

DEUTSCHE GEODÄTISCHE KOMMISSION
bei der Bayerischen Akademie der Wissenschaften

Reihe C

Dissertationen

Heft Nr. 563

Svetozar Petrovif

**Parameterschätzung
für unvollständige funktionale Modelle in der Geodäsie**

München 2003

**Verlag der Bayerischen Akademie der Wissenschaften
in Kommission bei der C. H. Beck'schen Verlagsbuchhandlung München**

ISSN 0065-5325

ISBN 3 7696 5002 6

DEUTSCHE GEODÄTISCHE KOMMISSION
bei der Bayerischen Akademie der Wissenschaften

Reihe C

Dissertationen

Heft Nr. 563

Parameterschätzung
für unvollständige funktionale Modelle in der Geodäsie

Habilitationsschrift
an der Fakultät VI – Bauingenieurwesen und Angewandte Geowissenschaften
der Technischen Universität Berlin
Lehrgebiet: Theoretische und Angewandte Geodäsie

vorgelegt von

Dr. Svetozar Petrović

München 2003

Verlag der Bayerischen Akademie der Wissenschaften
in Kommission bei der C. H. Beck'schen Verlagsbuchhandlung München

Adresse der Deutschen Geodätischen Kommission:

Deutsche Geodätische Kommission

Marstallplatz 8 • D – 80 539 München

Telefon (089) 23 031 1113 • Telefax (089) 23 031 – 283/ – 100

E-mail hornik@dgfi.badw.de • <http://www.dgfi.badw.de/dgk/dgk.html>

Gutachter: Prof. Dr. Dieter Lelgemann
Prof. Dr. Lothar Gründig
Prof. Dr. Erik Grafarend

Eröffnung des Verfahrens: 05.12.2001

Verleihung der Lehrbefähigung: 30.10.2002

© 2003 Deutsche Geodätische Kommission, München

Alle Rechte vorbehalten. Ohne Genehmigung der Herausgeber ist es auch nicht gestattet,
die Veröffentlichung oder Teile daraus auf photomechanischem Wege (Photokopie, Mikrokopie) zu vervielfältigen

Inhaltsverzeichnis

Einleitung	1
1 Motivation	3
1.1 Unvollständige Modelle	3
1.2 Behandlung systematischer Fehler in der Literatur	4
1.3 Zum Nachweis der Existenz systematischer Einflüsse in den Daten	7
1.4 Zur Ausgleichung unvollständiger Modelle	9
2 Kurze Darstellung . . .	11
2.1 Allgemeine Bemerkungen	11
2.2 Korrelationskoeffizient	13
2.3 Ausgleichung nach maximaler Korrelation	15
2.3.1 Definition	15
2.3.2 Verallgemeinerung auf gewichtete und korrelierte Beobachtungen	15
2.3.3 Mehrdeutigkeit der Lösung	18
2.3.4 Lösung des Ausgleichungsproblems nach maximaler Korrelation	19
2.3.5 Lineare Probleme	21
2.3.6 Beziehung zur Methode der kleinsten Quadrate	24
3 Anwendung . . . bei unvollständigen Modellen	25
3.1 Allgemeine Strategie	25
3.2 Residuenanalyse	26
3.2.1 Aufstellen zusätzlicher Bedingungen	28
3.2.2 Eindimensionaler Fall	30
4 Synthetische Beispiele	35
4.1 Beispiel ohne zufällige Fehler	35
4.2 Einführung zufälliger Fehler	38
4.3 Einfluß von Datenlücken	38
4.4 Nicht-eindeutige Aufgabenstellungen	42
5 Schlußbetrachtung	45
Literaturverzeichnis	49
Anhang	53
A.1 „Total least squares“ und Verallgemeinerungen davon	53
A.2 „Hidden bias“ in „Observational studies“	56
A.3 Literatur	59

Einleitung

Die Tatsache, daß die Realität sehr kompliziert aufgebaut ist und keine vollständige Beschreibung zuläßt, ist grundsätzlich allgemein bekannt. In Wissenschaften wie der Geodäsie, in denen die Beschreibung der Gegebenheiten auf Messungsergebnissen beruht und auf einem mathematischen Wege erfolgt, ist die Spaltung zwischen der Realität und dem Modell besonders offensichtlich. Alle Berechnungen und Datenmanipulationen werden in einem Modell durchgeführt, der Bezug zur Realität ist prinzipiell mehr oder weniger fragwürdig.

Die Feststellung, daß es streng genommen vollständige Modelle weder gibt noch geben kann, ist zunächst deprimierend. Sie ist jedoch gleichzeitig auch sehr hilfreich um einzusehen, daß die Betrachtungen auf eine relative Vollständigkeit eingeschränkt werden müssen. Da die Meßdaten selbst nur eine eingeschränkte Auflösung und Genauigkeit besitzen können, kann jede Beschreibung der geometrischen oder physikalischen Realität nur innerhalb der gleichen Grenzen erfolgen.

Mit der oben eingeführten Einschränkung wird prinzipiell eine Modellbildung möglich, die im Rahmen der gegebenen oder gestellten Auflösungs- und Genauigkeitsgrenzen als vollständig betrachtet werden kann. In der vorliegenden Arbeit wird deshalb der Begriff *Vollständigkeit* in diesem Sinne benutzt. In den praktischen Anwendungen ist häufig eine derartig vollständige Modellbildung tatsächlich möglich. Das Problem der Auflösung wird meist gar nicht erörtert, statt dessen wird implizit postuliert, daß eine ausreichende Auflösung gewährleistet ist (obwohl dies nicht immer zu rechtfertigen ist) und das Problem der Genauigkeit wird durch Aufstellung eines stochastischen Modells für die zufälligen Fehler gelöst.

Oftmals ist aber eine Modellbildung, die innerhalb der gegebenen Grenzen als vollständig angesehen werden könnte, gar nicht möglich. Das ist dann der Fall, wenn unbekannte systematische Anteile vorhanden sind und keine Hypothesen im voraus vorliegen, die eine Einbeziehung dieser Einflüsse in die mathematische Beschreibung ermöglichen würden. Dies ist ein sehr schwieriges Problem, welches gar nicht in vollem Umfang gelöst werden kann.

Unter gewissen Umständen ist es jedoch möglich, ausgehend von einem unvollständigen Modell, die fehlenden systematischen Anteile zu identifizieren und eine Modellerweiterung oder sogar -vervollständigung zu erreichen. Zu den in der Literatur dokumentierten Verfahren wird in dieser Arbeit eine weitere Möglichkeit vorgeschlagen, welche die Aussichten zur erfolgreichen Lösung des Problems erweitert und auf einer Ausgleichung nach maximaler Korrelation basiert.

Die Zielsetzung der Ausgleichung nach maximaler Korrelation ist eine Bestimmung der Parameter des unvollständigen Modells derart, daß der Korrelationskoeffizient zwischen den Daten und dem unvollständigen Modell maximiert wird. Dazu ist zunächst eine Aufhebung der Einschränkung, daß der Korrelationskoeffizient nur zwischen stochastischen und nicht zwischen beliebigen Funktionen berechnet werden darf, unumgänglich. Dann kann dieser auch zum Vergleich der geometrischen Formen eingesetzt werden.

Zu den ersten Ansätzen für die Einführung einer Ausgleichung nach maximaler Korrelation kam es während meiner Beteiligung an den Forschungsarbeiten der Geodätischen Fakultät in Zagreb (Kroatien), die unter Leitung des im vorigen Jahr zu früh verstorbenen Professors Krešimir Čolić durchgeführt wurden. Dabei wurde festgestellt, daß in einem Gebiet, in dem Isostasie überwiegend vom Airy-Typ herrschen sollte (junge Faltgebirge), aus vorhandenen Daten für die Undulationen des Geoids und für die Tiefen der Mohorovičić-Diskontinuität niedrige Beträge für den Korrelationskoeffizienten resultierten, s. [Čolić und Petrović 1984]. Da die Qualität der Lagerung des benutzten kleinräumigen astro-geodätischen Geoidmodells unsicher war, ergab sich die Frage, ob eine Änderung dieser Orientierung beträchtlichere Änderungen des Korrelationskoeffizienten hervorrufen würde. Die Antwort war positiv und der maximal erreichte Betrag des Korrelationskoeffizienten konnte als

Maß für die Formenverwandschaft der betrachteten Flächen interpretiert werden, s. [Čolić u.a. 1987], [Čolić u.a. 1988], [Čolić u.a. 1989b], [Petrović u.a. 1989], [Petrović u.a. 1993] und [Čolić u.a. 1994]. Die Hauptzielsetzung dieser Untersuchungen war jedoch nicht die Einführung einer neuen Art von Ausgleichung, sondern die Überprüfung der Lagerung eines lokalen Geoidmodells sowie anschließende geophysikalisch-geodätische Fragestellungen, wie z.B. eine Abschätzung des Dichtesprunges an der Mohorovičić-Diskontinuität, s. [Petrović und Čolić 1994].

Professor Helmut Moritz hat mich daraufhin ermutigt, die Maximierung des Korrelationskoeffizienten allgemeiner zu betrachten und eine Ausgleichung nach maximaler Korrelation auszuarbeiten. Dies wurde in meiner Dissertation [Petrović 1991] durchgeführt, s. auch [Petrović 1993].

Professor Dieter Lelgemann hat mich motiviert, die Ausgleichung nach maximaler Korrelation auch von anderen Standpunkten aus zu betrachten und Möglichkeiten einer Anwendung bei der Untersuchung von Unzulänglichkeiten in der Modellbildung zu überprüfen. Die Ergebnisse erster Versuche dazu wurden in [Petrović 1995], [Petrović 1996] und [Petrović 1997] veröffentlicht. Ein Vorschlag für eine systematische Vorgehensweise wird nun in der vorliegenden Arbeit präsentiert.

Kapitel 1

Motivation

1.1 Unvollständige Modelle

Die Zielsetzung der Geodäsie besteht grundsätzlich darin, Teile der physikalisch-geometrischen Realität zu beschreiben. Dabei handelt es sich vor allem um eine quantitative Beschreibung, die auf Meßergebnissen beruht. Zu diesem Zwecke soll ein mathematisches Modell aufgestellt werden, dessen Parameter zu bestimmen sind. In der Regel ist es nicht möglich, alle Parameter, die die physikalische oder geometrische Realität beschreiben, direkt zu beobachten. Meßbar sind statt dessen andere Größen, die mit den gesuchten Parametern in einer gewissen Beziehung stehen.

Die Betrachtungen in dieser Arbeit werden auf den Fall beschränkt, in dem mindestens theoretisch ein funktionaler Zusammenhang zwischen den Parametern (Vektor x) und den meßbaren Größen (Vektor L) existiert:

$$L = F(x). \quad (1.1)$$

Das funktionale Modell F , welches zunächst nur theoretisch existiert, sei vollständig und habe keine Unzulänglichkeiten. Leider ist die vollständige Gesetzmäßigkeit eines solchen Modells häufig nicht bekannt, sondern nur seine Hauptkomponente $f(u)$ ¹. Dies ist beispielsweise dann der Fall, wenn systematische Einflüsse (s_f) in $f(u)$ nicht einbezogen werden, weil entweder deren Existenz oder deren funktionales Gesetz nicht bekannt ist.

Andererseits enthalten die gemessenen Daten l , neben den meßbaren Größen selbst (Sollwerte), mindestens noch zufällige Fehler (z) und eventuell grobe Fehler (g). Die Gleichung (1.1) geht dementsprechend in den folgenden Ausdruck über:

$$l = \underbrace{\{f(u) + s_f\}}_{\text{Funktionales Modell}} + \underbrace{\{g + z\}}_{\text{Stochastisches Modell}}. \quad (1.2)$$

Das erste Problem besteht darin, diejenigen Daten zu eliminieren, die grobe Fehler enthalten. Im allgemeinen wird das mit den bekannten Methoden versucht, die auf überschüssigen Beobachtungen, robusten Ausgleichungsverfahren und statistischen Tests basieren (s. z.B. [Koch 1980], [Kampmann 1986], [Koch und Yang 1998], [Konak u.a. 1999], [Hekimoglu und Koch 1999]). Tatsächlich aber suchen alle diese Methoden nach Ausreißern und nicht nach groben Fehlern, da sie nicht die Natur, sondern den Betrag der Widersprüche zugrunde legen². Somit entscheidet letztendlich ein stochastisches Modell darüber, welche Beträge als zu groß und inakzeptabel betrachtet werden. Trotzdem können diese Methoden als ausreichend oder gar als erfolgreich betrachtet werden. Die Ausreißer werden dadurch aufgedeckt und beseitigt; ob übriggebliebene „kleine grobe Fehler“ von den zufälligen überhaupt zu unterscheiden sind, ist eher eine linguistische Frage.

¹Die gesuchten (unbekannten) Parameter wurden absichtlich mit u und nicht mit x bezeichnet, weil u möglicherweise nicht alle in x vorhandenen Parameter enthält. Darüberhinaus ist es möglich, daß u sogar einige falsche Parameter beinhaltet kann, die es in dem theoretisch existierenden wirklichen Zusammenhang überhaupt nicht gibt.

²Bei der Definition grober Fehler wird in der Regel die Entstehungsgeschichte, d.h. ihre Natur, zugrunde gelegt (s. z.B. [Gotthardt 1968] oder [Wolf 1968]). Die Entstehungsgeschichte ist aber mathematisch nicht beschreibbar. Deshalb verzichten einige Autoren auf eine derartige Definition grober Fehler und setzen diese von Anfang an mit den Ausreißern gleich. Diese Vorgehensweise ist schon in [Helmert 1872] zu finden und ist bei einigen späteren Autoren erhalten geblieben. Beispielsweise ist in [Reissmann 1962] zu lesen: „*Grobe Fehler* sind solche, die weit über der dem Messungsvorgang überschlägig zugeordneten Genauigkeit liegen“.

Die Annahme, daß das Datenmaterial von Ausreißern gesäubert werden kann, ist also zulässig und für die weitere Bearbeitung der Meßdaten notwendig³. Somit vereinfacht sich die Gleichung (1.2) zu

$$l = \{f(u) + s_f\} + z. \quad (1.3)$$

Die weiteren Betrachtungen in der vorliegenden Arbeit werden auf diesen Fall beschränkt.

Vorausgesetzt, daß darüber hinaus auch keine systematischen Fehler vorhanden sind ($s_f = 0$), gibt es einen akzeptablen Lösungsweg. Die sogenannten „Verbesserungen“ $v = -z$ werden eingeführt, worauf sich die Gleichung (1.3) in

$$l + v = f(u) \quad (1.4)$$

umwandelt. Gegebenenfalls können noch Gewichte p einbezogen werden. Eine Ausgleichung nach der Methode der kleinsten Quadrate, die auf der Bedingung

$$\sum pvv \longrightarrow \min \quad (1.5)$$

basiert, liefert eine Lösung, die oft als „bestmöglich“ betrachtet werden darf⁴.

Leider sind häufig unbekannte systematische Fehler existent ($s_f \neq 0$). Die einfachste Idee (die in der Praxis bewußt oder unbewußt Anwendung findet) besteht darin, die Existenz systematischer Fehler im ersten Schritt zu ignorieren und eine Ausgleichung nach kleinsten Quadraten mit dem unvollständigen Modell gemäß Gleichung (1.4) und (1.5) durchzuführen, was einer neuen Bedeutung der Residuen („Verbesserungen“ v), nämlich $v = -(s_f + z)$, entsprechen würde. Die auf diese Weise bestimmten Residuen können im besten Falle als

$$v = -(v_f + v_z) \quad (1.6)$$

interpretiert werden, wobei es sehr fragwürdig ist, inwieweit v_f und v_z die tatsächlichen s_f und z repräsentieren.

Trotzdem kann nach einer solchen Ausgleichung versucht werden, die Residuen v zu interpretieren, um das Modell zu erweitern.

Leider löst diese Vorgehensweise bekanntlich das Problem nicht in geeigneter Weise, weil sowohl die Schätzwerte für die Modellparameter als auch die Residuen im allgemeinen verzerrt sind (s. z.B. [Helmert 1924], [Wolf 1968], [Koch 1980], [Čolić u.a. 1989a]). Das Ergebnis verwundert wenig, denn eine erfolgreiche Residuenanalyse erfordert nicht minimale Residuen, sondern Residuen, die den fehlenden systematischen Anteil repräsentieren.

1.2 Behandlung systematischer Fehler in der Literatur

In der traditionellen Literatur zur Ausgleichungsrechnung wird der Begriff der systematischen Fehler meist in einem der ersten Abschnitte erwähnt oder definiert und knapp diskutiert. So ist beispielsweise in [Gotthardt 1968] folgendes zu finden:

„Systematische Fehler ergeben sich, wenn physikalische Gesetzmäßigkeiten außer acht gelassen werden. . . Weil diese Fehler dadurch entstehen, daß man die Wirklichkeit durch ein allzu vereinfachtes mathematisches Modell ersetzt, den mit der Temperatur veränderlichen Maßstab durch einen solchen konstanter Länge, den veränderlichen Luftdruck durch einen gleichbleibenden, spricht man auch von Modellfehlern. Sie lassen sich durch das Anbringen entsprechender Korrekturen beseitigen.“

Eine derartige Definition der systematischen Fehler als Modellfehler kann ohne Zweifel akzeptiert werden. Fragwürdig ist aber die Annahme, daß die Benutzung eines unvollständigen Modells allein auf einer *bewußten Entscheidung* basiert, so daß entsprechende Korrekturen, falls gewünscht, immer angebracht werden könnten. In diesem Falle würde sich das gesamte Problem auf die bloße Berücksichtigung zufälliger Fehler beschränken.

³ „Große Fehler können nicht der Gegenstand einer Ausgleichung sein, sondern müssen vorher beseitigt werden“, s. [Reissmann 1962]. Ähnliche Aussagen sind bei anderen Autoren zu finden, wobei der Begriff „Ausgleichung“ meistens im engeren Sinne einer Ausgleichung nach kleinsten Quadraten verwendet wird.

⁴ Diese Behauptung kann statistisch begründet werden, die Begründung ist in jedem Lehrbuch zur Ausgleichungsrechnung zu finden. Es ist jedoch zu beachten, daß die Behauptung nur im Kontext der beschriebenen Einschränkungen sinnvoll ist.

Die oben erwähnte Annahme wird in einem überwiegenden Teil der Literatur implizit postuliert und nur die Theorie der Ausgleichung allein im Vorhandensein zufälliger Fehler detailliert ausgearbeitet⁵. Die Frage der Identifizierung unbekannter systematischer Fehler wird selten erörtert.

Scheinbar allgemeiner ist der Vorschlag in [Wolf 1968]: „Schließlich besteht noch die Möglichkeit, . . . , das Gewicht einzelner Beobachtungen wegen eventueller systematischer Fehler herabzusetzen, so daß im ganzen ihr Einfluß auf das Ausgleichungsergebnis entsprechend zurückgedrängt wird.“ Die Durchführung dieser Strategie ist wiederum von den *a priori* Kenntnissen darüber abhängig, welche Beobachtungen wie stark von systematischen Fehlern betroffen sind. Weiterhin kann eine solche Strategie nur dann zum Erfolg führen, wenn die systematischen Fehler bloß wenige einzelne Beobachtungen wesentlich verfälschen⁶. Bei einem andersartigen Defizit des Modells ist das Herabsetzen der Gewichte aller Beobachtungen offensichtlich nicht zu empfehlen. Dementsprechend kann ein Aufdecken der im voraus nicht bekannten systematischen Einflüsse auf diesem Wege nicht erwartet werden.

Ein vergleichbarer Vorschlag in der Literatur rät, *die systematischen Anteile oder Trendanteile zu schätzen* (s. z.B. [Koch 1980]). Dazu muß aber offensichtlich ein Modell für diese Anteile im voraus entweder vorhanden sein oder postuliert werden.

Ähnlich wird in [Chatfield 1982] vorgegangen, wobei die Erfahrung zur Hilfe genommen wird, um zu entscheiden, bei welchen Anwendungen welche Arten der mathematischen Modellierung von systematischen Anteilen in Frage kämen. Auch in diesem Falle handelt es sich darum, ein Modell für systematische Einflüsse zu postulieren.

Aufgrund des Titels⁷ der Veröffentlichung [Yang u.a. 1999] wäre zu erwarten, daß im Artikel eine Methode vorgestellt wird, welche mittels eines robusten Ausgleichungsverfahrens die Form (funktionaler Zusammenhang) der systematischen Fehler identifizieren kann. Es stellt sich aber heraus, daß auch in diesem Falle ein Modell für die systematischen Anteile vorgegeben wurde (eine Gerade im ersten und Polynome vom Grade m im zweiten Schritt des vorgeschlagenen Verfahrens), während die robusten Schätzer den Einfluß der Ausreißer minimieren sollten.

Aufgrund dieser kleinen Auswahl einiger charakteristischen Stellen aus der Literatur läßt sich feststellen, daß das Problem der systematischen Fehler auch ein terminologisches Problem ist. Häufig werden gewisse Einflüsse (z.B. solche, die aus den Eigenschaften des Meßsystems folgen) *systematische Fehler* genannt⁸. Später wird neben dem „Hauptmodell“ auch ein Modell für diese Anteile postuliert (welches selbstverständlich perfekt, akzeptabel oder fragwürdig sein kann). Auf diese Weise verliert der Begriff *systematische Fehler* seine Bedeutung als *Modellfehler* – die sogenannten systematischen Fehler sind von Anfang an ein Teil des Modells, mit dem eine Ausgleichung durchgeführt wird. Vorausgesetzt, daß die gesamte Modellbildung korrekt und vollständig sei, sollte es nicht verwundern, wenn dadurch alle Modellparameter zufriedenstellend bestimmt werden, inklusive jener, die sich auf angebliche systematische Fehler beziehen.

Die beschriebene Umbenennung eines Begriffes kann aber nicht die Frage beantworten, wie ein unvollständiges Modell, dessen fehlende Anteile völlig unbekannt sind, vervollständigt werden kann. *Diese Frage ist aber das eigentliche Problem.* Deshalb wird in dieser Arbeit der Begriff „*systematische Fehler*“ nur im Sinne von „*Modellfehler*“ benutzt⁹. Sobald systematische Fehler identifiziert worden

⁵Einige Autoren betonen die *Notwendigkeit* der Beseitigung systematischer Fehler, sind aber etwas vorsichtiger bezüglich der *Möglichkeit*, diese Beseitigung durchzuführen. Beispielsweise ist in [Reissmann 1962] zu finden: „Die Ausschaltung der systematischen Fehler ist eine oft viel Überlegung erfordern Notwendigkeit, denn die in die Ausgleichung einzuführenden Messungsergebnisse müssen von ihnen befreit sein.“

⁶weil z.B. nur die Modellierung eines Teils des Meßsystems unvollständig ist

⁷„Robust estimation of systematic errors of satellite laser range“

⁸Sollte beispielsweise der Titel „On the elimination of biases in processing differential GPS observations“ des hervorragenden Artikels [Vaníček u.a. 1985] durch „Zur Elimination systematischer Fehler bei der Bearbeitung von differentiellen GPS Beobachtungen“ übersetzt werden, dann sollte nicht vergessen werden, daß der Begriff „systematische Fehler“ in einem völlig anderen Sinne als in der vorliegenden Arbeit benutzt wird. In dem zitierten Artikel handelt es sich um bekannte „Biases“, die entweder modelliert oder eliminiert („Differences“ und „double differences“ in der Bearbeitung von GPS-Daten beruhen grundsätzlich auf der gleichen Idee wie die Differenzbildung bei Winkelbeobachtungen in zwei Lagen in der traditionellen Vermessung.) werden und nicht um die Identifizierung von Modellunvollständigkeiten unbekannter Herkunft und Auswirkungen.

⁹Ein zusätzliches Problem besteht darin, daß es manchmal schwierig ist, eine wohldefinierte Grenze zwischen systematischen Fehlern und Ausreißern zu ziehen. Ein Beispiel dazu liefert die Überwachung von Punktbewegungen. Die Natur dieser Bewegungen kann sehr verschiedenartig sein. Manchmal gibt es offensichtliche Gemeinsamkeiten in den Bewegungen benachbarter Punkte, so daß eine systematische Gesetzmäßigkeit, die diese Bewegungen teilweise oder vollständig beschreibt bzw. erklärt, zu erwarten ist. Dies wäre der Fall z.B. bei der Überwachung tektonischer Plattenbewegungen, bei denen alle Punkte einer Gruppe grundsätzlich die gleiche Bewegung aufweisen („homogene

sind, können sie modelliert werden und dadurch Bestandteil des Modells werden.

Vom gerade beleuchteten Standpunkt aus betrachtet, ist die Identifizierung der systematischen Fehler von zwei Fragestellungen geprägt:

1. Nachweis des Vorhandenseins systematischer Einflüsse in den Daten und
2. Aufstellen eines zufriedenstellenden (mathematischen) Modells zur Beschreibung der bis dahin fehlenden Modellanteile, oder zumindest eine realistische Parameterschätzung für das vorhandene unvollständige Modell.

Diese zwei Probleme sollen klar voneinander getrennt werden, nicht nur deshalb, weil sie von wesentlich unterschiedlichem Schwierigkeitsgrad sind.

Als Grundlage für den Nachweis der systematischen Fehler können Residuen aus einer Ausgleichung mit einem unvollständigen Modell dienen. Diese kann z.B. eine Ausgleichung nach der Methode der kleinsten Quadrate sein. Obwohl die daraus resultierenden Residuen für die fehlenden Modellanteile im allgemeinen nicht repräsentativ sind, weisen sie beim Vorhandensein systematischer Fehler in der Regel zu hohe Beträge und eine nicht zufällige Struktur auf. Einige Vorschläge für eine entsprechende Residuenanalyse finden sich schon in [Helmert 1872]. Weitere Möglichkeiten können beispielsweise in [Chatfield 1982] gefunden werden. Alle derartigen Verfahren basieren auf statistischen Untersuchungen der Verteilung der Residuen. Eine kurze Darstellung findet sich im nächsten Abschnitt.

Der Nachweis der Existenz systematischer Fehler wird schwieriger, wenn die Beträge dieser Einflüsse, verglichen mit den Beträgen der in Daten enthaltenen zufälligen Fehler, relativ klein werden. Unter Einschränkung auf sehr spezielle Problemstellungen (nämlich solche, bei denen gewisse konkrete Eigenschaften der vorhandenen systematischen Einflüsse als bekannt vorausgesetzt werden können) ist der Nachweis systematischer Anteile trotzdem manchmal möglich. Beispielsweise wird in [Fritzensmeier u.a. 1986] eine Methode vorgestellt, die mit statistischen Mitteln systematische Fehler in wiederholt gemessenen geodätischen Netzen erfolgreich aufdecken kann. Die beste Erklärung dafür geben die Autoren selbst: „In diesem Beitrag soll für eine spezielle Konstellation von Meßdaten und von systematischen Effekten ein Verfahren vorgestellt werden, mit dem der Nachweis auch kleinerer Systematiken im Datenmaterial möglich erscheint: Liegen wiederholt ausgemessene geodätische Netze vor und können in etwa gleiche systematische Verfälschungen in allen Epochen angenommen werden, so ist es durch Anwendung einfacher Teststatistiken auf die Verbesserungen der Einzelpochenausgleichungen möglich, diese systematischen Effekte nachzuweisen.“ Die gleichen Autoren warnen aber: „ . . . aufgrund der heterogenen Struktur und der unterschiedlichen Ursachen für systematische Verfälschungen im Datenmaterial kann jedoch bisher keine Globallösung zum Aufdecken oder gar zum Ausschalten systematischer Fehler angegeben werden.“

Trotz allem scheint die Behauptung gerechtfertigt zu sein, daß die bestehende Literatur ausreichende Verfahren zum Nachweis der *Existenz* nicht allzu kleiner systematischer Fehler anbietet. Sind diese Einflüsse so klein, daß ein zuverlässiger Existenzbeweis nicht mehr möglich ist, stellt sich die Frage, ob vom Datenmaterial mehr verlangt werden darf, als aufgrund des Informationsgehalts wirklich möglich ist.

Das zweite Problem ist wesentlich komplizierter und offensichtlich noch nicht zufriedenstellend gelöst: die Aufstellung eines Modells zur Beschreibung der fehlenden Modellanteile, oder eine akzeptable Parameterschätzung im gegebenen unvollständigen Modell.

Nachdem die Existenz der systematischen Fehler im konkreten Beispiel nachgewiesen wurde (oder zumindest der Verdacht erhärtet wurde), wird versucht ein passendes Modell zu „erraten“. Es wird beispielsweise versucht, aufgrund der verfälschten Residuen aus einer Ausgleichung auf mögliche Ursachen zu schließen und den möglichen Zusammenhang mathematisch zu beschreiben. Eine derartige Vorgehensweise muß nicht unbedingt scheitern. Die aus einer mit einem unvollständigen Modell durchgeföhrten Ausgleichung resultierenden Residuen können beträchtlich, oder auch nur geringfügig verfälscht sein. Das eigentliche Problem besteht darin, daß unbekannt ist, wie stark die Verfälschung tatsächlich ist. Aus diesem Grund kann eine derartige Residuenanalyse sowohl in einer korrekten Identifizierung der systematischen Effekte, als auch in grob falschen Vorstellungen enden.

Deformationen“). Andererseits können in der Literatur zur Deformationsanalyse Beispiele gefunden werden, in denen jeder bewegliche Punkt eine völlig eigenständige Verschiebung zeigt, s. z.B. [HSBw 1979] oder [Reinking 1994] („inhomogene Deformationen“). Sollte in einem solchen Falle nicht sinnvollerweise von Ausreißern statt von systematischen Einflüssen gesprochen werden?

Ein anderer Weg besteht darin, eine rein mathematische Erweiterung des Modells vorzunehmen¹⁰, mit der Zielsetzung, die Residuen durch die Modellerweiterung so zu verringern, daß eine weitere Interpretation nicht sinnvoll erscheint. Es ist nicht zu bestreiten, daß ein solches Gesamtmodell die benutzten Daten reproduzieren kann. Es ist aber nicht zu erwarten, daß eine polynomiale Reihe bzw. ein anderes mathematisches und künstliches Gebilde der physikalischen Natur des Problems entsprechen kann. Daher kommt es häufig zu einer „Überparametrisierung“ und der „die Systematik beschreibende“ Teil eines solchen Modells ist physikalisch nicht interpretierbar¹¹. Die Konsequenz ist leider, daß dadurch auch eine physikalische (oder eine andere, der Natur des Problems entsprechende) Interpretation des Hauptteils zweifelhaft wird, weil auf diese Weise seine Parameter wesentlich verfälscht sein können.

Zusammenfassend läßt sich sagen, daß das Problem der Identifizierung der unbekannten systematischen Anteile noch nicht vollständig gelöst wurde, obwohl sich zahlreiche Autoren damit beschäftigt haben. Die Ursache ist die Tatsache, daß eine eindeutige Trennung zwischen deterministischen und zufälligen Anteilen nur dann relativ einfach ist, wenn von Anfang an ein vollständiges Modell für die deterministischen Anteile vorliegt.

1.3 Zum Nachweis der Existenz systematischer Einflüsse in den Daten

Die Voraussetzung für die Durchführung einer Ausgleichung nach kleinsten Quadraten ist, daß das Datenmaterial nur zufällige Fehler enthält. Ist diese Voraussetzung erfüllt, sind als Ergebnis zufällig¹² verteilte Residuen zu erwarten. Diese Tatsache ist schon lange bekannt, s. [Helmert 1872].

Sollte die Voraussetzung nicht erfüllt sein und das Datenmaterial systematische Einflüsse enthalten, dann können aus einer Ausgleichung nach kleinsten Quadraten Residuen erwartet werden, die nicht einer Vorstellung von zufällig verteilten Größen entsprechen. Obwohl die resultierende Verteilung eine realistische Rekonstruktion der systematischen Einflüsse nicht garantiert, bietet sie zumindest einen Nachweis ihrer Existenz, d.h. der Tatsache, daß das zugrundegelegte Modell nicht vollständig war.

Die Grundsätze einer derartigen Untersuchung der Verteilung von Residuen wurden schon in [Helmert 1872] aufgestellt und in der späteren, erweiterten Auflage [Helmert 1924]¹³ weiter ausgearbeitet. Diese Grundsätze werden nun in den nächsten Absätzen kurz geschildert.

Zweck der Untersuchung in [Helmert 1924] ist „... zu entscheiden, ob die Rechnungsergebnisse wahrscheinlichste Ergebnisse sind oder nur größte Gewichte haben, oder ob sie den Beobachtungen nur möglichst angepaßt sind“¹⁴

Als erstes wird gezeigt, daß das Vorhandensein systematischer Einflüsse die Residuen ändern *kann*, aber nicht unbedingt ändern *muß*¹⁵. Ausgehend davon, daß die Residuen von jenen unterschiedlich sein können, die sich in Abwesenheit systematischer Einflüsse ergeben würden, soll die Verteilung der Residuen untersucht werden.

Zuerst wird überprüft, ob die Fehlerverteilung ein „gerades“ Gesetz befolgt. Dafür sollen „die gemessenen Größen gleichartig oder gleichsinnig eingeführt werden“. Der Erwartungswert für die Vorzeichensumme bei rein zufälligen Fehlern ist Null mit einem mittleren Fehler von $\pm\sqrt{n}$. „Ist die

¹⁰Zu dieser Gruppe gehört das schon erwähnte Beispiel ([Yang u.a. 1999]), in dem die Systematik durch Polynome modelliert wird.

¹¹Dies geschieht beispielsweise bei jedem Versuch, eine aus beliebigen Frequenzen zusammengesetzte physikalische Erscheinung mit Hilfe einer Fourier-Reihe zu modellieren, s. [Mautz 2001]. Das gleiche gilt für die Anwendung der Kugelfunktionsmodelle zur Darstellung physikalischer Felder.

¹²Vielleicht wäre der Begriff *pseudozufällig* besser geeignet.

¹³In diesem Werk wurde ein ganzes Kapitel dieser Untersuchung gewidmet. Die systematischen Einflüsse werden dort meistens *regelmäßige Fehlerursachen* genannt.

¹⁴Eine schöne Warnung an jene, die noch heutzutage daran glauben, daß eine Ausgleichung nach kleinsten Quadraten unabhängig von den Umständen immer „die plausibelste“ Lösung liefert.

¹⁵Obwohl ich dem zustimme, halte ich die im Buch von Helmert aufgeführten Beispiele nicht für bestens geeignet, weil die in die Gruppe der *improperly posed problems* eingeordnet werden könnten. Wenn die Daten keinen ausreichenden Informationsgehalt haben, um daraus die Parameter des „vollständigen“ Modells bestimmen zu können, dann ist es auch nicht verwunderlich, daß dies nicht gelingen kann. Viel wichtiger für die Anwendung ist die Situation, in der aus den Daten die Parameter eines vollständigen Modells ermittelt werden könnten, vorausgesetzt, daß eine vollständige Modellbildung vorhanden ist.

Vorzeichensumme s , absolut genommen, größer als $\pm\sqrt{n}$, so hat man daher ernste Veranlassung, systematische Fehlereinflüsse anzunehmen. ... Da systematische Fehler die eben behandelte Vorzeichenverteilung nicht immer ungünstig beeinflussen, so ist es weiter nötig, die *Vorzeichenwechsel* der Fehlerreihe zu untersuchen.“ Dazu sollen die Residuen (im mehrdimensionalen Falle) z.B. nach einer geeigneten Variablen (wie z.B. Zeit) geordnet werden¹⁶.

Um die Vorzeichenverteilung zu überprüfen, wird in [Helmert 1924] die Differenz der Vorzeichenfolgen f und der Vorzeichenwechsel w ,

$$f - w = V_1 V_2 + V_2 V_3 + \dots + V_{n-1} V_n \quad (1.7)$$

eingeführt, wobei $V_i = \pm 1$ einzelne Vorzeichen sind. Bei rein zufälligen Fehlern ergibt sich Null als Erwartungswert für die Differenz $f - w$ mit einem mittleren Fehler von $\pm\sqrt{n-1}$. „Ist diese Differenz, absolut genommen, größer als $\pm\sqrt{n-1}$, so ist ein systematischer Einfluß der Variablen, nach der die Beobachtungen geordnet sind, zu vermuten“.

Um die Symmetrie der Verteilung noch näher zu überprüfen wird weiterhin empfohlen, die Summe der Quadrate der positiven Residuen mit der gleichartigen Summe der negativen zu vergleichen. Als zusätzliche Möglichkeiten sind die Anwendung des Abbéschen Kriteriums¹⁷ und Vergleiche der aus den Residuengruppen berechneten Mittelwerte untereinander und mit dem aus allen Residuen berechneten Mittelwert aufgelistet.

Anschließend wird im Abschnitt „Nähere Prüfung des Verteilungsgesetzes der Fehler“ vorgeschlagen, einen Vergleich mit bekannten Fehlergesetzen für zufällige Verteilungen durchzuführen. Dieser Vergleich wird in [Helmert 1924] anhand graphischer Darstellungen visuell durchgeführt.

In der zitierten Arbeit wird festgestellt, daß bei einer Verteilung der Residuen, die kein zufälliges Fehlergesetz folgt, die Verteilung der wahren Fehler *in der Regel* erst recht diese Eigenschaft nicht besitzen kann. Der Autor warnt aber, daß das Umgekehrte leider nicht gilt und daß eine Benutzung der Residuen anstelle wahrer Fehler „nicht unbedenklich“ ist.

Zusammenfassend gesehen kann eine anomale Verteilung der Residuen als Nachweis der Existenz systematischer Einflüsse betrachtet werden; sollten die Residuen zufällig verteilt sein, ist das noch kein Beweis dafür, daß das benutzte Modell vollständig ist. In [Helmert 1924] wurde an einem Beispiel gezeigt, daß in einem solchen Fall, bei dem der Vergleich der Verteilung der Residuen mit einer bekannten zufälligen Verteilung keine Antwort geben kann, die davor dargestellten einfacheren Kriterien zu einer Entscheidung führen können. Deshalb sollen sie nicht bloß als primitive Notlösungen betrachtet werden, sondern als vollwertige Untersuchungsmöglichkeiten.

Neuere Literatur hat wenig neue Ideen zum Nachweis der Existenz systematischer Einflüsse gebracht. Die oben knapp dargestellten Grundsätze sind im Prinzip erhalten geblieben. Die Techniken der mathematischen Statistik wurden aber inzwischen sehr entwickelt und haben in allen Teilen der Geodäsie, insbesondere in der Ausgleichungsrechnung¹⁸, eine breite Anwendung gefunden, s. beispielsweise [Koch 1980] oder [Gründig 1999]. Die heutigen Darstellungen der Überprüfung der Verteilung von Residuen nehmen damit eine neue Form an.

Die Hypothesen über Verteilungen werden heutzutage mit statistischen Testverfahren überprüft (s. z.B. [Wolf 1968], [Bjerhammar 1973], [Koch 1980]), wodurch die Willkür bei den Entscheidungen

¹⁶Neben diesem Vorschlag aus [Helmert 1924] sind auch andere Möglichkeiten zur Ordnung von Residuen in einem ein- oder mehrdimensionalen Raum denkbar.

¹⁷Vergleiche auch die Darstellung dieses Kriteriums in [Linnik 1961].

¹⁸Häufig wird die Ausgleichungsrechnung sogar als Teil der Statistik betrachtet. Neben offensichtlicher Vorteile birgt diese Herangehensweise auch erhebliche Gefahren in sich. *Messungen* sind von ihrer Natur aus *keine stochastischen Größen*. Meßfehler können beispielsweise anscheinend näherungsweise normalverteilt sein. Eine wirkliche Normalverteilung ist aber *ausgeschlossen*. Es gibt nämlich keine Ablesungen mit einer unbeschränkten Auflösung im Kleinen. Weiterhin ist ein wahrer Fehler von 100 km bei der Messung einer 10 m langen Strecke statistisch sehr unwahrscheinlich, aber trotzdem möglich, bei wirklichen Messungen kann aber ein solcher zufälliger Fehler überhaupt nicht vorkommen, usw. Messungen können nur näherungsweise und nur bedingt als stochastische Größen betrachtet werden. Dies soll deswegen nicht automatisch durch Gleichsetzung der Ausgleichungsrechnung mit (einem Teil) der Statistik realisiert werden, da dadurch das Bewußtsein verlorengeht, daß sich der Charakter der Messungen in stochastischen Eigenschaften nicht völlig erschöpft. Sonst kann es passieren, daß versucht wird, mit Hilfe der Statistik auch die Erscheinungen zu erklären, die sich auf diese Weise nicht erklären lassen. Deshalb ist es sinnvoll, die Ausgleichungsrechnung als Anpassung mathematischer Modelle an empirische Daten (z.B. durch Minimierung einer Norm) zu betrachten; unabhängig davon aber stochastische Modelle auszuarbeiten und „Äquivalenzsätze“ aufzustellen, die die Ausgleichungsrechnung statistisch rechtfertigen, s. [Grafarend und Schaffrin 1993]. Dann ist der Einsicht nicht der Weg versperrt, daß die statistischen Rechtfertigungen und Interpretationen nur insoweit gelten, inwieweit die Stochastik die gesamte Natur der Messungen vertritt.

scheinbar verschwindet und der Eindruck von Objektivität erweckt wird. Tatsächlich tritt die Willkür durch die Wahl des Signifikanzniveaus¹⁹ in den Hintergrund. Neben einer angenehmeren und objektiveren Durchführung der Verteilungsuntersuchungen kann die Anwendung der Statistik also auch eine unkritische Betrachtungsweise mit sich bringen.

Ein weiteres Problem bei einem Vergleich mit einer theoretischen Verteilung besteht darin, daß die Parameter dieser Verteilung benötigt werden. Zu diesem Thema ist in [Koch 1980] folgendes zu finden: „... Häufig wird mit diesen Tests die Annahme der Normalverteilung für die Beobachtungen bei Parameterschätzungen geprüft. Diese Vorgehensweise ist dann problematisch, wenn die Schätzwerte der Parameter dazu benutzt werden, um die theoretische Normalverteilung zu berechnen. Sind die Schätzwerte durch die im Kapitel 329 behandelten Modellabweichungen²⁰ verfälscht, kann sich schon aus diesem Grund eine Abweichung von der Normalverteilung ergeben.“

Bei der Analyse der Residuen einer Zeitreihe wird heutzutage zusätzlich Autokorrelation benutzt, um zu überprüfen, ob benachbarte Residuen als untereinander unabhängig betrachtet werden dürfen, s. beispielsweise [Wiener 1970] oder [Chatfield 1982]. Aus dieser Vorgehensweise resultieren häufig nützliche Informationen.

Stellt das untersuchte Problem keine Zeitreihe dar, dann ist ein Einsatz von Autokorrelationsfunktionen trotzdem möglich. Dazu sollen die Residuen nach einer geeigneten Variablen geordnet werden²¹. Diese Vorgehensweise wurde in [Vaníček und Craymer 1983] beispielhaft durchgeführt. Als Grunddatenmaterial wurden Widersprüche zwischen den Nivellementsmeßungen in beiden Richtungen für jedes Segment einer Linie oder eines Netzes verwendet²². Diese Widersprüche wurden nach verschiedenen Parametern geordnet, wie z.B. nach den dazugehörigen Höhen, Höhenunterschieden, der Länge der Segmente oder der Anzahl der Instrumentenstandpunkte. Die Anwendung der Autokorrelation auf die konstruierten Reihen zeigte in einigen Fällen signifikante und interpretierbare Trends. In der Schlußfolgerung betonen die Autoren, daß in den Nivellementsdaten eine Menge an Informationen enthalten ist, welche nur darauf wartet, systematisch benutzt zu werden. Weiterhin: „Clearly, the more kinds of auxiliary data we can get, the more complete picture on the character of the levelling we can assemble²³.“ Diese Warnungen erinnern an eine häufig vergessene Tatsache: Normalerweise werden die originären Meßwerte zuerst komprimiert und nur die komprimierten Daten für die weitere Bearbeitung verwendet. Neben offensichtlichen Vorteilen hat diese Vorgehensweise auch Nachteile: Ein Teil des vorhandenen Informationsgehalts in den Daten wird vernachlässigt²⁴. *Eine vollständigere Nutzung des Datenmaterials kann auch eine vollständigere Modellbildung ermöglichen.*

Trotz aller erwähnten Schwierigkeiten, die bei dem Nachweis der Existenz systematischer Einflüsse auftreten, scheint dieser Teil des Problems wesentlich einfacher zu sein, als die Identifizierung der Formen und Ursachen unbekannter systematischer Einflüsse.

Die Anhaltspunkte in der Literatur für einen erfolgreichen Existenznachweis können für die meisten Anwendungsbeispiele als hinreichend betrachtet werden.

1.4 Zur Ausgleichung unvollständiger Modelle

Eine Ausgleichung nach kleinsten Quadraten setzt ein vollständiges Modell voraus (vgl. Abschnitt 1.1). Ist das Modell zu Beginn nicht vollständig, kann zunächst mit dem unvollständigen Modell eine geeignete Ausgleichung durchgeführt werden und zwar mit dem Ziel, systematische Fehler, d.h. fehlende Modellanteile zu identifizieren und möglichst die Parameter des unvollständigen Modells abzuschätzen. Diese Identifizierung kann auf einer Residuenanalyse basieren. Bekanntlich ist eine

¹⁹ „Über dieses sog. ‚Produzenten-Risiko‘ α muß frei (= willkürlich) entschieden werden.“ „... die Willkür, die in der Wahl des Zahlenwertes für die statistische Sicherheit S liegt, so daß man u.U. ganz unterschiedliche Testergebnisse erhält, je nachdem, welcher S -wert gewählt wurde, ...“ ([Wolf 1968])

²⁰ Die im Kapitel 329 in [Koch 1980] behandelten Modellabweichungen entsprechen einer unvollständigen Modellbildung.

²¹ Das ist die gleiche Idee, die in [Helmert 1924] zur Untersuchung der Vorzeichenverteilung der Residuen vorgeschlagen wurde, vgl. S. 8.

²² Insgesamt wurden drei derartige Datensätze analysiert: zwei Nivellementslinien aus dem Schweizerischen Netz erster Ordnung und ein Netz zweiter Ordnung aus Norddeutschland.

²³ Offensichtlich, um so mehr Arten von zusätzlichen Daten wir bekommen können, desto vollständiger ist das Bild des Charakters des Nivellements, welches wir daraus zusammensetzen können.

²⁴ Am Beispiel eines Nivellements sind es neben den Differenzen aus den Messungen in beiden Richtungen auch die Anzahl der Standpunkte in einzelnen Segmenten, die Länge der Segmente usw.

Ausgleichung nach kleinsten Quadraten nicht am besten geeignet²⁵, um diese Residuen zu erzeugen; es sollen daher auch andere Wege erörtert werden.

Im Kapitel 3 wird eine andere Möglichkeit für eine Residuenanalyse vorgeschlagen: die Durchführung einer *Ausgleichung nach maximaler Korrelation (MCA)*²⁶ mit dem ursprünglichen unvollständigen Modell. Die damit verbundene geometrische Vorstellung ist die Identifizierung der Formen dieses Modells in den Daten. Sollte diese Identifizierung erfolgreich sein, ist auch eine akzeptable Schätzung der Parameter des vorhandenen unvollständigen Modells denkbar.

Die Bedingung der Korrelationsmaximierung ist in einem gewissen Sinne (s. Abschnitt 2.3.3 und 2.3.6) schwächer als die Bedingung der Minimierung der Quadratsumme der „Verbesserungen“. Die Folge ist, daß auch in den Fällen²⁷, in denen eine Ausgleichung nach kleinsten Quadraten eine eindeutige Lösung liefert, eine Ausgleichung nach maximaler Korrelation meistens in einer ganzen Klasse von Lösungen resultiert. Diese induziert wiederum eine Klasse der Residuensätze (s. Abschnitt 3.1).

Die bei der Suche nach der vollständigen Lösung unerwünschte Mehrdeutigkeit kann jedoch für die Residuenanalyse von großem Vorteil sein: In der Klasse aller Residuensätze befinden sich solche, die für die Form systematischer Fehler sowohl mehr als auch weniger repräsentativ sind.

Der letzte Schritt wäre, einen der Residuensätze auszuwählen, welcher im Hinblick auf die systematischen Einflüsse am besten interpretierbar ist. Dabei stellt sich die nicht triviale Frage der mathematischen Beschreibung der systematischen Fehler. Die auftretenden Schwierigkeiten sind von einer ähnlichen Art, wie bei der mathematischen Beschreibung grober Fehler, wobei sich alles noch etwas komplizierter gestaltet. Siehe dazu Abschnitt 3.2; dort werden auch einige Möglichkeiten zur Auflösung dieses Problems geschildert. Es scheint aber, daß eine allgemeine mathematische Beschreibung, die alles umfaßt, was in der Literatur „Systematik“ genannt wird, nicht möglich ist. Aus diesem Grund wird in Abschnitt 3.2 auch ein Versuch der Begriffsklärung und -festlegung vorgenommen, zusammen mit einer Formalisierung der Beschreibung systematischer Einflüsse, die in vielen für die Praxis relevanten Fällen als sinnvoll erscheint.

Nach der Auswahl eines repräsentativen Residuensatzes werden von der dazugehörigen Lösung akzeptable Parameterschätzungen für das ursprüngliche unvollständige Modell erwartet. Trotzdem soll versucht werden, den gefundenen systematischen Einfluß zu modellieren. Wie dieses *Postprocessing* erfolgen soll, läßt sich nicht auf eine allgemeine und rein formelle mathematische Weise beschreiben. Die Modellierung ist von der Natur des Problems abhängig. An dieser Stelle müssen die möglichen (z.B. physikalischen) Ursachen in Erwägung gezogen werden.

Nach der Vervollständigung des Modells kann eine Ausgleichung nach der Methode der kleinsten Quadrate durchgeführt werden. Sollten daraus ausreichend kleine und zufällig verteilte Residuen resultieren, dann ist eine weitere Analyse nicht zu rechtfertigen und das Ergebnis soll als endgültig betrachtet werden. Bei noch immer zu großen und/oder nicht zufällig verteilten Residuenbeträgen bietet sich eine Wiederholung des in vorgehenden Abschnitten vorgeschlagenen Vorgangs an.

²⁵ Andererseits liefern die aus einer Ausgleichung nach kleinsten Quadraten resultierenden Residuen nicht unbedingt ein völlig verzerrtes Bild. Sie sind manchmal für die nichtmodellierten systematischen Anteile teilweise repräsentativ. Leider bleibt immer unbekannt, wie groß die Verzerrungen tatsächlich sind.

²⁶MCA = Maximum Correlation Adjustment

²⁷die am häufigsten vorkommen

Kapitel 2

Kurze Darstellung der Ausgleichung nach maximaler Korrelation

2.1 Allgemeine Bemerkungen

Die Grundidee der hier beschriebenen Ausgleichung nach maximaler Korrelation besteht darin, ausgehend vom einzig vorhandenen, nämlich dem unvollständigen Modell (1.4), die Parameter weder unter der Bedingung (1.5), welche der Minimierung der L_2 -Norm entspricht, noch unter einer Bedingung, die auf einer anderen Norm basiert (L_1, L_∞, \dots) zu bestimmen, sondern unter der Forderung, daß der Korrelationskoeffizient zwischen dem unvollständigen Modell und den Daten maximal wird:

$$r^2(f(u), l) \longrightarrow \max. \quad (2.1)$$

Dadurch wird nicht eine Funktion der Residuen (z.B. eine Norm) minimiert und somit können diese nicht als „in einem gewissen Sinne minimal“ angesehen werden. Wie bekannt, führt eine Minimierung von Residuen in den Fällen, in denen die Daten systematische Einflüsse enthalten, einerseits zu nicht interpretierbaren Residuen und andererseits zu nicht realistischen Lösungen für die Modellparameter.

Von einer Ausgleichung nach maximaler Korrelation wird erwartet, daß sie in den Daten nur die Formen des unvollständigen Modells $f(u)$ identifiziert und die Residuen nicht zum Minimum zwingt. Somit sollte beides, sowohl besser interpretierbare Residuen als auch realistischere Modellparameter gefunden werden.

Aus den oben genannten Gründen sehe ich keine Möglichkeit, die Ausgleichung nach maximaler Korrelation in dem Sinne und der Form, wie diese in der vorliegenden Arbeit eingeführt und betrachtet wird, statistisch zu begründen – meine Begründung ist geometrischer Art: *der Vergleich von Formen*.

An dieser Stelle sei erwähnt, daß in der Literatur auch geometrische Begründungen der Ausgleichung nach kleinsten Quadraten zu finden sind¹. Geometrische Deutungen dieser Methode kommen noch häufiger vor, entweder als diesem Aspekt gewidmete Artikel (s. z.B. [Moritz 1966]) oder als Parallel- oder Zusatzbetrachtungen innerhalb einer grundsätzlich von Statistik geprägten Betrachtung (s. z.B. [Kolmogorov 1946], [Bjerhammar 1973] oder [Grafarend und Schaffrin 1993]). Weiterhin ist in der Literatur eine ganze Reihe von Abhandlungen zu finden, in denen die Ausgleichung nach kleinsten Quadraten mit den Prinzipien der technischen Mechanik oder der Physik in Bezug gesetzt wird. Beispielsweise werden in [Gauß 1829] und in [Jäger 1988] Analogien aufgestellt, in [Wellisch 1904] und [Friedrich 1943]² wird die Methode der kleinsten Quadrate aus der Mechanik hergeleitet und in [Linkwitz 1961] und [Milev 2000] werden mechanische Überlegungen und Ausgleichsrechnung gegenübergestellt. Schließlich gab es auch Versuche, robuste Schätzverfahren physikalisch zu begründen, s. z.B. [Saleh 2000].

¹Letztendlich handelt es sich auch dann um Minimierung einer Vektornorm, wenn diese Minimierung statistisch gerechtfertigt wird.

²„Durch die nachstehenden Darlegungen wird bewiesen: Man kann die M.d.kl.Qu. ohne wahrscheinlichkeitstheoretische Erwägungen völlig allgemeingültig ... aus den Grundlagen der Mechanik herleiten ...“

In der Literatur gibt es Betrachtungen der maximalen Korrelation, die mit der hier beschriebenen mehr oder weniger verwandt sind. Ein Beispiel dafür liefert die Literatur zur Mustererkennung, s. z.B. [Avi-Itzhak u.a. 1995]. Bei dieser Aufgabe entsprechen aber oft andere Kriterien besser der Natur des Problems, so daß die Maximierung der Korrelation in einigen Literaturquellen zur Mustererkennung nur beiläufig als Möglichkeit erwogen wird, s. z.B. [Young u.a. 1986] oder [Gyergyek u.a. 1988].

Beispiele zur Anwendung der Mustererkennung gibt es in der Geodäsie vor allem in der „Very Long Baseline Interferometry“ (VLBI), s. z.B. [Moritz und Mueller 1987] und in der digitalen Bildverarbeitung, s. z.B. [Kreiling 1976], [Boochs 1984], [Göpfert 1987] oder [Mather 1987]. Die diskrete Natur der Daten in diesen Anwendungen bewirkte, daß nur solche Algorithmen erarbeitet wurden, die durch systematisches Ausprobieren und Vergleichen nach Lösungen suchen, welche die höchste Korrelation liefern. Das Gleiche gilt für die Anwendung der maximalen Korrelation bei digitalen Nivellierungen ([Gächter u.a. 1984], [Ingensand 1990]) und in der Navigation ([Shokhin und Bykova 1998], [Bykov und Gritchanyuk 1998]). Eine Theorie der Ausgleichung nach maximaler Korrelation wurde dabei nicht entwickelt.

In [Stephens 1979] wird die Korrelation zwischen zwei Sätzen gerichteter Größen mit gemeinsamem Ursprung als Korrelation zwischen den dazugehörigen Einheitsvektoren, d.h. in einem gewissen Sinne der Korrelation zwischen Richtungen, definiert. Die maximale Korrelation wird durch Drehung von einem der beiden Sätze gesucht. Die dadurch eingeführte Ausgleichung nach maximaler Korrelation kann als ein Spezialfall der in dieser Arbeit betrachteten Ausgleichung verstanden werden.

In [Baczkowski und Mardia 1990] wird die Prädiktion aufgrund der Maximierung des quadrierten Korrelationskoeffizienten betrachtet, wobei die Korrelation im gleichen Sinne wie in der vorliegenden Arbeit verstanden wird. Die Zielsetzung ist dort aber eine andere (Dateninterpolation), weshalb ein stochastisches Modell postuliert wird, worauf auch die dort vorgeschlagene Auflösung der bei einer Ausgleichung nach maximaler Korrelation auftretenden Mehrdeutigkeit basiert.

Weiterhin soll ein Mißverständnis vermieden werden. Es bietet sich an, die Frage zu stellen, ob die Ausgleichung nach maximaler Korrelation nicht etwas sehr ähnliches oder gar das Gleiche wie die Maximum-Likelihood-Methode ist. Dazu ist anzumerken, daß bei der Maximum-Likelihood-Methode (s. z.B. [Bjerhammar 1973] oder [Koch 1980]) eine Likelihood-Funktion, d.h. im Grunde eine Wahrscheinlichkeitsdichte maximiert wird. Die Voraussetzung für die Anwendung dieser Methode ist die Kenntnis der Dichtefunktion. Im Falle normalverteilter Beobachtungen liefert die Methode im Gauß-Markoff-Modell gleiche Lösungen wie die Methode der kleinsten Quadrate. Im Gegensatz dazu benötigt eine Ausgleichung nach maximaler Korrelation keine statistischen Annahmen³. Deshalb ist ihre Beziehung mit der Ausgleichung nach kleinsten Quadraten von stochastischen Annahmen völlig unabhängig. Diese Beziehung hängt allein von den Eigenschaften des *deterministischen* Modells $f(u)$ ab (s. Abschnitt 2.3.6).

Bei der Quasi-Likelihood-Methode (s. z.B. [Wedderburn 1974], [McCullagh 1983]) weist die Quasi-Likelihood-Funktion höchste (maximale) quadrierte Korrelation mit der unbekannten wahren Likelihood-Funktion (s. [Li 1993]) auf. Diese Methode kann deshalb als eine Verallgemeinerung der Maximum-Likelihood-Methode betrachtet werden. Trotz maximaler Korrelation hat auch diese Methode wenig Gemeinsames mit der hier vorgestellten Ausgleichung nach maximaler Korrelation.

Weiterhin sind in der Literatur aus verschiedenen Fachgebieten mehrere Arbeiten vorhanden, in denen der Begriff *maximale Korrelation* zu finden ist. Interessanterweise kann daraus erssehen werden, daß der Begriff *Korrelation* in einem fast beliebigen Sinne benutzt werden darf⁴. So wird beispielsweise in [Feng u.a. 1998] die Korrelation zwischen einer Booleschen Funktion und Boolescher (linearen oder nichtlinearen) Kombinationen einer Untermenge ihrer Variablen betrachtet und maximiert, wobei die Korrelation als Differenz zweier Wahrscheinlichkeiten definiert wird. In [Zhang 1994] wird der Begriff „Korrelation“ ebenso in einem modifizierten Sinne benutzt und die maximale Korrelation wird eingesetzt, um eine „optimierte“ Wahl der Knotenpunkte bei der Modellierung mit Splines zu erzielen, in [Perico und Pratolongo 1997] geht es um die Wahl der Variablen, die berücksichtigt werden.

Anlaß für die Entwicklung der hier betrachteten Ausgleichung nach maximaler Korrelation war ein Anwendungsbeispiel aus der physikalischen Geodäsie, s. *Einleitung*.

³Die Minimierung der L_2 -Norm bei der Methode der kleinsten Quadrate ist durchführbar, unabhängig davon, ob es überhaupt zu einer statistischen Interpretation kommen soll. Genauso ist eine Maximierung des Korrelationskoeffizienten von einem möglichen statistischen Kontext nicht abhängig.

⁴Das Gleiche gilt für viele andere Begriffe, die in der wissenschaftlichen Terminologie auftreten. Dieser Umstand ist eine der wichtigsten Quellen aller Mißverständnisse in wissenschaftlichen Diskussionen.

Die Darstellung in der vorliegenden Arbeit basiert jedoch auf [Petrović 1991], ist aber mit der dortigen Betrachtungsweise nicht mehr in allen Punkten identisch. In der zitierten Arbeit wurde nur die Ausgleichung nach maximaler Korrelation ohne Gewichte eingeführt. Nun ist es gelungen, diese Ausgleichung auf gewichtete bzw. korrelierte Beobachtungen auszudehnen.

Noch wichtiger ist die Tatsache, daß ich damals versuchte, durch zusätzliche, mehr oder weniger logische, letztendlich aber mathematische und deshalb der Natur des Problems nicht immer entsprechende zusätzliche Bedingungen⁵ stets eine eindeutige Lösung zu erzwingen. Aufgrund der Einsicht, daß es vorteilhaft sein kann, die ganze Lösungsklasse zu erzeugen (s. Abschnitt 1.4 und 3.1), muß nun die allgemeine Strategie zur Lösung des Problems ergänzt und modifiziert werden (s. insbesondere Abschnitte 2.3.4 und 2.3.5).

2.2 Korrelationskoeffizient

Unter „Korrelation“ ist in dieser Arbeit immer die lineare Korrelation gemeint, die nicht nur in der geodätischen Literatur von großer Bedeutung ist⁶. Der Korrelationskoeffizient beschreibt die Beziehung zwischen zwei in numerischer Form ausgedrückte Eigenschaften von Objekten. Dementsprechend werden jedem Objekt zwei Werte zugeordnet. Alle Objekte werden in einer Menge zusammengefaßt. Diese Menge ist die Domäne (Definitionsbereich) zweier reellen Funktionen⁷, deren Werte in einer 1–1 Zuordnung stehen. Die gemeinsame Domäne dient im Prinzip nur diesem Zweck, der Realisierung der genannten 1–1 Zuordnung, und wird aus diesem Grund in den meisten gängigen Definitionen nicht explizit erwähnt (s. z.B. [Rainsford 1957], [Wolf 1968], [Großmann 1969], [Koch 1980]). Dennoch ist die Domäne immer implizit vorhanden. Die Definition des Korrelationskoeffizienten kann somit in folgende Form überführt werden, s. auch [Petrović 1991].

Definition 2.1 Für eine beliebige Menge $\mathcal{X} \subseteq \mathbb{R}^m$ und zwei beliebige reelle Funktionen $f : \mathcal{X} \rightarrow \mathbb{R}$ und $g : \mathcal{X} \rightarrow \mathbb{R}$ wird der Korrelationskoeffizient r zwischen den Mengen $\mathcal{Y} = f(\mathcal{X}) = \{Y : Y = f(X), X \in \mathcal{X}\}$ und $\mathcal{Z} = g(\mathcal{X}) = \{Z : Z = g(X), X \in \mathcal{X}\}$ durch die Gleichung

$$r^2 = r^2(\mathcal{Y}, \mathcal{Z}) = \frac{\left(\int (Y - \bar{Y})(Z - \bar{Z}) dX \right)^2}{\int (Y - \bar{Y})^2 dX \int (Z - \bar{Z})^2 dX} \quad (2.2)$$

definiert, wobei

$$\bar{Y} = \frac{\int Y dX}{\int dX}, \quad \bar{Z} = \frac{\int Z dX}{\int dX}. \quad (2.3)$$

Die Definition 2.1 ist nur dann anwendbar, wenn alle Integrale in den Ausdrücken (2.2) und (2.3) existieren. In praktischen Anwendungen ist die Anzahl der Daten immer endlich, wodurch sich die obige Definition wie folgt spezialisieren läßt.

Definition 2.2 Für eine beliebige Menge $\mathcal{X} = \{X_1, X_2, \dots, X_n\}$ und zwei beliebige reelle Funktionen $f : \mathcal{X} \rightarrow \mathbb{R}$ und $g : \mathcal{X} \rightarrow \mathbb{R}$ wird der Korrelationskoeffizient r zwischen den Mengen $\mathcal{Y} = f(\mathcal{X}) = \{Y_i : Y_i = f(X_i), X_i \in \mathcal{X}\}$ und $\mathcal{Z} = g(\mathcal{X}) = \{Z_i : Z_i = g(X_i), X_i \in \mathcal{X}\}$ durch die Gleichung

$$r^2 = r^2(\mathcal{Y}, \mathcal{Z}) = \frac{\left(\sum_{i=1}^n (Y_i - \bar{Y})(Z_i - \bar{Z}) \right)^2}{\sum_{i=1}^n (Y_i - \bar{Y})^2 \sum_{i=1}^n (Z_i - \bar{Z})^2} \quad (2.4)$$

⁵Die im Abschnitt 3.2.1 der vorliegenden Arbeit innerhalb der Residuenanalyse behandelten zusätzlichen Bedingungen sind von einer anderen Art. Diese kommen nicht aus rein mathematischen Überlegungen hervor, sondern dienen dem Versuch, einige Eigenschaften der systematischen Einflüsse näherungsweise zu beschreiben.

⁶Im Abschnitt 2.1 wurde dokumentiert, daß der Begriff *Korrelation* in der Literatur auch in anderen Bedeutungen zu finden ist.

⁷Dabei wird keine Einschränkung auf stochastische oder eine andere Art von Funktionen mit reellen Werten vorgenommen.

definiert, wobei

$$\bar{Y} = \frac{\sum_{i=1}^n Y_i}{n}, \quad \bar{Z} = \frac{\sum_{i=1}^n Z_i}{n}. \quad (2.5)$$

In der Literatur wird zum Teil noch die „nicht normierte Korrelation“ betrachtet, die durch

$$r = \sum_{i=1}^n (Y_i - \bar{Y})(Z_i - \bar{Z}) \quad (2.6)$$

oder gar durch

$$r = \sum_{i=1}^n Y_i Z_i \quad (2.7)$$

definiert wird, s. z.B. [Wiener 1970].

Wird weiterhin für die beiden Funktionen f und g ein stochastischer Charakter angenommen, kann für den Korrelationskoeffizienten (2.4) eine entsprechende Fehlerschätzung hergeleitet werden. Im Falle, daß noch dazu Ausreißer (aber keine systematischen Fehler) im Datenmaterial vermutet werden, bietet es sich an, in Analogie zu den robusten Ausgleichungsverfahren, einen „robusten Korrelationskoeffizienten“ einzuführen. Derartige Aufgabenstellungen werden u.a. in [Steiner und Hajagos 2001] erörtert.

In dieser Arbeit wird davon ausgegangen, daß das Datenmaterial bereits von Ausreißern gesäubert wurde, die systematischen Fehler aber nicht auszuschließen sind. Dementsprechend wird die lineare Korrelation im Sinne der Definition 2.2 benutzt, ein stochastischer Charakter der beiden Funktionen f und g wird aber nicht vorausgesetzt.

Werden die Bezeichnungen

$$\begin{aligned} c_1 + c_2 \mathcal{Y} &= \{c_1 + c_2 Y_i : Y_i = f(X_i), X_i \in \mathcal{X}\}, \\ c_3 + c_4 \mathcal{Z} &= \{c_3 + c_4 Z_i : Z_i = g(X_i), X_i \in \mathcal{X}\} \end{aligned} \quad (2.8)$$

eingeführt, dann ist der Beweis einfach zu erbringen, daß eine Verschiebung oder Streckung der Menge \mathcal{Y} (oder \mathcal{Z}) den Wert des Korrelationskoeffizienten für beliebige reelle Zahlen c_1, c_2, c_3 und c_4 nicht ändert (s. z.B. [Baczkowski und Mardia 1990] oder [Petrović 1991]):

$$r^2(c_1 + c_2 \mathcal{Y}, c_3 + c_4 \mathcal{Z}) = r^2(\mathcal{Y}, \mathcal{Z}). \quad (2.9)$$

Die Definitionen 2.1 und 2.2 ermöglichen es, im Falle $\mathcal{X} \subseteq \Re^m$, den Korrelationskoeffizienten r als Merkmal der Beziehung zwischen zwei geometrischen Konfigurationen im \Re^{m+1} zu betrachten; und zwar als Beziehung der Graphen der Funktionen f und g zueinander. Auf diese Weise wird r^2 (oder $|r|$) ein Maß dafür, wie verwandt die Formen der beiden Konfigurationen sind⁸. Läßt sich eine der beiden Konfigurationen durch eine Translation und eine Streckung (in die Richtung der $(m+1)$ -ten Achse) in die andere überführen, dann ist der Betrag des Korrelationskoeffizienten gleich 1. Ist diese Überführung nur näherungsweise möglich, dann beschreibt der Betrag von r^2 , wie gut die transformierte Konfiguration der anderen entspricht.

Die Definition 2.2 ermöglicht darüber hinaus, den Korrelationskoeffizienten zwischen den Vektoren l und $f(u)$ (s. Gleichung (1.4)) einzuführen. Es handelt sich immer um endlich viele, genauer gesagt n Beobachtungen, daher haben beide Vektoren n Komponenten:

$$l = \begin{bmatrix} l_1 \\ l_2 \\ \vdots \\ l_n \end{bmatrix}, \quad f(u) = \begin{bmatrix} f(u)_1 \\ f(u)_2 \\ \vdots \\ f(u)_n \end{bmatrix}. \quad (2.10)$$

⁸Die Verwendung des Korrelationskoeffizienten beim Vergleich geometrischer Formen (angewendet beispielsweise auf charakteristische Flächen des Erdkörpers und des Erdschwerefeldes, s. [Čolić und Petrović 1984], [Čolić u.a. 1987], [Vučetić u.a. 1990], [Petrović 1993]) und nicht unbedingt nur bei einem Vergleich stochastischer Variablen oder Funktionen, hat die Erarbeitung der in der vorliegenden Arbeit vorgestellten Art der Ausgleichung nach maximaler Korrelation veranlaßt. Die Zielsetzung hat sich aber geändert; es werden nicht mehr nur die Formen physikalisch definierter Flächen, sondern beliebige in den Daten und Modellen enthaltene Formen verglichen.

Die Domäne \mathcal{X} wird

$$\mathcal{X} = \{i : i = 1, \dots, n\} = \{1, 2, \dots, n\} \quad (2.11)$$

und dementsprechend

$$X_i = i, \quad (2.12)$$

$$Y_i = f(u)_i, \quad (2.13)$$

$$Z_i = l_i, \quad (2.14)$$

woraus folgt

$$r^2(f(u), l) = \frac{\left(\sum_{i=1}^n (f(u)_i - \bar{f}(u))(l_i - \bar{l}) \right)^2}{\sum_{i=1}^n (f(u)_i - \bar{f}(u))^2 \sum_{i=1}^n (l_i - \bar{l})^2}. \quad (2.15)$$

Dieser Korrelationskoeffizient vergleicht die Formen des unvollständigen Modells $f(u)$ mit den in den Daten l enthaltenen Formen. Bei einem vollständigen Modell⁹ und fehlerfreien Messungen ($s_f = 0$ und $z = 0$) ergäbe sich für r^2 ein Wert von 1.

2.3 Ausgleichung nach maximaler Korrelation

2.3.1 Definition

Die Idee der Ausgleichung nach maximaler Korrelation besteht darin, statt minimale Residuen gemäß (1.5) zu verlangen, den Korrelationskoeffizienten (2.15) zwischen dem unvollständigen Modell und den Daten zu maximieren:

$$r^2(f(u), l) \rightarrow \max. \quad (2.16)$$

Auf diese Weise werden *nicht* die unbekannten Modellparameter gesucht, indem *die Residuen zum Minimum gezwungen* werden. Es werden lediglich die Formen des vorhandenen, unvollständigen Modells in den Daten identifiziert. Dementsprechend lässt sich eine solche Ausgleichung folgendermaßen einführen.

Definition 2.3 Sei $\mathcal{X} = \{1, 2, \dots, n\}$, $g : \mathcal{X} \rightarrow \mathbb{R}$ definiert durch $g(i) = l_i$, $f(u) : \mathcal{X} \rightarrow \mathbb{R}$ definiert durch $f(u)(i) = f(u)_i$, weiterhin $\mathbf{F} \subseteq \{f(u) : u \in \mathbb{R}^m\}$ die gegebene Klasse reeller Funktionen und $\mathcal{Y} = f(u)(\mathcal{X})$, $\mathcal{Z} = g(\mathcal{X})$. Falls es ein Element $f(u) \in \mathbf{F}$ gibt, so daß $r^2(\mathcal{Y}, \mathcal{Z}) = r^2(f(u), l)$ den maximal möglichen Wert bezüglich aller Elemente der Klasse \mathbf{F} annimmt, dann ist die Funktion $f(u)$ die Lösung des Problems der Ausgleichung nach maximaler Korrelation.

Diese Definition führt eine Ausgleichung nach maximaler Korrelation ein, in der alle Beobachtungen gleichgewichtig sind. In [Petrović 1991] wurde nur dieser Spezialfall betrachtet. Inzwischen hat sich gezeigt, daß in Analogie zur Ausgleichung nach kleinsten Quadraten auch gewichtete und korrelierte Beobachtungen einführt werden können.

2.3.2 Verallgemeinerung auf gewichtete und korrelierte Beobachtungen

Jede Ausgleichung nach kleinsten Quadraten mit Gewichten lässt sich bekanntlich in eine andere Ausgleichung¹⁰ ohne Gewichte überführen¹¹, s. z.B. [Linnik 1961]. Aus (1.4) folgt

$$p_i v_i^2 = p_i (f(u)_i - l_i)^2 = (\sqrt{p_i} f(u)_i - \sqrt{p_i} l_i)^2. \quad (2.17)$$

⁹mit bekannten und fehlerfreien Parametern

¹⁰Diese andere Ausgleichung entspricht weder der ursprünglichen Fragestellung, noch einer interpretierbaren Ersatzaufgabe.

¹¹Am Anfang war dies auch die eigentliche Methode zur Durchführung einer Ausgleichung mit gewichteten Beobachtungen, s. [Bessel und Baeyer 1838]. Die Darstellungsform im zitierten Werk ist ziemlich umständlich und dementsprechend ist nicht gleich ersichtlich, daß es sich grundsätzlich um die betrachtete einfache Transformation handelt.

Nun können das Ersatzmodell $\tilde{f}(u)$ und die Ersatzbeobachtungen \tilde{l} eingeführt werden:

$$\tilde{l} = \begin{bmatrix} \sqrt{p_1}l_1 \\ \sqrt{p_2}l_2 \\ \vdots \\ \sqrt{p_n}l_n \end{bmatrix}, \quad \tilde{f}(u) = \begin{bmatrix} \sqrt{p_1}f(u)_1 \\ \sqrt{p_2}f(u)_2 \\ \vdots \\ \sqrt{p_n}f(u)_n \end{bmatrix}, \quad \tilde{v} = \begin{bmatrix} \sqrt{p_1}v_1 \\ \sqrt{p_2}v_2 \\ \vdots \\ \sqrt{p_n}v_n \end{bmatrix}. \quad (2.18)$$

Mit

$$R = \begin{bmatrix} \sqrt{p_1} & 0 & \cdots & 0 \\ 0 & \sqrt{p_2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sqrt{p_n} \end{bmatrix}. \quad (2.19)$$

folgt aus (2.18)

$$\tilde{f}(u) = Rf(u), \quad (2.20)$$

$$\tilde{l} = Rl, \quad (2.21)$$

$$\tilde{v} = Rv. \quad (2.22)$$

Anstelle von (1.4) ergibt sich nun

$$\tilde{l} + \tilde{v} = \tilde{f}(u). \quad (2.23)$$

Diese Gleichung induziert eine Ausgleichung nach kleinsten Quadraten ohne Gewichte, die mit der ursprünglichen Ausgleichung mit Gewichten äquivalent¹² ist. Es ist ersichtlich, daß diese Transformation unabhängig davon, ob ein Ausgleichungsproblem linear oder nichtlinear ist, durchgeführt werden kann.

Wird nun eine Matrix P durch

$$P = R^T R = \begin{bmatrix} p_1 & 0 & \cdots & 0 \\ 0 & p_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & p_n \end{bmatrix} \quad (2.24)$$

definiert, dann ist diese Matrix die übliche, aus der Ausgleichung nach kleinsten Quadraten bekannte Gewichtsmatrix.

Im Linearfall (vergleiche Abschnitt 2.3.5) ist

$$f(u) = Au, \quad (2.25)$$

und die Gleichung (2.20) nimmt die Form

$$\tilde{A} = RA \quad (2.26)$$

an. Die Gleichungen (2.21) und (2.22) bleiben unverändert und das transformierte Modell (2.23) für die Ausgleichung nach kleinsten Quadraten wird

$$\tilde{l} + \tilde{v} = \tilde{A}u. \quad (2.27)$$

Eine Ausgleichung nach maximaler Korrelation mit Gewichten kann auf gleiche Weise eingeführt werden wie bei einer Ausgleichung nach kleinsten Quadraten. Dazu muß lediglich die Bedeutung von \mathcal{Y} und \mathcal{Z} in Definition 2.3 gemäß (2.23) geändert werden:

$$r^2(\mathcal{Y}, \mathcal{Z}) = r^2(\tilde{f}(u), \tilde{l}). \quad (2.28)$$

Eine weitere Verallgemeinerung auf korrelierte Beobachtungen ist sowohl im Linearfall als auch im allgemeinen Fall problemlos möglich. Die Diagonalmatrix P aus (2.24) muß dann durch eine entsprechende vollbesetzte Matrix ersetzt werden. Das einzige Problem ist die Darstellung einer solchen Matrix P in der Form

$$P = R^T R. \quad (2.29)$$

¹²Im Sinne, daß sie gleiche Ergebnisse liefert.

Falls P positiv definit ist, läßt sich die Zerlegung (2.29) durchführen¹³, beispielsweise mit Hilfe des Cholesky-Verfahrens. Definition 2.3 zusammen mit (2.20)–(2.21) legt dann eine Ausgleichung nach maximaler Korrelation fest.

Die Verallgemeinerung der Ausgleichung nach maximaler Korrelation auf gewichtete und korrelierte Beobachtungen läßt sich auch auf eine andere Weise interpretieren, die auf einer Verallgemeinerung der Definition des Korrelationskoeffizienten beruht und mit der vorgestellten Transformation korrelierter in nichtkorrelierter Beobachtungen äquivalent ist.

Mit den Bezeichnungen

$$L = \begin{bmatrix} l_1 - \bar{l} \\ l_2 - \bar{l} \\ \vdots \\ l_n - \bar{l} \end{bmatrix}, \quad F = \begin{bmatrix} f(u)_1 - \bar{f(u)} \\ f(u)_2 - \bar{f(u)} \\ \vdots \\ f(u)_n - \bar{f(u)} \end{bmatrix} \quad (2.30)$$

kann die Definition des Korrelationskoeffizienten (2.15) als

$$r^2(f(u), l) = \frac{(F^T L)^2}{(F^T F)(L^T L)} \quad (2.31)$$

dargestellt werden.

Nach der Verallgemeinerung auf korrelierte Messungen ergibt sich für das transformierte Problem

$$r^2(\tilde{f}(u), \tilde{l}) = \frac{(\tilde{F}^T \tilde{L})^2}{(\tilde{F}^T \tilde{F})(\tilde{L}^T \tilde{L})}. \quad (2.32)$$

Diese Darstellung schließt (2.31) als Spezialfall ein: Ist P (und somit auch R) eine Einheitsmatrix, dann ist $\tilde{l} = l$ usw., bzw. $\tilde{L} = L$ und $\tilde{F} = F$.

Die Gleichung (2.32) läßt sich weiter in die Form

$$r^2(\tilde{f}(u), \tilde{l}) = \frac{(f^T \tilde{P} l)^2}{(f^T \tilde{P} f)(l^T \tilde{P} l)}. \quad (2.33)$$

überführen, wobei \tilde{P} als transformierte Gewichtsmatrix interpretiert werden kann, mit der sowohl die Zentrierung als auch die Berücksichtigung von Gewichten und Korrelationen erfolgt. Diese Transformation wurde in [Neitzel 2002] ausgearbeitet und im Detail diskutiert.

Nun kann (2.33) als eine verallgemeinerte Definition des Korrelationskoeffizienten angesehen werden.

Formell betrachtet ist eine derartige verallgemeinerte Darstellung auch in einiger Literatur zur Statistik zu finden¹⁴, meistens unter dem Namen „coefficient of determination“, s. [Searle 1971], [Seber 1977] oder [Draper und Smith 1981]. In der neueren geodätischen Literatur wird diese Möglichkeit, soweit mir bekannt, nur in [Perović 1998]¹⁵ eingeführt und diskutiert.

Nach der Festlegung des Korrelationskoeffizienten zwischen dem Modell und den Beobachtungen durch (2.33), anstelle von (2.15), kann die Definition 2.3 der Ausgleichung nach maximaler Korrelation ohne zusätzliche Modifikationen weiterhin verwendet werden. Die Ausgleichung korrelierter Beobachtungen nach maximaler Korrelation bedarf keiner speziellen weiteren Betrachtungen. Es ist einfach einzusehen, daß die auf diese Weise definierte Ausgleichung im Falle einer Einheitsmatrix P mit der davor eingeführten Ausgleichung ohne Gewichte und Korrelationen zusammenfällt und im Falle einer Matrix P , die Gewichte und Korrelationen enthält, mit dem Ergebnis der Transformation korrelierter in nichtkorrelierter Beobachtungen identisch ist.

¹³ Die Transformation (2.20)–(2.21) korrelierter in nichtkorrelierter Beobachtungen ist bei einer Ausgleichung nach maximaler Korrelation also die gleiche wie bei einer Ausgleichung nach kleinsten Quadraten und basiert auf der Zerlegung (2.29), s. beispielsweise [Wolf 1968], [Bjerhammar 1973], [Koch 1980]. Die dazu notwendige Bedingung, daß die Matrix P positiv definit sein soll, ist bei den meisten für die Anwendung relevanten Fragestellungen erfüllt.

¹⁴ Dort mit einer anderen Zielsetzung und Interpretation eingeführt.

¹⁵ Der „coefficient of determination“ war auch mir bis vor kurzem unbekannt. Professor Gligorije Perović hat mich auf das Vorhandensein dieser Verallgemeinerung des Korrelationskoeffizienten in der Literatur aufmerksam gemacht.

2.3.3 Mehrdeutigkeit der Lösung

Aus Definition 2.3 ist ersichtlich, daß als Lösung f nicht eine völlig beliebige Funktion gesucht wird, sondern, wie bei allen Arten der Ausgleichung, nur eine Funktion f aus einer im voraus festgelegten Klasse \mathbf{F} .

Mit Hilfe der Gleichung (2.9) läßt sich die Frage, unter welchen Umständen die Lösung *nicht* eindeutig sein kann, beantworten. Der folgende Satz ergibt sich unmittelbar aus (2.9), vergleiche auch [Baczkowski und Mardia 1990]¹⁶.

Satz 2.4 Falls die Lösung f der Ausgleichung nach maximaler Korrelation in einer Klasse \mathbf{F} mit der Eigenschaft

$$f \in \mathbf{F} \quad \& \quad c_1, c_2 \in \mathfrak{R} \quad \Rightarrow \quad c_1 + c_2 f \in \mathbf{F} \quad (2.34)$$

gesucht wird, dann gilt für jede Lösung $f_0 \in \mathbf{F}$, daß alle Funktionen $c_1 + c_2 f_0 \in \mathbf{F}$ auch Lösungen des gleichen Problems sind.

Dies bedeutet, daß die Lösung nach maximaler Korrelation im Falle (2.34) nicht nur mehrdeutig ist, sondern aus einer ganzen Unterkategorie besteht:

$$\{f : f = c_1 + c_2 f_0 \quad \forall c_1, c_2 \in \mathfrak{R}\} \subseteq \mathbf{F}. \quad (2.35)$$

Speziell, falls nur

$$f \in \mathbf{F} \quad \& \quad c_1 \in \mathfrak{R} \quad \Rightarrow \quad c_1 + f \in \mathbf{F}, \quad (2.36)$$

dann generiert jede Lösung f_0 eine Klasse der Lösungen

$$\{f : f = c_1 + f_0 \quad \forall c_1 \in \mathfrak{R}\} \subseteq \mathbf{F} \quad (2.37)$$

und im Falle

$$f \in \mathbf{F} \quad \& \quad c_2 \in \mathfrak{R} \quad \Rightarrow \quad c_2 f \in \mathbf{F} \quad (2.38)$$

ergibt sich

$$\{f : f = c_2 f_0 \quad \forall c_2 \in \mathfrak{R}\} \subseteq \mathbf{F}. \quad (2.39)$$

Im Falle der erstgenannten und allgemeineren Gleichung (2.34) kann bewiesen werden, daß die Lösung nach kleinsten Quadraten zur Klasse aller Lösungen nach maximaler Korrelation gehört, s. [Petrović 1991]. Damit ist die Lösung nach maximaler Korrelation „mehrdeutiger“ als die Lösung nach kleinsten Quadraten. Sogar bei Aufgabenstellungen, bei denen die Ausgleichung nach kleinsten Quadraten eine eindeutige Lösung liefert¹⁷, ist die Lösung nach maximaler Korrelation nicht eindeutig, was bei der Interpretation der Residuen von großem Vorteil sein kann, s. Abschnitt 3.2.

In der Literatur zur Statistik, s. z.B. [Rao 1973], ist häufig der Beweis zu finden, daß die durch eine Ausgleichung nach kleinsten Quadraten gefundenen Modellparameter einer linearen Regression den Betrag des linearen Korrelationskoeffizienten maximieren. Die Frage, ob es weitere Lösungen gibt, die diesen Betrag genauso maximieren, wird nicht diskutiert. Deshalb wird aufgrund dieser Erörterungen keine Ausgleichung nach maximaler Korrelation erarbeitet, sondern grundsätzlich nur eine Eigenschaft der Ausgleichung nach kleinsten Quadraten festgestellt.

Alle drei Konstellationen, (2.34), (2.36) und (2.38), sind zumindest theoretisch denkbar. Die erste und die dritte kommen in den Anwendungen auf alle Fälle vor, s. Abschnitt 2.3.5, wobei die zweite in der Praxis nur schwer vorstellbar ist. Der Suche nach der Lösung f in einer Klasse \mathbf{F} , die keine der drei genannten Eigenschaften besitzt, scheint keine Anwendung zuzuordnen zu sein¹⁸.

¹⁶“Obviously, this is not unique as squared correlation is invariant under change of scale and origin. This point is glossed over in the literature.“ („Offensichtlich ist dies nicht eindeutig, da die quadrierte Korrelation unter Maßstabs- und Ursprungsänderung invariant bleibt. Dieser Punkt wird in der Literatur vertuscht.“)

¹⁷Bekanntlich werden solche Aufgabenstellungen am häufigsten betrachtet.

¹⁸Diese Tatsache ist keine spezifische Eigenschaft der Ausgleichung nach maximaler Korrelation, sondern eine allgemeine Eigenschaft der in der Geodäsie betrachteten Fragestellungen.

2.3.4 Lösung des Ausgleichungsproblems nach maximaler Korrelation

Die Klasse \mathbf{F} der Funktionen in Definition 2.3 ist völlig willkürlich. Das ermöglicht eine sehr allgemeine Definition der Ausgleichung nach maximaler Korrelation.

Ohne Einführung von Einschränkungen ist es aber nicht möglich, eine mathematisch fundierte Lösungsstrategie auszuarbeiten. Dieses gilt natürlich nicht nur für eine Ausgleichung nach maximaler Korrelation, sondern genauso für eine Ausgleichung nach kleinsten Quadraten, bzw. für jede denkbare Form der Ausgleichung.

Bei jeder praktischen Anwendung gibt es ein (vollständiges oder unvollständiges) *mathematisch beschreibbares* Modell. Diesem Wissen folgend wird nun die gleiche Annahme akzeptiert, die auch bei allen anderen Arten der Ausgleichung gemacht wird, nämlich daß eine analytische Formel (mathematischer Ausdruck) existiert, die alle Funktionen $f \in \mathbf{F}$ beschreibt. Mit anderen Worten, es gibt eine reelle Funktion $F : \mathcal{X} \times \mathbb{R}^m \rightarrow \mathbb{R}$ mit der Eigenschaft, daß jede Auswahl eines Parametersatzes $\{u_1, u_2, \dots, u_m\}$ einer der Funktionen $f \in \mathbf{F}$ entspricht:

$$f(X) = F(X, u_1, u_2, \dots, u_m) \quad \forall X \in \mathcal{X}. \quad (2.40)$$

Umgekehrt, jede Funktion $f \in \mathbf{F}$ soll eine derartige Darstellung haben. Der folgende Satz kann bewiesen werden, s. [Petrović 1991].

Satz 2.5 Sei $\mathcal{X} = \{i : i = 1, \dots, n\}$ und die Funktion $F : \mathcal{X} \times \mathbb{R}^m \rightarrow \mathbb{R}$, die die Klasse \mathbf{F} definiert (Definition 2.3) sei mindestens einmal nach u_k ($k = 1, \dots, m$) differenzierbar. Weiterhin wird bezeichnet

$$\left. \begin{aligned} Z_i &= g(i) = l_i, \\ F_i &= F(i, u_1, u_2, \dots, u_m), \\ \overline{Z} &= \frac{\sum_{i=1}^n Z_i}{n} = \frac{\sum_{i=1}^n l_i}{n} = \overline{l}, \\ \overline{F} &= \frac{\sum_{i=1}^n F(i, u_1, u_2, \dots, u_m)}{n} = \frac{\sum_{i=1}^n F_i}{n}. \end{aligned} \right\} \quad (2.41)$$

Ist $f_0(u) \in \mathbf{F}$ eine Lösung, dann erfüllen die dazugehörigen Werte der Parameter u_k das Gleichungssystem

$$\left. \begin{aligned} \sum_{i=1}^n \frac{\partial(F_i - \overline{F})}{\partial u_k} (l_i - \overline{l}) \sum_{i=1}^n (F_i - \overline{F})^2 \\ - \sum_{i=1}^n (F_i - \overline{F})(l_i - \overline{l}) \sum_{i=1}^n (F_i - \overline{F}) \frac{\partial(F_i - \overline{F})}{\partial u_k} = 0 \\ (k = 1, 2, \dots, m). \end{aligned} \right\} \quad (2.42)$$

Mit den Abkürzungen

$$W_i = l_i - \overline{l} \quad (i = 1, 2, \dots, n) \quad (2.43)$$

und durch eine kleine Umformung unter Berücksichtigung der Tatsache daß

$$\sum_{i=1}^n W_i = 0, \quad (2.44)$$

läßt sich das Gleichungssystem (2.42) in die folgende Form bringen:

$$\left. \begin{aligned} \left(\sum_{i=1}^n W_i \frac{\partial F_i}{\partial u_k} \right) \left(\sum_{i=1}^n F_i^2 - n \overline{F}^2 \right) - \left(\sum_{i=1}^n W_i F_i \right) \left(\sum_{i=1}^n F_i \frac{\partial F_i}{\partial u_k} - n \overline{F} \frac{\partial \overline{F}}{\partial u_k} \right) = 0 \\ (k = 1, 2, \dots, m). \end{aligned} \right\} \quad (2.45)$$

Das Gleichungssystem (2.42) bzw. (2.45) ist im allgemeinen nichtlinear. Die daraus resultierende Schwierigkeit ist mit der bei einer Ausgleichung nach kleinsten Quadraten gleichwertig. In den meisten für die Anwendung relevanten Fällen resultiert eine Ausgleichung nach kleinsten Quadraten ebenso in einem nichtlinearen Gleichungssystem der „Normalgleichungen“, siehe beispielsweise [Sansò 1973], [Teunissen 1990], [Grafarend und Schaffrin 1991] und [Dermanis 1998].

Wie wird das Problem der Nichtlinearität bei einer Ausgleichung nach kleinsten Quadraten gelöst?

Im selten auftretenden Spezialfall¹⁹ (z.B. bei der Ausgleichung eines Nivellementsnetzes) sind die „Beobachtungsgleichungen“ (1.4) linear. Die Bedingung (1.5) resultiert dann in linearen „Normalgleichungen“, die mit den bekannten Methoden direkt gelöst werden können.

Aus dieser Tatsache ergibt sich eine mögliche allgemeine Vorgehensweise (in der Literatur meistens *Gauß–Newton–Verfahren* genannt): Die Beobachtungsgleichungen werden vor Einführung der Bedingung (1.5) linearisiert. Dieses Verfahren ist sehr verbreitet, einige Literaturquellen erwecken sogar den Eindruck, daß es keine anderen Möglichkeiten gibt, das Problem zu lösen, s. z.B. [Reissmann 1962]²⁰, [Wolf 1968]²¹, oder [Großmann 1969]. Das Fragwürdige bei dieser Methode besteht darin, daß gleich am Anfang die eigentliche Fragestellung durch eine andere (lineare) ersetzt wird und die Ausgleichung mit der neuen Aufgabenstellung durchgeführt wird. Die Ergebnisse sind davon abhängig, wie genau das lineare Modell das ursprüngliche repräsentiert. Die Qualität dieser Approximation hängt wiederum von der Güte der eingeführten, für die Linearisierung notwendigen, Näherungswerte für die Lösungen ab. Dabei ist es sehr schwierig, im voraus abzuschätzen, wie gut die Näherungswerte sein müssen. Dies ergibt sich aus der Tatsache, daß unterschiedliche nichtlineare Probleme nicht „in gleichem Maße“ oder nicht „gleich stark“ nichtlinear sind. Deshalb wird in [Liu 2000] die Frage erörtert, welche nichtlinearen Probleme linearisiert werden dürfen und welche nicht. Dabei wird die Krümmung als Maß für die Nichtlinearität eines Problems benutzt.

Trotz allem funktioniert die beschriebene Vorgehensweise in vielen, für die Praxis wichtigen Fällen recht gut; insbesondere, wenn sie iterativ durchgeführt wird. Die durchzuführende Iteration unterscheidet sich bei dieser Interpretation des Gauß–Newton–Verfahrens von den „normalen“, bei numerischen Verfahren üblichen Iterationen. Ein Iterationsschritt besteht dann aus dem *Ersetzen* eines Modells und anschließender Ausgleichung mit dem Ersatzmodell.

Ein Nachteil dieser Vorgehensweise besteht darin, daß es Problemstellungen gibt, bei denen die Iteration nie konvergiert, unabhängig davon, wie gut die Näherungswerte sind. In [Schwetlick 1979] ist ein Beispiel zu finden, bei dem die Konvergenz nur dann gewährleistet ist, wenn die Näherungswerte exakt mit der gesuchten Lösung übereinstimmen. Da die Lösung nie im voraus bekannt ist, wurden in [Schwetlick 1979] die Bedingungen für die Konvergenz des Gauß–Newton–Verfahrens detailliert untersucht und mehrere Modifikationen²² des Verfahrens vorgeschlagen, die die Konvergenz garantieren²³.

Andererseits läßt sich aber zeigen, s. z.B. [Teunissen 1990], daß sich diese Vorgehensweise auch anders interpretieren läßt, nämlich als ein iteratives Verfahren zur Lösung des ursprünglichen Problems. Die entsprechenden Analysen, die auf differentialgeometrischen Untersuchungen basieren, ermöglichen eine Modifikation der Schrittweite, die eine lokale Konvergenz garantiert²⁴.

Da aber kein beliebiger kritischer (stationärer) Punkt gesucht wird, sondern das globale Minimum, reicht die lokale Konvergenz nicht aus. Der Erfolg ist weiterhin von den Näherungswerten abhängig. Es gibt also auch Aufgabenstellungen, bei denen diese Strategie wegen unzureichender Näherungswerte versagt, s. z.B. [Mautz 2001].

¹⁹„Linearity is rather the exception than the rule in modeling reality, . . .“ ([Dermanis und Sansó 1995])

²⁰„Für die weitere Rechnung ist es erforderlich, daß die Verbesserungsgleichungen in linearer Form vorliegen. In vielen Fällen wird dies bei (1) noch nicht zutreffen. Man muß dann die ursprünglichen Verbesserungsgleichungen linearisieren.“

²¹„Eine Besonderheit der Ausgleichungsrechnung ist es, daß sie in (210,1) immer nur *lineare* Beziehungen voraussetzt. Trifft dies von Haus aus nicht zu, so muß man unter Einführung von beliebig wählbaren Näherungswerten x_0 (für die Unbekannten x) eine Zerlegung

$$x_j = x_{0j} + \delta x_j, \quad (j = 1, 2, \dots, u) \quad (210,3)$$

herbeiführen, wobei die Zuschläge δx_j hinreichend klein sein sollen und ihre numerische Ermittlung dann das Ziel der Ausgleichung darstellt. Nimmt man sodann eine Entwicklung nach dem Satz von Taylor . . .“

²²„Gedämpftes Gauß–Newton–Verfahren mit akzeptierten Schätzwert“, „Gedämpftes Gauß–Newton–Verfahren mit modifizierter Cholesky-Zerlegung“, „Regularisiertes Gauß–Newton–Verfahren“, „Ableitungsfreies gedämpftes Gauß–Newton–Verfahren mit modifizierter Cholesky-Zerlegung“.

²³Es scheint aber, daß die geodätische Standardsoftware solche Modifikationen selten enthält.

²⁴Daraus ergibt sich wieder eine Variante des gedämpften Gauß–Newton–Verfahrens.

Eine zweite Möglichkeit besteht darin, die Bedingung (1.5) auf die ursprünglichen nichtlinearen Beobachtungsgleichungen anzuwenden. Daraus resultieren nichtlineare Normalgleichungen. Diese können nur in einigen Fällen direkt gelöst werden²⁵. Wenn es nur endlich viele kritische (stationäre) Punkte gibt, sind keine Näherungswerte notwendig: alle kritischen Punkte können bestimmt werden, entsprechende Werte der Zielfunktion berechnet und dadurch das globale Minimum gefunden werden. Beispiele für eine direkte Lösung nichtlinearer Probleme sind unter anderem in [Sansò 1973], [Jovićić u.a. 1982b], [Teunissen 1988] und [Marchenko u.a. 2001] zu finden.

Sind die nichtlinearen Normalgleichungen nicht direkt lösbar, kann zuerst versucht werden, sie weitestgehend zu vereinfachen und erst dann die numerischen, iterativen Methoden anzuwenden. Diese Vorgehensweise ist beispielsweise in [York 1966], [Bopp und Krauss 1977], [Jovićić u.a. 1982a] und [Lohse 1994] zu finden. Eine systematische Vorgehensweise zur Zerlegung nichtlinearer Probleme, die gewisse, relativ allgemeine Eigenschaften haben, in ein lineares und ein nichtlineares Problem, so daß die Lösung in zwei Schritten erfolgen kann, ist in [Teunissen 1988] zu finden.

Ein Spezialfall dieser Vorgehensweise ist das Newton–Raphson–Verfahren (s. z.B. [Lohse 1994]), bei dem die nichtlinearen Normalgleichungen ohne vorherige Umformungen (weil beispielsweise keine Vereinfachung möglich war) linearisiert werden. Immer wenn numerische iterative Methoden eingesetzt werden, stellt sich die Frage nach einer Strategie für die Bereitstellung der Näherungswerte und so ist es auch in diesem Falle.

Eine dritte Möglichkeit besteht darin, bei der Suche nach dem Minimum auf die Benutzung der Differenzierbarkeit zu verzichten und durch systematisch-stochastisches Ausprobieren (sog. heuristische Verfahren) solche Lösungen für die Modellparameter zu finden, die die Zielfunktion $\sum pvv$ minimieren, s. z.B. [Mautz 1999], [Mautz 2001].

Bei einer Ausgleichung nach maximaler Korrelation sind die gleichen drei Lösungsansätze, wie bei einer Ausgleichung nach kleinsten Quadraten, möglich.

Bei linearen Problemen gibt es eine einfache Vorgehensweise, die direkt zur Lösung führt, s. Abschnitt 2.3.5.

Bei nichtlinearen Problemen kann das ursprüngliche Modell durch ein lineares ersetzt werden und mit ihm die Ausgleichung durchgeführt werden. Die entstehenden Probleme der Näherungswerte, der iterativen Methode, usw. sind völlig identisch mit den Problemen bei der Ausgleichung nach kleinsten Quadraten und die vorgeschlagene Vorgehensweise funktioniert in der Regel genauso gut wie bei einer Ausgleichung nach kleinsten Quadraten.

Direktes oder numerisches Lösen des nichtlinearen Gleichungssystems ist ebenso bei der Maximierung der Korrelation anwendbar. Ein Beispiel dafür findet sich in [Neitzel 1998].

Schließlich ist auch ein Einsatz der heuristischen Verfahren möglich, wenn lediglich die Zielfunktion $\sum vv$ (bzw. $\sum pvv$) durch $r^2(f(u), l)$ (bzw. $r^2(\tilde{f}(u), \tilde{l})$, s. (2.23)) ersetzt wird und statt des globalen Minimums das globale Maximum gesucht wird.

2.3.5 Lineare Probleme

Unter „linearen Problemen“ wird, in Analogie zur Ausgleichung nach der Methode der kleinsten Quadrate, der Fall bezeichnet, in dem das ursprüngliche Modell $f(u)$ bereits linear ist:

$$f(u) = Au, \quad (2.46)$$

d.h.

$$l + v = Au. \quad (2.47)$$

Hier sind

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1m} \\ a_{21} & a_{22} & \cdots & a_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nm} \end{bmatrix}, \quad (2.48)$$

²⁵ Nichtlinearität bedeutet *nicht* automatisch *Nichtlösbarkeit*.

bzw.

$$u = \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_m \end{bmatrix}, \quad l = \begin{bmatrix} l_1 \\ l_2 \\ \vdots \\ l_n \end{bmatrix}, \quad v = \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix}. \quad (2.49)$$

Dementsprechend repräsentiert (2.47) die linearen Beobachtungsgleichungen für vermittelnde Beobachtungen in dem in der Literatur üblichen Sinne. Wenn nun die Bedingung

$$r^2(Au, l) \longrightarrow \max \quad (2.50)$$

aufgestellt wird, verwandelt sich das Gleichungssystem (2.45) wegen

$$F_i = \sum_{j=1}^m a_{ij} u_j \quad (i = 1, 2, \dots, n), \quad (2.51)$$

$$\frac{\partial F_i}{\partial u_k} = a_{ik} \quad (i = 1, 2, \dots, n; k = 1, 2, \dots, m) \quad (2.52)$$

und

$$\frac{\partial \bar{F}}{\partial u_k} = \frac{1}{n} \sum_{i=1}^n a_{ik} \quad (k = 1, 2, \dots, m) \quad (2.53)$$

in

$$\left. \begin{aligned} & \left(\sum_{i=1}^n a_{ik} (l_i - \bar{l}) \right) \left(\sum_{i=1}^n \left(\sum_{j=1}^m a_{ij} u_j \right)^2 - \frac{1}{n} \left(\sum_{i=1}^n \sum_{j=1}^m a_{ij} u_j \right)^2 \right) \\ & - \left(\sum_{i=1}^n \sum_{j=1}^m a_{ij} (l_i - \bar{l}) u_j \right) \left(\sum_{i=1}^n \sum_{j=1}^m a_{ij} a_{ik} u_j - \frac{1}{n} \left(\sum_{i=1}^n a_{ik} \right) \left(\sum_{i=1}^n \sum_{j=1}^m a_{ij} u_j \right) \right) = 0 \end{aligned} \right\} \quad (2.54)$$

(k = 1, 2, ..., m).

Im Gegensatz zur Ausgleichung nach kleinsten Quadraten ist dieses System nicht linear und scheint auf ersten Blick nicht elementar lösbar zu sein. Es gibt aber einen Weg, die Lösung auf eine einfache Weise zu gewinnen, ohne das Gleichungssystem (2.54) direkt zu lösen. Dieser Weg hat zusätzliche Vorteile.

Aus (2.47)–(2.49) ist ersichtlich, daß das lineare Modell die Eigenschaft (2.38) automatisch besitzt, weil für jede beliebige Auswahl der Parameter u_1, u_2, \dots, u_m die dazugehörige Auswahl $\tilde{u}_1, \tilde{u}_2, \dots, \tilde{u}_m$ ($\tilde{u}_j = c_2 u_j$) existiert, so daß

$$c_2 A u = A \tilde{u}. \quad (2.55)$$

Dementsprechend ist die Lösung nach maximaler Korrelation nicht eindeutig, sondern eine ganze Klasse.

Das Modell hat aber nicht immer die stärkere Eigenschaft (2.34), so daß die zugeordnete Lösung nach kleinsten Quadraten nicht unbedingt eine der Lösungen nach maximaler Korrelation sein muß.

Im Spezialfall, wenn eine Spalte von A gleiche Elemente enthält, d.h.

$$\exists j \in \{1, \dots, m\} : a_{ij} = a \quad \forall i \in \{1, \dots, n\}, \quad (2.56)$$

dann hat das Modell offensichtlich ebenfalls die Eigenschaft (2.34). Die Lösung läßt sich dann sehr einfach finden: alle Lösungen nach maximaler Korrelation bilden die Klasse

$$\{f : f = c_1 + c_2 f_0 \quad \forall c_1, c_2 \in \mathbb{R}\} \subseteq \mathbf{F}, \quad (2.57)$$

wobei f_0 eine beliebige aus der Menge aller Lösungen sein kann. Die Klasse aller Lösungen enthält in diesem Falle auch die Lösung nach kleinsten Quadraten (s. Abschnitt 2.3.3), daher kann diese Lösung f_0 Grundlösung genannt werden und daraus die Klasse (2.57) generiert werden. Ein Beispiel für diesen Spezialfall ist die Anpassung eines Polynomes an einen vorhandenen Datensatz.

Im Allgemeinfall, bei dem die Bedingung (2.56) nicht erfüllt ist, bietet sich folgende Vorgehensweise an: Da $r^2(Au, l) = r^2(k + Au, l)$, hat

$$l + v = B\tilde{u}, \quad (2.58)$$

mit

$$B = \begin{bmatrix} 1 & a_{11} & a_{12} & \cdots & a_{1m} \\ 1 & a_{21} & a_{22} & \cdots & a_{2m} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & a_{n1} & a_{n2} & \cdots & a_{nm} \end{bmatrix}, \quad \tilde{u} = \begin{bmatrix} k \\ u_1 \\ u_2 \\ \vdots \\ u_m \end{bmatrix}, \quad (2.59)$$

die gleiche Lösung (nach maximaler Korrelation) für u_1, \dots, u_m wie (2.47). Das Gleichungssystem (2.58) besitzt aber die wichtige Eigenschaft (2.34), so daß sich seine Lösung nach maximaler Korrelation aus der Lösung \tilde{f}_0 nach kleinsten Quadraten erzeugen läßt. Darüberhinaus ist f_0 (gleich wie \tilde{f}_0 , aber ohne k) eine Lösung des ursprünglichen Problems (2.47) und die Generierung der ganzen Klasse kann unmittelbar durchgeführt werden.

Interessanterweise gibt es Problemstellungen, in denen anscheinend nur die Bedingung (2.38) erfüllt ist und die Lösung nach kleinsten Quadraten trotzdem zur Klasse der Lösungen nach maximaler Korrelation gehört. Das ist genau dann der Fall, wenn die Matrizen A und B von gleichem Rang sind. Dann spielt es keine Rolle, ob die Grundlösung f_0 direkt aus (2.47) gewonnen wird, oder aus dem erweiterten System (2.58). Das erweiterte Gleichungssystem liefert bei der Ausgleichung nach kleinsten Quadraten eine singuläre Normalgleichungsmatrix, so daß eine ihrer Lösungen (nämlich diejenige bei der $k = 0$) mit der Lösung des ursprünglichen Systems übereinstimmt. Die Bedingung (2.34) ist auch in diesem Falle erfüllt, auch wenn es auf den ersten Blick nicht immer sichtbar ist. Ein Beispiel dafür ist die Helmerttransformation in der Ebene:

$$\begin{bmatrix} X \\ Y \end{bmatrix} = \begin{bmatrix} X_0 \\ Y_0 \end{bmatrix} + \begin{bmatrix} \cos \phi & -\sin \phi \\ \sin \phi & \cos \phi \end{bmatrix} \begin{bmatrix} m & 0 \\ 0 & m \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} \quad (2.60)$$

mit den Beobachtungsgleichungen

$$\begin{aligned} X_i + v_{X_i} &= 1 \cdot X_0 + x_i \cdot (m \cos \phi) - y_i \cdot (m \sin \phi) \\ Y_i + v_{Y_i} &= 1 \cdot Y_0 + y_i \cdot (m \cos \phi) + x_i \cdot (m \sin \phi) \end{aligned} \quad (i = 1, \dots, n). \quad (2.61)$$

Nach der Einführung neuer Unbekannte für $m \sin \phi$ und $m \cos \phi$ (s. [Großmann 1969]) nimmt die Matrix der Beobachtungsgleichungen bekanntlich die folgende Form an:

$$A = \begin{bmatrix} 1 & 0 & x_1 & -y_1 \\ 0 & 1 & y_1 & x_1 \\ 1 & 0 & x_2 & -y_2 \\ 0 & 1 & y_2 & x_2 \\ \vdots & \vdots & \vdots & \vdots \\ 1 & 0 & x_n & -y_n \\ 0 & 1 & y_n & x_n \end{bmatrix}. \quad (2.62)$$

Die dazugehörige erweiterte Matrix

$$B = \begin{bmatrix} 1 & 1 & 0 & x_1 & -y_1 \\ 1 & 0 & 1 & y_1 & x_1 \\ 1 & 1 & 0 & x_2 & -y_2 \\ 1 & 0 & 1 & y_2 & x_2 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & 1 & 0 & x_n & -y_n \\ 1 & 0 & 1 & y_n & x_n \end{bmatrix} \quad (2.63)$$

hat offensichtlich auch den Rang 4, weil die erste Spalte eine Linearkombination der zweiten und der dritten ist.

Die in diesem Abschnitt beschriebene Vorgehensweise ermöglicht es, bei linearen Problemen, die Klasse aller Lösungen nach maximaler Korrelation zu erzeugen, ohne Software speziell zu diesem Zweck entwerfen zu müssen. Für die zugeordnete Ausgleichung nach kleinsten Quadraten ist bereits eine Vielzahl Software vorhanden. Es sind nur der einfache Übergang zu einer solchen Ausgleichung und anschließend die Erzeugung der Lösungsklasse nach maximaler Korrelation aus der Grundlösung f_0 zu implementieren.

2.3.6 Beziehung zur Methode der kleinsten Quadrate

Aus der Diskussion in den Abschnitten 2.3.3 und 2.3.5 wurde deutlich, daß die Klasse der Lösungen nach maximaler Korrelation in vielen Fällen die Lösung nach kleinsten Quadraten enthält. Das ist genau dann der Fall, wenn das betrachtete Modell \mathbf{F} die Eigenschaft (2.34) besitzt. Die Methode der kleinsten Quadrate ist dann offensichtlich ein Spezialfall der Ausgleichung nach maximaler Korrelation. Die Bedingung $\sum pvv \rightarrow \min$ kann auch als zusätzliche Bedingung interpretiert werden, die dazu dienen kann, aus der Klasse aller Lösungen nach maximaler Korrelation eine auszuwählen. Dadurch wird die Einführung jeder anderen Zusatzbedingung ebenso denkbar, falls keine Rechtfertigung zur Minimierung von Residuenquadern vorhanden ist.

Ist nur die Bedingung (2.38) erfüllt, kann die Lösung nach kleinsten Quadraten nicht direkt als Spezialfall der Lösung nach maximaler Korrelation betrachtet werden. Zumindest bei linearen Problemen ist es jedoch in einem erweiterten Sinne möglich, s. Abschnitt 2.3.5. Dann ist die Lösung nach kleinsten Quadraten des leicht modifizierten Problems eine der Lösungen nach maximaler Korrelation.

An dieser Stelle soll ein mögliches Mißverständnis vermieden werden. Die Ausgleichung nach kleinsten Quadraten basiert auf dem Minimieren von $\sum pvv$. Dieses Minimieren läßt sich immer durch eine Suche nach einem Maximum ersetzen, die in der gleichen Lösung resultiert. Hier wird nur die Ausführung für den einfachsten Fall

$$\sum vv \rightarrow \min, \quad (2.64)$$

d.h. für eine Ausgleichung ohne Gewichte, geschildert.

Ausgehend von (1.4) gilt

$$v = f(u) - l, \quad (2.65)$$

woraus folgt

$$\begin{aligned} \sum_{i=1}^n v_i v_i &= \sum_{i=1}^n (f(u)_i - l_i)^2 \\ &= \sum_{i=1}^n (f(u)_i)^2 - 2 \sum_{i=1}^n f(u)_i l_i + \sum_{i=1}^n l_i^2 \\ &= \sum_{i=1}^n l_i^2 - \left(2 \sum_{i=1}^n f(u)_i l_i - \sum_{i=1}^n (f(u)_i)^2 \right). \end{aligned} \quad (2.66)$$

Da

$$\begin{aligned} \sum_{i=1}^n v_i v_i &> 0, \\ \sum_{i=1}^n l_i l_i &> 0, \end{aligned} \quad (2.67)$$

wobei $\sum ll$ konstant (nicht von u abhängig) und im voraus bekannt ist, ist (2.64) äquivalent mit

$$2 \sum_{i=1}^n f(u)_i l_i - \sum_{i=1}^n (f(u)_i)^2 \rightarrow \max. \quad (2.68)$$

Der Ausdruck (2.68) unterscheidet sich wesentlich von dem Korrelationskoeffizienten (2.15), so daß (2.68) bzw. (2.64) nicht mit (2.16) äquivalent ist. *Die Ausgleichung nach maximaler Korrelation liefert nicht die gleichen Lösungen wie die Ausgleichung nach der Methode der kleinsten Quadrate.* Es verbleibt nur die in diesem Abschnitt festgestellte Beziehung zwischen den beiden Ausgleichungsarten.

Kapitel 3

Anwendung der Ausgleichung nach maximaler Korrelation bei unvollständigen Modellen

In diesem Kapitel wird eine Strategie zur Anwendung der Ausgleichung nach maximaler Korrelation beschrieben, welche in häufig auftretenden Fällen, in denen keine vollständige Modellbildung von Anfang an vorhanden ist, zu einer Identifizierung der fehlenden Modellanteile beitragen kann.

Alle Betrachtungen werden auf den Fall (1.3):

$$l = \{f(u) + s_f\} + z$$

beschränkt, in dem das Datenmaterial keine groben Fehler enthält. Im Gegensatz dazu können systematische Fehler s_f vorhanden sein, so daß das aufgestellte Modell $f(u)$ nicht als vollständig betrachtet werden kann.

3.1 Allgemeine Strategie

Die ersten Ansätze zur Erarbeitung der hier vorgestellten Strategie sind in [Petrović 1995], [Petrović 1996] und [Petrović 1997] zu finden.

Ausgehend von (1.4):

$$l + v = f(u)$$

wird eine Ausgleichung nach maximaler Korrelation gemäß Definition 2.3 durchgeführt. Diese basiert nicht auf der Annahme, daß die Residuen v nur zufällige Anteile enthalten, sondern vergleicht die Formen des unvollständigen Modells mit den gegebenen Daten.

In den für die Praxis wichtigen Fällen (s. Abschnitt 2.3.3) liefert die Ausgleichung nach maximaler Korrelation keine eindeutigen Lösungen. Die Klasse aller Lösungen enthält entweder alle Funktionen der Form

$$f = c_1 + c_2 f_0, \quad c_1, c_2 \in \Re, \tag{3.1}$$

oder der Form

$$f = c_2 f_0, \quad c_2 \in \Re. \tag{3.2}$$

Dabei ist f_0 eine beliebige Lösung des Ausgleichungsproblems nach maximaler Korrelation.

Jede einzelne Lösung aus (3.1) oder (3.2) generiert einen dazugehörigen Residuenvektor

$$\varepsilon = -(s_f + z) = f - l \tag{3.3}$$

d.h.

$$\varepsilon = c_1 + c_2 f_0 - l, \tag{3.4}$$

bzw.

$$\varepsilon = c_2 f_0 - l. \tag{3.5}$$

In beiden Fällen (3.1) und (3.2) gibt es dementsprechend auch eine ganze Klasse von Residuenvektoren.

Die verbleibende Aufgabe ist es, eine der Lösungen aus (3.1) bzw. (3.2) (d.h. die reellen Zahlen c_1 und c_2 bzw. nur c_2) so auszuwählen, daß die dazugehörigen Residuen (3.4) bzw. (3.5) interpretierbar, also für die Form systematischer Fehler repräsentativ, werden. Sollte dies gelingen, dann wäre eine Modellierung systematischer Fehler und eine Erweiterung des ursprünglichen Modells denkbar. Nach der Vervollständigung des Modells könnte als *Postprocessing* eine Ausgleichung nach kleinsten Quadraten durchgeführt werden.

Eine Modellierung systematischer Fehler, die *in den diskreten Punkten* rekonstruiert wurden, setzt eine Untersuchung der möglichen, z.B. physikalischen, Ursachen voraus. Es gibt deshalb keine Garantie dafür, daß eine vollkommene Erweiterung des Modells immer erzielt werden kann. Die Vervollständigung des ursprünglichen Modells und die abschließende Ausgleichung nach kleinsten Quadraten sind somit nicht immer gewährleistet.

Unabhängig von einer eventuell erfolgreichen Modellierung systematischer Fehler und einer abschließenden Ausgleichung nach kleinsten Quadraten liefert schon die gefundene Lösung nach maximaler Korrelation, der realistische Residuen zugeordnet sind, akzeptable Modellparameter. Die Parameterschätzung könnte an dieser Stelle abgeschlossen werden¹. Rein formell lassen sich aus Residuen (die sowohl zufällige, als auch systematische Einflüsse enthalten) mittlere Fehler berechnen. Sie werden größer sein, als die aus einer Ausgleichung nach kleinsten Quadraten resultierenden mittleren Fehler². Da weder ein stochastisches Modell für derartige Residuen vorliegt, noch vorliegen kann, werden die berechneten mittleren Fehler nicht statistisch oder wahrscheinlichkeitstheoretisch interpretierbar sein³. Trotzdem werden ihre Größenordnungen für die *Abweichung des unvollständigen Modells von den Daten* besser repräsentativ sein, als die unter einer falschen Annahme⁴ berechneten mittleren Fehler. Statistisch interpretierbare Fehlerschätzungen werden erst nach der Identifizierung der systematischen Einflüsse, Vervollständigung des Modells und einer abschließenden Ausgleichung nach kleinsten Quadraten möglich.

Wie soll aber die Auswahl einer Lösung erfolgen, die repräsentative Residuen generiert?

3.2 Residuenanalyse

Aus der Residuenklasse (3.4) oder (3.5) soll ein „repräsentativer“ Residuenvektor ausgewählt werden.

Die einfachste Idee wäre, „alle“ Residuenvektoren zu erzeugen und darzustellen. Danach könnte die Entscheidung getroffen werden, welcher von diesen Vektoren am besten einen möglichen systematischen Einfluß repräsentiert. In der durch (3.4) oder (3.5) beschriebenen Klasse gibt es aber unendlich viele Residuenvektoren, somit ist die geschilderte Idee streng genommen nicht realisierbar. Daten haben immer nur eine begrenzte Genauigkeit und die Lösung wird ebenso innerhalb eines bestimmten Intervalls und einer gewissen Genauigkeit gesucht, daher läßt sich die vorgeschlagene Vorgehensweise trotzdem durchführen, indem die reellen Zahlen c_1 und c_2 nur in einem begrenztem Bereich und mit einer realistischen Schrittweite variiert werden. Diese Methode wurde in [Neitzel 1998] erfolgreich angewandt.

¹Diese Tatsache ist besonderes wichtig im Falle, daß eine physikalische Modellierung (oder eine andere, die auf Ursachen basiert) nicht erfolgreich oder nur unvollständig durchgeführt werden kann.

²Die Methode der kleinsten Quadrate ist nichts anderes, als die Methode, die, unabhängig von den statistischen oder anderen Eigenschaften der Beobachtungen, immer minimale mittlere Fehler liefert, s. z.B. [Helmert 1924], [Linnik 1961] oder [Moritz 1966]. Ob diese Anforderung im konkreten Fall der eigentlichen Zielsetzung entspricht, ist eine andere Frage.

³Vom statistischen Standpunkt aus können sowohl die so berechneten Fehlerschätzungen, als auch die, die aus einer Ausgleichung nach kleinsten Quadraten bei Ignorieren der Unvollständigkeit des Modells hervorgehen würden, nur als gefälscht betrachtet werden. In [Koch 1997] wurde gezeigt, daß auch beim Vorhandensein „systematischer Fehler“ eine korrekte Fehlerfortpflanzung möglich ist. Dabei soll beachtet werden, daß im zitierten Artikel nur solche „systematische Fehler“ betrachtet werden, die aus dem vorhandenen oder zusätzlichen Datenmaterial (z.B. Kalibrierungsmessungen) innerhalb einer Gesamtausgleichung bestimmt werden können. Die Modellbildung ist dementsprechend vollständig und der Begriff „systematische Fehler“ wird in einem anderen Sinne als in der vorliegenden Arbeit benutzt. Es ist unmittelbar einsichtig, daß solche „systematische Fehler“, für die ein ausreichendes Modell im voraus vorhanden ist, keine Schwierigkeiten bereiten können und mit den unbekannten systematischen Einflüssen, die erst identifiziert werden sollen, wenig zu tun haben.

⁴daß keine systematischen Fehler im Datenmaterial vorkommen

Die Wahl eines Residuenvektors basierend auf der oben beschriebenen Vorgehensweise beinhaltet gewisse Vorstellungen über „systematisches Verhalten“. Aus diesem Grund soll versucht werden, diese Vorstellungen zumindest teilweise mathematisch zu beschreiben, um die Menge der zu betrachtenden Residuenvektoren einzuschränken. Gelingt dies, dann wäre es kein Problem, die Wahl des entsprechenden Residuenvektors automatisch oder fast automatisch (mit Hilfe eines Rechnerprogramms) durchzuführen.

Darüber hinaus bietet sich die Idee an, die Klasse der Residuenvektoren überhaupt nicht zu generieren, sondern die mathematische Beschreibung des systematischen Verhaltens als eine zusätzliche Bedingung zu benutzen, um die reellen Zahlen c_1 und c_2 zu ermitteln.

Bei der mathematischen Formulierung des Begriffes „systematisch“ gibt es gleichartige Schwierigkeiten wie bei der mathematischen Beschreibung der „groben Fehler“. Die Definition von beiden Begriffen basiert meistens auf *Ursachen*⁵, vergleiche Abschnitt 1.2. Die Ursachen bleiben in jedem konkreten Fall unbekannt, weshalb eine exakte mathematische Beschreibung nicht möglich ist.

Beim Auftreten grober Fehler kann das Problem so gelöst werden, daß statt grober Fehler Ausreißer mathematisch definiert werden. Diese Definition ignoriert die Ursachen und nimmt als Kriterium den Betrag und die Verteilung der Widersprüche. Nach der Festlegung eines Grenzwertes (was weder eindeutig noch ganz unproblematisch ist) ermöglicht eine passende Ausgleichung⁶ die Identifizierung der Ausreißer. Die gefundenen Ausreißer werden dann zu groben Fehlern erklärt. Eine Trennung der übriggebliebenen „kleinen groben Fehler“ von den zufälligen Fehlern ist auf diese Weise nicht möglich. Deshalb wird darauf verzichtet.

Die gleiche Idee ist bei systematischen Fehlern anwendbar. Die unbekannten Ursachen können nicht mathematisch beschrieben werden, sondern nur gewisse Eigenschaften der Residuen. In diesem Falle reicht es aber nicht aus, nur den Betrag der Residuen zu betrachten, obwohl auch ihm eine Bedeutung zukommt. Sollten die Residuen nach einer Ausgleichung⁷ „klein“ (relativ zur Datengenauigkeit) und zufällig verteilt⁸ sein, dann kann keine weitere Interpretation gerechtfertigt sein. Finden sich in den Residuen einige wenige Ausreißer von beträchtlichem Betrag, dann liegt es nahe, diese als Ausdruck grober Fehler zu betrachten und entsprechende Daten zu entfernen. Weiterhin bleibt noch der Fall zu betrachten, in dem die Residuen interpretierbare Beträge haben, aber keine zufällige Verteilung und keine eindeutigen Ausreißer aufweisen. Nach der Durchführung der im Abschnitt 1.3 beschriebenen Untersuchungen kann von nichtmodellierten systematischen Einflüssen, d.h. systematischen Fehlern, ausgegangen werden.

Angenommen, die Residuen sind für systematische Fehler repräsentativ, d.h.

$$\varepsilon \doteq -(s_f + z). \quad (3.6)$$

Sie sind dann aus zwei Komponenten zusammengesetzt, einer vom Betrag kleineren unregelmäßigen Komponente z , und der „systematischen“ Komponente s_f . Gemäß allgemein akzeptierter Vorstellungen über eine Systematik kann von dem Anteil s_f ein „regelmäßiger“ Verlauf erwartet werden. An erster Stelle könnte es entweder ein *glatter*⁹ oder ein *treppenförmiger* Verlauf sein¹⁰. Dementsprechend wird in der Residuenklasse (3.4) bzw. (3.5) entweder ein Residuenvektor gesucht, welcher am besten einen glatten Verlauf (in Zeit oder Raum, abhängig von der Natur des Problems) aufweist, oder ein Residuenvektor, welcher eine Treppenstruktur am besten repräsentiert.

⁵Eine der Ausnahmen ist in [Reissmann 1962] zu finden, wo der Autor die Eigenschaften und nicht die Ursachen in Vordergrund stellt: „Die zweite Gattung sind die *systematischen Fehler*, die meist von bestimmten äußeren Einflüssen abhängen und das Meßergebnis in regelmäßiger, oft gesetzmäßiger Weise beeinflussen.“ In [Helmert 1924] basiert die Definition systematischer Fehler auf den Ursachen, die Eigenschaften werden aber auch betont, indem diese Fehler „regelmäßige Fehler“ genannt werden.

⁶z.B. sogenannte „robuste Verfahren“

⁷An dieser Stelle wird nicht unbedingt eine Ausgleichung nach kleinsten Quadraten gemeint, sondern eine beliebige Art der Ausgleichung oder Residuenerzeugung, z.B., aber auch nicht ausschließlich, eine Ausgleichung nach maximaler Korrelation.

⁸z.B. normalverteilt

⁹Der Begriff „glatt“ wird hier nicht unbedingt nur in einem streng mathematischen Sinne der Stetigkeit von Ableitungen verwendet. Es handelt sich eher um die Vorstellung eines nicht sehr oszillierenden, ruhigen Verlaufs.

¹⁰Eine Kombination aus diesen beiden Eigenschaften ist ebenso denkbar. Es ist aber nicht auszuschließen, daß systematische Einflüsse auch andere Merkmale haben könnten. In der vorliegenden Arbeit wird die Betrachtung auf die beiden genannten Fälle eingeschränkt.

3.2.1 Aufstellen zusätzlicher Bedingungen

Um einen glatten oder treppenförmigen Verlauf der Residuen mathematisch beschreiben zu können, müssen die Residuen, d.h. einzelne Komponenten des Residuenvektors als Werte einer Funktion der Zeit oder der Lage im Raum betrachtet werden. Im einfachsten Falle einer Zeitreihe ist es nicht schwierig, die Residuen als eine Funktion der Zeit zu interpretieren. Äquivalent beziehen sich die Residuen in einem geodätischen Netz auf Punkte, denen die Lage im Raum zugeordnet ist. Die räumlichen Änderungen der Residuen könnten wiederum zu einer Funktion zusammengefaßt werden. Bei anderen Beispielen sollte eine ähnliche Interpretation zu finden sein. Die ausgewählte Funktion ist dementsprechend von der Natur des Problems abhängig. Auf diese Weise werden die Komponenten des Residuenvektors (für feste Werte c_1 und c_2) die Werte einer Funktion

$$\epsilon : \Re^k \longrightarrow \Re \quad (3.7)$$

in entsprechenden Zeit- oder Raumpunkten t_i :

$$\varepsilon_i = \epsilon(t_i), \quad t_i \in \Re^k. \quad (3.8)$$

Wenn nun die reellen Zahlen c_1 und c_2 nicht mehr als fest betrachtet werden, dann hängt die Funktion ϵ ebenfalls von diesen beiden Parametern ab:

$$\epsilon : \Re^{k+2} \longrightarrow \Re, \quad (3.9)$$

bzw.

$$\epsilon = \varepsilon(c_1, c_2, t_i). \quad (3.10)$$

Als zusätzliche Bedingung zur Auswahl einer Lösung wird die maximale „Glattheit“ oder „Treppenförmigkeit“ der Funktion ϵ verlangt.

Ein Maß für den glatten Verlauf kann die integrale Krümmung:

$$K = \int \kappa^2 dt, \quad (3.11)$$

sein, wobei die Krümmung κ auf übliche Weise als Gauß'sche Krümmung der der Funktion (3.9) zugehörigen Hyperfläche definiert wird. Ob die auf diese Weise im streng mathematischen Sinne definierte Krümmung wirklich in allen Fällen einen „ruhigen“ Verlauf der Residuen am besten repräsentiert, läßt sich nicht ohne weiteres entscheiden. Andere, vereinfachte, Definitionen einer Krümmung sind genauso denkbar. Die einfachste oder „primitivste“ dieser basiert auf der Annahme, daß die benachbarten Residuen beim Vorhandensein glatter systematischer Fehler gemeinsame Anteile enthalten. Daraus ergeben sich Differenzen zwischen benachbarten Residuen als Maß für eine Art Krümmung. Beide Extremfälle der Einführung von Glattheit werden im Abschnitt 3.2.2 zusammen mit weiteren Zwischenvarianten im einfachsten, eindimensionalen Fall detailliert ausgearbeitet.

Die zusätzliche Bedingung zur Bestimmung von c_1 und c_2 lautet somit:

$$K \rightarrow \min. \quad (3.12)$$

Die Treppenförmigkeit der Funktion (3.9) manifestiert sich in ihren ersten Ableitungen, die überall (dort, wo sie definiert sind) den Wert 0 annehmen und an den Bruchstellen überhaupt nicht existieren.

Die beiden zusätzlichen Bedingungen (für die Glattheit bzw. für die Treppenförmigkeit) verlangen die Ableitungen der Funktion ϵ . Diese Funktion ist aber unbekannt. Es stehen nur ihre diskreten Werte (3.8) zur Verfügung, daher lassen sich die zusätzlichen Bedingungen in obiger Form nicht realisieren. Realisierbar ist jedoch die Differenzbildung zwischen „benachbarten“ Residuen, woraus eine Näherung für die Ableitungen berechnet werden kann. Im eindimensionalen Fall können benachbarte Residuen eindeutig definiert werden. In mehrdimensionalen Fällen können z.B. alle Punktpaare innerhalb einer Kugel in einem n -dimensionalen Raum als benachbart betrachtet werden. Da immer mit unregelmäßig verteilten Punkten zu rechnen ist, ist eine Gewichtung der berechneten Näherungen für die Ableitungen, bzw. für die Krümmung sinnvoll. Die Näherungen für die Ableitungen sind um so besser, je näher die benutzten Punkte beieinander liegen. Die Gewichte für erste Ableitungen sollen dementsprechend mit wachsender Entfernung abnehmen, sie können also als Funktionen der reziproken Entfernung d_{ij} zwischen den Punkten i und j eingeführt werden:

$$\bar{p}_{ij} = \bar{p}_{ij} \left(\frac{1}{d_{ij}} \right). \quad (3.13)$$

Die einfachste solche Möglichkeit ist

$$\bar{p}_{ij} = \frac{1}{d_{ij}}. \quad (3.14)$$

Andererseits repräsentieren die aus den nahe gelegenen Punkten berechneten Näherungen für die Ableitungen eher lokale, kleinräumige Schwankungen der Residuen, die bei der Suche nach systematischen Fehlern weniger von Interesse sind. In praktischen Anwendungen sind immer neben systematischen Einflüssen noch zufällige (unregelmäßige) Fehler in den Residuen vorhanden. In denen wird diejenige Komponente gesucht, die den globalen „ruhigen“ Verlauf darstellt. Aus diesem Grund bietet sich zusätzlich eine weitere Gewichtung der berechneten Näherungen für die Ableitungen, nämlich so eine, bei der die Gewichte für erste Ableitungen mit wachsender Entfernung wachsen:

$$\underline{p}_{ij} = \underline{p}_{ij} (d_{ij}), \quad (3.15)$$

z.B.:

$$\underline{p}_{ij} = d_{ij}. \quad (3.16)$$

Aus (3.13) und (3.15) ergibt sich:

$$p_{ij} = \bar{p}_{ij} \underline{p}_{ij}. \quad (3.17)$$

Im Spezialfall, (3.14) mit (3.16), folgt:

$$p_{ij} = 1. \quad (3.18)$$

Wenn keine zufälligen Fehler vorhanden sind, verliert der Gewichtsfaktor (3.15) offensichtlich an Bedeutung, falls die zufälligen Anteile erheblichere Beträge annehmen (im Vergleich mit der systematischen Komponente), wächst auch die Bedeutung dieses Faktors. Ohne zusätzliche Informationen darüber, wie stark die beiden Komponenten (systematische und zufällige¹¹) sind, kann das „richtige“ Verhältnis zwischen den beiden Gewichtsfaktoren, (3.13) und (3.15) nicht bestimmt werden. Die zusätzlichen Informationen können aber nicht vorhanden sein, die Aufgabe besteht gerade darin die beiden Komponenten zu finden.

Aus den numerischen Untersuchungen im eindimensionalen Fall (vgl. Kapitel 4) hat sich (3.18) als ein vernünftiger *Kompromiß* erwiesen.

Nun läßt sich die Glattheitsbedingung folgendermaßen definieren. Aus jedem Punktpaar werden Näherungen für die ersten partiellen Ableitungen zwischen zwei Punkten berechnet. Zu diesen gehören die Gewichte p_{ij} . Aus allen ersten Ableitungen, die sich in einem Punkt treffen, können die ersten und die zweiten Ableitungen in einzelnen Punkten gewichtet berechnet werden und daraus die Krümmungen κ_i mit den entsprechenden Gewichten p_i . Die Bedingung (3.12) kann dann die Form

$$K = S(c_1, c_2) = \sum p_i \kappa_i^2 \longrightarrow \min \quad (3.19)$$

annehmen. Die Suche nach dem Minimum der Funktion $S(c_1, c_2)$ ergibt die Werte der reellen Zahlen c_1 und c_2 , für die die Residuen aus der Klasse (3.4) einen möglichst glatten Verlauf annehmen. Die Residuen enthalten auch die zufällige Komponente, deswegen wird dieser Verlauf nie perfekt glatt sein. Erwartet wird aber, daß aus ihm die Formen der systematischen Fehler erkannt werden können.

Die gleichen Probleme stellen sich auch bei einer mathematischen Beschreibung der Treppenbedingung. Die Ableitungen werden nur näherungsweise aus diskreten Punktkombinationen berechnet, deshalb werden die ersten Ableitungen nie (sogar wenn keine zufälligen Fehler vorhanden sind) den oben beschriebenen theoretischen horizontalen Verlauf aufweisen. In den Punktzwischenräumen, wo keine Bruchstellen liegen, die die einzelnen Treppen trennen, sind deshalb nur kleine Beträge der ersten Ableitungen zu erwarten. Dort wo die Grenze zwischen zwei Treppen verläuft, sollten erhebliche Beträge auftauchen. Angenommen, daß jede Treppe mehrere Datenpunkte enthält, so daß größere Beträge der ersten Ableitungen relativ selten auftreten, wandelt sich das Problem in die Suche nach Ausreißerstellen in den ersten Ableitungen. Die Auflösung des Problems kann dann beispielsweise mit den bekannten robusten Verfahren vorgenommen werden.

In den Fällen, in denen beide Arten systematischer Einflüsse gleichzeitig zu erwarten sind, bietet sich an, diese der Reihe nach zu identifizieren. Offensichtlich sind zuerst die „Treppen“ zu finden und zu modellieren, danach kann eine Anwendung der Glattheitsbedingung (in einer neuen Ausgleichung

¹¹ oder, in geometrischer Terminologie, die mit dem ruhigen Verlauf und die unregelmäßige

nach maximaler Korrelation) die Parameterschätzung abschließen. Zusätzlich kann noch ein *Postprocessing* versucht werden, um systematische Anteile auf Ursachen zurückzuführen und vollständig zu modellieren.

Das Hauptproblem bei der beschriebenen Vorgehensweise besteht darin, eine einheitliche mathematische Formulierung der zusätzlichen Bedingung für den allgemeinsten Fall zu finden. Es ist allerdings unklar, ob es das überhaupt geben kann. Deshalb kann die endgültige Lösung dieses Problems in dieser Arbeit noch nicht angeboten werden. Der in diesem Abschnitt vorgestellte Entwurf kann als Grundlage für eine Erweiterung der mathematischen Formulierung der zusätzlichen Bedingung dienen.

Ein von diesem Vorschlag etwas abweichender Weg zur Analyse der Residuenklasse im zweidimensionalen Fall (speziell bezogen auf die Anwendung in der Deformationsanalyse) wurde in [Neitzel 1999] vorgeschlagen.

Der einfachste, eindimensionale Fall einer auf der Ausgleichung nach maximaler Korrelation basierenden Residuenanalyse wird im nächsten Abschnitt ausführlicher behandelt.

3.2.2 Eindimensionaler Fall

Im einfachsten, eindimensionalen Fall, können die beiden betrachteten Arten der zusätzlichen Bedingungen in einer ziemlich allgemeinen Form mathematisch präzisiert werden, ohne die spezielle Natur des Problems berücksichtigen zu müssen.

Wenn ein glatter Verlauf der Funktion (3.4) bzw. (3.5) gesucht wird, bildet die übliche mathematische Definition der Krümmung

$$\kappa = \frac{\varepsilon''}{(1 + \varepsilon'^2)^{\frac{3}{2}}} \quad (3.20)$$

einen möglichen Ausgangspunkt für die Aufstellung der Zusatzbedingung.

Vorhanden sind jedoch nur n diskrete Werte der Funktion ε :

$$\varepsilon_i = \varepsilon(t_i) = c_1 + c_2 f_0(t_i) - l_i \quad (i = 1, \dots, n), \quad (3.21)$$

mit

$$t_i < t_{i+1} \quad \forall i \in \{1, \dots, n-1\}. \quad (3.22)$$

Daraus folgt:

$$\varepsilon' \left(\frac{t_{i+1} + t_i}{2} \right) = \delta_i \doteq \frac{\varepsilon_{i+1} - \varepsilon_i}{t_{i+1} - t_i} \quad (i = 1, \dots, n-1) \quad (3.23)$$

und

$$\varepsilon''(t_j) = \delta'(t_j) \doteq \frac{\delta_i - \delta_{i-1}}{\Delta t_i} \quad (i = 2, \dots, n-1). \quad (3.24)$$

Unter Berücksichtigung, daß sich der δ_i -Wert auf den Punkt $\frac{t_{i+1} + t_i}{2}$ und δ_{i-1} auf $\frac{t_i + t_{i-1}}{2}$ bezieht, folgt

$$\Delta t_i = \frac{t_{i+1} - t_{i-1}}{2} \quad (i = 2, \dots, n-1). \quad (3.25)$$

Es bleibt zu entscheiden, welchem Punkt t_j der Wert $\varepsilon''(t_j)$ zugeordnet werden soll. Es bietet sich an, einen Punkt zwischen $\frac{t_i + t_{i-1}}{2}$ und $\frac{t_{i+1} + t_i}{2}$ auszuwählen, wobei beide als unterschiedlich gewichtet zu betrachten sind. Eine Benutzung der Gewichte (3.14) resultiert in:

$$t_j = \frac{\frac{1}{t_{i+1} - t_i} \frac{t_{i+1} + t_i}{2} + \frac{1}{t_i - t_{i-1}} \frac{t_i + t_{i-1}}{2}}{\frac{1}{t_{i+1} - t_i} + \frac{1}{t_i - t_{i-1}}}. \quad (3.26)$$

Nach einer kleinen Umordnung ergibt sich

$$t_j = t_i, \quad (3.27)$$

womit die Gleichung (3.24) ihre endgültige Form annimmt:

$$\varepsilon''_i = \varepsilon''(t_i) = \delta'_i \doteq \frac{\delta_i - \delta_{i-1}}{\Delta t_i} \quad (i = 2, \dots, n-1). \quad (3.28)$$

Um den Ausdruck für die Krümmung zu bilden, werden *die ersten und die zweiten Ableitungen in gleichen Punkten* benötigt. Vorhanden sind bisher nur die Werte

$$\varepsilon' \left(\frac{t_{i+1} + t_i}{2} \right) \quad (i = 1, \dots, n-1) \quad (3.29)$$

und

$$\varepsilon''(t_i) \quad (i = 2, \dots, n-1), \quad (3.30)$$

s. (3.23) und (3.28). Eine der beiden Größen muß umgerechnet werden, z.B.

$$\varepsilon'(t_i) = \frac{\frac{1}{t_{i+1} - t_i} \varepsilon' \left(\frac{t_{i+1} + t_i}{2} \right) + \frac{1}{t_i - t_{i-1}} \varepsilon' \left(\frac{t_i + t_{i-1}}{2} \right)}{\frac{1}{t_{i+1} - t_i} + \frac{1}{t_i - t_{i-1}}}, \quad (3.31)$$

woraus folgt

$$\varepsilon'_i = \varepsilon'(t_i) = \frac{(\varepsilon_{i+1} - \varepsilon_i)(t_i - t_{i-1})^2 + (\varepsilon_i - \varepsilon_{i-1})(t_{i+1} - t_i)^2}{(t_{i+1} - t_{i-1})(t_{i+1} - t_i)(t_i - t_{i-1})}. \quad (3.32)$$

Für gleichabständige Daten nehmen (3.32) und (3.28) die erwartete einfache Form an:

$$\left. \begin{aligned} t_{i+1} - t_i &= \Delta t \quad \forall i \in \{1, \dots, n-1\}, \\ \varepsilon'_i &= \varepsilon'(t_i) \doteq \frac{\varepsilon_{i+1} - \varepsilon_{i-1}}{2\Delta t}, \\ \varepsilon''_i &= \varepsilon''(t_i) \doteq \frac{\varepsilon_{i+1} - 2\varepsilon_i + \varepsilon_{i-1}}{(\Delta t)^2}. \end{aligned} \right\} \quad (3.33)$$

Unter Benutzung von (3.28) und (3.32) lässt sich die Krümmung in einzelnen Punkten darstellen:

$$\kappa_i = \kappa(t_i) \doteq \frac{\varepsilon''_i}{(1 + \varepsilon_i'^2)^{\frac{3}{2}}} \quad (i = 2, \dots, n-1). \quad (3.34)$$

Eine mögliche Alternative dazu bietet die häufig benutzte noch einfachere Näherung

$$\kappa_i = \kappa(t_i) \doteq \varepsilon''_i \quad (i = 2, \dots, n-1), \quad (3.35)$$

bei der lediglich die zweite Ableitung als ein Maß für die Krümmung angesehen wird. Diese Betrachtungsweise ist in der Literatur sehr verbreitet.

Die Ableitungen werden numerisch aus diskreten, nicht immer sehr dicht beieinander liegenden Punkten berechnet, wobei die ursprünglich bestimmte erste Ableitung noch auf entsprechende Punkte umgerechnet wird. Deshalb muß die einfachere Approximation (3.35) nicht unbedingt schlechter als die theoretisch bessere Näherung (3.34) sein.

Dazu kommen noch die im Abschnitt 3.2.1 erörterten Argumente. Je exakter (im mathematischen Sinne) die Definition einer Krümmung ist, desto besser repräsentiert sie die lokale Glattheit und die kleinräumigen Schwankungen der Residuen. Dadurch ist sie möglicherweise weniger geeignet, um systematisches Verhalten zu identifizieren.

Aus diesen Gründen bieten sich die zusätzlichen, mathematisch nicht ganz korrekten, intuitiv aber akzeptablen, noch einfacheren Beschreibungen der Glattheit an. Die integrale Krümmung ist offensichtlich kleiner, wenn der Verlauf der Residuen weniger oszilliert, deshalb liegt es nahe, die Neigungsänderung als repräsentativ für die Krümmung einzuführen:

$$\kappa_i = \kappa(t_i) \doteq \Delta \varepsilon' = \varepsilon' \left(\frac{t_{i+1} + t_i}{2} \right) - \varepsilon' \left(\frac{t_i + t_{i-1}}{2} \right) \quad (i = 2, \dots, n-1). \quad (3.36)$$

Diese Näherung hat den gleichen Vorteil wie (3.35), sie bedarf auch keiner Umrechnung der ersten Ableitungen.

Nicht nur die Neigungsänderung, sondern sogar die Neigungen selbst sind für die Glattheit repräsentativ. Natürlich nicht im Sinne der Neigungen in einzelnen Punkten, sondern als integrale

Neigung, die bei einer erheblich schwankender Residuenkurve einen größeren Betrag annehmen wird. Somit bietet sich

$$\kappa_i = \varepsilon' \left(\frac{t_{i+1} + t_i}{2} \right) \quad (i = 1, \dots, n - 1) \quad (3.37)$$

an als nächste, noch einfachere Möglichkeit.

Letztendlich (vgl. Abschnitt 3.2.1) ist die einfachste Vorstellung von der Bedeutung systematischer Einflüsse die, daß die benachbarten Residuen eine gemeinsame Komponente enthalten. Dann sollen ihre Differenzen kleiner sein als dann, wenn keine systematischen Anteile vorhanden sind. Somit soll auch die Möglichkeit

$$\kappa_i = \kappa \left(\frac{t_i + t_{i+1}}{2} \right) \doteq \Delta \varepsilon = \varepsilon(t_{i+1}) - \varepsilon(t_i) \quad (i = 1, \dots, n - 1). \quad (3.38)$$

berücksichtigt werden.

Unabhängig davon, welche der fünf Definitionen (3.34) (3.35), (3.36), (3.37) oder (3.38) für die Krümmung in einzelnen Punkten akzeptiert wird, kann die Glattheitsbedingung in Form

$$K = S(c_1, c_2) = \sum p_i \kappa_i^2 \longrightarrow \min \quad (3.39)$$

formuliert werden.

Zusammenfassend ergeben sich folgende fünf Möglichkeiten:

$$\sum p(\Delta \varepsilon)^2 \longrightarrow \min, \quad (3.40)$$

$$\sum p(\varepsilon')^2 \longrightarrow \min, \quad (3.41)$$

$$\sum p(\Delta \varepsilon')^2 \longrightarrow \min, \quad (3.42)$$

$$\sum p(\varepsilon'')^2 \longrightarrow \min, \quad (3.43)$$

$$\sum p \frac{\varepsilon''^2}{(1 + \varepsilon'^2)^3} \longrightarrow \min. \quad (3.44)$$

Welche von diesen die besten Aussichten zur Identifizierung systematischer Einflüsse bietet, lässt sich nicht ohne weiteres entscheiden. Eine Einsicht bieten die in dem Kapitel 4 durchgeführten numerischen Untersuchungen.

Unabhängig davon, welche der Näherungen (3.38), (3.37), (3.36), (3.35) oder (3.34) für die Krümmung benutzt wird, steigt offensichtlich die Genauigkeit der Approximation κ_i für die Werte $\kappa(t_i)$ mit sinkendem Punktabstand der benutzten Punkte t_j und mit einer Steigerung der Regelmäßigkeit ihrer Verteilung.

Die Regelmäßigkeit der Datenverteilung hat noch eine Konsequenz. Im Grenzfall, wenn die Daten völlig regelmäßig, d.h. gleichabständig verteilt sind, sind die Bedingungen (3.40) und (3.41) äquivalent und führen auf dieselbe Lösung. Bei äquidistanten Daten haben nämlich alle $t_{i+1} - t_i$ den gleichen Betrag:

$$t_{i+1} - t_i = \Delta t \quad \forall i \in \{1, \dots, n - 1\} \quad (3.45)$$

und somit

$$\sum p_i (\varepsilon'_i)^2 = \sum p_i \left(\frac{\Delta \varepsilon_i}{\Delta t} \right)^2 = \frac{1}{(\Delta t)^2} \sum p_i (\Delta \varepsilon_i)^2. \quad (3.46)$$

Auf eine ähnliche Weise lässt sich zeigen, dass bei gleichabständigen Daten auch die Bedingungen (3.42) und (3.43) äquivalent sind.

Die Frage der Gewichte p_i wurde schon im Abschnitt 3.2.1 diskutiert. Einerseits sind die einzelnen Werte κ_i , die aus näher gelegenen Werten berechnet wurden, genauer. Das führt auf die Gewichte (3.13) oder speziell (3.14). Andererseits sind die so berechneten Werte für den globalen Verlauf der Residuen weniger repräsentativ, was zusätzliche Gewichte gemäß (3.15), oder speziell (3.16) induziert. Aus beiden ergibt sich (3.17), wobei

$$p_i = 1 \quad (3.47)$$

als ein vernünftiger Kompromiss betrachtet werden kann.

Nach der Aufstellung der Bedingung (3.39), d.h. einer der Bedingungen (3.40)–(3.44), könnte grundsätzlich die Bestimmung der reellen Zahlen c_1 und c_2 als eine zusätzliche einfache Ausgleichung nach der Methode der kleinsten Quadrate interpretiert werden¹². Tatsächlich kann aber nur c_2 ermittelt werden, c_1 bleibt weiterhin unbestimmt¹³. Aus (3.23) und (3.24) folgt nämlich daß $\Delta\varepsilon$, ε' , $\Delta\varepsilon'$ und ε'' nur von c_2 abhängen, obwohl ε eine Funktion von beiden Parametern ist¹⁴. Dann sind gemäß (3.34)–(3.38) auch alle κ_i nur von c_2 abhängig und deshalb

$$K = S(c_1, c_2) = S(c_2) = \sum p_i \kappa_i^2 \rightarrow \min. \quad (3.48)$$

Dieses Ergebnis verwundert wenig; eine Verschiebung um den Betrag c_1 kann keinen Einfluß auf die Krümmung und auf die Identifizierung der Formen der systematischen Fehler haben.

An dieser Stelle ist noch anzumerken, daß die Bedingung (3.44) zu einer zusätzlichen nichtlinearen Ausgleichung nach kleinsten Quadraten führt. Im Gegensatz dazu führen die Bedingungen (3.40)–(3.43) zu linearen Ausgleichungsproblemen. Diese Tatsache ist eine zusätzliche Rechtfertigung ihrer Einbeziehung in die Betrachtung¹⁵.

Ausgehend von den Überlegungen in Abschnitt 3.2.1 kann auch die Treppenbedingung für den eindimensionalen Fall formalisiert werden.

Die Parameter c_1 und c_2 (eigentlich wieder nur c_2 , da eine Verschiebung um c_1 keinen Einfluß auf die Formen hat) sollen so bestimmt werden, daß die ersten Ableitungen, bis auf wenige Ausreißer, möglichst kleine Beträge annehmen. Eigentlich werden diese Ausreißer gesucht, da sie die Grenzen von einzelnen Treppen definieren. Dazu bietet sich die in solchen Situationen häufig benutzte Minimierung der L_1 -Norm:

$$\sum p_j |\varepsilon'_j| \rightarrow \min, \quad (3.49)$$

$$\varepsilon'_j = \varepsilon' \left(\frac{t_{i+1} + t_i}{2} \right) \quad (i = 1, \dots, n-1; j = i). \quad (3.50)$$

Als eine einfachere Alternative zur Ausreißersuche in den Ableitungen bietet sich an, den Sprung in den ε -Werten als repräsentativ für die Treppen zu betrachten und die Bedingung (3.49) durch

$$\sum p_j |\Delta\varepsilon_j| \rightarrow \min \quad (l \leq 1), \quad (3.51)$$

$$\Delta\varepsilon_j = \varepsilon(t_j) - \varepsilon(t_{j-1}) \quad (i = 1, \dots, n-1; j = i) \quad (3.52)$$

zu ersetzen.

In beiden Fällen (3.49) und (3.51) stellt sich wieder die Frage der Gewichtsauswahl, wobei die gleiche Argumentation gilt, wie im Falle der Glattheitsbedingung. Einerseits sind die entsprechenden Größen (ε'_j bzw. $\Delta\varepsilon_j$), die aus näher gelegenen Werte berechnet wurden, genauer, was auf Gewichte (3.13) oder speziell (3.14) führt. Andererseits sind die so berechneten Werte für den globalen Verlauf der Residuen weniger repräsentativ, was zusätzliche Gewichte gemäß (3.15), oder speziell (3.16) induziert. Aus beiden Argumenten ergibt sich (3.17) und $p_i = 1$ kann auch in diesem Falle als ein vernünftiger Kompromiß betrachtet werden.

Nach der Aufstellung der Bedingung (3.49) oder (3.51) bedarf es nur noch einer entsprechenden Ausgleichung, um den Parameter c_2 zu bestimmen.

Nachdem entweder mit Hilfe der Glattheits- oder der Treppenbedingung (bzw. durch Anwendung der beiden) der Wert von c_2 bestimmt wurde, liefert (3.1) bzw. (3.2) die gesuchte Lösung, die im erstgenannten Falle die noch unbestimmte Konstante c_1 enthält.

¹²Im Falle (3.1) werden tatsächlich c_1 und c_2 gesucht, wobei im Falle (3.2) nur c_2 als Unbekanntes erscheint.

¹³Im Falle (3.1) ist die zugeordnete Ausgleichsaufgabe singulär.

¹⁴bei einer Lösungsklasse der Form (3.1)

¹⁵Rechtfertigung, weil bei einer „Lösung“ eines nichtlinearen Ausgleichungsproblems, welches nur mit numerischen Methoden gelöst werden kann, *nie* zu behaupten ist, daß es sich wirklich um eine Lösung handelt, die die Zielfunktion global minimiert.

Kapitel 4

Synthetische Beispiele

Die Anwendungsmöglichkeiten der Ausgleichung nach maximaler Korrelation unter Berücksichtigung zusätzlicher Bedingungen beim Vorhandensein systematischer Fehler werden in diesem Kapitel an einfachen eindimensionalen synthetischen Beispielen illustriert. Bei derartigen Beispielen ist die „Wahrheit“ im voraus bekannt, so daß ein Vergleich der Ergebnisse mit dem vorgegebenen „Sollzustand“ möglich ist. Es wird sich aber auch zeigen, daß in gewissen Konstellationen gar keine eindeutige „Wahrheit“ vorhanden sein kann.

4.1 Beispiel ohne zufällige Fehler

Das erste Beispiel ist in Abb. 4.1 dargestellt. Das unvollständige Modell besteht aus einer einfachen Sinuskurve

$$f(t) = A + B \sin(\omega t + C).$$

Diese Gleichung läßt sich problemlos in die folgende Form überführen:

$$f(t) = a \sin(\omega t) + b \cos(\omega t) + c. \quad (4.1)$$

Diese Form ist für die Durchführung einer Ausgleichung bekanntlich besser geeignet. Der Parameter ω , der die Frequenz beschreibt, wird als bekannt angenommen. Die Aufgabe besteht in der Bestimmung der unbekannten Modellparameter a , b und c . Das Vorhandensein systematischer Einflüsse wird in Form einer Parabel angenommen, die zufälligen Fehler werden zunächst zu Null gesetzt (s. Abb. 4.1). Weiterhin seien die Daten gleichabständig verteilt.

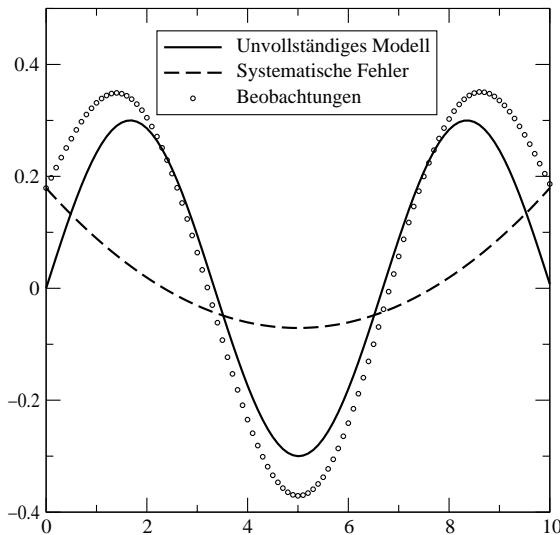


Abbildung 4.1: Beispiel 1 – Zusammensetzung der synthetischen Daten

Wird die Unvollständigkeit des Modells ignoriert und eine Ausgleichung nach kleinsten Quadraten

(LSA¹) durchgeführt, ergeben sich erheblich verfälschte Modellparameter (vgl. die ersten zwei Zeilen in Tabelle 4.1). Die dazugehörigen Residuen sind für die systematischen Einflüsse nicht repräsentativ, s. Abb. 4.2. Trotzdem ist aus diesen Residuen ersichtlich, daß die Daten systematisch verfälscht sind. Eine formelle Überprüfung dieser Behauptung ist mit den Methoden möglich, die im Abschnitt 1.3 dargestellt wurden.

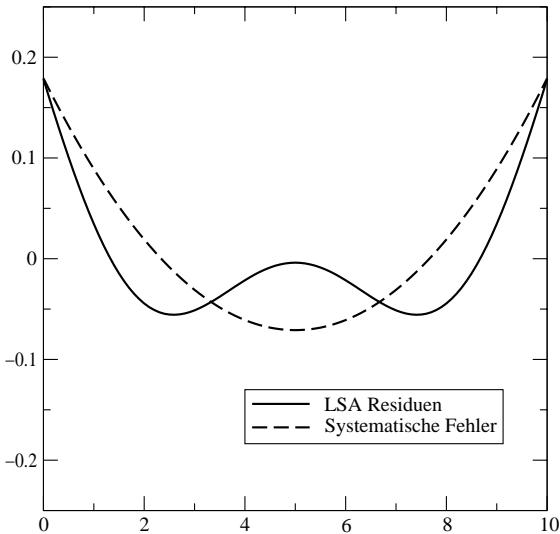


Abbildung 4.2: Beispiel 1 – Vergleich postulierter systematischer Fehler mit den Residuen aus der Ausgleichung nach kleinsten Quadraten

Eine Anwendung der Ausgleichung nach maximaler Korrelation gemäß Kapitel 3 resultiert in einer ganzen Klasse von Lösungen, die in diesem Falle auch die Lösung nach kleinsten Quadraten als Spezialfall beinhaltet, vgl. Abschnitt 2.3.3. Die dazugehörigen Residuen für einige der Lösungen sind in Abb. 4.3 dargestellt. Die Klasse enthält offensichtlich auch solche Residuensätze, die sich von den eingebauten tatsächlichen systematischen Einflüssen nicht wesentlich unterscheiden.

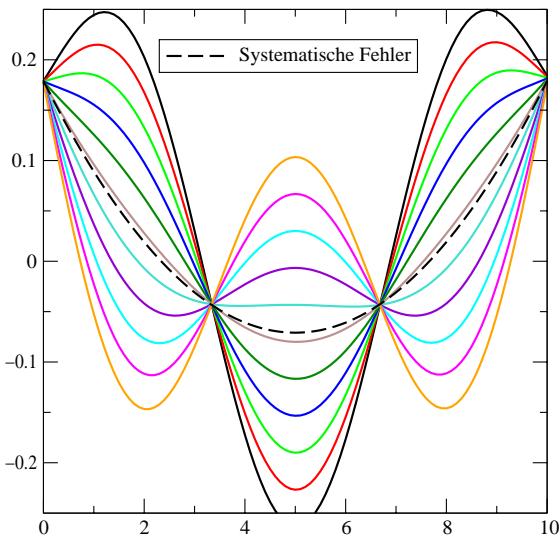


Abbildung 4.3: Beispiel 1 – Residuen für unterschiedliche Lösungen nach maximaler Korrelation

Die Anwendung eines Glattheitskriteriums kann nun dazu dienen, eine der Lösungen nach maximaler Korrelation auszuwählen. Im Abschnitt 3.2.2 wurden fünf derartige Kriterien erörtert. Für die fünf auf diese Weise ausgewählten Lösungen werden folgende Bezeichnungen² eingeführt:

¹LSA = Least Squares Adjustment

²Auch die schon eingeführten zwei Abkürzungen (MCA = Ausgleichung nach maximaler Korrelation und LSA = Ausgleichung nach kleinsten Quadraten) werden in diesem Kapitel an mehreren Stellen benutzt.

- GLATT1 = die Lösung basierend auf Minimierung der Quadratsumme der Residuendifferenzen, (3.40),
- GLATT2 = die Lösung basierend auf Minimierung der Quadratsumme der ersten Ableitungen, (3.41),
- GLATT3 = die Lösung basierend auf Minimierung der Quadratsumme der Differenzen der ersten Ableitungen, (3.42),
- GLATT4 = die Lösung basierend auf Minimierung der Quadratsumme der zweiten Ableitungen, (3.43),
- GLATT5 = die Lösung basierend auf Minimierung der Quadratsumme der Krümmungen, (3.44).

Im betrachteten Beispiel sind die Datenpunkte gleichabständig verteilt, so daß jeweils zwei Kriterien in gleichen Ergebnissen resultieren (GLATT1=GLATT2; GLATT3=GLATT4), vergleiche Abschnitt 3.2.2. Somit ergeben sich insgesamt drei voneinander unterschiedliche Lösungen (GLATT1, GLATT3, GLATT5). Die resultierenden Parameter des unvollständigen Modells sind in Tabelle 4.1 aufgelistet (die Verschiebung c bleibt dabei unbestimmt³, vergleiche Abschnitt 3.2.2). Die Residuensätze, die diesen drei Lösungen entsprechen, sind in Abb. 4.4 dargestellt.

Tabelle 4.1: Beispiel 1: Parameter des unvollständigen Modells

	a	b	c
Soll	0.300	0.000	0.000
LSA	0.367	0.001	0.000
GLATT1=GLATT2	0.291	0.001	0.000
GLATT3=GLATT4	0.290	0.001	0.000
GLATT5	0.264	0.001	0.000

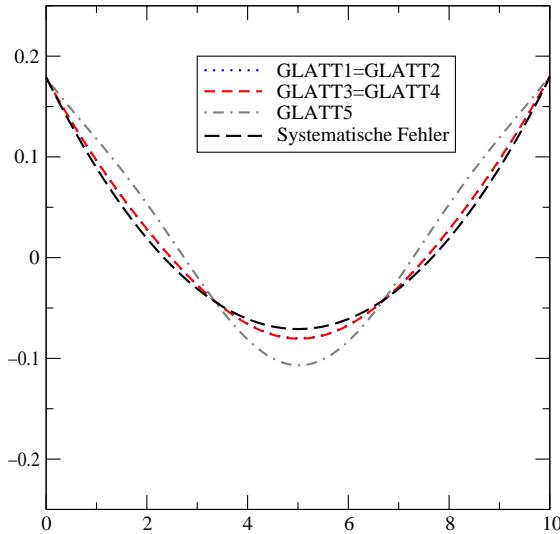


Abbildung 4.4: Beispiel 1 – Vergleich postulierter systematischer Fehler mit den Residuen für die drei ausgewählten Lösungen nach maximaler Korrelation

Zwei der drei Lösungen (GLATT1 und GLATT3) sind fast identisch; die dritte (GLATT5) weicht etwas ab. Alle drei Lösungen liefern bessere Parameterschätzungen für das unvollständige Modell und repräsentativere Residuen, als die Lösung nach kleinsten Quadraten. Die beiden fast identischen

³Der Parameter c kann problemlos nachträglich bestimmt werden, indem die schon bestimmten Parameter a und b festgehalten werden. Die in Tabelle 4.1 aufgeführten Werte für c wurden auf diese Weise ermittelt.

Lösungen (in Abb. 4.4 sind die entsprechenden Residuensätze nicht visuell zu unterscheiden) ermöglichen eine bessere Rekonstruktion der Komponenten des synthetischen Beispiels als die dritte. Das scheint kein Zufall zu sein, das letzte Kriterium (GLATT5), welches auf einer streng mathematischen Beschreibung⁴ der Krümmung basiert, hat sich bei den meisten (in der Arbeit angeführten und zusätzlich getesteten) Beispielen als ziemlich instabil erwiesen. Darüberhinaus war die Konvergenz bei der iterativen Berechnung (diese Formulierung der Glattheitsbedingung führt zu einem nichtlinearen Ausgleichungsproblem, vergleiche Abschnitt 3.2.2) häufig extrem langsam.

4.2 Einführung zufälliger Fehler

Als nächstes werden die Daten aus dem ersten Beispiel übernommen und um zufällige normalverteilte Anteile erweitert, s. Abb. 4.5. Alle Ergebnisse und Schlußfolgerungen sind praktisch gleich wie im Beispiel ohne zufällige Fehler. Zum Vergleich sind in Abb. 4.6 auch die Residuen aus der Ausgleichung nach kleinsten Quadraten unter Nichtberücksichtigung der Unvollständigkeit des Modells dargestellt. Auf eine graphische Darstellung der Residuenklasse nach maximaler Korrelation kann verzichtet werden. Bei Anwendung der Glattheitskriterien werden drei verschiedene, aber nicht wesentlich unterschiedliche Lösungen erhalten, die dazugehörigen Residuensätze sind ebenso in Abb. 4.6 dargestellt. Der einzige feststellbare Unterschied zu Abb. 4.2 und Abb. 4.4 ist das Vorhandensein zufälliger Fehler.

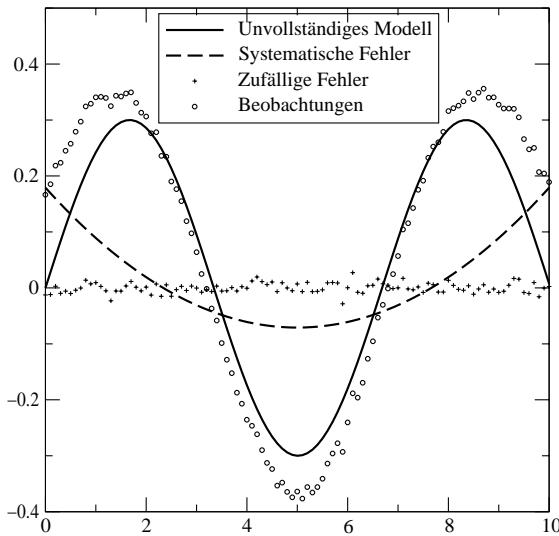


Abbildung 4.5: Beispiel 2 – Zusammensetzung der synthetischen Daten

4.3 Einfluß von Datenlücken

Bisher war die Datenverteilung gleichabständig. Als nächstes bietet sich an, eine unregelmäßige Datenverteilung zu berücksichtigen. Gleichzeitig wurde die Anzahl der Datenpunkte halbiert, s. Abb. 4.7.

Die Ergebnisse der Ausgleichung nach kleinsten Quadraten haben sich dadurch nicht wesentlich geändert, weder bei der Parameterschätzung, noch bezüglich der Form des Residuensatzes. Das gleiche gilt für die Klasse aller Lösungen nach maximaler Korrelation.

Die Anwendung der Glattheitskriterien liefert nun aber etwas geänderte Ergebnisse⁵. Die Datenverteilung ist diesmal unregelmäßig und die fünf unterschiedlichen Kriterien resultieren in fünf unterschiedlichen Lösungen. Die dazugehörigen Residuensätze sind in Abb. 4.8 dargestellt.

⁴Diese Beschreibung verwandelt sich bei der Anwendung in eine *scheinbar* strenge Beschreibung, weil nur diskrete Daten verfügbar sein können.

⁵Wie im Abschnitt 3.2.2 diskutiert, ist eine Einführung der Gewichte möglich, die durch die fehlende Gleichabständigkeit der Daten induziert sind. Es gibt keine eindeutigen Kriterien, um diese auszuwählen. Numerische Experimente mit Gewichteveränderungen haben die Argumentation aus Abschnitt 3.2.2 bestätigt, daß gleiche Gewichte (3.47) als ein akzeptabler Kompromiß betrachtet werden können. Deshalb sind im folgenden Ergebnisse dargestellt, die auf dieser Wahl basieren.

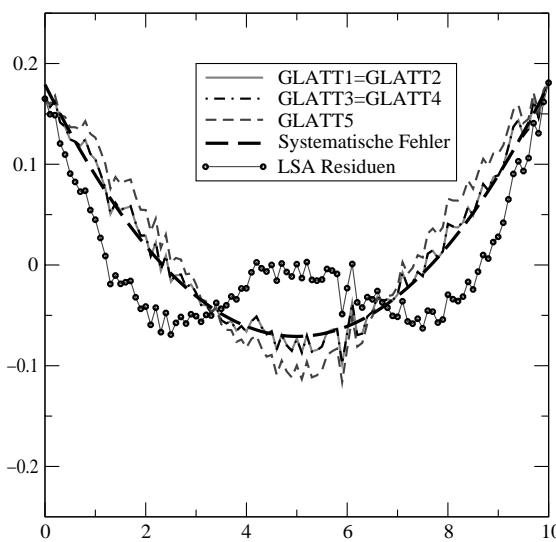


Abbildung 4.6: Beispiel 2 – Vergleich postulierter systematischer Fehler mit den Residuen der Lösung nach kleinsten Quadraten und denen der drei ausgewählten Lösungen nach maximaler Korrelation

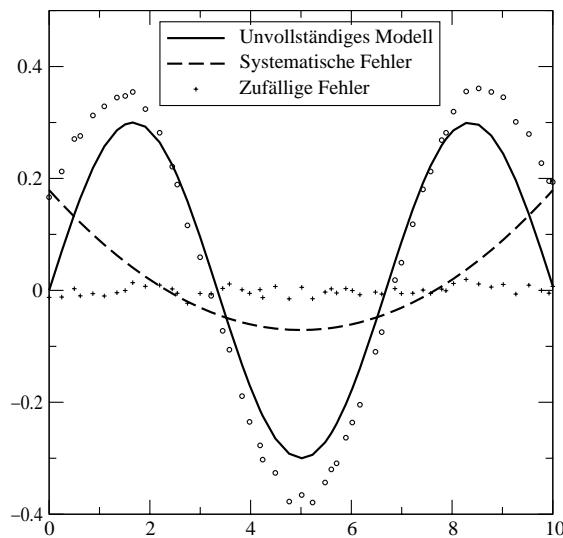


Abbildung 4.7: Beispiel 3 – Zusammensetzung der synthetischen Daten

Die beiden ersten Kriterien (basierend auf Differenzen von benachbarten Residuen, GLATT1, bzw. auf den ersten Ableitungen, GLATT2) liefern fast identische Parameter für das unvollständige Modell, welche die Soll-Werte sehr gut reproduzieren (nicht schlechter als das erste Kriterium in Tabelle 4.1). Die beiden dazugehörigen Residuensätze sind in Abb. 4.8 kaum zu unterscheiden und zeigen eine sehr gute Übereinstimmung mit den vorgegebenen systematischen Fehlern.

Die dritte Lösung (GLATT3, basierend auf Differenzen der ersten Ableitungen) weicht von den ersten beiden geringfügig ab, liefert aber ähnlich gute Schätzungen für die Modellparameter und für die – die Systematik enthaltenden – Residuen.

Das vierte Kriterium (GLATT4, zweite Ableitungen) resultiert schon in erheblich abweichenden Ergebnissen: Die Parameterschätzung ist in ihrer Qualität der bei der Anwendung einer Ausgleichung nach kleinsten Quadraten auf das unvollständige Modell gleichzusetzen, die Form des Residuensatzes scheint jedoch repräsentativer zu sein.

Das fünfte Kriterium (GLATT5, Krümmung im exakt mathematischen Sinne) liefert in diesem Falle viel zu kleine Werte für die Modellparameter, woraus die Residuen resultieren, die mit dem unvollständigen Modell stark korreliert sind.

Um das Verhalten der unterschiedlichen Glattheitskriterien noch deutlicher zu veranschaulichen, wird nun bei gleichem unvollständigen Modell und systematischen Anteilen die Unregelmäßigkeit der Datenverteilung wesentlich verstärkt, so daß auch größere Datenlücken entstehen, s. Abb. 4.9.

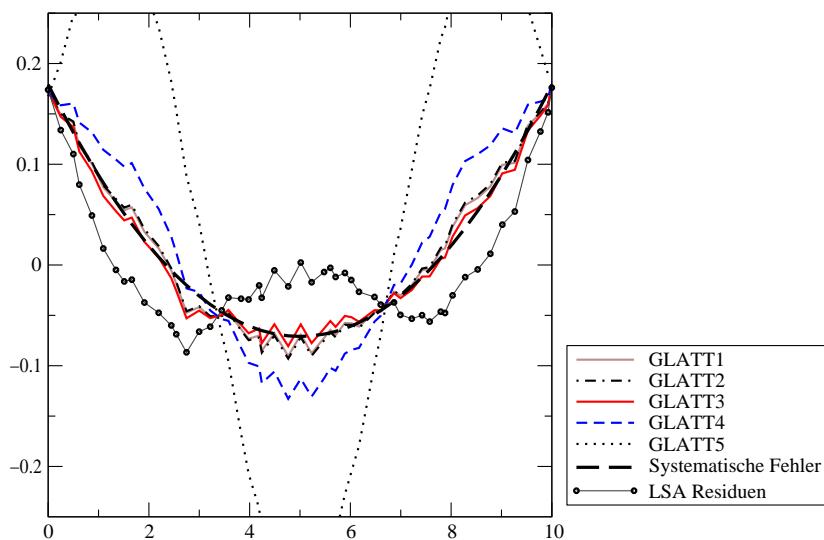


Abbildung 4.8: Beispiel 3 – Vergleich postulierter systematischer Fehler mit den Residuen der Lösung nach kleinsten Quadraten und denen der fünf ausgewählten Lösungen nach maximaler Korrelation



Abbildung 4.9: Beispiel 4 – Verteilung der Datenpunkte

Genauso wie im vorherigen Beispiel, haben sich weder die Ergebnisse der Ausgleichung nach kleinsten Quadrate noch die Klasse aller Lösungen nach maximaler Korrelation wesentlich geändert.

Die Auswirkungen der zusätzlichen „Glattheitsbedingungen“ sind aber nicht mehr gleich. Die ersten zwei Kriterien (Residuendifferenzen und erste Ableitungen, GLATT1 und GLATT2) liefern noch immer wesentlich bessere Parameterschätzungen und repräsentativere Residuensätze als die Methode der kleinsten Quadrate, die restlichen drei Kriterien (GLATT3, GLATT4 und GLATT5) resultieren in schlechten Schätzungen für die Parameter des unvollständigen Modells und in zu großen Residuen, die mit dem unvollständigen Modell stark korreliert sind, s. Abb. 4.10.

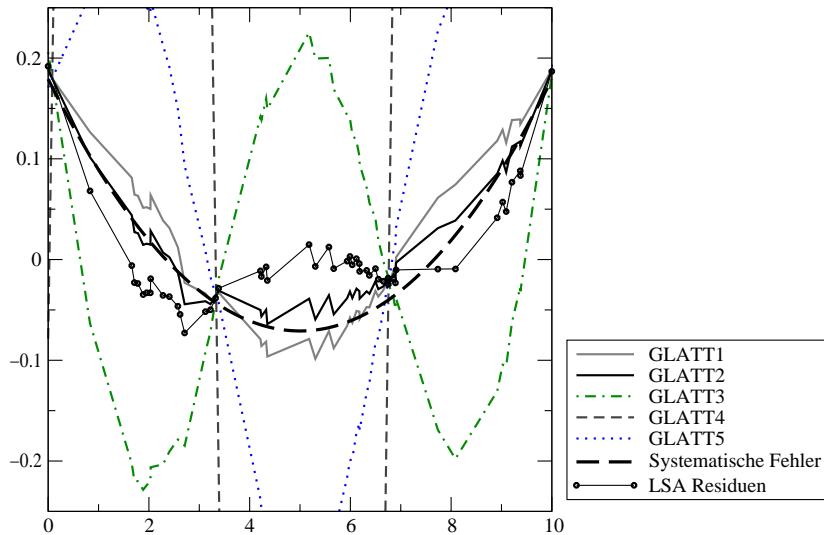


Abbildung 4.10: Beispiel 4 – Vergleich postulierter systematischer Fehler mit den Residuen der Lösung nach kleinsten Quadraten und denen der fünf ausgewählten Lösungen nach maximaler Korrelation

Dieses Ergebnis ist nicht nur für das ausgewählte Beispiel charakteristisch, sondern für alle untersuchten Beispiele. Von den aufgestellten fünf Kriterien verhalten sich die anscheinend „primitiveren“ stabiler und entsprechen offensichtlich besser der Vorstellung der systematischen Anteile.

Die Lösung GLATT1 entspricht der Idee, daß benachbarte Residuen beim Vorhandensein sys-

matischer Einflüsse gemeinsame Anteile haben sollen und damit der Vorstellung einer „langwelligen Glattheit“. Diese Lösung ist deshalb nicht besonders empfindlich auf kurzwellige, lokale Schwankungen. Die andere extreme Möglichkeit, GLATT5, die auf einer exakten mathematischen Beschreibung der Krümmung basiert, repräsentiert auch die Glattheit in einem entsprechenden Sinne, deshalb reagiert sie auf kleine lokale Schwankungen sehr empfindlich. Die restlichen drei untersuchten Möglichkeiten bilden den Übergang zwischen diesen beiden Extremfällen.

Da in praktischen Anwendungen immer mit zufälligen Fehleranteilen und somit mit lokalen Schwankungen zu rechnen ist, hat die erste, einfachste, weniger empfindliche Bedingung die größte Bedeutung⁶.

Trotzdem sind die empfindlicheren, komplizierteren Bedingungen nicht endgültig auszuschließen. Unter Umständen (z.B. wenn zufällige Anteile sehr klein sind und die Datenverteilung sehr regelmäßig ist) können sie auch bessere Lösungen liefern, als die einfachste Bedingung. Da aber die einfachste Lösung gleichzeitig die stabilste ist, kommen die komplizierteren Bedingungen nur dann in Frage, wenn ihre Ergebnisse grundsätzlich mit denen der einfachen übereinstimmen.

In den vier bisher vorgestellten Varianten eines einfachen synthetischen Beispiels funktioniert die auf der maximalen Korrelation basierende vorgeschlagene Vorgehensweise recht gut und liefert sowohl bessere Schätzungen für die Parameter des unvollständigen Modells als auch aussichtsreichere Möglichkeiten zur Identifizierung systematischer Fehler als eine Ausgleichung nach kleinsten Quadrataten.

Als nächstes ist zu klären, ob ähnliche Verhältnisse bei anderen denkbaren Anwendungsbeispielen zu erwarten sind.

Die Datenverteilung in Abb. 4.11 ist noch unregelmäßiger und das unvollständige Modell besteht nun aus mehreren Frequenzen, s. Abb. 4.12.

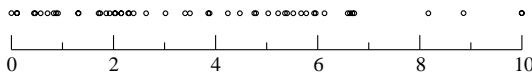


Abbildung 4.11: Beispiel 5 – Verteilung der Datenpunkte

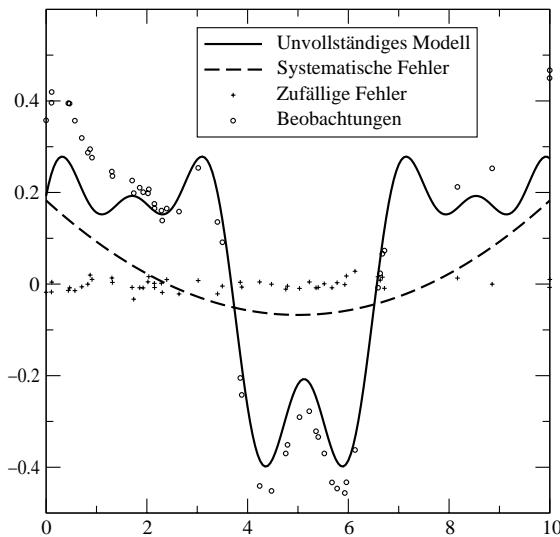


Abbildung 4.12: Beispiel 5 – Zusammensetzung der synthetischen Daten

Die Lösung GLATT1 aus der Klasse aller Lösungen nach maximaler Korrelation liefert noch immer die besten Schätzungen für das unvollständige Modell, GLATT2 verhält sich nur etwas besser als die Lösung nach kleinsten Quadrataten, die restlichen drei Glattheitskriterien resultieren in schlechten Parameterschätzungen und in Residuen, die eine sehr starke Korrelation mit dem unvollständigen Modell aufweisen. Diese Umstände werden am besten durch Residuenbilder veranschaulicht, s. Abb. 4.13.

⁶Eigentlich repräsentiert sie auch am besten das, was unter einem „glatten“ oder „ruhigen“ Verlauf der übriggebliebenen systematischen Anteile intuitiv zu verstehen sei.

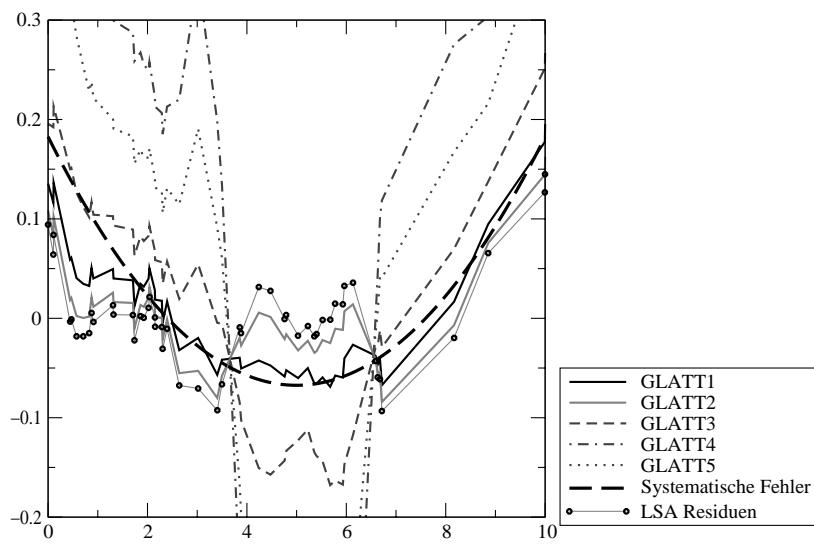


Abbildung 4.13: Beispiel 5 – Vergleich postulierter systematischer Fehler mit den Residuen der Lösung nach kleinsten Quadraten und denen der fünf ausgewählten Lösungen nach maximaler Korrelation

Weiterhin ist zu bemerken, daß selbstverständlich auch solche Beispiele konstruierbar sind, in denen trotz Unvollständigkeit des Modells eine Ausgleichung nach kleinsten Quadraten ebenfalls gute Näherungen für die Soll-Parameter und repräsentative Residuen liefert. In solchen Fällen erwiesen sich keine praktischen Unterschiede zwischen der Lösung nach kleinsten Quadraten und der Lösung durch Anwendung eines Glattheitskriteriums auf die aus der Ausgleichung nach maximaler Korrelation resultierende Klasse.

4.4 Nicht-eindeutige Aufgabenstellungen

Schließlich stellt sich die Frage, ob es überhaupt generell möglich ist, eine „absolute Wahrheit“ aus dem vorhandenen Datenmaterial zu extrahieren. Bei synthetischen Beispielen ist dieses Problem mit der Frage äquivalent, ob eine Methode denkbar ist, die immer genau das herausfiltert, was bei der Konstruktion des Beispiels eingebaut wurde. Bei praktischen Anwendungen lautet diese Frage, ob es immer möglich ist, vollständig das zu rekonstruieren, was die unbekannten (z.B. physikalischen) Ursachen zu den Daten beigetragen haben.

Dieses Problem wird wieder an einem sehr einfachen synthetischen Beispiel dargestellt, s. Abb. 4.14. Einfachheitshalber wurden zufällige Anteile zu Null gesetzt, das unvollständige Modell hat die Form einer simplen Parabel und die systematischen Einflüsse die einer stückweise geraden Linie.

Weder die Lösung nach kleinsten Quadraten, noch die Lösungen nach maximaler Korrelation, die aus der Klasse aller Lösungen mittels Glattheitsbedingungen ausgesucht wurden, liefern besonders gute Schätzungen für die Parameter des unvollständigen Modells. Die dazugehörigen Residuen veranschaulichen diese Tatsache, s. Abb. 4.15. Alle dargestellten Residuensätze haben einen weniger steilen Verlauf als die vorgegebenen systematischen Einflüsse.

Nun stellt sich die Frage, ob die Anwendung der maximalen Korrelation mit zusätzlichen Glattheitsbedingungen in diesem Falle erfolglos war. Die Antwort ist positiv, vorausgesetzt, daß die Entstehungsweise der synthetisch erzeugten Beobachtungen als „absolute Wahrheit“ betrachtet wird.

Hingegen muß bei weitergehender Interpretation der unterschiedlichen Ergebnisse (s. Abb. 4.15) festgestellt werden, daß die gleichen Beobachtungen auch auf eine andere Weise entstehen können, nämlich wie in Abb. 4.16 dargestellt. Ausgehend von dieser Zusammensetzung der synthetischen Daten läßt sich die Lösung GLATT3 (Abb. 4.15) als eine perfekte Rekonstruktion der Komponenten des synthetischen Beispiels betrachten.

Das unvollständige Modell und die systematische Komponente sind in diesem Falle nicht eindeutig trennbar. Daher kann es auch keine mathematisch beschreibbare Methode geben, die diejenige Zerlegung liefert, die *erwünscht ist*, sondern nur eine Lösung, die gewisse mathematisch beschreibbare Eigenschaften hat. Die Zusammensetzung in Abb. 4.16 entspricht einfach besser der Vorstellung,

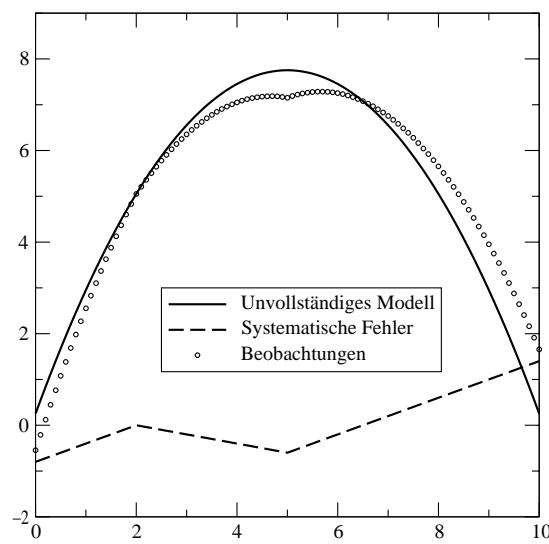


Abbildung 4.14: Beispiel 6 – Zusammensetzung der synthetischen Daten

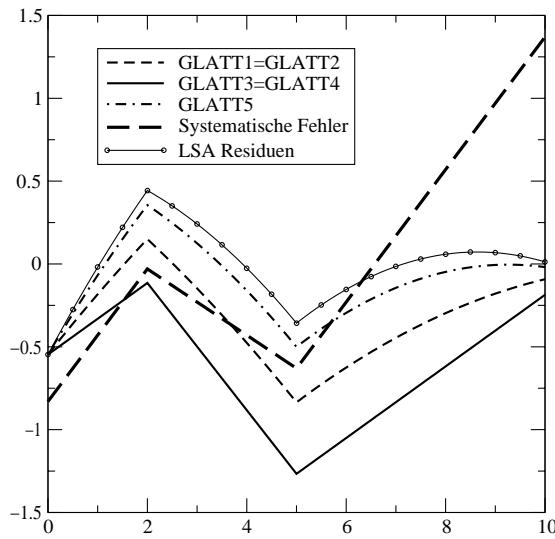


Abbildung 4.15: Beispiel 6 – Vergleich postulierter systematischer Fehler mit den Residuen der Lösung nach kleinsten Quadraten und denen der fünf ausgewählten Lösungen nach maximaler Korrelation

daß die systematischen Anteile einen „*ruhigen Verlauf*“ haben sollen, als die Zusammensetzung in Abb. 4.14.

Dementsprechend ist weder von der beschriebenen Auswertestrategie noch von irgendeiner anderen zu erwarten, daß sie auf eine „mystische“ Weise immer die „Wahrheit“ findet. Dies trifft insbesondere auf praktische Anwendungen zu, vor allem, wenn das unvollständige Modell und die systematischen Einflüsse gegensätzliche Anteile enthalten. Dann kann es nicht möglich sein, allein aufgrund der Daten eine perfekte Rekonstruktion der einzelnen Komponenten durchzuführen.

Trotzdem liefert die beschriebene Strategie eine Möglichkeit, solche Zerlegungen zu finden, die in einem gewissen Sinne plausibel sind – in dem Sinne, daß die Residuen einen „*ruhigen Verlauf*“ annehmen. Erst eine weitgehende Analyse der erhaltenen Residuenformen kann eine mögliche Erklärung für die gefundenen Formen liefern, indem mögliche Ursachen in Betracht gezogen werden.

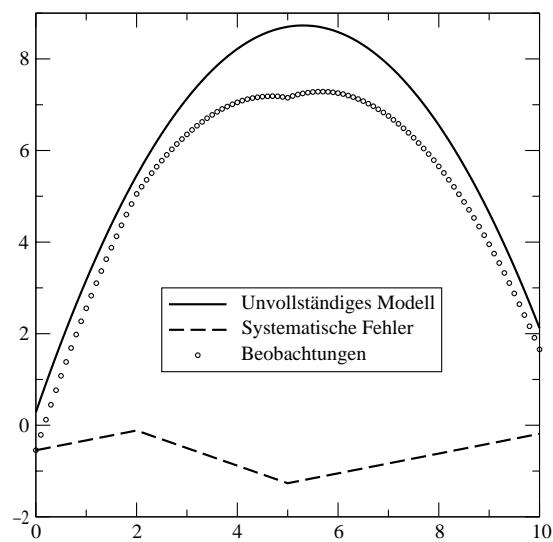


Abbildung 4.16: Beispiel 6 – Mögliche andere Zusammensetzung der synthetischen Daten

Kapitel 5

Schlußbetrachtung

Das Vorhandensein systematischer Fehler, welches in dieser Arbeit im Sinne der Unmöglichkeit, ein vollständiges Modell aufstellen zu können, interpretiert wird, bereitet große Schwierigkeiten bei der Bearbeitung nicht nur geodätischer Daten. Ein Ignorieren der Tatsache, daß eine Modellbildung nicht vollständig ist, resultiert in der Aufstellung von Annahmen, die mit dem eigentlichen Problem im Widerspruch stehen. Fast alle in der Literatur dokumentierten Methoden zur Bearbeitung geodätischer Daten (speziell Ausgleichung) basieren jedoch auf der Annahme, daß die Beobachtungen einen gewissen stochastischen Charakter aufweisen (d.h., daß keine systematischen Fehler bzw. Modellunzulänglichkeiten vorkommen). Auf diese Weise gelingt es immer, eindeutige Lösungen zu finden. Der Bezug des Modells zum ursprünglichen Problem wird aber fragwürdig und die Ergebnisse für die Modellparameter sind oftmals so verfälscht, daß aus den dazugehörigen Residuen gar nicht ersichtlich ist, wie groß diese Verfälschungen sind und wie die enthaltenen systematischen Einflüsse aussehen könnten.

Die zu lösenden zwei Probleme, die sich bei einer unvollständigen Modellbildung ergeben, sind von unterschiedlichem Schwierigkeitsgrad.

Das erste, kleinere Problem besteht darin, den Nachweis zu erbringen, daß im vorhandenen Datenmaterial systematische Einflüsse enthalten sein könnten. Die bestehende Literatur bietet mehrere geeignete Möglichkeiten, die beispielsweise auf den aus einer Ausgleichung nach kleinsten Quadraten resultierenden Residuen beruhen und die als ausreichend angesehen werden können. Diese Methoden bieten aber keine Möglichkeiten zur Analyse von Einzelresiduen, sondern nur eine Antwort auf die Frage, ob unbekannte systematische Anteile vermutet werden können.

Das zweite Problem ist wesentlich schwieriger: Die Form oder andere Eigenschaften der in den Residuen enthaltenen systematischen Einflüsse zu erkennen, um mögliche Ursachen und Erklärungen zu finden, was dann zu einer Erweiterung des ursprünglichen Modells führen könnte. Erst nach einer ausreichenden Vervollständigung der Modellbildung werden die statistischen Annahmen völlig unbedenklich und ermöglichen als Postprocessing eine endgültige Lösung des Problems, beispielsweise durch eine Ausgleichung nach kleinsten Quadraten.

Die in der Literatur vorgeschlagenen Lösungen für das zweite Problem müssen als nicht ausreichend betrachtet werden. Leider wird es auch nie eine vollständige Lösung dieses Problems geben. Wie in der vorliegenden Arbeit an einem sehr einfachen synthetischen Beispiel illustriert, ist die in den Daten enthaltene „einzige Wahrheit“ selbst nicht immer eindeutig. Die gleichen Daten können manchmal auf mehrere unterschiedliche Weise in die Komponenten unvollständiges Modell, fehlende systematische Einflüsse und kleine zufällige Anteile zerlegt werden. Welche der Zerlegungen als „Wahrheit“ betrachtet werden soll, ist von der Vorstellungen über „systematische Fehler“ abhängig. Intuitiv sind diese Vorstellungen klar und größtenteils allgemein akzeptiert. Zwischen diesen Vorstellungen und jedem Versuch, diese exakt mathematisch zu beschreiben, bleibt immer eine Grauzone, die manchmal vernachlässigbar und manchmal erheblich sein kann.

Die vorliegende Arbeit versucht weder die Tatsache zu bestreiten, daß unter gewissen Umständen die vorhandenen Methoden eine erfolgreiche Identifizierung der im voraus unbekannten systematischen Anteile ermöglichen, noch ein „Patentrezept“ anzubieten, das ein im allgemeinen unlösbare Problem immer erfolgreich löst. Statt dessen handelt es sich um eine Erweiterung der zur Verfügung stehenden Methoden zur Analyse derjenigen Probleme, bei denen keine vollständige Modellbildung von Anfang an möglich ist.

Als eine Möglichkeit wird in der vorliegenden Arbeit die Anwendung einer Ausgleichung nach maximaler Korrelation vorgeschlagen. Diese Ausgleichung wird ohne statistische Annahmen rein geometrisch begründet. Die Parameter des unvollständigen Modells werden so bestimmt, daß die in diesem Modell enthaltenen Formen so gut wie möglich an die in den Daten enthaltenen Formen angepaßt werden. Dies geschieht, indem der Korrelationskoeffizient zwischen dem unvollständigen Modell und den Daten maximiert wird. Auf diese Weise entsteht in allen für die Anwendung relevanten Fällen eine ganze Klasse der Lösungen und die dazugehörige Klasse der Residuensätze.

Die Ausgleichung nach maximaler Korrelation kann als eine Verallgemeinerung der Ausgleichung nach kleinsten Quadraten betrachtet werden. Falls nämlich die Klasse der Funktionen, in der die Lösung gesucht wird, gewisse einfache Eigenschaften hat, was meistens auch der Fall ist, dann enthält die Unterklasse der Lösungen nach maximaler Korrelation speziell auch die Lösung nach der Methode der kleinsten Quadrate. Im zweiten, für die Anwendung mitunter interessanten Fall, ist die Lösung nach kleinsten Quadraten eines leicht modifizierten Problems eine der Lösungen nach maximaler Korrelation.

Die auf den ersten Blick unerwünschte Mehrdeutigkeit der Lösung nach maximaler Korrelation ist bei der erörterten Zielsetzung jedoch von großem Vorteil: Die Lösungsklasse enthält nicht nur die Lösung, bei der Residuen zum Minimum gezwungen wurden, sondern auch solche, deren Residuen keine Minimalitätsbedingung erfüllen und statt dessen den Verlauf der systematischen Anteile besser repräsentieren. Um eine solche Lösung aus der ganzen Klasse auszuwählen, sind Vorstellungen über die Eigenschaften dieser Anteile notwendig. Intuitiv ist unmittelbar einsichtig welche Vorstellungen geeignet sind. Eine exakte mathematische Beschreibung des Begriffes „systematisch“ im allgemeinen kann es aber nicht geben.

In der vorliegenden Arbeit wurden zwei Aspekte der „Systematik“ herausgegriffen: mögliche (langwellige) Glattheit im weitesten Sinne des Wortes („ruhiger Verlauf“) und die „Treppenförmigkeit“. Für diese zwei Fälle wurden entsprechende mathematische Beschreibungen ausgearbeitet. Diese können dann benutzt werden, um aus der ganzen Residuenklasse die Lösung auszuwählen, die derartige systematische Anteile am besten repräsentiert. Für den einfachsten, eindimensionalen Fall wurde die Ausarbeitung dieser zusätzlichen Bedingungen detailliert durchgeführt und an synthetischen Beispielen verifiziert. Es hat sich gezeigt, daß die vorgeschlagene Vorgehensweise Möglichkeiten bietet, die Form der enthaltenen systematischen Anteile „realistisch“ zu identifizieren, auch dann, wenn die Methoden, die die Residuen minimieren, keine repräsentativen Residuenbilder zur Verfügung stellen.

Eine direkte Anwendung der beschriebenen Methodologie und des dazugehörigen Formelapparats ist im eindimensionalen Fall ohne weiteres möglich, z.B. auf Zeitreihen.

Die ersten Ansätze und Untersuchungen zum Einsatz der Ausgleichung nach maximaler Korrelation in der Formenerkennung zur Rekonstruktion von Trassierungselementen wurden in [Neitzel 1998] gemacht. Die Untersuchung wurde auf die eindimensionale Darstellung der betrachteten Kurvenabschnitte eingeschränkt und die Auswahl einer entsprechenden Lösung aus der Klasse aller Lösungen nach maximaler Korrelation – ohne Aufstellung der in dieser Arbeit beschriebenen oder modifizierten zusätzlichen Bedingungen – visuell vorgenommen. Dabei hat sich erwiesen, daß die Idee erfolgversprechend ist, weshalb eine Fortsetzung dieser Untersuchungen geplant ist.

Die ersten Versuche einer Anwendung der beschriebenen Methodologie auf mehrdimensionale Probleme der Deformationsanalyse wurden ebenso von meinem Kollegen Frank Neitzel vorgenommen, s. z.B. [Neitzel 1999] oder [Neitzel 2001]. Das Hauptproblem bei der Deformationsanalyse besteht darin, daß in vielen Fällen die vorhandenen Deformationserscheinungen keinen systematischen Charakter aufweisen. In solchen Fällen kann von der in dieser Arbeit vorgeschlagenen Vorgehensweise zur Identifizierung *systematischer Einflüsse* nicht zu viel erwartet werden. Trotzdem bleibt auch dann die Korrelation als ein mögliches Maß für die Ähnlichkeit der in den unterschiedlichen Punktgruppen in verschiedenen Epochen enthaltenen Formen¹. Dabei ermöglicht die Ausgleichung nach maximaler Korrelation dieses Beurteilungskriterium vom Einfluß einer willkürlichen Datumsfestlegung zu befreien, so daß der Betrag der Korrelationskoeffizienten ein von der Lagerung unabhängiges Maß für das Verhältnis zwischen geometrischen Formen in verschiedenen Epochen wird. In den Fällen, in denen die Deformationen einen systematischen Charakter aufweisen (wie etwa in der Plattentektonik) bietet sich eine direktere Anwendung der in dieser Arbeit vorgeschlagenen Vorgehensweise an.

¹In der bestehenden Literatur zur Deformationsanalyse wird diese Untersuchung der Ähnlichkeit meistens Kongruenzanalyse genannt (auch wenn der Maßstab unbekannt ist) und basiert immer auf metrischen Kriterien.

Weiterhin haben die in der vorliegenden Arbeit beschriebenen Untersuchungen die Erarbeitung einer neuen Methode zur Generalisierung kartographischer Linienelemente motiviert, s. [Vučetić 2001]. Bei allen in der Literatur zu findenden Verfahren zur Reduktion der Punktmenge, die eine Linie in Digitalform beschreibt, basiert diese Reduktion auf metrischen Kriterien. Neben offensichtlichen Vorteilen (einfache Kontrolle der erforderlichen Darstellungsgenauigkeit) weist eine solche Vorgehensweise auch erhebliche Mängel auf: Die in der Kartographie sehr wichtige Erhaltung des gleichen visuellen Eindrucks geht dabei oftmals verloren. Das neue Verfahren selektiert diejenige Untermenge der liniendefinierenden Punkte, die die Eigenschaft hat, daß die generalisierte Linie mit der ursprünglichen maximale Korrelation aufweist. Die Versuche mit praktischen Beispielen haben gezeigt, daß die charakteristischen Formen der Linie auf diese Weise wesentlich besser erhalten werden als bei herkömmlichen Verfahren. Der Unterschied der primären Zielsetzung der vorliegenden Arbeit und der gerade beschriebenen Anwendung besteht darin, daß im ersten Falle unerwünschte systematische Anteile eliminiert werden sollen und im zweiten die „systematischen“ Liniencharakteristika zu bewahren sind.

Weitere Anwendungen des Prinzips der maximalen Korrelation sind denkbar, sowohl jene, die als eine direkte Anwendung der in der vorliegenden Arbeit vorgeschlagenen Vorgehensweise angesehen werden können als auch solche, die eine entsprechende Modifikation voraussetzen.

Literaturverzeichnis

- [Avi-Itzhak u.a. 1995] Avi-Itzhak, H.I.; Van Mieghem, J.A.; Rub, L.: Multiple subclass pattern recognition: A maxmin correlation approach. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 17 (Nr. 4), S. 418–431.
- [Baczkowski und Mardia 1990] Baczkowski, A.J.; Mardia, K.V.: Prediction based upon maximizing squared correlation for stationary processes and simple kriging. *Journal of Applied Statistics*, 17 (Nr. 1), S. 159–164.
- [Bessel und Baeyer 1838] Bessel, F.W.; Baeyer: *Gradmessung in Ostpreußen und ihre Verbindung mit Preußischen und Russischen Dreiecksketten*. Königliche Akademie der Wissenschaften, Berlin.
- [Bjerhammar 1973] Bjerhammar, A.: *Theory of Errors and Generalized Matrix Inverses*. Elsevier Scientific Publishing Company, Amsterdam – London – New York.
- [Boochs 1984] Boochs, F.: *Ein Verfahren zur Herstellung digitaler Höhenmodelle aus photogrammetrischen Stereomodellen mit Hilfe der flächenhaften Korrelation in digitalen Bildern*. Dissertation, Deutsche Geodätische Kommission, Reihe C, Nr. 299, München.
- [Bopp und Krauss 1977] Bopp, H.; Krauss, H.: Die strenge Bestimmung einer ausgleichenden Geraden bei korrelierten Beobachtungen nach der Methode der kleinsten Quadrate. *Zeitschrift für Vermessungswesen*, Nr. 6, S. 261–270.
- [Bykov und Gritchanyuk 1998] Bykov, V.N.; Gritchanyuk, A.M.: Size effect of the reference and current images upon operation of maximum correlation air navigation systems. *Telecommunication and Radio Engineering*, 51(6-7), S. 12–14.
- [Chatfield 1982] Chatfield, C.: *Analyse von Zeitreihen*. Carl Hansen Verlag, München.
- [Čolić und Petrović 1984] Čolić, K.; Petrović, S.: Correlation between gravity anomalies, geoid heights and Mohorovičić discontinuity depths in the Dinaric-Pannonian region of Yugoslavia. *Proceedings of the International association of geodesy (IAG) symposia*. Columbus: Ohio State University, S. 137–146.
- [Čolić u.a. 1987] Čolić, K.; Petrović, S.; Aljinović, B.: Testing the existing absolute geoid heights in Yugoslavia using information offered by the Mohorovičić discontinuity, the Bouguer gravity anomalies and the relief heights. *Bollettino di geodesia e scienze affini*, Heft 4, S. 309–322.
- [Čolić u.a. 1988] Čolić, K.; Vučetić, N.; Petrović, S.: The implications of the correlations between gravity anomalies, the Mohorovičić discontinuity and the geoid. *Proceedings of the seminary “Contemporary problems in gravimetry”*. Prague: Geophysical institute of ČSAV, S. 327–337.
- [Čolić u.a. 1989a] Čolić, K.; Petrović, S.; Bašić, T.: What happens in the least squares fitting of small parts of various geoid models? *Bollettino di geodesia e scienze affini*, Heft 1, S. 29–43.
- [Čolić u.a. 1989b] Čolić, K.; Vučetić, N.; Petrović, S.: The geoid and the Mohorovičić discontinuity. *Akademie der Wissenschaften der DDR, Veröffentlichungen des Zentralinstituts für Physik der Erde*, Nr. 102, Teil II, S. 168–179.
- [Čolić u.a. 1994] Čolić, K.; Petrović, S.; Sünkel, H.; Burda, M.; Bielik, M.; Vyskočil, V.: Similarities of and differences between three mountain belts at the border of the Pannonian basin due to the behaviour of the Mohorovičić discontinuity. *Studia geophysica & geodætica*, Heft 2, S. 131–139.
- [Dermanis 1998] Dermanis, A.: Generalized inverses of nonlinear mappings and the nonlinear geodetic datum problem. *Journal of Geodesy*, 72, S. 71–100.
- [Dermanis und Sansó 1995] Dermanis, A.; Sansó, F.: Nonlinear estimation problems for nonlinear models. *Manuscripta Geodaetica*, 20, S. 110–122.
- [Draper und Smith 1981] Draper, N.R.; Smith, H.: *Applied Regression Analysis*. 2nd ed., John Wiley & Sons, New York - Chichester - Brisbane - Toronto - Singapore.
- [Feng u.a. 1998] Feng, D.; Pei, D.; Xiao, G.: Maximum correlation analysis of nonlinear combining functions. *Science in China (Series E)*, Band 41, Nr. 1, S. 31–41.
- [Friedrich 1943] Friedrich, K.: Zwei aus den Grundgesetzen der Mechanik abgeleitete Beweise für die Richtigkeit der Methode der kleinsten Quadrate nebst praktischen und erkenntnistheoretischen Folgerungen. *Zeitschrift für Vermessungswesen*, Nr. 5/6, S. 97–112.

- [Fritzensmeier u.a. 1986] Fritzensmeier, K.; Niemeier, W.; Brunner, F. K.: Zur Aufdeckung kleiner systematischer Fehler in wiederholt gemessenen geodätischen Netzen. *Allgemeine Vermessungs-Nachrichten*, 93. Jg., Heft 4, S. 125–139.
- [Gächter u.a. 1984] Gächter, B.; Braunecker, B.; Müller, F.: *Meßeinrichtung zum Erfassen einer Relativposition zwischen zwei Teilen*. Patentschrift DE 3424806 A 1.
- [Gauß 1829] Gauß, C. F.: Über ein neues allgemeines Grundgesetz der Mechanik. *Journal für die reine und angewandte Mathematik (Borchardt's Journal für Mathematik)*, de Gruyter, Berlin, Bd. 4, S. 232–235.
- [Göpfert 1987] Göpfert, W.: *Raumbezogene Informationssysteme*. Herbert Wichmann Verlag GmbH, Karlsruhe.
- [Gotthardt 1968] Gotthardt, E.: *Einführung in die Ausgleichungsrechnung*. Herbert Wichmann Verlag, Karlsruhe.
- [Grafarend und Schaffrin 1991] Grafarend, E.W.; Schaffrin, B.: The planar trisection problem and the impact of curvature on non-linear least-squares estimation. *Computational Statistics & Data Analysis*, 12, S. 187–199.
- [Grafarend und Schaffrin 1993] Grafarend, E.W.; Schaffrin, B.: *Ausgleichungsrechnung in linearen Modellen*. BI-Wiss.-Verlag, Mannheim – Leipzig – Wien – Zürich.
- [Großmann 1969] Großmann, W.: *Grundzüge der Ausgleichungsrechnung*. Springer-Verlag, Berlin – Heidelberg – New York.
- [Gründig 1999] Gründig, L.: *Statistische Testverfahren*. Vorlesungsskript, TU Berlin.
- [Gyergyek u.a. 1988] Gyergyek, L.; Pavešić, N.; Ribarić, S.: *Uvod u Raspoznavanje Uzoraka (Einführung in die Mustererkennung)*, Kroatisch. Tehnička knjiga, Zagreb.
- [Hekimoglu und Koch 1999] Hekimoglu, Ş.; Koch, K.R.: How can reliability of the robust methods be measured? *Proceedings of the Third Turkish-German Joint Geodetic Days*, Istanbul, S. 179–196.
- [Helmert 1872] Helmert, F. R.: *Die Ausgleichungsrechnung nach der Methode der kleinsten Quadrate*. Teubner Verlag, Leipzig, XI+348 S.
- [Helmert 1924] Helmert, F. R.: *Die Ausgleichungsrechnung nach der Methode der kleinsten Quadrate*. 3. Auflage, Teubner Verlag, Leipzig, XVIII+597 S.
- [HSBw 1979] Seminar über Deformationsanalysen gehalten an der Hochschule der Bundeswehr München mit Beiträgen von E. Baumann, W. Caspary, H. Dupraz / W. Niemeier / H. Pelzer, E. Kuntz / G. Schmitt, W. Welsch. Schriftenreihe, Heft 4, Wissenschaftlicher Studiengang Vermessungswesen, Hochschule der Bundeswehr München.
- [Ingensand 1990] Ingensand, H.: Das WILD NA2000 – Das erste digitale Nivellier der Welt. *Allgemeine Vermessungs-Nachrichten*, 97. Jg., Heft 6, S. 201–210.
- [Jäger 1988] Jäger, R.: *Analyse und Optimierung geodätischer Netze nach spektralen Kriterien und mechanischen Analogien*. Deutsche Geodätische Kommission, Reihe C, Nr. 342, München.
- [Jovičić u.a. 1982a] Jovičić, D.; Lapaine, M.; Petrović, S.: Prilagodavanje kružnice skupu točaka ravnine. *Zbornik rada 4. međunarodnog simpozija „Kompjuter na sveučilištu“*. (Anpassung eines Kreises an eine Punktmenge in der Ebene. Sammelband des 4. internationalen Symposiums „Computer an der Universität“; in kroatischer Sprache). Cavtat: Sveučilišni računski centar Zagreb, S. 469–476.
- [Jovičić u.a. 1982b] Jovičić, D.; Lapaine, M.; Petrović, S.: Prilagodavanje pravca skupu točaka prostora (Anpassung einer Geraden an eine Punktmenge im Raum; in kroatischer Sprache). *Geodetski list*, 1982, 10–12, S. 260–266.
- [Kampmann 1986] Kampmann, G.: Robuster Ausreißertest mit Hilfe der L1-Norm-Methode. *Allgemeine Vermessungs-Nachrichten*, 93. Jg., Heft 4, S. 139–147.
- [Koch 1980] Koch, K. R.: *Parameterschätzung und Hypothesentests in linearen Modellen*. Dümmler Verlag, Bonn, XII+296 S.
- [Koch 1997] Koch, K. R.: Bemerkungen zu „Was ist Genauigkeit“. VR 1997, S. 212, *Vermessungswesen und Raumordnung*, 59. Jg., Heft 7, S. 362–370.
- [Koch und Yang 1998] Koch, K.R.; Yang, Y.: Robust Kalman filter for rank deficient observation models. *Journal of Geodesy*, 72, S. 436–441.
- [Kolmogorov 1946] Колмогоров, А. Н.: К обоснованию метода наименьших квадратов (Zur Begründung der Methode der kleinsten Quadrate, Russisch). Успехи математических наук, Jg. 1 (neue Serie), Nr. 1, S. 57–70.
- [Konak u.a. 1999] Konak, K.; Dilaver, A.; Öztürk, E.: Fuzzy logic and detection of outliers. *Proceedings of the Third Turkish-German Joint Geodetic Days*, Istanbul, S. 169–178.
- [Kreiling 1976] Kreiling, W.: *Automatische Herstellung von Höhenmodellen und Orthophotos aus Stereobildern durch digitale Korrelation*. Dissertation, Karlsruhe.

- [Li 1993] Li, B.: A Deviance Function for the Quasi-Likelihood method. *Biometrika*, Jhg. 80, S. 741–753.
- [Linkwitz 1961] Linkwitz, K.: *Fehlertheorie und Ausgleichung von Streckennetzen nach der Theorie elastischer Systeme*. Deutsche Geodätische Kommission, Reihe C, Nr. 46, München.
- [Linnik 1961] Linnik, Yu. V.: *Method of Least Squares and Principles of the Theory of Observations*. Pergamon Press, Oxford – London – New York – Paris.
- [Liu 2000] Liu, G. L.: Nonlinear curvature measures of strength and nonlinear diagnosis. *Allgemeine Vermessungsnachrichten*, Nr. 3, S. 109–111.
- [Lohse 1994] Lohse, P.: *Ausgleichungsrechnung in nichtlinearen Modellen*. Deutsche Geodätische Kommission, Reihe C, Nr. 429, München.
- [Marchenko u.a. 2001] Marchenko, A.N.; Kravtsov, O.N.; Abrikosov, O.A.: Closed expressions for the parameters of Helmert transformation. *Bollettino di geodesia e scienze affini*, LX, Nr. 1, S. 17–29.
- [Mather 1987] Mather, P. M.: *Computer Processing of Remotely-Sensed Images*. John Wiley & Sons, Chichester – New York – Brisbane – Toronto – Singapore.
- [Mautz 1999] Mautz, R.: Solving non-linear adjustment problems by heuristic optimisation. *Proceedings of the Third Turkish-German Joint Geodetic Days*, Istanbul, S. 791–797.
- [Mautz 2001] Mautz, R.: *Zur Lösung nichtlinearer Ausgleichungsprobleme bei der Bestimmung von Frequenzen in Zeitreihen*. Deutsche Geodätische Kommission, Reihe C, Nr. 532, München.
- [McCullagh 1983] McCullagh, P.: Quasi-likelihood functions. *The Annals of Statistics*, 11, Nr. 1, S. 59–67.
- [Milev 2000] Milev, I.: *Integrierte Modelle zur physikalischen Interpretation geodätischer Deformationsuntersuchungen*. Deutsche Geodätische Kommission, Reihe C, Nr. 540, München.
- [Moritz 1966] Moritz, H.: Zur geometrischen Deutung der Minimumsprinzipien der Ausgleichungsrechnung. *Zeitschrift für Vermessungswesen*, Nr. 8, S. 293–296.
- [Moritz und Mueller 1987] Moritz, H.; Mueller, I. I.: *Earth Rotation: Theory and Observation*. Ungar, New York.
- [Neitzel 1998] Neitzel, F.: *Untersuchungen zur Formenerkennung durch Ausgleichung nach maximaler Korrelation*. Seminararbeit, Technische Universität Berlin, 43 S. Auszug in: W. Freeden (editor), *Progress in Geodetic Science at GW 98*, Shaker Verlag, Aachen, S. 299–306.
- [Neitzel 1999] Neitzel, F.: *Ausgleichung nach maximaler Korrelation in der geometrischen Deformationsanalyse*. Diplomarbeit, Technische Universität Berlin.
- [Neitzel 2001] Neitzel, F.: Ausgleichung nach maximaler Korrelation in der geometrischen Deformationsanalyse. *Vermessung, Photogrammetrie, Kulturtchnik*, 99, Nr. 3, S. 140–145.
- [Neitzel 2002] Neitzel, F.: *Zur Transformation und Kongruenzuntersuchung geodätischer Netze* (Arbeitstitel). Dissertation, Technische Universität Berlin, in Vorbereitung.
- [Perico und Pratolongo 1997] Perico, A.; Pratolongo, R.: Maximum-correlation mode-coupling approach to the Smoluchowski dynamics of polymers. *Macromolecules*, 30, Nr. 19, S. 5958–5969.
- [Perović 1998] Perović, G.: *Metod najmanjih kvadrata, Monografija 1 (Die Methode der kleinsten Quadrate, Monografie 1)*; Serbokroatisch. Belgrad (im Druck).
- [Petrović 1991] Petrović, S.: *Geometry of the correlation coefficient and its application in geodesy*. Mitteilungen der geodätischen Institute der Technischen Universität Graz, Folge 71, Graz 1991, VII+100 S.
- [Petrović 1993] Petrović, S.: Maximal correlation adjustment in physical geodesy. In: H. Montag and Ch. Reigber (Eds.): *Geodesy and Physics of the Earth, International Association of Geodesy Symposium No. 112*, Springer-Verlag, S. 242–245.
- [Petrović 1995] Petrović, S.: Systematic errors and adjustment. *Proceedings of the First Turkish-German Joint Geodetic Days*, Istanbul, S. 123–131.
- [Petrović 1996] Petrović, S.: Zur Ausgleichung beim Vorliegen systematischer Fehler. *Erste Geodätische Woche*, Stuttgart, 7.-12.10.1996. Zusammenfassung in: *Universität Stuttgart, Schriftenreihe der Institute des Fachbereichs Vermessungswesen, Report Nr. 1997.1* (herausgegeben von A. Gilbert und E.W. Grafarend), S. 49.
- [Petrović 1997] Petrović, S.: Maximum correlation adjustment and incomplete models. *Proceedings of the Second Turkish-German Joint Geodetic Days*, Berlin, S. 687–694.
- [Petrović und Čolić 1994] Petrović, S.; Čolić, K.: Estimating the density jump at the Mohorovičić discontinuity from the correlation with the geoid. *Manuscripta geodætica*, 20, S. 46–51.

- [Petrović u.a. 1989] Petrović, S.; Vučetić, N.; Čolić, K.: Absolute orientation of the astrogeodetic geoid model for the Yugoslav territory achieved by the method of the maximal linear correlation coefficients. *Akademie der Wissenschaften der DDR, Veröffentlichungen des Zentralinstituts für Physik der Erde*, Nr. 102, Teil II, S. 271–285.
- [Petrović u.a. 1993] Petrović, S.; Čolić, K.; Vučetić, N.; Medak, D.: Comparing various solutions for the Mohorovičić discontinuity shape in the Dinaric–Pannonian region. In: H. Montag and Ch. Reigber (Eds.): *Geodesy and Physics of the Earth, International Association of Geodesy Symposium No. 112*, Springer-Verlag, S. 159–162.
- [Rainsford 1957] Rainsford, H.F.: *Survey Adjustment and Least Squares*. Constable & Company Ltd., London.
- [Rao 1973] Rao, C.R.: *Linear Statistical Inference and Its Applications*. 2nd ed., John Wiley & Sons, New York - London - Sydney - Toronto.
- [Reinking 1994] Reinking, J.: *Geodätische Analyse inhomogener Deformationen mit nichtlinearen Transformationsfunktionen*. Deutsche Geodätische Kommission, Reihe C, Nr. 413, München.
- [Reissmann 1962] Reissmann, G.: *Die Ausgleichungsrechnung*. VEB Verlag für Bauwesen, Berlin.
- [Saleh 2000] Saleh, J.: Robust estimation based on energy minimization principles. *Journal of Geodesy*, 74, S. 291–305.
- [Sansò 1973] Sansò, F.: An exact solution of the roto-translation problem. *Photogrammetria*, Bd. 29, S. 203–216.
- [Schwetlick 1979] Schwetlick, H.: *Numerische Lösung nichtlinearer Gleichungen*. VEB Deutscher Verlag der Wissenschaften, Berlin 1979.
- [Searle 1971] Searle, S.R.: *Linear Models*. John Wiley & Sons, New York - London - Sydney - Toronto.
- [Seber 1977] Seber, G. A. F.: *Linear Regression Analysis*. John Wiley & Sons, New York - London - Sydney - Toronto.
- [Shokhin und Bykova 1998] Shokhin, M.G.; Bykova, T.V.: Increasing the speed of maximum correlation air navigation systems. *Telecommunication and Radio Engineering*, 51(6-7), S. 22–24.
- [Steiner und Hajagos 2001] Steiner, F. und Hajagos, B.: Reliability of the estimations of the correlation coefficient. *Acta geodaetica et geophysica Hungarica*, 36, Nr. 4, S. 399–416.
- [Stephens 1979] Stephens, M.A.: Vector Correlation. *Biometrika*, Jhg. 66, S. 41–48.
- [Teunissen 1988] Teunissen, P. J. G.: The non-linear 2D symmetric Helmert transformation: an exact non-linear least-squares solution. *Bulletin Géodésique*, Vol. 62, S. 1–15.
- [Teunissen 1990] Teunissen, P. J. G.: Nonlinear least squares. *Manuscripta geodætica*, Vol. 15, S. 137–150.
- [Vaníček und Craymer 1983] Vaníček, P.; und Craymer, M.: Autocorrelation functions in the search for systematic errors in levelling. *Manuscripta Geodaetica*, 8, S. 321–341.
- [Vaníček u.a. 1985] Vaníček, P.; Kleusberg, A.; Langley, R.B.; Santerre, R. und Wells, D.E.: On the elimination of biases in processing differential GPS observations. *Proceedings of the First International Symposium on Precise Positioning with the Global Positioning System, Rockville, Maryland, 15–19. April 1985*, National Geodetic Information Center, NOAA, S. 315–324. Defense
- [Vučetić 2001] Vučetić, N.: *Generalizacija linijskih elemenata karte po kriteriju maksimalne sličnosti* (*Generalisierung von linienhaften Kartenelementen nach dem Kriterium der maximalen Ähnlichkeit*, in Kroatisch). Dissertation, Geodätische Fakultät der Universität in Zagreb.
- [Vučetić u.a. 1990] Vučetić, N.; Petrović, S.; Čolić, K.; Burda, M.; Bielik, M.; Vyskočil, V.: Correlations between some geodetic and geophysical parameters in the West Carpathians, and the basic implications of the results. *Mitteilungen der geodätischen Institute der Technischen Universität Graz*, Folge 67, Graz, S. 215–231.
- [Wedderburn 1974] Wedderburn, R. W. M.: Quasi-likelihood functions, generalized linear models, and the Gauss-Newton Method. *Biometrika*, Jhg. 61, Nr. 3, S. 439–447.
- [Wellisch 1904] Wellisch, S.: Fehlerausgleichung nach der Theorie des Gleichgewichtes elastischer Systeme. *Österreichische Zeitschrift für Vermessungswesen*, Jhg. 2, Nr. 12, S. 181–190, Nr. 13, S. 197–210, Nr. 14, S. 213–219, Nr. 15, S. 229–235, Nr. 16, S. 246–253.
- [Wiener 1970] Wiener, N.: *Time Series*. The M.I.T. Press, Cambridge, Massachusetts, 1970, X+163 S.
- [Wolf 1968] Wolf, H.: *Ausgleichungsrechnung nach der Methode der kleinsten Quadrate*. Ferd. Dümmlers Verlag, Bonn.
- [Yang u.a. 1999] Yang, Y.; Cheng, M.K.; Shum, C.K.; Tapley, B.D.: Robust estimation of systematic errors of satellite laser range. *Journal of Geodesy*, 73, S. 345–349.
- [York 1966] York, D.: Least squares fitting of a straight line. *Canadian Journal of Physics*, S. 1079–1086.
- [Young u.a. 1986] Young, T.Y.; Fu, K.-S. (Hrsg.): *Handbook of Pattern Recognition and Image Processing*. Academic Press, San Diego, London, etc.
- [Zhang 1994] Zhang, H.: Maximal correlation and adaptive splines. *Technometrics*, 36, Nr. 2, S. 196–201.

Anhang

A.1 „Total least squares“ und Verallgemeinerungen davon

In einem Teil der Literatur zur mathematischen Statistik (beispielsweise in [Golub und van Loan 1980] oder [Van Huffel und Zha 1993]) wird zwischen “Least squares” (LS) und “Total least squares” (TLS), auch „errors-in-variables model“ genannt (s. [Fierro u.a. 1997], [Huwang und Huang 2000] oder [Chandrasekaran u.a. 2001]), unterschieden¹. Betrachtet wird ein überbestimmtes lineares Modell

$$l \approx Ax. \quad (\text{A-1})$$

Wegen unvermeidlicher Meßfehler (in der Terminologie der zitierten Literaturquellen: „perturbances“, d.h. Störungen) kann das Gleichungssystem (A-1) nur näherungsweise erfüllt werden. Angenommen, daß sich alle Meßfehler auf die Komponenten des Vektors l beziehen, normalverteilt sind und die gleichen Verteilungsparameter haben, bietet es sich an, den Vektor der „Verbesserungen“ v einzuführen:

$$l + v = Ax \quad (\text{A-2})$$

und die Zielfunktion

$$\sum vv \quad (\text{A-3})$$

zu minimieren². Entstammen die Meßfehler für die einzelnen Komponenten des Vektors l unterschiedlichen Normalverteilungen oder sind die Messungen noch dazu korreliert, so wird auf bekannte Weise eine entsprechende Gewichtsmatrix P eingeführt. In der zitierten Literatur heißt dieses Ausgleichungsmodell “Least squares” (LS).

Ist aber die Gleichung (A-1) nur näherungsweise erfüllt, weil nicht allein der Vektor l Störungen enthält, sondern ebenfalls die Matrix A , so ist die Modellierung durch (A-2) und die Minimierung der Zielfunktion (A-3) offensichtlich nicht angemessen. Es bietet sich dann an, auch die Elemente der Matrix A mit Verbesserungen zu versehen:

$$l + v = (A + V)x \quad (\text{A-4})$$

und die Zielfunktion

$$\sum(vv + VV) \quad (\text{A-5})$$

zu minimieren. Auf bekannte Weise können gegebenenfalls entsprechende Gewichtsmatrizen sowohl für l als auch für A eingeführt werden³. In der zitierten Literatur heißt dieses Ausgleichungsmodell “Total least squares” (TLS).

Es liegt nahe, TLS folgendermaßen zu interpretieren: Die Matrix A in (A-1) vertritt das Modell, der Vektor l die Messungen. Dementsprechend ist TLS eine Methode, die Störungen gleichzeitig in den Messungen und in dem Modell sucht.

¹In dieser Arbeit werden beide Begriffe in der englischen Originalform benutzt. Wie im vorliegenden Abschnitt gezeigt wird, können diese Begriffe nicht gerechtfertigt werden. Deshalb wird auf eine Übersetzung verzichtet.

²Die Formeln aus der zitierten Literatur wurden hier mit Absicht in die traditionelle (konservative) Notation umgeschrieben. Diese Notation weicht zwar von der heutzutage verwendeten statistischen Notation ab, ist jedoch besser dazu geeignet, Mißverständnisse zu vermeiden.

³Da es sich an dieser Stelle in erster Linie darum handelt, mögliche Mißverständnisse auszuräumen, ist es angemessen, die Formeln für den einfachsten Fall zu betrachten und die Verallgemeinerungen im Auge zu behalten. Bei einer anderen Zielsetzung wäre es vielleicht günstiger, die Formeln für den allgemeinsten Fall anzuführen und die Spezialfälle zu erläutern.

Vielelleicht lassen sich auf diese Weise systematische Fehler im Modell identifizieren. Falls ja, dann wäre TLS von höchster Relevanz für diese Arbeit.

Die Antwort ergibt sich aus (A-4) und (A-5): *Sowohl für die Störungen von l als auch für die von A wurde postuliert, daß sie einen stochastischen Charakter haben⁴.* Von einer Anwendung der TLS zur Identifizierung systematischer Fehler kann also keine Rede sein.

Obwohl im Rahmen der Problematik, die in dieser Arbeit behandelt wird, „Total least squares“ keinen Beitrag liefern kann, haben sich aus der Auseinandersetzung mit der Literatur zu TLS einige hochinteressante Erkenntnisse ergeben, die nachfolgend kurz geschildert werden.

Als erstes stellt sich die Frage: Obwohl TLS nicht zur Identifizierung systematischer Fehler geeignet ist, ist diese Vorgehensweise trotzdem eine Methode zur „Verbesserung“ des Modells? *Welche Art der „Modellverbesserung“ ergibt sich aus TLS?*

Interessanterweise: keine. TLS hat mit „Modellverbesserung“, „Modellveränderung“ oder „Modelluntersuchung“ nichts zu tun, obwohl es auf den ersten Blick so erscheinen mag. Um dies einzu-sehen, soll die Frage gestellt werden, in welchem Sinne die Matrix A das Modell vertritt. Am besten kann dies am einfachsten Beispiel veranschaulicht werden, welches in der Literatur zur TLS als Paradebeispiel benutzt wird (s. [Golub und van Loan 1980] oder [Van Huffel und Zha 1993]), um den Unterschied zwischen LS und TLS zu illustrieren. Es handelt sich um die Anpassung einer Geraden an eine gegebene Punktmenge, s. Abb. 1

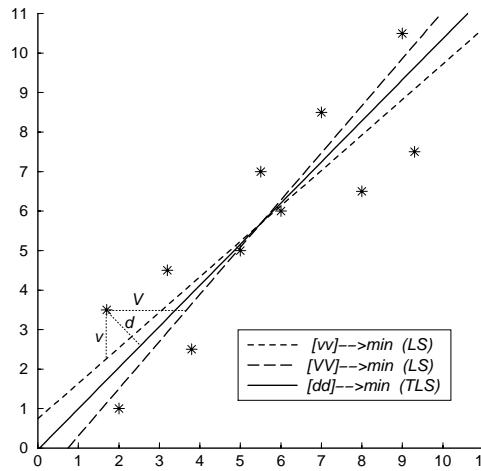


Abbildung 1: Gegebene Punkte und die ausgleichenden Geraden nach LS und TLS

Die Gleichung einer Geraden kann bekanntlich in der Form

$$Y = aX + b \quad (\text{A-6})$$

dargestellt werden. Das Gleichungssystem (A-1) besteht demnach aus folgenden Gleichungen

$$Y_i = aX_i + b \quad i = 1, \dots, n \quad (\text{A-7})$$

wobei n die Anzahl der Punkte bezeichnet. Die Matrizen in (A-1) haben deshalb folgende Bedeutung:

$$A = \begin{bmatrix} X_1 & 1 \\ X_2 & 1 \\ \vdots & \vdots \\ X_n & 1 \end{bmatrix}, \quad x = \begin{bmatrix} a \\ b \end{bmatrix}, \quad l = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix}. \quad (\text{A-8})$$

Werden Störungen⁵ ausschließlich für die Messungen Y_i zugelassen, so ergibt sich die LS-Lösung. Die Einführung zusätzlicher Störungen für die Matrix A bedeutet nichts anderes als die Berücksichtigung der Meßfehler auch für die Messungen X_i . Deshalb dürfen Störungen (Verbesserungen) nur für

⁴ Alle in den zitierten Literaturquellen ausgearbeiteten Lösungswege basieren auf dieser Annahme.

⁵ Der stochastische Charakter der Störungen wird genauso wie in den zitierten Literaturquellen postuliert.

die erste Spalte der Matrix A eingeführt werden⁶. Als TLS-Lösung ergibt sich die „orthogonal regression“ (s. [Van Huffel und Zha 1993]), d.h. die Gerade, für die die Summe der quadrierten Abstände von den gegebenen Punkten minimal ist.

Nun kann gezeigt werden, daß LS und TLS nichts anderes sind als *Spezialfälle einer korrekten Anwendung der Methode der kleinsten Quadrate*.

Sind nur die Y_i -Werte fehlerbehaftet, dann ist

$$Y_i + v_i = aX_i + b \quad i = 1, \dots, n \quad (\text{A-9})$$

ein angemessenes Grundmodell. Ein Beispiel für eine derartige Aufgabenstellung ist eine Signalregistrierung in gewissen Zeitabständen, wobei die Zeitmessung wesentlich genauer als die Signalregistrierung ist und deshalb als praktisch fehlerfrei betrachtet werden darf. Das Gleichungssystem (A-9) ist linear und kann als System der Beobachtungsgleichungen betrachtet werden; eine Ausgleichung nach vermittelnden Beobachtungen liefert dann die Lösung, die mit der LS-Lösung identisch ist.

Völlig analog ergibt sich eine LS-Lösung für den Fall, in dem nur die X_i -Werte fehlerbehaftet sind.

Sind sowohl die X_i -Werte als auch die Y_i -Werte fehlerbehaftet, dann kann

$$Y_i + v_i = a(X_i + V_i) + b \quad i = 1, \dots, n \quad (\text{A-10})$$

ein angemessenes Grundmodell sein. Dieses Modell entspricht beispielsweise einer Aufgabenstellung, bei der die Koordinaten X_i und Y_i von einer Reihe von Punkten eines geradlinigen Objektes durch Orthogonalverfahren mit gleichen Genauigkeiten bestimmt wurden. Das Gleichungssystem (A-10) kann nicht unmittelbar als System der Beobachtungsgleichungen betrachtet werden. Deshalb muß zuerst eine Umformung durchgeführt werden, woraus sich ein nichtlineares Ausgleichungsproblem ergibt. Im Falle nichtkorrelierter und gleichgewichtiger Beobachtungen ergibt sich eine quadratische Gleichung, s. [Adcock 1878], die direkt gelöst werden kann. Sind den Beobachtungen unterschiedliche Gewichte zugeordnet, ergibt sich eine „pseudokubische“ Gleichung, die iterativ gelöst werden kann, s. [York 1966]. Diese „pseudokubische“ Gleichung wurde in [York 1969] auf korrelierte Messungen verallgemeinert⁷ und auch in eine „pseudoquadratische“ Gleichung überführt⁸.

Alternativ kann das Problem nach der Transformation in ein nichtlineares Ausgleichungsproblem durch Linearisierung und Iteration gelöst werden, s. [Bopp und Krauss 1977]. In beiden Fällen ergibt sich die Lösung, die mit der TLS-Lösung identisch ist.

Die interessante Konsequenz ist, daß LS und TLS nicht zwei unterschiedliche Methoden sind, sondern Anwendungen der gleichen Methode (Ausgleichung nach kleinsten Quadraten) auf zwei unterschiedliche Aufgabenstellungen. Dadurch erübrigt sich jede Diskussion, welche der beiden „Methoden“ besser ist⁹. Es ist nur notwendig, immer die vorliegende Aufgabenstellung zu modellieren und nicht etwas völlig anderes¹⁰.

Die in diesem Abschnitt erörterte Literatur aus der mathematischen Statistik ist aber aus anderen Gründen für geodätische Fragestellungen von höchster Relevanz.

Nach der Aufstellung des Modells wird in der geodätischen Praxis das Problem in der Regel durch Linearisierung gelöst. Diese Vorgehensweise funktioniert meist recht gut, hat manchmal aber auch ausgeprägte Nachteile. In der zitierten statistischen Literatur wurde eine Vielzahl von Lösungswegen ausgearbeitet, wobei besondere Aufmerksamkeit den Stabilitätsproblemen, der numerischen Güntigkeit und der Effizienz geschenkt wurde, s. z.B. [Van Huffel und Zha 1993] oder [Huwang und Huang 2000].

⁶Dies läßt sich durch eine entsprechende Gewichtung für die Matrix A regeln. In [Golub und van Loan 1980] und [Van Huffel und Zha 1993] wird dies erreicht, indem *vor der Anwendung von TLS* die Punktwolke implizit zum Schwerpunkt translatiert wird, wodurch die zweite Spalte der Matrix A verschwindet. Es kann gezeigt werden, daß die beiden Vorgehensweisen äquivalent sind.

⁷Eigentlich auf einen Spezialfall der korrelierten Messungen, nämlich den Fall, in dem nur die zugeordneten X_i - und Y_i -Werte miteinander korreliert sind.

⁸Diese pseudoquadratische Gleichung ergibt sich auch direkt als Verallgemeinerung der Lösung von [Adcock 1878]. Warum York, dem die Lösung von Adcock bekannt war, überhaupt in [York 1966] nur die „pseudokubische“ Gleichung angeboten hat, ist unverständlich und nicht nachvollziehbar.

⁹Derartige Diskussionen sind in der Literatur leider zu finden, beispielsweise in [Golub und van Loan 1980], [Van Huffel und Zha 1993], [Huwang und Huang 2000] und [Watson 2001].

¹⁰In einem anderen Kontext ist in [Steiner und Hajagos 2001] die folgende allgemeingültige Aussage zu finden: „Every researcher of each practical discipline, however, well knows, that statistics is to be used always appropriately according to the demands of the given task of the discipline and never automatically“.

Dazu gehören unterschiedliche Verallgemeinerungen und Modifikationen von TLS, wie z.B. robuste TLS-Methoden ([He und Liang 2000])

Besonders interessant sind die Verallgemeinerungen, bei denen für die Störungen zusätzlich zum stochastischen Modell noch Schranken eingeführt werden („bounded errors-in-variables model“, s. z.B. [Chandrasekaran u.a. 2001]). In vielen Fragestellungen der Geodäsie wäre es sehr sinnvoll, in die Formulierung des Problems zusätzlich zu den Varianzen noch Toleranzen einzuführen. Es scheint, daß die Untersuchungen und daraus resultierenden Algorithmen zum „bounded errors-in-variables model“ entsprechende Möglichkeiten bieten. Eine Übertragung auf geodätische Fragestellungen ist eine zukünftige Aufgabe für die wissenschaftliche Forschung auf dem Gebiet der geodätischen Modellbildung, deren Bedeutung kaum überschätzt werden kann.

Hochrelevant ist ebenfalls die Untersuchung der Beziehungen zwischen TLS und Regularisierung, woraus eine Vielfalt neuer Regularisierungsmethoden und dazugehörigen effizienten numerischen Algorithmen hervorgegangen ist, s. z.B. [Fierro u.a. 1997] und [Golub u.a. 1999]. Bekanntlich ist das Problem der Regularisierung bei riesigen Gleichungssystemen eines der brennenden Probleme der heutigen Satellitengeodäsie, in der die Rekonstruktion von Kugelfunktionskoeffizienten des Erdschwerfeldes aus Satellitenmeßdaten in Gestalt eines hochgradig instabilen Problems auftritt.

A.2 „Hidden bias“ in „Observational studies“

Einen hervorragenden Überblick über das Thema „Hidden bias in Observational studies“ mit 77 Literaturstellen gibt Paul R. Rosenbaum unter dem Titel „Observational studies and nonrandomized experiments“, s. [Rosenbaum 1996].

Es stellt sich selbstverständlich die Frage, in welchem Zusammenhang „hidden bias“¹¹ zum Thema dieser Arbeit steht.

„Observational studies“ finden vor allem in der Epidemiologie (der Zweig der Medizin, der sich grundsätzlich mit den Ursachen der Erscheinung oder Nichterscheinung einer Krankheit in einer gewissen Population beschäftigt) und verwandten Fragestellungen¹² Anwendung. Deshalb kommen die meisten Beispiele in den zitierten Literaturquellen aus diesem Bereich.

Um eine realitätsnahe quantitative Klassifizierung der Reaktionen eines menschlichen Organismus auf eine Wirkung („treatment“) durchzuführen, sind „randomized experiments“ fast ideal. Die einzige Annahme, die gemacht werden muß, ist die Annahme, daß alle Kombinationen von denkbaren und undenkbaren Eigenschaften eines Individuums und der zusätzlichen Wirkungen auf dieses, die einen Einfluß auf den Ausgang eines Experimentes haben können, in dem untersuchten Teil der Bevölkerung zufällig verteilt sind. Diese Annahme kann offensichtlich nicht hundertprozentig erfüllt sein; daß sie jedoch in einem hohen Maße erfüllt ist, erscheint aber als evident. In einem derartigen Experiment werden Individuen nach dem Zufallsprinzip ausgewählt. Nach dem gleichen Prinzip wird die Entscheidung getroffen, ob ein Individuum der zu untersuchenden Wirkung ausgesetzt wird oder nicht. Eintreffen oder nicht Eintreffen (entweder so, binär, in Stufen oder kontinuierlich quantifiziert) der untersuchten Reaktion wird dokumentiert. Zumaldest in einem statistischen Sinne kann erwartet werden, daß der Gesamteinfluß aller anderen Wirkungen vernachlässigbar ist (Erwartungswert gleich Null), so daß eine hohe Frequenz der Koinzidenz einer Wirkung mit einer Reaktion als ein kausaler Zusammenhang zwischen den beiden interpretiert werden darf.

Obwohl „randomized experiments“ das beste von den vorhandenen Mitteln ist, um die Kausalität zu erforschen, kommen sie meist nicht in Frage. Um einen kausalen Zusammenhang zwischen Radioaktivität und Leukemie festzustellen, wäre es beispielsweise notwendig, zufällig ausgewählte gesunde Individuen nach einem Zufallsprinzip entweder der Radioaktivität auszusetzen oder nicht und zu dokumentieren, in welchen Fällen es zu einer Entwicklung der Leukemie gekommen ist.

¹¹Sowohl der Artikel [Rosenbaum 1996] als auch die in ihm zitierten Literaturstellen und die zusätzlich konsultierte Literatur stammen aus dem englischen Sprachraum. Diese Tatsache bietet die Möglichkeit an, auf die Übersetzung einiger Begriffe zu verzichten. Dies betrifft insbesondere jene Begriffe, die in einer übersetzten Form mit ähnlichen Begriffen aus dieser Arbeit verwechselt werden könnten. Auf diese Weise kann eine der wichtigsten Quellen aller Mißverständnisse in wissenschaftlichen Diskussionen vermieden werden. Es ist eher die Regel als eine Ausnahme, daß gleiche oder ähnliche Begriffe in unterschiedlichen Wissenschaften verschiedene Bedeutungen haben, die sich wiederum alle von der Bedeutung im täglichen Leben unterscheiden.

¹²Ohne Zweifel gehören die Fragen, wie ein Mensch auf spezifische Schadstoffe, eine Therapie oder eine spezielle Art der Ausbildung reagiert, in die gleiche Kategorie.

Als Ersatzmittel bieten sich „observational studies“ an, s. beispielsweise [Cochran 1965] und [Rubin 1974]. Untersucht wird eine Gruppe von Individuen, für welche bekannt ist, daß sie einer bestimmten Wirkung ausgesetzt wurde oder bei der bestimmte Symptome, eine spezifische Krankheit oder eine andere Reaktion bereits festgestellt wurde. Solche Gruppen sind einfach zu lokalisieren, beispielsweise unter den Patienten eines Krankenhauses, wobei es zusätzlich günstig ist, daß in diesem Falle viele Informationen über die „Untersuchungsobjekte“ in den Akten bereits vorhanden sind. Das Problem bei einer derartigen Herangehensweise liegt darin, daß eine solche Gruppe für die allgemeine Bevölkerung nicht besonders repräsentativ ist. Im betrachteten Beispiel kann nicht davon ausgegangen werden, daß der durchschnittliche Gesundheitszustand einer Gruppe von Patienten den durchschnittlichen Gesundheitszustand der Bevölkerung repräsentiert. Aus diesem Grund werden eine oder mehrere zusätzliche Gruppen gebildet, die Kontrollgruppen genannt werden.

Eine Kontrollgruppe muß mehrere widersprüchliche Anforderungen erfüllen. Einerseits sollte sie die „Normalbevölkerung“ repräsentieren, andererseits sollte sie im Hinblick auf die Faktoren, die einen Einfluß auf den Ausgang des Experimentes haben könnten, ähnliche Eigenschaften wie die untersuchte Gruppe aufweisen. Da es nicht möglich ist, alle Forderungen gleichzeitig zu erfüllen, werden Kompromißlösungen gesucht. Voraussetzung dafür ist, daß im voraus bekannt ist, welche Faktoren für den Ausgang relevant sind. Die einfachste Idee besteht darin, jedem Individuum der Untersuchungsgruppe ein Individuum der Kontrollgruppe zuzuordnen, nach dem Kriterium, daß die beiden Individuen die gleichen relevanten Faktoren haben. Da dies meistens nicht realisierbar ist, wurde eine ganze Reihe unterschiedlicher Methoden entwickelt, um beispielsweise die relevanten Faktoren für die beiden Gruppen im Mittel anzugeleichen oder alle einzelnen Differenzen zu minimieren. Ausführliche Darstellungen, sowohl zu verschiedenen Möglichkeiten der Realisierung einer sinnvollen Zuordnung zwischen den Individuen einer Untersuchungsgruppe und den Individuen einer oder mehrerer Kontrollgruppen, als auch zu den dabei auftretenden Problemen sind z.B. in [Cochran und Rubin 1973], [Rubin 1973], [Carpenter 1977], [Rosenbaum und Rubin 1985] und [Ming und Rosenbaum 2001] zu finden.

Ist es bekannt, daß die beiden Gruppen in Bezug auf die relevanten Faktoren zu unterschiedlich sind (Voraussetzung dafür sind die Kenntnisse, welche Faktoren alle relevant sind und das Vorhandensein von Informationen über alle diese Faktoren für alle Individuen beider Gruppen.), wird von einem „overt bias“ geredet. Fehlen die Informationen über relevante Faktoren oder die Kenntnisse darüber, welche alle Faktoren relevant sind, dann wird ein „hidden bias“ befürchtet¹³. Diese Befürchtung ist im Prinzip immer gerechtfertigt. Das Hauptproblem liegt eigentlich im Übergang von „association“ (Zusammenhang, aber in einem nichtkausalen Sinne) zu „causation“ (Ursachenfeststellung), s. beispielsweise [Cochran 1965], [Schlesselmann 1978], [Rosenbaum 1984], [Holland 1986] oder [Cox 1992].

Ein schönes Beispiel für die Probleme bei der Ursachenforschung ist eine Diskussion, die vor 50 Jahren in den USA entflammt, ob das Rauchen als Ursache für die Entstehung von Lungenkrebs betrachtet werden kann. Es gab zahlreiche Studien mit unterschiedlichsten Schlußfolgerungen. Auf dem Niveau der „association“ gab es keine ausgeprägten Mißverständnisse: Der Anteil von Zigarettenrauchern bei den Lungenkrebspatienten war wesentlich erhöht. Die Erörterungen zur „causation“ haben aber immer einen eher spekulativen als statistischen Charakter. Deshalb verwundert wenig, daß die Schlußfolgerungen der verschiedenen Studien widersprüchlich waren¹⁴.

In [Cornfield u.a. 1959] wurde eine Synthese der Ergebnisse aus den genannten widersprüchlichen Studien gemacht, die in der Literatur als erste formale „sensitivity analysis“ betrachtet wird. Am diskutierten Beispiel war der Gedankengang wie folgt: Ist der Anteil der Lungenkrebskrankungen bei den Zigarettenrauchern neun mal höher als bei den Nichtrauchern, dann ist das Risiko in der ersten Gruppe neun mal höher als in der anderen. Sollte das Rauchen nicht die Ursache sein, dann sollte die wahre Ursache bei den Zigarettenrauchern neun mal stärker vertreten sein als bei den Nichtrauchern¹⁵. *Die Entscheidung, ob eine derartige Ursache existieren oder nicht existieren kann, kann nicht auf dem statistischen Wege getroffen werden.* Deshalb enthalten alle derartigen Abhandlungen

¹³Terminologisch besteht die Idee darin, im ersten Falle von einem „offenen“ (sichtbaren, offensichtlichen) und im zweiten von einem „versteckten“ (unsichtbaren) Bias zu reden.

¹⁴Diese reichten von der Behauptung, daß der erhöhte Anteil als endgültiger Beweis für den kausalen Zusammenhang akzeptiert werden muß, bis zu der Idee, daß es völlig andere Faktoren gibt, die die Ursache sowohl für die Entwicklung der Neigung zum Rauchen als auch für die Anfälligkeit für Lungenkrebs sind.

¹⁵Diese Überlegung wurde im zitierten Artikel ausgehend von „odds ratio“ (Chancenverhältnis) anschließend mathematisch dargestellt, die in der späteren Literatur weiter ausgearbeitet wurden. Innerhalb dieses kurzen Kommentars wird auf eine formelle Darstellung verzichtet.

viele philosophische Aspekte¹⁶, aber auch die Berufung auf das Gefühl und die Vorstellbarkeit.

Trotzdem gelten die Schlußfolgerungen vieler „sensitivity analysis“ nach vielen Jahren noch immer als sinnvoll. Die Möglichkeit, daß die Schlußfolgerungen später widerlegt werden, weil die Existenz des Unvorstellbaren doch nachgewiesen wird, wird in der Literatur ebenso akzeptiert.

Wie schon festgestellt, erörtert eine „sensitivity analysis“ die Frage, wie hoch ein „hidden bias“ sein soll, um die Schlußfolgerungen der Studie zu ändern¹⁷. Eine derartige Analyse liefert aber keine Hinweise zum möglichen Vorhandensein eines „hidden bias“. Diese Tatsache verwundert wenig: „hidden bias“ bezieht sich vor allem auf die mit der „causation“ verbundenen Spekulationen und weniger auf die durch Zahlen dokumentierte „association“.

Anstrengungen zur Aufdeckung von „hidden bias“ sind durch die Art der vorhandenen zusätzlichen Informationen und durch den Typ der Bias, welche anhand dieser Art von Informationen hoffentlich entdeckt werden könnten, charakterisiert¹⁸. Im Idealfall wird die Studie derartig gestaltet, daß zu erwarten ist, daß die Auswirkungen der Behandlung („treatment“) und des „hidden bias“ auf die beobachteten Daten als unterschiedlich erscheinen, s. [Rosenbaum 1996]. Daraus resultieren die Vorschläge, so viel wie möglich unterschiedliche Faktoren zu beobachten und zu vergleichen. Dies betrifft nicht nur diejenige Faktoren, die in der konkreten Studie vom primären Interesse sind. Es ist im voraus gründlich zu überlegen, welche alle relevanten Faktoren denkbar sind und diese jedenfalls zu beobachten. Des weiteren sollten zwei oder mehrere Kontrollgruppen mit unterschiedlichen Eigenschaften verwendet und auch untereinander verglichen werden usw. Genauso können mehrere unterschiedliche Reaktionen auf die gleiche Wirkung (zwischen denen jedoch ein Zusammenhang vermutet wird) beobachtet und auf Kohärenz überprüft werden, s. [Rosenbaum 1994]. Aus allen diesen Erörterungen können sich Indizien für das mögliche Vorhandensein von „hidden bias“ ergeben. Eine weitere Möglichkeit besteht darin, die Schlußfolgerungen aus einer ganzen Reihe einzelner „observational studies“ zu ziehen, die unter unterschiedlichsten Bedingungen durchgeführt wurden, um „hidden bias“ nach Möglichkeit zu eliminieren, s. [Rosenbaum 2001].

Es soll nun versucht werden, die Frage zu beantworten, in welchem Zusammenhang die Behandlung von „hidden bias“ in „observational studies“ zum Problem der systematischen Fehler in geodätischen Fragestellungen steht.

Auf dem höchsten abstrakten Niveau entstammen beide Probleme der gleichen Quelle: Unmöglichkeit einer wirklich vollständigen Modellbildung, die alle Erscheinungen eindeutig erklärt.

Die Unterschiede in den beiden Fragestellungen sind aber auch sehr ausgeprägt.

In der Geodäsie wird mit fast vollständigen Modellen operiert. Es geht darum, noch die letzten Unstimmigkeiten zu erklären („Wissenschaft von der siebten Stelle“). In den „observational studies“ sind die Modelle meist sehr unvollständig und basieren häufiger auf reinen Spekulationen. Es handelt sich nicht selten darum, nur eine grobe Schätzung zu geben, ob eine Ursache als möglich betrachtet werden darf. Vorausgesetzt, daß auf eine mechanistische oder fatalistische Erklärung des menschlichen Verhaltens und der Reaktionen verzichtet wird, kommt noch die Tatsache dazu, daß unterschiedliche Individuen mit nicht unterscheidbaren Eigenschaften auf die gleiche äußere Wirkung unterschiedlich reagieren können, so daß die Modellbildung streng genommen keine deterministischen Komponenten beinhalten kann¹⁹.

In der Geodäsie werden seit langem grundsätzlich überbestimmte Probleme behandelt. Die Suche nach einer Kompromißlösung resultiert in entsprechenden Residuen. Weisen diese zu hohe Beträge und einen anormalen Verlauf auf, deutet dies auf das Vorhandensein systematischer Fehler hin. In „observational studies“ werden in der Regel stark unterbestimmte Probleme behandelt, bei denen die Ergebnisse keine Zeichen für ein Vorhandensein von „hidden bias“ enthalten können. Die Suche danach erfolgt grundsätzlich auf dem spekulativen Wege, indem unterschiedlichste zusätzliche Hypothesen aufgestellt werden. Diese Vorgehensweise kann auch zum Erfolg führen: Eine Erörterung verschiedener Hypothesen resultiert häufig in der Einführung weiterer Daten und führt dadurch zur

¹⁶S. z.B. [Greenhouse 1982], [Holland 1986] oder [Rubin 1974].

¹⁷Ein unlösbare Problem besteht dabei darin, daß den verschiedenenartigen unbekannten Faktoren unterschiedlichste unbekannte nicht vergleichbare Einheiten zugeordnet sind, so daß eine realitätsnahe Gewichtung fast ausgeschlossen ist.

¹⁸vgl. [Rosenbaum 1989]

¹⁹Einige Aufgabenstellungen aus „observational studies“ ähneln der fiktiven geodätischen Aufgabe, bei der in einer fremden Welt, wo die Summe der drei Winkel eines Dreiecks einen beliebigen Betrag annehmen kann, eine Triangulation durchzuführen ist.

Erhöhung des Informationsgehalts. Nur in der neueren Literatur sind systematische Bestrebungen zu finden, die beispielsweise die Einführung mehrerer Kontrollgruppen oder mehrerer Kontrollindividuen für jedes Untersuchungsindividuum vom Standpunkt aus der Überbestimmung betrachten, um die resultierenden eventuell anormalen Abweichungen als Anzeichen für ein mögliches Vorhandensein von „hidden bias“ zu interpretieren²⁰, s. [Rosenbaum 1987] und [Rosenbaum 1988].

In geodätischen Fragestellungen werden Ausreißer als unerwünschte Störfaktoren betrachtet und nach Möglichkeit mit geeigneten Methoden eliminiert. In „observational studies“ sind Ausreißer häufig die eigentlichen Informationsträger, deren Elimination zu verfälschten Schlussfolgerungen führen kann, s. z.B. [Greenhouse 1982].

Die Schlussfolgerung aus dem geschilderten Vergleich zwischen den Fragestellungen und Vorgehensweisen in der Geodäsie und in den „observational studies“ ist, daß die Natur der Fragestellungen und der Gehalt der zu Verfügung stehenden Informationen zu unterschiedlichen Schwierigkeitsgraden und Methoden führen. Es ist erstaunlich, daß „observational studies“ durch eine Mischung der Statistik, Philosophie und Spekulation in einigen fast hoffnungslosen Fällen sinnvolle Schätzungen liefern.

Obwohl die Erfahrungen aus „observational studies“ kaum einen Beitrag zur Behandlung des Problems der systematischen Fehler in den geodätischen Fragestellungen liefern können, hoffe ich, daß der geneigte Leser mir verzeihen wird, eine mehrseitige Darstellung dieser Studien in diesem Anhang vorgelegt zu haben. Eine Auseinandersetzung mit dieser Thematik kann jedenfalls hilfreich sein, um die Ursachen für häufige Mißinterpretationen der Ergebnisse von „observational studies“ im täglichen Leben besser zu verstehen.

A.3 Literatur

- [Adcock 1878] Adcock, R.J.: A problem in least squares. *The Analyst*, 5, Nr. 2, S. 53–54.
- [Bopp und Krauss 1977] Bopp, H.; Krauss, H.: Die strenge Bestimmung einer ausgleichenden Geraden bei korrelierten Beobachtungen nach der Methode der kleinsten Quadrate. *Zeitschrift für Vermessungswesen*, Nr. 6, S. 261–270.
- [Carpenter 1977] Carpenter, R.G.: Matching when covariates are normally distributed. *Biometrika*, 64, Nr. 2, S. 299–307.
- [Chandrasekaran u.a. 2001] Chandrasekaran, M.G.; Sayed, A.H. und Schubert, K.E.: The degenerate bounded errors-in-variables model. *SIAM Journal on Matrix Analysis and Applications*, 23, Nr. 1, S. 138–166.
- [Cochran 1965] Cochran, W.G.: The planning of observational studies of human populations (with discussion). *Journal of the Royal Statistical Society, Series A*, 128, Part 2, S. 234–255.
- [Cochran und Rubin 1973] Cochran, W.G. und Rubin, D.B.: Controlling bias in observational studies: A review. *Sankhyā: The Indian Journal of Statistics, Series A*, 35, S. 417–446.
- [Cornfield u.a. 1959] Cornfield, J.; Haenszel, W.; Hammond, E.C.; Lilienfeld, A.M.; Shimkin, M.B. und Wynder, E.L.: Smoking and lung cancer: Recent evidence and a discussion of some questions. *Journal of the National Cancer Institute*, 22, Nr. 1, S. 173–203.
- [Cox 1992] Cox, D.R.: Causality: Some statistical aspects. *Journal of the Royal Statistical Society, Series A*, 155, Part 2, S. 291–301.
- [Fierro u.a. 1997] Fierro, R.D.; Golub, G.H.; Hansen, P.C. und O’Leary, D.P.: Regularization by truncated total least squares. *SIAM Journal on Scientific Computing*, 18, Nr. 4, S. 1223–1241.
- [Golub und van Loan 1980] Golub, G.H. und van Loan, C.F.: An analysis of the total least squares problem. *SIAM Journal on Numerical Analysis*, 17, Nr. 6, S. 883–893.
- [Golub u.a. 1999] Golub, G.H.; Hansen, P.C. und O’Leary, D.P.: Tikhonov regularization and total least squares. *SIAM Journal on Matrix Analysis and Applications*, 21, Nr. 1, S. 185–194.
- [Greenhouse 1982] Greenhouse, S.W.: Jerome Cornfield’s contributions to epidemiology. *Biometrics Supplement: Current Topics in Biostatistics and Epidemiology*, 28, S. 33–45.
- [He und Liang 2000] He, X. und Liang, H.: Quantile regression estimates for a class of linear and partially linear errors-in-variables models. *Statistica Sinica*, 10, S. 129–140.
- [Holland 1986] Holland, P.W.: Statistics and causal inference (with discussion). *Journal of the American Statistical Association*, 81, Nr. 396, S. 945–970.

²⁰Diese in der Geodäsie altbewährte Herangehensweise findet in den „observational studies“ erst in letzter Zeit Anwendung.

- [Huwang und Huang 2000] Huwang, L. und Huang, Y.H.S.: On errors-in-variables in polynomial regression — Berkson case. *Statistica Sinica*, 10, S. 923–936.
- [Ming und Rosenbaum 2001] Ming, K. und Rosenbaum, P.R.: A note on optimal matching with variable controls using the assignment algorithm. *Journal of Computational and Graphical Statistics*, 10, Nr. 3, S. 455–463.
- [Rosenbaum 1984] Rosenbaum, P.R.: (1984c). From association to causation in observational studies: The role of tests of strongly ignorable treatment assignment. *Journal of the American Statistical Association*, 79, Nr. 385, S. 41–48.
- [Rosenbaum 1987] Rosenbaum, P.R.: (1987b). The role of a second control group in an observational study (with discussion). *Statistical Science*, 2, Nr. 3, S. 292–316.
- [Rosenbaum 1988] Rosenbaum, P.R.: Sensitivity analysis for matching with multiple controls. *Biometrika*, 75, Nr. 3, S. 577–581.
- [Rosenbaum 1989] Rosenbaum, P.R.: The role of known effects in observational studies. *Biometrics*, 45, S. 557–569.
- [Rosenbaum 1994] Rosenbaum, P.R.: Coherence in observational studies. *Biometrics*, 50, S. 368–374.
- [Rosenbaum 1996] Rosenbaum, P.R.: Observational studies and nonrandomized experiments. In: S. Ghosh und C.R. Rao (Eds.): *Handbook of Statistics, Band 13: Design and Analysis of Experiments*, Elsevier, Amsterdam-Lausanne-New York-Oxford-Shannon-Tokyo, S. 181–197.
- [Rosenbaum 2001] Rosenbaum, P.R.: Replicating Effects and Biases. *The American Statistician*, 55, Nr. 3, S. 223–227.
- [Rosenbaum und Rubin 1985] Rosenbaum, P.R. und Rubin, D.B.: The bias due to incomplete matching. *Biometrics*, 41, S. 103–116.
- [Rubin 1973] Rubin, D.B.: Matching to remove bias in observational studies. *Biometrics*, 29, S. 159–183. Correction: 30 (1974), S. 728.
- [Rubin 1974] Rubin, D.B.: Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of Educational Psychology*, 66, Nr. 5, S. 688–701.
- [Schlesselmann 1978] Schlesselmann, J.J.: Assessing effects of confounding variables. *American Journal of Epidemiology*, 108, Nr. 1, S. 3–8.
- [Steiner und Hajagos 2001] Steiner, F. und Hajagos, B.: Reliability of the estimations of the correlation coefficient. *Acta geodaetica et geophysica Hungarica*, 36, Nr. 4, S. 399–416.
- [Van Huffel und Zha 1993] Van Huffel, S. und Zha, H.: The total least squares problem. In: C.R. Rao (Ed.): *Handbook of Statistics, Band 9: Computational Statistics*, North-Holland, Amsterdam-London-New York-Tokyo, S. 377–408.
- [Watson 2001] Watson, G.A.: Data fitting problems with bounded uncertainties in the data. *SIAM Journal on Matrix Analysis and Applications*, 22, Nr. 4, S. 1274–1293.
- [York 1966] York, D.: Least squares fitting of a straight line. *Canadian Journal of Physics*, S. 1079–1086.
- [York 1969] York, D.: Least squares fitting of a straight line with correlated errors. *Earth and Planetary Science Letters*, 5, S. 320–324.