

Forschungszentrum Jülich

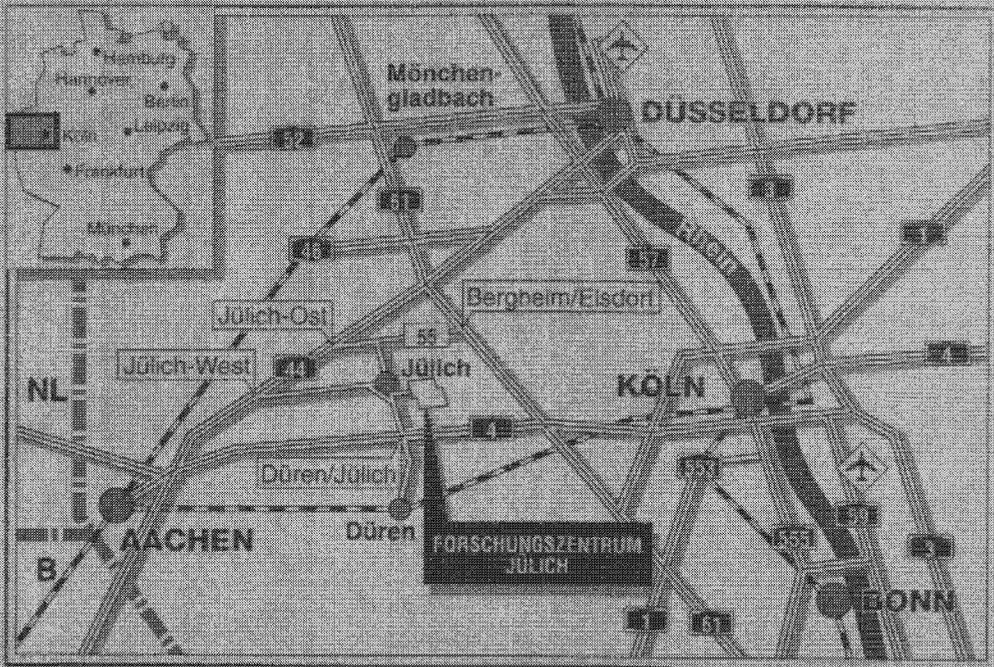


Zentralinstitut für Angewandte Mathematik

***Parallele Simulation
der Schadstoffausbreitung im Boden:
Lösung einer nichtlinearen para-
bolischen Differentialgleichung
mittels Gebietszerlegung***

Joannes van der Linden

Jüli-3356



Berichte des Forschungszentrums Jülich ; 3356
ISSN 0944-2952
Zentralinstitut für Angewandte Mathematik Jüli-3356

Zu beziehen durch: Forschungszentrum Jülich GmbH · Zentralbibliothek
D-52425 Jülich · Bundesrepublik Deutschland
☎ 02461/61-6102 · Telefax: 02461/61-6103 · e-mail: zb-publikation@fz-juelich.de

Zusammenfassung

Zur Simulation des Grundwasserflusses und der Schadstoffausbreitung im Boden sind unter anderem nichtlineare Dispersions-Advektions-Gleichungen mit stark inhomogenen Koeffizienten zu lösen.

Bei der numerischen Linienmethode werden solche Differentialgleichungen durch eine konstante lokale Diskretisierung in ein System steifer nichtlinearer gewöhnlicher Differentialgleichungen überführt, das dann mit Standardverfahren zur Behandlung von Anfangswertproblemen gewöhnlicher Differentialgleichungen gelöst wird. Der Vorteil dieses Vorgehens liegt darin, daß Verfahren mit adaptiver Steuerung der Ordnung und Schrittweite gewählt werden können, zum Beispiel Extrapolationsverfahren.

Die semi-implizite Mittelpunkregel als Grundverfahren führt auf eine h^2 -Extrapolation, in der für die Jacobi-Matrix des Verfahrens eine Vorkonditionierungsmatrix, nämlich ein Gebietszerlegungs-Vorkonditionierer, eingesetzt werden kann. Dadurch erhält man ein Verfahren, in dem lineare Gleichungssysteme nur noch lokal auf den Rechenknoten gelöst werden müssen.

In dieser Arbeit werden die Gleichungen, die den Grundwasserfluß und den Transport gelöster Stoffe beschreiben, sowie eine konservative Ortsdiskretisierung kurz vorgestellt; anschließend werden die Voraussetzungen analysiert, unter denen eine stabile Lösung der semidiskreten Aufgabe auch mit einer Approximation der Jacobi-Matrix mittels einer Gebietszerlegung möglich ist.

Inhaltsverzeichnis

Symbolverzeichnis	vii
Einleitung	1
1. Stofftransport durch poröse Medien	5
1.1. Die zugrundeliegenden Gleichungen	5
1.1.1. Klassische Formulierung	6
1.1.2. Schwache Formulierung	7
1.2. Semidiskretisierung und Linienmethode	8
1.2.1. Finite-Volumen-Verfahren	8
1.2.2. Linienmethode	10
1.3. Eigenschaften des semidiskreten Systems	11
1.3.1. Steifigkeit	11
1.3.2. Stabilität	14
2. Die numerische Lösung des semidiskreten Systems	19
2.1. Die semi-implizite Mittelpunkt-Regel	19
2.2. Extrapolation	21
2.3. Kontraktivität und Stabilität im nichtlinearen Fall	24
2.4. Ordnungs- und Schrittweitensteuerung	26
2.4.1. Schrittweiten	26
2.4.2. Minimierung des relativen Aufwandes	29
2.5. Zur Realisierung des Verfahrens	30
3. Gebietszerlegungsverfahren als Vorkonditionierer	33
3.1. Iterative Lösung linearer Gleichungssysteme	33

3.1.1. Lineare Iterationsverfahren	34
3.1.2. Polynomiale Beschleunigung	35
3.1.3. Vorkonditionierungstechniken	35
3.2. Gebietszerlegungsverfahren	37
3.3. Vorkonditionierer im semi-impliziten Verfahren	40
4. Implementierung	43
4.1. Message Passing	43
4.2. Kriterien zur Bewertung paralleler Programme	45
4.3. Realisierung	45
5. Simulationsergebnisse	49
Zusammenfassung und Ausblick	53
A. Einiges über Grundwasser-Hydrologie	55
A.1. Grundwasserfluß	55
A.1.1. Das Darcy-Gesetz	55
A.1.2. Massenerhaltung	59
A.1.3. Randwerte	61
A.2. Stofftransport	61
Literaturverzeichnis	65

Symbolverzeichnis

Die folgenden Bezeichnungen werden auf den angegebenen Seiten definiert. Sie sind hier nach den Sachgebieten sortiert.

In Kapitel 1 werden die Differentialgleichungen und die Diskretisierung dargestellt. Auch einige Begriffe aus der Numerik gewöhnlicher Differentialgleichungen werden in Kapitel 1 behandelt. Dabei treten folgende Bezeichnungen auf:

Ω	Gebiet, auf dem die Differentialgleichungen erklärt sind, S. 6
\mathbf{x}	Ortsvektor $\mathbf{x} \in \Omega$, S. 6
t	Zeitkomponente $t \in (0, T)$, S. 6
$\partial\Omega$	Rand des Gebiets Ω , S. 6
\mathbb{R}	Körper der reellen Zahlen, S. 6
\mathbb{R}_+	Menge der positiven reellen Zahlen, S. 6
$C^\infty(\Omega)$	Raum der auf Ω beliebig oft stetig differenzierbaren Funktionen, S. 7
$L^2(\Omega)$	Lebesgue-Raum der quadratisch integrierbaren Funktionen, S. 7
$H^1(\Omega)$	Sobolev-Raum der einmal schwach differenzierbaren L^2 -Funktionen, S. 7
$H(\text{div}, \Omega)$	Sobolev-Raum der vektorwertigen Funktionen mit Divergenz in L^2 , S. 7
$G_{i,j}, V_{i,j}, U_{i,j}$	Teilgebiet; Volumenelement im Finite-Volumen-Verfahren, S. 8
h	Orts-Schrittweite, hier: Kantenlänge eines Volumenquadrats, S. 8
$\chi_V(\mathbf{x})$	Charakteristische Funktion von V auf Ω , S. 8
\mathbf{e}_i	i -ter Standard-Einheitsvektor in \mathbb{R}^d (in kartesischen Koordinaten), S. 8
\mathbf{e}_g	Einheitsvektor in Gravitationsrichtung, S. 6
$\lambda(J)$	Spektrum der Matrix J , Menge der Eigenwerte, S. 11
λ_1, λ_n	betragskleinster und betragsgrößter Eigenwert, S. 12

$\langle \cdot, \cdot \rangle$	(beliebiges) Skalarprodukt, S. 14
$\ \cdot \ $	Vektor- bzw. Matrix-Norm, S. 14
$\mu[A]$	logarithmische Matrixnorm der Matrix A , S. 15

In Kapitel 2 wird das semi-implizite Verfahren vorgestellt. Folgende Bezeichnungen sind dabei von größerem Interesse:

$f_y(y_0)$	Jacobi-Matrix der rechten Seite f , S. 19
$A \approx f_y(y_0)$	Stabilisierungsmatrix, S. 20
h	Schrittweite im lokalen Verfahren (semi-implizite MPR)
H	Schrittweite im globalen Extrapolationsverfahren
$\{n_i\}$	Schrittweitenfolge für die lokalen Schrittweiten $h_i = \frac{H}{n_i}$
$S(t_0 + H, h)$	Näherungslösung aus der semi-impliziten Mittelpunkregel, S. 20
$T_{i,k}$	Näherung im Extrapolationsverfahren, S. 21
$R(z)$	rationale Stabilitätsfunktion, S. 22
$\varepsilon_{i,k}, \bar{\varepsilon}_{i,k}$	Fehler, Fehlerschätzer, S. 26
$W_{i,k}$	relativer Aufwand pro Einheits-Zeitschritt, S. 29

In Kapitel 3 werden Gebietszerlegungs-Vorkonditionierer vorgestellt. Dabei werden durchgängig folgende Bezeichnungen gewählt:

$\Omega = \bigcup_{i=1}^p \Omega_i$	Gebietszerlegung, S. 40
Q_i, Q_i^T	Restriktion, Prolongation zum Teilgebiet Ω_i , S. 39

Im Kapitel 4 werden unter anderem folgende Begriffe zur Bewertung paralleler Programme eingeführt:

t_p	Ausführungszeit eines Programms auf p Prozessoren, S. 45
$S(p)$	Speedup (proportionaler Zeitgewinn) auf p Knoten, S. 45
E	parallele Effizienz, S. 45

In der Grundwasser-Hydrologie (Anhang A) sind folgende Bezeichnungen üblich (in Klammern die jeweiligen Dimensionen):

v	Filtergeschwindigkeit, spezifischer Durchfluß (LT^{-1}), S. 58
$K, K_{sat}, K(\psi), K(\Theta)$	hydraulische Leitfähigkeit (LT^{-1}), S. 58

z	geodätische Höhe, Gravitationspotential (L), S. 57
ψ	Druckhöhe, hydrostatisches Potential (L), S. 57
$h = \psi + z$	Standrohrspiegelhöhe, totales Potential (L), S. 57
∇h	hydraulischer Gradient, S. 58
Θ	Wassergehalt (L^3L^{-3}), S. 58
n	Porosität des Mediums (L^3L^{-3}), S. 58
S	Sättigungsgrad (L^3L^{-3}), S. 58
F	hydraulischer Speicherkoeffizient, S. 60
c	Stoffkonzentration im Wasser (ML^{-3}), S. 62
s	Masse des sorbierten Stoffes (pro Einheitsmasse Boden; MM^{-1}), S. 62
D	Dispersionstensor (L^2T^{-1}), S. 63
D_l, D_t	Dispersionskoeffizienten, longitudinal und transversal (L^2T^{-1}), S. 63
a_l, a_t	Dispersivitäten (L), S. 63
ρ	Dichte, für Wasser bei $10^\circ C$ etwa $9.997 \cdot 10^2 \text{ kg m}^{-3}$
μ	dynamische Viskosität, für Wasser bei $10^\circ C$ etwa $1.31 \cdot 10^{-3} \text{ kg s}^{-1} \text{ m}^{-1}$
g	Erdbeschleunigung, etwa 9.81 m s^{-2}

Einleitung

Laut einer UNO-Analyse zum vierten „Weltwassertag“ am 22. März 1996 ist im kommenden Jahrhundert damit zu rechnen, daß eine extreme Wasserknappheit die meisten Städte der Entwicklungsländer und auch einige Städte in den Industrieländern bedrohen wird. Die Trinkwasserknappheit in städtischen Ballungsräumen werde durch unzureichende Wasserversorgung, Verschwendung und die Verschmutzung des Wassers durch Fäkalien und Industrieabfälle verursacht (*Süddeutsche Zeitung* vom 19. März 1996, S.1).

Vor diesem Hintergrund einer globalen Bedrohung durch einen Mangel an trinkbarem Wasser sind Projekte zu sehen, die sich um die drastische Senkung der Wasserverschwendung, um die Berücksichtigung der Versorgung mit Trinkwasser im Wohnungs- und Siedlungswesen und um die Sicherung der Trinkwasserqualität bemühen.

Im Zusammenhang mit der Qualitätssicherung kommt hydrogeologischen und chemischen Untersuchungen von Grundwasserlagern, Grundwasserfluß und Schadstoffausbreitung im Grundwasser eine große Bedeutung zu. Probleme wie zum Beispiel die Uferinfiltration bei stark verschmutzten Flüssen, in deren Nachbarschaft Trinkwasserbrunnen liegen, die Grundwasserbelastung durch Düngemittel oder Verunreinigungen des Bodens durch Chemieabfälle erfordern zu ihrer Bewältigung eine möglichst genaue Kenntnis des Flußverhaltens des Grundwassers sowie der Transport- und Reaktionsmechanismen im Grundwasser. Diese gewinnt man durch Laborexperimente und Feldversuche, aber auch durch numerische Simulationen.

Am Institut für Erdöl und organische Geochemie (ICG-4) des Forschungszentrums Jülich werden solche Untersuchungen zur Ausbreitung von Schadstoffen in Boden und Grundwasser durchgeführt. Das geschieht einerseits durch Freiland-Experimente auf einem Versuchsgelände in Krauthausen in der Nachbarschaft des Forschungszentrums, andererseits durch Computer-Simulationen der Transportvorgänge.

Bei den Versuchen in Krauthausen handelt es sich um sogenannte Tracer-Experimente, bei denen durch eine Bohrung ein Farbstoff (der *Tracer*) in den Boden eingebracht wird, dessen Konzentration an weiteren Bohrungen zu verschiedenen Zeiten gemessen wird. Durch diese Messungen erhofft man sich Erkenntnisse über den Zusammenhang von Wasserfluß, Schadstoffkonzentration, Permeabilität (Durchlässigkeit) des Bodens und Sorption (Reaktion des Stoffes mit dem Medium). Aus diesem Grunde werden verschiedene, reaktive und nichtreaktive, Tracer verwendet. In der Flußrichtung des Grundwassers gibt es insgesamt etwa 60 Meßbohrungen in Entfernungen um 20 – 30 Meter, um 60 – 70 Meter und um 120

Meter von den Injektions-Bohrungen¹. Die Flußgeschwindigkeit des Grundwassers beträgt etwa 1 Meter pro Tag.

Parallel zu diesen aufwendigen Experimenten werden die Vorgänge auch anhand von Simulationen untersucht; die dabei auftretenden nichtlinearen parabolischen Differentialgleichungen können mit vertretbarem Zeitaufwand und in der erforderlichen räumlichen Auflösung nur auf Parallelrechnern gelöst werden. Dazu wird am ICG-4 ein Computer-Programm (*TRACE*, *TRAN*sport of *CONT*aminants in *ENV*ironmental systems) [41] für einen massiv-parallelen Rechner (Intel Paragon XP/S 10) der KFA Jülich entwickelt. Für die Simulation des Grundwasserflusses existiert seit längerem ein skalares Programm (auf der Basis einer Finite-Element-Diskretisierung des räumlichen Gebiets), das mittels verschiedener Verfahren für den Intel-Paragon parallelisiert werden konnte [6, 19, 42, 44]. Für den Schadstoff-Transport gibt es neben dem Finite-Element-Programm auch eine Simulation mit einem stochastischen Verfahren (*ParTrace*, Particle-Tracking) [30].

In [44] wurde 1995 eine Partitionierungsstrategie für das Schwarzsche Verfahren der Gebietszerlegung vorgestellt und implementiert. Dabei ging es vor allem um die gleichmäßige Verteilung der Rechenlast, möglichst geringe Anzahl der mehrfach berechneten Punkte (in den überlappenden Bereichen) und Minimierung des Kommunikationsaufwandes. Die globale Struktur sowie wesentliche Teile des ursprünglichen Simulationsprogramms blieben erhalten.

In dieser Arbeit wird die Idee verfolgt, die bei der Simulation von Grundwasserfluß und Schadstofftransport auftretenden Gleichungen mittels der numerischen Linienmethode [40] zu lösen und dabei Gebietszerlegungs-Vorkonditionierer zur Berechnung approximierter inverser Matrizen in einem semi-impliziten Verfahren einzusetzen.

Im folgenden ersten Kapitel wird das hier verwendete mathematische Modell vorgestellt, sowie die räumliche Diskretisierung mittels Finiter Volumina für die Linienmethode. Das dabei entstehende System gewöhnlicher Differentialgleichungen ist steif und unter Voraussetzungen, wie sie auch für die vollständige Diskretisierung parabolischer Aufgaben bekannt sind, stabil.

Das zweite Kapitel ist der Darstellung der semi-impliziten Mittelpunkregel von Bader und Deuffhard gewidmet, die hier zur Lösung des aus der Semi-Diskretisierung gewonnenen Anfangswertproblems benutzt wird. Von besonderem Interesse ist dabei die Tatsache, daß das Verfahren auch für gewisse Approximationen der Jacobi-Matrix stabil ist.

Im dritten Kapitel werden Gebietszerlegungsverfahren als Vorkonditionierungs-Techniken vorgestellt. Gebietszerlegungs-Vorkonditionierer werden beispielsweise in linearen Iterationsverfahren und in Krylov-Unterraum-Verfahren als Vorkonditionierer eingesetzt. Hier wird die Vorkonditionierung zur Konstruktion einer approximierten Inversen für das semi-implizite Verfahren benutzt.

Anmerkungen zur Implementierung folgen im vierten Kapitel. Die Realisierung eines parallelen Programms konnte zwar nicht abgeschlossen werden, hier werden aber die grundsätzlichen Erwägungen zur Implementierung des Verfahrens auf einem Parallelrechner do-

¹Auch wo die Regeln der Orthographie es dem Autor nicht nahelegen, wird er — in dem Glauben, dadurch das Verständnis zu erleichtern — auf die im Deutschen üblichen Wortzusammenballungen verzichten und übersichtlichere Bindestrich-Formen wählen.

kumentiert.

Die Darstellung einiger Simulationsergebnisse und eine Zusammenfassung nebst Ausblick schließen den inhaltlichen Teil der Arbeit ab.

1. Stofftransport durch poröse Medien

In diesem ersten Kapitel werden die mathematischen Hintergründe beleuchtet, vor denen sich die Computer-Simulation hydrologischer Vorgänge abspielt. Zunächst wird das für die Simulation der Schadstoffausbreitung in Boden und Grundwasser zu lösende Differentialgleichungs-System vorgestellt. Da im allgemeinen klassische Lösungen für solche Gleichungen nicht existieren, wird anschließend die schwache Formulierung der Aufgabe angegeben. Zur Existenz und Eindeutigkeit von Lösungen solcher Systeme werden Ergebnisse aus der Literatur zitiert.

Es folgt die Darstellung eines Finite-Volumen-Verfahrens zur Diskretisierung der Gleichungen bezüglich der Ortskoordinaten. Es wurde ein einfaches konservatives Verfahren gewählt, da die Genauigkeit der Ortsdiskretisierung nicht Gegenstand dieser Arbeit ist. Die Diskretisierung mit Finiten Volumen nimmt eine Zwischenstellung zwischen den Differenzenverfahren und Finite-Element-Methoden ein, weil hier einerseits die räumlichen Ableitungen durch Differenzenquotienten approximiert werden, andererseits aber die Diskretisierung auf der schwachen Formulierung der Gleichungen aufbaut.

Der Beschreibung des Volumen-Verfahrens folgen als Motivation für das in dieser Arbeit gewählte Vorgehen einige Überlegungen zum bei der Linienmethode entstehenden System von gewöhnlichen Differentialgleichungen. Dabei gilt das Hauptaugenmerk der Steifigkeit solcher semidiskreten Systeme, weil diese die Wahl geeigneter Verfahren für Systeme gewöhnlicher Anfangswertaufgaben wesentlich einschränkt.

1.1. Die zugrundeliegenden Gleichungen

Für den Fluß einer Flüssigkeit in einem porösen Medium und die Ausbreitung eines in der Flüssigkeit gelösten Stoffes findet man in der Literatur [15, 41]¹ ein System parabolischer Differentialgleichungen, einer Gleichung vom Typ der Wärmeleitungs-Gleichung für das Potential, in dem sich das Fluid bewegt, und einer Konvektions-Diffusionsgleichung für den Stofftransport.

¹zu einigen Hintergründen und zur Herleitung der klassischen Formulierung siehe Anhang A: „Einiges über Grundwasser-Hydrologie“, S. 55

1.1.1. Klassische Formulierung

Folgendes System nichtlinearer parabolischer Gleichungen ist zu lösen:

$$F(\psi) \frac{\partial \psi}{\partial t} + \operatorname{div} \mathbf{v} = 0 \quad \text{mit } \mathbf{v} := -\mathbf{K}(\psi) \operatorname{grad}(\psi + z) \quad (1.1)$$

$$R(c, \psi) \frac{\partial c}{\partial t} + \operatorname{div}(c\mathbf{v} - \mathbf{q}) = 0 \quad \text{mit } \mathbf{q} := \mathbf{D}(\mathbf{v}) \operatorname{grad} c \quad (1.2)$$

$$\text{in } \Omega \times (0, T)$$

$$\psi(\mathbf{x}, 0) = \psi_0(\mathbf{x}) \quad (1.3)$$

$$c(\mathbf{x}, 0) = c_0(\mathbf{x}) \quad (1.4)$$

$$\text{für } \mathbf{x} \in \Omega$$

$$\psi(\mathbf{x}, t) = \psi_D(\mathbf{x}, t) \quad \text{für } \mathbf{x} \in \Gamma_D, t \in (0, T) \quad (1.5)$$

$$\mathbf{n}(\mathbf{x}) \cdot \mathbf{v}(\mathbf{x}, t) = v_N(\mathbf{x}, t) \quad \text{für } \mathbf{x} \in \Gamma_N, t \in (0, T) \quad (1.6)$$

$$\text{mit } \Gamma_D \cup \Gamma_N = \partial\Omega \text{ und } \Gamma_D \cap \Gamma_N = \emptyset$$

$$c(\mathbf{x}, t) = 0 \quad \text{für } \mathbf{x} \in \partial\Omega, t \in (0, T) \quad (1.7)$$

Hierbei bezeichne $\Omega \subseteq \mathbb{R}^d$ mit $d \in \mathbb{N}, d \leq 3$ ein polygonal berandetes beschränktes Gebiet, $\mathbf{n}(\mathbf{x})$ den äußeren Normalenvektor in dem Randpunkt $\mathbf{x} \in \partial\Omega$, und der Endzeitpunkt des Zeitintervalls $(0, T)$ der Simulation sei positiv und endlich ($0 < T < \infty$).

Die gesuchten Lösungen sind reellwertige Funktionen auf $\Omega \times (0, T)$, nämlich $\psi(\mathbf{x}, t)$, im Falle der Grundwassersimulation das hydrostatische Potential und $c(\mathbf{x}, t)$, die Konzentration des gelösten Tracers im Grundwasser. Der Vektor $\mathbf{v} \in \mathbb{R}^d$ bezeichnet den sogenannten Filterfluß (nach dem Gesetz von Darcy) und z die geodätische Höhe des Punktes $\mathbf{x} \in \Omega$, also $z = -\mathbf{x} \mathbf{e}_g$, mit dem Einheitsvektor \mathbf{e}_g in Richtung der Gravitation. Die positive Diagonalmatrix \mathbf{K} mit $K_{i,j} : \mathbb{R} \times \Omega \rightarrow \mathbb{R}$ bezeichnet die anisotrope hydraulische Leitfähigkeit des Bodens; bezüglich des hydrostatischen Drucks ψ sind die Komponenten von \mathbf{K} glatte Funktionen. Der sogenannte hydraulische Speicherkoeffizient $F : \mathbb{R} \rightarrow \mathbb{R}_+$ bezeichnet im Falle des ungesättigten Mediums die Änderung des Wassergehalts mit dem Druck; da dieser Term bei Sättigung verschwindet und dadurch die Gleichung (1.1) elliptisch wird, hat man beim Übergang von der gesättigten zur ungesättigten Zone ein freies Randwertproblem zu lösen. Da dieses über das in dieser Arbeit zu leistende hinausgeht, werden im folgenden nur Probleme betrachtet, bei denen der freie Rand nicht auftritt. Die Rand- und Anfangswerte werden so gewählt, daß im ganzen Gebiet keine Sättigung erreicht wird, genauer: es gilt für die gesamte Simulation $0 < F_0 < F(\psi)$. Eine Alternative bietet die Simulation im gesättigten Medium, in diesem Falle ist die Gleichung (1.1) elliptisch. Der Transportfluß $c\mathbf{v} - \mathbf{q} \in \mathbb{R}^d$ setzt sich aus einem konvektiven und einem diffusiven Anteil zusammen. Der in dem Diffusionsteil \mathbf{q} auftretende Koeffizient \mathbf{D} bezeichnet eine positiv definite Matrix mit $D_{i,j} : \mathbb{R}^d \rightarrow \mathbb{R}$, die für den Dispersionstensor (vgl. Anhang A, S. 63) steht. Der Koeffizient $R : \mathbb{R}_+ \times \mathbb{R} \rightarrow \mathbb{R}$ beschreibt die durch die Sorption des gelösten Stoffes hervorgerufene Retardierung des Transportvorgangs im Boden, und es gilt $0 < \Theta_r < R(c, \psi)$ für alle $c > 0$ und $\psi \in \mathbb{R}$.

Die Anfangs- und Randbedingungen (1.3)–(1.7) seien konsistent, ferner seien die Funktionen $\psi_0 : \Omega \rightarrow \mathbb{R}$ und $c_0 : \Omega \rightarrow \mathbb{R}$ glatt in Ω . Da die Leitfähigkeit \mathbf{K} auch von der Permeabilität des Mediums abhängt, die möglicherweise lokal starke Sprünge aufweist, kann die Existenz klassischer Lösungen nicht zugesichert werden. Aus diesem Grunde haben wir zur schwachen Formulierung der Gleichungen überzugehen.

1.1.2. Schwache Formulierung

Die sogenannte schwache Form der Differentialgleichungen beruht auf der Einführung von verallgemeinerten (schwachen) Ableitungen und Sobolevräumen, die aus den Funktionen mit schwacher Ableitung bestehen [22]. Hier werden speziell die Sobolevräume $H^1(\Omega)$ und $H(\operatorname{div}, \Omega) := \{\mathbf{u} \in [L^2(\Omega)]^d; \operatorname{div} \mathbf{u} \in L^2(\Omega)\}$ benutzt.

Aus den Gleichungen (1.1) und (1.2) erhält man unter Berücksichtigung der Randbedingungen (1.5) – (1.7) die folgende schwache Formulierung:

Finde $\psi \in \{\varphi \in H^1(\Omega); \varphi|_{\Gamma_D} = \psi_D\}$ und $c \in H_0^1(\Omega)$, so daß für beliebige Testfunktionen $\varphi_1, \varphi_2 \in C^\infty(\Omega)$

$$\int_{\Omega} \varphi_1 F(\psi) \frac{\partial \psi}{\partial t} d\mathbf{x} + \int_{\partial\Omega} \varphi_1 \mathbf{n} \cdot \mathbf{v} d\sigma - \int_{\Omega} \operatorname{grad} \varphi_1 \cdot \mathbf{v} d\mathbf{x} = 0 \quad (1.8)$$

und

$$\int_{\Omega} \varphi_2 R(c, \psi) \frac{\partial c}{\partial t} d\mathbf{x} + \int_{\partial\Omega} \varphi_2 \mathbf{n} \cdot (c\mathbf{v} - \mathbf{q}) d\sigma - \int_{\Omega} \operatorname{grad} \varphi_2 \cdot (c\mathbf{v} - \mathbf{q}) d\mathbf{x} = 0 \quad (1.9)$$

gelten. Dabei wird die Randbedingung zweiter Art (1.6) als natürliche Randbedingung in dem Randintegral von (1.8) berücksichtigt. Auch für die Flüsse werden schwache Formulierungen angegeben: die Vektoren $\mathbf{v}, \mathbf{q} \in [H^1(\Omega)]^d$ erfüllen mit beliebigen Testfunktionen $\mathbf{u}_1, \mathbf{u}_2 \in H(\operatorname{div}, \Omega)$ die Integralgleichungen

$$\int_{\Omega} \mathbf{u}_1 \cdot \mathbf{K}^{-1}(\psi) \cdot \mathbf{v} d\mathbf{x} = - \int_{\partial\Omega} \mathbf{n} \cdot \mathbf{u}_1 (\psi + z) d\sigma + \int_{\Omega} \operatorname{div} \mathbf{u}_1 (\psi + z) d\mathbf{x} \quad (1.10)$$

sowie

$$\int_{\Omega} \mathbf{u}_2 \cdot \mathbf{D}^{-1}(\mathbf{v}) \cdot \mathbf{q} d\mathbf{x} = \int_{\partial\Omega} \mathbf{n} \cdot \mathbf{u}_2 c d\sigma - \int_{\Omega} \operatorname{div} \mathbf{u}_2 c d\mathbf{x}. \quad (1.11)$$

Entsprechende Gleichungen mit angepaßten Randwerten definieren die Lösungen auch auf jedem polygonal berandeten Teilgebiet von Ω .

Die Existenz schwacher Lösungen für das System von Fluß und Transport im porösen Medium ist nicht offensichtlich. Für Systeme, die aus der Modellierung des Flusses von mehreren Fluid-Phasen in einem porösen Medium entstehen, werden in [1] Existenz- und Eindeutigkeitsätze formuliert.

Dazu werden die Gleichungen mittels einer sog. Kirchhoff-Transformation in eine spezielle Form gebracht, an die dann Bedingungen für die Existenz und Eindeutigkeit schwacher Lösungen gestellt werden.

1.2. Semidiskretisierung und Linienmethode

Die Diskretisierung des Systems (1.8)–(1.11) mit den entsprechenden Anfangs- und Randbedingungen (1.3)–(1.7) wird in zwei Schritten vorgenommen: durch eine Finite-Volumen-(FV)-Diskretisierung des räumlichen Gebiets, die auf obiger schwachen Formulierung der Gleichungen aufbaut, erhält man ein steifes System von Anfangswertaufgaben gewöhnlicher Differentialgleichungen, die dann mit einem entsprechenden numerischen Verfahren (siehe Kapitel 2) gelöst werden können. Dieses Vorgehen ist bekannt als (vertikale) Linienmethode (*method of lines, MOL*).

1.2.1. Finite-Volumen-Verfahren

Wähle nun speziell $d = 2$, es wird also nur der räumlich zweidimensionale Fall betrachtet. Das Gebiet Ω ist im folgenden ein Rechteckgebiet in der (x, z) -Ebene, es werden die kartesischen Standard-Koordinaten verwendet, und zwar derart, daß $\mathbf{e}_2 = -\mathbf{e}_g$ ist (\mathbf{e}_g der Einheitsvektor in Gravitationsrichtung), also die geodätische Höhe $z = -\mathbf{x} \mathbf{e}_g = \mathbf{x} \mathbf{e}_2$ in (1.10) die zweite Ortskoordinate bezeichnet.

Zur Diskretisierung wird das Gebiet Ω in eine Anzahl von polygonal berandeten Volumestückchen (Kontrollvolumina) $G \subset \Omega$ zerlegt, in denen jeweils die Erhaltungsgleichungen (1.8) und (1.9) aufgestellt werden [28]. Für die Formulierung dieser Gleichungen in den Finiten Volumina wählt man als spezielle Testfunktion auf einem Kontrollvolumen G die charakteristische Funktion $\varphi = \chi_G$, und erhält so die Kontinuitätsgleichungen

$$\int_G F(\psi) \frac{\partial \psi}{\partial t} d\mathbf{x} = - \int_{\partial G} \mathbf{n} \mathbf{v} d\sigma \quad \text{und}$$

$$\int_G R(c, \psi) \frac{\partial c}{\partial t} d\mathbf{x} = - \int_{\partial G} \mathbf{n} (c \mathbf{v} - \mathbf{q}) d\sigma.$$

Im folgenden wird eine einfache reguläre Zerlegung durch ein Quadratgitter zugrundegelegt. Die Kontrollvolumina des Verfahrens sind Quadrate $G_{i,j}$ ($i = 1, \dots, n_x; j = 1, \dots, n_z$) mit der Seitenlänge h . Außerdem werden die Lösungen und die Parameter der Gleichungen innerhalb der quadratischen Volumen-Zellen als konstant angenommen (dazu Auswertung von ortsabhängigen Funktionen im Schwerpunkt $\mathbf{x}_{i,j}$ von $G_{i,j}$), das Randintegral der Flüsse setzt sich zusammen aus den Flüssen durch die vier Quadratseiten von $G_{i,j}$. Dadurch erhält man für die obigen Gleichungen die Form

$$F_{i,j} \frac{d}{dt} \psi_{i,j} h^2 = -(v_{i+1,j} - v_{i,j} + u_{i,j+1} - u_{i,j}) h \quad \text{und} \quad (1.12)$$

$$R_{i,j} \frac{d}{dt} c_{i,j} h^2 = -c(v_{i+1,j} - v_{i,j} + u_{i,j+1} - u_{i,j}) h$$

$$+ (q_{i+1,j} - q_{i,j} + p_{i,j+1} - p_{i,j}) h, \quad (1.13)$$

wobei die waagerechten Flüsse durch die rechten und linken Ränder von $G_{i,j}$ mit $v_{i+1,j}$ resp. $v_{i,j}$ und $q_{i+1,j}$ resp. $q_{i,j}$, die senkrechten Flüsse durch die oberen und unteren Ränder mit $u_{i+1,j}$ resp. $u_{i,j}$ und $p_{i+1,j}$ resp. $p_{i,j}$ bezeichnet werden.

Bei der Formulierung des ersten Summanden in (1.13) ist noch von Bedeutung, wie die jeweiligen Werte für c ermittelt werden. Ein einfaches stabiles Schema bietet da die upwind-Diskretisierung, bei der jeweils die Konzentration in dem Volumen-Element zugrundegelegt wird, wo der betreffende Fluß ausfließt. Auf diese Weise wird bei der Berechnung der Konzentration die Konzentrationsänderung aufgrund von Dispersion vernachlässigt, dafür erhält man aber ein für jeden Volumen-Querschnitt h stabiles Schema.

Man erhält die Flußkomponenten durch die Quadratseiten aus Diskretisierungen der Gleichungen (1.10) und (1.11) auf zwei versetzten Gittern von quadratischen Volumina $V_{i,j}$ für die waagerechten und $U_{i,j}$ für die senkrechten Flüsse mit $\mathbf{u}_k^{(l)} = \mathbf{e}_l$ ($l = 1, \dots, d, k = 1, 2$) anstelle der Testfunktionen. In diesen versetzten Volumina werden für die Approximation wiederum $\mathbf{v} = (v_{i,j}, u_{i,j})$ und $\mathbf{q} = (q_{i,j}, p_{i,j})$ als konstant angenommen.

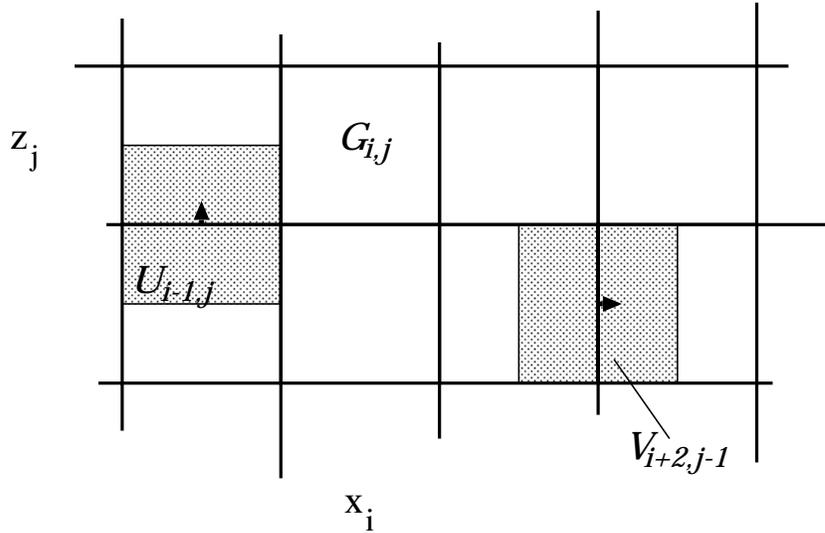


Abbildung 1.1.: Das versetzte Gitter des Volumenverfahrens

Beispielsweise gilt für die senkrechte Flußgröße $u_{i,j}$ zwischen den FV-Quadraten $G_{i,j-1}$ und $G_{i,j}$

$$\int_{U_{i,j}} \mathbf{e}_2 \cdot \mathbf{K}^{-1}(\psi) \cdot \mathbf{v} \, d\mathbf{x} = - \int_{\partial U_{i,j}} \mathbf{n} \cdot \mathbf{e}_2 (\psi + z) \, d\sigma$$

und mit den Annahmen zur Approximation folgt

$$\int_{U_{i,j} \cap G_{i,j-1}} \mathbf{e}_2 \cdot \mathbf{K}^{-1}(\psi) \cdot \mathbf{v} \, d\mathbf{x} + \int_{U_{i,j} \cap G_{i,j}} \mathbf{e}_2 \cdot \mathbf{K}^{-1}(\psi) \cdot \mathbf{v} \, d\mathbf{x} = (\psi_{i,j-1} + z_{j-1} - \psi_{i,j} - z_j) h \quad \text{bzw.}$$

$$u_{i,j} \left(\frac{1}{K_{2,2}(\psi_{i,j-1})} + \frac{1}{K_{2,2}(\psi_{i,j})} \right) \frac{h^2}{2} = (\psi_{i-1,j} - \psi_{i,j} - h) h. \quad (1.14)$$

Hieraus ergibt sich die Berechnung von \mathbf{v} auf dem versetzten Gitter mittels harmonisch gemittelter Leitfähigkeits-Koeffizienten.

Für die waagerechte Flußgröße $q_{i,j}$ zwischen den FV-Zellen $G_{i-1,j}$ und $G_{i,j}$ formuliert man

$$\int_{V_{i,j}} \mathbf{e}_1 \cdot \mathbf{D}^{-1}(\mathbf{v}) \cdot \mathbf{q} \, d\mathbf{x} = - \int_{\partial V_{i,j}} \mathbf{n} \cdot \mathbf{e}_1 c \, d\sigma$$

und erhält

$$q_{i,j} h^2 = \mathbf{D}(\mathbf{v})(c_{i-1,j} - c_{i,j})h. \quad (1.15)$$

Da man für die Berechnung von $\mathbf{D}(\mathbf{v})$ sämtliche Komponenten des Filterflusses \mathbf{v} benötigt, müssen die Komponenten in der jeweils fehlenden Richtung aus den Flüssen der entsprechenden überlappenden Zellen gemittelt werden.

Zu den Randwerten mag ein kurzer Hinweis genügen, sie werden auf möglichst einfache Weise in die Gleichungen eingebracht. Dirichlet-Randbedingungen werden durch die Einführung virtueller äußerer Zellen behandelt, deren Funktionswert durch lineare Fortsetzung über den Rand ermittelt wird; der durch Randbedingungen zweiter Art vorgegebene Fluß wird unmittelbar in die diskreten Erhaltungsgleichungen (1.12) und (1.13) eingesetzt.

1.2.2. Linienmethode

Nach dieser Semi-Diskretisierung wird das Anfangs-Randwert-Problem (1.1)–(1.7) durch ein System gewöhnlicher Differentialgleichungen approximiert. Mit der kanonischen Abzählung der Volumen-Zellen $k = (j-1)n_x + i$ ($i = 1, \dots, n_x, j = 1, \dots, n_z$) läßt sich dieses System folgendermaßen schreiben:

$$\begin{aligned} F^h(\Psi^h) \frac{d}{dt} \Psi^h &= A^h(\Psi^h) \quad \text{und} \\ R^h(c^h, \Psi^h) \frac{d}{dt} c^h &= B^h(c^h, \Psi^h) \end{aligned} \quad (1.16)$$

für $\Psi^h = (\Psi_k)_{k=1}^{n_x n_z}$ und $c^h = (c_k)_{k=1}^{n_x n_z}$ mit den Anfangsbedingungen

$$\begin{aligned} \Psi_{(j-1)n_x+i}(0) &= \Psi_0(x_i, z_j) \quad \text{und} \\ c_{(j-1)n_x+i}(0) &= c_0(x_i, z_j). \end{aligned} \quad (1.17)$$

Dabei bezeichnen $F^h(\Psi^h)$ und $R^h(c^h, \Psi^h)$ positive (und insb. nichtsinguläre) Diagonalmatrizen und $A^h(\Psi^h)$ sowie $B^h(c^h, \Psi^h)$ die den obigen Gleichungen entsprechenden diskreten elliptischen Differentialoperatoren.

Mit $y := (\Psi_1, \dots, c_{n_x n_z})^T$ und

$$f(y) := \begin{pmatrix} f_1(y) \\ f_2(y) \end{pmatrix} := \begin{pmatrix} (F^h(\Psi^h))^{-1} A^h(\Psi^h) \\ (R^h(c^h, \Psi^h))^{-1} B^h(c^h, \Psi^h) \end{pmatrix}$$

schreiben wir das Anfangswertproblem (1.16), (1.17) als autonomes System gewöhnlicher Anfangswertaufgaben

$$\dot{y} = f(y), \quad y(0) = y_0. \quad (1.18)$$

Dieses zusammengesetzte System ist im folgenden simultan zu lösen; aus seiner speziellen Form resultiert allerdings, daß die Lösung in zwei Einzelschritten durchgeführt werden kann. Die Jacobi-Matrix des Systems hat nämlich die spezielle Form

$$J = f_y = \begin{pmatrix} J_{1,1} & 0 \\ J_{2,1} & J_{2,2} \end{pmatrix} \quad (1.19)$$

mit Quadrat-Blöcken der Größe $n_x n_z$. Lineare Gleichungssysteme mit der Jacobi-Matrix, wie sie in dem semi-impliziten Lösungsverfahren (siehe Kapitel 2) auftreten, können also reduziert werden, so daß dann sukzessive zwei kleinere Systeme gelöst werden, die jeweils den einzelnen Gleichungen für den Wasserfluß resp. den Stofftransport entsprechen.

Außerdem kann das Spektrum der Jacobischen Matrix, das in dem nächsten Abschnitt zur Bewertung der Stabilität der Gleichung benutzt wird, als

$$\lambda(J) = \lambda(J_{1,1}) \cup \lambda(J_{2,2})$$

beschrieben werden.

1.3. Eigenschaften des semidiskreten Systems

Nachdem im vorigen Abschnitt die Semidiskretisierung bezüglich der räumlichen Koordinaten durchgeführt wurde, stellt sich die Frage nach den Eigenschaften des dadurch entstehenden Systems gewöhnlicher Differentialgleichungen.

Eine für die numerische Lösung sehr schwerwiegende Eigenschaft des Systems (1.18) ist seine Steifigkeit. Sie führt dazu daß explizite numerische Verfahren wegen ihrer mangelhaften Stabilität für die Behandlung dieser Aufgabe nicht geeignet sind. Allerdings eignen sich implizite oder semi-implizite Verfahren (wie das in Kapitel 2 vorgestellte Extrapolationsverfahren) für die Behandlung steifer Anfangswertprobleme.

Entscheidend für die sinnvolle numerische Behandlung eines Differentialgleichungssystems ist seine Stabilität, das heißt die Robustheit der Lösung gegen Störungen der Anfangswerte. Aus diesem Grunde folgen auch einige Bemerkungen zur Stabilität von Systemen, die mittels der numerischen Linienmethode aus parabolischen Differentialgleichungen gewonnen werden. Dazu werden einige Begriffe eingeführt, die auch bei der Bewertung des Extrapolationsverfahrens eine Rolle spielen werden.

1.3.1. Steifigkeit

Die Einschränkung der Zeitschrittweite h_t bei der numerischen Lösung von Anfangswertproblemen gewöhnlicher Differentialgleichungen kann auf zwei Erfordernissen beruhen: einerseits mag die Schrittweite aus Gründen der Genauigkeit der Approximation, andererseits zur Erhaltung der numerischen Stabilität nach oben beschränkt sein. Explizite Lösungsverfahren für gewöhnliche Anfangswertprobleme haben einen sehr eingeschränkten Stabilitätsbereich S , und für alle Eigenwerte λ der Jacobi-Matrix $f_y(y)$, die negativen

Realteil haben, muß die Zeitschrittweite die Bedingung $h_t \lambda \in S$ erfüllen, damit die Zeit-Iteration stabil bleibt.

Steife Systeme haben die Eigenschaft, daß die Eigenwerte der Jacobischen mit negativem Realteil weit gestreut sind. Dadurch schränken gerade die Eigenwerte mit stark negativem Realteil, die auf das Lösungsverhalten nur einen geringen Einfluß haben, die Schrittweiten ein.

Bei Systemen, die durch die Semidiskretisierung parabolischer Differentialgleichungen entstehen, werden die Realteile der betragsgrößten Eigenwerte sehr klein, sie gehen für $h \rightarrow 0$ gegen $-\infty$, so daß für kleinere Ortsschrittweiten h die Steifigkeit des Systems zunimmt.

In [35] wird dieser Sachverhalt exemplarisch für die eindimensionale lineare Diffusions-Konvektions-Gleichung

$$u_t + v u_x = D u_{xx} \quad (1.20)$$

für $0 \leq x \leq 1$ und $0 \leq t \leq T$

mit Anfangswerten $u(x, 0)$ und Dirichlet-Randwerten $u(0, t)$ und $u(1, t)$

mit Konstanten $v \geq 0$ und $D \geq 0$ für den konvektiven resp. diffusiven Transport dargestellt. Da diese auch als simples Modell für das hier betrachtete System gelten kann, werden einige Aspekte im folgenden kurz aufgeführt. Sobald überhaupt Diffusion stattfindet, also $D > 0$ ist, zeigt sich, daß die für die Stabilität der Ortsdiskretisierung dieser Gleichung bedeutsame sogenannte Gitter-Reynoldszahl (oder Peclet-Zahl) $R = vh/D$ auch im Zusammenhang mit der Linienmethode eine gewichtige Rolle spielt.

Verschwindet die Diffusion, d.h. $D = 0$, ändert sich der Charakter der Gleichung: sie hat nun hyperbolischen Typ, und Gleichungen dieses Typs werden immer explizit gerechnet, denn Schrittweiten-Beschränkungen für hyperbolische Aufgaben beruhen eher auf der geforderten Genauigkeit als auf der Stabilität [35]. Da in diesem Sinne die Steifigkeit wesentlich aus dem Diffusions-Anteil folgt, können semi-implizite Systeme auf eine einfache Weise so partitioniert werden, daß bei der numerischen Lösung der steife Anteil implizit und der konvektive Anteil explizit gerechnet werden.

Sei also im folgenden $D > 0$. Es ist bekannt, daß bei der Diskretisierung der ersten Ableitung u_x mit zentralen Differenzen die Ortsschrittweite h so gewählt werden muß, daß $R \leq 2$ ist, um Stabilität zu gewährleisten. Eine Upwind-Diskretisierung ist stabil für beliebige Peclet-Zahlen R .

Lemma 1 (Shampine)

(a) Wird die Diffusions-Konvektions-Gleichung (1.20) mit $D > 0$ mittels zentraler Differenzen diskretisiert und ist dabei die Ortsschrittweite $h = 1/(n + 1)$ so gewählt, daß die Peclet-Zahl $vh/D < 2$ ist, so sind alle Eigenwerte der Jacobi-Matrix des semi-diskreten Systems reell und negativ, und es gilt für den betragsgrößten Eigenwert

$$\lambda_n = -4Dh^{-2} + O(1).$$

(b) Wird (1.20) mit $D > 0$ mit Upwind-Differenzen diskretisiert, so sind die Eigenwerte der Jacobi-Matrix reell und negativ für alle Peclet-Zahlen R , und es gilt für den betragsgrößten

Eigenwert

$$\lambda_n = -4Dh^{-2} - \nu h^{-1} + O(1).$$

Beweis. Beide Differenzenverfahren führen auf Tridiagonalmatrizen; mit der Schreibweise der Jacobi-Matrix $A = \text{tridiag}(c, a, b)$ werden die konstanten Einträge in der Hauptdiagonalen (a), sowie in der unteren (c) bzw. oberen (b) Nebendiagonalen bezeichnet. Für die Eigenwerte einer solchen Tridiagonalmatrix wird in [35] die Formel

$$\lambda_s = a + 2\sqrt{bc} \cos\left(\frac{s\pi}{n+1}\right) \quad \text{für } s = 1, \dots, n \quad (*)$$

angegeben.

(a) Die Jacobi-Matrix hat im Fall der zentralen Differenzen für den Konvektions-Term die Form

$$J = \text{tridiag}\left(\frac{D}{h^2} + \frac{\nu}{2h}, -\frac{2D}{h^2}, \frac{D}{h^2} - \frac{\nu}{2h}\right).$$

Gilt für die Peclet-Zahl $R < 2$, so ist die Matrix J quasi-symmetrisch ($J_{k-1,k}J_{k,k-1} > 0$ für $2 \leq k \leq n-1$), das heißt es existiert eine Diagonalmatrix C , so daß CJC^{-1} symmetrisch ist. Folglich sind alle Eigenwerte reell.

Mit (*) und $h = 1/(n+1)$ folgt

$$\lambda_n = -\frac{2D}{h^2} + 2\frac{D}{h^2} \sqrt{1 - \frac{R^2}{4}} \cos(n\pi h),$$

und wegen $\cos(n\pi h) = \cos(\pi(1-h)) = -1 + O(h^2)$ folgt die Behauptung.

(b) Im Falle der Upwind-Differenzen für den Konvektions-Term erhält die Jacobi-Matrix die Form

$$J = \text{tridiag}\left(\frac{D}{h^2} + \frac{\nu}{h}, -\frac{2D}{h^2} - \frac{\nu}{h}, \frac{D}{h^2}\right).$$

Da $D > 0$ ist, gilt hier $J_{k-1,k}J_{k,k-1} > 0$ unabhängig von der Wahl von h , also ist für jedes R die Jacobi-Matrix quasi-symmetrisch und besitzt nur reelle Eigenwerte.

Wie oben kann man auch in diesem Falle den betragsgrößten Eigenwert explizit angeben:

$$\lambda_n = -\frac{2D}{h^2} - \frac{\nu}{h} + 2\frac{D}{h^2} \sqrt{1+R} \cos(\pi(1-h)).$$

□

Die Steifigkeit des resultierenden Systems ist also unabhängig davon, ob man zentrale Differenzen oder ein Upwind-Schema für die Ortsdiskretisierung wählt. In beiden Fällen wachsen die Beträge der betragsgrößten Eigenwerte mit derselben führenden Potenz der Ortschaftweite h .

Mit den hier benutzten Formeln kann man auch die Größenordnung der maximalen Eigenwerte angeben. Allerdings sind diese nicht sehr aufschlußreich und haben für die Bewertung der Steifigkeit keine besondere Bedeutung. Maximale Eigenwerte der zugeordneten Matrizen $1/2(J+J^T)$ spielen bei der Bewertung der Stabilität (siehe Lemma 5) eine Rolle.

1.3.2. Stabilität

Steife Systeme gewöhnlicher Anfangswertaufgaben

$$\dot{y} = f(t, y), \quad y(0) = y_0 \quad (1.21)$$

haben die Eigenschaft, daß die in der klassischen Stabilitätstheorie verwendete Lipschitz-Konstante sehr groß wird; bei semi-diskreten parabolischen Aufgaben wächst sie sogar unbeschränkt für $h \rightarrow +0$ (vgl. Lemma 1, für die Eigenwerte einer lokalen Jacobi-Matrix gilt $|\lambda| \leq L$). Aus diesem Grunde wird hier eine Verallgemeinerung der Lipschitz-Stetigkeit angegeben, die zudem ein hinreichendes Kriterium für die Stabilität nicht-linearer gewöhnlicher Differentialgleichungen liefert. Im Gegensatz zur klassischen Lipschitz-Konstanten kann die einseitige Lipschitz-Bedingung negativ werden.

Für die Untersuchung der Stabilität von Anfangswertproblemen ist auch die sogenannte logarithmische Matrix-Norm ein wichtiges Hilfsmittel; es handelt sich dabei nicht um eine Norm im üblichen Sinne, da sie auch negative Werte annehmen kann².

Im folgenden werden einige grundlegende Definitionen und Sätze aus der Stabilitätstheorie zitiert, die auch für die Bewertung des in Kapitel 2 behandelten Verfahrens von Bedeutung sind [39].

Eine hinreichende Bedingung für die Stabilität nichtlinearer Anfangswertprobleme ist ihre Kontraktivität. Sie besagt, daß zwei Lösungen zu verschiedenen Anfangswerten sich im Zeitverlauf nicht voneinander entfernen.

Definition 1 (Kontraktivität, dissipative Systeme)

Ein Differentialgleichungssystem (1.21) nennt man kontraktiv bezüglich der betrachteten Norm, falls für zwei Lösungen $u(t)$ und $v(t)$

$$\|u(t_2) - v(t_2)\| \leq \|u(t_1) - v(t_1)\| \quad \forall t_1, t_2 \text{ mit } 0 \leq t_1 \leq t_2 < \infty.$$

Systeme mit dieser Eigenschaft nennt man auch dissipativ.

Kontraktive Anfangswertprobleme haben stabile Lösungen.

Definition 2 (einseitige Lipschitz-Bedingung)

Seien $\langle \cdot, \cdot \rangle$ ein Skalarprodukt in \mathbb{R}^n mit der zugehörigen (induzierten) Norm $\|\cdot\|$ für $x \in \mathbb{R}^n$ und $l: \mathbb{R}_+ \rightarrow \mathbb{R}$ stückweise stetig. Dann genügt die Funktion f einer einseitigen Lipschitz-Bedingung, wenn gilt

$$\langle f(t, u) - f(t, v), u - v \rangle \leq l(t) \|u - v\|^2 \quad \forall t \in \mathbb{R}_+, u, v \in \mathbb{R}^n.$$

Die einseitige Lipschitz-Bedingung ist eine Verallgemeinerung der klassischen Lipschitz-Bedingung im folgenden Sinne:

²In den Stabilitätsbetrachtungen dieses und des folgenden Kapitels wird sich zeigen, daß man sich sogar wünscht, daß sie negativ wird.

Erfüllt die rechte Seite f eine klassische Lipschitz-Bedingung mit der Konstanten L , so gilt mit der Schwarz-Ungleichung für das Skalarprodukt $\langle \cdot, \cdot \rangle$

$$\begin{aligned} \langle f(t, u) - f(t, v), u - v \rangle &\leq \|f(t, u) - f(t, v)\| \|u - v\| \\ &\leq L \|u - v\|^2, \end{aligned}$$

das heißt, f erfüllt auch die einseitige Lipschitz-Bedingung mit derselben Konstante.

Satz 2

(a) Die Funktion $f(y)$ genüge einer einseitigen Lipschitz-Bedingung

$$\langle f(u) - f(v), u - v \rangle \leq l_0 \|u - v\|^2 \quad \forall u, v \in \mathbb{R}^n;$$

dann besitzt das Anfangswertproblem

$$\dot{y} = f(y), \quad y(0) = y_0$$

genau eine stetig differenzierbare Lösung $y(t)$ für jedes $t \in \mathbb{R}_+$.

(b) Sei $l(t)$ eine einseitige Lipschitz-Konstante für f , dann gilt für zwei beliebige Lösungen $u(t)$ und $v(t)$ von (1.21) zu verschiedenen Anfangswerten u_0 und v_0 die folgende Abschätzung in der induzierten Vektor-Norm:

$$\|u(t) - v(t)\| \leq \exp\left(\int_0^t l(\tau) d\tau\right) \|u_0 - v_0\| \quad \forall t \in \mathbb{R}_+.$$

(c) Sei $l(t) \leq 0$ für alle $t \in \mathbb{R}_+$. Dann ist das Differentialgleichungssystem (1.21) dissipativ und folglich jede Lösung $y(t)$ stabil³.

Beweis. Siehe z.B. [39], S. 198 f. □

Definition 3 (logarithmische Matrixnorm)

Sei $\|\cdot\|$ eine beliebige Vektornorm im \mathbb{R}^n , und bezeichne dasselbe Symbol auch eine zugeordnete Matrixnorm. Der Grenzwert

$$\mu[A] := \lim_{h \rightarrow +0} \frac{\|I + hA\| - 1}{h}$$

heißt die zugeordnete logarithmische Matrixnorm der Matrix $A \in \mathbb{R}^{n \times n}$.

Im Zusammenhang mit steifen Differentialgleichungen ist die logarithmische Matrixnorm deswegen von Interesse, weil sie für eine lineare Funktion $f(x) = Ax$ die kleinste einseitige Lipschitz-Konstante ist.

³Gilt sogar $l(t) \leq l_0 < 0$, so ist jede Lösung exponentiell und damit auch asymptotisch stabil [39].

Lemma 3

(a) In der Zeilensummen-Norm (Maximum-Norm) $\|\cdot\|_\infty$ gilt für die entsprechende logarithmische Matrixnorm

$$\mu_\infty[A] = \max_{i=1,\dots,n} \left(a_{i,i} + \sum_{j \neq i} |a_{i,j}| \right)$$

(b) Verwendet man zur Definition der logarithmischen Matrixnorm das euklidische Skalarprodukt, so gilt

$$\mu_2[A] = \lambda_{\max} \left(\frac{1}{2} (A + A^T) \right).$$

(c) Ist die Vektornorm $\|\cdot\|$ induziert von dem Skalarprodukt $\langle \cdot, \cdot \rangle$, so gilt

$$\mu[A] = \max_{x \neq 0} \frac{\langle Ax, x \rangle}{\langle x, x \rangle}$$

Die logarithmische Matrixnorm einer Matrix $A \in \mathbb{R}^{n \times n}$ ist demnach die kleinste reelle Zahl μ , für die gilt

$$\langle Ax, x \rangle \leq \mu \|x\|^2 \quad \forall x \in \mathbb{R}^n$$

Beweis. Siehe [39] □

Satz 4

(a) Seien $\|\cdot\|$ eine Norm in \mathbb{R}^n und $l(t)$ eine in \mathbb{R}_+ stetige Funktion. Ferner sei f stetig differenzierbar mit beschränkten Ableitungen und

$$\mu[f_y(t, y)] \leq l(t) \quad \text{für } t \geq 0 \text{ und } y \in \mathbb{R}^n.$$

Dann gilt für zwei Lösungen $u(t)$ und $v(t)$ die Abschätzung von Satz 2 (b).

(b) Die Funktion f sei stetig differenzierbar und die Norm $\|\cdot\|$ sei durch $\langle \cdot, \cdot \rangle$ induziert. Dann sind die folgenden Aussagen (1.22) und (1.23) äquivalent:

$$\mu[f_y(t, y)] \leq l(t) \quad \forall t \in \mathbb{R}_+, y \in \mathbb{R}^n \tag{1.22}$$

$$\langle f(t, u) - f(t, v), u - v \rangle \leq l(t) \|u - v\|^2 \quad \forall t \in \mathbb{R}_+, u, v \in \mathbb{R}^n \tag{1.23}$$

(c) Gilt $\mu[f_y(t, y)] \leq 0$, so ist das System (1.21) dissipativ und also stabil.

Beweis. Siehe [39]. □

Im folgenden wird es also um die Frage der Dissipativität der semi-diskreten Gleichungen gehen.

In der Maximum-Norm sind die logarithmischen Matrix-Normen der Jacobi-Matrizen aus Lemma 1 unter den dortigen Voraussetzungen gleich null, wie man mit der Formel aus Lemma 3 leicht nachrechnet.

Da man allerdings in Situationen geraten kann, in denen eine echt negative logarithmische Norm benötigt wird (vgl. Abschnitt 2.3), wird hier das folgende Lemma 5 bewiesen.

Im Falle der Ortsdiskretisierung mittels zentraler Differenzen wird weiterhin die Schrittweite so gewählt, daß für die Gitter-Peclet-Zahl $R = vh/D < 2$ gilt, auch wenn v bei der Berechnung der „euklidischen“ logarithmischen Matrixnorm keine Rolle spielt.

Lemma 5

(a) Wird zur Diskretisierung der Modellaufgabe (1.20) mit $D > 0$ ein zentrales Differenzenschema benutzt, so ist das entstehende semi-diskrete System dissipativ; dabei ist

$$\mu_2[J] = -D\pi^2 + O(h^2)$$

der maximale Eigenwert der Jacobischen für die entsprechende Wärmeleitungsaufgabe (mit Leitfähigkeits-Koeffizienten D).

(b) Wird (1.20) mit $D > 0$ mittels Upwind-Differenzen diskretisiert, so ist das entstehende semi-diskrete System dissipativ in der euklidischen Norm, und es gilt für die logarithmische Matrixnorm der Jacobi-Matrix:

$$\mu_2[J] = -D\pi^2 - \frac{v\pi^2 h}{2} + O(h^2).$$

Beweis. (a) Nach Lemma 3 berechnet sich die logarithmische Matrixnorm als der maximale Eigenwert der symmetrischen Matrix

$$J_s = \frac{1}{2}(J + J^T) = \text{tridiag}\left(\frac{D}{h^2}, -\frac{2D}{h^2}, \frac{D}{h^2}\right),$$

also mit der in dem obigen Beweis angegebenen Formel für die Eigenwerte einer Tridiagonalmatrix (und den ersten Gliedern der Entwicklung des Cosinus):

$$\begin{aligned} \mu[J] = \lambda_1(J_s) &= -\frac{2D}{h^2} + \frac{2D}{h^2} \cos(\pi h) \\ &= \frac{2D}{h^2} \left(-1 + 1 - \frac{\pi^2 h^2}{2} + \frac{\pi^4 h^4}{24} - O(h^6)\right) \\ &= -D\pi^2 + O(h^2). \end{aligned}$$

(b) Im Falle der Upwind-Differenzen gilt

$$J_s = \frac{1}{2}(J + J^T) = \text{tridiag}\left(\frac{D}{h^2} + \frac{v}{2h}, -\frac{2D}{h^2} - \frac{v}{2h}, \frac{D}{h^2} + \frac{v}{2h}\right),$$

und auch hier wird wie oben eingesetzt

$$\begin{aligned} \mu[J] = \lambda_1(J_s) &= -\frac{2D}{h^2} - \frac{v}{h} + \left(\frac{2D}{h^2} + \frac{v}{h}\right) \cos(\pi h) \\ &= \left(\frac{2D}{h^2} + \frac{v}{h}\right) \left(-1 + 1 - \frac{\pi^2 h^2}{2} + \frac{\pi^4 h^4}{24} - O(h^6)\right) \\ &= -D\pi^2 - \frac{v\pi^2 h}{2} + O(h^2). \quad \square \end{aligned}$$

Inwieweit lassen sich nun die Betrachtungen dieses Abschnittes auf das nichtlineare semi-diskrete System aus den Modellgleichungen für die Grundwasser- und Schadstoff-Simulation übertragen?

Zunächst ist es hilfreich, daß das Spektrum der Jacobi-Matrix des vollständigen Systems auf natürliche Weise in die einzelnen Spektren für die Grundwasser- und die Schadstoffgleichung zerfällt. Dadurch wird es möglich, die Gleichungen gesondert zu betrachten.

Beide Gleichungen sind elliptisch-parabolisch mit nicht-verschwindenden Diffusionskoeffizienten \mathbf{K} bzw. \mathbf{D} (vgl. S. 6). Es wird eine stabile upwind-Diskretisierung für den konvektiven Transport in der Diskretisierung der Konvektions-Diffusions-Gleichung (1.13) eingesetzt, damit lokal aufgrund der Ortsdiskretisierung die Stabilität gewahrt wird.

2. Die numerische Lösung des semidiskreten Systems

Wie im vorigen Kapitel dargestellt wurde, sind die bei der Linienmethode für parabolische Differentialgleichungen auftretenden Systeme von Anfangswertproblemen gewöhnlicher Differentialgleichungen im allgemeinen steif. Es wird außerdem im weiteren Verlauf angenommen, daß das zu lösende semidiskrete System kontraktiv ist.

Im folgenden wird ein Verfahren vorgestellt, das für die Lösung solcher steifer Anfangswertsysteme geeignet ist und zudem einen interessanten Zugang zur Parallelisierung des Lösungsverfahrens eröffnet.

Verfahren zur Behandlung steifer Anfangswertprobleme mit einer differenzierbaren rechten Seite beruhen auf der Lösung eines einzigen oder mehrerer linearer Gleichungssysteme mit der Jacobi-Matrix f_y des Systems

$$\dot{y} = f(y) \text{ und } y(0) = y_0 \quad (2.1)$$

in jedem Zeitschritt. Dieses bedeutet einen erhöhten Aufwand bei der Berechnung der Zwischenlösungen, aber man erhält dadurch Verfahren mit unbeschränkten Stabilitätsbereichen.

2.1. Die semi-implizite Mittelpunkt-Regel

Die Motivation der hier vorgestellten Methode von Bader und Deuffhard [4] besteht in der Erwartung, durch die Transformation

$$y(t) = e^{At} c(t)$$

mit einer Näherung der lokalen Jacobi-Matrix im Anfangspunkte

$$A \approx f_y(y_0)$$

die Steifigkeit neutralisieren zu können [26], S.145. Das zugehörige transformierte System

$$\dot{c} = e^{-At} (f(e^{At} c(t)) - A e^{At} c(t)) ,$$

bzw.

$$\dot{y} - Ay = \bar{f}(y) \quad \text{mit} \quad \bar{f}(y) := f(y) - Ay \quad (2.2)$$

wird mit der expliziten Mittelpunkregel diskretisiert, wobei man die auftretenden Matrix-Exponentialfunktionen $e^{\pm Ah}$ durch die ersten Glieder ihrer Taylorentwicklung um den Nullpunkt $E(Ah) = I \pm hA$ ersetzt. Das auf diese Weise gewonnene Verfahren heißt die semi-implizite Mittelpunkt-Regel. Hier und im folgenden bezeichne h die Zeitschrittweite des Lösungsverfahrens.

Die Neutralisierung der Steifigkeit — ergo die Stabilisierung des expliziten Verfahrens — beruht darauf, daß nach der Transformation die Funktion \bar{f} eine möglichst kleine Lipschitz-Konstante hat. Diese Lipschitz-Konstante des nichtlinearen Anteils, der nach der formalen Linearisierung (2.2) übrigbleibt, ist zugleich ein Maß für die Linearität der ursprünglichen Aufgabe. Es gilt nämlich mit der Lipschitz-Bedingung

$$\|\bar{f}(u) - \bar{f}(v)\| \leq \bar{L} \|u - v\| \quad \forall u, v$$

für $A = f_y(y_0)$ die Ungleichung

$$\|f_y(y) - f_y(y_0)\| = \|f_y(y) - A\| = \|\bar{f}_y\| \leq \bar{L}, \quad (2.3)$$

woraus für den linearen Fall die mögliche Wahl $\bar{L} = 0$ folgt. Für die Stabilität des Verfahrens ist nun von entscheidender Bedeutung, daß im nichtlinearen Fall die Lipschitz-Konstante \bar{L} in einer Umgebung von y_0 klein genug bleibt.

Mit einem semi-impliziten Euler-Schritt als Startschritt und einer abschließenden Glättung erhält man durch folgende Iteration aus der semi-impliziten Mittelpunkt-Regel ein Verfahren, welches auf eine h^2 -Extrapolation führt:

$$\begin{aligned} \eta_0 &:= y(t_0) \\ (I - hA)\eta_1 &= \eta_0 + h\bar{f}(\eta_0) \\ (I - hA)\eta_{k+1} &= (I + hA)\eta_{k-1} + 2h\bar{f}(\eta_k) \quad \text{für } k = 1, \dots, 2m \\ S_{2m} &:= \frac{1}{2}(\eta_{2m+1} + \eta_{2m-1}). \end{aligned} \quad (2.4)$$

Es ist dann $S(t_0 + H, h) := S_{2m}$ eine Näherung für $y(t_0 + H)$ mit der globalen Schrittweite $H = 2mh$.

Für diese Näherung gilt der folgende

Satz 6 (Bader/Deuffhard)

Seien A eine beliebige Matrix und $f(y) \in C^{2N+2}$ für eine natürliche Zahl $N \in \mathbb{N}$. Für festes $t = t_0 + H$ hat die durch (2.4) angegebene numerische Näherung für $0 \leq h \leq H$ eine asymptotische Fehlerentwicklung der Form

$$y(t) - S(t, h) = \sum_{j=1}^N \zeta_j(t) h^{2j} + C_{N+1}(t, h) h^{2N+2} \quad (2.5)$$

mit gleichmäßig beschränktem $C_{N+1}(t, h)$.

Für $A = 0$ erhält man das entsprechende explizite Verfahren und es gilt in diesem Falle $\zeta_j(t_0) = 0$. Im allgemeinen sind die $\zeta_j(t_0) \neq 0$.

Beweis. Siehe [4] bzw. [26], S. 146 f. □

Die Entwicklung des Extrapolationsverfahrens aus (2.4) wird im folgenden Abschnitt 2.2 dargestellt. Dort werden auch die Konsistenz und die grundlegenden linearen Eigenschaften des Verfahrens mit der exakten Jacobi-Matrix angegeben.

Da der Satz 6 für beliebige Matrizen A eine asymptotische Entwicklung (2.5) des Fehlers in geraden Potenzen der Schrittweite h zusichert, erhält man für Näherungen der Jacobi-Matrix jeweils konsistente Extrapolationsverfahren, die „zwischen“ dem semi-impliziten Verfahren mit der exakten Jacobi-Matrix und dem expliziten Verfahren liegen. Diese Matrizen A dienen der Stabilisierung der Methode, sodaß man umso stabilere Verfahren erwarten kann, je besser die Matrix A die exakte Jacobische approximiert, wie schon die Ungleichung (2.3) nahelegt. Für diesen (auch im nichtlinearen Fall) stabilisierenden Effekt kann man hinreichende Bedingungen formulieren, die in Abschnitt 2.3 dargestellt werden.

2.2. Extrapolation

Das Extrapolationsverfahren beruht auf der Durchführung des obigen Verfahrens (2.4) mit verschiedenen, sukzessive abnehmenden, Teilschrittweiten für denselben globalen Zeitschritt von t_0 nach $t_0 + H$. Man erhält dann eine Folge von Näherungswerten $S(t_0 + H, h_i)$.

Dazu definiert man eine Folge von Teilschrittweiten

$$h_i := \frac{H}{n_i} \quad \text{mit } i = 1, 2, \dots$$

durch die Angabe der streng monoton wachsenden Folge von natürlichen Zahlen $\{n_i\}$, wobei im Falle der h^2 -Extrapolation mit der semi-impliziten Mittelpunkregel die n_i wie in (2.4) gerade gewählt werden sollten [4].

Auf diese Weise erhält man eine Folge von Näherungslösungen, die jeweils eine asymptotische Fehlerentwicklung der Form (2.5) besitzen. Mit diesen Werten definiert man auf folgende Weise neue Näherungen höherer Konsistenzordnung:

$$\begin{aligned} T_{i,1} &:= S(t_0 + H, h_i) \quad \text{für } i = 1, 2, \dots \\ T_{i,k} &:= T_{i,k-1} + \frac{T_{i,k-1} - T_{i-1,k-1}}{(n_i/n_{i-k+1})^2 - 1} \quad \text{für } k = 2, \dots, i \text{ (falls } i \geq 2). \end{aligned} \quad (2.6)$$

Mittels dieser Rekursionsformel wird der Wert eines Interpolationspolynoms in h^2 durch die Stützpunkte $(h_i, T_{i,1})$ an der Stelle $h = 0$ ausgewertet (Interpolation im Nullpunkt nach Neville-Aitken, [38]). Üblicherweise werden die so gewonnenen Näherungswerte in einem dreieckigen Schema der folgenden Form angeordnet:

$$\begin{array}{cccc} T_{1,1} & & & \\ T_{2,1} & T_{2,2} & & \\ T_{3,1} & T_{3,2} & T_{3,3} & \\ \vdots & \vdots & \vdots & \ddots \end{array} \quad (2.7)$$

Das folgende Lemma besagt, daß in diesem Schema die Ordnung der Näherungslösungen von Spalte zu Spalte um zwei ansteigt. Außerdem werden durch das Extrapolationsverfahren, insbesondere bei einer geschickten Wahl der Schrittweitenfolge, $A(\alpha)$ -stabile Verfahren mit recht großen Winkeln definiert: $\alpha \geq 86^\circ$ für die ersten sieben Spalten des Tableaus (2.7) [10]. Für $i \leq 3$ bzw. $k = 1$ sind die durch $T_{i,k}$ definierten Verfahren A-stabil. Hat ein Verfahren die Stabilitätsfunktion $R(z)$, so gilt für einen Schritt des Verfahrens $y_1 = R(h\lambda)y_0$. Das Stabilitätsgebiet ist dann definiert als $S = \{z \in \mathbb{C}; |R(z)| \leq 1\}$, das heißt für Schrittweiten mit $h\lambda \in S$ ist das Verfahren stabil. Bei beschränkten Stabilitätsgebieten bedeutet dies eine Einschränkung für die Zeitschrittweite (vgl. Absatz 1.3).

Lemma 7 (Konsistenzordnung und lineare Stabilität)

Die k -te Spalte des Extrapolations-Tableaus (2.7) repräsentiert unabhängig von der Wahl der Matrix A Verfahren der Ordnung $2k - 1$.

Setzt man als Matrix A die exakte Jacobische Matrix ein, so ist die Stabilitätsfunktion für $T_{i,1}$ ($i \geq 1$) gegeben durch

$$R_{i,1}(z) = \frac{1}{(1 - z/n_i)^2} \left(\frac{1 + z/n_i}{1 - z/n_i} \right)^{n_i/2-1}. \quad (2.8)$$

Beweis. Die Konsistenzordnung von $T_{i,k}$ gewinnt man aus dem Vergleich des Interpolationspolynoms der Werte $T_{i-k+1,1}, \dots, T_{i,1}$ mit der Fehlerentwicklung aus Satz 6. Die Interpolationsbedingung $P(h_j) = T_{j,1}$ für das Polynom $P(h) = e_0 - e_1 h^2 - \dots - e_{k-1} h^{2k-2}$ liefert ein lineares Gleichungssystem für $e_0, e_1 H^2, \dots, e_{k-1} H^{2k-2}$, die asymptotische Entwicklung in der Form

$$T_{i,1} = S(t_0 + H, h_i) = y(t_0 + H) - \zeta_1(t_0 + H) h_i^2 - \dots - \zeta_{k-1}(t_0 + H) h_i^{2k-2} + \Delta_i$$

$$\Delta_i = \zeta_k(t_0 + H) h_i^{2k} - C_{k+1}(t_0 + H, h_i) h_i^{2k+2} = O(H^{2k})$$

ergibt dasselbe Gleichungssystem für die Koeffizienten $y(t_0 + H)$ und $\zeta_j(t_0 + H) H^{2j}$ für $j = 1, \dots, k - 1$ mit einer Störung der Ordnung $O(H^{2k})$ (das ist das Restglied Δ_i) auf der rechten Seite. Die Koeffizientenmatrix beider Systeme,

$$A = \begin{pmatrix} I & n_i^{-2} I & \dots & n_i^{-2k+2} I \\ \vdots & \vdots & \vdots & \vdots \\ I & n_{i-k+1}^{-2} I & \dots & n_{i-k+1}^{-2k+2} I \end{pmatrix},$$

ist regulär. Die quadratischen Blöcke haben die Größe des Systems (Anzahl der Elemente von y und $T_{i,k}$). Durch Subtraktion erhält man

$$|y(x_0 + H) - e_0| \leq \|A^{-1}\|_\infty \max |\Delta_i| = O(H^{2k}).$$

Die Stabilitätsfunktion gewinnt man aus der Betrachtung der semi-impliziten Mittelpunkt-Regel (2.4) für die Testgleichung $\dot{y} = \lambda y$ [24, 26]. □

Die Stabilitätsfunktionen der $T_{i,k}$ erhält man durch Einsetzen in die Rekursionsformel (2.6), zum Beispiel mit einem Formelmanipulationssystem.

Im folgenden wird es um die Frage der numerischen Stabilität des semi-impliziten Verfahrens gehen, zunächst weiterhin im linearen Fall, im Abschnitt 2.3 dann für nichtlineare gewöhnliche Differentialgleichungen. Sei also das lineare Anfangswertproblem

$$\dot{y} = Ay \quad y(0) = y_0 \quad (2.9)$$

gegeben.

Gegenstand der numerischen Stabilitätsuntersuchung ist die Frage, ob die Kontraktivität der exakten Lösung die Kontraktivität der numerischen Näherung nach sich zieht [24].

Bezeichnet μ die logarithmische Matrixnorm, d.h. μ sei die kleinste reelle Zahl mit

$$\langle Ax, x \rangle \leq \mu \|x\|^2 \quad \forall x \in \mathbb{R}^n,$$

dann erfüllt die exakte Lösung von (2.9)

$$\|y(t+h)\| \leq e^{\mu h} \|y(t)\|$$

(vgl. Satz 4, S. 16).

Ein Analogon für numerische Verfahren mit der rationalen Stabilitätsfunktion $R(z)$ wird in [24] angegeben:

Satz 8 (Hairer/Bader/Lubich)

Sei $R(z)$ eine rationale Funktion, $A \in \mathbb{R}^{n \times n}$ eine Matrix mit

$$\langle Ax, x \rangle \leq \mu \|x\|^2 \quad \forall x \in \mathbb{R}^n.$$

Dann gilt

$$\|R(A)\| \leq \varphi_R(\mu),$$

wobei die Funktion

$$\varphi_R(\mu) := \sup\{R(z); \operatorname{Re} z \leq \mu\}.$$

Beweis. Siehe [24]. □

Mit $y_1 = R(hA)y_0$ gilt nun auch $y_1 \leq \varphi_R(\mu[A]h)y_0$ für die logarithmische Matrixnorm $\mu[A]$.

Für die Stabilitätsfunktionen $R_{i,1}$ aus Lemma 7 gilt [24]:

$$\begin{aligned} \varphi_{i,1} &:= \varphi_{R_{i,1}} = \sup \left\{ \frac{1}{(1-z/n_i)^2} \left(\frac{1+z/n_i}{1-z/n_i} \right)^{n_i/2-1}; \operatorname{Re}(z) \leq x \right\} \\ &= \begin{cases} R_{i,1}(x) & \text{für } x \leq x_1 \text{ oder } x_2 \leq x \leq n_i \\ c_i/|x| & \text{für } x_1 \leq x \leq x_2 \end{cases}, \end{aligned}$$

wobei

$$x_{1,2} = -\frac{n_i^2}{2} \left(1 \pm \sqrt{1 - \frac{4}{n_i^2}} \right)$$

und c_i so gewählt sei, daß $\varphi_{i,1}$ stetig ist.

2.3. Kontraktivität und Stabilität im nichtlinearen Fall

Die in dem vorigen Kapitel (Abschnitt 1.3.2) dargestellten Stabilitätskriterien finden auch hier, in der nichtlinearen Stabilitätsanalyse des Verfahrens, eine Anwendung. Es sind Kriterien dafür anzugeben, daß die Dissipativität der semidiskreten Gleichungen sich auf die diskretisierten Gleichungen überträgt.

Vor allem sind Kriterien für die Stabilität des Verfahrens für den Fall anzugeben, daß die Matrix A in dem Verfahren lediglich eine Approximation der Jacobischen ist. Hintergrund der folgenden Untersuchung ist die Stabilisierung des expliziten Verfahrens gemäß (2.2) durch formale Linearisierung.

Von grundlegender Bedeutung sind hier die logarithmische Matrixnorm der Stabilisierungsmatrix A und die klassische Lipschitz-Konstante für den „Rest“ \tilde{f} :

Lemma 9 ([24])

Für die formal linearisierte Aufgabe

$$\dot{y} = Ay + \tilde{f}(y) \quad (2.10)$$

gilt mit

$$\begin{aligned} \langle Ax, x \rangle &\leq \mu \|x\|^2 \quad \forall x \in \mathbb{R}^n \\ \|\tilde{f}(u) - \tilde{f}(w)\| &\leq \bar{L} \|u - w\| \quad \forall u, w \in \mathbb{R}^n \end{aligned} \quad (2.11)$$

die Abschätzung

$$\|u(t+h) - w(t+h)\| \leq e^{(\mu + \bar{L})h} \|u(t) - w(t)\|$$

für zwei Lösungen u und w der Aufgabe (2.10).

Beweis. Sei $g(t) := \|u(t) - w(t)\|^2 = \langle u(t) - w(t), u(t) - w(t) \rangle$, dann ist g stetig differenzierbar mit

$$\begin{aligned} \dot{g}(t) &= 2 \langle \dot{u}(t) - \dot{w}(t), u(t) - w(t) \rangle \\ &= 2 \langle Au(t) + \tilde{f}(u(t)) - Aw(t) - \tilde{f}(w(t)), u(t) - w(t) \rangle \\ &= 2 \langle Au(t) - Aw(t), u(t) - w(t) \rangle + 2 \langle \tilde{f}(u(t)) - \tilde{f}(w(t)), u(t) - w(t) \rangle \\ &\leq 2(\mu + \bar{L}) \|u(t) - w(t)\|^2, \end{aligned}$$

da \bar{L} insbesondere eine einseitige Lipschitz-Konstante ist. Es gilt also insgesamt:

$$\dot{g}(t) \leq 2(\mu + \bar{L})g(t). \quad (2.12)$$

Nun sei ferner $E(t) := \exp(-2(\mu + \bar{L})t)$; dann folgt mit der Abschätzung (2.12)

$$\begin{aligned} \frac{d}{dt}(g(t)E(t)) &= \dot{g}(t)E(t) + g(t)\dot{E}(t) \\ &= \dot{g}(t)E(t) - 2(\mu + \bar{L})g(t)E(t) \\ &= E(t)(\dot{g}(t) - 2(\mu + \bar{L})g(t)) \\ &\leq 0. \end{aligned}$$

Also ist die Funktion $g(t)E(t)$ monoton fallend, ergo

$$g(t+h)E(t+h) \leq g(t)E(t) \quad \text{für } h > 0,$$

und damit

$$g(t+h) \leq g(t) \frac{E(t)}{E(t+h)} = \exp(2(\mu + \bar{L})h)g(t).$$

Die obige Behauptung folgt nun aus dieser Ungleichung mit den Quadratwurzeln der beiden Seiten. \square

Zu dem Beweis ist noch zu bemerken, daß in [24] zwar eine klassische Lipschitz-Konstante für \tilde{f} vorausgesetzt wurde, daß sie aber nur im Sinne der einseitigen Lipschitz-Konstante gebraucht wird.

Aus diesem Satz folgt nun direkt, daß die Bedingung

$$\mu + \bar{L} \leq 0 \tag{2.13}$$

hinreichend für die Kontraktivität des transformierten Systems ist. Verschwindet die Lipschitz-Konstante des Rests nicht (das ist bei nichtlinearen Aufgaben der Regelfall — vgl. (2.3)), so muß eine echt negative logarithmische Matrix-Norm für A gefunden werden.

Es ist nun von Interesse, unter welchen Bedingungen die numerische Lösung ebenfalls kontrahierend ist, und insbesondere: welcher Art sind die Schrittweitenbeschränkungen, die eine stabile Lösung gewährleisten?

In [24] wird auch für die Bedingung (2.13) ein numerisches Analogon angegeben.

Definition 4

Es werde für ein semi-implizites Verfahren eine Menge W definiert als die Menge aller positiven Zahlen ω , für die gilt:

erfüllt das formal linearisierte System (2.10) die Bedingung (2.11) mit $\mu + \omega\bar{L} \leq 0$ für ein $\omega \in W$, dann existiert eine Konstante C , die nur von dem Verfahren und von ω abhängt, nicht aber von der Steifigkeit der Differentialgleichung, so daß das Verfahren stabil ist für Schrittweiten H mit

$$H\bar{L} \leq C. \tag{2.14}$$

Als ω_0 wird dann das Infimum von W bezeichnet; ist $W = \emptyset$, dann setzt man $\omega_0 = \infty$.

Die Autoren formulieren aufgrund einer ausführlichen Analyse für die extrapolierte semi-implizite Mittelpunkregel die Bedingung

$$\mu + \omega_{i,k}\bar{L} \leq 0 \tag{2.15}$$

als hinreichendes Kriterium für die Kontraktivität der numerischen Lösung.

Die empirischen minimalen Werte für ω werden angegeben mit $\omega_{i,1} = 1.0$ für alle i , $\omega_{2,2} = 1.14$, $\omega_{3,2} = 1.97$, $\omega_{3,3} = 2.09$ für die von Bader und Deuffhard vorgeschlagene Schrittweitenfolge F_α [24, 4]. Für die sogenannte harmonische Schrittweitenfolge $F_H = \{2, 4, 6, \dots\}$

errechnet man leicht größere Werte: $\omega_{2,2} = 1.37$, $\omega_{3,2} = 2.23$, $\omega_{3,3} = 2.64$. Die Werte für ω zu Näherungen, die außerhalb der ersten Spalte des Tableaus stehen, sind in jedem Falle größer als eins: $\omega_{i,k} > 1$ für $k \geq 2$.

Die Verfahren der ersten Spalte und der ersten drei Zeilen des Extrapolationstableaus besitzen also eine nichtleere Menge W im obigen Sinne (daraus folgt ihre A -Stabilität).

2.4. Ordnungs- und Schrittweitensteuerung

Durch das Extrapolationsverfahren wird eine Methode zur Lösung von Systemen gewöhnlicher Anfangswertaufgaben definiert, die nicht nur variable Zeitschrittweiten, sondern auch eine Anpassung der Konsistenzordnung des Verfahrens ermöglicht. Bei der Steuerung der Ordnung wird eine Schätzung des Aufwandes zugrundegelegt, die auf der Zahl der Funktionsauswertungen in der extrapolierten semi-impliziten Mittelpunkregel basiert. Dazu wird im Anschluß an die folgenden allgemeineren Erörterungen über Fehlerschätzung und Schrittweiten-Vorhersage die Methode zur Minimierung des relativen Aufwandes (bezogen auf einen Einheits-Zeitschritt) dargestellt.

2.4.1. Schrittweiten

Eine Schrittweitensteuerung nimmt man vor, weil bei zu großen Schrittweiten ein viel zu großes Extrapolations-Schema berechnet werden muß beziehungsweise eine erwartete Genauigkeit bei einem beschränkten Tableau gar nicht erreicht wird. Zu kleine Schrittweiten führen zu unnötigen Rundungsfehlern und zu längeren Laufzeiten, weil mehr Schritte gerechnet werden als nötig.

Grundlage der Schrittweitensteuerung ist die Schätzung der Größenordnung des Fehlers auf der Basis der asymptotischen Entwicklung des globalen Fehlers zur lokalen Schrittweite h nach dem Satz von Bader und Deuffhard (Satz 6). Der globale Fehler der Iteration (2.4) ist nämlich im Extrapolationsverfahren der zu schätzende lokale Fehler zur Schrittweite H .

Nach Satz 6 gilt folgende Fehlerschätzung für jede Matrix A in dem semi-impliziten Verfahren [9]:

Lemma 10 (Fehlerschätzung)

Sei $\varepsilon_{i,k} := \|T_{i,k} - y(t_0 + H)\|$ der Diskretisierungsfehler in einer skalierten Norm. Für ein Verfahren mit der asymptotischen Fehlerentwicklung (2.5) aus Satz 6 mit polynomialer Extrapolation gemäß (2.6) gilt folgende Fehlerschätzung:

$$\varepsilon_{i,k} \doteq (n_{i-k+1} \cdots n_i)^{-2} C(\alpha) \|C_k(t_0 + H, h)\| H^{2k}, \quad (2.16)$$

wobei $C(\alpha)$ die sogenannte Toeplitz-Konstante [20] zu der Schrittweitenfolge $\{n_i\}$ ist.

Bemerkung. Eine notwendige und hinreichende Bedingung für die Konvergenz des Extra-

polationsschemas, die Toeplitz-Bedingung (vgl. [16], S. 337), fordert

$$\alpha := \sup_{n \geq 0} \frac{n_i}{n_{i+1}} < 1,$$

woraus für die Toeplitz-Konstante

$$C(\alpha) := \sup_{n \geq 0} \sum_{m=0}^n |c_m^{(n)}| < \infty$$

folgt, mit den Koeffizienten der aus (2.6) zu gewinnenden Linearkombination für die Diagonalelemente

$$T_{k,k} = \sum_{i=1}^k c_i^{(k)} T_{i,1}.$$

Diese Koeffizienten werden allerdings nicht wirklich berechnet.

Nach [4, 9] wird $C_k(t_0 + H, h) = \mathcal{O}(H)$ angesetzt, sodaß man statt (2.16) die folgende Abschätzung erhält:

$$\varepsilon_{i,k} \doteq (n_{i-k+1} \cdots n_i)^{-2} \tau_k H^{2k+1} \quad (2.17)$$

mit einer geeigneten positiven Zahl τ_k .

Beweis (Lemma 10). [9, 20] □

Diese Schätzung des Fehlers beruht auf der asymptotischen Entwicklung des Diskretisierungsfehlers und auf der Approximationsordnung gemäß Lemma 7 (a). Sie beruht außerdem auf der stillschweigenden Annahme, daß das Verfahren höherer Ordnung (also durch das in einer Zeile weiter rechts stehende T repräsentierte Verfahren) die deutlich bessere Approximation liefert. Die Fehlerschätzung ist unabhängig von der konkreten Wahl der Stabilisierungsmatrix A .

Für die Diskretisierungsfehler einer Spalte folgt aus der Fehlerschätzung (2.17), daß sie sich durch Faktoren unterscheiden, die nur von der Schrittweitenfolge $\{n_i\}$ abhängen,

$$\varepsilon_{i+1,k} \doteq \left(\frac{n_{i-k+1}}{n_{i+1}} \right)^2 \varepsilon_{i,k}, \quad (2.18)$$

da τ_k nur von dem Spaltenindex abhängt.

Hat man im Schema die k -te Zeile berechnet, so steht mit dem Element $T_{k,k}$ die Approximation der höchsten Ordnung zur Verfügung. Die besten Approximationen niedrigerer Ordnung sind die Elemente $T_{k,k-1}$ und $T_{k-1,k-1}$. Von diesen wird aus zwei Gründen das Subdiagonalelement $T_{k,k-1}$ gewählt: zunächst ist der subdiagonale Näherungswert genauer (vgl. (2.18)), sodann ist er auch in der konkreten Implementierung — wo nur die jeweils aktuelle Zeile des Tableaus im Arbeitsspeicher gehalten wird — ohne weiteren Speicherplatzbedarf zu erhalten.

Obwohl die so gewonnene Schätzung des lokalen Fehlers nur für das Subdiagonalelement $T_{k,k-1}$ gilt, wird man trotzdem mit der genaueren Approximation $T_{k,k}$ weiterrechnen, wenn die Fehlerschätzung die vorgegebene Toleranz bedient.

Mit anderen Worten: Ein globaler subdiagonaler Fehlerschätzer des Extrapolationsverfahrens (wieder in einer skalierten Norm)

$$\varepsilon_{k,k-1} \doteq \bar{\varepsilon}_{k,k-1} = \|T_{k,k-1} - T_{k,k}\|$$

bedient die Toleranz eps, falls folgendes Kriterium erfüllt ist:

$$\bar{\varepsilon}_{k,k-1} \leq \text{eps}.$$

Das Verfahren kann dann mit der Näherung $T_{k,k}$ für $y(t_0 + H)$ fortgesetzt werden [9].

Die Fehlerschätzung wird allerdings nicht nur verwendet, um eine Validierung des Ergebnisses vorzunehmen, sie wird auch zur Bestimmung einer Schrittweite H für den nächsten globalen Schritt des Extrapolationsverfahrens benutzt.

Dazu stellt man zunächst Bedingungen für das Konvergenzverhalten des Extrapolationsverfahrens; diese haben etwa die Form

$$\varepsilon_{i,k} \leq \alpha_{i,k} \tag{2.19}$$

für gewisse vorgeschriebene Werte $\alpha_{i,k}$. In [9] werden solche Bedingungen an das Konvergenzverhalten aus einer informations-theoretischen Betrachtung des Extrapolationsverfahrens hergeleitet. Das der Extrapolation zugrundeliegende Verfahren wird dabei als eine Art „Black Box“ angesehen; lediglich die asymptotische Fehlerentwicklung (Satz 6) und die Eigenschaften des Extrapolationsverfahrens spielen bei der Fehlerschätzung eine Rolle.

Auf der Grundlage von (2.17) folgt nun

$$\begin{aligned} \frac{\alpha_{i,k}}{\varepsilon_{i,k}} &\doteq \left(\frac{H_{i,k}}{H}\right)^{2k+1} \quad \text{bzw.} \\ H_{i,k} &\doteq H \sqrt[2k+1]{\frac{\alpha_{i,k}}{\varepsilon_{i,k}}} \end{aligned} \tag{2.20}$$

für eine „optimale“ Schrittweite $H_{i,k}$, bei der in (2.19) Gleichheit erfüllt ist. Dabei sind die $\varepsilon_{i,k}$ nicht exakt bekannt, und die $\alpha_{i,k}$ stammen aus Annahmen über das ideale Konvergenzverhalten des Extrapolationsschemas. Die $\alpha_{i,k}$ sind so zu wählen, daß sie die Bedingung (2.18) erfüllen, also

$$\alpha_{i+1,k} \doteq \left(\frac{n_{i-k+1}}{n_{i+1}}\right)^2 \alpha_{i,k}$$

und damit auch $H_{i+1,k} \doteq H_{i,k}$ gilt (vgl. (2.20)). Aus diesen Beobachtungen folgt, daß nur eine Schrittweiteschätzung pro Spalte des Tableaus gemacht werden muß.

Mit der subdiagonalen Fehlerschätzung $\bar{\varepsilon}_{k,k-1}$ erhält man für die Spalten mit den Indizes $k = 1, 2, \dots$ die Schrittweiteschätzung

$$\bar{H}_{k,k-1} := \sqrt[2k-1]{\frac{\alpha_{k,k-1}}{\bar{\varepsilon}_{k,k-1}}}. \tag{2.21}$$

2.4.2. Minimierung des relativen Aufwandes

Durch die Angabe von Fehlerbedingungen wird nicht nur die Schrittweite, sondern auch die Ordnung des Verfahrens beeinflusst.

Dazu reicht es, als Toleranz in der Fehlerbedingung ein $\text{eps} = \alpha_{i,k}$ aus dem Konvergenz-Schema zu wählen. Mit der theoretischen Schätzung (2.17) der Fehlerordnung erhält man dann

$$H_{i,k} \doteq \sqrt[2k+1]{\frac{\text{eps}}{\gamma_{i,k} \tau_k}} \quad (2.22)$$

mit $\gamma_{i,k} := (n_{i-k+1} \cdots n_i)^{-2}$. Dieses theoretische Ergebnis läßt sich allerdings auf die Schätzung der idealen Zeitschrittweite übertragen.

Bei einer optimalen Wahl der Konsistenzordnung wird das Ergebnis in der gewünschten Genauigkeit mit minimalem Aufwand berechnet. Da der größte Teil des Aufwandes im Extrapolationsverfahren in der Berechnung der ersten Spalte des Extrapolationsschemas durch die Iteration der semi-impliziten Mittelpunkregel besteht, wird der Aufwand der Berechnung des Elementes $T_{i,i}$ gut durch den Aufwand der dazu insgesamt benötigten Iterationen im Grundverfahren beschrieben. Als weitere Vereinfachung wird dieser Aufwand durch die Anzahl der benötigten Funktions-Auswertungen beschrieben. Diese Überlegungen führen zu der

Definition 5 (relativer Aufwand)

Als relativen Aufwand pro Einheits-Zeitschritt bezeichnet man die Größe

$$W_{i,k} := \frac{H}{H_{i,k}} A_i$$

mit der tatsächlichen globalen Zeitschrittweite H und der „idealen“ Schrittweite $H_{i,k}$. Dabei bezeichne A_i den absoluten Aufwand zur Berechnung der Näherung $T_{i,i}$ im Extrapolationstableau.

Da die Berechnung der rechten Seiten des Systems f einen großen Anteil dieses Aufwandes ausmacht, wird die Anzahl der Funktions-Auswertungen als Maß für den absoluten Aufwand gewählt. Es gilt dann

$$\begin{aligned} A_1 &:= n_1 + 1 \\ A_{i+1} &:= A_i + n_{i+1} \quad \text{für } i = 1, 2, \dots \end{aligned}$$

Die theoretische Schrittweite $H_{k,k-1}$ ersetzt man nun durch die berechenbare Schätzung $\bar{H}_{k,k-1}$ aus (2.21) und erhält für den relativen Aufwand die Schätzung

$$\bar{W}_{k,k-1} := A_k \sqrt[2k-1]{\frac{\bar{\epsilon}_{k,k-1}}{\text{eps}}} \quad (2.23)$$

mit dem subdiagonalen Fehlerschätzer $\bar{\epsilon}_{k,k-1}$.

Ein geschätzter optimaler Spaltenindex q kann nun bestimmt werden aus

$$\bar{W}_{q,q-1} = \min_{k=1,\dots,k_f} \bar{W}_{k,k-1}$$

mit einem maximalen Spaltenindex k_f . Hat man auf diese Weise q bestimmt, wählt man als Schrittweiten-Vorhersage für den nächsten Schritt

$$\bar{H}_{q,q-1} = H \sqrt[2q-1]{\frac{\text{eps}}{\bar{\epsilon}_{q,q-1}}}. \quad (2.24)$$

2.5. Zur Realisierung des Verfahrens

Im folgenden sind einige Anmerkungen gesammelt, die sich auf die Realisierung des Verfahrens auf einem Rechner beziehen. Es handelt sich dabei um Anmerkungen von allgemeinerer Art bzw. um Bemerkungen zu der Realisierung in diesem speziellen Fall, die an anderer Stelle keinen Platz hatten.

- Zur Vermeidung unnötiger Matrix-Vektorprodukte, und um die explizite Berechnung von \bar{f} zu umgehen, wird üblicherweise das Verfahren (2.4) in der folgenden Form implementiert:

mit den Differenzen $\Delta_k = \eta_{k+1} - \eta_k$ berechnet man die Näherung durch

$$\begin{aligned} (I - hA) \Delta_0 &= h f(y_0) \\ \eta_1 &:= y_0 + \Delta_0 \end{aligned}$$

für $k = 1, 2, \dots, 2m - 1$

$$\begin{aligned} (I - hA) (\Delta_k - \Delta_{k-1}) &= 2(h f(\eta_k) - \Delta_{k-1}) \\ \eta_{k+1} &:= \eta_k + \Delta_k \end{aligned}$$

$$\begin{aligned} (I - hA) \bar{\Delta}_{2m} &= h f(\eta_{2m}) - \Delta_{2m-1} \\ S_{2m} &:= \eta_{2m} + \bar{\Delta}_{2m} \end{aligned}$$

- Für das Extrapolationsverfahren mit der semi-impliziten Mittelpunkregel nimmt man üblicherweise eine der beiden Schrittweitenfolgen
 - doppelte harmonische Folge

$$F_{2H} = \{2, 4, 6, 8, \dots\}$$

- Schrittweitenfolge von Bader/Deuffhard

$$F_\alpha = \{2, 6, 10, 14, 22, 34, 50, \dots\}$$

- Eine skalierte Norm für ein System aus n Komponenten erhält man mit einer Diagonalmatrix (Skalierungsmatrix)

$$D = \begin{pmatrix} \sigma_1 & & & \\ & \sigma_2 & & \\ & & \ddots & \\ & & & \sigma_n \end{pmatrix} \quad \text{durch}$$

$$\bar{\epsilon}_{k,k-1} = \|D^{-1}(T_{k,k-1} - T_{k,k})\|,$$

d. h. durch komponentenweises Skalieren. Neben festen Skalierungen mit konstanten σ_i benutzt man häufig relative Skalierungen, etwa $\sigma_i = |\bar{y}_i|$, Betrag der i -ten Komponente der aktuellen Näherungslösung [11].

Außerdem sind „glatte“ Normen, wie etwa die euklidische Norm vorzuziehen, weil sie meist einen glatteren Verlauf der Schrittweitensteuerung ergeben. Die Maximum-Norm etwa führt häufig zu Sprüngen in den Schrittweiten.

- Das semi-implizite Verfahren bleibt stabil, wenn statt der Jacobi-Matrix des ganzen Systems die Jacobische des steifen Anteils als Stabilisierungsmatrix gewählt wird. Bei der Diffusions-Konvektions-Aufgabe bietet es sich an, die symmetrische und positiv definite Jacobi-Matrix des Diffusionsanteils zu wählen und die Konvektion rein explizit zu rechnen.
- Wo — wie hier — die Jacobi-Matrix nicht explizit zur Verfügung steht oder analytisch nicht zu gewinnen ist, approximiert man ihre Einträge per Differenzenverfahren. Dabei sollte man die dünne Besetzungsstruktur und eventuell Symmetrie berücksichtigen.
- In dem hier vorliegenden konkreten Fall wird die Jacobi-Matrix des Grundwasser- und Schadstoff-Systems reduziert, das heißt die drei Blöcke von

$$J = \begin{pmatrix} J_{1,1} & 0 \\ J_{2,1} & J_{2,2} \end{pmatrix}$$

werden auch getrennt berechnet:

$$J_{1,1} = \frac{\partial}{\partial \psi} f_1(y) \quad J_{2,1} = \frac{\partial}{\partial \psi} f_2(y) \quad J_{2,2} = \frac{\partial}{\partial c} f_2(y)$$

und zwar in jedem globalen Schritt des Extrapolationsverfahrens.

Im semi-impliziten Verfahren werden dann lineare Gleichungssysteme mit der Jacobi-Matrix J auf folgende Weise berechnet:

$$\left[\begin{pmatrix} I & 0 \\ 0 & I \end{pmatrix} - h \begin{pmatrix} J_{1,1} & 0 \\ J_{2,1} & J_{2,2} \end{pmatrix} \right] \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = \begin{pmatrix} f_1 \\ f_2 \end{pmatrix}$$

reduziert sich zu den beiden linearen Gleichungssystemen

$$(I - hJ_{1,1})y_1 = f_1$$

$$(I - hJ_{2,2})y_2 = f_2 - J_{2,1}y_1.$$

Das erste entspricht dabei der Gleichung für den Wasserfluß, das zweite entspricht dem Transportproblem für die Schadstoffausbreitung.

3. Gebietszerlegungsverfahren als Vorkonditionierer

Wie im vorigen Kapitel gezeigt wurde, werden in der semi-impliziten Mittelpunkt-Regel lineare Gleichungssysteme mit einer Approximation der Jacobi-Matrix gelöst, um das Verfahren zu stabilisieren.

Die Tatsache, daß nicht die exakten Jacobischen Matrizen benötigt werden, sondern lediglich Approximationen, die allerdings gewisse Bedingungen erfüllen müssen (vgl. Lemma 9, S. 24), führt zu der Idee, einen sogenannten Vorkonditionierer der Jacobischen als Stabilisierungsmatrix einzusetzen. Als Vorkonditionierer werden hier Matrizen bezeichnet, deren Inverse leichter (oder billiger) zu berechnen ist als die Inverse der ursprünglichen Matrix, die aber gleichzeitig in einem gewissen Sinne nahe bei der ursprünglichen Matrix liegen.

Wie Vorkonditionierer in Lösungsverfahren für große dünn besetzte lineare Gleichungssysteme eingesetzt werden und welche konkreten Verfahren dabei eine Rolle spielen, wird im folgenden kurz dargestellt. Dabei liegt das Hauptaugenmerk auf Gebietszerlegungs-Vorkonditionierern, die von besonderem Interesse in der parallelen Lösung linearer Gleichungssysteme sind.

3.1. Iterative Lösung linearer Gleichungssysteme

Zur Lösung eines linearen Gleichungssystems

$$Ax = b \tag{3.1}$$

mit einer regulären Koeffizientenmatrix A gibt es zwei Klassen von Verfahren, direkte und iterative Methoden. Während die direkten Verfahren auf einer expliziten Zerlegung (Faktorisierung) der Koeffizientenmatrix beruhen, die ihre dünne Besetzungsstruktur zerstört, verwenden iterative Lösungsverfahren nur die Matrixelemente, die von Null verschieden sind und verändern A nicht. Als unvollständige Zerlegung oder Block-Faktorisierung treten allerdings direkte Techniken auch innerhalb von iterativen Verfahren auf.

3.1.1. Lineare Iterationsverfahren

Die elementaren Iterationsverfahren zur Lösung linearer Gleichungssysteme (3.1) sind lineare Verfahren der Form

$$Cx_{l+1} = Rx_l + b \quad (3.2)$$

mit einer sogenannten Vorkonditionierungsmatrix C und der zugehörigen Defektmatrix $R = C - A$. Die Iterierten dieses Verfahrens sind nur dann wohldefiniert, wenn C regulär ist.

Iterationsverfahren dieser Form werden auch als Defektkorrektur-Methoden formuliert, in denen die Defektmatrix nicht explizit vorkommt. Löst man (3.2) nach der Korrektur (Verbesserung) $d_{l+1} = x_{l+1} - x_l$ auf, so erkennt man, daß zur ihrer Berechnung der sogenannte Defekt (das Residuum) $r_l = Ax_l - b$ der Näherungslösung x_l verwendet wird:

$$Cd_{l+1} = -r_l, \quad x_{l+1} = x_l + d_{l+1}. \quad (3.3)$$

In den Iterationsverfahren dieser Typs wird also das lineare Gleichungssystem (3.1) sukzessive so gelöst, daß jeweils eine Korrektur als Lösung eines Gleichungssystems mit der Vorkonditionierungsmatrix C berechnet wird. In diesem Sinne charakterisiert die Wahl des Vorkonditionierers C das Verfahren.

Sei etwa x^* ein Fixpunkt der Iteration (3.2), also $Cx^* = (C - A)x^* + b$, so löst x^* auch das lineare Gleichungssystem $Ax^* = b$. Umgekehrt verschwindet der Defekt für die exakte Lösung, d.h. Iterationsverfahren der Form (3.2) sind für beliebige reguläre Vorkonditionierer C konsistent. Sie konvergieren für alle Anfangsnäherungen x_0 genau dann, wenn der Spektralradius der Iterationsmatrix $C^{-1}R = I - C^{-1}A$ kleiner als eins ist:

$$\rho(C^{-1}R) < 1. \quad (3.4)$$

Das Konvergenzkriterium (3.4) ist hinreichend und notwendig. Für beliebige aber miteinander verträgliche Vektor- und Matrixnormen $\|\cdot\|$ ist ein hinreichendes Kriterium für die Konvergenz, das gleichzeitig eine Fehlerabschätzung liefert, gegeben durch [23]

$$\begin{aligned} \|C^{-1}R\| < 1, \text{ und für die Fehler gilt} \\ \|x_{l+1} - x^*\| &\leq \|C^{-1}R\| \|x_l - x^*\| \quad \text{und} \\ \|x_l - x^*\| &\leq \|C^{-1}R\|^l \|x_0 - x^*\|. \end{aligned} \quad (3.5)$$

Die Konvergenzgeschwindigkeit wird in der Anzahl von Iterationen gemessen, die benötigt werden, um einen gewissen relativen Fehler sicher zu erreichen; zum Beispiel werden für $\|x^* - x_k\| / \|x^* - x_0\| \leq \varepsilon$ bis zu

$$k \approx \frac{\ln(1/\varepsilon)}{\ln(1/\|C^{-1}R\|)}$$

Iterationen gebraucht [3].

Ist $\|C^{-1}R\|$ nahe bei eins, so benötigt man sehr viele Iterationen, bis man in der entsprechenden Vektornorm den Fehler auf ein gewünschtes Maß reduziert hat. Daher ist es von Interesse, daß die Effizienz von Iterationsverfahren (3.2) auf zwei Weisen verbessert werden kann:

- durch eine polynomiale Beschleunigung, die auf sogenannte semi-iterative Verfahren führt, oder
- durch geschickte Wahl der Vorkonditionierungsmatrix C , die sowohl eine „möglichst gute Approximation“ für die Koeffizientenmatrix A als auch „leicht zu invertieren“ sein soll.

Bevor im nächsten Abschnitt einige Methoden der Vorkonditionierung dargestellt werden, folgen hier einige Bemerkungen über polynomiale Iterationsverfahren.

3.1.2. Polynomiale Beschleunigung

Polynomiale Iterationsverfahren haben die Form

$$x_{l+1} = x_l - \tau_l C^{-1} r_l \quad (3.6)$$

mit dem Defekt $r_l = Ax_l - b$ wie oben (3.3), C regulär. Die Korrektur $d_{l+1} = -C^{-1}r_l$ wird also mit einem Parameter τ_l gewichtet (dabei wird der degenerierte Fall $\tau_l = 0$ ausgeschlossen). Falls für alle $l = 0, 1, 2, \dots$ derselbe Parameter $\tau_l = \tau$ gewählt wird, spricht man von einer stationären Iteration, ansonsten von einem semi-iterativen oder instationären Verfahren.

Bezeichnet man mit $e_l = x^* - x_l$ den Fehler im l -ten Schritt, so erhält man

$$\begin{aligned} e_{l+1} &= x^* - x_l + \tau_l C^{-1}(Ax_l - b) \\ &= e_l + \tau_l C^{-1}A(x_l - x^*) \\ &= (I - \tau_l C^{-1}A)e_l \end{aligned} \quad (3.7)$$

für $l = 1, 2, \dots$, sowie $e_m = p_m(C^{-1}A)e_0$. Dabei ist $p_m(\lambda) = \prod_{l=0}^m (1 - \tau_l \lambda)$ ein durch die Folge der τ_l definiertes Polynom m -ten Grades mit Nullstellen bei $1/\tau_l$. Außerdem sind Iterationsverfahren der Form (3.6) für beliebige Parameterfolgen $\{\tau_l; l = 0, 1, 2, \dots\}$ konsistent.

Als spezielle polynomiale beschleunigte Iterationsverfahren mit adaptiver Wahl der Parameter können vorkonditionierte CG-Verfahren und verwandte Methoden betrachtet werden. Allerdings werden diese meist nicht im Kontext linearer Iterationsverfahren behandelt, sondern bilden eine eigene Klasse iterativer Methoden für dünn besetzte große Gleichungssysteme [18, 5].

3.1.3. Vorkonditionierungstechniken

Hier kann nur kurz angerissen werden, welche Vielfalt von Vorkonditionierungstechniken angewandt werden, um eine konvergente Zerlegung $A = C - R$ der Koeffizientenmatrix zu

gewinnen. Bei manchen Verfahren wird die Matrix C angegeben, die zu einer „approximierten Inversen“ von A führt, aber das ist nicht erforderlich, da im Grunde nur die Wirkung der Inversen C^{-1} , also eine Approximation der Inversen der Koeffizientenmatrix A in dem Verfahren benötigt wird.

Die bekanntesten Vorkonditionierer sind die Gauß-Seidel und die Jacobi-Vorkonditionierer. Dazu wird durch die Zerlegung

$$A = D - L - U$$

der Koeffizientenmatrix in ihren Diagonalanteil D und die oberen (U) bzw. unteren (L) Dreiecksmatrizen zugrundegelegt. Jacobi- und Block-Jacobi-Verfahren beruhen auf der Invertierung des Diagonalanteils der Koeffizientenmatrix, der Jacobi (bzw. Block-Jacobi-) Vorkonditionierer (Gesamtschrittverfahren) hat die Form

$$C = D.$$

Das Gauß-Seidel-Verfahren (Einzelschrittverfahren) erhält man bei der Wahl des unteren Dreiecks-Anteils der Koeffizientenmatrix

$$C = D - L,$$

und ein allgemeineres Relaxationsverfahren (*successive (over-) relaxation method, SOR*) durch die Wahl

$$C = \omega^{-1}D - L \quad \text{mit } \omega \in (0, 2),$$

bzw. ein entsprechendes symmetrisches Verfahren (*symmetric SOR, SSOR*)

$$C = (\omega^{-1}D - L)(\omega^{-1}D)^{-1}(\omega^{-1}D - U) \quad \text{mit } \omega \in (0, 2)$$

durch Einbeziehen eines Relaxationskoeffizienten.

Eine wichtige Klasse von Verfahren zur Gewinnung von Vorkonditionierern beruht auf unvollständigen Zerlegungen der Matrix A (*incomplete LU/Cholesky factorization, ILU/IC*). Dabei werden verschiedene Techniken verwendet, um eine Zerstörung der dünnen Besetzungsstruktur der Matrix zu verhindern oder den Eintrag von neuen Elementen zu begrenzen. Polynomiale Vorkonditionierer beruhen auf einer Approximation der Inversen A^{-1} durch Polynome von Matrizen [3, 23]. Alle diese Vorkonditionierungs-Verfahren sind auch auf ihre Verwendbarkeit in parallelen Algorithmen auf Rechnern mit verteiltem Speicher untersucht worden [6, 7, 33].

Bekommt man das zu lösende lineare Gleichungssystem aus der Diskretisierung einer partiellen Differentialgleichung, so kann man für die Ermittlung des Vorkonditionierers Informationen über die Geometrie des zugrundeliegenden Gebiets verwenden. Auf diese Weise erhält man Gebietszerlegungsverfahren zur Lösung linearer Gleichungssysteme. Nicht-überlappende Gebietszerlegungsmethoden (siehe den nächsten Abschnitt 3.2) sind Block-Jacobi- oder Block-Gauß-Seidel-Verfahren. Das verwendete Splitting (im Grunde eine Umnummerierung der Unbekannten) beruht dann auf der Wahl der Teilgebiete. Bei den überlappenden Gebietszerlegungen kann die Vorkonditionierungsmatrix im allgemeinen nicht angegeben werden. Die Wirkung der approximierten Inversen C^{-1} wird dann durch eine geeignete Kombination der Teilgebiets-Lösungen hergestellt.

3.2. Gebietszerlegungsverfahren

Gebietszerlegungsmethoden (*domain decomposition methods*) sind in den vergangenen Jahren zu einem sehr aktiven Feld in der Numerik partieller Differentialgleichungen geworden [29], vor allem als natürlicher Ansatz zur Parallelisierung [21].

Gebietszerlegungsverfahren werden heute bereits in Anwendungsprogrammen eingesetzt, zum Beispiel in einem expliziten Verfahren zur parallelen Simulation von Crash-Tests [43]. Häufig wird das Schwarzsche Verfahren in seiner additiven Form dort eingesetzt, wo bestehende Programme auf Parallelrechnern oder Rechnerverbunden ohne große Änderungen weiter verwendet werden sollen.

Im weiteren Sinne gehören auch Partitionierungstechniken für Graphen und irreguläre Gitter [36] zu den Gebietszerlegungsmethoden. Den Hintergrund für diese Verfahren bildet die Notwendigkeit, unstrukturierte Gitter (z.B. adaptive FE-Gitter aus der Strömungssimulation) für die Verarbeitung auf einer vorgegebenen Anzahl von Prozessoren zu zerlegen.

Heute werden numerische Gebietszerlegungsverfahren im engeren Sinne als Lösungsverfahren für die bei der Diskretisierung partieller Differentialgleichungen auftretenden linearen oder nichtlinearen Gleichungssysteme betrachtet [37].

Die ursprüngliche Idee der Gebietszerlegung von H. A. Schwarz [34] war eine Methode, partielle Differentialgleichungen auf komplexen Gebieten zu lösen, die als Vereinigung einfacherer Gebiete aufgefaßt wurden. Das Problem auf dem Gesamtgebiet wurde zerlegt in gekoppelte Probleme auf den Teilgebieten; die Kopplung wurde durch Überlappungen der Teilgebiete gewährleistet. In Abb. 3.1 beispielsweise ist das Gesamtgebiet Ω in zwei

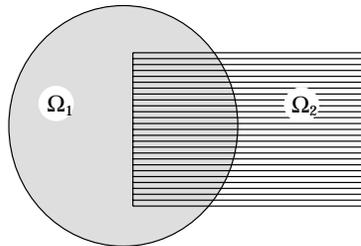


Abbildung 3.1.: Eine einfache Zerlegung in zwei überlappende Teilgebiete

Teilgebiete Ω_1 und Ω_2 zerteilt, die jeweils eine einfache Geometrie haben und sich in einem gewissen Bereich im Inneren von Ω überlappen. Der alternierende Schwarz-Prozeß für die Poisson-Gleichung

$$\begin{aligned} -\Delta u &= f & \text{in } \Omega &= \Omega_1 \cup \Omega_2, \\ u &= g & \text{auf } \partial\Omega, \end{aligned}$$

auf dem Gesamtgebiet besteht darin, ausgehend von einer Anfangs-Näherung $u^{(0)}$ die jeweils nächste Iterierte $u^{(n+1)}$ durch sukzessive Lösung der Probleme auf den Teilgebieten zu berechnen. Hierbei reicht es, für $u^{(0)}$ die Anfangswerte auf $\partial\Omega_1 \setminus \partial\Omega$ vorzugeben. Diese Werte auf den inneren Teilgebietsrändern (die „Pseudo-Randwerte“ auf $\partial\Omega_1 \setminus \partial\Omega$ bzw. auf

$\partial\Omega_2 \setminus \partial\Omega$) werden in den folgenden Iterationen aus dem überlappenden Bereich genommen:

$$\begin{aligned} -\Delta u^{(n+1/2)} &= f && \text{in } \Omega_1, \\ u^{(n+1/2)} &= u^{(n)} && \text{auf } \partial\Omega_1, \end{aligned}$$

sowie

$$\begin{aligned} -\Delta u^{(n+1)} &= f && \text{in } \Omega_2, \\ u^{(n+1)} &= u^{(n+1/2)} && \text{auf } \partial\Omega_2. \end{aligned}$$

Bei diesem alternierenden Verfahren handelt es sich um eine Iteration in Einzelschritten, da in jedem Iterationsschritt die bereits berechneten Näherungen aus dem gleichen Schritt gewählt werden und aus dem vorigen Schritt nur die Näherungen benutzt werden, die noch nicht neu berechnet wurden. Ein Gesamtschritt-Verfahren entsteht, wenn in jeder Iteration sämtliche Schätzungen für die Pseudo-Randwerte aus der vorigen Iteration genommen werden. Wie bei der Gauß-Seidel-Relaxation wird beim multiplikativen Schwarz-Verfahren eine Parallelisierung durch „Färbung“ der Teilgebiete erreicht; das additive Verfahren ist (wie das Jacobi-Relaxationsverfahren) bereits vom Ansatz her ein paralleles Verfahren.

Im folgenden geht es um Gebietszerlegungsverfahren zur Lösung linearer Gleichungssysteme, um Gebietszerlegungs-Vorkonditionierer, die eine approximierte Inverse der Koeffizientenmatrix auf der Basis von (in den Teilgebieten) lokalen Systemen liefern.

Man kann die numerischen Gebietszerlegungsverfahren (DD-Methoden für *domain decomposition*) grob in überlappende und nichtüberlappende Verfahren einteilen [37].

Überlappende Verfahren (Schwarz-Verfahren) sind Iterationsverfahren, die den Relaxationsverfahren für lineare Gleichungssysteme entsprechen (vgl. Abschnitt 3.1). Betrachtet man die lokalen Lösungen als Korrekturen, erkennt man ihre Verwandtschaft mit Mehrgitter-Verfahren; häufig werden auch Multilevel-Schwarz-Verfahren eingesetzt.

Wenn man die bezüglich ihrer lokalen Variablen diskretisierte Aufgabe betrachtet, so läßt sich eine Aufteilung in Teilgebiete als Partitionierung der Gitterpunkte beziehungsweise ihrer Indexmenge verstehen [23]. Der Restriktionsoperator Q zu einem gewissen Teilgebiet ist die Matrix, die eine Näherungslösung auf jene Komponenten reduziert, die in einem gewissen Teilgebiet liegen. Die Transponierte Q^T ist der entsprechende Prolongationsoperator.

Sei wieder x^* die exakte Lösung des linearen Gleichungssystems (3.1) und $x^{(n)}$ eine Näherungslösung, die man mit der Prolongation $Q^T c$ einer lokalen Korrektur verbessern möchte. Dann ist c die Lösung der restringierten Defektgleichung

$$(QAQ^T)c = Q(Ax^* - Ax^{(n)}),$$

beziehungsweise

$$Q^T c = Q^T (QAQ^T)^{-1} Q(b - Ax^{(n)}). \quad (3.8)$$

Die Matrix (QAQ^T) ist genau der zu den Knoten des betrachteten Teilgebiets gehörende Block der Koeffizientenmatrix A .

Im folgenden wird nun ein einfacher DD-Vorkonditionierer vorgestellt, der einfach zu konstruieren ist und eine direkte Parallelisierung ermöglicht. Es handelt sich um den mit dem Block-Jacobi-Verfahren verwandten additiven Schwarz-Vorkonditionierer [37].

Für die Darstellung der nichtüberlappenden additiven Gebietszerlegung ist es hilfreich, die Unbekannten gebietsweise abzuzählen, d.h. die Unbekannten eines Teilgebiets nacheinander; die entsprechenden Restriktionsoperatoren haben dann die folgende Blockgestalt:

$$Q_i = (0 \dots 0I_i0 \dots 0).$$

Eine Kombination der Teilgebietslösungen im obigen Sinne führt auf den Block-Jacobi-Vorkonditionierer, und die entsprechende approximierte Inverse von A ist dann zum Beispiel bei drei Teilgebieten

$$\begin{aligned} C^{-1} &= \begin{pmatrix} A_1^{-1} & 0 & 0 \\ 0 & A_2^{-1} & 0 \\ 0 & 0 & A_3^{-1} \end{pmatrix} \\ &= Q_1^T A_1^{-1} Q_1 + Q_2^T A_2^{-1} Q_2 + Q_3^T A_3^{-1} Q_3. \end{aligned} \quad (3.9)$$

Im Schwarz-Verfahren werden im allgemeinen überlappende Gebietszerlegungen gewählt, weil diese eine Kopplung der Teilprobleme liefern, die für die korrekte Behandlung partieller Differentialgleichungen unverzichtbar ist. Im numerischen Verfahren werden die überlappenden Differenzen-Knoten oder Volumen- bzw. Element-Zellen als Dirichlet-Randwerte behandelt.

Hat man etwa das Integrationsgebiet in p Teilgebiete von etwa gleicher Größe zerlegt, so ergibt sich analog zu (3.9) die zugehörige approximierte Inverse als

$$C^{-1} = \sum_{i=1}^p Q_i^T A_i^{-1} Q_i;$$

im Unterschied zum Block-Jacobi-Verfahren kann man hier die Matrix C nicht mehr explizit angeben. Das ist aber auch nicht erforderlich, da sowohl im Zusammenhang mit linearen Iterationen und vorkonditionierten CG-Verfahren, als auch in der Anwendung als Stabilisierungsmatrix letzten Endes nur die Wirkung einer approximierten Inversen C^{-1} benötigt wird. Auch diese wird nicht explizit berechnet, sondern ihre Wirkung auf einen Vektor v auf folgende Weise realisiert:

- $v_i := Q_i v$ Restriktion
- $A_i w_i = v_i$ Lösung eines linearen Gleichungssystems in jedem Teilgebiet (parallel)
- $C^{-1} v := \sum_{i=1}^p Q_i^T w_i$ Prolongation und Addition der Teilgebietslösungen (Man beachte, daß hierbei eine Summe von Vektoren berechnet wird, und nicht eine Summe über Matrizen.)

Verwendet man das Schwarz-Verfahren zur Parallelisierung, so ist zu beachten, daß die Werte von v in den jeweiligen überlappenden Bereichen zu Beginn eines Iterationsschritts übereinstimmen müssen. Meist werden einzelne Teilgebiete jeweils einem Prozessor zugeordnet, sodaß die Anpassung eine Kommunikation zwischen den Prozessoren erfordert.

Da die Kopplung der Teilprobleme bei den Schwarz-Verfahren nur lokal stattfindet, wird häufig eine Grobgitter-Beschleunigung mit einem Gitter in der Größe der Teilgebiete durchgeführt. Auf diese Weise erhält man überlappende Zweigitter- oder Mehrgitter-Schwarz-Verfahren mit einer wesentlich verbesserten Konvergenz. Ist etwa Q_c der Restriktionsoperator für das Grobgitter, so erhält man einen einfachen Zweigitter-Schwarz-Vorkonditionierer durch

$$C^{-1} = Q_c^T A_c^{-1} Q_c + \sum_{i=1}^p Q_i^T A_i^{-1} Q_i.$$

Andere Gebietszerlegungsverfahren mit globaler Kommunikation basieren auf der Block-Elimination von nichtüberlappenden Zerlegungen, wo aber die sogenannten Interface-Knoten, die zwischen den Teilgebieten etwa eine Schicht von Volumen-Zellen bilden, einen eigenen Block A_I erhalten. Die Teilgebiete sind dann durch eine A_I -Schicht voneinander separiert. Durch eine gebietsweise Abzählung, bei der A_I zuletzt gezählt wird, erhält die Systemmatrix eine Block-Form, in der der weitaus größte Teil A_D Block-Diagonalgestalt hat:

$$\begin{pmatrix} A_D & A_{DI} \\ A_{ID} & A_I \end{pmatrix} \begin{pmatrix} x_D \\ x_I \end{pmatrix} = \begin{pmatrix} b_D \\ b_I \end{pmatrix}$$

Das dabei z.B. bei zwei Teilgebieten entstehende Schur-Komplement-System

$$(A_I - \sum_{i=1}^p A_{Ii} A_i^{-1} A_{iI}) x_I = b_I - \sum_{i=1}^p A_{Ii} A_i^{-1} b_i$$

ist verhältnismäßig klein, und die darin vorkommenden Teilgebetslösungen sind unabhängig voneinander. Die Schur-Komplement-Methode (Unterstruktur-Methode) erfordert allerdings eine spezielle Vorkonditionierung; etwa einen sogenannten Neumann-Dirichlet- oder Neumann-Neumann-Vorkonditionierer [37].

3.3. Vorkonditionierer im semi-impliziten Verfahren

In diesem Abschnitt wird dargestellt, wie sich ein additives Gebietszerlegungsverfahren in die semi-implizite Mittelpunkregel einfügen läßt.

Sei durch

$$\Omega = \bigcup_{i=1}^p \Omega_i$$

eine Gebietszerlegung definiert. Dann ist — wie im vorigen Absatz dargestellt — der zugehörige additive Vorkonditionierer C definiert durch

$$C^{-1} = \sum_{i=1}^p Q_i^T (Q_i A Q_i^T)^{-1} Q_i.$$

Im semi-impliziten Verfahren (der semi-impliziten Euler-Methode oder der semi-impliziten Mittelpunkregel) ist das folgende lineare Gleichungssystem zu lösen (vgl. (2.4)):

$$(I - hA)y = f. \quad (3.10)$$

Dabei bezeichne A die Jacobi-Matrix $f_y(y_0)$.

Da für die Jacobi-Matrix A nur eine Approximation \hat{A} gesucht wird, an die allerdings die aus Kapitel 2 bekannten Stabilitäts-Voraussetzungen zu stellen sind, wird man zunächst versuchen, einen entsprechenden Gebietszerlegungs-Vorkonditionierer zu verwenden; dadurch erhält man — Stabilität vorausgesetzt — ein äußerst gut parallelisierendes Verfahren, weil die Lösung der linearen Gleichungssysteme, die den Großteil des Rechenaufwandes ausmacht, rein lokal geschehen könnte.

Das lineare Gleichungssystem (3.10) wird dann zu

$$Cy = f,$$

beziehungsweise

$$\begin{aligned} y &= C^{-1} f \\ &= \sum_{i=1}^p Q_i^T (Q_i (I - hA) Q_i^T)^{-1} Q_i f \\ &= \sum_{i=1}^p Q_i^T (I_i - hA_i)^{-1} f_i \end{aligned} \quad (3.11)$$

mit den jeweils lokalen Werten der rechten Seite f_i , den lokalen Jacobi-Matrizen A_i und den entsprechenden Einheitsmatrizen I_i für die Teilgebiete Ω_i .

Sind die Stabilitätsbedingungen verletzt, kann man durch einige wenige Iterationen des linearen Defektkorrektur-Verfahrens die Approximation — und damit die Stabilität verbessern.

Durch die in (3.11) dargestellte Gebietszerlegung wird folgendes lineare Gleichungssystem mit der approximierenden Jacobi-Matrix \hat{A} gelöst:

$$(I - h\hat{A})y = f.$$

Formal entsprechen sich also die Matrizen

$$(I - h\hat{A})^{-1} = \sum_{i=1}^p Q_i^T (I_i - hA_i)^{-1} Q_i,$$

aus dieser Darstellung ist aber eine geschlossene Repräsentation für \hat{A} nicht zu gewinnen.

Für die oben (Kapitel 2) dargestellten Stabilitätsbedingungen benötigt man allerdings Schätzer für die logarithmische Matrixnorm der unbekanntten Stabilisierungsmatrix \hat{A} sowie für die Lipschitz-Konstante des nichtlinearen Rests $\tilde{f}(y) = f(y) - \hat{A}y$. Da dies vorderhand aussichtslos erscheint, muß man einen anderen Weg finden, die Stabilität durch einen Test zu überprüfen.

4. Implementierung

Im Rahmen dieser Arbeit wurde ein Simulationsprogramm entworfen, das später auf einem massiv-parallelen Rechner, dem Paragon XP/S 10 der Firma Intel, ausgeführt werden sollte; allerdings sollte es auch möglich sein, das Programm ohne großen Aufwand auf andere Parallelrechner mit verteiltem Speicher (etwa das Cray-System T3E) zu übertragen. Aus diesem Grunde wurde z.B. statt der Kommunikationsbibliothek des Herstellers das „Message Passing Interface“ (MPI) benutzt, das für verschiedene Plattformen verfügbar ist.

In diesem Kapitel werden die Begriffe vorgestellt, welche die Programmierung dieser Parallelrechner betreffen. Danach werden die Bewertungskriterien für parallele Programme erläutert und die grundlegenden Entscheidungen bei der Realisierung des Simulationsprogramms dargestellt.

4.1. Message Passing

Es handelt sich bei dem Intel Paragon um einen skalierbaren massiv-parallelen Multiprozessor mit verteiltem Speicher; in der bekannten Klassifikation nach Flynn [8] gilt er als MIMD-Rechner (*Multiple Instruction Multiple Data Stream*).

Er besteht aus insgesamt 138 Rechenknoten, die über ein zweidimensionales rechteckiges Netzwerk miteinander verbunden sind. Jeder dieser Knoten verfügt unter anderem über je einen Intel i860 RISC-Prozessor für die Berechnungen und die Kommunikation, sowie über 32MB Arbeitsspeicher. Zusätzliche Knoten, die im wesentlichen den gleichen Aufbau haben, sind für Input/Output oder Service reserviert.

Für die Programmierung ist von Belang, daß die einzelnen Rechenknoten nicht auf einen gemeinsamen Speicherbereich zugreifen können, sondern jeder seinen eigenen lokalen Speicher hat, der für die anderen Knoten unsichtbar ist. Da allerdings in parallelen Programmen die Teilprozesse Informationen austauschen müssen, hat dies durch das Versenden von Daten von einem Knoten zu einem anderen zu geschehen. In dem hier entwickelten Programm geschieht diese Kommunikation durch explizites Senden und Empfangen von Nachrichten (*message passing*); eine andere Möglichkeit ist, daß mittels entsprechender Bibliotheken ein virtueller gemeinsamer Adreßraum zur Verfügung gestellt wird (dessen Realisierung allerdings auf dem Versenden von Daten zwischen den Knoten beruht).

Im Paragon wird diese Kommunikation über ein rechteckiges Verbindungsnetzwerk (*mesh*)

verwirklicht, dessen Topologie aber bei „wenigen“ Prozessoren bezüglich der Funktion oder der Effizienz des Informationsaustausches nicht sehr stark ins Gewicht fällt. Dieses Netz ist auf verschiedene Anzahlen von Prozessoren skalierbar. Auch vom Benutzer können Partitionen mit einer beliebigen Anzahl freier Prozessoren verwendet werden.

Kenngrößen des Rechners, die auf die Effizienz paralleler Programme einen großen Einfluß haben, sind die Latenzzeit (die Zeit, die zur Initialisierung eines Kommunikationsvorgangs benötigt wird), die Bandbreite des Verbindungsnetzwerks sowie die Möglichkeit nichtblockierender Kommunikation (also der zeitlichen Überlagerung von Berechnung und Kommunikation).

Die Leistungsdaten des Paragon XP/S, wie er in der KFA zur Verfügung steht, werden vom Hersteller wie folgt angegeben:

- Maximale Rechenleistung pro Prozessor:
100 MFLOPS SINGLE PRECISION
75 MFLOPS DOUBLE PRECISION
- Latenzzeit (*latency*) bei Message Passing:¹
37 μ s NIC A
32 μ s NIC B
- Bandbreite des Kommunikationsnetzes:
91 MB/s NIC A
138 MB/s NIC B

Benchmarks, die im ZAM durchgeführt wurden ergaben zum Beispiel für die BLAS-3-Routine DGEMM (mit $n = 500$) auf einem einzelnen Prozessor eine Rechenleistung von 44 MFLOPS [14].

Das bevorzugte Programmiermodell für den Paragon ist das SPMD-Modell (*Single Program Multiple Data*), bei dem dasselbe Programm auf alle p benutzten Rechenknoten geladen wird. In diesem Programm wird dann der aktuelle Knoten anhand einer relativen Knotennummer zwischen 0 und $p - 1$ identifiziert, so daß in Abhängigkeit von dieser Nummer der Programmablauf kontrolliert werden kann.

Parallele Algorithmen werden häufig danach unterschieden, ob der Parallelismus auf einer Verteilung der Daten beruht, auf denen im wesentlichen die gleichen Operationen gleichzeitig ausgeführt werden (datenparallele Algorithmen), oder ob verschiedene unabhängige Funktionen gleichzeitig auf verschiedenen Prozessoren berechnet werden (funktionsparallele Algorithmen). Wird eine Gebietszerlegung lediglich als Grundlage für ein Datenverteilungs-Schema benutzt, so gehört ein entsprechender Algorithmus sicher zu der ersten Gruppe. Aber in Fällen, wo Eigenschaften der zugrundeliegenden Differentialgleichung in Gebietszerlegungsmethoden eingehen, ist die Unterscheidung so eindeutig nicht mehr.

¹Mit NIC A und NIC B sind verschiedene Versionen der Kommunikationshardware bezeichnet. Im KFA-Paragon steht die Version NIC A zur Verfügung, neuere Maschinen, etwa in der ETH Zürich, sind mit NIC-B-Chips ausgestattet.

4.2. Kriterien zur Bewertung paralleler Programme

Bei der Ausführung eines parallelen Programms werden Zeitmessungen durchgeführt, um den Effekt der parallelen Ausführung quantitativ fassen zu können. Im optimalen Fall läßt sich ein Programm gleichmäßig auf p Prozessoren mit disjunkten Teilaufgaben aufteilen, so daß jeder Prozessor die gleiche Bearbeitungszeit für seinen Teil hat. In diesem Fall einer bestmöglichen Lastverteilung (*load balancing*) ohne Kommunikationsaufwand erwartet man ein Verhältnis zwischen der Ausführungszeit t_1 des sequentiellen Programms und der Ausführungszeit t_p des parallelen Programms von

$$\frac{t_1}{t_p} = p. \quad (4.1)$$

Im allgemeinen wird man allerdings nur

$$S(p) := \frac{t_1}{t_p} \leq p \quad (4.2)$$

erwarten, da in parallelen Programmen ein zusätzlicher Aufwand durch das Zusammenfassen der Teillösungen, durch redundante Berechnungen und Datenaustausch entsteht und selten von einem optimalen Lastausgleich ausgegangen werden kann. Die in (4.2) definierte Größe $S(p)$ ist der proportionale Zeitgewinn (*speedup*).

Die (parallele) Effizienz

$$E = \frac{S(p)}{p} \in (0, 1] \quad (4.3)$$

wird auch häufig in Prozent angegeben und ist ein Maß für die Parallelisierbarkeit auf unterschiedlich vielen Rechenknoten. Ist die Effizienz für eine beliebige Anzahl von Prozessoren hoch, so wird dies als Skalierbarkeit bezeichnet.

4.3. Realisierung

Zur Implementierung des zunächst sequentiellen Programms wurde die Programmiersprache C++ gewählt. Sie bietet die gewünschte Flexibilität und ermöglicht die Einbindung öffentlich verfügbarer Klassenbibliotheken, die das Programmieren wesentlich erleichtern.

Von Anfang an wurde die SparseLib++-Bibliothek von R. Pozo et.al. [32]² genutzt, die verschiedene komprimierte Speicherformate für dünnbesetzte Matrizen und entsprechende Matrix-Vektor-Operationen zur Verfügung stellt. Die MV++-Bibliothek für Vektoren und dichtbesetzte Matrizen [31] gehört zum Umfang der SparseLib++.

Aufgrund der Entscheidung, im semi-impliziten Verfahren nur die Jacobi-Matrix des steifen, dispersiven Anteils der Transportgleichung zu benutzen, die symmetrisch und positiv definit ist, wurde das CG-Verfahren aus der IML++-Bibliothek [12] als Löser für das

²URL <http://gams.nist.gov/acmd/Staff/RPozo/>

lineare Gleichungssystem eingefügt. IML++ stellt die wichtigsten „Templates“-Algorithmen [5] als C++-Templates zur Verfügung. Ein wichtiger Vorteil dieser Bibliothek ist auch, daß nach Bedarf das CG-Verfahren durch eine andere „Templates“-Routine leicht ersetzt werden kann. Die Definitionen für die Übergabeparameter sind sehr einfach und der Templates-Code sehr übersichtlich, so daß modifizierte oder neue Verfahren leicht auf gleiche Weise implementiert werden können.

Für die Realisierung des Programms auf einem Parallelrechner (hier etwa auf dem Paragon XP/S im ZAM) wurde die Kommunikationsbibliothek MPI (*Message Passing Interface*) gewählt. Zum einen wegen ihrer Einheitlichkeit und Verfügbarkeit auf verschiedenen Rechnerplattformen; zum anderen aber auch, weil die C++-Bindings von MPI eine Kommunikationsumgebung im objekt-orientierten Stil von C++ zur Verfügung stellen.

Das objekt-orientierte Paradigma von C++ wurde auch insofern genutzt, daß die Funktionen zur Aktualisierung der Jacobi-Matrix f_y , zur Berechnung der rechten Seite $f(y)$ der Differentialgleichung und zur Ein- und Ausgabe von Parametern und Werten in Objektklassen gekapselt wurden. Dieses geschah noch nicht sehr weitgehend, eröffnete aber bereits Perspektiven auf die Möglichkeiten einer völlig transparenten Kombination von Methoden und Modellen.

Im folgenden wird der Aufbau des Programmkerns, die Durchführung eines globalen Schrittes im Extrapolationsverfahren mit der Schrittweite H , in einem vereinfachten Schema dargestellt; dabei sind vor allem die Synchronisationsstellen im Programm von Interesse.

Synchronisation im parallelen Programm bedeutet hier, daß die parallelen Prozesse Daten austauschen und aus diesem Grunde aufeinander warten müssen.

In der folgenden stark vergrößernden Darstellung sind die Stellen, an denen jeweils synchronisiert werden muß, durch Unterstreichen markiert.

EXTRAPOL(t_0, H)

- 1 erzeuge Jacobi-Matrix
- 2 SIMPR t_0, H, n_1 **für** $i = 2, 3, \dots$
- 3 SIMPR t_0, H, n_i
- 4 **berechne** neue Zeile im Extrapolationsschema
- 5 Fehlerbedingung
- 6 **Ordnungs- und Schrittweiten-Vorhersage**

SIMPR(t_0, H, n)

- 1 **für** $j = 2, 3, \dots, n$
- 2 berechne rechte Seite
- 3 **Lösung** des lokalen linearen Gleichungssystems
- 4 evtl. Defekt-Korrektur
- 5 **Glättungsschritt**

Die Erzeugung der globalen Jacobi-Matrix. Da hier die Daten von jedem Gebiet (plus

zugehörigen Überlappungsbereichen) jeweils auf einem Prozessor liegen, müssen die Anschluß-Daten an den Gebietsgrenzen und im Überlappungsbereich zwischen den Prozessoren (Rechenknoten) ausgetauscht werden.

Fehlerbedingung. Da eine globale Fehlerschätzung (also eine für das ganze Gebiet) benötigt wird, ist Kommunikation nötig: Berechnung und Abgleich des Fehlerschätzers. Da die Ordnungs- und Schrittweitevorschau nur von dem Fehlerschätzer und von Parametern des Programms abhängt, ist dafür keine neuerliche Synchronisation erforderlich.

Berechnung der rechten Seite. In den überlappenden Bereichen muß die rechte Seite jeweils abgeglichen werden.

Defektkorrektur. Sollte eine Defektkorrektur zur Stabilisierung erforderlich sein, so ist dazu ein globales Matrix-Vektor-Produkt erforderlich.

5. Simulationsergebnisse

Das während der Bearbeitungszeit entworfene C++-Programm, das auf der Basis der Liniemethode und der extrapolierten semi-impliziten Mittelpunkregel die Gleichungen zur Simulation des Grundwasserflusses und Stofftransports löst, konnte aus Zeitgründen nur in einer ersten, sequentiellen Version implementiert werden.

Einige Ergebnisse lassen sich hier kurz zusammenfassen, obwohl eine systematische Analyse von hinreichend vielen verschiedenen Testläufen nicht mehr möglich war.

Die folgenden Bilder wurden mit MATLAB aus den Ausgabedateien des Programmes erzeugt; sie zeigen den Transport einer Schadstoffwolke in einem stationären Flußfeld. Simuliert wurde die Ausbreitung eines Tracers innerhalb von zwei Tagen in einem Feld mit einer Ausdehnung von $2\text{m} \times 2\text{m}$ (Abb. 5.3). Die mittlere Fließgeschwindigkeit beträgt etwa 1m am Tag.

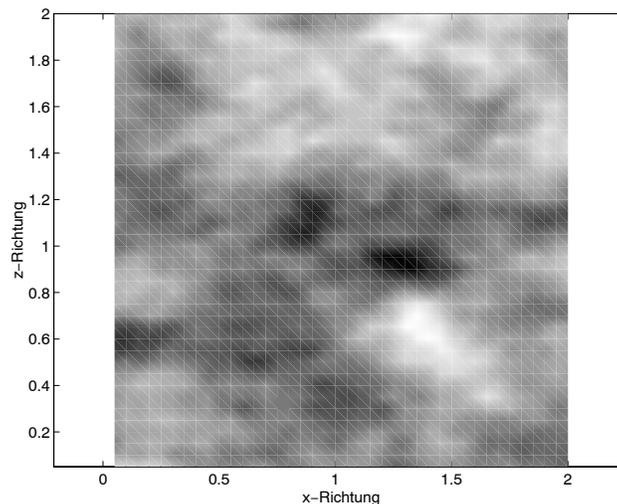


Abbildung 5.1.: Gesättigte hydraulische Leitfähigkeit (statistisch)

Die Daten für die gesättigte hydraulische Leitfähigkeit (siehe Anhang für die hydrologischen Zusammenhänge) wurden mit einem Programm der Arbeitsgruppe im ICG-4 aus den statistischen Parametern für mäßig durchlässigen Boden (z.B. sandiger Lehm) generiert (vgl. Abbildung 5.1).

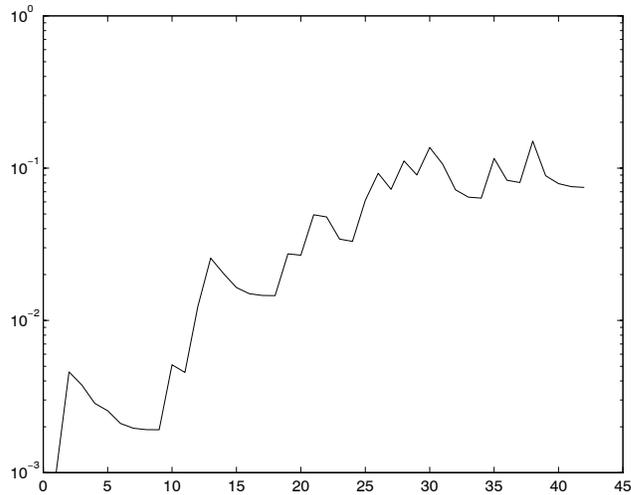


Abbildung 5.2.: Schrittweiten-Steuerung (x-Achse: Zeitschritte)

Die Zeitschrittsteuerung reflektiert die Inhomogenität des Mediums (vgl. Abbildung 5.2); bei homogenen Medien traten um etwa zwei Größenordnungen höhere Schrittweiten auf.

Im Verlaufe der Erstellung dieser Arbeit wurden auch kleine vereinfachte Modelle numerisch untersucht, um das weitere Vorgehen zu motivieren. Diese Experimente wurden in MATLAB durchgeführt, das eine sehr komfortable Umgebung für das Experimentieren mit Algorithmen bietet.

Es wurden Modelle sowohl mit als auch ohne Gebietszerlegung gerechnet, wobei in den Gebietszerlegungsmodellen zunächst immer die exakte Lösung des linearen Problems im Vordergrund stand.

Später ergab sich dann, daß nur wenige Iterationen des Defekt-Korrektur-Verfahrens benötigt wurden, um ein stabileres Verhalten der Schrittweitensteuerung des Extrapolationsverfahrens zu erreichen. Auch später durchgeführte erste Versuche mit dem Vorkonditionierer (ohne Iteration des Korrekturverfahrens) waren vielversprechend.

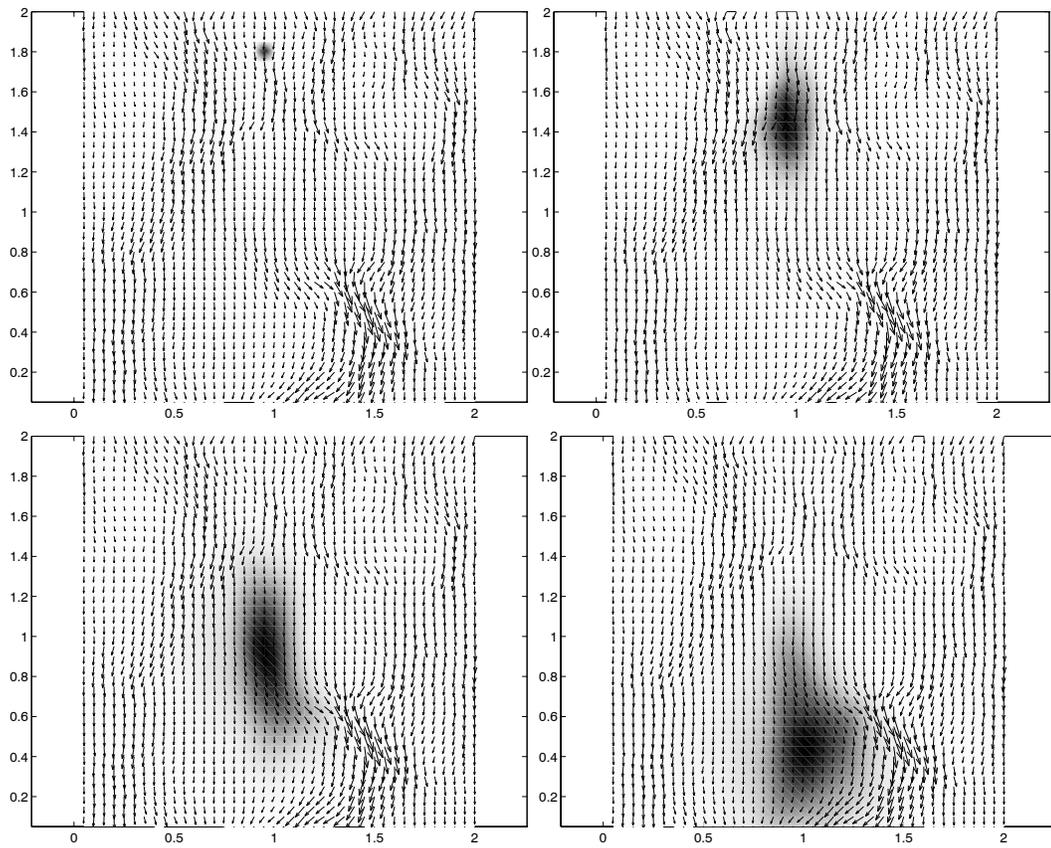


Abbildung 5.3.: Einige Stadien aus einer Simulation: Tracerwolke im Flußfeld des Grundwassers

Zusammenfassung und Ausblick

Die Simulation von Grundwasserfluß und Schadstofftransport im Boden gehört zu den großen Herausforderungen im wissenschaftlichen Rechnen. In genügender Auflösung und auf großen physikalischen Gebieten sind solche Simulationen nur von massiv-parallelen Computern zu bewältigen. Daher ist die Konstruktion neuer paralleler Algorithmen und die Integration bestehender Verfahren in eine parallele Umgebung eine vordringliche Aufgabe für die Anwendungswissenschaften, die Informatik und die praktische Mathematik.

Die Parallelisierung durch Zerlegung des physikalischen Gebiets ist anschaulich, flexibel und gestattet sehr weitgehend die Integration von bestehenden sequentiellen Programmen. Daher werden Gebietszerlegungs-Verfahren sehr häufig z.B. bei der Parallelisierung bestehender Programme verwendet.

In dieser Arbeit wurde die Idee verfolgt, Gebietszerlegungs-Vorkonditionierer zur Berechnung approximierter inverser Matrizen in einem semi-impliziten Verfahren einzusetzen. Dazu wurden die bei der Simulation zu lösenden Anfangs-Randwert-Probleme parabolischer Differentialgleichungen mit der numerischen Linienmethode behandelt. Bei dieser Semi-Diskretisierung im räumlichen Gebiet erhält man ein Anfangswertproblem gewöhnlicher Differentialgleichungen, das mit Standard-Verfahren zur numerischen Lösung steifer Anfangswertprobleme gelöst werden kann. Über die Steifigkeit und Stabilität solcher aus der Semidiskretisierung von Konvektions-Diffusions-Gleichungen entstehender Anfangswertprobleme gewöhnlicher Differentialgleichungen konnten einige Ergebnisse dargestellt werden.

Die semidiskreten Gleichungen wurden mit einem Extrapolationsverfahren von Bader und Deuffhard behandelt. Dieses auf der semi-impliziten Mittelpunkregel basierende Verfahren ist sehr robust und erlaubt eine automatische adaptive Steuerung der Schrittweite und der Ordnung des Verfahrens während der Integration in Zeitrichtung. Im Gegensatz zu impliziten Verfahren kann bei den semi-impliziten Methoden die Jacobi-Matrix des Problems vereinfacht werden, ohne Einbußen in der Robustheit der Verfahren hinnehmen zu müssen. Diese Approximationen der Jacobi-Matrix haben allerdings gewisse Stabilitätsbedingungen zu erfüllen. Dieser Sachverhalt wurde ausführlich dargestellt.

Durch die Extrapolation der semi-impliziten Methode erhält man zudem ein Verfahren, welches — auf einem konstanten Ortsgitter — eine adaptive Steuerung der Schrittweite und der Konsistenzordnung in der Zeit-Integration bietet. Gerade bei derart dissipativen Prozessen wie der Ausbreitung einer Tracer-Wolke im porösen Medium erweist sich diese Technik als sehr effizient. Es zeigt sich, daß die Schrittweiten im Verlauf der Simulation

sehr groß werden können.

Die Parallelisierung mittels eines Gebietszerlegungsverfahrens läßt zwei Optionen offen, von denen die eine zu einem sehr gut parallelisierenden Verfahren führt und die andere im Bedarfsfalle zur Stabilisierung eingesetzt werden kann. Eine hervorragende Parallelisierung kann man sich von einer Approximation der Jacobi-Matrix durch ein Gebietszerlegungsverfahren versprechen; allerdings waren Schätzungen für die Stabilität des dabei entstehenden Verfahrens bisher nicht zu gewinnen. Stabile Verfahren erhält man durch einige wenige Iterationen des auf der Gebietszerlegung basierenden Defektkorrektur-Verfahrens. Auch die hierbei zu erreichende Effizienz wird als zufriedenstellend eingeschätzt.

Neben der weiterhin offenen Frage nach einer Stabilitätstheorie des additiven Schwarz-Vorkonditionierers in der semi-impliziten Mittelpunkregel erscheint insbesondere die Integration weiterer paralleler Verfahren zur Gewinnung von approximierten Inversen für die Zukunft interessant. Allerdings zeigte sich auch im Verlaufe dieser Arbeit die Stabilität als nicht zu vernachlässigende Einschränkung.

Die Verwendung eines modernen Lösungsverfahrens für gewöhnliche Anfangswertprobleme hat sich uneingeschränkt bewährt. Die vertikale numerische Linienmethode scheint recht selten in der Praxis angewandt zu werden; woran das liegt, ist nach den hier gemachten guten Erfahrungen — was sowohl die Programmierung der Verfahren als auch die Erfahrungen mit der automatischen Schrittweitensteuerung angeht — unverständlich.

A. Einiges über Grundwasser-Hydrologie

Die in Kapitel 1 angegebene klassische Darstellung der hier zugrundeliegenden Gleichungen, stammt aus der Hydrologie. Hinter dieser Darstellung verbirgt sich eine andere Formulierung, die auch vom physikalischen Hintergrund her verständlicher ist, nämlich als Erhaltungsgleichungen, die der schwachen Form der Differentialgleichungen, wie sie für die Volumenmethode entwickelt wurde, entspricht. Da man dennoch in der Literatur meist die klassische Form findet, wurde diese zu Beginn vorgestellt.

In diesem zusätzlichen Kapitel sollen die Herleitung der klassischen Gleichungen aus empirischen Modellen und dem Prinzip der Massenerhaltung und der physikalische Zusammenhang kurz erläutert werden.

A.1. Grundwasserfluß

Die der Simulation der Grundwasserbewegung zugrundeliegende parabolische Differentialgleichung, die Richards-Gleichung, beruht auf dem Bewegungsgesetz von Darcy und auf dem Kontinuitätsprinzip (Massenerhaltung).

A.1.1. Das Darcy-Gesetz

Aufgrund von Experimenten formulierte Darcy¹ im Jahre 1856 die nach ihm benannte Bewegungsgleichung für Flüssigkeiten in porösen Medien [17].

Anhand des in Abbildung A.1 (a) dargestellten Versuchsaufbaus kann man das Darcysche Modell veranschaulichen: ein Zylinder mit kreisrundem Querschnitt A ist mit Sand gefüllt, an den Enden verschlossen und mit je einem Ein- und Ausflußrohr versehen; außerdem sind in dem Rohr zwei Manometer angebracht. Nun wird von oben her in einem konstan-

¹Henry Philibert Gaspard Darcy, Dijon 1803 – Paris 1858, war nach seinem Studium in Paris mit der Entwicklung und dem Bau eines Trinkwasser-Versorgungssystems für seine Heimatstadt Dijon beauftragt. Unter anderem plante er, Wasser durch Filterung in Sand zu reinigen. Mit seinen Experimenten, bei denen er Wasser durch mit Sand gefüllte Röhren leitete, entdeckte er das nach ihm benannte lineare Bewegungsgesetz für den laminaren Fluß von Flüssigkeiten in homogenen gesättigten porösen Medien. Damit wurde er zum Begründer der Grundwasser-Hydrologie als einer quantitativen Disziplin (*Encyclopaedia Britannica*).

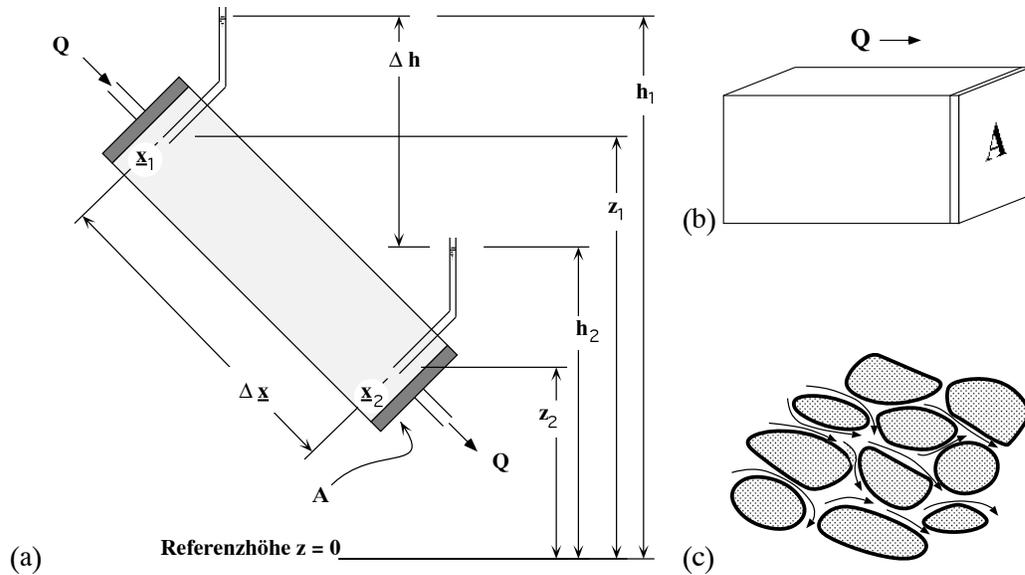


Abbildung A.1.: Darcy-Experiment (a), Filterfluß (b) und Porenfluß (c)

ten Strom Q Wasser eingefüllt, das den gesamten Porenraum in dem Zylinder ausfüllen soll; diese Sättigung ist erreicht, wenn am unteren Ende das Wasser mit demselben Massenstrom Q ausströmt. Die Referenzhöhe $z = 0$ ist beliebig gewählt, allerdings üblicherweise so, daß die Höhen z_1 und z_2 der Manometer-Eingänge (der Messpunkte x_1 und x_2 in dem Rohr) positiv sind. Die Druckhöhen der Manometer (der Druck wird in Metern Wassersäule angegeben) sind $\psi_1 = h_1 - z_1$ und $\psi_2 = h_2 - z_2$. Der Abstand (in der Richtung des Wasserflusses) zwischen den Meßpunkten ist $\Delta x = |x_1 - x_2|$.

Der Fluß, die sogenannte Filtergeschwindigkeit (spezifischer Durchfluß, *specific discharge*), ist definiert als

$$v = \frac{Q}{A}. \quad (\text{A.1})$$

Die Filtergeschwindigkeit muß man von der (mikroskopischen) Porengeschwindigkeit v_p unterscheiden, die bei einer gegebenen Porosität n des Mediums im Mittel $v_p = v/n$ beträgt; die Porosität ist der Anteil des Porenvolumens am Gesamtvolumen $n = V_p/V$. Zu der Unterscheidung vergleiche Abbildung A.1 (b) und (c).

Darcys Experimente zeigten, daß v bei konstantem Abstand Δx proportional zu $\Delta h = h_2 - h_1$ und umgekehrt proportional zu Δx ist, wenn Δh fest bleibt. Er formulierte also das Bewegungsgesetz

$$v = -K \frac{\Delta h}{\Delta x},$$

wobei der Proportionalitätsfaktor eine Konstante für das durchlässige Medium (und das Fluid) ist, die sogenannte hydraulische Leitfähigkeit.

Für die Beschreibung des lokalen Verhaltens einer Flüssigkeit in einem porösen Medium geht man zu dem Grenzwert

$$v = -K \frac{dh}{d\underline{x}} \quad (\text{A.2})$$

über.

Der sogenannte hydraulische Gradient $dh/d\underline{x}$ ist tatsächlich der Gradient, an dem entlang der Wasserfluß stattfindet; dies soll im folgenden kurz begründet werden. Üblicherweise fragt man nach dem Potential, denn das Wasser fließt immer von Stellen höheren Potentials zu Stellen mit niedrigerem Potential. Dieses Potential ist im Falle des Flusses durch ein poröses Medium die mechanische Energie pro Einheitsmasse des Fluids, die allerdings mit der Standrohrspiegelhöhe h identifiziert wird. Denn die Energie setzt sich bei inkompressiblen Flüssigkeiten (mit Dichte ρ) zusammen aus der Arbeit

$$w_1 = mgz, \quad (\text{A.3})$$

die aufgewendet werden muß, um eine Masse m um eine Höhe z zu heben (Gravitationspotential), und der Arbeit

$$w_2 = \frac{p - p_0}{\rho}, \quad (\text{A.4})$$

die bei der Erhöhung des Drucks von p_0 (dem Normaldruck in Referenzhöhe) auf p entsteht (hydrostatisches Potential). In diesem Sinne ist also das totale Potential (bezogen auf eine Einheitsmasse)

$$\Phi = gz + \frac{p - p_0}{\rho}. \quad (\text{A.5})$$

Da nun Darcy den Spiegelstand in seinen Manometern (vgl. Abb. A.1 (a)) als Höhe über der Bezugshöhe angegeben hat, wird der Druck auch als Druckhöhe angegeben; dabei gilt die Beziehung

$$p = \rho g(h - z) + p_0. \quad (\text{A.6})$$

Insgesamt folgt also, daß in der Tat das totale Potential Φ und die Standrohrspiegelhöhe proportional sind:

$$\Phi = gz + \frac{\rho g(h - z) + p_0 - p_0}{\rho} = gh. \quad (\text{A.7})$$

In der Grundwasserhydrologie wird also $h = z + \psi$ als das totale Potential bezeichnet, wobei z , die geodätische Höhe, mit dem Gravitationspotential und ψ , die Druckhöhe, mit dem hydrostatischen Potential identifiziert werden.

Natürlich gilt das Darcy-Gesetz nur in Größenordnungen, die groß sind bezüglich des (mittleren) Porendurchmessers. Hingegen bei sehr kleiner Leitfähigkeit unter einem sehr

geringen hydraulischen Gradienten beziehungsweise bei starkem Fluß in sehr durchlässigen Medien gilt das lineare Bewegungsgesetz nicht mehr.

Beim Grundwasserfluß liegt die sogenannte Reynoldszahl R_e im allgemeinen um Größenordnungen unter 10, und in diesem Bereich wird allgemein das Darcy-Gesetz als gültig angenommen. Die Reynoldszahl ist eine dimensionslose Zahl, welche die Neigung des Flusses zu Turbulenzen beschreibt. Im Falle des Flusses durch ein poröses Medium ist sie folgendermaßen definiert:

$$R_e = \frac{\rho v d}{\mu}; \quad (\text{A.8})$$

dabei sind ρ die Dichte und μ die Viskosität des Fluids, v die Filtergeschwindigkeit und d ein für das poröse Medium charakteristisches Längenmaß, etwa die Größe der Poren oder der mittlere Korndurchmesser.

Für das dreidimensionale Modell wird die Darcy-Gleichung in der Form

$$v = -K \nabla h \quad (\text{A.9})$$

verwendet, wobei der hydraulische Gradient $\nabla h = (\frac{\partial h}{\partial x}, \frac{\partial h}{\partial y}, \frac{\partial h}{\partial z})$ ist.

In einem ungesättigten oder inhomogenen Medium ist die hydraulische Leitfähigkeit K allerdings nicht mehr für das ganze Untersuchungsgebiet konstant, sie variiert mit dem Ort und mit dem Wassergehalt; um dies beschreiben zu können, schreibt man die gesättigte Leitfähigkeit als

$$K_{sat} = \frac{k \rho g}{\mu}, \quad (\text{A.10})$$

mit der nun allein vom Medium abhängigen sogenannten Permeabilität k und der Dichte ρ und Viskosität μ der fließenden Flüssigkeit. Die Inhomogenität des Mediums schlägt sich also in der Permeabilität nieder, die allerdings im Normalfall große Sprünge, meist um mehrere Größenordnungen, aufweist und in einem Gebiet etwa logarithmisch-normalverteilt ist (das heißt $Y = \log k$ ist normalverteilt). Die Permeabilität des Mediums ist nicht nur stark inhomogen (unterschiedlich an verschiedenen Stellen), sondern auch anisotrop (also an einer Stelle in verschiedenen Richtungen unterschiedlich). Aus diesem Grunde wird die hydraulische Leitfähigkeit als ortsabhängiger Tensor beschrieben.

Für die hydraulische Leitfähigkeit im ungesättigten Medium gibt es verschiedene empirische Modelle. Der Wassergehalt $\Theta = nS$, das Produkt aus der Porosität n und dem Sättigungsgrad S (der den Anteil des mit Wasser gefüllten leeren Volumens angibt),

$$\Theta = nS = \frac{V_p}{V} \frac{V_w}{V_p} = \frac{V_w}{V},$$

genügt im ungesättigten Boden zum Beispiel der sogenannten van-Genuchten-Gleichung

$$S_e = \frac{\Theta - \Theta_r}{\Theta_s - \Theta_r} = (1 + (\alpha \psi)^p)^{-q} \quad (\text{A.11})$$

$$0 < \Theta_r < \Theta_s < 1.$$

Dabei sind Θ_s der Sättigungsgehalt (der bei hohem Druck nicht überschritten wird) und Θ_r der Restgehalt (der bei niedrigem Druck nicht unterschritten wird) sowie $\alpha > 0$, $p > 0$ und $q := 1 - 1/p$ Fitting-Parameter, die aus Experimenten gewonnen werden. Den Quotienten S_e bezeichnet man als effektive Sättigung. Da durch die Beziehung (A.11) die Druckhöhe und der Wassergehalt zusammenhängen, kann man die hydraulische Leitfähigkeit auch als Funktion des hydrostatischen Druckes ψ auffassen.

Nach dem Modell von Mualem wird sie beispielsweise als

$$K(\Theta) = K_{sat} S_e^l (1 - (1 - S_e^{\frac{1}{q}})^q)^2 = K(\psi) \quad (\text{A.12})$$

mit S_e wie in (A.11) berechnet. In neueren Modellen wird noch das Verhältnis von ψ und Θ aus der van-Genuchten-Gleichung dahingehend korrigiert, daß ein Hysterisis-Effekt beim Befeuchten und Trocknen des Mediums berücksichtigt wird: bei Trocknung ist der Wassergehalt zu einem bestimmten Druck etwas höher als beim Befeuchten [41].

A.1.2. Massenerhaltung

Auch das Prinzip der Massenerhaltung kann man sich an dem Fluß durch den Zylinder (Abbildung A.1 (a)) klarmachen. Man betrachte den Fluß durch einen Abschnitt des Zylinders, z. B. zwischen den beiden Meßpunkten \underline{x}_1 und \underline{x}_2 . Da der Zylinder einen konstanten Querschnitt hat, beschreiben wir diesen Abschnitt als ein Intervall in \mathbb{R} . Da in dem Zylinder keine Quellen und Senken existieren (wo Wasser hinzukommt oder verschwindet), gilt das Prinzip der Massenerhaltung in der einfachsten Form:

Die zeitliche Änderung der Massendichte $\Theta(\underline{x}, t)$ des Wassers in dem Intervall $[\underline{x}_1, \underline{x}_2]$ ist gleich der Differenz des Flusses von Material an der Stelle \underline{x}_1 in das Intervall hinein und des Flusses an der Stelle \underline{x}_2 aus dem Intervall heraus.

Häufig beschreibt man dieses Prinzip,

$$\left(\frac{d}{dt} \int_0^1 \Theta(\underline{x}_1 + s(\underline{x}_2 - \underline{x}_1), t) ds \right) A = (v(\underline{x}_1, t) - v(\underline{x}_2, t))A,$$

bei hinreichend glatten Integranden Θ und v auch als Kontinuitätsgleichung

$$\frac{\partial}{\partial t} \Theta(\underline{x}, t) = - \frac{\partial}{\partial \underline{x}} v(\underline{x}, t). \quad (\text{A.13})$$

Im allgemeinen Fall der Strömung des Grundwassers (im dreidimensionalen Gelände) beschreibt man die Kontinuitätsgleichung bezüglich eines Kontrollvolumens (vgl. Abbildung A.2)

$$V = [\underline{\xi}, \bar{\xi}] \times [\underline{\eta}, \bar{\eta}] \times [\underline{\zeta}, \bar{\zeta}], \quad (\text{A.14})$$

das groß ist im Verhältnis zu den einzelnen Bodenporen, aber klein gegenüber der Größe des gesamten betrachteten Gebiets [13].

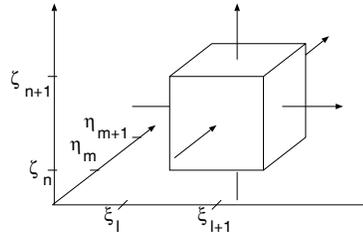


Abbildung A.2.: Kontrollvolumen mit Fluß längs den Flächennormalen

Dann gilt für den Wassergehalt Θ und die Filtergeschwindigkeit $v = (v_1, v_2, v_3)$ aus (A.9) die Beziehung

$$\begin{aligned}
\frac{d}{dt} \iiint_V \Theta dx dy dz &= \int_{\zeta_n}^{\zeta_{n+1}} \int_{\eta_m}^{\eta_{m+1}} v_1((\xi_t, y, z), t) - v_1((\xi_{t+1}, y, z), t) dy dz \\
&+ \int_{\zeta_n}^{\zeta_{n+1}} \int_{\xi_t}^{\xi_{t+1}} v_2((x, \eta_m, z), t) - v_2((x, \eta_{m+1}, z), t) dx dz \\
&+ \int_{\eta_m}^{\eta_{m+1}} \int_{\xi_t}^{\xi_{t+1}} v_3((x, y, \zeta_n), t) - v_3((x, y, \zeta_{n+1}), t) dx dy,
\end{aligned} \tag{A.15}$$

beziehungsweise, wiederum bei hinreichend glatten Funktionen,

$$\frac{\partial}{\partial t} \Theta = -\nabla v \cdot \cdot \tag{A.16}$$

Mit

$$\frac{\partial}{\partial t} \Theta = \frac{\partial \Theta}{\partial \psi} \frac{\partial \psi}{\partial t} =: F(\psi) \frac{\partial}{\partial t} \psi \tag{A.17}$$

(wobei man die Änderung des Wassergehalts mit dem Druck, $F(\psi)$, als den hydraulischen Speicherkoeffizienten bezeichnet) und nach Einfügen eines Terms Q für Quellen und Senken erhält man die Richards-Gleichung, eine nichtlineare parabolische Differentialgleichung, in der sogenannten Druckhöhen-Form (*pressure-head form*)

$$F(\psi) \frac{\partial}{\partial t} \psi = \nabla \cdot (K(\psi) \nabla(\psi + z)) + Q. \tag{A.18}$$

Die entsprechende elliptische Randwertaufgabe für den stationären Fluß erhält man mit $\frac{\partial}{\partial t} \psi = 0$. Bei der Annäherung des Wassergehalts an Θ_s (Sättigung) wird $F(\psi)$ sehr klein, inkompressibler gesättigter Fluß (Inkompressibilität des Mediums und des Fluids) ist stationär. Der Übergang von der belüfteten zur gesättigten Zone führt also auf ein Problem mit freien Rändern, zu einem Übergang vom parabolischen zum elliptischen Problem.

A.1.3. Randwerte

Verschiedene physikalische Situationen am Rande werden mit vier verschiedenen Typen von Randbedingungen modelliert. An Boden-Wasser-Grenzen, zum Beispiel bei Seen und Flüssen, werden Dirichlet-Randwerte (Randbedingung erster Art) vorgegeben, also

$$\psi(x, t) = \psi_D(x, t) \quad \text{auf } \Gamma_D.$$

Durch eine Neumann-Bedingung (Randbedingung zweiter Art) wird horizontaler Fluß vorgegeben, der an Boden-Boden-Rändern auftritt, das heißt

$$-\underline{n}(K\nabla\psi) = v_N(x, t) \quad \text{auf } \Gamma_N,$$

wobei \underline{n} den äußeren Normalenvektor des Randes Γ_N bezeichnet.

Durch eine CauchyBedingung, ebenfalls eine Randbedingung zweiter Art,

$$-\underline{n}(K\nabla(\psi + z)) = v_C(x, t) \quad \text{auf } \Gamma_C$$

wird ein vertikaler Fluß, zum Beispiel die Infiltration von oben und das Aussickern nach unten, beschrieben.

Für die komplexe Situation an der Erdoberfläche (Boden-Luft-Grenze) werden häufig gemischte Randbedingungen (Randbedingungen dritter Art) mit zeitabhaengigen Parametern vorgegeben [41].

A.2. Stofftransport

Die Berechnung des Grundwasserflusses mithilfe der Richards-Gleichung ist der erste Schritt in der Simulation der Ausbreitung gelöster Substanzen im Grundwasser sowie deren Reaktion mit dem Boden. Im folgenden wird die Ausbreitung eines gelösten Stoffes (anhand seiner Konzentration im Grundwasser) als Advektion und Dispersion beschrieben. Als Advektion bezeichnet man den Transport des gelösten Stoffes durch den Fluß, das heißt mit einer der Filtergeschwindigkeit entsprechenden Rate. Das Abweichen und Auseinanderlaufen von dem durch die Advektion vorgezeichneten Weg nennt man (hydrodynamische) Dispersion. Zudem kann man die Reaktion des Stoffes mit dem Boden als Retardierung (Sorption) beschreiben.

Auch bei der Stoffausbreitung geht man von dem Prinzip der Massenerhaltung aus,

$$\begin{aligned} \frac{d}{dt} \iiint_V \Theta c + \rho_b s \, dx dy dz = & \int_{\zeta_n}^{\zeta_{n+1}} \int_{\eta_m}^{\eta_{m+1}} q_1((\xi_l, y, z), t) - q_1((\xi_{l+1}, y, z), t) \, dy dz \\ & + \int_{\zeta_n}^{\zeta_{n+1}} \int_{\xi_l}^{\xi_{l+1}} q_2((x, \eta_m, z), t) - q_2((x, \eta_{m+1}, z), t) \, dx dz \quad (\text{A.19}) \\ & + \int_{\eta_m}^{\eta_{m+1}} \int_{\xi_l}^{\xi_{l+1}} q_3((x, y, \zeta_n), t) - q_3((x, y, \zeta_{n+1}), t) \, dx dy, \end{aligned}$$

oder, falls klassische Lösungen existieren,

$$\frac{\partial}{\partial t} (\Theta c + \rho_b s) = -\nabla \cdot q, \quad (\text{A.20})$$

mit der Konzentration c des gelösten Stoffes im Grundwasser, dem Wassergehalt Θ , der Masse s des sorbierten Stoffes (pro Einheitsmasse Boden) und der Dichte ρ_b des Bodens. Hier werden Effekte wie radioaktiver Zerfall des gelösten oder sorbierten Stoffes, chemische Reaktionen mit anderen im Grundwasser transportierten Stoffen oder mit dem Gas in der ungesättigten Zone sowie weitere Quellen und Senken nicht berücksichtigt. Der advective und der dispersive Transport verbergen sich in dem Flußvektor q :

$$q = q_{disp} + q_{adv}. \quad (\text{A.21})$$

Für die Advektion schreibt man einfach

$$q_{adv} = v c, \quad (\text{A.22})$$

die Dispersion setzt sich allerdings aus verschiedenen Komponenten (Bewegungen im mikroskopischen Bereich) zusammen, die in der Abbildung A.3 skizziert sind, nämlich verschiedene Geschwindigkeiten in den einzelnen Poren-Kanälen, die durch Haftung an den Porenwänden (a) und durch unterschiedliche Porenquerschnitte (b) hervorgerufen werden, sowie die Mischung durch das Auseinanderlaufen und Zusammenfließen von Porenkanälen (c). Hydrodynamische Dispersion ist ein Mischungsvorgang im mikroskopischen Bereich,

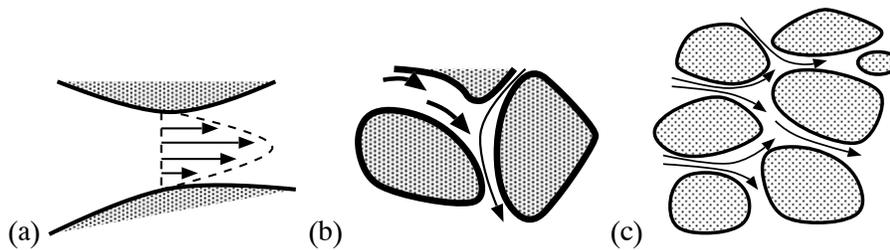


Abbildung A.3.: Hydrodynamische Dispersion

der eng mit der Porengeschwindigkeit zusammenhängt. Auch sie ist anisotrop: in Richtung des Flusses (longitudinale Dispersion) ist sie wesentlich stärker als senkrecht dazu (transversale Dispersion).

Man setzt üblicherweise die hydrodynamische Dispersion als einen diffusiven Prozeß an,

$$q_{disp} = -\Theta D \nabla c, \quad (\text{A.23})$$

mit dem Dispersionstensor D , der sich aus einer longitudinalen und einer transversalen Komponente zusammensetzt. Während es im allgemeinen üblich ist, den Tensor der hydraulischen Leitfähigkeit im Standardkoordinatensystem anzugeben (d.h. der Tensor K in (A.18) hat die Form einer 3×3 -Diagonalmatrix), ist dies im Falle der Dispersion nicht möglich. Im dreidimensionalen Fall definiert das Flußfeld krummlinige Koordinaten, nämlich in jedem Punkt den Einheitsvektor s_l in Flußrichtung und ein orthogonales Paar von Vektoren s_{t_1} und s_{t_2} , die transversal zum Fluß liegen. Es gilt in diesen Koordinaten

$$\nabla \cdot (\Theta D \nabla c) = \frac{\partial c}{\partial s_l} \left(D_l \frac{\partial c}{\partial s_l} \right) + \frac{\partial c}{\partial s_{t_1}} \left(D_t \frac{\partial c}{\partial s_{t_1}} \right) + \frac{\partial c}{\partial s_{t_2}} \left(D_t \frac{\partial c}{\partial s_{t_2}} \right)$$

mit den Koeffizienten D_s und D_t für die longitudinale resp. transversale Dispersion.

Durch Ausrechnen der Transformation

$$\Theta D = S \begin{pmatrix} D_l & 0 & 0 \\ 0 & D_t & 0 \\ 0 & 0 & D_t \end{pmatrix} S^T \quad \text{mit } S = (s_l, s_{t_1}, s_{t_2}), \quad S S^T = I$$

$$\text{und } s_l = (v_1^2 + v_2^2 + v_3^2)^{-\frac{1}{2}} (v_1, v_2, v_3)^T$$

erhält man unter Ausnutzung der Orthogonalität von S den Dispersions-Tensor ΘD als symmetrische 3×3 -Matrix:

$$\Theta D_{ij} = D_t \delta_{ij} + (D_l - D_t) \frac{v_i v_j}{|v|^2} \quad \text{für } i, j \in \{1, 2, 3\}, \quad (\text{A.24})$$

$$|v| = (v_1^2 + v_2^2 + v_3^2)^{\frac{1}{2}}, \quad \delta_{ij} \text{ Kronecker-Symbol.}$$

In [41] wird der Dispersions-Tensor mittels der sogenannten Dispersivitäten a_l und a_t beschrieben, die die Dimension der Länge haben; die Dispersions-Koeffizienten haben dann die Form $D_l = a_l |v|$ bzw. $D_t = a_t |v|$, und der Tensor in (A.24) wird beschrieben als

$$\Theta D_{ij} = a_t |v| \delta_{ij} + (a_l - a_t) \frac{v_i v_j}{|v|}.$$

Die molekulare Diffusion im Wasser ist im Vergleich ziemlich gering und wird hier vernachlässigt.

Die Sorption ist die von der Konzentration abhängige Anlagerung des gelösten Stoffes (etwa eines reaktiven Tracers) im Boden. Die Geschwindigkeit, mit der diese Reaktion stattfindet, wird als groß gegenüber der Filtergeschwindigkeit angenommen, so daß ein chemisches Gleichgewicht herrscht. An dieser Stelle werden zwei Isothermen (Masse des

sorbierten Stoffes in Abhängigkeit von der Konzentration bei gleichbleibender Temperatur) vorgestellt. Für kleine Konzentrationen gilt eine lineare Beziehung

$$s(c) = kc \quad (\text{A.25})$$

mit einem empirischen Koeffizienten k , Abweichungen von dem linearen Verhältnis werden durch die Freundlich-Isotherme,

$$s(c) = kc^j \quad (\text{A.26})$$

mit einem positiven Exponenten $j \leq 1$, beschrieben.

Mit

$$\begin{aligned} \frac{\partial}{\partial t} (\Theta c + \rho_b s(c)) &= \Theta \left(1 + \frac{\rho_b}{\Theta} \frac{\partial s}{\partial c} \right) \frac{\partial c}{\partial t} \\ &=: \Theta R \frac{\partial c}{\partial t} \end{aligned}$$

erhält man für die Transportgleichung (A.20) folgende Form:

$$\Theta R \frac{\partial c}{\partial t} = -\nabla \cdot (vc - \Theta D \nabla c) \quad (\text{A.27})$$

mit dem Retardierungsfaktor R , der für die oben angegebenen Isothermen die Werte

$$R_{linear} = 1 + \frac{\rho_b}{\Theta} k \quad \text{bzw.} \quad R_{freundlich} = 1 + \frac{\rho_b}{\Theta} k j c^{j-1}$$

hat.

Im Falle der linearen Isotherme ist also die Retardierung des Transports durch die Sorption von der Stoffkonzentration unabhängig. Legt man die Freundlich-Isotherme zugrunde, erhält man einen von der Konzentration abhängigen Retardierungsfaktor und also eine nichtlineare parabolische Differentialgleichung.

Für die Simulation eines Tracer-Experiments, bei dem ein Tracer in der Konzentration \hat{c} an der Stelle x_0 injiziert wird, wählt man als Anfangsbedingung beispielsweise

$$c_0 = \frac{\hat{c}}{n} \delta(x - x_0),$$

wobei δ eine Approximation der Dirac-Verteilung ist [41]. Für die Simulation eines Tracer-Experiments auf einem hinreichend großen Gebiet kann man homogene Dirichlet-Bedingungen auf dem Rand vorgeben.

Die Transportgleichung (A.27) ist über den Wassergehalt und die Filtergeschwindigkeit an die Richards-Gleichung (A.18) gekoppelt. Man berechnet also Θ und v zunächst mittels der in Abschnitt A.1 dargestellten Gleichungen und löst dann die Gleichung (A.27); auf eine Rückkopplung wird verzichtet, da im allgemeinen die Stoffkonzentration gegenüber der Flüssigkeitsdichte nicht ins Gewicht fällt [41].

Literaturverzeichnis

- [1] H. W. Alt and S. Luckhaus, *Quasilinear elliptic-parabolic differential equations*, Mathematische Zeitschrift **183** (1983), 311–341.
- [2] T. M. Apostol, *Mathematical Analysis*, second ed., Addison-Wesley, Reading, Massachusetts, 1974.
- [3] O. Axelsson, *A survey of preconditioned iterative methods for linear systems of algebraic equations*, BIT **25** (1985), 166–187.
- [4] G. Bader and P. Deuffhard, *A semi-implicit mid-point rule for stiff systems of ordinary differential equations*, Numerische Mathematik **41** (1983), 373–398.
- [5] R. Barret, M. Berry, T. Chan, J. Demmel, J. Donato, J. Dongarra, V. Eijhout, R. Pozo, C. Romine, and H. van der Vorst, *Templates for the Solution of Linear Systems: Building Blocks for Iterative Methods*, SIAM, 1993.
- [6] A. Basermann, *Iterative Verfahren für dünnbesetzte Matrizen zur Lösung technischer Probleme auf massiv-parallelen Systemen*, Berichte des Forschungszentrums Jülich Jül-3015, Zentralinstitut für Angewandte Mathematik, Forschungszentrum Jülich, Januar 1995.
- [7] A. Basermann, B. Reichel, and C. Schelthoff, *Preconditioned CG methods for sparse matrices on massively parallel machines*, Interner Bericht KFA-ZAM-IB-9531, Zentralinstitut für Angewandte Mathematik, Forschungszentrum Jülich, December 1995.
- [8] Th. Bemmeerl, *Programmierung skalierbarer Multiprozessoren*, Reihe Informatik, vol. 84, Bibliographisches Institut, Mannheim, 1992.
- [9] P. Deuffhard, *Order and stepsize control in extrapolation methods*, Numerische Mathematik **41** (1983), 399–422.
- [10] ———, *Recent progress in extrapolation methods for ordinary differential equations*, SIAM Review **27** (1985), no. 4, 505–535.
- [11] P. Deuffhard and F. Bornemann, *Numerische Mathematik II*, Walter de Gruyter, Berlin-New York, 1994.
- [12] J. Dongarra, A. Lumsdaine, R. Pozo, and K. A. Remington, *IML++ v. 1.2 Iterative Methods Library reference guide*, February 1995.

- [13] S. Dyck and G. Peschke, *Grundlagen der Hydrologie*, VEB Verlag für Bauwesen, Berlin, 1983.
- [14] R. Esser and R. Knecht, *Intel Paragon XP/S — architecture and software environment*, Interner Bericht KFA-ZAM-IB-9305, Zentralinstitut für Angewandte Mathematik, Forschungszentrum Jülich, April 1993.
- [15] R. E. Ewing, R. D. Lazarov, J. E. Pasciak, and A. T. Vassilev, *Mathematical modeling, numerical techniques, and computer simulation of flows and transport in porous media*, Technical Report ISC-95-03-MATH, Institute for Scientific Computation, ISC Texas A & M University, August 1995.
- [16] G. M. Fichtenholz, *Differential- und Integralrechnung*, neunte Auflage, VEB Deutscher Verlag der Wissenschaften, Berlin, 1986.
- [17] R. A. Freeze and J. A. Cherry, *Groundwater*, Prentice-Hall, Englewood Cliffs, N. J., 1979.
- [18] R. W. Freund, G. H. Golub, and N. M. Nachtigal, *Iterative Solution of Linear Systems*, Acta Numerica (1992), 1–44.
- [19] M. Gerndt, O. Neuendorf, J. Prümmer, and H. Vereecken, *A comparison of two parallelization strategies for TRACE*, Interner Bericht KFA-ZAM-IB-9425, Zentralinstitut für Angewandte Mathematik, Forschungszentrum Jülich, November 1994.
- [20] W. Gragg, *On extrapolation algorithms for ordinary differential initial value problems*, SIAM Journal of Numerical Analysis **2 B** (1965), 384–403.
- [21] *Große Gleichungssysteme auf Parallelrechnern. Numerik und Anwendungen*, Bergische Universität Gesamthochschule Wuppertal, 6. und 7. Oktober 1994.
- [22] Ch. Großmann and H.-G. Roos, *Numerik partieller Differentialgleichungen*, Teubner, Stuttgart, 1992.
- [23] W. Hackbusch, *Iterative Solution of Large Sparse Systems of Equations*, Springer-Verlag, New York, 1994.
- [24] E. Hairer, G. Bader, and Ch. Lubich, *On the stability of semi-implicit methods for ordinary differential equations*, BIT **22** (1982), 211–232.
- [25] E. Hairer, S. P. Nørsett, and G. Wanner, *Solving Ordinary Differential Equations I*, Springer-Verlag, Berlin, Heidelberg, New York, 1987.
- [26] E. Hairer and G. Wanner, *Solving Ordinary Differential Equations II*, second ed., Springer-Verlag, Berlin, Heidelberg, New York, 1991.
- [27] P. R. Halmos, *Finite-Dimensional Vector Spaces*, second ed., Springer-Verlag, New York, reprinted 1974.
- [28] J. J. I. M. van Kan and A. Segal, *Numerik partieller Differentialgleichungen für Ingenieure*, Teubner, Stuttgart, 1995.

- [29] D. E. Keyes, *Domain decomposition: A bridge between nature and parallel computers*, Preprint, Draft Version 29.05.1992.
- [30] O. Neuendorf, *Numerische 3D-Simulation des Stofftransports in einem heterogenen Aquifer*, Dissertation am Lehrstuhl für Geologie, Geochemie und Lagerstätten des Erdöls und der Kohle (RWTH Aachen); angefertigt am Institut für Erdöl und organische Geochemie (Forschungszentrum Jülich), Rheinisch-Westfälische Technische Hochschule Aachen, 1996.
- [31] R. Pozo, *MV++ v. 1.5a Matrix / Vector Class reference guide*, November 1995.
- [32] R. Pozo, K. A. Remington, and A. Lumsdaine, *SparseLib++ v. 1.5 Sparse Matrix Class Library reference guide*, November 1995.
- [33] C. Schelthoff, *Vergleich von parallelen Verfahren zur Vorkonditionierung für die Methode der konjugierten Gradienten*, Berichte des Forschungszentrums Jülich Jül-2913, Zentralinstitut für Angewandte Mathematik, Forschungszentrum Jülich, März 1994.
- [34] H. A. Schwarz, *Ueber einen Grenzübergang durch alternirendes Verfahren*, Gesammelte mathematische Abhandlungen (Berlin), vol. 2, Springer, Berlin, 1890, pp. 133–143.
- [35] L. F. Shampine, *ODE solvers and the method of lines*, Numerical Methods for Partial Differential Equations **10** (1994), 739–755.
- [36] H. D. Simon, *Partitioning of unstructured problems for parallel processing*, Computing Systems in Engineering **2** (1991), no. 3, 135–148.
- [37] B. F. Smith, *Domain decomposition methods for partial differential equations*, to appear in D. E. Keyes, A. Sameh and V. Venkatakrisnan, editors, “Parallel Numerical Algorithms”, proceedings of an ICASE/LaRC workshop, Kluwer, Dordrecht 1995.
- [38] J. Stoer, *Numerische Mathematik 1*, fünfte Auflage, Springer-Verlag, Berlin-Heidelberg-New York, 1989.
- [39] K. Strehmel and R. Weiner, *Numerik gewöhnlicher Differentialgleichungen*, Teubner, Stuttgart, 1995.
- [40] M. D. Tocci, C. T. Kelley, and C. T. Miller, *Accurate and economical solution of the pressure-head form of Richards' equation by the method of lines*, submitted to Water Resources Research, August 1995.
- [41] H. Vereecken, G. Lindenmayr, O. Neuendorf, U. Döring, and R. Seidemann, *TRACE. A mathematical model for reactive transport in 3D variably saturated porous media*, Internal Report KFA/ICG-4 No. 501494, Institut für Erdöl und organische Geochemie (ICG-4), Forschungszentrum Jülich, August 1994.

- [42] H. Vereecken, O. Neuendorf, G. Lindenmayr, and A. Basermann, *A parallel Schwarz domain decomposition method for the numerical solution of transient water flow in heterogeneous porous media*, Internal Report KFA/ICG-4 No. 500795, Institut für Erdöl und organische Geochemie (ICG-4), Forschungszentrum Jülich, May 1995.
- [43] D. Vinckier, B. Saddak, and A. Basermann, *A parallel industrial code for deformations with contact*, Interner Bericht KFA-ZAM-IB-9603, Zentralinstitut für Angewandte Mathematik, Forschungszentrum Jülich, Februar 1996.
- [44] R. Wimmershoff, *Entwicklung und Implementierung einer dreidimensionalen Partitionierungsstrategie für das Programm TRACE auf einem massiv-parallelen Rechner*, Diplomarbeit, Rheinisch-Westfälische Technische Hochschule Aachen, 1995.

Danksagung

Dem Direktor des Zentralinstituts für Angewandte Mathematik (ZAM) der Forschungszentrum Jülich GmbH, Herrn Prof. Dr. F. Hoßfeld, danke ich herzlich für die Betreuung der Arbeit und für die Möglichkeit, in der inspirierenden Umgebung eines interdisziplinären Instituts an einem so wichtigen und interessanten Thema der praktischen Mathematik zu arbeiten.

Herr Privatdozent Dr. H. Jarausch vom Lehrstuhl für Numerische Mathematik der RWTH Aachen hat freundlicherweise das Korreferat für diese Arbeit übernommen. Darüberhinaus hat er auch die Formulierung des Themas und die Gestaltung der Arbeit mit großem Engagement begleitet. Dafür und für seine stete Gesprächsbereitschaft danke ich ihm herzlich.

Herrn Dr. P. Weidner, dem Leiter der Abteilung Mathematik des ZAM, danke ich für seine kritisch-freundliche Begleitung der Arbeit. Nicht nur seine konkreten Hinweise, sondern auch seine ruhige Art waren mir in dieser Zeit eine große Hilfe.

Den Mitarbeitern des ZAM und des ICG-4, die in zahllosen Gesprächen und Diskussionen an dieser Arbeit Anteil genommen haben, danke ich an dieser Stelle für ihre Bereitschaft, sich mit meinen Fragen und Einwendungen auseinanderzusetzen. Namentlich nennen möchte ich Dr.-Ing. A. Basermann, der sich am Fortgang der Arbeit immer sehr interessiert gezeigt hat.

Allen Diplomandinnen und Doktorandinnen des ZAM (die Männer sind natürlich mitgemeint) danke ich für die nette Atmosphäre und gegenseitige Hilfsbereitschaft.

Meine Familie, Freunde und Freundinnen haben mich in der Zeit des Studiums und der Diplomarbeit begleitet; dies ist der Ort, ihnen zu danken.

Besonders danke ich meiner Freundin Stephanie, die alle Hoch- und Tiefpunkte der letzten Zeit mit mir durchlebt hat.

Meinem Vater danke ich für seine stete Unterstützung, er hat mir das Studium ermöglicht und ließ nie einen Zweifel an dem Vertrauen, das er in mich setzte.

In Dankbarkeit gedenke ich meiner lieben Mutter,
Margaretha Anna van der Linden-Rijnja.