum_{\lambda_j \in S_\xi} \xi^{\frac{1}{\gamma}}(\lambda_j)} \\
&= \frac{\xi_{\max}^{\frac{1}{\gamma}} \sum_{\lambda_j \in B_\xi} \left(\frac{\xi'(\lambda_j)}{\xi_{\max}} - 1 \right)}{\xi_{\max}^{\frac{1}{\gamma}} \left(N_{B_\xi} + \sum_{\lambda_j \in S_\xi \setminus B_\xi} \left(\frac{\xi(\lambda_j)}{\xi_{\max}} \right)^{\frac{1}{\gamma}} \right)} \\
&\quad + \frac{\xi_{\max}^{\frac{1}{\gamma}} \sum_{\lambda_j \in S_\xi \setminus B_\xi} \left(\frac{\xi(\lambda_j)}{\xi_{\max}} \right)^{\frac{1}{\gamma}} \left(\frac{\xi'(\lambda_j) - \xi(\lambda_j)}{\xi(\lambda_j)} \right)}{\xi_{\max}^{\frac{1}{\gamma}} \left(N_{B_\xi} + \sum_{\lambda_j \in S_\xi \setminus B_\xi} \left(\frac{\xi(\lambda_j)}{\xi_{\max}} \right)^{\frac{1}{\gamma}} \right)} \\
&\xrightarrow{\gamma \rightarrow 0} \frac{1}{N_{B_\xi}} \sum_{\lambda_j \in B_\xi} \left(\frac{\xi'(\lambda_j)}{\xi_{\max}} \right) - 1.
\end{aligned}$$

Dabei ist zu beachten, daß der eingekästelte Teil für $\gamma \rightarrow 0$ verschwindet, da die Faktoren $\frac{\xi(\lambda_j)}{\xi_{\max}}$ für alle $\lambda_j \in S_\xi \setminus B_\xi$ echt kleiner als Eins sind und die beteiligten Summen endlich sind. Zusammen ist damit die Behauptung gezeigt. \square

Bemerkung 5.3. Das qualitativ gleiche Verhalten von $\phi_{t_0}^\gamma(\xi, \xi')$ für gegen Null konvergierende Werte γ ist zu beobachten, wenn statt der ersten Variante für die Wahl von $\sigma^2(j, \gamma, \xi)$ die zweite oder dritte gewählt wird (siehe Anhang B).

Die genannte Richtungsableitung fällt also in linearer Ordnung für $\gamma \rightarrow 0$. Eine vernünftige Kenngröße, welche die Abhängigkeit der Richtungsableitung von ξ' erfaßt, ist der Faktor, welcher zu dem linearen Term von $\phi_{t_0}^\gamma(\xi, \xi')$ in γ gehört. In [2] wird dieser Faktor Sensitivity Function genannt und ist dort Ausgangspunkt eines iterativen Algorithmus zur Annäherung an ein optimales Design für lokal gewichtete Regression (siehe auch in [15] und [37]). Dieser Faktor sei fortan mit ϕ_{t_0} bezeichnet und schreibt sich wie folgt:

$$\begin{aligned}
\phi_{t_0}(\xi, \xi') &= \lim_{\gamma \rightarrow 0} \frac{1}{\gamma} \phi_{t_0}^\gamma(\xi, \xi') \\
&= - \sum_{i,j,l: \lambda_i, \lambda_j, \lambda_l \in S_\xi} \boldsymbol{\Sigma}_{t_0}(s_\xi)_i^\top \boldsymbol{\Sigma}^{-1}(S_\xi)_{il} \Xi(\xi, \xi')_{il} \boldsymbol{\Sigma}^{-1}(S_\xi)_{lj} \boldsymbol{\Sigma}_{t_0}(s_\xi)_j \quad (5.11)
\end{aligned}$$

mit

$$\Xi(\xi, \xi') = \text{diag} \left(\sum_{\lambda_m \in S_\xi} \frac{\xi'(\lambda_m)}{N_{B_\xi} \xi_{\max}} - \frac{\xi'(\lambda_l)}{\xi(\lambda_l)} : \lambda_l \in S_\xi \right). \quad (5.12)$$

Für ein atomares Designmaß $\xi' = \delta_{\lambda_{j_0}}$ mit $\lambda_{j_0} \in S_\xi$ kann die allgemeine Form in (5.11) spezifiziert werden. Man erhält die folgende Fallunterscheidung:

$$\begin{aligned} \Phi_{t_0}(\xi, \delta_{\lambda_{j_0}}) & \tag{5.13} \\ = & \begin{cases} -\frac{1}{N_{B_\xi}} \left(\boldsymbol{\Sigma}_{t_0}(S_\xi)^\top \boldsymbol{\Sigma}_\lambda^{-1}(S_\xi) \right)^2 + \frac{1}{\xi_{\max}} \left(\left(\boldsymbol{\Sigma}_{t_0}(S_\xi)^\top \boldsymbol{\Sigma}_\lambda^{-1}(S_\xi) \right)_{j_0} \right)^2 & \lambda_{j_0} \in B_\xi \\ \frac{1}{\xi(\lambda_{j_0})} \left(\left(\boldsymbol{\Sigma}_{t_0}(S_\xi)^\top \boldsymbol{\Sigma}_\lambda^{-1}(S_\xi) \right)_{j_0} \right)^2 & \lambda_{j_0} \notin B_\xi. \end{cases} \end{aligned}$$

Bemerkung 5.4. Das Designmaß ξ' geht linear in das Kriterium $\Phi_{t_0}(\xi, \xi')$ ein. Damit ist dieses Kriterium im Sinne der Definition in [38] linear, d.h. es gilt für ein beliebiges Designmaß ξ' mit $S_{\xi'} \subset T$:

$$\Phi_{t_0}(\xi, \xi') = \sum_{\lambda_j \in T} \Phi_{t_0}(\xi, \delta_{\lambda_j}) \xi'(\lambda_j).$$

5.1.3. Ein Algorithmus und Beispiele

Anhand der Funktion in (5.13) soll nun ein erster, einfacher Algorithmus dargestellt werden, der iterativ die informativsten Samplefunktionale bestimmt, bei Eintreten eines Meßfehlers. In den darauf folgenden Beispielberechnungen für diesen Algorithmus werden zunächst nur gewöhnliche Deltafunktionale als Samplefunktionale verwendet. Das Verfahren ist aber ausdrücklich keinen Beschränkungen bezüglich der Wahl der Samplefunktionale unterworfen. Skizziere den Algorithmus wie folgt:

```

 $\xi_1$  sei ein beliebiges Designmaß, mit  $S_{\xi_1} = T$ 
for i = 1 to  $i_{\text{end}}$  do
  finde  $\lambda^{(i)*}$  mit  $\lambda^{(i)*} = \arg \max_{\lambda_j \in T} \Phi_{t_0}(\xi_i, \delta_{\lambda_j})$ 
   $\xi_{i+1} := \frac{i}{i+1} \xi_i + \frac{1}{i+1} \delta_{\lambda^{(i)*}}$ 

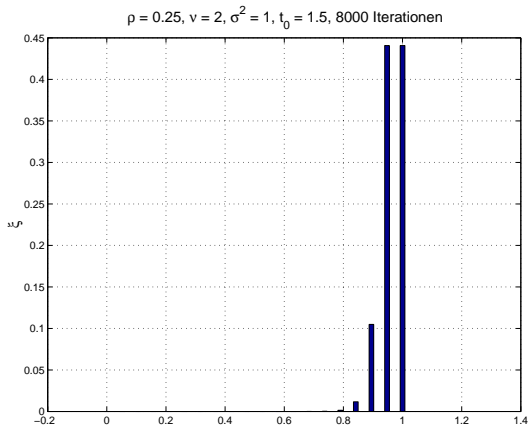
```

Die Abbruchkonstante i_{end} ist vorab zu wählen und stellt einen Kompromiß zwischen der Schnelligkeit, mit der das Verfahren abläuft, und der Stabilität des Ergebnisses dar. Generell konvergiert der Algorithmus stets für $i \rightarrow \infty$, was durch

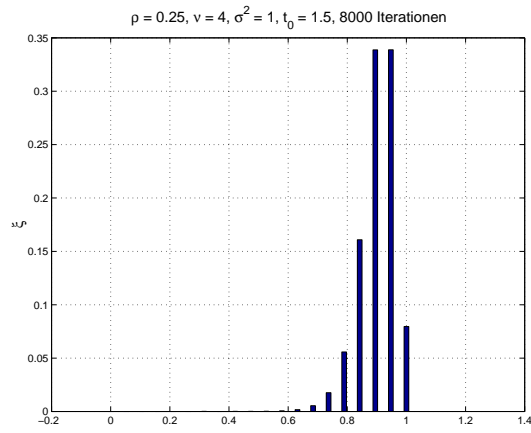
die Art der Definition der ξ_{i+1} begründet ist. So verschwindet für $i \rightarrow \infty$ in der jeweiligen Iteration der Einfluß des Maßes $\delta_{\lambda^{(i)*}}$, da dieses mit dem Faktor $1/i$ gewichtet wird.

Wie sich in numerischen Experimenten gezeigt hat, ist das Ergebnis des Verfahrens unabhängig von der Wahl des Startdesigns ξ_1 . In den Beispielen ist dies jeweils ein zufällig gewähltes Designmaß.

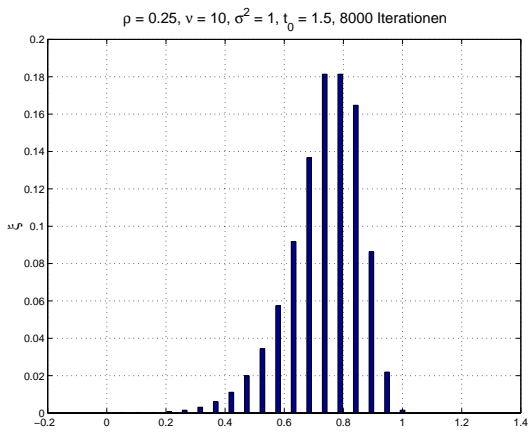
In Abbildung (5.2) ist der genannte Algorithmus auf ein Vorhersageproblem $\text{Pre}(t_0)$ mit $t_0 = 1.5$ für eine zufällige eindimensionale Funktion mit bekannter Erwartungswertfunktion angewendet worden. Dabei basiert die Vorhersage im quadratischen Mittel auf jeweils 20 äquidistanten Funktionswertmessungen im Intervall $[0, 1]$. Bei den Abbildungen ist mit Ausnahme von Bild (d) die zugrundeliegende Kovarianzfunktion die Matérnsche (siehe Beispiel 2.5(2)), wobei die Varianz identisch Eins und der Parameter der Reichweite $\rho = 0.25$ bzw. $\rho = 0.6$ gewählt wurde. Ferner wurde für die Wahl der Varianz des Meßfehlers die erste Variante von den auf Seite 63 vorgeschlagenen gewählt.



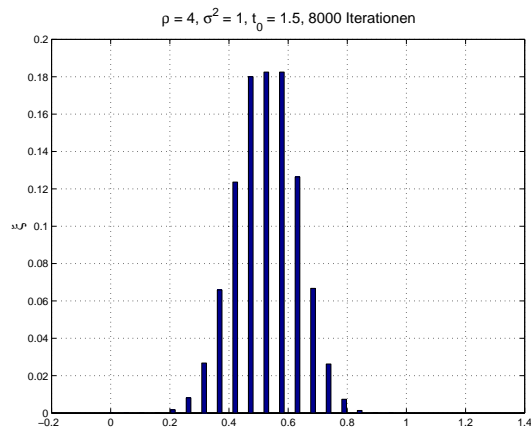
(a) Matérnsche Kovarianzfunktion, $l = 0$



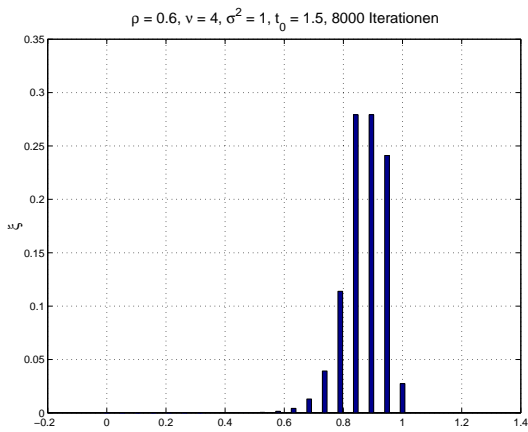
(b) Matérnsche Kovarianzfunktion, $l = 0$



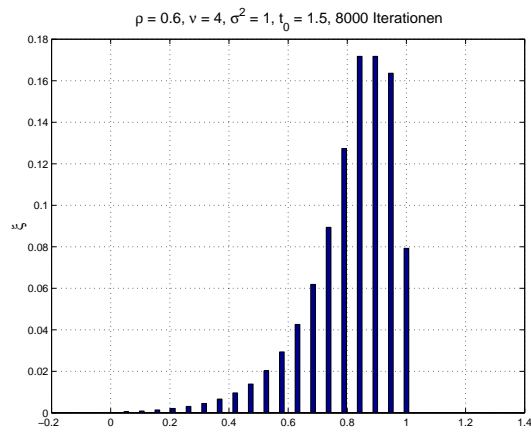
(c) Matérnsche Kovarianzfunktion, $l = 0$



(d) Gaußsche Kovarianzfunktion, $l = 0$



(e) Matérnsche Kovarianzfunktion, $l = 0$



(f) Matérnsche Kovarianzfunktion, $l = 2$

Abbildung 5.2.: Anwendung des Verfahrens auf ein Vorhersageproblem bei $t_0 = 1.5$ mit 20 auf dem Intervall $[0, 1]$ äquidistant verteilten Funktionsmessungen.

Bei den Bildern (a)–(c) fällt auf, daß sich bei zunehmender Differenzierbarkeit der zugrundeliegenden zufälligen Funktion die gemäß des Verfahrens informativsten Punkte von dem Vorhersageort zunehmend entfernen. Ein Extrembeispiel ist in Bild (d) dargestellt. Dort ist für die zugrundeliegende zufällige Funktion die Gaußsche Kovarianzfunktion gewählt worden (siehe Beispiel 2.5(1)). Dabei ist der Parameter der Reichweite hier gerade der Kehrwert von ρ . Somit sind die Parameter der Reichweite in diesem und in den Fällen (a)–(c) vergleichbar. Wie schon in den zitierten Beispielen angemerkt wurde, entspricht die Gaußsche Kovarianzfunktion dem Spezialfall einer Matérnschen mit Differenzierbarkeitsparameter $\nu = \infty$. Die Lage der informativsten Punkte hat sich hier noch weiter vom Vorhersageort entfernt, sie liegen nun symmetrisch um den Punkt 0.5.

Bei zunehmender Differenzierbarkeit gewinnen die vom Vorhersageort weiter entfernten Punkte an Bedeutung im Hinblick auf die Qualität der Vorhersage. Dieses Phänomen ist im Prinzip auch ein Kritikpunkt hinsichtlich der Verwendung der Gaußschen Kovarianzfunktion, als Extremfall einer unendlich oft differenzierbaren Kovarianzfunktion. Denn ist bei einem Gaußschen Prozeß mit ebendieser Kovarianzfunktion der Prozeß in einer infinitesimalen Umgebung eines Punktes bekannt, so läßt sich der Prozeß an jeder beliebig weit entfernten Stelle beliebig genau vorhersagen (siehe in [56]). Gleichzeitig jedoch werden die dem Vorhersageort am nächsten liegenden Punkte immer weniger stark gewichtet. Bedingt durch den auftretenden Meßfehler beeinflussen sie die Qualität der Vorhersage stets stark negativ, d.h. sie erhöhen den Vorhersagefehler. Nur bei einer relativ geringen Differenzierbarkeit scheint dieser Effekt kompensiert zu werden, so daß diese Punkte dennoch die für die Vorhersage informativsten Punkte sind (siehe Bild (a)).

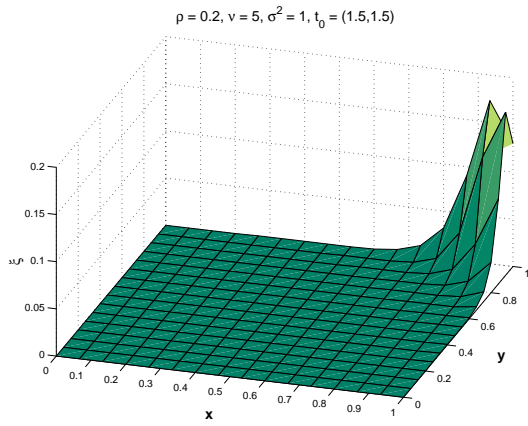
Wird der Parameter der Reichweite ρ erhöht, so ist der Effekt ähnlich, wie wenn der Grad der Differenzierbarkeit wächst (siehe in Bild (e)). Die Begründung ist dabei ebenfalls analog. Schließlich wird in Bild (f) für die Varianz des Meßfehlers die dritte Variante von den auf Seite 63 vorgeschlagenen, verwendet. Der Fall, daß der Parameter $l = 0$ ist, entspricht gerade der ersten Variante (siehe Anhang B) und durch die Verwendung von $l = 2$ wird in diesem Fall die Unterscheidung zwischen relativ informativeren und uninformativeren Punkten im Design „weicher“ gemacht (siehe dazu auch S. 63f.).

Für eine zu oben analoge Vorhersage, jedoch für eine zweidimensionale zufällige Funktion siehe Abbildung 5.3. Mit Ausnahme von Bild (d) wurde auch hier die Matérnsche Kovarianzfunktion gewählt. In Bild (c) dagegen, zum Vergleich wie oben, die Gaußsche Kovarianzfunktion, wobei auch hier der Range-Parameter so gewählt wurde, daß er mit den anderen Fällen vergleichbar ist. Die Vorhersage nach mittleren Quadraten wird jeweils für den Punkt $t_0 = (1.5, 1.5)$ gemacht und basiert auf Funktionswerten, die im Einheitsquadrat $[0, 1]^2$ auf einem uniformen 16×16 -Gitter gemessen werden.

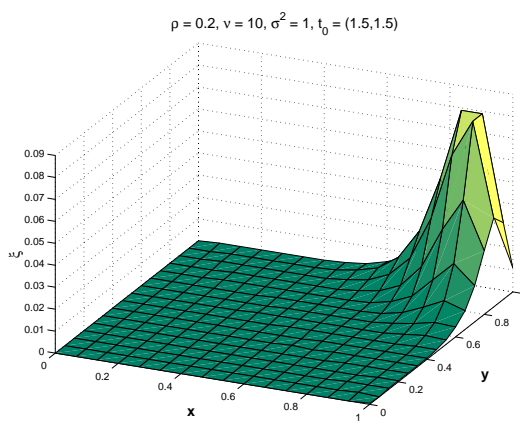
Die Ergebnisse spiegeln das bereits für den eindimensionalen dargestellte und interpretierte Verhalten bei Veränderung des Differenzierbarkeitsparameters analog wider.

5.1.4. Zusammenfassung und weitere Verwendungsmöglichkeiten

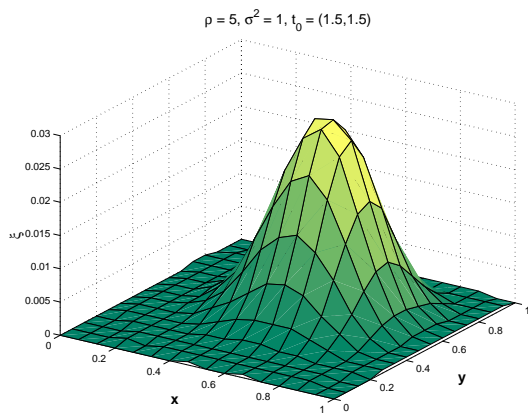
Der oben dargestellte Algorithmus sowie die dazugehörigen Beispiele sind nur eine erste einfache Anwendung des vorgestellten Prinzips der gestörten Kovarianzmatrizen. Zurückblickend war es das Ziel, eine Versuchsplanung für den Fall endlicher Versuchsdesigns zu



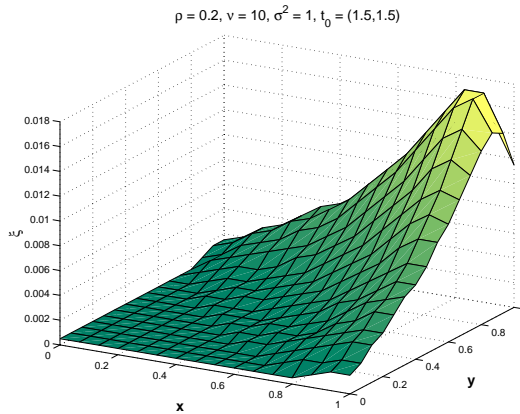
(a) Matérnsche Kovarianzfunktion, $l = 0$



(b) Matérnsche Kovarianzfunktion, $l = 0$



(c) Gaußsche Kovarianzfunktion, $l = 0$



(d) Matérnsche Kovarianzfunktion, $l = 4$

Abbildung 5.3.: Anwendung des Verfahrens auf ein Vorhersageproblem bei $t_0 = (1.5, 1.5)$ mit 256 auf dem Einheitsquadrat $[0, 1]^2$ gitterförmig verteilten Funktionsmessungen.

ermöglichen, wobei das zugrundeliegende Experiment eine Vorhersage mit Fehlerevaluierung nach mittleren Quadraten war. Bisher wird diese Art der Versuchsplanung praktisch ausschließlich mittels numerischer Optimierungsverfahren behandelt, die in der Regel direkt versuchen, ein Design zu finden, welches den mittleren quadratischen Fehler in (5.2) unter gewissen Nebenbedingungen (z.B. die Größe oder die Zusammensetzung des Designs betreffend) minimiert (siehe auch Abschnitt 3.3). Dagegen wurde mit dem hier vorgestellten Prinzip eine Möglichkeit aufgezeigt, das Problem tiefergehend analytisch zu behandeln, um dann einen einfachen, wie auch schnellen Algorithmus für die Suche nach einem optimalen Design zu verwenden. In gewisser Weise ist das Vorgehen durch die Methodik in der klassischen Versuchsplanung motiviert. Im Rahmen des Theorems von Whittle über die Minimierung konvexer Fehlerfunktionale wird ebenfalls ein iterativer Algorithmus zur Auffindung eines optimalen Designs vorgeschlagen, der auf der Richtungsableitungen des Fehlerfunktionals

basiert (siehe in [38]). Die Differenzierbarkeitseigenschaft des \square -Funktional geht jedoch im nichtklassischen, also dem Fall korrelierter Samplefunktionale, verloren, wie oben gezeigt wurde. Daher wurde das ursprüngliche \square -Funktional differenzierbar modifiziert. In dem gezeigten Beispiel wurde dieses Funktional für das Auffinden der informativsten Punkte innerhalb einer Menge von Samplefunktionalen bei Einsetzen von Meßfehlern verwendet.

Eine mögliche weitere Anwendung ist das Auffinden eines optimalen Teildesigns $T_s \subset T$, wobei T_s nur $s < k$ Samplefunktionale enthält. In [37] und [41] wird eine Verallgemeinerung des oben dargestellten Algorithmus vorgestellt, der es ermöglicht, die beabsichtigte Designgröße s bei den Iterationsschritten zu berücksichtigen. Dazu werden in jedem Iterationsschritt die Designmaße ξ_i in der Höhe $1/s$ gekappt. Es wird also statt mit ξ_i mit den Maßen $\xi_i^\# := \max(0, \xi_i - 1/s)$ gearbeitet. Es gilt somit bereits $\xi_i^\#(\lambda_j) = 0$, falls $\xi_i^\#(\lambda_j) \leq 1/s$. Das Samplefunktional λ_j wird also in diesem Fall bereits als nicht informativ in bezug auf die relative Maximierung des \square -Funktional eingestuft, obwohl vielleicht sogar $\xi_i(\lambda_j) > 0$ gilt. Um wiederum eine differenzierbare Modifikation des \square -Funktional zu erhalten, werden in den zitierten Arbeiten die gekappten Designmaße $\xi_i^\#$ analog zu der Approximation von ξ_{\max} , differenzierbar approximiert.

Literaturverzeichnis

- [1] ABRAMOWITZ, M. UND STEGUN, I. A. *Pocketbook of mathematical functions*. Verlag Harri Deutsch, Frankfurt a.M., 1984.
- [2] AD G. MONTEPIEDRA, V. V. F. UND NACHTSHEIM, C. J. Designs of experiments for locally weighted regression. *Journal of Statistical Planning and Inference* 81 (1999), 363–382.
- [3] AITKEN, A. C. *Determinanten und Matrizen*. BI Hochschultaschenbücher, 1964.
- [4] ALBRECHT, M. Geostatistische Meßnetzplanung. Techn. Bericht, TU Bergakademie Freiberg, 1996. Forschungsteilbericht zum Thema: Markscheiderisch-geologische Komplexauswertung unter geostatistischen Modellannahmen, DFG III 03-Me 1111/2-1.
- [5] ALBRECHT, M. Geostatistische Meßnetzplanung für das Gradientenkriging. Techn. Bericht, TU Bergakademie Freiberg, 1998. Forschungsteilbericht zum Thema: Markscheiderisch-geologische Komplexauswertung unter geostatistischen Modellannahmen, DFG III 03-Me 1111/2-1.
- [6] ATTEIA, M. *Hilbertian Kernels and Spline Functions*. North Holland, Amsterdam, 1992.
- [7] BATES, R. A., BUCK, R. J., RICCOMAGNO, E. UND WYNN, H. P. Experimental Design and Observation for Large Systems. *J. R. Statist. Soc. B* 58, 1 (1996), 77–94.
- [8] BAUER, H. *Wahrscheinlichkeitstheorie*, 4. Aufl. De Gruyter, Berlin, 1990.
- [9] CHRISTENSEN, R. *Linear Models for Multivariate, Time Series, and Spatial Data*. Springer-Verlag, New York, 1991.
- [10] CHRISTENSEN, R. *Plane Answers to Complex Questions: The Theory of Linear Models*, 2. Aufl. Springer-Verlag, New York, 1996.
- [11] CLEVELAND, W. S. UND LOADER, C. Smoothing by Local Regression: Principles and Methods. In *Statistical Theory and Computational Aspects of Smoothing*, W. Hardle and M. G. Schimek, Hrsg. Physica-Verlag, Heidelberg, 1996.
- [12] CRESSIE, N. A. C. *Statistics for Spatial Data*. Wiley, New York, 1991.
- [13] CURRIN, C., MITCHELL, T., MORRIS, M. UND YLVIKAKER, D. Bayesian Prediction of Deterministic Functions, With Application to the Design and Analysis of Computer Experiments. *Journal of the American Statistical Society* 86, 416 (1991), 953–963.

- [14] DOOB, J. L. *Stochastic Processes*. Wiley, New York, 1953.
- [15] FEDOROV, V. V. UND HACKL, P. *Model-Oriented design of experiments*. Springer-Verlag, New York, 1997.
- [16] GAO, H., WANG, J. UND ZHAO, P. The Updated Kriging Variance and Optimal Sample Design. *Mathematical Geology* 28, 3 (1996), 295–313.
- [17] GASPARI, G. UND COHN, S. E. Construction of correlation functions in two and three dimensions. *Q. J. R. Meteorol. Soc.* 125 (1999), 723–757.
- [18] GNEITING, T. Correlation Functions for atmospheric data analysis. *Q. J. R. Meteorol. Soc.* 125 (1999), 2449–2464.
- [19] GOOVAERTS, P. *Geostatistics for Natural Resources Evaluation*. Oxford University Press, New York, 1997.
- [20] HANDCOCK, M. S. UND WALLIS, J. R. An approach to statistical spatial-temporal modeling of meteorological fields (with discussion). *J. Amer. Statist. Assoc.*, 89 (1994), 368–390.
- [21] JÄKEL, M. Anwendung des Gradientenkrigings bei der Erkundung des 2. Miozänen Flözhorizontes in der Niederlausitz. Techn. Bericht, TU Bergakademie Freiberg, 1998. Forschungsteilbericht zum Thema: Markscheiderisch-geologische Komplexauswertung unter geostatistischen Modellannahmen, DFG III 03-Me 1111/2-1.
- [22] JANTSCHER, L. *Hilberträume*. Akademische Verlagsgesellschaft, Wiesbaden, 1977.
- [23] JOURNAL, A. G. UND HUIJBREGTS, C. J. *Mining Geostatistics*. Academic Press, London, 1978.
- [24] KENT, J. T. UND MARDIA, M. V. The Link Between Kriging and Thin-plate Splines. In *Probability, Statistics and Optimisation*, F. P. Kelly, Hrsg. Wiley, New York, 1994.
- [25] LANCASTER, P. UND ŠALKAUSKAS, K. *Curve and Surface Fitting: An Introduction*. Academic Press, London, 1986.
- [26] LASINGER, R. Integration of covariance kernels and stationarity. *Stochastic Processes and their Application* 45 (1993), 309–318.
- [27] LOÉVE, M. Fonctions aléatoires du second ordre. In *Processus stochastiques et mouvement Brownien*, P. Lévy, Hrsg. Gauthier-Villars, Paris, 1965.
- [28] MARDIA, K. V., KENT, J. T., GOODALL, C. R. UND LITTLE, J. A. Kriging and splines with derivative information. *Biomatrika* 83, 1 (1996), 207–221.
- [29] MARRON, J. S. A Personal View of Smoothing and Statistics. In *Statistical Theory and Computational Aspects of Smoothing*, W. Härdle and M. G. Schimek, Hrsg. Physica-Verlag, Heidelberg, 1996.
- [30] MATÉRN, B. *Spatial Variation*, 2. Aufl. Springer-Verlag, New York, 1986.

- [31] MATHERON, G. Principles of geostatistics. *Economic Geology* 58 (1963), 1246–1266.
- [32] MENZ, J. Geostatistische Vorhersage des Schichtenverlaufs im Gebirge auf der Grundlage von Bohrungen, Stoßbemusterungen und geophysikalischen Messungen. *Das Markscheidewesen* 98, 2 (1991), 70–73.
- [33] MENZ, J. Anwendung der Geostatistik zur Gebirgs- und Lagerstätten-geometrisierung. In *Neue Wege in Exploration und Produktion* (1996), 16. Minitrop-Seminar der DGMK Deutsche Wissenschaftliche Gesellschaft für Erdöl, Erdgas und Kohle e.V.
- [34] MITCHELL, T., MORRIS, M. UND YLVISAKER, D. Existence Of Smoothed Stationary Processes On An Interval. *Stochastic Processes and their Applications* 35 (1990), 109–119.
- [35] MONTÉS, O. Kriging interpolation of a Bèzier curve. *Computer aided design* 23, 10 (1991), 713–716.
- [36] MORRIS, M. D., MITCHELL, T. J. UND YLVISAKER, D. Bayesian Design and Analysis of Computer Experiments: Use of Derivatives in Surface Prediction. *Technometrics* 35, 3 (1993), 243–255.
- [37] MÜLLER, W. G. *Collecting Spatial Data*. Physica-Verlag, Heidelberg, 1998.
- [38] NÄTHER, W. *Effective Observation of Random Fields*. Teubner, Leipzig, 1985.
- [39] NOVAK, E. On the Power of Adaption. *Journal Of Complexity* 12 (1996), 199–237.
- [40] PÁZMAN, A. UND MÜLLER, W. G. A New Interpretation of Design Measures. In *MO-DA5 — Advances in Model-Oriented Data Analysis and Experimental Design*, A. C. Atkinson, L. Pronzato, and H. P. Wynn, Hrsg. Physica-Verlag, Heidelberg, 1998.
- [41] PÁZMAN, A. UND MÜLLER, W. G. Properties of Design Measures for Optimal Observations from Correlated Processes. Report No. 411, Augsburg, 1999.
- [42] RAO, C. R. UND TOUTENBURG, H. *Linear Models. Least Squares and Alternatives*. Springer-Verlag, New York, 1995.
- [43] REDDY, B. D. *Introductory functional analysis: with applications to boundary value problems and finite elements*. Springer-Verlag, New York, 1998.
- [44] RITTER, K. *Average Case Analysis of Numerical Problems*. Springer-Verlag, New York, 2000.
- [45] RITTER, K. UND MÜLLER-GRONBACH, T. Spatial Adaption For Predicting Random Functions. Preprint Nr. A-6-97, Erlangen, 1997.
- [46] RITTER, K., WASILKOWSKI, G. W. UND WOŹNIAKOWSKI, H. Multivariate Integration And Approximation For Random Fields Satisfying Sacks-Ylvisaker Conditions. *Ann. Appl. Prob.* 5 (1995), 518–540.

- [47] SACKS, J. UND SCHILLER, S. Spatial Designs. In *Statistical Decision Theory and Related Topics IV*, S. S. Gupta and J. O. Berger, Hrsg., vol. 2. Springer-Verlag, New York, 1988.
- [48] SACKS, J., WELCH, W. J., MITCHELL, T. J. UND WYNN, H. P. Design and Analysis of Computer Experiments. *Statistical Science* 4, 4 (1989), 409–435.
- [49] SACKS, J. UND YLVISAKER, D. Designs For Regression Problems With Correlated Errors. *Ann. Math. Statist.* 37 (1966), 66–89.
- [50] SACKS, J. UND YLVISAKER, D. Designs For Regression Problems With Correlated Errors; Many Parameters. *Ann. Math. Statist.* 39 (1968), 49–69.
- [51] SACKS, J. UND YLVISAKER, D. Designs For Regression Problems With Correlated Errors III. *Ann. Math. Statist.* 41, 6 (1970), 2057–2074.
- [52] SCHMITZ, N. *Vorlesungen über Wahrscheinlichkeitstheorie*. Teubner, Stuttgart, 1996.
- [53] SCHOTT, J. R. *Matrix Analysis for Statistics*. Wiley, New York, 1997.
- [54] SEIFERT, B. UND GASSER, T. Variance Properties of Local Polynomials and Ensuing Modifications. In *Statistical Theory and Computational Aspects of Smoothing*, W. Hardle and M. G. Schimek, Hrsg. Physica-Verlag, Heidelberg, 1996.
- [55] SHEWRY, M. C. UND WYNN, H. P. Maximum Entropy Sampling. *J. Appl. Stat.* 14, 2 (1987), 165–170.
- [56] STEIN, M. L. *Interpolation of spatial data: some theory for kriging*. Springer-Verlag, New York, 1999.
- [57] SU, Y. Estimation of random fields by piecewise constant estimators. *Stoch. Proc. Appl.* 71 (1997), 145–163.
- [58] TONG, Y. L. . *The Multivariate Normal Distribution*. Springer-Verlag, New York, 1990.
- [59] TRAUB, J. F., WASILKOWSKI, G. W. UND WOŹNIAKOWSKI, H. *Information-based complexity*. Academic Press, New York, 1988.
- [60] WAHBA, G. *Spline Models for observational Data*. CBMS-NSF Regional Conference Series in Applied Mathematics. SIAM, 1990.
- [61] YAGLOM, A. M. *Correlation Theory of Stationary and Random Functions, Volume 1: Basic Results*. Springer-Verlag, New York, 1987.
- [62] YAGLOM, A. M. *Correlation Theory of Stationary and Random Functions, Volume 2: Supplementary Notes and References*. Springer-Verlag, New York, 1987.

A. Anhang

Lemma. Sei $(H, \langle \cdot, \cdot \rangle_H)$ ein Hilbertraum und $B(H) = \{z \in H : \|z\|_H \leq 1\}$ die abgeschlossene Einheitskugel in H . Die Norm $\|\cdot\|_H$ ist dabei die durch das Skalarprodukt erzeugte. Es gilt für alle $x \in H$:

$$\|x\|_H = \sup_{h \in B(H)} |\langle x, h \rangle_H|.$$

Beweis. Sei o.B.d.A. $x \in H \setminus 0$.

(i) Es gilt:

$$\begin{aligned} \|x\|_H^2 &= \langle x, x \rangle_H = \|x\|_H \langle x, \frac{x}{\|x\|_H} \rangle_H \\ \Rightarrow \|x\|_H &= \left| \langle x, \frac{x}{\|x\|_H} \rangle_H \right|. \end{aligned}$$

Da $\frac{x}{\|x\|_H} \in B(H)$ ist, gilt also $\|x\|_H \leq \sup_{h \in B(H)} |\langle x, h \rangle_H|$.

(ii) Sei $x \in H$ beliebig gewählt. Es gilt dann $x = \|x\|_H e$ mit dem Einheitsvektor $e = \frac{x}{\|x\|_H}$. Damit folgt:

$$\begin{aligned} \sup_{h \in B(H)} |\langle x, h \rangle_H| &= \sup_{h \in B(H)} \|\|x\|_H \langle e, h \rangle_H\| \\ &= \|x\|_H \sup_{h \in B(H)} |\langle e, h \rangle_H| \\ &\stackrel{(*)}{\leq} \|x\|_H \sup_{h \in B(H)} (\|e\|_H \|h\|_H) = \|x\|_H. \end{aligned}$$

Dabei ist die mit (*) gekennzeichnete Ungleichung gerade die Cauchy-Schwartsche.

Aus (i) und (ii) folgt die Behauptung. \square

B. Anhang

B.1. Konvergenzverhalten von alternativen Varianztermen

$$\sigma^2(j, \gamma, \xi) = \left(\left(\frac{1}{\xi(j)} \right)^\gamma - \left(\frac{1}{\xi^{(\gamma)}} \right)^\gamma \right):$$

Zur Berechnung der Richtungsableitung $\phi_{t_0}^\gamma(\xi, \xi')$ werte den Differentiationsterm aus (5.8) aus. Sei ein beliebiges aber festes m so gewählt, daß $\lambda_m \in S_\xi$ liegt, dann gilt:

$$\begin{aligned} \left(\frac{\partial}{\partial \alpha} \mathbf{W}^\gamma(\xi_\alpha) \right)_{mm} &= -\gamma \left(\frac{\xi'(\lambda_m) - \xi(\lambda_m)}{\xi_\alpha^{\gamma+1}(\lambda_m)} - \frac{1}{\xi_\alpha^{(\gamma)\gamma+1}} \frac{\partial}{\partial \alpha} \xi_\alpha^{(\gamma)} \right) \\ &= -\gamma \left(\frac{\xi'(\lambda_m) - \xi(\lambda_m)}{\xi_\alpha^{\gamma+1}(\lambda_m)} - \frac{\sum_{\lambda_j \in S_\xi} \xi_\alpha^{\frac{1}{\gamma}}(\lambda_j) \frac{\xi'(\lambda_j) - \xi(\lambda_j)}{\xi_\alpha(\lambda_j)}}{\xi_\alpha^{(\gamma)\gamma} \sum_{\lambda_j \in S_\xi} \xi_\alpha^{\frac{1}{\gamma}}(\lambda_j)} \right) \\ &\xrightarrow{\alpha \rightarrow 0} -\gamma \left(\frac{\xi'(\lambda_m) - \xi(\lambda_m)}{\xi^{\gamma+1}} - \frac{\sum_{\lambda_j \in S_\xi} \xi^{\frac{1}{\gamma}}(\lambda_j) \frac{\xi'(\lambda_j) - \xi(\lambda_j)}{\xi(\lambda_j)}}{\xi^{(\gamma)\gamma} \sum_{\lambda_j \in S_\xi} \xi^{\frac{1}{\gamma}}(\lambda_j)} \right). \end{aligned}$$

Die Richtungsableitung $\phi_{t_0}^\gamma(\xi, \xi')$ ist also ebenso von linearer Ordnung in γ für $\gamma \rightarrow 0$. Zur Berechnung des Grenzwertes $\lim_{\gamma \rightarrow 0} \frac{1}{\gamma} \phi_{t_0}^\gamma(\xi, \xi')$ schreibe:

$$\begin{aligned} &\lim_{\gamma \rightarrow 0} \lim_{\alpha \rightarrow 0} \frac{1}{\gamma} \left(\frac{\partial}{\partial \alpha} \mathbf{W}^\gamma(\xi_\alpha) \right)_{mm} \\ &= -\lim_{\gamma \rightarrow 0} \left(\frac{\xi'(\lambda_m) - \xi(\lambda_m)}{\xi^{\gamma+1}} - \frac{1}{\xi^{(\gamma)\gamma}} \frac{\sum_{\lambda_j \in S_\xi} \xi^{\frac{1}{\gamma}}(\lambda_j) \frac{\xi'(\lambda_j) - \xi(\lambda_j)}{\xi(\lambda_j)}}{\sum_{\lambda_j \in S_\xi} \xi^{\frac{1}{\gamma}}(\lambda_j)} \right) \\ &\stackrel{(*)}{=} \left(\sum_{\lambda_j \in B_\xi} \frac{\xi'(\lambda_j)}{N_{B_\xi} \xi_{\max}} - \frac{\xi'(\lambda_m)}{\xi(\lambda_m)} \right). \end{aligned}$$

An der mit (*) gekennzeichneten Stelle ist zu beachten, daß $\frac{1}{\xi^{(\gamma)\gamma}}$ für $\gamma \rightarrow 0$ gegen 1 konvergiert.

Damit stimmt die Funktion $\phi_{t_0}(\xi, \xi')$ überein mit der in (5.13) berechneten.

$$\sigma^2(j, \gamma, \xi) = \frac{\gamma}{l} \left(\left(\frac{1}{\xi(j)} \right)^l - \left(\frac{1}{\xi(\gamma)} \right)^l \right):$$

Verwende die gleiche Indizeswahl wie oben und erhalte damit:

$$\begin{aligned} \left(\frac{\partial}{\partial \alpha} \mathbf{W}^\gamma(\xi_\alpha) \right)_{\text{mm}} &= -\gamma \left(\frac{\xi'(\lambda_m) - \xi(\lambda_m)}{\xi_\alpha^{l+1}(\lambda_m)} - \frac{\sum_{\lambda_j \in S_\xi} \xi_\alpha^{\frac{1}{\gamma}}(\lambda_j) \frac{\xi'(\lambda_j) - \xi(\lambda_j)}{\xi_\alpha(\lambda_j)}}{\xi_\alpha^{(\gamma)l} \sum_{\lambda_j \in S_\xi} \xi_\alpha^{\frac{1}{\gamma}}(\lambda_j)} \right) \\ &\xrightarrow{\alpha \rightarrow 0} -\gamma \left(\frac{\xi'(\lambda_m) - \xi(\lambda_m)}{\xi^{l+1}(\lambda_m)} - \frac{\sum_{\lambda_j \in S_\xi} \xi^{\frac{1}{\gamma}}(\lambda_j) \frac{\xi'(\lambda_j) - \xi(\lambda_j)}{\xi(\lambda_j)}}{\xi^{(\gamma)l} \sum_{\lambda_j \in S_\xi} \xi^{\frac{1}{\gamma}}(\lambda_j)} \right). \end{aligned}$$

Auch hier ist somit die Richtungsableitung $\Phi_{t_0}^\gamma(\xi, \xi)$ von linearer Ordnung in γ für $\gamma \rightarrow 0$. Berechne schließlich in Analogie zum obigen Fall den folgenden Grenzwert:

$$\lim_{\gamma \rightarrow 0} \lim_{\alpha \rightarrow 0} \frac{1}{\gamma} \left(\frac{\partial}{\partial \alpha} \mathbf{W}^\gamma(\xi_\alpha) \right)_{\text{mm}} = \left(\sum_{\lambda_j \in B_\xi} \frac{\xi'(\lambda_j)}{N_{B_\xi} \xi_{\max}^{l+1}} - \frac{\xi'(\lambda_m)}{\xi^{l+1}(\lambda_m)} - \left(\frac{1}{\xi_{\max}^l} - \frac{1}{\xi^l(\lambda_j)} \right) \right).$$

Die in (5.11) eingeführte Größe soll abschließend für den Fall, daß ξ' ein atomares Designmaß ist, dargestellt werden. Sei also $\xi' = \delta_{\lambda_{j_0}}$ gewählt, wobei $\lambda_{j_0} \in S_\xi$ ist. Es gilt dann:

$$\begin{aligned} &\Phi_{t_0}(\xi, \delta_{\lambda_{j_0}}) \\ &= \begin{cases} -\frac{1}{N_{B_\xi} \xi_{\max}^{l+1}} \left(\boldsymbol{\Sigma}_0^T(S_\xi) \boldsymbol{\Sigma}_\lambda^{-1}(S_\xi) \right)^2 + \frac{1}{\xi_{\max}^{l+1}} \left(\left(\boldsymbol{\Sigma}_0^T(S_\xi) \boldsymbol{\Sigma}_\lambda^{-1}(S_\xi) \right)_{j_0} \right)^2 + A & \lambda_{j_0} \in B_\xi \\ \frac{1}{\xi^{l+1}(\lambda_{j_0})} \left(\left(\boldsymbol{\Sigma}_0^T(S_\xi) \boldsymbol{\Sigma}_\lambda^{-1}(S_\xi) \right)_{j_0} \right)^2 + A & \lambda_{j_0} \notin B_\xi \end{cases} \end{aligned}$$

mit der Größe

$$A = \frac{1}{\xi_{\max}^{l+1}} \left(\boldsymbol{\Sigma}_0^T(S_\xi) \boldsymbol{\Sigma}_\lambda^{-1}(S_\xi) \right)^2 - \left(\boldsymbol{\Sigma}_0^T(S_\xi) \boldsymbol{\Sigma}_\lambda^{-1}(S_\xi) \Xi^l(\xi, \xi') \boldsymbol{\Sigma}_\lambda^{-1}(S_\xi) \boldsymbol{\Sigma}_0^T(S_\xi) \right).$$

In dem in Kapitel 5 vorgeschlagenen Algorithmus zur iterativen Bestimmung der informativsten Samplefunktionale bei Eintreten von Meßfehlern wird in jedem Iterationsschritt nach demjenigen Funktional $\lambda_{j_0} \in S_\xi$ gesucht, welches das Funktional $\Phi_{t_0}(\xi, \delta_{\lambda_{j_0}})$ maximiert. Daher kann der Summand A im obigen Ausdruck weggelassen werden, da er in beiden der zwei zu unterscheidenden Fälle auftritt.

C. Anhang

C.1. Zerlegung der Schätzvarianz bei lokaler Regression

Gegeben sei der lokale Regressionsschätzer $\hat{g}(t_0) = L_{\text{lok}}^T(t_0)Z$ mit dem zugehörigen Samplevektor $Z^T = (Z_1, \dots, Z_k, Z_{k+1})$. Nehme ferner an, daß die Varianz σ^2 der Schätzfehler bekannt sei und wähle sie o.B.d.A. identisch 1. Die Schätzvarianz erhält dann die Form:

$$\text{Var } \hat{g}(t_0) = f^T(t_0) \left(\mathbf{F}^T \mathbf{W}(t_0) \mathbf{F} \right)^{-1} \mathbf{F}^T \mathbf{W}(t_0)^2 \mathbf{F} \left(\mathbf{F}^T \mathbf{W}(t_0) \mathbf{F} \right)^{-1} f(t_0).$$

Fortan soll der Übersichtlichkeit halber in der Bezeichnungsweise die Abhängigkeit der Matrix \mathbf{W} vom Schätzzort t_0 fortgelassen werden. Mit den Bezeichnungen $\mathbf{a} = \sqrt{w_{k+1}} \lambda_{k+1}(f)$ und $\mathbf{a}' = w_{k+1} \lambda_{k+1}(f)$ sind die folgenden Zerlegungen von Matrizen im obigen Ausdruck möglich:

$$\begin{aligned} \mathbf{F}^T \mathbf{W} \mathbf{F} &= \mathbf{S} + \mathbf{a} \mathbf{a}^T := \tilde{\mathbf{F}}^T \tilde{\mathbf{W}} \tilde{\mathbf{F}} + \mathbf{a} \mathbf{a}^T \\ \mathbf{F}^T \mathbf{W}^2 \mathbf{F} &= \mathbf{S}' + \mathbf{a}' (\mathbf{a}')^T := \tilde{\mathbf{F}}^T \tilde{\mathbf{W}}^2 \tilde{\mathbf{F}} + \mathbf{a}' (\mathbf{a}')^T. \end{aligned}$$

Die Matrix $\tilde{\mathbf{F}}$ ist ganz analog zu der in Abschnitt 4.2 definiert, und es gelte ferner $\tilde{\mathbf{W}} = \text{diag}(w_1(t_0), \dots, w_k(t_0))$. Damit läßt sich die Varianz selbst wie folgt zerlegen:

$$\begin{aligned} \text{Var } \hat{g}(t_0) &= f^T(t_0) \left(\mathbf{S}^{-1} - \mathbf{V} \right) \left(\mathbf{S}' + \mathbf{a}' (\mathbf{a}')^T \right) \left(\mathbf{S}^{-1} - \mathbf{V} \right) f(t_0) \\ &= \text{Var } \hat{g}(t_0) + f^T(t_0) \left(\mathbf{S}^{-1} \mathbf{a}' (\mathbf{a}')^T \mathbf{S}^{-1} - \mathbf{V} \mathbf{S}' \mathbf{S}^{-1} - \mathbf{V} \mathbf{a}' (\mathbf{a}')^T \mathbf{S}^{-1} \right. \\ &\quad \left. - \mathbf{S}^{-1} \mathbf{S}' \mathbf{V} - \mathbf{S}^{-1} \mathbf{a}' (\mathbf{a}')^T \mathbf{V} + \mathbf{V} \mathbf{S}' \mathbf{V} + \mathbf{V} \mathbf{a}' (\mathbf{a}')^T \mathbf{V} \right) f(t_0). \end{aligned}$$

Dabei sei $\mathbf{V} = \frac{\mathbf{S}^{-1} \mathbf{a} \mathbf{a}^T \mathbf{S}^{-1}}{1 + \mathbf{a}^T \mathbf{S}^{-1} \mathbf{a}}$ definiert, und \hat{g} sei der entsprechende lokale Regressionsschätzer unter Verwendung der Samples Z_1, \dots, Z_k .