Identifizierung der hydraulischen Funktionen poröser Medien unter Verwendung formfreier Ansätze

Den Naturwissenschaftlichen Fakultäten
der Friedrich-Alexander-Universität Erlangen-Nürnberg
zur
Erlangung des Doktorgrades

vorgelegt von
Sandro Bitterlich
aus Annaberg-Buchholz

Als Dissertation genehmigt von den Naturwissenschaftlichen Fakultäten
der Universität Erlangen-Nürnberg

Tag der mündlichen Prüfung:	18. Juni 2003
Vorsitzender der Promotionskommision:	Prof. Dr. H. Kisch
Erstberichterstatter:	Prof. Dr. P. Knabner
Zweitberichterstatter:	Dr. B. Kaltenbacher

Vorwort

Ich möchte mich bei allen Personen bedanken, die mir bei der Erstellung dieser Arbeit geholfen haben. An erster Stelle ist dabei natürlich mein Doktorvater Prof. Dr. P. Knabner zu nennen, der mir vielfältige Anregungen für meine Arbeit gab. Besonderer Dank gilt auch Herrn Prof. Dr. W. Durner für seine nützlichen Hinweise und die Bereitstellung umfangreicher experimenteller Daten aus Säulenexperimenten. Auch Frau Dr. B. Kaltenbacher danke ich für die Übernahme des Zweitgutachtens. Darüberhinaus bedanke ich mich bei allen Mitarbeiterinnen und Mitarbeitern des Lehrstuhls I für Angewandte Mathematik der Friedrich-Alexander-Universität für das sehr angenehme Arbeitsklima.

Erlangen, im März 2003,

Sandro Bitterlich

Inhaltsverzeichnis

Kapitel 1

Einleitung

Für die Altlastensanierung gewinnt die numerische Simulation von (reaktiven) Transportprozessen zunehmend an Bedeutung. Die Simulation am Computer kann dazu dienen, ein tieferes Verständnis über die in Böden ablaufenden Prozesse zu erlangen. Sie ermöglichen auch die Identifikation spezieller am Gesamtprozess beteiligter Teilprozesse und anhand von Fallstudien können die Auswirkungen unterschiedlichster Einflussnahmen auf den Gesamtprozess untersucht werden.

Dies bildet jedoch nur einen ersten Schritt und das eigentliche Anliegen ist mithilfe der numerischen Simulation eine Langzeitprognose über die Schadstoffausbreitung zu erhalten. Die Simulation soll darüber Aussagen liefern, ob und inwieweit ein Eingriff in den Altlastenstandort notwendig ist oder bereits das natürliche Reinigungsvermögen einen Eintrag von Schadstoffen z. B. in das Grundwasserreservoir verhindert. Die Zuverlässigkeit der aus numerischen Simulationen gewonnenen Erkenntnisse hängt davon ab, wie die Realität im Computer abgebildet wird. Das zugrunde liegende mathematische Modell muss alle für den Gesamtprozess relevanten Teilprozesse berücksichtigen.

Ein grundlegendes Problem bei der Beschreibung von Transportprozessen in Böden ist die Modellierung des Wasser- bzw. allgemeiner Fluidtransports. Oft wird der Fluidtransport selbst nicht simuliert, sondern es wird ein stationäres Fließregime angesetzt. Dies ist z. B. gerechtfertigt, wenn Transportprozesse im Grundwasserbereich simuliert werden sollen. In vielen Fällen wird die Annahme eines stationären Flusses nicht adäquat sein und die Notwendigkeit bestehen, den Transport unter ungesättigten Bedingungen zu modellieren.

Das Modell für die (un-)gesättigte Strömung wird durch die *Richards-Gleichung* beschrieben. Diese Gleichung ist ein vereinfachtes Zweiphasenmodell, bei dem die Bewegung der Gasphase nicht explizit berücksichtigt wird, son-

dern nur ihr Einfluss auf die Leitfähigkeit des porösen Mediums. Der Druck der Gasphase wir als konstant angenommen.

Von der Richards-Gleichung existieren mehrere Versionen, die sich darin unterscheiden, welche physikalische Größe als primäre Unbekannte betrachtet wird. In der Sättigungsformulierung wird die Sättigung (Fluidgehalt) als zu bestimmende Unbekannte betrachtet, während die Druckformulierung den Druck als Unbekannte verwendet. Die Sättigungsformulierung erlaubt die Modellierung eines verschwindenden Fluidgehalts im Boden. Dagegen ist sie im gesättigten Bereich nicht anwendbar. Mit der Druckformulierung kann neben dem ungesättigten auch der gesättigte Bereich beschrieben werden. Im Gegensatz zur Sättigungsformulierung ist hier die Modellierung einer Austrocknung des Bodens nicht möglich. Welche Formulierung Verwendung findet hängt vom jeweiligen Anwendungsproblem ab. In einigen Arbeiten wird auch eine gemischte Formulierung der Richards-Gleichung benutzt, die sowohl den Druck als auch die Sättigung als Unbekannte enthält. Wir werden hier mit der Druckformulierung der Richards-Gleichung arbeiten.

In die Richards-Gleichung gehen die Materialeigenschaften des Bodens in Form von nichtlinearen Koeffizientenfunktionen ein. Im Detail sind dies die Retentionsfunktion, die einen Zusammenhang zwischen dem Grad der Sättigung und des vorherrschenden Drucks herstellt, sowie die hydraulische Leitfähigkeit. Diese beiden Beziehungen werden wir im Weiteren als hydraulische Funktionen bezeichnen.

Eine Simulation der ungesättigten Strömung setzt die Kenntnis dieser Funktionen voraus. Im begrenzten Umfang direkt messbar ist nur die Retentionsfunktion. Aus gemessenen Druck-Fluidgehalts-Datenpaaren kann mithilfe von Interpolationen die Retentionsfunktion approximiert werden. In der Regel werden die hydraulischen Funktionen jedoch mit einer indirekten Methode bestimmt. Dazu werden im Labor *Säulenexperimente* durchgeführt. Das experimentelle Setup ist wie folgt (siehe Abbildung 1 und [53] für die Versuchsanlage der Universität Bayreuth): Der zu untersuchende Boden befindet sich in einem vertikal orientierten Zylinder (Bodensäule, 1). Der obere Rand der Bodensäule ist gegenüber der Umwelt abgeschlossen. Der Boden der Säule ist mit einer keramischen Platte versehen und mit einer Unterdruckanlage verbunden. Die Bodensäule wird bis zur Sättigung mit dem Fluid aufgefüllt und mithilfe der Gravitation in einen Gleichgewichtszustand (kein Fluss) gebracht. Mit diesem Gleichgewicht als Anfangszustand wird durch das Anlegen eines Unterdrucks am Boden der Säule ein Ausfluss des Fluids aus der Säule erzeugt (*Ausflussexperiment*). Im Gegenzug bewirkt eine Druckerhöhung einen Rückfluss des Fluids in die Säule (*Rückflussexperiment*). Aus- und Rückfluss können auch periodisch gekoppelt werden. In die Säule sind zwei Tensiometer (2a) und zwei TDR-Sensoren (2b) eingelassen.

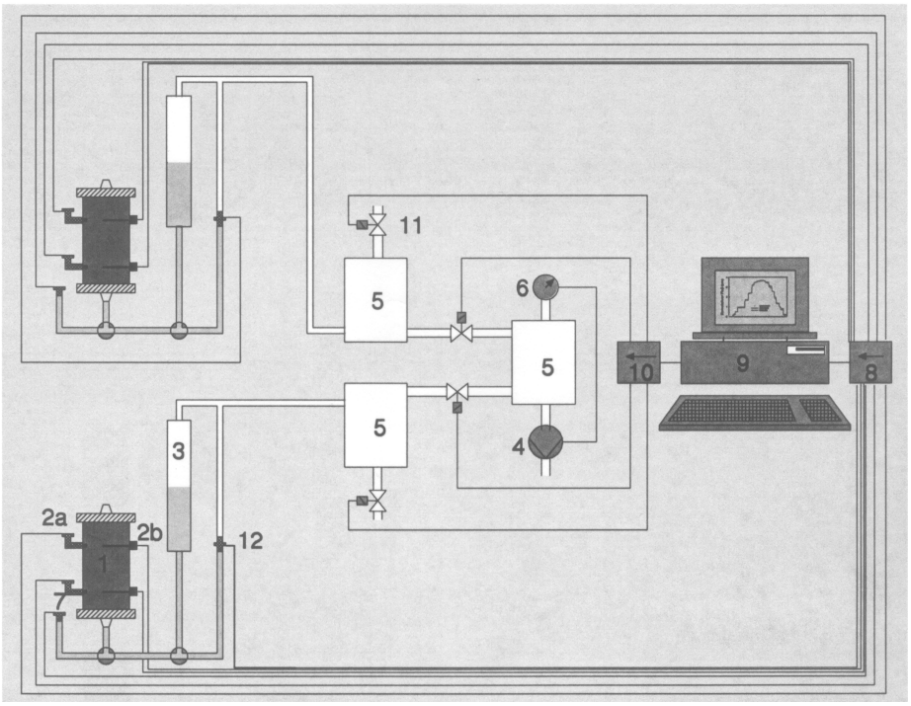


Abbildung 1.1: Aufbau der Versuchsanlage zur Bestimmung der hydraulischen Kenngrößen von Bodensäulen. 1: Bodensäule, 2a: Mikro-Tensiometer mit Druckaufnehmer, 2b: TDR, 3: Bürette, 4: Vakuumpumpe, 5: Vorratsbehälter, 6: Manometer, 7: Druckaufnehmer, 8: Multiplexer und Analog/Digital-Converter, 9: Computer, 10: Digital/Analog-Converter, 11: Magnetventil, 12: Differenzdruckaufnehmer. Reproduktion mit Genehmigung von Prof. Dr. W. Durner, Abteilung Bodenkunde und Bodenphysik der Technischen Universität Braunschweig.

Mit diesen kann der Druck bzw. der Fluidgehalt gemessen werden. Auch der Druck am oberen Rand ist messbar. Ebenso ist eine Messung des kumulative Ausfluss im Auffangbehälter (3) möglich. Die Steuerung des Experiments erfolgt durch einen PC (9). Weiterführende Informationen über Versuchsaufbau und Durchführung des Experiments sind in [48] und [53] zu finden.

Wenn die hydraulischen Funktionen gegeben sind, so liefert eine Simulation des Säulenexperiments entsprechende simulierte Messdaten. Da die Bodensäule senkrecht zur vertikalen Flussrichtung homogen gepackt wird, ist eine räumlich eindimensionale Modellierung ausreichend. Durch Invertieren des Simulationsprozesses können die hydraulischen Funktionen aus gegebenen Messdaten rekonstruiert werden.

Diese Methode der *inversen Modellierung* ist seit langem Standard für die

Bestimmung von Materialeigenschaften. In [41] wird die Methode zur Identifizierung von nichtlinearen Diffusionskoeffizienten in der Poröse-Medien-Gleichung eingesetzt. [27] behandelt die numerische Parameteridentifizierung im Mehrphasenfluss. Viele Arbeiten (z. B. [14], [15], [47]), die sich mit der Identifizierung der hydraulischen Funktionen beschäftigen, unterscheiden sich häufig nur durch die verwendeten Parametrisierungen dieser Funktionen und der Diskretisierungstechnik für die Modellgleichung. Im Gegensatz zu den erwähnten Arbeiten, wo von einer fixierten Form der hydraulischen Funktionen ausgegangen wird, werden wir stückweise polynomiale Ansätze zur Parametrisierung verwenden. Die vorliegende Arbeit wurde hauptsächlich durch die Arbeit von B. Igler [30] motiviert, wo nichtlineare Sorptionscharakteristiken in Form von stückweise linearen Funktionen aus Säulendurchbruchsexperimenten identifiziert werden.

Durch Änderungen des Experimentdesigns wird versucht die Zuverlässigkeit der bestimmten Parameter zu erhöhen. In [59] wird z. B. ein *Closed-Flow*-Design für Säulendurchbruchsexperimente untersucht. Das Experimentdesign bei der Identifizierung der hydraulischern Funktionen bezieht sich größtenteils auf die Variation des am Ausflussrand der Bodensäule angelegten Unterdrucks ([14], [53], [54]). In [8] wird das Experimentdesign aus mathematischer Sicht betrachtet.

Neben den anwendungsorientierten Arbeiten existieren auch Arbeiten, die sich aus theoretischer Sicht mit den hier betrachteten inversen Problemen beschäftigen. So werden z. B. in [6], [9], [10], [11] und [12] Eindeutigkeitsresultate bewiesen.

Diese Arbeit ist in 5 Kapitel unterteilt. Das nächste Kapitel beginnt mit einer allgemeinen Einführung in die Behandlung von inversen Problemen. Insbesondere wird die Parameteridentifizierung besprochen. Das Identifizierungsproblem wird in ein Minimierungsproblem eines Fehlerfunktionals überführt. Es werden Methoden zur Brechnung des Gradienten im kontinuierlichen und diskreten Fall beschrieben. Das zweite Kapitel schließt mit Betrachtungen zur eindeutigen Identifizierbarkeit. Diese Kapitel orientiert sich im Kern an den Darstellungen in [30].

Im dritten Kapitel wird das Modell zur Beschreibung der Säulenexperimente vorgestellt. Anhand von Beispielen wird die Schlechtgestelltheit des inversen Problems erläutert. Aufbauend auf das Kapitel 2 wird die eindeutige Identifizierbarkeit der hydraulischen Funktionen aus den Säulenexperimenten im kontinuierlichen Modell untersucht. Anschließend wird eine hybrid-gemischte Finite-Elemete-Methode zur Diskretisierung der Richards-Gleichung in ihren Grundzügen beschrieben. Für das diskrete Modell werden Methoden zur Berechnung des Gradienten des Fehlerfunktionals angegeben. Hierzu wird u. a. ein adjungiertes Problem aufgestellt.

Den Kern der vorliegenden Arbeit bildet das Kapitel 4. Dort werden verschiedene Spline-Parametrisierungstechniken für die Nichtlinearitäten vorgestellt. Entsprechend der hierarchischen Struktur der Parametrisierungen wird die Identifizierung in ein Multi-Level-Verfahren eingebettet und mit einer Sensitiviätsanalyse gekoppelt. Der zweite Abschnitt von Kapitel 4 führt eine Reihe von Fallstudien auf, anhand derer die Identifizierung der hydraulischen Funktionen auf Stabilität und Zuverlässigkeit untersucht wird. Dabei werden auch experimentelle Daten betrachtet. Zur Charakterisierung des inversen Problems dienen die Sensitivitäten, die auch für die Einbringung von Adaptivität in das Multi-Level-Verfahren (adaptive Verfeinerung, adaptives Fehlerfunktional) genutzt werden. Zusätzlich können in die Identifizirung a priori Informationen einbezogen werden.

Das fünfte Kapitel gibt einen Einblick in das Experimentdesign. Zur Motivation werden zunächst zwei Beispiele für ein *Multistep*-Experiment (siehe z. B. [13]) untersucht. Anschließend erfolgt eine mathematische Formulierung des optimalen Experimentdesigns. Das Kapitel schließt mit Beispielen, in denen die Variation der Position der Druckbeobachtung und die Variation der Multistep-Funktion untersucht werden.

Die Anhänge A und B enthalten eine Zusammenfassung und eine Darstellung der in der Arbeit verwendeten Funktionenräume, die nicht an anderer Stelle definiert werden.

Die Implemetierung des numerischen Identifizierungsverfahrens erfolgte in der Programmiersprache C und wurde in das Softwarepaket RICHY1D [38] eingebaut. Neben der Simulation von Transportvorgängen ist in RICHY1D damit die Identifizierung der hydraulischen Funktionen aber auch von Sorptionscharakteristiken aus Säulenexperimenten möglich.

Kapitel 2

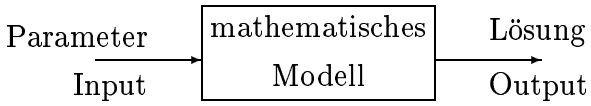
Zur Lösung inverser Probleme

2.1 Identifizierungsprobleme

2.1.1 Korrektheit mathematischer Aufgaben

Das mathematische Modell eines Experiments enthält häufig eine Anzahl von Parametern, die gewisse physikalische Eigenschaften bzw. Materialeigenschaften beschreiben. Solche Modelle sind beispielsweise partielle Differentialgleichungen mit den entsprechenden Anfangs- und Randbegingungen, in denen diese (Material-)Parameter in Form von Koeffizientenfunktionen (oder auch als rechte Seite) eingehen.

Für vorgegebene Parameter, die als Input des Modells betrachtet werden können, erhalten wir durch das Lösen des Modells den Output. Der Lösungsprozess, die Simulation des Experiments, wird auch als das *direkte Problem* bezeichnet.



Durch das Invertieren des direkten Problems gelangen wir zu einem Parameteridentifizierungsproblem. Es handelt sich hierbei um ein *inverses Problem*, bei dem aus einer Beobachtung des Experiments, d. h. aus einer Beobachtung (Messung) der Lösung (des Outputs) des direkten Problems, die unbekannten Parameter des nur der Form nach bekannten Modells zu bestimmen sind.

Derartige inverse Probleme sind im Allg. inkorrekt gestellt. Dabei heißt eine mathematische Aufgabe *korrekt gestellt* im Sinne von Hadamard (siehe [24]), wenn sie den folgenden 3 Bedingungen genügt:

1. Es existiert eine Lösung der Aufgabe (Existenz).
2. Die Lösung der Aufgabe ist eindeutig bestimmt (Eindeutigkeit).
3. Die Lösung der Aufgabe hängt stetig von den Parametern der Aufgabe ab (Stabilität).

Wenn eine dieser Bedingungen verletzt ist, so spricht man von einem *inkorrekt gestellten* Problem.

Das direkte Problem können wir durch eine (nichtlineare) Abbildung $\mathcal{A}$, den direkten Lösungsoperator, von der Menge P der Parameter in die Menge U der Lösungen beschreiben:

$$\mathcal{A} : P \to U.$$

Dabei seien der Parameterraum P Teilmenge eines Banachraumes und der Lösungsraum U ein Banachraum. Die Erfüllung der Forderungen 1–3 für das direkte Problem, d. h. die Wohldefiniertheit von $\mathcal{A}$, hängt wesentlich von der Wahl der Räume U und P und der zugehörigen Metriken ab.

Es sei angenommen, dass die Abbildung $\mathcal{A}$ auf ganz P definiert und stetig in den Metriken der Räume P und U ist. Dann ist das direkte Problem korrekt gestellt. Die Lösung des inversen Problems entspricht dem Finden der Umkehrabbildung $\mathcal{A}^{-1} : U \to P$ von $\mathcal{A}$. Damit das inverse Problem korrekt gestellt ist, müssen die Bedingungen

1. $\mathcal{A}$ ist bijektiv und
2. $\mathcal{A}^{-1}$ ist stetig in den Metriken der Räume U und P

erfüllt sein.

2.1.2 Verallgemeinerte Inverse

Die Vorgabe für das inverse Problem $u_0 = \mathcal{A}(p_0)$ $(p_0 \in P, u_0 \in U)$ kann durch Messungen meist nicht exakt aus dem Experiment bestimmt werden. Deshalb werden nur gestörte Daten u_ε zur Verfügung stehen, die mit einem gewissen Datenfehler

$$\|u_\varepsilon - u_0\| \leq \varepsilon$$

behaftet sind. Somit ist es möglich, dass diese gestörte Größe u_ε nicht im Bildraum des direkten Lösungsoperators $\mathcal{A}$ enthalten ist $(u_\varepsilon \in U \setminus \mathcal{A}(P))$. In diesem Fall existiert keine (klassische) Lösung $p_\varepsilon \in P$ von

$$\mathcal{A}(p_\varepsilon) = u_\varepsilon. \tag{2.1}$$

Es ist also ein allgemeinerer Lösungsbegriff für Operatorgleichungen der Form (2.1) erforderlich. Dazu folgen wir den Darstellungen in [17] und [31] (bzw. [39] für den linearen Fall).

Für injektives $\mathcal{A}$ ist die *verallgemeinerte Lösung* von (2.1) dasjenige Element $p_\varepsilon \in P$, welches das *Fehlerfunktional*

$$\mathcal{J}(p) := \|\mathcal{A}(p) - u_\varepsilon\| \tag{2.2}$$

minimiert. Anderenfalls ist die verallgemeinerte Lösung gegeben durch das Element $p_\varepsilon \in P$, welches unter allen Lösungen von

$$\min_{p \in P} \mathcal{J}(p)$$

die kleinste Norm in P besitzt. Die eindeutige Existenz der jeweiligen Minima ist durch die Eigenschaften von $\mathcal{A}$ bzw. P sicherzustellen. Hinreichende Bedingungen hierfür sind, dass A ein stetiger Operator ist und P eine kompakte Teilmenge eines streng normierten Banachraumes darstellt.

Der hierdurch definierte Operator

$$\mathcal{A}^\dagger : U \to P$$

wird als *verallgemeinerte Inverse* bezeichnet. Wenn $u_\varepsilon \in \mathcal{A}(P)$ gilt, dann fällt für injektives $\mathcal{A}$ die verallgemeinerte Lösung mit der gewöhnlichen Lösung zusammen.

Bei den praktisch interessanten Aufgabenstellungen ist die verallgemeinerte Inverse häufig nicht stetig, was das Hauptproblem bei der Behandlung inverser Aufgaben darstellt. Deshalb werden zur Stabilisierung des Lösungsprozesses *Regularisierungsverfahren* verwendet, bei denen die verallgemeinerte Inverse durch eine Familie von stetigen Operatoren

$$\mathcal{R}_\alpha : U \to P$$

mit der punktweisen Konvergenz

$$\lim_{\alpha \to 0} \mathcal{R}_\alpha[\mathcal{A}(p)] = p \tag{2.3}$$

für alle $p \in P$ ersetzt wird. Hierbei ist α der *Regularisierungsparameter*. Als Näherungslösung des inversen Problems bei der Vorgabe u_ε verwenden wir nun

$$p_\varepsilon^\alpha = \mathcal{R}_\alpha[u_\varepsilon].$$

Hieraus ergibt sich für den *Identifizierungsfehler*

$$\begin{aligned} \|p_\varepsilon^\alpha - p_0\| &= \|\mathcal{R}_\alpha[u_\varepsilon] - \mathcal{R}_\alpha[u_0] + \mathcal{R}_\alpha[\mathcal{A}(p_0)] - p_0\| \\ &\leq \|\mathcal{R}_\alpha[u_\varepsilon] - \mathcal{R}_\alpha[u_0]\| + \|\mathcal{R}_\alpha[\mathcal{A}(p_0)] - p_0\|. \end{aligned}$$

Für lineare Operatoren $\mathcal{R}_\alpha$ gilt speziell

$$\|p_\varepsilon^\alpha - p_0\| \leq \varepsilon\|\mathcal{R}_\alpha\| + \|\mathcal{R}_\alpha[\mathcal{A}(p_0)] - p_0\|.$$

Damit haben wir für den Identifizierungsfehler eine obere Schranke erhalten, die sich aus einem durch den Datenfehler bestimmten Anteil $\|\mathcal{R}_\alpha[u_\varepsilon] - \mathcal{R}_\alpha[u_0]\|$ und dem Regularisierungsfehler $\|\mathcal{R}_\alpha[\mathcal{A}(p_0)] - p_0\|$ zusammensetzt.

Für beliebiges festes α geht $\|\mathcal{R}_\alpha[u_\varepsilon] - \mathcal{R}_\alpha[u_0]\|$ wegen der Stetigkeit der Operatoren $\mathcal{R}_\alpha$ für $\varepsilon \to 0$ gegen Null. Wenn $\varepsilon > 0$ fest gewählt wird, dann geht wegen Bedingung (2.3) zwar der Regularisierungsfehler $\|\mathcal{R}_\alpha[\mathcal{A}(p_0)] - p_0\|$ für $\alpha \to 0$ gegen Null aber nicht notwendig $\|\mathcal{R}_\alpha[u_\varepsilon] - \mathcal{R}_\alpha[u_0]\|$. Im linearen Fall bedeutet dies, dass im Allg. gilt

$$\lim_{\alpha \to 0} \|\mathcal{R}_\alpha\| = \infty.$$

Es wird also im Allg. für jedes $\varepsilon > 0$ in Abhängigkeit von ε und u_ε ein $\alpha_{\varepsilon,u_\varepsilon} > 0$ existieren, sodass für $\alpha < \alpha_{\varepsilon,u_\varepsilon}$ der Identifizierungsfehler wieder wächst, d. h. der Datenfehler gewinnt an Einfluss. Daraus ergibt sich das Problem der Wahl eines optimalen Regularisierungsparameters α.

Es seien hier einige gebräuchliche Regularisierungsverfahren genannt:

1. Bei der Regularisierungsmethode von Tikhonov wird zum Fehlerfunktional (2.2) ein so genanntes *stabilisierendes Funktional* (Strafterm) addiert:
$$\mathcal{J}_\alpha(p) := \mathcal{J}(p) + \alpha\|p - p^*\|^2$$
mit einer Norm $\|.\|$ in P und einer a priori Schätzung p^* von p_0 und dem Regularisierungsparameter α.

2. Es können Iterationsverfahren zur Regularisierung herangezogen werden, wie z. B. die Landweber-Iteration oder das Verfahren der konjugierten Gradienten.

3. Auch Projektionsverfahren eignen sich zur Regularisierung inverser Probleme.

4. Ausgehend von der Spektralzerlegung des Operators $\mathcal{A}$ werden so genannte Filtermethoden gewonnen.

Es existiert eine Vielzahl von Literatur zum Thema der Regularisierung inverser Probleme. Es sei hier auf [17], [23], [29] und [39] verwiesen. In [16], [26], [42], [43], [49], [50] und [55] sind Regularisierungsstrategien für nichtlineare Operatoren und insbesondere die Landweber-Iteration untersucht worden. Die dabei erforderlichen Voraussetzungen werden bei unseren Identifizierungsproblemem im Allg. jedoch nicht erfüllt sein.

2.1.3 Problemformulierung

Wenn es bei der Durchführung des Experiments nicht möglich ist, die komplette Lösung des direkten Problems zu beobachten, dann werden nur Teile dieser Lösung oder durch diese Lösung bestimmte Größen gemessen. Bei instationären Problemen wie parabolischen Differentialgleichungen ist es üblich einige charakteristische Größen während eines Zeitintervalls $[0, T]$ zu beobachten.

Wir bezeichnen im Weiteren die Beobachtungen mit $\omega_k, k = 1, ..., \kappa$, und führen die zugehörigen Beobachtungsoperatoren $\mathcal{B}_k$ ein, die auch explizit von den unbekannten Parametern abhängen können:

$$\mathcal{B} := \begin{pmatrix} \mathcal{B}_1 \\ \vdots \\ \mathcal{B}_\kappa \end{pmatrix} : U \times P \to \begin{pmatrix} W_1 \\ \vdots \\ W_\kappa \end{pmatrix} =: W$$

$$(u, p) \mapsto \begin{pmatrix} \mathcal{B}_1(u, p) \\ \vdots \\ \mathcal{B}_\kappa(u, p) \end{pmatrix} = \mathcal{B}(u, p) =: \omega = \begin{pmatrix} \omega_1 \\ \vdots \\ \omega_\kappa \end{pmatrix}.$$

Dabei seien für $k = 1, ..., \kappa$ die Beobachtungsräume W_k Banachräume. Wenn wir für u den direkten Lösungsoperator $\mathcal{A}(p)$ einsetzen, so kann $\mathcal{B}$ als Abbildung vom Parameterraum P in den Beobachtungsraum W aufgefasst werden:

$$\mathcal{B} : P \to W$$

$$p \mapsto \omega.$$

Die Aufgabe besteht nun darin, einen Parameter $p_0 \in P$ mit

$$\mathcal{B}(\mathcal{A}(p_0), p_0) = \omega_0 \tag{2.4}$$

zu bestimmen. ω_0 bezeichnet die ungestörte Beobachtung. Der Begriff der Identifizierbarkeit der Parameter $p \in P$ aus einer Beobachtung wird nun wie folgt definiert:

Definition 2.1 *Die Parameter p der Menge P heißen* identifizierbar *aus der Beobachtung ω, wenn der zugehörige Operator $\mathcal{B}$ als Abbildung von P in den Beobachtungsraum W injektiv ist.*

Die Eigenschaft der Identifizierbarkeit hängt demnach ab von

- der Wahl des Raumes P,

- Art und Anzahl der Beobachtungen (Definition des Beobachtungsoperators $\mathcal{B}$) und
- dem experimentellen Setup, insbesondere Anfangs- und Randbedingungen (Definition des direkten Lösungsoperators $\mathcal{A}$).

Diese Einflussfaktoren sind derart festzulegen, dass die Identifizierbarkeit gewährleistet wird. Dabei ist die technische Durchführbarkeit des Experiments und der Beobachtungen zu berücksichtigen. Insbesondere sollte die kleinst mögliche Anzahl von Beobachtungen gewählt werden.

Infolge von Messfehlern wird anstelle von ω_0 eine gestörte Beobachtung ω_ε mit

$$\|\omega_\varepsilon - \omega_0\| \leq \varepsilon$$

gegeben sein. Deshalb folgen wir dem Prinzip der verallgemeinerten Inversen und betrachten anstelle des Identifizierungsproblems (2.4) das folgende Minimierungsproblem:
Gesucht wird ein $p_\varepsilon \in P$ mit

$$\mathcal{J}_\varepsilon(p_\varepsilon) = \min_{p \in P} \mathcal{J}_\varepsilon(p). \tag{2.5}$$

Das Fehlerfunktional $\mathcal{J}_\varepsilon$ sei gegeben als $\mathcal{J}_\varepsilon = \tilde{\mathcal{J}}_\varepsilon \circ \mathcal{B} \circ (\mathcal{A}(p), p)$ für ein Funktional $\tilde{\mathcal{J}}_\varepsilon$ der Form

$$\tilde{\mathcal{J}}_\varepsilon(\omega) = \sum_{k=1}^{\kappa} \tilde{\mathcal{J}}_{\varepsilon,k}(\omega_k)$$

mit stetigen Funktionalen $\tilde{\mathcal{J}}_{\varepsilon,k} : W_k \to \mathbb{R}_+$, wobei die Bedingung

$$\tilde{\mathcal{J}}_\varepsilon(\omega) = 0 \qquad \Longleftrightarrow \qquad \omega = \omega_\varepsilon$$

zu erfüllen ist. Das Funktional $\mathcal{J}_\varepsilon$ bestimmt den Grad der Abweichung der zum Parameter $p \in P$ gehörenden Beobachtung ω von den mit Messfehlern behafteten experimentellen Daten ω_ε.

Es ist wieder die eindeutige Existenz eines Minimums von (2.5) sicherzustellen. Hinreichend etwa für die Existenz einer Lösung von (2.5) ist die Kompaktheit des Parameterraumes P.

Wenn die Lösung p_ε von (2.5) im Inneren von P liegt, so muss für ein (Fréchet- oder Gâteaux-) differenzierbares Funktional $\tilde{\mathcal{J}}_\varepsilon$

$$\tilde{\mathcal{J}}'_\varepsilon[p_\varepsilon] = 0$$

gelten. Wenn dagegen p_ε ein Element des Randes von P ist, so sind Lagrange-Bedingungen zu erfüllen (siehe z. B. Theorem 3.17 in [32] und Theorem 20.3 in [21]).

2.2 Differentiation des Fehlerfunktionals

Bei der Lösung des Identifizierungsproblems ist also ein nichtlineares Funktional zu minimieren. Optimalitätskriterien und effiziente numerische Minimierungsalgorithmen benötigen die Ableitung (Gradient) dieses Funktionals nach den Parametern. Wir werden in diesem Abschnitt zeigen, wie diese Ableitung mithilfe eines adjungierten Problems berechnet werden kann.

2.2.1 Der Fall des kontinuierlichen Problems

Das Modell des direkten Problems sei gegeben durch eine nichtlineare Gleichung bzw. durch ein nichtlineares Gleichungssystem

$$\mathcal{G}(u,p) = 0 \tag{2.6}$$

für eine nichtlineare Abbildung

$$\mathcal{G} : U \times P \to V^*$$

$$(u,p) \mapsto \mathcal{G}(u,p)$$

mit Banachräumen U, V und einem normierten Raum P. Die Gleichung (2.6) kann als eine allgemeine (schwache) Formulierung von partiellen Differentialgleichungen aufgefasst werden. Dabei bezeichnen wieder U den Lösungsraum, P den Parameterraum und V sei der Raum der Testfunktionen.

Wenn wir voraussetzen, dass eine nichtleere Teilmenge $\tilde{P} \subset P$ existiert, sodass für jedes $p \in \tilde{P}$ die Gleichung (2.6) genau eine Lösung $u \in U$ besitzt, dann können wir den direkten Lösungsoperator $\mathcal{A} : \tilde{P} \to U$ durch

$$\mathcal{G}(\mathcal{A}(p), p) = 0$$

definieren. $\mathcal{A}$ ist demnach die implizite Funktion von $\mathcal{G} = 0$.

Die Existenz und Glattheit dieses direkten Lösungsoperators hängt von der Glattheit der Abbildung $\mathcal{G}$ ab. Dies zeigt der bekannte Satz über die implizite Funktion.

Satz 2.2 (Satz über die implizite Funktion)
Seien $(U, \|.\|_U)$, $(P, \|.\|_P)$ und $(V^*, \|.\|_{V^*})$ normierte Räume, D eine Umgebung von $(u_0, p_0) \in U \times P$ und die Abbildung $\mathcal{G} : U \times P \to V^*$ genüge den Bedingungen

1. *$\mathcal{G}$ ist stetig in D mit $\mathcal{G}(u_0, p_0) = 0$,*
2. *$\frac{\partial \mathcal{G}}{\partial u}$ und $\frac{\partial \mathcal{G}}{\partial p}$ existieren als partielle Fréchet-Ableitungen in D und diese sind in D stetig und*

3. *der inverse Operator*

$$\left(\frac{\partial \mathcal{G}}{\partial u}[u_0, p_0]\right)^{-1} : V^* \to U$$

existiert und ist linear und beschränkt.

Dann existieren eine Umgebung $\tilde{P}$ *von* p_0 *und ein Operator* $\mathcal{A} : \tilde{P} \to U$, *sodass gilt:*

1. $\mathcal{G}(\mathcal{A}(p), p) = 0$ *für alle* $p \in \tilde{P}$,
2. $\mathcal{G}(u, p) = 0$ *impliziert* $u = \mathcal{A}(p)$ *für alle* $p \in \tilde{P}$ *und*
3. $\mathcal{A}$ *ist Fréchet-differenzierbar in* $\tilde{P}$ *mit*

$$\mathcal{A}'[p] = -\left(\frac{\partial \mathcal{G}}{\partial u}[\mathcal{A}(p), p]\right)^{-1} \frac{\partial \mathcal{G}}{\partial p}[\mathcal{A}(p), p] \tag{2.7}$$

für alle $p \in \tilde{P}$.

Beweis: Siehe z. B. Theorem 12.4.1 und Corollary 1 in [62]. □

Falls die Voraussetzungen des obigen Satzes erfüllt sind, dann existiert die Ableitung $\mathcal{J}' = \frac{d\mathcal{J}}{dp}$ des Fehlerfunktionals

$$\mathcal{J} = \tilde{\mathcal{J}} \circ \mathcal{B} \circ (\mathcal{A}(p), p),$$

wenn der Operator $\mathcal{B}$ und das Funktional $\tilde{\mathcal{J}}$ Fréchet-differenzierbar sind. Diese Ableitung kann mittels eines adjungierten Problems berechnet werden. Der folgende Satz ist eine Verallgemeinerung eines Resultats aus [30] für lineare und von den Koeffizienten unabhängige Beobachtungsoperatoren.

Satz 2.3 (Adjungiertes Problem)
Die Voraussetzungen von Satz 2.2 seien erfüllt. Des Weiteren seien $\tilde{\mathcal{J}}$ *und* $\mathcal{B}$ *Fréchet-differenzierbar (Gâteaux-differenzierbar). Dann gelten:*

1. *Das Funktional* $\mathcal{J} : P \to \mathbb{R}_+$ *ist wohldefiniert und Fréchet-differenzierbar (Gâteaux-differenzierbar) in einer Umgebung* $\tilde{P}$ *von* p_0.
2. *Sei* p *Element der offenen Menge* $\tilde{P}$, *in der* $\mathcal{A}$ *und* $\mathcal{J}$ *Fréchet-differenzierbar sind (siehe 1.), und* $u = \mathcal{A}(p)$. *Wenn* $\eta \in V$ *für alle* $\delta u \in U$ *das adjungierte Problem*

$$\left\langle \eta, \frac{\partial \mathcal{G}}{\partial u}[u, p]\delta u \right\rangle = \left\langle \tilde{\mathcal{J}}'\left[\mathcal{B}(u, p)\right], \frac{\partial \mathcal{B}}{\partial u}[u, p]\delta u \right\rangle \tag{2.8}$$

löst, dann ist die Ableitung von $\mathcal{J}$ gegeben durch

$$\langle \mathcal{J}'[p], \delta p \rangle = \left\langle \tilde{\mathcal{J}}' \left[\mathcal{B}(u,p) \right], \frac{\partial \mathcal{B}}{\partial p}[u,p]\delta p \right\rangle - \left\langle \eta, \frac{\partial \mathcal{G}}{\partial p}[u,p]\delta p \right\rangle \qquad (2.9)$$

für alle $\delta p \in P$.

Beweis:

1. Die Behauptung folgt sofort aus dem Satz über die implizite Funktion und der Kettenregel.

2. Aus der Definition des Funktionals $\mathcal{J}$ folgt unter Anwendung der Kettenregel zunächst

$$\begin{aligned}
\langle \mathcal{J}'[p], \delta p \rangle &= \left\langle \tilde{\mathcal{J}}' \left[\mathcal{B}(u,p) \right], \frac{d}{dp} \left(\mathcal{B} \circ (\mathcal{A}(p), p) \right) \delta p \right\rangle \\
&= \left\langle \tilde{\mathcal{J}}' \left[\mathcal{B}(u,p) \right], \left(\frac{\partial \mathcal{B}}{\partial p}[u,p] + \frac{\partial \mathcal{B}}{\partial u}[u,p] \frac{d\mathcal{A}}{dp} \right) \delta p \right\rangle \\
&= \left\langle \tilde{\mathcal{J}}' \left[\mathcal{B}(u,p) \right], \frac{\partial \mathcal{B}}{\partial p}[u,p]\delta p \right\rangle \\
&\qquad + \left\langle \tilde{\mathcal{J}}' \left[\mathcal{B}(u,p) \right], \frac{\partial \mathcal{B}}{\partial u}[u,p] \frac{d\mathcal{A}}{dp} \delta p \right\rangle
\end{aligned}$$

   Da η das adjungierte Problem (2.8) löst, erhalten wir

$$\begin{aligned}
\langle \mathcal{J}'[p], \delta p \rangle &= \left\langle \tilde{\mathcal{J}}' \left[\mathcal{B}(u,p) \right], \frac{\partial \mathcal{B}}{\partial p}[u,p]\delta p \right\rangle \\
&\qquad + \left\langle \eta, \frac{\partial \mathcal{G}}{\partial u}[u,p] \frac{d\mathcal{A}}{dp} \delta p \right\rangle .
\end{aligned} \qquad (2.10)$$

   Vollständiges Differenzieren von $\mathcal{G}\left(\mathcal{A}(p), p\right) = 0$ liefert

$$0 = \frac{d\mathcal{G}}{dp}\left(\mathcal{A}(p), p\right) = \frac{\partial \mathcal{G}}{\partial u}[u,p] \frac{d\mathcal{A}}{dp} + \frac{\partial \mathcal{G}}{\partial p}[u,p],$$

   woraus

$$\frac{\partial \mathcal{G}}{\partial u}[u,p] \frac{d\mathcal{A}}{dp} = -\frac{\partial \mathcal{G}}{\partial p}[u,p]$$

   folgt. Das Einsetzen in (2.10) ergibt die Behauptung

$$\langle \mathcal{J}'[p], \delta p \rangle = \left\langle \tilde{\mathcal{J}}' \left[\mathcal{B}(u,p) \right], \frac{\partial \mathcal{B}}{\partial p}[u,p]\delta p \right\rangle - \left\langle \eta, \frac{\partial \mathcal{G}}{\partial p}[u,p]\delta p \right\rangle . \qquad (2.11)$$

□

Bemerkung 2.4 Unter Aufspaltung des Fehlerfunktional $\mathcal{J}$ in

$$\mathcal{J}(p) = \sum_{k=1}^{\kappa} \tilde{\mathcal{J}}_k \circ \mathcal{B}_k \circ (\mathcal{A}(p), p)$$

ist die Ableitung unter den entsprechenden Voraussetzungen gegeben durch

$$\langle \mathcal{J}'[p], \delta p \rangle = \sum_{k=1}^{\kappa} \left\langle \tilde{\mathcal{J}}_k' \left[\mathcal{B}_k(u,p) \right], \frac{\partial \mathcal{B}_k}{\partial p}[u,p]\delta p \right\rangle - \left\langle \eta, \frac{\partial \mathcal{G}}{\partial p}[u,p]\delta p \right\rangle$$

für alle $\delta p \in P$, wenn $\eta \in V$ Lösung von

$$\left\langle \eta, \frac{\partial \mathcal{G}}{\partial u}[u,p]\delta u \right\rangle = \sum_{k=1}^{\kappa} \left\langle \tilde{\mathcal{J}}_k' \left[\mathcal{B}_k(u,p) \right], \frac{\partial \mathcal{B}_k}{\partial u}[u,p]\delta u \right\rangle$$

für alle $\delta u \in U$ ist.

2.2.2 Der Fall des diskreten Problems

Für eine numerische Lösung wird die unendlichdimensionale Modellgleichung (2.6) durch ein Diskretisierungsverfahren in ein endlichdimensionales Problem überführt. Deshalb betrachten wir ein nichtlineares Gleichungssystem

$$\mathcal{G}(u,p) = 0 \tag{2.12}$$

für eine nichtlineare Abbildung

$$\mathcal{G} : \mathbb{R}^m \times \mathbb{R}^r \to \mathbb{R}^m$$

$$(u,p) \mapsto \mathcal{G}(u,p)$$

mit den Dimensionen $m, r \in \mathbb{N}$ und stetigen partiellen Ableitungen $\frac{\partial \mathcal{G}}{\partial u}$ und $\frac{\partial \mathcal{G}}{\partial p}$ als vereinfachtes Modell des direkten Problems. Die Lösung und die Parameter sind nun als reelle Vektoren $u \in \mathbb{R}^m$ bzw. $p \in P \subset \mathbb{R}^r$ gegeben.

Wir setzen wieder voraus, dass für eine offene Teilmenge $\tilde{P}$ von P das Gleichungssystem (2.12) für jedes $p \in \tilde{P}$ genau eine Lösung $u \in \mathbb{R}^m$ besitzt. Somit kann der direkte Lösungsoperator

$$\mathcal{A} : \tilde{P} \to \mathbb{R}^m$$

$$p \mapsto u$$

als implizite Funktion gemäß

$$\mathcal{G}(\mathcal{A}(p), p) = 0$$

definiert werden. Das Fehlerfunktional $\mathcal{J}$ sei analog zum kontinuierlichen Fall definiert als

$$\mathcal{J} : P \to \mathbb{R}_+$$

$$p \mapsto \tilde{\mathcal{J}} \circ \mathcal{B} \circ (\mathcal{A}(p), p) = \sum_{k=1}^{\kappa} \tilde{\mathcal{J}}_k \circ \mathcal{B} \circ (\mathcal{A}(p), p)$$

mit stetig differenzierbaren Beobachtungsoperatoren

$$\mathcal{B}_k : \mathbb{R}^m \times \mathbb{R}^r \to \mathbb{R}^{n_k}$$

$$(u, p) \mapsto \omega_k$$

und stetig differenzierbaren Funktionalen

$$\tilde{\mathcal{J}}_k : \mathbb{R}^{n_k} \to \mathbb{R}_+$$

für $k = 1, ..., \kappa$.

Der Gradient $\frac{d\mathcal{J}}{dp}$ des Fehlerfunktionals kann wie folgt berechnet werden: Zunächst wenden wir wieder die Kettenregel an und erhalten

$$\begin{aligned}
\frac{d\mathcal{J}}{dp} &= \frac{d}{dp}\left(\tilde{\mathcal{J}} \circ \mathcal{B} \circ (\mathcal{A}(p), p)\right) \\
&= \sum_{k=1}^{\kappa} \frac{d\tilde{\mathcal{J}}_k}{d\omega_k} \frac{d}{dp} \left(\mathcal{B}_k \circ (\mathcal{A}(p), p)\right) \\
&= \sum_{k=1}^{\kappa} \frac{d\tilde{\mathcal{J}}_k}{d\omega_k} \left(\frac{\partial \mathcal{B}_k}{\partial p} + \frac{\partial \mathcal{B}_k}{\partial u}\frac{d\mathcal{A}}{dp}\right) \\
&= \sum_{k=1}^{\kappa} \frac{d\tilde{\mathcal{J}}_k}{d\omega_k} \frac{\partial \mathcal{B}_k}{\partial p} + \sum_{k=1}^{\kappa} \frac{d\tilde{\mathcal{J}}_k}{d\omega_k} \frac{\partial \mathcal{B}_k}{\partial u} \frac{d\mathcal{A}}{dp}. \qquad (2.13)
\end{aligned}$$

Aus $\mathcal{G}(\mathcal{A}(p), p) = 0$ folgt durch vollständiges Differenzieren

$$0 = \frac{d\mathcal{G}}{dp} = \frac{\partial \mathcal{G}}{\partial u}\frac{d\mathcal{A}}{dp} + \frac{\partial \mathcal{G}}{\partial p}. \qquad (2.14)$$

Wenn $\frac{\partial \mathcal{G}}{\partial u}$ regulär ist, dann erhalten wir $\frac{d\mathcal{A}}{dp}$ aus (2.14) als

$$\frac{d\mathcal{A}}{dp} = -\left(\frac{\partial \mathcal{G}}{\partial u}\right)^{-1} \frac{\partial \mathcal{G}}{\partial p}$$

und folglich

$$\frac{d\mathcal{J}}{dp} = \sum_{k=1}^{\kappa} \frac{d\tilde{\mathcal{J}}_k}{d\omega_k} \left(\frac{\partial \mathcal{B}_k}{\partial p} - \frac{\partial \mathcal{B}_k}{\partial u}\left(\frac{\partial \mathcal{G}}{\partial u}\right)^{-1} \frac{\partial \mathcal{G}}{\partial p}\right).$$

Eine andere Möglichkeit besteht darin, wie im kontinuierlichen Fall ein adjungiertes Problem zu betrachten:
Finde ein $\eta \in \mathbb{R}^m$, welches

$$\eta^T \frac{\partial \mathcal{G}}{\partial u} = \sum_{k=1}^{\kappa} \frac{d\tilde{\mathcal{J}}_k}{d\omega_k} \frac{\partial \mathcal{B}_k}{\partial u} \tag{2.15}$$

löst.

Nach (2.14) gilt

$$\frac{\partial \mathcal{G}}{\partial u} \frac{d\mathcal{A}}{dp} = -\frac{\partial \mathcal{G}}{\partial p}.$$

Aus dieser Gleichung, (2.13) und (2.15) folgt somit

$$\begin{aligned} \frac{d\mathcal{J}}{dp} &= \sum_{k=1}^{\kappa} \frac{d\tilde{\mathcal{J}}_k}{d\omega_k} \frac{\partial \mathcal{B}_k}{\partial p} + \sum_{k=1}^{\kappa} \frac{d\tilde{\mathcal{J}}_k}{d\omega_k} \frac{\mathcal{B}_k}{\partial u} \frac{d\mathcal{A}}{dp} \\ &= \sum_{k=1}^{\kappa} \frac{d\tilde{\mathcal{J}}_k}{d\omega_k} \frac{\partial \mathcal{B}_k}{\partial p} + \eta^T \frac{\partial \mathcal{G}}{\partial u} \frac{d\mathcal{A}}{dp} \\ &= \sum_{k=1}^{\kappa} \frac{d\tilde{\mathcal{J}}_k}{d\omega_k} \frac{\partial \mathcal{B}_k}{\partial p} - \eta^T \frac{\partial \mathcal{G}}{\partial p}. \end{aligned} \tag{2.16}$$

Dieses Ergebnis kann auch direkt aus Bemerkung 2.4 abgeleitet werden.

2.3 Verallgemeinerte Integralidentitäten und Identifizierbarkeit

Die Methode der Integralidentitäten wurde in [6] zur Untersuchung der Identifizierbarkeit im Sinne von Definition 2.1 von Koeffizientenfunktionen bei partiellen Differentialgleichungen entwickelt. Es wurde die Äquivalenz von Optimierungs- und Identifizierungsproblemen gezeigt. Siehe hierzu [9], [10], [11], [12]. Eine etwas allgemeinere Form dieser Methode ist in [30] dargelegt.

Das Prinzip der Methode besteht darin, einen nichtlinearen Operator $\mathcal{G}$, welcher als schwache Formulierung einer partiellen Differentialgleichung angesehen werden kann, in zwei Operatoren $\delta\mathcal{G}_p$ und $\delta\mathcal{G}_u$ aufzuspalten. Diese Operatoren, Finite-Differenzen-Versionen von $\frac{\partial \mathcal{G}}{\partial p}\delta p$ und $\frac{\partial \mathcal{G}}{\partial u}\delta u$, werden zur Definition adjungierter Probleme herangezogen, über welche die Identifizierbarkeit gezeigt wird.

Wir verwenden im Weiteren die Bezeichnungen aus Unterabschnitt 2.2.1.

Definition 2.5 *Wir definieren für $k = 1, ..., \kappa$ die Operatoren*

$$\delta\mathcal{B}_k : \tilde{P} \times \tilde{P} \to W_k$$

durch

$$\delta\mathcal{B}_k(p_1, p_2) := \mathcal{B}_k(\mathcal{A}(p_1), p_1) - \mathcal{B}_k(\mathcal{A}(p_2), p_2)$$

mit $p_1, p_2 \in \tilde{P}$. Der Operator

$$\delta\mathcal{B} : \tilde{P} \times \tilde{P} \to W$$

ist definiert durch

$$\delta\mathcal{B}(p_1, p_2) := \omega_1 - \omega_2 = \mathcal{B}(\mathcal{A}(p_1), p_1) - \mathcal{B}(\mathcal{A}(p_2), p_2) = \begin{pmatrix} \delta\mathcal{B}_1(p_1, p_2) \\ \vdots \\ \delta\mathcal{B}_\kappa(p_1, p_2) \end{pmatrix}.$$

$\delta\mathcal{B}$ ist eine Finite-Differenzen-Version von $\frac{d\mathcal{B}}{dp}\delta p$ und die Identifizierbarkeit der Parameter $p \in \tilde{P}$ aus einer Beobachtung $\omega = \mathcal{B}(u, p)$ bedeutet, dass die Implikation

$$p_1 \neq p_2 \to \delta\mathcal{B}(p_1, p_2) \neq 0$$

für alle $p_1, p_2 \in \tilde{P}$ gilt. Dabei ist zu beachten:

- $\delta\mathcal{B}(p_1, p_2) \neq 0 \iff \exists k \in \{1, ..., \kappa\} : \delta\mathcal{B}_k(p_1, p_2) \neq 0$
- $\delta\mathcal{B}(p_1, p_2) = 0 \iff \forall k \in \{1, ..., \kappa\} : \delta\mathcal{B}_k(p_1, p_2) = 0$

Definition 2.6 *Seien $p_1, p_2 \in \tilde{P}$ zulässige Parameter und $u_1 = \mathcal{A}(p_1), u_2 = \mathcal{A}(p_2)$ die zugehörigen Lösungen.*

1. *Der (nichtlineare) Operator*

$$\delta\mathcal{G}_p : \tilde{P} \times \tilde{P} \to V^*$$

wird definiert durch

$$\begin{aligned} \delta\mathcal{G}_p(p_1, p_2) &:= \mathcal{G}(u_2, p_1) - \mathcal{G}(u_2, p_2) \\ &= \mathcal{G}(\mathcal{A}(p_2), p_1) - \mathcal{G}(\mathcal{A}(p_2), p_2) \end{aligned}$$

für alle $p_1, p_2 \in \tilde{P}$.

2. *Ein (nichtlinearer) Operator*

$$\mathcal{N} : \tilde{P} \times \tilde{P} \to V^*$$

heißt $\mathcal{G}$-u-adjungierter Operator*, falls gilt*

$$\begin{aligned} \mathcal{N}(p_1, p_2) &= \mathcal{G}(u_1, p_1) - \mathcal{G}(u_2, p_1) \\ &= \mathcal{G}(\mathcal{A}(p_1), p_1) - \mathcal{G}(\mathcal{A}(p_2), p_1). \end{aligned}$$

3. *Sei $\delta\mathcal{G}_u$ ein $\mathcal{G}$-u-adjungierter Operator. Dann ist $\eta_k \in V$ eine* Lösung des $\delta\mathcal{G}_u$-adjungierten Problems *für $\mathcal{B}_k$ und ein $\omega_k^* \in W_k^*$, wenn gilt*

$$\langle \delta\mathcal{G}_u(p_1, p_2), \eta_k \rangle = \langle \omega_k^*, \delta\mathcal{B}_k(p_1, p_2) \rangle. \tag{2.17}$$

Bemerkung 2.7 Wenn $\frac{\partial \mathcal{G}}{\partial u}$ stetig ist, so erhält man einen $\mathcal{G}$-u-adjungierten Operator durch

$$\delta\mathcal{G}_u(p_1, p_2) := \int_0^1 \frac{\partial \mathcal{G}}{\partial u}[u_2 + s(u_1 - u_2), p_1]\, ds \cdot (u_1 - u_2).$$

Siehe z. B. Theorem 12.1.5 in [62].

Mit den obigen Definitionen kann nun ein Lemma formuliert werden.

Lemma 2.8 (Verallgemeinerte Integralidentitäten)
Seien $p_1, p_2 \in \tilde{P}$ zulässige Parameter, $u_1 = \mathcal{A}(p_1)$, $u_2 = \mathcal{A}(p_2)$ die zugehörigen Lösungen und $\delta\mathcal{G}_u$ ein $\mathcal{G}$-u-adjungierter Operator. Wenn eine Lösung $\eta_k \in V$ des $\delta\mathcal{G}_u$-adjungierten Problems (2.17) für $\mathcal{B}_k$ und ein $\omega_k^* \in W_k^*$ existiert, so gilt

$$\langle \delta\mathcal{G}_p(p_1, p_2), \eta_k \rangle = -\langle \omega_k^*, \delta\mathcal{B}_k(p_1, p_2) \rangle. \tag{2.18}$$

Beweis: Wegen $u_1 = \mathcal{A}(p_1)$ und $u_2 = \mathcal{A}(p_2)$ gilt zunächst

$$\mathcal{G}(u_2, p_2) = 0 = \mathcal{G}(u_1, p_1).$$

Aus der Definition von $\delta\mathcal{G}_p$ und $\delta\mathcal{G}_u$ folgt damit

$$\begin{aligned} \langle \delta\mathcal{G}_p(p_1, p_2), \eta_k \rangle &= \langle \mathcal{G}(u_2, p_1) - \mathcal{G}(u_2, p_2), \eta_k \rangle \\ &= \langle \mathcal{G}(u_2, p_1) - \mathcal{G}(u_1, p_1), \eta_k \rangle \\ &= -\langle \delta\mathcal{G}_u(p_1, p_2), \eta_k \rangle \\ &= -\langle \omega_k^*, \delta\mathcal{B}_k(p_1, p_2) \rangle, \end{aligned}$$

denn η_k ist Lösung von (2.17). □

Bemerkung 2.9 Die Bezeichnung von Gleichung (2.18) als Integralidentität ist dadurch motiviert, dass in [9], [10], [11] und [12], wo die Gültigkeit dieser Gleichung für Spezialfälle gezeigt wird, die beiden Seiten der Gleichung durch Integrale repräsentiert werden.

Mit (2.18) haben wir eine Gleichung erhalten, die eine Relation zwischen Änderungen in den Parametern, repräsentiert durch $\delta\mathcal{G}_p$, und Änderungen in den Beobachtungen, repräsentiert durch $\delta\mathcal{B}_k$, darstellt. Hierauf beruht die Grundidee des Beweises der Invertierbarkeit von „Koeffizienten-zu-Daten"-Abbildungen. Die Hauptschwierigkeit liegt darin, geeignete adjungierte Probleme gemäß Definition 2.6.3 zu finden. Wenn derartige adjungierte Probleme existieren, dann können hinreichende Bedingungen für die Identifizierbarkeit angegeben werden.

Satz 2.10 (Identifizierbarkeit)

1. *Zwei Lösungen $u_1 = \mathcal{A}(p_1)$ und $u_2 = \mathcal{A}(p_2)$ des direkten Problems für die Parameter $p_1, p_2 \in \tilde{P}$ sind voneinander verschieden ($u_1 \neq u_2$), wenn gilt*

$$\delta\mathcal{G}_p(p_1, p_2) \neq 0.$$

2. *Es sei $\delta\mathcal{G}_u$ ein $\mathcal{G}$-u-adjungierter Operator und für jedes $k \in \{1, ..., \kappa\}$ existiere eine Lösung $\eta_k \in V$ des $\delta\mathcal{G}_u$-adjungierten Problems (2.17) für $\mathcal{B}_k$ und ein $0 \neq \omega_k^* \in W_k^*$.*

   (a) *Falls für ein $k \in \{1, ..., \kappa\}$ und alle $p_1, p_2 \in \tilde{P}$ mit $p_1 \neq p_2$*

$$\langle \delta\mathcal{G}_p(p_1, p_2), \eta_k \rangle \neq 0$$

   gilt, dann sind die Parameter $p \in \tilde{P}$ aus der Beobachtung $\omega_k = \mathcal{B}_k(u, p)$ und damit auch aus der Beobachtung $\omega = \mathcal{B}(u, p)$ identifizierbar.

   (b) *Wenn für alle $p_1, p_2 \in \tilde{P}$ aus*

$$\langle \delta\mathcal{G}_p(p_1, p_2), \eta_k \rangle = 0 \quad \text{für alle } k \in \{1, ..., \kappa\}$$

   $p_1 = p_2$ folgt, so sind die Parameter $p \in \tilde{P}$ aus der Beobachtung $\omega = \mathcal{B}(u, p)$ identifizierbar.

Beweis:

1. Aus der Definition von $\delta\mathcal{G}_p$ und $\mathcal{G}(u_1, p_1) = \mathcal{G}(u_2, p_2) = 0$ folgt

$$\begin{aligned} 0 \neq \delta\mathcal{G}_p(p_1, p_2) &= \mathcal{G}(u_2, p_1) - \mathcal{G}(u_2, p_2) \\ &= \mathcal{G}(u_2, p_1) - \mathcal{G}(u_1, p_1) \\ &= \mathcal{G}(u_2, p_1) \end{aligned}$$

   und damit $u_1 \neq u_2$.

2. (a) Nach Lemma 2.8 gilt für $p_1, p_2 \in \tilde{P}$ mit $p_1 \neq p_2$

$$0 \neq \langle \delta \mathcal{G}_p(p_1, p_2), \eta_k \rangle = - \langle \omega_k^*, \delta \mathcal{B}_k(p_1, p_2) \rangle ,$$

also $\delta \mathcal{B}_k(p_1, p_2) \neq 0$ und damit auch $\delta \mathcal{B}(p_1, p_2) \neq 0$.

(b) Es sei angenommen, dass die Parameter der Menge $\tilde{P}$ nicht identifizierbar sind aus der Beobachtung $\omega = \mathcal{B}(u, p)$. Dann existieren Parameter $p_1, p_2 \in \tilde{P}$ mit $p_1 \neq p_2$ und $\delta \mathcal{B}_k(p_1, p_2) = 0$ für alle $k \in \{1, ..., \kappa\}$. Aus Lemma 2.8 folgt

$$\langle \delta \mathcal{G}_p(p_1, p_2), \eta_k \rangle = 0 \quad \text{für alle } k \in \{1, ..., \kappa\},$$

womit nach Voraussetzung $p_1 = p_2$ gilt. Dies ist ein Widerspruch zur Annahme.

□

Bemerkung 2.11 Die Aussage 2.a aus Satz 2.10 ist eine Spezialisierung von Aussage 2.b. Wenn die Voraussetzung von 2.a gilt, so ist automatisch auch die Voraussetzung von 2.b erfüllt.

Der obige Satz kann wie folgt interpretiert werden: Wenn geeignete adjungierte „Steuerungen“ $\omega_k^* \in W_k^*$ existieren, sodass die Lösungen η_k der zugehörigen adjungierten Probleme gewisse Bedingungen erfüllen, dann spiegeln sich Änderungen der Parameter in den Beobachtungen wider. Es besteht eine Art Äquivalenz zwischen der „Beobachtbarkeit“ des direkten Problems (Identifizierbarkeit) und der „Steuerbarkeit“ des adjungierten Problems.

Wenn es sich bei den Parametern $p \in \tilde{P}$ um Koeffizientenfunktionen handelt, so wird eine Änderung der Funktionen nur dann zu sichtbaren Änderungen in den Beobachtungen führen, wenn diese Funktionen auf eine sinnvolle Weise vergleichbar und unterscheidbar sind. In unseren Anwendungen betrachten wir stetige und skalare Funktionen, bei denen die Begriffe der Vergleichbarkeit und Unterscheidbarkeit gemäß Definition 3.5.5 in [30] geeignet sind.

Definition 2.12 (Vergleich- und Unterscheidbarkeit von Funktionen)
Es seien gegeben eine zusammenhängende Teilmenge I von $\mathbb{R}$ und zwei reelle Zahlen $a < b$, $a, b \in I$.

1. *Zwei Funktionen $f, g \in C(I)$ sind* vergleichbar *auf $[a, b]$, wenn die Menge*

$$\{x \in [a, b] \mid f(x) = g(x)\}$$

eine endliche Vereinigung von zusammenhängenden Teilmengen von $[a, b]$ ist.

2. *Zwei Funktionen* $f, g \in C(I)$ *sind* unterscheidbar *auf* $[a, b]$, *wenn sie*

   (a) *vergleichbar auf* $[a, b]$ *sind und*

   (b) *für mindestens ein* $x \in [a, b]$

   $$f(x) \neq g(x)$$

   gilt.

Da für stetige Funktionen f und g die Menge

$$\{x \in [a, b] \mid f(x) = g(x)\}$$

abgeschlossen und demzufolge

$$\{x \in (a, b) \mid f(x) \neq g(x)\}$$

eine offene Menge ist, erhalten wir

Folgerung 2.13 *Es seien* $f, g \in C(I)$ *unterscheidbar auf* $[a, b]$. *Dann existiert eine endliche Anzahl von offenen und zusammenhängenden Intervallen* $I_j \subset [a, b]$ $(j = 1, ..., n)$, *sodass* $f(x) = g(x)$ *für alle* $x \in [a, b] \setminus \bigcup_{j=1}^{n} I_j$ *und für alle* $j = 1, ..., n$ *entweder* $f > g$ *auf* I_j *oder* $f < g$ *auf* I_j *gilt.*

Bemerkung 2.14 Wenn f und g vergleichbare Funktionen sind, so schneidet bzw. berührt die Funktion $f - g$ die Gerade $y = 0$ nur in endlich vielen Stellen.

Beispiel 2.15 Funktionen, die nicht unterscheidbar sind, müssen nicht notwendig in allen Punkten ihres Definitionsbereichs gleich sein. Ein Beispiel sind die beiden Funktionen $f(x) = 0$ und $g(x) = x^n \sin\left(\frac{1}{x}\right)$ auf einem Intervall $[a, b]$ mit $0 \in [a, b]$. Obwohl beide Funktionen auf $[a, b]$ beschränkte und stetige Ableitungen bis zur Ordnung $n - 1$ besitzen, sind f und g auf $[a, b]$ nicht vergleichbar. Vergleichbarkeit wird also keineswegs durch Glattheitseigenschaften impliziert.

Beispiel 2.16 Wir betrachten auf dem Intervall $[0, 1]$ die Funktionen $f(x) = \sin\left(\frac{1}{x}\right)$, $g(x) = \sin\left(\frac{1}{x}\right) + 2$ und $h(x) = 0$. Die Funktionen f und g sind ebenso wie g und h auf $[0, 1]$ vergleichbar. Die Funktionen f und h sind aber nicht vergleichbar auf $[0, 1]$. D. h., die Vergleichbarkeit auf einem Intervall $[a, b]$ erzeugt keine Äquivalenzrelation in $C[a, b]$.

Im Folgenden definieren wir Funktionenräume, in denen die Vergleichbarkeit zu einer Äquivalenzrelation führt.

Definition 2.17 *Eine Funktion $f \in C[a,b]$ heißt* stückweise holomorph, *wenn eine endliche Zerlegung $a = x_0 < ... < x_n = b$ von $[a,b]$ existiert, sodass für alle $i = 1, ..., n$ die Einschränkung von f auf das Teilintervall $[x_{i-1}, x_i]$ holomorph ist. Demgemäß definieren wir*

$$C_{\text{pwa}}[a,b] := \{f \in C[a,b] \mid f \text{ ist stückweise holomorph}\}$$

und

$$C^1_{\text{pwa}}[a,b] := \left\{f \in C^1[a,b] \mid f' \text{ ist stückweise holomorph}\right\}.$$

Es ist klar, dass gilt $C^1_{\text{pwa}}[a,b] \subset C_{\text{pwa}}[a,b]$. Darüber hinaus gelten die folgenden Aussagen:

Satz 2.18

1. *Zwei beliebige Funktionen $f, g \in C_{\text{pwa}}[a,b]$ sind vergleichbar auf $[a,b]$.*
2. *Zwei Funktionen $f, g \in C_{\text{pwa}}[a,b]$ sind genau dann unterscheidbar auf $[a,b]$, wenn $f \neq g$ in $[a,b]$ im üblichen Sinne.*
3. *Seien $f, g \in C^1_{\text{pwa}}[a,b]$ unterscheidbar auf $[a,b]$ mit $f(\tilde{x}) = g(\tilde{x})$ für ein $\tilde{x} \in [a,b]$. Dann sind deren Ableitungen f' und g' ebenfalls unterscheidbar auf $[a,b]$.*
4. *Seien $f, g \in C_{\text{pwa}}[a,b]$ unterscheidbar auf $[a,b]$. Dann sind die Funktionen $F(x) := \int_a^x f(y)dy$ und $G(x) := \int_a^x g(y)dy$ unterscheidbar auf $[a,b]$.*

Beweis:

1. Da f und g stückweise holomorph auf $[a,b]$ sind, existiert eine endliche Zerlegung $a = x_0 < ... < x_n = b$, sodass für alle $i = 1, ..., n$ die Einschränkung $h_i := (f-g)|_{[x_{i-1},x_i]}$ holomorph ist. Nach einem Resultat aus der Funktionentheorie (siehe z. B. Kapitel 8, §1.3 in [45]) ist damit h_i entweder konstant oder die Menge $\{x \in [x_{i-1}, x_i] \mid h_i(x) = \alpha\}$ ist für jedes $\alpha \in \mathbb{R}$ endlich. Daraus folgt aber, dass die Menge

$$\{x \in [x_{i-1}, x_i] \mid f(x) = g(x)\}$$

entweder leer, gleich dem Intervall $[x_{i-1}, x_i] \subseteq [a,b]$ oder eine endliche Vereinigung von zusammenhängenden, abgeschlossenen Mengen $\left\{x^k\right\} \subset [a,b]$ ist. Also ist $\{x \in [a,b] \mid f(x) = g(x)\}$ eine endliche Vereinigung von zusammenhangenden Teilmengen von $[a,b]$, womit f und g auf $[a,b]$ vergleichbar sind.

2. Die Behauptung folgt sofort aus der Vergleichbarkeit der beiden Funktionen.

3. Nach 1. sind f' und g' auf $[a,b]$ vergleichbar. Angenommen $f'(x) = g'(x)$ für alle $x \in [a,b]$. Dann ist $f - g$ konstant, also entweder

   (a) $f(x) = g(x)$ für alle $x \in [a,b]$ oder

   (b) $f(x) \neq g(x)$ für alle $x \in [a,b]$.

   Der Fall (a) kann nicht eintreten, da f und g auf $[a,b]$ unterscheidbar sind. Der Fall (b) steht im Widerspruch zur Voraussetzung $f(\tilde{x}) = g(\tilde{x})$ für ein $\tilde{x} \in [a,b]$. Also folgt, dass $f'(x) \neq g'(x)$ für mindestens ein $x \in [a,b]$, d. h. f' und g' sind unterscheidbar auf $[a,b]$.

4. Zunächst folgt $F, G \in C^1_{\text{pwa}}[a,b]$, womit F und G vergleichbar sind. Da f und g auf $[a,b]$ unterscheidbar sind, existiert eine endliche Zerlegung $a = x_0 < ... < x_n = b$ von $[a,b]$, sodass in jedem offenen Teilintervall (x_{i-1}, x_i), $i = 1, ..., n$, entweder $f = g$, $f > g$ oder $f < g$ und in mindestens einem offenen Teilintervall die Ungleichheit gilt. O. B. d. A. sei $f > g$ in (x_0, x_1). Dann folgt

$$F(x) - G(x) = \int_a^x (f(y) - g(y))dy > 0$$

   für $x \in (x_0, x_1]$. Folglich sind F und G unterscheidbar auf $[a,b]$.

□

Bemerkung 2.19 Die oben definierten Räume enthalten insbesondere alle stetigen bzw. stetig differenzierbaren, stückweisen Polynomfunktionen.

Kapitel 3

Modell zur Identifizierung der hydraulischen Funktionen

3.1 Mathematische Beschreibung der Säulenexperimente

Der Fluidtransport durch das in der Säule befindliche poröse Medium wird beschrieben durch die *Richards-Gleichung* in der Druckformulierung. Dieses Modell setzt sich zusammen aus der Volumenerhaltungsgleichung

$$\partial_t \Theta(\psi(x,t)) + \nabla \cdot q(x,t) = 0 \tag{3.1}$$

und dem Darcy-Gesetz

$$q(x,t) = -K(\psi(x,t))\nabla(\psi(x,t) + z) \tag{3.2}$$

für $(x,t) \in Q_T$, $Q_T := \Omega \times (0,T)$. Hierbei bezeichnen

- ψ den Druck, skaliert auf die Höhe einer entsprechenden Wassersäule [Länge],
- q die volumetrische Fließrate [Länge/Zeit] und
- z die Höhe entgegen der Gravitationsrichtung [Länge].

Die Koeffizientenfunktionen $\Theta(\psi)$ und $K(\psi)$ beschreiben die hydraulischen Eigenschaften des porösen Mediums. Die Retentionsfunktion $\Theta(\psi)$ [-] stellt eine Beziehung zwischen Druck und Fluidgehalt θ her. $K(\psi)$ [Länge/Zeit] gibt die hydraulische Leitfähigkeit in Abhängigkeit vom Druck wieder.

Der ungesättigte Bereich, wo neben dem Fluid auch Luft im Porenraum vorhanden ist, wird charakterisiert durch negative Drücke $\psi < 0$. In diesem Bereich sind die hydraulischen Funktionen Θ und K nichtnegativ und monoton wachsend. Im gesättigten Bereich, wo $\psi \geq 0$ ist, sind diese Funktionen

konstant:

$$\left.\begin{array}{l}\Theta(\psi) = \theta_{\text{sat}} \\ K(\psi) = K_{\text{sat}}\end{array}\right\} \text{ für } \psi \geq 0.$$

Folglich handelt es sich bei der Richards-Gleichung um eine elliptisch-parabolisch degenerierte Gleichung. Die Hysterese der hydraulischen Funktionen (vgl. z. B. [58]), die in der Natur vorzufinden ist, wird in unserem Modell nicht berücksichtigt.

Das Modell zur Beschreibung des Säulenexperiments wird vervollständigt durch die Angabe einer Anfangsbedingung

$$\psi(x,0) = \psi_0(x) \quad \text{in } \Omega \tag{3.3}$$

und zwei Typen von Randbedingungen auf dem Rand $\Gamma = \partial\Omega$. Der Ausflussrand, bezeichnet mit Γ_D, wird durch eine Dirichlet-Randbedingung modelliert:

$$\psi(x,t) = g(x,t) \quad \text{auf } \Gamma_{DT} := \Gamma_D \times (0,T). \tag{3.4}$$

Für den restlichen Teil des Randes, bezeichnet mit Γ_F, wird eine homogene Fluss-Randbedingung angesetzt, um die Isolation dieses Teils des Randes zu beschreiben:

$$q(x,t) \cdot \nu = 0 \quad \text{auf } \Gamma_{FT} := \Gamma_F \times (0,T). \tag{3.5}$$

Dabei bezeichnet ν die Außennormale.

Die Beobachtung sind gegeben durch den Druck auf $\tilde{\Gamma}_F \subseteq \Gamma_F$

$$\omega_{\tilde{\Gamma}_{FT}}(x,t) = \psi(x,t) \quad \text{auf } \tilde{\Gamma}_{FT} := \tilde{\Gamma}_F \times (0,T) \tag{3.6}$$

oder im Inneren des Gebietes

$$\omega_{\tilde{Q}_T}(x,t) = \psi(x,t) \quad \text{in } \tilde{Q}_T := \tilde{\Omega} \times (0,T) \tag{3.7}$$

mit $\tilde{\Omega} \subset \Omega$ sowie den kumulativen Ausfluss auf $\tilde{\Gamma}_D \subset \Gamma_D$

$$\omega_{\tilde{\Gamma}_{DT}}(x,t) = \int_0^t q(x,\tau) \cdot \nu \, d\tau \quad \text{auf } \tilde{\Gamma}_{DT} := \tilde{\Gamma}_D \times (0,T). \tag{3.8}$$

Das Lösen des Modells (3.1–3.5) für gegebene Funktionen Θ und K liefert eindeutige Beobachtungen $\omega_{\tilde{\Gamma}_D}$, $\omega_{\tilde{Q}_T}$ und $\omega_{\tilde{\Gamma}_F}$ und wird als das direkte Problem bezeichnet, die Simulation des Säulenexperiments. Die experimentellen Bedingungen erlauben es, sich bei der Simulation auf das entsprechende räumlich eindimensionale Modell zu beschränken. Wir werden im Weiteren dennoch das räumlich mehrdimensionale Modell und dessen Diskretisierung betrachten und uns nur dort auf den eindimensionalen Fall beschränken, wo es erforderlich ist. Für hinreichend glatte Koeffizienten Θ und K hat das

direkte Probleme eine eindeutige glatte Lösung, sofern die Lösung aufgrund der Anfangs- und Randbedingungen $\psi \leq \underline{\psi}$ für ein $\underline{\psi} < 0$ erfüllt, d. h. überall ungesättigt ist (siehe z. B. [37]). Im allgemeinen gesättigt-ungesättigten Fall kann wegen der elliptisch-parabolischen Degeneration nur eine eindeutige schwache Lösung geringerer Regularität garantiert werden ([2], [44]). Speziell die Zeitableitungen besitzen nur eine L^2-Regularität.

Die hydraulischen Funktionen werden häufig durch empirische Modelle beschrieben. Eines der in der Bodenphysik gängigsten Modelle stammt von van Genuchten und Mualem [60]. Für $\psi < 0$ lautet dies:

$$\Theta(\psi) = \theta_{\text{res}} + (\theta_{\text{sat}} - \theta_{\text{res}}) \left(\frac{1}{1 + (-\alpha\psi)^n} \right)^{\frac{n-1}{n}} \tag{3.9}$$

$$K(\psi) = K_{\text{sat}} \frac{\left(1 - (-\alpha\psi)^{n-1}(1 + (-\alpha\psi)^n)^{\frac{1-n}{n}}\right)^2}{(1 + (-\alpha\psi)^n)^{\frac{n-1}{2n}}}. \tag{3.10}$$

Hierbei sind α und n rein empirisch motivierte Parameter. Daneben existiert eine Vielzahl weiterer Modelle (siehe z. B. [28]). Durch die Auswahl eines dieser Modelle wird die prinzipielle Form, insbesondere das Krümmungsverhalten, der hydraulischen Funktionen a priori festgelegt. Um dies zu vermeiden, werden wir einen allgemeineren formfreien Ansatz verwenden, bei dem Θ und K als Funktionen identifiziert werden. Es werden nur die folgenden (physikalisch gerechtfertigten) Annahmen gemacht, die im Weiteren stets erfüllt sein sollen.

Annahme 3.1

1. $\Theta \in C^1(-\infty, 0]$ *mit* $0 \leq \Theta(\psi) \leq \theta_{\text{sat}}$ *und* $0 \leq \Theta'(\psi) \leq c$ *mit* $c > 0$.
2. $K \in C^1_{\text{pw}}(-\infty, 0]$ *mit* $0 < K(\psi) \leq K_{\text{sat}}$ *und* $0 \leq K'(\psi) \leq k$ *mit* $k > 0$.
3. $\psi_0 \in C^1(\Omega)$ *mit* $\psi_0 \geq 0$ *und* $\nabla(\psi_0 + z) = 0$ *in* Ω *(d. h. verschwindender Fluss im Anfangszustand).*
4. $g \in C^1(\Gamma_{DT})$ *mit* $g(x, 0) = \psi_0(x)$ *auf* Γ_D.

3.2 Schlechtgestelltheit des inversen Problems

Das inverse Problem besteht darin, aus experimentellen Daten für die Beobachtungen $\omega_{\tilde{\Gamma}_F}$, $\omega_{\tilde{Q}_T}$ und $\omega_{\tilde{\Gamma}_D}$ die zugehörigen Koeffizienten Θ und K zu rekonstruieren. Im Folgenden seien einige Beispiele für die Schlechtgestelltheit des inversen Problems aufgeführt. Hierzu betrachten wir das räumlich

eindimensionale Modell

$$\begin{aligned} \partial_t \Theta(\psi) - \partial_x \left(K(\psi)(\partial_x \psi - 1)\right) &= 0 \quad \text{in } (0,L)\times(0,T), && (3.11)\\ \psi(x,0) &= x \quad \text{in } (0,L), && (3.12)\\ \psi(L,t) &= g(t) \quad \text{in } (0,T), && (3.13)\\ K(\psi(0,t))(\partial_x \psi(0,t) - 1) &= 0 \quad \text{in } (0,T). && (3.14) \end{aligned}$$

Hier ist $\Omega = (0, L)$ entgegen der Gravitationsrichtung orientiert, d. h. $x = 0$ ist das obere, $x = L$ das untere Ende der vertikal ausgerichteten Säule.

Beispiel 3.2 Es sei ψ Lösung von (3.11–3.14) für die Koeffizienten Θ_1 und K. Dann ist ψ ebenfalls Lösung von (3.11–3.14) für die Koeffizienten $\Theta_2 := \Theta_1 + a$ und K, wobei $a \in \mathbb{R}$ beliebig wählbar ist. Denn es gilt $\partial_t \Theta_1(\psi) = \partial_t \Theta_2(\psi)$.

Dies hat unmittelbar zur Konsequenz, dass aus den Beobachtungen $\omega_{\tilde{\Gamma}_F}$, $\omega_{\tilde{Q}_T}$ und $\omega_{\tilde{\Gamma}_D}$ nicht die Retentionsfunktion Θ identifizierbar ist, sondern nur die Kapazitätsfunktion $\Theta'(\psi)$. Um $\Theta(\psi)$ zu erhalten, muss z. B. der Wert θ_{sat} vorgegeben werden. Eine andere Möglichkeit besteht darin, Fluidgehaltsmessungen bei der Identifizierung zu berücksichtigen.

Beispiel 3.3 Sei $g'(t) < 0$ in $(0, T)$. Dann kann für beliebige Koeffizienten (Θ, K), die den Annahmen 3.1 genügen, wie in Lemma 3.4 in [9] gezeigt werden, dass für die Lösung ψ von (3.11–3.14)

$$g(T) + x - L \le \psi(x,t) \le x \quad \text{für } (x,t) \in Q_T \tag{3.15}$$

gilt. Wenn wir beispielsweise zwei Paare (Θ, K_1) und (Θ, K_2) mit

$$\begin{aligned} K_1(\psi) &= K_2(\psi) \quad \text{für } \psi \in [\underline{\psi}, \infty),\\ K_1(\psi) &\neq K_2(\psi) \quad \text{für } \psi \in (-\infty, \underline{\psi}) \end{aligned}$$

betrachten, wobei $\underline{\psi} := g(T) - L$, so folgt aus (3.15), dass der Druckbereich auf dem sich K_1 und K_2 unterscheiden für die Lösung des Problems (3.11–3.14) nicht relevant ist und somit für die zugehörigen Lösungen

$$\psi_1 = \psi_2$$

gilt. Die Identifizierung der hydraulischen Funktionen muss daher auf das Intervall $[\underline{\psi}, \infty)$ bzw. $[\underline{\psi}, 0]$ eingeschränkt werden.

Das folgende Beispiel zeigt, dass auch Messfehler eine Schlechtgestelltheit des inversen Problems verursachen können.

Beispiel 3.4 Es sei wieder $g'(t) < 0$ in $(0,T)$. Dann kann unter Annahme 3.1 gemäß Lemma 3.1 und Lemma 3.4 in [9] für (3.11–3.14) gezeigt werden, dass die Beobachtungen $\omega_{\Gamma_F}(t) = \psi(0,t)$ und $\omega_{\Gamma_D}(t) = \int_0^t q(L,\tau)\, d\tau$ den Bedingungen

$$\partial_t \omega_{\Gamma_F}(t) < 0 \quad \text{und} \quad \partial_t \omega_{\Gamma_D}(t) > 0$$

genügen. Für $\omega_{\tilde{Q}_T}(x,t) = \psi(x,t)$ folgt nach Lemma 3.3 aus [9]

$$\partial_t \omega_{\tilde{Q}_T}(t) < 0 \quad \text{(fast überall in } \tilde{Q}_T).$$

Wenn Messfehler zu nichtmonotonen Daten für den Druck und kumulativen Ausfluss führen, so existieren keine Funktionen Θ und K mit den Eigenschaften 3.1, die diese Messdaten reproduzieren können.

3.3 Kontinuierliches Modell

3.3.1 Variationsformulierung

Zunächst wird die Richards-Gleichung auf die Form gebracht, wie sie in der Arbeit von Alt und Luckhaus [2] verwendet wurde. Hierzu wird eine Kirchhoff-Transformation auf die neue Unbekannte

$$u = \mathcal{K}(\psi) := \int_0^\psi K(s) ds \tag{3.16}$$

durchgeführt. Wegen Annahme 3.1 ist $\mathcal{K}(\psi)$ streng monoton und damit invertierbar. Nun definieren wir

$$\begin{aligned} b(u) &:= \Theta \circ \mathcal{K}^{-1}(u), \\ k(s) &:= K \circ \Theta^{-1}(s), \\ b^0 &:= b(\mathcal{K}(\psi_0)). \end{aligned}$$

Weiterhin werden folgende Annahmen gemacht:

Annahme 3.5

1. *$\Omega \subset \mathbb{R}^N$ sei ein beschränktes, konvexes Gebiet mit Lipschitzrand $\partial\Omega = \Gamma_D \cup \Gamma_F$ mit Γ_D messbar.*
2. *Es existiere ein $u_D \in L^2((0,T); H^1(\Omega)) \cap L^\infty(Q_T)$ mit $u_D|_{\Gamma_D} = \mathcal{K}(g)$ und $\partial_t u_D \in L^1((0,T); L^\infty(\Omega))$.*
3. *Es existiere eine meßbare Funktion $u^0 \in L^\infty(\Omega)$.*

4. Die Funktionen b und k seien Lipschitz-stetig.

Bezeichnung 3.6 $H^1_{0,D}(\Omega) := \{u \in H^1(\Omega) \mid u|_{\Gamma_D} = 0\}$.

Gemäß Definition 3.9 in [52] und Definition 1.1 in [18] wird eine schwache Lösung des direkten Problems definiert:

Definition 3.7 *Eine Funktion u wird als schwache Lösung des direkten Problems (3.1–3.5) bezeichnet, falls*

1. $u - u_D \in L^2((0,T); H^1_{0,D}(\Omega))$,
2. $b(u) \in L^\infty(Q_T)$,
3. $$\int_{Q_T} (b(u) - b(u^0))\partial_t \eta - \int_{Q_T} (\nabla u + k(b(u))\nabla z) \cdot \nabla \eta = 0 \qquad (3.17)$$
   für alle $\eta \in L^2((0,T); H^1_{0,D}(\Omega))$ mit $\partial_t \eta \in L^\infty(Q_T)$ und $\eta(.,T) = 0$ fast überall in Ω.

Für die Existenz und Eindeutigkeit einer schwachen Lösung erhaltem wir nach Alt und Luckhaus ([2]) die folgende Aussage (vgl. Theorem 3.10 und Theorem 3.11 in [52]).

Lemma 3.8 *Unter den Annahmen 3.1 und 3.5 existiert eine eindeutige schwache Lösung des direkten Problems.*

Beweis: Siehe [2] Theorem 1.7 und Theorem 2.4. □

3.3.2 Gemischte Variationsformulierung

Im transformierten Problem gilt

$$q = -(\nabla u + k(b(u))\nabla z)$$

mit

$$\int_0^\cdot q \, d\tau \in H^{1,1}(Q_T)$$

nach Lemma 3.14 in [52]. Mithilfe dieser Flussvariablen wird nun eine gemischte Variationsformulierung für das rücktransformierte Problem aufgestellt.

Bezeichnung 3.9

1. $L^2_{0,D}((0,T);H^1(\Omega)) := \{\psi \in L^2((0,T);H^1(\Omega)) \mid \psi|_{\Gamma_D} = 0\}$.
2. $L^2_{g,D}((0,T);H^1(\Omega)) := \{\psi \in L^2((0,T);H^1(\Omega)) \mid \psi - g \in L^2_{0,D}((0,T);H^1(\Omega))\}$.
3. $H_{0,F}(\mathrm{div};\Omega) := \{q \in H(\mathrm{div};\Omega) \mid q \cdot \nu|_{\Gamma_F} = 0\}$.

Definition 3.10 *(u,q) ist eine schwache gemischte Lösung des direkten Problems, wenn*

1. $(\psi,q) \in L^2_{g,D}((0,T);H^1(\Omega)) \times L^2((0,T);H_{0,F}(\mathrm{div};\Omega))$,
2. $$-\int_\Omega \Theta(\psi_0)\eta(.,0) - \int_{Q_T} \Theta(\psi)\partial_t\eta + \int_{Q_T} \nabla \cdot q\eta = 0 \tag{3.18}$$
   für alle $\eta \in L^2(Q_T)$ mit $\partial_t\eta \in L^\infty(Q_T)$, $\eta(.,0) \in L^\infty(\Omega)$ und $\eta(.,T) = 0$ fast überall in Ω und
   $$\int_{Q_T} q \cdot v + \int_{Q_T} K(\psi)\nabla\psi \cdot v + \int_{Q_T} K(\psi)\nabla z \cdot v = 0 \tag{3.19}$$
   für alle $v \in (L^2(Q_T))^N$.

Nun können die in Kapitel 2 eingeführten Bezeichnungen für das kontinuierliche Modell spezifiziert werden.

Bezeichnung 3.11

1. $U := L^2_{g,D}((0,T);H^1(\Omega)) \times L^2((0,T);H_{0,F}(\mathrm{div};\Omega))$ *(Lösungsraum).*
2. $V := \{\eta \in L^2(Q_T) \mid \partial_t\eta \in L^\infty(Q_T), \eta(.,0) \in L^\infty(\Omega)\} \times (L^2(Q_T))^N$ *(Raum der Testfunktionen).*
3. (a) *$\tilde{\Gamma}_{FT}$ sei messbare Teilmenge von Γ_{FT}, $W_{\tilde{\Gamma}_{FT}} := L^2(\tilde{\Gamma}_{FT})$, $\beta_1 \in L^\infty(\Gamma_{FT})$ mit $\beta_1|_{\Gamma_{FT}\setminus\tilde{\Gamma}_{FT}} = 0$ und $\beta_1|_{\tilde{\Gamma}_{FT}} > 0$.*
   (b) *$\tilde{Q}_T$ sei messbare Teilmenge von Q_T, $W_{\tilde{Q}_T} := L^2(\tilde{Q}_T)$, $\beta_2 \in L^\infty(Q_T)$ mit $\beta_2|_{Q_T\setminus\tilde{Q}_T} = 0$ und $\beta_2|_{\tilde{Q}_T} > 0$.*
   (c) *$\tilde{\Gamma}_{DT}$ sei messbare Teilmenge von Γ_{DT}, $W_{\tilde{\Gamma}_{DT}} := L^2(\tilde{\Gamma}_{DT})$, $\gamma \in L^\infty(\Gamma_{DT})$ mit $\gamma|_{\Gamma_{DT}\setminus\tilde{\Gamma}_{DT}} = 0$ und $\gamma|_{\tilde{\Gamma}_{DT}} > 0$.*
   (d) *$\tilde{W}_{\tilde{\Gamma}_{DT}} := L^2(\tilde{\Gamma}_{DT})$.*

   Hierbei sind β_1, β_2 und γ die Gewichte der einzelnen Beobachtungen.

4. *Der Koeffizienten- bzw. Parameterraum P ist der Raum aller Funktionenpaare (Θ, K), die der Annahme 3.1 genügen.*

Die Beobachtungsoperatoren

$$\begin{aligned}
\mathcal{B}_{\tilde{\Gamma}_{FT}} : U &\to W_{\tilde{\Gamma}_{FT}} \\
(\psi, q) &\mapsto \beta_1 \psi|_{\tilde{\Gamma}_{FT}} \\
\mathcal{B}_{\tilde{Q}_T} : U &\to W_{\tilde{Q}_T} \\
(\psi, q) &\mapsto \beta_2 \psi|_{\tilde{Q}_T} \\
\mathcal{B}_{\tilde{\Gamma}_{DT}} : U &\to W_{\tilde{\Gamma}_{DT}} \\
(\psi, q) &\mapsto -\gamma \int_0^{\cdot} q|_{\tilde{\Gamma}_{DT}} \cdot \nu \, d\tau \\
\tilde{\mathcal{B}}_{\tilde{\Gamma}_{DT}} : U &\to \tilde{W}_{\tilde{\Gamma}_{DT}} \\
(\psi, q) &\mapsto \gamma q|_{\tilde{\Gamma}_{DT}} \cdot \nu
\end{aligned}$$

seien wohldefiniert.

Bemerkung 3.12 Ein Vorteil der eingeführten gemischten Formulierung besteht darin, dass wir ausschließlich lineare Beobachtungen berücksichtigen müssen, die explizit nicht von den Koeffizienten Θ und K abhängen.

Annahme 3.13

1. $\Theta^*, K^*, \mathcal{K}^* \in L^\infty(Q_T)$.
2. $F^* \in L^2(Q_T)$.
3. $g^* \in L^2(\Gamma_{DT})$.
4. $h^* \in L^2(\Gamma_{FT})$.

Ausgehend von der gemischten Form wird mit diesen Annahmen ein adjungiertes Problem definiert.

Definition 3.14 *(η, v) ist eine schwache gemischte Lösung des zum direkten Problem adjungierten Problems, falls*

1. $(\eta, v) \in V$ *mit* $\eta(., T) = 0$ *fast überall in* Ω,
2. $$-\int_{Q_T} \Theta^* \phi \partial_t \eta + \int_{Q_T} \nabla(\mathcal{K}^* \phi) \cdot v + \int_{Q_T} K^* \phi \nabla z \cdot v = \int_{Q_T} F^* \phi + \int_{\Gamma_{FT}} h^* \phi \tag{3.20}$$

für alle $\phi \in L^2_{0,D}((0,T);H^1(\Omega))$ *und*

$$\int_{Q_T} w \cdot v + \int_{Q_T} \nabla \cdot w\eta = \int_{\Gamma_{DT}} g^* w \cdot \nu \tag{3.21}$$

für alle $w \in L^2((0,T);H_{0,F}(\mathrm{div};\Omega))$.

Lemma 3.15 *Die nichtlineare Abbildung* $\mathcal{G} : U \times P \to V^*$ *sei gegeben durch*

$$\begin{aligned}\langle \mathcal{G}((u,q),(\Theta,K)),(\eta,v)\rangle = \\ -\int_{Q_T} \Theta(\psi)\partial_t\eta + \int_{Q_T} \nabla\cdot q\eta + \int_{Q_T} q\cdot v + \int_{Q_T} K(\psi)\nabla\psi\cdot v \\ + \int_{Q_T} K(\psi)\nabla z\cdot v - \int_\Omega \Theta(\psi_0)\eta(.,0)\end{aligned} \tag{3.22}$$

für $(\psi,q)\in U$, $(\Theta,K)\in P$ *und für alle* $(\eta,v)\in V$ *mit* $\eta(.,T)=0$ *fast überall in* Ω. *Dann ist* (ψ,q) *schwache gemischte Lösung des direkten Problems für die Koeffizienten* (Θ,K) *genau dann, wenn* (ψ,q) *die Gleichung*

$$\mathcal{G}((\psi,q),(\Theta,K)) = 0 \tag{3.23}$$

erfüllt.

Beweis: Sei (ψ,q) schwache gemischte Lösung des direkten Problems und $(\eta,v)\in V$ mit $\eta(.,T)=0$ fast überall in Ω. Wenn wir (3.18) und (3.19) addieren, so erfüllt die resultierende Gleichung (3.23).

Wenn (3.23) für alle Testfunktionen $(\eta,v)\in V$ mit $\eta(.,T)=0$ fast überall in Ω gültig ist, so auch für $(\eta,0)$ und $(0,v)$. Hieraus folgt (3.18) und (3.19). □

Bemerkung 3.16 In der Notation von Kapitel 2 sei jetzt stets $(\psi_i,q_i) = \mathcal{A}(\Theta_i,K_i)$, $i=1,2$.

Lemma 3.17

1. *Die Abbildung* $\delta\mathcal{G}_{(\Theta,K)} : P\times P \to V^*$ *erfüllt*

$$\begin{aligned}\left\langle \delta\mathcal{G}_{(\Theta,K)}((\Theta_1,K_1),(\Theta_2,K_2)),(\eta,v)\right\rangle = \\ -\int_{Q_T} (\Theta_1(\psi_2)-\Theta_2(\psi_2))\,\partial_t\eta + \int_{Q_T} (K_1(\psi_2)-K_2(\psi_2))\,\nabla\psi_2\cdot v \\ +\int_{Q_T} (K_1(\psi_2)-K_2(\psi_2))\,\nabla z\cdot v - \int_\Omega (\Theta_1(\psi_0)-\Theta_2(\psi_0))\,\eta(.,0)\end{aligned} \tag{3.24}$$

für alle $(\eta,v)\in V$ *mit* $\eta(.,T)=0$ *fast überall in* Ω.

2. Sei

$$\begin{aligned}\Theta^* &:= \int_0^1 \Theta_1'(\psi_2 + s(\psi_1 - \psi_2))\, ds,\\ K^* &:= \int_0^1 K_1'(\psi_2 + s(\psi_1 - \psi_2))\, ds,\\ \mathcal{K}^* &:= \int_0^1 K_1(\psi_2 + s(\psi_1 - \psi_2))\, ds.\end{aligned}$$

Die Abbildung $\delta\mathcal{G}_{(\psi,q)} : P \times P \to V^*$*, die durch*

$$\begin{aligned}&\left\langle \delta\mathcal{G}_{(\psi,q)}((\Theta_1, K_1), (\Theta_2, K_2)), (\eta, v)\right\rangle =\\ &\quad - \int_{Q_T} \Theta^*(\psi_1 - \psi_2)\partial_t\eta + \int_{Q_T} \nabla\cdot(q_1 - q_2)\eta\\ &\quad + \int_{Q_T} (q_1 - q_2)\cdot v + \int_{Q_T} \nabla\left(\mathcal{K}^*(\psi_1 - \psi_2)\right)\cdot v\\ &\quad + \int_{Q_T} K^*(\psi_1 - \psi_2)\nabla z \cdot v \qquad (3.25)\end{aligned}$$

für alle $(\eta, v) \in V$ *mit* $\eta(., T) = 0$ *fast überall in* Ω *definiert wird, ist ein* $\mathcal{G}$*-*(ψ, q)*-adjungierter Operator.*

Beweis:

1. Siehe Definition 2.6.1.

2. Nach Theorem 2.18 in [30] gilt

$$\begin{aligned}\Theta^*(\psi_1 - \psi_2) &= \Theta_1(\psi_1) - \Theta_1(\psi_2)\\ K^*(\psi_1 - \psi_2) &= K_1(\psi_1) - K_1(\psi_2)\\ \mathcal{K}^*(\psi_1 - \psi_2) &= \mathcal{K}_1(\psi_1) - \mathcal{K}_1(\psi_2)\end{aligned}$$

und mit Definition 2.6.2 folgt unter Beachtung von

$$\nabla\mathcal{K}(\psi) = K(\psi)\nabla\psi$$

die Behauptung.

□

Satz 3.18 *Sei (η, v) eine schwache Lösung des adjungierten Problems für*

$$\begin{aligned}\Theta^* &= \int_0^1 \Theta_1'(\psi_2 + s(\psi_1 - \psi_2))\, ds,\\ K^* &= \int_0^1 K_1'(\psi_2 + s(\psi_1 - \psi_2))\, ds,\\ \mathcal{K}^* &= \int_0^1 K_1(\psi_2 + s(\psi_1 - \psi_2))\, ds.\end{aligned}$$

1. *Wenn $F^* = \omega^*_{\tilde{Q}_T}\beta_2$, $0 \neq \omega^*_{\tilde{Q}_T} \in W_{\tilde{Q}_T}$, $g^* = 0$ fast überall auf Γ_{DT} und $h^* = 0$ fast überall auf Γ_{FT}, dann ist (η, v) eine Lösung des $\delta\mathcal{G}_{(\psi,q)}$-adjungierten Problems für den Beobachtungsoperator $\mathcal{B}_{\tilde{Q}_T}$ und $\omega^*_{\tilde{Q}_T}$.*

2. *Wenn $F^* = 0$ fast überall in Q_T, $g^* = 0$ fast überall auf Γ_{DT} und $h^* = \omega^*_{\tilde{\Gamma}_{FT}}\beta_1$, $0 \neq \omega^*_{\tilde{\Gamma}_{FT}} \in W_{\tilde{\Gamma}_{FT}}$, dann ist (η, v) eine Lösung des $\delta\mathcal{G}_{(\psi,q)}$-adjungierten Problems für den Beobachtungsoperator $\mathcal{B}_{\tilde{\Gamma}_{FT}}$ und $\omega^*_{\tilde{\Gamma}_{FT}}$.*

3. *Wenn $F^* = 0$ fast überall in Q_T, $g^* = \tilde{\omega}^*_{\tilde{\Gamma}_{DT}}\gamma$, $0 \neq \tilde{\omega}^*_{\tilde{\Gamma}_{DT}} \in \tilde{W}_{\tilde{\Gamma}_{DT}}$ und $h^* = 0$ fast überall auf Γ_{FT}, dann ist (η, v) eine Lösung des $\delta\mathcal{G}_{(\psi,q)}$-adjungierten Problems für den Beobachtungsoperator $\tilde{\mathcal{B}}_{\tilde{\Gamma}_{DT}}$ und $\tilde{\omega}^*_{\tilde{\Gamma}_{DT}}$.*

Beweis:

1. Zunächst gilt

$$\begin{aligned}\left\langle \omega^*_{\tilde{Q}_T}, \delta\mathcal{B}_{\tilde{Q}_T}((\Theta_1, K_1), (\Theta_2, K_2)) \right\rangle &= \int_{\tilde{Q}_T} \omega^*_{\tilde{Q}_T}\beta_2(\psi_1 - \psi_2)\\ &= \int_{Q_T} F^*(\psi_1 - \psi_2).\end{aligned}$$

   Da $\psi_1 - \psi_2 \in L^2_{0,D}((0,T); H^1(\Omega))$ und $q_1 - q_2 \in L^2((0,T); H_{0,F}(\mathrm{div}; \Omega))$, gelten (3.20) für $\phi = \psi_1 - \psi_2$ und (3.21) für $w = q_1 - q_2$. Summation von (3.20) und (3.21) liefert, dass für alle $(\Theta_1, K_1), (\Theta_2, K_2) \in P$

   $$\left\langle \delta\mathcal{G}_{(\psi,q)}((\Theta_1, K_1), (\Theta_2, K_2)), (\eta, v) \right\rangle = \left\langle \delta\omega^*_{\tilde{Q}_T}, \delta\mathcal{B}_{\tilde{Q}_T}((\Theta_1, K_1), (\Theta_2, K_2)) \right\rangle$$

   erfüllt ist.

2. Analog zu 1. folgt die Behauptung mit

$$\left\langle \omega^*_{\tilde{\Gamma}_{FT}}, \delta\mathcal{B}_{\tilde{\Gamma}_{FT}}((\Theta_1, K_1), (\Theta_2, K_2)) \right\rangle = \int_{\Gamma_{DT}} h^*(\psi_1 - \psi_2).$$

3. Die Behauptung folgt analog mit

$$\left\langle \tilde{\omega}^*_{\tilde{\Gamma}_{DT}}, \delta\tilde{\mathcal{B}}_{\tilde{\Gamma}_{DT}}((\Theta_1, K_1), (\Theta_2, K_2)) \right\rangle = \int_{\Gamma_{DT}} g^*(q_1 - q_2) \cdot \nu.$$

□

3.3.3 Identifizierbarkeit

Neben den Annahmen 3.1 seien in diesem Unterabschnitt noch die folgenden Annahmen gültig.

Annahme 3.19

1. $N = 1$, $\Omega = (0, L)$, $L > 0$, $\tilde{\Gamma}_F = \{0\}$ *und* $\tilde{\Gamma}_D = \{L\}$.
2. *Der Wert* $\Theta(0) = \theta_{\text{sat}}$ *sei fest vorgegeben.*
3. $g'(t) < 0$ *für alle* $t \in (0, T)$.
4. $\nabla z = -1$ *(vertikale Orientierung entgegen der Gravitationsrichtung).*
5. $\tilde{Q}_T = \tilde{\Omega} \times (0, T)$ *mit* $\text{meas}(\tilde{\Omega}) > 0$.
6. *Die Lösung des direkten Problems* (ψ, q) *sei hinreichend glatt.*

Bezeichnung 3.20

1. $\underline{\psi} := g(T) - L$.
2. $\tilde{P}_\Theta := \left\{ f \in C^1_{\text{pwa}}[\underline{\psi}, 0] \mid f, f' \geq 0,\ f, f' \not\equiv 0 \right\}$.
3. $\tilde{P}_K := \left\{ f \in C_{\text{pwa}}[\underline{\psi}, 0] \mid f \in C^1_{\text{pw}}[\underline{\psi}, 0],\ f, f' \geq 0,\ f, f' \not\equiv 0 \right\}$.

Aus Lemma 3.4 in [11] wissen wir, dass für die Lösung des direkten Problems gilt (vgl. auch Beispiel 3.3)

$$\underline{\psi} \leq \psi \leq L.$$

Der $\mathcal{G}$-(ψ, q)-adjungierte Operator aus Lemma 3.17.2 führt nach Satz 3.18 auf das adjungierte Problem aus Definition 3.14 mit den Koeffizienten Θ^*, K^*

und $\mathcal{K}^*$ gemäß Satz 3.18. Im eindimensionalen Fall lautet dieses adjungierte Problem in der klassischen Formulierung

$$\begin{aligned}
\Theta^*\partial_t\eta + \mathcal{K}^*\partial_x v + K^* v &= -F^* && && (3.26)\\
v &= \partial_x \eta && \text{in } Q_T, && (3.27)\\
\eta(.,T) &= 0 && \text{in } \Omega, && (3.28)\\
\mathcal{K}^*(0,.)v(0,.) &= -h^* && \text{in } (0,T), && (3.29)\\
\eta(L,.) &= g^* && \text{in } (0,T), && (3.30)
\end{aligned}$$

bzw.

$$\begin{aligned}
\Theta^*\partial_t\eta + \mathcal{K}^*\partial_{xx}\eta + K^*\partial_x\eta &= -F^* && \text{in } Q_T, && (3.31)\\
\eta(.,T) &= 0 && \text{in } \Omega, && (3.32)\\
\mathcal{K}^*(0,.)\partial_x\eta(0,.) &= -h^* && \text{in } (0,T), && (3.33)\\
\eta(L,.) &= g^* && \text{in } (0,T),. && (3.34)
\end{aligned}$$

Für die Lösung dieses Problems erhalten wir die folgenden Aussagen.

Lemma 3.21 *Seien Θ^*, K^* und $\mathcal{K}^*$ gemäß Lemma 3.17.2 definiert. Seien (η, v) die Lösung von (3.26–3.30) und $\tau \in (0,T)$, sodass Θ^*, K^* und $\mathcal{K}^*$ auf $(0,\tau)$ nicht identisch verschwinden. Dann gelten:*

1. *Wenn $g^* \in C[0,T]$ mit*

$$\begin{aligned}
g^* &> 0 && \text{in } [0,\tau),\\
g^* &= 0 && \text{in } [\tau,T]
\end{aligned}$$

und $h^* = 0$, $F^* = 0$, so folgt

$$\begin{aligned}
&\eta > 0 \quad \text{und} \quad v > 0 && \text{in } \Omega\times[0,\tau),\\
&\eta = 0 \quad \text{und} \quad v = 0 && \text{in } \Omega\times[\tau,T].
\end{aligned}$$

2. *Wenn $h^* \in C[0,T]$ mit*

$$\begin{aligned}
h^* &< 0 && \text{in } [0,\tau),\\
h^* &= 0 && \text{in } [\tau,T]
\end{aligned}$$

und $g^* = 0$, $F^* = 0$, so folgt

$$\begin{aligned}
&\eta < 0 \quad \text{und} \quad v > 0 && \text{in } \Omega\times[0,\tau),\\
&\eta = 0 \quad \text{und} \quad v = 0 && \text{in } \Omega\times[\tau,T].
\end{aligned}$$

3. *Wenn* $F^* \in C(Q_T)$ *mit*

$$\begin{aligned} F^* &> 0 \qquad \text{in } \tilde{\Omega} \times [0,\tau), \\ F^* &= 0 \qquad \text{sonst} \end{aligned}$$

und $g^* = 0$, $h^* = 0$, *so folgt*

$$\begin{aligned} \eta &> 0 \qquad \text{in } \tilde{\Omega} \times [0,\tau), \\ \eta &= 0 \qquad \text{sonst.} \end{aligned}$$

Beweis: Da die Koeffizienten Θ^*, K^* und $\mathcal{K}^*$ nichtnegativ sind, folgen die Aussagen jeweils aus einem Maximumprinzip (vgl. [5], [11] und [12]). □

Satz 3.22

1. *Sei* K *gegeben. Dann ist die Koeffizientenfunktion* $\Theta \in \tilde{P}_\Theta$ *identifizierbar aus einer der Beobachtungen* $\tilde{\omega}_{\Gamma_{DT}}$, $\omega_{\Gamma_{FT}}$ *oder* $\omega_{\tilde{Q}_T}$.

2. *Sei* Θ *gegeben. Dann ist die Koeffizientenfunktion* $K \in \tilde{P}_K$ *sowohl aus der Beobachtung* $\tilde{\omega}_{\Gamma_{DT}}$ *als auch aus der Beobachtung* $\omega_{\Gamma_{FT}}$ *identifizierbar.*

3. *Das Koeffizientenpaar* $(\Theta, K) \in \tilde{P}_\Theta \times \tilde{P}_K$ *ist aus den Beobachtungen* $\tilde{\omega}_{\Gamma_{DT}}$ *und* $\omega_{\Gamma_{FT}}$ *identifizierbar.*

Beweis:

1. Seien $\Theta_1, \Theta_2 \in \tilde{P}_\Theta$ mit $\Theta_1 \neq \Theta_2$. Da $\Theta_1(0) = \Theta_2(0) = \theta_{\text{sat}}$, folgt für $\delta\Theta(\psi) := \Theta_1(\psi) - \Theta_2(\psi)$ nach Satz 2.18 und Folgerung 2.13, dass o. B. d. A. $\xi_1, \xi_2 \in \mathbb{R}$, $\xi_1 < \xi_2$, existieren mit $\delta\Theta'(\psi) > 0$ in $(\xi_1, \xi_2) \subset [\underline{\psi}, 0]$ und $\delta\Theta'(\psi) \geq 0$ in $[\xi_2, 0]$. Sei

$$\tau := \max\{t \in (0,T) \mid \psi_2(.,t) \geq \xi_1 \text{ fast überall in } Q_t\}.$$

Für den Operator $\delta\mathcal{G}_{(\Theta,K)}((\Theta_1,K),(\Theta_2,K))$ erhalten wir nach partieller Integration

$$\left\langle \delta\mathcal{G}_{(\Theta,K)}((\Theta_1,K),(\Theta_2,k)),(\eta,v)\right\rangle = \int_{Q_T} \delta\Theta'(\psi_2)\partial_t\psi_2\eta.$$

Nach Lemma 3.3 in [11] gilt, dass

$$\partial_t\psi_2 < 0 \qquad \text{fast überall in } Q_T. \tag{3.35}$$

(η, v) sei die Lösung von (3.26–3.30) mit den Koeffizienten gemäß Satz 3.18.

(a) Wenn g^*, h^* und F^* gemäß Lemma 3.21.1 bzw. 3.21.3 gewählt werden, so folgt mit $\delta\Theta'(\psi_2) > 0$ in einer Nichtnullmenge von Q_τ

$$\left\langle \delta\mathcal{G}_{(\Theta,K)}((\Theta_1,K),(\Theta_2,K)),(\eta,v)\right\rangle < 0.$$

Mit Satz 3.18 und Satz 2.10.2.a folgt die Identifizierbarkeit von Θ aus der Beobachtung $\omega_{\Gamma_{FT}}$ bzw. $\omega_{\tilde{Q}_T}$.

(b) Wenn g^*, h^* und F^* gemäß Lemma 3.21.2 gewählt werden, so folgt mit $\delta\Theta'(\psi_2) > 0$ in einer Nichtnullmenge von Q_τ

$$\left\langle \delta\mathcal{G}_{(\Theta,K)}((\Theta_1,K),(\Theta_2,K)),(\eta,v)\right\rangle > 0.$$

Die Sätze 3.18 und 2.10.2.a liefern damit die Identifizierbarkeit von Θ aus $\tilde{\omega}_{\Gamma_{DT}}$.

2. Seien $K_1, K_2 \in \tilde{P}_K$ mit $K_1 \neq K_2$. Für $\delta K(\psi) := K_1(\psi) - K_2(\psi)$ gibt es analog $\xi_1, \xi_2 \in \mathbb{R}$, $\xi_1 < \xi_2$, sodass o. B. d. A. $\delta K(\psi) > 0$ in $(\xi_1, \xi_2) \subset [\underline{\psi}, 0]$ und $\delta K(\psi) \geq 0$ in $[\xi_2, 0]$. τ sei wie oben definiert. Der Operator $\delta\mathcal{G}_{(\Theta,K)}((\Theta,K_1),(\Theta,K_2))$ ist gegeben durch

$$\left\langle \delta\mathcal{G}_{(\Theta,K)}((\Theta,K_1),(\Theta,K_2)),(\eta,v)\right\rangle = \int_{Q_T} \delta K(\psi_2)(\partial_x\psi_2 - 1)v.$$

Nach Lemma 3.2 in [11] gilt

$$\partial_x\psi_2 - 1 < 0 \qquad \text{fast überall in } Q_T. \tag{3.36}$$

(η, v) sei die Lösung von (3.26–3.30) mit den Koeffizienten gemäß Satz 3.18 und g^*, h^* und F^* analog Lemma 3.21.1 bzw. 3.21.2 definiert. Dann folgt mit $\delta K(\psi_2) > 0$ in einer Nichtnullmenge von Q_τ

$$\left\langle \delta\mathcal{G}_{(\Theta,K)}((\Theta_1,K),(\Theta_2,K)),(\eta,v)\right\rangle < 0.$$

Hieraus folgt mit den Sätzen 3.18 und 2.10.2.a die Identifizierbarkeit von $K(\psi)$ aus $\omega_{\Gamma_{FT}}$ und $\tilde{\omega}_{\Gamma_{DT}}$.

3. Seien $\Theta_1, \Theta_2 \in \tilde{P}_\Theta$ und $K_1, K_2 \in \tilde{P}_K$. Nach Satz 2.18 sind Θ_1' und Θ_2' sowie K_1 und K_2 vergleichbar. Es existieren also eine zu Θ_1' und Θ_2' gehörende Zerlegung

$$\underline{\psi} < \xi_1 < \xi_2 < \ldots < \xi_m < 0$$

und eine zu K_1 und K_2 gehörende Zerlegung

$$\underline{\psi} < \tilde{\xi}_1 < \tilde{\xi}_2 < \ldots < \tilde{\xi}_n < 0,$$

sodass $\delta\Theta'(\psi)$ und $\delta K(\psi)$ ihr Vorzeichen auf den jeweiligen Teilintervallen nicht ändern. Wir setzen $\hat{\xi} := \max\left\{\xi_m, \tilde{\xi}_n\right\}$ und

$$\tau := \max\left\{t \in (0,T) \mid \psi_2(.,t) \geq \hat{\xi} \text{ fast überall in } Q_t\right\}.$$

Für den Operator $\delta\mathcal{G}_{(\Theta,K)}((\Theta_1,K_1),(\Theta_1,K_2))$ gilt nach partieller Integration

$$\begin{aligned}\left\langle \delta\mathcal{G}_{(\Theta,K)}((\Theta_1,K_1),(\Theta_1,K_2)),(\eta,v)\right\rangle &= \int_{Q_T} \delta\Theta'(\psi_2)\partial_t\psi_2\eta \\ &+ \int_{Q_T} \delta K(\psi_2)(\partial_x\psi_2 - 1)v.\end{aligned}$$

Sei (η_1, v_1) Lösung von (3.26–3.30) mit den Koeffizienten aus Satz 3.18 und g^*, h^* und F^* gemäß Lemma 3.21.1. Sei (η_2, v_2) Lösung von (3.26–3.30) mit den Koeffizienten aus Satz 3.18 und g^*, h^* und F^* gemäß Lemma 3.21.2. Aus

$$\begin{aligned}\left\langle \delta\mathcal{G}_{(\Theta,K)}((\Theta_1,K_1),(\Theta_1,K_2)),(\eta_1,v_1)\right\rangle &= 0 \\ \left\langle \delta\mathcal{G}_{(\Theta,K)}((\Theta_1,K_1),(\Theta_1,K_2)),(\eta_2,v_2)\right\rangle &= 0,\end{aligned}$$

d. h.

$$\begin{aligned}\int_{Q_T} \delta\Theta'(\psi_2)\partial_t\psi_2\eta_1 &= -\int_{Q_T} \delta K(\psi_2)(\partial_x\psi_2 - 1)v_1 \\ \int_{Q_T} \delta\Theta'(\psi_2)\partial_t\psi_2\eta_2 &= -\int_{Q_T} \delta K(\psi_2)(\partial_x\psi_2 - 1)v_2,\end{aligned}$$

folgt mit (3.35), (3.36) und Lemma 3.21 bei Beachtung, dass in $[\hat{\xi}, 0]$ kein Vorzeichenwechsel von $\delta\Theta'$ and δK stattfinden kann,

$$\delta\Theta'(\psi_2) = 0 \text{ und } \delta K(\psi_2) = 0 \text{ in } [\hat{\xi}, 0].$$

Da $\Theta_1(0) = \Theta_2(0)$ folgt $\delta\Theta(\psi_2) = 0$. Somit gilt also

$$\Theta_1 = \Theta_2 \text{ und } K_1 = K_2 \text{ in } [\hat{\xi}, 0].$$

Analoges Vorgehen für das jeweilige Restintervall $[\underline{\psi}, \hat{\xi}]$ liefert nach endlich vielen Schritten die Gleichheit in $[\underline{\psi}, 0]$. Durch Anwendung von Satz 3.18 und Satz 2.10.2.b folgt hieraus die Identifizierbarkeit von Θ und K aus den Beobachtungen $\tilde{\omega}_{\Gamma_{DT}}$ und $\omega_{\Gamma_{FT}}$.

□

Aus Satz 2.18 folgt unmittelbar

Lemma 3.23 *Sei $\tilde{\omega}_{\Gamma_{DT}} \in C_{\text{pwa}}[0,T]$. Dann bleiben die Aussagen von Satz 3.22 über die Identifizierbarkeit gültig, wenn die Beobachtung $\tilde{\omega}_{\Gamma_{DT}}$ ersetzt wird durch die Beobachtung $\omega_{\Gamma_{DT}}$.*

3.4 Diskretes Modell

3.4.1 Diskretisierung

Mithilfe der in [52] beschriebenen hybrid-gemischten Finite-Elemente-Methode wird die Richards-Gleichung in der Druckformulierung in ein diskretes Modell überführt. Dabei wird von einer angepassten gemischten Variationsformulierung ausgegangen, die sich aus der Formulierung in Definition 3.10 durch partielle Integration in (3.19) und entsprechende Anpassung der Funktionenräume ergibt.

Zunächst wird durch das implizite Euler-Verfahren in der Zeit diskretisiert. Dazu werden eine endliche Zerlegung

$$0 = t^0 < t^1 < ... < t^m = T$$

des Zeitintervalls $[0, T]$ und die Bezeichnungen

$$\begin{aligned} \Delta t^i &:= t^i - t^{i-1} \\ \psi^i(.) &:= \psi(., t^i) \\ q^i(.) &:= q(., t^i) \end{aligned}$$

eingeführt.

Für die räumliche Diskretisierung zerlegen wir das Gebiet Ω in eine endliche Anzahl von Elementen T mit $\bigcup T = \Omega$. $\mathcal{T}$ bezeichne die Menge aller Elemente und $\mathcal{E}$ sei die Menge aller Kanten (bzw. Knoten im räumlich eindimensionalen Fall). Das Gebiet Ω sei polygonal berandet, sodass eine disjunkte Aufteilung

$$\mathcal{E} = \mathcal{E}_I \cup \mathcal{E}_\Gamma = \mathcal{E}_I \cup \mathcal{E}_D \cup \mathcal{E}_F$$

mit

$$\Gamma_D = \mathcal{E}_D \,, \qquad \Gamma_F = \mathcal{E}_F \qquad \text{und} \qquad \mathcal{E}_I = \text{innere Kanten}$$

möglich ist.

Nach Einführung der diskreten Funktionenräume

$$\begin{aligned} M^0_{-1}(\mathcal{T}) &:= \left\{\eta \in L^2(\Omega) \mid \eta|_T \in P^0(T)\ \forall T \in \mathcal{T}\right\} \\ N^0_{-1,0}(\mathcal{E}) &:= \left\{\lambda \in L^2(\mathcal{E}) \mid \lambda|_E \in P^0(E)\ \forall E \in \mathcal{E}, \lambda|_E = 0\ \forall E \in \mathcal{E}_D\right\} \\ RT^0_{-1}(\mathcal{T}) &:= \left\{q \in \left(L^2(\Omega)\right)^N \mid q|_T \in RT^0(T)\ \forall T \in \mathcal{T}\right\} \end{aligned}$$

(P^0 bezeichnet den Raum der konstanten Funktionen und RT^0 den Raviart-Thomas-Raum niedrigster Ordnung) gelangen wir mit der Hybridisierung, bei der anstelle von $q \in H(\mathrm{div}, \Omega)$ nur noch $q|_T \in H(\mathrm{div}, T)$ gefordert wird, zu folgender volldiskreter hybrid-gemischter Variationsformulierung:

Für alle Zeitpunkte t^i, $i \in \{1, ..., m\}$, ist
$(\psi_h^i, \lambda_h^i, q_h^i) \in M_{-1}^0(\mathcal{T}) \times N_{-1,0}^0(\mathcal{E}) \times RT_{-1}^0(\mathcal{T})$ die Lösung von

$$\int_\Omega \Theta(\psi_h^i)\eta_h + \Delta t^i \int_\Omega \nabla \cdot q_h^i \eta_h = \int_\Omega \Theta(\psi_h^{i-1})\eta_h \quad \forall \eta_h \in M_{-1}^0(\mathcal{T}) \tag{3.37}$$

$$\int_\Omega K^{-1}(\psi_h^i) q_h^i \cdot v_h - \int_\Omega \psi_h^i \nabla \cdot v_h + \int_\Omega \nabla z \cdot v_h + \sum_{T \in \mathcal{T}} \int_{\partial T} \lambda_h^i v_h \cdot \nu = -\int_{\Gamma_{DT}} g\, v_h \cdot \nu \quad \forall v_h \in RT_{-1}^0(\mathcal{T}) \tag{3.38}$$

$$\sum_{T \in \mathcal{T}} \int_{\partial T} \mu_h^i q_h^i \cdot \nu = 0 \quad \forall \mu_h \in N_{-1,0}^0(\mathcal{E}). \tag{3.39}$$

Bemerkung 3.24 λ_h in (3.38) bezeichnen die durch die Hybridisierung eingeführten Freiheitsgrade (Lagrange-Multiplikatoren) und die Gleichung (3.39) die zusätzlich zu stellende Forderung (vgl. Abschnitte 3.1 und 3.2 in [52]).

Für die Räume $M_{-1}^0(\mathcal{T})$ und $N_{-1,0}^0(\mathcal{E})$ verwenden wir jeweils die Basisfunktionen

$$\varphi_S(x) = \begin{cases} 1 & \text{falls } x \in S, \\ 0 & \text{falls } x \notin S \end{cases} \qquad \text{mit} \quad S \in \{E, T\}$$

und für den Raum $RT_{-1}^0(\mathcal{T})$ wählen wir die Basisfunktionen w_{TE}, deren Träger jeweils das Element T ist und die

$$\int_{E'} w_{TE} \cdot \nu \, ds = \delta_{EE'}$$

erfüllen. Für die diskrete Lösung haben wir nun die Darstellungen

$$\begin{aligned} \psi_h^i(x) &:= \sum_{T \in \mathcal{T}} \psi_T^i \varphi_T(x), \\ \lambda_h^i(x) &:= \sum_{E \in \mathcal{E}} \lambda_E^i \varphi_T(x), \\ q_h^i(x) &:= \sum_{T \in \mathcal{T}} \sum_{E \subset T} q_{TE}^i w_{TE}(x). \end{aligned}$$

Mit den abkürzenden Schreibweisen

$$z_{TE} := \int_T \nabla z \cdot w_{TE}, \quad B_{TEE'} := \int_T w_{TE} \cdot w_{TE'}, \quad b_T := \sum_{E' \subset T} B_{TEE'}^{-1}$$

erhalten wir nach Auswertung der Integrale und anschließender statischer Kondensation (vgl. Abschnitt 3.3 in [52]) ein System von diskreten Gleichungen.

3.4.2 Diskretes direktes Problem

Definition 3.25 *$(\lambda_h, \psi_h, q_h) := (\lambda_h^i, \psi_h^i, q_h^i)_{i=0,...,m}$ mit $\lambda_h^i = (\lambda_E^i)_{E\in\mathcal{E}} \in \mathbb{R}^{|\mathcal{E}|}$, $\psi_h^i = (\lambda_T^i)_{T\in\mathcal{T}} \in \mathbb{R}^{|\mathcal{T}|}$ und $q_h^i = (q_{TE}^i)_{T\in\mathcal{T}, E\subset T} \in \mathbb{R}^{|\mathcal{E}_\Gamma|+2|\mathcal{E}_I|}$ ist Lösung des diskreten direkten Problems, wenn*

$$\begin{aligned}
\lambda_E^0 &= \psi_{0,E} \qquad \forall E \in \mathcal{E} \\
\psi_T^0 &= \psi_{0,T} \qquad \forall T \in \mathcal{T} \\
q_{TE}^0 &= K(\psi_T^0) \sum_{E'\subset T} B_{TEE'}^{-1}(\psi_T^0 - \lambda_{E'}^0 - z_{TE}) \\
&\qquad\qquad \forall T \in \mathcal{T}, E \subset T \text{ mit } E \notin \mathcal{E}_F
\end{aligned}$$

und für alle $i = 1, ..., m$

$$\sum_{T\supset E} K(\psi_T^i) \sum_{E'\subset T} B_{TEE'}^{-1}(\psi_T^i - \lambda_{E'}^i - z_{TE'}) = 0 \qquad \forall E \in \mathcal{E} \setminus \mathcal{E}_D \tag{3.40}$$

$$K(\psi_T^i)\left(\tilde{N}\psi_T^i - \sum_{E\subset T} \lambda_E^i\right) + |T|\frac{\Theta(\psi_T^i) - \Theta(\psi_T^{i-1})}{\Delta t^i b_T} = 0 \ \forall T \in \mathcal{T} \tag{3.41}$$

$$K(\psi_T^i) \sum_{E'\subset T} B_{TEE'}^{-1}(\psi_T^i - \lambda_{E'}^i - z_{TE'}) = q_{TE}^i \qquad \forall T \in \mathcal{T}, E \subset T, E \notin \mathcal{E}_F \tag{3.42}$$

sowie

$$\begin{aligned}
\lambda_E^i &= g_E^i \qquad \forall E \in \mathcal{E}_D \\
q_{TE}^i &= 0 \qquad \forall E \in \mathcal{E}_F, T \supset E
\end{aligned}$$

gilt, wobei $\tilde{N}$ die Anzahl der Kanten pro Element ist.

Die Gleichungen (3.41) sind die Diskretisierung der Kontinuitätsgleichung, die Gleichungen (3.42) sind das diskrete Darcy-Gesetz und die Gleichungen (3.40) beschreiben zusammen mit (3.42) die Erhaltung des Flusses über die Kanten. Die Gleichungen (3.42) könnten aus dem Gleichungssystem eliminiert werden. Wir behalten sie jedoch bei, da wir zur Lösung des inversen Problems Beobachtungen des (kumulativen) Flusses verwenden. Als Ausdruck von Druck und den Lagrange-Multiplikatoren wären diese Beobachtungen nichtlinear. Die Lagrange-Multiplikatoren können als Approximationen des Druckes auf den Kanten interpretiert werden.

Bezeichnung 3.26

1. *Der Lösungsraum ist gegeben als* $U_h := \mathbb{R}^{|\mathcal{E}|} \times \mathbb{R}^{|\mathcal{T}|} \times \mathbb{R}^{|\mathcal{E}_\Gamma|+2|\mathcal{E}_I|}$.

2. $V_h := U_h^* = \mathbb{R}^{|\mathcal{E}|} \times \mathbb{R}^{|\mathcal{T}|} \times \mathbb{R}^{|\mathcal{E}_\Gamma|+2|\mathcal{E}_I|}$ *ist der Raum der Testfunktionen.*

3. $\tilde{P} := \tilde{P}_\Theta \times \tilde{P}_K$.

Zur Vereinfachung der Notation wird angenommen, dass die diskreten Beobachtungszeitpunkte in der Menge $\{t^0, t^1, ..., t^m\}$ enthalten sind. Die diskreten Beobachtungen sind wie folgt definiert:

(a) $W_{F,h} := \mathbb{R}^{n_F}$, $n_F \in \mathbb{N}$,

$$\begin{aligned} \mathcal{B}_{F,h} : U &\rightarrow W_{F,h} \\ (\psi_h, \lambda_h, q_h) &\mapsto \omega_{F,h} := \begin{pmatrix} \lambda_{E_1}^{i_1} \\ \vdots \\ \lambda_{E_{n_F}}^{i_{n_F}} \end{pmatrix} \end{aligned}$$

für n_F verschiedene Paare $(i_k, E_k) \in \{0, ..., m\} \times \mathcal{E}_F$.

(b) $W_{Q,h} := \mathbb{R}^{n_Q}$, $n_Q \in \mathbb{N}$,

$$\begin{aligned} \mathcal{B}_{Q,h} : U &\rightarrow W_{Q,h} \\ (\psi_h, \lambda_h, q_h) &\mapsto \omega_{Q,h} := \begin{pmatrix} \psi_{T_1}^{i_1} \\ \vdots \\ \psi_{T_{n_Q}}^{i_{n_Q}} \end{pmatrix} \end{aligned}$$

für n_Q verschiedene Paare $(i_k, E_k) \in \{0, ..., m\} \times \mathcal{T}$. Wenn der Beobachtungsort einer Kante der Diskretisierung entspricht, dann können auch die Lagrange-Multiplikatoren verwendet werden.

(c) $W_{D,h} := \mathbb{R}^{n_D}$, $n_D \in \mathbb{N}$,

$$\begin{aligned} \tilde{\mathcal{B}}_{D,h} : U &\rightarrow W_{D,h} \\ (\psi_h, \lambda_h, q_h) &\mapsto \tilde{\omega}_{D,h} := \begin{pmatrix} q_{TE_1}^{i_1} \\ \vdots \\ q_{TE_{n_D}}^{i_{n_D}} \end{pmatrix} \end{aligned}$$

bzw. bei der Approximation des kumulativen Ausflusses durch die Trapezregel

$$\begin{aligned} \mathcal{B}_{D,h} : U &\to W_{D,h} \\ (\psi_h, \lambda_h, q_h) &\mapsto \omega_{D,h} := \begin{pmatrix} \frac{1}{2}\sum_{i=0}^{i_1-1} \Delta^{i+1}(q^i_{TE_1} + q^{i+1}_{TE_1}) \\ \vdots \\ \frac{1}{2}\sum_{i=0}^{i_{n_D}-1} \Delta^{i+1}(q^i_{TE_{n_D}} + q^{i+1}_{TE_{n_D}}) \end{pmatrix} \end{aligned}$$

für n_D verschiedene Paare (i_k, TE_k) mit $i_k \in \{0, ..., m\}$ und $E_k \in \mathcal{E}_D, T \supset E_k$.

Wir definieren folgende Abbildungen:

$$\mathcal{H} : U_h \times \tilde{P} \to \mathbb{R}^{(m+1)\cdot|\mathcal{E}|}$$

durch

$$\begin{aligned} \mathcal{H}(\lambda_h, \psi_h, q_h, \Theta, K) &:= \left(\left(\mathcal{H}^i_E(\lambda^i_h, \psi^i_h, K) \right)_{E\in\mathcal{E}} \right)_{i=0,...,m} \\ \mathcal{H}^i_E(\lambda^i_h, \psi^i_h, K) &:= \\ \begin{cases} \psi_{0,E} - \lambda^0_E, & i = 0, \\ \sum_{T\supset E} K(\psi^i_T) \sum_{E'\subset T} B^{-1}_{TEE'}(\psi^i_T - \lambda^i_{E'} - z_{TE'}), & i \neq 0, E \notin \mathcal{E}_D, \\ g^i_E - \lambda^i_E, & i \neq 0, E \in \mathcal{E}_D, \end{cases} \end{aligned}$$

$$\mathcal{G} : U_h \times \tilde{P} \to \mathbb{R}^{(m+1)\cdot|\mathcal{T}|}$$

durch

$$\begin{aligned} \mathcal{G}(\lambda_h, \psi_h, q_h, \Theta, K) &:= \left(\left(\mathcal{G}^i_T(\lambda^i_h, \psi^{i-1}_h, \psi^i_h, \Theta, K) \right)_{T\in\mathcal{T}} \right)_{i=0,...,m} \\ \mathcal{G}^i_T((\lambda^i_h, \psi^{i-1}_h, \psi^i_h, \Theta, K) &:= \\ \begin{cases} \psi^0_T - \psi_{0,T}, & i = 0, \\ K(\psi^i_T) \left(\tilde{N}\psi^i_T - \sum_{E\subset T} \lambda^i_E \right) + |T| \frac{\Theta(\psi^i_T) - \Theta(\psi^{i-1}_T)}{\Delta t^i b_T}, & i \neq 0 \end{cases} \end{aligned}$$

und

$$\mathcal{F} : U_h \times \tilde{P} \to \mathbb{R}^{(m+1)\cdot(|\mathcal{E}_\Gamma|+2|\mathcal{E}_I|)}$$

durch

$$\begin{aligned} \mathcal{F}(\lambda_h, \psi_h, q_h, \Theta, K) &:= \left(\left(\mathcal{F}^i_{TE}(\lambda^i_h, \psi^i_h, q^i_h, K) \right)_{\substack{T\in\mathcal{T} \\ E\subset T}} \right)_{i=0,...,m} \\ \mathcal{F}^i_{TE}((\lambda^i_h, \psi^i_h, q^i_h, K) &:= \\ \begin{cases} K(\psi^i_T) \sum_{E'\subset T} B^{-1}_{TEE'}(\psi^i_T - \lambda^i_{E'} - z_{TE'}) - q^i_{TE}, & E \notin \mathcal{E}_F \\ -q^i_{TE}, & E \in \mathcal{E}_F. \end{cases} \end{aligned}$$

Damit ist (ψ_h, λ_h, q_h) genau dann Lösung des diskreten direkten Problems für $\Theta \in \tilde{P}_\Theta$ und $K \in \tilde{P}_K$, wenn das nichtlineare Gleichungssystem

$$\left.\begin{array}{rcl} \mathcal{H}(\lambda_h, \psi_h, q_h, \Theta, K) &=& 0 \\ \mathcal{G}(\lambda_h, \psi_h, q_h, \Theta, K) &=& 0 \\ \mathcal{F}(\lambda_h, \psi_h, q_h, \Theta, K) &=& 0 \end{array}\right\} \tag{3.43}$$

erfüllt ist. Die Gleichung $\mathcal{G}(\lambda_h, \psi_h, q_h, \Theta, K) = 0$ kann gedeutet werden als eine implizite Darstellung des Drucks auf einem Element. In den Abschnitten 3.3 und 3.4 in [52] ist erläutert, dass für eine hinreichend feine Diskretisierung (in Zeit und Raum) die Ableitungen $\frac{\partial \mathcal{G}_T^i}{\partial \psi_T^i}$ streng positiv und $\mathcal{G}_T^i$ nach dem Druck aufgelöst werden kann:

$$\psi_T^i = G^i(\sum_{E \in T} \lambda_E^i, \psi_T^{i-1}). \tag{3.44}$$

Wenn (3.44) in $\mathcal{H}(\lambda_h, \psi_h, q_h, \Theta, K) = 0$ eingesetzt wird, so erhalten wir für alle $i = 1, ..., m$ ein globales nichtlineares System für die Lagrange-Multiplikatoren, welches mit einem Newton-Verfahren gelöst wird. Im räumlich eindimensionalen Fall ist die direkte Lösung der in diesem iterativen Verfahren auftretenden linearen Gleichungssysteme für die relevanten Gebiete überschaubarer Länge mit dem Gauß-Verfahren in vernünftiger Zeit möglich. Im mehrdimensionalen Fall kann hierzu ein Mehrgitterverfahren eingesetzt werden (siehe [52]). Durch Lösung der lokalen Probleme (3.41) und (3.42) erhalten wir dann den Druck auf den Elementen und den Fluss über die Kanten. Der Algorithmus zur Lösung des diskreten direkten Problems lautet wie folgt:

Algorithmus 3.27 (Diskretes direktes Problem)
Setze $\lambda_h^0 = \psi_0|_\varepsilon$, $\psi_h^0 = \psi_0|_\mathcal{T}$.
Berechne q_h^0 *aus* $\mathcal{F}^0(\lambda_h^0, \psi_h^0, q_h^0, K) = 0$.
Für $i = 1, ..., m$

> *Löse* $\mathcal{H}^i\left(\lambda_h^i, G^i\left(\sum_{E \subset T} \lambda_E^i, \psi_h^{i-1}\right), K\right) = 0$.
> *Berechne* $\psi_h^i = G^i\left(\sum_{E \subset T} \lambda_E^i, \psi_h^{i-1}\right)$.
> *Berechne* q_h^i *aus* $\mathcal{F}^i(\lambda_h^i, \psi_h^i, q_h^i, K) = 0$.

Indem wir (3.43) schreiben als

$$\begin{aligned}(\eta_h^T, \xi_h^T, \upsilon_h^T)\tilde{\mathcal{G}}(\lambda_h, \psi_h, q_h, \Theta, K) &= (\eta_h^T, \xi_h^T, \upsilon_h^T)\begin{pmatrix} \mathcal{H}(\lambda_h, \psi_h, q_h, \Theta, K) \\ \mathcal{G}(\lambda_h, \psi_h, q_h, \Theta, K) \\ \mathcal{F}(\lambda_h, \psi_h, q_h, \Theta, K) \end{pmatrix} \\ &= \eta_h^T \mathcal{H}(\lambda_h, \psi_h, q_h, \Theta, K) \\ +\xi_h^T \mathcal{G}(\lambda_h, \psi_h, q_h, \Theta, K) + \upsilon_h^T \mathcal{F}(\lambda_h, \psi_h, q_h, \Theta, K) = 0 &\qquad \forall (\eta_h, \xi_h, \upsilon_h) \in V_h,\end{aligned}$$

erhalten wir eine schwache Formulierung des diskreten direkten Problems.

3.4.3 Adjungiertes Problem

Definition 3.28 $(\eta_h, \xi_h, \upsilon_h) \in \mathbb{R}^{|\mathcal{E}|} \times \mathbb{R}^{|\mathcal{T}|} \times \mathbb{R}^{|\mathcal{E}_\Gamma|+2|\mathcal{E}_I|}$ *ist eine Lösung des zum diskreten direkten Problem adjungierten Problems, wenn für alle* $i = 1, ..., m-1$

$$-\sum_{T \supset E} \beta_T^{*^i} \left(\sum_{E' \subset T} B^{-1}_{TEE'} \eta^i_{E'} + \sum_{\substack{E' \subset T \\ E' \notin \mathcal{E}_F}} B^{-1}_{TEE'} \upsilon^i_{TE'} + \xi^i_T \right) = h_E^{*^i}$$

$$\forall E \in \mathcal{E} \setminus \mathcal{E}_D \qquad (3.45)$$

$$\sum_{\substack{E \subset T \\ E \notin \mathcal{E}_D}} \left(\beta_T^{*^i} b_T + \gamma_{TE}^{*^i} \right) \eta^i_E + \left(\beta_T^{*^i} \tilde{N} + \delta_T^{*^i} + \alpha_T^{*^i} \frac{|T|}{\Delta t^i b_T} \right) \xi^i_T$$

$$+ \sum_{\substack{E \subset T \\ E \notin \mathcal{E}_F}} \left(\beta_T^{*^i} b_T + \gamma_{TE}^{*^i} \right) \upsilon^i_{TE} - \alpha_T^{*^i} \frac{|T|}{\Delta t^{i+1} b_T} \xi^{i+1}_T = F_T^{*^i} \quad \forall T \in \mathcal{T} \qquad (3.46)$$

$$-\upsilon^i_{TE} = g_{TE}^{*^i} \quad \forall T \in \mathcal{T}, E \subset T \qquad (3.47)$$

und

$$-\eta^i_E - \sum_{T \supset E} \left\{ \beta_T^{*^i} \left(\xi^i_T + \sum_{\substack{E' \subset T \\ E' \notin \mathcal{E}_F}} B^{-1}_{TEE'} \upsilon^i_{TE'} \right) \right\} = 0 \quad \forall E \in \mathcal{E}_D, \qquad (3.48)$$

für $i = 0$

$$-\eta^0_E - \sum_{T \supset E} \beta_T^{*^0} \sum_{\substack{E' \subset T \\ E' \notin \mathcal{E}_F}} B^{-1}_{TEE'} \upsilon^0_{TE'} = h_E^{*^0} \quad \forall E \in \mathcal{E} \quad (3.49)$$

$$\sum_{\substack{E \subset T \\ E \notin \mathcal{E}_F}} \left(\beta_T^{*^0} b_T + \gamma_T^{*^0} \right) \upsilon^0_{TE} + \xi^0_T - \alpha_T^{*^0} \frac{|T|}{\Delta t^1 b_T} \xi^1_T = F_T^{*^0} \quad \forall T \in \mathcal{T} \quad (3.50)$$

und Gleichung (3.47) sowie für $i = m$ *die Gleichungen (3.45) und (3.46) ohne den Term mit* ξ_T^{m+1} *und die Gleichungen (3.47) und (3.48) erfüllt sind.*

Dieses adjungierte Problem stellt ein lineares Gleichungssystem dar, welches direkt in der Form aus Definition 3.28 gelöst werden kann. Analog zum direkten Problem kann aber auch hier die Dimension des zu lösenden Gleichungssystems reduziert werden. Die Werte von υ^i_{TE} sind direkt durch (3.47)

gegeben und können in den anderen Gleichungen entsprechend ersetzt werden. Die Gleichungen (3.46), (3.48) und (3.50) können nach ξ_T^i aufgelöst und damit die entsprechenden Terme in (3.45) substituiert werden, vorausgesetzt die Koeffizienten von ξ_T^i sind verschieden von Null. Das Resultat ist ein lineares Gleichungssystem für η_h der folgenden Form:

$$\sum_{T\supset E}\left\{\frac{1}{\beta_T^{*i}\tilde{N}+\delta_T^{*i}+\alpha_T^{*i}\frac{|T|}{\Delta t^i b_T}}\sum_{\substack{E'\subset T\\ E'\notin\mathcal{E}_D}}(\beta_T^{*i}b_T+\gamma_{TE'}^{*i})\eta_{E'}^i-\beta_T^{*i}\sum_{E'\subset T}B_{TEE'}^{-1}\eta_{E'}^i\right\}$$

$$=h_E^{*i}-\sum_{T\supset E}\left\{\beta_T^{*i}\sum_{\substack{E'\subset T\\ E'\notin\mathcal{E}_F}}g_{TE}^{*i}\right.$$

$$\left.-\frac{1}{\beta_T^{*i}\tilde{N}+\delta_T^{*i}+\alpha_T^{*i}\frac{|T|}{\Delta t^i b_T}}\left[F_T^{*i}+\sum_{\substack{E'\subset T\\ E'\notin\mathcal{E}_F}}(\beta_T^{*i}b_T+\gamma_{TE'}^{*i})g_{TE'}^{*i}+\alpha_T^{*i}\frac{|T|}{\Delta t^{i+1}b_T}\xi_T^{i+1}\right]\right\}$$

$$\forall E\in\mathcal{E}\setminus\mathcal{E}_D \qquad (3.51)$$

und

$$-\eta_E^i+\sum_{T\supset E}\frac{1}{\beta_T^{*i}\tilde{N}+\delta_T^{*i}+\alpha_T^{*i}\frac{|T|}{\Delta t^i b_T}}\sum_{\substack{E'\subset T\\ E'\notin\mathcal{E}_D}}(\beta_T^{*i}b_T+\gamma_{TE'}^{*i})\eta_{E'}^i \;=$$

$$\sum_{T\supset E}\left\{\frac{1}{\beta_T^{*i}\tilde{N}+\delta_T^{*i}+\alpha_T^{*i}\frac{|T|}{\Delta t^i b_T}}F_T^{*i}+\sum_{\substack{E'\subset T\\ E'\notin\mathcal{E}_F}}(\beta_T^{*i}b_T+\gamma_{TE'}^{*i})g_{TE'}^i\right.$$

$$\left.+\alpha_T^{*i}\frac{|T|}{\Delta t^{i+1}b_T}\xi_T^{i+1}-\beta_T^{*i}\sum_{\substack{E'\subset T\\ E'\notin\mathcal{E}_F}}B_{TEE'}^{-1}g_{TE'}^{*i}\right\}\quad\forall E\in\mathcal{E}_D \qquad (3.52)$$

für $i=1,...,m-1$,

$$-\eta_E^0=h_E^{*0}-\sum_{T\supset E}\beta_T^{*0}\sum_{\substack{E'\subset T\\ E'\notin\mathcal{E}_F}}B_{TEE'}^{-1}g_{TE'}^{*0}\qquad\forall E\in\mathcal{E} \qquad (3.53)$$

und für $i=m$ Gleichungen (3.51) und (3.52) ohne den Term mit ξ_T^{m+1}.

Für ξ_h gilt

$$\xi_T^i \;=\; \frac{1}{\beta_T^{*i}\tilde{N}+\delta_T^{*i}+\alpha_T^{*i}\frac{|T|}{\Delta t^i b_T}}\left\{F_T^{*i}+\sum_{\substack{E\subset T\\ E\notin\mathcal{E}_F}}(\beta_T^{*i}b_T+\gamma_{TE}^{*i})g_{TE}^{*i}\right.$$

$$+\alpha_T^{*^i}\frac{|T|}{\Delta t^{i+1}b_T}\xi_T^{i+1} - \sum_{\substack{E\subset T\\ E\notin\mathcal{E}_D}}(\beta_T^{*^i}+\gamma_{TE}^{*^i})\eta_E^i\Bigg\} \quad \forall T\in\mathcal{T} \qquad (3.54)$$

für $i = 1, ..., m-1$,

$$\xi_T^0 = F_T^{*^0} + \alpha_T^{*^0}\frac{|T|}{\Delta t^1 b_T}\xi_T^1 + \sum_{\substack{E\subset T\\ E\notin\mathcal{E}_F}}(\beta_T^{*^0}b_T+\gamma_{TE}^{*^0})g_{TE}^{*^0} \qquad \forall T\in\mathcal{T} \qquad (3.55)$$

und für $i = m$ (3.54) ohne den Term mit ξ_T^{m+1}. Gegenüber dem direkten Problem muss das adjungierte Problem allerdings zeitinvers gelöst werden. Eine Aussage zur Lösbarkeit werden wir später treffen (siehe Unterabschnitt 3.4.4).

Algorithmus 3.29 (Adjungiertes Problem)
Für $i = m, ..., 1$

> *Löse das lineare Gleichungssystem (3.51,3.52).*
> *Setze* η_h^i *in (3.54) ein.*
> *Berechne* ξ_h^i *aus (3.54).*
> *Setze* $v_h^i = -g^{*^i}$.

Berechne η_h^0 *aus (3.53).*
Berechne ξ_h^0 *aus (3.55).*
Setze $v_h^0 = -g^{*^0}$.

Lemma 3.30

1. *Die Abbildung* $\delta\tilde{\mathcal{G}}_{(\Theta,K)} : \tilde{P}\times\tilde{P}\to V_h^*$ *erfüllt*

$$\eta_h^T\delta\mathcal{H}_{(\Theta,K)} + \xi_h^T\delta\mathcal{G}_{(\Theta,K)} + v_h^T\delta\mathcal{F}_{(\Theta,K)} =$$
$$\sum_{i=1}^{m}\sum_{E\in\mathcal{E}\setminus\mathcal{E}_D}\eta_E^i\sum_{T\supset E}\delta K_T^i\sum_{E'\subset T}B_{TEE'}^{-1}(\psi_{T,2}^i - \lambda_{E',2}^i - z_{TE'})$$
$$+\sum_{i=1}^{m}\sum_{T\in\mathcal{T}}\xi_T\left\{\delta K_T^i\left(\tilde{N}\psi_{T,2}^i - \sum_{E\subset T}\lambda_{E,2}^i\right) + \frac{|T|}{\Delta t^i b_T}(\delta\Theta_T^i - \delta\Theta_T^{i-1})\right\}$$
$$+\sum_{i=0}^{m}\sum_{T\in\mathcal{T}}\sum_{\substack{E\subset T\\ E\notin\mathcal{E}_F}}v_{TE}^i\delta K_T^i\sum_{E'\subset T}B_{TEE'}^{-1}(\psi_{T,2} - \lambda_{E',2} - z_{TE'})$$

mit $\delta\Theta_T^i := \Theta_1(\psi_{T,2}^i) - \Theta_2(\psi_{T,2}^i)$ *und* $\delta K_T^i := K_1(\psi_{T,2}^i) - K_2(\psi_{T,2}^i)$.

2. *Für die adjungierten Beobachtungsoperatoren gilt:*

   (a) $\mathcal{B}_{F,h}^T : \mathbb{R}^{n_F} \to \mathbb{R}^{(m+1)\cdot|\mathcal{E}|}$ *mit*

   $$\left(\mathcal{B}_{F,h}^T\right)_E^i = \begin{cases} 1 & \text{falls } i = i_k, E = E_k \text{ für ein } k \in \{1, ..., n_F\}, \\ 0 & \text{sonst.} \end{cases}$$

   (b) $\mathcal{B}_{Q,h}^T : \mathbb{R}^{n_Q} \to \mathbb{R}^{(m+1)\cdot|\mathcal{T}|}$ *mit*

   $$\left(\mathcal{B}_{Q,h}^T\right)_T^i = \begin{cases} 1 & \text{falls } i = i_k, T = T_k \text{ für ein } k \in \{1, ..., n_Q\}, \\ 0 & \text{sonst.} \end{cases}$$

   (c) $\tilde{\mathcal{B}}_{D,h}^T : \mathbb{R}^{n_D} \to \mathbb{R}^{(m+1)\cdot(|\mathcal{E}_\Gamma|+|\mathcal{E}_I|)}$ *mit*

   $$\left(\tilde{\mathcal{B}}_{D,h}^T\right)_{TE}^i = \begin{cases} 1 \text{ falls } i = i_k, TE = TE_k \text{ für ein } k \in \{1, ..., n_D\}, \\ 0 \text{ sonst.} \end{cases}$$

Beweis:

1. Siehe Definition 2.6.
2. Die Behauptung folgt aus der Definition der Beobachtungsoperatoren.

□

Satz 3.31 *$(\eta_h, \xi_h, \upsilon_h) \in V_h$ sei eine Lösung des zum diskreten direkten Problem adjungierten Problems für die Koeffizienten*

$$\begin{aligned}
\alpha_T^{*^i} &:= \int_0^1 \Theta_1' \left(\psi_{T,2}^i + s(\psi_{T,1}^i - \psi_{T,2}^i)\right) ds, \\
\beta_T^{*^i} &:= K_1(\psi_{T,1}^i), \\
\gamma_{TE}^{*^i} &:= \sum_{E' \subset T} B_{TEE'}^{-1} (\psi_{T,2}^i - \lambda_{E',2}^i - z_{TE'}) \int_0^1 K_1' \left(\psi_{T,2}^i + s(\psi_{T,1}^i - \psi_{T,2}^i)\right) ds, \\
\delta_T^{*^i} &:= \left(\tilde{N}\psi_{T,2}^i - \sum_{E \subset T} \lambda_{E,2}^i\right) \int_0^1 K_1' \left(\psi_{T,2}^i + s(\psi_{T,1}^i - \psi_{T,2}^i)\right) ds.
\end{aligned}$$

1. *Wenn $F^* = 0$, $g^* = 0$ und*

   $$h^* = \mathcal{B}_{F,h}^T \omega^*$$

   mit $\omega^* \in \mathbb{R}^{n_F}$, so ist $(\eta_h, \xi_h, \upsilon_h)$ eine Lösung des $\delta\tilde{\mathcal{G}}_{(\lambda_h, \psi_h, q_h)}$-adjungierten Problems für den Beobachtungsoperator $\mathcal{B}_{F,h}$ und ω^*.

2. *Wenn*
$$F^* = \mathcal{B}_{Q,h}^T \omega^*$$
mit $\omega^* \in \mathbb{R}^{n_Q}$, $g^* = 0$ *und* $h^* = 0$, *so ist* $(\eta_h, \xi_h, \upsilon_h)$ *eine Lösung des* $\delta\tilde{\mathcal{G}}_{(\lambda_h,\psi_h,q_h)}$*-adjungierten Problems für den Beobachtungsoperator* $\mathcal{B}_{Q,h}$ *und* ω^*.

3. *Wenn* $F^* = 0$,
$$g^* = \tilde{\mathcal{B}}_{D,h}^T \omega^*$$
mit $\omega^* \in \mathbb{R}^{n_D}$ *und* $h^* = 0$, *so ist* $(\eta_h, \xi_h, \upsilon_h)$ *eine Lösung des* $\delta\tilde{\mathcal{G}}_{(\lambda_h,\psi_h,q_h)}$*-adjungierten Problems für den Beobachtungsoperator* $\tilde{\mathcal{B}}_{D,h}$ *und* ω^*.

Beweis:

$$\begin{aligned}
&\delta\tilde{\mathrm{G}}_{(\lambda_h,\psi_h,q_h)}((\Theta_1, K_1), (\Theta_2, K_2)) \\
&\quad = \tilde{\mathcal{G}}((\lambda_{h,1}, \psi_{h,1}, q_{h,1}), (\Theta_1, K_1)) - \tilde{\mathcal{G}}((\lambda_{h,2}, \psi_{h,2}, q_{h,2}), (\Theta_1, K_1))
\end{aligned}$$

ist linear in $(\lambda_{h,1} - \lambda_{h,2}, \psi_{h,1} - \psi_{h,2}, q_{h,1} - q_{h,2})$. Mit Umformungen der Art

$$\begin{aligned}
&K_1(\psi_{T,1}^i)\psi_{T,1}^i - K_1(\psi_{T,2}^i)\psi_{T,2}^i \\
&\quad = K_1(\psi_{T,1}^i)\psi_{T,1}^i - K_1(\psi_{T,1}^i)\psi_{T,2}^i + K_1(\psi_{T,1}^i)\psi_{T,2}^i K_1(\psi_{T,2}^i)\psi_{T,2}^i \\
&\quad = \left\{ K_1(\psi_{T,1}^i) + \psi_{T,2}^i \int_0^1 K_1' \left(\psi_{T,2}^i + s(\psi_{T,1}^i - \psi_{T,2}^i)\right) \, ds \right\} (\psi_{T,1}^i - \psi_{T,2}^i)
\end{aligned}$$

erhalten wir
$$(\eta_h^T, \xi_h^T, \upsilon_h^T)\, \delta\tilde{\mathrm{G}}_{(\lambda_h,\psi_h,q_h)} = \mathcal{B}_{\cdot,h}^T \omega^*.$$

□

3.4.4 Berechnung des diskreten Gradienten

Im numerischen Verfahren werden die Koeffizientenfunktionen Θ und K ebenfalls diskretisiert. Geeignete Methoden werden im nächsten Kapitel beschrieben. Hier sei jetzt angenommen, dass die Funktionen Θ durch einen r_Θ-dimensionalen und K durch einen r_K-dimensionalen Parametervektor gegeben sind. Damit haben wir einen diskreten Parameterraum $\tilde{P} \subseteq \mathbb{R}^{\hat{r}}$ mit $\hat{r} := r_\Theta + r_K$. Es sei weiter angenommen, dass die Koeffizientenfunktionen nach den einzelnen Parametern differenzierbar sind.

Im Folgenden werden drei Methoden zur Berechnung des diskreten Gradienten des Fehlerfunktionals für einen gegebenen Parametervektor $p \in \tilde{P}$ beschrieben. Dabei verwenden wir ein Fehlerfunktional der Form

$$\tilde{\jmath}_h(\omega_h) = \tilde{\jmath}_{F,h}(\omega_{F,h}) + \tilde{\jmath}_{Q,h}(\omega_{Q,h}) + \tilde{\jmath}_{D,h}(\tilde{\omega}_{D,h}) + \tilde{\jmath}_{D,h}(\omega_{D,h}). \tag{3.56}$$

Wenn nicht alle Teilbeobachtungen berücksichtigt werden sollen, so sind im Weiteren die entsprechenden Terme jeweils durch Null zu ersetzen.

Approximation durch Differenzenquotienten

Die partiellen Ableitungen des Fehlerfunktionals nach den einzelnen Parametern können durch einen Differenzenquotienten angenähert werden. Bei geeigneter Schrittweite $\tau > 0$ lautet die Berechnungsvorschrift für den rechtsseitigen Differenzenquotienten

$$\frac{d\jmath_h}{dp}[p] \cdot e_j \approx \frac{\jmath_h(p + \tau e_j) - \jmath_h(p)}{\tau}$$

für $j = 1, ..., \hat{r}$, wobei e_j den j-ten Einheitsvektor bezeichnet.

Adjungierte Methode

Bei der adjungierten Methode wird der Gradient des Fehlerfunktionals auf dem im Abschnitt 2.2 beschriebenen Weg über die Lösung eines adjungierten Problems berechnet. Dessen Lösung entspricht der Lösung des adjungierten Problems aus Definition 3.28 mit den Koeffizienten

$$\left.\begin{array}{rcl} \alpha_T^{*^i} & := & \Theta'(\psi_T^i) \\ \beta_T^{*^i} & := & K(\psi_T^i) \\ \gamma_{TE}^{*^i} & := & K'(\psi_T^i) \sum_{E' \subset T} B^{-1}_{TEE'}(\psi_T^i - \lambda_{E'}^i - z_{TE'}) \\ \delta_T^{*^i} & := & K'(\psi_T^i) \left(\tilde{N}\psi_T^i - \sum_{E \subset T} \lambda_E^i \right) \end{array}\right\} \tag{3.57}$$

Hierbei bezeichnet (λ_h, ψ_h, q_h) die zugehörige Lösung des diskreten direkten Problems und die rechten Seiten sind gegeben durch

$$\begin{aligned} F_T^{*^i} &= \tilde{\jmath}'_{Q,h}[\omega_{Q,h}] \left(\mathcal{B}^T_{Q,h}\right)^i_T , \\ g_{TE}^{*^i} &= \tilde{\jmath}_{D,h}[\omega_{D,h}] \left(\mathcal{B}^T_{D,h}\right)^i_E \\ &\quad + \tilde{\jmath}'_{D,h}[\tilde{\omega}_{D,h}] \sum_{k=1}^{n_D} \frac{1}{2} \delta_{TE_k,TE} \left\{ \left(\Delta t^i\right)_{0<i\le i_k} + \left(\Delta t^{i+1}\right)_{0\le i<i_k} \right\}, \\ h_E^{*^i} &= \tilde{\jmath}'_{F,h}[\omega_{F,h}] \left(\mathcal{B}^T_{F,h}\right)^i_E . \end{aligned}$$

Da in der gemischten Formulierung die Beobachtungen nicht explizit von den Koeffizienten abhängen, gilt nach (2.16) für den Gradienten

$$\frac{d\mathcal{J}}{dp}[p] = -\left(\eta_h^T, \xi_h^T, \upsilon_h^T\right) \begin{pmatrix} \frac{\partial \mathcal{H}}{\partial p}[\lambda_h, \psi_h, q_h, p] \\ \frac{\partial \mathcal{G}}{\partial p}[\lambda_h, \psi_h, q_h, p] \\ \frac{\partial \mathcal{F}}{\partial p}[\lambda_h, \psi_h, q_h, p] \end{pmatrix}$$

$$= -\begin{pmatrix} \eta_h^T \frac{\partial \mathcal{H}}{\partial p_1}[\lambda_h, \psi_h, q_h, p] + \xi_h^T \frac{\partial \mathcal{G}}{\partial p_1}[\lambda_h, \psi_h, q_h, p] + \upsilon_h^T \frac{\partial \mathcal{F}}{\partial p_1}[\lambda_h, \psi_h, q_h, p] \\ \vdots \\ \eta_h^T \frac{\partial \mathcal{H}}{\partial p_{\hat{r}}}[\lambda_h, \psi_h, q_h, p] + \xi_h^T \frac{\partial \mathcal{G}}{\partial p_{\hat{r}}}[\lambda_h, \psi_h, q_h, p] + \upsilon_h^T \frac{\partial \mathcal{F}}{\partial p_{\hat{r}}}[\lambda_h, \psi_h, q_h, p] \end{pmatrix}$$

mit

$$\eta_h^T \frac{\partial \mathcal{H}}{\partial p_j}[\lambda_h, \psi_h, q_h, p] + \xi_h^T \frac{\partial \mathcal{G}}{\partial p_j}[\lambda_h, \psi_h, q_h, p] + \upsilon_h^T \frac{\partial \mathcal{F}}{\partial p_j}[\lambda_h, \psi_h, q_h, p] =$$

$$\sum_{i=0}^{m} \eta_h^{i^T} \frac{\partial \mathcal{H}^i}{\partial p_j}[\lambda_h, \psi_h, q_h, p] + \xi_h^{i^T} \frac{\partial \mathcal{G}^i}{\partial p_j}[\lambda_h, \psi_h, q_h, p] + \upsilon_h^{i^T} \frac{\partial \mathcal{F}^i}{\partial p_j}[\lambda_h, \psi_h, q_h, p].$$

Durch vollständiges Differenzieren erhalten wir für $j = 1, ..., \hat{r}$

$$\frac{\partial J}{\partial p_j}[p] = -\sum_{E \in \mathcal{E} \setminus \mathcal{E}_F} \sum_{T \supset E} \upsilon_{TE}^0 \frac{\partial K}{\partial p_j}[\psi_T^0] \sum_{E' \subset T} B_{TEE'}^{-1} \left(\psi_T^0 - \lambda_{E'}^0 - z_{TE'}\right)$$

$$-\sum_{i=1}^{m} \Bigg\{ \sum_{E \in \mathcal{E} \setminus \mathcal{E}_D} \eta_E^i \sum_{T \supset E} \frac{\partial K}{\partial p_j}[\psi_T^i] \sum_{E' \subset T} B_{TEE'}^{-1} (\psi_E^i - \lambda_{E'}^i - z_{TE'})$$

$$+\sum_{T \in \mathcal{T}} \xi_T^i \left[\frac{\partial K}{\partial p_j}[\psi_T^i] \left(\tilde{N} \psi_T^i - \sum_{E \subset T} \lambda_E^i \right) + \frac{|T|}{\Delta t^i b_T} \left(\frac{\partial \Theta}{\partial p_j}[\psi_T^i] - \frac{\partial \Theta}{\partial p_j}[\psi_T^{i-1}] \right) \right]$$

$$+ \sum_{E \in \mathcal{E} \setminus \mathcal{E}_F} \sum_{T \supset E} \upsilon_{TE}^i \frac{\partial K}{\partial p_j}[\psi_T^i] \sum_{E' \subset T} B_{TEE'}^{-1} (\psi_T^i - \lambda_{E'}^i - z_{TE'}) \Bigg\}. \tag{3.58}$$

Die Systemmatrix des dimensionsreduzierten adjungierten Problems mit den Koeffizienten (3.57) entspricht gerade der Jacobimatrix, die auch innerhalb des Newton-Verfahrens zur Lösung des diskreten direkten Problems benötigt wird. Damit folgt die Lösbarkeit des Gleichungssystems stets aus der Lösbarkeit des diskreten direkten Problems. Entsprechend den Ausführungen in [52] ist damit das allgemeine adjungierte Problem gemäß Definition 3.28 für eine hinreichend feine Diskretisierung eindeutig lösbar, falls die auftretenden Koeffizienten positiv und monoton wachsend sind. Unter diesen Bedingungen ist dann auch die Dimensionsreduzierung möglich.

Algorithmus 3.32 (Berechnung des Gradienten über die adjungierte Methode)
Löse das diskrete direkte Problem mit Algorithmus 3.27.
Speichere (λ_h, ψ_h, q_h).
Löse das adjungierte Problem mit Algorithmus 3.29.
Speichere $(\eta_h, \xi_h, \upsilon_h)$.
Für $j = 1, ..., \hat{r}$

Berechne $\frac{\partial \mathcal{J}}{\partial p_j}$ *nach (3.58).*

Bemerkung 3.33 Zur Berechnung des Gradienten des kontinuierlichen Problems ist das folgende adjungierte Problem zu lösen:

$$\left.\begin{array}{rcll} \Theta'(\psi)\partial_t\eta + \nabla\cdot\upsilon + \frac{K'(\psi)}{K(\psi)^2}q\cdot\upsilon &=& -F^*(x,t) & \\ \upsilon &=& K(\psi)\nabla\eta & \text{in } Q_T, \\ \eta(x,T) &=& 0 & \text{in } \Omega, \\ \upsilon(x,t)\cdot\nu &=& h^*(x,t) & \text{auf } \Gamma_{FT}, \\ \eta(x,t) &=& g^*(x,t) & \text{auf } \Gamma_{DT}. \end{array}\right\} \tag{3.59}$$

Eine Diskretisierung des kontinuierlichen adjungierten Problems führt auf ein diskretes Problem, welches im Allg. nicht äquivalent mit dem zum diskreten direkten Problem adjungierten Problem ist. Wenn wir in unserem Fall das kontinuierliche adjungierte Problem analog zum direkten Problem mit der hybrid-gemischten Finite-Elemente-Methode diskretisieren, so können wir durch geringfügige Modifikationen im erhaltenen diskreten adjungierten Problem zum adjungierten Problem aus Definition 3.28 gelangen.

Direkte Methode

Bei der direkten Methode wird nicht zuerst ein adjungiertes Problem gelöst, sondern die partiellen Ableitungen werden direkt aus (2.14) berechnet. Aus (3.43) erhalten wir durch vollständiges differenzieren

$$\begin{aligned} \frac{\partial \mathcal{H}^i}{\partial \lambda_h^i}\frac{\lambda_h^i}{\partial p_j} + \frac{\partial \mathcal{H}^i}{\partial \psi_h^i}\frac{\psi_h^i}{\partial p_j} &= -\frac{\partial \mathcal{H}^i}{\partial p_j} \\ \frac{\partial \mathcal{G}^i}{\partial \lambda_h^i}\frac{\lambda_h^i}{\partial p_j} + \frac{\partial \mathcal{G}^i}{\partial \psi_h^i}\frac{\psi_h^i}{\partial p_j} &= -\frac{\partial \mathcal{G}^i}{\partial \psi_h^{i-1}}\frac{\psi_h^{i-1}}{\partial p_j} - \frac{\partial \mathcal{G}^i}{\partial p_j} \\ \frac{\partial \mathcal{F}^i}{\partial \lambda_h^i}\frac{\lambda_h^i}{\partial p_j} + \frac{\partial \mathcal{F}^i}{\partial \psi_h^i}\frac{\psi_h^i}{\partial p_j} + \frac{\partial \mathcal{F}^i}{\partial q_h^i}\frac{q_h^i}{\partial p_j} &= -\frac{\partial \mathcal{F}^i}{\partial p_j}. \end{aligned}$$

Zur Vereinfachung der Notation schreiben wir für die partiellen Ableitungen $\frac{\partial}{\partial p_j}$ den Index "p_j" und erhalten für $j = 1, ..., \hat{r}$

$$\sum_{T \supset E} \Bigg\{ \Big(K'(\psi_T^i)\psi_{T,p_j}^i + K_{p_j}(\psi_T^i) \Big) \sum_{E' \subset T} B_{TEE'}^{-1}(\psi_T^i - \lambda_{E'}^i - z_{TE'}) + K(\psi_T^i) \sum_{E' \subset T} B_{TEE'}^{-1}(\psi_{T,p_j}^i - \lambda_{E',p_j}^i) \Bigg\} = 0 \quad \forall E \in \mathcal{E} \setminus \mathcal{E}_D \tag{3.60}$$

$$\big(K'(\psi_T^i) + K_{p_j}(\psi_T^i) \big) \left(\tilde{N}\psi_T^i - \sum_{E \subset T} \lambda_E^i \right) + K(\psi_T^i) \left(\tilde{N}\psi_{T,p_j}^i - \sum_{E \subset T} \lambda_{T,p_j}^i \right) + \frac{|T|}{\Delta^i b_T} \Big(\Theta'(\psi_T^i)\psi_{T,p_j}^i - \Theta'(\psi_T^{i-1})\psi_{T,p_j}^{i-1} + \Theta_{p_j}(\psi_T^i) - \Theta_{p_j}(\psi_T^{i-1}) \Big) = 0 \quad \forall T \in \mathcal{T} \tag{3.61}$$

$$\Big(K'(\psi_T^i)\psi_{T,p_j}^i + K_{p_j}(\psi_T^i) \Big) \sum_{E' \subset T} B_{TEE'}^{-1}(\psi_T^i - \lambda_{E'}^i - z_{TE'}) + K(\psi_T^i) \sum_{E' \subset T} B_{TEE'}^{-1}(\psi_{T,p_j}^i - \lambda_{E',p_j}^i) = q_{TE,p_j}^i \quad \forall T \in \mathcal{T}, E \subset T, E \notin \mathcal{E}_F \tag{3.62}$$

und $\lambda_{E,p_j}^i = 0 \; \forall E \in \mathcal{E}_D$, $q_{TE,p_j}^i = 0 \; \forall E \in \mathcal{E}_F$ für $i = 1, ..., m$, sowie $\lambda_{E,p_j}^0 = 0 \; \forall E \in \mathcal{E}$, $\psi_{T,p_j}^0 = 0 \; \forall T \in \mathcal{T}$ und (3.62) für $i = 0$.

Auch hierbei handelt es sich um ein lineares Gleichungssystem, welches analog zum diskreten direkten Problem und zum adjungierten Problem zu einem System für die Unbekannten λ_{E,p_j}^i reduziert wird. Die Systemmatrix entspricht auch hier der Jacobimatrix aus dem Newton-Verfahren. Analog zum adjungierten Problem ist damit das Gleichungssystem (3.60–3.62) immer lösbar, wenn das diskrete direkte Problem lösbar ist. Für die einzelnen Parameter p_j ändert sich hier jeweils nur die rechte Seite.

Die partiellen Ableitungen der Beobachtungen sind gegeben durch

$$\begin{aligned}
\frac{\partial \omega_{F,E_k}^{i_k}}{\partial p_j} &= \lambda_{E_k,p_j}^{i_k}, \qquad k = 1, ..., n_F \\
\frac{\partial \omega_{Q,T_k}^{i_k}}{\partial p_j} &= \psi_{T_k,p_j}^{i_k}, \qquad k = 1, ..., n_Q \\
\frac{\partial \tilde{\omega}_{D,TE_k}^{i_k}}{\partial p_j} &= q_{TE_k,p_j}^{i_k}, \qquad k = 1, ..., n_D \\
\frac{\partial \omega_{D,TE_k}^{i_k}}{\partial p_j} &= \frac{1}{2} \sum_{i=0}^{i_k - 1} \Delta t^{i+1} \left(q_{TE_k,p_j}^i + q_{TE_k,p_j}^{i+1} \right), \qquad k = 1, ..., n_D
\end{aligned}$$

und werden später in der Sensitivitätsmatrix zusammengefasst (siehe Unterabschnitt 4.1.5). Der Gradient setzt sich zusammen aus

$$\frac{\partial \mathcal{J}_h}{\partial p_j}[p] = \frac{\partial \tilde{\mathcal{J}}_{F,h}}{\partial \omega_{F,h}} \mathcal{B}_{F,h}(\lambda_{h,p_j}, \psi_{h,p_j}, q_{h,p_j}) + \frac{\partial \tilde{\mathcal{J}}_{Q,h}}{\partial \omega_{Q,h}} \mathcal{B}_{Q,h}(\lambda_{h,p_j}, \psi_{h,p_j}, q_{h,p_j}) + \frac{\partial \tilde{\mathcal{J}}_{D,h}}{\partial \tilde{\omega}_{D,h}} \tilde{\mathcal{B}}_{D,h}(\lambda_{h,p_j}, \psi_{h,p_j}, q_{h,p_j}) + \frac{\partial \tilde{\mathcal{J}}_{D,h}}{\partial \omega_{D,h}} \mathcal{B}_{D,h}(\lambda_{h,p_j}, \psi_{h,p_j}, q_{h,p_j}) \qquad (3.63)$$

für $j = 1, ..., \hat{r}$.

Algorithmus 3.34 (Berechnung des Gradienten mit der direkten Methode)
Setze $\lambda_h^0 = \psi_0|_{\mathcal{E}}$, $\psi_h^0 = \psi_0|_{\mathcal{T}}$.
Berechne q_h^0 *aus* $\mathcal{F}^0(\lambda_h^0, \psi_h^0, q_h^0, K) = 0$.
Für $j = 1, ..., \hat{r}$

> *Setze* $\lambda_{h,p_j}^0 = 0$, $\psi_{h,p_j}^0 = 0$.
> *Berechne* q_{h,p_j}^0 *aus (3.62) mit* $i = 0$.

Für $i = 1, ..., m$

> *Löse* $\mathcal{H}^i\left(\lambda_h^i, G^i\left(\sum_{E \subset T} \lambda_E^i, \psi_h^{i-1}\right), K\right) = 0$.
> *Berechne* $\psi_h^i = G^i\left(\sum_{E \subset T} \lambda_E^i, \psi_h^{i-1}\right)$.
> *Berechne* q_h^i *aus* $\mathcal{F}^i(\lambda_h^i, \psi_h^i, q_h^i, K) = 0$.
> *Für* $j = 1, ..., \hat{r}$
>
> > *Substituiere* ψ_{h,p_j}^i *in (3.60) mithilfe von (3.61).*
> > *Löse das erhaltene lineare Gleichungssystem für* λ_{h,p_j}^i.
> > *Berechne* ψ_{h,p_j}^i *aus (3.61).*
> > *Berechne* q_{h,p_j}^i *gemäß (3.62).*
> > *Berechne und (speichere)* ω_{F,h,p_j}^i, ω_{Q,h,p_j}^i, $\tilde{\omega}_{D,h,p_j}^i$ *und* ω_{D,h,p_j}^i.
> > *Datiere den Gradienten nach (3.63) auf.*

Bei der Berechnung des Gradienten mit der direkten Methode ist also bei der Lösung des diskreten direkten Problems lediglich ein *Postprocessing* erforderlich, welches im Wesentlichen aus der Lösung von linearen Systemen besteht. Wenn die Ableitungen der Beobachtungen nicht für andere Zwecke als für die Gradientenberechnung benötigt werden, so müssen diese nicht abgespeichert werden.

Bemerkung 3.35 Die partiellen Ableitungen der Beobachtungen nach p_j können auch über das adjungierte Problem berechnet werden. Dazu sind die Funktionale $\tilde{\mathcal{J}}_{F,h}$, $\tilde{\mathcal{J}}_{Q,h}$ und $\tilde{\mathcal{J}}_{D,h}$ als Identität zu wählen und die Beobachtungsoperatoren $\mathcal{B}_{F,h}$, $\mathcal{B}_{Q,h}$, $\tilde{\mathcal{B}}_{D,h}$ und $\mathcal{B}_{D,h}$ so zu variieren, dass sie gerade den gewünschten Teil von $\omega_{F,h}$, $\omega_{Q,h}$, $\tilde{\omega}_{D,h}$ bzw. $\omega_{D,h}$ herausfiltern.

Vergleich der Methoden

Die Berechnung des Gradienten mit der Finite-Differenzen-Methode erfordert den meisten Rechenaufwand. Bei Verwendung des einseitigen Differenzenquotienten ist insgesamt $\hat{r}+1$-mal das diskrete direkte Problem, also ein nichtlineares Gleichungssystem, zu lösen. Um mit dem zentralen Differenzenquotienten eine höhere Approximationsgenauigkeit zu erhalten, erhöht sich der Aufwand auf die $2\hat{r}+1$-malige Lösung des diskreten direkten Problems.

Bei Anwendung der adjungierten Methode ist neben dem diskreten direkten Problem das zugehörige adjungierte Problem zu lösen. Da dies jedoch rückwärts in der Zeit gelöst werden muss und dessen Koeffizienten von der Lösung des diskreten direkten Problems abhängen, muss diese komplett abgespeichert werden. Das adjungierte Problem ist linear, sodass in jedem Zeitschritt lediglich ein lineares Gleichungssystem zu lösen ist, um den Gradienten des Fehlerfunktionals zu berechnen. Der Rechenzeitaufwand ist also geringer als die zweifache Lösung des diskreten direkten Problems mit dem Newton-Verfahren.

Bei der direkten Methode werden in jedem Zeitschritt $\hat{r}$ zusätzliche lineare Gleichungssysteme gelöst. Für den Fall das $\hat{r}$ nicht zu groß ist, führt die direkte Methode ungefähr zu einer Verdopplung der Rechenzeit gegenüber dem direkten Problem. Im Gegensatz zur adjungierten Methode ist es jedoch nicht notwendig die komplette Lösung des diskreten direkten Problems abzuspeichern. In jedem Zeitschritt wird nur die diskrete direkte Lösung im aktuellen Zeitpunkt und der Druck auf einem Element für den vorhergehenden Zeitschritt benötigt.

Da zur Identifizierung der hydraulischen Funktionen aus Säulenexperimenten nur das räumlich eindimensionale Modell betrachtet wird, werden wir die auftretenden linearen Gleichungssysteme durch Gauß-Elimination lösen.

Wenn neben dem Gradienten des Fehlerfunktionals auch sämtliche partiellen Ableitungen der Beobachtungen nach den Parametern p_j von Interesse sind, so erhöht sich der Rechenaufwand in der adjungierten Methode, während bei der direkten Methode diese Ableitungen ohnehin schon berechnet werden. In der adjungierten Methode sind dann insgesamt $\hat{\kappa} := n_F + n_Q + 2n_D$ adjungierte Probleme mit variierender rechter Seite zu lösen. Dies führt insgesamt zu $\hat{\kappa}$ zusätzlichen linearen Gleichungssystemen pro Zeitschritt. Somit ist die direkte Methode effizienter zur Berechnung der partiellen Ableitungen der Beobachtungen, wenn gilt

$$\hat{r} < \hat{\kappa}. \tag{3.64}$$

Im anderen Fall wäre die adjungierte Methode vorzuziehen. Bei den hier betrachteten inversen Problemen sollte die Gesamtzahl der diskreten Beob-

achtungen jedoch die Anzahl der Parameter in den Koeffizientenfunktionen übersteigen, sodass (3.64) erfüllt und damit die direkte Methode am effizientesten für die Berechnung der Ableitungen der Beobachtungen nach den Parametern in den Koeffizientenfunktionen ist.

Kapitel 4

Numerische Behandlung des inversen Problems und Beispiele

4.1 Numerisches Verfahren für die Identifizierung

4.1.1 Fehlerfunktional und Beobachtungen

Im Weiteren sei mit S diejenige Teilmenge des räumlichen Gebietes Ω bezeichnet, auf der die jeweilige Beobachtung definiert ist. Durch eine gewichtete L^2-Norm gelangen wir für eine Beobachtung $\omega \in L^2(S_T)$ zu einem Fehlerfunktional

$$\tilde{\mathcal{J}}_\varepsilon : L^2(S_T) \to \mathbb{R}_+$$

mit

$$\tilde{\mathcal{J}}_\varepsilon(\omega) = \int_{S_T} (\omega - \omega_\varepsilon)(x,t)\alpha(x,t)(\omega - \omega_\varepsilon)(x,t)$$

und einer Wichtungsfunktion $\alpha(x,t) \in L^\infty(S_T)$, $\alpha(x,t) > 0$, welches Fréchet-differenzierbar ist mit

$$\left\langle \tilde{\mathcal{J}}'_\varepsilon[\omega], \delta\omega \right\rangle = 2 \int_{S_T} (\omega - \omega_\varepsilon)(x,t)\alpha(x,t)\delta\omega(x,t)$$

und die Bedingung

$$\tilde{\mathcal{J}}_\varepsilon(\omega) = 0$$

genau dann erfüllt, wenn die simulierte Beobachtung ω mit der tatsächlichen Beobachtung ω_ε übereinstimmt. Wenn eine endliche Anzahl von Beobachtungen $\omega_k \in L^2(S_{k,T})$, $k = 1, ..., \kappa$ an fixierten Ortskoordinaten durchgeführt

wird, so minimieren wir das Funktional

$$\tilde{\mathcal{J}}_\varepsilon(\omega) = \sum_{k=1}^{\kappa} \int_{S_{k,T}} \alpha_k(t) \left(\omega_k(t) - \omega_{\varepsilon,k}(t)\right)^2 .$$

Bei der Modellierung der auftretenden Messfehler wird davon ausgegangen, dass die exakte Beobachtung $\omega_{0,k}(t)$ durch ein weißes Rauschen überlagert wird. Ein mögliches Modell ist z. B.

$$\omega_{\varepsilon,k}(t) = \omega_{0,k}(t) + \zeta\left(t; \frac{\varepsilon}{3}\right) \sup_{t\in[0,T]} |\omega_{0,k}(t)|, \tag{4.1}$$

wobei $\varepsilon \in [0,1]$ und $\zeta\left(t; \frac{\varepsilon}{3}\right)$, $t \in [0,T]$ ein Gaußscher Prozess ist mit der Erwartungswertfunktion

$$\mathrm{E}\zeta\left(t; \frac{\varepsilon}{3}\right) = 0 \quad \forall t \in [0,T]$$

und der Kovarianzfunktion

$$\mathrm{Cov}\left(\zeta\left(t_1; \frac{\varepsilon}{3}\right), \zeta\left(t_2; \frac{\varepsilon}{3}\right)\right) = \frac{\varepsilon^2}{9}\delta(t_1 - t_2)$$

für $t_1, t_2 \in [0,T]$ mit

$$\delta(t) := \begin{cases} 1 & \text{für } t = 0, \\ 0 & \text{sonst.} \end{cases}$$

Hierbei gilt für jedes $\hat{t} \in [0,T]$

$$|\omega_{\varepsilon,k}(\hat{t}) - \omega_{0,k}(\hat{t})| < \varepsilon \sup_{t\in[0,T]} |\omega_{0,k}(t)|$$

mit einer Wahrscheinlichkeit von 0.9973. Bei diesem Modell werden Korrelationen, d. h. Abhängigkeiten, zwischen Beobachtungen zu verschiedenen Zeitpunkten also ebenso ausgeschlossen, wie die einzelnen Beobachtungen ω_k als voneinander unabhängig betrachtet werden.

Beispiel 4.1 Wenn der kumulative Ausfluss, welcher zu diskreten Zeiten t^i beobachtet wird, in die Identifizierung der hydraulischen Funktionen eingeht, so kann dieser prinzipiell auf zwei Wegen ermittelt werden. Einmal kann zu jedem Zeitpunkt t^i direkt das komplette Ausflussvolumen seit dem Beginn des Experiments gemessen werden, womit wir sofort den kumulativen Ausfluss erhalten. In diesem Fall ist das obige Fehlermodell anwendbar. Andererseits wäre es auch möglich zu jedem Zeitpunkt t^i das während des Zeitintervalls $[t^{i-1}, t^i)$ ausgeflossene Volumen zu messen und den kumulativen

Ausfluss zum Zeitpunkt t^i durch Addition des erhaltenen Messwertes zum kumulativen Ausfluss zum Zeitpunkt t^{i-1} zu bestimmen. In diesem Fall ist obiges Fehlermodell jedoch nicht mehr korrekt, da wir hier eine Fehleraddition erhalten und damit der Wert des kumulativen Ausflusses zum Zeitpunkt t^i mit den Werten zu den vorhergehenden Zeitpunkten $t^{i-1}, t^{i-2}, ...$ korreliert wäre. Dieser Sachverhalt sollte dann im Fehlerfunktional durch entsprechend Gewichte berücksichtigt werden.

Für eine numerische Optimierungsprozedur zur Lösung von (2.5) werden die entsprechenden diskreten Versionen des direkten Lösungsoperators $\mathcal{A}$, der Beobachtungsoperatoren $\mathcal{B}_k, \mathcal{B}$ und des Funktionals $\tilde{\mathcal{J}}_\varepsilon$ benötigt, die wir analog Kapitel 3 mit $\mathcal{A}_h$, $\mathcal{B}_{k,h}, \mathcal{B}_h$ und $\tilde{\mathcal{J}}_{\varepsilon,h}$ bezeichnen. Die Diskretisierung unseres speziellen Problems wurde bereits in Abschnitt 3.4 beschrieben. Die diskrete Version des Funktional $\tilde{\mathcal{J}}_\varepsilon$ lautet

$$\tilde{\mathcal{J}}_{\varepsilon,h}(\omega_h) = \sum_{k=1}^{\kappa} \sum_{i=1}^{n_k} \alpha_k^i \left(\omega_k^i - \omega_{\varepsilon,k}^i\right)^2 \tag{4.2}$$

mit dem Gradienten

$$\nabla \tilde{\mathcal{J}}_{\varepsilon,h}(\omega_h) = \left(2\alpha_k^i \left(\omega_k^i - \omega_{\varepsilon,k}^i\right)\right)_{\substack{i=1,\dots,n_k \\ k=1,\dots,\kappa}}$$

für die zu diskreten Zeitpunkten $\tilde{t}_k^i$, $i = 1, .., n_k$ stattfindenden Beobachtungen $\omega_{k,h}$, $k = 1, ..., \kappa$. Um möglichen Unterschieden in den Größenordnungen der betrachteten Beobachtungen Rechnung zu tragen, ist es vorteilhaft die von den Messzeitpunkten unabhängigen Wichtungsfaktoren

$$\alpha_k = \left(\frac{1}{n_k} \sum_{i=1}^{n_k} \omega_{\varepsilon,k}^i\right)^{-2} \tag{4.3}$$

zu verwenden (vgl. [22] und [35]).

In einem nächsten Schritt werden wir nun die unbekannten Nichtlinearitäten diskretisieren, d. h. durch eine endliche Anzahl von reellen Parametern beschreiben. Für diese Parametrisierung verwenden wir Spline-Ansätze, bei denen im Gegensatz zu sonst üblichen Modellen (van Genuchten-Mualem, etc.) nicht von einer starren Form ausgegangen wird.

4.1.2 Spline-Parametrisierungen

Zur Parametrisierung einer nichtlinearen Funktion $f \in P = C_{\text{pwa}}(\mathbb{R})$ bzw. $C^1_{\text{pwa}}(\mathbb{R})$ (Definition analog zu Definition 2.17) wählen wir uns eine Folge

endlichdimensionaler Unterräume P_r der Dimension $r \in \mathbb{N}_+$ mit

$$P_r \subset P_{r'} \quad \text{für } r \leq r' \tag{4.4}$$

und

$$\overline{\bigcup_{r \in \mathbb{N}_+}} P_r = P.$$

Diese Unterräume repräsentieren wir durch eine Basis $\{\Phi_{j,r}\}_{j=1}^r$:

$$P_r = \text{span}\{\Phi_{j,r}\}_{j=1}^r.$$

Da jede Funktion aus diesem Unterraum einer Darstellung der Form

$$f_r(\psi) = \sum_{j=1}^{r} p_{j,r} \Phi_{j,r}(\psi) \tag{4.5}$$

mit $p_{j,r} \in \mathbb{R},\ j = 1, ..., r$ genügt, sind die Funktionen aus P_r eindeutig durch einen Parametervektor $p_r = (p_{1,r}, ..., p_{r,r})^T \in \mathbb{R}^r$ definiert. Infolge dieser eineindeutigen Zuordnung können die Räume P_r mit den zugehörigen diskreten Parameterräumen identifiziert werden. Die Dimension r bezeichnet die Anzahl der Freiheitsgrade. Die Glattheitseigenschaften der Formfunktionen Φ_j bestimmen die Glattheitseigenschaften des Raumes P_r.

Die Projektionsoperatoren

$$\Pi_r : P \to P_r,$$

für die

$$\lim_{r \to \infty} \|\Pi_r f - f\| = 0$$

gelten soll, werden durch geeignete Interpolationen definiert. Derartige Projektionsverfahren sind auch gängige Stabilisierungsverfahren, bei denen die Dimension der Unterräume die Rolle des Regularisierungsparameters übernimmt.

Zunächst legen wir ein Grundintervall $[\underline{\psi}, \overline{\psi}] \subset \mathbb{R}$ und eine Zerlegung dieses Intervalls fest:

$$\underline{\psi} = \psi_{1,\tilde{r}} < \psi_{2,\tilde{r}} ... < \psi_{\tilde{r},\tilde{r}} = \overline{\psi} \quad \text{(Stützstellen)}.$$

Damit die Funktion $f_r \in P_r$ auf dem gesamten Raum $\mathbb{R}$ definiert ist, können wir diese Funktion außerhalb des Intervalls $[\underline{\psi}, \overline{\psi}]$ als konstant annehmen. Nun können grundlegend zwei Konzepte verfolgt werden:

Lokale Basen

Hierbei handelt es sich um eine Parametrisierung mit den üblichen B-Splines als Basisfunktionen (vergl. z. B. [25]). Bei einem stückweise linearen Ansatz entsteht die *Lagrange-Basis* (Hutfunktionen), welche definiert ist durch

$$\Phi_{j,\tilde{r}}(\psi_{i,\tilde{r}}) = \delta_{ij} \quad \text{für} \quad 1 \leq i,j \leq \tilde{r}. \tag{4.6}$$

Dabei bezeichnet δ_{ij} das Kronecker-Symbol. Für diese Basis, die bei $\tilde{r}$ Stützstellen aus $r = \tilde{r}$ Basisfunktionen besteht, ist die Interpolation einer Funktion f im Raum P_r leicht zu lösen: $f_r \in P_r$, sodass

$$f_r(\psi_{i,r}) = f(\psi_{i,r}) \quad \text{für alle } i = 1, ..., r$$

gilt, ist durch

$$p_{j,r} = f(\psi_{j,r}), \quad j = 1, ..., r$$

gegeben. Die Elemente des so definierten Raumes P_r sind in den Stützstellen nicht differenzierbar. Sie können nur als links- oder rechtsseitig stetig differenzierbar angenommen werden. Deshalb sind entsprechende Parametrisierungen mit B-Splines höherer Ordnung einzusetzen, wenn stärkere Differenzierbarkeitseigenschaften benötigt werden.

Die Interpolationsaufgabe bei solchen höheren Ansätzen erfordert die Lösung eines linearen Gleichungssystems zur Bestimmung der Parameter $p_{j,r}$. So ist beispielsweise im Raum, der durch quadratische B-Splines gebildet wird, eine Interpolation definiert durch

$$f_r(\xi_{i,\tilde{r}}) = f(\xi_{i,\tilde{r}})$$

für die Interpolationsstellen

$$\xi_{i,\tilde{r}} := \frac{1}{2}(\psi_{i,\tilde{r}} + \psi_{i+1,\tilde{r}}), \quad i = 1, .., \tilde{r} - 1$$

sowie $\xi_{0,\tilde{r}} = \psi_{1,\tilde{r}}$ und $\xi_{\tilde{r},\tilde{r}} = \psi_{\tilde{r},\tilde{r}}$. Dies führt auf ein Tridiagonalsystem. Eine Parametrisierung mit quadratischen B-Splines besteht bei $\tilde{r}$ Stützstellen aus $r = \tilde{r} + 1$ Freiheitsgraden. Bei kubischen B-Splines kann eine Interpolation mit natürlichen Endbedingungen oder mit Hermite-Endbedingungen gewählt werden (siehe z. B. [25], Kap.6, §3).

Bemerkung 4.2 Als Basisfunktionen ungeeignet sind stückweise konstante Funktionen, da diese in den Stützstellen unstetig sind und in allen Stetigkeitspunkten deren Ableitung verschwindet.

Anstelle der lokalen Basen sind auch Basen einsetzbar, wie sie z. B. in [46] für Wavelets konstruiert worden sind.

Hierarchische Basen

Das Prinzip der hierarchischen Basen beruht auf einer skalenweisen Parametrisierung. Die Stufe der Parametrisierung wird durch den *Skalenindex* s angegeben. Beginnend mit dem Skalenindex 0 gelangt man durch Hinzunahme eines kompletten Satzes neuer Basisfunktionen auf die jeweils nächsthöhere Stufe der Parametrisierung. Dazu wählt man eine *Skalierungsfunktion* φ mit dem Träger $[0,1]$. Diese kann ein B-Spline sein. Der Ansatzraum P^0 zum Skalenindex 0 ist definiert durch die lokale Basis für die 2 Stützstellen $\underline{\psi}$ und $\overline{\psi}$:

$$P^0 = \text{span}\{\Phi_{j,r_0}\}_{j=1}^{r_0}.$$

(Dabei ist r_0 abhängig vom Grad der Skalierungsfunktion φ.)

Die hierarchische Basis zum Skalenindex $s \geq 1$ besteht nun aus den Basisfunktionen von P^0 und den Funktionen

$$\varphi_i^j(\psi) = \varphi\left(2^{j-1}\frac{\psi - \underline{\psi}}{\overline{\psi} - \underline{\psi}} - (i-1)\right) \quad \text{für } j = 1, ..., s, i = 1, ..., 2^{j-1}.$$

Damit ist der Ansatzraum P^s darstellbar als eine direkte Summe

$$\begin{aligned} P^s &= P^0 \oplus V^1 \oplus ... \oplus V^s \\ &= P^{s-1} \oplus V^s \end{aligned}$$

der von den Funktionen φ_i^j, $i = 1, ..., 2^{j-1}$, aufgespannten Räume

$$V^j := \text{span}\{\varphi_i^j\}_{i=1}^{2^{j-1}}.$$

Die Abbildung 4.1 zeigt die Basen der Skalen 0, 1 und 2 für den linearen B-Spline als Skalierungsfunktion.

Eine Funktion $f^s \in P^s$ besitzt $r = 2^s + r_0 - 1$ Freiheitsgrade zu $\tilde{r} = 2^s + 1$ Stützstellen und ist darstellbar durch

$$f^s(\psi) = f_r(\psi) = \sum_{i=1}^{r_0} p_i^0 \Phi_{i,r_0}(\psi) + \sum_{j=1}^{s}\sum_{i=1}^{2^{j-1}} p_i^j \varphi\left(2^{j-1}\frac{\psi - \underline{\psi}}{\overline{\psi} - \underline{\psi}} - (i-1)\right). \quad (4.7)$$

Bei äquidistanten Stützstellen besitzen alle Basisfunktionen einer lokalen Basis von P_r (mit Ausnahme von $\Phi_{1,r}$ und $\Phi_{r,r}$) die gleiche L^2-Norm. Dies ist bei den hierarchischen Basen nicht der Fall. Alle Elemente des Raumes V^j haben zwar die gleiche L^2-Norm, diese wird jedoch mit wachsendem Index j kleiner, d. h.

$$\|\varphi_{i_2}^{j_2}\|_2 < \|\varphi_{i_1}^{j_1}\|_2$$

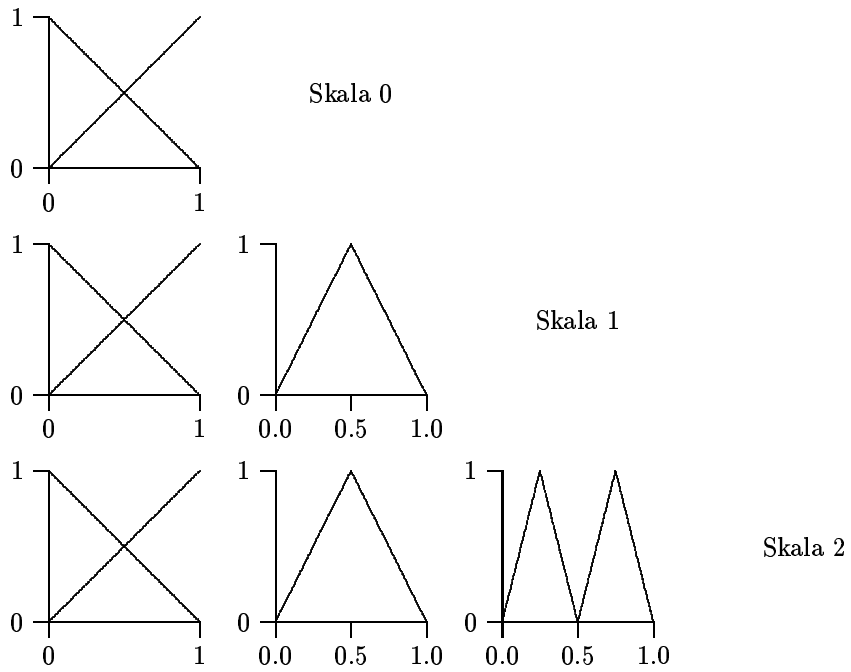


Abbildung 4.1: Hierarchische Basen der Skalen 0, 1 und 2 für den linearen B-Spline über das Intervall $[0,1]$.

für $1 \leq j_1 < j_2, i_1 = 1, ..., 2^{j_1-1}, i_2 = 1, ..., 2^{j_2-1}$. Die Funktionen der Räume P^0 und V^1 beinflussen die Koeffizientenfunktionen global, während sich die Beinflussung durch Funktionen der Räume $V^j, j > 1$ infolge der kleiner werdenden Träger mit wachsendem j immer stärker lokalisiert. Sämtliche Elemente einer lokalen Basis besitzen dagegen den gleichen lokalen Charakter.

Es seien $\{\Phi_{j,r}^{\text{hier}}\}_{j=1}^{r=2^s+r_0-1}$ die hierarchische Basis zum Skalenindex s für den linearen B-Spline und $\{\Phi_{j,r}^{\text{loc}}\}_{j=1}^{r}$ die lokale Basis für den linearen B-Spline bei äquidistanten Stützstellen, jeweils für r Freiheitsgrade. p_r^{hier} und p_r^{loc} seien die zugehörigen Parametervektoren einer Funktion $f \in P_r$. Diese Parametervektoren können durch

$$p_r^{\text{loc}} = M_r^{\text{hier}} p_r^{\text{hier}}$$

mit einer regulären Matrix $M_r^{\text{hier}} \in \mathbb{R}^{r \times r}$ ineinander transformiert werden:

$$f_r(\psi) = \sum_{j=1}^{r} p_{j,r}^{\text{loc}} \Phi_{j,r}^{\text{loc}}(\psi) = \sum_{j=1}^{r} p_{j,r}^{\text{hier}} \Phi_{j,r}^{\text{hier}}(\psi)$$

$$\Longleftrightarrow$$

$$f_r(\psi_{i,r}) = \sum_{j=1}^{r} p_{j,r}^{\text{loc}} \Phi_{j,r}^{\text{loc}}(\psi_{i,r}) = \sum_{j=1}^{r} p_{j,r}^{\text{hier}} \Phi_{j,r}^{\text{hier}}(\psi_{i,r}) \quad \text{für alle } i = 1, ..., r,$$

wobei $\psi_{1,r} < ... < \psi_{r,r}$ die äquidistanten Stützstellen sind. Damit ist die Matrix M_r^{hier} gegeben durch

$$M_r^{\text{hier}} = \left(\Phi_{j,r}^{\text{hier}}(\psi_{i,r})\right)_{i,j=1}^{r}.$$

Die Regularität dieser Matrix folgt aus der Definition der Basisfunktionen $\{\Phi_{j,r}^{\text{hier}}\}$.

Für Basen bezüglich höherer B-Splines kann gemäß der Interpolation für lokale Basen eine analoge Transformation definiert werden:

$$M_r^{\text{loc}} p_r^{\text{loc}} = M_r^{\text{hier}} p_r^{\text{hier}}.$$

Von den gesuchten Funktionen ist bekannt, dass sie aus physikalischen Gründen monoton wachsend sind. Indem wir aus dieser Monotonieeigenschaft Bedingungen für die Parameter ableiten, wird eine zusätzliche Stabilisierung des Identifizierungsverfahrens erreicht (siehe hierzu Abschnitt 4.2).

Wenn eine Parametrisierung (4.5) mit linearen Splines verwendet wird, so sind die Bedingungen

$$p_{j,r} \leq p_{j+1,r} \quad \text{für } j = 1, ..., r-1. \tag{4.8}$$

hinreichend dafür, dass $f_r(\psi)$ monoton wachsend ist. Für Splines höherer Ordnung kann die Monotonie durch die linearen Nebenbedingungen

$$\sum_{j=1}^{r} p_{j,r}\Phi_{j,r}(\psi_i) \leq \sum_{j=1}^{r} p_{j,r}\Phi_{j,r}(\psi_{i+1}) \quad \text{für } i = 1, ..., l-1 \tag{4.9}$$

berücksichtigt werden, wobei die Zerlegung $\underline{\psi} = \psi_1 < \psi_2 < ... < \psi_l = \overline{\psi}$ eine höhere Feinheit besitzt als diejenige Zerlegung, welche zur Definition der Ansatzfunktionen $\Phi_{j,r}$ verwendet wird (d. h. $l > \tilde{r}$).

Eine Möglichkeit die Monotonie allein durch die Bedingungen (4.8) zu gewährleisten und dennoch eine höhere Glattheit zu erhalten, wird im folgenden Unterabschnitt beschrieben.

4.1.3 Monotoner stückweise kubischer Ansatz

Wir gehen aus von einer Zerlegung $\underline{\psi} = \psi_{1,r} < \psi_{2,r} < ... < \psi_{r,r} = \overline{\psi}$. Die Freiheitsgrade einer Funktion $f_r \in \overline{P}_r$ sollen den Funktionswerten in den Stützstellen entsprechen:

$$f_r(\psi_{j,r}) = p_{j,r}, \quad j = 1, ..., r.$$

Gemäß der Darstellung in [20] wird $f_r \in P_r$ in jedem Teilintervall $I_{j,r} := [\psi_{j,r}, \psi_{j+1,r}]$, $j = 1, ..., r-1$, definiert durch

$$f_r(\psi) = p_{j,r} H_1(\psi) + p_{j+1,r} H_2(\psi) + d_{j,r} H_3(\psi) + d_{j+1,r} H_4(\psi). \tag{4.10}$$

Hierbei sind

$$\begin{aligned}
H_1(\psi) &= 3\left(\frac{\psi_{j+1,r} - \psi}{\psi_{j+1,r} - \psi_{j,r}}\right)^2 - 2\left(\frac{\psi_{j+1,r} - \psi}{\psi_{j+1,r} - \psi_{j,r}}\right)^3 \\
H_2(\psi) &= 3\left(\frac{\psi - \psi_{j,r}}{\psi_{j+1,r} - \psi_{j,r}}\right)^2 - 2\left(\frac{\psi - \psi_{j,r}}{\psi_{j+1,r} - \psi_{j,r}}\right)^3 \\
H_3(\psi) &= \frac{(\psi_{j+1,r} - \psi)^2}{\psi_{j+1,r} - \psi_{j,r}} - \frac{(\psi_{j+1,r} - \psi)^3}{(\psi_{j+1,r} - \psi_{j,r})^2} \\
H_4(\psi) &= \frac{(\psi - \psi_{j,r})^3}{(\psi_{j+1,r} - \psi_{j,r})^2} - \frac{(\psi - \psi_{j,r})^2}{\psi_{j+1,r} - \psi_{j,r}}
\end{aligned}$$

die üblichen kubischen Hermite-Basisfunktionen. Die Koeffizienten $d_{j,r}$ in (4.10) bestimmen die Ableitungen von f_r in den Stützstellen:

$$f_r'(\psi_{j,r}) = d_{j,r}, \quad j = 1, ..., r.$$

Die Werte von $d_{j,r}$, $j = 1, ..., r$, sind in Abhängigkeit vom Parametervektor p_r so festzulegen, dass f_r auf jedem Teilintervall $I_{j,r}$ monoton ist. Zusammen mit den Bedingungen (4.8) folgt dann die Monotonie von f_r auf $[\underline{\psi}, \overline{\psi}]$.

Mit den Notationen

$$\Delta_{j,r} := \frac{p_{j+1,r} - p_{j,r}}{\psi_{j+1,r} - \psi_{j,r}}, \ \alpha_{j,r} := \frac{d_{j,r}}{\Delta_{j,r}} \text{ und } \beta_{j,r} := \frac{d_{j+1,r}}{\Delta_{j,r}}$$

gelten folgende Aussagen:

Lemma 4.3

1. *Falls $\Delta_{j,r} = 0$, dann ist $f_r(\psi)$ monoton auf $I_{j,r}$ genau dann, wenn $d_{j,r} = d_{j+1,r} = 0$ gilt.*

2. *Wenn die Bedingungen*

$$\left.\begin{array}{c} \operatorname{sign}(d_{j,r}) = \operatorname{sign}(d_{j+1,r}) = \operatorname{sign}(\Delta_{j,r}) \\ \alpha_{j,r}^2 + \beta_{j,r}^2 - 9 \leq 0 \end{array}\right\} \tag{4.11}$$

erfüllt sind, dann ist $f_r(\psi)$ auf $I_{j,r}$ monoton.

Beweis:

1. Siehe [20].

2. Falls $\alpha_{j,r} + \beta_{j,r} - 2 \leq 0$, so folgt die Behauptung aus Lemma 1 in [20]. Wenn $\alpha_{j,r} + \beta_{j,r} - 2 > 0$, so liefert Lemma 2 in [20] die Behauptung.

□

Im Weiteren sei vorausgesetzt, dass die Bedingungen (4.8) erfüllt sind. Dann ist $\Delta_{j,r} \geq 0$ und aus der zweiten Aussage von Lemma 4.3 folgt:

Folgerung 4.4 *Wenn $d^*_{j,r}$ und $d^*_{j+1,r}$ die Bedingungen (4.11) erfüllen, dann ist die Funktion $f_r(\psi)$ für alle $d_{j,r} \in [0, d^*_{j,r}]$ und alle $d_{j+1,r} \in [0, d^*_{j+1,r}]$ auf $I_{j,r}$ monoton wachsend.*

Die Werte von $d_{j,r}$, $j = 1, ..., r$ werden nun nach folgender Vorschrift gebildet.

Algorithmus 4.5 (Monotone kubische Interpolation)

(i) Setze

$$\begin{aligned} d_{1,r} &= \frac{p_{2,r} - p_{1,r}}{\psi_{2,r} - \psi_{1,r}}, \\ d_{j,r} &= \frac{1}{2}\left(\frac{p_{j+1,r} - p_{j,r}}{\psi_{j+1,r} - \psi_{j,r}} + \frac{p_j - p_{j-1,r}}{\psi_{j,r} - \psi_{j-1,r}}\right) \text{ für } j = 2, ..., r-1, \\ d_{r,r} &= \frac{p_{r,r} - p_{r-1,r}}{\psi_{r,r} - \psi_{r-1,r}}. \end{aligned}$$

(ii) Für $j = 1, ..., r-1$

Falls $\Delta_{j,r} = 0$, so setze $d_{j,r} = d_{j+1,r} = 0$.

(iii) Für $j = 1, ..., r-1$

Falls $\alpha^2_{j,r} + \beta^2_{j,r} - 9 > 0$, so modifiziere $d_{j,r}$, $d_{j+1,r}$ zu $d^*_{j,r} = \alpha^*_{j,r}\Delta_{j,r}$ und $d^*_{j+1,r} = \beta^*_{j,r}\Delta_j$, wobei $\alpha^*_{j,r} = \tau_{j,r}\alpha_{j,r}$, $\beta^*_{j,r} = \tau_{j,r}\beta_{j,r}$ und $\tau_{j,r} = 3\left(\alpha^2_{j,r} + \beta^2_{j,r}\right)^{-\frac{1}{2}}$.

Bemerkung 4.6 Der Punkt $(\alpha^*_{j,r}, \beta^*_{j,r})$ ist so konstruiert, dass er dem Schnittpunkt der Geraden, die den Ursprung mit $(\alpha_{j,r}, \beta_{j,r})$ verbindet, und der Kreislinie um den Ursprung mit dem Radius 3 entspricht.

Wegen

$$\beta_{j-1,r}\Delta_{j-1,r} = d_{j,r} = \alpha_{j,r}\Delta_{j,r}$$

sind bei der Modifizierung im Schritt (iii) des obigen Algorithmus die Interaktionen zwischen benachbarten Teilintervallen zu beachten. Wenn $d_{j,r}$ auf $I_{j,r}$ modifiziert wird, dann wird $\beta_{j-1,r}$ ebenfalls modifiziert. Folgerung 4.4 gewährleistet aber, dass die Monotonie auf dem Intervall $I_{j-1,r}$ erhalten bleibt.

Bemerkung 4.7 Die Werte $d_{j,r}$ für $j = 1, ..., r$ sind keine Freiheitsgrade, die zu bestimmen sind, sondern in Abhängigkeit von den Werten $p_{j,r}$, $j = 1, ..., r$ durch die Berechnungsvorschrift im Algorithmus 4.5 eindeutig festgelegt.

4.1.4 Multi-Level-Algorithmus

Durch die oben diskutierten Parametrisierungsmethoden sind wir zu einem endlichdimensionalen Problem gelangt, welches darin besteht, Parametervektoren $p^{\nu}_{\varepsilon,h,r_\nu} \in P^{\nu}_{r_\nu}, \nu = 1, ..., M$ (M = Anzahl der unbekannten Koeffizientenfunktionen) zu bestimmen, die das Optimierungsproblem

$$\mathcal{J}_{\varepsilon,h}\left(p^1_{\varepsilon,h,r_1}, ..., p^M_{\varepsilon,h,r_M}\right) = \min_{\substack{p^{\nu}_{r_\nu} \in P^{\nu}_{r_\nu} \\ \nu=1,...,M}} \mathcal{J}_{\varepsilon,h}\left(p^1_{r_1}, ..., p^M_{r_M}\right) \tag{4.12}$$

mit $\mathcal{J}_{\varepsilon,h} = \tilde{\mathcal{J}}_{\varepsilon,h} \circ \mathcal{B}_h \circ \left(\mathcal{A}_h(p^1_{r_1}, ..., p^M_{r_M}), p^1_{r_1}, ..., p^M_{r_M}\right)$ lösen. Hierbei werden die Parameterräume $P^{\nu}_{r_\nu}$ durch die Monotoniebedingungen (4.8) oder (4.9) eingeschränkt, sodass es sich bei (4.12) um ein restringiertes Minimierungsproblem handelt. Dabei muss nicht notwendig für jede unbekannte Koeffizientenfunktion die gleiche Art der Parametrisierung verwendet werden. Die Gesamtzahl der Freiheitsgrade beträgt $\hat{r} = \sum_{\nu=1}^{M} r_\nu$.

Die grundlegenden Probleme, die bei den in Unterabschnitt 4.1.2 beschriebenen Parametrisierungen auftreten, sind zum einen die geeignete Festlegung des Grundintervalls $[\underline{\psi}, \overline{\psi}]$ und zum anderen die Frage nach der richtigen Anzahl der Freiheitsgrade, d. h. der Dimension r_ν der diskreten Parameterräume $P^{\nu}_{r_\nu}$. Das Intervall $[\underline{\psi}, \overline{\psi}]$ wird meist durch das Experiment selbst festgelegt, da die Koeffizientenfunktionen nur über ein beschränktes Intervall, welches durch die Anfangs- und Randbedingungen bestimmt ist, identifizierbar sind (siehe Kapitel 3). Schwieriger ist es eine geeignete Anzahl von Freiheitsgraden zu finden. Denn mit wachsender Anzahl von Freiheitsgraden steigt die Komplexität des Optimierungsproblems und die Flexibilität der Koeffizientenfunktionen nimmt zu. Eine zu geringe Flexibilität kann u. U. eine unzureichende Reproduktion der Messdaten zur Folge haben. Eine zu hohe Flexibilität führt häufig zu Effekten in den identifizierten Funktionen, die durch

Messfehler verursacht werden. Wie wir später in Beispielen sehen werden, zeigen die Identifizierungsfehler das für inverse Probleme typische Verhalten. Sowohl zu wenige als auch zu viele Freiheitsgrade bewirken größere Identifizierungsfehler. Das Optimum (d. h. der minimale Identifizierungsfehler) liegt irgendwo dazwischen. Das Optimierungsproblem ist außerdem umso schlechter konditioniert, je höher die Anzahl der Freiheitsgrade ist. Eine wachsende Anzahl von Freiheitsgraden wird die Konvergenzgeschwindigkeit des Optimierungsverfahrens verringern. Es besteht auch die Gefahr eines Abbruchs in einem Nebenminimum.

In Anbetracht dieser Probleme ist es wünschenswert möglichst gute Startnäherungen zu besitzen. Deshalb beginnen wir damit, das Problem (4.12) für die kleinst möglichen r_ν zu lösen. Anschließend berechnen wir für die optimalen Parametervektoren $p^\nu_{\varepsilon,h,r_\nu} \in P^\nu_{r_\nu}$ die zugehörigen Parameterdarstellungen $\tilde{p}^\nu_{\varepsilon,h,r_\nu+\Delta r_\nu} \in P^\nu_{r_\nu+\Delta r_\nu}$ und verwenden diese als Startwerte für die Optimierung in den Räumen $P^\nu_{r_\nu+\Delta r_\nu}$. Dabei erhalten wir $\tilde{p}^\nu_{\varepsilon,h,r_\nu+\Delta r_\nu}$ durch das Lösen einer entsprechenden Interpolationsaufgabe (siehe Unterabschnitt 4.1.2). Bei der Definition der Räume $P^\nu_{r_\nu+\Delta r_\nu}$ ist zu beachten, dass die Bedingung (4.4) erfüllt wird. Wenn diese Vorgehensweise sukzessive fortgesetzt wird, so gelangen wir zu einem *Multi-Level-Algorithmus*, wie er auch in [30] beschrieben ist.

Algorithmus 4.8 (Multi-Level-Algorithmus)
Wähle Startwerte $p^\nu_{\varepsilon,h,r_{\nu,\text{start}}-\Delta r_\nu}, \nu = 1, ..., M.$
$r_\nu := r_{\nu,\text{start}}$ *für* $\nu = 1, ..., M.$
DO

> *Berechne die Parameterdarstellung* $\tilde{p}^\nu_{\varepsilon,h,r_\nu}$ *von* $p^\nu_{\varepsilon,h,r_\nu-\Delta r_\nu}$ *im Raum* $P^\nu_{r_\nu}$ *für* $\nu = 1, ..., M.$
> *Löse das Optimierungsproblem (4.12) mit den Startwerten* $\tilde{p}^\nu_{\varepsilon,h,r_\nu}, \nu = 1, ..., M.$
> *Speichere die Ergebnisse in* $p^\nu_{\varepsilon\, h,r_\nu}, \nu = 1, ..., M$ *ab.*
> $r_\nu := r_\nu + \Delta r_\nu$ *für* $\nu = 1, ..., M.$

WHILE $\varepsilon \cdot \mu_{\hat{r}} \leq \mu_{\text{tol}}$ *und* $\max_{\nu=1,...,M} r_\nu \leq r_{\max}$

Dabei sind Δr_ν und r_{max} geeignet zu wählen. Δr ist bezüglich der Stufe des Multi-Level-Algorithmus variabel wählbar. $\mu_{\hat{r}}$ bezeichnet einen noch zu definierende Fehlerindikator und μ_{tol} eine Toleranzgrenze.

Bei Spline-Parametrisierungen mit hierarchischen Basen wird die Schrittweite $\Delta s = 1$ ($r = 2^s + r_0 - 1$) gewählt und obiger Algorithmus stellt eine Mehrskalenoptimierung dar. Ein Vorteil der hierarchischen Basen ist, dass beim Übergang zur nächsthöheren Skala nur die neuen Parameterwerte mit

0 initialisiert und die alten Parameterwerte beibehalten werden. Es ist also im Unterschied zu lokalen Basen keine eigentliche Interpolationsaufgabe zu lösen.

Ein einfaches Abbruchkriterium für den Multi-Level-Algorithmus ist durch das z. B. in [31] oder [49] dargestellte Diekrepanzkriterium gegeben, welches in unserer Notation lautet: Die „optimale “ Anzahl von Freiheitsgraden ist erreicht, sobald gilt

$$\mathcal{J}_{\varepsilon,h}\left(p^1_{\varepsilon,h,r_1},...,p^M_{\varepsilon,h,r_M}\right) \leq \varepsilon^2 < \mathcal{J}_{\varepsilon,h}\left(p^1_{\varepsilon,h,r'_1},...,p^M_{\varepsilon,h,r'_M}\right) \tag{4.13}$$

für $r'_\nu < r_\nu,\ \nu = 1,...,M$, wobei ε hier den Datenfehler in der zugehörigen gewichteten L^2-Norm angibt.

Aus der Bedingung (4.4) an die Parameterräume $P^\nu_{r_\nu}$ folgt für die Optimalwerte des Fehlerfunktionals

$$0 \leq \mathcal{J}_{\varepsilon,h}\left(p^1_{\varepsilon,h,r'_1},...,p^M_{\varepsilon,h,r'_M}\right) \leq \mathcal{J}_{\varepsilon,h}\left(p^1_{\varepsilon,h,r_1},...,p^M_{\varepsilon,h,r_M}\right)$$

für $r_\nu \leq r'_\nu,\ \nu = 1,...,M$. Dabei wird im Allg. der Fehlerreduktionsfaktor

$$\frac{\mathcal{J}_{\varepsilon,h}\left(p^1_{\varepsilon,h,r_1-\Delta r_1},...,p^M_{\varepsilon,h,r_M-\Delta r_M}\right)}{\mathcal{J}_{\varepsilon,h}\left(p^1_{\varepsilon,h,r_1},...,p^M_{\varepsilon,h,r_M}\right)} \geq 1$$

im Multi-Level-Algorithmus von Schritt zu Schritt kleiner werden und gegen 1 gehen. Daher können wir in Abwandlung des Diskrepanzkriteriums (4.13) den Multi-Level Algorithmus auch solange fortsetzen, wie mit $\mu_{\text{tol}} > 0$ gilt

$$\begin{aligned}\mathcal{J}_{\varepsilon,h}\left(p^1_{\varepsilon,h,r_1-\Delta r_1},...,p^M_{\varepsilon,h,r_M-\Delta r_M}\right) &- J_{\varepsilon,h}\left(p^1_{\varepsilon,h,r_1},...,p^M_{\varepsilon,h,r_M}\right)\\ &< \varepsilon^2\mu_{\text{tol}}\mathcal{J}_{\varepsilon,h}\left(p^1_{\varepsilon,h,r_1-\Delta r_1},...,p^M_{\varepsilon,h,r_M-\Delta r_M}\right)\end{aligned}$$

bzw.

$$1 - \frac{\mathcal{J}_{\varepsilon,h}\left(p^1_{\varepsilon,h,r_1},...,p^M_{\varepsilon,h,r_M}\right)}{\mathcal{J}_{\varepsilon,h}\left(p^1_{\varepsilon,h,r_1-\Delta r_1},...,p^M_{\varepsilon,h,r_M-\Delta r_M}\right)} < \varepsilon^2\mu_{\text{tol}}\,.$$

D. h. wir beenden den Multi-Level-Algorithmus dann, wenn der Fehlerreduktionsfaktor einen vorgegebenen Wert unterschreitet. Falls ε den zugehörigen relativen Datenfehler bezeichnet, so kann z. B. $\mu_{\text{tol}} = 1$ gewählt werden.

4.1.5 Sensitivitätsanalyse

Aufgrund der Datenfehler ist ein fallender Optimalwert

$$\mathcal{J}_{\varepsilon,h}\left(p^1_{\varepsilon,h,r_1},...,p^M_{\varepsilon,h,r_M}\right)$$

nicht notwendig gleichbedeutend mit fallenden Identifizierungsfehlern (vgl. Unterabschnitt 2.1.2). Deshalb führen wir eine Sensitivitätsanalyse durch (vgl. hierzu auch [7]). Dazu benötigen wir die folgende Definition:

Definition 4.9 *Die Matrix*

$$\frac{d\left(\mathcal{B}_h \circ (\mathcal{A}_h(p_{\varepsilon,h,\hat{r}}), p_{\varepsilon,h,\hat{r}})\right)}{dp_{\hat{r}}} =: \mathcal{S}_{h,\hat{r}} \in \mathbb{R}^{\hat{\kappa}\times\hat{r}}$$

($\hat{\kappa} = \sum_{k=1}^{\kappa} n_k$) wird Sensitivitätsmatrix *von $\mathcal{B}_h(\mathcal{A}_h(p_{\hat{r}}), p_{\hat{r}})$ in $p_{\varepsilon,h,\hat{r}}$ genannt.*

Im Folgenden setzen wir voraus, dass $\hat{\kappa} \geq \hat{r}$ gilt und die Matrix $\mathcal{S}_{h,\hat{r}}$ den Rang $\hat{r}$ besitzt. Des Weiteren werden wir die Monotonierestriktionen für die Parameterräume vernachlässigen.

Ausgehend von der Optimallösung $p_{\varepsilon,h,\hat{r}} = \left((p^1_{\varepsilon,h,r_1})^T, ..., (p^M_{\varepsilon,h,r_M})^T\right)^T \in \mathbb{R}^{\hat{r}}$ zum Beobachtungsvektor $\omega_{\varepsilon,h} = \left(\omega^T_{\varepsilon,1,h}, ..., \omega^T_{\varepsilon,\kappa,h}\right)^T \in \mathbb{R}^{\hat{\kappa}}$ wollen wir untersuchen, wie sich kleine Änderungen $\delta\omega_{\varepsilon,h}$ in diesem Beobachtungsvektor auf die zu bestimmenden Parameter auswirken, d. h. wir suchen den zugehörigen Korrekturvektor $\delta p_{\hat{r}} \in \mathbb{R}^{\hat{r}}$ für $p_{\varepsilon,h,\hat{r}}$. Dieser bestimmt sich aus dem folgenden Problem:

$$\min_{\delta p_{\hat{r}} \in \mathbb{R}^{\hat{r}}} \sum_{k=1}^{\kappa} \sum_{i=1}^{n_k} \alpha^i_k \left(\omega^i_k(p_{\varepsilon,h,\hat{r}} + \delta p_{\hat{r}}) - (\omega^i_{\varepsilon,k} + \delta\omega^i_{\varepsilon,k})\right)^2 . \tag{4.14}$$

Auf das Minimierungsproblem (4.14) wenden wir einen Schritt des Gauß-Newton-Verfahrens an (siehe z. B. [56], S. 195). Nach der Linearisierung erhalten wir zunächst

$$\min_{\delta p_{\hat{r}} \in \mathbb{R}^{\hat{r}}} \sum_{k=1}^{\kappa} \sum_{i=1}^{n_k} \alpha^i_k \left(\omega^i_k(p_{\varepsilon,h,\hat{r}}) + \frac{d\omega^i_k(p_{\varepsilon,h,\hat{r}})}{dp_{\hat{r}}}\delta p_{\hat{r}} - (\omega^i_{\varepsilon,k} + \delta\omega^i_{\varepsilon,k})\right)^2 . \tag{4.15}$$

Dieses lineare Minimierungsproblem ist äquivalent mit

$$\begin{aligned} \mathcal{S}^T_{h,\hat{r}}\mathcal{S}_{h,\hat{r}}\delta p_{\hat{r}} &= \mathcal{S}^T_{h,\hat{r}}\left(\omega_{\varepsilon,h} - \omega_h(p_{\varepsilon,h,\hat{r}}) + \delta\omega_{\varepsilon,h}\right) \\ &= \mathcal{S}^T_{h,\hat{r}}\left(\omega_{\varepsilon,h} - \omega_h(p_{\varepsilon,h,\hat{r}})\right) + \mathcal{S}^T_{h,\hat{r}}(\delta\omega_{\varepsilon,h}) . \end{aligned}$$

Da $p_{\varepsilon,h,\hat{r}}$ optimal für $\omega_{\varepsilon,h}$ ist, gilt (bei Vernachlässigung der Parameterrestriktionen)

$$\mathcal{S}^T_{h,\hat{r}}\left(\omega_{\varepsilon,h} - \omega_h(p_{\varepsilon,h,\hat{r}})\right) = 0$$

und somit

$$\mathcal{S}^T_{h,\hat{r}}\mathcal{S}_{h,\hat{r}}\delta p_{\hat{r}} = \mathcal{S}^T_{h,\hat{r}}(\delta\omega_{\varepsilon,h}) . \tag{4.16}$$

Die Gleichung (4.16) entspricht schließlich dem linearen Minimierungsproblem:

$$\min_{\delta p_{\hat{r}} \in \mathbb{R}^{\hat{r}}} \sum_{k=1}^{\kappa} \sum_{i=1}^{n_k} \alpha^i_k \left(\frac{d\omega^i_k(p_{\varepsilon,h,\hat{r}})}{dp_{\hat{r}}}\delta p_{\hat{r}} - \delta\omega^i_{\varepsilon,k}\right)^2 . \tag{4.17}$$

(4.17) kann betrachtet werden als verallgemeinerter Ansatz zur Lösung des folgenden linearen Gleichungssystems im Parameterraum:

$$\frac{d\left(\mathcal{B}_h \circ \left(\mathcal{A}_h(p_{\varepsilon,h,\hat{r}}), p_{\varepsilon,h,\hat{r}}\right)\right)}{dp_{\hat{r}}} \delta p_{\hat{r}} = \delta\omega_{\varepsilon,h}. \tag{4.18}$$

Änderungen $\delta p_{\hat{r}}$ in den Parametern und Änderungen $\delta\omega_{\varepsilon,h}$ in den Beobachtungen sind also durch die Sensitivitätsmatrix $\mathcal{S}_{h,\hat{r}}$ gekoppelt. Mithilfe der verallgemeinerten Inversen der Sensitivitätsmatrix ist die Gleichung (4.18) auflösbar:

$$\delta p_{\hat{r}} = \mathcal{S}^{\dagger}_{h,\hat{r}} \delta\omega_{\varepsilon,h}.$$

Wenn wir eine Singulärwertzerlegung der Sensitivitätsmatrix durchführen, so erhalten wir

$$\delta p_{\hat{r}} = \sum_{j=1}^{\hat{r}} \sigma^{-1}_{j,\hat{r}} \left\langle \delta\omega_{\varepsilon,h}, u_{j,\hat{r}} \right\rangle v_{j,\hat{r}}. \tag{4.19}$$

Dabei bezeichnet $\{\sigma_{j,\hat{r}}, u_{j,\hat{r}}, v_{j,\hat{r}}\}_{j=1,\ldots,\hat{r}}$ ein singuläres System der Sensitivitätsmatrix $\mathcal{S}_{h,\hat{r}}$ (siehe z. B. [39], S. 147 ff.). Wenn die Matrix $\mathcal{S}_{h,\hat{r}}$ nicht den Rang $\hat{r}$ besitzt, dann ist in (4.19) und der folgenden Definition 4.10 $\hat{r}$ entsprechend durch den Rang zu ersetzen.

Als Folgerung aus der Gleichung (4.19) ergibt sich, dass Fehler in den Daten in Abhängigkeit von den Singulärwerten $\sigma_{j,\hat{r}}$ auf die Parameter übertragen werden. Kleine Werte von $\sigma_{j,\hat{r}}$ führen zu einer großen Fehlerverstärkung. Das Wachstumsverhalten dieser Singulärwerte dient häufig auch zur Klassifizierung inverser Probleme (siehe z. B. [39] für lineare Probleme). Damit können wir im Multi-Level-Algorithmus Fehlerindikatoren $\mu_{\hat{r}}$ verwenden, die aus den Singulärwerten gewonnen werden.

Definition 4.10

$$\mu_{\text{cond}}\left(\mathcal{S}_{h,\hat{r}}\right) := \max_{j,k=1,\ldots,\hat{r}} \frac{\sigma_{j,\hat{r}}}{\sigma_{k,\hat{r}}}$$

heißt die Spektralkondition *von* $\mathcal{B}_h(\mathcal{A}_h(p_{\hat{r}}), p_{\hat{r}})$ *in* $p_{\varepsilon,h,\hat{r}}$.

$$\mu_{\max}\left(\mathcal{S}_{h,\hat{r}}\right) := \max_{j=1,\ldots,\hat{r}} \frac{1}{\sigma_{j,\hat{r}}}$$

wird Maximum-Charakteristik *von* $\mathcal{B}_h(\mathcal{A}_h(p_{\hat{r}}), p_{\hat{r}})$ *in* $p_{\varepsilon,h,\hat{r}}$ *genannt.*

Bei der Maximum-Charakteristik handelt es sich um dem maximalen Faktor der Fehlerverstärkung und bei der Spektralkondition um das maximale Verhältnis zwischen den Fehlerverstärkungsfaktoren. Zur Bestimmung dieser

Größen ist mithilfe des Algorithmus 3.34 in jedem Schritt des Multi-Level-Algorithmus die Sensitivitätsmatrix aufzustellen. Das Abbruchkriterium im Multi-Level-Algorithmus kann dahingehend interpretiert werden, dass das Verfahren selbst eine geeignete Flexibilität der Koeffizientenfunktionen bestimmt. Die Berechnung der Singulärwerte fällt aufgrund der geringen Größe von $\hat{r}$ nicht sonderlich ins Gewicht.

Bemerkung 4.11 Die obige Sensitivitätsanalyse ist auch partiell für einzelne Beobachtungen und Koeffizientenfunktionen durchführbar. Außerdem kann anstelle von $S_{h,\hat{r}}$ auch die mit den Faktoren $\sqrt{\alpha_k^i}$ gewichtete Sensitivitätsmatrix betrachtet werden. D. h., in (4.14) und damit auch in (4.17) werden beide Anteile der Differenz unter dem Quadrat mit der Wurzel des entsprechenden Wichtungsfaktors multipliziert.

4.2 Fallstudien

Zur numerischen Untersuchung des Multi-Level-Algorithmus simulieren wir ein Experiment mit den van Genuchten-Mualem-Funktionen (3.9) und (3.10) mit $\theta_{\text{sat}} = 0.52$, $\theta_{\text{res}} = 0.218$, $K_{\text{sat}} = 1.3167$ cm/h, $\alpha = 0.0115$ cm^{-1} und $n = 2.03$ in einer Raumdimension für eine Säule der Länge $L = 15.0$ cm ($\Omega = (0, L)$). Die verwendeten Parameter entsprechen einem Lehm und sind aus [60] entnommen. Am unteren Rand der Säule wird der Druck während des Zeitintervalls $[0, T]$ mit $T = 40$ h linear von 15.0 cm auf -200.0 cm abgesenkt. Zu den Zeitpunkten $\tilde{t}^i = (i-1)\frac{T}{n}$, $i = 1, ..., n$, finden am oberen Rand ($x = 0$, Flussrand) die Druckmessung und am unteren Rand ($x = L$, Dirichlet-Rand) die Messung des kumulativen Ausflusses statt. Diese exakten Beobachtungen stören wir gemäß Modell (4.1) mit einen $\varepsilon > 0$ und bezeichnen die gestörten Beobachtungen mit $\omega_{\varepsilon,\psi}$ (Druck) und $\omega_{\varepsilon,q}$ (kumulativer Ausfluss). Die Diskretisierung des Raumes bei der Erzeugung der simulierten Daten erfolgt mit $h = \frac{L}{3000}$ und in der Zeit verwenden wir die Schrittweite $\Delta t = 0.05$. Bei den inversen Rechnungen werden wir eine gröbere Diskretisierung verwenden.

Bei der Identifizierung der hydraulischen Funktionen Θ und K bilden wir das Funktional (4.2) aus der Druckmessung ω_ψ und dem kumulativen Ausfluss ω_q. Wenn die Retentionsfunktion Θ bekannt und nur die Leitfähigkeit K zu bestimmen ist, dann genügt es im Fehlerfunktional nur den kumulativen Ausfluss ω_q zu berücksichtigen.

Der kleinste Wert des Drucks, der bei diesem speziellen Beispiel auftreten kann, beträgt $\underline{\psi} = -215$ cm (siehe Beispiel 3.3). Zur Sicherheit geben wir noch 5 cm hinzu und parametrisieren die hydraulischen Funktionen über das Intervall $[-220, 0]$. Wir verwenden die in 4.1.2 und 4.1.3 beschriebenen Pa-

rametrisierungen und setzen dabei $\Theta(0) = \theta_{\text{sat}}$. Da die gesättigte Leitfähigkeit K_{sat} leicht durch direktes Messen bestimmt werden kann, geben wir zur Stabilisierung der Identifizierung außerdem $K(0) = K_{\text{sat}}$ vor. Zusätzlich wird der Parameterraum durch die entsprechenden Monotoniebedingungen restringiert.

Wenn wir die hydraulischen Funktionen mittels einer lokalen Basis diskretisieren, so legen wir eine äquidistante Zerlegung zugrunde. In jedem Verfeinerungsschritt wird dann jedes Teilintervall halbiert. Dies entspricht einem zu den hierarchischen Basen analogen skalenweisen Vorgehen.

Zur Minimierung glatter Fehlerfunktionale setzen wir die SQP-Methode (siehe [51]) ein.

4.2.1 Einfluss des Diskretisierungsparameters

Ein Vergleich der Sensitivitäten für die räumlichen Diskretisierungen $h = \frac{L}{50}$, $h = \frac{L}{100}$ und $h = \frac{L}{1000}$ (Abbildung 4.2) ergibt, dass mit zunehmender Anzahl von Freiheitsgraden die Werte von μ_{cond} und $\mu_{\max}$ zwar ansteigen, dies aber weitestgehend unabhängig vom Diskretisierungsparameter h geschieht. Die räumliche Diskretisierung hat anscheinend keinen entscheidenden Einfluss auf die Identifizierungsergebnisse, insofern nicht zu grob diskretisiert wird.

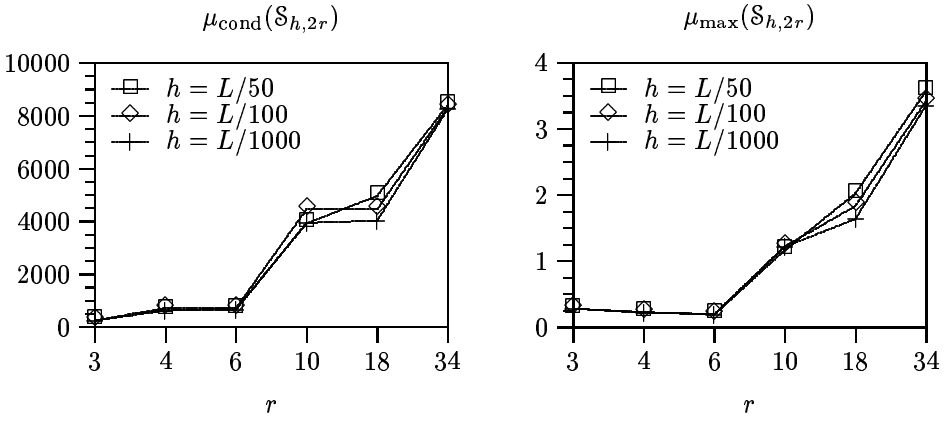


Abbildung 4.2: Sensitivitäten zu den räumlichen Diskretisierungen mit $h = \frac{L}{50}$, $h = \frac{L}{100}$ und $h = \frac{L}{1000}$ bei $n = 50$, $\varepsilon = 5\%$ und einer quadratischen lokalen Parametrisierung von Θ und K.

Die Zeitschrittweiten für die Simulation des direkten Problems bestimmen wir aus den Messzeitpunkten $\tilde{t}^i$, $i = 0, ..., n$ ($\tilde{t}^0 = t^0$). Dabei gehen wir so vor, dass ausgehend vom Zeitpunkt $\tilde{t}^i$ der nachfolgende Messzeitpunkt $\tilde{t}^{i+1}$ mit

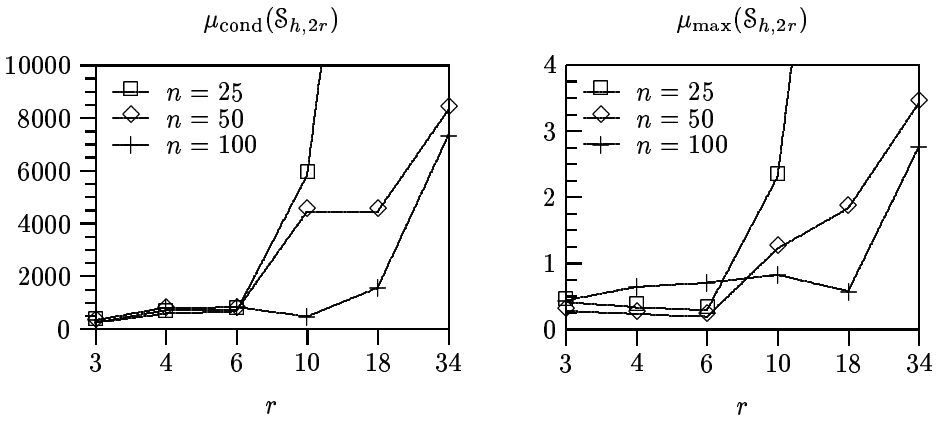


Abbildung 4.3: μ_{cond} und $\mu_{\max}$ für $n = 25$, 50 und 100 mit $h = \frac{L}{100}$ bei einem quadratischen lokalen Spline-Ansatz und $\varepsilon = 5\%$.

mindestens einem Zwischenschritt direkt angesteuert wird. Es gilt also

$$\max_{i=1,\ldots,m} \Delta t^i \leq \max_{i=1,\ldots,n} \frac{1}{2} \left(\tilde{t}^i - \tilde{t}^{i-1} \right).$$

Es existiert ein kritischer Wert von r, ab dem für μ_{cond} und $\mu_{\max}$ ein *blow-up*-Effekt eintritt. Dieser kritische Wert von r ist neben der Problemabhängigkeit auch von der Anzahl der betrachteten Messdaten abhängig. Dabei kommt es für kleine n früher zum blow-up als für größere n (Abb. 4.3).

4.2.2 Konvergenz des Multi-Level-Algoritmus und Abhängigkeit von der Art der Parametrisierung

Zur Untersuchung der Multi-Level-Identifizierung parametrisieren wir die Koeffizientenfunktionen Θ und K zunächst mithilfe von quadratische Splines bei Verwendung der lokalen Basis. Die nach dem van Genuchten-Mualem-Modell gewählten Funktionen Θ_{vGM} und K_{vGM} wurden zur Gewinnung synthetischer Messdaten mit einem Fehler von $\varepsilon = 5\%$ verwendet. Die Identifizierungsergebnisse, die der Multi-Level-Algorithmus für $n = 50$ (d. h. insgesamt 100 Datenpunkte) und $h = \frac{L}{100}$ liefert, sind in den Abbildungen 4.4–4.15 dargestellt. $\Theta_{\varepsilon,h,r}$ und $K_{\varepsilon,h,r}$ bezeichnen die identifizierte Retentionsfunktion bzw. die identifizierte Leitfähigkeit bei r Freiheitsgraden.

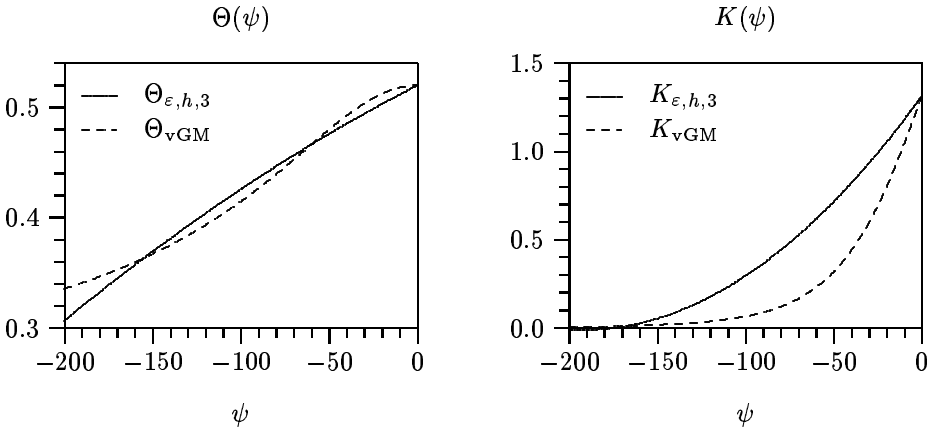


Abbildung 4.4: Hydraulische Funktionen, $r = 3$, $\varepsilon = 5\%$, $h = \frac{L}{100}$, $n = 50$.

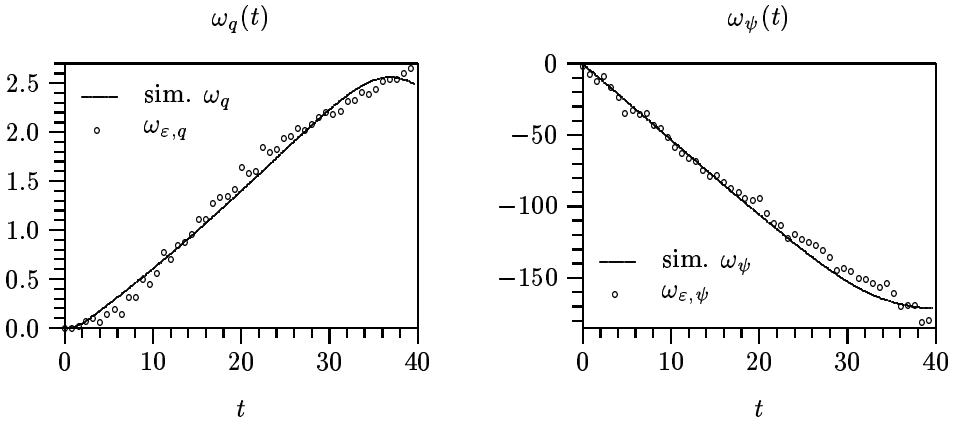


Abbildung 4.5: Beobachtungen zu Abbildung 4.4.

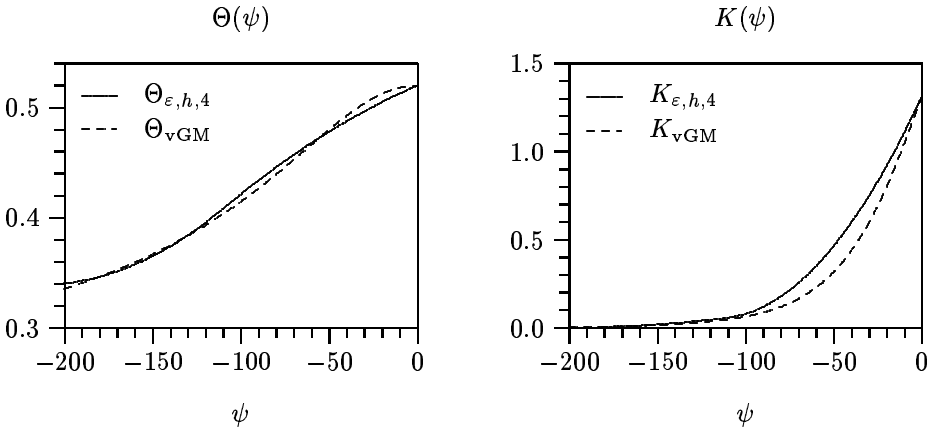


Abbildung 4.6: Hydraulische Funktionen, $r = 4$, $\varepsilon = 5\%$, $h = \frac{L}{100}$, $n = 50$.

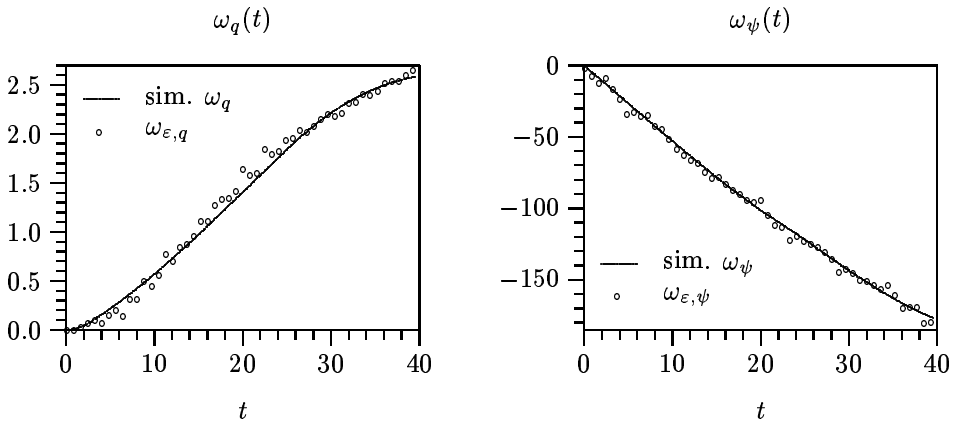


Abbildung 4.7: Beobachtungen zu Abbildung 4.6.

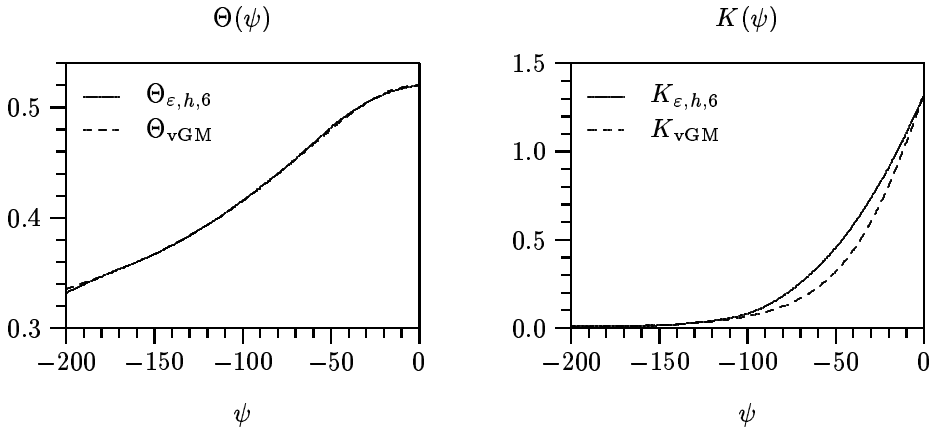


Abbildung 4.8: Hydraulische Funktionen, $r = 6$, $\varepsilon = 5\%$, $h = \frac{L}{100}$, $n = 50$.

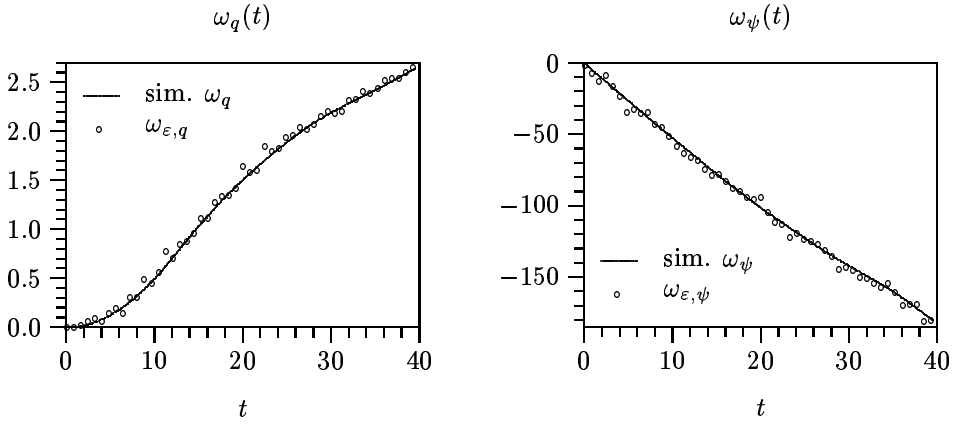


Abbildung 4.9: Beobachtungen zu Abbildung 4.8.

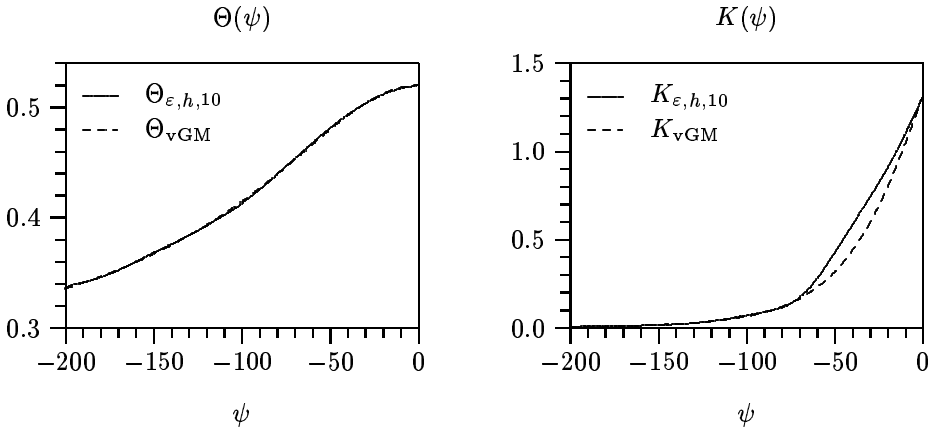


Abbildung 4.10: Hydraulische Funktionen, $r = 10$, $\varepsilon = 5\%$, $h = \frac{L}{100}$, $n = 50$.

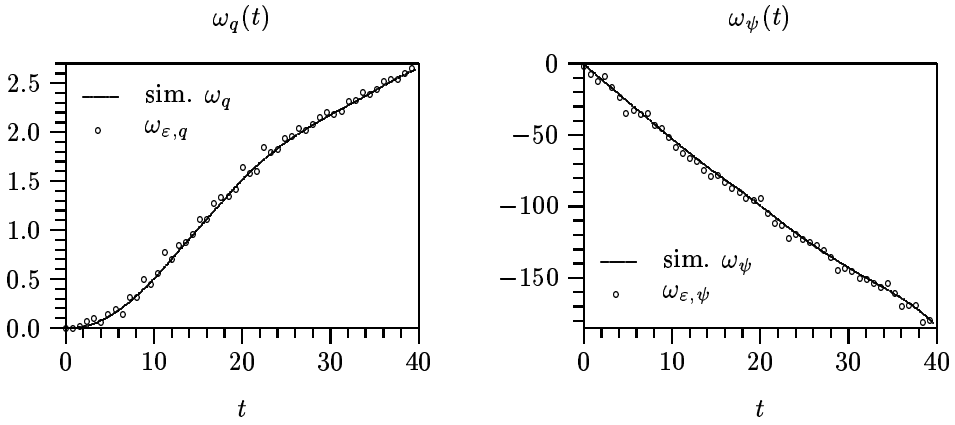


Abbildung 4.11: Beobachtungen zu Abbildung 4.10.

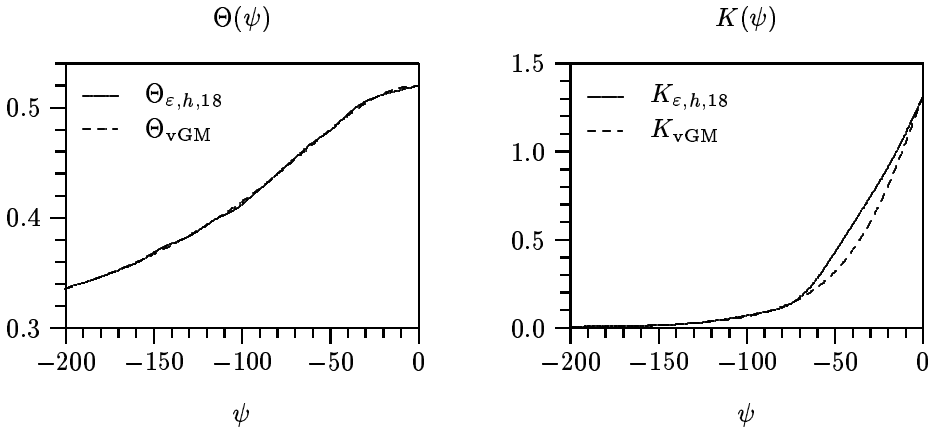


Abbildung 4.12: Hydraulische Funktionen, $r = 18$, $\varepsilon = 5\%$, $h = \frac{L}{100}$, $n = 50$.

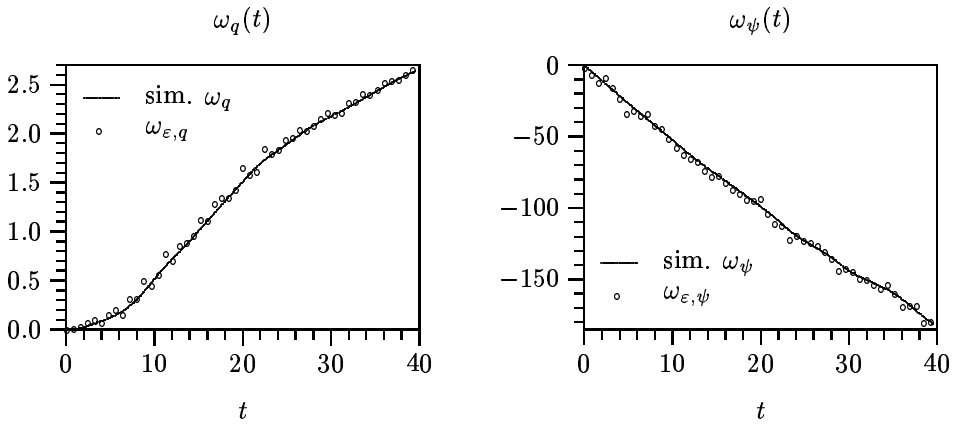


Abbildung 4.13: Beobachtungen zu Abbildung 4.12.

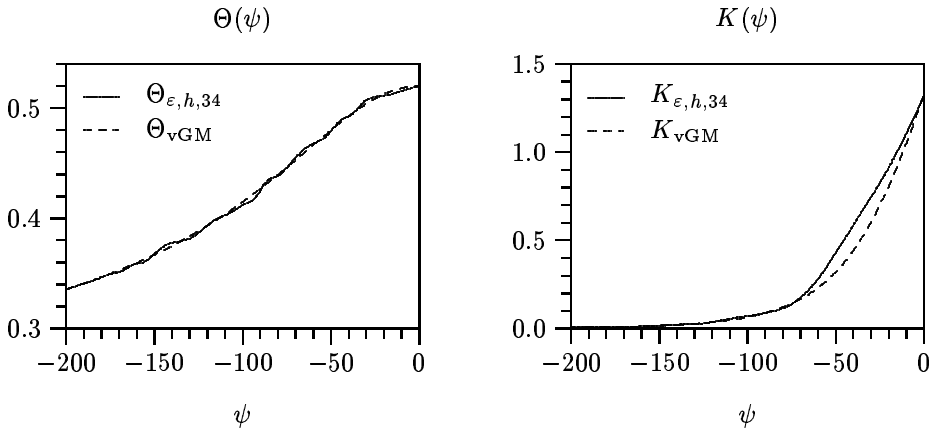


Abbildung 4.14: Hydraulische Funktionen, $r = 34$, $\varepsilon = 5\%$, $h = \frac{L}{100}$, $n = 50$.

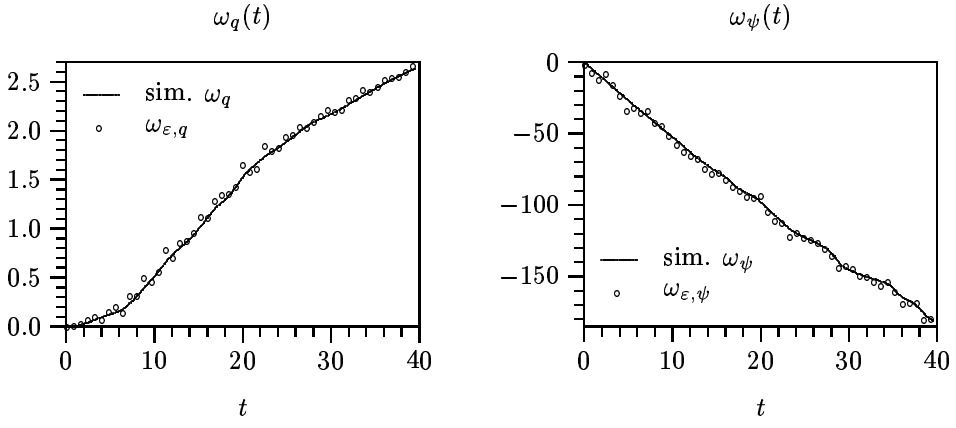


Abbildung 4.15: Beobachtungen zu Abbildung 4.14.

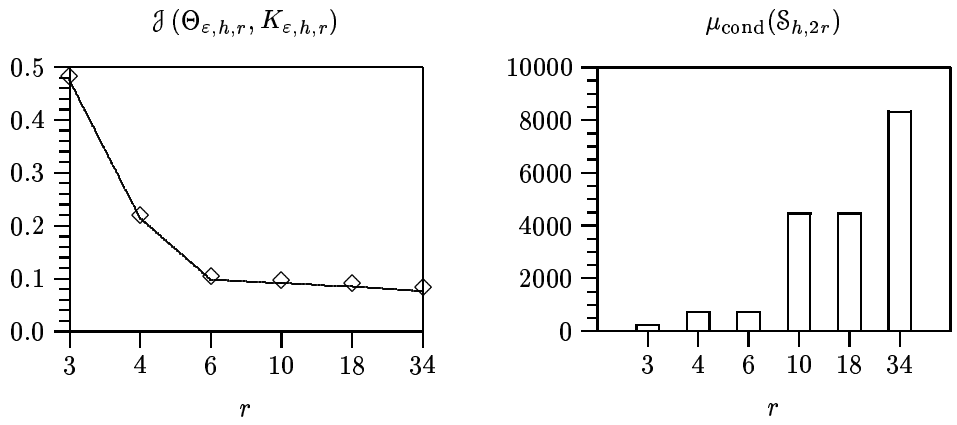


Abbildung 4.16: Fehlerfunktional und Spektralkondition, $\varepsilon = 5\%$, $h = \frac{L}{100}$, $n = 50$.

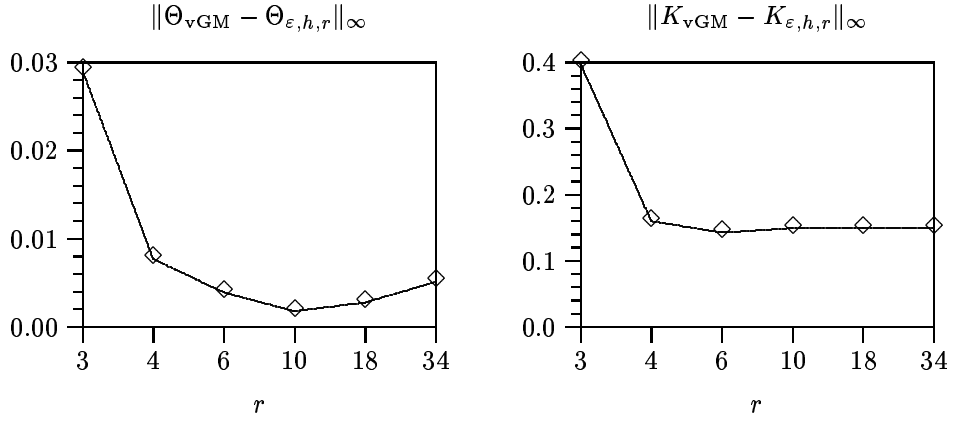


Abbildung 4.17: Identifizierungsfehler, $\varepsilon = 5\%$, $h = \frac{L}{100}$, $n = 50$.

Wie zu erwarten fällt das Fehlerfunktional (Abbildung 4.16) für wachsendes r, während die in der Maximumsnorm gemessenen Identifizierungsfehler $\|\Theta_{\text{vGM}} - \Theta_{\varepsilon,h,r}\|_\infty$ und $\|K_{\text{vGM}} - K_{\varepsilon,h,r}\|_\infty$ zunächst fallen, dann aber wieder ansteigen. Dabei bleibt die identifizierte Leitfähigkeit wesentlich stabiler als die identifizierte Retentionsfunktion (Abbildung 4.17). Bei der Funktion $\Theta_{\varepsilon,h,34}$ ist der Einfluss der Störung ε deutlich zu sehen. Die Werte von μ_{cond} und $\mu_{\max}$ wachsen mit zunehmender Anzahl von Freiheitsgraden (Abbildung 4.16).

Je genauer die Eingangsdaten der Identifizierung sind, desto besser sind die Identifizierungsergebnisse. In Abbildung 4.18 sind die Identifizierungsfehler für Störungen mit $\varepsilon = 2\%$ und $\varepsilon = 5\%$ abgebildet.

Die Ergebnisse sind auch davon abhängig, welche Art der Parametrisierung verwendet wird. So wird für eine geringe Anzahl von Freiheitsgraden

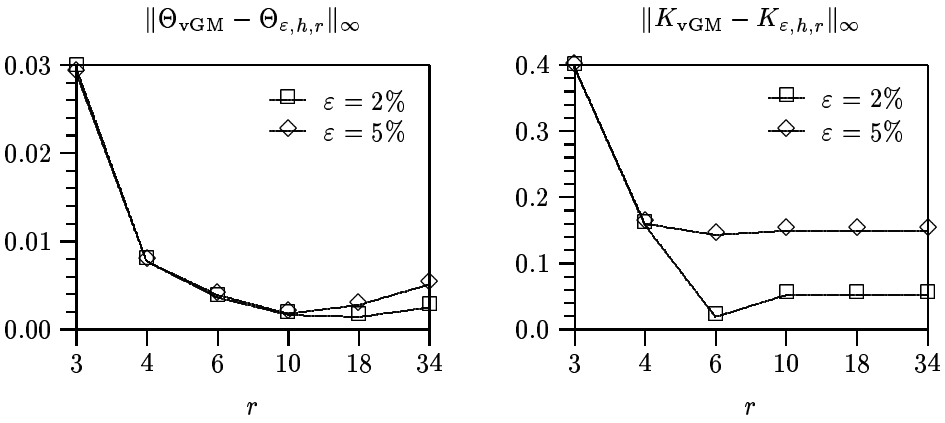


Abbildung 4.18: Identifizierungsfehler in Abhängigkeit vom Datenfehler, $h = \frac{L}{100}$, $n = 50$.

eine quadratische Parametrisierung einer linearen Parametrisierung im Allg. überlegen sein. Allerdings gewinnt der Datenfehler beim quadratischen Ansatz mit steigender Anzahl von Freiheitsgraden schnell an Einfluss. Die Parametrisierung mit hierarchischen Basen erweist sich besonders beim linearen Ansatz häufig als stabiler. Die Identifizierungsergebnisse für die Parametrisierung mit monotonen kubischen Ansatzfunktionen (MKP) bei $r = 3, 5, 9, 33$ sind in den Abbildungen 4.19–4.26 dargestellt.

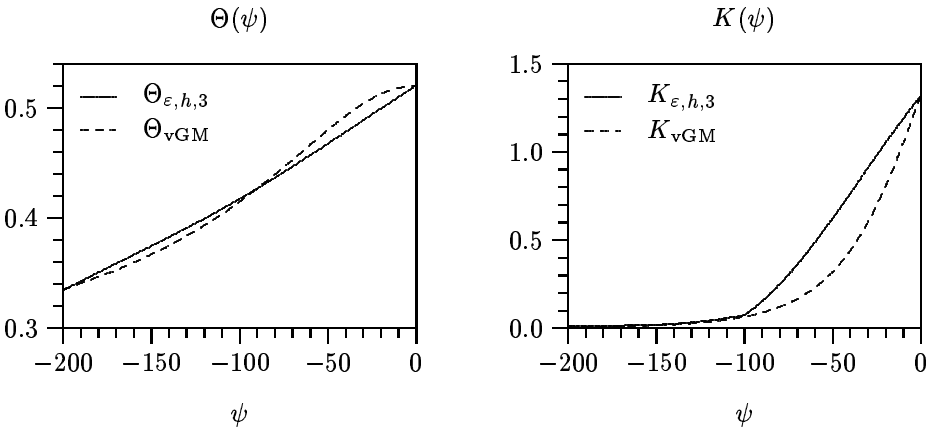


Abbildung 4.19: MKP, Hydraulische Funktionen, $r = 3$, $\varepsilon = 5\%$, $h = \frac{L}{100}$, $n = 50$.

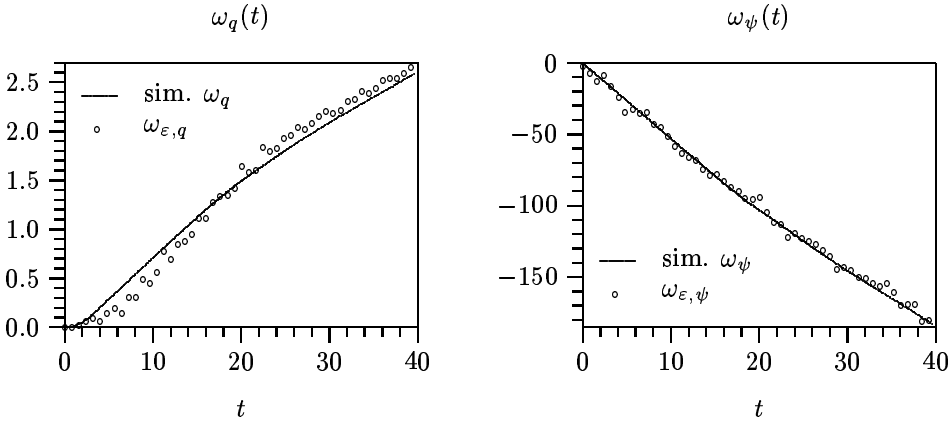


Abbildung 4.20: Beobachtungen zu Abbildung 4.19.

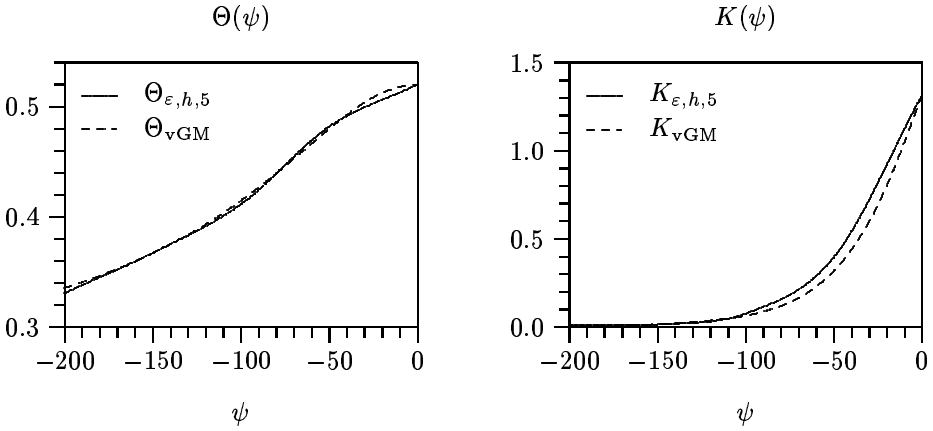


Abbildung 4.21: MKP, Hydraulische Funktionen, $r = 5$, $\varepsilon = 5\%$, $h = \frac{L}{100}$, $n = 50$.

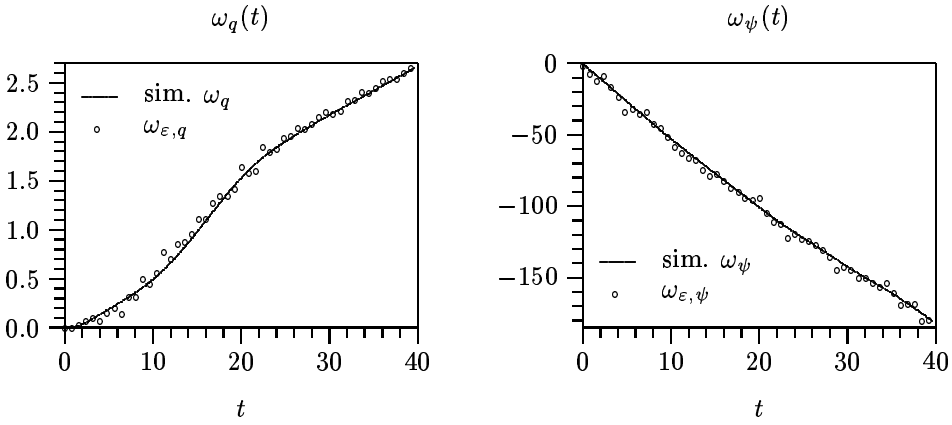


Abbildung 4.22: Beobachtungen zu Abbildung 4.21.

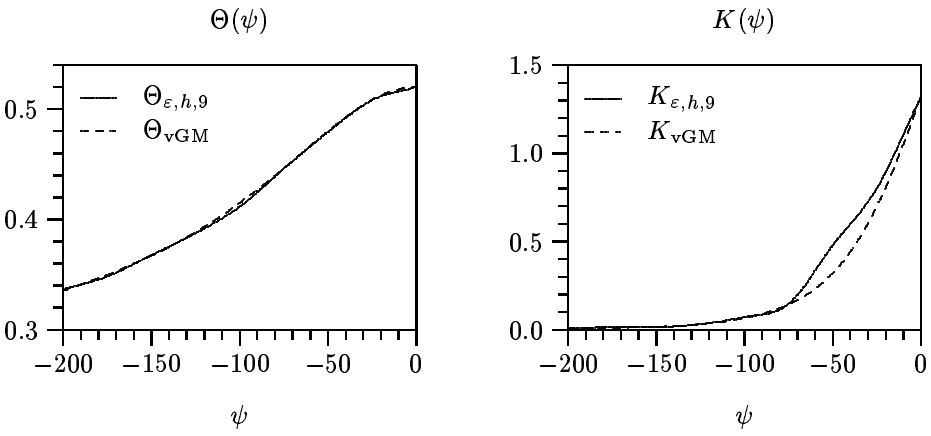


Abbildung 4.23: MKP, Hydraulische Funktionen, $r = 9$, $\varepsilon = 5\%$, $h = \frac{L}{100}$, $n = 50$.

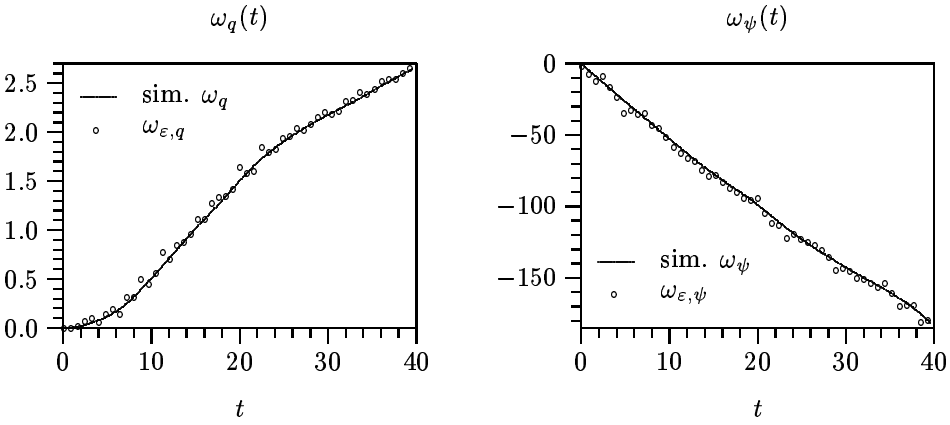


Abbildung 4.24: Beobachtungen zu Abbildung 4.23.

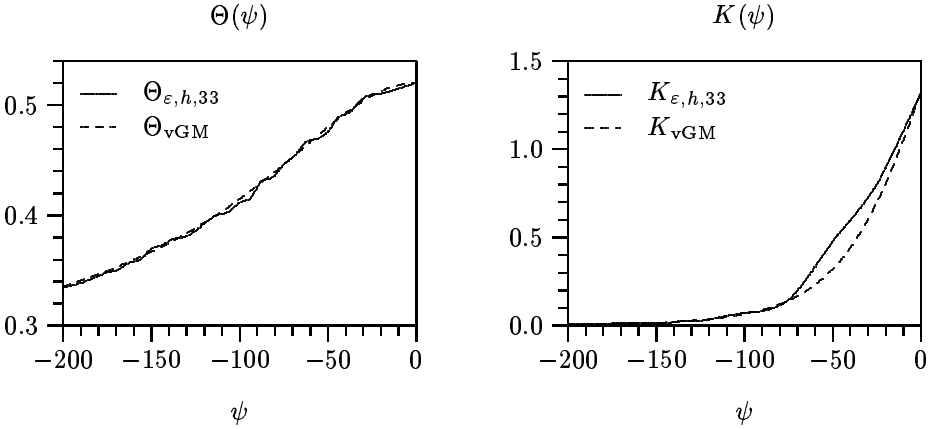


Abbildung 4.25: MKP, Hydraulische Funktionen, $r = 33$, $\varepsilon = 5\%$, $h = \frac{L}{100}$, $n = 50$.

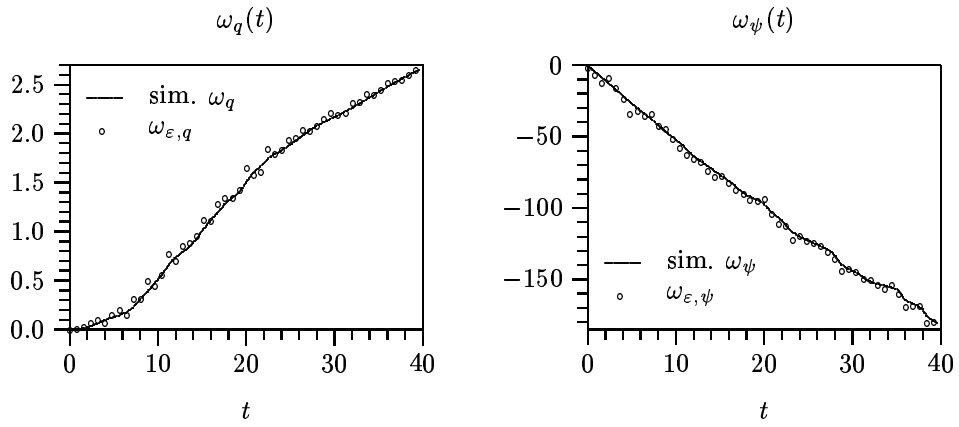


Abbildung 4.26: Beobachtungen zu Abbildung 4.25.

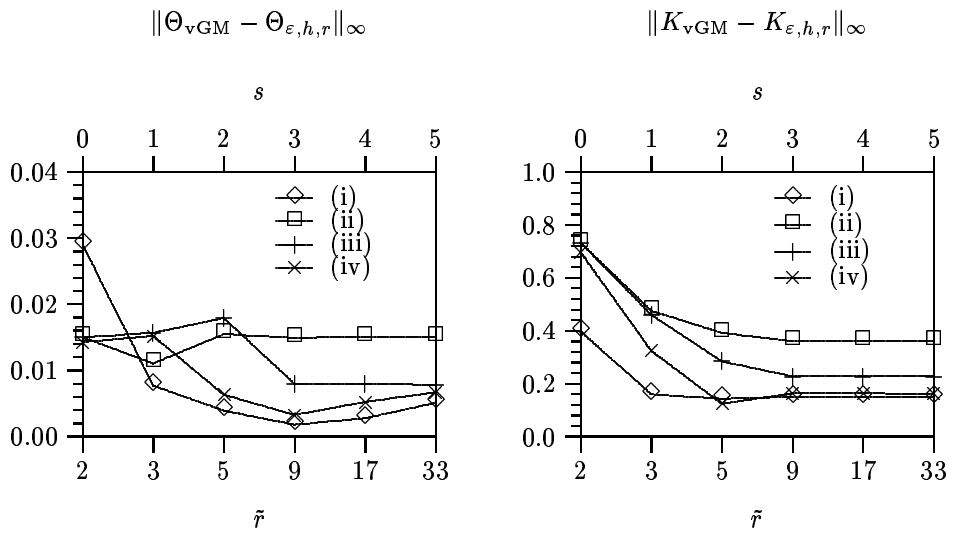


Abbildung 4.27: Identifizierungsfehler für die Parametrisierungen (i)–(iv), $\varepsilon = 5\%$, $h = \frac{L}{100}$, $n = 50$.

In den Abbildungen 4.27 und 4.28 werden die folgenden 4 Parametrisierungen gegenübergestellt:

(i) Θ und K quadratisch mit lokaler Basis,

(ii) Θ quadratisch und K linear mit lokaler Basis,

(iii) Θ quadratisch und K linear mit hierarchischer Basis und

(iv) monotoner kubischer Ansatz für Θ und K.

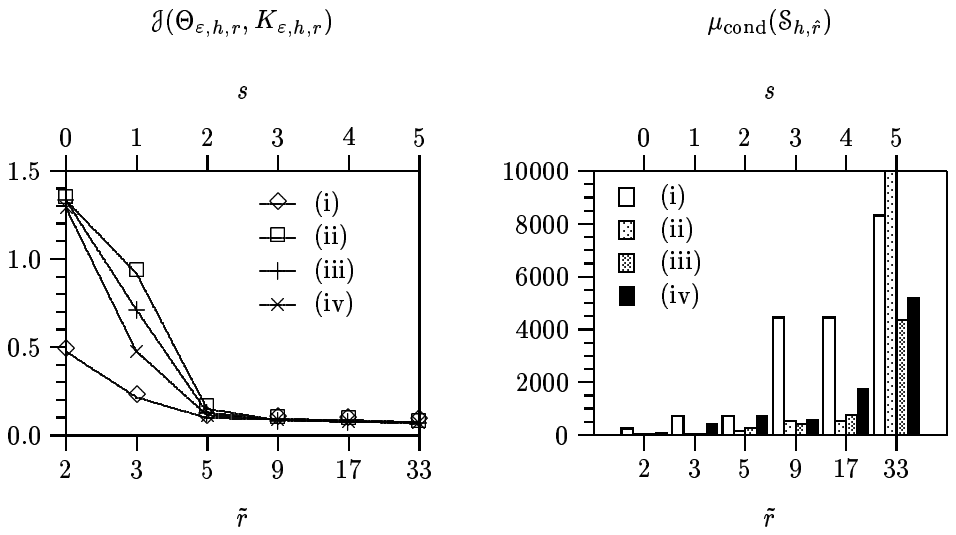


Abbildung 4.28: Fehlerfunktional und Spektralkondition für die Parametrisierungen (i)–(iv), $\varepsilon = 5\%$, $h = \frac{L}{100}$, $n = 50$.

Zu beachten ist hierbei, dass $\tilde{r}$ die Anzahl der Stützstellen der Parametrisierung bezeichnet. Da in den Fällen (ii) und (iii) Θ mit quadratischen Splines und K mit linearen Splines parametrisiert sind, enthält Θ jeweils einen Freiheitsgrad mehr als K.

Indem wir anstelle der Funktion K die Funktion $\ln K$ parametrisieren. können wir u. U. erreichen, dass sich die Sensitivität verbessert. So ist z. B. bei linearer Parametrisierung von $\ln K$ und quadratischer Parametrisierung

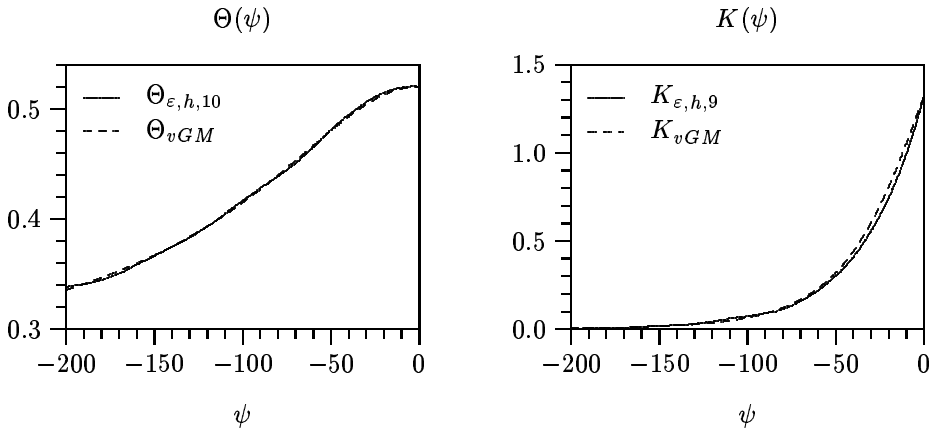


Abbildung 4.29: Identifizierte hydraulische Funktionen bei quadratischem Ansatz von Θ und linearem Ansatz von $\ln K$ mit $\tilde{r} = 9$ Stützstellen, $\varepsilon = 5\%$, $h = \frac{L}{100}$, $n = 50$.

von Θ (Abbildung 4.29) eine deutliche Verbesserung der identifizierten Leitfähigkeit gegenüber dem quadratischen Ansatz für Θ und K (Abbildung 4.10) festzustellen.

Bei einem anderen Bodentyp mit den van Genuchten-Mualem-Parametern $\theta_{\text{sat}} = 0.35$, $\theta_{\text{res}} = 0.005$, $K_{\text{sat}} = 1.2$ cm/h, $\alpha = 0.01$ cm^{-1} und $n = 4.0$ (entsprechen den typischen Werten eines Sandes) liefert die quadratische lokale Parametrisierung der Retentionsfunktion und des Logarithmus der Leitfähigkeit die Resultate mit den kleinsten Identifizierungsfehlern. Einige Ergebnisse aus dem Multi-Level-Algorithmus ($r = 4, 9, 34$) sind in den Abbildungen 4.30–4.35 dargestellt. In diesem Beispiel wurde der Druck am unteren Rand langsamer abgesenkt und als Endzeit $T = 100$ h gewählt. Bei der identifizierten Retentionsfunktion sieht man, dass die Beobachtungen noch nicht genug Informationen für den Druckbereich $[-200, -180]$ enthalten. Am Ausflussrand liegt zum Zeitpunkt $t = 100$ h zwar ein Druck von -200 cm an, dieser hat jedoch noch nicht ausreichend Einfluss innerhalb der Säule.

Bei der SQP-Methode, welche zur Minimierung des Fehlerfunktionals eingesetzt wird, handelt es sich um ein iteratives Verfahren. Die Ergebnisse hängen damit von der Anzahl der Iterationen ab. Die Optimierungsresultate einer Stufe des Multi-Level-Algorithmus bestimmen den Startwert für die Optimierung in der nächste Stufe. Unterschiedliche Startwerte können jedoch zu unterschiedlichen Optimierungsresultaten führen. Somit ist es möglich, dass sich die Multi-Level-Identifizierung in einem Nebenminimum festsetzt, wenn in einem Schritt des Multi-Level-Algorithmus zu wenige Iterationen durchgeführt werden, d. h. das Optimum nur unzureichend approximiert wird. Dies kann anhand des in den Abbildungen 4.30–4.35 betrachteten Beispiels illustriert werden. Die Optimierung im zweiten Schritt des Multi-Level-Algorithmus ($r = 4$) wird nach 9 Iterationen abgebrochen. Das Optimalitätskriterium (siehe [51]) ist dann noch nicht erfüllt und wir erhalten andere identifizierte hydraulische Funktionen, insbesondere die Leitfähigkeiten unterscheiden sich deutlich (vgl. Abbildungen 4.36 und 4.37). Wenn die Multi-Level-Identifizierung mit diesen Startwerten fortgesetzt wird (wobei die Minimierung wieder erst nach Erfüllung des Optimalitätskriteriums abgebrochen wird), so erhalten wir auch für die folgenden Parametrisierungsstufen andere Ergebnisse (vgl. 4.38 und 4.39 für $r = 10$). Ebenfalls wieder stark betroffen ist in diesem Beispiel die Leitfähigkeit. Obwohl jetzt im Gegensatz zu Abbildung 4.30 in der Leitfähigkeit keine Monotonieverletzung auftritt und teilweise der Wert des Fehlerfunktionals (siehe Abbildung 4.40) kleiner ist, erhalten wir letztendlich einen größeren Identifizierungfehler. Die Spektralkondition verhält sich annähernd gleich. Das Fazit hieraus ist, dass in jedem Schritt des Multi-Level-Algorithmus das Optimum möglichts „gut“ approximiert werden sollte.

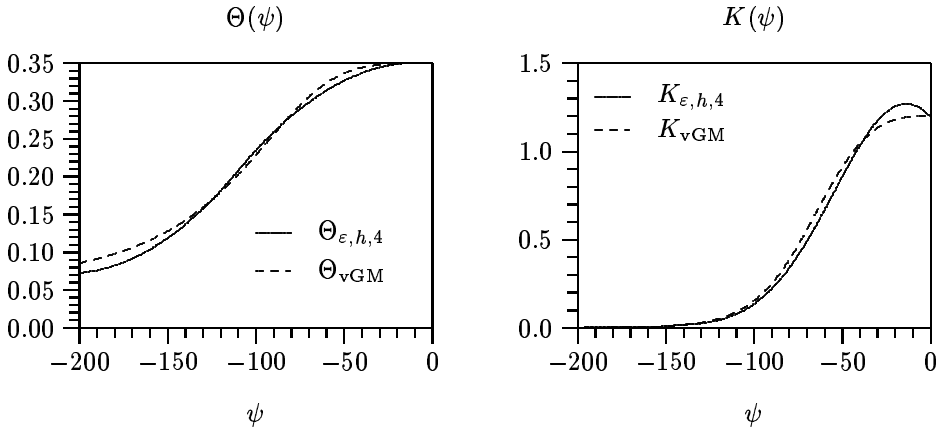


Abbildung 4.30: Hydraulische Funktionen, $r = 4$, $\varepsilon = 5\%$, $h = \frac{L}{100}$, $n = 50$.

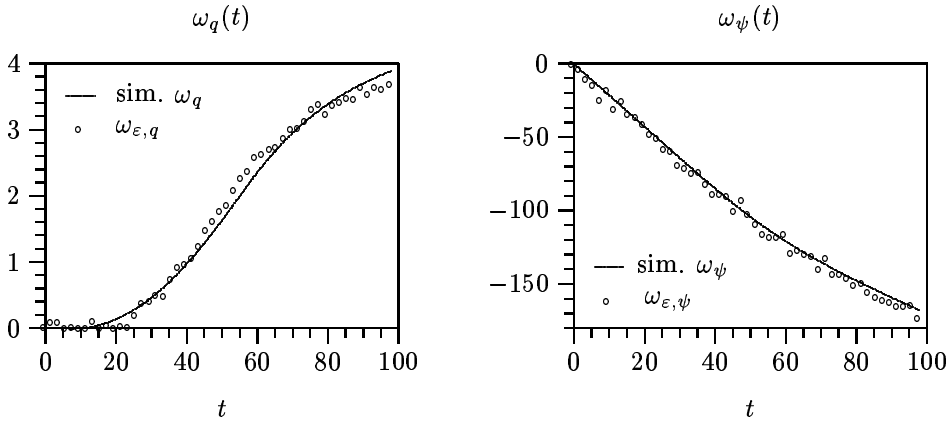


Abbildung 4.31: Beobachtungen zu Abbildung 4.30.

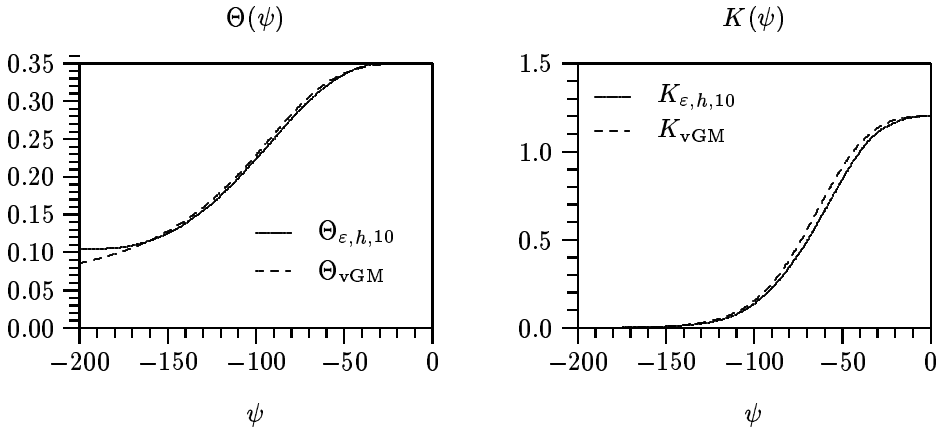


Abbildung 4.32: Hydraulische Funktionen, $r = 10$, $\varepsilon = 5\%$, $h = \frac{L}{100}$, $n = 50$.

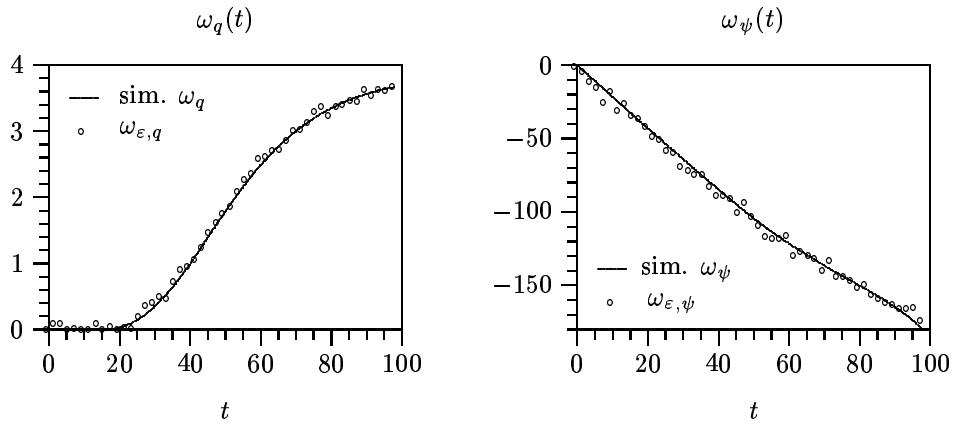


Abbildung 4.33: Beobachtungen zu Abbildung 4.32.

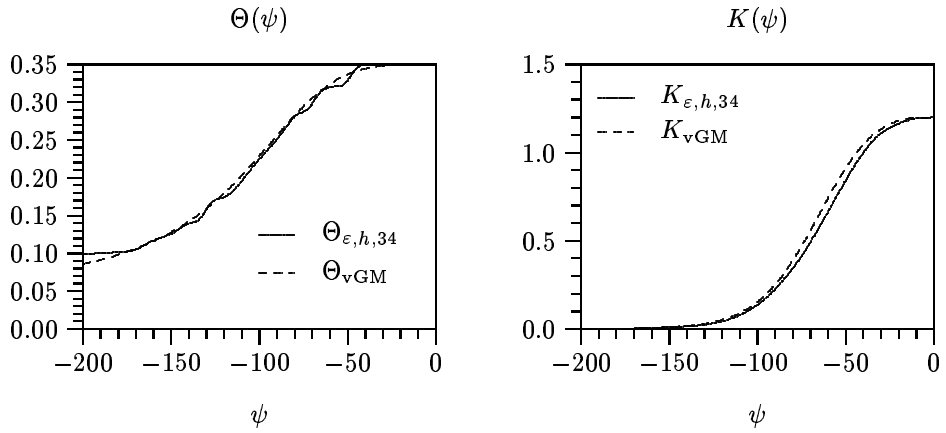


Abbildung 4.34: Hydraulische Funktionen, $r = 34$, $\varepsilon = 5\%$, $h = \frac{L}{100}$, $n = 50$.

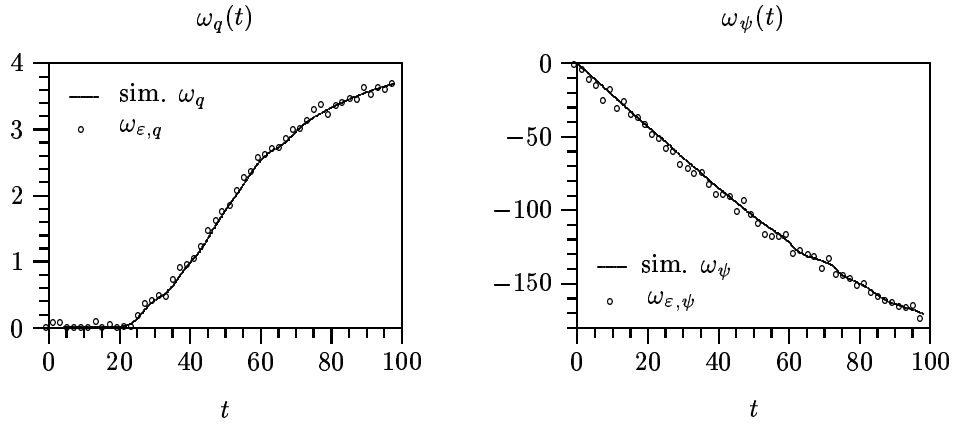


Abbildung 4.35: Beobachtungen zu Abbildung 4.34.

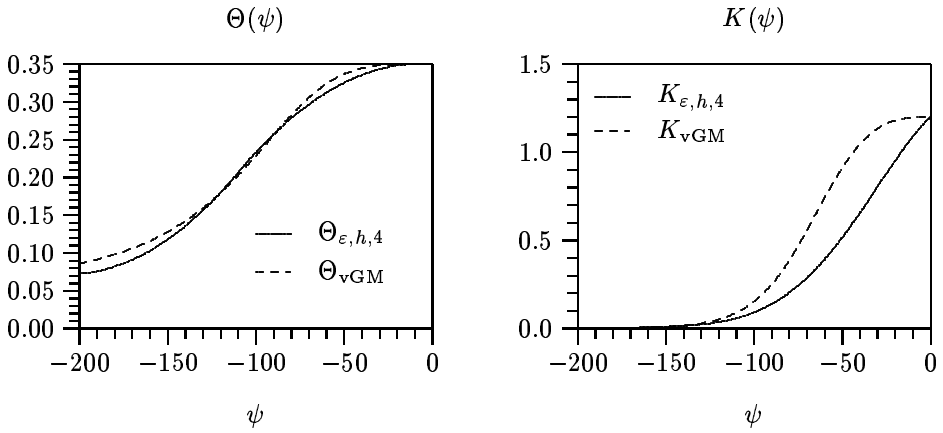


Abbildung 4.36: Hydraulische Funktionen, nichtoptimal, $r = 4$.

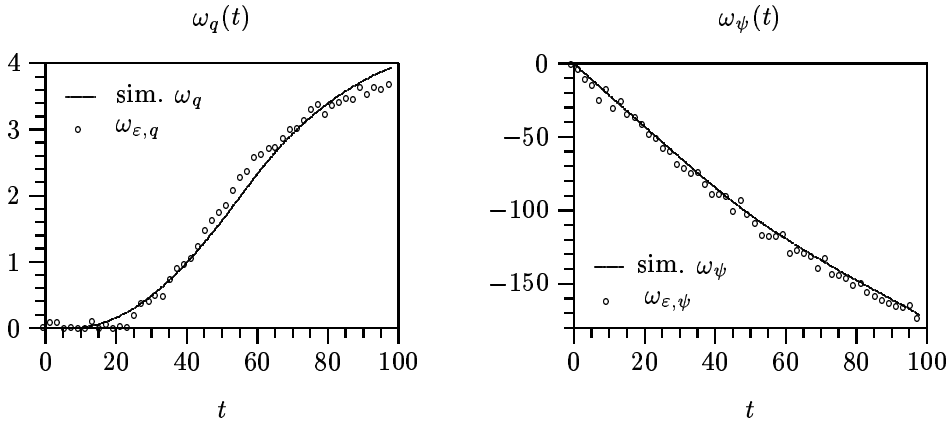


Abbildung 4.37: Beobachtungen zu Abbildung 4.36.

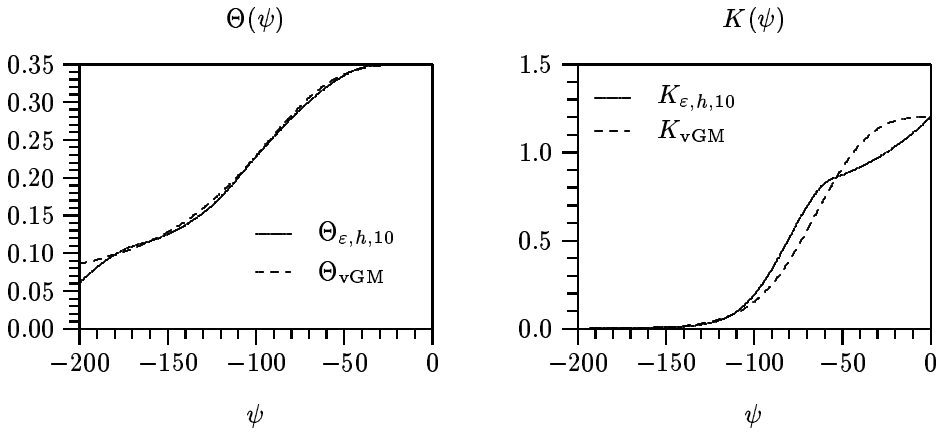


Abbildung 4.38: Hydraulische Funktionen, nichtoptimal, $r = 10$.

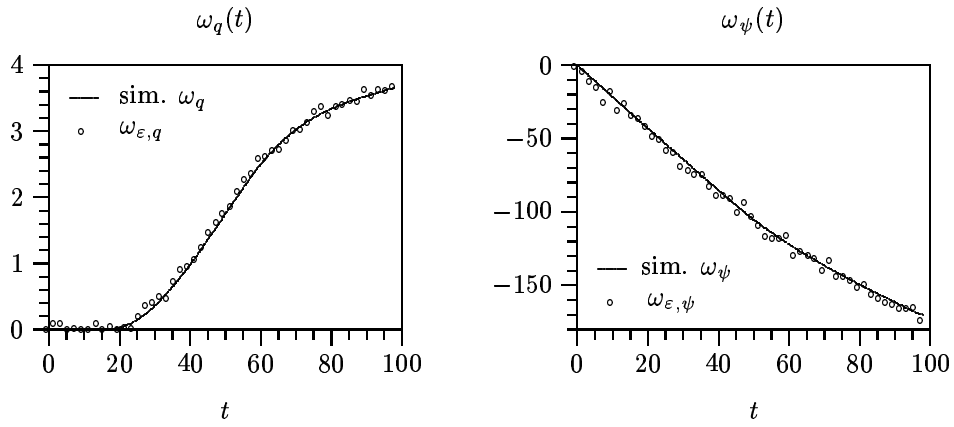


Abbildung 4.39: Beobachtungen zu Abbildung 4.38.

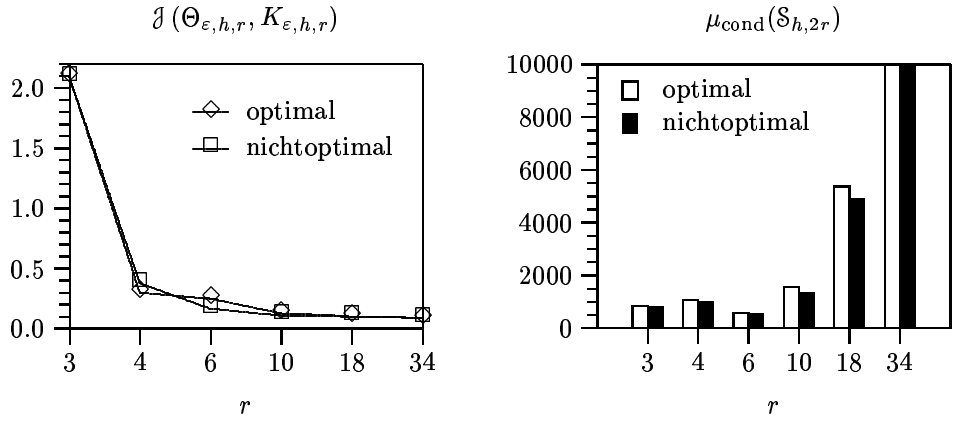


Abbildung 4.40: Fehlerfunktional und Spektralkondition, optimal und nichtoptimal.

Bevor wir das Identifizierungsverfahren an experimentellen Daten testen, sei noch das Folgende bemerkt. Die insbesondere bei der identifizierten Retentionsfunktion beobachteten oszillatorischen Störungen können sich bei Nichtbeachtung der Monotoniebedingungen derart verstärken, dass die identifizierten Funktionen nicht mehr monoton sind. Dies kann dazu führen, dass das direkte Problem nicht mehr lösbar ist (siehe Bemerkungen zur Lösbarkeit in [52] und Abschnitt 3.4).

4.2.3 Experimentelle Daten

Hydraulische Funktionen für Bayreuther sandigen Lehm

Für einen Bayreuther sandigen Lehm (BSL) wurde an der Universität Bayreuth ein Ausflussexperiment durchgeführt, bei dem eine Säule der Länge

$L = 15.0$ cm durch Absenken des am Ausflussrand angelegten Drucks von 15.0 cm auf -60.0 cm entwässert wurde. Die Identifizierung der hydraulischen Funktionen erfolgte mit quadratischer lokaler Parametrisierung von Retentionsfunktion und logarithmisierter Leitfähigkeit. Die Ergebnisse der Identifizierung sind für $r = 3, 4, 6, 10, 18, 34$ in den Abbildungen 4.41–4.53 dargestellt. Offensichtlich tritt im Zeitintervall $[18, 40]$ im kumulativen Ausfluss ein deutlicherer Messfehler auf als in den übrigen Daten.

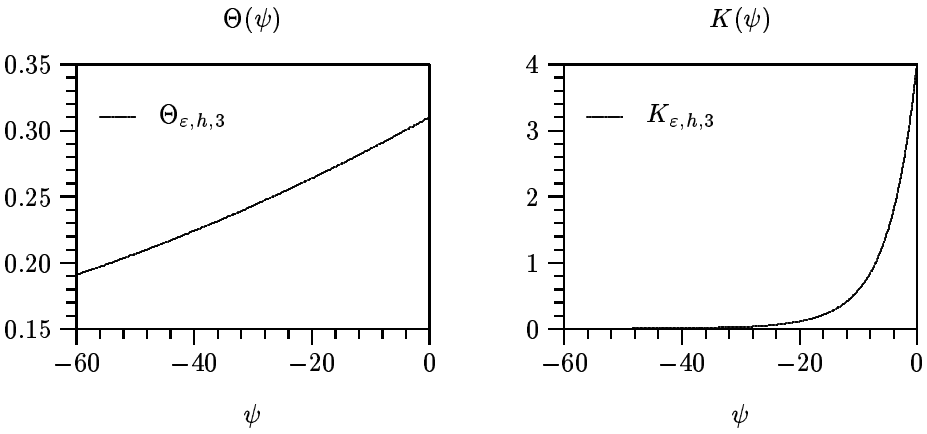


Abbildung 4.41: Hydraulische Funktionen für BSL, $r = 3$, $h = \frac{L}{100}$, $n = 157$.

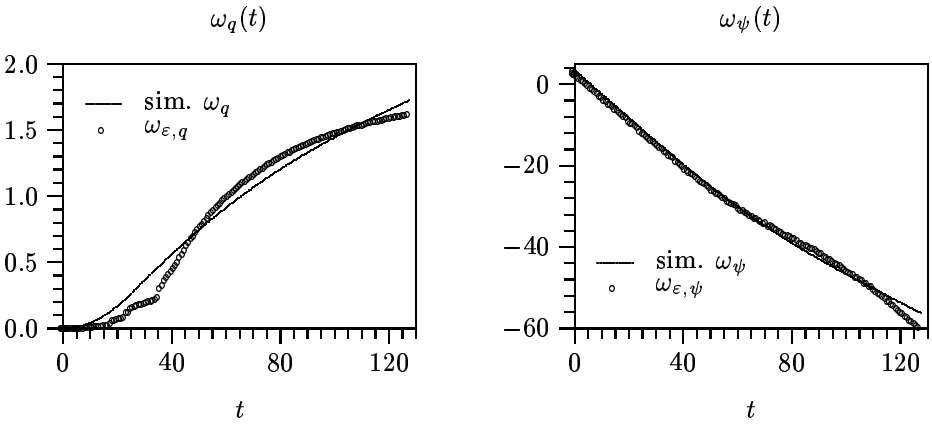


Abbildung 4.42: Beobachtungen zu Abbildung 4.41.

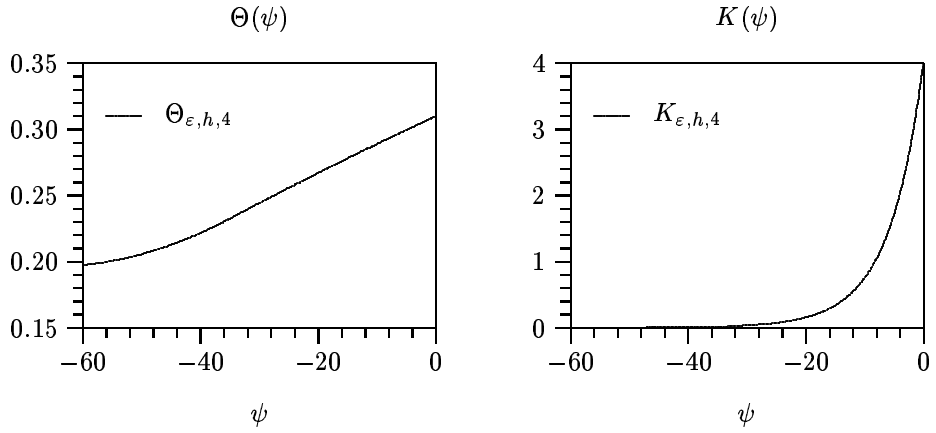


Abbildung 4.43: Hydraulische Funktionen für BSL, $r = 4$, $h = \frac{L}{100}$, $n = 157$.

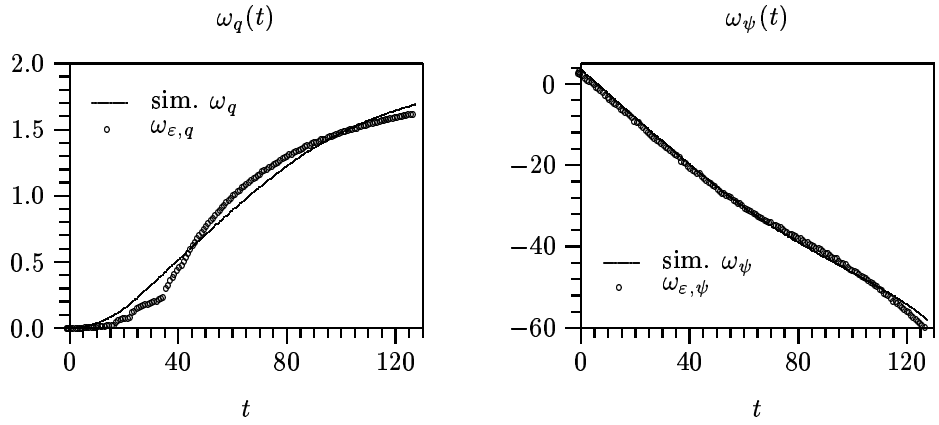


Abbildung 4.44: Beobachtungen zu Abbildung 4.43.

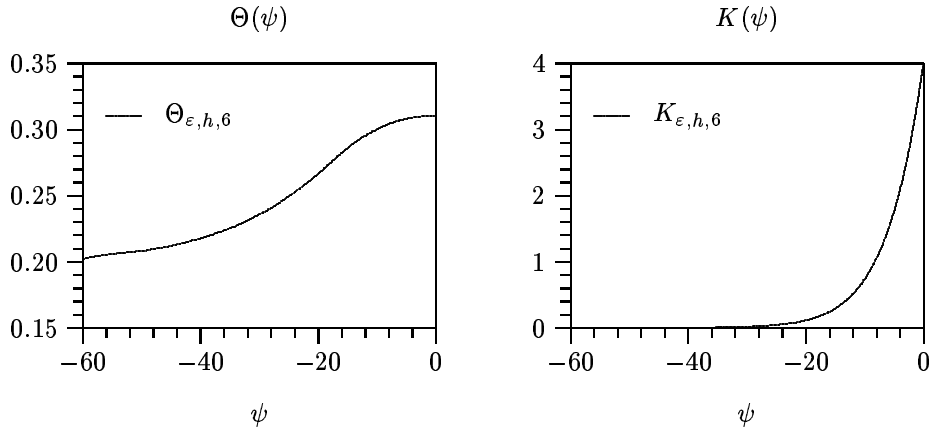


Abbildung 4.45: Hydraulische Funktionen für BSL, $r = 6$, $h = \frac{L}{100}$, $n = 157$.

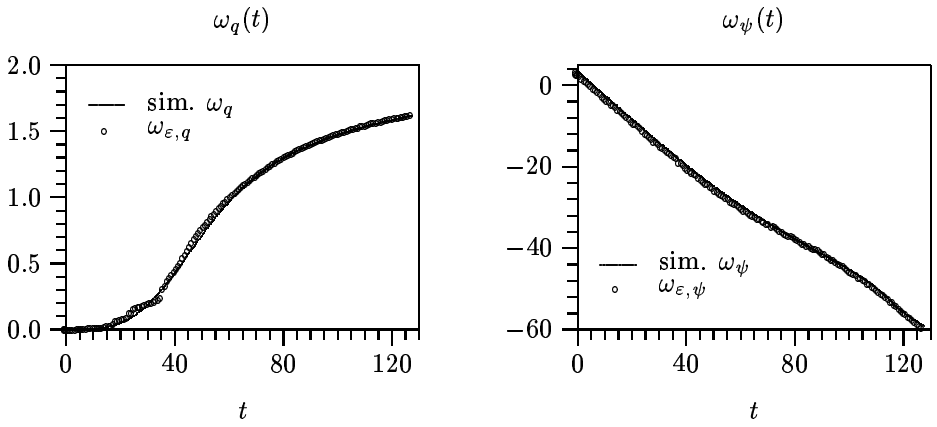


Abbildung 4.46: Beobachtungen zu Abbildung 4.45.

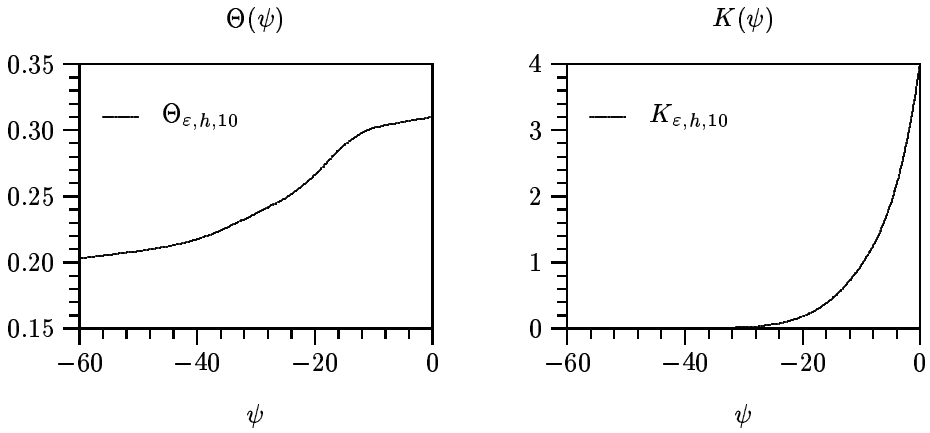


Abbildung 4.47: Hydraulische Funktionen für BSL, $r = 10$, $h = \frac{L}{100}$, $n = 157$.

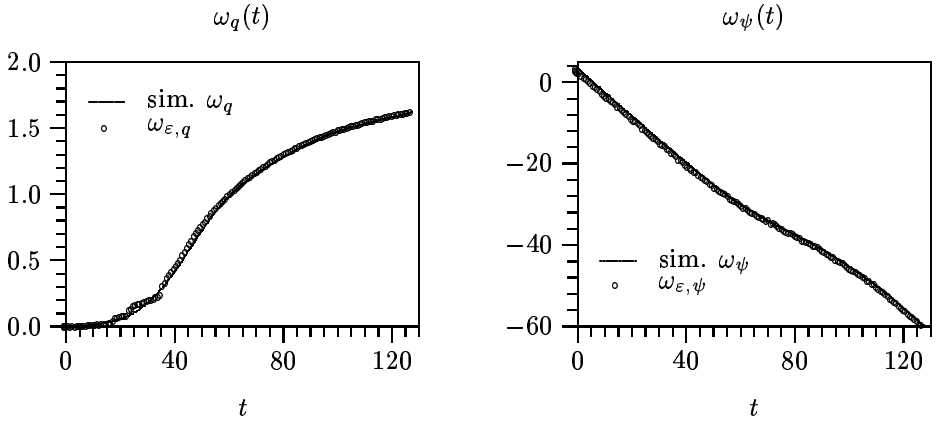


Abbildung 4.48: Beobachtungen zu Abbildung 4.47.

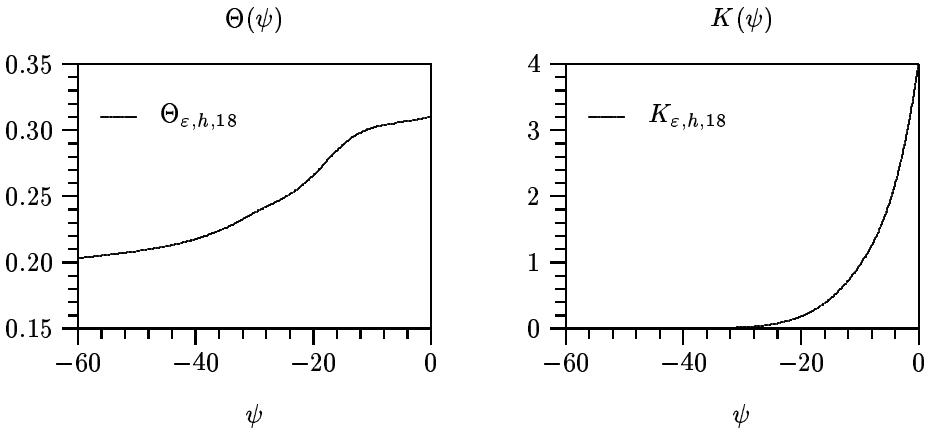


Abbildung 4.49: Hydraulische Funktionen für BSL, $r = 18$, $h = \frac{L}{100}$, $n = 157$.

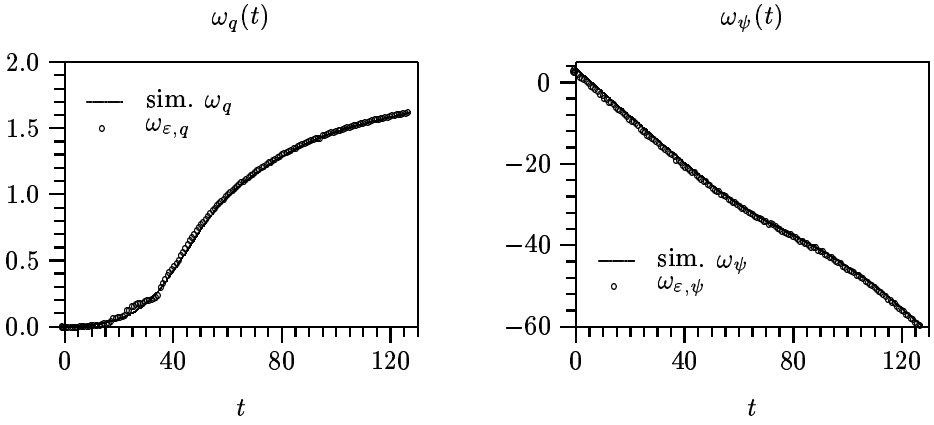


Abbildung 4.50: Beobachtungen zu Abbildung 4.49.

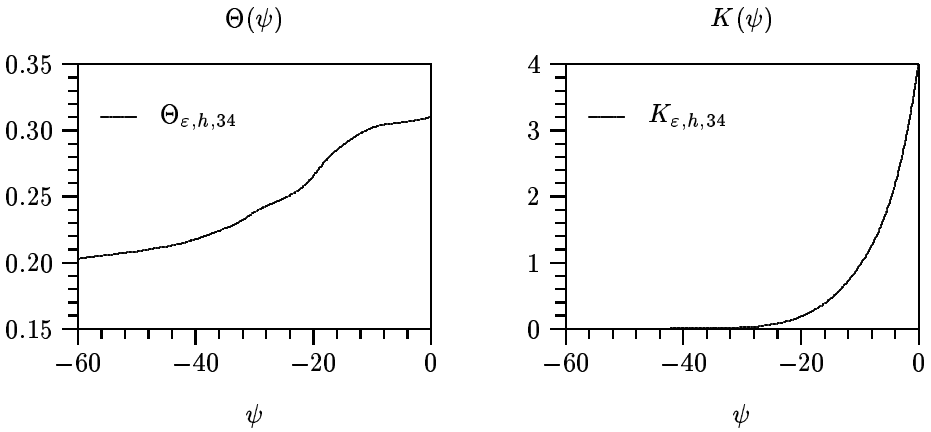


Abbildung 4.51: Hydraulische Funktionen für BSL, $r = 34$, $h = \frac{L}{100}$, $n = 157$.

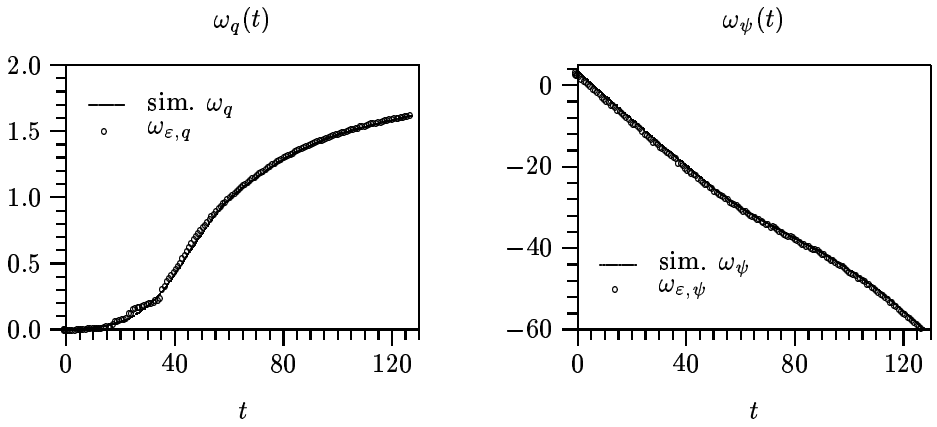


Abbildung 4.52: Beobachtungen zu Abbildung 4.51.

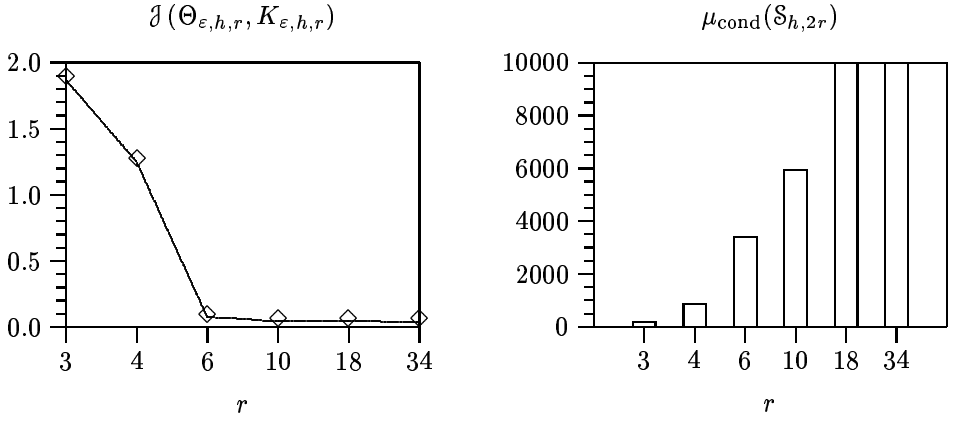


Abbildung 4.53: Fehlerfunktional und Spektralkondition für BSL, $h = \frac{L}{100}$, $n = 157$.

Für dieses Beispiel sollen die Sensitivitäten der Beobachtungen, d. h. deren partiellen Ableitungen, bezüglich der Parameter im jeweiligen Optimalwert betrachtet werden. Hierzu sind stückweise lineare Parametrisierungen für die hydraulischen Funktionen verwendet worden. In den Abbildungen 4.54 und 4.55 sind die Sensitiviäten für $r = 9$ Freiheitsgrade abgebildet. Dabei zeigt ein Diagramm jeweils die Sensitivitäten der Druckbeobachtung bzw. der Ausflussbeobachtung bezüglich der Parameter der Retentionsfunktion bzw. der Leitfähigkeit. Die t-Achse entspricht der Zeitachse der Beobachtungen und die ψ-Achse stellt den Definitionsbereich der hydraulischen Funktionen dar.

Deutlich zu erkennen ist zunächst einmal die starke Lokalität der einzelnen Parameter. Die lokalen Maxima bzw. Minima der Sensitivitäten korrespondieren jeweils mit einer Stützstelle im ψ-Bereich. Weiterhin ist zu erkennen, dass eine Änderung der Parameter kaum einen Einfluss auf den Druck na-

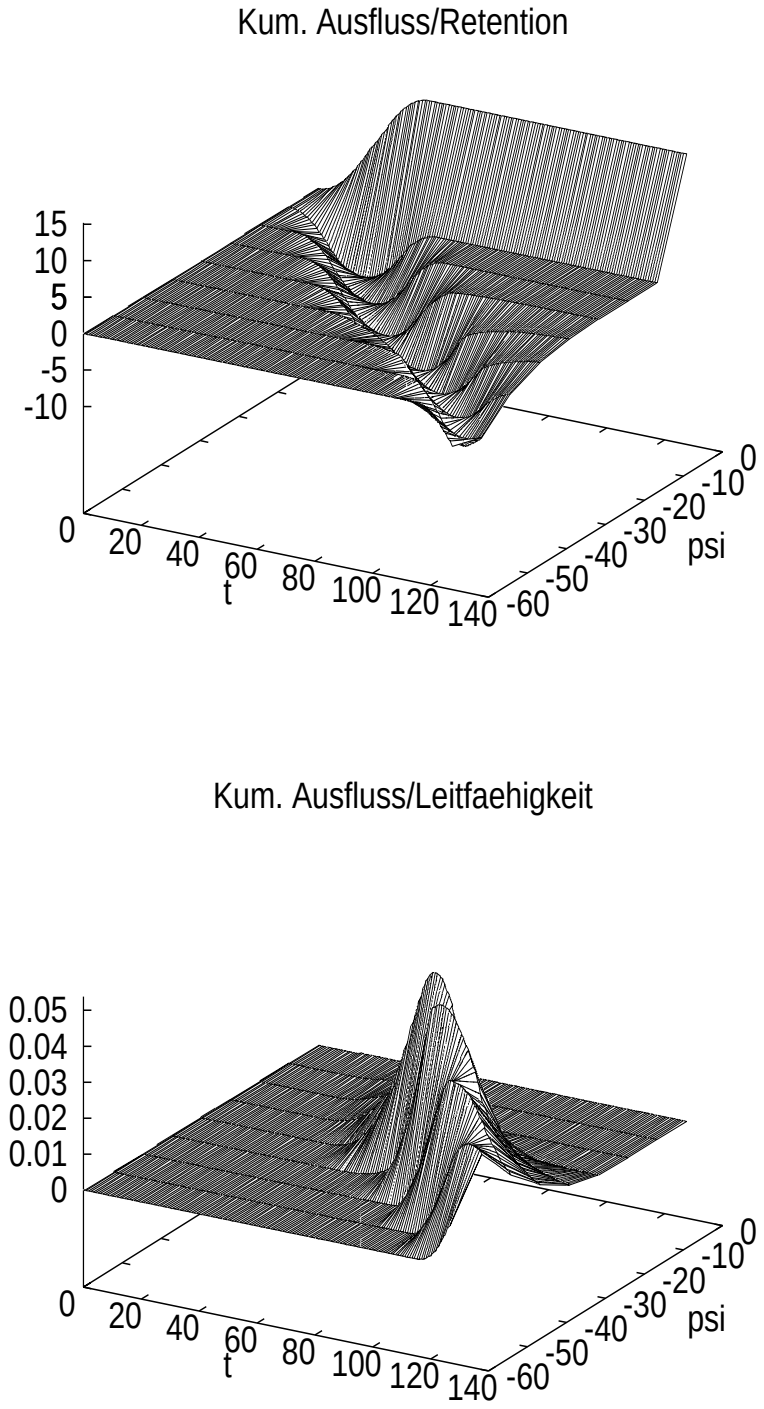


Abbildung 4.54: Sensitivitäten des kumulativen Ausflusses für BSL bei stückweise linerem Ansatz mit $r = 9$, $h = \frac{L}{100}$, $n = 157$.

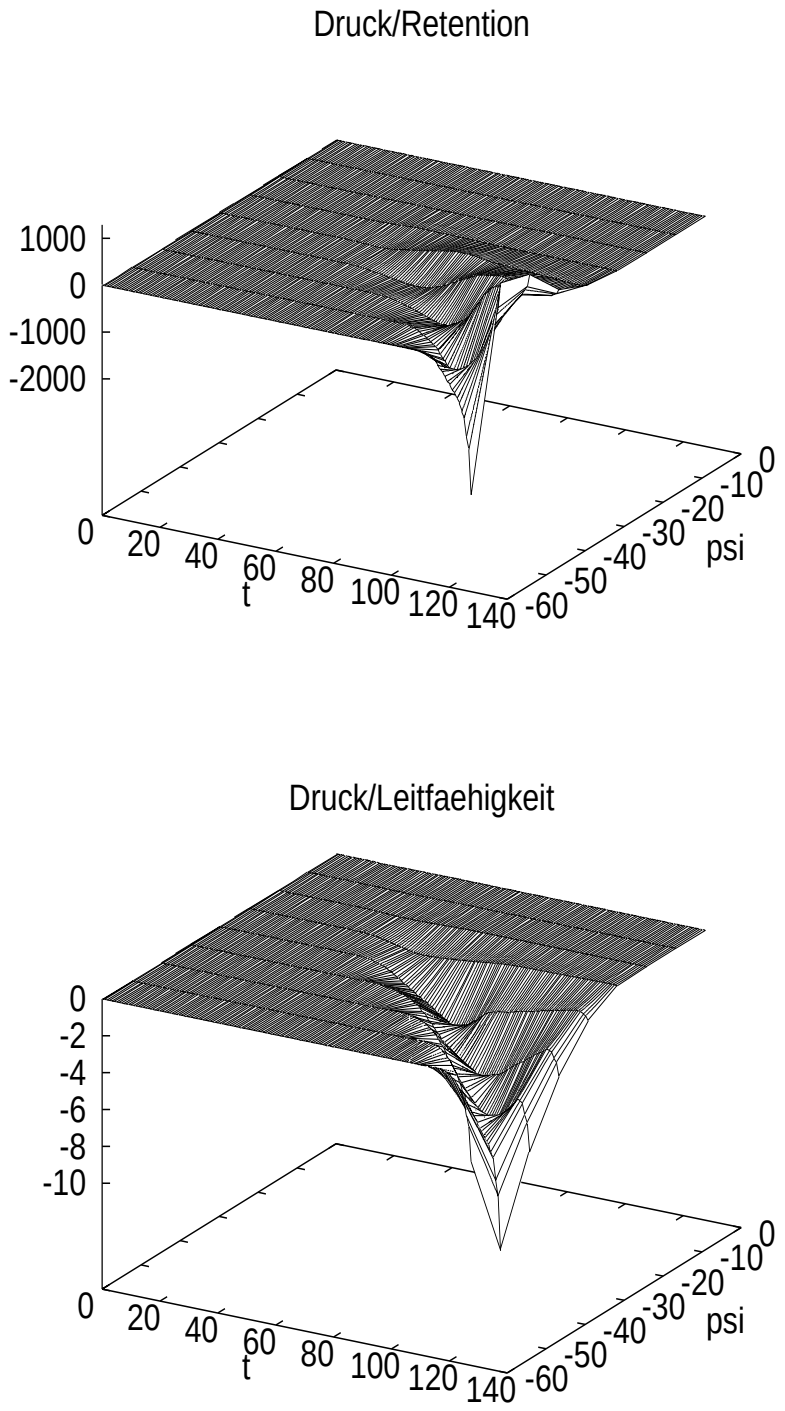


Abbildung 4.55: Sensitivitäten der Druckbeobachtung für BSL bei stückweise linearem Ansatz mit $r = 9$, $h = \frac{L}{100}$, $n = 157$.

he der Sättigungsgrenze ausübt. Dies entspricht ungefähr dem Zeitintervall $[0, 40]$. Auch bewirkt eine Veränderung des Wertes der Leitfähigkeit im Bereich der Sättigung oder nahe Sättigung kaum eine Änderung in den zugehörigen Beobachtungen. Für die Optimierung ist es deshalb sinnvoll mit fixiertem K_{sat}-Wert zu arbeiten.

Die Sensitivitäten für die Parametrisierungen mit anderer Anzahl von Freiheitsgraden führen zu ähnlichen Darstellungen wie in den Abbildungen 4.54 und 4.55, wobei die Anzahl der lokalen Maxima bzw. Minima entsprechend der Anzahl der Freiheitsgrade variiert.

Hydraulische Funktionen für Forchheimer Sand

Für einen weiteren Boden, einen Forchheimer Sand (FS), erfolgte ein analoges Experiment. Auch hier wurde ein quadratischer lokaler Ansatz zur Parametrisierung von Θ und $\ln K$ gewählt. Die Identifizierungsergebnisse für $r = 4, 6, 10$ sind in den Abbildungen 4.56–4.62 aufgeführt. Obwohl in den Daten für den kumulativen Aussfluss und den Druck Messfehler enthalten sind, treten diese nicht so deutlich hervor, wie dies teilweise in den Ausflussdaten für den Bayreuther sandigen Lehm der Fall war.

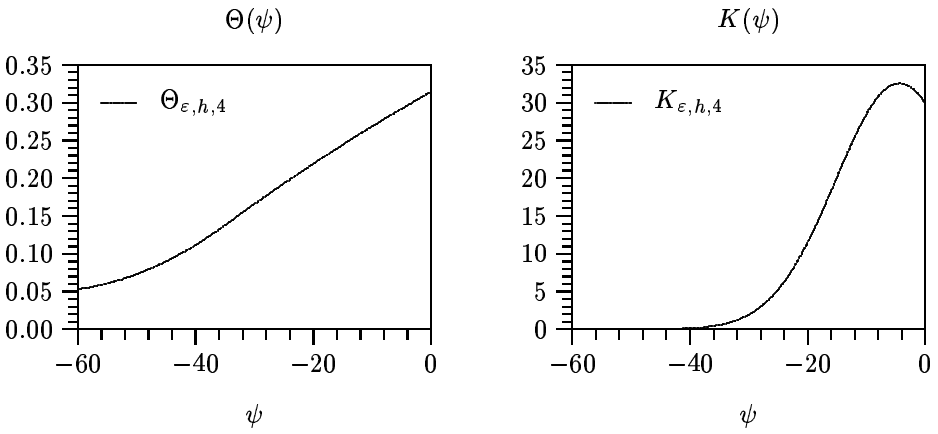


Abbildung 4.56: Hydraulische Funktionen für FS, $r = 4$, $h = \frac{L}{100}$, $n = 160$.

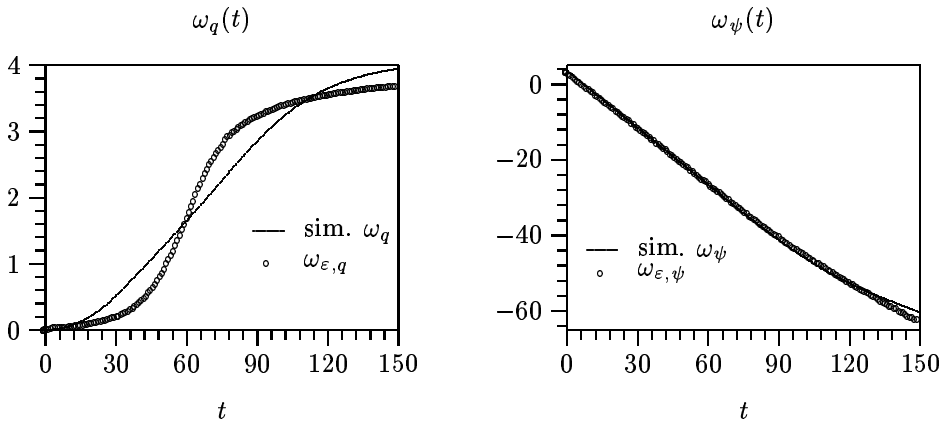


Abbildung 4.57: Beobachtungen zu Abbildung 4.56.

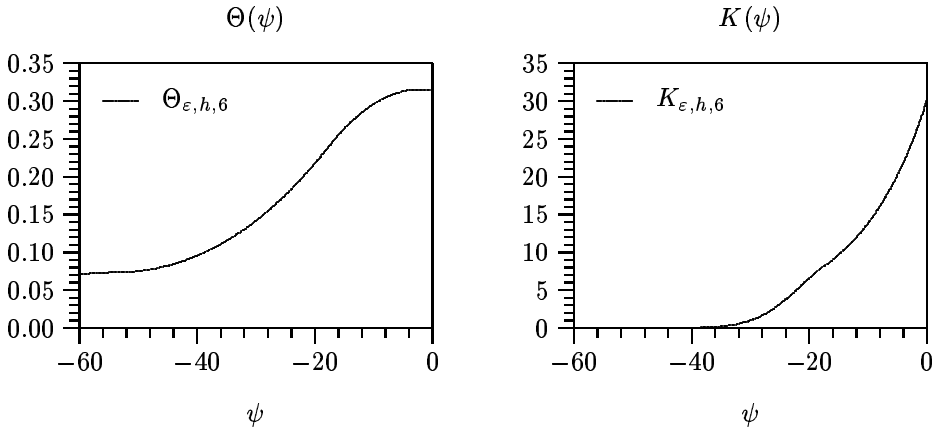


Abbildung 4.58: Hydraulische Funktionen für FS, $r = 6$, $h = \frac{L}{100}$, $n = 160$.

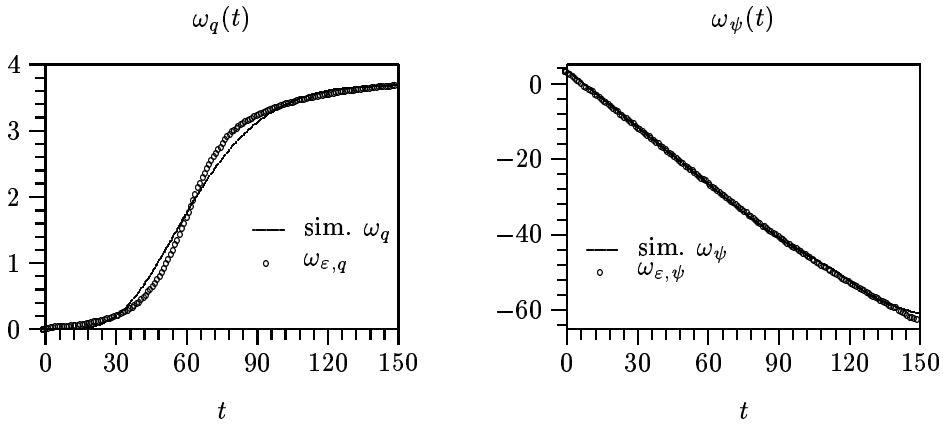


Abbildung 4.59: Beobachtungen zu Abbildung 4.58.

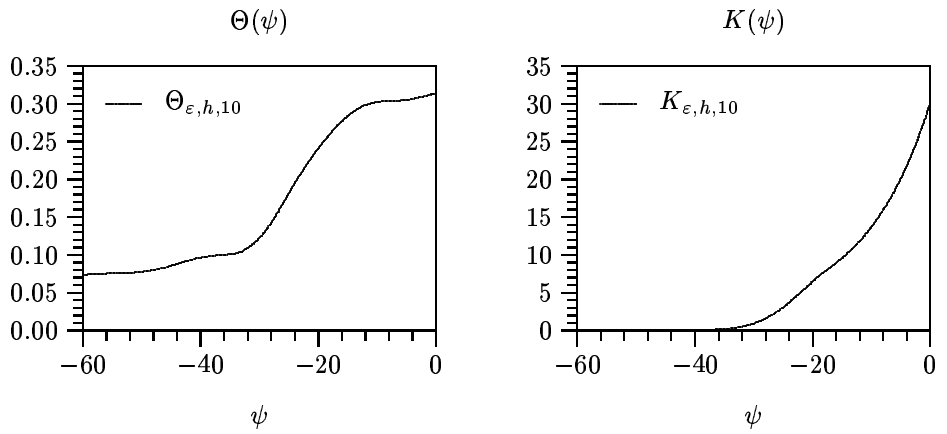


Abbildung 4.60: Hydraulische Funktionen für FS, $r = 10$, $h = \frac{L}{100}$, $n = 160$.

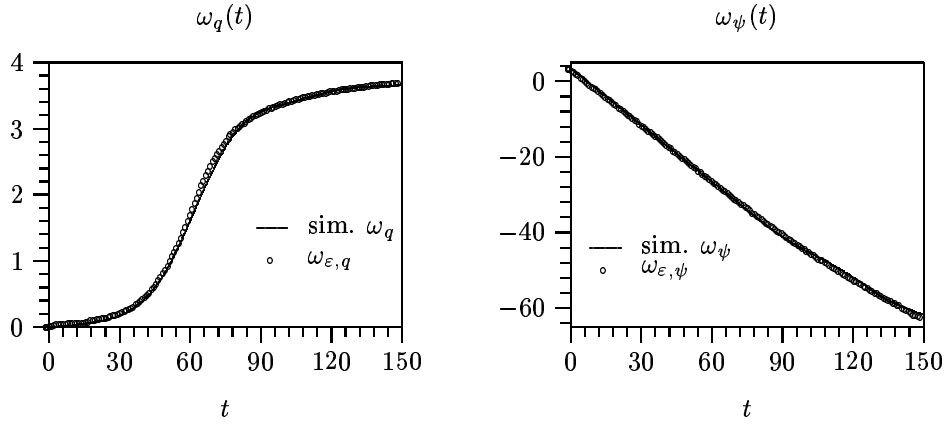


Abbildung 4.61: Beobachtungen zu Abbildung 4.60.

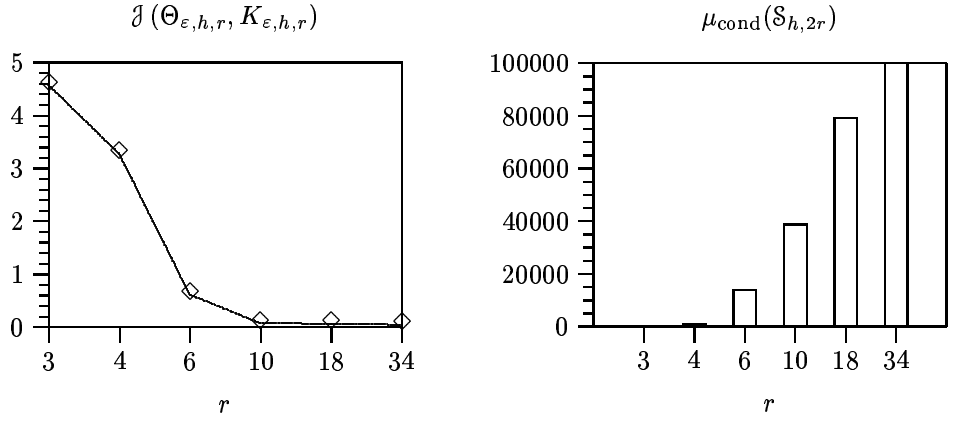


Abbildung 4.62: Fehlerfunktional und Spektralkondition für FS, $h = \frac{L}{100}$, $n = 160$.

4.3 Adaptivität im Multi-Level-Algorithmus

4.3.1 Eine adaptive Verfeinerungsstrategie

Das Vorgehen, wie wir im Multi-Level-Algorithmus zur nächsten Stufe der Parametrisierung gelangen, wird als *Verfeinerungsstrategie* bezeichnet. Bisher sind wir entsprechend dem Prinzip der hierarchischen Basen skalenweise vorgegangen. Wenn wir mit einer linearen lokalen Basis parametrisieren, so können wir verfeinern, indem wir zu den alten Stützstellen eine oder mehrere neue Stützstellen hinzufügen. Eine Wahlmöglichkeit besteht hierbei nicht nur in der Anzahl der neuen Stützstellen, sondern auch in deren Lage. Die alten Stützstellen werden beibehalten, damit die Bedingung (4.4) erfüllt wird.

Ausgangspunkt sei die Parametrisierung $p^{\nu}_{r_\nu}$ ($\nu \in \{1, ..., M\}$ fest) einer Koeffizientenfunktion mittels einer linearen lokalen Basis zur Zerlegung

$$\psi_{1,r_\nu} < \psi_{2,r_\nu} < ... < \psi_{r_\nu,r_\nu} \tag{4.20}$$

mit r_ν Freiheitsgraden. Unser Ziel ist es, eine Parametrisierung mit einem zusätzlichen Freiheitsgrad zu erhalten. Wir setzen also $\Delta r_\nu = 1$ und suchen zur Zerlegung (4.20) eine zusätzliche Stützstelle:

$$\psi_{\text{neu}} = ?$$

Die (diskreten) Beobachtungen seien gegeben durch die Wertepaare $\left(\tilde{t}^i, \omega^i_\varepsilon\right)$, $i = 1, ..., \hat{\kappa}$. Die zur Optimallösung $p_{\varepsilon,h,\hat{r}}$ gehörende simulierte Beobachtung $\omega_h(p_{\varepsilon,h,\hat{r}})$ bezeichnen wir zur Vereinfachung mit $\omega_{h,\text{opt}}$. Das diskrete Funktional (4.2) ist darstellbar in der Form

$$\tilde{\mathcal{J}}_{\varepsilon,h}(\omega_h) = \sum_{i=1}^{\hat{\kappa}} \tilde{\mathcal{J}}_{\varepsilon,h,i}\left(\omega^i\right)$$

mit $\tilde{\mathcal{J}}_{\varepsilon,h,i}(\omega^i) = \alpha^i \left(\omega^i - \omega^i_\varepsilon\right)^2$.

Zur Gewinnung einer *adaptiven Verfeinerungsstrategie* legen wir zunächst eine Teilmenge von $\{1, ..., \hat{\kappa}\}$ fest:

$$I \subseteq \{1, ..., \hat{\kappa}.\}$$

Wenn wir eine (lokale) Verbesserung der simulierten Beobachtung an der „schlechtesten“ Stelle erreichen wollen, so wählen wir I als die einelementige Menge $I = \{k\}$, welche aus demjenigen Index $k \in \{1, ..., \hat{\kappa}\}$ besteht, für den

$$\left|\tilde{\mathcal{J}}_{\varepsilon,h,k}\left(\omega^k_{\text{opt}}\right)\right| = \max_{1 \le i \le \hat{\kappa}} \left|\tilde{\mathcal{J}}_{\varepsilon,h,i}\left(\omega^i_{\text{opt}}\right)\right|$$

gilt. $I = \{1, ..., \hat{\kappa}\}$ oder andere Auswahlkriterien sind ebenso möglich.

Die neue Stützstelle platzieren wir nun dort, wo sich die Sensitivität der zur Indexmenge I gehörenden (Teil-)Beobachtung bezüglich $p^{\nu}_{\varepsilon,h,r_\nu}$ am stärksten ändert. Dazu betrachten wir den zur ν-ten Koeffizientenfunktion gehörenden Teil der Sensitivitätsmatrix $\mathcal{S}_{h,\hat{r}}$

$$\left(\frac{\partial \omega^i_{\text{opt}}}{\partial p^{\nu}_{j,r_\nu}}\right)_{\substack{i=1,\dots,\hat{\kappa}\\ j=1,\dots,r_\nu}}$$

und bestimmen einen Index $l \in \{1, ..., r_\nu - 1\}$, für den

$$\sum_{i\in I}\left|\frac{\frac{\partial \omega^i_{\text{opt}}}{\partial p^{\nu}_{l+1,r_\nu}} - \frac{\partial \omega^i_{\text{opt}}}{\partial p^{\nu}_{l,r_\nu}}}{\psi_{l+1,r_\nu} - \psi_{l,r_\nu}}\right| = \max_{1\le j\le r_\nu-1}\sum_{i\in I}\left|\frac{\frac{\partial \omega^i_{\text{opt}}}{\partial p^{\nu}_{j+1,r_\nu}} - \frac{\partial \omega^i_{\text{opt}}}{\partial p^{\nu}_{j,r_\nu}}}{\psi_{j+1,r_\nu} - \psi_{j,r_\nu}}\right| \tag{4.21}$$

gilt. $\frac{\frac{\partial \omega^i_{\text{opt}}}{\partial p^{\nu}_{j+1,r_\nu}} - \frac{\partial \omega^i_{\text{opt}}}{\partial p^{\nu}_{j,r_\nu}}}{\psi_{j+1,r_\nu} - \psi_{j,r_\nu}}$ ist eine Approximation der ersten Ableitung von $\frac{\partial \omega^i_{\text{opt}}}{\partial p^{\nu}_{r_\nu}}$ nach ψ im Intervall $[\psi_{j,r_\nu}, \psi_{j+1,r_\nu}]$. Falls für den Index l

$$|\psi_{l+1,r_\nu} - \psi_{l,r_\nu}| > \delta \cdot \max_{1\le j\le r_\nu-1} |\psi_{j+1,r_\nu} - \psi_{j,r_\nu}| \tag{4.22}$$

mit einem $\delta > 0$ gilt, so wird im Intervall $[\psi_{l,r_\nu}, \psi_{l+1,r_\nu}]$ verfeinert:

$$\psi_{\text{neu}} = \frac{1}{2}\left(\psi_{l,r_\nu} + \psi_{l+1,r_\nu}\right). \tag{4.23}$$

Anderenfalls ist ein neuer Index gemäß (4.21) für $j \in \{1, ..., r_\nu - 1\} \setminus \{l\}$ zu bestimmen. Durch die Bedingung (4.22) können wir verhindern, dass sich die Stützstellen in einem Teilgebiet häufen, während in einem anderen Teilgebiet keine oder nur wenige Stützstellen liegen. Die verfeinerte Zerlegung für die Parametrisierung mit $r_\nu + 1$ Freiheitsgraden ist damit gegeben durch

$$\psi_{j,r_\nu+1} := \begin{cases} \psi_{j,r_\nu} & \text{für } j = 1, ..., l, \\ \psi_{\text{neu}} & \text{für } j = l+1, \\ \psi_{j-1,r_\nu} & \text{für } j = l+2, ..., r_\nu + 1. \end{cases} \tag{4.24}$$

Beispiel 4.12 Wir testen die oben beschriebene adaptive Verfeinerungsstrategie an der Identifizierung der Leitfähigkeit K aus dem kumulativen Ausfluss ω_q mit einem Datenfehler von $\varepsilon = 5\%$ bei bekannter Retentionsfunktion Θ. Zum Vergleich wiederholen wir die Identifizierung und benutzen nichtadaptive Strategien, bei denen wir gemäß (4.23) und (4.24) verfeinern, wobei

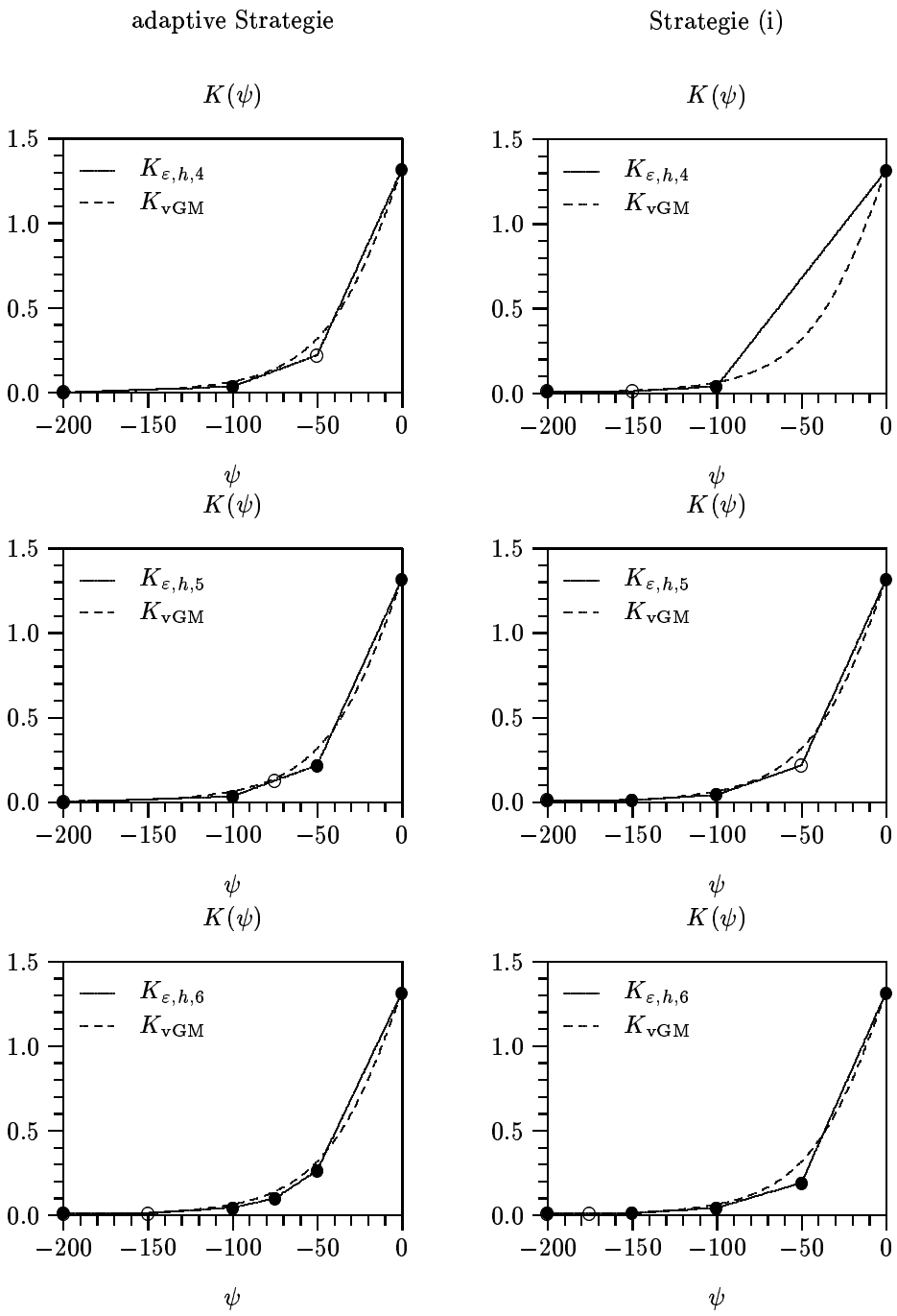


Abbildung 4.63: Leitfähigkeiten bei adaptiver und nichtadaptiver Verfeinerungsstrategie für $r = 4, 5, 6$, $\varepsilon = 5\%$, $h = \frac{L}{100}$, $n = 50$.

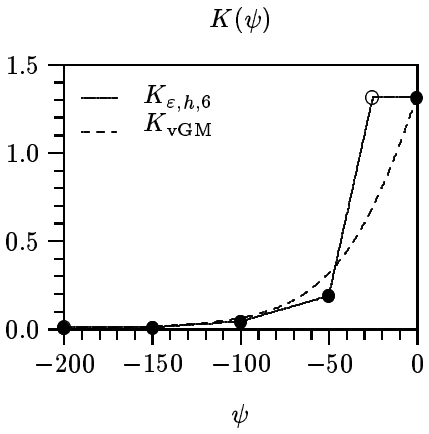


Abbildung 4.64: Leitfähigkeit bei der Verfeinerungsstrategie (ii) für $r = 6$.

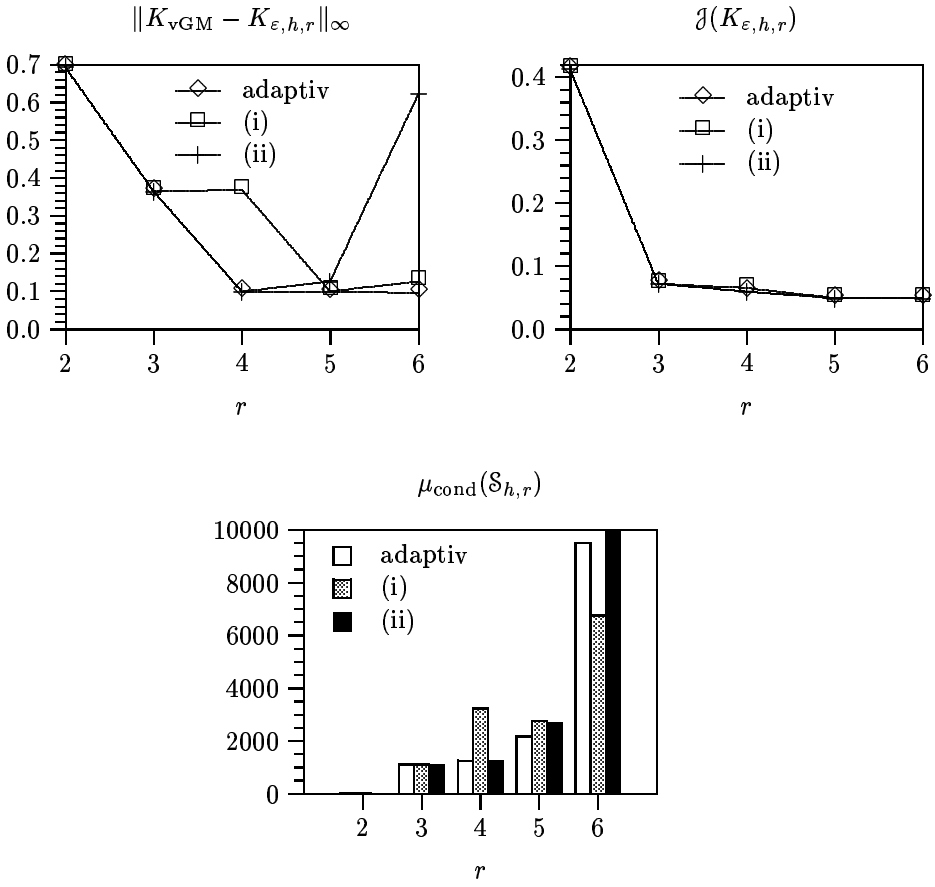


Abbildung 4.65: Identifizierungsfehler, Fehlerfunktional und Spektralkondition für die Strategien adaptiv, (i) und (ii), $\varepsilon = 5\%$, $h = \frac{L}{100}$, $n = 50$.

(i)

$$l = \min\left\{\hat{j} \in \{1, ..., r-1\} \;\middle|\; |\psi_{\hat{j}+1,r} - \psi_{\hat{j},r}| = \max_{1\le j\le r-1} |\psi_{j+1,r} - \psi_{j,r}|\right\}$$

(ii)

$$l = \max\left\{\hat{j} \in \{1, ..., r-1\} \;\middle|\; |\psi_{\hat{j}+1,r} - \psi_{\hat{j},r}| = \max_{1\le j\le r-1} |\psi_{j+1,r} - \psi_{j,r}|\right\}$$

gesetzt wird.

Insgesamt liefert die adaptive Verfeinerungsstrategie stabilere Ergebnisse als die nichtadaptiven Strategien (i) und (ii). Bei der Anwendung von Strategie (ii) erhalten wir z. B. für 6 Freiheitsgrade ein der Form nach etwas anderes Ergebnis. (Abbildungen 4.63–4.65)

Bemerkung 4.13 Bei Verwendung der monotonen kubischen Parametrisierung ist die adaptive Verfeinerungsstrategie sofort übertragbar. Im Fall einer Parametrisierung mit quadratischen B-Splines (oder B-Splines höherer Ordnung) ist dies nicht unmittelbar möglich. Im Gegensatz zu den linearen B-Splines, bei denen nach (4.6)

$$f_{r_\nu}(\psi_{j,r_\nu}) = p_{j,r_\nu} \quad \text{für} \quad j = 1, ..., \tilde{r} = r_\nu$$

gilt, entsprechen hier die Parameter p_{j,r_ν} nicht direkt den Funktionswerten in den Stützstellen ψ_{j,r_ν}. Bei der Parametrisierung mit den quadratischen B-Splines gilt

$$f_{r_\nu}(\psi_{j,r_\nu}) = \frac{1}{2}(p_{j,r_\nu} + p_{j+1,r_\nu}) \quad \text{für} \quad j = 1, ..., \tilde{r}. \tag{4.25}$$

Die Sensitivität der Funktion f_{r_ν} auf einem Interval $[\psi_{j,r_\nu}, \psi_{j+1,r_\nu}]$ ist damit bei der Parametrisierung mit quadratischen B-Splines nicht wie bei der stückweise linearen Parametrisierung durch die zu p_{j,r_ν} und p_{j+1,r_ν} gehörenden Einträge der Sensitivitätsmatrix bestimmt, sondern es muss auch die Sensitivität des Parameters p_{j+2,r_ν} berücksichtigt werden. Eine Idee zur Übertragung der adaptiven Verfeinerung auf diese Parametrisierung besteht deshalb darin (4.21) zu ersetzen durch

$$\sum_{i\in I} \left| \frac{\frac{1}{2}\left(\frac{\partial \omega^i_{\text{opt}}}{\partial p^\nu_{l+1,r_\nu}} + \frac{\partial \omega^i_{\text{opt}}}{\partial p^\nu_{l+2,r_\nu}}\right) - \frac{1}{2}\left(\frac{\partial \omega^i_{\text{opt}}}{\partial p^\nu_{l,r_\nu}} + \frac{\partial \omega^i_{\text{opt}}}{\partial p^\nu_{l+1,r_\nu}}\right)}{\psi_{l+1,r_\nu} - \psi_{l,r_\nu}} \right| =$$

$$\max_{1\le j\le r_\nu - 1} \sum_{i\in I} \left| \frac{\frac{1}{2}\left(\frac{\partial \omega^i_{\text{opt}}}{\partial p^\nu_{j+1,r_\nu}} + \frac{\partial \omega^i_{\text{opt}}}{\partial p^\nu_{j+2,r_\nu}}\right) - \frac{1}{2}\left(\frac{\partial \omega^i_{\text{opt}}}{\partial p^\nu_{j,r_\nu}} + \frac{\partial \omega^i_{\text{opt}}}{\partial p^\nu_{j+1,r_\nu}}\right)}{\psi_{j+1,r_\nu} - \psi_{j,r_\nu}} \right|. \tag{4.26}$$

D. h. f_{r_ν} und $\frac{\partial \omega^i_{\text{opt}}}{\partial p^\nu_{r_\nu}}$ werden entsprechend 4.25 zu linearen Funktionen vereinfacht und dann wie im linearen Fall behandelt.

4.3.2 Adaptivität im Fehlerfunktional

Im Multi-Level-Algorithmus minimieren wir das Funktional (4.2). Die Wichtungsfaktoren α^i_k haben wir dabei unabhängig von i (Zeitpunkt der Messung) als Skalierungsfaktoren (4.3) gewählt. Nun ist es aber so, dass infolge unterschiedlicher Sensitivitäten die einzelnen Daten $\omega^i_k, i = 1, ..., n_k$ einer Beobachtung einen unterschiedlichen Einfluss auf die Identifizierung ausüben. Im Allg. werden bei der Minimierung zunächst die simulierten Beobachtungen ω^i_k mit hohen Sensitivitäten an die zugehörigen Messwerte $\omega^i_{\varepsilon,k}$ angepasst. Wenn eine gute Anpassung der Beobachtungen mit hohen Sensitivitäten erreicht wurde, der Wert des Fehlerfunktionals jedoch weiter verringert werden soll, d. h. auch die Übereinstimmung von simulierten Beobachtungen mit geringeren Sensitivitäten mit den Messdaten soll verbessert werden, so sind größere Änderungen in den Koeffizientenfunktionen vorzunehmen. Hierdurch können sich aber auch die Beobachtungen mit hohen Sensitivitäten wieder ändern. Daher kann es zu einer Vergrößerung des Fehlers in diesen Beobachtungen kommen. Wegen der höheren Sensitivitäten kann diese Fehlerverstärkung (in Beobachtungen mit hohen Sensitivitäten) die Fehlerreduktion (in Beobachtungen mit kleinen Sensitivitäten) übersteigen. Der Wert des Fehlerfunktionals würde zunehmen. Dies hat zur Folge, dass im Multi-Level-Algorithmus anfänglich größtenteils nur Beobachtungen mit hohen Sensitivitäten das Identifizierungsergebnis bestimmen.

Da sich die Sensitivitäten während des Multi-Level-Algorithmus ändern, berechnen wir die Wichtungsfaktoren α^i_k für das Funktional (4.2) auf jeder Stufe des Multi-Level-Algorithmus neu und gelangen so zu einem *adaptiven Fehlerfunktional*. Dazu stellen wir die Sensitivitätsmatrix

$$\left(\frac{\partial \omega^i_k}{\partial p^\nu_{j,r_\nu}} \right)_{\substack{k=1,\dots,\kappa, i=1,\dots,n_k \\ \nu=1,\dots,N, j=1,\dots,r_\nu}}$$

für die Startwerte von $p^\nu_{r_\nu}, \nu = 1, ..., N$, auf und bilden für $k = 1, ..., \kappa$, $i = 1, ..., n_k$ die Gewichte α^i_k aus deren Einträgen:

$$\alpha^i_k = \left(\frac{\sum_{\nu=1}^{N} \sum_{j=1}^{r_\nu} \left| \frac{\partial \omega^i_k}{\partial p^\nu_{j,r_\nu}} \right|}{\sum_{i=1}^{n_k} \sum_{\nu=1}^{N} \sum_{j=1}^{r_\nu} \left| \frac{\partial \omega^i_k}{\partial p^\nu_{j,r_\nu}} \right|} \right)^{-1} .$$

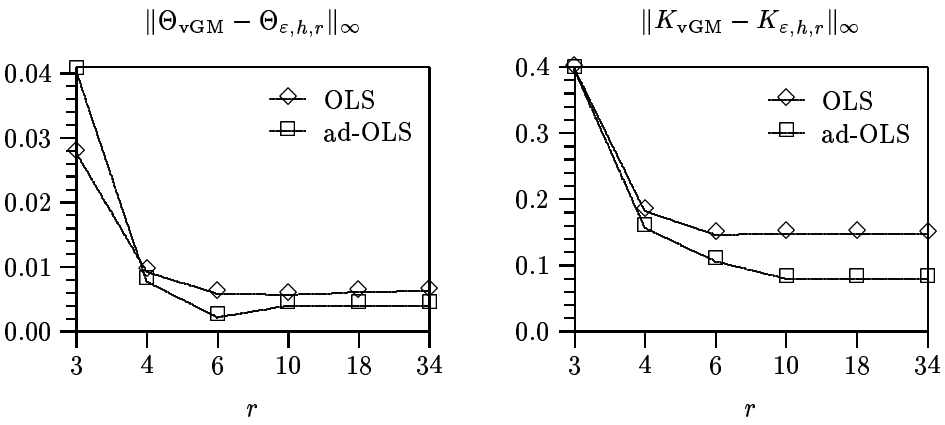


Abbildung 4.66: Identifizierungsfehler bei der Identifizierung mit adaptivem und nichtadaptivem Fehlerfunktional, $\varepsilon = 5\%$, $h = \frac{L}{100}$, $n = 50$.

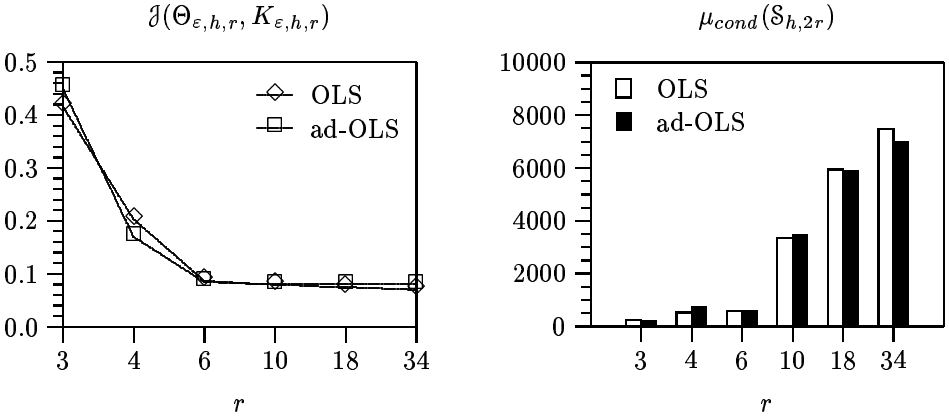


Abbildung 4.67: Fehlerfunktional und Spektralkondition bei der Identifizierung mit adaptivem und nichtadaptivem Fehlerfunktional, $\varepsilon = 5\%$, $h = \frac{L}{100}$, $n = 50$.

Wir verwenden als Gewichte also den reziproken Wert der relativen Sensitivitäten. Um eventuelle Unterschiede in den Größenordnungen der verschiedenen Beobachtungsarten auszugleichen, können wir die α_k^i noch mit den Skalierungsfaktoren (4.3) multiplizieren.

Beispiel 4.14 Wir vergleichen die Identifizierung mit dem Fehlerfunktional mit fixierten Wichtungsfaktoren (4.3) (OLS) und die Identifizierung mit dem adaptiven Fehlerfunktional (ad-OLS) an einem Beispiel mit $\varepsilon = 5\%$, $h = \frac{L}{100}$ und $n = 50$ bei quadratischer lokaler Parametrisierung. Um die Reduktion des Fehlers zwischen den simulierten und gemessenen Beobachtungen bei der Anwendung des adaptiven Fehlerfunktional zu erfassen, berechnen wir zusätzlich auch den Wert des nichtadaptiven Funktionals. Wie in Abbildung

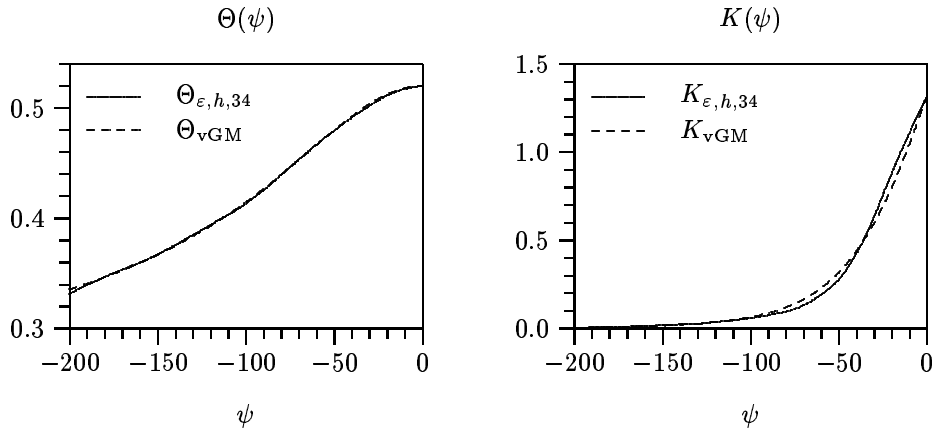


Abbildung 4.68: Hydraulische Funktionen bei adaptiven Fehlerfunktional, $r = 34$, $\varepsilon = 5\%$, $h = \frac{L}{100}$, $n = 50$.

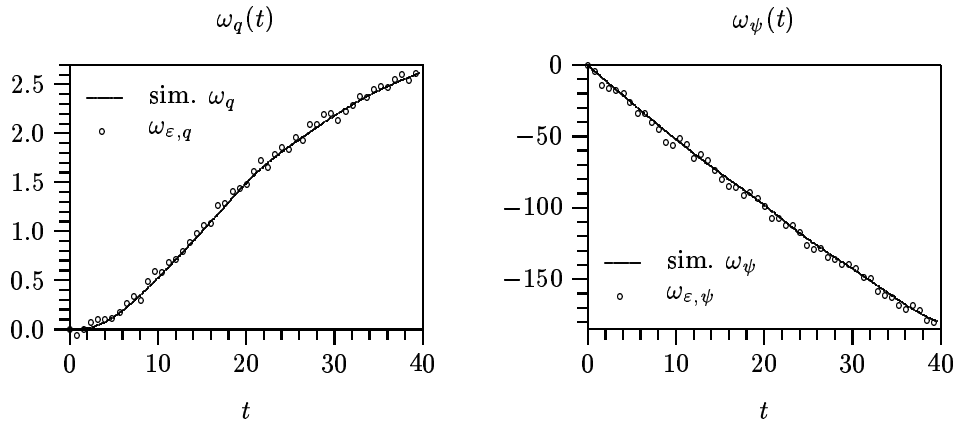


Abbildung 4.69: Beobachtungen zu Abbildung 4.68.

4.66 zu sehen ist, liefert das adaptive Fehlerfunktional deutlich kleinere Identifizierungsfehler (besonders bei der Leitfähigkeit). Außerdem treten beim adaptiven Fehlerfunktional für $r = 34$ kaum oszillatorische Störungen auf (Abbildungen 4.68 und 4.69). Für mehr als $r = 10$ Freiheitsgrade pro Funktion kommt es nicht mehr zu sichtbaren Änderungen in den Identifizierungsergebnissen.

4.4 Berücksichtigung weiterer a priori Informationen

Wenn neben einer Druckmessung im Inneren der Säule an derselben Stelle gleichzeitig die Messung des Fluidgehalts θ durchgeführt wird, so erhalten

wir Datenpaare

$$\left(\psi^i, \theta_i^*\right), \quad i = 1, ..., n,$$

die wir als a priori Schätzungen der Retentionsfunktion Θ in den Punkten ψ^i auffassen. Zur weiteren Stabilisierung der Identifizierung der Retentionsfunktion können wir damit einen Tikhonov-Term

$$\mathcal{J}_T(\Theta) := \sum_{i=1}^{n} \left|\Theta(\psi^i) - \theta_i^*\right|^2$$

zum Fehlerfunktional addieren:

$$\mathcal{J}_{\varepsilon,\alpha,h}(\Theta) := \mathcal{J}_{\varepsilon,h}(\Theta) + \alpha \mathcal{J}_T(\Theta).$$

Für die Leitfähigkeit ist dies nicht so leicht möglich. Wenn man jedoch die Mualem-Beziehung

$$\begin{aligned} K(\theta) &= K_{\text{sat}} \cdot K_{\text{res}}^*(\theta) \\ &= K_{\text{sat}} \cdot \left(\frac{\theta - \theta_{\text{res}}}{\theta_{\text{sat}} - \theta_{\text{res}}}\right)^{1/2} \left[\frac{\int_{\theta_{\text{res}}}^{\theta} \frac{1}{\psi(s)} ds}{\int_{\theta_{\text{res}}}^{\theta_{\text{sat}}} \frac{1}{\psi(s)} ds}\right]^2 \\ &\quad \text{für } \theta \in [\theta_{\text{res}}, \theta_{\text{sat}}] \end{aligned} \tag{4.27}$$

als ein realistisches Modell betrachtet, so kann man aus der Retentionsfunktion Θ eine Näherung für die Leitfähigkeit und damit einen geeigneten Strafterm für das Fehlerfunktional gewinnen. Dazu benötigen wir die zu $\Theta(\psi)$ inverse Abbildung $\psi(\theta)$ über dem Intervall $[\theta_{\text{res}}, \theta_{\text{sat}}]$. Diese existiert, da $\Theta(\psi)$ zwischen θ_{res} und θ_{sat} aus physikalischen Gründen streng monoton ist. Damit wir bei der Identifizierung der Retentionsfunktion das komplette Intervall $[\theta_{\text{res}}, \theta_{\text{sat}}]$ berücksichtigen, parametrisieren wir Θ wie bisher über ein Intervall $[\underline{\psi}, 0]$ und setzen zusätzlich entweder

(a) $\theta_{\text{res}} = \Theta\left(\underline{\psi}\right)$ oder

(b) $\theta_{\text{res}} = 0$ und approximieren die Retentionsfunktion für $\psi < \underline{\psi}$ exponentiell durch

$$\exp\left(\frac{\ln \Theta\left(\underline{\psi}\right)}{\underline{\psi}} \psi\right).$$

(Wegen $\underline{\psi} < 0$ und $\Theta\left(\underline{\psi}\right) < 1$ ist $\frac{\ln \Theta\left(\underline{\psi}\right)}{\underline{\psi}} > 0$, womit $\Theta(\psi)$ für $\psi < \underline{\psi}$ streng monoton wachsend ist.)

Dabei ist $\underline{\psi}$ so zu wählen, dass die residuale Sättigung nicht bereits für ein $\psi > \underline{\psi}$ erreicht wird.

Zum Fehlerfunktional addieren wir einen Strafterm, der die Abweichung bestimmt, die zwischen der identifizierten residualen Leitfähigkeit und der residualen Mualem-Leitfähigkeit K^*_{res} zur aktuellen Retentionsfunktion Θ besteht:

$$\mathcal{J}_{\varepsilon,\alpha,\mu}(\Theta, K) := \mathcal{J}_{\varepsilon}(\Theta, K) + \alpha \int_{\theta_{\text{res}}}^{\theta_{\text{sat}}} S_\mu \left(\frac{K(\theta)}{K_{\text{sat}}} - K^*_{\text{res}}(\theta) \right) d\theta. \tag{4.28}$$

$\alpha > 0$ spielt die Rolle des Regularisierungsparameters und $\mu \geq 0$ ist ein Parameter, der zur Festlegung eines Toleranzbereiches dient, wie z. B. im quadratischen Strafterm

$$S_\mu(x) = \begin{cases} (-\mu - x)^2 & \text{für } x < -\mu, \\ 0 & \text{für } -\mu \leq x \leq \mu, \\ (x - \mu)^2 & \text{für } x > \mu. \end{cases}$$

Wegen der strengen Monotonie von $\Theta(\psi)$ kann durch Werte

$$\underline{\psi} = \psi_0 < \psi_1 < \ldots < \psi_{n-1} < \psi_n$$

mit $\psi_n := \min\{\psi \mid \Theta(\psi) = \theta_{\text{sat}}\}$ eine Zerlegung

$$\theta_{\text{res}} \leq \Theta\left(\underline{\psi}\right) = \Theta\left(\psi_0\right) < \Theta\left(\psi_1\right) < \ldots < \Theta\left(\psi_n\right) = \theta_{\text{sat}}$$

von $[\theta_{\text{res}}, \theta_{\text{sat}}]$ definiert werden. Auf dem Intervall $[\underline{\psi}, \psi_n]$ ist $K(\psi)$ ebenfalls streng monoton und es gilt für $\psi \in [\underline{\psi}, \psi_n]$

$$K(\theta) = K(\psi) \qquad \Longleftrightarrow \qquad \theta = \Theta(\psi).$$

Die diskrete Version von (4.28) lautet damit

$$\mathcal{J}_{\varepsilon,\alpha,\mu,h}(\Theta, K) := \mathcal{J}_{\varepsilon,h}(\Theta, K) + \alpha \sum_{i=1}^{n} \gamma_i S_\mu \left(\frac{K\left(\psi_i\right)}{K(0)} - K^*_{\text{res}}\left(\Theta\left(\psi_i\right)\right) \right) \tag{4.29}$$

mit Integrationsgewichten γ_i, die z. B. Eins gesetzt werden können.

Die bei der Berechnung von $K^*_{\text{res}}\left(\Theta\left(\psi_i\right)\right)$ auszuwertenten Intergrale können durch eine Quadraturformel approximiert werden, z. B.

$$\begin{aligned} \int_{\theta_{\text{res}}}^{\Theta(\psi_i)} \frac{1}{\psi(s)} ds &= I_0 + \sum_{j=1}^{i} \int_{\Theta(\psi_{j-1})}^{\Theta(\psi_j)} \frac{1}{\psi(s)} ds \\ &\approx I_0 + \sum_{j=1}^{i} \sum_{k=1}^{m_j} \frac{\Theta\left(\psi_{j_k}\right) - \Theta\left(\psi_{j_{k-1}}\right)}{2} \left(\frac{1}{\psi_{j_k}} + \frac{1}{\psi_{j_{k-1}}} \right) \end{aligned}$$

mit

$$I_0 := \begin{cases} 0 & \text{für Fall (a),} \\ \int_0^{\Theta(\underline{\psi})} \frac{\ln \Theta(\underline{\psi})}{\underline{\psi} \ln s} ds & \text{für Fall (b)} \end{cases}$$

bei Zerlegungen

$$\psi_{j-1} = \psi_{j_0} < \psi_{j_1} < ... < \psi_{j_{m_j}} = \psi_j$$

der Intervalle $[\psi_{j-1}, \psi_j]$, $j = 1, ..., n$ bzw. der zugehörigen θ-Zerlegungen von $[\Theta(\psi_{j-1}), \Theta(\psi_j)]$.

Bei gleichzeitiger Identifizierung von Leitfähigkeit und Retentionsfunktion ist in den Fehlerfunktionalen (4.28) und (4.29) durch die Mualem-Beziehung auch eine Rückwirkung von der Leitfähigkeit zur Retentionsfunktion vorhanden. Durch die Steuerung der beiden Parameter α und μ können wir gezielten Einfluss auf die Identifizierung ausüben. Obwohl man durch obige Strafmethode wieder zu einer starren Form der Parametrisierung zurückkehrt, ist diese Methode geeignet, damit sich das Krümmungsverhalten der hydraulischen Funktionen bereits in den ersten Schritten des Multi-Level-Algorithmus richtig herausbildet. Mit dem weiteren Fortschreiten im Multi-Level-Algorithmus kann durch eine Verkleinerung des Regularisierungsparameters α der Einfluss des Strafterms verringert werden. Durch den Parameter μ kann außerdem die Breite der zulässigen Abweichung vom Mualem-Modell gesteuert werden. Wenn das Mualem-Modell nicht geeignet erscheint, dann kann für $K^*_{\text{res}}(\theta)$ auch ein anderes Modell verwendet werden.

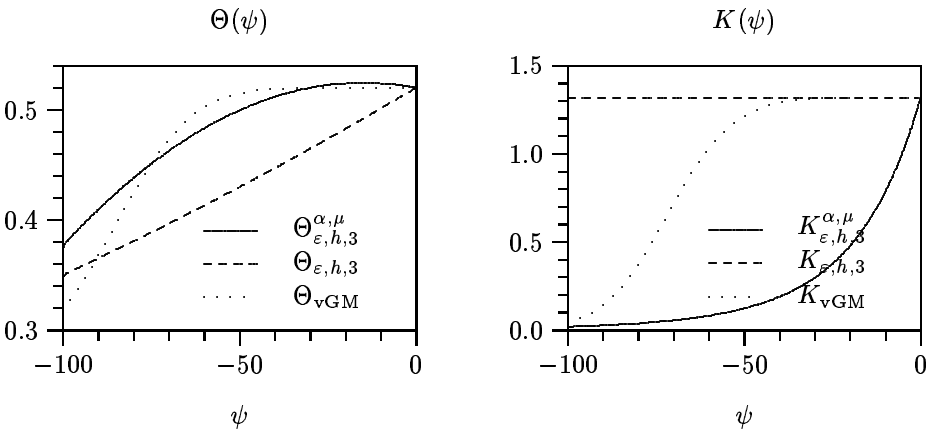


Abbildung 4.70: Hydraulische Funktionen, $r = 3$, $\varepsilon = 5\%$, $h = \frac{L}{100}$, $n = 50$.

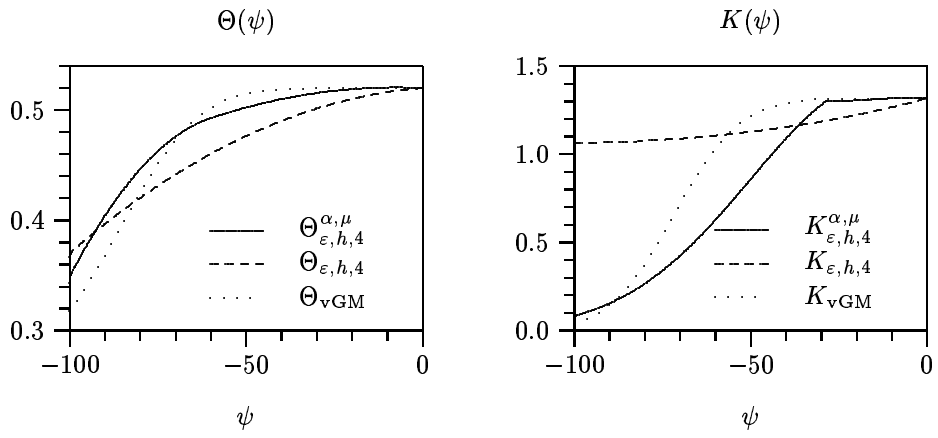


Abbildung 4.71: Hydraulische Funktionen, $r = 4$, $\varepsilon = 5\%$, $h = \frac{L}{100}$, $n = 50$.

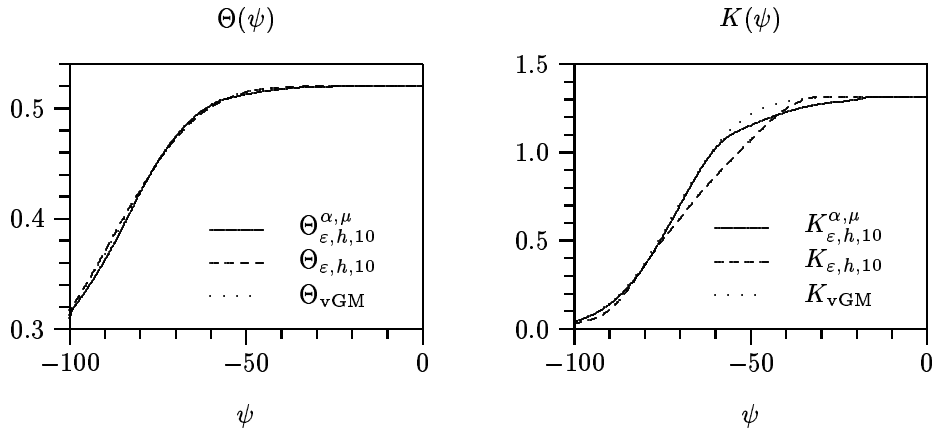


Abbildung 4.72: Hydraulische Funktionen, $r = 10$, $\varepsilon = 5\%$, $h = \frac{L}{100}$, $n = 50$.

Beispiel 4.15 Wir betrachten dazu ein Beispiel, bei dem mit den bisherigen Methoden die Identifizierungsergebnisse noch große Fehler aufweisen. Durch die Verwendung des Fehlerfunktionals (4.29) mit $\mu = 0.1$ und

$$\begin{aligned} \alpha &= 1.0 \quad \text{für } r = 3 \\ \alpha &= 0.5 \quad \text{für } r = 4 \\ \alpha &= 0.1 \quad \text{für } r = 6 \\ \alpha &= 0.05 \quad \text{für } r = 10 \end{aligned}$$

haben sich die Ergebnisse der Identifizierung verbessert (siehe Abbildungen 4.70 - 4.72). Dabei ist zu beachten, daß wir als Original eine Mualem-Leitfähigkeit zugrunde gelegt haben. Die Parametrisierung erfolgte quadratisch lokal für Θ (mit Fall (b)) und $\ln K$.

Eine weitere Anwendungsmöglichkeit der oben beschriebenen Strafmethode zeigt das folgende Beispiel.

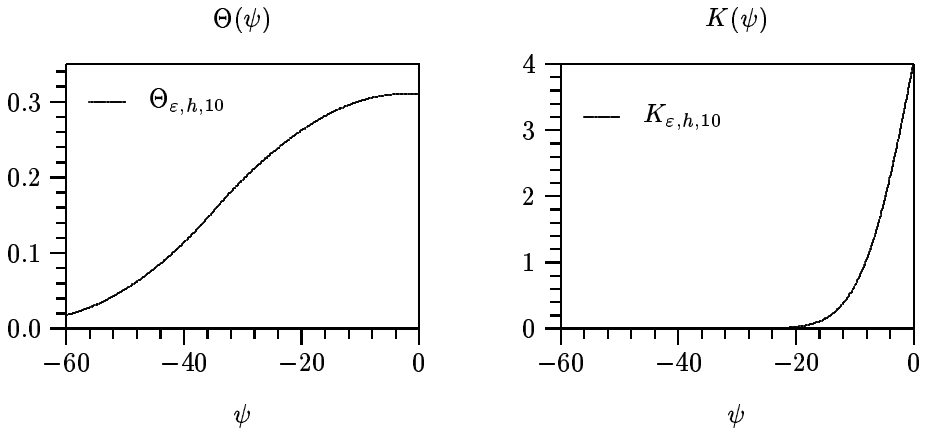


Abbildung 4.73: Hydraulische Funktionen für BSL, $r = 10$, $h = \frac{L}{100}$, $n = 157$, ohne Berücksichtigung des Druckes im Fehlerfunktional.

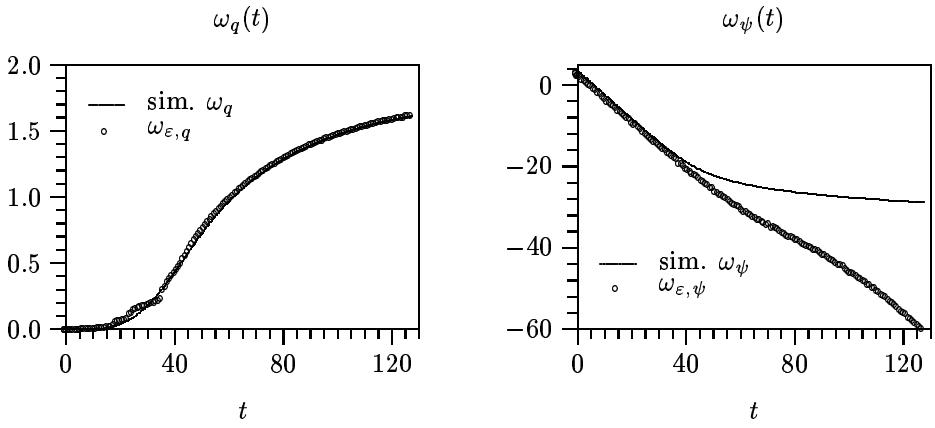


Abbildung 4.74: Beobachtungen zu Abbildung 4.73.

Beispiel 4.16 Am Beispiel des im Unterabschnitt 4.2.3 betrachteten Bayreuther sandigen Lehms (BSL) können wir demonstrieren, dass es bei den hier verwendeten polynomialen Parametrisierungen nicht aussreicht, nur den kumulativen Ausfluss im Fehlerfunktional zu berücksichtigen. Die Abbildungen 4.73 und 4.74 zeigen, dass wir in diesem Fall völlig andere Identifizierungsresultate erhalten. Obwohl der kumulative Ausfluss gut angepasst wird, stimmt die jetzige identifizierte Retentionsfunktion nicht mit derjenigen überein, die wir bei Identifizierung mit zusätzlichen Druckdaten erhalten. Entsprechend deutlich sind auch die Unterschiede zwischen dem gemessenen Druck und dem simulierten Druck. Wenn wir aber die oben beschriebene Mualem-Regularisierung (mit Fall (b)) in die Identifizierung einbeziehen, so gelangen wir näherungsweise zu den Ergebnissen, wie sie die Identifizierung mit Druckdaten liefert (vgl. Abbildungen 4.75 und 4.76).

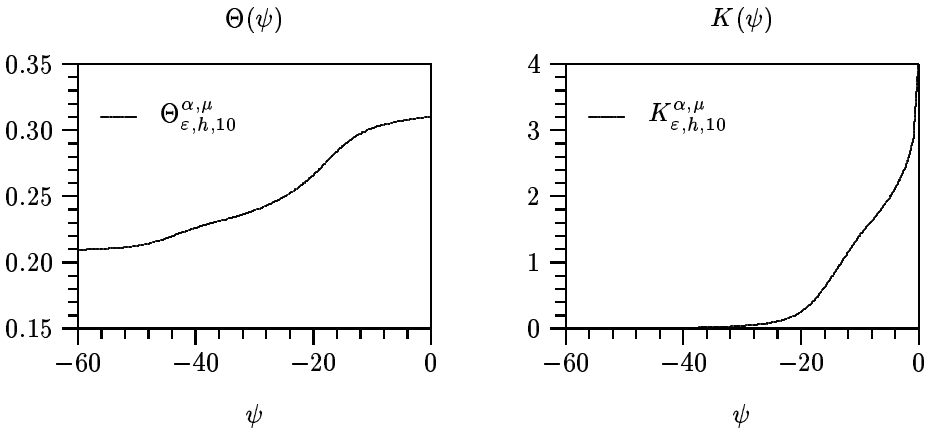


Abbildung 4.75: Hydraulische Funktionen für BSL, $r = 10$, $h = \frac{L}{100}$, $n = 157$, Fehlerfunktional mit Mualem-Strafterm ($\alpha = 0.25$, $\mu = 0.1$) und ohne Berücksichtigung des Druckes.

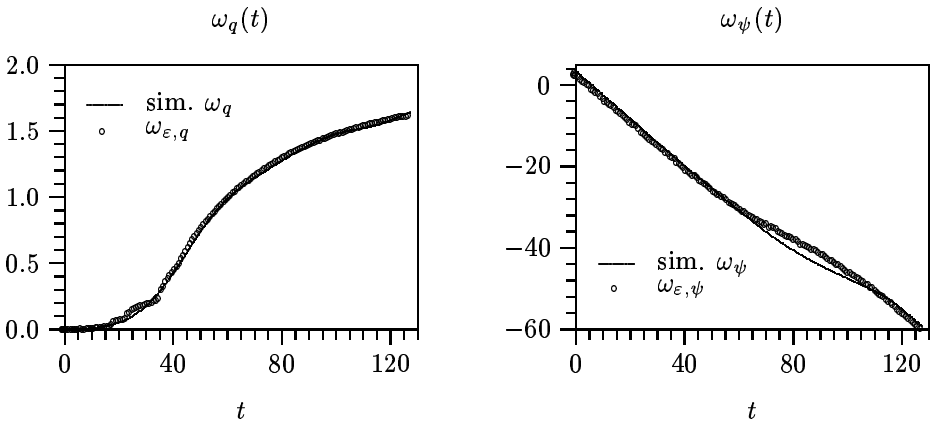


Abbildung 4.76: Beobachtungen zu Abbildung 4.75.

4.5 Zusammenfassung der Ergebnisse

Abschließend wollen wir noch einmal die wesentlichsten Erkenntnisse aus den betrachteten Beispielen zusammentragen:

- Wenn eine hinreichende Feinheit der räumlichen Diskretisierung erreicht ist, so ist deren Einfluss auf die Ergebnisse der Identifizierung vernachlässigbar. Eine weitere Verfeinerung der räumlichen Diskretisierung führt zu keiner wesentlichen Verbesserung der Stabilität des inversen Problems.

- Ein entscheidender Einflussfaktor auf die „Güte“ der Identifizierungsergebnisse ist die Anzahl der Messdaten (sowie deren Verteilung im

Wertebereich). Je geringer die Anzahl der Daten ist, desto weniger Informationen liefern diese über die hydraulischen Funktionen und desto stärker ist der Einfluss möglicher Datenfehler auf die Identifizierung.

- Die Ergebnisse der Identifizierung werden teilweise von der Art der verwendeten Parametrisierung beeinflusst. Welche Parametrisierung geeignet ist, kann vom betrachteten Bodentyp abhängen. Im Allg. sind höhere Ansätze (quadratische oder kubische Splines) gegenüber einer Parametrisierung mit linearen Splines vorteilhafter, insbesondere da sie stärkere Glattheitseigenschaften besitzen.

- Wenn anstelle der Leitfähigkeit der Logarithmus der Leitfähigkeit parametrisiert wird, so kann häufig eine Verbesserung der Sensitivität erreicht werden.

- Durch adaptive Elemente im Multi-Level-Algorithmus, die die aktuellen Sensitivitäten der Beobachtungen bezüglich der einzelnen Parameter berücksichtigen, kann eine größere Stabilität der Identifizierung gewonnen werden. Hierdurch ist eine Verringerung der Identifizierungsfehler möglich.

- Die Verwendung zusätzlicher, geeigneter Strafterme im Fehlerfunktional erlaubt die Identifizierung der hydraulischen Funktionen auch dann, wenn nur reduzierte Messdaten vorliegen.

- Die betrachteten Beispiele zeigen deutlich, dass bei den hier benutzten Parametrisierungstechniken (und Experimentdesign) keine Aussicht besteht, den Wert der Leitfähigkeit im Bereich der Sättigung (K_{sat}-Wert) mitzubestimmen. Dieser muss also a priori vorgegeben werden.

Kapitel 5

Betrachtungen zum optimalen Experimentdesign

5.1 Motivation

In dem Bereich der Anwendungswissenschaften, in dem man sich mit der Identifizierung von Materialeigenschaften beschäftigt, tritt immer mehr das Experimentdesign in den Mittelpunkt. Man begnügt sich nicht mehr damit ein Experimentdesign zur Verfügung zu haben, welches die Identifizierung der gewünschten Materialeigenschaften in einem gewissen Toleranzbereich ermöglicht, sondern man versucht dieses Experimentdesign zu optimieren. Diese Optimierung bezieht sich auf die Erhöhung des Informationsgehaltes - der Indikativität, wie es in [59] genannt wird - des Experiments bezüglich der zu identifizierenden Materialgrößen. Dabei kann es auch darum gehen, Effekte aus Prozessen, die im mathematischen Modell keine Berücksichtigung finden, weitestgehend zu eliminieren. Oder es wird versucht die technischen Apparaturen zu verbessern, um eine Optimierung der Messmethoden zu erreichen.

Bei den in dieser Arbeit betrachteten Säulenexperimenten haben wir verschiedene Möglichkeiten das Experimentdesign zu verändern. Eine Möglichkeit besteht darin, zu einem neuen Messplan überzugehen. Andererseits können auch die Anfangs- und Randbedingungen variiert werden. So wird z. B. in [11] ein Säulenexperiment betrachtet, bei dem die Gravitation die treibende Kraft ist. Jedoch erlaubt diese Experiment die Identifizierung der hydraulischen Funktionen maximal nur für einen Druck aus dem Bereich $[-L, 0]$, wobei L der Länge der Säule entspricht (vgl. Lemma 1.4 in [11]).

Den häufigsten Ansatzpunkt im Experimentdesign bieten die Randbedingungen, wobei die Bedingung $g'(t) < 0$ (notwendig für den Beweis der Iden-

tifizierbarkeit) vernachlässigt wird. Das im Abschnitt 4.2 betrachtete Experimentdesign wird in der Hydrologie als *kontinuierliches* Ausflussexperiment bezeichnet. Ein anderes Experimetdesign verwenden die *Onestep*- und *Multistep*-Verfahren, bei denen der am Ausflussrand angelegte Druck $g(t)$ in einer bzw. mehreren Stufen abgesenkt wird. Der Dirichlet-Rand wird hierbei durch eine stückweise konstante Funktion $g(t)$ beschrieben. Diese drei Methoden sind Gegenstand der Untersuchungen in [13], [14], [15], [53] und [54]. In diesen Arbeiten werden die hydraulischen Funktionen mit dem van Genuchten-Mualem-Modell oder Modifikationen diese Modells (bimodale Retentionsfunktion) parametrisiert. Es hat sich herausgestellt, dass die Onestep-Methode ungeeignet, die Multistep- und die kontinuierliche Methode dagegen gleichermaßen geignet sind zur Identifizierung der hydraulischen Parameter. Andererseits führt die Methode mit linearer Veränderung des Druckes am Ausflussrand zu den geringsten Unsicherheiten bei der Parameterbestimmung. Mit den in dieser Arbeit verwendeten Parametrisierungen gelangen wir zu ähnlichen Aussagen. Wir wollen hierzu zwei Beispiele betrachten.

Beispiel 5.1 Zunächst konstruieren wir wie für das kontinuierliche Experiment synthetische Daten der Druck- und Ausflussbeobachtungen mit $\varepsilon = 5\%$ für ein Multistep-Ausflussexperiment (Säulenlänge $L = 15$ cm) bei dem der Druck am Ausflussrand in 5 Stufen von 15 cm auf -100 cm innerhalb des Zeitintervalls $[0, 20]$ abgesenkt wird. Die hydraulischen Funktionen, die für die Simulation verwendet wurden, entsprechen denen in den Abbildungen 4.4–4.15. Bei der Identifizierung verwenden wir eine quadratische lokale Spline-Parametrisierung der hydraulischen Funktionen.

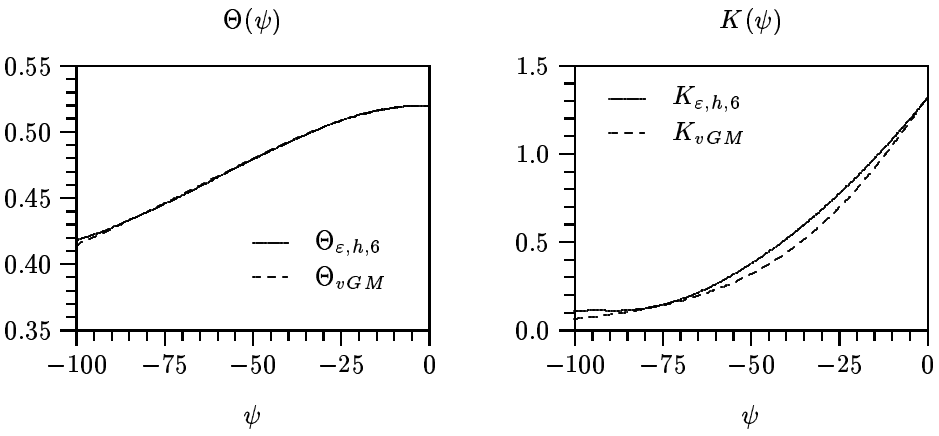


Abbildung 5.1: Hydraulische Funktionen für Multistep-Experiment, $r = 6$, $\varepsilon = 5\%$, $h = \frac{L}{100}$, $n = 50$.

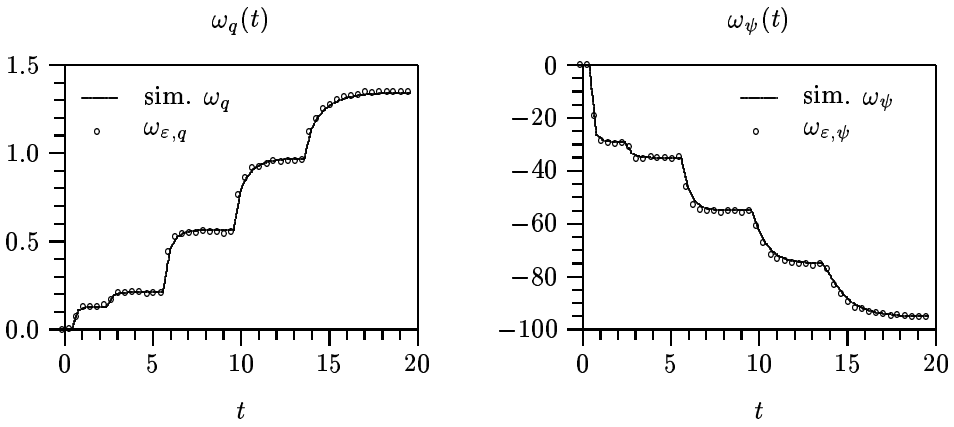


Abbildung 5.2: Beobachtungen zu Abbildung 5.1.

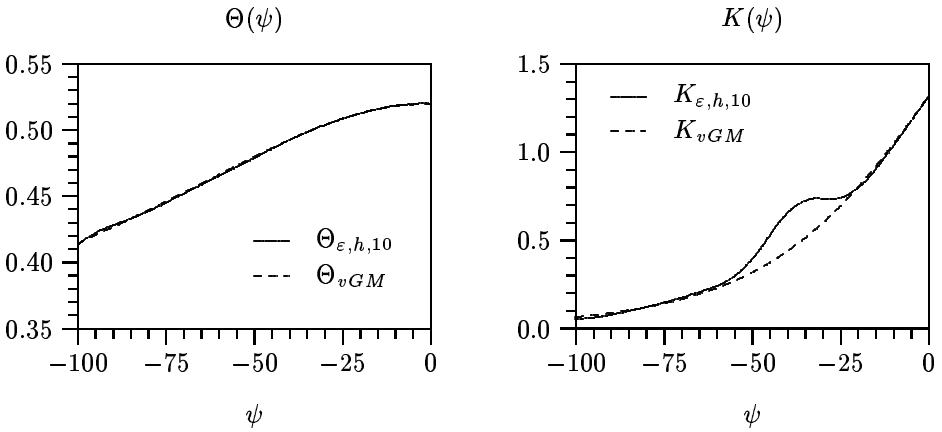


Abbildung 5.3: Hydraulische Funktionen für Multistep-Experiment, $r = 10$, $\varepsilon = 5\%$, $h = \frac{L}{100}$, $n = 50$.

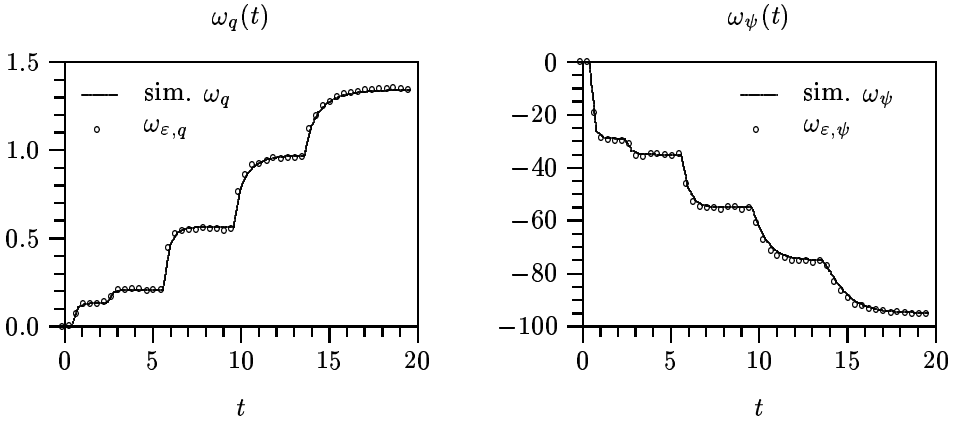


Abbildung 5.4: Beobachtungen zu Abbildung 5.3.

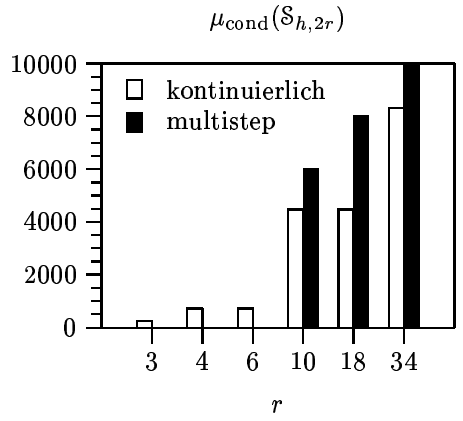


Figure 5.5: Vergleich der Spektralkonditionen für kontinuierliches und Multistep-Ausflussexperiment.

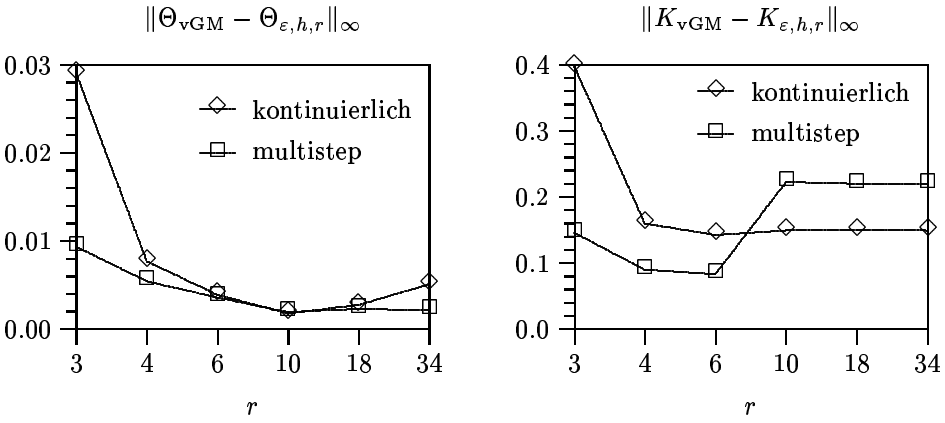


Abbildung 5.6: Vergleich der Identifizierungsfehler für kontinuierliches und Multistep-Ausflussexperiment

In den ersten Schritten des Multi-Level-Algorithmus erhalten wir ähnliche Ergebnisse wie beim kontinuierlichen Ausflussexperiment. Als Beispiel sind in den Abbildungen 5.1 und 5.2 die Ergebnisse für 6 Freiheitsgrade pro Funktion angegeben. Auf der nächsten Parametrisierungsstufe (10 Freiheitsgrade) wird jedoch ein deutlicher Identifizierungsfehler in der hydraulischen Leitfähigkeit produziert (Abbildungen 5.3 und 5.4). Dieser bleibt auch in den weiter Schritten des Multi-Level-Algorithmus erhalten. Die Spektralkondition der Sensitivitätsmatrix nimmt für $r \geq 10$ z. T. deutlich höhere Werte an als im kontinuierlichen Ausflussexperiment, während die Werte für $r < 10$ kleiner als im kontinuierlichen Experiment sind (vgl. Abbildung 5.5). Ent-

sprechend sind im Multistep-Experiment auch die Identifizierungsfehler der hydraulischen Funktionen für $r < 10$ kleiner, für $r \geq 10$ liefert das kontinuierliche Experiment kleinere Identifizierungsfehler in der hydraulischen Leitfähigkeit (vgl. Abbildung 5.6).

Aus diesem und weiteren betrachteten Beispielen folgern wir, dass einerseits die Multistep-Methode zu verbesserten Identifizierungsergebnissen führen kann, andererseits aber mögliche Datenfehler früher im Multi-Level-Algorithmus stärkeren Einfluss auf die Identifizierung ausüben können.

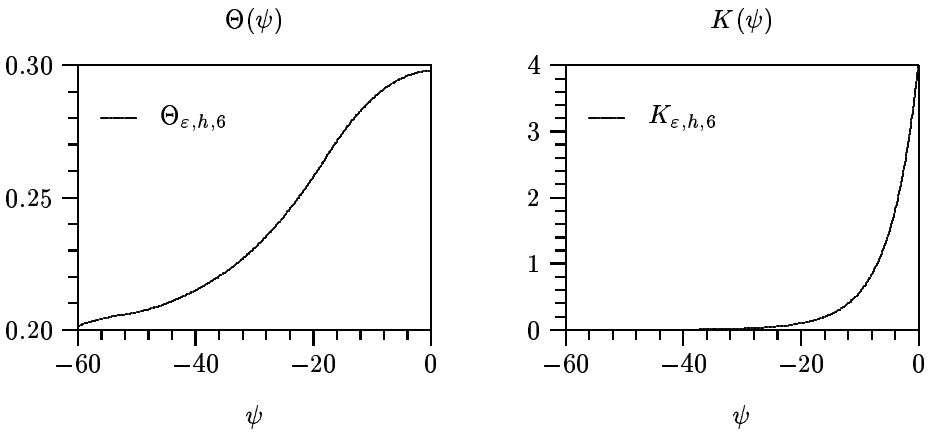


Abbildung 5.7: Hydraulische Funktionen für BSL, multistep, $r = 6$, $h = \frac{L}{100}$, $n = 146$.

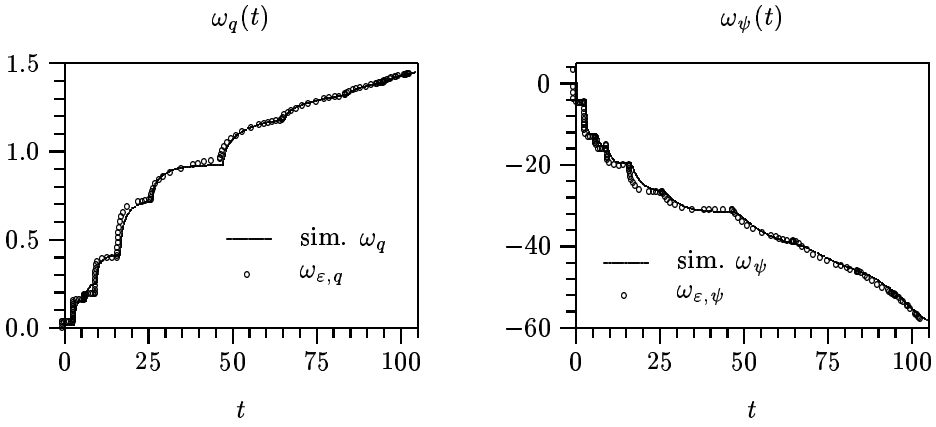


Abbildung 5.8: Beobachtungen zu Abbildung 5.7.

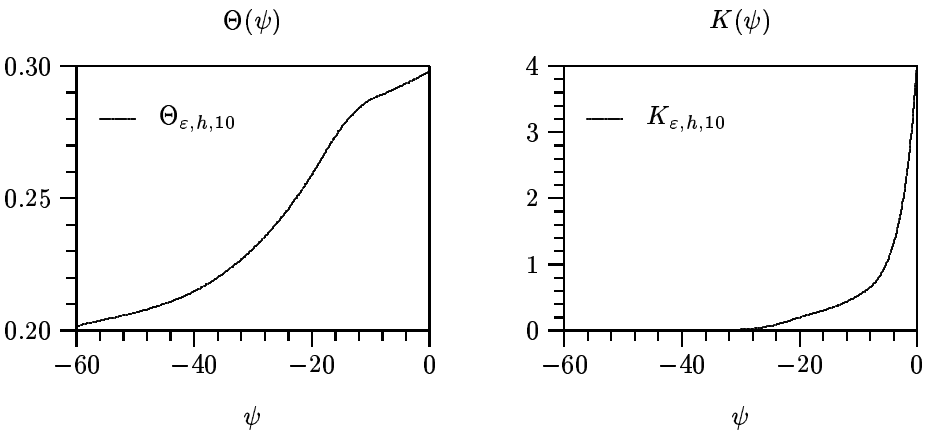


Abbildung 5.9: Hydraulische Funktionen für BSL, multistep, $r = 10$, $h = \frac{L}{100}$, $n = 146$.

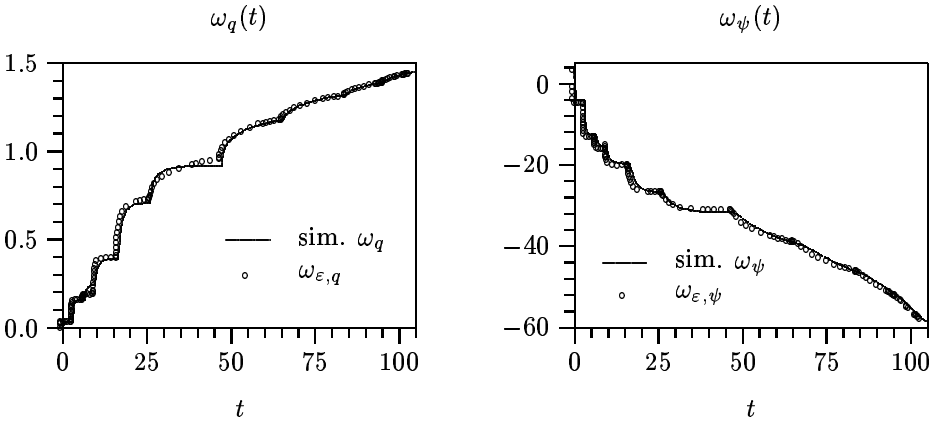


Abbildung 5.10: Beobachtungen zu Abbildung 5.9.

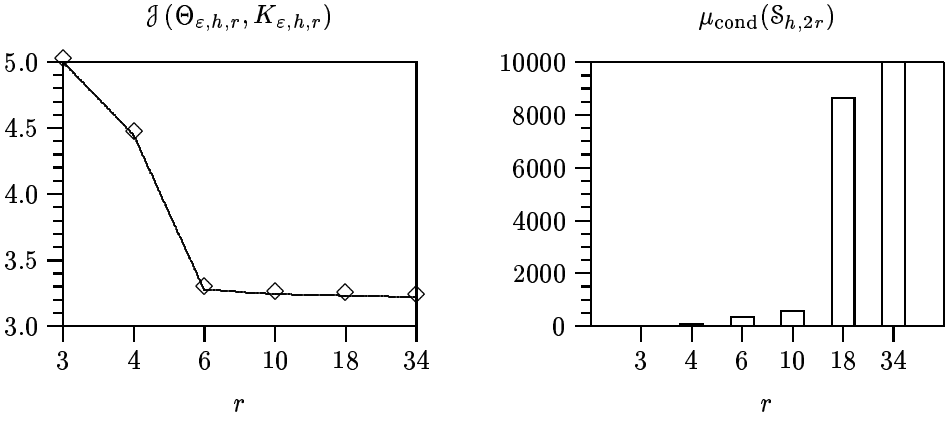


Abbildung 5.11: Fehlerfunktional und Spektralkondition für BSL, multistep.

Beispiel 5.2 Als weiteres Beispiel untersuchen wir ein 10-Stufen-Multistep-Ausflussexperiment für den bereits in Unterabschnitt 4.2.3 betrachteten Bayreuther sandigen Lehm (BSL). Auch hier wurde wieder die quadratische lokale Parametrisierung der Retentionsfunktion und des Logarithmus der Leitfähigkeit gewählt. Die Identifizierungsergebnisse für $r = 6$ und $r = 10$ sowie die Werte des Fehlerfunktionals und der Spektralkondition sind in den Abbildungen 5.7–5.11 dargestellt. Wir stellen fest, dass gegenüber dem Experiment mit dem kontinuierlichen Ausfluss die Werte der Spektralkondition der Sensitivitätsmatrix für $r \leq 18$ hier deutlich kleiner sind. Obwohl die Werte des Fehlerfunktionals für verschiedene Experimente untereinander nicht vergleichbar sind (unterschiedliche Anzahl und Verteilung der Messdaten), ist der Wert gegen den das Fehlerfunktional strebt hier doch signifikant größer (ca. um den Faktor 100) als im Fall des kontinuierlichen Ausflusses. Dass dies nicht allein durch die Messdaten verursacht wird, ist aus der Anpassung des kumulativen Ausflusses ersichtlich. Es treten teilweise große Abweichungen des simulierten kumulativen Ausflusses von den Messdaten auf (Abbildung 5.10), welche auch im weiteren Verlauf des Multi-Level-Algorithmus nicht wesentlich verringert werden. Eine offene Frage ist, ob es sich bei diesen Abweichungen um Messfehler handelt oder diese durch fehlende Flexibilität der Parametrisierung verursacht werden.

5.2 Problemformulierung

Das Ergebnis des Experimentdesigns wird davon abhängen, welches Ziel für die Untersuchungen vorgegeben wird. In der mathematischen Statistik sind im Zusammenhang mit der optimalen Versuchsplanung verschiedene Optimalitätsbegriffe eingeführt worden, so z. B. die A-, D- oder E-Optimaliät ([3], [19]). Die bei diesen Optimalitätskriterien wesentliche Größe ist die Kovarianzmatrix oder die Informationsmatrix.

In den Beispielen des letzten Kapitels haben wir die Sensitivitätsmatrix dazu verwendet Aussagen über die Schlechtgestelltheit des inversen Problems zu gewinnen. Da das Ziel des Experimentdesigns die Maximierung der Zuverlässigkeit der bestimmten Parameter bzw. die Minimierung der Schlechtgestelltheit sein wird, werden wir das Zielfunktional aus der Sensitivitätsmatrix bilden und mit

$$\mu(\mathcal{S}_{h,\hat{r}}) \tag{5.1}$$

bezeichnen. Geeignet für μ sind die Spektralkondition und die Maximum-Charakteristik. Auch andere Größen wie z. B. die Determinante oder die Spur der Sensitivitätsmatrix wären denkbar.

Die Sensitivitätsmatrix hängt von der Wahl des Experimentszenarios ab, d. h.

$$\mathcal{S}_{h,\hat{r}} = \mathcal{S}_{h,\hat{r}}(\xi) \qquad \text{mit } \xi \in \Xi.$$

Dabei bezeichnet Ξ die Menge aller zulässigen Experimentszenarien. Diese Menge wird durch die Festlegung von Kontroll- bzw. Optimierungsvariablen genauer spezifiziert. Dies erfolgt jeweils in den später betrachteten Beispielen.

Da die Identifizierung der hydraulischen Funktionen mit einem Multi-Level-Verfahren erfolgt und nicht mit einer festen Anzahl von Parametern in den hydraulischen Funktionen gearbeitet wird, ist eine zusätzliche Abhängigkeit der Zielfunktion (5.1) von der Anzahl der Parameter (Stufe der Parametrisierung) gegeben. Für ein gegebenes Szenario $\xi \in \Xi$ betrachten wir die Anzahl $\hat{r}$ der Parameter mit welcher der Multi-Level-Algorithmus abbricht (entsprechend der Wahl von $r_{\max}$ und μ_{tol}) als die „optimale“ Anzahl von Parametern für das Szenario ξ:

$$\hat{r}_{\text{opt}} = \hat{r}_{\text{opt}}(\xi).$$

Zur Vereinfachung ist es auch möglich $\hat{r}_{\text{opt}}$ unabhängig von ξ als Konstante vorzugeben.

Beim *Problem des optimalen Experimentdesigns* ist damit ein Experimentszenario $\xi^* \in \Xi$ gesucht, welches der Bedingung

$$\mu\left(\mathcal{S}_{h,\hat{r}_{\text{opt}(\xi^*)}}(\xi^*)\right) = \inf_{\xi\in\Xi} \mu\left(\mathcal{S}_{h,\hat{r}_{\text{opt}(\xi)}}(\xi)\right) \tag{5.2}$$

genügt.

Der Typ der Parametrisierung ist vorher festzulegen und darf während des Experimentdesigns nicht geändert werden. Denn unterschiedliche Arten der Parametrisierung können auch zu unterschiedlichen Resultaten im Experimentdesign führen.

Bemerkung 5.3 Wenn die Messdaten ω_ε dem Gaußschen Fehlermodell (siehe Unterabschnitt 4.1.1) genügen und im Fehlerfunktional (4.2) die Kehrwerte der Varianzen als Wichtungsfaktoren α_k^i $(k = 1, ..., \kappa,\ i = 1, ..., n_k)$ verwendet werden, so wissen wir, dass im Sinne der mathematischen Statistik der aus der Minimierung diese Fehlerfunktionals erhaltene Parametervektor $(p^1_{\varepsilon,h,r_1}, ..., p^M_{\varepsilon,h,r_M})$ eine *Maximum-Likelihood-Schätzung* darstellt (siehe z. B. [4]). Als Zufallsvektor aufgefasst ist dieser asymptotisch erwartungstreu und normalverteilt. Für die Kovarianzmatrix erhalten wir aus der Linearisierung gemäß Unterabschnitt 4.1.5 (vgl. z. B. auch [4])

$$\text{Cov}\left(p^1_{\varepsilon,h,r_1}, ..., p^M_{\varepsilon,h,r_M}\right) \approx \left(\mathcal{S}^T_{h,\hat{r}}\mathcal{S}_{h,\hat{r}}\right)^{-1}$$

(unter Annahme des vollen Ranges der Sensitivitätsmatrix). Die inversen Quadrate der Singulärwerte der Sensitivitätsmatrix approximieren also die Eigenwerte der Kovarianzmatrix. Damit gelangen wir zur E-Optimalität, wenn wir im Experimentdesign die Maximum-Charakteristik der Sensitivitätsmatrix als Zielfunktional verwenden.

In den nachfolgenden Beispielen wird jeweils mit der gewichteten Sensitivitätsmatrix gerechnet, auch wenn dies nicht explizit erwähnt wird. Die Gewichte sind dabei gemäß (4.3) gewählt. Prinzipiell bleibt das Verhalten der Spektralkondition und der Maximum-Charakteristik gleich, wenn die gewichtete Senitivitätsmatrix betrachtet wird. Geringfügige Unterschiede treten dennoch auf, insbesondere kommt es bei einer große Anzahl von Parametern z. T. zu erhebliche Unterschieden in der Größenordnung.

Eine besondere Schwierigkeit der Formulierung (5.2) besteht darin, dass die dort benutzte Zielfunktion im Allg. in ξ unstetig sein wird. Die Unstetigkeit wird durch die Funktion

$$\hat{r}_{\text{opt}} : \Xi \to \mathbb{N}_+$$

verursacht, die einen diskreten Wertebereich besitzt und bei der es sich um eine Sprungfunktion handelt. Oft wird $\hat{r}_{\text{opt}(\xi)}$ eine stückweise konstante Funktion sein. Damit ist die Stetigkeit (bzw. auch stärkere Glattheit) der Funktionen $\mu(.)$ und $\mathcal{S}_{h,\hat{r}}(.)$ nicht ausreichend für die Stetigkeit von $\mu\left(\mathcal{S}_{h,\hat{r}_{\text{opt}(\xi)}}(\xi)\right)$ in ξ. Die Glattheit der Zielfunktion in (5.2) wird wesentlich durch die Glattheit der Funktion $\hat{r}_{\text{opt}(\xi)}$ bestimmt. Stetigkeit der Zielfunktion wird nur dann zu erwarten sein, wenn $\hat{r}_{\text{opt}(\xi)}$ konstant ist. Diese Problematik wird auch durch das nachfolgende Beispiel illustriert.

Beispiel 5.4 In einem ersten Beispiel wollen wir untersuchen, welchen Einfluss die Position der Druckbeobachtung auf die Identifizierung der hydraulischen Funktionen ausübt. Wie bereits erwähnt, kann der Druck außer am oberen Rand der Säule auch an einer beliebigen Stelle im Inneren der Säule gemessen werden. Um die Abhängigkeit der Druckbeobachtung ω_ψ von der Ortskoordinate x_d, an der die Messung stattfindet, zu charakterisieren, schreiben wir $\omega_\psi = \omega_\psi(x_d)$. Diese Koordinate x_d ist hier die Optimierungsvariable und die Menge der zulässigen Experimentszenarien wird geschrieben als

$$\Xi_d := \{\{(\omega_\psi(x_d), \omega_q)\} \mid 0 \leq x_d < L\}\,. \tag{5.3}$$

Die hybrid-gemischte Finite-Elemente-Methode, die zur Lösung des direkten Problems eingesetzt wird, approximiert den Druck konstant auf jedem Element. Zusätzlich sind im eindimensionalen Fall durch die Lagrange-Multiplikatoren noch Druckdaten in den Knoten gegeben. Wenn eine äquidistante

Anordnung der Knoten mit der Schrittweite h zugrundegelegt wird, haben wir insgesamt $2\frac{L}{h}+1$ (L =Länge der Säule) Werte für den Druck und die Menge der zulässigen Experimentszenarien kann im diskreten Fall eingeschränkt werden auf

$$\Xi_{d,h} := \left\{ \{(\omega_\psi(x_d), \omega_q)\} \mid x_d = i \cdot \frac{h}{2}, i = 0, ..., \frac{2L}{h} - 1 \right\}. \tag{5.4}$$

(Der rechte Rand, d. h. $i = \frac{2L}{h}$, wird in (5.4) nicht berücksichtigt, da dort der Druck vorgegeben wird.) Um eine glatte Approximation für den Druck zu erhalten und (5.3) beibehalten zu können, muss entweder ein höherer Diskretisierungsansatz benutzt werden oder mit den Lagrange-Multiplikatoren eine stückweise lineare Darstellung über die gesamte Säule gebildet werden. Wenn die räumliche Diskretisierung jedoch fein genug gewählt wird, bringt dies keine wesentlichen Änderungen und wir beschränken uns auf die Menge $\Xi_{d,h}$ gemäß (5.4).

Insbesondere haben wir damit eine endliche Anzahl von Experimentszenarien und wir benötigen keinen Optimierungsalgorithmus zur Lösung von (5.2), sondern sämtliche Szenarien können mit einem vertretbaren Zeitaufwand nacheinander abgearbeitet werden.

Wir beziehen uns auf das bereits mehrfach betrachtete Testbeispiel mit der quadratischen lokalen Parametrisierung und einer räumlichen Diskretisierung mit 100 Elementen. Die Abbildung 5.12 zeigt die Entwicklung der Spektralkondition und der Maximum-Charakteristik im Multi-Level-Algorithmus für 6 verschiedene Positionen der Druckbeobachtung. Beide Größen weisen qualitativ das gleiche Verhalten auf. Die Werte dieser Größen wachsen unabhängig von der Anzahl der Freiheitsgrade mit der Verringerung des Abstandes der Koordinate x_d zum Ausflussrand.

In Abbildung 5.13 sind die Werte der Zielfunktion $\mu\left(\mathcal{S}_{h,\hat{r}_{\text{opt}(\xi)}}(\xi)\right)$ für $\mu = \mu_{\text{cond}}$ und $\mu = \mu_{\max}$ für alle $\xi \in \Xi_{d,h}$ abgebildet. $\hat{r}_{\text{opt}(\xi)}$ ist die „optimale“ Anzahl von Parametern, die der Multi-Level-Algorithmus für $r_{\max} = 10$ und $\frac{1}{\varepsilon} \cdot \mu_{\text{tol}} = 10000$ für μ_{cond} bzw. $\frac{1}{\varepsilon} \cdot \mu_{\text{tol}} = 100$ für $\mu_{\max}$ liefert. Hierbei wurde $\varepsilon = 5\%$ gewählt. Beide hydraulischen Funktionen wurden gleichartig parametrisiert und wir erhalten für die „optimale“ Anzahl von Freiheitsgraden pro Funktion für das Zielfunktional μ_{cond}

$$r_{\text{opt}(\xi)} = \begin{cases} 10 & x_{\text{d}} \in [0, 14.25], \\ 4 & x_{\text{d}} \in (14.25, 14.55), \\ 3 & x_{\text{d}} \in [14.55, 15) \end{cases}$$

und für das Zielfunktional $\mu_{\max}$

$$r_{\text{opt}(\xi)} = \begin{cases} 10 & x_{\text{d}} \in [0, 14.25], \\ 3 & x_{\text{d}} \in (14.25, 15). \end{cases}$$

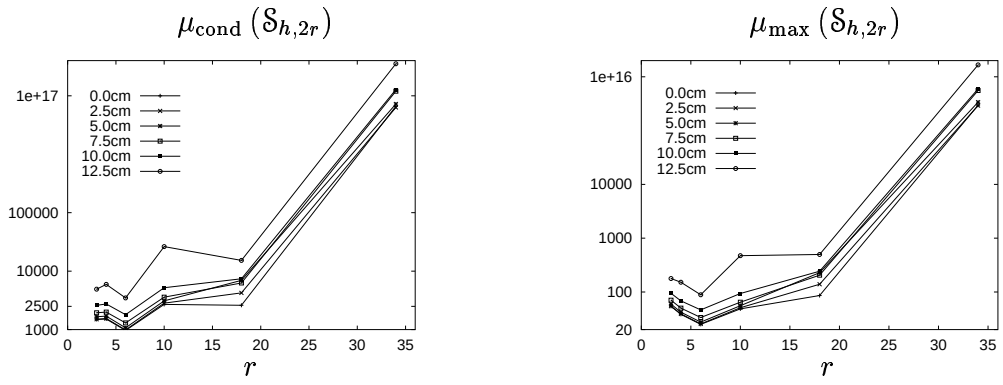


Abbildung 5.12: Spektralkondition und Maximum-Charakteristik in Abhängigkeit von der Koordinate x_d der Druckbeobachtung für eine Säule der Länge $L = 15$ cm.

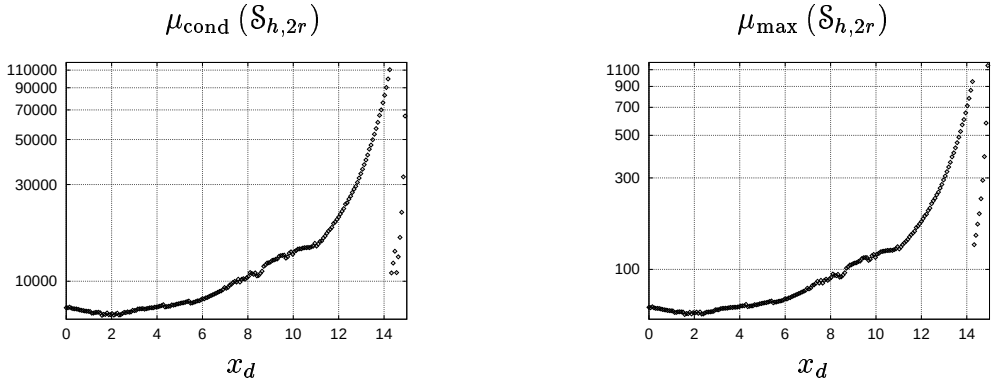


Abbildung 5.13: Spektralkondition und Maximum-Charakteristik bei „optimaler“Anzahl von Freiheitsgraden in Abhängigkeit von der Position der Druckbeobachtung für eine Säule der Länge $L = 15$ cm, äquisistante räumliche Diskretisierung mit 100 Elementen.

Dementsprechende Sprünge sind auch in den zugehörigen Zielfunktionswerten zu beobachten.

Diese Beispiel zeigt deutlich, dass eine Druckbeobachtung in der oberen Hälfte der Säule ($0 \leq x_d < \frac{L}{2}$) für die Identifizierug der hydraulischen Funktionen besser geeignet ist als eine Druckbeobachtung in der unteren Hälfte der Säule. Das Optimum befindet sich in diesem Beispiel knapp unterhalb des oberen Randes (≈ 2 cm). Dieses Ergebnis, das weitere Tests mit synthetischen Daten bestätigt haben, entspricht auch den Erfahrungen der Experimentatoren.

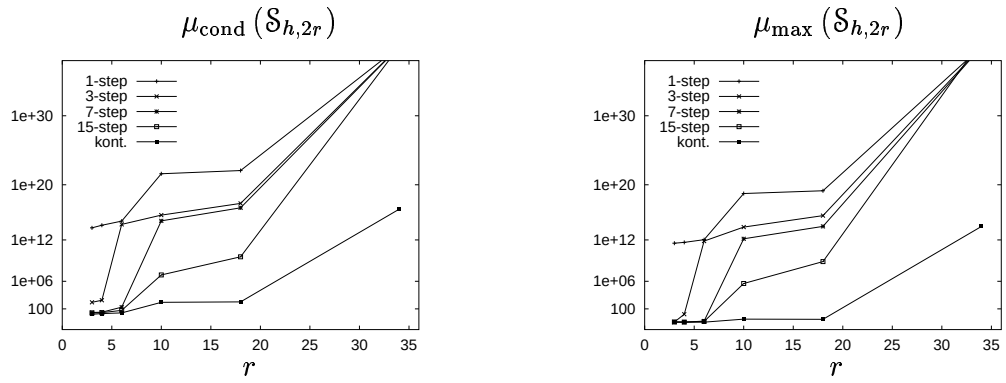


Abbildung 5.14: Spektralkondition und Maximum-Charakteristik bei Variation der Anzahl der Druckstufen in der Dirichlet-Randbedingung.

Im Weiteren wollen wir die Multistep-Experimente etwas genauer betrachten.

Beispiel 5.5 Zunächst untersuchen wir, wie die Anzahl der Druckstufen in der Multistep-Funktion $g(t)$ (Dirichlet-Randbedingung) die Identifizierung der hydraulischen Funktionen beeinflusst. Dazu konstruieren wir uns wieder synthetische Daten ($\varepsilon = 5\%$) für eine Säule der Länge $L = 15$ cm. Der Druck am Dirichlet-Rand soll über dem Zeitintervall $[0, 40]$ von 15 cm auf -100 cm abgesenkt werden. Wir beginnen mit einer Multistep-Funktion, bei der der Druck in einer Stufe abgesenkt wird (Onestep-Experiment), d. h. wir haben zwei Teilintervalle $[0, 20)$ und $[20, 40]$ in denen der Druck jeweils konstant 15 cm bzw. -100 cm ist. Die nächste Multistep-Funktion erhalten wir indem in den Mittelpunkten beider Teilintervalle eine neue Druckstufe eingeführt wird. Wir erhalten also 3 Druckstufen, wobei die Druckdifferenz jeweils den gleichen Betrag haben soll. Analog bilden wir die nachfolgenden Multistep-Funktionen. In der Abbildung 5.14 sind die Spektralkondition und die Maximum-Charakteristik der Sensitivitätsmatrix für die ersten 4 Multistep-Funktionen und dem Grenzfall des kontinuierlichen Experiments (unendliche Anzahl von Druckstufen) dargestellt. Die Onestep-Methode liefert unabhängig von der Anzahl der Freiheitsgrade die größten Werte für die Spektralkondition und die Maximum-Charakteristik. Bereits im Multistep-Experiment mit 3 Druckstufen sind diese Werte z. T. deutlich kleiner. Diese Entwicklung setzt sich bei wachsender Anzahl von Druckstufen fort und die Werte der Spektralkondition und der Maximum-Charakteristik nähern sich den für das Experiment mit kontinuierlichen Ausfluss erhaltenen Werten an. Für eine geringe Anzahl von Freiheitgraden ($r = 3, 4, 6$) bewegen sich die Werte für das Experiment mit 7 Druckstufen bereits in der Größenordung

der Werte für das kontinuierliche Experiment. Für $r \geq 10$ bestehen jedoch noch Unterschiede von mehreren Größenordnungen. Im Beispiel 5.1 hatten wir für $r = 3, 4, 6$ im Multistep-Experiment mit 5 Druckstufen kleinere Werte für die Spektralkondition erhalten als für das kontinuierliche Experiment. Dort wurden die Werte jedoch auf die nichtgewichtete Sensitivitätsmatrix bezogen.

5.3 Optimierung der Multistep-Funktion

Zunächst können wir festhalten, dass bei der Optimierung durch Variation der Anzahl der Druckstufen in der Dirichlet-Randbedingung eine eindeutige Tendenz zum Experiment mit kontinuierlichem Ausfluss zu beobachten ist. Im Weiteren wollen wir die Optimierung der Multistep-Funktion bei fixierter Anzahl von möglichen Druckstufen betrachten. Auch die Zeitpunkte an denen eine Druckänderung erfolgen kann sollen fest vorgegeben werden. Dazu wird die Funktion $g(t)$ stückweise konstant definiert. Die Menge der zulässigen Experimentszenarien kann auf unterschiedlichste Art formuliert werden. Zu beachten ist, dass die betrachteten Spline-Parametrisierungen die Identifizierung der hydraulischen Funktionen nur über dem Druckbereich ermöglichen, welcher durch das Experiment abgedeckt wird. D. h., wenn die hydraulischen Funktionen im Interval $[\underline{\psi}, 0]$ identifiziert werden sollen, dann muss gelten

$$\inf_{t\in(0,T)} \psi(.,t) \leq \underline{\psi}.$$

Dies kann z. B. durch Berücksichtigung der Nebenbedingung

$$\min_{t\in[0,T]} g(t) = \underline{\psi} \tag{5.5}$$

erreicht werden. Die Kompatibilität von $g(t)$ mit der Anfangsbedingung wird weiterhin angenommen. Die Menge der zulässigen Experimentszenarien kann wie folgt definiert werden:

$$\begin{aligned} \Xi_g^1 \;:=\; \Big\{ \{g(t)\} \;\Big|\; & g(t)|_{(t^i,t^{i+1})} \in P^0(t^i, t^{i+1}), i = 0, ..., n-1, \\ & g(0) = \psi_0(L), \min_{t\in[0,T]} g(t) = \underline{\psi} \Big\}, \end{aligned} \tag{5.6}$$

wobei $0 = t^0 < t^1 < ... < t^n = T$ eine vorgegebene Zerlegung von $[0, T]$ in n Teilintervalle ist und P^0 den Raum der konstanten Funktionen bezeichnet. Indem wir fordern, dass das Minimum für $t = T$ angenommen wird, können wir die Minimum-Bedingung in (5.6) vereinfachen:

$$\begin{aligned} \Xi_g^2 \;:=\; \Big\{ \{g(t)\} \;\Big|\; & g(t)|_{(t^i,t^{i+1})} \in P^0(t^i, t^{i+1}), i = 0, ..., n-1, \\ & g(0) = \psi_0(L),\; g(T) = \underline{\psi},\; \underline{\psi} \leq g(t) \leq \psi_0(L) \Big\}. \end{aligned} \tag{5.7}$$

In der Menge Ξ_g^2 sind auch Funktionen enthalten, die neben der Druckabsenkung eine Erhöhung des Druckes vornehmen. In der Realität kommt in diesem Fall die Hysterese ins Spiel, die wir vernächlässigen. Durch die Einbeziehung weiterer Nebenbedingungen kann eine Drucherhöhung auch ausgeschlossen werden.

Eine Schwierigkeit bei der Optimierung der Multistep-Funktion ergibt sich aus dem numerischen Verfahren zur Lösung des direkten Problems. Bei großen Druckgradienten wird häufig eine Nichtkonvergenz des Newton-Verfahrens beobachtet. Diese läßt sich auch durch eine Verfeinerung der Diskretisierung (Ort und Zeit) nicht beseitigen. Einen wesentlichen Einfluss scheint hierbei auch die Parametrisierung der hydraulischen Funktionen auszuüben.

Das Designproblem für die Menge Ξ_g^2 wurde erfolgreich mit einem stochastischen Verfahren bearbeitet. Hierbei handelt es sich um ein *Verfahren der simulierten Abkühlung* (*Simulated Annealing*). Details zu derartigen Verfahren sind in [34], [36] und [57] zu finden.

Algorithmus 5.6 (Verfahren der simulierten Abkühlung)

(i) Initialisierung

$k := 0$
Wähle „Anfangstemperatur“ $\tau_k = \tau_{\text{start}} > 0$, *Schrittweite* $\Delta\tau > 0$ *und* $l > 0$.
Erzeuge ein (zufälliges) Anfangsszenario $\xi \in \Xi$.

(ii) Erzeugung von Nachbarschaftsszenarien und Reduzierung der „Temperatur“

WHILE $\tau_k > 0$
  Führe folgende Schritte l*-mal aus.*
    Erzeuge ein zufälliges Szenario $\xi' \in U_{\tau_k}(\xi) \subset \Xi$.
    Bestimme $\Delta\mu = \mu\left(\mathcal{S}_{\hat{r}_{\text{opt}}(\xi')}(\xi')\right) - \mu\left(\mathcal{S}_{\hat{r}_{\text{opt}}(\xi)}(\xi)\right)$.
    Wenn $\Delta\mu < 0$ *oder* $\exp^{-\Delta\mu/\tau_k} \geq RANDOM[0,1]$,
    setze $\xi = \xi'$.
  $k := k + 1$
  $\tau_k := \tau_{k-1} - \Delta\tau$

Die jeweils zu untersuchenden Szenarien ξ' werden zufällig aus einer Umgebung $U_{\tau_k}(\xi)$ von ξ erzeugt. Das neue Szenario ξ' wird akzeptiert, wenn sich ein verbesserter Zielfunktionswert ergibt. Im Falle einer Zielfunktionswertverschlechterung wird ξ' mit der Wahrscheinlichkeit $\exp^{-\Delta\mu/\tau_k}$ akzeptiert. Diese

Wahrscheinlichkeit verringert sich mit zunehmendem $\Delta\mu$. Durch die Zulässigkeit von Zielfunktionswertverschlechterungen soll verhindert werden, dass die Optimierung in einem lokalen Tal von μ liegen bleibt. Im Prinzip müssen für jede Temperatur solange Nachbarschaftsszenarien betrachtet werden bis sich ein Gleichgewichtszustand einstellt. Hier wird vereinfacht stets nach l Schritten die Temperatur reduziert. Eine fallende Temperatur bewirkt eine Verringerung der Akzeptanzwahrscheinlichkeit von Zielfunktionswertverschlechterungen bis bei $\tau_k = 0$ die Lösung „eingefroren“ wird. Einflussgrößen, die das Endszenario mitbestimmen, sind die Schrittweite $\Delta\tau$, die Anzahl der für jede Temperatur τ_k zu untersuchenden Nachbarschaftsszenarien und die Nachbarschaftsumgebung U_{τ_k}. Diese Umgebung wird so gebildet, dass die oben erwähnte Nichtkonvergenz des Newton-Verfahrens nicht eintritt.

Beispiel 5.7 Wir beziehen uns auf Beispiel 5.5 und setzen $L = 15$, $\underline{\psi} = -100$ cm und $T = 40$ h. Das Zeitinterval $(0, T)$ wird äquidistant in 16 Teilintervalle unterteilt:

$$t^i := i\frac{T}{16} \qquad , i = 0, ..., 16.$$

Die Kontrollparameter im Verfahrem werden wie folgt gewählt:

$$\tau_{\text{start}} = 1000, \quad \Delta\tau = 20, \quad l = 20.$$

Für die Nachbarschaftsumgebung verwenden wir

$$\begin{aligned} U_{\tau_k}(g(t)) \quad := \quad & \left\{ \tilde{g}(t) \in \Xi_g^2 : g(t) - \gamma\tau_k \frac{\psi_0(L) - \underline{\psi}}{\tau_{\text{start}}} \leq \tilde{g}(t) \right. \\ & \left. \leq g(t) + \gamma\tau_k \frac{\psi_0(L) - \underline{\psi}}{\tau_{\text{start}}} \right\} \end{aligned} \tag{5.8}$$

mit $\gamma = 0.25$. Die Nachbarschaftsszenarien werden nach der Gleichverteilung erzeugt. Das Start- und Endszenario, welches wir mit dem Algorithmus 5.6 erhalten, sind in der Abbildung 5.15 dargestellt. Dabei wurde als Zielfunktional die Konditionszahl der Sensitivitätsmatrix verwendet. Im Abbruchkriterium des Multi-Level-Algorithmus wurde ebenfalls die Konditionszahl betrachtet mit $\frac{1}{\varepsilon}\mu_{\text{tol}} = 10000$ und $r_{\max} = 10$.

Wenn wir die im Verfahren auftretenden Werte des Zielfunktionals betrachten (Abbildung 5.16), so ist eine Verringerung vom Start- zum Endszenario festzustellen. Aber der Gesamtbereich, in dem sich die Werte bewegen, ist verhältnismäßig klein. Die Differenz zwischen Anfangs- und Endwert ist kleiner als 100. Auch werden bereits für die Starttemperatur Szenarien betrachtet, deren Zielfunktionswert fast schon dem Endwert entsprechen. Wesentlich beeinflusst wird dies durch die Definition der Umgebung U_{τ_k}. Im Gegensatz

dazu haben wir bei der Optimierung der Position der Druckmessung beispielsweise eine deutlich größe Variation der Zielfunktionswerte erhalten.

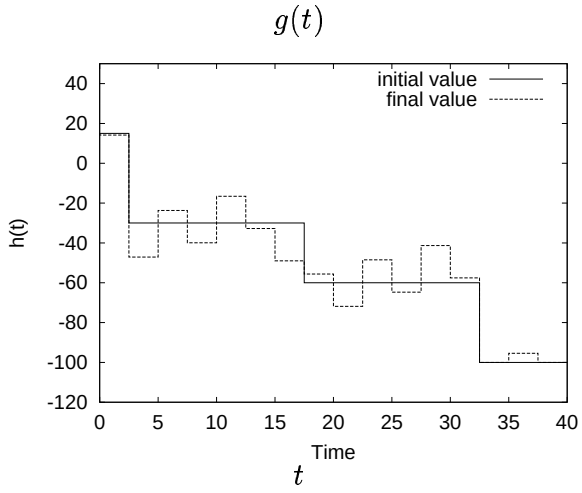


Abbildung 5.15: Anfangs- und Endwert der Dirichlet-Randbedingung beim Verfahren der simulierten Abkühlung.

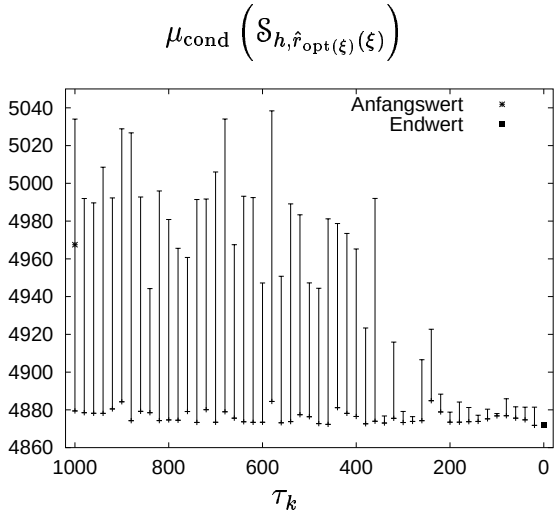


Abbildung 5.16: Zielfunktionswerte im Verfahren der simulierten Abkühlung.

5.4 Schlussfolgerungen

Anhand von einigen Beispielen haben wir Probleme des optimalen Experimentdesigns für die Identifizierung der hydraulischen Funktionen betrachtet. Das optimale Experimentdesign ist damit noch nicht erschöpfend behandelt, aber wir können schon einige Ergebnisse formulieren. Auf die Frage nach der optimalen Position der Messung des Drucks kann die Empfehlung gegeben werden, dass die Messung in der Nähe des oberen Randes der Säule erfolgen sollte. Durch den Vergleich von Experimenten mit Multistep- und kontinuierlichen Ausfluss haben wir festgestellt, dass im Allg. der kontinuierliche Ausfluss eine stabilere Identifizierung gewährleistet. Die Ergebnisse aus einem stochastischen Suchverfahren deuten darauf hin, dass die Stabilität bei begrenzten Änderungen der einzelnen Druckgradienten nicht wesentlich variiert. Wichtig für eine stabile Identifizierung im gesamten Intervall $[\underline{\psi}, 0]$ ist die hinreichende Berücksichtigung aller Druckwerte aus diesem Intervall in der Identifizierung. Insgesamt lassen die erzielten Resultate den Schluss zu, dass bei einer formfreien Parametrisierung der hydraulischen Funktionen die kontinuierlichen Ausflussexperimente den Multistep-Ausflussexperimenten vorzuziehen sind.

Anhang A

Zusammenfassung

Diese Arbeit beschäftigt sich mit der Identifizierung der hydraulischen Funktionen poröser Medien aus Säulenexperimenten. Besonderes Augenmerk wird hierbei auf das numerische Identifizierungsverfahren gerichtet.

In einem einleitenden Kapitel wird die Thematik dieser Arbeit vorgestellt. Es erfolgt eine kurze Beschreibung der zur Identifizierung der hydraulischen Funktionen geeigneten Säulenversuche.

Das zweite Kapitel beginnt mit einer Einführung in die Problematik der inversen Probleme. Es wird ein Überblick über Methoden zur Lösung und Stabilisierung inverser Probleme gegeben. Dem Prinzip der verallgemeinerten Inversen folgend wird das Parameteridentifizierungsproblem durch die Minimierung eines Fehlerfunktionals gelöst. Effiziente Optimierungsverfahren benötigen die Ableitung (Gradient) des Fehlerfunktionals. Diese Ableitung kann mithilfe eines adjungierten Problems berechnet werden. Zunächst wird der unendlichdimensionale Fall betrachtet. Anschließend werden die Resultate auf den endlichdimensionalen Fall übertragen. Die in [6] entwickelte und in [30] in verallgemeinerter Form beschriebene Methode der Integralidentitäten wird in einer angepassten Form dargestellt. Es wird ein adjungiertes Problem betrachtet, welches in enger Beziehung zu dem adjungierten Problem steht, das zur Berechnung des Gradienten des Fehlerfunktionals dient. Die Identifizierbarkeit nichtlinearer Koeffizientenfunktionen kann in Abhängigkeit von den Eigenschaften des direkten und adjungierten Problems gezeigt werden.

Das dritte Kapitel beginnt mit der Darstellung des Modells zur Beschreibung der Säulenexperimente. Der Fluidtransport wird durch die Richards-Gleichung modelliert. Anhand von Beispielen wird die Schlechtgestelltheit des inversen Problems demonstriert. Es wird eine Variationsformulierung der Richards-Gleichung vorgestellt, für die Existenz- und Eindeutigkeitsresultate bekannt sind. Für die weiteren Betrachtungen wird eine gemischte Variationsformulierung aufgestellt. Analog zu den in Kapitel 1 beschriebenen Metho-

den wird hierür ein adjungiertes Problem formuliert. Die Identifizierbarkeit der hydraulischen Funktionen anhand von Säulenexperimenten wird bewiesen. Im letzten Abschnitt des dritten Kapitels wird das diskrete Modell betrachtet. Die in [52] entwickelte hybrid-gemischte Finite-Elemente-Methode wird zur Diskretisierung der Richards-Gleichung eingesetzt und kurz skizziert. Nach Bemerkungen zur Lösung des diskreten direkten Problems wird analog zum kontinuierlichen Problem ein adjungiertes Problem betrachtet. Es werden drei Methoden zur Berechnung des diskreten Gradienten des Fehlerfunktionals beschrieben: die Differenzenmethode, die adjungierte Methode und die direkte Methode. Das Kapitel 3 schließt mit einem Vergleich der drei Methoden im Hinblick auf ihre Rechenzeit- und Speichereffizienz.

Das vierte Kapitel widmet sich dem numerischen Verfahren zur Identifizierung von Nichtlinearitäten. Hierbei wird wie im zweiten Kapitel eine allgemeine Formulierung verwendet. Nach der Präzisierung des Fehlerfunktionals werden verschiedene Methoden zur Parametrisierung der Nichtlinearitäten mithilfe von Spline-Funktionen vorgestellt. Entsprechend dem hierarchischen Konzept dieser Parametrisierungen erfolgt die numerische Identifizierung in einem Multi-Level-Verfahren. Die Multi-Level-Identifizierung wird mit einer Sensitivitätsanalyse gekoppelt. Im zweiten Abschnitt des vierten Kapitels wird die numerische Identifizierung anhand von Fallstudien untersucht. Hierbei wird sich wieder konkret auf die Identifizierung der hydraulischen Funktionen bezogen. Es wird der Einfluss des Diskretisierungsparameters und der Art der Parametrisierung auf die Identifizierung untersucht. Die spezielle Charakteristik des Multi-Level-Verfahrens wird demonstriert. Dabei wird die Identifizierung auch an experimentellen Daten getestet. Es bestehen Möglichkeiten Adaptivität in das Multi-Level-Verfahren einzubringen. So kann für bestimmte Parametrisierungsarten eine adaptive Verfeinerungsstrategie eingesetzt werden. Mithilfe der Sensitivitäten können die Wichtungsfaktoren im Fehlerfunktional adaptiv angepasst werden. Zusätzliche a priori Informationen über die hydraulischen Funktionen (wie z. B. die Mualem-Beziehung) können nach der Methode von Tikhonov in die Identifizierung eingebunden werden. Hierzu werden jeweils Beispiele betrachtet.

Das fünfte und abschließende Kapitel gibt einen Einblick in das optimale Experimentdesign. Zur Motivation werden zunächst Beispiele für Multistep-Experimente betrachtet. Anschließend erfolgt eine mathematische Darstellung des Experimentdesignproblems. Das optimale Experimentdesign liefert eine klare Aussage darüber, wo in der Säule eine Druckbeobachtung durchgeführt werden sollte, um eine gute Stabilität der Identifizierung zu gewährleisten. Die Optimierung der Multistep-Funktion wird mit einem stochastischen Verfahren bearbeitet.

Die Anhänge A und B enthalten diese Zusammenfassung und eine Dar-

stellung der insbesondere in Kapitel 2 benutzten Funktionenräume, insofern diese nicht an anderer Stelle definiert werden.

Anhang B

Funktionenräume

Im Folgenden werden die benutzten Funktionenräume aufgeführt, insofern diese nicht an anderer Stelle definiert werden.

(i) Räume stetiger Funktionen

$C^k(X;Y)$, $k \in \mathbb{N}$, ist der Raum aller stetigen Funktionen $F : X \to Y$, welche stetige Ableitungen bis zur k-ten Ordnung besitzen.
$C^k(X) := C^k(X;\mathbb{R})$.
$C^k[a,b] := C^k([a,b])$ mit $a, b \in \mathbb{R}$.
$C(X;Y) := C^0(X;Y)$.

(ii) Lebesgue-Räume

$L^p(X)$, $1 \leq p \leq \infty$, besteht aus allen messbaren Funktionen, für welche die Norm

$$\|f\|_p := \begin{cases} \left(\int_X |f|^p \, dx\right)^{\frac{1}{p}} & \text{für} \quad 1 \leq p < \infty, \\ \operatorname{ess\,sup}_X |f| & \text{für} \quad p = \infty \end{cases}$$

endlich ist.

$X_T := X \times (0,T)$.
$L^{p,q}(X_T)$, $p, q \geq 1$, ist der Raum aller messbaren Funktionen, für welche die Norm

$$\|f\|_{p,q} := \left(\int_0^T \left(\int_X |f|^p\right)^{\frac{q}{p}}\right)^{\frac{1}{q}}$$

endlich ist.

$L^{p,q}(X_T)$ für $1 \leq p, q \leq \infty$ ist analog definiert, wobei die ∞-Norm durch das wesentliche Supremum des Absolutwertes gegeben ist.

$L^p(X_T) := L^{p,p}(X_T), 1 \leq p \leq \infty$.

$L^p((0,T);X), 1 \leq p \leq \infty$ ist der Raum der messbaren Funktionen, für welche die Norm

$$\|f\|_{L^p((0,T);X)} := \begin{cases} \left(\int_J \|f(t)\|_X^p \, dt\right)^{\frac{1}{p}} & \text{für} \quad 1 \leq p < \infty, \\ \inf\limits_{\substack{\|f(t)\|_X \leq M \\ \text{für fast alle } t \in (0,T)}} M & \text{für} \quad p = \infty \end{cases}$$

endlich ist.

(iii) Sobolev-Räume

Sei $X \subset \mathbb{R}^N$, $\alpha \in \mathbb{R}^N$ Multi-Index, $|\alpha| := \sum_{i=1}^N \alpha_i$ und $D^\alpha := \frac{\partial^{\alpha_1+\ldots+\alpha_N}}{\partial x_1^{\alpha_1} \ldots \partial x_N^{\alpha_N}}$ die verallgemeinerte Ableitung.

$H^1(X) \subset L^2(X)$ und $H^{1,1}(X_T) \subset L^2(X_T)$ sind Hilbert-Räume mit den Normen

$$\|f\|_{H^1(X)} := \left(\sum_{|\alpha| \leq 1} \|D^\alpha f\|_2^2\right)^{\frac{1}{2}},$$

$$\|f\|_{H^{1,1}(X_T)} := \left(\sum_{i+|\alpha| \leq 1} \|\partial_t^i D^\alpha f\|_{2,2}^2\right)^{\frac{1}{2}}.$$

(iv) Vektorwertige Funktionenräume

Sei $X \subset \mathbb{R}^N$.

$$H(\text{div}; X) := \left\{ f \in (L^2(X))^N \mid \nabla \cdot f \in L^2(X) \right\}$$

mit der Norm

$$\|f\|_{\text{div}} := \left(\|f\|_2^2 + \|\nabla \cdot f\|_2^2\right)^{\frac{1}{2}}.$$

Literaturverzeichnis

[1] H. W. Alt. *Lineare Funktionalanalysis.* Springer, Berlin, 1992.

[2] H. W. Alt and S. Luckhaus. Quasilinear elliptic-parabolic differential equations. *Math. Z.*, 183:311–341, 1983.

[3] H. Bandemer and A. Bellmann. *Statistische Versuchsplanung.* B. G. Teubner, Leipzig, 1979.

[4] S. Brandt. *Datenanalyse.* B. I.-Wissenschaftsverlag, Mannheim, Wien, Zürich, 1981.

[5] J. R. Cannon. *The One Dimensional Heat Equation.* Addison-Wesley, Reading, MA, 1984.

[6] J. R. Cannon and P. DuChateau. Indirect determination of hydraulic properties of porous media. *International Series of Numerical Methematics*, 114:37–50, 1993.

[7] G. Chavent, J. Zhang, and C. Chardaire-Riviere. Estimation of mobilities and capillary pressure from centrifuge experiments. In H. D. Bui, M. Tanaka, M. Bonnet, H. Maigre, E. Luzzato, and M. Reynier, editors, *Inverse Problems in Engineering Mechanics*, pages 265–272. Balkema, Rotterdam, 1994.

[8] A. Dieses. *Numerical Methods for Optimization Problems in Water Flow and Reactive Solute Transport Processes of Xenobiotics in Soils.* PhD thesis, Ruprecht-Karls-Universität Heidelberg, Germany, 2001.

[9] P. DuChateau. An inverse problem for the hydraulic properties of porous media. In *Proc. of the 1994 groundwater and modeling conference*, pages 95–103. Colorado State University, 1994.

[10] P. DuChateau. An introduction to inverse problems in partial differential equations for engineers, physicists and mathematicians, a tutorial. In

J. Gottlieb and P. DuChateau, editors, *Proceedings of the workshop on parameter identification and inverse problems in hydrology, geology and ecology*, pages 3–50. Kluwer Aca. Publ., 1995.

[11] P. DuChateau. An inverse problem for the hydraulic properties of porous media. *SIAM J. Math. Anal.*, 28(3):611–632, May 1997.

[12] P. DuChateau. Monotonicity and invertibility of coefficient-to-data mappings for parabolic inverse problems. *SIAM J. Math. Anal.*, 26(6):1473–1487, November 1995.

[13] W. Durner, E. Priesack, H.-J. Vogel, and T. Zurmühl. Determination of parameters for flexible hydraulic functions by inverse modeling. In M. Th. van Genuchten, F. J. Leij, and L. Wu, editors, *Proc. Int. Workshop on Characterization and Measurement of the Hydraulic Properties of Unsaturated Porous Media, October 22–24, 1997*, pages 817–829, Riverside, CA, 1999. University of California.

[14] W. Durner, B. Schultze, and T. Zurmühl. State-of-the-art in inverse modeling of inflow/outflow experiments. In M. Th. van Genuchten, F. J. Leij, and L. Wu, editors, *Proc. Int. Workshop on Characterization and Measurement of the Hydraulic Properties of Unsaturated Porous Media, October 22–24, 1997*, pages 661–681, Riverside, CA, 1999. University of California.

[15] W. Durner and T. Zurmühl. Determination of parameters for bimodal hydraulic functions by inverse modeling. *Soil Sci. Soc. Am. J.*, 62:874–880, 1998.

[16] H. W. Engl, K. Kunisch, and A. Neubauer. Convergence rates for Tikhonov regularization of non-linear ill-posed problems. *Inverse Problems*, 5:523–540, 1989.

[17] H. W. Engl and A. Neubauer M. Hanke. *Regularization of Inverse Problems*. Kluwer, Dordrecht, 1996.

[18] R. Eymard, M. Gutnic, and D. Hilhorst. The finite volume method for Richards equation. *Computational Geoscience 3*, 3:259–294, 1999.

[19] V. V. Fedorov. *Theory of Optimal Experiments*. Academic Press, New York, 1972.

[20] F.N . Fritsch and R. E. Carlson. Monotone piecewise cubic interpolation. *SIAM Num. Anal.*, 17:238–246, 1980.

[21] S. Fuĉík and A. Kufner. *Nonlinear Differential Equations.* Elsevier Scientific Publishing Company, Amsterdam-Oxford-New York, 1980.

[22] M. M. Gribb. Parameter estimation for determining hydraulic properties of a fine sand from transient flow measurements. *Water Resources Research*, 32(7):1965–1974, 1996.

[23] C. W. Groetsch. *Inverse Problems in the Mathematical Sciences.* Vieweg, Braunschweig, 1993.

[24] J. Hadamard. *Lectures on the Cauchy Problem in Linear Partial Differential Equations.* Yale University Press, New Haven, 1923.

[25] G. Hämmerlin and K.-H. Hoffmann. *Numerische Mathematik.* Springer, Berlin, 1994.

[26] M. Hanke, A. Neubauer, and O. Scherzer. A convergence analysis of the Landweber iteration for nonlinear ill-posed problems. *Numerische Mathematik*, 72:21–37, 1995.

[27] S. B. Hazra and V. Schulz. Numerical parameter identification in multiphase flow through porous media. *Computing and Visualization in Science*, 5(2):107–113, 2002.

[28] D. Hillel. *Environmental Soil Physics.* Academic Press, San Diego, London, 1998.

[29] B. Hofmann. *Mathematik inverser Probleme.* Teubner, Leipzig, 1999.

[30] B. A. Igler. *Identification of Nonlinear Coefficient Functions in Reactive Transport through Porous Media.* PhD thesis, Friedrich-Alexander-Universität Erlangen-Nürnberg, Germany, 1998.

[31] V. Isakov. *Inverse Problems for Partial Differential Equations.* Springer, New York, 1998.

[32] J. Jahn. *Introduction to the Theory of Nonlinear Optimization.* Springer, Berlin, 1994.

[33] K. Katayama and N. Narihisa. Performance of simulated annealing-based heuristic for the unconstrained binary quadratic programming problem. *European Journal of Operational Research*, 134:103–119, 2001.

[34] S. Kirkpatrick, C. D. Jr. Gelatt, and M. P. Vecchi. Optimization by simulated annealing. *Science*, 220:671–680, 1983.

[35] J. B. Kool, J. C. Parker, and M. Th. van Genuchten. Parameter estimation for unsaturated flow and tranport models–A review. *J. Hydrol.*, 91:255–293, 1987.

[36] H. Kuhn. Heuristische Suchverfahren mit simulierter Abkühlung. *Wirtschaftswissenschaftliches Studium*, 8:387–391, 1992.

[37] A. Ladyzhenskaya, V. A. Solonnikov, and N. N. Uralceva. *Linear and Quasi-linear Equations of Parabolic Type.* Transl. Math. Monographs 23, AMS, RI, 1969.

[38] Lehrstuhl für Angewandte Methematik I, Friedrich-Alexander-Universität Erlangen-Nürnberg. RICHY*'s manual.* http://www.uni-erlangen.de/am1/software/RichyDocumentation/Main.html.

[39] A. K. Louis. *Inverse und schlecht gestellte Probleme.* Teubner, Stuttgart, 1989.

[40] P. H. Müller, editor. *Lexikon der Stochastik.* Akademie Verlag, Berlin, 1991.

[41] R. Nabokov. *An Inverse Problem for the Porous Medium Equation: Identification of Permeability.* PhD thesis, Ruprecht-Karls-Universität Heidelberg, Germany, 1996.

[42] A. Neubauer. Tikhonov regularization for non-linear ill-posed problems: Optimal convergence rates and finite-dimensional approximation. *Inverse Problems*, 5:541–557, 1989.

[43] A. Neubauer. On Landweber iteration for nonlinear ill-posed problems in Hilbert scales. *Numerische Mathematik*, 85:309–328, 2000.

[44] F. Otto. L^1-contraction and uniqueness for unstationary saturated-unsaturated porous media flow. *Adv. Math. Sci. Appl.*, 7(2):537–553, 1997.

[45] R. Remmert. *Funktionentheorie.* Springer, Berlin, 1995.

[46] A. Rieder. A wavelet multilevel method for ill-posed problems stabilized by Tikhonov regularization. *Numerische Mathematik*, 75:501–522, 1997.

[47] N. Romano and A. Santini. Determining soil hydraulic functions from evaporation experiments by a parameter estimation approach: Experimental verifications and numerical studies. *Water Resources Research*, 35:3343–3359, 1999.

[48] R. Scheibke. Entwicklung einer hochauflösenden Datenerfassung zur Bestimmung der hydraulischen Eigenschaften ungestörter Bodensäulen. Master's thesis, Lehrstuhl für Hydrologie, Universität Bayreuth, 1990.

[49] O. Scherzer. An iterative multi level algorithm for solving nonlinear ill-posed problems. *Numerische Mathematik*, 80:579–600, 1998.

[50] O. Scherzer, H. W. Engl, and K. Kunisch. Optimal a posteriori parameter choice for Tikhonov regularization for solving nonlinear ill-posed problems. *SIAM J. Numer. Anal.*, 30:1796–1838, 1989.

[51] K. Schittkowski. *Trends in Mathematical Optimization*, chapter Solving constrained nonlinear least squares problems by a general purpose SQP-method. Birkhäuser, 1988.

[52] E. Schneid. *Hybrid-Gemischte Finite-Elemente-Diskretisierung der Richards-Gleichung.* PhD thesis, Friedrich-Alexander-Universität Erlangen-Nürnberg, Germany, 2000.

[53] B. Schultze and W. Durner. *Optimierung von Meßmethoden zur hydraulischen Charakterisierung von Böden.* Lehrstuhl für Hydrologie, Universität Bayreuth.

[54] B. Schultze, T. Zurmühl, and W. Durner. Ein Vergleich von Onestep-, Multistep- und Kontinuierlichen Gradientenverfahren zur Bestimmung der hydraulischen Funktionen von Bodensäulen. *Mitteilungen der Deutschen Bodenkundlichen Gesellschaft*, 76:157–160, 1995.

[55] I. Seidman and C. R. Vogel. Well posedness and convergence of some regularization methods for non-linear ill posed problems. *Inverse Problems*, 5:227–238, 1989.

[56] P. Spellucci. *Numerische Verfahren der nichtlinearen Optimierung.* Birkhäuser, Basel, 1993.

[57] A. Tarantola. *Inverse Problem Theory.* Elsevier, Amsterdam, 1987.

[58] G. C. Topp. Soil-water hysteresis measured in a sandy loam and compared with the hysteretic domain model. *Soil Sci. Soc. Am. Proc.*, 33:645–651, 1969.

[59] K. U. Totsche. *Reaktiver Stofftransport in Böden: Optimierte Experimentdesigns zur Prozeßidentifikation.* Habilitationschrift, Bayreuther Bodenkundliche Berichte, Band 75, 2001.

[60] M. Th. van Genuchten. A closed-form equation for the predicting the hydraulic conductivity of unsaturated soils. *Soil Sci. Soc. Am. J.*, 44:892–898, 1980.

[61] D. Werner. *Funktionalanalysis.* Springer, Berlin, 1995.

[62] A. Wouk. *A course of applied functional analysis.* John Wiley and Sons, New York, 1979.

[63] E. Zeidler. *Nonlinear Functional Analysis and its Applications I-IV.* Springer, Berlin, 1984-93.

Lebenslauf

Persönliche Daten

Name:	Sandro Bitterlich
Geburtsdatum/-ort:	22. Januar 1974 in Annaberg-Buchholz
Familienstand:	ledig
Staatsangehörigkeit:	deutsch

Bildungsweg

Schulausbildung

Sept. 1980 – Juni 1990	Polytechnische Oberschule „E. Weinert“ in Neudorf
Juni 1990	Abschluss 10. Klasse
Sept. 1990 – Juni 1992	Erweiterte Oberschule „Joh. R. Becher“ in Annaberg-Buchholz
Juni 1992	Abschluss mit Abitur

Hochschulausbildung

Okt. 1992 – Dez. 1998	Studium der Mathematik an der TU Bergakademie Freiberg
Dez. 1998	Abschluss als Diplom-Mathematiker
seit Feb. 1999	Promotion am Institut für Angewandte Mathematik der Friedrich-Alexander-Universität Erlangen-Nürnberg

Zivildienst

März 1994 – Mai 1995	Jugendherberge in Neudorf

Berufliche Tätigkeiten

seit Feb. 1999	Wissenschaftlicher Mitarbeiter (BAT IIa/2) in Drittmittelprojekten am Institut für Angewandte Mathematik der Friedrich-Alexander-Universität Erlangen-Nürnberg