

Regularisierung unter Berücksichtigung von Resduentoleranzen

vorgelegt von
Diplom-Ingenieur
Julia Kaschenz
aus Berlin

Von der Fakultät VI
der Technischen Universität Berlin
zur Erlangung des akademischen Grades

Doktor der Ingenieurwissenschaften
- Dr.-Ing. -

genehmigte Dissertation

Promotionsausschuss:

Vorsitzender: Prof. Dr. D. Lelgemann
Gutachter: Prof. Dr. L. Gründig
Gutachter: Prof. Dr. Ch. Reigber (GFZ)
Gutachter: PD Dr. S. Petrovic (GFZ)

Tag der wissenschaftlichen Aussprache: 30. Januar 2006

Berlin 2006
D 83

Zusammenfassung

Bei einer Vielzahl von Aufgaben z.B. in der mathematischen Physik und der Geophysik wird man mit sog. inversen Problemen konfrontiert. Oft sind gerade diese inversen Probleme inkorrekt gestellt. Für eine große Klasse von inkorrekt gestellten Problemen liegen zusätzliche Informationen vor, welche Aussagen über die Natur der Lösung enthalten. Unter Nutzung dieser zusätzlichen Informationen wurden bereits von verschiedenen Wissenschaftlern Auswerteverfahren erarbeitet, die allgemein zur Gruppe der Regularisierungsverfahren gehören. Diese Regularisierungsverfahren können in drei Untergruppen eingeteilt werden: die Regularisierung nach A.N. Tikhonov, das regularisierende Verfahren und die stochastische Regularisierung. Bei der Regularisierung nach A.N. Tikhonov werden qualitative Informationen bzgl. der Lösung in den Prozess der Lösungsbestimmung mit einbezogen. Das regularisierende Verfahren benötigt keine zusätzlichen Informationen bzgl. der Lösung des zu untersuchenden Problems, sondern stellt ein „gedämpftes“ Iterationsverfahren dar. Bei der stochastischen Regularisierung werden quantitative zusätzliche Informationen bzgl. der Lösung verwendet, z.B. durch Einbeziehung einer bereits vorhandenen Lösung sowie deren Genauigkeit in die Lösungsbestimmung.

Da diese in der Anwendung und Bedeutung unterschiedlichen Regularisierungsverfahren allerdings eine sehr ähnliche mathematische Formulierung haben, kam es in den letzten Jahren häufig zu unkritischen Anwendungen der Verfahren sowie zu fehlerhaften Interpretationen von erhaltenen Ergebnissen. Lediglich das regularisierende Verfahren kann als unkritisch betrachtet werden. Bei den anderen beiden Regularisierungsverfahren lässt sich der Einfluss der zusätzlichen Informationen auf die zu bestimmende Lösung nur sehr schlecht oder überhaupt nicht einschätzen. Dies kann dazu führen, dass die Residuen der ausgewerteten Daten unzumutbar groß sind, die neue Lösung aber sehr gut zu der zusätzlichen Information bzgl. der Lösung passt. Die zusätzlichen Informationen entstammen aber oft älteren Modellen, welche auf älteren ungenaueren Daten basieren. Da es somit nicht sinnvoll erscheint, die Informationen der neu auszuwertenden Daten auf Kosten der zusätzlichen Informationen zu vernachlässigen, wurden an einigen Stellen in der Literatur Überlegungen dahingehend durchgeführt, wie der Einfluss der zusätzlichen Informationen „optimal“ gewählt werden kann. Bislang scheint es dafür allerdings noch keine wirklich „optimale“ Methode zu geben.

Somit erschien es zweckmäßig, in dieser Arbeit ein Regularisierungsverfahren zu erarbeiten, welches nicht die genannten Nachteile aufweist. Bei diesem können zusätzlich zum allgemeinen Regularisierungsproblem Bedingungen für die Residuen eingeführt werden. Diese können z.B. bewirken, dass sämtliche Residuen nach der Auswertung innerhalb der in der Gewichtsmatrix festgelegten Genauigkeitsgrenzen der Daten liegen. Damit kann gewährleistet werden, dass so viele Informationen wie möglich aus den neu auszuwertenden Daten gewonnen werden können. Da es sich bei den eingeführten Bedingungen nicht um sog. Gleichheitsbedingungen handelt sondern um Ungleichheitsbedingungen, war es notwendig, einen geeigneten Auswertalgorithmus zur Lösung des somit erweiterten Regularisierungsproblems zu erarbeiten. Dieser Algorithmus wurde anhand von fünf synthetischen Beispielen getestet und für die Auswertung von Radiookkultationsdaten verwendet. Dabei zeigte sich, dass das erarbeitete Verfahren nicht nur grundsätzlich funktioniert, sondern tatsächlich die erwähnten Nachteile nicht besitzt. Gleichzeitig ist es mit diesem Verfahren möglich, zu überprüfen, ob bereits vorhandene Lösungen des untersuchten Problems mit den neuen Daten innerhalb ihrer Genauigkeiten konsistent sind und ob und wie stark die neuen Daten zur Verbesserung der ursprünglichen Lösung beitragen können.

Inhaltsverzeichnis

Abbildungsverzeichnis	7
Tabellenverzeichnis	8
1 Einleitung	9
1.1 Motivation und Zielsetzung	9
1.2 Überblick	10
2 Bestehende Regularisierungsverfahren	12
2.1 Korrekt und inkorrekt gestellte Probleme	12
2.2 Regularisierungsverfahren	13
2.2.1 Regularisierung nach A.N. Tikhonov	14
2.2.2 Regularisierendes Verfahren	18
2.2.3 Stochastische Regularisierung	22
2.3 Verfahren zur Bestimmung des Regularisierungsparameters	26
2.4 Diskussion	27
3 Möglichkeiten zur Einführung von Toleranzenbedingungen für Residuen	28
3.1 Veranschaulichung der allgemeinen Idee	28
3.2 Möglichkeiten zur Einführung von Toleranzenbedingungen	29
3.2.1 Variante 1: $v^2 \leq t^2$	31
3.2.2 Variante 2: $v \leq t$ und $-v \leq t$	32
3.3 Ausarbeitung der Algorithmen	32
3.3.1 Transformation eines LSI-Problems in ein LDP-Problem	34
3.3.2 Transformation eines LDP-Problems in ein NNLS-Problem	36
3.3.3 Lösungsalgorithmus für NNLS-Probleme	37
3.3.4 Beweis, dass die direkte Transformation eines LSI-Problems in ein NNLS-Problem nicht gültig ist	39
3.3.5 Veranschaulichung der Ausgleichung mit Toleranzenbedingungen	40
4 Vergleich der unterschiedlichen Bedingungsungleichungen anhand synthetischer Beispiele	43
4.1 Beispiele: nicht regularisiert und gleichgewichtet	43
4.2 Beispiele: regularisiert und gleichgewichtet	52
4.3 Beispiele: regularisiert und unterschiedlich gewichtet	55
5 Anwendungsbeispiel	60
5.1 Prinzip der Radiookkultation	60
5.2 Ein Standardverfahren zur Ableitung der spezifischen Feuchte und Feuchttemperatur	62
5.3 Regularisierung mit Residuentoleranzen	64
5.4 Ausgewählte Beispiele	66
6 Schlussbetrachtung	78
Literaturverzeichnis	80

A	Anhang	85
A.1	Funktionalmodelle und Daten der Beispiele 1-5	85
A.2	Liste der verwendeten Radiookkultations-Inputdateien	87
A.3	Globale Genauigkeitsschätzungen für die Radiookkultations-Refraktivitäten	89
A.4	Varianzen der Vorinformationen	93
A.5	Zusätzliche Abbildungen zum Kapitel 5	95

Abbildungsverzeichnis

3.1	Zwei mögliche Lösungen mit unterschiedlichen Residuen	29
4.1	Zielfunktion für Beispiel 1 und 1b)	45
4.2	Zielfunktion für Beispiel 2 und 2b)	46
4.3	Zielfunktion für Beispiel 3 und 3f)	49
4.4	Zielfunktion für Beispiel 5 und den „Grenzfall“ 5e)	49
4.5	Zielfunktion für das Beispiel 4	50
4.6	Zielfunktion für die „Grenzfälle“ 4a) und 4b)	50
4.7	Zielfunktion für Beispiel 4c) und 4d)	51
5.1	Prinzip des Radiookkultations-Verfahrens	60
5.2	Schematischer Auswerteablauf	61
5.3	Verteilung der Okkultationsprofile	66
5.4	Residuenplots und Fehlerannahmen	69
5.5	Temperaturvergleich	70
5.6	Vergleich der spezifischen Feuchten	71
5.7	Residuenplots und Fehlerannahmen bei Übergewichtung der Pseudobeobachtungen	73
5.8	Vergleich der spezifischen Feuchten aus 1DVAR	74
5.9	Vergleich der spezifischen Feuchten aus dem erweiterten Regularisierungsverfahren	75
A.1	Temperaturvergleich	95
A.2	Vergleich der spezifischen Feuchten	96
A.3	Vergleich der spezifischen Feuchten aus 1DVAR	97
A.4	Vergleich der spezifischen Feuchten aus dem erweiterten Regularisierungsverfahren	98

Tabellenverzeichnis

4.1	Ergebnisse der Ausgleichung nach kleinsten Quadraten für alle Beispiele	44
4.2	Ergebnisse für Beispiel 1 mit unterschiedlichen Toleranzen für die Residuen . . .	44
4.3	Ergebnisse für Beispiel 2 mit unterschiedlichen Toleranzen für die Residuen . . .	45
4.4	Ergebnisse für Beispiel 3 mit unterschiedlichen Toleranzen für die Residuen . . .	46
4.5	Ergebnisse für Beispiel 4 mit unterschiedlichen Toleranzen für die Residuen . . .	47
4.6	Ergebnisse für Beispiel 5 mit unterschiedlichen Toleranzen für die Residuen . . .	48
4.7	Ergebnisse der Ausgleichung nach kleinsten Quadraten (mit Regularisierung) . .	52
4.8	Ergebnisse für Beispiel 1 mit Regularisierung bei unterschiedlichen Toleranzen für die Residuen	53
4.9	Ergebnisse für Beispiel 2 mit Regularisierung bei unterschiedlichen Toleranzen für die Residuen	53
4.10	Ergebnisse für Beispiel 3 mit Regularisierung bei unterschiedlichen Toleranzen für die Residuen	54
4.11	Ergebnisse für Beispiel 4 mit Regularisierung bei unterschiedlichen Toleranzen für die Residuen	54
4.12	Ergebnisse für Beispiel 5 mit Regularisierung bei unterschiedlichen Toleranzen für die Residuen	55
5.1	Kritische Höhenbereiche bei den ausgewählten Profilen	72

1. Einleitung

1.1 Motivation und Zielsetzung

In der Geodäsie, Geophysik und vielen anderen Wissenschaftsbereichen trifft man sehr häufig auf sog. inkorrekt gestellte Probleme. Darunter werden zum einen Probleme verstanden, deren Lösungsbestimmung durch das Vorliegen von Messfehlern in den Ausgangsdaten erschwert ist, d.h. die berechnete Lösung ändert sich bei geringfügig kleinen Änderungen der Messfehler signifikant. Zum anderen versteht man unter inkorrekt gestellten Problemen auch oft Probleme, die für sich allein (ohne Verwendung zusätzlicher Informationen) unlösbar sind, da in den auszuwertenden Daten nicht die Informationen enthalten sind, die zur Ableitung der gesuchten Parameter notwendig wären. Um dennoch inkorrekt gestellte Probleme lösen zu können, werden häufig Regularisierungsverfahren angewendet. Dabei werden zusätzliche Informationen oder Bedingungen (Vorinformationen) bzgl. der gesuchten Lösung innerhalb des Prozesses der Lösungsbestimmung berücksichtigt. So können z.B. aus der physikalischen Bedeutung der gesuchten Lösung Eigenschaften dieser Lösung abgeleitet werden, welche dann meist die zusätzlichen Informationen darstellen. Es können aber auch aus bereits vorhergehenden Untersuchungen Werte für einzelne oder alle gesuchten Parameter vorliegen, für die meist ebenso Genauigkeiten bekannt sind. Diese vorhandenen Parameter und Genauigkeiten werden vor allem bei den eigentlich unlösbaren Problemen als zusätzliche Informationen einbezogen.

Da viele gerade für die Geowissenschaften interessanten Probleme, wie z.B. die Ableitung des Gravitationspotenzials der Erde aus Satellitenmessungen, die Bestimmung von Formationen im Erdinnern aus seismischen Wellen oder auch die Untersuchung der Erdatmosphäre anhand von Radiowellen, inkorrekt gestellte Probleme darstellen und nur durch die Verwendung von Regularisierungsverfahren lösbar gemacht werden können, entstand die Motivation zur näheren Untersuchung dieser Verfahren sowie der Festlegung des Regularisierungseinflusses. Diese Zielsetzung wurde verstärkt, da vielfach festgestellt werden musste, dass trotz einiger kritischer Bemerkungen, welche in der Literatur hinsichtlich der Wahl des Einflusses der Regularisierung zu finden sind, es oft zu einer unkritischen Anwendung der Regularisierungsverfahren kommt, wodurch die Gewinnung neuer Informationen aus aktuellen, häufig sehr genauen Messungen verhindert wird. So erfolgt in einigen Wissenschaftsbereichen durch eine unangemessene Wahl des Regularisierungsparameters meist nur eine mehr oder weniger gute Reproduktion bereits vorhandener Modellparameter, obwohl durch die heutigen neuen innovativen Beobachtungsverfahren (wie z.B. die Satellitenmessverfahren) wesentlich mehr Einblick in einzelne Wissenschaftsbereiche möglich wäre.

Das hauptsächliche Problem bei der Anwendung von Regularisierungsverfahren besteht in der geeigneten Wahl des Regularisierungsparameters. Dieser regelt den Einfluss der zusätzlichen Informationen bzw. Bedingungen innerhalb des Prozesses der Lösungsbestimmung. Oft wird in der Literatur zur Regularisierung kritisch darauf hingewiesen, dass bei zu großzügiger Wahl des Regularisierungsparameters der Einfluss der Regularisierung (der zusätzlichen Informationen bzw. Bedingungen) zu stark ist. Somit werden die in den auszuwertenden Daten enthaltenen Informationen nicht angemessen genutzt und mit der Wahl des Regularisierungsparameters die Ergebnisse faktisch vorgegeben. Ist andererseits der Einfluss der Regularisierung zu gering, so haben die Messfehler einen zu starken Einfluss - die Lösung zeigt ein unphysikalisches oder

unlogisches Verhalten.

Es existieren bereits eine Vielzahl von Verfahren zur Bestimmung des „optimalen“ Regularisierungsparameters (siehe z.B. [Kusche 2002]). „Optimal“ bedeutet hierbei, dass ein beliebig gewähltes Kriterium von der Lösung bzw. den Residuen erfüllt wird, welche unter Verwendung des „optimalen“ Regularisierungsparameters berechnet werden konnten. So liegt z.B. ein „optimaler“ Regularisierungsparameter vor, wenn die Summe der Residuenquadrate unterhalb eines vorgegebenen Schrankenwertes liegt (Diskrepanzkriterium nach Morozov) oder wenn die Varianzkomponenten unterschiedlicher Beobachtungsgruppen ein bestimmtes Verhältnis zueinander aufweisen (Varianzkomponenten-Schätzung). Der Wert des „optimalen“ Regularisierungsparameters kann somit je nach verwendetem Verfahren (Kriterium) für dessen Bestimmung sehr unterschiedlich sein, so dass diesem immer eine gewisse Willkür anhaftet.

Innerhalb dieser Arbeit wird der Einfluss der Regularisierung auf die Lösung nur dann als „optimal“ betrachtet, wenn sämtliche Residuen der Beobachtungen minimal sind, unter Berücksichtigung dass alle Residuen innerhalb von Toleranzen liegen. Dieses Kriterium kann bei Auswertungen mittels Regularisierung für alle denkbaren Messwerte genutzt werden. Die Ziele der vorliegenden Arbeit sind somit:

- die Zusammenstellung verbreiteter Missverständnisse bzgl. unterschiedlicher Regularisierungsverfahren,
- die Ausarbeitung eines effizienten Regularisierungsverfahrens, bei dem durch eine gezielte Berücksichtigung der Genauigkeiten der auszuwertenden Daten der Einfluss der Regularisierung auf die Lösung kontrolliert wird,
- die Untersuchung der Eigenschaften und Test des erarbeiteten Auswerteverfahrens anhand von synthetischen Beispielen sowie
- die Veranschaulichung der Vorteile des erarbeiteten Regularisierungsverfahrens an einem realen Anwendungsbeispiel.

Durch die Anwendung des erarbeiteten Regularisierungsverfahrens sollen einerseits möglichst viele neue Informationen aus den auszuwertenden Daten gewonnen werden können und andererseits sich für die erhaltenen Lösungen neue Interpretationsmöglichkeiten ergeben. Es ist zwar nicht möglich, auf die Einbeziehung von Vorinformationen bei der Lösung inkorrekt gestellter Probleme zu verzichten, aber mit angemessenen Bedingungen für die Residuen der auszuwertenden Daten soll eine Überbewertung der Vorinformationen verhindert werden.

1.2 Überblick

Innerhalb des **Kapitels 2** erfolgt zunächst die Definition von korrekt und inkorrekt gestellten Problemen sowie der Regularisierung. Es werden die drei unterschiedlichen Arten der Regularisierung dargestellt: die Regularisierung nach A.N. Tikhonov, das regularisierende Verfahren sowie die stochastische Regularisierung. Aufgrund umfangreicher Recherchen zu diesen Regularisierungsverfahren konnten ihre zeitliche Entwicklung, der mathematische Rahmen und einzelne Besonderheiten der Verfahren herausgearbeitet werden. Die Vor- und Nachteile der Verfahren werden beschrieben und vor allem wird auf die unterschiedliche Bedeutung des Regularisierungsparameters eingegangen. Ebenso wird kurz auf die allgemein zur Anwendung kommenden Verfahren eingegangen, mit denen die Bestimmung des „optimalen“ Regularisierungsparameters

möglich sein soll.

Das **Kapitel 3** stellt den Kern dieser Arbeit dar. Darin wird auf die allgemeine Idee der Einführung von Toleranzen (Bedingungen) für die Residuen der Beobachtungen eingegangen. Weiterhin erfolgt die allgemeine analytische Ausarbeitung zweier denkbarer mathematischer Realisierungen zur Einführung von Toleranzenbedingungen für die Residuen sowie des Algorithmus zur Lösung eines um die Toleranzenbedingungen erweiterten Ausgleichungsproblems. Abschließend wird zum besseren Verständnis der ausgearbeiteten Formeln und des Algorithmus die Auswertung eines einfachen Beispiels veranschaulichend erläutert.

An fünf verschiedenen synthetischen Beispielen werden dann im **Kapitel 4** die Eigenschaften der beiden unterschiedlichen Varianten der Realisierung der Toleranzenbedingung für die Residuen untersucht sowie die Stabilität und Korrektheit des erarbeiteten Auswertungsverfahrens. Für einen unabhängigen Vergleich der Lösungen werden sämtliche Berechnungen ebenso mittels eines heuristischen Verfahrens durchgeführt. Durch umfassende tabellarische und grafische Vergleiche zwischen den Ergebnissen der unterschiedlichen Varianten der Realisierung der Toleranzenbedingung und der Verfahren zur Lösungsbestimmung ist es möglich gewesen, Eigenschaften der Fläche der zu minimierenden Funktion abzuleiten, die einen positiven, negativen oder neutralen Einfluss auf die Lösungsbestimmbarkeit ausüben. Am Ende des Kapitels wird mittels eines geometrisch anschaulichen Beispiels an den numerischen Ergebnissen auf den möglichen negativen Einfluss der Regularisierung eingegangen, wobei eine beabsichtigte Übergewichtung der Vorinformationen erfolgte. Im Vergleich dazu dienen die Ergebnisse des ausgearbeiteten Regularisierungsverfahrens, welche sehr gut die Wirkungsweise der einbezogenen Residuentoleranzen verdeutlichen.

In **Kapitel 5** wird das neue Regularisierungsverfahren mit Residuentoleranzen auf die Ableitung von Feuchttemperaturen und spezifischen Feuchten aus Radiookkultations-Refraktivitäten angewendet, welche aus CHAMP¹-Messungen bestimmt wurden. Dazu erfolgt zunächst eine kurze Einführung in das Radiookkultationsverfahren. Anschließend wird ein Standardverfahren zur Ableitung der Feuchttemperaturen und spezifischen Feuchten aus Refraktivitäten (1DVAR) vorgestellt, wobei kritisch auf die aus den Formeln ersichtlichen Nachteile eingegangen wird. Ebenso werden die Formeln für die Anwendung der Regularisierung mit Residuentoleranzen zur Ableitung obiger Atmosphärenparameter dargestellt. Die Vorteile, Nachteile und Unterschiede dieser beiden Regularisierungsverfahren werden durch die Auswertung von 63 global verteilten Radiookkultationsprofilen veranschaulicht. Für sechs ausgewählte Profile werden detailliert die Unterschiede in den Residuen der Refraktivitäten, in den Feuchttemperaturen und den spezifischen Feuchten hervorgehoben, sowohl für den „korrekt“ gewichteten Fall als auch den übergewichteten Fall. Dabei wird besonders auf die Möglichkeiten bei der Interpretation der Ergebnisse eingegangen.

Im **Kapitel 6** folgen resümierende Schlussbetrachtungen.

¹CHALLENGING Minisatellite Payload

2. Bestehende Regularisierungsverfahren

Ausgehend von den Definitionen für korrekt und inkorrekt gestellte Probleme wird nachfolgend ein kurzer Überblick zu den bestehenden Regularisierungsverfahren mit ihren Vor- und Nachteilen gegeben. Abschließend wird auf Verfahren eingegangen, mit denen die Bestimmung des Regularisierungsparameters erfolgen kann.

2.1 Korrekt und inkorrekt gestellte Probleme

Bei einer Vielzahl von Aufgaben z.B. in der mathematischen Physik, der Geophysik, Mustererkennung und in der Astrophysik erfolgt die Unterteilung in sog. direkte und inverse Probleme. Was versteht man aber unter dem Begriff „inverse Probleme“? Zu was ist das jeweilige Problem invers?

Man nennt zwei Probleme invers zueinander, wenn die mathematische Formulierung des einen Problems zu der des anderen führt. Aus hauptsächlich historischen Gründen bezeichnet man das eine dieser Probleme (gewöhnlich das einfachere oder zuerst untersuchte) als das direkte Problem und das andere als inverses Problem ([Engl, Hanke und Neubauer 1996], aus dem Englischen übersetzt)¹.

In vielen Anwendungsbereichen sind oft gerade die inversen Probleme sog. inkorrekt gestellte Probleme. Ebenso können aber auch direkte Probleme zu der Gruppe inkorrekt gestellter Probleme gehören.

Anfang des 20. Jahrhunderts beschäftigte sich der französische Mathematiker J. Hadamard mit der Lösbarkeit von partiellen Differentialgleichungen. Werden diese Gleichungen mittels Integration gelöst, so erhält man aufgrund der Integrationskonstanten immer eine Klasse von Lösungen und nur durch die Hinzunahme zusätzlicher Randbedingungen ist die Auswahl einer Lösung möglich. In diesem Zusammenhang betrachtete Hadamard drei Fragen (Theorem von Cauchy-Kowalewski), die sich stets bzgl. der Lösbarkeit von Problemen stellen, hier bezogen auf das Problem von Cauchy ([Hadamard 1932], aus dem Französischen übersetzt):

- *Hat das Problem von Cauchy eine Lösung?*
- *Gibt es nicht nur eine einzige Lösung?*
(In anderen Worten: Ist das Problem korrekt gestellt?)
- *Und schließlich, wie kann man diese [Anm.: einzige] Lösung berechnen?*

Daraus lässt sich eine sehr allgemeine Definition für die Korrektheit oder Inkorrektheit eines Problems formulieren: Ein Problem heißt korrekt gestellt, wenn es eine eindeutige Lösung besitzt und diese bestimmt werden kann, ohne dass zusätzliche Informationen oder Annahmen jedweder Art einbezogen werden müssen. Anderenfalls liegt ein inkorrekt gestelltes Problem vor. Fälschlicherweise wird in der Literatur (siehe z.B. [Fedotov 1990]) häufig die dritte Frage bei Hadamard durch eine andere Frage ersetzt (z.B. Ist die Aufgabe stabil lösbar?) und trotzdem

¹Bei anderen Autoren (z.B. [Ilk 1993] und [Parker 1994]) ist im Falle des Modells $y = f(x)$ die Berechnung von y aus x das direkte und von x aus y das inverse Problem.

behauptet, dass Hadamard diese Fragen betrachtete.

Eine weitere Definition der Korrektheit eines Problems ist auch von Tikhonov bekannt ([Fedotov 1990], aus dem Russischen übersetzt):

Eine Aufgabe heißt korrekt gestellt, falls:

- *es im Voraus bekannt ist, dass mindestens eine Lösung existiert,*
- *diese Lösung eindeutig ist und*
- *unendlich kleine Änderungen der ursprünglichen Daten lediglich zu unendlich kleinen Änderungen der Lösung führen.*

Somit unterscheidet sich die Tikhonov'sche Definition nur im dritten Punkt von der von Hadamard. Dieser dritte Punkt bei Tikhonov besagt, dass die Lösung stetig von den ursprünglichen Daten abhängig sein soll.

Im Folgenden soll die aus den drei Fragen von Hadamard formulierte Definition als Definition für die Korrektheit eines Problems verwendet werden, da diese allgemein gefasst ist und sich der dritte Punkt bei Tikhonov eher auf die Stabilität eines Problems bezieht als auf die Korrektheit.

Für eine große Klasse von inkorrekt gestellten Problemen liegen zusätzliche Informationen vor, welche Aussagen über die Natur der Lösung enthalten, z.B. Monotonie, Glattheit, Konvexität. Werden diese Informationen genutzt, so können effiziente Algorithmen zur Lösung von inkorrekt gestellten Problemen erarbeitet werden ([Tikhonov et al. 1995]). Im folgenden Abschnitt soll die zeitliche Entwicklung derartiger Algorithmen aufgezeigt und auf ihre unterschiedlichsten Anwendungsmöglichkeiten aufmerksam gemacht werden.

2.2 Regularisierungsverfahren

In der Literatur findet man zum Begriff „Regularisierung“ beispielsweise:

„Der Übergang zu Näherungsverfahren, die gegenüber Ungenauigkeiten des Modells und der zur Verfügung stehenden Daten weniger empfindlich sind, soll Regularisierung (im weiteren Sinne) genannt werden.“ [Friedrich, Hofmann und Tautenhahn 1979]

Allgemein gesagt, ist Regularisierung die Approximation von inkorrekt gestellten Problemen durch eine Klasse benachbarter korrekt gestellter Probleme ([Engl, Hanke und Neubauer 1996], aus dem Englischen übersetzt).

Diese sehr allgemeinen Definitionen der Regularisierung schließen die Vielzahl von unterschiedlichen Verfahren, die in der Literatur zu finden sind, mit ein, welche meist rein formal nur schwer unterschieden werden können. Aufgrund dieser formalen Ähnlichkeit kam es in den vergangenen Jahren recht häufig zu Missverständnissen bei der Anwendung dieser unterschiedlichen Regularisierungsverfahren, welche anhand der folgenden Darstellungen der wesentlichen Unterschiede und Eigenschaften dieser einzelnen Verfahren möglichst behoben werden sollen. Dabei sollen im Wesentlichen drei Arten der Regularisierungsverfahren unterschieden werden: die Regularisierung nach A.N. Tikhonov, das regularisierende Verfahren und die stochastische Regularisierung.

In [Kusche 2002] u.a. Veröffentlichungen findet man für die Regularisierungsverfahren eine anscheinend endlose Vielfalt an Namen, wie z.B. Iterierte Tikhonov-Regularisierung, Truncated

Singular Value Decomposition und Kollokation². Tatsächlich resultiert diese Namensvielfalt nicht aus der Vielfalt von Regularisierungsverfahren, sondern aus den unterschiedlichen mathematischen Methoden, die für die Lösungsbestimmung bei regularisierten Problemen möglich sind. Da diese Methoden lediglich zu rechentechnischen Unterschieden führen sollen, wird darauf nicht näher eingegangen.

2.2.1 Regularisierung nach A.N. Tikhonov

Der russische Mathematiker A.N. Tikhonov kann als derjenige betrachtet werden, der sich als erster mit inkorrekt³ gestellten Problemen (bzw. inversen Problemen) und ihrer Lösung mittels Regularisierung befasste. Im Jahr 1943 verfasste er seine erste Veröffentlichung zu diesem Thema, in der zu finden ist ([Tikhonov 1943], aus dem Englischen übersetzt):

Der gewöhnliche Weg inverse Probleme zu lösen [Anm.: es wird keine Überbestimmung verlangt], besteht darin, eine gewisse Auswahl zu treffen. Für eine willkürlich gewählte (genügend große) Klasse von möglichen Strukturen eines Mediums [Anm.: eine Klasse möglicher Lösungen] wird das dazugehörige physikalische Feld [Anm.: die ausgeglichenen Beobachtungen] berechnet und dann die Lösung ausgewählt, so dass das berechnete physikalische Feld zu den Messungen die kleinsten Abweichungen hat. Um dieses Auswahlverfahren auf einer soliden Grundlage basieren lassen zu können, ist es notwendig, sich auf das Vorliegen bestimmter Regelmäßigkeiten zu beziehen, die das jeweilige Problem aufweisen muss:

- *Man muss zunächst das Eindeutigkeits-Theorem für eine direkte Übereinstimmung nachweisen, d.h. man muss prüfen, dass nicht zwei unterschiedliche Medientypen [Anm.: Lösungen] das gleiche dazugehörige physikalische Feld [Anm.: gleiche ausgeglichene Beobachtungen] besitzen. Wenn ja, dann kann man von umgekehrter Übereinstimmung sprechen. Ohne diese hat die Auswahlmethode keinen Sinn.*
- *Die Übereinstimmung zwischen dem berechneten und gemessenen Feld stellt keine absolute Übereinstimmung dar, da das Auswahlverfahren lediglich ein Approximationsverfahren ist. Daher muss zusätzlich die Stabilität des inversen Problems untersucht werden bzw. die Kontinuität der inversen Abbildung. D.h. man muss prüfen, ob bei geringfügiger Änderung des Behelfsfeldes [Anm.: gemessenen Feldes] die dazugehörige Struktur des Mediums [Anm.: die Lösung] große oder kleine Abweichungen zum aktuellen Medium aufweist.*

Mit dieser allgemeinen Darstellung der notwendigen Vorgehensweise bei der Behandlung inverser Probleme stellt Tikhonov die Idee für das Verfahren vor, welches er später als Regularisierung bezeichnete. Er geht dabei auf die Schwierigkeiten ein, auf die man bei der Lösung inkorrekt gestellter Probleme trifft: Uneindeutigkeit und Instabilität der Lösung. Besonders interessant ist, dass er hervorhebt, dass derartige Probleme lediglich approximativ gelöst werden können und dass er für die Auswahl der „besten“ Lösung die Abweichungen zu den Messungen heranzieht. Diese nachvollziehbaren und sehr wichtigen Sachverhalte gerieten in den letzten Jahren häufig in Vergessenheit.

Die beiden „Regelmäßigkeiten“, die nach Tikhonov vorliegen müssen, damit das Auswahlverfahren sinnvoll eingesetzt werden kann, stellen eine Methode dar, um überprüfen zu können, ob

²In [Rummel et al. 1979] wird der Zusammenhang und die Ähnlichkeit zwischen Kollokation und Regularisierung aufgezeigt.

³Dabei bezieht sich Tikhonov selbstverständlich immer auf seine Definition für inkorrekt gestellte Probleme (siehe Abschnitt 2.1).

tatsächlich ein inkorrekt gestelltes Problem vorliegt oder nicht (bezogen auf seine Definition).

An einigen Stellen in der Literatur findet man, dass S. Twomey und A.N. Tikhonov etwa zeitgleich (1963) das erste Verfahren veröffentlichten, welches Zwänge (bzw. zusätzliche Informationen) einbezieht und für die Lösung inkorrekt gestellter Probleme genutzt werden kann (siehe z.B. [Rodgers 2000]). Dies ist sicherlich korrekt, wenn man als erste Veröffentlichung diejenige betrachtet, in der die erste mathematische Formulierung des Auswerteverfahrens zu finden ist. Die Grundidee des Verfahrens wurde jedoch schon 1943 von Tikhonov [Tikhonov 1943] formuliert.

Seit den sechziger Jahren beschäftigte sich eine Vielzahl von Wissenschaftlern mit derartigen Auswerteverfahren, welche aber immer auf die Tikhonov'sche Idee zurückgeführt werden können. Insofern wird nachfolgend nur auf die Entwicklungsstufen des Tikhonov'schen Auswerteverfahrens und kurz auf Twomey's Abhandlungen eingegangen.

Tikhonov 1963:

Liegen fehlerfreie Daten l vor, so kann mit $\bar{x} = A^{-1}l$ die sog. „exakte“⁴ Lösung $\bar{x}$ des Problems berechnet werden, wobei die Matrix A den funktionalen Zusammenhang zwischen den Daten und der Lösung repräsentiert. Sind die Daten fehlerbehaftet (wahrer Fehler δ), so kann nur eine Näherungslösung x bestimmt werden, für die gilt [Tikhonov 1963a]:

$$\begin{aligned} Ax &= l + \delta, \\ x - \bar{x} &\leq \epsilon, \\ l + \delta &= \bar{l}. \end{aligned} \tag{2.1}$$

Ist ein Problem korrekt gestellt, so existiert eine Beziehung zwischen δ und ϵ , welche zeigt wie sich die Fehler δ auf die Lösung x übertragen. Das Problem ist somit stetig. Inkorrekt gestellte Probleme sind dagegen unstetig und es existiert keine einfache Beziehung zwischen δ und ϵ . Daher müssen derartige Probleme regularisiert werden, d.h. es muss ein sog. Ersatzproblem formuliert werden, welches stetig ist und eine günstige Fehlerübertragung von den Messungen auf die Lösung aufweist. Um ein Ersatzproblem mit diesen Eigenschaften aufzustellen, werden zusätzliche Informationen oder Zwänge bzgl. der Lösung des Problems folgendermaßen eingeführt [Tikhonov 1963b]:

$$M^\alpha[x, \bar{l}] = N[x, \bar{l}] + \alpha\Omega[x], \tag{2.2}$$

$$N[x, \bar{l}] = \int_a^b [Ax - \bar{l}]^2 dx, \tag{2.3}$$

wobei $M^\alpha[x, \bar{l}]$ das sog. glättende Funktional darstellt, $N[x, \bar{l}]$ die Summe der Residuenquadrate, $\Omega[x]$ die mathematische Formulierung der zusätzlichen Informationen bzw. Zwänge und α den Regularisierungsparameter.

Die Aufgabe besteht zum einen darin, geeignete zusätzliche Informationen oder Zwänge zu finden, wodurch das jeweilige Problem stabilisiert werden kann, z.B. indem verlangt wird, dass die Lösung möglichst glatt sein soll. Zum anderen muss der Regularisierungsparameter so gewählt werden, dass das Problem stabil bleibt, aber die Lösung x möglichst dicht an die „exakte“ Lösung $\bar{x}$ herankommt. Da $\bar{x}$ in der Regel nicht bekannt ist, ist die Wahl von α recht

⁴Bei Tikhonov und auch bei Twomey (z.B. in [Twomey 1963]) wird angenommen, dass genau so viele Messdaten vorliegen, wie Unbekannte gesucht werden. Die Grundidee der Regularisierung zur Stabilisierung eines inkorrekt gestellten Problems ist aber analog auf überbestimmte Probleme übertragbar.

schwierig und nie „perfekt“ möglich.

Twomey 1963:

In [Twomey 1963] wird als mögliche zusätzliche Information die Forderung nach einer glatten Lösung mathematisch wie folgt formuliert⁵ (entspricht der approximativen Minimierung des Integrals der zweiten Ableitung der Lösung x):

$$\Omega[x] = \sum_i (x_{i-1} - 2x_i + x_{i+1})^2. \quad (2.4)$$

Im Gegensatz zu Tikhonov betrachtet Twomey die Minimierung von $\Omega[x]$, also die Einbeziehung der zusätzlichen Informationen, als das Hauptziel bei der Minimierung von $M^\alpha[x, \bar{l}]$. Bei der Auswertung fehlerbehafteter Messdaten scheint es jedoch grundsätzlich sinnvoller zu sein, die Minimierung der Summe der Residuenquadrate in den Vordergrund zu stellen.

Weiterhin schlägt er vor, die Abweichung der aktuellen Lösung zu einer beliebigen vorgegebenen hypothetischen Lösung zu minimieren, die dritte Ableitung der Lösung x zu minimieren oder verschiedene Zwänge sinnvoll mit unterschiedlich starkem Einfluss zu kombinieren. Wie gelangt man aber an eine hypothetische Lösung, zu der die aktuelle Lösung kleinste Abweichungen haben soll? *Ist die allgemeine Form und Größenordnung der Lösung bekannt, so kann eine „Testlösung“ konstruiert werden ...* ([Twomey 1963], aus dem Englischen übersetzt).

Geht man dann so vor, wie es Twomey vorschlägt und betrachtet vorrangig die Minimierung des Terms $\Omega[x]$, so stellt man fest, dass die Lösung x nach der Minimierung von $M^\alpha[x, \bar{l}]$ sich nicht sehr von der „Testlösung“ unterscheiden wird. Welchen Sinn hat aber eine derartig bestimmte neue Lösung? Keinen, denn zum einen ist die „Testlösung“ nur recht grob und allgemein und zum anderen könnten durch die vorrangige Minimierung von $\Omega[x]$ die neuen Messdaten keine wesentliche Verbesserung / Veränderung der „Testlösung“ bewirken.

Leider hat sich gerade diese Art der Benutzung von zusätzlichen Informationen bzgl. der Lösung und diese Minimierungsvorgehensweise in den meisten praktischen Anwendungen durchgesetzt. Auch Twomey betont, dass die zusätzlichen Informationen geeignet gewählt werden müssen und der Regularisierungsparameter so gewählt werden muss, dass der Wert von $N[x, \bar{l}]$ (entspricht der Summe der Residuenquadrate) akzeptabel ist⁶.

Twomey 1965:

In [Twomey 1965] sind weitere mögliche zusätzliche Bedingungen bzgl. der Lösung x , die zur Stabilisierung eines Minimierungsproblems verwendet werden können, zu finden, wodurch das Ausmaß der Oszillation von x festgesetzt wird, z.B. mit:

$$\Omega[x] = \sum_i \left(x_i - \frac{1}{n} \sum_j x_j \right)^2 \quad (\text{für Probleme mit gleichartigen Unbekannten}), \quad (2.5)$$

$$\Omega[x] = \sum_i \frac{(x_i - p_i)^2}{p_i} \quad (\text{für allgemeine Probleme}). \quad (2.6)$$

⁵Es ist anzumerken, dass Twomey die mathematische Formulierung dieser Forderung aus [Phillips 1962] entnommen hat. In [Phillips 1962] wird auch erwähnt, dass diese Schreibweise nur dann zulässig ist, wenn sämtliche Unbekannten bzgl. einer gemeinsamen Variablen äquidistant sind. Weiterhin werden in [Phillips 1962] sehr anschauliche Beispiele aufgeführt.

⁶Die Ansicht von Tikhonov, die einzelnen Abweichungen zu den Messungen (Residuen) zu betrachten, scheint allerdings sinnvoller zu sein, da die Summe der Residuenquadrate einen akzeptablen Wert annehmen kann, auch wenn einzelne Residuen sehr groß sind.

Dabei stellen p_i eine geschätzte a-priori Lösung und n die Anzahl der zu bestimmenden Parameter dar. Durch Verwendung von Gleichung (2.5) als zusätzliche Bedingung wird die Abweichung der Lösung x vom Mittelwert aller Unbekannten minimiert, bei Verwendung von Gleichung (2.6) die relative Abweichung der Lösung x von der gewählten a-priori Lösung p .

Weiterhin findet man in [Twomey 1965] Vorschläge, wie der Regularisierungsparameter α bestimmt werden kann: Nach der Festlegung eines Schrankenwertes für die Summe der Residuenquadrate ($N[x, \bar{l}]$), welcher kleinst möglich und trotzdem zuverlässig sein soll, wird für unterschiedliche Werte des Regularisierungsparameters der dazugehörige Wert für $N[x, \bar{l}]$ berechnet. Der erste Wert für α , für den $N[x, \bar{l}]$ etwas unterhalb des festgelegten Schrankenwertes liegt, kann dann meistens für die endgültige Lösung des Problems verwendet werden. Falls α genauer bestimmt werden soll, empfiehlt er, $N[x, \bar{l}]$ bzgl. α grafisch darzustellen. Es kann dann ein Wert für α ausgewählt werden, so dass $N[x, \bar{l}]$ dem Schrankenwert exakt oder annähernd entspricht. Diese Varianten zur Ermittlung des Regularisierungsparameters weisen wiederum den Nachteil auf, dass nicht die Größe der einzelnen Residuen, sondern die Summe der Residuenquadrate betrachtet wird.

Tikhonov und Arsenin 1977:

Werden in [Tikhonov 1963a] und [Tikhonov 1963b] lediglich die Daten l als fehlerbehaftet betrachtet, so findet man in [Tikhonov und Arsenin 1977], dass ebenso der Operator A nur näherungsweise bekannt sein kann. Für die Gleichung (2.1) ergibt sich dann:

$$\begin{aligned}\bar{A}x &= \bar{l}, \\ A + \mu &= \bar{A},\end{aligned}\tag{2.7}$$

wobei μ die Ungenauigkeit des Operators A repräsentiert⁷. Werden derartige Probleme mittels Regularisierung gelöst, ist bei der Wahl des Regularisierungsparameters α nicht mehr nur der Fehler δ zu berücksichtigen, sondern ebenso die Ungenauigkeit μ des Operators A .

Weiterhin findet man in [Tikhonov und Arsenin 1977] eine Unterteilung der zusätzlichen Informationen in einerseits qualitative und andererseits quantitative Informationen. Dabei versteht man unter quantitativen Informationen z.B. Informationen über bereits existierende Lösungen des Problems und unter qualitativen Informationen z.B. Informationen über Eigenschaften der Lösung, wie Glattheit, Monotonie und Konvexität.

Eine mögliche Wahl für $\Omega[x]$, um die sog. Normallösung bestimmen zu können, ergibt sich mit:

$$\Omega[x] = \|x_N - x\|,\tag{2.8}$$

⁷In der Literatur werden derartige Problemstellungen als „Total Least Squares“ (TLS), „errors in variables model“ oder auch als „orthogonal regression“ bezeichnet, siehe z.B. [Van Huffel und Zha 1993], [Golub und van Loan 1980], [Golub et al. 1999], [Fierro et al. 1997]. Wie in [Petrović 2003] gezeigt wurde, handelt es sich bei TLS lediglich um einen Spezialfall der Anwendung der Methode der kleinsten Quadrate und nicht, wie an einigen Stellen der Literatur behauptet wird, um ein völlig neuartiges Minimierungsverfahren. Somit ist das zu lösende Problem nach der Einführung von Ungenauigkeiten bzgl. des Operators A ein anderes, als wenn nur Fehler bzgl. der Daten l einbezogen werden, d.h. die zu minimierenden Größen sind unterschiedlich, die Minimierung erfolgt aber auf gleichem Wege. Somit entfällt eine zusätzliche Betrachtung der Anwendung der Regularisierung auf TLS, da diese somit analog erfolgen kann wie bei der Anwendung der Methode der kleinsten Quadrate.

wobei x_N die Normallösung darstellt, welche man erhalten würde, wenn δ und μ gleich null wären. Die mathematische Schreibweise $\|\cdot\|$ bezeichnet eine beliebige Norm⁸, nach der die Minimierung des glättenden Funktionalen erfolgen kann. Tikhonov betrachtet neben der allgemein gebräuchlichen L_2 -Norm- ebenso die L_1 -Norm-Minimierung, bei der nicht die Summe von Quadraten, sondern die Summe von Beträgen minimiert wird. Ob eine derartige Normallösung für praktische Probleme verfügbar ist, ist fragwürdig. Vielfach wurde bei der praktischen Anwendung des Regularisierungsverfahrens stattdessen eine existierende Lösung für das jeweilige Problem verwendet. Diese Wahl für $\Omega[x]$ (Gleichung (2.8)) ist sehr ähnlich zu der in [Twomey 1963] betrachteten Minimierung der Abweichungen zu einer „Testlösung“.

Da die Wahl des Regularisierungsparameters einen sehr entscheidenden Schritt bei der Bestimmung der Näherungslösung des inkorrekt gestellten Problems darstellt, wurden an vielen Stellen der Literatur dahingehend Untersuchungen angestellt. Eine praktische Methode, um α festzulegen, findet man in [Tikhonov und Arsenin 1977]: Der Fehler der Daten sei weiterhin δ . Man erzeuge dann einen endlichen Ausschnitt der monotonen Sequenz $\alpha_0, \alpha_1, \alpha_2, \dots, \alpha_n$. Zum Beispiel können diese α 's mittels einer geometrischen Reihe der Form $\alpha_k = \alpha_0 q^k$ mit $k = 0, 1, 2, \dots, n$ und $q > 0$ konstruiert werden. Für jeden Wert α_k wird dann die Lösung x^α bestimmt, welche das glättende Funktional $M^\alpha[x^\alpha, \bar{l}]$ minimiert, und es werden sämtliche Residuen berechnet. Man wählt dann dasjenige α_k aus, für welches sämtliche Residuen innerhalb der Messgenauigkeit δ liegen.

Anhand dieser Vorgehensweise erkennt man sehr gut, dass man bei einer anderen Wahl von q , n oder α_0 für die geometrische Reihe zur Erzeugung der α 's ein anderes α_k als geeignet auswählen und somit unterschiedliche Lösungen des Problems erhalten würde, die wenig oder auch stark voneinander verschieden sein können. Daher wäre es besser, wenn man α so bestimmen könnte, dass sämtliche Residuen innerhalb der Messgenauigkeit liegen und minimal sind.

Eine Möglichkeit den Regularisierungsparameter grafisch zu bestimmen und die Größe der einzelnen Residuen zu berücksichtigen, könnte z.B. folgender Art sein: Für eine Sequenz von α 's berechnet man die Lösung des glättenden Funktionalen sowie alle Residuen v_i der Messungen. Ist nun für alle Messungen die Standardabweichung m_i bekannt, so lässt sich $k_i = \frac{v_i^2}{m_i^2}$ berechnen. Stellt man den maximalen Wert von k_i ($\max(k_i)$) bzgl. der α 's grafisch dar, so erhält man eine Kurve, aus der erkennbar ist, bei welchen Werten von α sämtliche Residuen innerhalb ihrer Messgenauigkeit liegen. Dies ist der Fall für alle α 's, für die das dazugehörige $\max(k_i) \leq 1$ ist.

In [Louis 1989] ist sehr anschaulich beschrieben, was unterschiedliche Regularisierungsparameter bewirken, wenn Glattheit der Lösung gefordert wird: „Große Werte von γ [Anm.: entspricht α] geben dem Strafterm [Anm.: $\Omega[x]$] mehr Gewicht, die Lösungen werden immer 'glatter'. Bei kleinen Werten von γ hat der Defekt $\|Af - g\|^2$ [Anm.: $N[x, \bar{l}]$] den größeren Einfluß auf die Lösung f_γ . Sind die Daten nur wenig gestört, dann wird man immer ein kleines γ wählen. Hat man dagegen wenig 'Vertrauen' in die Daten, so erhöht man den Einfluß der Zusatzinformation durch die Wahl eines größeren γ .“

2.2.2 Regularisierendes Verfahren

Zusätzlich zu dem im vorherigen Abschnitt beschriebenen Regularisierungsverfahren trifft man in der Literatur auch auf das sog. regularisierende Verfahren, siehe z.B. [Schwetlick 1979]. Wie

⁸Siehe dazu Kapitel 3, Formel (3.25).

die Bezeichnung bereits andeutet, besteht die Aufgabe dieses Verfahrens darin, ein singuläres Problem in ein reguläres umzuwandeln. Analog wie bei dem Regularisierungsverfahren nach Tikhonov wird auch hier das ursprünglich zu lösende Problem (meist ein Ausgleichungsproblem nach kleinsten Quadraten) durch die Einbeziehung zusätzlicher Bedingungen regularisiert und somit stabilisiert. Der Unterschied besteht aber darin, dass in diesem Fall die zusätzlichen Bedingungen bei allen Anwendungen des regularisierenden Verfahrens gleichartig sind und sich leicht aber grundsätzlich von den zusätzlichen Bedingungen bzw. Informationen des vorherigen Abschnitts unterscheiden. Anhand eines typischen allgemeinen Ausgleichungsproblems nach kleinsten Quadraten soll nun dieses Verfahren veranschaulicht werden.

Bekannt seien die nichtlinearen Beobachtungsgleichungen:

$$l_i + v_i = f(x) \quad i = 1, 2, \dots, n, \quad (2.9)$$

wobei l_i die Messungen darstellen, v_i die zufälligen Fehleranteile der Messungen, f den funktionalen Zusammenhang zwischen den Messungen l_i und den zu bestimmenden Unbekannten x und n die Anzahl der Messungen. Die linearisierten Beobachtungsgleichungen in Matrizenschreibweise lauten dann:

$$\begin{aligned} v &= A\Delta x - \tilde{l}, \\ A &= \left(\frac{\partial f}{\partial x} \right)_{x_r}, \\ \tilde{l} &= l - f(x_r), \\ \Delta x &= x_{r+1} - x_r, \end{aligned} \quad (2.10)$$

$$(2.11)$$

wobei x_r eine geeignet gewählte Näherungslösung für den r -ten Iterationsschritt bei der Minimierung von $\sum_i v_i^2$ darstellt⁹. Liegen z.B. mehr Messungen als Unbekannte vor, so besitzt die Gleichung (2.10) unendlich viele Lösungen. Es soll nun die Lösung bestimmt werden, für die die Summe der Residuenquadrate ($\sum_{i=1}^n v_i^2$) minimal wird. Man spricht dann von einer L_2 -Norm-Minimierung. Grundsätzlich kann bei dem regularisierenden Verfahren jede beliebige Norm minimiert oder maximiert werden, entscheidend ist die Zielsetzung des betrachteten Problems. Zur Minimierung nach anderen Normen siehe z.B. [Stichler 1984], [Casparly 1988], [Somogyi und Závoti 1990], [Kilian 2002], [Marx 2003]. Die Lösung x_{r+1} der $(r+1)$ -ten Iteration bei der Minimierung der Summe der Residuenquadrate kann mittels folgender Formel berechnet werden, siehe z.B. [Helmert 1872]:

$$\begin{aligned} x_{r+1} &= x_r + N^{-1}A^T\tilde{l}, \\ N &= A^T A. \end{aligned} \quad (2.12)$$

Ist die Matrix N aufgrund einer gewissen Instabilität des Problems, der Linearisierung oder auch wegen Rundungsfehlern bei der Berechnung singulär, so ist die notwendige Invertierung von N , um x_{r+1} berechnen zu können, nicht möglich.

Was bedeutet dies rein mathematisch? Ist eine Matrix G singulär, so ist mindestens ein Eigenwert λ_i dieser Matrix gleich Null. Addiert man zu allen Eigenwerten der Matrix einen konstanten positiven Wert k , so erhält man die Eigenwerte $\bar{\lambda}_i$ der Matrix $\bar{G}$, welche wie folgt definiert ist [Scales 1985]:

$$\bar{G} = G + kI, \quad (2.13)$$

⁹An dieser Stelle erfolgt keine explizite Berücksichtigung von Gewichten für die einzelnen Messungen. Eine Berücksichtigung unterschiedlicher Gewichte p_i ist allerdings problemlos möglich, so dass dann $\sum_i p_i v_i^2$ minimiert wird.

wobei I die Einheitsmatrix darstellt. Die Matrizen G und $\bar{G}$ besitzen die gleichen normierten Eigenvektoren, aber $\bar{G}$ ist eine reguläre und somit invertierbare Matrix, wenn k ausreichend groß gewählt wird. Zusätzlich ist die Konditionszahl¹⁰ K der Matrix $\bar{G}$ größer als die der Matrix G , wenn man als Konditionszahl einer beliebigen Matrix M den Wert

$$K = \frac{|\min(EW(M))|}{|\max(EW(M))|} \quad (2.14)$$

bezeichnet, wobei mit $EW(M)$ die Eigenwerte der Matrix M gemeint sind.

Je größer die Konditionszahl der Matrix N in Gleichung (2.12), desto unempfindlicher ist die Berechnung der Lösung x_{r+1} gegenüber Rundungsfehlern, [Scales 1985]. Somit ergeben sich unmittelbar zwei Vorteile für eine singuläre Matrix, wenn zu dieser eine Diagonalmatrix addiert wird - die resultierende Matrix ist regulär und besser konditioniert. Anstelle der Gleichung (2.12) ergibt sich dann:

$$x_{r+1} = x_r + (N + kI)^{-1} A^T \tilde{l}. \quad (2.15)$$

Was wird aber bei Verwendung der Gleichung (2.15) statt der Gleichung (2.12) minimiert? Verwendet man die Gleichung (2.12), so wird lediglich die Summe der Residuenquadrate $\sum_{i=1}^n v_i^2$ minimiert. Dagegen wird bei Verwendung der Gleichung (2.15) die Summe der Residuenquadrate und die gewichtete Summe der quadrierten Abweichungen zwischen der r -ten und der $(r+1)$ -ten Iterationslösung minimiert, also:

$$\varphi(\Delta x) = \sum_{i=1}^n v_i^2 + k \sum_{j=1}^m (x_{r+1}^j - x_r^j)^2, \quad (2.16)$$

wobei $\varphi(\Delta x)$ die zu minimierende Funktion darstellt und m die Anzahl der zu bestimmenden Unbekannten.

BEWEIS: Ausgehend von den Gleichungen (2.10) und (2.11) erhält man für die zu minimierende Funktion in Matrizenschreibweise:

$$\varphi(\Delta x) = v^T v + k \Delta x^T \Delta x. \quad (2.17)$$

Da von $\varphi(\Delta x)$ das Minimum gesucht ist, erfolgt zunächst die Ableitung der Funktion $\varphi(\Delta x)$ nach den Unbekannten Δx :

$$\frac{\partial \varphi}{\partial \Delta x} = 2v^T A + 2k \Delta x^T I. \quad (2.18)$$

Setzt man diese Ableitung zu Null und führt einige Umformungen durch, so erhält man:

$$\begin{aligned} 0 &= v^T A + k \Delta x^T I, \\ 0 &= A^T v + k I \Delta x, \\ 0 &= A^T (A \Delta x - \tilde{l}) + k I \Delta x, \\ 0 &= A^T A \Delta x - A^T \tilde{l} + k I \Delta x, \end{aligned}$$

¹⁰Die Definition der Konditionszahl einer Matrix wurde aus [Wilkinson 1969] entnommen. Es ist zu beachten, dass in der Literatur sehr viele unterschiedliche Definitionen für die Kondition einer Matrix zu finden sind. Allgemein wird damit aber immer die Empfindlichkeit einer Matrix gegenüber mathematischer Umformungen beschrieben (siehe z.B. [Parker 1994]).

$$\begin{aligned}
A^T A \Delta x + k I \Delta x &= A^T \tilde{l}, \\
(N + k I) \Delta x &= A^T \tilde{l}, \\
x_{r+1} &= x_r + (N + k I)^{-1} A^T \tilde{l}.
\end{aligned} \tag{2.19}$$

Bei der Extremwertbestimmung einer differenzierbaren Funktion durch Bestimmung der stationären Punkte muss geprüft werden, ob die Nullstellen der 1. Ableitung der Funktion einem Minimum, Maximum oder Sattelpunkt entsprechen, d.h. in diesem Fall ist zu prüfen, ob Δx tatsächlich einem Minimum entspricht:

$$\begin{aligned}
\frac{\partial \varphi}{\partial \Delta x} &= 2(A^T A \Delta x + k I \Delta x - A^T \tilde{l}), \\
\frac{\partial^2 \varphi}{\partial \Delta x^2} &= 2(A^T A + k I).
\end{aligned}$$

Da k stets positiv und $N = A^T A$ eine symmetrische Matrix ist, ist $(A^T A + k I)$ immer positiv definit, so dass Gleichung (2.19) die Iterationsformel zur Minimierung von $\varphi(\Delta x)$ darstellt. Damit ist der Beweis abgeschlossen - die Gleichungen (2.19) und (2.15) sind identisch.

Diese Art der Regularisierung dient somit zur Anpassung der Iterationsschrittweite bei der Minimierung einer beliebigen nichtlinearen Funktion, welche linearisiert wurde. Diese Idee geht auf K. Levenberg zurück und wurde von ihm 1944 erstmals in [Levenberg 1944] erörtert. Sein Ziel bestand darin, die Größe des Zuschlags Δx zur jeweiligen Näherungslösung x_r zu begrenzen („gedämpftes Ausgleichungsverfahren nach kleinsten Quadraten“), da die linearisierten Beobachtungsgleichungen (2.10) lediglich in der Umgebung der Näherungslösung x_r die nichtlinearen Beobachtungsgleichungen gut approximieren. Dabei lässt er zu, dass jede einzelne Komponente des Zuschlagvektors Δx unterschiedlich stark gedämpft werden kann, so dass statt $k I$ in Gleichung (2.15) eine Diagonalmatrix mit beliebigen unterschiedlichen positiven Diagonalelementen zulässig ist. Mit diesem Verfahren kann somit die Konvergenz bei der iterativen Minimierung der Summe der Residuenquadrate gewährleistet werden.

In [Marquardt 1963] wird der von Levenberg entwickelte Algorithmus dahingehend erweitert, dass man zu einem Verfahren gelangt, welches einen „optimalen“ Kompromiss zwischen dem Gauß-Newton-Verfahren und dem Gradientenverfahren¹¹ bei der Minimierung einer Funktion darstellt¹². Dadurch wird es möglich, die Vorteile des Gradientenverfahrens und die des Gauß-Newton-Verfahrens bestmöglich auszunutzen¹³. Gradientenverfahren konvergieren schnell zur Lösung, auch bei recht schlechten Näherungswerten, weisen allerdings in der Nähe des Minimums nur eine schlechte Konvergenz auf. Das Gauß-Newton-Verfahren ist dagegen sehr empfindlich auf schlechte Näherungswerte, konvergiert aber schnell in der Nähe des Minimums. Ausgehend von Formel (2.15), wobei k für jeden Iterationsschritt bei der Minimumsuche neu gewählt wird, so dass zu Beginn der Iteration möglichst das Gradientenverfahren zum Einsatz kommt und am Ende das Gauß-Newton-Verfahren, erhält man für $k = 0$ und $k \rightarrow \infty$:

$$k = 0 : \quad x_{r+1} = x_r + N^{-1} A^T \tilde{l}, \tag{2.20}$$

¹¹Zu den unterschiedlichen Iterationsverfahren siehe z.B. [Schwetlick 1979] und [Scales 1985].

¹²Dieses Verfahren ist unter dem Namen „Levenberg-Marquardt-Verfahren“ sehr bekannt.

¹³Als zusätzlichen Vorteil dieses Verfahrens betrachtet Marquardt, dass durch die Addition von k auf alle Diagonalelemente der Matrix N die Kondition der zu invertierenden Matrix verbessert wird.

$$k \rightarrow \infty : \quad x_{r+1} \approx x_r + \frac{1}{k} A^T \tilde{l} \quad \text{da } kI \gg N. \quad (2.21)$$

Für $k = 0$ entspricht die Formel (2.20) der Iterationsformel für das Gauß-Newton-Verfahren. Für $k \rightarrow \infty$ stellt die Formel (2.21) näherungsweise die Iterationsformel für das Gradientenverfahren dar. Variiert man nun k zwischen 0 und ∞ oder ∞ und 0 für jeden Iterationsschritt, so kann man bei geeigneter Wahl von k die Konvergenz zum Minimum wesentlich verbessern (zu speziellen Konvergenzeigenschaften siehe z.B. [Schwetlick 1973]). Leider gibt es für die Wahl von k lediglich die Möglichkeit, diesen Wert mittels einer „trial and error“ Prozedur festzulegen¹⁴.

Es ist besonders wichtig zu beachten, dass durch die Verwendung dieses regularisierenden Verfahrens nicht ein Ersatzproblem des ursprünglichen Minimierungsproblems gelöst wird, sondern lediglich die Konvergenz bei der Funktionsminimierung verbessert wird. Die Wahl des Regularisierungsparameters k ist hier nicht so kritisch zu betrachten wie die Wahl des Regularisierungsparameters α im vorherigen Abschnitt. Wird α nicht sachgerecht gewählt, so erhält man eine andere Lösung, wohingegen bei zu kleiner oder zu großer Wahl von k nur die Anzahl der Iterationsschritte bei der Minimumsbestimmung beeinflusst wird, die Endlösung im Falle der Konvergenz aber immer die gleiche ist.

2.2.3 Stochastische Regularisierung

Als dritte Gruppe der Regularisierungsverfahren soll die stochastische Regularisierung vorgestellt werden. Aufgrund der sehr großen Anzahl von Publikationen zum Thema Regularisierung und der vielfältigen und uneinheitlichen Bezeichnungsweise für identische Verfahren war es nicht möglich festzustellen, wer die stochastische Regularisierung zuerst ausgearbeitet und veröffentlicht hat. Genannt seien aber [Friedrich, Hofmann und Tautenhahn 1979], [Tarantola und Valette 1982], [Louis 1989] und [Rodgers 2000].

Der Unterschied zwischen der Regularisierung nach Tikhonov und dem regularisierenden Verfahren einerseits und der stochastischen Regularisierung andererseits besteht darin, dass die zusätzlichen Informationen nicht als zusätzliche Bedingungen zu dem ursprünglichen Minimierungsproblem betrachtet werden, sondern als zusätzliche Beobachtungen (im Weiteren: Pseudobeobachtungen) einbezogen werden, für welche in der Regel stochastische Informationen vorliegen. Man erhält somit beispielsweise folgende linearisierte Beobachtungsgleichungen

$$\begin{aligned} \tilde{l} + v &= A\Delta x, \\ \tilde{l}_p + v_p &= \Delta x, \\ \Delta x &= x_{r+1} - x_r, \end{aligned} \quad (2.22)$$

wobei $\tilde{l}_p$ die reduzierten Pseudobeobachtungen darstellt und v_p die dazugehörigen Residuen. Üblicherweise wird allen reduzierten Pseudobeobachtungen der Wert Null¹⁵ zugeordnet, so dass $v_p = \Delta x$ gilt, d.h. je größer die Residuen für die Pseudobeobachtungen sind, desto größer ist auch die Iterationsschrittweite Δx (wenn Δx nach Gleichung (2.22) definiert wurde).

Für die Beobachtungen $\tilde{l}$ sind meist die Standardabweichungen und möglicherweise auch die Korrelationen zwischen den einzelnen Beobachtungen bekannt. Um diese Informationen mit in das Minimierungsproblem einbeziehen zu können, wird die Gewichtsmatrix P_b der Beobachtungen

¹⁴In [Schwetlick 1979] sind dazu einige Strategien zu finden, nach denen die Wahl von k erfolgen kann.

¹⁵Durch eine geeignete Umformung kann dies immer erreicht werden.

$\tilde{l}$ wie folgt definiert:

$$P_b = \begin{bmatrix} P_b^{11} & P_b^{12} & \dots & P_b^{1n} \\ \vdots & \vdots & \ddots & \vdots \\ P_b^{n1} & P_b^{n2} & \dots & P_b^{nn} \end{bmatrix},$$

wobei n die Anzahl der Beobachtungen $\tilde{l}$ darstellt. Diese Gewichtsmatrix lässt sich durch Invertierung der Kofaktorenmatrix der Beobachtungen berechnen, welche unter Verwendung der Standardabweichungen der Beobachtungen, des mittleren Gewichtseinheitsfehlers sowie der Kovarianzen festgelegt werden kann¹⁶. Ist P_b eine Diagonalmatrix, so sind sämtliche Beobachtungen voneinander unabhängig, was in der Praxis meist angenommen wird, da entsprechende Informationen bzgl. der Kovarianzen oft fehlen. Die Diagonalelemente der Matrix P_b sind immer positiv und ungleich Null, denn wenn ein Diagonalelement Null wäre, würde die entsprechende Beobachtung überhaupt nicht berücksichtigt werden.

Analog zu dem regularisierenden Verfahren kann in diesem Beispiel die Gewichtsmatrix P_p für die Pseudobeobachtungen als Einheitsmatrix oder beliebige Diagonalmatrix (wobei alle Diagonalelemente positiv sind) gewählt werden, so dass die einzelnen Zuschläge Δx_i unterschiedlich stark gedämpft werden können.

Die zu minimierende Funktion bei der stochastischen Regularisierung unter Einbeziehung der Gewichtsmatrizen für die Beobachtungen und Pseudobeobachtungen lautet dann wie folgt:

$$\varphi(\Delta x) = v^T P_b v + v_p^T P_p v_p. \quad (2.23)$$

Durch null setzen der 1. Ableitung dieser Funktion erhält man:

$$\begin{aligned} \frac{\partial \varphi}{\partial \Delta x} = 0 &= 2v^T P_b A + 2v_p^T P_p, \\ 0 &= A^T P_b v + P_p v_p, \\ 0 &= A^T P_b (A \Delta x - \tilde{l}) + P_p (\Delta x - \tilde{l}_p), \\ 0 &= A^T P_b A \Delta x - A^T P_b \tilde{l} + P_p \Delta x - P_p \tilde{l}_p, \\ &\Downarrow \\ \Delta x &= (A^T P_b A + P_p)^{-1} (A^T P_b \tilde{l} + P_p \tilde{l}_p). \end{aligned} \quad (2.24)$$

Vergleicht man Gleichung (2.24) mit Gleichung (2.15) ($k = 1$) und beachtet, dass in diesem Fall $\tilde{l}_p = 0$ ist und beide Gewichtsmatrizen als Einheitsmatrix gewählt werden können, so ist ersichtlich, dass bei der stochastischen Regularisierung der Einfluss der Pseudobeobachtungen auf die Lösung durch das Verhältnis der Gewichtsmatrizen P_b und P_p zueinander geregelt wird. Dahingegen wurde der Einfluss der zusätzlichen Informationen (entspricht den Pseudobeobachtungen) bei dem regularisierenden Verfahren durch den Parameter k geregelt. Die Formel (2.24) zur Berechnung der Unbekannten mittels stochastischer Regularisierung ist z.B. in [Friedrich, Hofmann und Tautenhahn 1979], [Tarantola und Valette 1982] und [Rodgers 2000] zu finden, wobei $\tilde{l}_p$ dort eine andere Bedeutung zukommt und damit das Verhältnis von P_b zu P_p nicht mehr äquivalent zu k , sondern äquivalent zu α aus Abschnitt 2.2.1 ist.

In [Tarantola und Valette 1982] z.B. stellt $\tilde{l}_p$ die Differenz der Näherungslösung x_r zu einer bekannten Lösung (a-priori Lösung) dar. Diese a-priori Lösung kann ihrer Meinung nach aus

¹⁶Im Weiteren soll stets von der angemessenen Wahl von Gewichtsmatrizen ausgegangen werden, obwohl diese aus den Kofaktorenmatrizen berechnet werden, welche hier als nicht singulär angenommen werden.

bestehenden Modellen entnommen werden, das Ergebnis einer vorangegangenen Auswertung sein oder einen einfachen mittleren zu erwartenden Wert der Lösung darstellen. Das eine derartige Einführung von zusätzlichen Informationen bzw. a-priori Lösungen leicht zu einer unangemessenen Auswertung der eigentlichen Beobachtungen führen kann, wurde bereits diskutiert. Weiterhin wird in [Tarantola und Valette 1982] fälschlicherweise behauptet, dass lediglich eine zusätzliche Minimierung der Näherungslösung zu einer a-priori Lösung zweckmäßig ist und eine Iterationsschrittweiten-Dämpfung, wie sie im vorhergehenden Abschnitt beschrieben wurde, fehlerhaft ist. Das dies nicht der Fall ist, wurde mit den Erläuterungen im Abschnitt 2.2.2 ausreichend dargelegt.

So wie bei der Regularisierung nach Tikhonov die Wahl des Regularisierungsparameters besonders kritisch betrachtet wurde, muss bei der stochastischen Regularisierung die Wahl der Gewichtsmatrizen sehr gewissenhaft erfolgen. In [Tarantola und Valette 1982] findet man diesbezüglich folgendes (aus dem Englischen übersetzt): *Bei praktischen Problemen kann es passieren, dass obwohl das Problem mathematisch regulär ist, dieses aufgrund des benutzten Algorithmus numerisch singulär oder instabil wird. Dies bedeutet immer, dass die Vorinformationen [Anm.: Pseudobeobachtungen] die Lösung nicht ausreichend einschränken. Eine Verringerung der Varianzen in $C_{p_0p_0}$ [Anm.: entspricht einer Vergrößerung der Diagonalelemente von P_p] wird die Lösung stabilisieren, ... Doch es muss betont werden, dass eine Verringerung dieser Varianzen zur Stabilisierung der Lösung gleichbedeutend ist mit der Einführung von Vorinformationen, die tatsächlich nicht verfügbar sind, da die Varianzen in $C_{p_0p_0}$ die Vorkenntnisse über die Parameter beschreiben.*¹⁷ Mit diesen Aussagen wird deutlich, dass den Autoren die unterschiedlichen Anwendungsmöglichkeiten und Eigenschaften der Regularisierungsverfahren nicht bewusst sind. Die Einführung von zusätzlichen Informationen dient lediglich dazu, ein inkorrekt gestelltes Problem lösbar zu machen, so dass aus der Vielzahl von möglichen Lösungen eines inkorrekt gestellten Problems eine plausible Lösung ausgewählt werden kann. Um ein Problem zu stabilisieren, ist das regularisierende Verfahren zu verwenden, wodurch nicht das zu minimierende Problem verändert wird, sondern nur die Iterationsschrittweite geeignet gewählt wird, so dass eine Konvergenz zum gesuchten Minimum gewährleistet werden kann.

In [Rodgers 2000] findet man zum Thema Gewichtsmatrizen (aus dem Englischen übersetzt): *S_a und S_e [Anm.: entsprechen P_b und P_p] werden üblicherweise als 'Tuningparameter' verwendet und derart durch willkürliche Faktoren angepasst, dass man ein ästhetisch ansprechendes Ergebnis erhält. Es ist selbstverständlich, dass jede derartige Verwendung begründbar sein sollte und man soll nicht vergessen, dass z.B. eine Verdopplung von S_a und S_e nicht die Lösung verändert, sondern nur die Fehlerschätzung der Lösung.*¹⁸ Dies zeigt sehr anschaulich, wie wichtig das Verhältnis zwischen den Gewichtsmatrizen P_b und P_p ist, wobei die absolute Größe der Fehlerschätzungen doch recht nebensächlich ist. Leider scheint es, dass das Wort 'Tuningparameter' falsch interpretiert wird, d.h. es wird oft zu wenig auf das entscheidende Verhältnis zwischen den Gewichtsmatrizen geachtet, als viel mehr darauf, wie gut oder schlecht die neue Lösung zu bereits

¹⁷Im Original: „It may happen in actual problems that although mathematically regular, the algorithm may become numerically singular or unstable. This will always mean that the a priori information does not constrain the solution enough. A reduction of the variances in $C_{p_0p_0}$ will stabilize the solution, ... But we must emphasize that since variances in $C_{p_0p_0}$ describe our a priori knowledge on parameters, reducing these variances in order to stabilize the solution means that we are introducing a priori information that, in fact, we do not possess.“

¹⁸Im Original: „Both S_a and S_e are commonly used as 'tuning parameters' and adjusted by arbitrary factors to obtain an aesthetically pleasing retrieval. It goes without saying that any such use should be justifiable, and please remember that e.g. doubling both S_a and S_e will not change the retrieval, but only the error estimates.“

vorhandenen Lösungen passt (in dem Fall, wo die Pseudobeobachtungen die Differenz zwischen der Näherungslösung und einer a-priori Lösung darstellen). Dies resultiert aber auch daraus, dass die Wahl der Gewichtsmatrizen in vielen praktischen Anwendungen nur unzuverlässig möglich ist, da nicht ausreichende Kenntnisse dafür verfügbar sind. Dazu [Rodgers 2000] (aus dem Englischen übersetzt): *Weniger realistisch, aber praktisch, ist die Beschreibung der Vorinformationen über x durch eine Gauß'sche Wahrscheinlichkeitsdichtefunktion.*¹⁹ Dass eine unrealistisch gewählte Gewichtsmatrix für die Vorinformationen nicht zwangsläufig zu unrealistischen Ergebnissen führen muss, ist daraus ersichtlich, da die zusätzliche Festlegung der Gewichtsmatrix für die Beobachtungen letztendlich erst das Verhältnis zwischen den beiden Gewichtsmatrizen festsetzt. Da aber für die Gewichtsmatrix der Beobachtungen oft die relativ gut bekannten Standardabweichungen der Beobachtungen verwendet werden, wird durch eine unrealistisch gewählte Gewichtsmatrix der Vorinformationen die Lösung dann doch meist ebenso unrealistisch. Somit sollte man nicht das 'Tuning' der Fehlerschätzung für die Lösung in den Vordergrund stellen oder die Abweichungen zu einer a-priori Lösung, sondern beide Gewichtsmatrizen anhand aller verfügbaren Kenntnisse so sachgerecht wie möglich festlegen.

Liegt ein lineares Ausgleichungsproblem vor, so lauten die Beobachtungsgleichungen für die Beobachtungen l und die Pseudobeobachtungen l_p :

$$\begin{aligned} l + v &= Ax, \\ l_p + v_p &= x. \end{aligned}$$

Durch Substitution mit $x = x_r + \Delta x$ erhält man:

$$\begin{aligned} l + v &= Ax_r + A\Delta x, \\ l_p + v_p &= x_r + \Delta x, \\ &\Downarrow \\ \tilde{l} + v &= A\Delta x \rightarrow \tilde{l} = l - Ax_r, \\ \tilde{l}_p + v_p &= \Delta x \rightarrow \tilde{l}_p = l_p - x_r. \end{aligned}$$

Häufig entsprechen die Pseudobeobachtungen l_p gerade der Näherungslösung x_r , so dass alle reduzierten Pseudobeobachtungen den Wert Null erhalten. In der Literatur werden diese „Null-Pseudobeobachtungen“ oft als „stochastische Vorinformationen“ bezeichnet [Schwintzer 1990], da für diese stochastische Informationen vorliegen (entspricht den Genauigkeitsangaben für die Näherungslösung x_r). Die Gewichtsmatrix P_p muss in diesem Fall nicht zwangsläufig eine Diagonalmatrix sein, sondern es können ebenso Korrelationen zwischen den Pseudobeobachtungen berücksichtigt werden. Häufig täuscht die Translation der Pseudobeobachtungen in „Null-Pseudobeobachtungen“ darüber hinweg, dass hierbei eine Verbesserung vorhandener Lösungen/Modelle x_r aufgrund der Beobachtungen nur sehr eingeschränkt möglich ist, da $v^T P_b v$ und $v_p^T P_p v_p$ minimiert werden. Lediglich bei einer extremen Untergewichtung der Pseudobeobachtungen können die eigentlichen Beobachtungen eine signifikante Veränderung der bestehenden Lösung x_r bewirken. Somit muss unabhängig davon, ob die reduzierten Pseudobeobachtungen den Wert Null haben oder nicht, das Verhältnis zwischen den Gewichtsmatrizen P_b und P_p angemessen festgelegt werden, wenn der Informationsgehalt der eigentlichen Beobachtungen bzgl.

¹⁹Im Original: „Less realistic, but convenient, is to describe prior knowledge of x by a Gaussian probability density function.“

der gesuchten Lösung genutzt werden soll²⁰.

2.3 Verfahren zur Bestimmung des Regularisierungsparameters

Wie im Abschnitt 2.2 betont wurde, ist die Festlegung des Regularisierungsparameters bzw. Gewichtsverhältnisses bei der Anwendung von Regularisierungsverfahren für eine erfolgreiche Auswertung neuer Beobachtungen von extremer Wichtigkeit. So sind bereits sehr viele Verfahren zur Bestimmung des „optimalen“ Regularisierungsparameters erarbeitet worden. Einen guten Überblick darüber gibt z.B. [Kusche 2002]. Auf einige verbreitete Verfahren soll an dieser Stelle kurz eingegangen werden.

Varianzkomponenten-Schätzung:

Bei der stochastischen Regularisierung benötigt man für die Beobachtungen und Pseudobeobachtungen jeweils eine Gewichtsmatrix (P_b und P_p), in die der mittlere Fehler der Gewichtseinheit m_0 eingeht. Dieser mittlere Fehler stellt die durchschnittliche Genauigkeit der Beobachtungen bzw. Pseudobeobachtungen dar, so dass m_0 für die Matrizen P_b und P_p in der Regel unterschiedlich gewählt werden wird. Die Berechnung der Lösung x kann dann nach Formel (2.24) erfolgen. Mittels dieser ermittelten Lösung lassen sich dann neue mittlere Fehler der Gewichtseinheit (= Varianzkomponenten) aus den resultierenden Residuen berechnen (siehe [Schwintzer 1990], [Koch und Kusche 2002]). Unter Verwendung der neuen Varianzkomponenten werden die Matrizen P_b und P_p neu aufgestellt und wiederum die Lösung x berechnet. Es folgt ein iterativer Prozess aus Neuberechnung der Varianzkomponenten und Lösungsbestimmung, welcher abgebrochen wird, wenn sich zwei aufeinander folgend bestimmten Lösungen nicht signifikant voneinander unterscheiden. Es liegt dann ein „ausgeglichenes“ Verhältnis zwischen der Einbeziehung der Beobachtungen und den Pseudobeobachtungen bei der Lösungsbestimmung vor. Das Verhältnis zwischen den so ermittelten Varianzkomponenten stellt den Regularisierungsparameter dar.

Generalisierte Kreuzvalidation (GKV):

Das Verfahren der GKV ermittelt denjenigen Regularisierungsparameter, so dass die damit berechenbare Lösung eine (weggelassene) Einzelbeobachtung im quadratischen Mittel am besten vorhersagen kann [Kusche 2002]. Bisherige Untersuchungen haben allerdings gezeigt, dass das Verfahren der GKV sich wenig robust verhält, wenn die Daten nicht modellierbares farbiges Rauschen enthalten bzw. wenn Korrelationen zwischen den Beobachtungen vernachlässigt werden. Für mehr Details siehe [Golub et al. 1979] und [Krakauer et al. 2004].

L-Kurven-Verfahren:

Mittels des L-Kurven-Verfahrens kann ein geeigneter Regularisierungsparameter grafisch ermittelt werden, [Hansen und O’Leary 1993] und [Kusche 2002]. Dabei wird die gewichtete Summe der Residuenquadrate der Beobachtungen bzgl. der gewichteten Summe der Residuenquadrate der Vorinformationen für unterschiedliche Werte des Regularisierungsparameters doppeltlogarithmisch aufgetragen. Die entstehende Kurve nimmt dann meist eine charakteristische L-Form an. Der Regularisierungsparameter, der dem „Knickpunkt“ der Kurve entspricht, wird als „optimaler“ Regularisierungsparameter betrachtet.

²⁰Es lässt sich zeigen, dass das Ergebnis von der durchgeführten Substitution unabhängig ist und allein von dem Gewichtsverhältnis abhängt.

Diskrepanzkriterium nach Morozov:

Bei diesem Verfahren wird der „optimale“ Regularisierungsparameter so gewählt, dass die gewichtete Summe der Residuenquadrate der Beobachtungen kleiner oder gleich einem Schrankenwert δ ist, [Kusche 2002]. Dieser Schrankenwert kann anhand der Genauigkeit der auszuwertenden Beobachtungen abgeschätzt werden. Meist wird der Regularisierungsparameter numerisch oder grafisch so festgesetzt, dass die gewichtete Summe der Residuenquadrate der Beobachtungen gerade δ entspricht. Dieses Verfahren ist in der geophysikalischen Literatur unter dem Namen „fitting within the tolerance“ bekannt, [Parker 1994].

2.4 Diskussion

Bei Verwendung der im Abschnitt 2.2 vorgestellten Regularisierungsverfahren zur Lösungsbestimmung von inkorrekt gestellten Problemen ist die Bestimmung des Regularisierungsparameters bzw. des Gewichtsverhältnisses zwischen den Beobachtungen und Pseudobeobachtungen von entscheidender Bedeutung für die endgültige Lösung (das regularisierende Verfahren soll hierbei ausgeschlossen werden). In der Literatur zur Regularisierung findet man zwar sehr viele Verfahren, mit denen der Regularisierungsparameter bzw. das Gewichtsverhältnis „optimal“ bestimmbar sein soll, doch selbst bei den ausgereiftesten Verfahren bleibt immer eine gewisse Willkür hinsichtlich der damit erhaltenen endgültigen Lösung. Einerseits können durch zu großen Einfluss der zusätzlichen Informationen oder Bedingungen die Beobachtungen zu wenig Einfluss auf die Lösung haben und somit einzelne oder mehrere Residuen unerwünscht groß sein. Andererseits kann der Einfluss der zusätzlichen Informationen oder Bedingungen zu gering gewesen sein, so dass die Lösung ein unphysikalisches Verhalten aufweist.

Die im Abschnitt 2.3 erläuterten Verfahren zur Bestimmung des „optimalen“ Regularisierungsparameters zeigen sehr deutlich, dass bei jedem Verfahren mit „optimal“ eine andere Eigenschaft der Lösung bzw. der Residuen verknüpft ist. Bei keinem Verfahren wird die Größe der einzelnen Residuen der Beobachtungen nach Ermittlung der „optimalen“ Lösung betrachtet. Lediglich die gewichtete Summe der Residuenquadrate der Beobachtungen dient gegebenenfalls als Gütekriterium.

Aufgrund dieser Tatsache soll innerhalb des folgenden Kapitels ein Ausgleichungsverfahren nach kleinsten Quadraten (schließt die Regularisierungsverfahren mit ein) ausgearbeitet werden, bei dem nicht nur die Summe der Residuenquadrate minimiert wird, sondern zusätzlich auch für die Residuen der Beobachtungen angemessene Grenzwerte (Toleranzen) berücksichtigt werden können. Dadurch sollte es möglich werden, diejenige Lösung eines inkorrekt gestellten Problems zu bestimmen, die bestmöglich den Beobachtungen und ihren Genauigkeiten entspricht und gleichzeitig bestmöglich die zusätzlichen Informationen oder Bedingungen in den Lösungsprozess einbezieht. Eine derartige „bestmögliche“ Lösung ist in vielen Anwendungsgebieten von großem Interesse, da diese Lösung möglichst viele neue Informationen aus den neu ausgewerteten Daten enthält. Modellverbesserungen bzw. Verbesserungen der Näherungslösung sind somit möglich und nachweisbar. Selbstverständlich kann es aber auch Anwendungsgebiete geben, bei denen die „bestmögliche“ Lösung andere Eigenschaften aufweisen sollte.

3. Möglichkeiten zur Einführung von Toleranzenbedingungen für Residuen

Zunächst soll innerhalb dieses Kapitels auf die allgemeine Idee der Einführung von Toleranzen für Residuen eingegangen werden. Anschließend folgt die Ausarbeitung zweier möglicher mathematischer Realisierungen für die allgemeine Toleranzenbedingung für Residuen sowie des Algorithmus zur Lösung eines um Toleranzenbedingungen erweiterten Ausgleichungsproblems. Anhand eines Beispiels wird abschließend die Anwendung des erarbeiteten Ausgleichungsverfahrens veranschaulicht.

3.1 Veranschaulichung der allgemeinen Idee

An einem einfachen Ausgleichungsproblem soll die allgemeine Idee der Einführung von Toleranzen für Residuen veranschaulicht werden. Dazu wird von drei Messungen y_i (2.5; 2.5; 5.5) ausgegangen, die eine Gerade $y_i = ax_i + b$ beschreiben. Die Werte x_i (1.5; 4.0; 7.0) werden als fehlerfrei und bekannt angenommen und die Werte y_i als gleich genau betrachtet. Durch Minimierung der Summe der Residuenquadrate, ohne Bedingungen bzgl. der Residuen zu berücksichtigen, erhält man die Parameter a und b der ausgleichenden Geraden und den dazugehörigen Residuenvektor (siehe dazu Abbildung 3.1, durchgezogene Linie und Residuen 1-3):

$$\begin{aligned} a &= 0.56044, \\ b &= 1.16484, \\ v &= \begin{pmatrix} -0.49451 \\ 0.90659 \\ -0.41209 \end{pmatrix}. \end{aligned}$$

Es stellt sich nun die Frage: Kann die Gerade derart gedreht und/oder verschoben werden, dass alle Residuen einen Wert z.B. ≤ 0.7 annehmen? Eine mögliche Lösung ist in der Abbildung 3.1 grafisch dargestellt. Darin sind die Datenpunkte, die ausgleichende Gerade (durchgezogene Linie) sowie die ausgleichende Gerade unter Berücksichtigung von maximal zulässigen Werten für die Residuen (gestrichelte Linie) dargestellt. Durch eine Parallelverschiebung und leichte Drehung der Geraden ist es hier möglich, dass sämtliche Residuen ≤ 0.7 werden. Die Parameter der gestrichelt dargestellten Geraden und der dazugehörige Residuenvektor sind:

$$\begin{aligned} a &= 0.56000, \\ b &= 0.96000, \\ v &= \begin{pmatrix} -0.70000 \\ 0.70000 \\ -0.62000 \end{pmatrix}. \end{aligned}$$

Ebenso kann der Betrag der Residuen ausschließlich durch Drehungen oder Parallelverschiebungen der Geraden verändert werden. Es ist aber auch ersichtlich, dass, wenn man den Betrag aller Residuen sehr stark verkleinern möchte, ab einer gewissen Grenze keine Lösung mehr

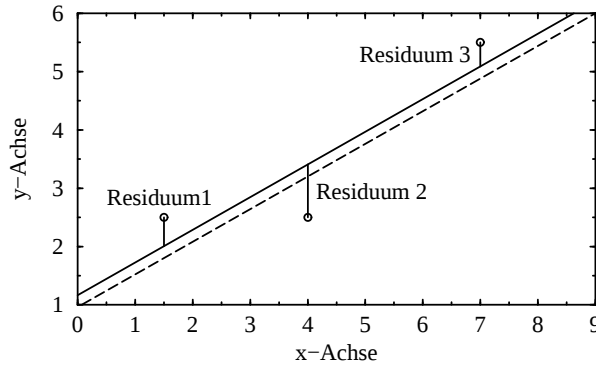


Abbildung 3.1: Zwei mögliche Lösungen mit unterschiedlichen Residuen:
 durchgezogene Linie ... ausgleichende Gerade,
 gestrichelte Linie ... ausgleichende Gerade unter Berücksichtigung von Residuengrenzwerten.

bestimmbar ist. So kann in diesem Beispiel die Gerade beliebig gedreht und verschoben werden, ohne dass alle Residuen einen Wert ≤ 0.4 annehmen können.

Die im vorherigen Kapitel erläuterten Regularisierungsverfahren stellen allgemein eine Ausgleichung nach vermittelnden Beobachtungen mit Bedingungen für die Unbekannten dar. Derartige Ausgleichungsprobleme können in der Regel in eine reine Ausgleichung nach vermittelnden Beobachtungen transformiert werden [Helmert 1872], wobei dann wie bei der stochastischen Regularisierung zwischen Beobachtungen und Pseudobeobachtungen unterschieden werden kann. Die Einbeziehung der Pseudobeobachtungen bewirkt somit allgemein die Regularisierung des Problems. Aufgrund dieses Zusammenhangs wird im Folgenden stets von Ausgleichungsproblemen gesprochen, wobei immer die Regularisierungsverfahren eingeschlossen sind. Um die regularisierende Wirkung der Pseudobeobachtungen steuern zu können, soll nun nicht ein Regularisierungsparameter bestimmt werden, sondern Toleranzen für die Residuen der Beobachtungen (nicht für die Pseudobeobachtungen!) berücksichtigt werden (entspricht einer Regularisierung unter Berücksichtigung von Residuentoleranzen).

Es sind nun folgende Fragen zu klären: Welche Möglichkeiten bestehen, um Bedingungen für Residuen in eine Ausgleichung einzubeziehen? Welche Algorithmen sind denkbar, um die unbekannten Parameter eines Ausgleichungsproblems derart zu bestimmen, dass die Residuen innerhalb sinnvoll gewählter Intervalle liegen?

3.2 Möglichkeiten zur Einführung von Toleranzenbedingungen

Ein beliebiges Ausgleichungsproblem wird in linearisierter Form durch die Funktionalmatrix A , den Vektor der ursprünglichen oder reduzierten Messungen $\tilde{l}$ und durch eine Gewichtsmatrix P beschrieben. Die Gewichtsmatrix berechnet sich wie folgt aus der Kovarianzmatrix der Beobachtungen C_u und einem mittleren Gewichtseinheitsfehler m_0 :

$$P = m_0^2 \cdot C_u^{-1}. \quad (3.1)$$

Auf der Hauptdiagonalen der Kovarianzmatrix der Beobachtungen stehen die quadrierten mittleren Fehler der Beobachtungen. Je kleiner der mittlere Fehler einer Beobachtung ist, desto größer ist das Gewicht dieser Beobachtung innerhalb der Ausgleichung. Die Berechnung der unbekannten Parameter x des Ausgleichungsproblems unter Minimierung der Summe der Residuenquadrate

erfolgt unter Verwendung der allgemein üblichen Formeln, wobei x_0 den Vektor mit den Näherungswerten für die Unbekannten darstellt (siehe dazu [Helmert 1872], dort in der traditionellen Notation):

$$\begin{aligned}\Delta x &= (A^T P A)^{-1} \cdot A^T P \tilde{l}, \\ x &= x_0 + \Delta x.\end{aligned}\tag{3.2}$$

Daraus ist unmittelbar ersichtlich, dass die Lösung eines Ausgleichungsproblems stark variiert werden kann und somit auch der resultierende Residuenvektor, wenn die mittleren Fehler der Beobachtungen variiert werden. So könnte man durch Variation der Gewichtsmatrix iterativ nach einer Lösung suchen, bei der sämtliche oder einzelne Residuen innerhalb bestimmter Grenzen liegen. Auf welchem Wege muss aber die Gewichtsmatrix verändert werden und wie kann man abschätzen, inwieweit Änderungen der Gewichtsmatrix die resultierenden Residuen verändern? Weiterhin sollte es möglich sein, dass nicht nur eine beliebige mögliche Lösung bestimmt wird, sondern diejenige, die z.B. die kleinst mögliche Summe der Residuenquadrate aufweist und gleichzeitig Toleranzen für die Residuen erfüllt. Durch eine Variation der Gewichtsmatrix lässt sich all dies nicht realisieren. Daher wird im Weiteren gezeigt, welche Möglichkeiten bestehen, um auf eleganterem Wege derartige Probleme zu lösen, z.B. indem zusätzliche Bedingungsungleichungen zum ursprünglichen Ausgleichungsproblem formuliert werden.

Die zu erfüllenden Toleranzenbedingungen für die Residuen haben im Allgemeinen die Form: $|v_i| \leq t_i$, d.h. der Betrag der einzelnen Residuen v_i soll kleiner oder gleich einem vorgegebenen Wert (Toleranz) t_i sein. Bei einem heuristischen¹ Auswerteverfahren ist eine derartige Einführung von Toleranzenbedingungen leicht möglich. Dabei wird für sämtliche zu bestimmenden Unbekannten jeweils ein Intervall vorgegeben, in dem die jeweilige Unbekannte vermutet wird. Diese Intervalle stellen den potenziellen Lösungsraum für das betrachtete Problem dar. Um möglichst zuverlässig nach dem globalen Minimum (entspricht der besten Lösung bei Minimierung einer beliebig ausgewählten Norm, hier wurde die L_2 -Norm verwendet) suchen zu können, wird zunächst ein recht großes Intervall für alle Unbekannten als möglicher Lösungsraum vorgegeben. Es wird dann mehrmals (z.B. 100 Millionen mal) ein beliebiger Wert für die Unbekannten aus diesen Intervallen pseudozufällig vorgegeben und der Wert der zu minimierenden Funktion (Summe der Residuenquadrate) als Gütekriterium der Unbekannten berechnet. Die Unbekanntenkombination, für die der Wert der zu minimierenden Funktion (Zielfunktion) am kleinsten ist und für die sämtliche geforderten Residuentoleranzen erfüllt sind, stellt die beste gefundene Lösung dar. Um die Zuverlässigkeit der heuristischen Unbekanntenbestimmung zu erhöhen, sollte der Zufallsgenerator zur Erzeugung der pseudozufälligen Unbekanntenwerte z.B. zehn Mal neu initialisiert werden, so dass zehn Mal unterschiedliche Pseudo-Zufallszahlen-Folgen verwendet werden. Nachdem die beste Lösung unter Verwendung der grob vorgegebenen Intervalle für die Unbekannten bestimmt wurde, werden die Intervalle unter Berücksichtigung der bereits ermittelten Lösung verkleinert. Für diese verkleinerten Intervalle werden dann wieder z.B. 100 Millionen mögliche Unbekanntenkombinationen untersucht für jede der zehn verschiedenen Initialisierungen des Zufallsgenerators. Verwendet man zehn Initialisierungen des Zufallsgenerators, 100 Millionen Generierungen von pseudozufälligen Werten für die Unbekannten und zwei unterschiedlich große Intervalle für die Unbekannten, so sind etwa drei Nachkommastellen der Lösung signifikant.

Aufgrund der beschriebenen Vorgehensweise bei der Bestimmung der Unbekannten mit dem heuristischen Auswerteverfahren ist gewährleistet, dass die Minimierung der Zielfunktion zum

¹vom griechischen „heuriskein“ = finden oder entdecken

globalen Minimum führt und nicht in ein lokales Minimum. Dass sich dieses Verfahren zur globalen Minimumsuche gut eignet, wurde bereits in verschiedenen Arbeiten diskutiert, z.B. [Mautz 2001] und [Kaschenz 2003]. Ebenso wurde dort auch der entscheidende Nachteil dieses Verfahrens erörtert - der Rechenaufwand. Um eine zuverlässige Lösung zu erhalten, ist es einerseits notwendig, sehr viele Unbekanntenkombinationen pseudozufällig zu generieren und andererseits auch mehrere Initialisierungen des Zufallsgenerators zu nutzen. Somit ist der Rechenaufwand um so größer, je zuverlässiger die Lösung sein soll. Aufgrund der geringen Effizienz des heuristischen Verfahrens soll dieses innerhalb der vorliegenden Arbeit lediglich als ein mögliches zusätzliches unabhängiges Kontrollverfahren betrachtet werden.

Um Bedingungsungleichungen für die Residuen innerhalb eines Ausgleichungsproblems berücksichtigen zu können, muss die allgemeine Form der Toleranzbedingung $|v_i| \leq t_i$ derart transformiert werden, dass eine funktionale Darstellung der Bedingungen möglich wird. Zum einen besteht die Möglichkeit darin, die Bedingungen in der Form $v_i^2 \leq t_i^2$ einzuführen und zum anderen für jedes einzelne Residuum zwei Bedingungen zu formulieren: $v_i \leq t_i$ und $-v_i \leq t_i$. Beide Varianten der mathematischen Realisierung der Toleranzbedingung für die Residuen sollen im Weiteren aufgezeigt werden.

3.2.1 Variante 1: $v^2 \leq t^2$

Zunächst soll die Variante betrachtet werden, bei der die Bedingungen derart formuliert werden, dass jedes einzelne Residuumquadrat v_i^2 kleiner oder gleich einer bestimmten quadrierten Toleranz t_i^2 sein soll:

$$v_i^2 \leq t_i^2. \quad (3.3)$$

Die nichtlinearen Beobachtungsgleichungen eines Ausgleichungsproblems haben die bekannte Form:

$$l_i + v_i = f_i(x_j). \quad (3.4)$$

Werden diese an der Stelle der Näherungswerte x_0 linearisiert, so erhält man:

$$l_i + v_i = \left[\left(\frac{\partial f_i}{\partial x_j} \right)_{x_0} \right] \Delta x + f_i(x_0). \quad (3.5)$$

Durch Zusammenfassen folgt:

$$v_i = A\Delta x - \tilde{l}_i, \quad (3.6)$$

mit

$$\begin{aligned} A &= \left[\left(\frac{\partial f_i}{\partial x_j} \right)_{x_0} \right], \\ \tilde{l}_i &= l_i - f_i(x_0). \end{aligned}$$

Somit folgt für die nichtlinearen Gleichungen für die Residuenquadrate:

$$v_i^2 = (f_i(x_j) - l_i)^2 = g_i(x_j, l_i) \quad (3.7)$$

und in linearisierter Form:

$$v_i^2 = B_1\Delta x + b_1, \quad (3.8)$$

mit

$$\begin{aligned} B_1 &= \left[\left(\frac{\partial g_i}{\partial x_j} \right)_{x_0} \right], \\ b_1 &= [g_i(x_0, l_i)]. \end{aligned}$$

Daraus lassen sich die linearen Bedingungsungleichungen formulieren:

$$v^2 \leq t^2 \Leftrightarrow B_1 \Delta x + b_1 \leq t^2. \quad (3.9)$$

3.2.2 Variante 2: $v \leq t$ und $-v \leq t$

Die zweite Variante, Bedingungen für die Residuen zu formulieren, resultiert aus der Forderung, dass alle Residuen v_i und $-v_i$ kleiner oder gleich einer bestimmten Toleranz $t_i > 0$ sein sollen. Somit erhält man nach Formel (3.4) für jede Beobachtung zwei nichtlineare Gleichungen für die Residuen:

$$\begin{aligned} v_i &= f_i(x_j) - l_i = h_i(x_j, l_i), \\ -v_i &= l_i - f_i(x_j) = -h_i(x_j, l_i), \end{aligned}$$

und in linearisierter Form:

$$\pm v_i = B_2 \Delta x + b_2, \quad (3.10)$$

mit

$$\begin{aligned} B_2 &= \left[\left(\frac{\partial(\pm h_i)}{\partial x_j} \right)_{x_0} \right], \\ b_2 &= [\pm h_i(x_0, l_i)]. \end{aligned}$$

Daraus folgen wieder die linearen Bedingungsungleichungen:

$$v \leq t \text{ und } -v \leq t \Leftrightarrow B_2 \Delta x + b_2 \leq t. \quad (3.11)$$

Es ist sofort ersichtlich, dass die partiellen Ableitungen $\frac{\partial h}{\partial x}$ bis auf das Vorzeichen identisch sind zu den partiellen Ableitungen $\frac{\partial f}{\partial x}$. Bei Verwendung der Variante 2 der Formulierung der Toleranzbedingung benötigt man daher nur eine Art der Linearisierung, bei der Variante 1 dagegen zwei verschiedene Arten. Somit kann man vermuten, dass sich die Variante 2 der Bedingungsungleichungen als günstiger erweisen wird als die Variante 1.

3.3 Ausarbeitung der Algorithmen

Im Weiteren muss geklärt werden, wie ein Ausgleichungsproblem unter Einbeziehung der dargestellten Bedingungsungleichungen gelöst werden kann. In vielen gängigen Büchern zur Ausgleichungsrechnung (z.B. [Großmann 1953], [Reißmann 1968], [Gotthardt 1968]) findet man zwar Algorithmen zur Lösung der sog. „vermittelnden Ausgleichung mit Bedingungen zwischen den Unbekannten“ und zur „bedingten Ausgleichung mit Unbekannten“, wo die Einführung von Bedingungsgleichungen möglich ist, aber in beiden Fällen handelt es sich um „Gleichheits-Bedingungen“. Bei den vorliegenden Bedingungen wird aber nicht die exakte Gleichheit gefordert, sondern v_i^2 bzw. $\pm v_i$ soll kleiner oder gleich einem Wert sein. Somit können

diese Algorithmen nicht verwendet werden.

In [Lawson und Hanson 1974] wird ein sog. LSI (**L**east **S**quares with **I**nequality constraints)-Problem für lineare Ausgleichungsprobleme betrachtet und anhand eines Beispiels mit Ungleichheitsbedingungen für die Unbekannten veranschaulicht. Die Definition des LSI-Problems lautet dort wie folgt (aus dem Englischen übersetzt):

E sei eine $(m_2 \times n)$ -Matrix, f ein m_2 -Vektor, G eine $(m \times n)$ -Matrix und h ein m -Vektor. Das entsprechende Ausgleichungsproblem nach kleinsten Quadraten mit linearen Ungleichheitsbedingungen kann dann wie folgt formuliert werden:

$$\min \|Ex - f\| \text{ unter Berücksichtigung von } Gx \geq h.$$

Des Weiteren werden von Lawson und Hanson zwei Spezialfälle des LSI-Problems betrachtet:

NNLS (**N**on**N**egative **L**east **S**quares)-Problem:
 $\min \|Ex - f\| \text{ unter Berücksichtigung von } x \geq 0.$

LDP (**L**east **D**istance **P**rogramming)-Problem:
 $\min \|x\| \text{ unter Berücksichtigung von } Gx \geq h.$

Es stellt sich nun die Frage, ob das vorliegende Ausgleichungsproblem mit Toleranzenbedingungen für Residuen in ein von Lawson und Hanson betrachtetes Problem transformiert werden kann. Nach Gleichung (3.6) lautet das zu minimierende Problem

$$\min \|A\Delta x - \tilde{l}\|. \quad (3.12)$$

Die ermittelten linearen Bedingungsungleichungen, mit denen Toleranzen für die Residuen berücksichtigt werden können, lauten nach den Formeln (3.9) und (3.11)

$$B_1\Delta x + b_1 \leq t^2 \text{ bzw.} \quad (3.13)$$

$$B_2\Delta x + b_2 \leq t. \quad (3.14)$$

Man erkennt, dass das vorgestellte LSI-Problem eine ähnliche Form aufweist, wie das zu untersuchende Ausgleichungsproblem mit Toleranzenbedingungen. Mit den Substitutionen $A = E$ und $\tilde{l} = f$ folgt aus Gleichung (3.12)

$$\min \|E\Delta x - f\|. \quad (3.15)$$

Eine Substitution mit $\Delta x = x$ wurde bewusst nicht durchgeführt, da in [Lawson und Hanson 1974] lediglich lineare Probleme betrachtet werden, hier aber gleich auf nichtlineare Probleme übergegangen werden soll.

Die Gleichungen (3.13) und (3.14) können nun wie folgt umgestellt werden

$$\begin{aligned} B_1\Delta x &\leq t^2 - b_1, \\ B_2\Delta x &\leq t - b_2, \\ &\Updownarrow \\ -B_1\Delta x &\geq -t^2 + b_1, \\ -B_2\Delta x &\geq -t + b_2, \\ &\Updownarrow \\ G_1\Delta x &\geq h_1, \end{aligned} \quad (3.16)$$

$$G_2\Delta x \geq h_2, \quad (3.17)$$

wobei die Substitutionen $G_1 = -B_1$ bzw. $G_2 = -B_2$ durchgeführt wurden sowie $h_1 = b_1 - t^2$ bzw. $h_2 = b_2 - t$ ist.

Somit ist das hier betrachtete Ausgleichungsproblem mit Toleranzenbedingungen in das LSI-Problem transformierbar, so dass eine Nutzung von in [Lawson und Hanson 1974] vorgestellten Transformationen und Algorithmen möglich wird.

Da in [Lawson und Hanson 1974] lediglich für das NNLS-Problem ein Lösungsalgorithmus angegeben ist, wird dort zusätzlich die Transformation eines LSI-Problems in ein NNLS-Problem beschrieben. Der in [Lawson und Hanson 1974] beschriebene Lösungsalgorithmus kann zwar nur für lineare Probleme verwendet werden, doch ist eine Erweiterung dieses Algorithmus auf nicht-lineare Probleme vorstellbar. Daher soll im Folgenden die Transformation eines LSI-Problems in ein NNLS-Problem aufgezeigt werden, welche sich in zwei Transformationsschritte unterteilt (dass eine direkte Transformation nicht möglich ist, wird im Abschnitt 3.3.4 gezeigt): 1. Transformation des LSI-Problems in ein LDP-Problem und 2. Transformation des LDP-Problems in ein NNLS-Problem. Anschließend wird der für nichtlineare Probleme erweiterte Algorithmus zur Bestimmung der Lösung eines NNLS-Problems dargelegt.

3.3.1 Transformation eines LSI-Problems in ein LDP-Problem

Gegeben sei ein LSI-Problem durch die Matrizen E ($m_2 \times n$) mit $\text{Rang}(E) = n$ und G ($m \times n$) und die Vektoren f ($m_2 \times 1$) und h ($m \times 1$). Zunächst wird die Matrix E mittels einer Singulärwertzerlegung (siehe [Press et al. 1992]) in die Matrizen H , $\bar{R}$ und K zerlegt:

$$\begin{aligned} E &= H\bar{R}K^T, \\ E &= H \begin{bmatrix} R \\ 0 \end{bmatrix} K^T, \end{aligned} \quad (3.18)$$

wobei H eine ($m_2 \times m_2$)-Orthogonalmatrix ist, K eine ($n \times n$)-Orthogonalmatrix, R eine ($n \times n$)-nichtsinguläre Matrix und $\bar{R}$ eine ($m_2 \times n$)-Matrix. Für die zu minimierende Funktion $\varphi(x)$ erhält man:

$$\begin{aligned} \varphi(x) &= (\|Ex - f\|)^2 \\ &= (\|H\bar{R}K^T x - f\|)^2 \\ &= \left(\left\| H \begin{bmatrix} R \\ 0 \end{bmatrix} K^T x - f \right\| \right)^2. \end{aligned} \quad (3.19)$$

Führt man die Substitution $x = Ky$ durch und berücksichtigt, dass $K^T K = K^{-1} K = \text{Einheitsmatrix}$ ist, dann folgt:

$$\varphi(x) = \left(\left\| H \begin{bmatrix} R \\ 0 \end{bmatrix} y - f \right\| \right)^2. \quad (3.20)$$

Die Gleichung (3.20) wird im nächsten Schritt mit $(\|H^T\|)^2$ multipliziert. Dabei wird berücksichtigt, dass $(\|H^T\|)^2 = 1$ ist, da H eine Orthogonalmatrix ist, und dass $\|H\|\|w\| = \|Hw\|$ ist, wobei w einen beliebigen Vektor darstellt (siehe dazu [Lawson und Hanson 1974], S. 5). In der Gleichung (3.20) stellt der Term innerhalb $(\|\dots\|)^2$ einen ($m_2 \times 1$)-Vektor dar.

$$\begin{aligned} (\|H^T\|)^2 \varphi(x) &= \left(\|H^T\| \cdot \left\| H \begin{bmatrix} R \\ 0 \end{bmatrix} y - f \right\| \right)^2 \\ &\Downarrow \end{aligned}$$

$$\begin{aligned}
\varphi(x) &= \left(\left\| H^T H \begin{bmatrix} R \\ 0 \end{bmatrix} y - H^T f \right\| \right)^2 \\
&= \left(\left\| \begin{bmatrix} R \\ 0 \end{bmatrix} y - H^T f \right\| \right)^2 \\
&= \left(\left\| H^T f - \begin{bmatrix} R \\ 0 \end{bmatrix} y \right\| \right)^2.
\end{aligned} \tag{3.21}$$

Da $\bar{R}$ eine $(m_2 \times n)$ -Matrix, R eine $(n \times n)$ -Matrix und y einen $(n \times 1)$ -Vektor darstellen, folgt, dass $\bar{R}y = \begin{bmatrix} Ry : 0 \end{bmatrix}^T$ ist. Weiterhin wird $H^T f$ wie folgt zerlegt:

$$H^T f = \begin{bmatrix} H_1^T f \\ H_2^T f \end{bmatrix}, \tag{3.22}$$

wobei $H_1^T f$ einen $(n \times 1)$ -Vektor darstellt und $H_2^T f$ einen $((m_2 - n) \times 1)$ -Vektor. Somit kann die Gleichung (3.21) umgeschrieben werden in:

$$\varphi(x) = \left(\left\| \begin{bmatrix} H_1^T f \\ H_2^T f \end{bmatrix} - \begin{bmatrix} Ry \\ 0 \end{bmatrix} \right\| \right)^2 \tag{3.23}$$

$$= (\|H_1^T f - Ry\|)^2 + (\|H_2^T f\|)^2. \tag{3.24}$$

Dabei ist der Übergang von Gleichung (3.23) nach Gleichung (3.24) nur möglich, da der Term in Gleichung (3.23) innerhalb von $(\|\dots\|)^2$ einen Vektor darstellt und eine beliebige p -Norm eines Vektors definiert ist durch (siehe [Wilkinson 1969]):

$$(\|w\|)^p = (|w_1|^p + |w_2|^p + \dots + |w_n|^p)^{1/p} \quad (p = 1, 2, \dots, \infty), \tag{3.25}$$

wobei unter $(\|w\|)^\infty \max |w_i|$ verstanden wird.

Der Term $(\|H_2^T f\|)^2$ ist bei linearen und nichtlinearen (hier nur während einer Iteration) Problemen konstant und muss somit bei der Minimierung der Funktion $\varphi(x)$ nicht mit berücksichtigt werden. Substituiert man $z = Ry - H_1^T f$, so erhält man die endgültige zu minimierende Funktion des LDP-Problems:

$$\varphi(x) = (\|z\|)^2. \tag{3.26}$$

Durch Einsetzen von $x = Ky$ und $y = R^{-1}(z + H_1^T f)$ in die Bedingungsungleichung $Gx \geq h$ des LSI-Problems erhält man abschließend die Bedingungsungleichung des LDP-Problems:

$$\begin{aligned}
Gx &\geq h \\
&\Downarrow \\
GKy &\geq h \\
&\Downarrow \\
GKR^{-1}(z + H_1^T f) &\geq h \\
&\Downarrow \\
\tilde{G}z &\geq \tilde{h}
\end{aligned} \tag{3.27}$$

mit

$$\begin{aligned}
\tilde{G} &= GKR^{-1}, \\
\tilde{h} &= h - \tilde{G}H_1^T f.
\end{aligned}$$

Nach Berechnung der LDP-Lösung z kann über $y = R^{-1}(z + H_1^T f)$ und $x = Ky$ die LSI-Lösung x berechnet werden.

3.3.2 Transformation eines LDP-Problems in ein NNLS-Problem

Im Folgenden wird die Transformation eines LDP-Problems in ein NNLS-Problem ohne Herleitung angegeben. Es wird lediglich der Zusammenhang zwischen der LDP-Lösung z und der NNLS-Lösung u aufgezeigt.

Gegeben sei ein LDP-Problem $\min \|z\|$ mit $\tilde{G}z \geq \tilde{h}$. Die Transformation in das dazugehörige NNLS-Problem $\min \|\tilde{E}u - \tilde{f}\|$ mit $u \geq 0$ erfolgt mit (siehe [Lawson und Hanson 1974]):

$$\tilde{E} = \begin{bmatrix} \tilde{G}^T \\ \tilde{h}^T \end{bmatrix}, \quad \tilde{f} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{bmatrix}. \quad (3.28)$$

Die NNLS-Lösung u besitzt nach dem Kuhn-Tucker-Theorem ([Lawson und Hanson 1974], S. 165/166) folgende Eigenschaften:

$$\begin{aligned} u_i &= 0 \quad \text{und} \quad p_i \geq 0 \quad \text{für} \quad i \in M_1, \\ u_i &> 0 \quad \text{und} \quad p_i = 0 \quad \text{für} \quad i \in M_2 \end{aligned}$$

mit

$$r = \tilde{E}u - \tilde{f} \quad \text{und} \quad p = \tilde{E}^T r,$$

wobei p den Gradienten der zu minimierenden Funktion des NNLS-Problems darstellt.

Wird das LDP-Problem derart gelöst, so muss gewährleistet sein, dass aus der NNLS-Lösung u die LDP-Lösung z berechnet werden kann und diese dann einerseits tatsächlich das Minimum der zu minimierenden Funktion des LDP-Problems darstellt und andererseits die LDP-Bedingungen erfüllt werden.

Gemäß der Kuhn-Tucker-Bedingungen für z bei der Minimierung von $\min \|z\|$ mit $\tilde{G}z \geq \tilde{h}$ ergibt sich die Forderung, dass der Gradientenvektor z als nicht-negative Linearkombination der Zeilen der Matrix $\tilde{G}$ darstellbar sein muss. Daraus folgt für z unter Berücksichtigung der Vorzeichenbedingungen für u ([Lawson und Hanson 1974], S. 160):

$$r = \tilde{E}u - \tilde{f} = \begin{bmatrix} \tilde{G}^T \\ \tilde{h}^T \end{bmatrix} u - \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{bmatrix} = \begin{bmatrix} \tilde{G}^T u \\ \tilde{h}^T u - 1 \end{bmatrix} = \begin{bmatrix} r_1 \\ \vdots \\ r_{n+1} \end{bmatrix} \quad (3.29)$$

$$\begin{aligned} \Downarrow \\ z &= \tilde{G}^T u (-r_{n+1})^{-1} \\ &= \begin{bmatrix} r_1 \\ \vdots \\ r_n \end{bmatrix} (-r_{n+1})^{-1}, \end{aligned} \quad (3.30)$$

so dass z damit tatsächlich das Minimum des LDP-Problems darstellt.

Im Weiteren wird geprüft, ob die LDP-Bedingungen erfüllt sind, wenn die NNLS-Lösung obige Eigenschaften aufweist. Dazu muss zunächst betrachtet werden, welche Eigenschaften $(\|r\|)^2$ besitzt:

$$\begin{aligned} p &= \tilde{E}^T r \Rightarrow p^T = r^T \tilde{E} \Rightarrow p^T u = r^T \tilde{E} u, \\ (\|r\|)^2 &= r^T r = r^T (\tilde{E} u - \tilde{f}) = r^T \tilde{E} u - r^T \tilde{f} = p^T u - r_{n+1}. \end{aligned} \quad (3.31)$$

Da entweder $u_i = 0$ oder $p_i = 0$ ist (siehe Eigenschaften der NNLS-Lösung), folgt:

$$(\|r\|)^2 = -r_{n+1}.$$

Der Wert von $(\|r\|)^2$ kann nach Formel (3.25) wie folgt berechnet werden:

$$(\|r\|)^2 = (r_1^2 + r_2^2 + \dots + r_{n+1}^2)^{1/2}. \quad (3.32)$$

Daraus folgt, dass $(\|r\|)^2 > 0$, wenn r keinen Nullvektor darstellt bzw. r_{n+1} negativ ist.

Als erstes erfolgt nun die Überprüfung der Gültigkeit der LDP-Bedingungen für u_i und p_i der M_1 -Menge, so dass gilt: $p_i \geq 0$ und $u_i = 0$.

$$p = \tilde{E}^T r = \begin{bmatrix} \tilde{G}^T \\ \tilde{h}^T \end{bmatrix}^T \begin{bmatrix} z \\ -1 \end{bmatrix} (-r_{n+1}) = (\tilde{G}z - \tilde{h})(-r_{n+1}) = (\tilde{G}z - \tilde{h})(\|r\|)^2. \quad (3.33)$$

Da $p_i \geq 0$ sein muss und $(\|r\|)^2 > 0$ gilt, folgt, dass $\tilde{G}z - \tilde{h} \geq 0$ sein muss, damit diese Eigenschaft der NNLS-Lösung erfüllt ist. Daher ist $\tilde{G}z \geq \tilde{h}$ für $p_i \geq 0$ und $u_i = 0$ erfüllt. Für $p_i = 0$ und $u_i > 0$ der M_2 -Menge gilt analog, dass $\tilde{G}z - \tilde{h} = 0$ sein muss, damit ebenso diese Eigenschaft der NNLS-Lösung erfüllt ist, d.h. $\tilde{G}z = \tilde{h}$. Die Bedingungen des LDP-Problems sind somit erfüllt, wenn $(\|r\|)^2 > 0$ ist. Daraus kann man schließen, dass, wenn $(\|r\|)^2 = 0$ ist, die Bedingungen nicht erfüllt sind.

Mit $r = \begin{bmatrix} z \\ -1 \end{bmatrix} (-r_{n+1})$ und $r = \tilde{E}u - \tilde{f}$ folgt:

$$\begin{aligned} (\|r\|)^2 &= r^T r = \begin{bmatrix} z^T & -1 \end{bmatrix} \left(\begin{bmatrix} \tilde{G}^T \\ \tilde{h}^T \end{bmatrix} u - \tilde{f} \right) (-r_{n+1}) \\ &= \left(z^T \tilde{G}^T u - \tilde{h}^T u + 1 \right) (-r_{n+1}) = \left[(z^T \tilde{G}^T - \tilde{h}^T) u + 1 \right] (-r_{n+1}) \\ &= (q^T u + 1) (-r_{n+1}), \end{aligned} \quad (3.34)$$

wobei $q = \tilde{G}z - \tilde{h} \geq 0$ ist. Da $u_i \geq 0$ gilt, muss $q^T < 0$ sein, damit $(\|r\|)^2 = 0$ werden kann. q darf aber nicht negativ werden, wenn die LDP-Bedingungen erfüllt werden sollen. Somit sind die LDP-Bedingungen nicht erfüllt, wenn $(\|r\|)^2 = 0$ ist.

Liegt die NNLS-Lösung u vor, so kann diese mittels $r = \tilde{E}u - \tilde{f}$ und der Formel (3.30) in die LDP-Lösung z zurücktransformiert werden.

3.3.3 Lösungsalgorithmus für NNLS-Probleme

Gegeben sei eine $(m_2 \times n)$ -Matrix $\tilde{E}$ und ein $(m_2 \times 1)$ -Vektor $\tilde{f}$. Des Weiteren definiert man zwei $(1 \times n)$ -Vektoren P und Z . Die Lösung des NNLS-Problems soll mit u bezeichnet werden und d als Dual-Vektor. Der Lösungsalgorithmus zur Bestimmung von u lautet dann wie folgt:

1. Setze alle Elemente von P und u zu null und $Z = (1, 2, \dots, n)$.
2. Berechne den Dual-Vektor $d = \tilde{E}^T(\tilde{f} - \tilde{E}u)$.
3. Sind alle Elemente von Z gleich null oder alle Elemente d_j kleiner oder gleich null für $j \in Z$, dann wird die Berechnung von u abgeschlossen.
4. Bestimme den Index $t \in Z$, für den gilt: $d_t = \max(d_j : j \in Z)$.
5. Verschiebe den Index t vom Vektor Z in den Vektor P .

6. Definiere eine $(m_2 \times n)$ -Matrix E_p mit:
 Spalte j von $E_p :=$ Spalte j von $\tilde{E}$ für $j \in P$ und 0 für $j \in Z$
 und berechne den $(n \times 1)$ -Vektor $\bar{z}$ des Ausgleichungsproblems nach kleinsten Quadraten $\min \|E_p \bar{z} - \tilde{f}\|$. Dabei ist zu berücksichtigen, dass lediglich die Komponenten $\bar{z}_j, j \in P$ berechnet werden können. Für $j \in Z$ wird $\bar{z}_j = 0$ gesetzt.
7. Ist $\bar{z}_j > 0$ für alle $j \in P$, dann setze $u = \bar{z}$ und gehe zu Schritt 2.

Dieser Lösungsalgorithmus wurde aus [Lawson und Hanson 1974] entnommen, wobei dort noch vier weitere Schritte zu finden sind. Mittels dieser zusätzlichen Schritte wird versucht, Rundungsfehler zu berücksichtigen, wobei die Autoren anmerken, dass auch damit nicht alle Probleme, verursacht durch Rundungsfehler, vermieden werden können. Daher wird auf diese zusätzlichen Schritte verzichtet, da, wenn Probleme durch Rundungen festgestellt werden, auf eine höhere Rechengenauigkeit (z.B. von doppelter Genauigkeit auf vierfache Genauigkeit bei Fortran 90) übergegangen werden kann.

Im Schritt 2 wird der Dual-Vektor d berechnet, welcher den Gradienten der zu minimierenden Funktion $(\|\tilde{E}u - \tilde{f}\|)^2$ darstellt. Sind alle Elemente dieses Vektors ≤ 0 , so bedeutet das, dass einerseits die zu minimierende Funktion nicht weiter minimiert werden kann - man befindet sich bereits im Minimum - und andererseits, dass keine weiteren Unbekannten $u_j > 0$ bestimmt werden können. Liegt $Z = \vec{0}$ vor, so konnten alle Unbekannten berechnet werden, wobei diese > 0 sind.

Grundsätzlich erfolgt die Berechnung der Unbekannten u_j sukzessive und zwar derart, dass immer die Unbekannte hinzugenommen wird, deren Anteil am Gradientenvektor d maximal ist, so dass für diese die Wahrscheinlichkeit, die zu minimierende Funktion weiter minimieren zu können, besonders hoch ist. Alle Unbekannten u_j , für die kein Wert > 0 bestimmt werden konnte, werden zu null gesetzt, so dass die NNLS-Bedingungen jedenfalls erfüllt sind.

Für die Berechnung der Elemente der Matrix $\tilde{E}$ und des Vektors $\tilde{f}$ benötigt man bei nichtlinearen Problemen (z.B. wenn Residuentoleranzen eingeführt werden) Näherungswerte x_0 für die LSI-Lösung x (in [Lawson und Hanson 1974] werden nur lineare Probleme betrachtet). Somit stellen die Schritte 1-7 lediglich den Iterationsalgorithmus zur Bestimmung der NNLS-Lösung dar. Es muss aber noch eine weitere Iteration erfolgen, um zu berücksichtigen, dass die ursprünglichen Beobachtungsgleichungen nichtlinear sind. Für diese Iteration wird zunächst eine Näherungslösung x_0 benötigt, um die Matrizen und Vektoren des LSI-Problems aufstellen zu können. Dieses LSI-Problem wird dann in das dazugehörige NNLS-Problem transformiert und gelöst. Die NNLS-Lösung u muss zurück in die LSI-Lösung Δx transformiert werden. Die erhaltene Lösung $x = x_0 + \Delta x$ wird dann als neue Näherungslösung betrachtet und sämtliche Berechnungen von neuem durchgeführt. Diese „äußere“ Iteration wird beendet, wenn die Summe der Residuenquadrate z.B. 10000 Mal innerhalb der Rechengenauigkeit gleich blieb bzw. eine maximal vorgegebene Anzahl von Iterationen durchgeführt wurde. Dies ist nur eine Möglichkeit zur Festlegung des Abbruchs der Iteration und kann keineswegs als immer vorteilhaft betrachtet werden. So ist es theoretisch möglich, dass die Konvergenz zum Minimum der zu minimierenden Funktion so extrem langsam erfolgt, dass mit diesem Abbruchkriterium der Iterationsprozess zu früh abgebrochen wird. Dieses Abbruchkriterium hat sich aber bei den getesteten Beispielen als vorteilhaft erwiesen und ist sicherlich viel geeigneter als das allgemein übliche Abbruchkriterium - Abbruch, wenn die neue Lösung x erstmalig von der Näherungslösung x_0 nicht signifikant verschieden ist.

3.3.4 Beweis, dass die direkte Transformation eines LSI-Problems in ein NNLS-Problem nicht gültig ist

Selbstverständlich stellt sich einem sofort die Frage, warum eine derart umständliche Transformation des LSI-Problems in ein entsprechendes NNLS-Problem notwendig ist. Wesentlich einfacher scheint auf den ersten Blick folgende Transformation zu sein:

Gegeben sei ein LSI-Problem der Art $\min\|Ex - f\|$ mit den Bedingungen $Gx \geq h$. Dieses soll nun direkt in das dazugehörige NNLS-Problem $\min\|\tilde{E}u - \tilde{f}\|$ mit $u \geq 0$ transformiert werden. Aus $Gx \geq h$ folgt:

$$Gx - h \geq 0. \quad (3.35)$$

Daraus folgt, dass $u = Gx - h$ ist. Stellt man diese Gleichung nach x um, so erhält man:

$$u = Gx - h, \quad (3.36)$$

$$u + h = Gx, \quad (3.37)$$

$$G^T(u + h) = G^T Gx, \quad (3.38)$$

$$x = (G^T G)^{-1} G^T (u + h).$$

Wird nun Gleichung (3.38) in $Ex - f$ eingesetzt, erhält man:

$$Ex - f = E(G^T G)^{-1} G^T u + E(G^T G)^{-1} G^T h - f. \quad (3.39)$$

Daraus folgt für $\tilde{E}$ und $\tilde{f}$:

$$\tilde{E} = E(G^T G)^{-1} G^T, \quad (3.40)$$

$$\tilde{f} = f - E(G^T G)^{-1} G^T h. \quad (3.41)$$

Mittels des Lösungsalgorithmus für NNLS-Probleme könnte man nun die NNLS-Lösung u berechnen und diese anhand der Gleichung (3.38) zurück in die ursprüngliche LSI-Lösung x transformieren.

Leider ist diese sehr einfache Transformation nicht gültig, da die Gleichungen (3.36) und (3.37) nicht äquivalent zueinander sind. Das bedeutet in diesem Fall, dass zwar die Umformung von Gleichung (3.36) nach Gleichung (3.37) gilt, die umgekehrte Umformung aber nicht. Dies ist aber notwendig, da man die NNLS-Lösung zurück in die LSI-Lösung transformieren muss.

Zwei Gleichungen $A(x) = 0$ und $B(x) = 0$ heißen äquivalent zueinander, wenn folgendes gilt: Wenn für gewisse x die Gleichung $A(x) = 0$ erfüllt ist, dann muss auch $B(x) = 0$ für diese x -Werte gelten und umgekehrt.

Das bedeutet speziell auf die Gleichungen (3.36) und (3.37) bezogen, wenn die Gleichung $G^T(u + h) = G^T Gx$ erfüllt ist, dann muss auch die Gleichung $u + h = Gx$ erfüllt sein. Gibt es ein u und x , für die dies nicht gilt, so sind die Gleichungen (3.36) und (3.37) nicht äquivalent. Das folgende Beispiel zeigt, dass die Gleichungen (3.36) und (3.37) nicht äquivalent sind:

$$u = \begin{bmatrix} 458646.8131 \\ 0.0000 \\ 600000.2488 \end{bmatrix},$$

$$x = \begin{bmatrix} 0.0304 \\ 0.1248 \end{bmatrix},$$

$$h = \begin{bmatrix} -999999.5578 \\ -999999.5644 \\ -999999.4375 \end{bmatrix},$$

$$G = \begin{bmatrix} 1.9950 & 1.3300 \\ -5.2800 & -1.3200 \\ 10.5000 & 1.5000 \end{bmatrix}.$$

$G^T(u + h) = G^T Gx$ gilt, da:

$$\begin{bmatrix} 7.4932 \\ 1.4916 \end{bmatrix} = \begin{bmatrix} 7.4932 \\ 1.4916 \end{bmatrix},$$

$u + h = Gx$ gilt aber nicht, da:

$$\begin{bmatrix} -5.4135 \cdot 10^5 \\ -1.0000 \cdot 10^6 \\ -4.0000 \cdot 10^5 \end{bmatrix} \neq \begin{bmatrix} 0.2268 \\ -0.3255 \\ 0.5069 \end{bmatrix}.$$

Damit ist bewiesen, dass diese direkte Transformation eines LSI-Problems in ein entsprechendes NNLS-Problem nicht gültig ist und somit die aufwendigere Transformation, wie sie in den vorangegangenen Abschnitten beschrieben wurde, genutzt werden muss.

3.3.5 Veranschaulichung der Ausgleichung mit Toleranzenbedingungen

Zur besseren Übersicht wird im Folgenden zusammengestellt, wie ein Ausgleichungsproblem mit Toleranzenbedingungen für die Residuen aufgestellt und gelöst werden kann. Dies wird exemplarisch anhand eines einfachen Beispiels erfolgen.

Im Beispiel sollen die Parameter a und b einer Geraden $y = aw + b$ bestimmt werden, wobei die Messungen y_i mit $i = 1, 2, 3$ vorliegen, welche als gleich genau betrachtet werden. Somit entfällt die Einführung einer Gewichtsmatrix innerhalb dieser Ausgleichung. Die w_i -Werte werden als fehlerfrei und bekannt vorausgesetzt. Die Toleranzen für die Residuen v_i werden mit t_i bezeichnet.

Die Beobachtungsgleichungen lauten dann wie folgt:

$$\begin{aligned} y_1 + v_1 &= aw_1 + b, \\ y_2 + v_2 &= aw_2 + b, \\ y_3 + v_3 &= aw_3 + b. \end{aligned}$$

Substituiert man $a = a_0 + \Delta a$ und $b = b_0 + \Delta b$ und verwendet die Matrizenschreibweise, folgt:

$$\begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix} = E\Delta x - f, \tag{3.42}$$

mit

$$E = \begin{bmatrix} w_1 & 1 \\ w_2 & 1 \\ w_3 & 1 \end{bmatrix}, \quad f = \begin{bmatrix} y_1 - a_0 w_1 - b_0 \\ y_2 - a_0 w_2 - b_0 \\ y_3 - a_0 w_3 - b_0 \end{bmatrix}, \quad \Delta x = \begin{bmatrix} \Delta a \\ \Delta b \end{bmatrix}, \quad x_0 = \begin{bmatrix} a_0 \\ b_0 \end{bmatrix}.$$

Anhand der durchgeführten Substitution erfolgt für die Anwendung beider Varianten der Realisierung der Toleranzenbedingung der Übergang des hier linearen Ausgleichungsproblems

in ein nichtlineares Ausgleichungsproblem. Dies ist zwingend notwendig, wenn die Variante 1 der Bedingungsungleichungen für die Residuen verwendet werden soll, da diese Ungleichungen selbst bei linearen Ausgleichungsproblemen nichtlinear sind.

Damit lautet die zu minimierende Funktion des LSI-Problems: $\min \|E\Delta x - f\|$. Für die Bedingungsungleichungen gilt nach Abschnitt 3.2.1 und 3.2.2:

$$B_1\Delta x + b_1 \leq t^2 \quad \text{bzw.} \quad B_2\Delta x + b_2 \leq t, \quad (3.43)$$

mit

$$B_1 = \begin{bmatrix} 2(a_0w_1 + b_0 - y_1)w_1 & 2(a_0w_1 + b_0 - y_1) \\ 2(a_0w_2 + b_0 - y_2)w_2 & 2(a_0w_2 + b_0 - y_2) \\ 2(a_0w_3 + b_0 - y_3)w_3 & 2(a_0w_3 + b_0 - y_3) \end{bmatrix},$$

$$b_1 = \begin{bmatrix} (a_0w_1 + b_0 - y_1)^2 \\ (a_0w_2 + b_0 - y_2)^2 \\ (a_0w_3 + b_0 - y_3)^2 \end{bmatrix}, \quad t^2 = \begin{bmatrix} t_1^2 \\ t_2^2 \\ t_3^2 \end{bmatrix},$$

bzw.

$$B_2 = \begin{bmatrix} w_1 & 1 \\ -w_1 & -1 \\ w_2 & 1 \\ -w_2 & -1 \\ w_3 & 1 \\ -w_3 & -1 \end{bmatrix}, \quad b_2 = \begin{bmatrix} a_0w_1 + b_0 - y_1 \\ -(a_0w_1 + b_0 - y_1) \\ a_0w_2 + b_0 - y_2 \\ -(a_0w_2 + b_0 - y_2) \\ a_0w_3 + b_0 - y_3 \\ -(a_0w_3 + b_0 - y_3) \end{bmatrix}, \quad t = \begin{bmatrix} t_1 \\ t_1 \\ t_2 \\ t_2 \\ t_3 \\ t_3 \end{bmatrix}.$$

Die Bedingungsungleichungen $G\Delta x \geq h$ des LSI-Problems erhält man dann aus:

$$G_i = -B_i \text{ für } i=1, 2 \text{ und } h_1 = b_1 - t^2 \text{ bzw. } h_2 = b_2 - t.$$

Somit stehen nun die E -Matrix, der f -Vektor, die G -Matrix und der h -Vektor des LSI-Problems zur Verfügung. Die $\tilde{G}$ -Matrix und den $\tilde{h}$ -Vektor des LDP-Problems $\min \|z\|$ mit $\tilde{G}z \geq \tilde{h}$ berechnet man dann wie folgt:

$$\begin{aligned} \tilde{G} &= GK\bar{R}^{-1}, \\ \tilde{h} &= h - \tilde{G}H_1^T f, \end{aligned}$$

wobei K , $\bar{R}$ und H_1^T gemäß Abschnitt 3.3.1 bestimmt werden. Die Matrix $\tilde{E}$ und den Vektor $\tilde{f}$ des dazugehörigen NNLS-Problems $\min \|\tilde{E}u - \tilde{f}\|$ mit $u \geq 0$ erhält man aus:

$$\tilde{E} = \begin{bmatrix} \tilde{G}^T \\ \tilde{h}^T \end{bmatrix}, \quad \tilde{f} = \begin{bmatrix} 0 \\ \vdots \\ 1 \end{bmatrix}. \quad (3.44)$$

Durch Nutzung des Lösungsalgorithmus für NNLS-Probleme (Schritte 1-7) aus Abschnitt 3.3.3 erhält man den NNLS-Lösungsvektor u . Dieser wird mit:

$$\begin{aligned} r &= \tilde{E}u - \tilde{f}, \\ z_j &= -r_j/r_{n+1} \quad \text{mit } j = 1, \dots, n \end{aligned} \quad (3.45)$$

in die LDP-Lösung z umgerechnet, wobei n die Anzahl der Unbekannten des LDP-Problems darstellt. Abschließend erhält man mit:

$$\begin{aligned}\Delta x &= K\bar{R}^{-1}(z + H_1^T f), \\ x &= x_0 + \Delta x\end{aligned}$$

die Lösung x des LSI-Problems. Diese Lösung kann nun als neue Näherungslösung x_0 des LSI-Problems betrachtet und eine weitere Iteration gerechnet werden.

Für den Fall, dass die Annahme, dass alle Messungen y_i gleich genau sind, nicht gerechtfertigt ist, ist die Einführung einer Gewichtsmatrix P problemlos möglich. Zunächst erfolgt die Bestimmung einer Matrix S mittels z.B. einer Choleskyzerlegung derart, dass gilt (siehe [Petrović 2003]):

$$P = S^T S.$$

Dazu ist es notwendig, dass P positiv definit ist, was im Allgemeinen keine Einschränkung darstellt. Mittels dieser Matrix S kann nun das gewichtete Ausgleichungsproblem in ein ungewichtetes transformiert werden, so dass man für die Matrix E und den Vektor f des LSI-Problems erhält:

$$E_{neu} = SE; \quad f_{neu} = Sf; \quad v_{neu} = Sv.$$

Eine Transformation der Bedingungsungleichungen des LSI-Problems ist nicht notwendig, da Δx der transformierten Beobachtungsgleichungen identisch ist zum Δx der originalen Beobachtungsgleichungen. Statt E und f verwendet man somit im gewichteten Fall E_{neu} und f_{neu} , ohne dass sich weitere Aspekte bei der Lösung des LSI-Problems verändern. Es ist lediglich zu beachten, dass in diesem Fall $v_{neu}^T v_{neu}$ minimiert wird und sich die Toleranzen t bzw. t^2 auf die tatsächlichen v -Werte beziehen und nicht auf die durch die Transformation entstandenen v_{neu} -Werte.

4. Vergleich der unterschiedlichen Bedingungsungleichungen anhand synthetischer Beispiele

Im vorherigen Kapitel wurden zwei unterschiedliche mögliche mathematische Realisierungen der Toleranzenbedingung für Residuen erarbeitet sowie der Algorithmus zur Lösung von um Toleranzenbedingungen für die Residuen erweiterten Ausgleichungsproblemen. Innerhalb dieses Kapitels soll einerseits anhand von zwei linearen (Beispiele 1 und 2) und drei nichtlinearen Beispielen (Beispiele 3-5) gezeigt werden, welche Eigenschaften die beiden unterschiedlichen Varianten der Bedingungsungleichungen für die Residuen aufweisen, wie stabil beide Varianten sind und welche Variante als besser betrachtet werden kann. Andererseits sollen die Korrektheit und die Stabilität der Lösungsbestimmung bei Verwendung des erarbeiteten Auswertalgorithmus untersucht werden. Dazu werden zunächst im Abschnitt 4.1 einfache vermittelnde Ausgleichungsprobleme ohne Pseudobeobachtungen betrachtet. Im Abschnitt 4.2 werden dann diese Beispiele um Pseudobeobachtungen erweitert, so dass dann regularisierte Ausgleichungsprobleme vorliegen. Für sämtliche Beobachtungen und Pseudobeobachtungen werden in den Abschnitten 4.1 und 4.2 einheitliche Genauigkeiten innerhalb des jeweiligen Beispiels angenommen, so dass eine Berücksichtigung von Gewichtsmatrizen nicht notwendig ist. Im Abschnitt 4.3 wird exemplarisch anhand des Beispiels 5 gezeigt, wie vorteilhaft die Berücksichtigung der Variante 2 ($-v \leq t \ \& \ v \leq t$) der Bedingungsungleichungen für regularisierte Probleme sein kann, wenn für die eigentlichen Beobachtungen realistische Fehlerschätzungen und für die Pseudobeobachtungen zu optimistische Fehlerschätzungen angenommen werden.

4.1 Beispiele: nicht regularisiert und gleichgewichtet

Wie bereits festgelegt, handelt es sich bei den betrachteten Beispielen 1 und 2 um lineare und bei den Beispielen 3-5 um nichtlineare Ausgleichungsprobleme. Die verwendeten Funktionalmodelle und Daten ($y_1, y_2, \dots, y_n$ mit n als Anzahl der Beobachtungen) sind für sämtliche Beispiele im Anhang A.1 dargestellt. An dieser Stelle erfolgt lediglich die Zusammenstellung und Diskussion der Ergebnisse.

In den Tabellen 4.2 bis 4.6 sind die Ergebnisse (Unbekannte a , b und Summe der Residuenquadrate Svv) aufgelistet, welche mit den einzelnen Beispielen (1, 2, ..., 5) erzielt wurden, wobei unterschiedliche Toleranzen (gekennzeichnet mit a), b), ...) für die Residuen angesetzt wurden. Zum Vergleich sind in der Tabelle 4.1 für alle Beispiele die Ergebnisse zusammengestellt worden, die mittels einer Ausgleichung nach kleinsten Quadraten ohne Berücksichtigung von Toleranzenbedingungen für die Residuen erhalten wurden. Dabei stellt der Wert für Svv den kleinsten erreichbaren Wert dar. Die Differenzen zwischen diesen Werten für Svv und denen, die bei einer Ausgleichung mit Toleranzenbedingungen für die Residuen erhalten werden, können sehr klein aber auch sehr groß sein. Anhand dieser Differenzen kann man abschätzen, welchen mittleren „Genauigkeitsverlust“ man durch die Einführung von Toleranzenbedingungen akzeptieren muss. Grundsätzlich wurden einerseits bei allen Beispielen (außer Beispiel 4) alle Residuentoleranzen auf einen sinnvollen, meist einheitlichen, Wert gesetzt und andererseits wurde für jedes einzelne Residuum jeweils eine sehr kleine Toleranz vorgegeben, wobei die restlichen Toleranzen dann sehr großzügig gewählt wurden. Dieser zweite Fall wurde speziell gewählt, da somit die Grenzen des Verfahrens besonders gut repräsentiert werden können. Aufgelistet sind die Ergebnisse für beide verwendeten Varianten der mathematischen Realisierung der Toleranzenbedingung ($v^2 \leq t^2$ und

Tabelle 4.1: Ergebnisse der Ausgleichung nach kleinsten Quadraten für alle Beispiele

Bsp.nr.	Svv	Residuen $v_1/\dots/v_n$	Lösung
1	1.236264	$-0.4945/ 0.9066/ -0.4121$	$a = 0.560440$ $b = 1.164840$
2	0.040000	$-0.1333/ -0.1333/ 0.0667$	$a = 0.966667$ $b = 2.000000$
3	0.002000	$0.0100/ -0.0300/ 0.0300/$ -0.0100	$a = 1.456022$ $b = 0.531337$
4	0.002221	$-0.0014/ -0.0253/ 0.0370/$ -0.0144	$a = 0.093564$ $b = 1.772420$
5	$4.318722 \cdot 10^{-6}$	$0.0008/ 0.0013/ -0.0015/$ $-2.5 \cdot 10^{-5}/ 1.0 \cdot 10^{-5}$	$a = 30.002756$ $b = 49.998360$

Tabelle 4.2: Ergebnisse für Beispiel 1 mit unterschiedlichen Toleranzen für die Residuen

Bsp.nr.	Variante 1: $v^2 \leq t^2$	Variante 2: $-v \leq t \ \& \ v \leq t$	Heuristik	Toleranzen für $v_1/v_2/v_3$
1a)	$a = 0.560000$ $b = 0.960000$ $Svv = 1.364400$	$a = 0.560000$ $b = 0.960000$ $Svv = 1.364400$	$a = 0.559979$ $b = 0.960080$ $Svv = 1.364406$	$0.7/0.7/0.7$
1b)	$a = 0.453151$ $b = 1.815274$ $Svv = 1.534963$	$a = 0.453151$ $b = 1.815274$ $Svv = 1.534963$	$a = 0.453229$ $b = 1.815158$ $Svv = 1.534964$	$0.005/3.0/3.0$
1c)	$a = 0.563934$ $b = 1.044262$ $Svv = 1.270164$	$a = 0.563934$ $b = 1.044262$ $Svv = 1.270164$	$a = 0.563946$ $b = 1.044215$ $Svv = 1.270164$	$3.0/0.8/3.0$
1d)	$a = 0.584713$ $b = 1.107006$ $Svv = 1.250828$	$a = 0.584713$ $b = 1.107006$ $Svv = 1.250828$	$a = 0.584724$ $b = 1.106934$ $Svv = 1.250828$	$3.0/3.0/0.3$

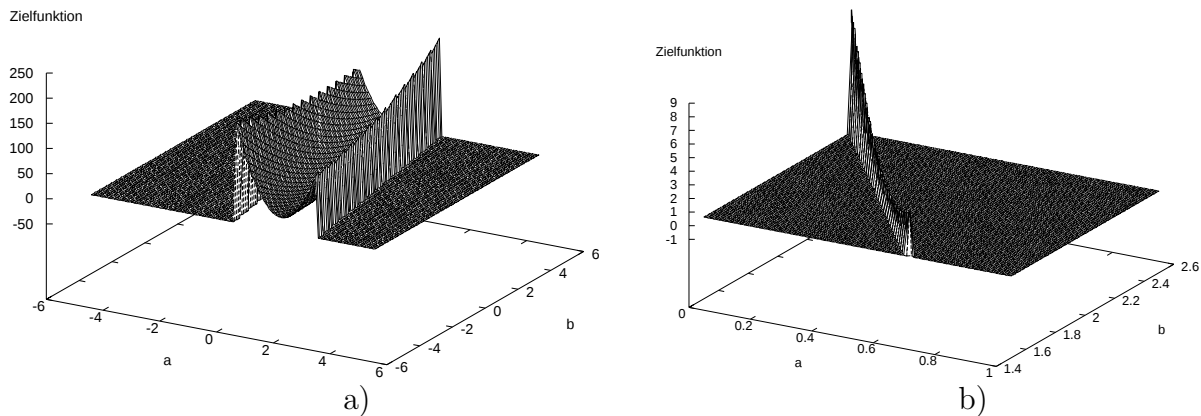
$-v \leq t \ \& \ v \leq t$) sowie die heuristisch ermittelten Ergebnisse¹. Die Definition der Toleranzbedingungen für die Residuen erfolgte für die Variante 1 wie sie im Abschnitt 3.2.1 beschrieben wurde und für die Variante 2 nach Abschnitt 3.2.2. Die Berechnung der Lösung wurde, wie im Abschnitt 3.3 beschrieben, durchgeführt. Die Bestimmung der heuristischen Lösung erfolgte nach

¹Wie im Abschnitt 3.2 bereits erläutert wurde, lassen sich mittels der Heuristik unabhängige Kontrollergebnisse berechnen, womit die Ergebnisse der Varianten 1 und 2 der Realisierung der Toleranzbedingung überprüft werden können. Da dieses Verfahren sehr rechenzeitintensiv arbeitet, wurde nicht die höchste Präzision bei der Bestimmung der Unbekannten gefordert, d.h. es sind nur etwa drei Nachkommastellen signifikant.

Tabelle 4.3: Ergebnisse für Beispiel 2 mit unterschiedlichen Toleranzen für die Residuen

Bsp.nr.	Variante 1: $v^2 \leq t^2$	Variante 2: $-v \leq t \ \& \ v \leq t$	Heuristik	Toleranzen für $v_1/v_2/v_3$
2a)	$a = 0.980000$ $b = 2.000000$ $S_{vv} = 0.043200$	$a = 0.980000$ $b = 2.000000$ $S_{vv} = 0.043200$	$a = 0.979994$ $b = 1.999996$ $S_{vv} = 0.043200$	0.12/0.12/0.12
2b)	$a = 0.975000$ $b = 2.075000$ $S_{vv} = 0.052500$	$a = 0.975000$ $b = 2.075000$ $S_{vv} = 0.052500$	$a = 0.975009$ $b = 2.074991$ $S_{vv} = 0.052500$	0.05/3.0/3.0
2c)	$a = 0.975000$ $b = 1.925000$ $S_{vv} = 0.052500$	$a = 0.975000$ $b = 1.925000$ $S_{vv} = 0.052500$	$a = 0.974983$ $b = 1.924983$ $S_{vv} = 0.052500$	3.0/0.05/3.0
2d)	$a = 0.950250$ $b = 2.000000$ $S_{vv} = 0.044850$	$a = 0.950250$ $b = 2.000000$ $S_{vv} = 0.044851$	$a = 0.950250$ $b = 1.999994$ $S_{vv} = 0.044851$	0.5/0.5/0.001

Abschnitt 3.2.

**Abbildung 4.1:** Grafische Darstellung der zu minimierenden Funktion: a) für das Beispiel 1 mit „riesigen“ Toleranzen; b) für das Beispiel 1b).

Bei den linearen Beispielen 1 und 2 (Tabellen 4.2 und 4.3) wurden bei allen Berechnungen jeweils mit beiden Varianten der Realisierung der Toleranzenbedingung die gleichen Lösungen erhalten und der Vergleich mit den heuristisch ermittelten Ergebnissen bestätigt, dass die Minimierung der Zielfunktion² zum globalen Minimum führte. In den Abbildungen 4.1 a) und 4.2 a) ist die zu minimierende Funktion für diese beiden Beispiele grafisch dargestellt, wobei sämtliche einzelnen Residuen ≤ 10.0 sind, d.h. die Toleranzen für die Residuen wurden sehr

²Mit Zielfunktion ist die Summe der Residuenquadrate gemeint.

Tabelle 4.4: Ergebnisse für Beispiel 3 mit unterschiedlichen Toleranzen für die Residuen

Bsp.nr.	Variante 1: $v^2 \leq t^2$	Variante 2: $-v \leq t \ \& \ v \leq t$	Heuristik	Toleranzen für $v_1/\dots/v_4$
3a)	$a = 1.449138$ $b = 0.559344$ $Svv = 0.002500$	$a = 1.449138$ $b = 0.559344$ $Svv = 0.002500$	$a = 1.449150$ $b = 0.559296$ $Svv = 0.002500$	0.025/0.025/ 0.025/0.025
3b)	$a = 1.457493$ $b = 0.522758$ $Svv = 0.002036$	$a = 1.457493$ $b = 0.522758$ $Svv = 0.002036$	$a = 1.457502$ $b = 0.522741$ $Svv = 0.002036$	0.005/10.0/ 10.0/10.0
3c)	$a = 1.453731$ $b = 0.550321$ $Svv = 0.002333$	$a = 1.453731$ $b = 0.550321$ $Svv = 0.002333$	$a = 1.453792$ $b = 0.550126$ $Svv = 0.002333$	10.0/0.02/ 10.0/10.0
3d)	$a = 1.453731$ $b = 0.531329$ $Svv = 0.002333$	$a = 1.453731$ $b = 0.531329$ $Svv = 0.002333$	$a = 1.453771$ $b = 0.531123$ $Svv = 0.002333$	10.0/10.0/ 0.02/10.0
3e)	$a = 1.457493$ $b = 0.527078$ $Svv = 0.002036$	$a = 1.457493$ $b = 0.527078$ $Svv = 0.002036$	$a = 1.457505$ $b = 0.526991$ $Svv = 0.002036$	10.0/10.0/ 10.0/0.005
3f)	keine stabile Lösung	$a = 1.458669$ $b = 0.523628$ $Svv = 0.002115$	$a = 1.458691$ $b = 0.523468$ $Svv = 0.002116$	0.0075/0.029/ 0.036/0.001

groß gewählt. Somit erhält man eine grafische Darstellung der zu minimierenden Funktion in der Umgebung des Minimums, ohne dass Toleranzenbedingungen wirken. Man erkennt, dass der Verlauf dieser Funktionen sehr gleichmäßig ist ohne Täler oder Sattelpunkte. In den Abbildungen 4.1 b) und 4.2 b) ist die zu minimierende Funktion für die Beispiele 1b) und 2b)

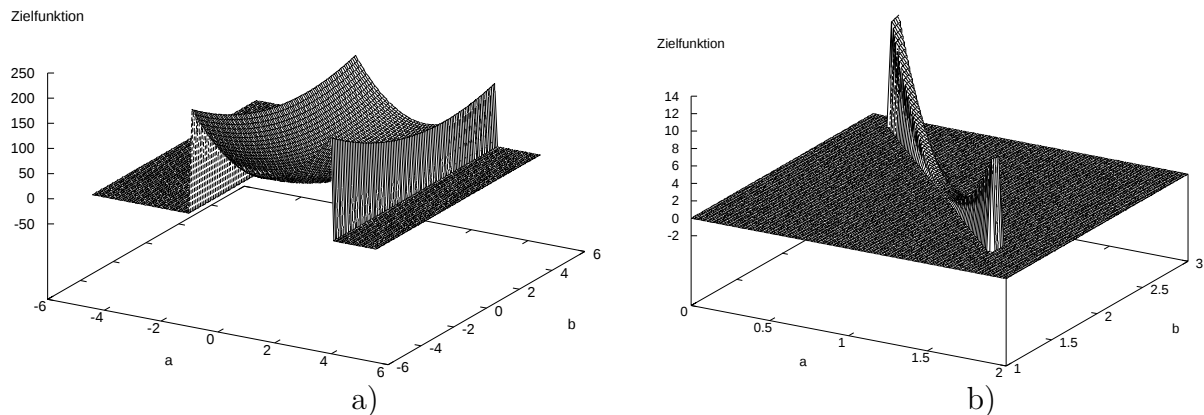
**Abbildung 4.2:** Grafische Darstellung der zu minimierenden Funktion: a) für das Beispiel 2 mit „riesigen“ Toleranzen; b) für das Beispiel 2b).

Tabelle 4.5: Ergebnisse für Beispiel 4 mit unterschiedlichen Toleranzen für die Residuen

Bsp.nr.	Variante 1: $v^2 \leq t^2$	Variante 2: $-v \leq t \ \& \ v \leq t$	Heuristik	Toleranzen für $v_1/\dots/v_4$
4a)	$a = 0.094287$ $b = 1.771866$ $Svv = 0.002221$	$a = 0.094980$ $b = 1.771530$ $Svv = 0.002222$	$a = 0.094236$ $b = 1.772005$ $Svv = 0.002221$	0.001/10.0/ 10.0/10.0
4b)	$a = 0.094109$ $b = 1.772034$ $Svv = 0.002221$	$a = 0.094032$ $b = 1.772126$ $Svv = 0.002221$	$a = 0.093989$ $b = 1.772177$ $Svv = 0.002221$	10.0/0.025/ 10.0/10.0
4c)	$a = 0.262095$ $b = 1.630295$ $Svv = 0.050720$	$a = 0.085748$ $b = 1.775122$ $Svv = 0.002372$	$a = 0.085772$ $b = 1.775104$ $Svv = 0.002373$	10.0/10.0/ 0.03/10.0
4d)	keine Lösung bestimmbar	$a = -0.595926$ $b = 2.036681$ $Svv = 0.018502$	$a = 0.091344$ $b = 1.774742$ $Svv = 0.002244$	10.0/10.0/ 10.0/0.01

grafisch dargestellt, wo die Toleranzenbedingungen wesentlich stärker wirken. Es sind somit nur die Bereiche der zu minimierenden Funktion aus den Abbildungen 4.1 a) und 4.2 a) dargestellt, für die die strengerer Toleranzenforderungen der Residuen erfüllt sind. Die resultierende Fläche ist zwar nur ein schmaler Streifen, in dem das Minimum gefunden werden muss, aber trotzdem ist es dort problemlos möglich gewesen, die richtige Lösung zu bestimmen.

Bei den Beispielen 3-5 traten dagegen vereinzelt Probleme auf (Tabellen 4.4, 4.5 und 4.6). So war die Bestimmung der Lösungen im Beispiel 3 mit beiden Varianten der Realisierung der Toleranzenbedingung gut möglich, bis auf das Beispiel 3f). Für dieses Beispiel konnte bei Verwendung der Variante 1 der Realisierung der Toleranzenbedingung keine stabile Lösung bestimmt werden, d.h. es gab im Laufe der Iteration vereinzelt Lösungen, für die die geforderten Toleranzen erfüllt worden wären, aber diese Lösungen stellten nicht die Lösung im Sinne des globalen Minimums dar. Mittels der Heuristik konnte wieder die Korrektheit der ermittelten Ergebnisse überprüft werden.

Für das Beispiel 5 war eine korrekte und einheitliche Bestimmung der Lösung mit beiden Varianten der Realisierung der Toleranzenbedingung in den Fällen 5a) bis 5e) möglich. Für den Fall 5f) konnte mittels der Variante 1 keine Lösung bestimmt werden - die geforderten Toleranzen wurden im Laufe der Iteration nie erfüllt. Weitere Probleme ergaben sich für die Fälle 5e) und 5f) im sog. Grenzbereich mit beiden Varianten der Realisierung der Toleranzenbedingung³. Damit ist Folgendes gemeint: Wurden die Berechnungen mit den angesetzten Toleranzen erfolgreich abgeschlossen, so wurden in einer weiteren Berechnung die Toleranzen, die vorher großzügig gewählt wurden, auf einen kleineren Wert gesetzt, der nur ein wenig ($\text{ca. } 1 \cdot 10^{-5}$) größer war, als das dazugehörige im Vorfeld berechnete Residuum⁴. Diese neuen Toleranzen

³Für die Beispiele 3a) bis 3f) und die Beispiele 5a) bis 5d) ergaben sich im „Grenzbereich“ keinerlei Probleme.

⁴Derartige „Grenzfälle“ wurden ebenso für alle betrachteten Fälle von Toleranzen für die Residuen der Beispiele

Tabelle 4.6: Ergebnisse für Beispiel 5 mit unterschiedlichen Toleranzen für die Residuen

Bsp.nr.	Variante 1: $v^2 \leq t^2$	Variante 2: $-v \leq t \ \& \ v \leq t$	Heuristik	Toleranzen für $v_1/\dots/v_5$
5a)	$a = 30.002506$ $b = 49.998510$ $S_{vv} = 4.403 \cdot 10^{-6}$	$a = 30.002506$ $b = 49.998510$ $S_{vv} = 4.403 \cdot 10^{-6}$	$a = 30.002493$ $b = 49.998493$ $S_{vv} = 4.406 \cdot 10^{-6}$	0.002/0.002/ 0.002/ $2 \cdot 10^{-5}$ / $2 \cdot 10^{-5}$
5b)	$a = 30.002600$ $b = 49.998400$ $S_{vv} = 4.347 \cdot 10^{-6}$	$a = 30.002600$ $b = 49.998400$ $S_{vv} = 4.347 \cdot 10^{-6}$	$a = 30.002600$ $b = 49.998402$ $S_{vv} = 4.347 \cdot 10^{-6}$	$6 \cdot 10^{-4}$ /10.0/ 10.0/10.0/ 10.0
5c)	$a = 30.003020$ $b = 49.997600$ $S_{vv} = 5.233 \cdot 10^{-6}$	$a = 30.003020$ $b = 49.997600$ $S_{vv} = 5.233 \cdot 10^{-6}$	$a = 30.003028$ $b = 49.997599$ $S_{vv} = 5.236 \cdot 10^{-6}$	10.0/ $5 \cdot 10^{-4}$ / 10.0/10.0/ 10.0
5d)	$a = 30.003254$ $b = 49.999191$ $S_{vv} = 6.198 \cdot 10^{-6}$	$a = 30.003254$ $b = 49.999191$ $S_{vv} = 6.198 \cdot 10^{-6}$	$a = 30.003261$ $b = 49.999188$ $S_{vv} = 6.200 \cdot 10^{-6}$	10.0/10.0/ $5 \cdot 10^{-4}$ / 10.0/10.0
5e)	$a = 30.002006$ $b = 49.998810$ $S_{vv} = 5.083 \cdot 10^{-6}$	$a = 30.002006$ $b = 49.998810$ $S_{vv} = 5.083 \cdot 10^{-6}$	$a = 30.002005$ $b = 49.998808$ $S_{vv} = 5.083 \cdot 10^{-6}$	10.0/10.0/ 10.0/ $1 \cdot 10^{-5}$ / 10.0
5f)	keine Lösung bestimmbar	$a = 30.002506$ $b = 49.998510$ $S_{vv} = 4.403 \cdot 10^{-6}$	$a = 30.002503$ $b = 49.998506$ $S_{vv} = 4.404 \cdot 10^{-6}$	10.0/10.0/ 10.0/10.0/ $5 \cdot 10^{-6}$

sollten somit erfüllbar sein. In den Fällen 5e) und 5f) konnte aber in diesen „Grenzfällen“ keine Lösung bestimmt werden, auch nicht, wenn statt der doppelten Rechengenauigkeit bei Fortran 90 die vierfache Rechengenauigkeit verwendet wurde. Um zu veranschaulichen, woher die Probleme kommen, ist in den Abbildungen 4.3 a) und 4.4 a) die zu minimierende Funktion für die Beispiele 3 und 5 grafisch dargestellt, wobei sämtliche Residuen wieder ≤ 10.0 sind. In der Abbildung 4.3 a) ist sehr deutlich zu erkennen, dass die Fläche der zu minimierenden Funktion sehr ungleichmäßig gekrümmt ist und einen „Sattel“ aufweist. Dies bringt selbst bei standardmäßig verwendeten Minimierungsverfahren Probleme mit sich, da die Minimierung in Richtung des negativen Gradienten der zu minimierenden Funktion erfolgt und man somit bei ungünstiger Lage des Näherungspunktes zum falschen Minimum gelangen kann. Die Fläche der zu minimierenden Funktion im Beispiel 5 (Abbildung 4.4 a)) weist analog zu den Beispielen 1 und 2 günstige Krümmungseigenschaften auf. In den Abbildungen 4.3 b) und 4.4 b) ist die verbleibende Fläche, in der das Minimum gefunden werden muss, für das Beispiel 3f) und den „Grenzfall“ 5e) abgebildet. Bei Betrachtung dieser Abbildungen verwundert es wenig, dass in diesen Fällen die Bestimmung einer Lösung nicht möglich oder problematisch war, da lediglich eine sehr kleine Fläche übrig bleibt, für die die Toleranzbedingungen der Residuen erfüllt sind. Für den „Grenzfall“ 5f) erhält man eine sehr ähnliche grafische Darstellung der zu

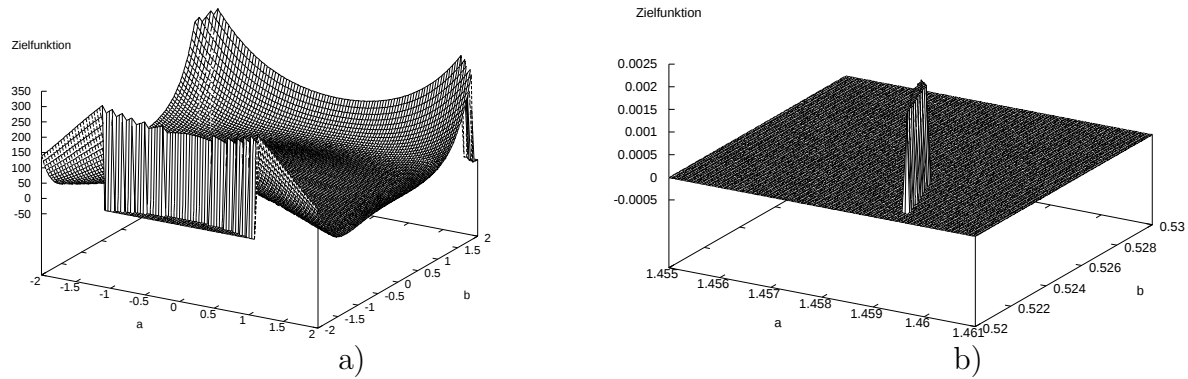


Abbildung 4.3: Grafische Darstellung der zu minimierenden Funktion: a) für das Beispiel 3 mit „riesigen“ Toleranzen; b) für das Beispiel 3f).

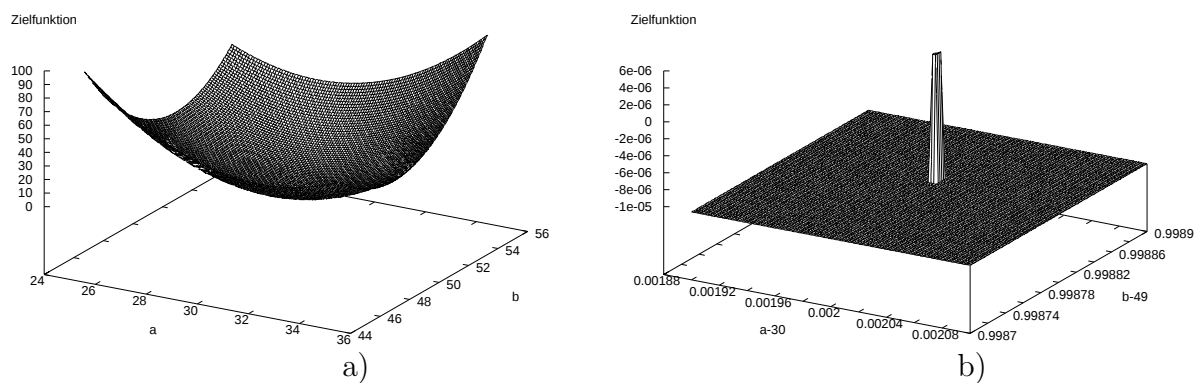


Abbildung 4.4: Grafische Darstellung der zu minimierenden Funktion: a) für das Beispiel 5 mit „riesigen“ Toleranzen; b) für den „Grenzfall“ 5e).

minimierenden Funktion wie für den „Grenzfall“ 5e), so dass es auch hier verständlich ist, dass keine Lösung bestimmt werden konnte. Da die Bestimmung einer Lösung nur in diesen extremen Fällen versagt, kann grundsätzlich nicht davon ausgegangen werden, dass der verwendete Lösungsalgorithmus instabil ist.

Als letztes bleibt nun noch das Beispiel 4 (Tabelle 4.5) zu beurteilen, welches sich als kompliziertestes erwiesen hat. In den Fällen 4a) und 4b) lieferten beide Varianten der Realisierung der Toleranzenbedingung die gleiche (bis auf Unterschiede durch die jeweilige Rechengenauigkeit) korrekte Lösung, jedoch versagen beide Varianten in beiden Fällen in den besagten „Grenzfällen“. Im Fall 4c) liefert die Variante 1 eine falsche Lösung, d.h. die geforderten Toleranzen wurden zwar erfüllt, aber das globale Minimum der Zielfunktion nicht gefunden. Betrachtet man die Ergebnisse für den Fall 4d), so erkennt man, dass die Variante 2 eine falsche Lösung liefert und die Variante 1 sogar völlig versagt. In den „Grenzfällen“ von 4c) und 4d) konnten ebenso keine Lösungen bestimmt werden. Es stellt sich nun die Frage, warum derartige Probleme nur bei diesem Beispiel so gehäuft auftreten? Dazu ist es wieder notwendig, die grafische Darstellung der Zielfunktion mit unterschiedlichen Toleranzenbedingungen für die Residuen zu betrachten. In der Abbildung 4.5 ist die zu minimierende Funktion für das Beispiel 4 um die Näherungslösung (Startpunkt) für die erste Iteration grafisch dargestellt ($a_0 = 0.1$ und $b_0 = 1.7$). Man erkennt, dass der Startpunkt nahezu in einer Ebene liegt. Die Abbildung 4.6 a) zeigt die grafische Darstellung der Zielfunktion für den „Grenzfall“

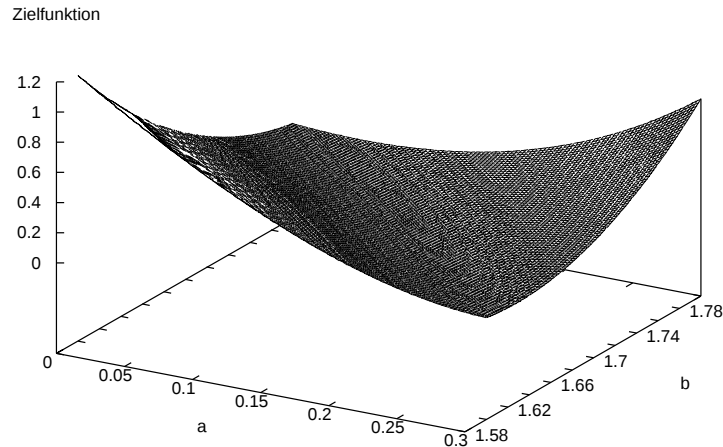


Abbildung 4.5: Grafische Darstellung der zu minimierenden Funktion für das Beispiel 4.

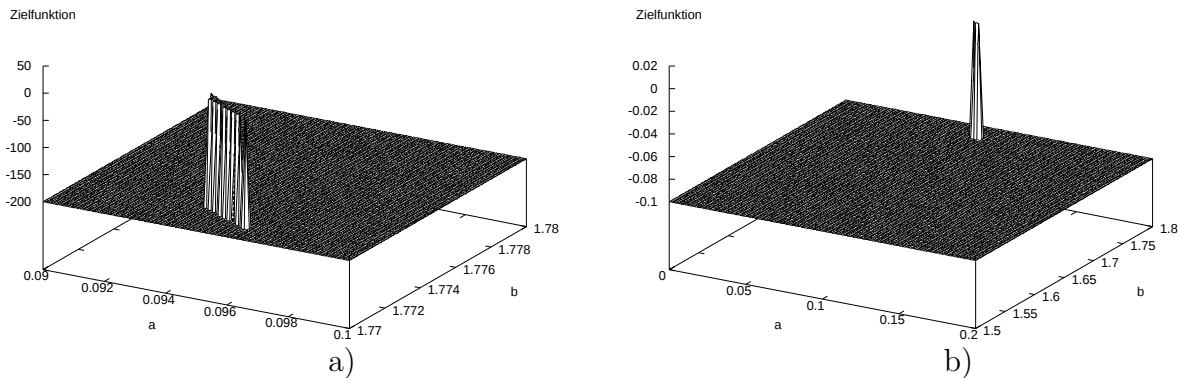


Abbildung 4.6: Grafische Darstellung der zu minimierenden Funktion: a) für den „Grenzfall“ 4a); b) für den „Grenzfall“ 4b).

4a). Die abgebildete Fläche ist sehr klein und liegt vom Startpunkt (a_0, b_0) aus betrachtet bei einem höheren Wert der Zielfunktion (vergleiche mit Abbildung 4.5). Damit ist es nicht möglich, eine Lösung für diesen Fall zu bestimmen, da die Minimierung der Zielfunktion in Richtung des negativen Gradienten der Zielfunktion erfolgt und dieser zeigt immer in die Richtung des maximalen Abstiegs. Die Abbildung 4.6 b) zeigt für den „Grenzfall“ 4b) die grafische Darstellung der Zielfunktion. Hier trifft das gleiche zu wie im „Grenzfall“ 4a) - der potenzielle Lösungsbereich ist sehr klein und liegt bei einem höheren Wert der Zielfunktion als der Startpunkt. In den Fällen 4c) und 4d), wo mit der Variante 1 bzw. mit beiden Varianten keine Lösung bestimmt werden konnte, weist die Zielfunktion unter Berücksichtigung der jeweiligen Toleranzen Lücken auf (siehe Abbildung 4.7 a) und 4.7 b)). Es ist somit auch hier nicht verwunderlich, dass bei derartigen Verläufen der Zielfunktion keine Lösung bestimmbar ist.

Erstellt man für alle betrachteten Beispiele und unterschiedlichen Toleranzenforderungen für die Residuen grafische Darstellungen der jeweiligen Zielfunktion, und analysiert diese sehr ausgiebig, so kann man Eigenschaften der Zielfunktionsflächen ermitteln, die positiv, negativ oder neutral auf die Lösungsbestimmbarkeit wirken. Folgende Eigenschaften konnten ermittelt werden (auf eine Zusammenstellung aller erstellter Grafiken wird verzichtet, da es der Übersichtlichkeit schaden würde).

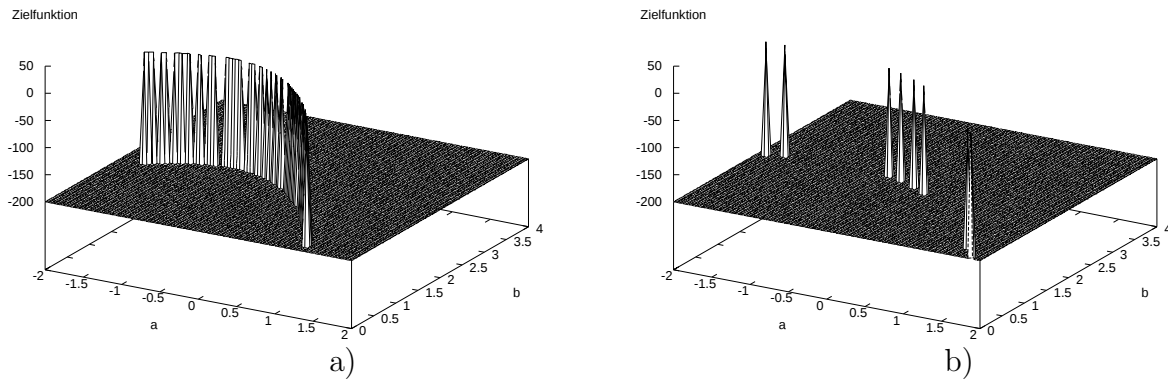


Abbildung 4.7: Grafische Darstellung der zu minimierenden Funktion: a) für das Beispiel 4c); b) für das Beispiel 4d).

Eigenschaften der Zielfunktionsfläche, wenn die Lösung bestimmt werden konnte:

- die gesamte Fläche hat einen gleichmäßigen Verlauf (Beispiele 1, 2, 5),
- der Startpunkt hat eine günstige Position innerhalb der gesamten Fläche (Beispiele 1, 2, 5),
- der Startpunkt liegt dicht an der potenziellen Zielfunktionsfläche bzw. dem -streifen, d.h. an dem Bereich der Zielfunktionsfläche, der nach Einbeziehung der Toleranzen übrig bleibt (Beispiele 1, 2, 4),
- die Zielfunktionsfläche fällt zum Lösungspunkt hin ab (Beispiele 1, 2, 3, 5),
- die Zielfunktionsfläche bzw. der -streifen besitzt gute Krümmungseigenschaften (Beispiele 1, 2, 3, 4, 5),
- die Zielfunktionsfläche bzw. der -streifen enthält keine Lücken (Beispiele 1, 2).

Eigenschaften der Zielfunktionsfläche, wenn die Lösung nicht bestimmt werden konnte:

- die Zielfunktionsfläche enthält einen relativ ebenen Bereich (Beispiel 4),
- geringe Krümmung der Zielfunktionsfläche bzw. des -streifens (Beispiele 3, 4, 5),
- ungünstige Lage des Startpunktes (Beispiele 3, 4),
- sehr kleine Zielfunktionsfläche bzw. -streifen (Peaks, Beispiele 4, 5).

Eigenschaften der Zielfunktionsfläche die in beiden Fällen auftreten (Lösung bestimmbar oder nicht bestimmbar):

- kleine bzw. kurze Zielfunktionsfläche bzw. -streifen (Beispiele 1, 3, 4),
- schmaler Streifen der Zielfunktion (Beispiele 1, 2, 3, 4, 5),
- Startpunkt weit entfernt vom Lösungspunkt (Beispiele 1, 2, 3, 4, 5),
- Lücken in der Zielfunktionsfläche bzw. in dem -streifen (Beispiele 1, 4, 5),

- Lösungspunkt am Rand der potenziellen Zielfunktionsfläche gelegen (Beispiele 1, 2, 4, 5).

Betrachtet man die ermittelten Eigenschaften der Zielfunktionsflächen und die aufgezeigten Ergebnisse, so kann man abschließend sagen, dass der Algorithmus zur Lösung von Ausgleichungsproblemen mit Residuentoleranzen sehr stabil ist, da dieser nur in sehr extremen Fällen versagt. Beim Vorliegen der ermittelten negativen Eigenschaften der Zielfunktionsfläche ist auch mit anderen Verfahren zur Lösung anderer Minimierungsprobleme eine effiziente Lösungsbestimmung nicht möglich. Dabei scheint es, dass die Hauptursache für die Probleme bei der Lösungsbestimmung in der ungünstigen Lage des Startpunktes zu sehen ist, d.h. die Näherungswerte wurden schlecht gewählt. Dies ist allerdings keine spezifische Schwäche des erarbeiteten Algorithmus, sondern ein allgemeines Problem bei der Verwendung von Iterationsverfahren.

Weiterhin ist eindeutig erkennbar, dass die Variante 2 der Realisierung der Toleranzenbedingung stabiler ist als die Variante 1, da mit dieser Variante fast immer eine Lösung ermittelt werden konnte. Das entspricht den Erwartungen, da bei dieser Variante die Linearisierung der Bedingungsungleichungen der Linearisierung der Beobachtungsgleichungen entspricht.

4.2 Beispiele: regularisiert und gleichgewichtet

Innerhalb dieses Abschnittes werden die Beispiele aus dem vorherigen Abschnitt nochmals ausgewertet, wobei aber sog. Pseudobeobachtungen für die Unbekannten mit in die Ausgleichung einbezogen werden. Man spricht dann von Regularisierung.

Tabelle 4.7: Ergebnisse der Ausgleichung nach kleinsten Quadraten (mit Regularisierung)

Bsp.nr.	Svv	Residuen $v_1/\dots/v_n$	Lösung
1	1.791211	-0.1649 / 1.0503 / -0.4914	$a = 0.486080$ $b = 1.606000$
2	3.718245	0.5877 / -0.7456 / 0.2842	$a = 1.021100$ $b = 2.666667$
3	5.033224	0.5490 / 0.2676 / 0.0863 / -0.1950	$a = 1.279581$ $b = 0.976209$
4	6.699587	0.5512 / 0.4913 / 0.3154 / -0.1766	$a = 0.628485$ $b = 1.367848$
5	4.095769	-1.0594 / 0.8979 / 0.2379 / 0.0234 / -0.0234	$a = 28.942556$ $b = 50.894969$

Bei allen betrachteten Beispielen stellen die Parameter a und b die Unbekannten dar, wobei diese allerdings in jedem Beispiel eine andere Bedeutung haben. Die Einführung von Pseudobeobachtungen erfolgt dann so, dass zu den im Anhang A.1 dargestellten Beobachtungsgleichungen noch folgende zwei Beobachtungsgleichungen hinzukommen:

$$\begin{aligned}\bar{a} + v_{\bar{a}} &= a, \\ \bar{b} + v_{\bar{b}} &= b.\end{aligned}$$

Dabei stellen $\bar{a}$ und $\bar{b}$ eine Art Vorinformation für die unbekannten Parameter a und b dar. Bei der Bestimmung der Lösung mit den Varianten 1 und 2 der Realisierung der Toleranzenbedingung wird wiederum die Summe der Residuenquadrate minimiert, einschließlich der Residuen

der Pseudobeobachtungen. Die Toleranzenbedingungen werden allerdings nur für die eigentlichen Beobachtungen eingeführt. Die Werte der Pseudobeobachtungen sind im Anhang A.1 aufgelistet.

Tabelle 4.8: Ergebnisse für Beispiel 1 mit Regularisierung bei unterschiedlichen Toleranzen für die Residuen

Bsp.nr.	Variante 1: $v^2 \leq t^2$	Variante 2: $-v \leq t \ \& \ v \leq t$	Heuristik	Toleranzen für $v_1/v_2/v_3$
1a)	$a = 0.533333$ $b = 1.066666$ $Svv = 2.470000$	$a = 0.533333$ $b = 1.066666$ $Svv = 2.470000$	$a = 0.533333$ $b = 1.066666$ $Svv = 2.470006$	0.7/0.7/0.7
1b)	$a = 0.459937$ $b = 1.805094$ $Svv = 1.866300$	$a = 0.459937$ $b = 1.805094$ $Svv = 1.866300$	$a = 0.459977$ $b = 1.805034$ $Svv = 1.866300$	0.005/3.0/3.0
1c)	$a = 0.426356$ $b = 1.294574$ $Svv = 2.887597$	$a = 0.426356$ $b = 1.294574$ $Svv = 2.887597$	$a = 0.426432$ $b = 1.294270$ $Svv = 2.887597$	3.0/0.5/3.0
1d)	$a = 0.554062$ $b = 1.521569$ $Svv = 1.991653$	$a = 0.554062$ $b = 1.521569$ $Svv = 1.991653$	$a = 0.554056$ $b = 1.521608$ $Svv = 1.991653$	3.0/3.0/0.1

Tabelle 4.9: Ergebnisse für Beispiel 2 mit Regularisierung bei unterschiedlichen Toleranzen für die Residuen

Bsp.nr.	Variante 1: $v^2 \leq t^2$	Variante 2: $-v \leq t \ \& \ v \leq t$	Heuristik	Toleranzen für $v_1/v_2/v_3$
2a)	$a = 0.980000$ $b = 2.000000$ $Svv = 5.083600$	$a = 0.980000$ $b = 2.000000$ $Svv = 5.083600$	$a = 0.979998$ $b = 2.000005$ $Svv = 5.083600$	0.12/0.12/0.12
2b)	$a = 0.968182$ $b = 2.331818$ $Svv = 4.107727$	$a = 0.968182$ $b = 2.331818$ $Svv = 4.107727$	$a = 0.968173$ $b = 2.331827$ $Svv = 4.107727$	0.2/3.0/3.0
2c)	$a = 1.054545$ $b = 2.454545$ $Svv = 3.874545$	$a = 1.054545$ $b = 2.454545$ $Svv = 3.874545$	$a = 1.054713$ $b = 2.454713$ $Svv = 3.874546$	3.0/0.5/3.0
2d)	$a = 0.951250$ $b = 2.666666$ $Svv = 3.810821$	$a = 0.951250$ $b = 2.666666$ $Svv = 3.810821$	$a = 0.951250$ $b = 2.666633$ $Svv = 3.810821$	3.0/3.0/0.005

Analog zum vorherigen Abschnitt werden die Ergebnisse der Varianten 1 und 2 der Realisierung der Toleranzenbedingung mit den Ergebnissen der Heuristik gegenübergestellt (Tabellen 4.8 bis 4.12). Zum Vergleich dient ebenso die Tabelle 4.7, in der für alle Beispiele die Ergebnisse der Ausgleichung nach kleinsten Quadraten ohne Berücksichtigung von Toleranzenbedingungen für die Residuen angegeben sind. Es werden darin lediglich die Residuen der Beobachtungen aufgeführt.

Tabelle 4.10: Ergebnisse für Beispiel 3 mit Regularisierung bei unterschiedlichen Toleranzen für die Residuen

Bsp.nr.	Variante 1: $v^2 \leq t^2$	Variante 2: $-v \leq t \ \& \ v \leq t$	Heuristik	Toleranzen für $v_1/\dots/v_4$
3a)	keine Lösung bestimmbar	$a = 1.449138$ $b = 0.559344$ $Svv = 6.262749$	$a = 1.449135$ $b = 0.559352$ $Svv = 6.262751$	0.025/0.025/ 0.025/0.025
3b)	$a = 1.360367$ $b = 0.780082$ $Svv = 5.430471$	$a = 1.360367$ $b = 0.780082$ $Svv = 5.430471$	$a = 1.360305$ $b = 0.780128$ $Svv = 5.430471$	0.2/3.0/ 3.0/3.0
3c)	$a = 1.285838$ $b = 0.907193$ $Svv = 5.164747$	$a = 1.285838$ $b = 0.907193$ $Svv = 5.164747$	$a = 1.285851$ $b = 0.907180$ $Svv = 5.164747$	3.0/0.1/ 3.0/3.0
3d)	$a = 1.270815$ $b = 0.975240$ $Svv = 5.037747$	$a = 1.270815$ $b = 0.975240$ $Svv = 5.037747$	$a = 1.270812$ $b = 0.975243$ $Svv = 5.037747$	3.0/3.0/ 0.05/3.0
3e)	$a = 1.317454$ $b = 0.957785$ $Svv = 5.069613$	$a = 1.317454$ $b = 0.957785$ $Svv = 5.069613$	$a = 1.317386$ $b = 0.957916$ $Svv = 5.069613$	3.0/3.0/ 3.0/0.05

Tabelle 4.11: Ergebnisse für Beispiel 4 mit Regularisierung bei unterschiedlichen Toleranzen für die Residuen

Bsp.nr.	Variante 1: $v^2 \leq t^2$	Variante 2: $-v \leq t \ \& \ v \leq t$	Heuristik	Toleranzen für $v_1/\dots/v_4$
4a)	keine Lösung bestimmbar	$a = 0.305438$ $b = 1.615953$ $Svv = 7.468865$	$a = 0.305430$ $b = 1.615989$ $Svv = 7.468865$	0.15/0.15/ 0.15/0.15
4b)	keine Lösung bestimmbar	$a = 0.355663$ $b = 1.582688$ $Svv = 7.265764$	$a = 0.355667$ $b = 1.582669$ $Svv = 7.265764$	0.2/3.0/ 3.0/3.0
4c)	keine stabile Lösung bestimmbar	$a = 0.502967$ $b = 1.444057$ $Svv = 6.862478$	$a = 0.502924$ $b = 1.444143$ $Svv = 6.862478$	3.0/0.3/ 3.0/3.0
4d)	keine Lösung bestimmbar	$a = 0.570851$ $b = 1.341268$ $Svv = 6.876815$	$a = 0.570758$ $b = 1.341377$ $Svv = 6.876815$	3.0/3.0/ 0.1/3.0
4e)	keine stabile Lösung bestimmbar	keine stabile Lösung bestimmbar	$a = 0.626316$ $b = 1.376274$ $Svv = 6.700568$	3.0/3.0/ 3.0/0.15

Tabelle 4.12: Ergebnisse für Beispiel 5 mit Regularisierung bei unterschiedlichen Toleranzen für die Residuen

Bsp.nr.	Variante 1: $v^2 \leq t^2$	Variante 2: $-v \leq t \ \& \ v \leq t$	Heuristik	Toleranzen für $v_1/\dots/v_5$
5a)	$a = 29.802000$ $b = 50.197100$ $Svv = 6.582091$	$a = 29.802000$ $b = 50.197100$ $Svv = 6.582091$	$a = 29.802000$ $b = 50.197094$ $Svv = 6.582110$	0.2/0.2/0.2/ 0.02/0.02
5b)	$a = 29.502000$ $b = 50.807750$ $Svv = 4.779497$	$a = 29.502000$ $b = 50.807750$ $Svv = 4.779497$	$a = 29.502000$ $b = 50.807753$ $Svv = 4.779497$	0.5/3.0/3.0/ 3.0/3.0
5c)	$a = 29.018664$ $b = 50.497100$ $Svv = 4.518916$	$a = 29.018664$ $b = 50.497100$ $Svv = 4.518916$	$a = 29.018891$ $b = 50.497100$ $Svv = 4.518916$	3.0/0.5/3.0/ 3.0/3.0
5d)	$a = 28.874406$ $b = 50.775127$ $Svv = 4.152788$	$a = 28.874406$ $b = 50.775127$ $Svv = 4.152788$	$a = 28.874472$ $b = 50.775090$ $Svv = 4.152788$	3.0/3.0/0.1/ 3.0/3.0
5e)	$a = 29.114491$ $b = 50.796549$ $Svv = 4.174448$	$a = 29.114491$ $b = 50.796549$ $Svv = 4.174448$	$a = 29.114465$ $b = 50.796503$ $Svv = 4.174448$	3.0/3.0/3.0/ 0.02/3.0
5f)	$a = 29.115252$ $b = 50.796111$ $Svv = 4.175147$	$a = 29.115252$ $b = 50.796111$ $Svv = 4.175147$	$a = 29.115323$ $b = 50.796234$ $Svv = 4.175147$	3.0/3.0/3.0/ 3.0/0.02

Grundsätzlich kann bei den einzelnen Beispielen das gleiche Verhalten beobachtet werden (anhand der tabellarisch aufgelisteten Ergebnisse sowie der grafischen Darstellungen der Zielfunktion, welche hier aufgrund der Übersichtlichkeit nicht aufgeführt werden) wie bei den nicht regularisierten Beispielen im vorherigen Abschnitt, sowohl bei den betrachteten Toleranzen⁵ für die Residuen als auch für die „Grenzfälle“. Die Ergebnisse zeigen allerdings, dass mit der Variante 1 recht häufig keine oder keine stabile Lösung ermittelt werden konnte. Somit ist auch nach diesen Vergleichen die Variante 2 der Realisierung der Toleranzenbedingung als die stabilere Variante zu betrachten.

4.3 Beispiele: regularisiert und unterschiedlich gewichtet

Abschließend werden alle betrachteten Beispiele mit Regularisierung und unter Verwendung einer Gewichtsmatrix für die Beobachtungen (einschließlich Pseudobeobachtungen) ausgewertet. Die Gewichtsmatrizen wurden derart aufgestellt, dass die einzelnen Beobachtungen leicht unterschiedliche Gewichte erhalten und miteinander korreliert sind, wenn es sich um Beobachtungen des gleichen Typs handelt (z.B. Strecken oder Winkel). Die Pseudobeobachtungen wurden immer wesentlich geringer gewichtet als die eigentlichen Beobachtungen. Da sich die Variante 2 der

⁵Es ist anzumerken, dass die innerhalb dieses Abschnitts gewählten Toleranzen für die Residuen bei den einzelnen Beispielen meist nicht den gewählten Toleranzen aus dem vorherigen Abschnitt entsprechen, da ein Vergleich der hier untersuchten Regularisierungsprobleme mit den Ausgleichungsproblemen des vorherigen Abschnitts nicht unmittelbar möglich ist.

Realisierung der Toleranzenbedingung bereits in den beiden vorherigen Abschnitten als stabilere Variante erwiesen hat, erfolgten sämtliche Berechnungen nur mit dieser Variante. Ein unabhängiger Vergleich wurde mit der heuristisch ermittelten Lösung durchgeführt. Es wurden wieder die gleichen Verhaltensweisen bei den einzelnen Beispielen wie in den vorherigen Abschnitten beobachtet, so dass hier auf eine tabellarische und grafische Zusammenstellung aller Ergebnisse verzichtet werden soll. Lediglich das Beispiel 5 soll an dieser Stelle näher betrachtet werden. Mit diesem kann sehr schön veranschaulicht werden, was im Falle zu optimistisch angenommener Genauigkeiten für die Pseudobeobachtungen passiert, wenn man eine „normale“ Regularisierung statt einer Regularisierung mit Toleranzenbedingungen für die Residuen durchführt. Bei dem Beispiel 5 handelt es sich um ein geometrisch anschauliches Beispiel - ein rechtwinkliges Dreieck, bei dem die drei Seiten und die beiden Winkel gemessen wurden. Gesucht ist die Länge der Katheten des Dreiecks unter Berücksichtigung aller Messungen. Die Genauigkeiten m_i der Beobachtungen und Pseudobeobachtungen wurden für dieses Beispiel wie folgt festgelegt, wobei die Genauigkeiten der eigentlichen Beobachtungen den simulierten Messgenauigkeiten entsprechen:

$$\begin{aligned} m_{a,b,c} &= \pm 0.003 \, m, \\ m_{\alpha,\beta} &= \pm 5 \cdot 10^{-5} \, rad, \\ m_{\bar{a},\bar{b}} &= \pm 0.045 \, m, \\ m_0 &= \pm 0.003 \, m. \end{aligned}$$

Wesentlich realistischere Genauigkeiten für die Pseudobeobachtungen wären $m_{\bar{a},\bar{b}} = \pm 2.0 \, m$. Verwendet wurde folgende gewählte Gewichtsmatrix:

$$P = \begin{pmatrix} 1.0 & 0.02 & 0.01 & 0.0 & 0.0 & 0.0 & 0.0 \\ 0.02 & 1.0 & 0.02 & 0.0 & 0.0 & 0.0 & 0.0 \\ 0.01 & 0.02 & 1.0 & 0.0 & 0.0 & 0.0 & 0.0 \\ 0.0 & 0.0 & 0.0 & 3600 \, m^2 & 0.01 & 0.0 & 0.0 \\ 0.0 & 0.0 & 0.0 & 0.01 & 3600 \, m^2 & 0.0 & 0.0 \\ 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.004 & 0.0 \\ 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.004 \end{pmatrix}. \quad (4.1)$$

Berechnet man nun die Unbekannten a und b mittels einer „normalen“ Regularisierung ohne Berücksichtigung von Toleranzenbedingungen für die Residuen und unter Verwendung der angegebenen Gewichtsmatrix (weitere Daten siehe Anhang A.1), so erhält man folgende Lösung und Residuen:

$$\begin{aligned} a &= 29.999541 \, m, & b &= 50.002015 \, m, \\ v &= \begin{pmatrix} v_a \\ v_b \\ v_c \\ v_\alpha \\ v_\beta \\ v_{\bar{a}} \\ v_{\bar{b}} \end{pmatrix} = \begin{pmatrix} -0.002459 \, m \\ 0.004915 \, m \\ 0.000011 \, m \\ 0.000055 \, rad \\ -0.000070 \, rad \\ 1.999541 \, m \\ -1.997985 \, m \end{pmatrix}. \end{aligned} \quad (4.2)$$

Obwohl innerhalb der Gewichtsmatrix P die Strecken a , b und c eine Genauigkeit von $\pm 0.003 \, m$ zugeordnet bekommen haben, ist das Residuum für b nach dieser Auswertung größer als die zur Beobachtung b dazugehörige Genauigkeit. Ebenso sind auch die Residuen für α und β größer als

die verwendete Genauigkeit innerhalb der Gewichtsmatrix.

Zum Vergleich dazu erhält man für das gleiche Beispiel ohne Regularisierung (also ohne Verwendung der Pseudobeobachtungen) unter Verwendung der gleichen Gewichte und Korrelationen für die Beobachtungen folgende Ergebnisse:

$$a = 30.002149 \text{ m}, \quad b = 49.998713 \text{ m},$$

$$v = \begin{pmatrix} v_a \\ v_b \\ v_c \\ v_\alpha \\ v_\beta \end{pmatrix} = \begin{pmatrix} 0.000149 \text{ m} \\ 0.001613 \text{ m} \\ -0.001479 \text{ m} \\ -0.000013 \text{ rad} \\ -0.000002 \text{ rad} \end{pmatrix}. \quad (4.3)$$

Daran erkennt man sehr gut, welchen großen Einfluss die eigentlich schlechten Pseudobeobachtungen auf das Ergebnis haben, denn ohne Einführung der Pseudobeobachtungen sind sämtliche Residuen kleiner als die für die Gewichtsmatrix angesetzten Genauigkeiten der einzelnen Beobachtungen⁶.

Verwendet man dagegen das ausgearbeitete Regularisierungsverfahren mit der Variante 2 der Realisierung der Toleranzenbedingung zur Bestimmung der Lösung, wobei die Toleranzen für die Strecken a , b und c 0.003 m betragen und für die Winkel α und β $5 \cdot 10^{-5} \text{ rad}$, so erhält man:

$$a = 29.999720 \text{ m}, \quad b = 50.000100 \text{ m},$$

$$v = \begin{pmatrix} v_a \\ v_b \\ v_c \\ v_\alpha \\ v_\beta \\ v_{\bar{a}} \\ v_{\bar{b}} \end{pmatrix} = \begin{pmatrix} -0.002280 \text{ m} \\ 0.003000 \text{ m} \\ -0.001539 \text{ m} \\ 0.000035 \text{ rad} \\ -0.000050 \text{ rad} \\ 1.999720 \text{ m} \\ -1.999900 \text{ m} \end{pmatrix}. \quad (4.4)$$

Die ermittelte Lösung ist nur geringfügig von der Lösung der „normalen“ Ausgleichung ohne Regularisierung verschieden ($\Delta a = 0.002 \text{ m}$, $\Delta b = 0.001 \text{ m}$). Die Abweichungen zur Lösung der „normalen“ Regularisierung sind größer ($\Delta a = 0.003 \text{ m}$, $\Delta b = 0.003 \text{ m}$). Allerdings liegen nun sämtliche Residuen der eigentlichen Beobachtungen innerhalb der Toleranzen und somit innerhalb der simulierten Messgenauigkeit. Somit konnte durch die Einbeziehung von Residuentoleranzen etwas erreicht werden, was ebenfalls durch eine Änderung der Verhältnisse zwischen den Gewichten der Beobachtungen und Pseudobeobachtungen theoretisch möglich sein sollte (wobei die Gewichtsverhältnis-Änderung nicht so leicht oder überhaupt nicht hätte abgeschätzt werden können). Die Einbeziehung von Residuentoleranzen ist somit mindestens gleichbedeutend mit der „bestmöglichen“ Wahl des Regularisierungsparameters bzw. Gewichtsverhältnisses. Die Eigenschaften dieser „bestmöglichen“ Lösung wurden bereits im Abschnitt 2.4 aufgezeigt. Erfolgt keine Einbeziehung von Residuentoleranzen, so erhält man je nach Wahl des Regularisierungsparameters bzw. Gewichtsverhältnisses zwischen den Beobachtungen und Pseudobeobachtungen eine mehr oder weniger willkürliche Lösung für das betrachtete Problem.

⁶An dieser Stelle muss angemerkt werden, dass dies bei „normalen“ Ausgleichungen ohne Regularisierung keinesfalls immer so ist. Selbst bei realistischer Wahl der Genauigkeiten der Beobachtungen für die Gewichtsmatrix können die resultierenden Residuen größer sein als die jeweils dazugehörigen Genauigkeiten der Beobachtungen.

Eine kontrollierte Beeinflussung der Residuen der Beobachtungen ist dabei nicht oder nur begrenzt möglich.

Welche Bedeutung hat eine Lösungsbestimmung unter Einbeziehung von Residuentoleranzen in der Praxis? In der Praxis hat man oft den Fall, dass die gesuchten Parameter bereits aus vorherigen Datenauswertungen mit einer gewissen Genauigkeit bekannt sind. Diese Genauigkeitsschätzungen sind aber meist zu optimistisch. Weiterhin hat man neue Beobachtungen, für die eine hohe Genauigkeit angenommen wird, mit denen der bereits vorhandene Parametersatz verbessert werden soll. So wird dann häufig eine Regularisierung durchgeführt (ohne Toleranzenbedingungen für die Residuen einzubeziehen), wobei die vorhandenen Parameter als Vorinformationen (Pseudobeobachtungen) eingehen mit den dazugehörigen meist zu optimistischen Genauigkeiten. Somit wird häufig eine Lösung bestimmt, die die gleichen Eigenschaften aufweist wie die für das hier betrachtete Beispiel 5 erhaltene Lösung der „normalen“ Regularisierung - die Residuen der eigentlichen Beobachtungen sind teilweise größer als die Messgenauigkeit der Beobachtungen. Die Vorinformationen wurden so hoch gewichtet, dass sie einen zu großen Einfluss auf die Lösung haben. Ziel sollte es aber sein, aus den neuen Beobachtungen so viel wie möglich an Informationen dazuzugewinnen. Dies kann nur dann erfolgen, wenn die neuen Beobachtungen ein angemessenes Gewicht innerhalb der Regularisierung erhalten und darauf geachtet wird, dass die Residuen der neuen Beobachtungen nach der Regularisierung auch tatsächlich innerhalb sinnvoll ausgewählter Intervalle (z.B. ± 2 facher oder 3 facher mittlerer Fehler) liegen. Dann enthalten die berechneten Parameter neue interpretierbare Informationen. Diese neuen Informationen können verschiedenartig sein. So kann festgestellt werden, dass sich die neuen Parameter überhaupt nicht von den alten Parametern unterscheiden, d.h. die neuen Daten brachten keine neuen Informationen. Dieser Fall ist sicherlich der schlechteste. In einem anderen Fall ist es möglich, dass die neuen Parameter geringe Abweichungen zu den alten Parametern aufweisen, d.h. die neuen Daten bewirkten eine Verbesserung der alten Parameter (unter Einbeziehung aller vorhandenen alten Informationen in Form der Pseudobeobachtungen). Dies ist der beste Fall. Es kann aber auch passieren, dass die neuen Parameter sich extrem von den alten Parametern unterscheiden. In diesem Fall ist es schwer, die Lösung zu beurteilen. Einerseits kann man daraus schlussfolgern, dass die neuen Daten vielleicht doch nicht so genau waren, wie angenommen wurde, andererseits könnte es auch sein, dass die alten Parameter sehr ungenau sind. Eine Entscheidung ist hier sicher schwer.

Anhand des Beispiels 5 konnte erfolgreich gezeigt werden, dass durch die Einbeziehung von Residuentoleranzen in den Regularisierungsprozess die Größe der resultierenden Residuen der Beobachtungen kontrolliert beeinflusst werden kann, so dass eine interpretierbare Lösung erhalten wird. Die Wirkung der Toleranzenbedingungen für die Residuen auf die erhaltene Lösung entspricht der eines Regularisierungsparameters. Eine Betrachtung gleichgewichteter und nicht regularisierter Beispiele war lediglich notwendig, um anhand von einfachen Beispielen die Varianten der Realisierung der Toleranzenbedingung für die Residuen und den Lösungsalgorithmus vergleichen und testen zu können und bei positiver Bewertung des Verfahrens auf komplexere Probleme übergehen zu können.

Abschließend lässt sich sagen, dass der ausgearbeitete Algorithmus sich als sehr stabil und leistungsfähig erwiesen hat und eine Anwendung auf praxisnahe Beispiele realistisch erscheint. Da sich die Variante 2 der Realisierung der Toleranzenbedingung für die Residuen im Vergleich zur Variante 1 als stabiler erwiesen hat, sollte man möglichst diese Variante verwenden, welche auch den Vorteil aufweist, dass die Linearisierung der Bedingungsungleichungen der Linearisierung der Beobachtungsgleichungen entspricht. Somit müssen vor allem

bei komplexeren Problemstellungen keine neuen partiellen Ableitungen gebildet werden, um die Matrix B aufstellen zu können. Als Nachteil der Variante 1 der Realisierung der Toleranzenbedingung für die Residuen ist ebenso die Tatsache zu betrachten, dass durch die Einbeziehung der Bedingungsungleichungen der Variante 1 lineare Minimierungsprobleme wie nichtlineare behandelt werden müssen. Dies bringt zwangsläufig numerische Probleme mit sich.

5. Anwendungsbeispiel - Ableitung von Atmosphärenparametern aus Radiookkultations-Refraktivitäten

Nachdem im vorherigen Kapitel anhand von synthetischen Beispielen gezeigt werden konnte, dass das erarbeitete Regularisierungsverfahren mit Toleranzenbedingungen für die Residuen zu den gewünschten Ergebnissen führt, soll nun innerhalb dieses Kapitels die Anwendung auf reale Daten erfolgen. Da sich die Variante 2 der Realisierung der Toleranzenbedingung (3.2.2) stabiler und leistungsfähiger erwiesen hat als die Variante 1 (3.2.1), wird lediglich diese Variante bei der Auswertung realer Daten benutzt.

5.1 Prinzip der Radiookkultation

Das Radiookkultationsverfahren wurde bereits in den sechziger Jahren zur Erforschung der Atmosphäre des Mars und der Venus angewendet ([Fjeldbo et al. 1967], [Fjeldbo und Eshleman 1968] und [Fjeldbo et al. 1970]) und diente seither zur Untersuchung fast aller Planeten unseres Sonnensystems sowie dazugehöriger Monde, [Wickert 2002]. Vorschläge zur Anwendung des Verfah-

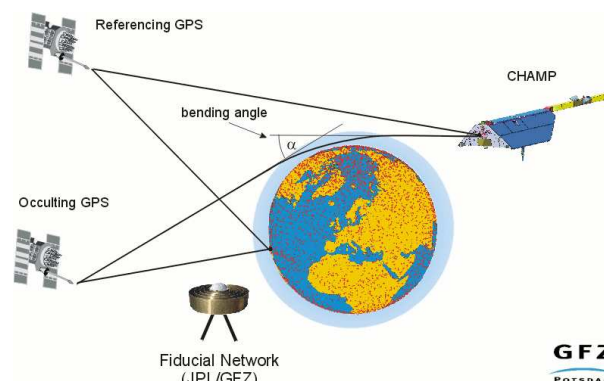


Abbildung 5.1: Prinzip des Radiookkultations-Verfahrens mit CHAMP.

rens zur Erforschung der Erdatmosphäre wurden zwar ebenso in den sechziger Jahren gegeben (z.B. [Fischbach 1965]), aber aufgrund von genaueren konventionellen Verfahren zur Untersuchung der Erdatmosphäre kam es bis in die neunziger Jahre nicht zum Einsatz. Erst durch den Einsatz von 2-Frequenz-GPS¹-Empfängern auf niedrig fliegenden Satelliten wurden die Voraussetzungen für die Nutzung der Radiookkultation zur Erforschung der Erdatmosphäre geschaffen, da somit die Satellitenpositionen und -geschwindigkeiten mit hoher Genauigkeit bereitgestellt werden konnten.

Ein erstes Konzept zur Demonstration des Potenzials der GPS-Radiookkultation für die Sondierung der Erdatmosphäre entstand Ende der achtziger Jahre, [Melbourne et al. 1994]. Im

¹Global Positioning System

Jahr 1995 wurde der wissenschaftliche Kleinsatellit Microlab-1 (GPS-Met) in seine Umlaufbahn gebracht und damit bis 1997 ca. 50000 Okkultationen für die Erdatmosphärensondierung aufgezeichnet. Aktuelle Okkultationen liefert der CHAMP-Satellit des GeoForschungsZentrums (GFZ), [Reigber et al. 2003] und [Reigber et al. 2005].

Das Prinzip der Radiookkultationsmessung ist in der Abbildung 5.1 am Beispiel des CHAMP-Satelliten dargestellt. Die Signale (elektromagnetische Wellen) zweier GPS Satelliten (Referencing GPS und Occulting GPS) werden mittels eines GPS-Empfängers auf einem niedrig fliegenden Satelliten (CHAMP) und von einer GPS-Bodenstation empfangen. Die Wellen, die sich vom Okkultations-GPS-Satelliten (Occulting GPS) zum CHAMP-Satelliten durch die Erdatmosphäre hindurch ausbreiten, werden gebrochen. Die Eigenschaften der Erdatmosphäre in unterschiedlichen Höhenbereichen können somit anhand des Brechungswinkels (bending angle) α charakterisiert werden. Die Variation des Brechungswinkels mit der Höhe ist hauptsächlich von der

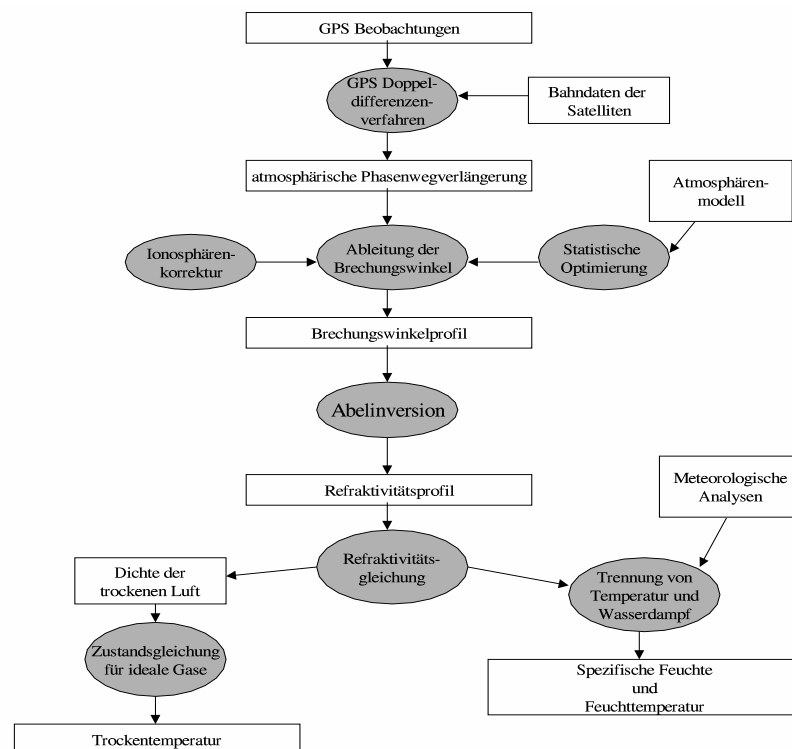


Abbildung 5.2: Schematischer Ablauf der Ableitung von Atmosphärenparametern aus den Radiookkultationsmessungen, [Wickert 2002].

vertikalen Veränderung des atmosphärischen Brechungsindex abhängig.

Durch das Doppeldifferenzenverfahren (siehe z.B. [Seeber 1993]) werden aus den GPS-Messungen die atmosphärischen Phasenwegverlängerungen abgeleitet. Aus der zeitlichen Ableitung der atmosphärischen Phasenwegverlängerung können dann die atmosphärischen Brechungswinkel berechnet werden. Die vertikalen Brechungswinkelprofile werden dann mittels der sog. Abelinversion in vertikale Refraktivitätsprofile transformiert, [Wickert 2002]. In höheren Atmosphärenschichten kann die Annahme gemacht werden, dass der Wasserdampfgehalt Null ist (die absolute Höhe, ab wo diese Annahme gerechtfertigt ist, schwankt stark zwischen Äquator- und Polgebieten), so dass aus den Refraktivitäten direkt die Trockentemperatur abgeleitet werden kann. Ist diese Annahme nicht gerechtfertigt, so muss eine Trennung in Feuchttemperatur und spezifische Feuchte erfolgen.

Diese ist jedoch nicht ohne Verwendung von zusätzlichen Informationen oder Annahmen, z.B. aus bestehenden meteorologischen Modellen, durchführbar (vergleiche mit Formel (5.2)). Dieser Teilabschnitt innerhalb der Radiookkultations-Auswertung (Abbildung 5.2, rechte Verzweigung nach „Refraktivitätsgleichung“) soll innerhalb dieses Kapitels detailliert betrachtet werden. Im Folgenden soll in groben Zügen das Verfahren vorgestellt werden, mit welchem die Trennung in Feuchttemperatur und spezifische Feuchte im Allgemeinen erfolgt. Weiterhin soll das erarbeitete Regularisierungsverfahren mit Residuentoleranzen alternativ zur Trennung dieser Atmosphärenparameter verwendet werden.

Zur Veranschaulichung der gesamten Radiookkultations-Auswertung dient die Abbildung 5.2. Für mehr Details siehe z.B. [Gorbunov und Sokolovskiy 1993], [Melbourne et al. 1994], [Steiner 1998] und [Wickert 2002].

5.2 Ein Standardverfahren zur Ableitung der spezifischen Feuchte und Feuchttemperatur (1DVAR)

Als ein Standardverfahren zur Ableitung der spezifischen Feuchte und der Feuchttemperatur aus Refraktivitäten wird in der Literatur zur Radiookkultation ein Regularisierungsverfahren empfohlen (z.B. in [Healy und Eyre 2000]), welches allgemein als eindimensionales Variationsverfahren (1DVAR) bezeichnet wird. Dieses soll im Folgenden kurz erläutert werden.

Die Eigenschaften der Atmosphäre werden mittels des Brechungsindex n beschrieben. Da dieser für die Erdatmosphäre bei nahezu 1 liegt, ist es zweckmäßig, diesen in die Refraktivität N umzurechnen

$$N = (n - 1) \cdot 10^6. \quad (5.1)$$

Die Refraktivität kann näherungsweise wie folgt als Funktion des Luftdrucks P [hPa], der Feuchttemperatur T [K] und der spezifischen Feuchte Q [g/kg] geschrieben werden:

$$N(z) = \frac{c_1 \cdot P(z)}{T(z)} \left[1 + \frac{k \cdot Q(z)}{T(z)} \right], \quad (5.2)$$

mit den Konstanten $c_1 = 7.76 \cdot 10^{-5} \frac{K}{hPa}$ und $k = 7.72918 \frac{Kkg}{g}$ [Healy und Eyre 2000].

Im Falle der Radiookkultations-Auswertung am GFZ Potsdam liegen Refraktivitäten vom Erdboden bis etwa 35 km Höhe (und häufig darüber hinaus) mit einer vertikalen Auflösung von 200 m vor. Die unbekannten Parameter $T(z)$ und $Q(z)$ werden nicht für alle Höhen z , für die Refraktivitäten vorliegen, bestimmt, sondern lediglich für 43 festgelegte Höhengniveaus². Die spezifische Feuchte wird nur für die unteren 26 Höhengniveaus bestimmt, da für höhere Atmosphärenschichten angenommen wird, dass die Feuchte Null ist. Der Luftdruck $P(z)$ wird als fehlerfrei und bekannt vorausgesetzt, nur der Erdoberflächen-Luftdruck wird als unbekannt betrachtet. Somit liegt die Anzahl der unbekannten Parameter bei 70 - zusammengesetzt aus 43 unbekannten Temperaturen, 26 unbekannten spezifischen Feuchten und einem unbekannten Erdoberflächen-Luftdruck. Die Anzahl der Unbekannten ist damit offensichtlich kleiner als die Anzahl „beobachteter“ Refraktivitäten (durchschnittlich ca. 170), so dass anscheinend³ eine Redundanz vorliegt. Tatsächlich ist

²Diese 43 Höhengniveaus reichen von 1013.25 hPa bis 0.1 hPa.

³Im Falle einer Ausgleichung nach vermittelnden Beobachtungen ist die Redundanz die Differenz zwischen der Anzahl der Beobachtungen und der Anzahl der Unbekannten, vorausgesetzt, dass ein korrekt gestelltes Problem vorliegt.

aber keine Ausgleichung nach kleinsten Quadraten durchführbar, ohne dass zusätzliche Informationen oder Annahmen einbezogen werden, da die Trennung in Feuchttemperatur und spezifische Feuchte nicht eindeutig möglich ist. Als Vorinformationen dienen beim 1DVAR-Verfahren häufig Feuchttemperaturen, spezifische Feuchten und der Erdoberflächen-Luftdruck aus dem ECMWF⁴-Modell⁵.

Die Beobachtungsgleichungen für das zu lösende Regularisierungsproblem haben für die Refraktivitäten folgende Form [Healy und Eyre 2000]:

$$N(z) + v(z) = \frac{c_1 \cdot P(z)}{T(z)} \left[1 + \frac{k \cdot Q(z)}{T(z)} \right] \quad (5.3)$$

mit

$$P(z) = P_i \left(1 + \frac{\gamma_i}{T_i^V} (z - Z_i) \right)^{(-g/(R\gamma_i))} \quad (5.4)$$

$$Q(z) = Q_i \cdot \exp[\alpha_i(z - Z_i)] \quad (5.5)$$

$$T(z) = T_i + \beta_i(z - Z_i) \quad (5.6)$$

$$T_i^V = T_i \cdot (1 + 0.608 \cdot Q_i) \quad (5.7)$$

$$Z_{i+1} = Z_i + \frac{R}{g} \cdot \frac{T_{i+1}^V - T_i^V}{\log_e T_{i+1}^V - \log_e T_i^V} \cdot \log_e \left(\frac{P_i}{P_{i+1}} \right), \quad (5.8)$$

wobei z die geopotentielle Höhe darstellt, g die Schwerebeschleunigung, R die Gaskonstante, $\alpha_i, \beta_i, \gamma_i$ die Gradienten der spezifischen Feuchte, der Temperatur und des Luftdrucks bezogen auf die geopotentielle Höhe und T_i, Q_i, P_i, Z_i die Temperatur, die spezifische Feuchte, der Luftdruck und die geopotentielle Höhe bzgl. der festgesetzten Höhenniveaus (mit $P_1 = P_{EOF} =$ Erdoberflächen-Luftdruck). Hinzu kommen die Beobachtungsgleichungen für die sog. Pseudobeobachtungen, wodurch die Vorinformationen in den Ausgleichungsprozess mit einbezogen werden und die Regularisierung erfolgt:

$$T_i^b + v_i^T = T_i, \quad (5.9)$$

$$Q_i^b + v_i^Q = Q_i, \quad (5.10)$$

$$P_{EOF}^b + v^P = P_{EOF}, \quad (5.11)$$

wobei T_i^b, Q_i^b, P_{EOF}^b Werte für die unbekannten Parameter P_{EOF}, T_i, Q_i aus dem ECMWF-Modell sind. Die unbekannten Parameter können dann durch Anwendung einer Ausgleichung nach kleinsten Quadraten (entspricht der bereits vorgestellten stochastischen Regularisierung, siehe Abschnitt 2.2.3) berechnet werden, d.h. die gewichtete Summe aller Residuenquadrate wird minimiert:

$$\left\{ \sum_{i=1}^m p_i v_i^2 + \sum_{i=1}^{43} p_i^T (v_i^T)^2 + \sum_{i=1}^{26} p_i^Q (v_i^Q)^2 + p^P (v^P)^2 \right\} \rightarrow \min, \quad (5.12)$$

⁴European Centre for Medium-Range Weather Forecasts

⁵Für mehr Details zum ECMWF-Modell siehe z.B. <http://www.ecmwf.int>.

wobei die p_i, p_i^T, p_i^Q, p^P die Gewichte der einzelnen Beobachtungen bzw. Pseudobeobachtungen darstellen, v_i, v_i^T, v_i^Q, v^P die Residuen der Refraktivitäten und Pseudobeobachtungen und m die Anzahl der auszuwertenden Refraktivitäten. Als erste Näherungslösung für die iterative Lösungsbestimmung dienen dabei die Modellparameter T_i^b, Q_i^b und P_{EOF}^b .

Anhand der Formel (5.12) lässt sich unmittelbar der Nachteil des 1DVAR-Verfahrens ablesen: Werden die Gewichte für die Pseudobeobachtungen im Verhältnis zu den Gewichten der eigentlichen Beobachtungen zu groß gewählt, so werden bei der Minimierung der gewichteten Summe der Residuenquadrate hauptsächlich die Abweichungen v_i^T, v_i^Q, v^P der neuen Lösung zur bestehenden Modelllösung minimiert. Die Residuen der Refraktivitäten können dahingegen unerwünscht und unkontrollierbar groß werden. Eine derartige Übergewichtung der Vorinformationen geschieht recht häufig, da einerseits zu vorhandenen Modellen oft sehr optimistische Fehlerschätzungen für die Modellparameter angegeben werden und andererseits wird oft absichtlich das vorhandene Modell übergewichtet, um das zu lösende Problem „besser“ zu stabilisieren. Dies ist eine allgemein akzeptierte Vorgehensweise bei der Lösung regularisierter Probleme, die unabhängig von der konkreten Anwendung propagiert wird. So findet man z.B. im Zusammenhang mit der Schwerefeldbestimmung in [Gruber 2000]: „... für eine Dekorrelation der stark korrelierten Koeffizienten [ist] eine Übergewichtung der stochastischen Vorinformationen notwendig. Im Falle von GRIM4-Satellitenlösungen wird eine Übergewichtung der Gradvarianzen mit dem Faktor 100 eingeführt ...“

Liegen neue Messdaten vor (wie hier die CHAMP-Refraktivitäten, die eine neue Datenqualität aufweisen sollen), so sollte das Ziel sein, aufgrund der neuen genauen Daten neue Informationen für die unbekannten Parameter zu ermitteln, d.h. das bestehende Modell (ECMWF), welches aus anderen Daten abgeleitet wurde, zu verbessern. Durch die Nutzung eines Regularisierungsverfahrens wie das 1DVAR-Verfahren kann dieses Ziel meist nicht erreicht werden oder nur zufallsbedingt bei einer für die Refraktivitäten günstigen Wahl der Gewichtsverhältnisse. Häufig besteht die Auffassung, dass die erhaltene Lösung nur dann gut ist, wenn die Abweichungen der neuen Lösung zu der bestehenden Modelllösung möglichst klein sind. Dabei bleiben die Residuen der Refraktivitäten meist völlig ohne Beachtung bzw. es werden lediglich mittlere quadratische Residuen als Qualitätskriterium für die Lösungen verwendet. Diese mittleren quadratischen Residuen können allerdings keine Aussagen darüber enthalten, wie die Verteilung der einzelnen Residuen ist. Durch Variation der Gewichtsverhältnisse sollte es zwar grundsätzlich möglich sein, eine sinnvolle Verteilung der Residuen zu erreichen, diese Vorgehensweise ist aber sehr mühsam, da es im Voraus nicht möglich ist, die Wirkung der Gewichtsvariation auf die Residuen abzuschätzen. Somit soll das innerhalb dieser Arbeit erarbeitete Regularisierungsverfahren mit Toleranzbedingungen für die Residuen als alternatives Verfahren zum 1DVAR-Verfahren für die Auswertung der Refraktivitäten verwendet werden.

5.3 Regularisierung mit Residuentoleranzen zur Ableitung der spezifischen Feuchte und Feuchttemperatur

Werden zu dem im vorherigen Abschnitt beschriebenen Regularisierungsproblem (1DVAR) lediglich zusätzliche Bedingungen für die Residuen eingeführt, so kann der Nachteil des 1DVAR-Verfahrens - die Unkontrollierbarkeit der Größe der Residuen der Refraktivitäten - behoben werden. Die Toleranzbedingungen für die Residuen sind $v_i \leq t_i$ und $-v_i \leq t_i$, d.h. für jede „beobachtete“ Refraktivität werden zwei Bedingungen bzgl. des Residuums eingeführt, wobei t_i einen sinnvollen Grenzwert für das Residuum v_i darstellt. Die Residuen ergeben sich aus der

Formel (5.3) zu:

$$\begin{aligned} v_i &= \frac{c_1 \cdot P(z_i)}{T(z_i)} \left[1 + \frac{k \cdot Q(z_i)}{T(z_i)} \right] - N(z_i), \\ -v_i &= N(z_i) - \frac{c_1 \cdot P(z_i)}{T(z_i)} \left[1 + \frac{k \cdot Q(z_i)}{T(z_i)} \right]. \end{aligned}$$

Durch Linearisierung der Toleranzenbedingungen und Beobachtungsgleichungen (siehe Abschnitt 3.2) und Verwendung des im Abschnitt 3.3 erarbeiteten Algorithmus ist es möglich, die unbekannten Parameter (Temperaturen, spezifische Feuchten und Erdoberflächen-Luftdruck) aus den Refraktivitäten abzuleiten. Da innerhalb der Auswertung Grenzen (Toleranzen) für die Residuen der Refraktivitäten berücksichtigt werden (entsprechen z.B. sinnvollerweise den Genauigkeitsannahmen, die auch für die Gewichtsmatrix der Beobachtungen verwendet werden), ergeben sich Lösungen, die einerseits gut zum bestehenden Modell passen können, aber andererseits die Refraktivitäten nicht innerhalb ihrer Genauigkeit verletzen. Als weiterer Vorteil dieses erweiterten Regularisierungsverfahrens ergibt sich, dass die Lösung bzw. die Abweichungen der neuen Lösung zum bestehenden Modell interpretiert werden können.

Bei dem betrachteten Beispiel (Ableitung von Atmosphärenparametern aus Refraktivitäten) handelt es sich um ein nicht eindeutig lösbares Problem, d.h. ein inkorrekt gestelltes Problem. Es existieren unendlich viele mögliche Paare von Temperaturen und spezifischen Feuchten, die die Refraktivitäten gleich gut repräsentieren können. Was bringt dann eine Auswertung der Refraktivitäten mittels Regularisierung mit Toleranzenbedingungen für die Residuen? Um inkorrekt gestellte Probleme lösen zu können, sind zusätzliche Hypothesen oder Beobachtungen notwendig, die hier durch die Hinzunahme von Pseudobeobachtungen erbracht werden. Diese stellen grundsätzlich nicht zu vernachlässigende Informationen dar, sollten allerdings bei der Auswertung nicht überbewertet werden. Somit werden mittels des erweiterten Regularisierungsverfahrens Lösungen erzwungen (es existiert keine eindeutige Lösung für inkorrekt gestellte Probleme), bei denen sämtliche Residuen der Beobachtungen innerhalb der Genauigkeitsvorgaben für die einzelnen Refraktivitäten liegen. Da aber gleichzeitig in den Auswerteprozess die bestehenden Modellparameter eingegangen sind (ohne diese wäre eine Lösungsbestimmung bei diesem Beispiel unmöglich), ist die ermittelte Lösung nicht nur von den neuen „beobachteten“ Refraktivitäten abhängig, sondern ebenso von diesen vorhandenen Modellparametern. Daraus ist ersichtlich, dass mit dem erweiterten Regularisierungsverfahren überprüft werden kann, ob die neuen Daten zum bestehenden Modell passen. Dies stellt somit die eigentliche Aufgabenstellung bei der Auswertung der Refraktivitäten dar: Überprüfung und, wenn möglich, gezielte Verbesserung des vorhandenen Modells. Da vorhandene Modelle in der Regel nicht grob falsch sein sollten, sind meistens nur recht kleine Abweichungen der neuen Lösung zu der bestehenden Lösung zu erwarten. Diese Abweichungen können dann als neue Informationen (gewonnen aus den Refraktivitäten) für das bestehende Modell betrachtet werden, da dieses indirekt an der Lösungsbestimmung beteiligt war.

Bei Anwendung des 1DVAR-Verfahrens können die erhaltenen Abweichungen zwischen der neuen und der bestehenden Lösung nicht derart interpretiert werden, da die Residuen der Refraktivitäten beliebig groß sein können. Nur für den Fall, dass sämtliche Residuen zufallsbedingt innerhalb der Genauigkeitsgrenzen der Refraktivitäten liegen, ist diese Interpretation gerechtfertigt.

5.4 Ausgewählte Beispiele

Um die Vor- und Nachteile der beiden zuvor vorgestellten Regularisierungsverfahren zur Ableitung der spezifischen Feuchte und der Feuchttemperatur aus Refraktivitäten untersuchen zu können, wurden 63 global verteilte Refraktivitätsprofile beliebig ausgewählt und ausge-

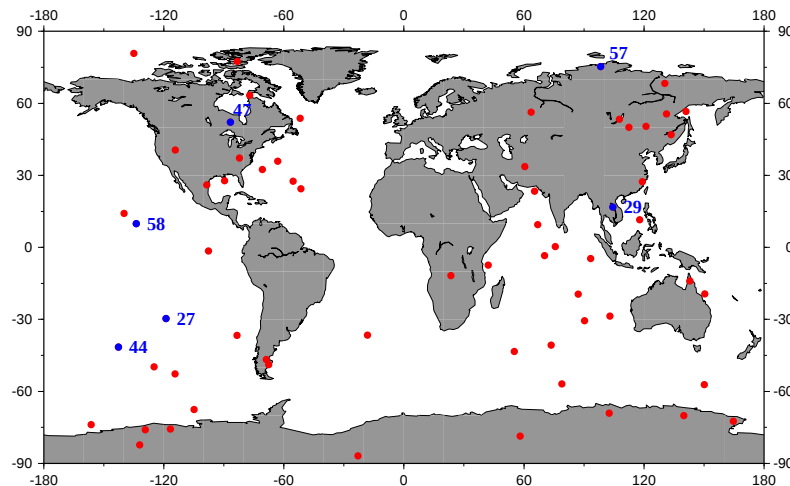


Abbildung 5.3: Globale Verteilung der 63 ausgewählten Okkultationsprofile; blaue Punkte = speziell ausgewählte Profile.

wertet, siehe rote und blaue Punkte in Abbildung 5.3. Diese Profile wurden am GFZ aus CHAMP-Okkultationsmessungen abgeleitet und sind im CHAMP-ISDC⁶ unter den Dateinamen abgespeichert, wie sie im Anhang A.2 aufgelistet sind. Die für die Regularisierung notwendigen Vorinformationen wurden dem ECMWF-Modell entnommen. Dieses Modell stellt für 60 Druckniveaus den Luftdruck, die spezifische Feuchte und die Feuchttemperatur zur Verfügung. Da das 1DVAR-Programm am GFZ für die Ableitung der Atmosphärenparameter auf 43 Druckniveaus konzipiert wurde, welche mit den ECMWF-Druckniveaus nicht identisch sind, ist es notwendig, die aus dem ECMWF-Modell bezogenen Vorinformationen auf die gewünschten 43 Druckniveaus zu interpolieren. Diese Interpolationen erfolgen logarithmisch bzgl. des Luftdrucks. Für die Genauigkeiten der Refraktivitäten wurde eine globale Genauigkeitsschätzung des GFZ verwendet [Wickert 2004], siehe Anhang A.3. Korrelationen zwischen den einzelnen Refraktivitäten wurden innerhalb dieser Arbeit vernachlässigt, da genaue Kenntnisse darüber fehlen. Eine global gültige⁷ Varianz-Kovarianzmatrix für die Vorinformationen wurde von S.B. Healy vom Meteorologischen Institut in Großbritannien bereitgestellt, worin auch Kovarianzen verschieden von null angegeben wurden. Diese Kovarianzen wurden für sämtliche im Rahmen dieser Arbeit durchgeführten Berechnungen zu null gesetzt. Die verwendeten Varianzen für die

⁶Information System and Data Center

⁷An dieser Stelle ist anzumerken, dass eine derartige Varianz-Kovarianzmatrix nur fiktiv existieren kann, da einerseits keine gleichmäßige globale Überdeckung von Messungen vorliegt, aus denen die Vorinformationen abgeleitet werden und andererseits die Qualität der Messungen in unterschiedlichen Regionen der Erde unterschiedlich ist.

Vorinformationen sind im Anhang A.4 aufgelistet.

Da es sich bei dem 1DVAR-Verfahren wie auch bei dem erweiterten Regularisierungsverfahren um Iterationsverfahren handelt, war es ebenso notwendig, Abbruchkriterien für die Iteration festzusetzen. Dazu wurden zwei Kriterien verwendet: Die Iteration wird abgebrochen, wenn die Summe der Residuenquadrate (aller Residuen) fünf mal in Folge innerhalb der Rechengenauigkeit gleich blieb bzw. wenn die maximale Iterationsanzahl von 50 Iterationen erreicht wurde⁸.

Wählt man beim erweiterten Regularisierungsverfahren die Toleranzen für die Residuen der Refraktivitäten extrem groß (z.B. mit 10^6 N-Einheiten), so haben diese auf die Berechnung der Lösung keinen Einfluss, so dass man die gleiche Lösung erhält, wie mit dem 1DVAR-Verfahren. Dies konnte durch Vergleichsberechnungen bestätigt werden. Aus diesem Grund wurden sämtliche im Folgenden mit 1DVAR bezeichneten Ergebnisse mittels des erweiterten Regularisierungsverfahrens mit festgesetzten Toleranzen von 10^6 berechnet. Dies hat den Vorteil, dass beim Vergleich der Ergebnisse zwischen 1DVAR und dem erweiterten Regularisierungsverfahren keine extra Betrachtung der unterschiedlichen Rechenschärfen erfolgen muss.

Verwendet man das 1DVAR-Verfahren zur Berechnung der spezifischen Feuchten und Feuchttemperaturen, so liegen lediglich bei 22 (von 63) Profilen die Residuen der Refraktivitäten innerhalb der für die Gewichtsmatrix der Refraktivitäten verwendeten Genauigkeiten. Werden Toleranzen für die Residuen der Refraktivitäten eingeführt, welche den Fehlerannahmen der Refraktivitäten für die Gewichtsmatrix entsprechen⁹, so können für insgesamt 40 Profile die Residuen der Refraktivitäten in den Toleranzenbereich gezwungen werden. Bei den restlichen 23 Profilen lagen unterschiedliche Gründe für die Nichterfüllbarkeit der Toleranzen vor. So konnte bei vier Profilen keine Konvergenz des Iterationsverfahrens erzielt werden. Bei acht Profilen lagen fast sämtliche Residuen der Refraktivitäten nach der 1DVAR-Auswertung außerhalb der Fehlerannahmen für die Refraktivitäten. Um die Residuen mit dem erweiterten Regularisierungsverfahren innerhalb vorgegebener Toleranzen zwingen zu können, ist es allerdings notwendig, dass einzelne Residuen nach der Auswertung ohne Toleranzen kleiner als die dazugehörigen Fehlerannahmen der Beobachtungen waren und nur einzelne zu groß. Nur dann ist eine Umverteilung der Residuen möglich, so dass bei diesen acht Profilen die Nichterfüllbarkeit der Toleranzen gerechtfertigt ist. Mit Umverteilung der Residuen ist gemeint, dass einzelne zu große Residuen verkleinert werden, andere sehr kleine Residuen dagegen größer werden. Darauf wird nochmals ausführlich am Ende dieses Abschnittes eingegangen. Bei den restlichen elf Profilen lagen bereits die „kleinst-möglichen“ Residuen der Refraktivitäten außerhalb der Fehlerannahmen für die Refraktivitäten. Diese „kleinst-möglichen“ Residuen konnten näherungsweise dadurch ermittelt werden, indem die Gewichte für die Pseudobeobachtungen möglichst klein gewählt wurden (hier wurden die ursprünglichen Gewichte durch 50.0 dividiert¹⁰), so dass die Fehlerannahmen für die Pseudobeobachtungen möglichst groß wurden. Somit wurden bei der Minimierung der Summe der Residuenquadrate (bei Verwendung des 1DVAR-Verfahrens) hauptsächlich die Residuen der Refraktivitäten minimiert, wohingegen

⁸Diese Abbruchkriterien wurden aufgrund numerischer Tests festgelegt.

⁹Die hier verwendeten Fehlerschätzungen für die Refraktivitäten sind recht pessimistisch, so dass diese durchaus dem 2- oder 3fachen wahren Fehler entsprechen können.

¹⁰Der Wert 50.0 wurde anhand von numerischen Tests festgelegt.

die Residuen der Pseudobeobachtungen sehr groß werden könnten¹¹. Die „kleinst-möglichen“ Residuen repräsentieren die innere Genauigkeit ([Wolf 1979]) der Refraktivitäten sehr gut, d.h. sie zeigen, wie gut die einzelnen Refraktivitäten zueinander passen. Somit ist auch für diese elf Profile die Nichterfüllbarkeit der Toleranzen gerechtfertigt.

Die aufgetretenen Probleme bei der Konvergenz des Iterationsverfahrens könnten z.B. dadurch behoben werden, dass das im Abschnitt 2.2.2 vorgestellte „gedämpfte Iterationsverfahren“ verwendet wird¹². Bei den Profilen, bei denen die Residuen bzw. die „kleinst-möglichen“ Residuen oberhalb der Fehlerannahmen lagen, ist es notwendig, die Genauigkeiten der Vorinformationen und der Refraktivitäten bzw. nur die Genauigkeiten der Refraktivitäten neu zu überdenken und neu festzulegen. Mit den neuen verbesserten Fehlerannahmen sollte dann die Erfüllbarkeit der Toleranzen (= Fehlerannahmen für die Refraktivitäten) ohne weitere Probleme möglich sein. An dieser Stelle soll aber von derartigen Untersuchungen abgesehen werden, da es innerhalb dieses Anwendungskapitels nicht um die Festlegung der Genauigkeiten für die Refraktivitäten gehen soll, sondern um den Vergleich der beiden vorgestellten Regularisierungsverfahren. Eine Erfolgsquote von 63 Prozent (entspricht den 40 Profilen) für das erweiterte Regularisierungsverfahren ist für einen aussagekräftigen Vergleich völlig ausreichend.

Anhand von sechs ausgewählten Profilen soll nun der Unterschied zwischen den Lösungen der beiden Regularisierungsverfahren veranschaulicht werden. Die gewählten Profile sind in der Abbildung 5.3 als blaue Punkte mit einer Punktnummer dargestellt, die der Nummer der Inputdatei nach Anhang A.2 entspricht.

In der Abbildung 5.4 sind die für die Gewichtsmatrix der Refraktivitäten verwendeten Fehlerschätzungen (rot) dargestellt, die Residuen der Refraktivitäten nach Anwendung des 1DVAR-Verfahrens (grün) sowie die bereits erläuterten „kleinst-möglichen“ Residuen der Refraktivitäten (blau). Man erkennt, dass die innere Genauigkeit der Refraktivitäten bei den gewählten sechs Profilen recht unterschiedlich ist. Die Profile 47 und 57 weisen auf sehr konsistente Refraktivitäten hin, wohingegen gerade beim Profil 29 die innere Genauigkeit stark schwankt. Bei allen sechs Profilen liegen einige 1DVAR-Residuen oberhalb der Fehlerschätzungen der Refraktivitäten und weisen geradezu „Spitzen“ auf. Man könnte nun annehmen, dass es sich bei den einzelnen „Spitzen“ um Residuen von Refraktivitäten handelt, die grobe Fehler aufweisen. Dass dies nicht der Fall ist, zeigen die „kleinst-möglichen“ Residuen. Diese sind an den Stellen der „Spitzen“ wesentlich kleiner als die 1DVAR-Residuen. Somit resultieren diese „Spitzen“ nicht aus Fehlern in den Refraktivitäten, sondern aus einer Übergewichtung der Vorinformationen gegenüber den Refraktivitäten. Werden diese „herausstechenden“ Residuen gezwungen, innerhalb der Fehlerannahmen der Refraktivitäten (= Toleranzen) zu liegen, so ist dies gleichbedeutend damit, dass speziell aus den dazugehörigen Refraktivitäten Informationen für die Lösung (Feuchttemperatur und spezifische Feuchte) bezogen werden und die Vorinformationen gezielt weniger Einfluss auf die Lösung erhalten. Somit ist die Lösung (Profil der Feuchttemperatur und der spezifischen Feuchte) vor allem in den Höhenbereichen, wo die „herausstechenden“ Residuen zu beobachten waren, bei Verwendung des erweiterten Regularisierungsverfahrens deutlich erkennbar von der 1DVAR-Lösung (und somit auch von der Vorinformation) verschieden, Abbildungen 5.5 und 5.6. Sämtliche Residuen der Refraktivitäten

¹¹Siehe dazu Formel (5.12). Sind die Gewichte der Pseudobeobachtungen p_i^T, p_i^Q, p^P sehr klein, so können die Residuen der Pseudobeobachtungen v_i^T, v_i^Q, v^P groß werden, ohne dass die gesamte Summe der Residuenquadrate davon stark beeinflusst wird.

¹²Diese Konvergenzprobleme stellen keine für dieses Verfahren spezifischen Probleme dar, sondern allgemeine Probleme bei der Verwendung von Iterationsverfahren. Allerdings werden diese Probleme selten festgestellt, da die Behandlung der Konvergenz oft „unsauber“ erfolgt.

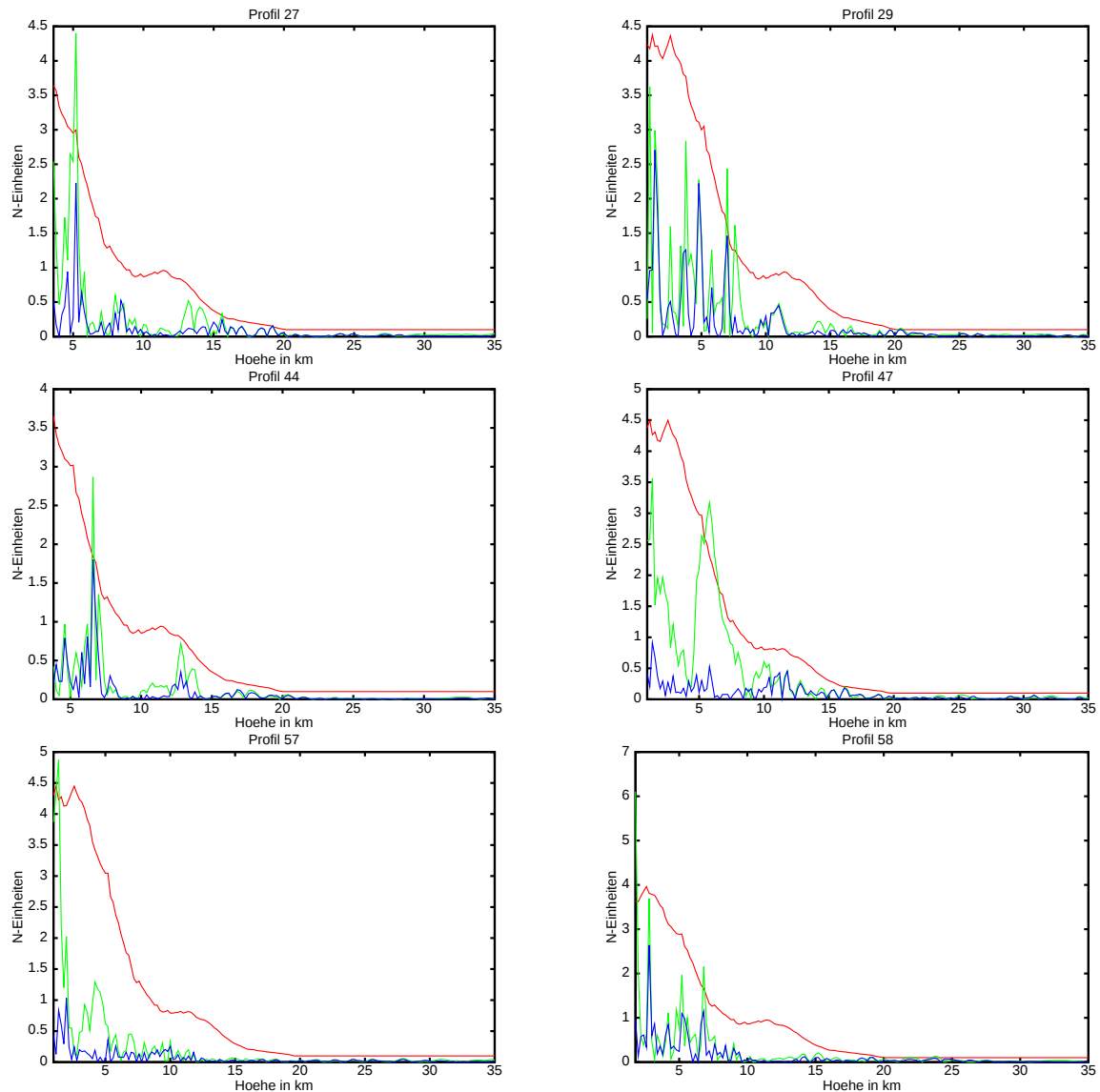


Abbildung 5.4: Residuenplots und Fehlerannahmen:

- rot ... Fehlerschätzungen der Refraktivitäten für die Gewichtsmatrix;
- grün ... 1DVAR-Residuen der Refraktivitäten;
- blau ... „kleinst-mögliche“ Residuen der Refraktivitäten.

der sechs Profile liegen dann im Bereich der Fehlerannahmen für die Refraktivitäten. Die Ergebnisse für die Feuchttemperaturen sind in Abbildung 5.5 dargestellt, die Ergebnisse für die spezifischen Feuchten in Abbildung 5.6. Die grüne Kurve stellt jeweils die Differenzenbeträge der Vorinformationen (ECMWF-Modell) zur 1DVAR-Lösung dar, die blaue Kurve die Differenzenbeträge der Lösung des erweiterten Regularisierungsverfahrens zur 1DVAR-Lösung und die rote Kurve die Differenzenbeträge der ECMWF-Vorinformation zur Lösung des erweiterten Regularisierungsverfahrens. Die größten Abweichungen zwischen den beiden Lösungen (1DVAR und erweitertes Regularisierungsverfahren) lassen sich für die spezifischen Feuchten bei den Profilen 29 und 44 in einem Höhenbereich von 6-8 km mit Abweichungen von ca. 0.3 g/kg beobachten. Bei den Profilen 27, 57 und 58 lassen sich kleinere Abweichungen bei den spezifischen Feuchten in Höhenbereichen von 4-6 km, 1-2 km bzw. 5-10 km feststellen. Bei dem Profil 47 sind die Abweichungen bei der spezifischen Feuchte nahezu Null, dafür sind aber die Abweichungen in

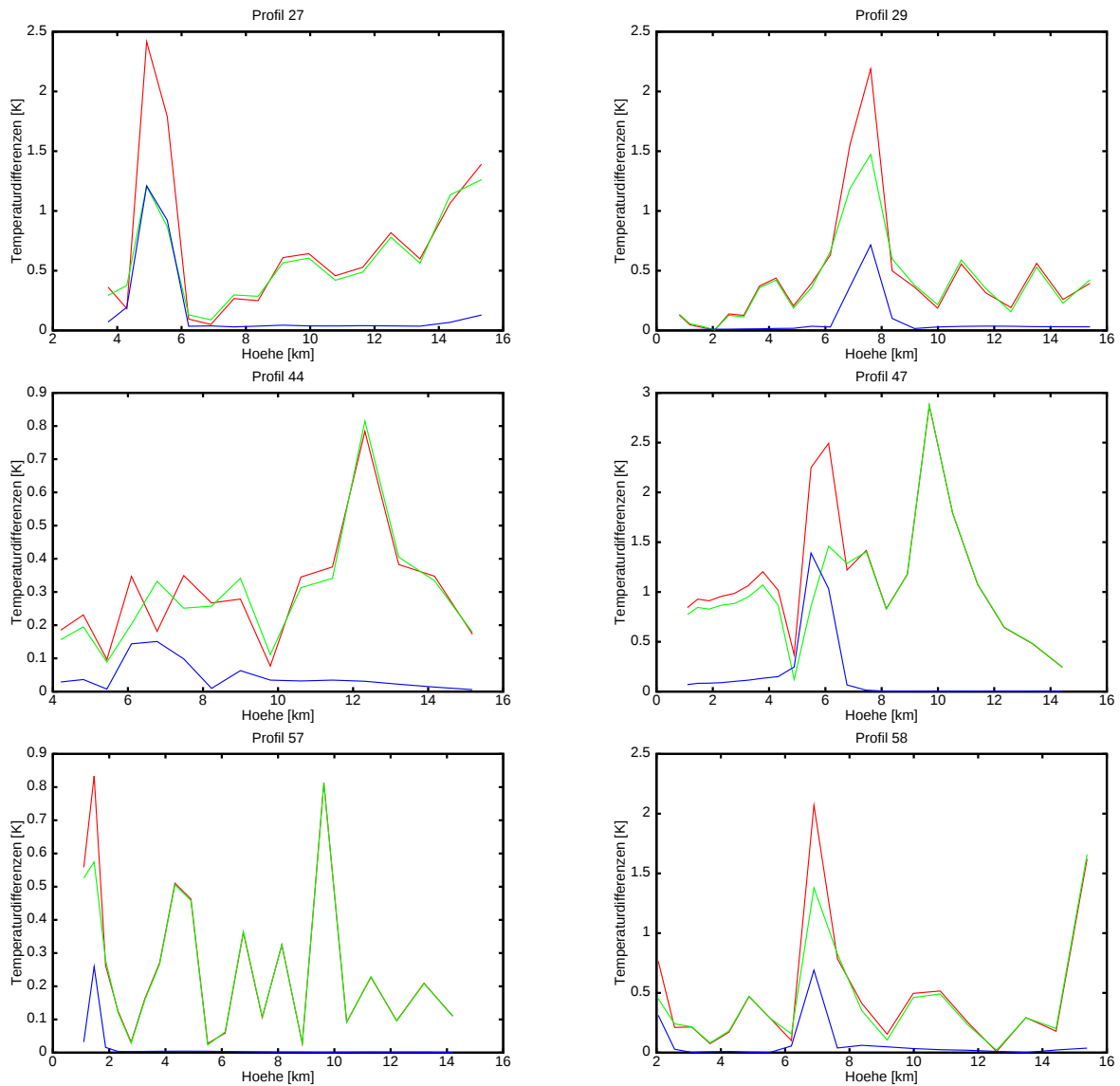


Abbildung 5.5: Differenzenbeträge der Feuchttemperaturen:

rot ... ECMWF-Vorinformation - Lösung des erweiterten Regularisierungsverfahrens;

grün ... ECMWF-Vorinformation - 1DVAR-Lösung;

blau ... Lösung des erweiterten Regularisierungsverfahrens - 1DVAR-Lösung.

den Feuchttemperaturen für dieses Profil am größten (ca. 1.4 K).

Die Profile 47 und 57 weisen nur einen sehr geringen Anteil an spezifischer Feuchte auf¹³, was an ihrer nördlichen Lage - Kanada bzw. Sibirien - liegt. Den größten Anteil an spezifischer Feuchte zeigt das Profil 29, welches über Thailand liegt.

Wie bereits erläutert wurde, stellen die beim erweiterten Regularisierungsverfahren zu beobachtenden Abweichungen zwischen der Lösung für die Feuchttemperatur bzw. der spezifischen Feuchte und den Vorinformationen neue Informationen dar, die hier aus den CHAMP-Okkultationen gewonnen werden konnten. In Höhenbereichen bis ca. 5 km sollte allerdings bei

¹³Die absoluten Größen der berechneten Feuchttemperaturen und spezifischen Feuchten sind im Anhang A.5 in den Abbildungen A.1 und A.2 dargestellt.

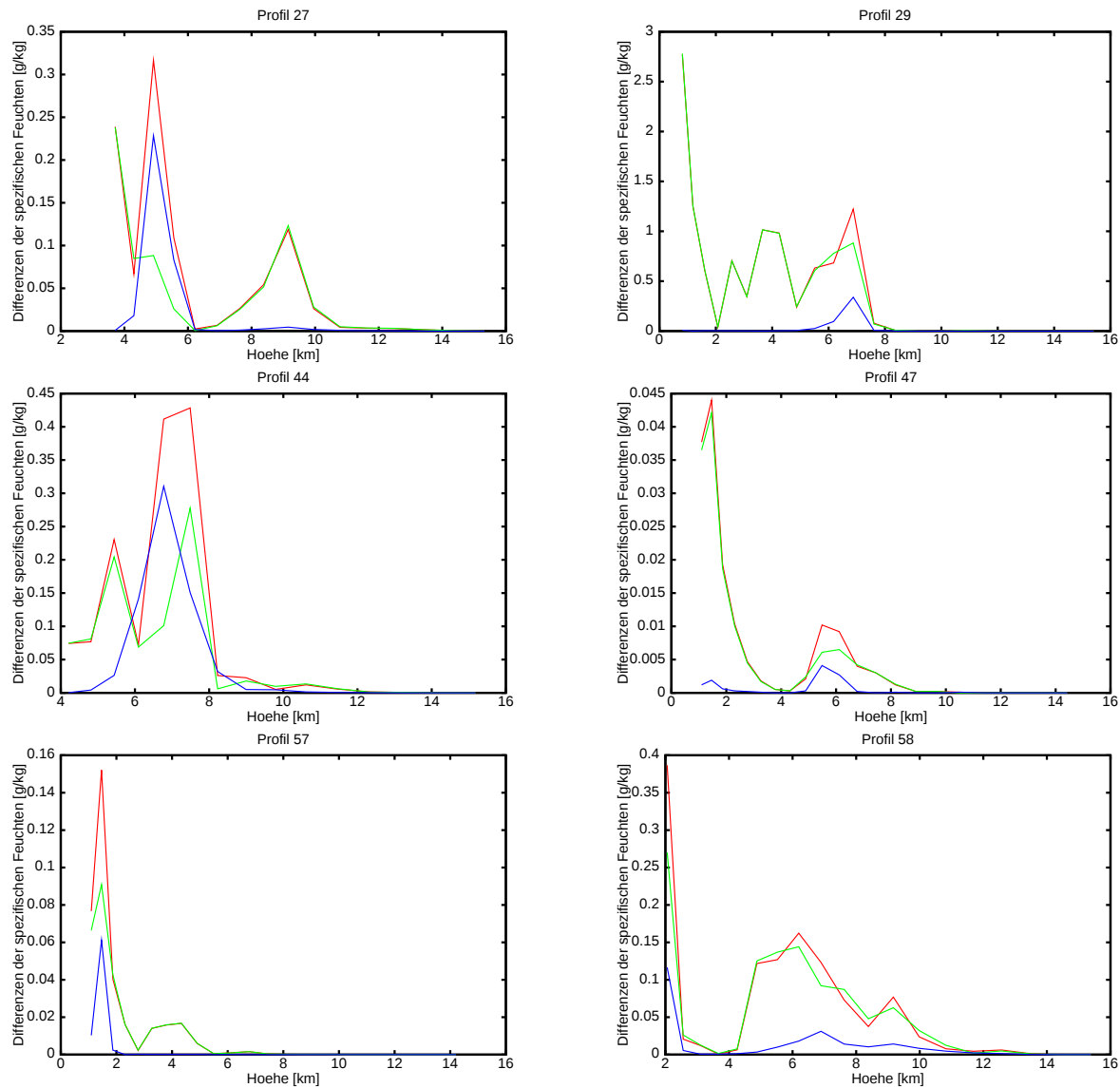


Abbildung 5.6: Differenzenbeträge der spezifischen Feuchten:

rot ... ECMWF-Vorinformation - Lösung des erweiterten Regularisierungsverfahrens;
 grün ... ECMWF-Vorinformation - 1DVAR-Lösung;
 blau ... Lösung des erweiterten Regularisierungsverfahrens - 1DVAR-Lösung.

der Interpretation der Abweichungen berücksichtigt werden, dass einige Okkultationsprofile in diesem Bereich aufgrund von kritischer Refraktion und Fehlern bei der Signalverfolgung recht ungenau sind, siehe dazu [Beyerle et al. 2003] und [Beyerle et al. 2004]. Dies betrifft am häufigsten Profile im tropischen Pazifik, Zentral Afrika, Indonesien und Indien ([Beyerle et al. 2004]), Gegenden, in denen die Luftfeuchtigkeit sehr hoch und somit das Auftreten von Störungen während der Signalausbreitung sehr wahrscheinlich ist. Bei den ausgewählten sechs Profilen kam es in den in der Tabelle 5.1 aufgeführten Höhenbereichen wahrscheinlich zu Signalverfolgungsfehlern bzw. kritischer Refraktion. Diese Höhenbereiche wurden aus den Amplitudenspektren der GPS-Signale abgeschätzt. Die Spektren wurden mittels der sog. Full Spectrum Inversion berechnet ([Jensen et al. 2003]; Beyerle 2004, persönliche Kommunikation). Daraus folgt, dass die beobachteten Abweichungen bei der spezifischen Feuchte bzw. der Feuchttemperatur zu den Vorinformationen in den kritischen Höhenbereichen nicht unbedingt als eine Verbesserung für

Tabelle 5.1: Kritische Höhenbereiche bei den ausgewählten Profilen

Profilnr.	kritischer Höhenbereich [km]
27	1.5 – 7.0
29	0.0 – 5.0
44	2.0 – 5.0
47	0.0 – 10.0
57	0.0 – 1.0
58	1.0 – 5.0

die vorhandenen Atmosphärenparameter des ECMWF-Modells interpretiert werden können, sondern auch durch Signalverfolgungsfehler und kritische Refraktion verursacht sein können. Bei allen sechs Profilen lassen sich aber auch in Höhenbereichen, welche als unkritisch eingestuft werden konnten, Abweichungen der Lösung des erweiterten Regularisierungsverfahrens zu den Vorinformationen beobachten, z.B. Abweichungen von 2.0 - 2.5 K bei den Profilen 29, 47 und 58 sowie eine Abweichung von 1.2 g/kg beim Profil 29. Somit kann man dort davon ausgehen, dass die Anwendung des erweiterten Regularisierungsverfahrens dazu führte, dass neue Informationen bzgl. der spezifischen Feuchte und der Feuchttemperatur aus den Radiookkultations-Refraktivitäten gewonnen werden konnten (gegenüber der 1DVAR-Lösung und gegenüber dem ECMWF-Modell).

In Gebieten (wie z.B. der Antarktis), wo die Luftfeuchtigkeit sehr niedrig ist und somit Signalverfolgungsfehler und kritische Refraktion sehr selten auftreten, scheint die Anwendung des erweiterten Regularisierungsverfahrens besonders vielversprechend zu sein. In der Antarktis ist zudem das ECMWF-Modell recht ungenau, da für dieses Gebiet nur wenige Messdaten (z.B. Radiosonden-Messungen) vorhanden sind, die für die Modellerstellung genutzt werden können (CHAMP-Daten gehen bislang nicht ein).

Von den 63 ausgewählten Profilen befinden sich 3 Profile über der Antarktis (ohne Küstenbereich), siehe Abbildung 5.3. Die für diese Profile ermittelten Abweichungen der mit dem erweiterten Regularisierungsverfahren berechneten Feuchttemperaturen und spezifischen Feuchten zu den entsprechenden ECMWF-Parametern liegen in der gleichen Größenordnung wie bei den sechs ausgewählten Profilen. Diese Abweichungen können dort allerdings recht zuverlässig über den gesamten Höhenbereich als Verbesserungen der ECMWF-Parameter interpretiert werden. Da der CHAMP-Satellit lediglich ca. 16 Radiookkultations-Profile pro Tag über der Antarktis liefert, ist durch Radiookkultations-Messungen mit CHAMP allein noch keine oder nur eine geringfügige Verbesserung des ECMWF-Modells über der Antarktis möglich. Allerdings liegen bereits erste Radiookkultationsmessungen vom GRACE¹⁴-Satelliten vor [Wickert et al. 2005], so dass eine Verbesserung des ECMWF-Modells über der Antarktis in der nächsten Zeit möglich erscheint.

Wie bereits im Abschnitt 5.2 erwähnt wurde, kommt es häufig bei der Verwendung von Regularisierungsverfahren absichtlich oder durch fehlende Kenntnisse der tatsächlichen Ge-

¹⁴Gravity Recovery And Climate Experiment

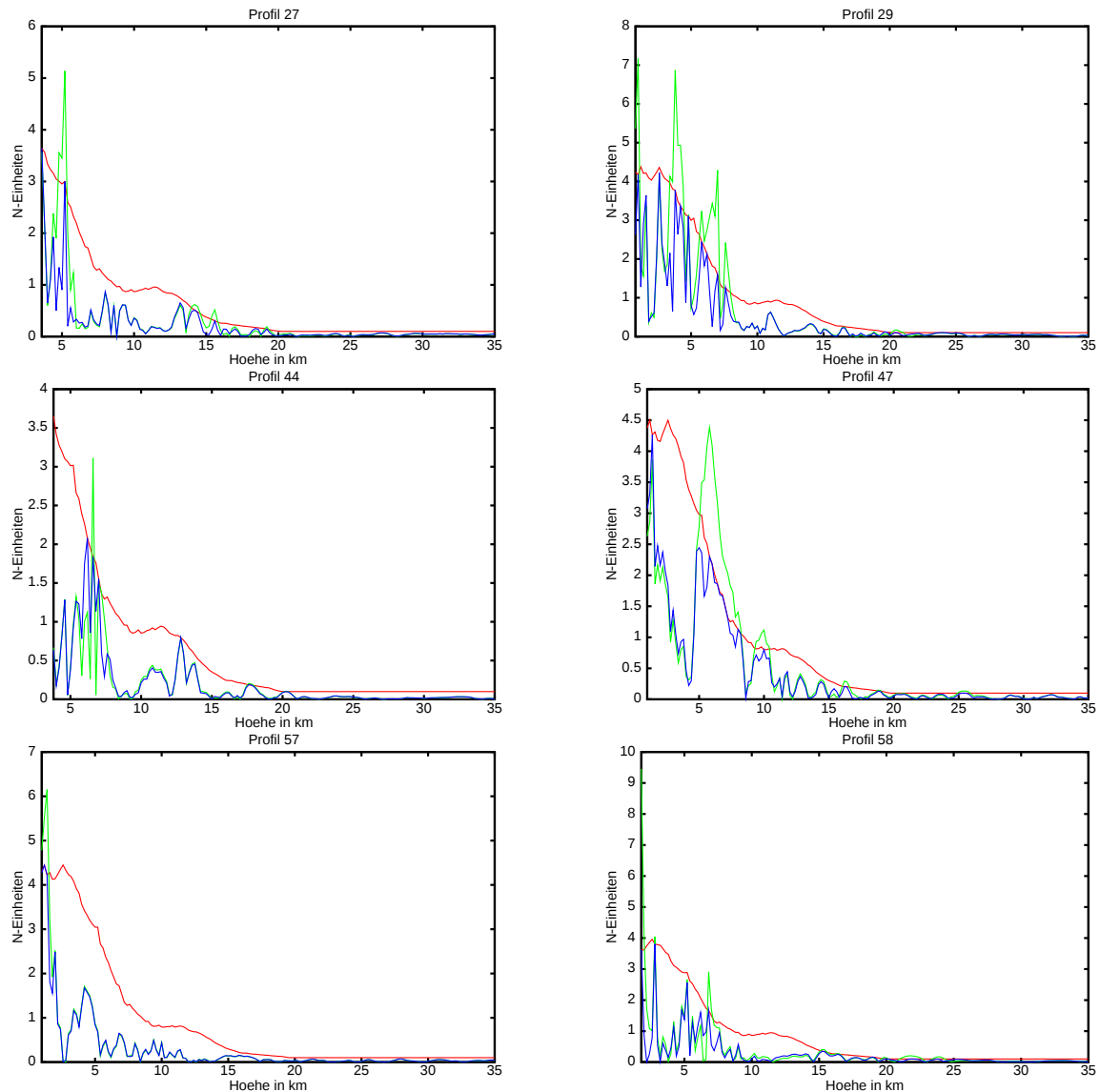


Abbildung 5.7: Residuenplots und Fehlerannahmen bei Übergewichtung der Pseudobeobachtungen:

- rot ... Fehlerschätzung der Refraktivitäten für die Gewichtsmatrix;
- grün ... 1DVAR-Residuen der Refraktivitäten;
- blau ... Residuen der Refraktivitäten aus dem erweiterten Regularisierungsverfahren.

wichtsverhältnisse zu einer Übergewichtung der Vorinformationen gegenüber den eigentlichen Beobachtungen. Es soll nun anhand der ausgewählten sechs Profile gezeigt werden, welche Auswirkungen eine Übergewichtung der Vorinformationen um den Faktor 10.0 (d.h. alle Gewichte der Pseudobeobachtungen werden mit 10.0 multipliziert) auf die Residuen und die Lösung bei Nutzung des 1DVAR-Verfahrens hat. Dazu wurde in Abbildung 5.7 wieder die Fehlerkurve der Refraktivitäten (rot) (entsprechen den Fehlerannahmen, welche für die Gewichtsmatrix der Refraktivitäten verwendet wurden) dargestellt, die Residuen der Refraktivitäten nach Anwendung des 1DVAR-Verfahrens (grün) und die Residuen der Refraktivitäten nach Verwendung des erweiterten Regularisierungsverfahrens (blau). Sehr deutlich ist erkennbar, dass die Größe der 1DVAR-Residuen im Vergleich zum nicht übergewichteten Fall (Abbildung 5.4) stark zunimmt und häufiger die Fehlerannahmen der Refraktivitäten überschritten werden. Daraus ergibt sich

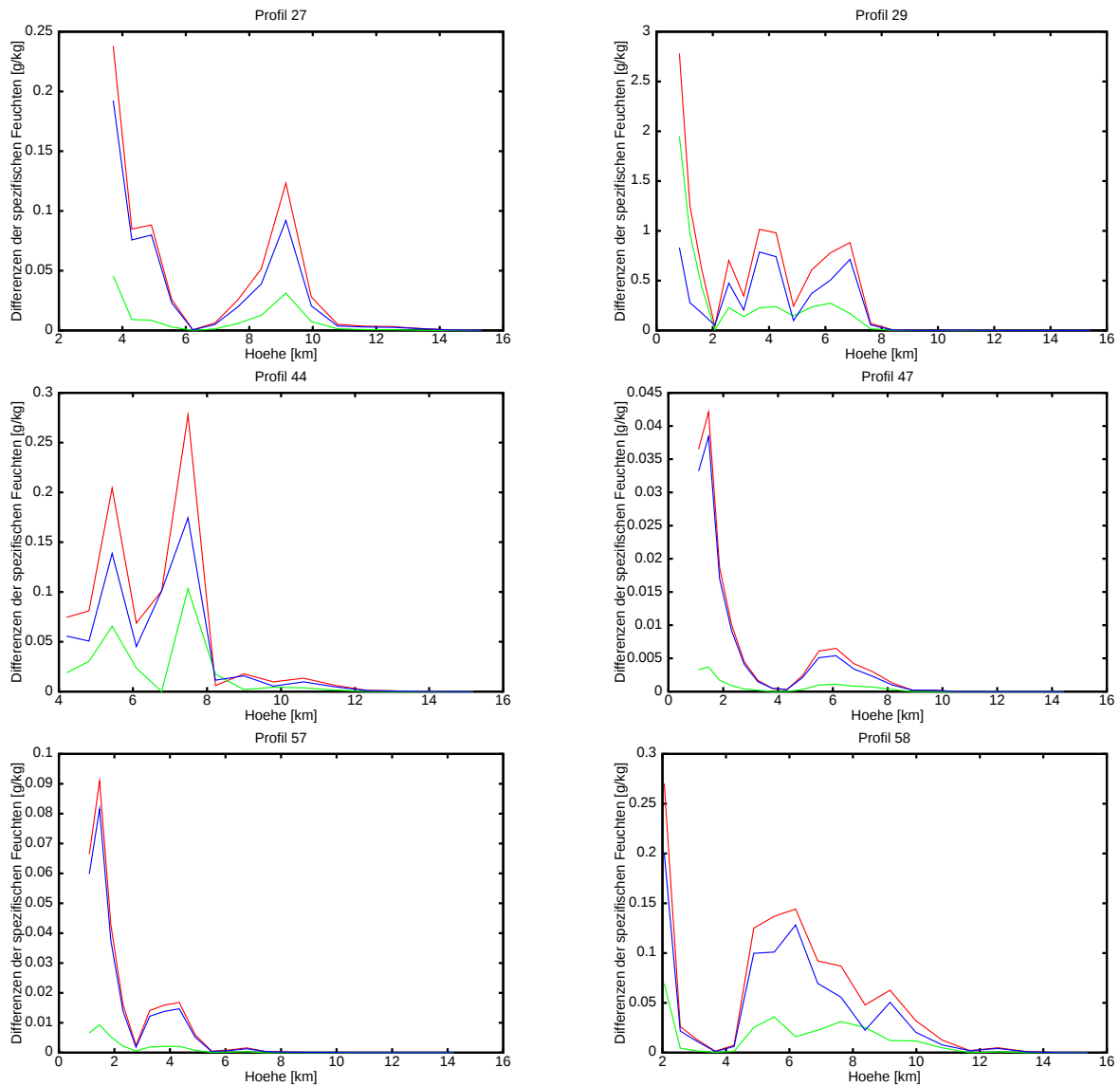


Abbildung 5.8: Differenzenbeträge der spezifischen Feuchten aus unterschiedlichen 1DVAR-Lösungen:

rot ... spezifische Feuchte des ECMWF - 1DVAR-spezifische Feuchte im nicht übergewichteten Fall;

grün ... spezifische Feuchte des ECMWF - 1DVAR-spezifische Feuchte im übergewichteten Fall;

blau ... 1DVAR-spezifische Feuchte im nicht übergewichteten Fall - 1DVAR-spezifische Feuchte im übergewichteten Fall.

unmittelbar für die berechneten 1DVAR-Feuchttemperaturen und -spezifischen Feuchten, dass diese nun sehr ähnlich sind zu den Temperaturen und Feuchten, die als Vorinformationen von ECMWF genutzt wurden¹⁵. Folglich werden aber auch weniger Informationen aus den eigentlich auszuwertenden Refraktivitäten genutzt. Die Ergebnisse für die 1DVAR-spezifischen Feuchten werden in der Abbildung 5.8 gezeigt. Dabei stellt die rote Kurve die Differenzenbeträge der spezifischen Feuchte des ECMWF zur 1DVAR-spezifischen Feuchte im nicht übergewichteten

¹⁵Siehe dazu auch Abschnitt 5.2 (Nachteil des 1DVAR-Verfahrens).

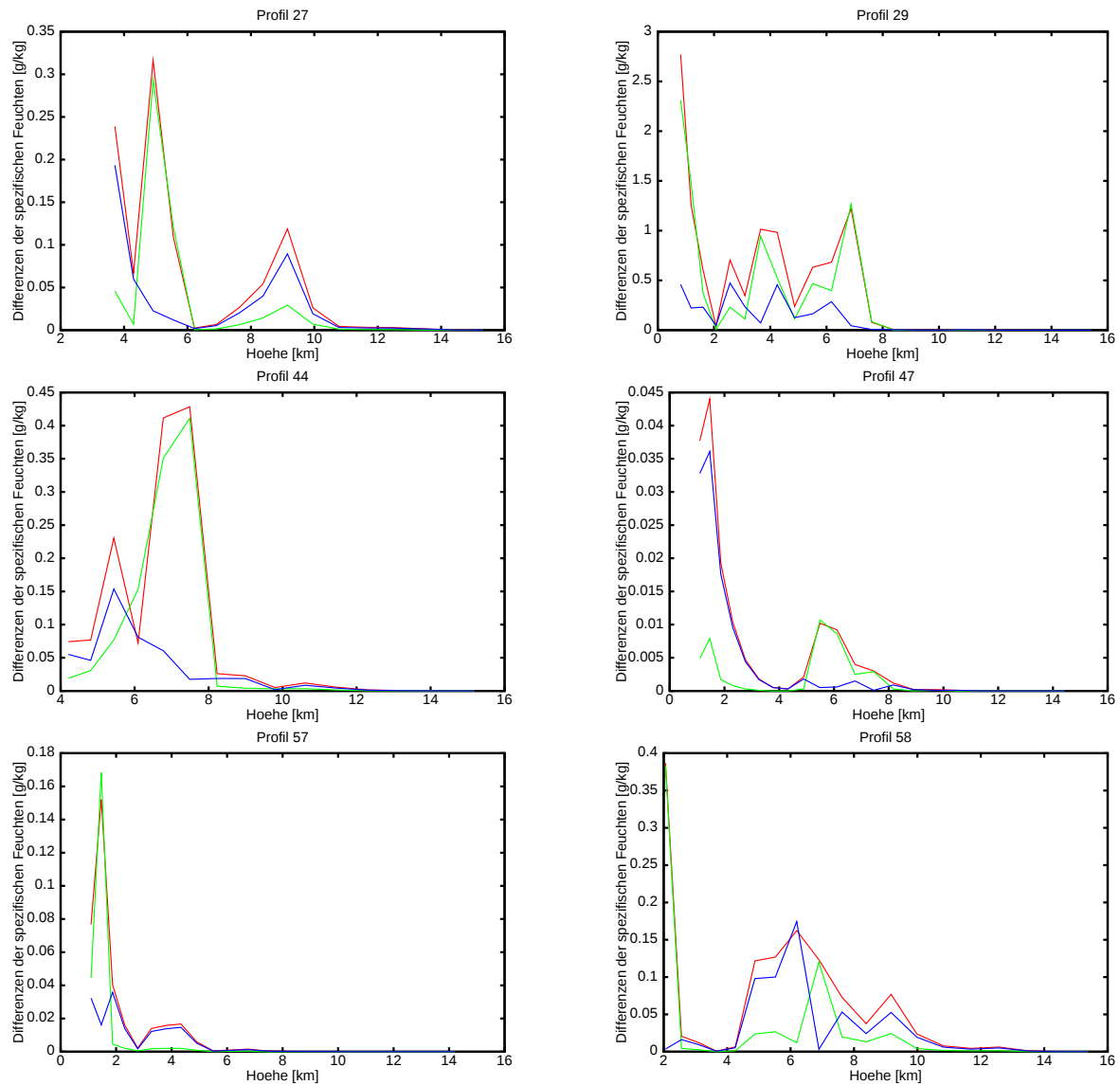


Abbildung 5.9: Differenzenbeträge der spezifischen Feuchten aus unterschiedlichen Lösungen des erweiterten Regularisierungsverfahrens:

rot ... spezifische Feuchte des ECMWF - Lösung des erw. Regularisierungsverfahrens im nicht übergewichteten Fall;

grün ... spezifische Feuchte des ECMWF - Lösung des erw. Regularisierungsverfahrens im übergewichteten Fall;

blau ... Lösung des erw. Regularisierungsverfahrens im nicht übergewichteten Fall - Lösung des erw. Regularisierungsverfahrens im übergewichteten Fall.

Fall dar, die grüne Kurve die Differenzenbeträge der spezifischen Feuchte des ECMWF zur 1DVAR-spezifischen Feuchte im übergewichteten Fall und die blaue Kurve die Differenzenbeträge der 1DVAR-spezifischen Feuchten des nicht übergewichteten und übergewichteten Falls. Die Abweichungen zwischen den berechneten spezifischen Feuchten aus dem übergewichteten und dem nicht übergewichteten Fall sind besonders groß bei den Profilen 27, 29 und 58 mit Abweichungen von 0.8 g/kg bzw. 0.2 g/kg. Die Abweichungen der Feuchttemperaturen zwischen den betrachteten Fällen sind wesentlich kleiner, so dass diese nicht grafisch dargestellt werden. Lediglich beim Profil 47 ist eine maximale Abweichung von 1.9 K zu beobachten. Grundsätzlich

lässt sich somit sagen, dass die 1DVAR-Lösung sehr stark von der Wahl der Gewichtsmatrix der Pseudobeobachtungen abhängig ist. Durch Änderung dieser Gewichtsmatrix lassen sich beliebig große oder kleine Abweichungen der 1DVAR-Lösung zu den Vorinformationen (ECMWF) erzeugen¹⁶.

Verwendet man dagegen das erweiterte Regularisierungsverfahren im übergewichteten Fall, so sind einerseits die Residuen der Refraktivitäten wieder im Bereich der Fehlerannahmen der Refraktivitäten (siehe Abbildung 5.7, blaue Kurve) und andererseits ist die erhaltene Lösung nicht so stark von der Lösung verschieden, welche für den nicht übergewichteten Fall ermittelt wurde. Dies lässt sich gut aus einem Vergleich der ermittelten spezifischen Feuchten in Abbildung 5.9 erkennen. Die rote Kurve stellt die Differenzenbeträge der spezifischen Feuchte des ECMWF zur Lösung des erweiterten Regularisierungsverfahrens im nicht übergewichteten Fall dar, die grüne Kurve die Differenzenbeträge der spezifischen Feuchte des ECMWF zur Lösung des erweiterten Regularisierungsverfahrens im übergewichteten Fall und die blaue Kurve die Differenzenbeträge der Lösungen des erweiterten Regularisierungsverfahrens des nicht übergewichteten und übergewichteten Falls. Die maximale Abweichung zwischen den Lösungen des erweiterten Regularisierungsverfahrens lässt sich für das Profil 29 beobachten mit 0.5 g/kg, welche somit nur etwa halb so groß ist wie die beobachtete 1DVAR-Abweichung von 0.8 g/kg¹⁷. Ebenso geht die maximale Abweichung zwischen den Lösungen für die Feuchttemperatur auf 1.0 K zurück (Profil 27) und ist somit auf etwa die Hälfte der maximalen 1DVAR-Abweichung gesunken. Da die Abweichungen der Lösungen für die Feuchttemperatur hier ebenso sehr klein sind, erfolgt keine grafische Darstellung der Ergebnisse. Bei Verwendung des erweiterten Regularisierungsverfahrens ist es somit möglich, die Lösungsbestimmung durch eine Übergewichtung der Vorinformationen zu stabilisieren, ohne dass dabei die Residuen der auszuwertenden Daten (hier Refraktivitäten) beliebig groß werden und ohne dass die Lösung zu sehr zu der vorhandenen Lösung (hier ECMWF-Modell) ähnlich wird. Besonders in dem Fall, dass genaue Kenntnisse der Genauigkeiten der Vorinformationen nicht vorhanden sind, stellt das erweiterte Regularisierungsverfahren eine gute Alternative zu allgemein verwendeten Regularisierungsverfahren dar.

Da in der Abbildung 5.7 die Residuen der Refraktivitäten für die 1DVAR-Auswertung und die Auswertung mit dem erweiterten Regularisierungsverfahren (Toleranzen = Fehlerannahmen für die Refraktivitäten) dargestellt sind, lässt sich daraus gut erkennen, wie die Umverteilung der Residuen erfolgt, wenn Toleranzen für diese eingeführt werden. An den Stellen, wo die Residuen zu groß waren, werden diese innerhalb der Fehlerannahmen gesenkt. An anderen Stellen, wo die Residuen wesentlich kleiner als die Fehlerannahmen waren, werden diese vergrößert. Am Profil 44 erkennt man z.B. im Höhenbereich von ca. 6.5 km einen starken Anstieg der Residuen und bei ca. 7 km Höhe eine starke Abnahme der Residuen. Nicht erkennbar ist, dass bei allen Profilen vor allem die Residuen in den Höhen von 18 km bis 35 km vergrößert werden. Ebenso nimmt bei allen Profilen die Größe der Residuen der Pseudobeobachtungen (für diese werden keine Toleranzen eingeführt) nach einer Umverteilung der Residuen der Beobachtungen zu. Daraus resultiert gerade die deutliche Abweichung der Lösung des erweiterten Regularisierungsverfahrens zu den Vorinformationen.

Abschließend lässt sich sagen, dass das erarbeitete Regularisierungsverfahren deutliche Vorteile

¹⁶Die absoluten Größen der berechneten 1DVAR-spezifischen Feuchten im übergewichteten und nicht übergewichteten Fall sind im Anhang A.5 in der Abbildung A.3 dargestellt.

¹⁷Die absoluten Größen der mittels des erweiterten Regularisierungsverfahrens berechneten spezifischen Feuchten im übergewichteten und nicht übergewichteten Fall sind im Anhang A.5 in der Abbildung A.4 dargestellt.

gegenüber allgemeinen Regularisierungsverfahren, wie das 1DVAR-Verfahren, bei der Auswertung realer Datensätze aufweist. Zum einen konnte gezeigt werden, dass die Lösung interpretierbar ist und grundsätzlich eine Überprüfung und Verbesserung des vorhandenen Modells erfolgen kann. Dahingegen kann mit dem 1DVAR-Verfahren eigentlich nur eine mehr oder weniger gute Modellreproduktion erfolgen, je nach Gewichtung der Vorinformationen in Bezug auf die eigentlichen Beobachtungen. Diese Tatsache ist besonders deutlich an der erhaltenen Lösung im übergewichteten Fall erkennbar. Zum anderen erwies sich das erweiterte Regularisierungsverfahren selbst bei einer Übergewichtung der Vorinformationen als stabiler bzgl. der Lösung und kann somit auch dann zum Einsatz kommen, wenn die Genauigkeitsinformationen bzgl. der Pseudo-beobachtungen nicht so zuverlässig sind.

6. Schlussbetrachtung

Bei einer Vielzahl geowissenschaftlicher Aufgabenstellungen handelt es sich um inkorrekt gestellte Probleme, welche nur durch die Verwendung zusätzlicher Informationen bzgl. der Lösung gelöst werden können. Die zur Lösung derartiger Probleme zur Anwendung kommenden Regularisierungsverfahren können in drei Gruppen unterteilt werden: in die Regularisierung nach A.N. Tikhonov, das regularisierende Verfahren und die stochastische Regularisierung. Dabei nimmt das regularisierende Verfahren eine Sonderstellung ein, da dieses zur Verbesserung der Konvergenzeigenschaften bei der iterativen Lösungsbestimmung korrekt und inkorrekt gestellter Probleme beiträgt und keine zusätzlichen Informationen bzgl. der Lösung einbezieht. Die Eigenschaften und Unterschiede der Regularisierungsverfahren wurden vor allem vor dem Hintergrund der häufigen Fehlanwendungen dieser Verfahren näher untersucht. Ebenso erfolgte eine analytische Darstellung und Bewertung der einzelnen Regularisierungsverfahren. Die Bestimmung des Regularisierungsparameters (als ein Parameter oder Gewichtsverhältnis) stellt für diese Verfahren den wichtigsten und zugleich kritischsten Punkt dar. Anhand von vier ausgewählten Verfahren zur Bestimmung des „optimalen“ Regularisierungsparameters konnte gezeigt werden, welche unterschiedlichen Anforderungen von diesen Verfahren an die „optimale“ Lösung eines regularisierten Problems gestellt werden.

Da die derzeit vorhandenen Verfahren zur Bestimmung des Regularisierungsparameters eine kontrollierte Beeinflussung der Residuen der eigentlichen Beobachtungen nicht vorsehen, wurde ein Regularisierungsverfahren erarbeitet, bei dem durch eine Einbeziehung von Toleranzbedingungen für die Residuen gerade dies ermöglicht wird. Für die analytische Einführung von Toleranzbedingungen ergaben sich zwei Möglichkeiten: $v_i^2 \leq t_i^2$ und $v_i \leq t_i \ \& \ -v_i \leq t_i$, wobei v_i die Residuen der Beobachtungen bezeichnet und t_i einen Grenzwert (Toleranz) für v_i darstellt. Die notwendigen Formeln, um diese Toleranzbedingungen für ein beliebiges Ausgleichungsproblem aufstellen zu können, wurden dargelegt. Weiterhin wurde ein Algorithmus ausgearbeitet, mit dem die Lösung von um Ungleichheitsbedingungen erweiterten Ausgleichungsproblemen bestimmt werden kann. Da dieser Algorithmus lediglich für Ausgleichungsprobleme mit speziellen Ungleichheitsbedingungen angewendet werden kann, wurden ebenso die Formeln bereitgestellt, die man für die Transformation eines allgemeinen Ausgleichungsproblems mit beliebigen Ungleichheitsbedingungen (z.B. Toleranzbedingungen für Residuen) in den entsprechenden Spezialfall benötigt.

Die Eigenschaften, die Stabilität und die Unterschiede der beiden Varianten der Realisierung der Toleranzbedingung für die Residuen konnten anhand von fünf synthetischen Beispielen abgeleitet werden. Umfangreiche numerische Tests ermöglichten eine Klassifizierung der Eigenschaften der Fläche der zu minimierenden Funktion, welche einen positiven, negativen oder neutralen Einfluss auf die Bestimmbarkeit der korrekten Lösung ausüben. Dazu wurde die Korrektheit der ermittelten Lösungen mittels eines heuristischen Verfahrens kontrolliert. Es stellte sich heraus, dass die Variante 2 ($v_i \leq t_i \ \& \ -v_i \leq t_i$) der Realisierung der Toleranzbedingung gegenüber der Variante 1 ($v_i^2 \leq t_i^2$) wesentlich zuverlässiger die korrekte Lösung der betrachteten Beispiele lieferte. Dies lässt sich darauf zurückführen, dass die partiellen Ableitungen bei der Linearisierung der Toleranzbedingungen denen des Funktionalmodells entsprechen. Insgesamt liegt somit ein stabiler und zuverlässiger Algorithmus zur Lösung von um Toleranzbedingungen erweiterten Ausgleichungsproblemen vor.

Die praktische Anwendbarkeit des erarbeiteten Regularisierungsverfahrens unter Berücksichtigung von Toleranzenbedingungen für Residuen konnte anhand der Bestimmung von Atmosphärenparametern aus CHAMP-Radiokkultations-Refraktivitäten des GFZ gezeigt werden. Im Vergleich zu allgemein üblichen Regularisierungsverfahren war es damit möglich, gezielt Informationen aus den Beobachtungen über die Atmosphärenparameter zu gewinnen und somit Aussagen über Verbesserungsmöglichkeiten des vorhandenen Modells zu treffen. Vor allem für die Polargebiete, wo die bestehenden Atmosphärenmodelle stark unterschiedliche Informationen über den Zustand der Atmosphäre liefern, bietet die Regularisierung mit Residuentoleranzen die Möglichkeit der Modellüberprüfung und -verbesserung. Ebenso erwies sich das erarbeitete Verfahren als vorteilhaft für den Fall, dass die zusätzlichen Informationen innerhalb des Ausgleichsprozesses übergewichtet wurden. Die ermittelte Lösung wurde wesentlich geringer von der Übergewichtung beeinflusst als die Lösung, die mittels eines allgemein üblichen Regularisierungsverfahrens bestimmt wurde. Da in der Praxis häufig die Genauigkeitsangaben bzgl. der zusätzlichen Informationen zu optimistisch sind oder gar fehlen, liefern die allgemeinen Regularisierungsverfahren meist Lösungen, die keine oder nur wenig neue Informationen aus den aktuellen Daten enthalten.

Die Einbeziehung von Ungleichheitsbedingungen innerhalb einer Ausgleichung kann sicherlich auch in einem erweiterten Sinne nützlich sein, d.h. nicht nur, um Toleranzen für Residuen zu berücksichtigen. So hat man z.B. im straßenbaulichen Bereich oft das Problem, dass eine neu zu planende Trasse bestimmte festgesetzte Punkte enthalten muss und die entsprechenden Trassierungselemente zwischen diesen Punkten aus straßenbaulicher Sicht optimal festgelegt werden müssen. Dabei liegen für die einzelnen unterschiedlichen Trassierungselemente (z.B. Kreis und Klottoide) Richtlinien in Form von Ungleichheitsbedingungen vor, mit denen die Größe oder die Form der Trassierungselemente begrenzt wird. Eine Verwendung der dargelegten Transformationen und des erarbeiteten Lösungsalgorithmus ist dafür denkbar, siehe z.B. [Klumpp 1973].

Literaturverzeichnis

- [Beyerle et al. 2003] Beyerle, G.; Gorbunov, M.E.; Ao, C.O.: *Simulation studies of GPS radio occultation measurements*. Radio Science, Vol. 38, No. 5.
- [Beyerle et al. 2004] Beyerle, G.; König-Langlo, G.; Wickert, J.; Schmidt, T.; Heise, S.; Kaschenz, J.: *An analysis of the negative refractivity bias detected in GPS radio occultation data: Results from simulation studies, aerological soundings and CHAMP observations*. Vortrag beim Joint CHAMP/GRACE Science Meeting, 6.-8. Juli 2004, Potsdam.
- [Caspary 1988] Caspary, W.: *Fehlerverteilungen, Methode der kleinsten Quadrate und robuste Alternativen*. ZfV 3/1988, S.123-133.
- [Collard und Healy 2002] Collard, A.D.; Healy, S.B.: *The combined impact of future space-based atmospheric sounding instruments on numerical weather prediction*. Q.J.Royal Meteorological Society, Volume 128, S.1-16.
- [Engl, Hanke und Neubauer 1996] Engl, H.W.; Hanke, M.; Neubauer, A.: *Regularization of Inverse Problems*. Kluwer Academic Publishers, Dordrecht / Boston / London.
- [Fedotov 1990] Федотов, А.М.: Некорректные задачи со случайными ошибками в данных. (*Übersetzung: Inkorrekt gestellte Aufgaben mit zufälligen Fehlern in den Daten.*) Наука, Новосибирск.
- [Fierro et al. 1997] Fierro, R.D.; Golub, G.H.; Hansen, P.C.; O'Leary, D.P.: *Regularization by truncated total least squares*. SIAM Journal on Scientific Computing, 18, Nr. 4, S.1223-1241.
- [Fischbach 1965] Fischbach, F.F.: *A satellite method for pressure and temperature below 24 km*. Bull.Am.Meteorol.Soc., Vol. 46, No. 9, S.528-532.
- [Fjeldbo et al. 1967] Fjeldbo, G.; Kliore, A.; Levy, G.S.; Cain, D.L.; Rasool, S.I.: *Atmosphere and Ionosphere of Venus from the Mariner V S-Band radio occultation measurement*. Science, S.1683-1688.
- [Fjeldbo und Eshleman 1968] Fjeldbo, G.; Eshleman, R.: *The atmosphere of Mars analyzed by integral inversion of the Mariner IV occultation data*. Planet.Space Sci., Volume 16, S.1035-1059.
- [Fjeldbo et al. 1970] Fjeldbo, G.; Kliore, A.J.; Eshleman, R.: *The neutral atmosphere of Venus as studied with the Mariner V radio occultation experiments*. The Astronomical Journal, Volume 76, Number 2, S.123-140.
- [Friedrich, Hofmann und Tautenhahn 1979] Friedrich, V.; Hofmann, B.; Tautenhahn, U.: *Möglichkeiten der Regularisierung bei der Auswertung von Meßdaten*. Wissenschaftliche Zeitschrift der Technischen Hochschule Karl-Marx-Stadt, VEB Kongreß- und Werbedruck Oberlungwitz, 10/1979.
- [Golub et al. 1979] Golub, G.H.; Heath, M.; Wahba, G.: *Generalized Cross-Validation as a Method for Choosing a Good Ridge Parameter*. Technometrics, Vol. 21, No. 2, S.215-223.

- [Golub et al. 1999] Golub, G.H.; Hansen, P.C.; O’Leary, D.P.: *Tikhonov regularization and total least squares*. SIAM Journal on Matrix Analysis and Applications, 21, Nr. 1, S.185-194.
- [Golub und van Loan 1980] Golub, G.H.; van Loan, C.F.: *An analysis of the total least squares problem*. SIAM Journal on Numerical Analysis, 17, Nr. 6, S.883-893.
- [Gorbunov und Sokolovskiy 1993] Gorbunov, M.E.; Sokolovskiy, S.V.: *Remote sensing of refractivity from space for global observations of atmospheric parameters*. Max-Planck-Institut für Meteorologie, Report No. 119, Hamburg.
- [Gotthardt 1968] Gotthardt, E.: *Einführung in die Ausgleichungsrechnung*. Herbert Wichmann Verlag, Karlsruhe.
- [Großmann 1953] Großmann, W.: *Grundzüge der Ausgleichungsrechnung nach der Methode der kleinsten Quadrate nebst Anwendungen in der Geodäsie*. Springer-Verlag, Berlin-Göttingen-Heidelberg.
- [Gruber 2000] Gruber, Th.: *Hochauflösende Schwerefeldbestimmung aus Kombination von terrestrischen Messungen und Satellitendaten über Kugelfunktionen*. Dissertation, Scientific Technical Report STR00/16, GFZ Potsdam.
- [Hadamard 1932] Hadamard, J.: *Le problème de Cauchy et les équations aux dérivées partielles linéaires hyperboliques*. (Übersetzung: *Das Problem von Cauchy und partielle lineare hyperbolische Differentialgleichungen*.) Hermann et Cie Éditeurs, Paris.
- [Hansen und O’Leary 1993] Hansen, P.C.; O’Leary, D.P.: *The use of the L-curve in the regularization of discrete ill-posed problems*. SIAM J.Sci.Comput., S.1487-1503.
- [Healy und Eyre 2000] Healy, S.B.; Eyre, J.R.: *Retrieving temperature, water vapour and surface pressure information from refractive-index profiles derived by radio occultation: A simulation study*. Q.J.R.Meteorol.Soc., Volume 126, S.1661-1683.
- [Helmert 1872] Helmert, F.R.: *Die Ausgleichungsrechnung nach der Methode der kleinsten Quadrate*. Verlag B.G.Teubner, Leipzig.
- [Ilk 1993] Ilk, K.H.: *Regularization for high resolution gravity field recovery by future satellite techniques*. Proceedings der Internationalen Konferenz in Potsdam, 30. August - 3. September, Mathematical Research - Inverse Problems: Principles and Applications in Geophysics, Technology, and Medicine, Volume 74, Akademie Verlag.
- [Jensen et al. 2003] Jensen, A.S.; Lohmann, M.; Benzon, H.-H.; Nielsen, A.: *Full spectrum inversion of radio occultation signals*. Radio Science, Vol.38, No.3.
- [Kaschenz 2003] Kaschenz, J.: *Datenmodellierung mittels verschiedener Frequenzreihen*. ZfV 4/2003, S.260-265.
- [Kilian 2002] Kilian, N.: *Zur Ausgleichung nach verschiedenen Normen*. Diplomarbeit an der Technischen Universität Berlin am Institut für Geodäsie und Geoinformationstechnik, Berlin.
- [Klumpp 1973] Klumpp, R.: *Achsberechnungen mit Hilfe der Ausgleichungs- und Optimierungsrechnung*. Deutsche Geodätische Kommission, Reihe C, Heft Nr. 194, München.

- [Koch und Kusche 2002] Koch, K.-R.; Kusche, J.: *Regularization of geopotential determination from satellite data by variance components*. Journal of Geodesy, Vol. 76, S.259-268.
- [Krakauer et al. 2004] Krakauer, N.Y.; Schneider, T.; Randerson, J.T.; Olsen, S.C.: *Linear inversion methods and generalized cross-validation*. Online supplement zu „Using generalized cross-validation to select parameters in inversions for regional carbon fluxes“.
- [Kusche 2002] Kusche, J.: *Inverse Probleme bei der Gravitationsfeldbestimmung mittels SST- und SGG-Satellitenmissionen*. Deutsche Geodätische Kommission, Reihe C, Heft Nr. 548, München.
- [Lawson und Hanson 1974] Lawson, C.L.; Hanson, R.J.: *Solving Least Squares Problems*. Society for Industrial and Applied Mathematics, Philadelphia, 1995, unveränderte Neuauflage von 1974.
- [Levenberg 1944] Levenberg, K.: *A method for the solution of certain non-linear problems in least squares*. Quart.Appl.Math. , S.164-168.
- [Louis 1989] Louis, A.K.: *Inverse und schlecht gestellte Probleme*. B.G. Teubner Verlag, Stuttgart.
- [Marquardt 1963] Marquardt, D.W.: *An algorithm for least-squares estimation of nonlinear parameters*. SIAM J.Appl.Math. , S.431-441.
- [Marx 2003] Marx, Ch.: *Untersuchungen zur robusten und resistenten Parameterschätzung*. Diplomarbeit an der Technischen Universität Berlin am Institut für Geodäsie und Geoinformationstechnik, Berlin.
- [Mautz 2001] Mautz, R.: *Zur Lösung nichtlinearer Ausgleichungsprobleme bei der Bestimmung von Frequenzen in Zeitreihen*. Deutsche Geodätische Kommission, Reihe C, Heft Nr. 532, München.
- [Melbourne et al. 1994] Melbourne, W.G.; Davis, E.S.; Duncan, C.B.; Hajj, G.A.; Hardy, K.R.; Kursinsky, E.R.; Meehan, T.K.; Young, L.E.; Yunck, T.P.: *The application of spaceborne GPS to atmospheric limb sounding and global change monitoring*. JPL Publication 94-18, California Institut of Technology, Pasadena.
- [Parker 1994] Parker, R.L.: *Geophysical Inverse Theory*. Princeton University Press, Princeton / New Jersey.
- [Petrović 2003] Petrović, S.: *Parameterschätzung für unvollständige funktionale Modelle in der Geodäsie*. Deutsche Geodätische Kommission, Reihe C, Heft Nr. 563, München.
- [Phillips 1962] Phillips, D.L.: *A technique for the numerical solution of certain integral equations of the first kind*. Journal ACM., S.84-97.
- [Press et al. 1992] Press, W.H.; Teukolsky, S.A.; Vetterling, W.T.; Flannery, B.P.: *Numerical Recipes in Fortran 77 - The Art of Scientific Computing*. Cambridge University Press, 2nd edition.
- [Reigber et al. 2003] Reigber, C.; Lühr, H.; Schwintzer, P. (Eds.): *First CHAMP Mission Results for Gravity, Magnetic and Atmospheric Studies*. Springer-Verlag.

- [Reigber et al. 2005] Reigber, C.; Lühr, H.; Schwintzer, P.; Wickert, J. (Eds.): *Earth Observation with CHAMP: Results from Three Years in Orbit*. Springer-Verlag.
- [Reißmann 1968] Reißmann, G.: *Die Ausgleichungsrechnung - Grundlagen und Anwendungen in der Geodäsie*. VEB Verlag für Bauwesen, Berlin.
- [Rodgers 2000] Rodgers, C.D.: *Inverse Methods For Atmospheric Soundings - Theory and Practice*. World Scientific Publishing Co.Pte.Ltd., Series on Atmospheric, Oceanic and Planetary Physics, Volume 2, Singapore.
- [Rummel et al. 1979] Rummel, R.; Schwarz, K.-P.; Gerstl, M.: *Least Squares Collocation and Regularization*. Bull. Geod. 53, S.343-361.
- [Scales 1985] Scales, L.E.: *Introduction to Non-Linear Optimization*. MacMillan Publishers LTD, London and Basingstoke.
- [Schwetlick 1973] Schwetlick, H.: *Zur Konvergenz regularisierter Gauß-Newton-Verfahren*. Журнал вычислительной математики и математической физики, Том 13, No. 6, S.1371-1382.
- [Schwetlick 1979] Schwetlick, H.: *Numerische Lösung nichtlinearer Gleichungen*. VEB Deutscher Verlag der Wissenschaften, Berlin.
- [Schwintzer 1990] Schwintzer, P.: *Sensitivity Analysis in Least Squares Gravity Field Modelling Redundancy Decomposition of Stochastic A Priori Information*. Interner Bericht des Deutschen Geodätischen Forschungsinstituts, München.
- [Seeber 1993] Seeber, G.: *Satellite Geodesy - Foundations, Methods, and Applications*. Walter de Gruyter Verlag, Berlin / New York.
- [Somogyi und Závoti 1990] Somogyi, J.; Závoti, J.: *Die Anwendung der L_p -Norm-Schätzung für Ähnlichkeitstransformationen*. ZfV 1/1990, S.28-36.
- [Steiner 1998] Steiner, A.K.: *High resolution sounding of key climate variables using the radio occultation technique*. Wissenschaftlicher Bericht No. 3/1998, Institut für Meteorologie und Geophysik, Karl-Franzens-Universität Graz.
- [Stichler 1984] Stichler, S.: *Zur Anwendung robuster statistischer Methoden in der Geodäsie*. Vermessungstechnik, 32. Jg. , Heft 11, S.379-381.
- [Tarantola und Valette 1982] Tarantola, A.; Valette, B.: *Generalized Nonlinear Inverse Problems Solved Using the Least Squares Criterion*. Reviews Of Geophysics And Space Physics, Vol. 20, No. 2, S.219-232.
- [Tikhonov 1943] Tikhonov, A.N.: *On the stability of inverse problems*. Dokl.Acad.Nauk USSR, S.195-198.
- [Tikhonov 1963a] Тихонов, А.Н.: О решении некорректно поставленных задач и методе регуляризации (*Übersetzung: Über die Lösung inkorrekt gestellter Aufgaben und die Methode der Regularisierung.*) Доклады Академии наук СССР, Том 151, No.3, S.501-504.
- [Tikhonov 1963b] Тихонов, А.Н.: О регуляризации некорректно поставленных задач (*Übersetzung: Über die Regularisierung inkorrekt gestellter Aufgaben.*) Доклады Академии наук СССР, Том 153, No.1, S.49-52.

- [Tikhonov und Arsenin 1977] Tikhonov, A.N.; Arsenin, V.Y.: *Solutions of ill-posed problems*. englische Übersetzung von „Методы решения некорректных задач“, V.H. Winston & Sons, Washington D.C..
- [Tikhonov et al. 1995] Tikhonov, A.N.; Goncharsky, A.V.; Stepanov, V.V.; Yagola, A.G.: *Numerical Methods for the Solution of Ill-Posed Problems*. Kluwer Academic Publishers, Dordrecht / Boston / London.
- [Twomey 1963] Twomey, S.: *On the numerical solution of Fredholm integral equation of the first kind by the inversion of the linear system produced by quadrature*. J.Ass.Comput.Mach., S.97-101.
- [Twomey 1965] Twomey, S.: *The application of numerical filtering to the solution of integral equations encountered in indirect sensing measurements*. Journal of The Franklin Institute, Vol. 279, No. 2, S.95-109.
- [Van Huffel und Zha 1993] Van Huffel, S.; 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.
- [Wickert 2002] Wickert, J.: *Das CHAMP-Radiookkultationsexperiment: Algorithmen, Prozessierungssystem und erste Ergebnisse*. Dissertation, Scientific Technical Report STR02/07, Potsdam.
- [Wickert 2004] Wickert, J.: *Comparison of vertical refractivity and temperature profiles from CHAMP with radiosonde measurements*. Scientific Technical Report STR04/19, Potsdam.
- [Wickert et al. 2005] Wickert, J.; Beyerle, G.; König, R.; Heise, S.; Grunwaldt, L.; Michalak, G.; Reigber, Ch.; Schmidt, T.: *GPS radio occultation with CHAMP and GRACE: A first look at a new and promising satellite configuration for global atmospheric sounding*. Annales Geophysicae, Vol. 23, S.653-658.
- [Wilkinson 1969] Wilkinson, J.H.: *Rundungsfehler*. Springer Verlag, Berlin-Heidelberg.
- [Wolf 1979] Wolf, H.: *Ausgleichsrechnung II - Aufgaben und Beispiele zur praktischen Anwendung*. Ferd. Dümmlers Verlag, Dümmlerbuch 7836, Bonn.

A. Anhang

A.1 Funktionalmodelle und Daten der Beispiele 1-5

Beispiel 1:

Beobachtungsgleichungen: $y_i + v_i = ax_i + b$, entspricht der Anpassung einer Geraden an drei gegebene Ordinaten

Daten:

$$x = \begin{bmatrix} 1.5 \\ 4.0 \\ 7.0 \end{bmatrix} \quad y = \begin{bmatrix} 2.5 \\ 2.5 \\ 5.5 \end{bmatrix} \quad (\text{A.1})$$

Unbekannte: a, b

Näherungslösung für die erste Iteration: $a_0 = 0.53$ und $b_0 = 1.04$

Pseudobeobachtungen: $\bar{a} = 1.0$ und $\bar{b} = 2.0$

Beispiel 2:

Beobachtungsgleichungen:

$$\begin{aligned} y_1 + v_1 &= a + b \\ y_2 + v_2 &= a - b \\ y_3 + v_3 &= 4a \end{aligned}$$

Daten:

$$y = \begin{bmatrix} 3.1 \\ -0.9 \\ 3.8 \end{bmatrix} \quad (\text{A.2})$$

Unbekannte: a, b

Näherungslösung für die erste Iteration: $a_0 = 1.1$ und $b_0 = 1.9$

Pseudobeobachtungen: $\bar{a} = 2.0$ und $\bar{b} = 4.0$

Beispiel 3:

Beobachtungsgleichungen: $y_i + v_i = a^2 x_i + b^3$

Daten:

$$x = \begin{bmatrix} 0.5 \\ 1.0 \\ 1.5 \\ 2.0 \end{bmatrix} \quad y = \begin{bmatrix} 1.2 \\ 2.3 \\ 3.3 \\ 4.4 \end{bmatrix} \quad (\text{A.3})$$

Unbekannte: a, b

Näherungslösung für die erste Iteration: $a_0 = 1.4$ und $b_0 = 0.5$

Pseudobeobachtungen: $\bar{a} = 2.0$ und $\bar{b} = 3.0$

Beispiel 4:

Beobachtungsgleichungen: $y_i + v_i = ax_i + a^2 + x_i^2 b$, entspricht der Anpassung von vier Ordinaten an eine spezielle Parabel

Daten:

$$x = \begin{bmatrix} 0.5 \\ 1.0 \\ 1.5 \\ 2.0 \end{bmatrix} \quad y = \begin{bmatrix} 0.5 \\ 1.9 \\ 4.1 \\ 7.3 \end{bmatrix} \quad (\text{A.4})$$

Unbekannte: a, b

Näherungslösung für die erste Iteration: $a_0 = 0.1$ und $b_0 = 1.7$

Pseudobeobachtungen: $\bar{a} = 3.0$ und $\bar{b} = 2.0$

Beispiel 5:

Beobachtungsgleichungen:

$$\begin{aligned} l_a + v_a &= a \\ l_b + v_b &= b \\ l_c + v_c &= \sqrt{a^2 + b^2} \\ l_\alpha + v_\alpha &= \arctan(b/a) \\ l_\beta + v_\beta &= \arctan(a/b) \end{aligned}$$

entspricht der Bestimmung der Katheten eines rechtwinkligen Dreiecks aus drei Strecken- und zwei Winkelmessungen

Daten:

$$\begin{bmatrix} l_a \\ l_b \\ l_c \\ l_\alpha \\ l_\beta \end{bmatrix} = \begin{bmatrix} 30.002 \text{ m} \\ 49.9971 \text{ m} \\ 58.311 \text{ m} \\ 1.030346827 \text{ rad} \\ 0.5404645 \text{ rad} \end{bmatrix} \quad (\text{A.5})$$

Unbekannte: a, b

Näherungslösung für die erste Iteration: $a_0 = 29.5 \text{ m}$ und $b_0 = 50.3 \text{ m}$

Pseudobeobachtungen: $\bar{a} = 28.0 \text{ m}$ und $\bar{b} = 52.0 \text{ m}$

A.2 Liste der verwendeten Radiookkultations-Inputdateien

Folgende Dateien waren Grundlage für die Berechnungen im Abschnitt 5.4 (Quelle: CHAMP-ISDC):

Nr.	Dateiname
01.	CH-AI-3-ATM+2004_001_21_0229_mobn_005.dat
02.	CH-AI-3-ATM+2004_002_20_0197_tidb_005.dat
03.	CH-AI-3-ATM+2004_003_20_0192_mobn_005.dat
04.	CH-AI-3-ATM+2004_003_23_0220_kokb_005.dat
05.	CH-AI-3-ATM+2004_004_22_0207_cro1_005.dat
06.	CH-AI-3-ATM+2004_005_20_0222_pimo_005.dat
07.	CH-AI-3-ATM+2004_005_23_0249_sant_005.dat
08.	CH-AI-3-ATM+2004_006_23_0247_sant_005.dat
09.	CH-AI-3-ATM+2004_007_19_0192_ous2_005.dat
10.	CH-AI-3-ATM+2004_007_22_0222_ous2_005.dat
11.	CH-AI-3-ATM+2004_008_19_0191_guam_005.dat
12.	CH-AI-3-ATM+2004_008_23_0227_sant_005.dat
13.	CH-AI-3-ATM+2004_009_20_0211_sutm_005.dat
14.	CH-AI-3-ATM+2004_010_21_0223_cro1_005.dat
15.	CH-AI-3-ATM+2004_010_23_0241_fair_005.dat
16.	CH-AI-3-ATM+2004_011_15_0155_sutm_005.dat
17.	CH-AI-3-ATM+2004_012_22_0230_bogt_005.dat
18.	CH-AI-3-ATM+2004_013_23_0231_sutm_005.dat
19.	CH-AI-3-ATM+2004_014_22_0230_nya2_005.dat
20.	CH-AI-3-ATM+2004_016_22_0233_ban2_005.dat
21.	CH-AI-3-ATM+2004_017_23_0229_mobn_005.dat
22.	CH-AI-3-ATM+2004_018_22_0236_ous2_005.dat
23.	CH-AI-3-ATM+2004_019_20_0199_mkea_005.dat
24.	CH-AI-3-ATM+2004_020_23_0230_bogt_005.dat
25.	CH-AI-3-ATM+2004_021_22_0230_gold_005.dat
26.	CH-AI-3-ATM+2004_022_23_0226_hrao_005.dat
27.	CH-AI-3-ATM+2004_024_23_0231_ous2_005.dat
28.	CH-AI-3-ATM+2004_025_23_0218_sant_005.dat
29.	CH-AI-3-ATM+2004_026_22_0220_nya2_005.dat
30.	CH-AI-3-ATM+2004_027_21_0233_usud_005.dat
31.	CH-AI-3-ATM+2004_028_22_0229_mkea_005.dat
32.	CH-AI-3-ATM+2004_029_22_0226_kokb_005.dat
33.	CH-AI-3-ATM+2004_031_22_0231_pimo_005.dat
34.	CH-AI-3-ATM+2004_032_22_0238_ous2_005.dat
35.	CH-AI-3-ATM+2004_033_22_0222_fair_005.dat
36.	CH-AI-3-ATM+2004_034_20_0212_cro1_005.dat
37.	CH-AI-3-ATM+2004_035_20_0210_guam_005.dat
38.	CH-AI-3-ATM+2004_036_21_0217_sutm_005.dat
39.	CH-AI-3-ATM+2004_037_22_0219_mkea_005.dat
40.	CH-AI-3-ATM+2004_038_21_0211_guam_005.dat
41.	CH-AI-3-ATM+2004_039_23_0227_hrao_005.dat
42.	CH-AI-3-ATM+2004_040_22_0220_potm_005.dat
43.	CH-AI-3-ATM+2004_041_20_0206_ous2_005.dat

44. CH-AI-3-ATM+2004_042_23_0257_tidb_005.dat
45. CH-AI-3-ATM+2004_043_23_0236_nya2_005.dat
46. CH-AI-3-ATM+2004_044_19_0196_gold_005.dat
47. CH-AI-3-ATM+2004_046_22_0161_godf_005.dat
48. CH-AI-3-ATM+2004_047_22_0193_sant_005.dat
49. CH-AI-3-ATM+2004_048_23_0227_nya2_005.dat
50. CH-AI-3-ATM+2004_049_21_0213_madr_005.dat
51. CH-AI-3-ATM+2004_050_22_0213_tidb_005.dat
52. CH-AI-3-ATM+2004_052_22_0231_godf_005.dat
53. CH-AI-3-ATM+2004_053_22_0209_hrao_005.dat
54. CH-AI-3-ATM+2004_054_21_0207_tidb_005.dat
55. CH-AI-3-ATM+2004_055_22_0205_ban2_005.dat
56. CH-AI-3-ATM+2004_056_22_0203_nya2_005.dat
57. CH-AI-3-ATM+2004_057_21_0201_gold_005.dat
58. CH-AI-3-ATM+2004_059_23_0228_lpgm_005.dat
59. CH-AI-3-ATM+2004_060_22_0206_ous2_005.dat
60. CH-AI-3-ATM+2004_061_23_0214_mobn_005.dat
61. CH-AI-3-ATM+2004_064_19_0162_cro1_005.dat
62. CH-AI-3-ATM+2004_065_21_0204_usud_005.dat
63. CH-AI-3-ATM+2004_066_22_0202_ban2_005.dat

A.3 Globale Genauigkeitsschätzungen für die Radiookkultations-Refraktivitäten

Nachfolgend sind die globalen Genauigkeitsschätzungen für die GFZ Radiookkultations-Refraktivitäten für Höhen von 0.4 km bis 35 km mit einer vertikalen Auflösung von 200 m aufgelistet, [Wickert 2004]. Diese Abschätzungen entstammen aus einem Vergleich der GFZ-Refraktivitäten (berechnet mit der Programmversion 004 unter Verwendung der Full Spectrum Inversion) mit 6-stündigen ECMWF-Analysen. Somit weisen diese Fehlerschätzungen den Nachteil auf, dass zum einen für deren Berechnung die ECMWF-Analysen als fehlerfrei betrachtet werden, was sicher nicht gerechtfertigt ist, und zum anderen hier keine zeitliche und räumliche Abhängigkeit für die Fehlerschätzung berücksichtigt wird.

In der ersten Spalte befindet sich die Höhenangabe in km, in der zweiten Spalte der Bias in Prozent und in der dritten Spalte die Standardabweichung in Prozent. Der Gesamtfehler einer beliebigen Refraktivität N berechnet sich dann wie folgt:

$$Fehler = \frac{Bias \cdot N}{100} + \frac{Standardabweichung \cdot N}{100}. \quad (A.6)$$

So ergibt sich z.B. für die 4. Zeile der nachfolgenden Tabelle (Höhe 1 km):

$$Fehler = \frac{-1.141 \cdot N}{100} + \frac{2.653 \cdot N}{100}. \quad (A.7)$$

Ist der somit berechnete Fehler einer Refraktivität kleiner als 0.1 N-Einheiten, so wird der entsprechende Fehler mit 0.1 N-Einheiten festgesetzt.

Hoehe (km)	Bias (%)	Standardabweichung (%)
0.400	-1.045	2.362
0.600	-1.045	2.362
0.800	-1.257	2.732
1.000	-1.141	2.653
1.200	-1.107	2.711
1.400	-1.116	2.682
1.600	-0.919	2.564
1.800	-0.746	2.384
2.000	-0.638	2.316
2.200	-0.551	2.323
2.400	-0.403	2.268
2.600	-0.332	2.291
2.800	-0.402	2.357
3.000	-0.326	2.278
3.200	-0.323	2.299
3.400	-0.316	2.283
3.600	-0.257	2.195
3.800	-0.211	2.146
4.000	-0.195	2.040
4.200	-0.162	1.976
4.400	-0.131	1.935
4.600	-0.051	1.838
4.800	-0.014	1.798
5.000	0.032	1.760
5.200	0.072	1.766
5.400	0.113	1.545
5.600	0.118	1.526
5.800	0.122	1.428
6.000	0.129	1.378
6.200	0.154	1.264

6.400	0.144	1.208
6.600	0.126	1.147
6.800	0.122	1.156
7.000	0.086	1.082
7.200	0.048	0.999
7.400	0.031	0.987
7.600	0.059	1.002
7.800	0.074	0.943
8.000	0.079	0.909
8.200	0.075	0.877
8.400	0.086	0.854
8.600	0.091	0.820
8.800	0.095	0.787
9.000	0.116	0.786
9.200	0.109	0.740
9.400	0.108	0.736
9.600	0.131	0.747
9.800	0.145	0.785
10.000	0.144	0.765
10.200	0.150	0.791
10.400	0.170	0.818
10.600	0.185	0.839
10.800	0.198	0.881
11.000	0.194	0.882
11.200	0.223	0.910
11.400	0.250	0.935
11.600	0.252	0.948
11.800	0.245	0.938
12.000	0.225	0.928
12.200	0.229	0.926
12.400	0.233	0.931
12.600	0.251	0.942
12.800	0.239	0.945
13.000	0.216	0.955
13.200	0.193	0.947
13.400	0.165	0.929
13.600	0.149	0.924
13.800	0.106	0.917
14.000	0.048	0.911
14.200	-0.000	0.919
14.400	-0.038	0.917
14.600	-0.084	0.925
14.800	-0.144	0.932
15.000	-0.194	0.952
15.200	-0.206	0.939
15.400	-0.220	0.932
15.600	-0.260	0.938
15.800	-0.299	0.941
16.000	-0.326	0.942
16.200	-0.294	0.920
16.400	-0.290	0.935
16.600	-0.327	0.956
16.800	-0.372	0.993
17.000	-0.374	1.001
17.200	-0.334	0.964
17.400	-0.305	0.926
17.600	-0.287	0.906
17.800	-0.270	0.895
18.000	-0.253	0.879
18.200	-0.251	0.873
18.400	-0.255	0.869

18.600	-0.257	0.876
18.800	-0.240	0.849
19.000	-0.216	0.837
19.200	-0.220	0.814
19.400	-0.232	0.789
19.600	-0.248	0.772
19.800	-0.259	0.764
20.000	-0.264	0.757
20.200	-0.266	0.750
20.400	-0.286	0.742
20.600	-0.313	0.745
20.800	-0.338	0.758
21.000	-0.355	0.779
21.200	-0.369	0.807
21.400	-0.373	0.837
21.600	-0.383	0.858
21.800	-0.398	0.864
22.000	-0.404	0.867
22.200	-0.406	0.872
22.400	-0.401	0.872
22.600	-0.387	0.864
22.800	-0.368	0.856
23.000	-0.359	0.842
23.200	-0.352	0.828
23.400	-0.346	0.813
23.600	-0.330	0.804
23.800	-0.306	0.801
24.000	-0.276	0.799
24.200	-0.246	0.801
24.400	-0.220	0.805
24.600	-0.199	0.819
24.800	-0.175	0.838
25.000	-0.142	0.848
25.200	-0.103	0.859
25.400	-0.064	0.873
25.600	-0.028	0.888
25.800	-0.006	0.899
26.000	0.010	0.912
26.200	0.019	0.929
26.400	0.028	0.951
26.600	0.040	0.967
26.800	0.050	0.964
27.000	0.055	0.981
27.200	0.045	1.005
27.400	0.026	1.027
27.600	0.006	1.059
27.800	-0.013	1.093
28.000	-0.033	1.118
28.200	-0.049	1.144
28.400	-0.066	1.170
28.600	-0.091	1.185
28.800	-0.121	1.203
29.000	-0.150	1.226
29.200	-0.175	1.261
29.400	-0.187	1.308
29.600	-0.193	1.352
29.800	-0.206	1.385
30.000	-0.224	1.408
30.200	-0.245	1.443
30.400	-0.268	1.480
30.600	-0.290	1.519

30.800	-0.301	1.550
31.000	-0.316	1.585
31.200	-0.326	1.613
31.400	-0.320	1.627
31.600	-0.310	1.638
31.800	-0.297	1.643
32.000	-0.275	1.649
32.200	-0.240	1.663
32.400	-0.206	1.674
32.600	-0.162	1.670
32.800	-0.124	1.647
33.000	-0.097	1.635
33.200	-0.066	1.641
33.400	-0.026	1.661
33.600	0.008	1.684
33.800	0.030	1.721
34.000	0.058	1.771
34.200	0.107	1.812
34.400	0.161	1.862
34.600	0.224	1.887
34.800	0.291	1.937
35.000	0.291	1.937

A.4 Varianzen der Vorinformationen

Die Varianzen für die Vorinformationen sind nachfolgend ausgehend vom niedrigsten Höhenlevel bis zum höchsten Höhenlevel aufgelistet, siehe auch [Collard und Healy 2002].

Varianzen der verwendeten 43 Temperatur-Vorinformationen in Kelvin:

Hoehe (hPa)	Varianz (K)
1013.25	2.09761
1005.43	1.61585
985.88	1.66328
957.44	1.46330
922.46	1.51633
882.80	1.49131
839.95	1.50704
795.09	1.51512
749.12	1.53012
702.73	1.64402
656.43	1.77045
610.60	1.92816
565.54	2.07605
521.46	2.15991
478.54	2.12293
436.95	2.05212
396.81	1.92416
358.28	1.93762
321.50	1.86246
286.60	1.94791
253.71	1.98609
222.94	1.93639
194.36	1.70310
167.95	1.39952
143.84	1.19405
122.04	1.17554
102.05	1.28121
85.18	1.16399
69.97	1.12857
56.73	1.16886
45.29	1.26673
35.51	1.30758
27.26	1.77068
20.40	2.22149
14.81	2.65703
10.37	2.81776
6.95	3.55308
4.40	4.52485
2.61	5.67889
1.42	6.04378
0.69	3.92136
0.29	9.89987
0.10	29.99920

Varianzen der verwendeten 26 spezifischen Feuchte-Vorinformationen in g/kg:

Hoehe (hPa)	Varianz (g/kg)
1013.25	0.280705
1005.43	0.213698
985.88	0.117540
957.44	0.097612
922.46	0.119475
882.80	0.139346
839.95	0.150451
795.09	0.156666
749.12	0.155679
702.73	0.166792
656.43	0.177762
610.60	0.182385
565.54	0.172915
521.46	0.146842
478.54	0.128501
436.95	0.141963
396.81	0.182271
358.28	0.228893
321.50	0.248839
286.60	0.273603
253.71	0.401877
222.94	0.775591
194.36	1.130570
167.95	1.146090
143.84	1.032600
122.04	1.174650

Varianz der verwendeten Erdoberflächen-Luftdruck-Vorinformation in hPa:

2.6896

A.5 Zusätzliche Abbildungen zum Kapitel 5

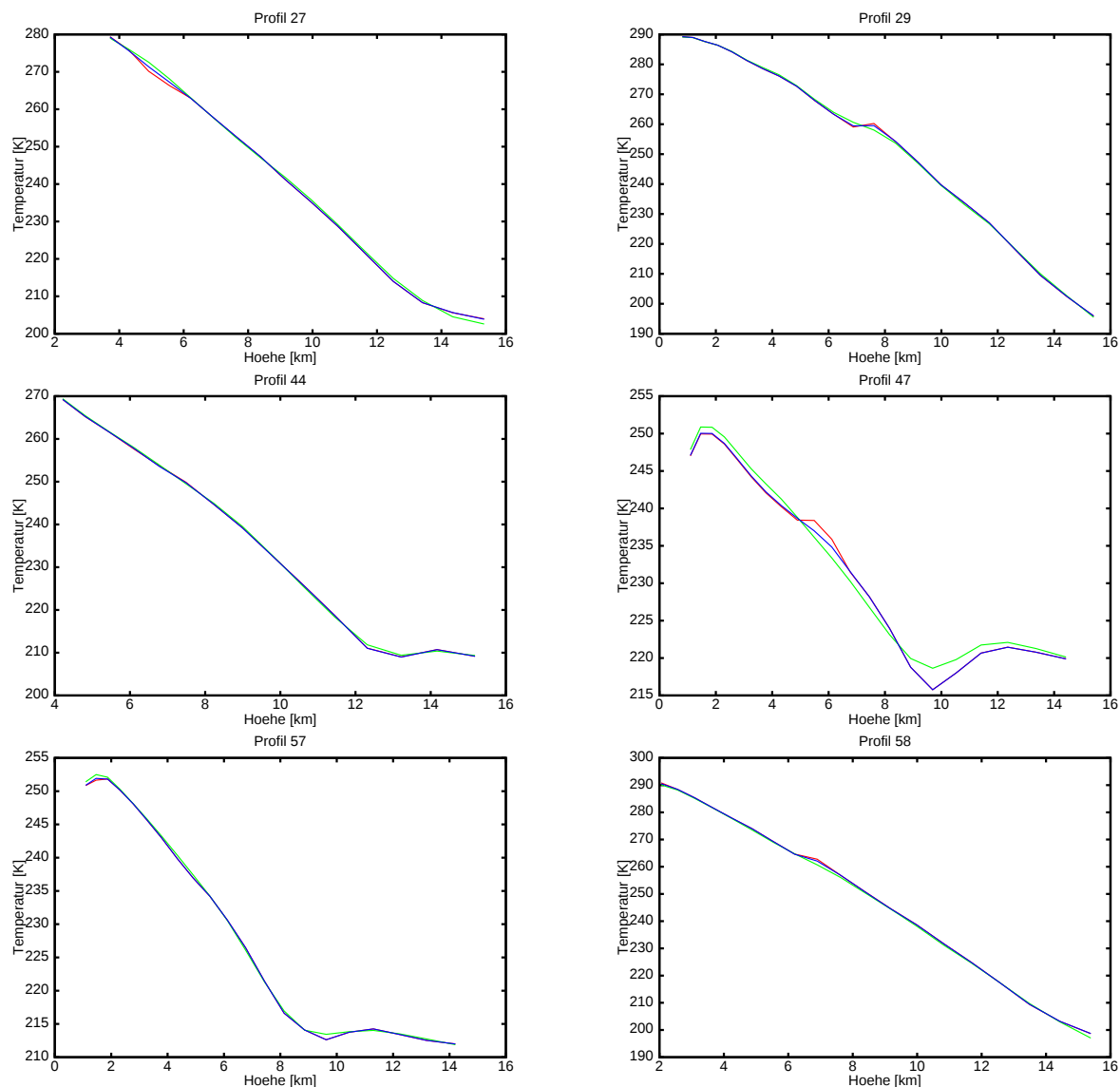


Abbildung A.1: Vergleich der Feuchttemperaturen:

rot ... Lösung des erweiterten Regularisierungsverfahrens;
 grün ... Profil der Vorinformationen aus ECMWF;
 blau ... 1DVAR-Lösung.

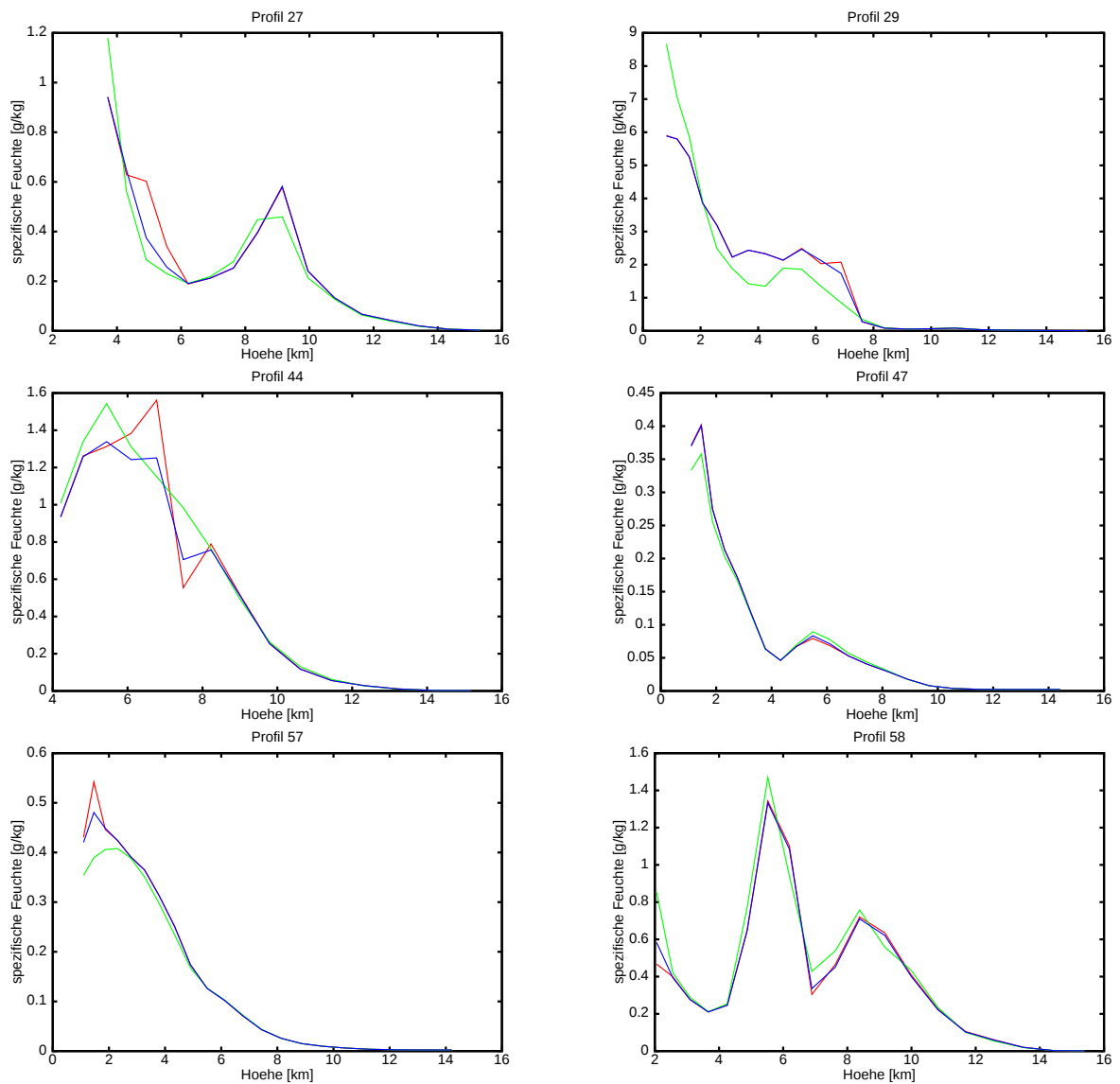


Abbildung A.2: Vergleich der spezifischen Feuchten:
 rot ... Lösung des erweiterten Regularisierungsverfahrens;
 grün ... Profil der Vorinformationen aus ECMWF;
 blau ... 1DVAR-Lösung.

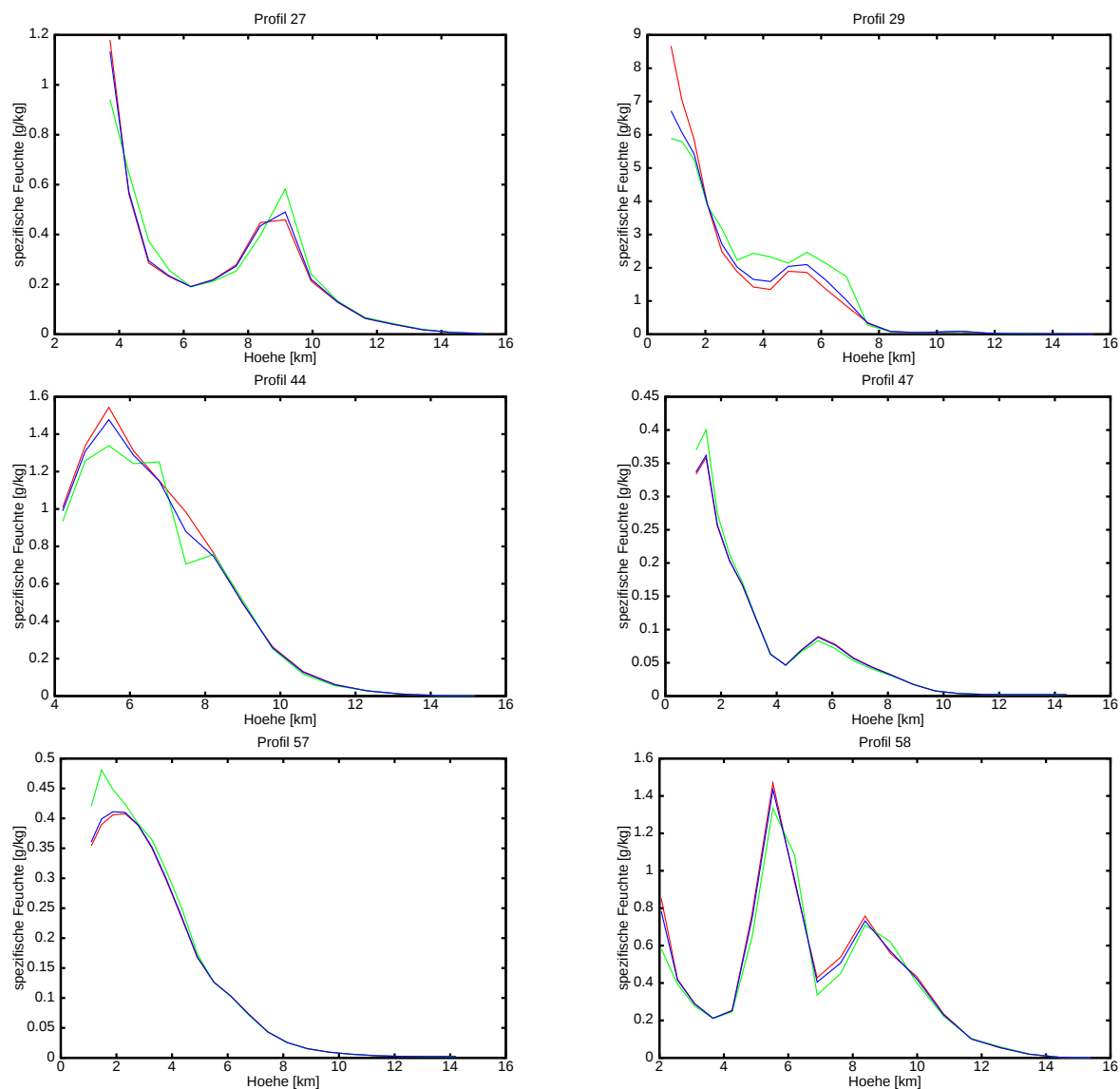


Abbildung A.3: Vergleich der spezifischen Feuchten aus unterschiedlichen 1DVAR-Lösungen:

rot ... spezifische Feuchten aus dem ECMWF-Modell;

grün ... 1DVAR-spezifische Feuchten im nicht übergewichteten Fall;

blau ... 1DVAR-spezifische Feuchten im übergewichteten Fall.

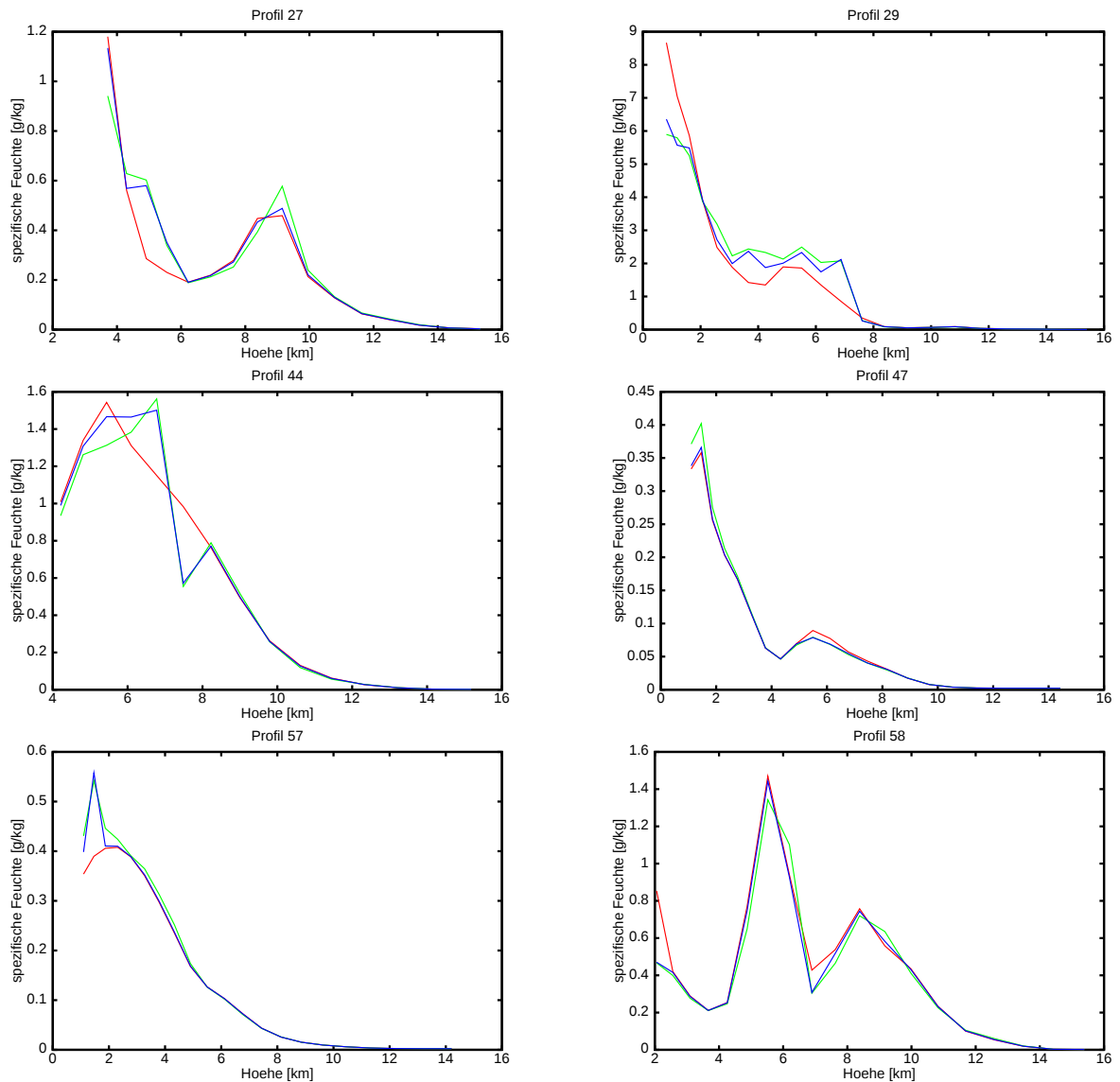


Abbildung A.4: Vergleich der spezifischen Feuchten aus unterschiedlichen Lösungen des erweiterten Regularisierungsverfahrens:

- rot ... spezifische Feuchten aus dem ECMWF-Modell;
- grün ... Lösung des erw. Regularisierungsverfahrens im nicht übergewichteten Fall;
- blau ... Lösung des erw. Regularisierungsverfahrens im übergewichteten Fall.