

DIPLOMARBEIT

**Konstruktion B–Spline–basierter
monotoner Mehrgitterverfahren
zur Bewertung
Amerikanischer Optionen**

Angefertigt am Institut für Angewandte Mathematik

Vorgelegt der

Mathematisch–Naturwissenschaftlichen Fakultät der
Rheinischen Friedrich–Wilhelms–Universität Bonn

Mai 2004

Von

Markus Holtz

aus Bonn

Inhaltsverzeichnis

Einleitung	1
1 Optionen und ihre Bewertung	7
1.1 Grundlagen der Optionsbewertung	7
1.1.1 Definitionen und Bezeichnungen	7
1.1.2 Motive für den Kauf von Optionen	9
1.1.3 Strukturaussagen	10
1.2 Das Black–Scholes Modell	12
1.2.1 Einleitung und Annahmen	12
1.2.2 Europäische Optionen und die Black–Scholes Gleichung	13
1.2.3 Die Bewertung Amerikanischer Optionen als freies Randwertproblem	15
1.3 Risikokennziffern – The Greek Letters	17
1.4 Verfahren zur Bewertung von Amerikanischen Optionen	19
1.4.1 Baumverfahren	19
1.4.2 Finite–Differenzen und Finite–Elemente	24
1.4.3 Analytische Approximationen und weitere Verfahren	25
2 Finite–Elemente Lösung des freien Randwertproblems	27
2.1 Formulierung des freien Randwertproblems: Einleitung und Übersicht	27
2.2 Herleitung der variationellen Form	29
2.2.1 Das freie Randwertproblem als lineares komplementäres Problem	29
2.2.2 Transformation und Lokalisation	31
2.2.3 Schwache Form und Reduktion auf homogene Randbedingungen	33
2.3 Diskretisierung	35
2.3.1 Diskretisierung in der Zeit	35
2.3.2 Diskretisierung im Ort	36
2.4 Fehlerabschätzungen	39
2.4.1 Elliptische Variationsgleichungen	40
2.4.2 Elliptische Variationsungleichungen	42

2.5	Implementierung des Verfahrens	44
2.5.1	Aufstellen von Massen-, Steifigkeitsmatrix und der rechten Seite	44
2.5.2	Bestimmung der Ableitungen und des freien Randes	47
2.6	Das projektive Gauss–Seidel–Verfahren	49
3	B–Spline basierte monotone Mehrgitterverfahren	53
3.1	Mehrgitteralgorithmen für Variationsungleichungen: Einleitung und Übersicht . .	53
3.2	Relaxationsverfahren	61
3.2.1	Erweiterte Relaxationsverfahren	61
3.2.2	Approximative erweiterte Relaxationsverfahren	63
3.2.3	Asymptotisches Konvergenzverhalten	65
3.2.4	Abgeschnittene Version der erweiterten Relaxationsverfahren	68
3.3	Monotone Mehrgitterverfahren (MMG)	70
3.3.1	Implementierung des Verfahrens	70
3.3.2	Asymptotische Fehlerabschätzungen	71
4	Konstruktion monotoner und quasioptimaler Grobgitterapproximationen	73
4.1	Monotone Grobgitterapproximationen	74
4.2	Quasioptimale Grobgitterapproximationen	78
4.3	Formulierung als lineare Optimierungsaufgabe	81
4.4	Lösung der linearen Optimierungsaufgabe	83
4.4.1	Lösung mit dem Simplex–Algorithmus	84
4.4.2	Direkte Lösung für Hutfunktionen mit Fourier–Motzkin–Elimination . . .	84
4.4.3	Näherungsweise Lösung für B–Splines höherer Ordnung mit dem Optimierten Grobgitterhindernis–Konstruktionsalgorithmus (OGK) . . .	86
5	Verallgemeinerung auf mehrere Dimensionen	95
5.1	Tensorprodukt–B–Splines	95
5.2	Monotone und quasioptimale Grobgitterapproximationen	97
5.3	Mehrdimensionale Optionspreistheorie	102
5.3.1	Amerikanische Basket–Optionen	102
5.3.2	Amerikanische Optionen mit stochastischer Volatilität	103
5.3.3	Pfadabhängige Amerikanische Optionen	104
6	Numerische Ergebnisse	107
6.1	Europäische und Amerikanische Optionspreise	107
6.2	Einfluss der Ordnung der Ansatzfunktionen auf die Qualität der Lösung	111
6.3	Einfluss der Parameter	115
6.4	Konvergenzverhalten des PSOR– und der Mehrgitterverfahren	116
6.5	Mehrgitter–Konvergenzraten	120

Zusammenfassung	120
A Mathematische Grundlagen	125
A.1 B-Splines	125
A.2 Sobolevräume und Bilinearformen	128
A.3 Variationsungleichungen und Minimierungsprobleme	131
A.4 Hindernisprobleme und lineare komplementäre Probleme	133
Notation	135
Bibliography	139

Einleitung

Seit den berühmten Arbeiten von F. Black, R. Merton und M. Scholes 1973 [BS] ist der systematische Handel mit *Optionen* weltweit massiv angewachsen und hat sich inzwischen als ein umfangreiches und eigenständiges Segment im Investmentgeschäft etabliert. Optionen sind Finanzverträge, die ihrem Halter das Recht einräumen, einen zugrundeliegenden Basiswert, z.B. eine Aktie oder eine Devisen, zu einem festgelegten Preis an einem zukünftigen Zeitpunkt kaufen oder verkaufen zu dürfen. Mittlerweile sind sie als Mittel zur Risikoabsicherung von Handelsgeschäften und Finanzanlagen oder als Mittel zur Spekulation auf den Finanzmärkten nicht mehr wegzudenken. Allein in Deutschland wurden im letzten Jahr Optionen im Wert von 150 Mrd. Euro umgesetzt (Quelle Dt. Bank). Sie werden sowohl in standardisierter Form an den Börsen angeboten, als auch individuell auf Kundenbedürfnisse zugeschnitten. Insbesondere im zweiten Fall kann sich der Preis eines solchen Finanzvertrages nicht aus dem Verhältnis von Angebot und Nachfrage ergeben, sondern es wird ein sogenannter *fairer Preis* durch eine geeignete mathematische Modellierung und Berechnung bestimmt.

Man unterscheidet zwischen *Europäischen* und *Amerikanischen* Optionen. Unter bestimmten Annahmen an den Finanzmarkt wird in [BS] nachgewiesen, dass der faire Preis einer Europäischen Option ein Randwertproblem zu einer instationären Diffusions-Konvektions-Reaktions-Gleichung zweiter Ordnung, der sogenannten *Black-Scholes Gleichung*, erfüllt. Für dieses Randwertproblem kann relativ einfach eine analytische Lösung angegeben werden. Amerikanische Optionen, die den mit Abstand größten Teil der an den Börsen gehandelten Optionen ausmachen, dürfen im Gegensatz zu Europäischen Optionen jederzeit ausgeübt werden. Bei der Bewertung wirkt sich dieses zusätzliche Recht insofern aus, dass nun ein Randwertproblem zur Black-Scholes Gleichung in einem Gebiet gelöst werden muss, dessen Rand als Teil des Problems mitzubestimmen ist. Man spricht dann von einem *freien Randwertproblem*, in dem der freie Rand den sogenannten optimalen Ausübungszeitpunkt der Option definiert. Für dieses Problem ist trotz der großen praktischen Relevanz bisher keine analytische Lösung im Sinne einer geschlossenen Lösungsformel bekannt und kann vielleicht auch nicht erreicht werden. Daher ist die Konstruktion und Analyse analytischer Approximationen und effizienter und flexibler numerischer Verfahren weiterhin Gegenstand aktiver Forschung (s. [Bl, BR, BJS, CP, DH, FV, HW, JLL, Oo, Me, Zh2]).

Freie Randwertprobleme tauchen in verschiedenen Gebieten der Angewandten Mathematik auf. Sie sind generell schwieriger zu behandeln als klassische Randwertprobleme und es sind nur in wenigen Ausnahmefällen analytische Lösungen bekannt. In vielen Fällen, z.B. bei *Hindernisproblemen*, zu denen auch die Bewertung Amerikanischer Optionen gehört, lassen sie sich so umformulieren, dass der freie Rand in der neuen Formulierung nicht mehr explizit auftaucht. Bei entsprechend schwacher Formulierung ist jetzt eine *Variationsungleichung* oder äquivalent dazu ein Minimierungsproblem über einer konvexen Menge zu lösen.

Bei einer geeigneten Diskretisierung lässt sich dieses iterativ mit einer projektiven Variante des klassischen Gauss-Seidel Verfahrens (s. [Cr]) lösen. Die Konvergenz des Verfahrens ist bei kleinen Gitterweiten zwar sehr langsam, sie lässt sich aber mit dem Einsatz von Mehrgittertechniken ähnlich wie bei Randwertproblemen beschleunigen (s. [BC, HM, Ho, Ko1, Ko2, Kr, Ma, Oo]).

Die prinzipielle Schwierigkeit bei der Konstruktion solcher Mehrgitterverfahren ist es, den freien Rand bzw. die Hindernisfunktion so auf gröberen Gittern darzustellen, dass die Grobgitterkorrektur möglichst effektiv ist, die korrigierte Lösung die konvexe Lösungsmenge aber nicht verlässt.

Bei den *monotonen Mehrgitterverfahren (MMG)* aus [Ko1] wird dies durch die Verwendung spezieller, sogenannter *monotoner und quasioptimaler Restriktionsoperatoren* für die Hindernisfunktion erreicht. Auf diese Weise kann nachgewiesen werden, dass das Verfahren besonders robust ist. Ein systematisches Anpassen der Basisfunktionen an die Position des freien Randes führt zur *abgeschnittenen Variante (TrMMG)* des monotonen Mehrgitterverfahrens und zu deutlich schnellerer Konvergenz. In numerischen Tests stellt man hier asymptotisch ähnliche Konvergenzraten wie bei Standard-Mehrgitterverfahren fest.

Die Arbeit [Ko1] bleibt, wie sämtliche mir bekannte Arbeiten zu Mehrgitterverfahren für freie Randwertprobleme, auf stetige, stückweise lineare Ansatzfunktionen zur Diskretisierung des kontinuierlichen Problems beschränkt. Diese werden häufig als Hutfunktionen bezeichnet. In der Finanzmathematik sind jedoch möglichst gute Näherungen an die Ableitungen der Lösung ebenso wichtig wie die Bestimmung der Lösung selbst. Die Ableitungen haben große Bedeutung als *Risikokennziffern*, die benötigt werden, um an die Marktsituation angepasste Absicherungsstrategien für Aktien- oder Options-Portfolios ableiten zu können. Die Verwendung glatter Ansatzfunktionen hat hierbei im Vergleich zu Hutfunktionen den Vorteil, dass die höheren Ableitungen wesentlich genauer bestimmt werden können.

Gegenstand dieser Arbeit ist es, die Konstruktion und die Theorie der monotonen Mehrgitterverfahren aus [Ko1] auf beliebig glatte Ansatzfunktionen zu verallgemeinern und die Verfahren zur Bestimmung von Werten und Risikokennziffern Europäischer und Amerikanischer Optionen anzuwenden. Als Basisfunktionen werden *B-Splines* verwendet. Aufgrund der Positivität der B-Spline-Basisfunktionen sind sie zur Konstruktion monotoner und quasioptimaler Restriktionsoperatoren weitaus besser geeignet als die durch Hermite Interpolation bestimmten nodalen C^1 -glatten Finite-Elemente-Funktionen. Außerdem sind sie aufgrund ihrer Verfeinerungseigenschaften für den Einsatz in Mehrgitterverfahren prädestiniert.

Das *erste Kapitel* beinhaltet die Grundlagen der Optionsbewertung im Rahmen des Black-Scholes Modells. In Abschnitt 1.1 werden zunächst die wichtigsten Arten von Optionsverträgen und das wichtige Prinzip eines arbitragefreien Marktes eingeführt. Man unterscheidet hierbei zwischen Kaufoptionen (Call) und Verkaufsoptionen (Put). Es wird veranschaulicht, wie Optionen zur Risikoabsicherung oder zur Spekulation verwendet werden können. Außerdem wird der Satz von Merton und die sogenannte Put-Call-Parität behandelt. Damit ist motiviert, warum im weiteren Verlauf dieser Arbeit nur Amerikanische Verkaufsoptionen und Europäische Kaufoptionen betrachtet werden. In Abschnitt 1.2 wird das Black-Scholes Modell eingeführt. Unter den Modellannahmen, zu denen im Wesentlichen die Annahmen eines arbitragefreien Marktes und von log-normal verteilten Aktienkursen gehören, wird zunächst das Randwertproblem hergeleitet, welches den fairen Wert einer Europäischen Optionen bestimmt und für das eine analytische Lösung in Form der berühmten Black-Scholes Formel bekannt ist. Danach wird das instationäre freie Randwertproblem vorgestellt, welches bei der Bewertung Amerikanischer Optionen auftritt und in den nachfolgenden Kapiteln numerisch gelöst wird. In Abschnitt 1.3 werden die in der Praxis wichtigen Risikokennziffern Delta, Gamma und Theta eingeführt, die den Ableitungen des Optionswertes entsprechen und den Einsatz von Verfahren höherer Ordnung zu ihrer präzisen Bestimmung erfordern. Eine Literaturübersicht über Verfahren zur Bewertung von Amerikanischen Optionen findet sich in Abschnitt 1.4. Zu den verbreitetsten Methoden zählen Baumverfahren, Finite-Differenzen bzw. Finite-Elemente-Verfahren und analytische Approximationen. Die Inhalte von Kapitel 1 sind im Wesentlichen den Büchern [Hu, Sa, WHD2, Se] entnommen.

Im *zweiten Kapitel* wird ein Finite-Elemente-Verfahren zur Lösung des freien Randwertproblems beschrieben, welches in Kapitel 1 zur Bewertung Amerikanischer Optionen hergeleitet wurde. Diese Abschnitte orientieren sich eng an den Darstellungen in [Se, WHD1]. Dort ist das Verfahren für Hutfunktionen beschrieben. Es wird in der vorliegenden Arbeit auf B-Spline-Ansatzfunktionen beliebiger Glattheit verallgemeinert. Es wird gezeigt, dass die damit verbundenen Schwierigkeiten bei der Modellierung der Randdaten umgangen werden können, indem die schwache Form des Problems im Unterschied zu [Se, WHD1] auf homogene Randdaten reduziert wird. In Abschnitt 2.2 wird zunächst die schwache Form des freien Randwertproblems hergeleitet. Die Black-Scholes Gleichung wird dafür in die Wärmeleitungsgleichung transformiert und auf homogene Randdaten reduziert. Finite-Differenzen Diskretisierung in der Zeit und Finite-Elemente Diskretisierung im Ort führen im Abschnitt 2.3 zur vollständig diskreten Form. Das hieraus resultierende Minimierungsproblem über einer konvexen Menge im \mathbb{R}^2 lässt sich nun mit dem projektiven Gauss-Seidel-Verfahren lösen. Es wird in Abschnitt 2.6 beschrieben. Massen-, Steifigkeitsmatrix und die rechte Seite lassen sich bei äquidistanten Gittern explizit aufstellen. Außerdem können die Ableitungen der Lösung durch direktes Ableiten genügend glatter Ansatzfunktionen bestimmt werden (Abschnitt 2.5). Mit quadratischen bzw. kubischen B-Splines erhält man auf diese Weise stückweise stetige Approximationen an die erste bzw. an die ersten beiden Ableitungen der Lösung.

Zur Konvergenzbeschleunigung des Finiten Elemente Verfahrens werden in *Kapitel 3* B-Spline-basierte monotone Mehrgitterverfahren behandelt. In Abschnitt 3.1 wird zunächst eine Literaturübersicht über Mehrgitterverfahren für freie Randwertprobleme gegeben. Insbesondere wird genauer auf die Mehrgitteralgorithmen PFAS aus [BC] und MMG aus [Ko1] eingegangen. In Abschnitt 3.2 wird die Theorie der monotonen Mehrgitterverfahren erläutert und von Hutfunktionen auf B-Spline-Basisfunktionen beliebiger Ordnung übertragen. Dazu werden zunächst erweiterte Relaxationsmethoden für Variationsungleichungen betrachtet. Dabei ergibt sich der Bedarf nach monotonen und quasioptimalen Approximationen der Hindernisfunktion auf größeren Gittern. Ein systematisches Anpassen der Basisfunktionen an den aktuellen freien Rand führt in Abschnitt 3.2.4 zur abgeschnittenen Version der Verfahren. In der Vorgehensweise halten sich diese Abschnitte eng an die Arbeit [Ko1]. Anstelle von Funktionswerten an Gitterpunkten wird jedoch auf die Entwicklungskoeffizienten der B-Spline-Basis zurückgegriffen.

Der wesentliche Baustein monotoner Mehrgitterverfahren ist die Konstruktion monotoner und quasioptimaler Restriktionsoperatoren zur Approximation der Hindernisfunktion auf größeren Gittern. Die Konstruktion wird in *Kapitel 4* diskutiert.

Die Hindernisfunktion ist hierbei durch einen B-Spline $\tilde{S} : I \rightarrow \mathbb{R}$ der Ordnung k mit \tilde{n} Freiheitsgraden auf einem Gitter $\Delta \subset I := [a, b]$ mit der Gitterweite h gegeben. Ziel ist es, einen B-Spline $S : I \rightarrow \mathbb{R}$ gleicher Ordnung mit n Freiheitsgraden auf einem größeren Gitter $T \subset \Delta$ mit Gitterweite $H := 2h$ zu finden, so dass gilt:

1. Für alle $x \in I$ gelte $S(x) \leq \tilde{S}(x)$ (Monotonie).
2. Für alle $x \in I$ gelte $S(x) \geq L_k(x)$ für eine noch zu spezifizierende untere Schranke $L_k(x)$ (Quasioptimalität).
3. Der Spline S sollte die Hindernisfunktion \tilde{S} möglichst gut in Hinblick auf eine noch zu spezifizierende Zielfunktion F approximieren (Optimalität der Approximation).
4. Der Aufwand zur Konstruktion der Funktion S soll $O(n)$ Operationen nicht überschreiten (optimale Komplexität).

Im Rahmen eines Mehrgitteralgorithmus sichert die erste Bedingung die Monotonie und Robustheit des Verfahrens, die zweite eine asymptotische Reduktion des Verfahrens zu einer linearen Relaxation, die dritte eine möglichst effiziente Grobgitterkorrektur und die vierte die Erhaltung der optimalen Komplexität des Mehrgitterverfahrens.

In Abschnitt 4.1 wird zunächst gezeigt, dass die Forderung nach Monotonie auf ein lineares Ungleichungssystem für die Entwicklungskoeffizienten der B-Spline-Basis führt. Damit kann in Abschnitt 4.2 die untere Schranke L_k konstruiert werden. Für den Fall der Ordnung $k = 2$ führt dies auf die Restriktionen, die in der Arbeit [Ma] angegeben sind. In Abschnitt 4.3 wird schließlich die Zielfunktion F spezifiziert. Zusammen mit den Ergebnissen der beiden vorhergehenden Abschnitte kann das Problem, eine optimale monotone und quasioptimale Restriktion S zu bestimmen, dann als Lösung einer linearen Optimierungsaufgabe formuliert und mit dem Simplex-Algorithmus gelöst werden. Da der Aufwand des Simplex-Algorithmus jedoch zu hoch ist, um ihn im Rahmen eines Mehrgitterverfahrens einsetzen zu können, wird in Unterabschnitt 4.4.2 mit der sogenannten Fourier-Motzkin-Elimination eine direkte Lösung der linearen Optimierungsaufgabe für den Fall $k = 2$ hergeleitet. Hierbei werden die mit geometrischen Überlegungen konstruierten Restriktionsoperatoren aus [Ko1] wiederentdeckt. Für den Fall $k > 2$ wird in Unterabschnitt 4.4.3 mit dem *optimierten Grobgitterhindernis-Konstruktionsalgorithmus (OGK)* ein neues näherungsweise Lösungsverfahren vorgeschlagen, welches die spezielle Struktur der Matrizen A_k ausnutzt und dadurch lediglich $O(n)$ Operationen benötigt.

In *Kapitel 5* wird ausgeführt, dass sich über einen Tensorproduktansatz alle Ergebnisse auf höhere Dimensionen übertragen lassen. Für den $2d$ -Fall ist sowohl die direkte Lösung mittels Fourier-Motzkin Elimination für Hutfunktionen als auch die näherungsweise Lösung des OGK-Algorithmus für C^1 -glatte B-Splines angegeben. Die dazu benötigten Eigenschaften von Tensorprodukt-B-Splines sind in Abschnitt 5.1 zusammengefasst. Als mögliche Anwendungen werden in Abschnitt 5.3 Optionsverträge diskutiert, deren Bewertung auf mehrdimensionale freie Randwertprobleme führt. Dazu gehören Amerikanische Basket-Optionen, Amerikanische Optionen mit stochastischer Volatilität und pfadabhängige Amerikanische Optionen.

Die numerischen Ergebnisse des Finite-Elemente-Verfahrens und der monotonen Mehrgitterverfahren bei ihrer Anwendung auf die Bewertung Europäischer und Amerikanischer Optionen sind in *Kapitel 6* für Ansatzfunktionen verschiedener Glattheit zusammengefasst.

Es zeigt sich, dass direktes Ableiten genügend glatter Basisfunktionen nach Abschnitt 2.5.2 zu erheblich genaueren Approximationen an die Ableitungen der Lösung und damit an die wichtigen Risikokennziffern der Option führt als durch numerische Differentiation von Hutfunktionen. Bei Verwendung von B-Splines der Ordnung k kann auf diese Weise die $(k - 2)$ -te Ableitung mit quadratischer Konvergenz bestimmt werden.

Bei einem Vergleich der Konvergenzverhalten verschiedener Varianten der monotonen Mehrgitterverfahren wird festgestellt, dass die Verfahren mit OGK-optimierten Restriktionsoperatoren generell schneller konvergieren als mit den einfacheren Restriktionsoperatoren aus Abschnitt 4.2 bzw. [Ma]. Dies ist in Übereinstimmung mit den numerischen Ergebnissen aus [Ko1] für den Fall $k = 2$ und spricht für $k > 2$ für die Qualität des OGK-Algorithmus. Bei der abgeschnittenen Version des monotonen Mehrgitterverfahrens werden die bekannt guten Konvergenzraten linearer Mehrgitterverfahren erreicht. Diese Aussagen gelten sowohl für Hut- als auch für glatte B-Spline-Ansatzfunktionen. Weiter wird festgestellt, dass der freie Rand bereits innerhalb weniger Iterationen richtig bestimmt ist und, dass es im Hinblick auf Genauigkeit und Aufwand bereits ausreicht, nur ein bis zwei Glättungsschritte auf jedem Verfeinerungslevel durchzuführen. In Hinblick auf den Einfluss der Ordnung der Ansatzfunktionen stellt sich heraus, dass bei Verwendung glatter Ansatzfunktionen sowohl die Lösung als auch der freie Rand mit weniger Iterationen und geringeren Aufwand bestimmt wird, als es mit Hutfunktionen der Fall ist.

Abschließend werden in *Anhang A* einige für diese Arbeit benötigte mathematische Grundlagen zusammengefasst. Dazu zählen die wichtigsten Eigenschaften von B-Splines, die Definition der Sobolevräume und Existenz- und Eindeutigkeitsaussagen für die Lösungen von Variationsungleichungen. Außerdem werden die Zusammenhänge von Variationsungleichungen mit beschränkten Minimierungs-, mit Hindernis- und mit linear komplementären Problemen erläutert.

Danksagung Mein Interesse an finanzmathematischen Fragestellungen verdanke ich Professorin A. Kunoth, die mich auf diesen Bereich aufmerksam gemacht hat und es mir ermöglicht hat, mich in dieses spannende Gebiet der Angewandten Mathematik einzuarbeiten und in diesem Bereich meine Diplomarbeit zu fertigen. Dafür und für die gute Betreuung und Unterstützung bei der Anfertigung dieser Arbeit bedanke ich mich ganz besonders. Die wesentliche Anregung zu dem Thema dieser Diplomarbeit stammt von Professor M. Griebel. Auch ihm möchte ich an dieser Stelle herzlich danken. Mein Dank gilt weiter Dr. T. Gerstner für viele hilfreiche Kommentare und ein sehr gutes Praktikum über "Computational Finance", bei dem ich viel über die Bewertung von Optionen lernen konnte. Bei Professor R. Krause möchte ich mich schließlich für die freundliche Beantwortung von Fragen im Zusammenhang mit monotonen Mehrgitterverfahren und für die Übernahme des Korreferates bedanken.

Kapitel 1

Optionen und ihre Bewertung

1.1 Grundlagen der Optionsbewertung

In diesem Kapitel sind die Grundlagen der Optionsbewertung zusammengefasst, die im weiteren Verlauf dieser Arbeit benötigt werden. Die wichtigsten Optionsverträge und ihre Anwendungen werden erläutert und aus der zentralen Annahme eines arbitragefreien Marktes werden die grundlegenden Eigenschaften von Optionswerten hergeleitet.

1.1.1 Definitionen und Bezeichnungen

Definition 1.1.1 (Option) *Bei einer Option handelt es sich um einen Finanzvertrag, bei dem der Halter das Recht erwirbt, ein zugrundeliegendes Wertpapier (den Basiswert) zu oder vor einem spezifizierten Zeitpunkt T (dem Fälligkeitszeitpunkt) zu einem festen Preis K (dem Ausübungspreis) zu kaufen oder zu verkaufen.*

Der Basiswert, der Fälligkeitszeitpunkt T und der Ausübungspreis K werden also vor Vertragsabschluss festgelegt. Der aktuelle Wert V eines solchen Finanzvertrages zum Zeitpunkt t variiert je nach dem aktuellen Preis des Basiswertes. Der Preis des Basiswertes ist ein Prozess $S = S(t)$ in der Zeit t . Sind Missverständnisse ausgeschlossen wird das Argument t der Übersichtlichkeit halber häufiger unterdrückt. Anstelle der exakten Bezeichnung $V = V(K, S(t), t, T)$ wird meist die kürzere Bezeichnung $V = V(S, t)$ verwendet. Der Ausübungspreis K und der Fälligkeitszeitpunkt T werden als feste Parameter behandelt.

Definition 1.1.2 (Kauf-/Verkaufsoption) *Eine Option, die das Recht zum Kaufen beinhaltet, heißt Kaufoption oder Call. Handelt es sich um das Recht zum Verkauf, so spricht man von einer Verkaufsoption oder einem Put.*

Definition 1.1.3 (Europäische-/Amerikanische Optionen) *Darf eine Option einmalig während ihrer Laufzeit ausgeübt werden, handelt es sich um eine Amerikanische Option. Darf sie nur zum Fälligkeitszeitpunkt ausgeübt werden, handelt es sich um eine Europäische Option.*

Da der Wert einer Option vom Preis des Basiswertes abhängt, werden Optionen neben anderen Finanzverträgen (z.B. Futures) auch als *Derivate* bezeichnet. Die zugrundeliegenden Basiswerte können von Aktien, Währungen und Rohstoffen, wie Gold oder Öl, bis zu Wetter oder Schneehöhen prinzipiell alles umfassen. Im folgenden wird angenommen, dass der zugrundeliegende Basiswert eine divididendengeschützte Aktie ist, d.h. eine Aktie, für die im Gültigkeitszeitraum der Option keine Dividenden ausgezahlt werden. Alle Ergebnisse lassen sich aber leicht auf allgemeine Basiswerte mit stetiger oder diskreter Dividendenausschüttung verallgemeinern.

Seit seiner Einführung 1973 am Chicago Board of Trade ist der Derivat Handel weltweit stark angewachsen. Allein in Deutschland wurden im letzten Jahr Derivate im Wert von etwa 150 Mrd. Euro umgesetzt (Quelle Dt. Bank). Die meisten an den Börsen gehandelten Optionen sind vom Amerikanischen Typ. Neben den oben definierten *Standard Optionen* gibt es eine Vielzahl weiterer Formen und Varianten von Optionen, sogenannte *Exotische Optionen*. Hier hängt die Auszahlung der Option oft nicht nur vom Preis des zugrundeliegenden Wertpapiers bei Ausübung ab, sondern von Teilen oder vom gesamten Kursverlauf. Beispiele sind *Barrier Optionen* oder *Asiatische Optionen*. Eine spezielle Art Amerikanischer Optionen wird als *Bermuda Option* bezeichnet. Solche Optionen dürfen nur zu bestimmten Zeitpunkten während der Laufzeit der Option vorzeitig ausgeübt werden.

Unter bestimmten Modellannahmen, insbesondere an die Entwicklung des Basiswertes, ist es möglich einen fairen Wert für Optionen anzugeben. Dieser wird im weiteren Verlauf dieser Arbeit als Optionswert oder Optionspreis bezeichnet. Hier spielen numerische Verfahren eine große Rolle, da nur in wenigen Fällen eine geschlossene Formel bekannt ist. Der faire Wert einer Option muss nicht gleich dem Börsenwert des Derivates sein, welcher aus Angebot und Nachfrage und der subjektiven Wertvorstellung von Käufern und Verkäufern entsteht, bietet aber für alle Marktteilnehmer ein wichtiges Hilfsmittel und hat den systematischen Handel mit Finanzderivaten erst ermöglicht. Europäische Optionen lassen sich im Allgemeinen deutlich leichter bewerten als Amerikanische, weil es hier nur einen einzigen möglichen Ausübungszeitpunkt gibt, den Fälligkeitszeitpunkt T .

Die zentrale Annahme zur Bewertung von Finanzverträgen ist der Ausschluss von Arbitragemöglichkeiten.

Definition 1.1.4 (Arbitragemöglichkeit) *Eine Arbitragemöglichkeit ist eine Handelsstrategie, die keine Anfangsinvestition benötigt und mit positiver Wahrscheinlichkeit einen Gewinn ergibt ohne das Risiko eines Verlustes zu beinhalten.*

Definition 1.1.5 (Arbitragefreier Markt) *Ein arbitragefreier Markt bezeichnet einen Finanzmarkt, in dem keine Arbitragemöglichkeiten bestehen.*

Beispiel 1.1.6 Eine Arbitragemöglichkeit liegt z.B. dann vor, wenn eine Aktie an der New Yorker Börse zum Preis von 125 Dollar und an der Frankfurter Börse zum Preis von 100 Euro gehandelt wird und die Umrechnungsrate 1.20 Dollar für einen Euro beträgt. Bei gleichzeitigem Kauf der Aktie in Frankfurt und Verkauf der Aktie in New York ergibt sich bei Vernachlässigung von Transaktionskosten ein risikoloser Gewinn von $125 - 1.2 * 100 = 5$ Dollar.

Der Ausschluss von Arbitrage impliziert sofort das *Gesetz des einheitlichen Preises*. Dieses besagt, dass Finanzverträge, die dieselbe Auszahlung ergeben, denselben Preis besitzen müssen. Der faire Preis einer Option ist also unabhängig von der Risikoneigung der Akteure eindeutig bestimmbar.

Im Zusammenhang mit Optionen werden im gesamten Verlauf dieser Arbeit die Bezeichnungen aus Tabelle 1.1 verwendet. Ist die Unterscheidung zwischen Kauf- und Verkaufsoption bzw. zwischen Europäischen und Amerikanischen Optionen nicht aus dem Kontext möglich, geschieht dies durch untere Indizes c und p bzw. obere Indizes eur und am an den entsprechenden Bezeichnungen. So bezeichnet V_p^{am} den Preis einer Amerikanischer Verkaufsoption.

Außerdem wird für $a \in \mathbb{R}$ die Abkürzung

$$a^+ := \max\{a, 0\}$$

verwendet.

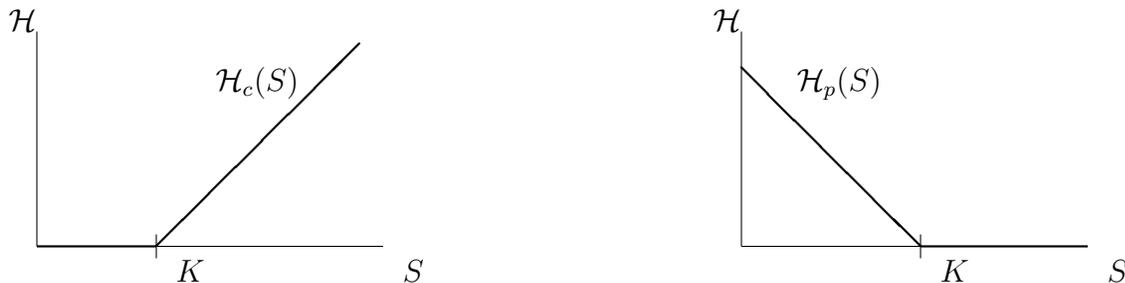
$T \in \mathbb{R}^+$	Fälligkeitszeitpunkt der Option
$S(t) \in \mathbb{R}^+$	Preis des Basiswertes zum Zeitpunkt t
$K \in \mathbb{R}^+$	Ausübungspreis der Option
$V(S, t) \in \mathbb{R}^+$	Wert der Option mit Fälligkeit T bei aktuellem Aktienkurs $S = S(t)$

Tabelle 1.1: Bezeichnungen im Zusammenhang mit Optionen.

Der Wert einer Option am Fälligkeitszeitpunkt T lässt sich sofort angeben. Wenn der Preis des Basiswertes $S = S(T)$ zu diesem Zeitpunkt größer als der Ausübungspreis K ist, wird der Halter einer Kaufoption sein mit der Option verbundenes Recht wahrnehmen, den Basiswert zum Ausübungspreis K zu kaufen und nach sofortigem Verkauf zum Preis S (bei Vernachlässigung von Transaktionskosten) den Gewinn $S - K$ realisieren. Ist der Preis des Basiswertes S dagegen kleiner als der Ausübungspreis K , wird er von seinem Recht kein Gebrauch machen und die Option wertlos verfallen lassen. Analog lässt sich für den Halter einer Verkaufsoption argumentieren, so dass man zum Zeitpunkt T erhält

$$(1.1) \quad \begin{aligned} V_c(S, T) &= (S - K)^+ =: \mathcal{H}_c(S) \quad \text{bei einer Kaufoption,} \\ V_p(S, T) &= (K - S)^+ =: \mathcal{H}_p(S) \quad \text{bei einer Verkaufsoption.} \end{aligned}$$

Die Auszahlung einer Option am Fälligkeitszeitpunkt T in Abhängigkeit vom Basiswert S wird als *Auszahlungsfunktion* $\mathcal{H} = \mathcal{H}(S)$ der Option bezeichnet. Je nachdem, ob es sich um eine Kauf- oder eine Verkaufsoption handelt, gilt $\mathcal{H} = \mathcal{H}_c$ bzw. $\mathcal{H} = \mathcal{H}_p$ (vgl. Abbildung 1.1). Zum Zeitpunkt $t < T$ definiert der Wert der Auszahlungsfunktion \mathcal{H} den sogenannten *inneren Wert* der Option.

Abbildung 1.1: Auszahlungsfunktionen einer Kaufoption (links) und einer Verkaufsoption (rechts) mit Ausübungspreis K .

1.1.2 Motive für den Kauf von Optionen

Optionen werden von Anlegern sowohl zur Spekulation als auch zur Absicherung (*Hedging*) von bereits eingegangenen Positionen gegenüber zukünftigen Entwicklungen eingesetzt. Bei relativ geringer Anfangsinvestition ermöglichen sie bei günstiger Entwicklung des Basiswertes aufgrund der sogenannten Hebelwirkung der Verträge (*Leverage Effekt*) sehr hohe Gewinne. Bei ungünstiger Entwicklung droht dagegen schnell der gesamte Verlust der Anfangsinvestition.

Beispiel 1.1.7 (Spekulation) Eine Europäische Kaufoption kostet $V(S, t) = 39$ zum Zeitpunkt t . Ihr liegt eine Aktie mit Kurs $S(t) = 666$ zugrunde. Der Ausübungspreis beträgt $K = 680$.

Ein Investor erwartet steigende Aktienkurse. Angenommen die Erwartung tritt ein und zum Ausübungszeitpunkt T beträgt der Aktienkurs $S(T) = 730$. Dann liefert die Kaufoption die Auszahlung

$$V(S, T) = (S(T) - K)^+ = 50,$$

so dass die Investition in die Kaufoption mit einer Rendite von

$$\frac{V(S, T) - V(S, t)}{V(S, t)} = 28\%$$

verbunden ist. Bei Kauf der Aktie hätte die Rendite dagegen nur

$$\frac{S(T) - S(t)}{S(t)} = 9.6\%$$

betragen. Tritt die subjektive Erwartung des Investors allerdings nicht ein, d.h. ist $S(T) < K$, dann verliert er seine gesamte Anfangsinvestition $V(S, t) = 39$.

Beispiel 1.1.8 (Hedging) Ein Investor besitzt eine Aktie mit Wert $S(t)$ zum Zeitpunkt t . Er möchte sich gegen Aktienkurse $S(T) < K$ absichern. Dies ist mit der Investition in eine Verkaufsoption möglich. Der Wert $P = P(S, V, t)$ eines Portfolios bestehend aus einer Aktie mit Wert $S(t)$ und einer Verkaufsoption mit Ausübungspreis K , Fälligkeitszeitpunkt T und Wert $V(S, t)$ liegt zum Zeitpunkt $t = T$ mit

$$P(S, V, T) = S(T) + V(S, T) = S(T) + (K - S(T))^+ = \max\{S(T), K\}$$

in jedem Fall über der unteren Schranke K . Ein solches Portfolio wird als *Protective Put* und die Kosten $V(S, t)$ als *Hedgекosten* bezeichnet.

1.1.3 Strukturaussagen

Die Möglichkeit eine Amerikanische Option jederzeit ausüben zu können, impliziert

$$(1.2) \quad V^{\text{am}}(S, t) \geq \mathcal{H}(S)$$

für alle $S \geq 0$ und alle $0 \leq t < T$. Der Wert einer Amerikanischen Option V ist also immer mindestens so hoch wie der Wert der Auszahlungsfunktion \mathcal{H} . Würde dies nicht gelten, könnte ein Händler die Option kaufen und durch sofortige Ausübung einen risikolosen Gewinn $\mathcal{H}(S) - V^{\text{am}}(S, t) > 0$ verbuchen. Dies widerspricht aber der Annahme, dass der Markt keine Arbitragemöglichkeiten zulässt.

Unter der zentralen Annahme eines arbitragefreien Marktes lassen sich weiter ohne sonstige Modellannahmen, z.B. an die Aktienkursdynamik, folgende a priori Schranken für den Optionswert angeben (s. [Se], Anhang A.7)

$$\begin{aligned} (S(t) - Ke^{-r(T-t)})^+ &\leq V_c^{\text{eur}}(S, t) \leq S(t) \\ (Ke^{-r(T-t)} - S(t))^+ &\leq V_p^{\text{eur}}(S, t) \leq Ke^{-r(T-t)} \\ (S(t) - K)^+ &\leq V_c^{\text{am}}(S, t) \leq S(t) \\ (K - S(t))^+ &\leq V_p^{\text{am}}(S, t) \leq K. \end{aligned}$$

Hierbei bezeichnet $r \in \mathbb{R}^+$ den konstanten Jahreszinssatz für eine sichere Anlage. Da es sich bei einer Option um ein Recht handelt, gilt insbesondere immer $V(S, t) \geq 0$. Da Amerikanische Optionen im Vergleich zu Europäischen das zusätzliche Recht beinhalten, dass die Option auch vor dem Fälligkeitszeitpunkt ausübt werden darf, gilt

$$V^{\text{eur}}(S, t) \leq V^{\text{am}}(S, t)$$

sowohl für Kauf- als auch für Verkaufsoptionen für alle $S > 0$ und alle $t \in [0, T]$.

Genauer lässt sich zeigen, dass der Preis einer Amerikanischen Kaufoption mit zugrundeliegenden dividendengeschützten Wertpapier immer gleich dem Preis einer entsprechenden Europäischen Option ist.

Satz 1.1.9 (Merton) *Zahlt das zugrundeliegende Wertpapier keine Dividenden aus, so ist es niemals optimal eine Amerikanische Kaufoption vor dem Fälligkeitszeitpunkt auszuüben und es gilt*

$$V^{\text{eur}}(S, t) = V^{\text{am}}(S, t)$$

sowohl für Kauf- als auch für Verkaufsoptionen für alle $S > 0$ und alle $t \in [0, T]$.

Beweisskizze: Wegen den obigen Strukturaussagen $V^{\text{am}}(S, t) \geq V^{\text{eur}}(S, t)$ und $V_c^{\text{eur}}(S, t) \geq (S - Ke^{-r(T-t)})^+ \geq (S - K)^+$ lohnt sich vorzeitiges Ausüben der Option nicht. Ausführlich findet sich dieses Argument in [Hu], S.175. \square

Wegen Satz 1.1.9 beschränkt sich die Betrachtung im weiteren auf Amerikanische Verkaufsoptionen. In diesem Fall kann vorzeitiges Ausüben optimal sein.

Satz 1.1.10 *Für eine Amerikanische Verkaufsoption gibt es immer einen Aktienkurs $S(t) > 0$, so dass vorzeitiges Ausüben zum Zeitpunkt $t < T$ optimal ist.*

Beweisskizze: Die maximale Auszahlung einer Amerikanischen Verkaufsoption am Fälligkeitszeitpunkt T ist K . Ein Händler wird eine Amerikanischen Verkaufsoption zum Zeitpunkt t vorzeitig ausüben, wenn die Auszahlung größer ist als die maximale diskontierte Auszahlung am Fälligkeitszeitpunkt T , d.h. wenn

$$K - S(t) > e^{-r(T-t)} K.$$

Hinreichend für ein vorzeitiges Ausüben der Option ist also $S(t) < (1 - e^{-r(T-t)}) K$. Dies ist leicht variiert der Beweis aus [Sa], Seite 45, Satz 2.3. \square

Bei Europäischen Optionen reicht es aus entweder nur Kauf- oder nur Verkaufsoptionen zu betrachten. Bei gegebenem Preis einer Europäischen Kaufoption erhält man den Preis einer entsprechenden Verkaufsoption sofort mit der sogenannten *Put-Call-Parität*.

Satz 1.1.11 (Put-Call-Parität) *Für eine Europäischen Verkaufsoption und eine Europäische Kaufoption auf ein dividendengeschütztes Wertpapier mit demselben Ausübungspreis K und gleicher Fälligkeit T gilt zum Zeitpunkt t*

$$(1.3) \quad S(t) + V_p^{\text{eur}}(S, t) - V_c^{\text{eur}}(S, t) = K e^{-r(T-t)}.$$

Beweisskizze: Aus (1.1) folgt $V_p^{\text{eur}}(S, T) - V_c^{\text{eur}}(S, T) = K - S(T)$. Damit impliziert das Gesetz des einheitlichen Preises $V_p^{\text{eur}}(S, t) - V_c^{\text{eur}}(S, t) = K e^{-r(T-t)} - S(t)$. Dieser Beweis findet sich in [Hu] auf Seite 174. \square

Für Amerikanische Optionen lässt sich dagegen nur

$$K e^{-r(T-t)} \leq S(t) + V_p^{\text{am}}(S, t) - V_c^{\text{am}}(S, t) \leq K$$

festhalten (s. [Se], S.210). Die obere Schranke gilt dabei nur, falls keine Dividenden gezahlt werden.

1.2 Das Black–Scholes Modell

1.2.1 Einleitung und Annahmen

Das 1973 von F. Black, R. Merton und M. Scholes entwickelte *Black–Scholes Modell* aus [BS] begründet die wesentlichen und grundlegenden Ideen des Financial Engineering und hat nachhaltig die Bewertung von Optionsverträgen im internationalen Finanzhandel beeinflusst. 1997 haben R. Merton und M. Scholes für ihr Modell den Nobelpreis für Wirtschaftswissenschaften erhalten. Eine gute Übereinstimmung mit tatsächlich beobachteten Aktienkursen und Optionspreisen wurde mehrfach empirisch bestätigt (siehe z.B. [Hu] S.448).

Die folgenden Annahmen werden in ihrem Modell und für den gesamten weiteren Verlauf dieser Arbeit getroffen.

1. Es handelt sich um einen *reibungsfreien Finanzmarkt*, d.h.
 - es werden keine Transaktionskosten und Steuern berücksichtigt,
 - der Aktien- bzw. Optionshandel ist zu jedem Zeitpunkt $t \in [0, T]$ möglich,
 - gewünschte Transaktionen können in beliebigem Umfang ohne Rückwirkung auf die Kursentwicklung durchgeführt werden,
 - Wertpapiere stehen in beliebig teilbaren Einheiten zur Verfügung.
2. Für die Aktien werden keine Dividenden gezahlt und der Zinssatz r für eine sichere Anlage ist konstant über die Zeit.
3. Der Markt lässt keine Arbitragemöglichkeiten zu.
4. Die Aktienkursdynamik $S(t)$ ist ein Prozess mit stetigen Pfaden, der durch die stochastische Differentialgleichung

$$(1.4) \quad dS(t) = \mu S(t) dt + \sigma S(t) dB(t)$$

mit konstanten Parametern $\mu \in \mathbb{R}$ und $\sigma \in \mathbb{R}^+$ gegeben ist. Dabei bezeichnet $B(t)$ die eindimensionale standardisierte Brownsche Bewegung. Eine Definition der Brownschen Bewegung findet sich z.B. in [Ba]. Der Prozess $S(t)$ wird als *geometrische Brownsche Bewegung* bezeichnet. Der Parameter μ wird als *Drift*, der Parameter σ als die *Volatilität* des Aktienkursprozesses bezeichnet.

Die beiden zentralen Annahmen des Black–Scholes Modell sind hierbei die letzten beiden Punkte. Verallgemeinerungen auf Aktien mit Dividendenzahlungen (stetig oder diskret) oder auf zeitabhängige Drift und Zinsraten sind leicht möglich. Ebenso gibt es zahlreiche Erweiterungen, die Transaktionskosten berücksichtigen oder stochastische Zinsraten bzw. Volatilitäten modellieren (s. z.B. [Hu, BR, Si]). Diese führen aber im Allgemeinen zu komplizierteren Modellen. Ebenso gibt es mittlerweile ein Vielzahl an Modellen, die durch eine allgemeinere Modellierung der Aktienkursdynamik (z.B. mit *Sprungprozessen*) versuchen eine verbesserte Übereinstimmung mit tatsächlich beobachteten Daten zu erreichen (s. [Eb, MCC, Ra, Zh2] zu den Modellen und [MPS] zur Numerik).

Mit dem Lemma von Itô lässt sich die Lösung der stochastischen Differentialgleichung aus (1.4) bestimmen. Man erhält

$$(1.5) \quad S(t) = S(0)e^{(\mu - \frac{\sigma^2}{2})t + \sigma B(t)}.$$

Hieran kann man ablesen, dass die normierten, logarithmierten Aktienkurse $\ln(S(t)/S(0))$ normalverteilt sind mit Erwartungswert $(\mu - \frac{\sigma^2}{2})t$ und Varianz $\sigma\sqrt{t}$. Für den Erwartungswert und die Varianz des stochastischen Prozesses $S(t)$ lässt sich dann zeigen, dass

$$(1.6) \quad E[S(t)] = e^{\mu t} S(0) \quad \text{und} \quad \text{Var}[S(t)] = S(0)^2 e^{2\mu t} (e^{\sigma^2 t} - 1).$$

Weiter wird angenommen, dass der Aktienkurs bereits alle relevanten Informationen für die Preisbildung der Option besitzt. Der Kursprozess einer Option kann somit als Funktion der Zeit t und des Aktienkurses S aufgefasst werden und wird folgerichtig wie bisher mit $V = V(S, t)$ bezeichnet.

1.2.2 Europäische Optionen und die Black-Scholes Gleichung

Eine Handelsstrategie heißt *selbstfinanzierend*, falls nach einer Anfangsinvestition weder Geldzuschüsse notwendig noch Geldüberflüsse möglich sind.

In der fundamentalen Arbeit [BS] wird gezeigt, dass es eine dynamische Handelsstrategie bestehend aus Δ Aktien und dem Geldbetrag B gibt, die selbstfinanzierend ist, und die unabhängig von der Entwicklung des Aktienkurses am Zeitpunkt T zu der gleichen Auszahlung führt wie eine Europäische Option. Auf einem arbitragefreien Markt folgt dann, dass der Preis der Option V der Anfangsinvestition $\Delta S(t) + B$ in das Duplizierungsportfolio entsprechen muss, d.h.

$$V(S, t) = \Delta S(t) + B.$$

Auf diese Weise wird gezeigt, dass der Optionspreis für $S \geq 0$ und $0 \leq t < T$ die berühmte **Black-Scholes Gleichung**

$$(1.7) \quad \mathcal{L}V := \frac{\partial}{\partial t} V + \frac{\sigma^2}{2} S^2 \frac{\partial^2}{\partial S^2} V + rS \frac{\partial}{\partial S} V - rV = 0$$

erfüllen muss. Die Black-Scholes Gleichung ist eine rückwärts parabolische partielle Differentialgleichung zweiter Ordnung mit Konvektions-, Diffusions- und Reaktionstermen. Anfangs- und Randbedingungen unterscheiden sich je nach Art der Option. Im Standardfall gilt wie in (1.1) motiviert

$$(1.8) \quad V_c(S, T) = \mathcal{H}_c(S) \quad \text{bzw.} \quad V_p(S, T) = \mathcal{H}_p(S).$$

Am Rand $S = 0$ (im Falle einer Kaufoption) bzw. $S \rightarrow \infty$ (im Falle einer Verkaufsoption) ist die Option wegen (1.1) offensichtlich wertlos, d.h. für alle $t < T$ gilt

$$(1.9) \quad V_c(S, t) = 0 \quad \text{für } S = 0 \quad \text{bzw.} \quad V_p(S, t) = 0 \quad \text{für } S \rightarrow \infty.$$

Über die Put-Call-Parität 1.1.11 erhält man die Randbedingung für $S \rightarrow \infty$ (im Falle einer Kaufoption) bzw. für $S = 0$ (im Falle einer Verkaufsoption)

$$(1.10) \quad V_c(S, t) = S - Ke^{-r(T-t)} \quad \text{für } S \rightarrow \infty \quad \text{bzw.} \quad V_p(S, t) = Ke^{-r(T-t)} - S \quad \text{für } S = 0.$$

Zusammenfassend lässt sich der Wert einer Europäischen Kaufoption durch das folgende Randwertproblem eindeutig charakterisieren.

Satz 1.2.1 (Bewertung Europäischer Kaufoptionen) *Der Wert $V = V(S, t)$ einer Europäischen Kaufoption erfüllt für $S \geq 0$ und $0 \leq t < T$ die Black-Scholes Gleichung*

$$\mathcal{L}V := \frac{\partial}{\partial t} V + \frac{\sigma^2}{2} S^2 \frac{\partial^2}{\partial S^2} V + rS \frac{\partial}{\partial S} V - rV = 0$$

mit den Randdaten

$$V(S, t) = 0 \quad \text{für } S = 0 \quad \text{und} \quad V(S, t) = S - K e^{-r(T-t)} \quad \text{für } S \rightarrow \infty$$

und den Enddaten

$$V(S, T) = (S - K)^+.$$

Eine ausführliche Herleitung der Black–Scholes Gleichung und der Randbedingungen für Europäische Optionen mit Arbitrage–Argumenten findet man z.B. in [Se], Abschnitt 4.4 und Anhang A.3. Man erkennt anhand der Black–Scholes Gleichung, dass der Optionspreis unabhängig von der Drift μ und damit unabhängig von der subjektiven Risikoneigung der Marktteilnehmer ist. Dies wird als Prinzip der *risikoneutralen Bewertung* bezeichnet.

Für die Randwertaufgabe 1.2.1 lässt sich eine analytische Lösungsformel angeben. Eine der vielen Möglichkeiten der Herleitung besteht darin, die Black–Scholes Gleichung, wie in Abschnitt 2.2.2 ausgeführt, in die Wärmeleitungsgleichung

$$\mathcal{W}u := \frac{\partial u}{\partial \tau} - \frac{\partial^2 u}{\partial x^2} = 0$$

zu transformieren. Diese besitzt für gegebene Anfangsdaten $u_0(x)$ bekanntermaßen die eindeutige Lösung

$$u(x, \tau) = \frac{1}{\sqrt{4\pi\tau}} \int_{-\infty}^{\infty} u_0(s) e^{-\frac{(x-s)^2}{4\tau}} ds.$$

Rücktransformation liefert dann die berühmte Black–Scholes Formel für Europäische Kaufoptionen.

Satz 1.2.2 (Black–Scholes Formel) *Der Wert $V = V(S, t)$ einer Europäischen Kaufoption ist gegeben durch*

$$(1.11) \quad V(S, t) = S N(d_1) - K e^{-r(T-t)} N(d_2).$$

Dabei bezeichnet

$$N(x) := \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{y^2}{2}} dy$$

die Verteilungsfunktion der Standardnormalverteilung und $d_1, d_2 \in \mathbb{R}$ sind gegeben durch

$$d_1 := \frac{\ln(S/K) + (r + \frac{1}{2}\sigma^2)(T-t)}{\sigma\sqrt{T-t}}$$

und

$$d_2 := d_1 - \sigma\sqrt{T-t}.$$

Einen ausführlichen Beweis findet man z.B. in [WHD2], Abschnitt 5.3. In obiger Form gilt die Formel nur für Kaufoptionen. Der Wert einer Europäischen Verkaufsoptionen lässt sich jedoch mit Hilfe der Put–Call–Parität 1.1.11 direkt aus dem Wert einer entsprechenden Kaufoption bestimmen.

1.2.3 Die Bewertung Amerikanischer Optionen als freies Randwertproblem

Im Unterschied zu Europäischen Optionen darf der Halter einer Amerikanischen Option die Option zu einem beliebigen Zeitpunkt $t < T$ ausüben. Da a priori der Zeitpunkt, an dem die Option ausgeübt wird, unbekannt ist, ist die Bewertung Amerikanischer Optionen schwieriger als die von Europäischen. Sie führt auf ein instationäres freies Randwertproblem zur Black-Scholes Gleichung mit nicht-glatten Anfangsdaten. Die wichtigsten Ergebnisse sollen im folgenden kurz zusammengefasst werden. Wegen Satz 1.1.9 werden zunächst Amerikanische Verkaufsoptionen betrachtet.

Nach Satz 1.1.10 gilt für beliebiges $t < T$, dass es einen Aktienkurs S gibt, so dass direktes Ausüben der Option mit der Auszahlung $K - S > 0$ optimal ist. Ist es für einen Aktienkurs S optimal die Option auszuüben, so muss dies auch für alle Kurse $\tilde{S} < S$ gelten, da die Auszahlung umso höher ist, je kleiner S ist. Damit gibt es für alle $t < T$ einen eindeutigen Aktienkurs $S_f(t)$, so dass für alle $S \leq S_f(t)$ Ausüben der Option optimal ist, wohingegen es für alle $S > S_f(t)$ günstiger ist, die Option zu halten und auf einen fallenden Aktienkurs und damit eine höhere Auszahlung zu hoffen. Der Wert $S_f(t)$ wird als *optimaler Ausübungspreis* der Option bezeichnet. Er definiert einen *freien Rand* S_f , der den Raum $\Omega_{\mathcal{L}} := \mathbb{R}^+ \times [0, T]$ in zwei Gebiete $\Omega_{\mathcal{L},1} := \{(S, t) \in \Omega_{\mathcal{L}} : S \leq S_f(t)\}$ und $\Omega_{\mathcal{L},2} := \{(S, t) \in \Omega_{\mathcal{L}} : S > S_f(t)\}$ aufteilt. Der freie Rand ist zeitabhängig und a priori unbekannt. Wegen Satz 1.1.10 gilt

$$(1.12) \quad 0 < S_f(t) < K.$$

In $\Omega_{\mathcal{L},1}$ beträgt der Optionswert $V(S, t) = K - S$ und erfüllt nicht mehr die Black-Scholes Gleichung. Einsetzen liefert

$$(1.13) \quad \mathcal{L}V := \frac{\partial V}{\partial t} + \frac{1}{2}\sigma^2 S^2 \frac{\partial^2 V}{\partial S^2} + rS \frac{\partial V}{\partial S} - rV = -rK < 0.$$

In $\Omega_{\mathcal{L},2}$ gilt dagegen $V(S, T) > \mathcal{H}(S)$ und der Optionswert erfüllt die Black-Scholes Gleichung.

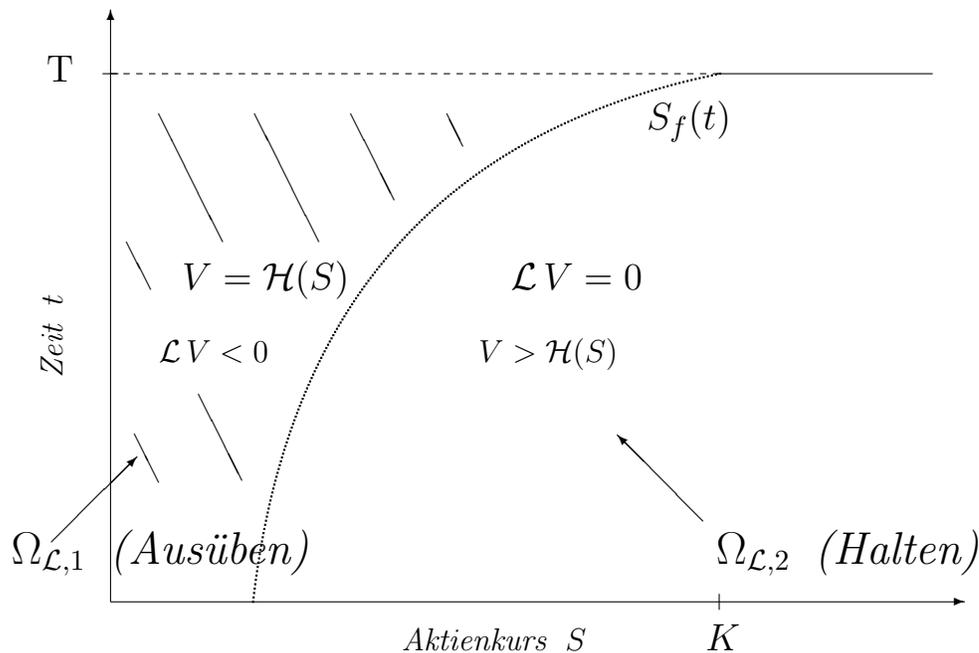


Abbildung 1.2: Die Bewertung Amerikanischer Optionen als freies Randwertproblem

Damit das freie Randwertproblem eindeutig lösbar ist, werden noch Randdaten am freien Rand benötigt. Da für $S = S_f$ Ausüben optimal ist, gilt

$$(1.14) \quad V(S_f, t) = (K - S_f)^+.$$

Einfache Arbitrageargumente implizieren außerdem (s. [WHD2], Abschnitt 7.4), dass der Optionspreis $V(S, t)$ einer Amerikanischen Verkaufsoption an $S = S_f$ stetig nach S differenzierbar ist und, dass dort

$$(1.15) \quad \frac{\partial V}{\partial S}(S_f, t) = -1$$

gilt. Weiter lässt sich zeigen (s. [Kw]), dass der freie Rand $S_f(t)$ eine stetige, monoton wachsende Funktion der Zeit ist und, dass für $t \rightarrow T$, d.h. gegen Ablauf der Option

$$(1.16) \quad \lim_{t \rightarrow T} S_f(t) = K$$

gilt. Zusammenfassend erhält man folgendes freies Randwertproblem, welches den Wert einer Amerikanischer Verkaufsoption eindeutig bestimmt.

Satz 1.2.3 (Bewertung Amerikanischer Verkaufsoptionen) *Der Wert $V = V(S, t)$ einer Amerikanischen Verkaufsoption erfüllt*

$$\mathcal{L}V := \frac{\partial V}{\partial t} + \frac{1}{2}\sigma^2 S^2 \frac{\partial^2 V}{\partial S^2} + rS \frac{\partial V}{\partial S} - rV = 0 \quad \text{für } S > S_f \text{ und } 0 \leq t < T,$$

$$V(S, t) = (K - S)^+ \quad \text{für } S \leq S_f \text{ und } 0 \leq t < T$$

mit den Randdaten

$$V(S, t) = 0 \quad \text{für } S \rightarrow \infty \text{ und } 0 \leq t < T,$$

den Enddaten

$$V(S, T) = (K - S)^+ \quad \text{für } S \geq 0$$

und den Bedingungen, dass V und $\frac{\partial V}{\partial S}$ an $S = S_f(t)$ stetig sind.

Da es im folgenden noch gebraucht wird, sei außerdem daran erinnert, dass wegen (1.2) und (1.13) für alle $S > 0$ gilt

$$(1.17) \quad \begin{aligned} \mathcal{L}V &\leq 0, \\ V(S, t) &\geq (K - S)^+. \end{aligned}$$

Ähnliche Überlegungen gelten für Amerikanische Kaufoptionen mit Dividenden und für Amerikanische Optionen mit allgemeineren, sogar pfadabhängigen Auszahlungsfunktionen (s. [WHD2], Abschnitt 7.5). Für Amerikanische Kaufoptionen mit stetiger Dividende $d > 0$ erhält man

$$V(S_f, t) = (S_f - K)^+ \quad \text{und} \quad \frac{\partial V}{\partial S}(S_f, t) = 1.$$

Der freie Rand ist in diesem Fall eine stetige, monoton fallende Funktion der Zeit und für $t \rightarrow T$ gilt

$$\lim_{t \rightarrow T} S_f(t) = \begin{cases} \frac{r}{d}K & \text{falls } d < r \\ K & \text{falls } d \geq r. \end{cases}$$

Bemerkung 1.2.4 Es bezeichne $H^2(\mathbb{R}^+)$ den in Anhang A.2 definierten Sobolevraum der zweimal schwach differenzierbaren Funktionen $v \in L^2(\mathbb{R}^+)$ und V die Lösung des freien Randwertproblems aus Satz 1.2.3. Dann gilt $V(\cdot, t) \in H^2(\mathbb{R}^+)$ für alle $t \in [0, T)$, und V besitzt eine einfache Sprungstelle in seiner zweiten Ableitung an freien Randpunkten.

Beweis: Nach (1.15) ist V einmal stetig differenzierbar nach S in \mathbb{R}^+ und jeweils beliebig oft differenzierbar in $\Omega_1 := [0, S_f)$ bzw. $\Omega_2 := [S_f, \infty)$. Damit folgt die Behauptung aus Satz A.2.3 mit $k = 2$. Für den Nachweis der zweiten Behauptung sei auf [BHR] verwiesen. \square

Im Gegensatz zur Bewertung Europäischer Optionen ist für die Bewertung Amerikanischer Optionen keine geschlossene Lösungsformel bekannt. Hier muss auf numerische Verfahren zurückgegriffen werden.

1.3 Risikokennziffern – The Greek Letters

Der Handel mit Optionen ermöglicht hohe Gewinne, beinhaltet aber auch ein hohes Verlustrisiko. Dem Verkäufer einer Kaufoption drohen bei steigendem Preis des Basiswertes sogar unbegrenzte Verluste. In solchen Fällen ist es unabdingbar sich gegen diese Risiken abzusichern. Da der Wert einer Option durch den gleichen stochastischen Prozess bestimmt ist wie der Wert des Basiswertes, kann man den Verkauf einer Option vollständig durch den Kauf einer bestimmten Menge an Basiswerten absichern.

Als Verkäufer einer Europäischen Kaufoption zum Zeitpunkt $t = 0$ könnte dies z.B. durch den Kauf einer Einheit des zugrundeliegenden Wertpapiers mit dem Wert $S(0)$ geschehen. Liegt der Aktienkurs zum Ausübungszeitpunkt T über den Ausübungspreis K ist man zwar zur Auszahlung $S(T) - K$ verpflichtet, verbucht aber auf der anderen Seite durch den Besitz der Aktie den Gewinn $S(T) - S(0)$ und begrenzt so sein Verlustrisiko. Wird die Option vom Käufer dagegen nicht ausgeübt, da der Aktienkurs zum Zeitpunkt T unter dem Ausübungspreis K liegt, ist diese Strategie mit den im Vergleich zum Optionspreis $V(S(0), 0)$ hohen Kosten $S(0)$ für den Kauf der Aktie verbunden.

Verwendet man anstelle dieser Art der statischen Absicherung eine dynamische Absicherungsstrategie, so lassen sich erheblich Kosten einsparen. In diesem Fall wird das zur Absicherung verwendete Portfolio an Basiswerten regelmäßig an die Aktienkursentwicklung angepasst. Die in dem Portfolio zu haltende Anzahl an Aktien beträgt dann (s. die Herleitung der Black-Scholes Gleichung in [Hu], Abschnitt 11.5)

$$\text{Delta} := \frac{\partial V}{\partial S}.$$

Zur Risikominimierung wird in einem Portfolio von Aktienoptionen eine delta-neutrale Position angestrebt. Dies bedeutet, dass das Portfolio möglichst so ausbalanciert wird, dass $\partial V / \partial S = 0$ gilt. Da sich Delta aber stetig ändert, muss das Portfolio regelmäßig an die neuen Werte angepasst werden. In der Praxis ist eine stetige Portfolioanpassung aber unmöglich und jede Portfolioänderung ist in der Regel mit Transaktionskosten verbunden. Aus diesem Grund muss bei den Abständen, in dem das Portfolio neu ausbalanciert wird, ein Kompromiss zwischen hoher Sicherheit und geringen Kosten gefunden werden. Hierfür wird

$$\text{Gamma} := \frac{\partial^2 V}{\partial S^2}$$

als Maß verwendet. Ein kleiner Gamma Wert ist mit einem geringem Delta, also einer geringen Sensitivität des Optionspreis in Hinblick auf Aktienkursänderungen verbunden. Die Abstände, in denen das Portfolio an die Aktienkursdynamik angepasst wird, können also verhältnismäßig groß gewählt werden. Solange der Betrag von Gamma dagegen groß ist, haben Aktienkursänderungen

einen starken Einfluss auf den Optionswert und es bietet sich an, die Abstände verhältnismäßig klein zu wählen. Neben Delta und Gamma gibt es weitere in der Praxis zur Risikoanalyse verwendete Risikokennziffern. Erwähnt sei lediglich noch

$$\text{Theta} := \frac{\partial V}{\partial t},$$

welches zur Näherung für Gamma verwendet werden kann.

Unter den Annahmen des Black-Scholes Modell erfüllt der Wert $V(S, t)$ einer Europäischen Optionen die Black-Scholes Gleichung aus (1.7). Daher muss die Beziehung

$$\text{Theta} + \frac{1}{2}\sigma^2 S^2 \text{Gamma} + rS \text{Delta} = rV$$

zwischen Delta, Gamma und Theta gelten. Außerdem können die Ableitungen des Optionswertes, d.h. die Risikokennziffern Delta, Gamma und Theta durch partielles Ableiten der Black-Scholes Formel direkt angegeben werden. Für eine Kaufoption erhält man (s. [Sa], Abschnitt 7.3.) mit der Verteilungsfunktion der Standardnormalverteilung N und den in (1.11) definierten $d_1 = d_1(S, t)$ und $d_2 = d_2(S, t)$ die Identitäten

$$\begin{aligned} \text{Delta} &:= \frac{\partial V(S, t)}{\partial S} = N(d_1), \\ \text{Gamma} &:= \frac{\partial^2 V(S, t)}{\partial S^2} = \frac{N'(d_1)}{S \sigma \sqrt{T-t}}, \\ \text{Theta} &:= \frac{\partial V(S, t)}{\partial t} = -Ke^{-r(T-t)} \left[\frac{\sigma N'(d_2)}{2\sqrt{T-t}} + r N(d_2) \right]. \end{aligned}$$

Da $0 \leq N(x) \leq 1$ und $N'(x) > 0$ für alle $x \in \mathbb{R}$ ist, folgen hieraus sofort die Abschätzungen

$$(1.18) \quad \begin{array}{lll} 0 \leq \text{Delta} \leq 1 & (\text{Der Optionspreis ist monoton steigend in } S), \\ 0 < \text{Gamma} & (\text{Der Optionspreis ist konvex in } S), \\ 0 > \text{Theta} & (\text{Der Optionspreis ist monoton fallend in } t). \end{array}$$

Diese Aussagen sind allein durch den Ausschluss von Arbitrage auch ohne Modellannahmen an die Aktienkursdynamik und auch für Amerikanische Optionen gültig.

Bei der Bestimmung der Risikokennziffern muss für Amerikanische Optionen allerdings auf numerische Verfahren zurückgegriffen werden.

1.4 Verfahren zur Bewertung von Amerikanischen Optionen

Da die Bewertung Amerikanischer Optionen weiterhin Gegenstand aktueller und aktiver Forschung ist (s. [Bl], [BR], [BjS], [CP], [DH], [FV], [HW], [JLL], [Me], [Oo], [Zh2]) und aufgrund der Vielzahl und Vielfalt unterschiedlicher Optionsformen und Modellierungsansätzen, ist eine vollständige Übersicht über Verfahren zur Bewertung von Amerikanischen Optionen nicht möglich. Die Übersicht in diesem Kapitel ist zum Teil den Büchern [Hu] und [Se] entnommen und bleibt auf die Modellierung des Black–Scholes Modell beschränkt.

Im Gegensatz zu Europäischen Optionen existieren für Amerikanische Optionen bis auf ganz wenige Ausnahmefälle keine geschlossenen Bewertungsformeln, so dass bei der Bewertung auf numerische Verfahren oder analytische Approximationen zurückgegriffen werden muss. In [GS] wird in einem Vergleich mehrerer numerischer Verfahren geschlossen, dass je nach der speziellen Struktur der Option jedes Verfahren ihre Vorteile und Nachteile hat und keines als allgemein bestes angesehen werden kann.

Bei der Optionsbewertung lassen sich prinzipiell zwei Ansätze unterscheiden. Im Martingal–Ansatz wird der Optionswert als diskontierter Erwartungswert der Auszahlung bezüglich eines bestimmten Maßes, dem sogenannten äquivalenten Martingalmaß dargestellt. Dies führt auf ein Integrationsproblem, welches z.B. mit Monte Carlo Verfahren gelöst werden kann. Wenn dies auch für viele Europäische Optionstypen, wie pfadabhängige Optionen oder Basket–Optionen ein guter Ansatz ist, so ist er nur bedingt geeignet, die Möglichkeit vorzeitiger Ausübung zu erfassen, die für Amerikanische Optionen charakteristisch ist. Hier bietet es sich eher an, den Wert der Option als Lösung einer partiellen Differentialgleichung darzustellen, indem in geeigneter Weise alle stochastischen Terme eliminiert werden. Die Differentialgleichung kann dann mit Finite–Differenzen oder Finite–Elemente–Verfahren gelöst werden. Auch Baumverfahren lassen sich in diese Klasse einordnen. Den Zusammenhang zwischen dem Integral–Ansatz und dem PDE–Ansatz liefert die Feynman–Kac Formel (s. [Du], S.237).

Für Amerikanische Option mit unendlicher Laufzeit ist eine analytische Lösung bekannt (s. [Mc]). Solche sogenannten Perpetual Optionen sind aber höchstens von theoretischem Interesse. Für den Wert einer Amerikanischer Kaufoptionen zum Zeitpunkt $t = 0$ mit Fälligkeitszeitpunkt T und n Dividendenzahlungen zu den Zeitpunkten $0 < t_1, \dots, t_n < T$ ist in [Bk] die Approximationen

$$V_{am}(0, T) \approx \max\{V_{eur}(0, T), V_{eur}(0, t_n)\}$$

vorgeschlagen. Eine exaktere Prozedur stammt aus [Ro], [Ge] und [Wh]. In diesen Arbeiten wird ausgenutzt, dass es höchstens an den Zeitpunkten t_i , $i = 1, \dots, n$ optimal sein kann, eine Amerikanische Kaufoptionen mit diskreter Dividende vorzeitig auszuüben. Dies ist eine verallgemeinerte Aussage von Satz 1.1.9. In den meisten Fällen ist ein vorzeitiges Ausüben sogar nur zum Zeitpunkt t_n optimal. In diesem Fall liefert die Formel aus [Ro], [Ge] und [Wh] sogar die exakte Lösung. Dies gilt aber generell nicht für Amerikanische Verkaufsoptionen oder für Amerikanische Kaufoptionen mit stetiger Dividende.

1.4.1 Baumverfahren

Baumverfahren simulieren zunächst in einem Vorwärtsschritt die zukünftige Aktienkursentwicklung bis zum Fälligkeitszeitpunkt T und rechnen hieraus rekursiv bis zum Optionswert am Zeitpunkt $t = 0$ zurück. Sie können als Diskretisierung des Black–Scholes Modell aufgefasst werden. Das einfachste und in der Praxis verbreitetste Baumverfahren ist das Binomialverfahren. Allgemeiner sind Multinomialverfahren und das Verfahren von Broad–Glasserman. Die Verfahren werden häufig in Verbindung mit der sogenannten Control Variate Technik eingesetzt. In vielen Fällen führt dies zu einer signifikanten Fehlerreduktion.

Binomialverfahren

Das 1979 ursprünglich zur anschaulichen Darstellung und Interpretation des Black–Scholes Modells aus [BS] eingeführte Binomialmodell von J.Cox, S.Ross und M.Rubinstein [CRR] hat schnell praktische Bedeutung dadurch erlangt, dass es einfach und intuitiv implementierbar und auf viele Exotische Optionen erweiterbar ist. Die Konvergenzeigenschaften sind aus numerischer Sicht allerdings nicht ausreichend. Die Konvergenz ist langsam und nicht monoton.

Das Binomialverfahren ist ausführlich in den Büchern [Se], Abschnitt 1.4, [Hu], Kapitel 9 und Abschnitt 16.1 und von [Sa], Kapitel 5 beschrieben. Das Prinzip und einige Erweiterungen werden im folgendem kurz wiedergegeben. Das Binomialverfahren wurde zum Testen des implementierten Finite–Elemente–Verfahrens verwendet.

Ziel des Verfahrens ist die Berechnung eines fairen Optionswert $V(S_0, 0)$ zu einem vorgegebenen Aktienkurs $S_0 := S(0)$ zum Zeitpunkt $t = 0$. Dafür wird das Zeitintervall $[0, T]$ in M Zeitintervalle $[t_i, t_{i+1}]$ der Länge Δt unterteilt.

Es wird angenommen, dass sich der Aktienkurs S in jedem Zeitintervall entweder mit der Wahrscheinlichkeit p um den Faktor $u > 1$ nach oben oder mit der Wahrscheinlichkeit $1 - p$ um den Faktor $1 > d > 0$ nach unten bewegt. Bei gegebenem Aktienkurs $S_i := S(t_i)$ beträgt der Aktienkurs zum Zeitpunkt t_{i+1} also

$$S_{i+1} = \begin{cases} u S_i & \text{mit Wahrscheinlichkeit } p \\ d S_i & \text{mit Wahrscheinlichkeit } 1 - p. \end{cases}$$

Weiter wird angenommen, dass für den Erwartungswert von S_{i+1} bei gegebenem Aktienkurs S_i gilt

$$(1.19) \quad E[S_{i+1}] = e^{r\Delta t} S_i.$$

Dabei bezeichnet r wie bisher den konstanten Jahreszinssatz für eine sichere Anlage. Damit entsprechen die Erwartungswerte des diskreten Binomialmodells den risikoneutralen Erwartungswerten aus (1.6) des kontinuierlichen Black–Scholes Modells. Wegen $E[S_{i+1}] = p S_i u + (1-p) S_i d$ folgt aus (1.19), dass

$$p = \frac{e^{r\Delta t} - d}{u - d}.$$

Außerdem bedeutet (1.19), dass der diskontierte Preisprozess $e^{-r\Delta t} S_i$ unter dem durch die Übergangswahrscheinlichkeit p eindeutig bestimmten Wahrscheinlichkeitsmaß P ein Martingal ist. Nach dem Prinzip der risikoneutralen Bewertung (s. z.B. [Se], Abschnitt 1.5.) lässt sich dann der faire Wert einer Europäischen Option zum Zeitpunkt t_i unter dem Wahrscheinlichkeitsmaß P als diskontierter Erwartungswert der Auszahlung zum Zeitpunkt t_{i+1} schreiben, d.h.

$$(1.20) \quad V(S_i, t_i) = e^{-r(t_{i+1}-t_i)} E_P[V(S_{i+1}, t_{i+1})].$$

Das auf den obigen Überlegungen basierende Binomialverfahren besteht nun aus einer Vorwärts- und einer Rückwärtsphase.

Die Menge aller Aktienkurse, die unter obigen Annahmen bis zur Zeit T auftreten können, definiert ein baumartiges Gitter $S_{j,i} = S_0 u^j d^{i-j}$ für $i = 0, \dots, M$ und $j = 0, \dots, i$, welches in der Vorwärtsphase des Binomialverfahren berechnet wird.

In der folgenden Rückwärtsphase werden dann zu jedem Gitterpunkt $S_{j,i}$ die Optionspreise $V_{j,i}$ berechnet. Zum Ausübungszeitpunkt T , d.h. für $i = M$, ist der Wert der Option $V_{j,M}$ für $j = 0, \dots, M$ durch die Auszahlungsfunktion \mathcal{H} als

$$V_{j,M} = h(S_{j,M})$$

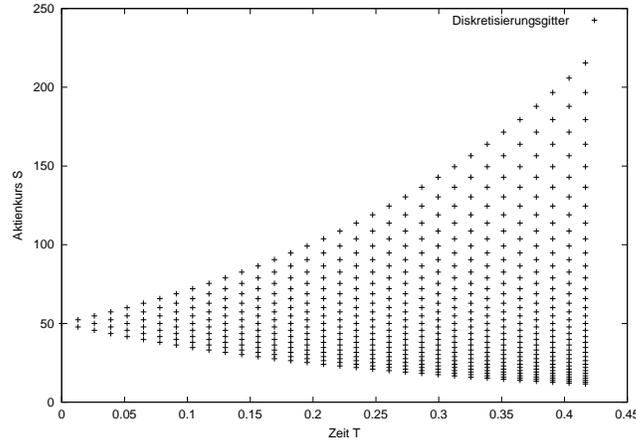


Abbildung 1.3: Diskretisierungsgitter der Binomialmethode für $S_0 = 50$, $T = \frac{1}{5}$ und $M = 32$.

gegeben. Je nachdem, ob es sich um eine Kauf- oder Verkaufsoption handelt, gilt $\mathcal{H} = h_c$ bzw. $\mathcal{H} = \mathcal{H}_p$ mit $\mathcal{H}_c, \mathcal{H}_p$ aus (1.1). Die aus dem $(i+1)$ -ten Zeitschritt bekannten Werten $V_{j,i+1}$ werden nun benutzt, um rückwärts die Werte $V_{j,i}$ für $j = 0, \dots, i$ des i -ten Zeitschritts als diskontierten Erwartungswert

$$(1.21) \quad V_{j,i} = e^{-r\Delta t} (p V_{j+1,i+1} + (1-p) V_{j,i+1}).$$

zu berechnen. Die Rekursion in (1.21) entspricht der Formel (1.20) in der Doppelindexnotation. Auf diese Weise werden in der Rückwärtsphase des Binomialverfahrens für $i = M-1, \dots, 0$ ausgehend von den Werten $V_{j,M}$ alle Werte $V_{j,i}$ für $j = 0, \dots, i$ rekursiv bestimmt. Der Wert $V_{0,0}$ ist die gesuchte Approximation an $V(S_0, 0)$.

Für Amerikanische Optionen muss in jedem Schritt zusätzlich noch überprüft werden, ob vorzeitiges Ausüben günstiger ist, als das Halten der Option. Anstelle von (1.21) erhält man

$$(1.22) \quad V_{j,i} = \max \{ h(S_{j,i}), e^{-r\Delta t} (p V_{j+1,i+1} + (1-p) V_{j,i+1}) \}.$$

Näherungen an die Risikokennziffern Delta, Gamma und Theta (s. Abschnitt 1.3) lassen sich als Differenzen von Optionswerten $V_{j,i}$ und Aktienkursen $S_{j,i}$ bestimmen. Zur Zeitpunkt t_1 erhält man z.B.

$$\text{Delta} \approx \frac{V_{11} - V_{01}}{S_{11} - S_{01}}.$$

Für die Wahl der beiden Parameter u und d gibt es verschiedene Möglichkeiten. In [CRR] werden die beiden Parameter u und d so gewählt, dass neben den Erwartungswerten auch die Varianzen vom diskreten und kontinuierlichen Modell übereinstimmen. Zusätzlich wird $ud = 1$ gefordert. Dies definiert zwei Gleichungen, aus denen die Parameter u und d bestimmt werden können. Man erhält

$$\begin{aligned} u &= \beta + \sqrt{\beta^2 - 1}, \\ d &= \beta - \sqrt{\beta^2 - 1} \end{aligned}$$

mit $\beta := \frac{1}{2} (e^{-r\Delta t} + e^{(r+\sigma^2)\Delta t})$. In diesem Fall kann man mit dem Zentralen Grenzwertsatz zeigen (s. [Sa], Abschnitt 5.4 oder [Du], Abschnitt 11B), dass das Binomialmodell für $M \rightarrow \infty$ gegen das kontinuierliche Black-Scholes Modell konvergiert.

Beispiel 1.4.1 Es wird angenommen, dass der Zinssatz für eine sichere Anlage $r = 2,5\%$ und der heutige Aktienkurs $S_0 = 9$ beträgt. Dann beträgt der Black–Scholes Preis $V_{BS}(S_0, 0)$ zum Zeitpunkt $t = 0$ für eine Europäische Kaufoption mit Ausübungspreis $K = 10$ bei einer Volatilität des Aktienkurses von $\sigma = 0.6$ und bei einer Laufzeit von einem Jahr

$$V_{BS}(S_0, 0) = 1.853859.$$

Ein Binomialverfahren mit obigen Parametern u, d und p liefert in Abhängigkeit der Anzahl der Zeitintervalle M die Werte aus Abbildung 1.4. Man erkennt, dass die Konvergenz nicht monoton und relativ langsam ist.

M	$V(S_0, 0)$
32	1.875648
128	1.851225
256	1.856415
512	1.853251
2048	1.853733

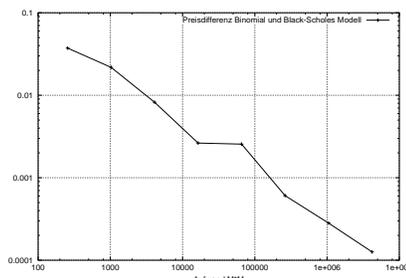


Abbildung 1.4: Konvergenz des Binomialmodells gegen das Black–Scholes Modell.

Das Binomialmodell ist erweiterbar auf viele pfadabhängige Optionen, z.B. Barrier Optionen. Außerdem lassen sich leicht Dividendenzahlungen und zeitabhängige Zinsraten berücksichtigen. In [Bn] wird eine beschleunigte Variante konstruiert, die die Anzahl der Schritte im Binomialmodell mittel Richardson Extrapolation reduziert. In [By] wird die Binomialmethode erweitert, so dass auch zwei zugrunde liegende Wertpapiere berücksichtigt werden können. In [FG] wird eine sogenannte adaptive Gittermethode vorgeschlagen, die die Bäume an die spezielle Struktur der Option anzupassen versucht, indem sie in geeigneter Weise Teile eines feinen Baums über einen groben Baum legen. Bei Amerikanischen Optionen werden so der Bereich in Nähe des Ausübungszeitpunktes und der Bereich in Nähe des Ausübungspreises feiner aufgelöst.

Trinomial– und Multinomialverfahren

Als Alternative zum Binomialverfahren können auch Trinomialverfahren oder allgemein Multinomialverfahren verwendet werden. Anstelle von nur zwei möglichen Aktienkursentwicklungen innerhalb eines Zeitintervalls nimmt man jetzt an, dass B verschiedene Aktienkurse auftreten können. Bei gegebenem Kurs S_i zum Zeitpunkt t_i gilt also in einem Trinomialmodell ($B = 3$)

$$S_{i+1} = \begin{cases} u S_i & \text{mit Wahrscheinlichkeit } p_u \\ m S_i & \text{mit Wahrscheinlichkeit } p_m \\ d S_i & \text{mit Wahrscheinlichkeit } p_d. \end{cases}$$

Dabei sind die Faktoren u, m und $d \in \mathbb{R}^+$ und es gilt $p_d + p_m + p_u = 1$.

Fordert man wie beim Binomialverfahren, dass Erwartungswerte und Varianzen vom diskreten und kontinuierlichen Modell übereinstimmen und nimmt man $u d = m = 1$ an, lassen sich die Parameter d, m, u, p_d, p_m und p_u eindeutig bestimmen. Unter Vernachlässigung von Termen höherer Ordnung als Δt erhält man

$$u = e^{\sigma\sqrt{3\Delta t}}, \quad d = \frac{1}{u}, \quad m = 1,$$

$$p_u = \beta + \frac{1}{6}, \quad p_d = -\beta + \frac{1}{6}, \quad p_m = \frac{2}{3}$$

mit

$$\beta := \sqrt{\frac{\Delta t}{12\sigma^2}} \left(r - \frac{\sigma^2}{2} \right).$$

Man kann zeigen, dass ein Trinomialverfahren einem expliziten Finite-Differenzen Verfahren und ein Multinomialverfahren mit B Zweigen pro Knoten einem impliziten Finite-Differenzen Verfahren mit $B + 1$ Gitterpunkten im Ort entspricht (s. [Hu], S.422).

Control Variate Technik

Die Control Variate Technik kann verwendet werden, um den Fehler zu reduzieren, der bei der Bestimmung Amerikanischer Optionswerte in einem Baumverfahren auftritt. Dazu werden die approximativen Werte B_{am} und B_{eur} einer Amerikanischen bzw. Europäischen Option mit dem Baumverfahren bestimmt und mit der Black-Scholes Formel der exakte Wert V_{eur} der Europäischen Option berechnet.

Man nimmt an, dass der Fehler, der bei der Bestimmung von B_{am} auftritt, dem Fehler entspricht, der bei der Bestimmung von B_{eur} entsteht.

Dann lässt sich der korrigierte Wert V_{am} der Amerikanischen Option angeben als

$$V_{am} := B_{am} + V_{eur} - B_{eur}.$$

Numerische Tests zeigen, dass hiermit in der Tat oft eine signifikante Fehlerreduktion möglich ist. Die Control Variate Technik kann auch in Zusammenhang mit anderen Verfahren, z.B. Monte Carlo Verfahren, verwendet werden. Ausführlichere Informationen über die Control Variate Technik in Zusammenhang mit der Optionsbewertung finden sich in [HIW].

Das Verfahren von Broadly-Glasserman

In der Arbeit [BG] wird vom Martingalansatz ausgegangen. Hier besagt der Satz von Girsanov (s. [Du], S.288), dass der Preis einer Amerikanischen Option als ein Erwartungswert bezüglich eines bestimmten Maßes, dem sogenannten äquivalenten Martingalmaß.

$$V(S, 0) = \sup_{t \leq T} e^{-rt} E^*[V(S, t)]$$

gegeben ist. Der Algorithmus von Broadly-Glasserman bestimmt ein Konfidenzintervall für den Optionspreis, indem er einen Erwartungswert mit zu hohem und einen mit zu niedrigem Bias erzeugt. Asymptotisch besitzen beide Erwartungswerte keinen Bias.

Hierzu wird zunächst ein Zufallsbaum mit B Zweigen je Knoten konstruiert. Zum Zeitpunkt t_j mit $j = 1, \dots, M$ sind dann B^j verschiedene Aktienkurse möglich, die mit $S_j^{i_1, i_2, \dots, i_j}$, $1 \leq i_1, \dots, i_j \leq B$ bezeichnet werden. Sie werden in einem Vorwärtsschritt als Random Walk

$$S_{j+1}^{i_1, i_2, \dots, i_{j+1}} = S_j^{i_1, i_2, \dots, i_j} e^{(r - \frac{1}{2}\sigma^2)\Delta t + \sigma\sqrt{\Delta t} z_{i_{j+1}}}$$

für $i_{j+1} = 1, \dots, B$ simuliert. Hierbei ist $z_{i_{j+1}}$ eine $N(0, 1)$ -verteilte Zufallsvariable.

In einem Rückwärtsschritt wird nun zu jedem Knoten i_1, i_2, \dots, i_j zum Zeitpunkt t_j der zu hohe Erwartungswert $\theta_{high, j}^{i_1, i_2, \dots, i_j}$ und der zu niedrige Erwartungswert $\theta_{low, j}^{i_1, i_2, \dots, i_j}$ aus den Werten des Zeitpunktes t_{j+1} berechnet.

Für $j = M$ gilt

$$\theta_{high,M}^{i_1,\dots,i_M} = \theta_{low,M}^{i_1,\dots,i_M} = V(S_M^{i_1,\dots,i_M}, t_M).$$

Hieraus bestimmt man rekursiv für $j = M - 1, \dots, 0$

$$\theta_{high,j}^{i_1,\dots,i_j} = \max \left\{ V(S_j^{i_1,\dots,i_j}, t_j), e^{-r\Delta t} \frac{1}{B} \sum_{i_{j+1}=1}^B \theta_{high,j+1}^{i_1,\dots,i_j,i_{j+1}} \right\}$$

und

$$\theta_{low,j}^{i_1,\dots,i_j} = \frac{1}{B} \sum_{k=1}^B \eta_j^{i_1,\dots,i_j,k}.$$

Hierbei ist $\eta_j^{i_1,\dots,i_j,k} = V(S_j^{i_1,\dots,i_j}, t_j)$, falls

$$V(S_j^{i_1,\dots,i_j}, t_j) \geq e^{-r\Delta t} \frac{1}{B-1} \sum_{i_{j+1}=1}^B \theta_{low,j+1}^{i_1,\dots,i_j,i_{j+1}} \quad \text{für alle } k = 1, \dots, B.$$

Ansonsten gilt $\eta_j^{i_1,\dots,i_j,k} = e^{-r\Delta t} \theta_{low,j+1}^{i_1,\dots,i_j,k}$.

Der gesuchte Optionspreis zum Zeitpunkt $t_0 = 0$ und Kurs S_0 beträgt dann

$$V(S_0, 0) = \frac{1}{2}(\theta_{high,0} + \theta_{low,0}).$$

Das Broady–Glassermann Verfahren hat in Vergleich zum Binomialverfahren den Vorteil, dass es viel leichter auf allgemeinere Optionstypen, wie pfadabhängige Optionen oder Basket–Optionen verallgemeinerbar ist.

1.4.2 Finite–Differenzen und Finite–Elemente

Sobald eine partielle Differentialgleichung und Randbedingungen für den Wert einer Option angegeben werden können, können Finite–Differenzen und Finite–Elemente Methoden zur Optionsbewertung verwendet werden. Für zahlreiche Optionstypen sind die partiellen Differentialgleichungen und geeignete Randbedingungen in [WHD1] aufgelistet. Da sich Integralansätze für die Bewertung Amerikanischer Optionen nur sehr bedingt eignen, bietet sich hier besonders der Weg über die Differentialgleichung an.

Werte von Europäischen Optionen erfüllen im Allgemeinen Randwertprobleme, wohingegen Werte Amerikanischer Optionen im Allgemeinen durch freie Randwertprobleme gegeben sind (s. Problem 1.2.1 und Problem 1.2.3).

Die Diskretisierung der partiellen Differentialgleichungen mit Finite–Differenzen oder Finite–Elementen führt bei Europäischen Optionen auf ein lineares Gleichungssystem und bei Amerikanischen Optionen auf ein diskretes, lineares komplementäres Problem (s. Abschnitt 2.2.1 und Anhang A.4). Beide können mit Standardverfahren gelöst werden. In Kapitel 2 ist dies ausführlich für den Finite–Elemente–Ansatz diskutiert. Um das diskrete, lineare, komplementäre Problem effizient zu lösen, können zusätzlich Mehrgittertechniken eingesetzt werden.

In [BS2] wird die Finite–Differenzen Methode zum ersten Mal auf die Bestimmung von Optionspreisen angewandt. Detaillierte Informationen über die Diskretisierung der Black–Scholes Gleichung findet sich in den Büchern [Hu], Abschnitt 16.8, [Se] und [WHD1], Kapitel 17–21. Durch die Anwendung nichtlinearer Transformationen lässt sich oft die Effizienz des Verfahrens erhöhen (s. [CP]) oder die Schwierigkeit umgehen, die Konvektionsterme in der partiellen

Differentialgleichung verursachen (s. [Se]). Einen Vergleich verschiedener Finite-Differenzen Verfahren findet sich in [GS]. Dort wird geschlossen, dass das explizite Finite-Differenzen Verfahren verbunden mit einer logarithmischen Transformation der Aktienkurse für viele Optionstypen die effizienteste Methode ist. In der aktuellen Arbeit [HW] wird gezeigt, dass durch die Einführung künstlicher Randbedingungen die Effizienz der Finite-Differenzen Methode weiter erhöht werden kann.

Im Vergleich zu Finite-Elementen haben Finite-Differenzen den Nachteil, dass sie nur schwer auf nicht-gleichmäßige Gitter und nur schwer auf höhere Ordnungen verallgemeinerbar sind. Außerdem sind Fehlerabschätzungen im Finite-Elemente Rahmen leichter zu treffen.

Die Anwendung Finite-Elemente-Verfahren auf die Black-Scholes Gleichung ist in [Se], Kapitel 5 und in [WHD1], Kapitel 21 und Anhang D für stückweise lineare Ansatzfunktionen beschrieben. In Kapitel 2 wird das Vorgehen auf Ansatzfunktionen beliebiger Ordnung verallgemeinert. Für generelle Informationen über Finite-Elemente-Verfahren siehe z.B. [Br].

1.4.3 Analytische Approximationen und weitere Verfahren

Als Alternative zu numerischen Verfahren sind eine Reihe analytischer Approximationen an den Wert einer Amerikanischen Option vorgeschlagen wurden.

Die bekannteste ist eine quadratische Approximation, die auf die Arbeit [Mi] zurückgeht und in [BW] erweitert wurde. Hier wird ausgenutzt, dass die Differenz v zwischen dem Wert einer Europäischen und einer Amerikanischen Option die Black-Scholes Gleichung 1.7 erfüllt. Die Substitutionen

$$\tau = T - t, \quad h(\tau) = 1 - e^{-r\tau}, \quad q = 2r/\sigma^2$$

und $v(S, t) = h(\tau)g(S, h)$ führen zur transformierten Gleichung

$$S^2 \frac{\partial^2 g}{\partial S^2} + qS \frac{\partial g}{\partial S} - \frac{q}{h}g - q(1-h) \frac{\partial g}{\partial h} = 0.$$

Die Approximation besteht jetzt darin, anzunehmen, dass der letzte Term auf der linken Seite gleich Null ist. Dies ist dadurch motiviert, dass $1-h$ nah an Null ist, wenn τ groß ist, und dass $\partial g/\partial h$ nah an Null ist, wenn τ klein ist (s. [BW]).

Für die vereinfachte Gleichung lässt sich eine analytische Lösung angeben und nach Rücktransformation erhält man den Preis einer Amerikanischen Verkaufsoption als

$$V_{am}(S, t) = \begin{cases} V_{eur}(S, t) + A(S_f) (S/S_f)^\gamma & \text{falls } S > S_f \\ K - S & \text{sonst.} \end{cases}$$

Der optimale Ausübungspreis S_f wird durch die iterativ zu lösende Gleichung

$$K - S_f = V_{eur}(S_f, t) + A(S_f)$$

bestimmt. Dabei bezeichnet N die Verteilungsfunktion der Standardnormalverteilung und die Variablen $\gamma = \gamma(h)$, $A = A(S_f)$ und $d = d(S)$ sind gegeben durch

$$\gamma := \frac{1}{2} \left(1 - q - \sqrt{(q-1)^2 + 4\alpha/h} \right),$$

$$A := -(1 - N[d(S_f)]) \frac{S_f}{\gamma},$$

$$d := \frac{\ln(S/K) + (r + \frac{1}{2}\sigma^2)(T-t)}{\sigma\sqrt{T-t}}.$$

Eine neuere Approximation stammt aus [BJS]. Sie ist etwas weniger rechenintensiv und genauer für Optionen mit langer Laufzeit. Beide Approximationen sind sowohl für Kauf- als auch für Verkaufsoptionen anwendbar und sind direkt auf Aktien mit stetiger Dividendenzahlung erweiterbar.

Interessant für diese Arbeit ist ein asymptotisches Modell zur Bewertung Amerikanischer Optionen mit stochastischer Volatilität aus [FPS]. Hierfür wird eine gute Approximation an die dritte Ableitung $\partial^3 V / \partial S^3$ der Lösung V eines eindimensionalen Black-Scholes Modell benötigt. Dies zeigt den Bedarf nach Black-Scholes Lösern, die auf Funktionen höherer Ordnung beruhen.

Weitere Lösungsverfahren

Die Bewertung Amerikanischer Optionen führt zu linearen komplementären Problemen. Diese sind eng verwandt mit restringierten Minimierungsproblemen (s. Abschnitt A.3). Damit kommen *Penalty Methoden* als Löser in Betracht. Hier wird ein geeignetes unrestringiertes Minimierungsproblem gelöst, welches einen Strafterm enthält, der sicherstellt, dass die Lösung im Grenzwert der Lösung des restringierten Problems entspricht. In der Arbeit [FV] wird für die Bewertung von Amerikanischen Optionen eine Penalty Methode mit quadratischer Konvergenz vorgeschlagen. In [DH] wird ebenfalls bei der Formulierung als Minimierungsproblem angesetzt. Dort werden Methoden der *linearen Optimierung* zur Lösung verwendet. Numerische Ergebnisse zeigen schnellere Konvergenz als Finite-Differenzen Methoden mit äquidistanten Gittern.

Kapitel 2

Finite–Elemente Lösung des freien Randwertproblems

2.1 Formulierung des freien Randwertproblems: Einleitung und Übersicht

In diesem Kapitel wird ein auf B–Spline–Basisfunktionen basierendes Finite–Elemente–Verfahren zur Lösung des zeitabhängigen, freien Randwertproblems vorgestellt, welches bei der Bewertung Amerikanischer Optionen im Black–Scholes Modell auftritt.

Es bezeichne $\Omega_{\mathcal{L}}$ das Gebiet

$$(2.1) \quad \Omega_{\mathcal{L}} := \mathbb{R}^+ \times [0, T)$$

und \mathcal{L} den Operator

$$(2.2) \quad \mathcal{L} := \frac{\partial}{\partial t} + \frac{1}{2}\sigma^2 S^2 \frac{\partial^2}{\partial S^2} + rS \frac{\partial}{\partial S} - r$$

zur Black–Scholes Gleichung aus (1.7) für gegebene Parameter $r \in \mathbb{R}^+$ und $\sigma \in \mathbb{R}^+$. Außerdem sei für $K \in \mathbb{R}^+$ die Funktion $\mathcal{H} = \mathcal{H}(S)$ in Analogie zu (1.1) definiert als

$$\mathcal{H}(S) := (K - S)^+.$$

Problem 2.1.1 (Freies Randwertproblem) Finde (V, S_f) mit $V = V(S, t)$ und $S_f = S_f(t)$, so dass gilt

$$\begin{aligned} \mathcal{L}V(S, t) &= 0 && \text{für } S > S_f \text{ und } 0 \leq t < T, \\ V(S, t) &= \mathcal{H}(S) && \text{für } S \leq S_f \text{ und } 0 \leq t < T, \end{aligned}$$

mit den Randdaten

$$V(S, t) = 0 \quad \text{für } S \rightarrow \infty \text{ und } 0 \leq t < T,$$

den Enddaten

$$V(S, T) = \mathcal{H}(S) \quad \text{für } S \geq 0$$

und den Bedingungen, dass V und $\partial V / \partial S$ an $S = S_f(t)$ stetig sind.

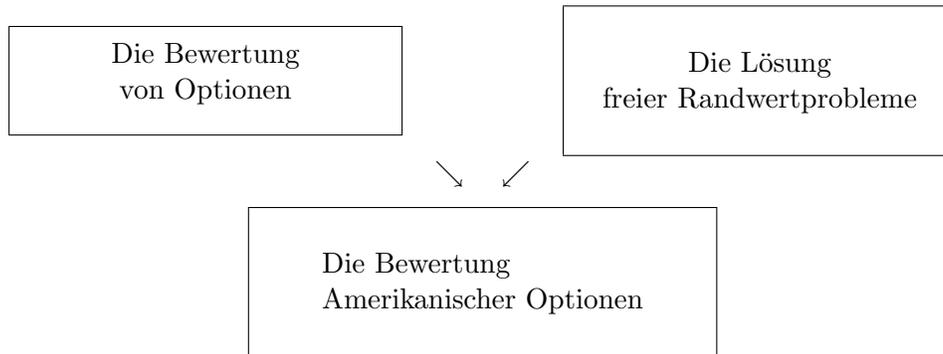


Abbildung 2.1: Zusammenhang von Optionsbewertung mit freien Randwertproblemen.

Wie in Abschnitt 1.2.3 hergeleitet wurde, beschreibt die Lösung V den *fairen Wert einer Amerikanischen Verkaufsoption*.

Das im folgenden beschriebene Lösungsverfahren lässt sich mit wenigen Modifikationen auch zur Lösung der parabolischen Black-Scholes Gleichung verwenden, welche bei der Bewertung Europäischer Optionen im Rahmen des Black-Scholes Modells auftritt (s. Abschnitt 1.2.2). Da hier im Gegensatz zu Problem 2.1.1 nicht die Restriktion $V(S, t) \geq \mathcal{H}(S)$ gilt, wird es im folgenden zur besseren Unterscheidung als *unrestringiertes Randwertproblem* bezeichnet.

Problem 2.1.2 (Unrestringiertes Randwertproblem) Finde $V = V(S, t)$, so dass für alle $(S, t) \in \Omega_{\mathcal{L}}$ die Black-Scholes Gleichung $\mathcal{L}V = 0$ mit den Randdaten

$$V(S, t) = 0 \quad \text{für } S = 0 \quad \text{und} \quad V(S, t) = S - Ke^{-r(T-t)} \quad \text{für } S \rightarrow \infty$$

und den Enddaten $V(S, T) = \mathcal{H}(S)$ gilt.

Wenn man nicht die Lösung für die gesamte (S, t) -Fläche benötigt, wird man zur Lösung des unrestringierten Problem 2.1.2 auf die geschlossene Lösungsformel aus (1.11) zurückgreifen. Im folgenden dient das unrestringierte Problem in erster Linie als Benchmarking für den Lösungsalgorithmus des schwierigeren freien Randwertproblems 2.1.1.

Die Finite-Elemente Methode besitzt eine Reihe von Vorteilen im Vergleich zu den anderen in Kapitel 1.4 vorgestellten numerischen Verfahren. Dazu zählen unter anderem die Möglichkeiten, dass auch nicht gleichmäßige Gitter leicht berücksichtigt werden können und, dass Basisfunktionen beliebiger Ordnung verwendet und damit höhere Approximationsordnungen erreicht werden können. Die Verwendung von glatten Basisfunktionen beinhaltet ferner den großen Vorteil, dass die Ableitungen der Lösung, die bei der Optionsbewertung die wichtigen Risikokennziffern liefern, direkt durch Ableiten der Basisfunktionen bestimmt werden können.

Die folgenden Abschnitte orientieren sich an den Darstellungen in [Se], Kapitel 5 und in [WHD1], Kapitel 21 und Anhang D. Dort ist jeweils das Verfahren für Hutansatzfunktionen beschrieben. Es wird hier auf B-Splines beliebiger Glattheit verallgemeinert. Um die damit neu auftretenden Schwierigkeiten bei der Modellierung der Randdaten zu umgehen, wird im Unterschied zu [Se] und [WHD1] die schwache Form des Problems auf homogene Randdaten reduziert.

In Abschnitt 2.2 wird zunächst die schwache Formulierung hergeleitet. Die Black-Scholes Gleichung wird dafür in die Wärmeleitungsgleichung transformiert und auf homogene Randdaten reduziert. Diskretisierung in der Zeit mit einem θ -Schema und Finite-Elemente Diskretisierung im Ort führen im Abschnitt 2.3 zur vollständig diskreten Form. Das hieraus resultierende Minimierungsproblem über einer konvexen Menge in \mathbb{R}^2 lässt sich nun mit dem *projektiven Gauss-Seidel-Verfahren* lösen. Es wird in Abschnitt 2.6 beschrieben. Massen-, Steifigkeitsmatrix und die rechte Seite lassen sich bei äquidistanten Gittern explizit aufstellen (Abschnitt

2.5.1). Abschließend wird in Abschnitt 2.5.2 die Bestimmung des freien Randes und die Bestimmung der Ableitungen durch direktes Ableiten des Basisfunktionen und diskutiert. Eine solche Lösung ist natürlich nur dann möglich, wenn Basisfunktionen verwendet werden, die ausreichend oft differenzierbar sind. Mit quadratischen bzw. kubischen Splines erhält man stückweise stetige Approximationen an die ersten beiden Ableitungen der Lösung.

Im folgenden Schema sind die Lösungsverfahren für das freie und das unrestringierte Randwertproblem, wie sie in den Abschnitten 2.2 bis 2.5.2 ausgeführt sind, gegenübergestellt.

Bewertung Europäischer Optionen	Bewertung Amerikanischer Optionen
↓	(Annahmen des Black–Scholes Modell) ↓
PDE (Black–Scholes Gleichung)	freies Randwertproblem
↓ (Transformation)	(Umformulierung + Transformation) ↓
PDE (Wärmeleitungsgleichung)	lineares komplementäres Problem
↓ (Multiplikation mit Testfunktion, Partielle Integration)	↓
Variationsgleichung	Variationsungleichung
↓ (Finite–Elemente + Zeitdiskretisierung)	↓
LGS	LGS mit Nebenbedingung
↓ (SOR + Rücktransformation)	(PSOR + Rücktransformation) ↓
Lösung	Lösung

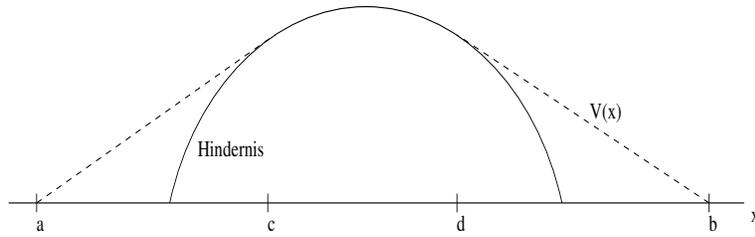
2.2 Herleitung der variationellen Form

Das Ergebnis dieses Abschnittes ist die variationelle Form des freien Randwertproblems 2.1.1. Zunächst wird die lineare komplementäre Formulierung des Problems hergeleitet. Durch eine geeignete Transformation werden anschließend die auftretenden Konvektions- und Reaktionsterme eliminiert. Zuletzt wird das Problem mit einer geeigneten Substitution auf homogene Rand- und Anfangsbedingungen reduziert.

2.2.1 Das freie Randwertproblem als lineares komplementäres Problem

Eine große Klasse von freien Randwertproblemen lässt sich in Form einer *Variationsungleichung* oder in Form eines *linearen komplementären Problems* (s. Anhang A.3 und A.4) formulieren. Eine solche Umformulierung ist oft günstig, um Schwierigkeiten bei der Behandlung des freien Randes zu umgehen. Man spricht dann von einem "fixed-domain" Ansatz, da das Gebiet, in dem das neue Problem formuliert ist, jetzt festgelegt ist. Der freie Rand ist vollständig in diesem Gebiet enthalten, taucht in der neuen Formulierung aber nicht mehr explizit auf. Er trennt die Menge K° der Punkte, an denen die Lösung die Differentialgleichung erfüllt, von der Menge K^\bullet an Punkten, an denen dies nicht gilt. Im zweiten Fall werden die Punkte als *Kontaktpunkte* oder auch als *aktive Punkte* bezeichnet. Liegt ein *Hindernisproblem* (s. Anhang A.4) vor, dann ist diese Bezeichnung dadurch motiviert, dass die Lösung an allen Punkten der Menge K^\bullet Kontakt mit der Hindernisfunktion besitzt. Dies soll an einem einfachen Beispiel erläutert werden.

Beispiel 2.2.1 Gegeben sei die Hindernisfunktion $\mathcal{H}(x) \in C^2(I)$ aus Abbildung 2.2 mit $I := [a, b]$ und $\mathcal{H}(a) = \mathcal{H}(b)$. Gesucht ist eine Funktion $V \in C^1(I)$, die $V \geq \mathcal{H}$ und $V(a) = V(b) = 0$ erfüllt und minimale Länge besitzt. Damit eine Lösung existiert muss $\mathcal{H}(a), \mathcal{H}(b) < 0$ gelten. Das Intervall $[c, d]$, in dem $V(x) = \mathcal{H}(x)$ gilt, wird als Kontaktmenge K^\bullet der Funktion

Abbildung 2.2: Hindernisfunktion \mathcal{H} und Lösung V .

\mathcal{H} bezeichnet. Alternativ lässt sich die Funktion V auch durch das folgende einfache *freie Randwertproblem* festlegen: Finde $V \in C^1(I)$ mit $V(a) = V(b) = 0$ und

$$V'' = 0 \quad \text{in } [a, c] \cup [d, b],$$

$$V = \mathcal{H} \quad \text{in } [c, d].$$

Die Punkte $c, d \in I$, an denen V das Hindernis \mathcal{H} tangential berührt, bilden in diesem Fall den a priori unbekanntem freien Rand. Offensichtlich gilt

$$V''(V - \mathcal{H}) = 0,$$

$$-V'' \geq 0, \quad V - \mathcal{H} \geq 0.$$

Dies wird als die *lineare komplementäre Formulierung* des Problems bezeichnet. Mit partieller Integration erhält man analog zu Abschnitt 2.2.3 die Formulierung als *Variationsungleichung*: Finde $V \in H_0^1(I)$ mit

$$\int_a^b V'(v - V)' dx \geq 0$$

für alle $v \in H_0^1(I)$ mit $v \geq \mathcal{H}$.

Das freie Randwertproblems 2.1.1 kann analog zu dem obigen Beispiel als ein Hindernisproblem aufgefasst werden. Die Funktion \mathcal{H} definiert wie im Beispiel die untere Hindernisfunktion für die Lösung V . Für dieses Problem wird zunächst die lineare komplementäre Formulierung hergeleitet.

In Abschnitt 1.2.3 wurde gezeigt, dass für $S \leq S_f(t)$

$$\mathcal{L}V < 0, \quad V(S, t) = \mathcal{H}(S)$$

und für $S > S_f(t)$

$$\mathcal{L}V = 0, \quad V(S, t) > \mathcal{H}(S)$$

gilt. Da also entweder $\mathcal{L}V = 0$ oder $V - \mathcal{H} = 0$ ist, gilt in jedem Fall

$$\mathcal{L}V \cdot (V(S, t) - \mathcal{H}(S)) = 0.$$

Dies liefert die lineare komplementäre Formulierung des freien Randwertproblems 2.1.1 im Gebiet $\Omega_{\mathcal{L}} = \mathbb{R}^+ \times [0, T)$.

Problem 2.2.2 (Lineares komplementäres Problem) Finde $V = V(S, t)$, so dass für $(S, t) \in \Omega_{\mathcal{L}}$ gilt

$$\mathcal{L}V \cdot (V(S, t) - \mathcal{H}(S)) = 0,$$

$$\mathcal{L}V \leq 0, \quad V(S, t) - \mathcal{H}(S) \geq 0,$$

mit den Randdaten

$$V(S, t) = 0 \quad \text{für } S \rightarrow \infty \quad \text{und} \quad V(S, t) = \mathcal{H}(S) \quad \text{für } S \rightarrow 0,$$

und den Enddaten

$$V(S, T) = \mathcal{H}(S).$$

Zur Äquivalenz dieser beiden Formulierungen sei auf den Anhang verwiesen.

2.2.2 Transformation und Lokalisation

Eine direkte Diskretisierung des Operators \mathcal{L} aus (2.2) führt in der Regel wegen des Konvektionsterms zu numerischen Schwierigkeiten. Mit den Transformationen

$$(2.3) \quad S = Ke^x, \quad t = T - \frac{2\tau}{\sigma^2}$$

und

$$(2.4) \quad V(S, t) = K e^{-\frac{x}{2}(q-1) - (\frac{1}{4}(q-1)^2 + q)\tau} u(x, \tau),$$

wobei $q := 2r/\sigma^2$, ist die rückwärts parabolische Differentialgleichung $\mathcal{L}V = 0$ in $V(S, t)$ aber äquivalent zur homogenen, vorwärts parabolischen Wärmeleitungsgleichung

$$(2.5) \quad \mathcal{W}u := \frac{\partial u}{\partial \tau} - \frac{\partial^2 u}{\partial x^2} = 0$$

in der neuen Unbekannten $u(x, \tau)$. Das Gebiet $\Omega_{\mathcal{L}}$ transformiert sich dabei auf

$$\Omega_{\mathcal{W}} := \mathbb{R} \times (0, \sigma^2 T/2].$$

Auf diese Weise umgeht man die numerischen Schwierigkeiten bei der Behandlung des Konvektionsterms und vereinfacht durch die nun konstanten Koeffizienten die Implementierung. Auf der anderen Seite wirkt sich die Transformation erst einmal unvorteilhaft auf das Diskretisierungsgitter aus. Dies kann bei einem Finite-Elemente-Ansatz durch die Verwendung gestreckter Gitter aber wieder kompensiert werden.

Wendet man obige Transformation auf die Funktion $\mathcal{H}(S, t) = \mathcal{H}(S) = (K - S)^+$ an, ergibt sich die transformierte Funktion $g : \Omega_{\mathcal{W}} \rightarrow \mathbb{R}^+$ als

$$(2.6) \quad g(x, \tau) := e^{(\frac{1}{4}(q+1)^2\tau)} \max\{e^{\frac{x}{2}(q-1)} - e^{\frac{x}{2}(q+1)}, 0\}.$$

Weiter erhält man folgende transformierte Formen des linear komplementären Problems 2.2.2 und des unrestringierten Randwertproblems 2.1.2.

Problem 2.2.3 (Transformiertes lineares komplementäres Problem) *Finde die Funktion $u = u(x, \tau)$, so dass für alle $x \in \Omega_{\mathcal{W}}$ gilt*

$$\mathcal{W}u \cdot (u - g) = 0$$

$$\mathcal{W}u \geq 0, \quad u \geq g$$

mit den Randdaten

$$\lim_{x \rightarrow \pm\infty} u(x, \tau) = \lim_{x \rightarrow \pm\infty} g(x, \tau)$$

und den Anfangsdaten

$$u(x, 0) = g(x, 0).$$

Problem 2.2.4 (Transformiertes unrestringiertes Randwertproblem) *Finde die Funktion $u = u(x, \tau)$, so dass für alle $(x, \tau) \in \Omega_{\mathcal{W}}$ die Wärmeleitungsgleichung $\mathcal{W}u = 0$ mit den Randdaten*

$$u(x, \tau) = 0 \quad \text{für } x \rightarrow -\infty \quad \text{und} \quad u(x, \tau) = r_c \quad \text{für } x \rightarrow +\infty$$

und den Anfangsdaten

$$u(x, 0) = e^{\frac{x}{2}(q-1)} \max\{e^x - 1, 0\}$$

gilt. Dabei bezeichnet

$$r_c := \left(e^{\frac{x}{2}(q+1)+q\tau} - e^{\frac{x}{2}(q-1)} \right) e^{\frac{\tau}{4}(q-1)^2}.$$

Exemplarisch, weil bis auf diese alle übrigen Transformationen in [Se], Kapitel 4 ausführlich durchgeführt sind, sei hier nur die transformierte Randbedingung $u(x, \tau) = r_c$ für $x \rightarrow +\infty$ hergeleitet:

Mit (1.10) und (2.3) gilt für $S \rightarrow \infty$

$$V(S, t) = S - Ke^{-r(T-t)} = Ke^x - Ke^{-r\frac{2\tau}{\sigma^2}} = K(e^x - e^{-q\tau}).$$

Einsetzen in (2.4) liefert

$$\begin{aligned} u(x, \tau) &= (e^x - e^{-q\tau}) e^{\frac{x}{2}(q-1) + (\frac{1}{4}(q-1)^2 + q)\tau} \\ &= (e^x - e^{-q\tau}) e^{\frac{x}{2}(q-1) + q\tau} e^{\frac{\tau}{4}(q-1)^2} \\ &= \left(e^{\frac{x}{2}(q+1) + q\tau} - e^{\frac{x}{2}(q-1)} \right) e^{\frac{\tau}{4}(q-1)^2} \\ &= r_c. \end{aligned}$$

□

Lokalisation des Gebietes

Für numerische Berechnungen muss das unbeschränkte Gebiet $\Omega_{\mathcal{W}}$ durch ein beschränktes Gebiet

$$(2.7) \quad \Omega := I_x \times I_\tau := [x_{\min}, x_{\max}] \times (0, \sigma^2 T/2] \subset \mathbb{R}^2$$

mit $x_{\min} < 0 < x_{\max}$ ersetzt werden. Für das unrestringierte Randwertproblem kann man zeigen (s. [MPS]), dass der Lokalisationsfehler mit steigendem x_{\max} bzw. fallendem x_{\min} exponentiell kleiner wird. Für das freie Randwertproblem ist in [JLL] ein gleichmäßig fallender Lokalisationsfehler nachgewiesen. Für die numerischen Berechnungen wurde [Oo] folgend $S_{\max} = 3K$ und $S_{\min} = \frac{1}{3}K$ gewählt. Aus den Transformationen 2.2.2 ergeben sich dann die Werte x_{\min} und x_{\max} .

Die Ausführungen in den Unterabschnitten 2.2.3 bis 2.3.1 beziehen sich nur auf das schwierigere freie Randwertproblem 2.2.3, da die Herleitung der schwachen Form des unrestringierten Randwertproblems 2.2.4 Standard ist und die Diskretisierung vollkommen analog verläuft. Die Endresultate sind für beide Probleme in Abschnitt 2.3.2 zusammengefasst.

2.2.3 Schwache Form und Reduktion auf homogene Randbedingungen

Zu dem Gebiet $\Omega = I_x \times I_\tau \subset \mathbb{R}^2$ aus (2.7) sei der Raum $H(\Omega)$ definiert als die Menge aller Funktionen $v = v(x, \tau)$ in $L^2(\Omega)$, die bezüglich x einmal schwach und bezüglich τ einmal stark stetig differenzierbar sind. Eine kurze Einführung der Sobolevräume und ihrer wichtigsten Eigenschaften findet sich in Anhang A.2.

Die Menge der Testfunktionen sei definiert als

$$\bar{\mathcal{K}} = \{v \in H(\Omega) : v(x, \tau) \geq g(x, \tau), v(x_{\max}, \tau) = g(x_{\max}, \tau), v(x_{\min}, \tau) = g(x_{\min}, \tau), \\ v(x, 0) = g(x, 0) \text{ für alle } (x, \tau) \in \Omega\}$$

und es bezeichne $u \in H(\Omega)$ die Lösung des transformierten linearen komplementären Problems 2.2.3. Dann ist $v - g \geq 0$ und $u - g \geq 0$, wegen $u, v \in \bar{\mathcal{K}}$. Subtrahieren von

$$\int_{x_{\min}}^{x_{\max}} \mathcal{W} u (v - g) dx \geq 0$$

und

$$\int_{x_{\min}}^{x_{\max}} \mathcal{W} u (u - g) dx \geq 0$$

liefert

$$\int_{x_{\min}}^{x_{\max}} \mathcal{W} u (v - u) dx \geq 0.$$

Bei anschließender partieller Integration verschwinden die Randterme wegen $u, v \in \bar{\mathcal{K}}$ und man erhält die schwache Form von Problem 2.2.3.

Problem 2.2.5 (Schwache Form mit inhomogenen Randdaten)

$$(2.8) \quad \text{Finde } u \in \bar{\mathcal{K}} : \int_{x_{\min}}^{x_{\max}} \frac{\partial u}{\partial \tau} (v - u) + \frac{\partial u}{\partial x} \left(\frac{\partial v}{\partial x} - \frac{\partial u}{\partial x} \right) dx \geq 0 \quad \text{für alle } v \in \bar{\mathcal{K}}.$$

Im weiteren Verlauf ist es günstig das Problem durch Subtraktion einer Funktion $g(x, \tau) \in H(\Omega)$, die an den Rändern $x = x_{\max}$ und $x = x_{\min}$ mit den Randbedingungen übereinstimmt, auf homogene Randbedingungen zu reduzieren. Hierzu kann offensichtlich die Funktion g aus (2.6) verwendet werden. Die Substitution

$$(2.9) \quad u = y + g$$

in (2.8) liefert das auf homogene Randdaten reduzierte Problem.

Problem 2.2.6 (Schwache Form mit homogenen Randdaten)

$$(2.10) \quad \text{Finde } y \in \mathcal{K} : \int_{x_{\min}}^{x_{\max}} \frac{\partial y}{\partial \tau} (v - y) + \frac{\partial y}{\partial x} \left(\frac{\partial v}{\partial x} - \frac{\partial y}{\partial x} \right) dx \geq \ell(v - y) \quad \text{für alle } v \in \mathcal{K}.$$

Dabei ist

$$\mathcal{K} := \{v \in H(\Omega) : v(x, \tau) \geq 0, v(x_{\min}, \tau) = v(x_{\max}, \tau) = 0, v(x, 0) = 0\}$$

und das lineare Funktional ℓ ist definiert als

$$(2.11) \quad \ell(v) := - \int_{x_{\min}}^{x_{\max}} \frac{\partial g}{\partial \tau} v + \frac{\partial g}{\partial x} \frac{\partial v}{\partial x} dx.$$

Es zeigt sich, dass die Lösung des neuen Problems nicht nur homogene Randdaten hat, sondern sogar in einer Umgebung des Randes $x = x_{\min}$ konstant gleich Null und in einer Umgebung von $x = x_{\max}$ beliebig klein ist, wenn x_{\max} genügend groß und x_{\min} genügend klein gewählt wird.

Lemma 2.2.7 *Es gilt $y(x, \tau) = 0$ für alle $\tau \in I_\tau$, falls $x < \ln(1 - \exp(-2r\tau/\sigma^2))$ oder falls $x \rightarrow \infty$.*

Beweis: Aus dem Beweis von Satz 1.1.10 folgt, dass

$$V(S, t) = K - S, \quad \text{falls } S < \left(1 - e^{-r(T-t)}\right) K.$$

Mit den Transformationen aus (2.3) bedeutet dies

$$V(S, t) = K(1 - e^x), \quad \text{falls } x < \ln(1 - e^{-2\frac{r\tau}{\sigma^2}}).$$

Wegen (2.9) und (2.4) folgt in diesem Fall

$$\begin{aligned} y + g &= e^{\frac{x}{2}(q-1) + (\frac{1}{4}(q-1)^2 + q)\tau} - e^{\frac{x}{2}(q+1) + (\frac{1}{4}(q-1)^2 + q)\tau} \\ &= e^{(\frac{1}{4}(q-1)^2 + q)\tau} \left(e^{\frac{x}{2}(q-1)} - e^{\frac{x}{2}(q+1)} \right). \end{aligned}$$

Ein Vergleich mit der Definition von g in (2.6), liefert schließlich $y(x, \tau) = 0$, sobald die Bedingung $x < \ln(1 - e^{-2\frac{r\tau}{\sigma^2}}) < 0$ gilt. Wegen

$$\lim_{x \rightarrow \infty} u(x, \tau) = \lim_{x \rightarrow \infty} g(x, \tau)$$

(s. Problem 2.2.3) und wegen (2.9) folgt außerdem $\lim_{x \rightarrow \infty} y(x, \tau) = 0$. □

Nach Lemma 2.2.7 erfüllt die Lösung y der Variationsungleichung aus (2.10) nicht nur homogene Randdaten, sondern ist sogar in einer Umgebung des Randes konstant gleich Null. Damit ist die Modellierung der Randdaten auch für B-Splines höherer Ordnung einfach durch Nullsetzen der Rand B-Splines ohne Verlust an Approximationsgüte möglich. Neben homogenen Randdaten erhält man außerdem anstelle der ursprünglich nicht-glatten Anfangsdaten verschwindende Anfangsdaten. Dies erleichtert die Konvergenzanalyse und man erspart sich die Konstruktion einer Approximation aus dem Finite-Elemente Raum an die Anfangsdaten. Ein weiterer Vorteil ist, dass anstelle der Nebenbedingung $u \geq g$ die einfachere Nebenbedingung $y \geq 0$ gilt. Dies vereinfacht den Einsatz von Mehrgitterverfahren, weil – im Gegensatz zur Funktion g – der Transfer der Nullfunktion von feineren auf gröbere Gitter trivial ist. Dafür erhält man jetzt allerdings eine rechte Seite. Für äquidistante Gitter kann diese aber ebenso wie Massen- und Steifigkeitsmatrix für Ansatzfunktionen beliebiger Ordnung explizit aufgestellt werden.

Mit der Bilinearform

$$\bar{a}(u, v) := \int_{x_{\min}}^{x_{\max}} \frac{\partial}{\partial x} u \frac{\partial}{\partial x} v \, dx$$

und mit dem L^2 -Skalarprodukt $\langle \cdot, \cdot \rangle$ in $I_x = [x_{\min}, x_{\max}]$ lässt sich die Variationsungleichung aus (2.10) schreiben als

$$(2.12) \quad \text{finde } y \in \mathcal{K} : \quad \left\langle \frac{\partial y}{\partial \tau}, v - y \right\rangle + \bar{a}(y, v - y) \geq \ell(v - y) \quad \text{für alle } v \in \mathcal{K}.$$

Dies ist eine parabolische Variationsungleichung der ersten Art. Es ist bekannt (s. [EO]), dass diese eine eindeutige Lösung besitzt, wenn die Bilinearform $\bar{a}(\cdot, \cdot)$ symmetrisch und positiv definit ist. Wegen Lemma A.2.6 ist dies erfüllt.

2.3 Diskretisierung

Die Variationsungleichung aus (2.10) soll mit äquidistanter Schrittweite in der Zeit mit Finite-Differenzen und mit variabler Schrittweite im Ort mit Finite-Elementen diskretisiert werden. Zunächst wird die Zeit diskretisiert, so dass in jedem Zeitschritt ein elliptisches Problem zu lösen ist. Als Finite-Elemente-Ansatzfunktionen dienen hierbei *B-Splines*. Die Definition und die wichtigsten Eigenschaften von B-Splines sind im Anhang A.1 zusammengestellt.

Ein ursprünglich gleichmäßiges Gitter wird durch die in Abschnitt 2.2.2 angewandte nichtlineare Transformation der Black-Scholes Gleichung verzerrt. Je kleiner der Basiswert, desto enger wird das Gitter. Diesem für die Genauigkeit der Näherungslösung nicht vorteilhaften Effekt kann mit der Verwendung von gestreckten Gittern a priori entgegengewirkt werden. Zusätzlich kann die Genauigkeit drastisch erhöht werden (s. [CP]), indem das Gitter in der Nähe des Ausübungspreises weiter verfeinert wird. Das ist der Bereich, der in der Praxis besonders interessant ist. Bei Amerikanischen Optionen bietet es sich weiterhin an, das Gitter in der Nähe des freien Randes vergleichsweise eng zu wählen.

2.3.1 Diskretisierung in der Zeit

Es sei $\tau_j := j\Delta\tau$ für $j = 0, \dots, N$ und $N \in \mathbb{N}$ eine äquidistante Diskretisierung des Zeitintervalls I_τ mit der Schrittweite $\Delta\tau := \frac{1}{2}\sigma^2 T/N$. Diskretisierung von (2.10) mit Finiten-Differenzen führt dann mit der Bezeichnung $y^\nu := y(x, \tau_\nu)$ auf die semidiskrete Form

$$(2.13) \quad \int_{x_{\min}}^{x_{\max}} \frac{y^{\nu+1} - y^\nu}{\Delta\tau} (v^{\nu+1} - y^{\nu+1}) + \left[\theta \frac{\partial y^{\nu+1}}{\partial x} + (1 - \theta) \frac{\partial y^\nu}{\partial x} \right] \frac{\partial (v^{\nu+1} - y^{\nu+1})}{\partial x} dx \\ \geq \ell(v^{\nu+1} - y^{\nu+1}) \quad \text{für alle } v \in \mathcal{K}^{\nu+1}.$$

Ein solches Schema, das zwischen einer expliziten und einer impliziten Darstellung interpoliert, wird als θ -Schema bezeichnet. Für $\theta = 0$ beschreibt die Zeitdiskretisierung in (2.13) ein *explizites Eulerverfahren*, für $\theta = \frac{1}{2}$ ein *Crank-Nicholson* und für $\theta = 1$ ein *implizites Eulerverfahren*.

Mit der Bilinearform

$$(2.14) \quad a(u, v) := \int_{x_{\min}}^{x_{\max}} u v + \theta \Delta\tau \frac{\partial u}{\partial x} \frac{\partial v}{\partial x} dx$$

und dem linearen Funktional

$$(2.15) \quad f(v) := \Delta\tau \ell(v) + \int_{x_{\min}}^{x_{\max}} y^\nu v + \left[(\theta - 1) \Delta\tau \frac{\partial y^\nu}{\partial x} \right] \frac{\partial v}{\partial x} dx$$

erhält man

Problem 2.3.1 (semidiskrete Form) *Finde $y^{\nu+1} \in \mathcal{K}^{\nu+1}$, so dass*

$$a(y^{\nu+1}, v^{\nu+1} - y^{\nu+1}) \geq f(v^{\nu+1} - y^{\nu+1}) \quad \text{für alle } v \in \mathcal{K}^{\nu+1}$$

mit

$$\mathcal{K}^\nu := \{v(\cdot, \tau_\nu) \in H_0^1(I_x) : v(x, \tau_\nu) \geq 0 \text{ für alle } x \in I_x\}.$$

Hierbei bezeichnet $H_0^1(I_x)$ den Raum der einmal schwach differenzierbaren Funktionen $v \in L^2(I_x)$ mit kompaktem Träger in I_x (s. Anhang A.2). Wegen Lemma A.2.6 ist die Bilinearform $a(\cdot, \cdot)$ aus (2.14) stetig, symmetrisch und positiv definit in $H^1(I_x)$ und elliptisch in $H_0^1(I_x)$. Nach dem Satz A.3.3 von Lax–Milgram existiert daher eine eindeutige Lösung von Problem 2.3.1.

Die elliptische Variationsungleichung aus Problem 2.3.1 lässt sich als ein Hindernisproblem auffassen. Vergleiche hierzu Anhang A.4. Die Hindernisfunktion ist in diesem Fall die Nullfunktion.

2.3.2 Diskretisierung im Ort

Sei $x_i := x_{\min} + i h_i$, $i = 1, \dots, M$ mit $M \in \mathbb{N}$ eine nicht notwendig äquidistante Diskretisierung des Intervalls $I_x := [x_{\min}, x_{\max}]$ zu vorgegebenen Schrittweiten h_i mit

$$\sum_{i=1}^{M-1} h_i = x_{\max} - x_{\min}.$$

Zu einer vorgegebenen Ordnung $k \in \mathbb{N}$ sei weiter

$$S_h := \mathcal{N}_{k, T_h} \subset H_0^1(I_x)$$

der endlich-dimensionale Raum der Splines der Ordnung k bezüglich der in (A.1) eingeführten erweiterten Knotenfolge $T_h := \{\theta_i\}_{i=1, \dots, n+k}$ zur Knotenfolge $\{x_i\}_{i=1, \dots, M}$.

Die Lösung $y^{\nu+1} \in H_0^1(I_x)$ des semidiskreten Problems 2.3.1 soll nun durch $y_h^{\nu+1} \in S_h$ approximiert werden. Ist $\{N_{i,k}\}_{i=1, \dots, n}$ die B-Spline-Basis der Ordnung k für den n -dimensionalen Raum S_h , dann gilt

$$n = M + k - 2$$

und $y_h^\nu \in S_h$ und $v_h^\nu \in S_h$ lassen sich darstellen als

$$(2.16) \quad y_h^\nu(x) = \sum_{i=1}^n w_i^\nu N_{i,k}(x), \quad v_h^\nu(x) = \sum_{i=1}^n v_i^\nu N_{i,k}(x)$$

mit eindeutigen Koeffizienten $w_i^\nu \in \mathbb{R}$ bzw. $v_i^\nu \in \mathbb{R}$ für $i = 1, \dots, n$.

Der diskrete Lösungsraum sei bezeichnet mit

$$(2.17) \quad \bar{\mathcal{K}}_h^\nu := \{v_h^\nu \in S_h : v_h^\nu(x) \geq 0 \text{ für alle } x \in I_x\} \subset H_0^1(I_x).$$

Es gelten die Definitionen

$$\mathbf{w}^\nu := (w_1^\nu, \dots, w_n^\nu)^T \in \mathbb{R}^n,$$

$$\mathbf{v}^\nu := (v_1^\nu, \dots, v_n^\nu)^T \in \mathbb{R}^n,$$

$$A = (A_{i,j}) \in \mathbb{R}^{n \times n} \text{ mit } A_{i,j} := \int_{x_{\min}}^{x_{\max}} N'_{i,k} N'_{j,k} dx \text{ die Steifigkeitsmatrix,}$$

$$B = (B_{i,j}) \in \mathbb{R}^{n \times n} \text{ mit } B_{i,j} := \int_{x_{\min}}^{x_{\max}} N_{i,k} N_{j,k} dx \text{ die Massenmatrix}$$

und es bezeichnet $N'_{i,k} := \partial N_{i,k} / \partial x$.

Einsetzen der Darstellung aus (2.16) in die Definition der Bilinearform $a(\cdot, \cdot)$ aus (2.14) liefert

$$\begin{aligned}
& a(y_h^{\nu+1}, v_h^{\nu+1} - y_h^{\nu+1}) \\
&= a\left(\sum_{i=1}^n w_i^{\nu+1} N_{i,k}, \sum_{j=1}^n (v_j^{\nu+1} - w_j^{\nu+1}) N_{j,k}\right) \\
(2.18) \quad &= \sum_{i,j=1}^n w_i^{\nu+1} (v_j^{\nu+1} - w_j^{\nu+1}) a(N_{i,k}, N_{j,k}) \\
&= \sum_{i,j=1}^n w_i^{\nu+1} (v_j^{\nu+1} - w_j^{\nu+1}) \int_{x_{\min}}^{x_{\max}} N_{i,k} N_{j,k} dx + \theta \Delta \tau \int_{x_{\min}}^{x_{\max}} N'_{i,k} N'_{j,k} dx \\
&= (\mathbf{v}^{\nu+1} - \mathbf{w}^{\nu+1})^T (B + \theta \Delta \tau A) \mathbf{w}^{\nu+1}.
\end{aligned}$$

Mit Einsetzen der Darstellung aus (2.16) in die Linearform ℓ aus (2.11) erhält man analog

$$\begin{aligned}
& \ell(v_h^{\nu+1} - y_h^{\nu+1}) \\
(2.19) \quad &= - \int_{x_{\min}}^{x_{\max}} \frac{\partial g}{\partial \tau} \sum_{j=1}^n (v_j^{\nu+1} - w_j^{\nu+1}) N_{j,k} + \frac{\partial g}{\partial x} \sum_{j=1}^n (v_j^{\nu+1} - w_j^{\nu+1}) N'_{j,k} dx \\
&= - \sum_{j=1}^n (v_j^{\nu+1} - w_j^{\nu+1}) \int_{x_{\min}}^{x_{\max}} \frac{\partial g}{\partial \tau} N_{j,k} + \frac{\partial g}{\partial x} N'_{j,k} dx \\
&= (\mathbf{v}^{\nu+1} - \mathbf{w}^{\nu+1})^T \mathbf{r}
\end{aligned}$$

mit $\mathbf{r} := (r_1, \dots, r_n)^T$ und

$$(2.20) \quad r_i := - \int_{x_{\min}}^{x_{\max}} \frac{\partial g}{\partial \tau} N_{i,k} + \frac{\partial g}{\partial x} N'_{i,k} dx.$$

Für den zweiten Teil der rechten Seite f aus (2.15) ergibt sich

$$\begin{aligned}
& \int_{x_{\min}}^{x_{\max}} y_h^{\nu} (v_h^{\nu+1} - y_h^{\nu+1}) + \left[(\theta - 1) \Delta \tau \frac{\partial y_h^{\nu}}{\partial x} \right] \frac{\partial (v_h^{\nu+1} - y_h^{\nu+1})}{\partial x} dx \\
(2.21) \quad &= \sum_{i,j=1}^n w_i^{\nu} (v_j^{\nu+1} - w_j^{\nu+1}) \left(\int_{x_{\min}}^{x_{\max}} N_{i,k} N_{j,k} dx + (\theta - 1) \Delta \tau \int_{x_{\min}}^{x_{\max}} N'_{i,k} N'_{j,k} dx \right) \\
&= (\mathbf{v}^{\nu+1} - \mathbf{w}^{\nu+1})^T (B + (\theta - 1) \Delta \tau A) \mathbf{w}^{\nu}.
\end{aligned}$$

Aus (2.18), (2.19) und (2.21) setzt sich die vollständig diskretisierte Form der semidiskreten Variationsungleichung aus Problem 2.3.1 in Matrixschreibweise zusammen als

$$\begin{aligned}
0 &\leq a(y_h^{\nu+1}, v_h^{\nu+1} - y_h^{\nu+1}) - f(v_h^{\nu+1} - y_h^{\nu+1}) \\
&= (\mathbf{v}^{\nu+1} - \mathbf{w}^{\nu+1})^T [(B + \theta \Delta \tau A) \mathbf{w}^{\nu+1} - \Delta \tau \mathbf{r} - (B + (\theta - 1) \Delta \tau A) \mathbf{w}^{\nu}]
\end{aligned}$$

oder mit den Abkürzungen

$$\begin{aligned}
C &:= B + \theta \Delta \tau A, \\
\mathbf{b} &:= \Delta \tau \mathbf{r} + (B + (\theta - 1) \Delta \tau A) \mathbf{w}^{\nu}
\end{aligned}$$

als

$$(2.22) \quad 0 \leq (\mathbf{v}^{\nu+1} - \mathbf{w}^{\nu+1})^T (C \mathbf{w}^{\nu+1} - \mathbf{b}).$$

Der konvexe Lösungsraum $\bar{\mathcal{K}}_h^\nu$ im ν -ten Zeitschritt aus (2.17) soll durch den leicht restriktiveren Raum

$$\mathcal{K}_h^\nu := \{v_h^\nu \in S_h : v_i^\nu \geq 0 \text{ für alle } i = 1, \dots, n\},$$

der einen Quader im \mathbb{R}^n beschreibt, ersetzt werden. Dabei sei erinnert, dass für feste ν die Darstellung aus (2.16) gilt. Die beiden diskreten Räume $\bar{\mathcal{K}}_h^\nu$ und \mathcal{K}_h^ν sind nur für stückweise lineare Funktionen äquivalent. Mit den Strukturaussagen aus Abschnitt 1.3 weiß man aber, dass die Lösung keine Minima in der Nähe der Hindernisfunktion aufweist und die diskrete Approximation damit nach Lemma 2.6.6 in der eingeschränkten Lösungsmenge \mathcal{K}_h^ν liegen muss.

Die Ergebnisse dieses Abschnittes lassen sich jetzt in Problem 2.3.2 mit $\mathbf{w} = \mathbf{w}^{\nu+1}$ und $\mathbf{v} = \mathbf{v}^{\nu+1}$ zusammenfassen.

Problem 2.3.2 (Diskrete Variationsungleichung) *Finde in jedem Zeitschritt ein $\mathbf{w} \in \mathbb{R}^n$ mit $\mathbf{w} \geq 0$, so dass*

$$(2.23) \quad (\mathbf{v} - \mathbf{w})^T (C \mathbf{w} - \mathbf{b}) \geq 0$$

für alle $\mathbf{v} \in \mathbb{R}^n$ mit $\mathbf{v} \geq 0$ gilt.

Man beachte, dass im Hinblick auf Problem 2.2.6 die Hindernisfunktion Null ist. Mit dem Satz 5.4. aus [Se], Seite 155 kann die diskrete Variationsungleichung in ein diskretes linear komplementäres Problem umformuliert werden.

Satz 2.3.3 *Die Lösung von Problem 2.3.2 ist äquivalent zur Lösung von Problem 2.3.4.*

Problem 2.3.4 (Diskretes lineares komplementäres Problem) *Finde in jedem Zeitschritt ein $\mathbf{w} \in \mathbb{R}^n$, so dass*

$$(2.24) \quad \begin{aligned} (\mathbf{w})^T (C \mathbf{w} - \mathbf{b}) &= 0, \\ C \mathbf{w} - \mathbf{b} &\geq 0, \quad \mathbf{w} \geq 0 \end{aligned}$$

gilt.

Die Ungleichungen sind dabei komponentenweise aufzufassen.

Die Existenz einer eindeutigen Lösung folgt aus Satz A.3.3 mit der Bilinearform $a(\mathbf{w}, \mathbf{v}) := \mathbf{w}^T C \mathbf{v}$, der Linearform $f(\mathbf{v}) := \mathbf{b}^T \mathbf{v}$ und der konvexen Menge $\mathcal{K} := \{\mathbf{v} \in \mathbb{R}^n : \mathbf{v} \geq 0\}$.

Die Lösung von Problem 2.3.4 kann nun iterativ mit dem *projektiven Gauss–Seidel–Verfahren* bestimmt werden. Dieses ist eine Variante des klassischen Gauss–Seidel–Verfahrens. Sie wird in Abschnitt 2.6 beschrieben.

Zum unrestringiertes Randwertproblem

Für das leichtere Randwertproblem 2.2.4 erhält man die schwache Form durch Multiplikation mit einer Testfunktion $v \in H_0^1(\Omega)$ und anschließender partieller Integration. Diese kann ebenfalls durch Subtraktion der Funktion

$$\tilde{g}(x, \tau) := e^{\frac{1}{4}(q-1)^2\tau} \max \left\{ e^{\frac{x}{2}(q+1)+q\tau} - e^{\frac{x}{2}(q-1)}, 0 \right\}$$

auf homogene Randbedingungen reduziert werden. Solange $x_{\min} < -q\tau$ erfüllt ist, gilt $\tilde{g}(x_{\min}, \tau) = 0$ und offensichtlich ist $\tilde{g}(x_{\max}, \tau) = r_c$. Da außerdem $\tilde{g}(x, 0) = u(x, 0)$ ist, erhält man auch hier anstelle nicht-glatte Anfangsdaten verschwindende Anfangsdaten. Diskretisierung in der Zeit liefert das semidiskrete Problem

$$(2.25) \quad \text{finde } u \in H_0^1(I_x) : \quad a(u, v) = f(v) \quad \text{für alle } v \in H_0^1(I_x).$$

Diskretisierung im Ort mit Finite-Elementen führt mit den gleichen Bezeichnungen wie oben zur vollständig diskreten Form des unrestringierten Randwertproblems 2.2.4.

Problem 2.3.5 (Diskretes unrestringiertes Randwertproblem) *Finde in jedem Zeitschritt ein $\mathbf{w} \in \mathbb{R}^n$, so dass*

$$(2.26) \quad C \mathbf{w} - \mathbf{b} = 0$$

gilt.

In jedem Zeitschritt ist ein lineares Gleichungssystem zu lösen. In dem bisher behandelten eindimensionalen Fall ist dies mit LR-Zerlegung in $O(n)$ Operationen möglich.

Bemerkung 2.3.6 Es sind auch Kombinationen von Problem 2.3.4 und Problem 2.3.5 denkbar in dem Sinn, dass an gewissen Zeitpunkten das lineare komplementäre Problem aus (2.24) gilt, während zu allen anderen Zeiten das Gleichungssystem aus (2.26) erfüllt sein muss. Solche Probleme, die bei der Bewertung von sogenannten *Bermuda Optionen* auftreten, können in dem vorgestellten Rahmen ebenfalls ohne Schwierigkeiten und mit nur minimalen Veränderungen am Lösungsalgorithmus behandelt werden.

2.4 Fehlerabschätzungen

In diesem Abschnitt sollen Fehlerabschätzungen für die semidiskrete Form

$$(2.27) \quad \text{finde } u \in H_0^1 : \quad a(u, v) = f(v) \quad \text{für alle } v \in H_0^1$$

(s. (2.25)) des unrestringierten Randwertproblems 2.2.4 und für die semidiskrete Form

$$(2.28) \quad \text{finde } u \in \mathcal{K} : \quad a(u, v - u) \geq f(v - u) \quad \text{für alle } v \in \mathcal{K}$$

mit der konvexen Menge

$$\mathcal{K} := \{v \in H_0^1 : v(x) \geq 0 \text{ für alle } x \in I_x\}$$

(s. Problem 2.3.1) des freien Randwertproblems 2.2.3 hergeleitet werden. Es sei daran erinnert, dass die Bilinearform $a(\cdot, \cdot)$ aus (2.14) nach dem Lemma A.2.6 stetig, symmetrisch, positiv definit und H_0^1 -elliptisch ist. Der Abschnitt 2.4.1 behandelt Fehlerabschätzungen für Variationsgleichungen und ist im Wesentlichen den Büchern [Br] und [Th] entnommen. Die Resultate aus Abschnitt 2.4.2 zu Hindernisproblemen stammen aus [BHR]. Bei den Abschätzungen wird im folgenden anstelle von $x \leq cy$ mit einer generischen Konstante c , die unabhängig von der Gitterweite h ist, die Schreibweise $x \lesssim y$ verwendet.

2.4.1 Elliptische Variationsgleichungen

Zunächst sei die Variationsgleichung aus (2.27) betrachtet. Die diskrete Form lautet

$$(2.29) \quad \text{finde } u_h \in S_h : a(u_h, v_h) = f(v_h) \quad \text{für alle } v_h \in S_h.$$

Grundlegend für die Fehlerabschätzung der Finite-Elemente Näherung u_h ist das folgende Lemma.

Lemma 2.4.1 (Céa) *Die Bilinearform $a(\cdot, \cdot)$ sei elliptisch in H_0^m und stetig. Dann ist*

$$\|u - u_h\|_m \lesssim \inf_{v_h \in S_h} \|u - v_h\|_m.$$

Beweis: Subtrahieren von $a(u, v) = f(v)$ für alle $v \in H_0^1$ und von $a(u_h, v_h) = f(v_h)$ für alle $v_h \in S_h$ liefert wegen $S_h \subset H_0^1$ die Galerkin-Orthogonalität

$$a(u - u_h, v_h) = 0 \quad \text{für alle } v_h \in S_h.$$

Aus der Elliptizität von $a(\cdot, \cdot)$, der Galerkin-Orthogonalität und der Stetigkeit von $a(\cdot, \cdot)$ folgt

$$\begin{aligned} \alpha \|u - u_h\|_m^2 &\leq a(u - u_h, u - u_h) \\ &= a(u - u_h, u - v_h) + a(u - u_h, v_h - u_h) \\ &\lesssim \|u - u_h\|_m \|u - v_h\|_m \end{aligned}$$

und nach Kürzen und Dividieren durch α die Behauptung. \square

Die Genauigkeit der Lösung hängt also wesentlich davon ab, wie gut die Lösung u in den diskreten Funktionenräumen S_h approximiert werden kann. Wählt man stückweise Polynome der Ordnung k zu einem äquidistanten Gitter mit Gitterweite h als Ansatzfunktionen und ist die Lösung u genügend regulär, dann lässt sich die Fehlerordnung

$$\|u - u_h\|_1 \leq O(h^{k-1})$$

nachweisen. Der Beweis beruht auf folgender Interpolationsabschätzung, die der Einfachheit halber nur für den Fall einer uniformen Intervallunterteilung formuliert wird. Für quasiuniforme Zerlegungen des Gebietes siehe [Br], S.75.

Satz 2.4.2 (Approximationssatz) *Es bezeichne $I_h : H^k \rightarrow S_h$ die Interpolation durch stückweise Polynome vom Grad $k - 1$. Dann gilt*

$$\|u - I_h u\|_{m,h} \lesssim h^{k-m} |u|_k \quad \text{für } u \in H^k, \quad 0 \leq m \leq k.$$

Die Definitionen der diskreten Sobolevnorm $\|\cdot\|_{m,h}$ und der Semimorm $|\cdot|_m$ für $m \in \mathbb{N}$ finden sich in Anhang A.2.

Beweis: Der Beweis beruht auf dem Lemma von Bramble-Hilbert. Vergleiche [Br] S.75. \square

Aus dem Lemma von Céa und aus dem Approximationssatz folgt jetzt die Fehlerabschätzung für das elliptische Variationsproblem aus (2.27), falls das Problem genügend regulär ist.

Definition 2.4.3 (Regularität) *Das Variationsproblem aus (2.27) heißt H^s -regulär, wenn es zu jedem $f \in H^{s-2}$ eine Lösung $u \in H^s$ gibt und*

$$\|u\|_s \lesssim \|f\|_{s-2}$$

gilt.

Für die Variationsgleichung aus (2.27) ist bekannt, dass das Problem H^2 – regulär ist, falls das Gebiet konvex ist, dass es H^s – regulär ist, wenn das Gebiet einen C^s Rand besitzt (s. [Br], S.85) und dass die Lösung im Inneren des Gebietes unendlich glatt ist (s. [Ev], S.59).

Satz 2.4.4 (Fehlerabschätzung in der Energienorm) *Das in (2.27) formulierte Variationsproblem sei H^k – regulär mit einer elliptischen und stetigen Bilinearform $a(\cdot, \cdot)$ und es sei $S_h \subset H_0^1$ der Raum der stückweisen Polynome der Ordnung k . Dann gilt für die Finite-Elemente Näherung $u_h \in S_h$ die Fehlerabschätzung*

$$\|u - u_h\|_1 = O(h^{k-1}).$$

Beweis: Aus dem Lemma von Céa, wegen $I_h \in S_h$, dem Approximationssatz und der H^k – Regularität folgt

$$\begin{aligned} \|u - u_h\|_1 &\lesssim \inf_{v_h \in S_h} \|u - v_h\|_1 \\ &\lesssim \|u - I_h u\|_1 \\ &\lesssim h^{k-1} \|u\|_k \\ &\lesssim h^{k-1} \|f\|_{k-2}. \end{aligned}$$

□

Die Fehlerabschätzung

$$\|u - u_h\|_0 = O(h^k)$$

in der L^2 – Norm lässt sich hieraus mit der Dualitätstechnik des Aubin–Nitsche Lemmas (s. [Br], S.87) ableiten. In der L^∞ – Norm erhält man mit einer inversen Abschätzung und dem Lemma vom Bramble–Hilbert (s. [Br], S.89)

$$\|u - u_h\|_\infty \lesssim h|u|_2$$

bzw. mit Hilfe gewichteter Normen die wesentlich stärkere Aussage

$$\|u - u_h\|_\infty \lesssim h^2 |\log h|^{3/2} \|D^2 u\|_\infty.$$

Wird das lineare Gleichungssystem nur näherungsweise z.B. bei Verwendung numerischer Quadratur aufgestellt, benötigt man eine Verallgemeinerung des Lemmas von Céa, um die zusätzlichen Fehler abzuschätzen.

Lemma 2.4.5 (Strang) *Sei anstelle von (2.29) das gestörte Variationsproblem*

$$(2.30) \quad \text{finde } u_h \in S_h : \quad a_h(u_h, v_h) = f_h(v_h) \quad \text{für alle } v_h \in S_h$$

mit einer stetigen und elliptischen Bilinearform $a_h(\cdot, \cdot)$ gegeben, dann gilt

$$\|u - u_h\|_m \leq c \inf_{v_h \in S_h} \left(\|u - v_h\|_m + \sup_{w_h \in S_h} \frac{|a(v_h, w_h) - a_h(v_h, w_h)|}{\|w_h\|_m} \right) + c \sup_{w_h \in S_h} \frac{|f(w_h) - f_h(w_h)|}{\|w_h\|_m}.$$

Wird z.B. die Bilinearform exakt aber die rechte Seite nur näherungsweise berechnet, dann folgt, dass eine Quadratur–Genauigkeit von $\|f - f_h\|_m = O(h^{k-m})$ ausreichend ist. Mit der Cauchy–Schwarz Ungleichung kann der entsprechende Term aus Lemma 2.4.5 durch

$$\sup_{w_h \in S_h} \frac{|f(w_h) - f_h(w_h)|}{\|w_h\|_m} \leq \|f - f_h\|_m$$

abgeschätzt werden.

Betrachtet man das ursprüngliche *parabolische* Problem

$$(2.31) \quad \text{finde } u \in H_0^1 : \left\langle \frac{\partial u}{\partial \tau}, v \right\rangle + a(u, v) = f(v) \quad \text{für alle } v \in H_0^1,$$

dann setzen sich die Fehlerabschätzungen aus den Fehlern zusammen, die bei der Finite-Elemente Diskretisierung im Ort und aus denen, die bei der θ -Diskretisierung der Zeit entstehen.

Für das implizite Eulerverfahren ($\theta = 1$) erhält man die Fehlerordnung $O(h^k + \Delta\tau)$ (s. [Th], S.13) und für das Crank-Nicholson Verfahrens ($\theta = \frac{1}{2}$) die Fehlerordnung $O(h^k + (\Delta\tau)^2)$ (s. [Th], S.15) jeweils unter der Annahme, dass die Lösung u von (2.31) genügend glatt ist. Hierbei bezeichnet $k \in \mathbb{N}$ wie bisher die Ordnung der Ansatzfunktionen im Ort und $\Delta\tau \in \mathbb{R}^+$ die Zeitschrittweite. Die Konvergenzordnungen bleiben auch dann erhalten, wenn die Gleichungssysteme in jedem Zeitschritt nicht mehr exakt, sondern nur noch iterativ z.B. mit Mehrgitterverfahren gelöst werden, solange dies ausreichend genau geschieht (s. Lemma 2.4.5 und [Th], Satz 11.1 und Satz 11.4).

2.4.2 Elliptische Variationsungleichungen

Die diskrete Form der Variationsungleichung aus (2.28) lautet

$$(2.32) \quad \text{finde } u_h \in \mathcal{K}_h : \quad a(u_h, v_h - u_h) \geq f(v_h - u_h) \quad \text{für alle } v_h \in \mathcal{K}_h$$

mit der konvexen Menge

$$\mathcal{K}_h := \{v_h \in S_h : v_h(x) \geq 0 \text{ für alle } x \in I_x\}.$$

Zunächst soll eine Verallgemeinerung des Lemmas 2.33 von Céa bewiesen werden. Dazu sei $A : H_0^1 \rightarrow H^{-1}$ der Riesz-Operator, der $\langle Au, v \rangle = a(u, v)$ für alle $v \in H_0^1$ erfüllt. Wie bisher bezeichnet S_h den Raum der B-Splines von vorgegebener Ordnung k .

Lemma 2.4.6 *Zu einer stetigen, H_0^m -elliptischen Bilinearform $a(\cdot, \cdot)$ seien $u \in H_0^1$ bzw. $u_h \in S_h$ die Lösung von (2.28) bzw. (2.32). Dann ist*

$$\|u - u_h\|_m \lesssim \inf_{v_h \in S_h} \left(\|u - v_h\|_m^2 + \langle Au - f, v_h - u \rangle \right)^{\frac{1}{2}}.$$

Beweis: Wie im Lemma von Céa folgt

$$(2.33) \quad \alpha \|u - u_h\|_m^2 \leq a(u - u_h, u - v_h) + a(u - u_h, v_h - u_h).$$

Der zweite Term verschwindet hier allerdings nicht. Wegen (2.32) lässt er sich aber nach oben abschätzen.

$$\begin{aligned} a(u - u_h, v_h - u_h) &\leq a(u, v_h - u_h) - f(v_h - u_h) \\ &= \langle Au - f, v_h - u_h \rangle \\ &= \langle Au - f, v_h - u \rangle + \langle Au - f, u \rangle + \langle Au - f, -u_h \rangle. \end{aligned}$$

Der zweite Term verschwindet und der dritte Term kann durch Null nach oben abgeschätzt werden. Ein Beweis hierfür findet sich im Anhang in Lemma A.4.6. Man beachte, dass die

Hindernisfunktion hier Null ist und, dass bei der Anwendung des Lemmas die Positivität der B-Splines eingeht. Damit ergibt sich zusammen mit (2.33) die Ungleichung

$$\alpha \|u - u_h\|_m^2 \leq a(u - u_h, u - v_h) + \langle Au - f, v_h - u \rangle.$$

Die Stetigkeit $a(u, v) \leq c \|u\|_m \|v\|_m$ und die Binomische Formel $ab \leq \frac{1}{2}a^2 + \frac{1}{2}b^2$ führen zu

$$\left(\alpha - \frac{c}{2}\right) \|u - u_h\|_m^2 \leq \frac{c}{2} \|u - v_h\|_m^2 + \langle Au - f, v_h - u \rangle.$$

Teilen durch $(\alpha - \frac{c}{2})$ und Wurzelziehen führen dann zur Behauptung. Der Beweis ist an [EO] angelehnt. \square

Bemerkung 2.4.7 Die Regularität $u \in H^2$ wurde für das hier relevante Problem bereits in Bemerkung 1.2.4 nachgewiesen. Für allgemeine Hindernisprobleme wie in (2.28), liefert eine genauere Analyse von F. Brezzi [BHR] $u \in H^{5/2-\epsilon}$ für ein beliebiges $\epsilon > 0$, falls $f \in H^1 \cap L^\infty$.

Satz 2.4.8 (Fehlerabschätzung für lineare Elemente) *Zu $f \in L^2$ sei $u \in H^2$ die Lösung der Variationsungleichung aus (2.28). Weiter sei $S_h \subset H_0^1$ speziell der Raum der stetigen stückweise linearen Splines. Dann gilt für die Finite-Elemente Näherung $u_h \in S_h$ aus (2.32) die Fehlerabschätzung*

$$\|u - u_h\|_1 = O(h).$$

Beweis: Nach der Cauchy-Schwarz Ungleichung folgt

$$(2.34) \quad \langle Au - f, v_h - u \rangle \leq \|Au - f\|_0 \|v_h - u\|_0 \lesssim \|I_h u - u\|_0$$

und Anwendung des Approximationssatz 2.4.2 liefert

$$(2.35) \quad \|I_h u - u\|_0 \lesssim h^2 \|u\|_2 \quad \text{und} \quad \|I_h u - u\|_1 \lesssim h \|u\|_2.$$

Aus Lemma 2.4.6, (2.34) und (2.35) folgt nun

$$\begin{aligned} \|u - u_h\|_1^2 &\lesssim \inf_{v_h \in S_h} (\|u - v_h\|_1^2 + \langle Au - f, v_h - u \rangle) \\ &\lesssim (\|u - I_h u\|_1^2 + \|I_h u - u\|_0) \\ &\lesssim (h^2 \|u\|_2^2 + h^2 \|u\|_2) \\ &= O(h^2). \end{aligned}$$

Der Beweis ist analog zu [BHR] geführt. \square

Satz 2.4.9 (Fehlerabschätzung für quadratische Elemente) *Zu $f \in H^1 \cap L^\infty$ sei $u \in H^{5/2-\epsilon}$ die Lösung der Variationsungleichung aus (2.28). Ist $S_h \subset H_0^1$ der Raum der stückweise quadratischen B-Splines, dann gilt für die Finite-Elemente Näherung $u_h \in S_h$ aus (2.32) die Fehlerabschätzung*

$$\|u_h - u\|_1 = O(h^{3/2-\epsilon}).$$

Beweis: Der Approximationssatz 2.4.2 und Bemerkung 2.4.7 implizieren

$$(2.36) \quad \|I_h u - u\|_0 \lesssim h^{5/2-\epsilon} \|u\|_{5/2-\epsilon} \quad \text{und} \quad \|I_h u - u\|_1 \lesssim h^{3/2-\epsilon} \|u\|_{5/2-\epsilon}.$$

Aus Lemma 2.4.6 und den Abschätzungen (2.34) und (2.36) folgt wie oben

$$\begin{aligned}
\|u - u_h\|_1^2 &\lesssim \inf_{v_h \in S_h} (\|u - v_h\|_1^2 + \langle Au - f, v_h - u \rangle) \\
&\lesssim (\|u - I_h u\|_1^2 + \|I_h u - u\|_0) \\
&\lesssim (h^{9/4-\epsilon} \|u\|_{5/2-\epsilon}^2 + h^{5/2-\epsilon} \|u\|_{5/2-\epsilon}) \\
&= O(h^{9/4-\epsilon}).
\end{aligned}$$

Wurzelziehen liefert schließlich $\|u_h - u\|_1 = O(h^{3/2-\epsilon})$. \square

Bemerkung 2.4.10 Werden anstelle stückweise quadratischer B-Splines stückweise quadratische durch Interpolation bestimmte Finite-Elemente Basisfunktionen verwendet, dann sind wegen der fehlenden Positivitätseigenschaft Lemma A.4.6 und damit auch Lemma 2.4.6 nicht anwendbar. In diesem Fall ist der Beweis deutlich komplizierter. Er findet sich in [BHR].

Bemerkung 2.4.11 Die beiden Fehlerabschätzungen sind optimal für die betrachteten Räume und angegebenen Regularitätsannahmen. Die Frage, ob eine optimale Abschätzung in der L^2 -Norm mittels einer Verallgemeinerung des Lemmas von Aubin-Nitsche möglich ist, ist noch offen. In der L^∞ -Norm gelten unter geeigneten Annahmen an die Regularität der Daten die Abschätzungen

$$\|u_h - u\|_\infty = O(h^{2-\epsilon})$$

und

$$\|u_h - u\|_\infty = O(h^2 |\log h|)$$

die auf [Bc] und [N] zurückgehen.

2.5 Implementierung des Verfahrens

2.5.1 Aufstellen von Massen-, Steifigkeitsmatrix und der rechten Seite

Im Fall von äquidistanten Gittern lassen sich die Einträge der Steifigkeitsmatrix A und der Massenmatrix B explizit angeben.

Lemma 2.5.1 Für $i = 1, \dots, n$ sei $N_{i,k}$ der i -te Basis-B-Spline der Ordnung k zu einem äquidistanten Gitter mit Gitterweite h . Dann gilt:

$$\begin{aligned}
B_{i,j} &:= \int_{\mathbb{R}} N_{i,k}(x) N_{j,k}(x) dx = h N_{0,2k-1}(k+i-j) =: h s_{i-j}^k, \\
A_{i,j} &:= \int_{\mathbb{R}} N'_{i,k}(x) N'_{j,k}(x) dx = \frac{1}{h} \left(2s_{i-j}^{k-1} - s_{i-j-1}^{k-1} - s_{i-j+1}^{k-1} \right) =: \frac{1}{h} d_{i-j}^k.
\end{aligned}$$

Beweis: Der Beweis findet sich in [Hö] auf Seite S.35. \square

Für lineare ($k = 2$), quadratische ($k = 3$) bzw. kubische ($k = 4$) B-Splines sind die Matrixeinträge in Tabelle 2.1 aufgeführt.

k	s_0^k	s_1^k	s_2^k	s_3^k	d_0^k	d_1^k	d_2^k	d_3^k
2	$\frac{2}{3}$	$\frac{1}{6}$			2	-1		
3	$\frac{11}{20}$	$\frac{13}{60}$	$\frac{1}{120}$		1	$-\frac{1}{3}$	$-\frac{1}{6}$	
4	$\frac{151}{315}$	$\frac{387}{1680}$	$\frac{1}{42}$	$\frac{1}{5040}$	$\frac{2}{3}$	$-\frac{1}{8}$	$-\frac{1}{5}$	$\frac{1}{120}$

Tabelle 2.1: Einträge der Massen- und Steifigkeitsmatrix.

Für den Fall $k = 2$ und nicht äquidistanter Gitter gilt (s. [Se], Seite 150)

$$A_{i,j} = \begin{cases} \frac{1}{h_i} + \frac{1}{h_{i+1}} & , \text{ falls } j = i \\ -\frac{1}{h_{i+1}} & , \text{ falls } j = i + 1 \\ -\frac{1}{h_i} & , \text{ falls } j = i - 1 \\ 0 & , \text{ sonst,} \end{cases} \quad B_{i,j} = \begin{cases} \frac{1}{3}(h_i + h_{i+1}) & , \text{ falls } j = i \\ -\frac{1}{6}h_{i+1} & , \text{ falls } j = i + 1 \\ -\frac{1}{6}h_i & , \text{ falls } j = i - 1 \\ 0 & , \text{ sonst.} \end{cases}$$

Alternativ kann natürlich auch numerische Quadratur zur näherungsweisen Berechnung der Matrixeinträge verwendet werden. Lemma 2.4.5 gibt hierbei die benötigte Genauigkeit $O(h^k)$ der Quadratur vor.

Berechnung der rechten Seite

Im Fall äquidistanter Gitter und der Funktion g aus (2.6) lässt sich die diskrete rechte Seite \mathbf{r} aus (2.20) explizit angeben. Es sei erinnert, dass die erweiterte Knotenfolge mit $T_h := \{\theta_i\}_{i=1, \dots, n+k}$ bezeichnet ist.

Lemma 2.5.2 *Es sei $g(x, \tau) := s(\tau) \max\{e^{\frac{x}{2}(q-1)} - e^{\frac{x}{2}(q+1)}, 0\}$ die Funktion aus (2.6). Dann gilt für die i -te Komponente der rechten Seite*

$$(2.37) \quad \begin{aligned} r_i &:= - \int_{x_{min}}^{x_{max}} \frac{\partial g}{\partial \tau} N_{i,k} + \frac{\partial g}{\partial x} N'_{i,k} dx \\ &= -s(\tau) \left(\frac{1}{4}(q+1)^2 [\alpha_{i,k} + (-1)^{k-1} a_{i,k}] + \beta_{i,k} + (-1)^{k-2} b_{i,k} \right) \end{aligned}$$

mit der Funktion

$$s(\tau) := \exp\left(\frac{1}{4}(q+1)^2 \tau\right),$$

den Konstanten

$$\begin{aligned} \alpha_{i,k} &:= \sum_{j=1}^{k-1} (-1)^{j+1} e_j(0) N_{i,k}^{(j-1)}(0), \\ \beta_{i,k} &:= \sum_{j=1}^{k-3} (-1)^j e_j(0) N_{i,k}^{(j+1)}(0), \\ a_{i,k} &= a_{1,k}(\theta_i, \theta_{i+1}, \dots, \theta_{i+k}) := \frac{1}{h^{k-1}} \sum_{j=0}^k (-1)^{j+1} \binom{k}{j} e_k(\theta_{i+j} \wedge 0), \\ b_{i,k} &= b_{2,k}(\theta_i, \theta_{i+1}, \dots, \theta_{i+k}) := \frac{1}{h^{k-1}} \sum_{j=0}^k (-1)^{j+1} \binom{k}{j} e_{k-2}(\theta_{i+j} \wedge 0) \end{aligned}$$

und der Funktion

$$e_j(x) = \left(\frac{2}{q-1}\right)^j e^{\frac{x}{2}(q-1)} - \left(\frac{2}{q+1}\right)^j e^{\frac{x}{2}(q+1)} \quad \text{für } j \in \mathbb{Z}.$$

Dabei bezeichnet $a \wedge b := \min\{a, b\}$.

Beweis: Wegen $g(x, \tau) = 0$ für $x > 0$ ist

$$r_i = - \int_{x_{\min}}^0 \frac{\partial}{\partial \tau} (s(\tau) e_0(x)) N_{i,k}(x) + \frac{\partial}{\partial x} (s(\tau) e_0(x)) N'_{i,k}(x) dx.$$

Die Funktionen e_j sind gerade so definiert, dass

$$\frac{\partial}{\partial x} e_j(x) = e_{j-1}(x) \quad \text{für } j \in \mathbb{Z}.$$

Daher gilt

$$r_i = -s(\tau) \left(\frac{1}{4}(q+1)^2 \int_{x_{\min}}^0 e_0 N_{i,k} dx + \int_{x_{\min}}^0 e_{-1} N'_{i,k} dx \right).$$

Integriert man das erste Integral $(k-1)$ -mal und das zweite Integral $(k-2)$ -mal partiell, so erhält man

$$(2.38) \quad \begin{aligned} r_i = & -s(\tau) \left(\frac{1}{4}(q+1)^2 \left[\alpha_{i,k} + (-1)^{k-1} \int_{x_{\min}}^0 e_{k-1} N_{i,k}^{(k-1)} dx \right] \right. \\ & \left. + \beta_{i,k} + (-1)^{k-2} \int_{x_{\min}}^0 e_{k-3} N_{i,k}^{(k-1)} dx \right). \end{aligned}$$

Wegen $N_{i,k} \in H_0^{k-1}(I_x)$ mit dem Intervall $I_x := [x_{\min}, x_{\max}]$ sind die Integrale wohldefiniert und die Randterme am linken Rand $x = x_{\min}$ gleich Null. Die Randterme am rechten Rand $x = 0$ sind mit $\alpha_{i,k}$ bzw. $\beta_{i,k}$ abgekürzt. Mit $(k-1)$ -facher Anwendung der Rekursion aus (A.4) folgt, dass

$$N_{i,k}^{(k-1)} = h^{1-k} \sum_{j=0}^{k-1} (-1)^j \binom{k-1}{j} N_{i+j,1}$$

und wegen

$$N_{i+j,1} = \chi_{[\theta_{i+j}, \theta_{i+j+1})} := \begin{cases} 1 & \text{falls } x \in [\theta_{i+j}, \theta_{i+j+1}) \\ 0 & \text{sonst} \end{cases}$$

folgt, dass

$$\begin{aligned} \int_{x_{\min}}^0 e_{k-1} N_{i,k}^{(k-1)} dx &= h^{1-k} \sum_{j=0}^{k-1} (-1)^j \binom{k-1}{j} \int_{\theta_{i+j} \wedge 0}^{\theta_{i+j+1} \wedge 0} e_{k-1} dx \\ &= h^{1-k} \sum_{j=0}^{k-1} (-1)^j \binom{k-1}{j} (e_k(\theta_{i+j+1} \wedge 0) - e_k(\theta_{i+j} \wedge 0)) \\ &= h^{1-k} \sum_{j=0}^k (-1)^{j+1} \binom{k}{j} e_k(\theta_{i+j} \wedge 0) = a_{i,k}. \end{aligned}$$

Die im letzten Schritt benötigte Identität

$$\sum_{j=0}^{k-1} (-1)^j \binom{k-1}{j} (x_{j+1} - x_j) = \sum_{j=0}^k (-1)^{j+1} \binom{k}{j} x_j$$

für $k \in \mathbb{N}$ und $x_j \in \mathbb{R}$ lässt sich leicht mit Induktion über k nachweisen. Analog gilt für das zweite Integral

$$\int_{x_{\min}}^0 e_{k-3} N_{i,k}^{(k-1)} dx = b_{i,k}.$$

Einsetzen in (2.38) liefert schließlich die Behauptung. \square

Bemerkung 2.5.3 Für kleine k sind auf diese Weise nur sehr wenige Funktionsauswertungen zur Berechnung der Einträge r_i der rechten Seite aus (2.37) notwendig. Die Funktionswerte $e_k(\theta_i)$ bzw. $e_{k-2}(\theta_i)$ für $i = 1, \dots, n+k$ und die Werte $e_j(0)$ für $j = 1, \dots, k-1$ müssen nur einmal am Programmanfang berechnet werden und die Konstanten $\alpha_{i,k}$ bzw. $\beta_{i,k}$ müssen nur berechnet werden, falls Null im Inneren des Trägers von $N_{i,k}$ liegt. So lässt sich im Falle $k = 3$ der Eintrag r_i mit nur zwei Funktionsauswertungen berechnen. Bei $k = 4$ sind vier Funktionsauswertungen notwendig.

Berechnung der rechten Seite bei nicht äquidistanten Gittern

Bei nicht äquidistanten Gittern kann die Funktion $g(x, \tau_\nu) \in H^1$ zum Zeitpunkt τ_ν durch eine Funktion $I_h g^\nu =: g_h^\nu \in S_h$ approximiert werden. Stellt man diese in der B-Spline-Basis mit dem Koeffizientenvektor $\mathbf{g}^\nu := (g_1^\nu, \dots, g_n^\nu)^T$ als

$$(2.39) \quad g_h^\nu(x) = \sum_{i=1}^n g_i^\nu N_{i,k}(x)$$

dar, erhält man vollkommen analog und mit den gleichen Bezeichnungen wie in Abschnitt 2.3 die diskrete rechte Seite

$$(2.40) \quad \mathbf{r} = (C \mathbf{g}^{\nu+1} - (B + (\theta - 1)\Delta\tau A) \mathbf{g}^\nu) / \Delta\tau.$$

Auf diese Weise ist die rechte Seite mit zwei Multiplikationen dünnbesetzter Matrizen berechenbar. Mit Lemma 2.4.5 folgt, dass dieser zusätzliche Approximationsfehler keinen Einfluss auf die Fehlerordnung hat. Alternativ kann natürlich auch numerische Integration zur Bestimmung der rechten Seite verwendet werden.

2.5.2 Bestimmung der Ableitungen und des freien Randes

Ein Vorteil bei der Verwendung glatter Basisfunktionen im Finite-Elemente-Ansatz ist, dass die Ableitungen der Lösung direkt durch Ableiten der Basisfunktionen und damit wesentlich genauer bestimmt werden können.

Lemma 2.5.4 *Es sei $k \geq 3$ und für festes τ sei $u(x, \tau)$ die Finite-Elemente Lösung des transformierten Problems 2.2.3 oder 2.2.4. Dann gilt für die erste Ableitung der Lösung $V(S, t)$ des ursprünglichen Problems 2.1.1 bzw. 2.1.2*

$$\frac{\partial V(S, t)}{\partial S} = e^{-\frac{\pi}{2}(q+1) - (\frac{1}{4}(q-1)^2 + q)\tau} \left(\frac{\partial u(x, \tau)}{\partial x} - \frac{1}{2}(q-1)u(x, \tau) \right).$$

Falls $k \geq 4$ ist, dann gilt außerdem für die zweite Ableitung

$$\frac{\partial^2 V(S, t)}{\partial S^2} = e^{-\frac{\pi}{2}(q+3) - (\frac{1}{4}(q-1)^2 + q)\tau} \left(\frac{\partial^2 u(x, \tau)}{\partial x^2} - q \frac{\partial u(x, \tau)}{\partial x} + \frac{1}{4}(q^2 - 1)u(x, \tau) \right) / K.$$

Beweis: Unter Ausnutzung der Identitäten aus (2.3) und aus (2.4) und wegen $\partial S/\partial x = Ke^x$ gilt

$$\begin{aligned} \frac{\partial V(S, t)}{\partial S} &= \frac{\partial V(S, t)}{\partial x} \frac{1}{Ke^x} \\ &= Ke^{-\frac{x}{2}(q-1) - (\frac{1}{4}(q-1)^2 + q)\tau} \left(\frac{\partial u(x, \tau)}{\partial x} - \frac{1}{2}(q-1)u(x, \tau) \right) \frac{1}{Ke^x} \\ &= e^{-\frac{x}{2}(q+1) - (\frac{1}{4}(q-1)^2 + q)\tau} \left(\frac{\partial u(x, \tau)}{\partial x} - \frac{1}{2}(q-1)u(x, \tau) \right). \end{aligned}$$

Falls $k \geq 4$ ist, erhält man durch erneutes Ableiten

$$\begin{aligned} \frac{\partial^2 V(S, t)}{\partial S^2} &= \frac{\partial}{\partial S} \left(\frac{\partial V(S, t)}{\partial S} \right) \\ &= \frac{\partial}{\partial x} \left[e^{-\frac{x}{2}(q+1) - (\frac{1}{4}(q-1)^2 + q)\tau} \left(\frac{\partial u(x, \tau)}{\partial x} - \frac{1}{2}(q-1)u(x, \tau) \right) \right] \frac{1}{Ke^x} \\ &= e^{-\frac{x}{2}(q+1) - (\frac{1}{4}(q-1)^2 + q)\tau} \left(\frac{\partial^2 u(x, \tau)}{\partial x^2} - q \frac{\partial u(x, \tau)}{\partial x} + \frac{1}{4}(q^2 - 1)u(x, \tau) \right) \frac{1}{Ke^x} \\ &= \frac{1}{K} e^{-\frac{x}{2}(q+3) - (\frac{1}{4}(q-1)^2 + q)\tau} \left(\frac{\partial^2 u(x, \tau)}{\partial x^2} - q \frac{\partial u(x, \tau)}{\partial x} + \frac{1}{4}(q^2 - 1)u(x, \tau) \right). \end{aligned}$$

□

Ausgehend von der Darstellung

$$u(x, \tau) = \sum_{i=1}^n w_i(\tau) N_{i,k}(x)$$

für die diskrete Lösung $u \in \mathcal{N}_{k, T_h}$ lassen sich die Funktionswerte $u(x, \tau)$ und die partiellen Ableitungen $\partial u/\partial x$ bzw. $\partial^2 u/\partial x^2$ mit den Rekursionsformeln für B-Splines aus (A.6) und der Rekursionsformel für deren Ableitungen aus (A.5) stabil und effizient bestimmen.

Erfüllt $V(S, t)$ die Differentialgleichung aus (1.7), dann lässt sich die Veränderung der Lösung in der Zeit bei bereits bestimmter erster und zweiter partieller Ableitung nach S über die durch die Differentialgleichung gegebene Beziehung

$$\frac{\partial V}{\partial t} = rV - \frac{1}{2}\sigma^2 S^2 \frac{\partial^2 V}{\partial S^2} - rS \frac{\partial V}{\partial S}$$

berechnen.

Bei der Verwendung von Hutansatzfunktionen können die Ableitungen lediglich durch numerische Differentiation der Lösung bestimmt werden. Die in numerischen Tests beobachteten Approximationsordnungen bei der Bestimmung der Ableitungen mit unterschiedlich glatten Ansatzfunktionen wird in Abschnitt 6.2 diskutiert.

Der freie Rand

Der freie Rand $S_f(t)$ aus Problem 2.1.1 besteht aus den Punkten, an denen die Lösung V die Hindernisfunktion \mathcal{H} tangential berührt. Durch Rücktransformation lässt er sich aus dem freien Rand $x_f(\tau)$ des transformierten Problems 2.2.3, d.h. aus den Punkten, an denen die Lösung u des transformierten Problems die transformierte Hindernisfunktion g tangential berührt, berechnen.

Der freie Rand $x_f(\tau)$ lässt sich innerhalb der Hauptprogramm Schleife berechnen. In jedem Zeitschritt ν muss lediglich derjenige Gitterpunkt x_j für ein $j \in \{1, \dots, M\}$ festgehalten werden, ab dem (von x_{\min} nach x_{\max} gesehen) zum ersten Mal

$$u(x_j, \tau_\nu) > g(x_j, \tau_\nu)$$

gilt. Das heißt, es ist

$$u(x_i, \tau_\nu) = g(x_i, \tau_\nu) \quad \text{für alle } 1 \leq i < j.$$

Die Approximation $x_f(\tau) \approx x_j$ kann noch leicht verbessert werden, wenn man $x_f(\tau)$ als den Schnittpunkt der Geraden durch die Punkte $u(x_1, \tau_\nu)$ und $u(x_{j-1}, \tau_\nu)$ und der Geraden durch $u(x_j, \tau_\nu)$ und $u(x_M, \tau_\nu)$ bestimmt.

2.6 Das projektive Gauss–Seidel–Verfahren

Das Gleichungssystem

$$C \mathbf{w} = \mathbf{b}$$

mit der symmetrischen und positiv definiten Matrix $C = (C_{i,j}) \in \mathbb{R}^{n \times n}$ aus Problem 2.3.5 ist äquivalent zum Minimierungsproblem

$$(2.41) \quad \text{finde } \mathbf{w} \in \mathbb{R}^n : \mathcal{J}(\mathbf{w}) \leq \mathcal{J}(\mathbf{v}) \quad \text{für alle } \mathbf{v} \in \mathbb{R}^n$$

mit dem quadratischen Funktional $\mathcal{J}(\mathbf{w}) := \frac{1}{2}a(\mathbf{w}, \mathbf{w}) - \langle \mathbf{b}, \mathbf{w} \rangle$ und der Bilinearform $a(\mathbf{u}, \mathbf{v}) := \mathbf{u}^T C \mathbf{v}$. Der Beweis ist im Anhang in Satz A.3.1 ausgeführt. Beim klassischen Gauss–Seidel–Verfahren wird im ν -ten Iterationsschritt beginnend mit einer Startiterierten \mathbf{w}^0 bei bereits bestimmten w_j^ν für $j < i$ die neue Komponente w_i^ν sukzessive als Lösung von

$$(2.42) \quad \mathcal{J}(w_1^\nu, \dots, w_{i-1}^\nu, w_i^\nu, w_{i+1}^{\nu-1}, \dots, w_n^{\nu-1}) \leq \mathcal{J}(w_1^\nu, \dots, w_{i-1}^\nu, v_i^\nu, w_{i+1}^{\nu-1}, \dots, w_n^{\nu-1})$$

für alle $v_i^\nu \in \mathbb{R}$ bestimmt. Es ist bekannt, dass die Lösung von (2.42) gegeben ist durch

$$(2.43) \quad \mathbf{w}^\nu = \mathbf{w}^{\nu-1} + (U + D)^{-1}(\mathbf{b} - C\mathbf{w}^{\nu-1})$$

mit der Zerlegung $C = L + U + D$ in untere Dreiecksmatrix L , Diagonalmatrix D und obere Dreiecksmatrix U . Weiter ist bekannt (s. [GLT]), dass das Verfahren gegen die Lösung von (2.41) konvergiert, falls C symmetrisch und positiv definit mit positiven Diagonaleinträgen ist.

Ist anstelle eines Gleichungssystems ein lineares komplementäres Problem

$$\begin{aligned} (\mathbf{w} - \mathbf{g})^T (C \mathbf{w} - \mathbf{b}) &= 0, \\ C \mathbf{w} - \mathbf{b} &\geq 0, \quad \mathbf{w} - \mathbf{g} \geq 0 \end{aligned}$$

gegeben, erhält man anstelle von (2.41) äquivalenterweise das beschränkte Minimierungsproblem

$$(2.44) \quad \text{finde } \mathbf{w} \in \mathcal{K} : \mathcal{J}(\mathbf{w}) \leq \mathcal{J}(\mathbf{v}) \quad \text{für alle } \mathbf{v} \in \mathcal{K}$$

mit der konvexen Menge $\mathcal{K} := \{\mathbf{v} \in \mathbb{R}^n : \mathbf{v} \geq \mathbf{g}\}$. Die Menge $\mathcal{K} \subset \mathbb{R}^n$ wird als *faktoriell* bezeichnet, da sie sich in der Form

$$(2.45) \quad \mathcal{K} = \prod_{i=1}^n \mathcal{K}_i$$

mit $\mathcal{K}_i := [g_i, \infty) \subset \mathbb{R}$ darstellen lässt. Die Äquivalenz von linear komplementären Problemen mit Variationsungleichungen folgt aus Satz 2.3.3 und die Äquivalenz von Variationsungleichungen mit beschränkten Minimierungsproblemen findet sich im Anhang in Satz A.3.2.

In Analogie zu (2.42) werden die neuen Komponenten w_i^ν nun bei bereits bestimmten w_j^ν für $j < i$ sukzessive als Lösung von

$$(2.46) \quad \mathcal{J}(w_1^\nu, \dots, w_{i-1}^\nu, w_i^\nu, w_{i+1}^{\nu-1}, \dots, w_n^{\nu-1}) \leq \mathcal{J}(w_1^\nu, \dots, w_{i-1}^\nu, v_i^\nu, w_{i+1}^{\nu-1}, \dots, w_n^{\nu-1})$$

für alle $v_i' \in \mathcal{K}_i$ bestimmt. Dafür wird in jedem Iterationsschritt die Lösung von (2.42) in den Quader \mathcal{K} projiziert. Bezeichnet $\mathcal{P}_{\mathcal{K}_i}$ die Projektion in die Menge \mathcal{K}_i , d.h.

$$\mathcal{P}_{\mathcal{K}_i}(v_i) := \max\{g_i, v_i\},$$

dann sind die beiden Schritte

$$(2.47) \quad \begin{aligned} \bar{\mathbf{w}}^\nu &:= \mathbf{w}^{\nu-1} + (U + D)^{-1}(\mathbf{b} - C\mathbf{w}^{\nu-1}) \quad \text{und} \\ w_i' &= \mathcal{P}_{\mathcal{K}_i}(\bar{w}_i^\nu) \quad \text{für } i = 1, \dots, n \end{aligned}$$

durchzuführen. Das Verfahren entspricht also bis auf den Projektionsschritt dem klassischen Gauss-Seidel-Verfahren und wird als *projektives Gauss-Seidel-Verfahren* bezeichnet.

Satz 2.6.1 *Es sei $\mathcal{K} \in \mathbb{R}^n$ eine faktorielle, konvexe Menge wie in (2.45).*

(i) *Falls die Matrix C symmetrisch und positiv definit ist, konvergiert das projektive Gauss-Seidel-Verfahren für eine beliebige Anfangsnäherung $\mathbf{w}^0 \geq \mathbf{g}$ gegen die Lösung \mathbf{w} des Minimierungsproblems aus (2.44).*

(ii) *Gilt strikte Komplementarität, d.h. $(\mathbf{w} - \mathbf{g})_i + (C\mathbf{w} - \mathbf{b})_i > 0$ für alle $i = 1, \dots, n$, dann gibt es eine Konstante ν_0 , so dass für alle $\nu \geq \nu_0$ und für alle $i = 1, \dots, n$ die i -te Komponente der ν -ten Iterierten w_i^ν echt größer als g_i ist, falls die exakte Lösung w_i echt größer als g_i ist, und es gilt $w_i^\nu = g_i$, falls $w_i = g_i$.*

Beweisskizze: Ausgehend von der Formulierung als Minimierungsproblem zeigt man, dass die Folge $\{\mathcal{J}(\mathbf{w}^\nu)\}_{\nu \in \mathbb{N}}$ eine monoton fallende Folge ist. Ein Kompaktheitsargument liefert dann die Konvergenz. Ausführlich findet man den Beweis in [Cr]. \square

Bemerkung 2.6.2 Ist die Menge \mathcal{K} konvex, aber nicht faktoriell, so konvergiert das projektive Gauss-Seidel-Verfahren im Allgemeinen nicht gegen die Lösung des Minimierungsproblems aus (2.44). Für ein Gegenbeispiel siehe [GLT], S.65.

Bemerkung 2.6.3 Der zweite Teil des Satzes besagt, dass ab einer bestimmten Anzahl an Iterationen die Menge K^\bullet der Kontaktpunkte, d.h. die Menge der $i \in \{1, \dots, n\}$ für die $w_i = g_i$ gilt, richtig bestimmt ist. Damit entspricht das Konvergenzverhalten des projektiven Gauss-Seidel-Verfahrens asymptotisch dem des klassischen Gauss-Seidel-Verfahrens für lineare Gleichungssysteme. Bei abnehmender Gitterweite h ergeben sich asymptotisch damit auch beim projektiven Gauss-Seidel-Verfahren nicht zufriedenstellende Konvergenzraten $1 - O(h^2)$. Das Verfahren besitzt allerdings ebenfalls die Eigenschaft, die hochfrequenten Fehleranteile zu glätten. Dies wird in Kapitel 3 mit der Anwendung von Mehrgittertechniken zur Konvergenzbeschleunigung ausgenutzt.

Bemerkung 2.6.4 In vielen Fällen ist es möglich, die Konvergenz des Verfahrens zu beschleunigen, indem man das etwas allgemeinere projektive SOR-Verfahren (PSOR) mit einem Relaxationsparameter $1 \leq \omega \leq 2$ verwendet. Das projektive Gauss-Seidel-Verfahren ist das PSOR-Verfahren mit $\omega = 1$. In Hinblick auf die Verwendung von Mehrgitterverfahren zur Konvergenzbeschleunigung bietet sich die sukzessive Überrelaxation aber nicht an, da die Glättungseigenschaft der hochfrequenten Fehleranteile verloren geht. Aus diesem Grund wird hier das PSOR-Verfahren und die damit verbundene Frage nach einem optimalen Relaxationsparameter nicht weiter betrachtet.

Bemerkungen zu Minimierungsproblemen aus der Diskretisierung von freien Randwertproblemen

In Abschnitt 2.3.2 wurde gezeigt, dass die Diskretisierung des freien Randwertproblems 2.1.1 auf eine diskrete elliptische Variationsungleichung führt. Nach Satz A.3.2 ist diese äquivalent zu einem quadratischen Minimierungsproblem der Form

$$(2.48) \quad \text{finde } y_h \in \bar{\mathcal{K}}_h : \mathcal{J}(y_h) \leq \mathcal{J}(v_h) \text{ für alle } v_h \in \bar{\mathcal{K}}_h$$

über der konvexen, aber nicht faktoriellen Menge

$$\bar{\mathcal{K}}_h := \{y_h \in S_h : y_h(x) \geq g(x) \text{ für alle } x \in I\}$$

mit einer gegebenen Hindernisfunktion $g \in H^1(I)$ und dem Intervall $I \subset \mathbb{R}$. S_h bezeichnet wie bisher den endlich-dimensionalen Raum der B-Splines vorgegebener Ordnung $k > 1$.

Damit die Nebenbedingung $y_h \geq g$ richtig berücksichtigt wird, ist bei der Lösung mit einem projektiven Verfahren zu klären, wie B-Splines $y_h \in S_h = \mathcal{N}_{k,T_h}$ der Ordnung k mit der Darstellung

$$(2.49) \quad y_h = \sum_{i=1}^n w_i N_{i,k}$$

in den Lösungsraum $\bar{\mathcal{K}}_h$ projiziert werden können. Es wird vorgeschlagen, die Hindernisfunktion $g \in H^1(I)$ ebenfalls (notfalls approximativ) in der B-Spline-Basis von \mathcal{N}_{k,T_h} zu entwickeln, d.h.

$$(2.50) \quad g(x) \approx I_h g(x) = \sum_{i=1}^n g_i N_{i,k}(x) =: g_h(x) \in S_h.$$

Aufgrund der Positivität der B-Splines kann auf diese Weise über einen Vergleich der Entwicklungskoeffizienten geprüft werden, ob y_h die (approximative) Hindernisfunktion erfüllt.

Lemma 2.6.5 *Ist $g_i \leq w_i$ für alle $i = 1, \dots, n$ mit den Darstellungen von $y_h, g_h \in S_h$ aus (2.49) bzw. (2.50), dann folgt $g_h(x) \leq y_h(x)$ für alle $x \in I$.*

Beweis: Einsetzen der Darstellungen aus (2.49) bzw. (2.50), die Voraussetzung und die Positivität der B-Spline-Basisfunktionen $N_{i,k}(x) \geq 0$ für alle $x \in I$ liefern

$$y_h(x) - g_h(x) = \sum_{i=1}^n (w_i - g_i) N_{i,k}(x) \geq 0$$

für alle $x \in I$ und damit die Behauptung. □

Auf diese Weise reicht es also anstelle von Funktionswerten die n Entwicklungskoeffizienten von y_h und g_h zu betrachten, um sicherzustellen, dass $y_h \in \bar{\mathcal{K}}_h$ gilt. Der diskrete Lösungsraum

$$\mathcal{K}_h := \{y_h \in S_h : w_i \geq g_i \text{ für alle } i = 1, \dots, n\} \subset \bar{\mathcal{K}}_h$$

beschreibt dann wie in (2.45) einen Quader im \mathbb{R}^n , so dass das projektive Gauss–Seidel–Verfahren ohne weitere Modifikationen angewendet werden kann.

Für $k = 2$ sind die beiden Räume \mathcal{K}_h und $\bar{\mathcal{K}}_h$ äquivalent. Für $k > 2$ gilt $\mathcal{K}_h \subset \bar{\mathcal{K}}_h$. Dass die umgekehrte Inklusion für $k > 2$ im Allgemeinen nicht richtig, ist in Abbildung 2.3 veranschaulicht. Dort ist ein B-Spline y_h dargestellt, der im gesamten Intervall echt oberhalb

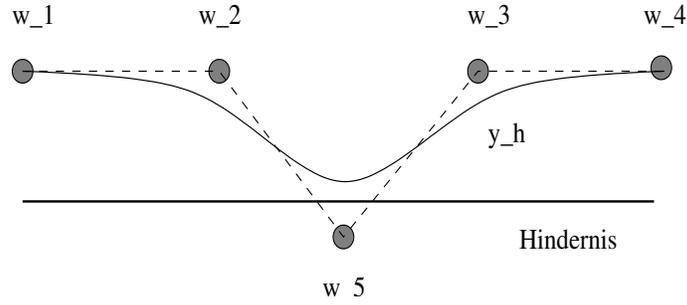


Abbildung 2.3: B-Spline y_h mit Entwicklungskoeffizienten w_i mit lokalem Minimum in der Nähe der Hindernisfunktion.

der konstanten Hindernisfunktion liegt. Der Entwicklungskoeffizient w_5 ist jedoch kleiner als der entsprechende Koeffizient der Hindernisfunktion.

Eine Einschränkung auf den leicht restriktiveren Lösungsraum \mathcal{K}_h ist natürlich nur dann möglich, wenn die Lösung des ursprünglichen Problems auch in \mathcal{K}_h enthalten ist und der oben diskutierte Fall ausgeschlossen werden kann. Dies ist auf jeden Fall dann gesichert, wenn die Lösung keine echten lokalen Minima in der Nähe der Hindernisfunktion aufweist.

Lemma 2.6.6 *Falls die diskrete Lösung $y_h \in \bar{\mathcal{K}}_h$ des Minimierungsproblems aus (2.48) keine echten lokalen Minima in der Nähe der Hindernisfunktion besitzt, dann gilt $y_h \in \mathcal{K}_h$.*

Beweis: Es sei $y_h \in \bar{\mathcal{K}}_h$ mit der Darstellung aus (2.49) die Lösung des Minimierungsproblems aus (2.48) und ohne Beschränkung der Allgemeinheit sei $g(x) = 0$ für alle $x \in I$. Anderenfalls betrachte $\tilde{y}_h := y_h - g$.

Angenommen es wäre $y_h \notin \mathcal{K}_h$. Dann ist $w_i < 0$ für ein $i \in \{1, \dots, n\}$. Damit $y_h \in \bar{\mathcal{K}}_h$ gilt, muss $w_{i-j_1} > 0$ und $w_{i+j_2} > 0$ gelten für $1 \leq j_1 < j_2 \leq n$ mit

$$\text{supp } N_{i,k} \subset (\text{supp } N_{j_1,k} \cup \text{supp } N_{j_2,k}).$$

Im Widerspruch zur Voraussetzung besitzt die Funktion y_h dann aber im Intervall $\text{supp } N_{j_1,k} \cup \text{supp } N_{j_2,k}$ ein echtes lokales Minimum $(x^*, y_h(x^*))$ (vgl. Abbildung 2.3). Aus der Konvexe-Hülle-Eigenschaft für B-Splines folgt, dass $y_h(x^*)$ in der konvexen Hülle der Entwicklungskoeffizienten c_{i-k+1}, \dots, c_i liegt. In diesem Sinne liegt das Minimum in der Nähe der Hindernisfunktion g . \square

Bemerkung 2.6.7 Im konkreten Problem 2.3.4 ist die Hindernisfunktion g konstant gleich Null. Wegen $0 \in \mathcal{N}_{k,T_h}$, ist hier also eine Approximation der Hindernisfunktion g durch eine Splinefunktion $I_h g \in \mathcal{N}_{k,T_h}$ nicht notwendig.

Kapitel 3

B–Spline basierte monotone Mehrgitterverfahren

Im ersten Abschnitt dieses Kapitels wird eine Literaturübersicht über Mehrgitteralgorithmen für freie Randwertprobleme gegeben, die sich in Form einer Variationsungleichung bzw. in Form eines linearen komplementären Problems darstellen lassen. Alle zitierten Arbeiten bleiben auf Hutfunktionen zur Diskretisierung des kontinuierlichen Problems beschränkt. Zum besseren Verständnis wird zunächst der vielen Verfahren zugrunde liegende Mehrgitteralgorithmus *PFAS* (*projective full approximation storage*) aus [BC] erläutert und mit einem Mehrgitteralgorithmus FAS (*full approximation storage*) für nichtlineare Differentialgleichungen zusammengestellt. Der für diese Arbeit zentrale *monotone Mehrgitteralgorithmus MMG* (*monotone multigrid method*) aus [Ko1] wird anschließend im Zusammenhang mit Mehrgitteralgorithmen für lineare Differentialgleichungen dargestellt.

In den nächsten beiden Abschnitten wird die Konstruktion und Theorie der monotonen Mehrgitterverfahren erläutert und von Hutfunktionen auf B–Spline–Basisfunktionen beliebiger Ordnung übertragen. Dazu werden wie in [Ko1, Ko2] zunächst erweiterte Relaxationsmethoden betrachtet.

3.1 Mehrgitteralgorithmen für Variationsungleichungen: Einleitung und Übersicht

Mehrgitterverfahren gehören zu den effizientesten Methoden zur Lösung von Randwertproblemen. Ihnen liegt die Beobachtung zugrunde, dass klassische Iterationsverfahren wie das Jakobi oder das Gauss–Seidel–Verfahren zwar nur langsam konvergieren, aber bereits nach wenigen Schritten sehr gute Approximationen an die hochfrequenten Anteile der Lösung liefern. Mehrgitterverfahren kombinieren dies mit einer komplementären Komponente, welche die niederfrequenten Anteile der Lösung anschließend auf einem gröberen Diskretisierungsgitter mit entsprechend geringerem Aufwand approximiert.

Die ersten Mehrgitteralgorithmen stammen von 1964 aus [F1, F2]. Dort wird für die Poisson–Gleichung im Einheitsquadrat gezeigt, dass der Rechenaufwand nur proportional mit der Anzahl der Unbekannten ansteigt. Diese Ergebnisse konnten 1966 in [Bv] auf Differenzenschemata für elliptische Differentialgleichungen zweiter Ordnung erweitert werden. Den obigen Arbeiten folgend entdeckte A. Brandt 1973, dass Mehrgitterverfahren für viele aktuelle Probleme anderen bekannten Verfahren überlegen sind (s. [B]). W. Hackbusch hat 1976 die Mehrgitteridee wieder aufgegriffen und die Konzepte mit neuen Ideen vereinfachen können (s. [H1]). Seitdem ist die

Mehrgittertechnik intensiv weiterentwickelt und auf immer komplexere und allgemeinere Probleme angewendet worden. Sie gehört mittlerweile zu den Standardverfahren zur Lösung von linearen und nichtlinearen Randwertproblemen.

Das prinzipielle Problem bei der Behandlung von freien Randwertproblemen mit Mehrgittertechniken ist, dass der freie Rand auf gröberen Gittern nur ungenau aufgelöst werden kann. Durch die notwendige Projektion im Glättungsschritt und eventuell irregulärer Verläufe des freien Randes ist der Fehler im Allgemeinen nicht so glatt wie bei Problemen mit festem Rand und kann daher zumindest in der Nähe des freien Randes durch die Grobgitterkorrektur in der Regel nur unzureichend approximiert werden.

Das PFAS Verfahren von A. Brandt und C.W. Cryer: Algorithmus, Varianten und aufbauende Arbeiten

A. Brandt und C.W. Cryer stellten 1980 in [BC] einen Mehrgitteralgorithmus für lineare komplementäre Probleme vor, welche aus der Diskretisierung freier Randwertprobleme entstehen. Ihr sogenanntes *PFAS* (*projective full approximation storage*) Verfahren entspricht im Wesentlichen einem Mehrgitterverfahren für nichtlineare Gleichungssysteme, bei dem das projektive Gauss-Seidel-Verfahren als Glätter verwendet wird. Sie zeigen, dass ihr Verfahren substantiell schneller als vorher bekannte Verfahren ist, geben aber keinen Konvergenzbeweis an. Der PFAS Algorithmus kann als eine Variante des *FAS* (*full approximation storage*) Algorithmus aus [B] für nichtlineare partielle Differentialgleichungen aufgefasst werden. Dies soll im folgenden für das lineare komplementäre Problem

$$(3.1) \quad \begin{aligned} \mathcal{L}u &\geq f, \\ u &\geq g, \\ (u - g)(\mathcal{L}u - f) &= 0 \end{aligned}$$

kurz erläutert werden. Es werden die Bezeichnungen aus Tabelle 3.1 verwendet.

Die Idee wird zunächst für zwei Gitter erklärt. Hierbei wird zwischen dem feinen Gitter mit Gitterweite h und dem groben Gitter mit der Gitterweite H unterschieden. Die Algorithmen werden dagegen gleich für den allgemeinen Fall von J Gittern beschrieben. Das größte Gitter auf der untersten Ebene $\ell = 1$ besitzt dann die Gitterweite h_1 . Die Gitterweite des feinsten Gitters auf der obersten Ebene $\ell = J$ beträgt $h_J = 2^{-J+1}h_1$.

\mathcal{L}	linearer Operator
\mathcal{N}	nichtlinearer Operator
S_ℓ	endlich-dimensionaler Raum stückweiser Polynome
\mathcal{S}	Glättungsoperator
η_1, η_2	Anzahl der Vor- bzw. Nachglättungsschritte
\mathcal{P}	Projektionsoperator
r, \tilde{r}	Restriktionsoperatoren
p	Prolongationsoperator
$u^{\nu, \ell, m}$	Näherung im ν -ten Zykel auf Ebene ℓ im m -ten Teilschritt

Tabelle 3.1: Bezeichnungen im Zusammenhang mit Mehrgitterverfahren.

Auch wenn es sich bei den linear komplementären Problem in (3.1) um einen linearen Operator \mathcal{L} handelt, ist das Problem wegen der Nebenbedingung $u \geq g$ als nichtlinear aufzufassen. Daher liegt es nahe zunächst Mehrgitterverfahren für nichtlineare partielle Differentialgleichungen

$$(3.2) \quad \mathcal{N}(u) = f$$

zu betrachten. Sie basieren auf dem gleichen Prinzip wie die für lineare Gleichungen. Der Fehler der Lösung wird erst auf einem feinen Gitter durch ein nichtlineares Analogon zum Jacobi oder Gauss–Seidel–Verfahren geglättet, so dass er dann gut auf einem gröberen Gitter approximiert werden kann. Es sei S_h ein endlich–dimensionaler Raum stückweiser Polynome. Weiter sei

$$\mathcal{N}_h(u_h) = f_h$$

die diskrete Form von (3.2) in S_h und es bezeichne $u_h^{\nu,1}$ die geglättete Näherung im ν –ten Zykel mit Fehler $v_h := u_h - u_h^{\nu,1}$. Dann gilt

$$\mathcal{N}_h(u_h^{\nu,1} + v_h) = \mathcal{N}_h(u_h^{\nu,1}) + d_h$$

mit dem Defekt $d_h := f_h - \mathcal{N}_h(u_h^{\nu,1})$. Diese Gleichung wird jetzt auf einem groben Gitter approximiert durch

$$\mathcal{N}_H(u_H + v_H) = \mathcal{N}_H(u_H) + d_H.$$

Dazu seien zwei Restriktionsoperatoren $r, \tilde{r} : S_h \rightarrow S_H$ für den Defekt d_h bzw. für die Näherung $u_h^{\nu,1}$ gegeben und es bezeichne $d_H := r d_h$ und $u_H := \tilde{r} u_h^{\nu,1}$. Damit erhält man in $w_H := u_H + v_H$ die auf dem groben Gitter zu lösende nichtlineare Gleichung

$$\mathcal{N}_H(w_H) = f_H$$

mit $f_H := \mathcal{N}_H(u_H) + d_H$. Aus der Lösung w_H der Grobgittergleichung bestimmt man die Grobgitterapproximation $v_H := w_H - u_H$ an den Fehler v_h . Diese wird mittels eines Prolongationsoperator p auf das feine Gitter zurück transformiert und zur Näherung $u_h^{\nu,1}$ addiert. Wendet man dieses Vorgehen nicht nur auf zwei Gittern, sondern rekursiv auf mehreren Ebenen an, erhält man den FAS Algorithmus für nichtlineare partielle Differentialgleichungen aus [B].

Algorithmus 3.1.1 FAS $_\ell$ (ν –ter Zyklus auf der Ebene $\ell \geq 1$)

Sei $u^{\nu,\ell}$ eine gegebene Näherung in S_ℓ .

1. *A priori–Glättung:* $u^{\nu,\ell,1} := \mathcal{S}^{\eta_1}(u^{\nu,\ell})$.
2. *Grob–Gitter–Korrektur:* $d_\ell := f_\ell - \mathcal{N}_\ell(u^{\nu,\ell,1})$,
 $u^{\ell-1,0} := \tilde{r} u^{\nu,\ell,1}$,
 $\mathcal{N}_{\ell-1} := r \mathcal{N}_\ell p$,
 $f_{\ell-1} := \mathcal{N}_{\ell-1} u^{\ell-1,0} + r d_\ell$.

Falls $\ell = 1$ ist, löse $\mathcal{N}_{\ell-1}(v) = f_{\ell-1}$ exakt und setze $v^{\ell-1} := v$.

Falls $\ell > 1$ ist, führe γ Schritte von **FAS** $_{\ell-1}$ mit Startwert $u^{\ell-1,0}$ und Ergebnis $v^{\ell-1}$ aus.

Setze $u^{\nu,\ell,2} := u^{\nu,\ell,1} + p(v^{\ell-1} - u^{\ell-1,0})$.

3. *A posteriori–Glättung:* $u^{\nu,\ell,3} := \mathcal{S}^{\eta_2}(u^{\nu,\ell,2})$.

Setze $u^{\nu,\ell+1} := u^{\nu,\ell,3}$.

Ist nun anstelle der nichtlinearen Differentialgleichung das lineare komplementäre Problem aus (3.1) mit der diskreten Form

$$(3.3) \quad \begin{aligned} \mathcal{L}_h u_h &\geq f_h, \\ u_h &\geq g_h, \\ (u_h - g_h)(\mathcal{L}_h u_h - f_h) &= 0 \end{aligned}$$

im endlich-dimensionalen Raum S_h zu lösen, erhält man analog für die geglättete Näherung $u_h^{\nu,1}$, für den Fehler $v_h := u_h - u_h^{\nu,1}$ und für den Defekt $d_h := f_h - \mathcal{L}_h u_h^{\nu,1}$ das Defektproblem

$$\begin{aligned}\mathcal{L}_h(u_h^{\nu,1} + v_h) &\geq \mathcal{L}_h u_h^{\nu,1} + d_h, \\ v_h + u_h^{\nu,1} &\geq g_h, \\ (v_h + u_h^{\nu,1} - g_h)(\mathcal{L}_h v_h - d_h) &= 0.\end{aligned}$$

Das Defektproblem wird ähnlich wie oben auf dem groben Gitter approximiert durch

$$\begin{aligned}\mathcal{L}_H(u_H + v_H) &\geq \mathcal{L}_H u_H + d_H, \\ v_H + u_H &\geq g_H, \\ (v_H + u_H - g_H)(\mathcal{L}_H v_H - d_H) &= 0.\end{aligned}$$

Dazu seien wieder Restriktionsoperatoren $r, \tilde{r} : S_h \rightarrow S_H$ für den Defekt d_h bzw. für die Näherung $u_h^{\nu,1}$ gegeben, und es ist $d_H := r d_h$ und $u_H := \tilde{r} u_h^{\nu,1}$. Die Diskretisierung der Funktion g in S_H wird mit g_H bezeichnet. Damit erhält man in $w_H := u_H + v_H$ das auf dem groben Gitter zu lösende lineare komplementäre Problem

$$\begin{aligned}\mathcal{L}_H w_H &\geq f_H, \\ w_H &\geq g_H, \\ (w_H - g_H)(\mathcal{L}_H w_H - f_H) &= 0\end{aligned}$$

mit $f_H := \mathcal{L}_H u_H + d_H$. Aus der Lösung w_H des Grobgitterproblems bestimmt man die Grobgitterapproximation $v_H := w_H - u_H$ an den Fehler v_h . Nachdem diese mittels eines Prolongationsoperators p auf das feine Gitter zurück transformiert ist, wird sie zur Näherung $u_h^{\nu,1}$ addiert. Wendet man dieses Vorgehen nicht nur auf zwei Gittern, sondern rekursiv auf mehreren Ebenen an, erhält man den PFAS Algorithmus für lineare komplementäre Probleme aus [BC].

Algorithmus 3.1.2 PFAS $_\ell$ (ν -ter Zyklus auf der Ebene $\ell \geq 1$)

Sei $u^{\nu,\ell}$ eine gegebene Näherung in S_ℓ .

1. *A priori-Glättung:* $\bar{u}^{\nu,\ell,1} := \mathcal{S}^m(u^{\nu,\ell})$.
Projektion: $u^{\nu,\ell,1} := \mathcal{P}\bar{u}^{\nu,\ell,1}$.
2. *Grob-Gitter-Korrektur:* $d_\ell := f_\ell - \mathcal{L}_\ell u^{\nu,\ell,1}$,
 $u^{\ell-1,0} := \tilde{r} u^{\nu,\ell,1}$,
 $\mathcal{L}_{\ell-1} := r \mathcal{L}_\ell p$,
 $f_{\ell-1} := \mathcal{L}_{\ell-1} u^{\ell-1,0} + r d_\ell$,
 $g_{\ell-1} := r g_\ell$.

Falls $\ell = 1$ ist, löse das lineare komplementäre Problem

$$\begin{aligned}\mathcal{L}_{\ell-1} v &\geq f_{\ell-1}, \\ v &\geq g_{\ell-1}, \\ (v - g_{\ell-1})(\mathcal{L}_{\ell-1} v - f_{\ell-1}) &= 0\end{aligned}$$

exakt und setze $v^{\ell-1} := v$.

Ist $\ell > 1$, führe γ Schritte von **PFAS $_{\ell-1}$** mit Startwert $u^{\ell-1,0}$ und Ergebnis $v^{\ell-1}$ aus.

Setze $u^{\nu,\ell,2} := u^{\nu,\ell,1} + p(v^{\ell-1} - u^{\ell-1,0})$.

3. *A posteriori-Glättung:* $\bar{u}^{\nu,\ell,3} := \mathcal{S}^{n_2}(u^{\nu,\ell,2})$.
Projektion: $u^{\nu,\ell,3} := \mathcal{P}\bar{u}^{\nu,\ell,3}$.

Setze $u^{\nu,\ell+1} := u^{\nu,\ell,3}$.

Numerische Tests zeigen, dass die Konvergenz deutlich beschleunigt werden kann, indem die Grobitterkorrektur nur weg vom Hindernis angewandt wird. Diese Variante aus [BC] wird als PDPFAS bezeichnet. Falls das Verfahren konvergiert, stellt man oft sehr gute Konvergenzraten fest. Allerdings gibt es weder für das PFAS noch für das PDPFAS Verfahren einen Konvergenzbeweis. Das Problem ist, dass die Näherung $u^{\nu,\ell,2}$ nach Grobitterkorrektur im Allgemeinen nicht mehr die durch die Hindernisfunktion gegebene Nebenbedingung

$$u^{\nu,\ell,2} \geq g_\ell$$

des linearen komplementären Problems erfüllt. Dies führt dazu, dass bei bestimmten Problemen Robustheitsprobleme auftreten (s. [Ko2]). Eine Klassifizierung von Problemen, für die das der Fall ist, lässt sich nur schwer festlegen. Verfahren, welche die obige Nebenbedingung einhalten, werden im folgenden als *monoton* bezeichnet. Solche Verfahren, die im zweiten Teil dieses Abschnitts beschrieben werden, sind deutlich robuster als das PFAS Verfahren. Außerdem existieren für sie Konvergenzbeweise.

Die Arbeit [BC] ist Grundlage zahlreicher weiterführender Arbeiten. Erwähnt seien die Arbeiten [HM, Ho, Ko1, Ma, CP, Oo], auf die im Rahmen einer Literaturübersicht im folgenden kurz eingegangen wird.

Alle Verfahren bleiben auf Hutfunktionen zur Diskretisierung des kontinuierlichen Operators beschränkt. Daher bezeichne S_h für den Rest dieses Abschnittes den endlich-dimensionalen Raum der Hutfunktionen. Weiter sei im folgenden $a : S_h \times S_h \rightarrow \mathbb{R}$ eine symmetrische, stetige und auf S_h positiv definite Bilinearform und $f : S_h \rightarrow \mathbb{R}$ ein lineares Funktional.

In [HM] wird das diskrete, nichtlineare Minimierungsproblem

$$(3.4) \quad \text{finde } u_h \in \mathcal{K}_h : \mathcal{F}_h(u_h) \leq \mathcal{F}_h(v_h) \quad \text{für alle } v_h \in \mathcal{K}_h$$

behandelt. Hierbei ist $\mathcal{F}_h : S_h \rightarrow \mathbb{R}$ ein Operator mit positiv definiten, selbstadjungierter zweiter Frechétableitung und

$$\mathcal{K}_h := \{v_h \in S_h : \underline{g}_h \leq v_h \leq \bar{g}_h\} \subset S_h$$

eine konvexe Menge mit gegebenen unteren und oberen Schranken $\underline{g}_h, \bar{g}_h \in S_h$. Es wird vorgeschlagen, das restringierte Minimierungsproblem aus (3.4) zu einer Folge von unrestringierten Minimierungsproblemen zu reduzieren. Der Algorithmus besteht dann aus zwei Phasen. In Phase A wird mit dem PFAS Verfahren aus [BC] die Menge $K^\bullet(u_h)$ der aktiven Punkte gesucht. Damit sind diejenigen Punkte gemeint, an denen die Lösung das Hindernis berührt. In Phase B wird das unrestringierte Minimierungsproblem für die Menge $K^\circ(u_h)$ der in Phase A bestimmten nicht aktiven Punkte mit einem Mehrgitterverfahren für nichtlineare Gleichungssysteme gelöst. Die Lösung liefert die optimale Abstiegsrichtung, in die aber höchstens so weit gegangen wird, wie es die Nebenbedingung zulässt. Unter der Annahme, dass "Phase B hinreichend dominant ist", wird die Konvergenz des Gesamtalgorithmus nachgewiesen.

In der Arbeit [Ho] wird von der Variationsungleichung

$$\text{finde } u_h \in \mathcal{K}_h : a(u_h, u_h - v_h) \leq f(u_h - v_h) \quad \text{für alle } v_h \in \mathcal{K}_h$$

ausgegangen. Dabei ist

$$\mathcal{K}_h := \{v_h \in S_h : v_h \leq g_h\} \subset S_h$$

eine konvexe Menge mit dem oberen Hindernis $g_h \in S_h$. Alternativ werden auch beidseitige Hindernisse und Hindernisse, die von der Lösung des Problems abhängen, betrachtet. Mit dem Riesz-Operator $A_h : S_h \rightarrow S_h$, der $a(u_h, v_h) = \langle A_h u_h, v_h \rangle$ erfüllt, lässt sich die Variationsungleichung äquivalent als komponentenweise aufzufassende Hamilton-Jacobi-Bellmann Gleichung

$$\max_{v_h \in S_h} \{\langle A_h v_h - f_h, v_h - g_h \rangle\} = 0$$

umformulieren, für die in [LM] ein iterativer Lösungsalgorithmus angegeben ist. Dieser wird als äußere Schleife verwendet. Die innere Schleife beinhaltet die Lösung eines um die aktiven Komponenten der gegebenen Iterierten verkleinerten Gleichungssystems. Die Lösung des Gleichungssystems erfolgt analog zu Phase B in [HM] mit einem Mehrgitterverfahren für nichtlineare Gleichungssysteme. Unter der Annahme, dass in der inneren Schleife exakt gelöst wird, wird monotone Konvergenz des Algorithmus nachgewiesen. Numerische Ergebnisse bestätigen die Konvergenz auch bei näherungsweise Lösung der Gleichungssysteme.

In den Arbeiten [CP, Oo] wird das PFAS Verfahren verwendet, um den Wert von Amerikanischen Optionen mit stochastischer Volatilität zu bestimmen. Die Bewertung führt auf ein instationäres, zweidimensionales freies Randwertproblem zu einer linearen Konvektions–Diffusions–Reaktions–Gleichung zweiter Ordnung (s. Unterabschnitt 5.3.2). In [CP] wird die lineare komplementäre Formulierung des Problems mit Finite–Differenzen im Ort und einem impliziten θ –Schema in der Zeit diskretisiert. Die Diskretisierung führt in jedem Zeitschritt auf ein diskretes lineares komplementäres Problem, welches mit dem PFAS Verfahren aus [BC] gelöst wird. Als Glätter dient das projektive Gauss–Seidel–Verfahren. Mittels einer analytischen Transformation des Gitters und adaptiv gewählten Zeitschrittweiten, kann der Aufwand des Verfahrens deutlich reduziert werden.

Im Unterschied zu [CP] wird in den neueren Arbeit [Oo] die Zeit mit dem BDF2–Verfahren diskretisiert. Das BDF2–Verfahren ist stabiler als ein θ –Schema, welches zu Oszillationen in den Ableitungen der Lösung führen kann. Die diskreten linearen komplementären Probleme werden mit der schnelleren Variante PDPFAS aus [BC] gelöst. Um die Robustheit und die Konvergenzgeschwindigkeit weiter zu verbessern, wird das Mehrgitterverfahren als Vorkonditionierer für eine Krylov Unterraummethode eingesetzt. Fourieranalyse und numerische Test zeigen schnelle Konvergenz des Verfahrens.

Das monotone Mehrgitterverfahren MMG von R. Kornhuber

Die Mehrgitterverfahren aus [Ma, Ko1, Ko2] unterscheiden sich von allen obigen Verfahren dadurch, dass durch spezielle Restriktionen der Hindernisfunktion sichergestellt wird, dass die Näherung $u^{\nu,\ell,2}$ nach Grobgitterkorrektur auf jeder Ebene $\ell \leq J$ des Verfahrens die Nebenbedingung $u^{\nu,\ell,2} \geq g_\ell$ einhält. Die Konstruktion solcher Restriktionsoperatoren ist ein zentrales Thema dieser Arbeit und wird in Kapitel 4 diskutiert. Die dadurch resultierende Monotonie der Verfahren führt dazu, dass die Verfahren besonders robust sind.

In [Ma] wird das restringierte Minimierungsproblem

$$\text{finde } u_h \in \mathcal{K}_h : \mathcal{J}(u_h) \leq \mathcal{J}(v_h) \quad \text{für alle } v_h \in \mathcal{K}_h$$

mit dem quadratischen Funktional

$$\mathcal{J}(v_h) := \frac{1}{2}a(v_h, v_h) - f(v_h)$$

und der konvexen Lösungsmenge

$$\mathcal{K}_h := \{v_h \in S_h : v_h \geq g_h\} \subset S_h$$

behandelt. Die Grundstruktur des Mehrgitterverfahrens aus [Ma] entspricht der des monotonen Mehrgitteralgorithmus aus [Ko1] und ist unten ausführlicher beschrieben. Es wird nachgewiesen, dass das Verfahren für alle Startwerte $u_h^0 \in \mathcal{K}_h$ konvergent ist. Weiter wird unter der Annahme, dass das Problem nicht–degeneriert (s. Definition 3.2.5) ist, gezeigt, dass die Kontaktmenge der Lösung nach einer bestimmten Anzahl an Iterationen richtig bestimmt ist und sich das Verfahren asymptotisch zu einer linearen Iteration reduziert.

Die weiterführenden Arbeiten [Ko1, Ko2] von R. Kornhuber können als Pionierarbeit für die Konvergenzanalyse von Mehrgitterverfahren für Variationsungleichungen über Teilraumzerlegungsmethoden betrachtet werden ([Tai]). Die Grundstruktur des sogenannten *monotonen Mehrgitteralgorithmus MMG (monotone multigrid method)* ist unten erläutert. Die Theorie wird ausführlich in den Abschnitten 3.2 und 3.3 behandelt. Die Restriktionsoperatoren für die Hinderisfunktion können in Vergleich zu [Ma] verbessert werden. Die Konstruktion wird in Kapitel 4 diskutiert. Mit Teilraumzerlegungsmethoden wird Mehrgitterkonvergenz des Verfahrens nachgewiesen. Ein systematisches Anpassen der Basisfunktionen an den aktuellen freien Rand führt zur *abgeschnittenen Variante TrMMG (truncated monotone multigrid method)* des Verfahrens. In numerischen Tests erreicht diese bei Anwendung auf Variationsungleichungen die bekanntermaßen guten Konvergenzraten linearer Mehrgitterverfahren.

Der Mehrgitteralgorithmus MMG aus [Ko1] zur Lösung von Variationsungleichungen bzw. von linear komplementären Problemen kann in Analogie zu einem klassischen Mehrgitteralgorithmus MG für lineare partielle Differentialgleichungen betrachtet werden. Dies wird im folgenden mit den Bezeichnungen aus Tabelle 3.1 kurz erläutert. Es sei

$$(3.5) \quad \mathcal{L}u = f$$

eine lineare Differentialgleichung mit diskreter Form

$$\mathcal{L}_h u_h = f_h$$

im endlich-dimensionalen Raum S_h und es bezeichne $u_h^{\nu,1}$ die geglättete Näherung im ν -ten Zykel mit Fehler $v_h := u_h - u_h^{\nu,1}$. Dann gilt

$$\mathcal{L}_h v_h = d_h$$

mit dem Defekt $d_h := f_h - \mathcal{L}_h u_h^{\nu,1}$. Diese Gleichung wird jetzt durch die in S_H zu lösende lineare Gleichung

$$\mathcal{L}_H v_H = d_H$$

approximiert. Dazu sei ein Restriktionsoperatoren $r : S_h \rightarrow S_H$ für den Defekt d_h gegeben und es bezeichne $d_H := r d_h$. Die Lösung v_H ist eine Näherung an v_h und wird mittels eines Prolongationsoperator p auf das feine Gitter zurück interpoliert. Wendet man dieses Vorgehen nicht nur auf zwei Gittern, sondern rekursiv auf mehreren Ebenen an, erhält man den Mehrgitteralgorithmus MG für lineare partielle Differentialgleichungen.

Algorithmus 3.1.3 \mathbf{MG}_ℓ (ν -ter Zyklus auf der Ebene $\ell \geq 1$)

Sei $u^{\nu,\ell} \in S_\ell$ eine gegebene Näherung.

1. *A priori*-Glättung: $u^{\nu,\ell,1} := \mathcal{S}^{\eta_1}(u^{\nu,\ell})$.
2. *Grob-Gitter-Korrektur*: $d_\ell := f_\ell - \mathcal{L}_\ell u^{\nu,\ell,1}$,
 $f_{\ell-1} := r d_\ell$,
 $\mathcal{L}_{\ell-1} := r \mathcal{L}_\ell p$.

Falls $\ell = 1$ ist, löse $\mathcal{L}_{\ell-1} v = f_{\ell-1}$ exakt und setze $v^{\ell-1} := v$.

Ist $\ell > 1$, führe γ Schritte von $\mathbf{MG}_{\ell-1}$ mit Startwert $u^{\ell-1,0} := 0$ und Ergebnis $v^{\ell-1}$ aus.

Setze $u^{\nu,\ell,2} := u^{\nu,\ell,1} + p v^{\ell-1}$.

3. *A posteriori*-Glättung: $u^{\nu,\ell,3} := \mathcal{S}^{\eta_2}(u^{\nu,\ell,2})$.

Setze $u^{\nu,\ell+1} := u^{\nu,\ell,3}$.

Für $\gamma = 1$ ist dies ein V-Zyklus für $\gamma = 2$ ein W-Zyklus.

Ist anstelle einer linearen Differentialgleichung das diskrete lineare komplementäre Problem aus (3.3) zu lösen, erhält man analog für die geglättete Näherung $u_h^{\nu,1} \in S_h$, für den Fehler $v_h := u_h - u_h^{\nu,1}$ und für den Defekt $d_h := f_h - \mathcal{L}_h u_h^{\nu,1}$ das Defektproblem

$$(3.6) \quad \begin{aligned} \mathcal{L}_h v_h &\geq d_h, \\ v_h &\geq g_h - u_h^{\nu,1}, \\ (v_h - g_h + u_h^{\nu,1})(\mathcal{L}_h v_h - d_h) &= 0. \end{aligned}$$

Ein glatter Fehler v_h kann ohne wesentlichen Informationsverlust in S_H approximiert werden. Das Grobgitterproblem lautet nun

$$(3.7) \quad \begin{aligned} \mathcal{L}_H v_H &\geq d_H, \\ v_H &\geq g_H, \\ (v_H - g_H)(\mathcal{L}_H v_H - d_H) &= 0. \end{aligned}$$

Es erweist sich als sinnvoll verschiedenen Restriktionsoperatoren $r, \tilde{r} : S_h \rightarrow S_H$ für den Defekt d_h bzw. für das Hindernis $g_h - u_h^{\nu,1}$ zu verwenden. Es bezeichnet $d_H := r d_h$ die Restriktion des Defektes. Bei Inspektion von (3.6) in Hinblick auf (3.7) ist zu beachten, dass g_H nicht die Diskretisierung der Funktion g in S_H ist, sondern definiert ist als

$$g_H := \tilde{r}(g_h - u_h^{\nu,1}).$$

Die Lösung v_H ist eine Näherung an den Fehler v_h und wird mittels eines Prolongationsoperator p auf das feine Gitter zurück transformiert und zur Näherung $u_h^{\nu,1}$ addiert. Wichtig ist, dass der Restriktionsoperator \tilde{r} so gewählt wird, dass die neue Näherung die Nebenbedingung

$$(3.8) \quad u_h^{\nu,2} := u_h^{\nu,1} + p v_H \geq g_h.$$

auf dem feinen Gitter erfüllt.

Wendet man dieses Vorgehen nicht nur auf zwei Gittern, sondern rekursiv auf mehreren Ebenen an, erhält man das monotone Mehrgitterverfahren MMG für lineare komplementäre Probleme.

Algorithmus 3.1.4 MMG $_\ell$ (ν -ter Zyklus auf der Ebene $\ell \geq 1$)

Sei $u^{\nu,\ell} \in S_\ell$ eine gegebene Näherung.

1. *A priori-Glättung:* $\bar{u}^{\nu,\ell,1} := \mathcal{S}^m(u^{\nu,\ell})$.
Projektion: $u^{\nu,\ell,1} := \mathcal{P}\bar{u}^{\nu,\ell,1}$.
2. *Grob-Gitter-Korrektur:* $d_\ell := f_\ell - \mathcal{L}_\ell u^{\nu,\ell,1}$,
 $f_{\ell-1} := r d_\ell$,
 $g_{\ell-1} := \tilde{r}(g_\ell - u^{\nu,\ell,1})$,
 $\mathcal{L}_{\ell-1} := r \mathcal{L}_\ell p$.

Falls $\ell = 1$ ist, löse das lineare komplementäre Problem

$$\begin{aligned} \mathcal{L}_{\ell-1} v &\geq f_{\ell-1}, \\ v &\geq g_{\ell-1}, \\ (v - g_{\ell-1})(\mathcal{L}_{\ell-1} v - f_{\ell-1}) &= 0 \end{aligned}$$

exakt und setze $v^{\ell-1} := v$.

Ist $\ell > 1$, führe γ Schritte von MMG $_{\ell-1}$ mit Startwert $u^{\ell-1,0} := 0$ und Ergebnis $v^{\ell-1}$ aus.

Setze $u^{\nu,\ell,2} := u^{\nu,\ell,1} + p v^{\ell-1}$.

3. *A posteriori*-Glättung: $\bar{u}^{\nu,\ell,3} := S^{\eta_2}(u^{\nu,\ell,2})$.
 Projektion: $u^{\nu,\ell,3} := \mathcal{P}\bar{u}^{\nu,\ell,3}$.

Setze $u^{\nu,\ell+1} := u^{\nu,\ell,3}$.

Die Bedingung aus (3.8) sichert, dass monotone Mehrgitterverfahren besonders robust sind. Damit diese Bedingung erfüllt ist, werden spezielle Restriktionsoperatoren \tilde{r} für die Hindernisfunktion benötigt. Die Konstruktion solcher Restriktionsoperatoren wird für B-Splines beliebiger Ordnung in Kapitel 4 diskutiert. Die Konvergenztheorie monotoner Mehrgitterverfahren wird in den folgenden Abschnitten 3.2 und 3.3 behandelt.

3.2 Relaxationsverfahren

Die Konvergenztheorie für das monotone Mehrgitterverfahren soll [Ko1] folgend über Teilraumzerlegungsmethoden geführt werden. Dafür werden zunächst erweiterte Relaxationsverfahren betrachtet. Die Vorgehensweise hält sich eng an [Ko1]. Die Beweise sind zum Teil in Anlehnung an [Kr] geführt. Während in den zitierten Arbeiten nodale Basisfunktionen verwendet werden, wird hier das Verfahren und die Theorie auf B-Spline-Basisfunktionen beliebiger Ordnung übertragen.

Zur besseren Übersicht ist die an [Ko1] angelehnte Notation in Tabelle 3.2 zusammengefasst. Iterationen werden mit oberen Indizes ν und Zwischeniterationen mit oberen Indizes j bezeichnet. Bei Entwicklung in einer Basis werden die Koeffizienten einer Funktion mit dem gleichen Buchstaben wie die Funktion, aber mit einem unteren Index i bezeichnet.

$S_\ell \subset S_J \subset H_0^1(\Omega)$	endlich-dimensionale Räume mit Dimension n_ℓ
$\Sigma_\ell \subset S_\ell$	Basis von S_ℓ mit Kardinalität n_ℓ
$\nu \in \mathbb{N}$	Iterationszähler
$j \in \mathbb{N}$	Zwischeniterationszähler
$M^\nu \subset S_J$	Menge der Suchrichtungen im ν -ten Iterationsschritt
$M_c^\nu \subset M^\nu$	Menge der Grobgitterfunktionen in M^ν
$\mu^{\nu,j} \in M^\nu$	j -te Suchrichtung im ν -ten Iterationsschritt
$D^{\nu,j}, D^{*,\nu,j} \subset \text{span}\{\mu^{\nu,j}\}$	eindimensionale Lösungsräume lokaler Teilprobleme
$v^{\nu,j} \in D^{\nu,j}, v^{*,\nu,j} \in D^{*,\nu,j}$	Korrekturen in Richtung von $\mu^{\nu,j}$
$\psi^{\nu,j} \in \text{span}\{\mu^{\nu,j}\}$	lokale, approximierte Hindernisfunktionen
$\bar{u}_J^\nu \in S_J$	geglättete Iterierte u_J^ν

Tabelle 3.2: Bezeichnungen im Zusammenhang mit Relaxationsverfahren.

Die zugrundeliegende Idee von Relaxationsverfahren ist es, das Minimierungsproblem im endlich-dimensionalen Raum S_J in endlich viele lokale Teilprobleme zu zerlegen. Beim Gauss-Seidel-Verfahren spannt z.B. jede Basisfunktion von S_J einen lokalen Teilraum auf, in dem dann sukzessive minimiert wird. Bei Mehrgitterverfahren wird diese Teilraumzerlegung um zusätzliche Räume ergänzt, die von Grobgitterfunktionen $\mu \in S_\ell$ mit $\ell < J$ aufgespannt werden.

3.2.1 Erweiterte Relaxationsverfahren

Es sei Ω ein Gebiet im \mathbb{R}^d und

$$S_1 \subset S_2 \subset \dots \subset S_J \subset H_0^1(\Omega)$$

eine geschachtelte Folge endlich-dimensionaler Räume.

Dabei sei $S_\ell := \mathcal{N}_{k,\Delta_\ell}$ für $1 \leq \ell \leq J$ der Raum der B-Splines der Ordnung k zum Gitter $\Delta_\ell \subset \Omega$ (s. Anhang A.1). Die Basis von S_J sei stets mit

$$\Sigma_J := \{\mu^1, \dots, \mu^{n_J}\} := \{N_{1,\Delta_J}, \dots, N_{n_J,\Delta_J}\}$$

bezeichnet. Hierbei wird die Abkürzung $N_{i,\Delta_\ell} := N_{i,k,\Delta_\ell}$ verwendet. Bezüglich dieser Basis besitzen die Funktionen $u, g \in S_J$ die Darstellungen

$$u(x) = \sum_{i=1}^{n_J} u_i N_{i,\Delta_J}(x), \quad g(x) = \sum_{i=1}^{n_J} g_i N_{i,\Delta_J}(x).$$

Bemerkung 3.2.1 Während es für Hutfunktionen $u, g \in \mathcal{N}_{2,\Delta_J}$ reicht nur Funktionswerte an Stützstellen $\theta \in \Delta_\ell$ zu betrachten, um die Bedingung $u(x) \leq g(x)$ für alle $x \in \Omega$ zu überprüfen, ist dies bei Funktionen höherer Ordnung nicht mehr ausreichend. Bei einer Darstellung in der B-Spline-Basis folgt die Monotonie aber wegen Lemma 2.6.5 aus $u_i \leq g_i$ für alle $i = 1, \dots, n_J$. Damit kann bei der Verwendung von B-Splines anstelle eines Vergleiches von Funktionswerten an Stützstellen auf den Vergleich von Entwicklungskoeffizienten zurückgegriffen werden. Der Vorteil ist, dass diese Vorgehensweise einfacher auf beliebig glatte Ansatzfunktionen übertragbar ist.

Im folgenden sei

$$(3.9) \quad \mathcal{J}(v) := \frac{1}{2}a(v, v) - \ell(v)$$

ein quadratischer Operator mit einer stetigen, symmetrischen und $H_0^1(\Omega)$ -elliptischen Bilinearform $a(\cdot, \cdot)$ und einem Funktional $\ell \in H^{-1}(\Omega)$.

Betrachte hierzu das diskrete Minimierungsproblem

$$(3.10) \quad \text{finde } u_J \in \mathcal{K}_J : \mathcal{J}(u_J) \leq \mathcal{J}(v) \quad \text{für alle } v \in \mathcal{K}_J$$

mit der abgeschlossenen und konvexen Menge

$$\mathcal{K}_J := \{v \in S_J : v_i \leq g_i \text{ für alle } i = 1, \dots, n_J\}.$$

Die Funktion $g \in S_J$ stellt ein oberes Hindernis für die diskrete Lösung u_J dar. Wegen $S_J \subset H_0^1(\Omega)$, muss dabei $g(x) \geq 0$ für alle $x \in \partial\Omega$ gelten, damit das Problem wohlgestellt ist. Die Behandlung unterer Hindernisse ist vollkommen analog möglich.

Nach den Sätzen A.3.3 und A.3.2 besitzt das Minimierungsproblem eine eindeutige Lösung und ist äquivalent zur Variationsungleichung

$$(3.11) \quad \text{finde } u_J \in \mathcal{K}_J : a(u_J, v - u_J) \geq \ell(v - u_J) \quad \text{für alle } v \in \mathcal{K}_J.$$

Iterativ kann eine Lösung mit dem projektiven Gauss-Seidel-Verfahren (s. Abschnitt 2.6) bestimmt werden. Dabei wird das Funktional \mathcal{J} sukzessive in Richtung der Feingitterbasisfunktionen $\mu^j \in \Sigma_J$ minimiert. Dies ist ein sogenanntes nichtlineares multiplikatives Schwarz-Verfahren mit der Zerlegung

$$S_J = \text{span } \Sigma_J.$$

Ausführliche Informationen über die Theorie der Schwarz-Verfahren finden sich z.B. in [GO, Zh1]. Aus der gegebenen Iterierten u_J^ν erhält man so die neue Iterierte

$$u_J^{\nu+1} := \bar{u}_J^\nu := \mathcal{S}(u_J^\nu).$$

Im folgenden werden die Funktionen μ^j auch als *Suchrichtung des Relaxationsverfahrens* bezeichnet. In Bemerkung 2.6.3 ist bereits darauf hingewiesen worden, dass die Konvergenzgeschwindigkeit des projektiven Gauss–Seidel–Verfahrens mit steigender Anzahl an Unbekannten schnell abnimmt. Wie bei linearen Mehrgitterverfahren soll die Konvergenz daher unter Ausnutzung der Glättungseigenschaft des projektiven Gauss–Seidel–Verfahrens mit Korrekturen beschleunigt werden, die durch Minimierung des Funktionals \mathcal{J} in Richtung von Funktionen mit großem Träger bestimmt werden. Für $\nu \in \mathbb{N}$ lautet die erweiterte Menge der Suchrichtungen dann

$$(3.12) \quad M^\nu := \Sigma_J \cup M_c^\nu$$

mit einer Menge $M_c^\nu := \{\mu^{\nu, n_J+1}, \dots, \mu^{\nu, m^\nu}\}$ an Grobgitterfunktionen, die in jedem Iterationsschritt ν variieren kann. Das Erzeugendensystem M^ν führt jetzt zu der Zerlegung

$$S_J = \text{span } \Sigma_J \cup \text{span } M_c^\nu$$

des diskreten Raumes S_J . Die durch die Menge M^ν induzierte *erweiterte Relaxationsmethode* (*extended relaxation*) lautet nun folgendermaßen:

Zu gegebener Iterierter $u_J^\nu =: w_0^\nu$ werden sukzessive die Zwischeniterierten $w^{\nu, j} = w^{\nu, j-1} + v^{*, \nu, j}$ für $j = 1, \dots, m^\nu$ bestimmt. Die Korrektur $v^{*, \nu, j}$ ist dabei die eindeutige Lösung des lokalen Teilproblems

$$(3.13) \quad \text{finde } v^{*, \nu, j} \in D^{*, \nu, j} : \mathcal{J}(w^{\nu, j-1} + v^{*, \nu, j}) \leq \mathcal{J}(w^{\nu, j-1} + v) \quad \text{für alle } v \in D^{*, \nu, j}$$

mit dem eindimensionalen Teilraum

$$D^{*, \nu, j} := \{v \in \text{span}\{\mu^{\nu, j}\} : v_i \leq g_i - w_i^{\nu, j-1} \text{ für alle } i = 1, \dots, n_J\}.$$

Hierbei bezeichnet der untere Index i die Koeffizienten bei Darstellung in der Basis Σ_J . Die neue Iterierte lautet dann

$$(3.14) \quad u_J^{\nu+1} := w^{\nu, m^\nu} = u_J^\nu + \sum_{j=1}^{m^\nu} v^{*, \nu, j} = \bar{u}_J^\nu + \sum_{j=n_J+1}^{m^\nu} v^{*, \nu, j}.$$

Man beachte, dass bei einem erweiterten Relaxationsverfahren über ein Erzeugendensystem minimiert wird. Dieses enthält alle Basisfunktionen des diskreten Raumes S_J . Es gilt $w^{\nu, n_J} = \bar{u}_J^\nu = \mathcal{S}(u_J^\nu)$. Das erweiterte Relaxationsverfahren kann insofern als ein restringiertes Linienverfahren aufgefasst werden, als über jede Richtung des Erzeugendensystems M^ν soweit minimiert wird, wie es das Hindernis zulässt.

3.2.2 Approximative erweiterte Relaxationsverfahren

Das erweiterte Relaxationsverfahren aus (3.14) ist monoton im Sinne, dass

$$\mathcal{J}(u_J^{\nu+1}) \leq \mathcal{J}(u_J^\nu)$$

für alle $u_J^\nu \in \mathcal{K}_J$ gilt. Für $j = n_J + 1, \dots, m^\nu$ zerstört das Überprüfen, ob $v \in \text{span}\{\mu^{\nu, j}\}$ auch in $D^{*, \nu, j}$ liegt, allerdings die optimale Komplexität des Verfahrens, weil v dazu zurück auf das feinste Gitter Δ_J transformiert werden muss. Um diese Problematik zu lösen, werden anstelle der *exakten Korrekturen* $v^{*, \nu, j}$ im folgenden *approximative Korrekturen* $v^{\nu, j}$ für $j = n_J + 1, \dots, m^\nu$ durch die Lösung des approximativen lokalen Teilproblems

$$(3.15) \quad \text{finde } v^{\nu, j} \in D^{\nu, j} : \mathcal{J}(w^{\nu, j-1} + v^{\nu, j}) \leq \mathcal{J}(w^{\nu, j-1} + v) \quad \text{für alle } v \in D^{\nu, j}$$

bestimmt. Der eindimensionale Lösungsraum $D^{\nu,j} \subset \text{span}\{\mu^{\nu,j}\}$ ist nun als

$$D^{\nu,j} := \{v \in \text{span}\{\mu^{\nu,j}\} : v_i \leq \psi_i^{\nu,j} \text{ für alle } i = 1, \dots, n_J\}$$

definiert. Die *approximative lokale Hindernisfunktion* $\psi^{\nu,j}$ wird dabei durch eine geeignete Restriktion der *exakten* lokalen Hindernisfunktion $g - w^{\nu,j-1}$ für die Korrektur $v^{\nu,j}$ konstruiert. Die Konstruktion wird in Kapitel 4 für B-Splines beliebiger Glattheit beschrieben. Für das exakte Hindernis gilt offensichtlich

$$(3.16) \quad g - w^{\nu,j-1} \geq 0.$$

Die neue Iterierte des *approximativen erweiterten Relaxationsverfahrens (approximated extended relaxation)* lautet dann

$$(3.17) \quad u_J^{\nu+1} := w^{\nu,m^\nu} = u_J^\nu + \sum_{j=1}^{n_J} v^{*,\nu,j} + \sum_{j=n_J+1}^{m^\nu} v^{\nu,j} = \bar{u}_J^\nu + \sum_{j=n_J+1}^{m^\nu} v^{\nu,j}.$$

Entscheidend ist, dass $\psi^{\nu,j}$ im eindimensionalen Raum $\text{span}\{\mu^{\nu,j}\}$ liegt. Daher muss jetzt nur noch ein einziger Entwicklungskoeffizient v_i mit $\psi_i^{\nu,j}$ verglichen werden, um festzustellen, ob $v \in D^{\nu,j}$ ist.

Die approximativen lokalen Hindernisse $\psi^{\nu,j}$ sollen so konstruiert sein, dass die mit diesem Hindernis bestimmten Korrekturen $v^{\nu,j}$ auch zulässig im Hinblick auf das exakte Hindernis $g - w^{\nu,j-1}$ sind. Dies motiviert die folgende Definition.

Definition 3.2.2 (Monotone lokale Hindernisse) *Das lokale Hindernis $\psi^{\nu,j} \in \text{span}\{\mu^{\nu,j}\}$ heißt monoton, falls $D^{\nu,j} \subset D^{*,\nu,j}$ gilt.*

Monotone Hindernisse führen also zu einem "damping" der exakten Korrekturen. Liegen monotone Hindernisse vor, dann folgt aus $u_J^\nu =: w^{\nu,0} \in \mathcal{K}_J$, dass $w^{\nu,j} \in \mathcal{K}_J$ für alle $j = 1, \dots, m^\nu$ gilt. Weiter lässt sich als Verallgemeinerung von Satz 2.6.1 (i) globale Konvergenz des approximativen erweiterten Relaxationsverfahren aus (3.17) nachweisen.

Satz 3.2.3 *Es sei $0 \in D^{\nu,j} \subset D^{*,\nu,j}$ für alle $\nu \geq 1$. Dann konvergiert das approximative erweiterte Relaxationsverfahren aus (3.17) mit $v^{*,\nu,j}$ und $v^{\nu,j}$ aus (3.13) bzw. (3.15) für alle Startiterierten $u_J^0 \in S_J$ gegen die Lösung u_J von (3.10).*

Beweis: Der Beweis ist aus [Kr], S.36 übernommen. Die Folge $\{u_J^\nu\}_{\nu \in \mathbb{N}}$ ist beschränkt, weil

$$\mathcal{J}(u_J^\nu) \leq \mathcal{J}(u_J^1) \quad \text{für alle } \nu \in \mathbb{N}$$

gilt und weil wegen (3.9) $\mathcal{J}(v^\nu) \rightarrow \infty$ für jede unbeschränkte Folge $v^\nu \in S_J$ gilt. Da der Raum S_J endlich-dimensional ist, reicht es zu zeigen, dass jede konvergente Teilfolge

$$(3.18) \quad u_J^{\nu_l} \rightarrow u_J^* \in S_J \quad \text{für } l \rightarrow \infty$$

von u_J^ν gegen die Lösung u_J von (3.10) konvergiert. Aus der Stetigkeit des Glättungsoperators \mathcal{S} folgt dann

$$(3.19) \quad \mathcal{S}(u_J^{\nu_l}) \rightarrow \mathcal{S}(u_J^*) \quad \text{für } l \rightarrow \infty$$

und die Monotonie der Iteration impliziert

$$(3.20) \quad \mathcal{J}(u_J^{\nu_l+1}) \leq \mathcal{J}(u_J^{\nu_l+1}) \leq \mathcal{J}(\mathcal{S}(u_J^{\nu_l})) \leq \mathcal{J}(u_J^{\nu_l}) \quad \text{für } \nu_l = 1, 2, \dots$$

Die Stetigkeit von \mathcal{J} und (3.18) – (3.20) führen schließlich zu

$$\mathcal{J}(\mathcal{S}(u_j^*)) = \mathcal{J}(u_j^*).$$

Man sieht leicht, dass dies zu $\mathcal{S}(u_j^*) = u_j^*$ äquivalent ist. Nach Satz 2.6.1 ist u_j aber der einzige Fixpunkt des projektiven Gauss–Seidel–Verfahrens. Damit folgt $u_j^* = u_j$. Wegen

$$\mathcal{J}(u_j^{\nu_l+1}) \leq \mathcal{J}(w_j^{\nu_l}) \leq \mathcal{J}(u_j^{\nu_l})$$

für $j = 1, \dots, m^\nu$ folgt sogar die Konvergenz der Zwischeniterierten

$$(3.21) \quad w^{\nu,j} \rightarrow u_j \quad \text{für } \nu \rightarrow \infty.$$

□

3.2.3 Asymptotisches Konvergenzverhalten

Im folgenden soll das asymptotische Konvergenzverhalten des approximativen erweiterten Relaxationsverfahrens aus (3.17) genauer untersucht werden.

Definition 3.2.4 (Kontaktmenge) Die Kontaktmenge von $v \in S_J$ ist definiert als

$$K_J^\bullet(v) := \{i \in \{1, \dots, n_J\} : v_i = g_i\}.$$

Die Menge ohne Kontakt wird mit

$$K_J^\circ(v) := \{i \in \{1, \dots, n_J\} : v_i \neq g_i\}$$

bezeichnet.

In der Menge $K_J^\bullet(u_J)$ berührt die Lösung u_J das Hindernis g , wohingegen sie in der Menge $K_J^\circ(u_J)$ die Variationsgleichung

$$a(u_J, v - u_J) = \ell(v - u_J) \quad \text{für alle } v \in \mathcal{K}_J$$

erfüllt. Im folgenden soll ausgeschlossen werden, dass es Punkte in der Kontaktmenge $K_J^\bullet(u_J)$ gibt, an denen u_J die Variationsgleichung erfüllt. Es wird also gefordert, dass die Hindernisfunktion an allen Punkten der Kontaktmenge einen restriktiven Einfluss auf die Lösung u_J besitzt. Anderenfalls wird das Problem als *degeneriert* bezeichnet. In Abbildung 3.1 ist der Unterschied zwischen nichtdegenerierten und degenerierten Hindernisproblemen veranschaulicht.

Definition 3.2.5 (Nicht–degeneriertes Minimierungsproblem) Wir nennen das diskrete Minimierungsproblem aus (3.10) nicht–degeneriert, falls

$$a(u_J, N_{i,\Delta_J}) < \ell(N_{i,\Delta_J}) \quad \text{für alle } i \in K_J^\bullet(u_J).$$

Als Verallgemeinerung von Satz 2.6.1 (ii) lässt sich nachweisen, dass die Kontaktmenge nach einer bestimmten Anzahl an Iterationen richtig bestimmt ist.

Lemma 3.2.6 Das diskrete Minimierungsproblem aus (3.10) sei nicht–degeneriert. Dann gibt es ein $\nu_0 \geq 0$, so dass

$$K_J^\bullet(u_J^\nu) = K_J^\bullet(u_J) \quad \text{für alle } \nu \geq \nu_0.$$

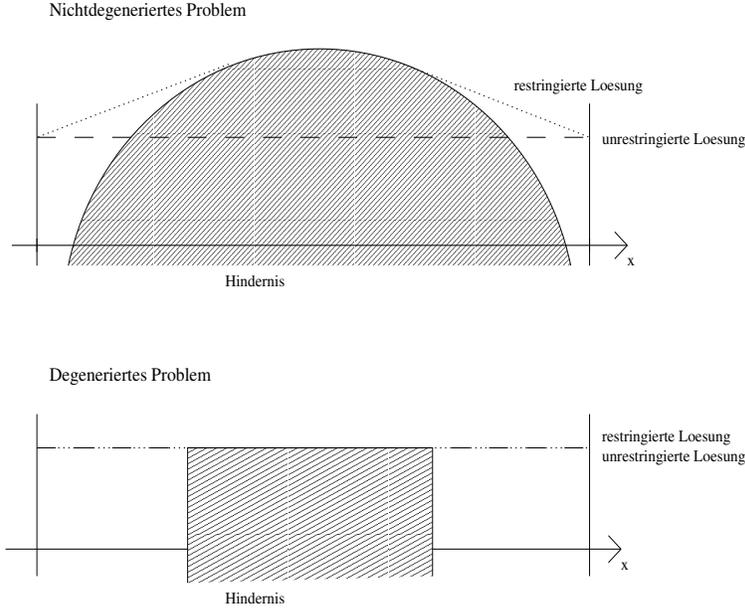


Abbildung 3.1: Beispiel für ein nichtdegeneriertes und für ein degeneriertes Hindernisproblem.

Beweis: Der Beweis ist in Anlehnung an [Kr], S.38 geführt. Es sei $i \in K_J^\circ(u_J)$. Wegen Satz 3.2.3 konvergiert $u_J^\nu \rightarrow u_J$ für $\nu \rightarrow \infty$ und damit ist $g_i - (u_J^\nu)_i > 0$ für alle $\nu \geq \nu_1$ für ein hinreichend großes $\nu_1 \geq 0$. Dies impliziert

$$(3.22) \quad K_J^\circ(u_J) \subset K_J^\circ(u_J^\nu) \quad \text{für alle } \nu \geq \nu_1.$$

Ist $i \in K_J^\bullet(u_J)$, dann besagt die Voraussetzung, dass das Problem nicht-degeneriert ist, dass

$$a(u_J, N_{i,\Delta_J}) < \ell(N_{i,\Delta_J}).$$

Wegen der Konvergenz der Zwischeniterierten aus (3.21) gilt dann auch

$$(3.23) \quad a(w^{\nu,j}, N_{i,\Delta_J}) < \ell(N_{i,\Delta_J}) \quad \text{für alle } \nu \geq \nu_2$$

und alle $j = 1, \dots, m^\nu$ mit einem hinreichend großem $\nu_2 \geq \nu_1$.

Angenommen es wäre $i \notin K_J^\bullet(w^{\nu,j})$, dann gilt für die Entwicklungskoeffizienten

$$w_i^{\nu,j} = w_i^{\nu,j-1} + v_i^{*,\nu,j} < g_i$$

für alle $j = 1, \dots, n_J$. Wegen (3.13) und Satz A.3.1 folgt hieraus, dass die Korrektur $v^{*,\nu,j} \in \text{span}\{\mu^{\nu,j}\}$ die Variationsgleichung

$$a(v^{*,\nu,j}, v) = \ell(v) - a(w^{\nu,j-1}, v) \quad \text{für alle } v \in \text{span}\{\mu^{\nu,j}\}$$

erfüllt. Wegen $w^{\nu,j} = w^{\nu,j-1} + v_j^*$ ist dann aber

$$a(w^{\nu,j}, v) = \ell(v) \quad \text{für alle } v \in \text{span}\{\mu^{\nu,j}\}$$

im Widerspruch zu (3.23). Damit ist $K_J^\bullet(u_J) \subset K_J^\bullet(w^{\nu,j})$ für alle $\nu \geq \nu_2$ und wegen $w^{\nu,n_J} = \bar{u}_J^\nu = \mathcal{S}(u_J^\nu)$ insbesondere für $j = n_J$ die Inklusion

$$K_J^\bullet(u_J) \subset K_J^\bullet(\bar{u}_J^\nu) \quad \text{für alle } \nu \geq \nu_2$$

gezeigt. Für $\nu \geq \nu_2$ werden Entwicklungskoeffizienten der geglätteten Näherung \bar{u}_J^ν mit Kontakt, d.h. Entwicklungskoeffizienten mit $(\bar{u}_J^\nu)_i = g_i$, in der Iteration nicht wieder inaktiviert. Daher folgt $K_J^\bullet(\bar{u}_J^\nu) \subset K_J^\bullet(u_J^{\nu+1})$ für $\nu \geq \nu_2$. Dies impliziert

$$(3.24) \quad K_J^\bullet(u_J) \subset K_J^\bullet(u_J^\nu) \quad \text{für alle } \nu \geq \nu_2 + 1.$$

Aus (3.22) und (3.24) folgt nun die Behauptung mit $\nu_0 := \nu_2 + 1$. \square

Definition 3.2.7 (Reguläre Suchmenge) *Es bezeichne ν_0 den Iterationsschritt aus Lemma 3.2.6, ab dem die Kontaktmenge der Lösung u_J richtig identifiziert ist. Eine Folge an Suchrichtungen $\{M^\nu\}_{\nu \in \mathbb{N}}$ heißt regulär, falls sie konstant ist, sobald die Kontaktmenge richtig identifiziert ist, d.h. $M^\nu = M = \Sigma_J \cup M_c$ für alle $\nu \geq \nu_0$ mit einer von ν unabhängigen Menge $M_c \subset S_J$.*

Sobald die Kontaktmenge der Lösung u_J identifiziert ist, lässt sich die Variationsungleichung aus (3.11) als *Variationsgleichung*

$$(3.25) \quad \text{finde } u_J \in S_J^\circ : a(u_J, v) = \ell(v) \quad \text{für alle } v \in S_J^\circ$$

mit dem *reduzierten Teilraum*

$$(3.26) \quad S_J^\circ := \{v \in S_J : v_i = 0 \quad \text{für } i \in K_J^\bullet(u_J)\} \subset S_J$$

schreiben. Die *reduzierte Menge an Suchrichtungen*

$$(3.27) \quad M^\circ := \{\mu \in M : \mu_i = 0 \quad \text{für } i \in K_J^\bullet(u_J)\} \subset M$$

induziert dann eine sogenannte lineare multiplikative Schwarz-Methode zur iterativen Lösung der Variationsgleichung aus (3.25). Die Korrektur $v^{\nu,j}$ in Richtung von $\mu^{\nu,j} \in M^\circ$ ist die Lösungen des linearen, lokalen Teilproblems

$$(3.28) \quad \text{finde } v^{\nu,j} \in \text{span}\{\mu^{\nu,j}\} : \mathcal{J}(w^{\nu,j-1} + v^{\nu,j}) \leq \mathcal{J}(w^{\nu,j-1} + v) \quad \text{für alle } v \in \text{span}\{\mu^{\nu,j}\}.$$

Falls das Problem nicht-degeneriert ist, reduziert sich das auf die Variationsungleichung aus (3.11) angewandte erweiterte Relaxationsverfahren wegen Satz 2.6.1 asymptotisch zur linearen erweiterten Relaxation aus (3.28). Soll dies auch für das approximative erweiterte Relaxationsverfahren gelten, müssen die lokalen approximativen Hindernisse $\psi^{\nu,j}$ so konstruiert sein, dass sie alle Korrekturen des linearen Schemas in Richtung von $\mu^{\nu,j} \in M^\circ$ zulassen. Solche Hindernisse werden als *quasioptimal* bezeichnet.

Definition 3.2.8 (Quasioptimale lokale Hindernisse) *Wir nennen eine Folge lokaler Hindernisse $\{\psi^{\nu,j}\}_{\nu \in \mathbb{N}}$ quasioptimal, falls die Konvergenz der Zwischeniterierten $w^{\nu,j} \rightarrow u_J$ und der zugehörigen Kontaktmengen $K_J^\bullet(w^{\nu,j}) \rightarrow K_J^\bullet(u_J)$ implizieren, dass es ein $c > 0$ und ein $\nu_0 \geq 0$ gibt, so dass für alle $\mu^{\nu,j} \in M^\circ$*

$$0 < c < \psi^{\nu,j} \quad \text{für alle } \nu \geq \nu_0$$

gilt.

Während monotone Hindernisse eine obere Schranke erfüllen, die die Robustheit des Verfahrens sicherstellt, erfüllen quasioptimale Hindernisse eine untere Schranke, die sicherstellt, dass sich das Verfahren asymptotisch zu einer linearen Relaxation reduziert. Der Begriff der Quasioptimalität definiert also eine Mindestanforderung, der die Hindernisse im folgenden genügen sollen. Während die Hindernisse aus [Ma] genau dieser Mindestanforderung entsprechen, liegen die Hindernisse in [Ko1] oft sogar oberhalb dieser Mindestanforderung. Die Konstruktion wird ausführlich in Kapitel 4 für B-Splines beliebiger Glattheit diskutiert.

Nach dem folgenden Satz besitzen das erweiterte Relaxationsverfahren und alle approximativen erweiterten Relaxationsverfahren, die von derselben regulären Folge $\{M^\nu\}_{\nu \in \mathbb{N}}$ an Suchrichtungen und beliebigen quasioptimalen Hindernissen induziert werden, das gleiche lineare asymptotische Konvergenzverhalten.

Satz 3.2.9 *Das diskrete Minimierungsproblem aus (3.10) sei nicht-degeneriert. Dann konvergiert das durch eine reguläre Folge $\{M^\nu\}_{\nu \in \mathbb{N}}$ von Suchrichtungen und monotone und quasioptimale Hindernisse $\{\psi^{\nu,j}\}_{\nu \in \mathbb{N}}$ induzierte approximative erweiterte Relaxationsverfahren für alle Startiterierten $u_J^0 \in S_J$ gegen die Lösung u_J von (3.10) und reduziert sich asymptotisch zum linearen erweiterten Relaxationsverfahren aus (3.28).*

Beweis: Die globale Konvergenz folgt sofort aus Satz 3.2.3. Nach Lemma 3.2.6 ist nach ν_0 Iterationen die Kontaktmenge $K_J^\bullet(u_J)$ identifiziert. Da $\{M^\nu\}_{\nu \in \mathbb{N}}$ regulär ist, gilt $M^\nu = M$ für alle $\nu \geq \nu_0$.

Betrachte für $\nu \geq \nu_0$ zunächst Korrekturen $v^{\nu,j}$ in Richtung von $\mu^{\nu,j} \in M^\circ$ ohne Einfluss auf die Kontaktmenge. Wegen $v^{\nu,j} \rightarrow 0$ für $\nu \rightarrow \infty$ und da $\{\psi^{\nu,j}\}_{\nu \in \mathbb{N}}$ quasioptimal ist, gibt es ein $\nu_1 \geq \nu_0$ mit

$$v^{\nu,j} < c < \psi^{\nu,j} \quad \text{für alle } \nu \geq \nu_1.$$

Die Korrekturen $v^{\nu,j}$ erfüllen damit für alle $\nu \geq \nu_1$ das unrestringierte Minimierungsproblem aus (3.28).

Ist dagegen $\mu^{\nu,j} \in M \setminus M^\circ$, dann sind die entsprechenden Korrekturen $v^{\nu,j} = 0$, sobald die Kontaktmenge $K_J^\bullet(u_J)$ identifiziert ist, d.h. für alle $\nu \geq \nu_0$. Der Beweis ist [K01], S. 10 entnommen. \square

Im Konvergenzverhalten lassen sich somit zwei Phasen unterscheiden. Dies ist zum einen die *transiente Phase*, die von der Suche nach der Kontaktmenge der Lösung dominiert ist, und zum anderen die *asymptotische Phase*, in der das reduzierte lineare Schema gelöst wird.

3.2.4 Abgeschnittene Version der erweiterten Relaxationsverfahren

Insbesondere wenn gute Startnäherungen zur Verfügung stehen, ist die Gesamtkonvergenz des Verfahrens häufig von Anfang an von der asymptotischen Phase dominiert. Dort wird das Verfahren um so effektiver sein, je größer die reduzierte Menge M° an Basisfunktionen ist, die zur Lösung des reduzierten linearen Systems verwendet werden können. Da alle Grobgitterfunktionen, die zu Korrekturen an Kontaktpunkten führen, bisher nicht in der Menge M° enthalten sind, ist die Konvergenz des Mehrgitterverfahrens unter Umständen deutlich schlechter als im unrestringierten Fall. Aus diesem Grund sollen im folgenden solche Grobgitterfunktionen in einer geeigneten Weise an die aktuelle Position der Kontaktmenge $K_J^\bullet(\bar{u}_J^\nu)$ angepasst werden. Dies führt zu einer größeren Menge an Suchrichtungen M° , was wiederum durch den Multiskaleneffekt mit schnellerer Konvergenz verbunden ist. Das resultierende Verfahren wird als *abgeschnittene Version (truncated version)* des erweiterten Relaxationsverfahrens bezeichnet. Die Menge an Suchrichtungen hängt dann in jedem Iterationsschritt von der Position der Kontaktmenge $K_J^\bullet(\bar{u}_J^\nu)$ ab, die durch die Basisfunktionen des feinsten Gitters bestimmt wird. Die Konstruktion der modifizierten Suchrichtungen erfolgt über einen Abschneideoperator. Dies wird im folgenden erläutert.

Präziser als in (3.12) wird die Menge der Grobgitterfunktionen M_c zunächst unabhängig vom Iterationsschritt ν als

$$(3.29) \quad M_c := \{\mu^{n_{J+1}}, \dots, \mu^m\} := \Sigma_{J-1} \cup \Sigma_{J-2} \cup \dots \cup \Sigma_1$$

mit den B-Spline Basen $\Sigma_\ell := \{N_{1,\Delta_\ell}, \dots, N_{n_\ell,\Delta_\ell}\} \subset S_\ell$ dargestellt. Dabei sei Δ_1 ein äquidistantes Gitter mit Gitterweite h_1 und Δ_ℓ ein entsprechend verfeinertes Gitter mit Gitterweite

$$h_\ell = 2^{-\ell+1} h_1.$$

Für $k = 2$, d.h. für Hutfunktionen, ist die Anzahl n der Entwicklungskoeffizienten gleich der Anzahl m der Stützstellen. Allgemein gilt

$$n = m + k - 2$$

für B-Splines der Ordnung k . Die Anzahl $n_{\ell,k}$ der Entwicklungskoeffizienten in Abhängigkeit vom Level ℓ und der Ordnung k beträgt

$$n_{\ell,k} = k 2^{\ell-1} + k - 1.$$

Die konstante Menge M_c wird über einen Abschneideoperator in jedem Iterationsschritt an die aktuelle Kontaktmenge angepasst. Die so modifizierte Menge an Suchrichtungen wird mit \tilde{M}_c^ν bezeichnet. Die Ausgangspunkt für die Konstruktion ist die Definition von Kontaktmengen auf dem Level $\ell < J$. Sie sind rekursiv in Abhängigkeit der Kontaktmenge $K_J^\bullet(\bar{u}_J^\nu)$ auf dem feinstem Level J (s. Definition 3.2.4) als

$$(3.30) \quad K_\ell^\bullet(\bar{u}_J^\nu) := \{i \in \{1, \dots, n_\ell\} : 2i - (k-1) \in K_{\ell+1}^\bullet(\bar{u}_J^\nu)\}$$

definiert. Für $v \in S_\ell$ sei weiter der Operator $t_\ell : S_\ell \rightarrow S_\ell^\nu$ definiert durch

$$(t_\ell v)_i = \begin{cases} 0 & \text{falls } i \in K_\ell^\bullet(\bar{u}_J^\nu) \\ v_i & \text{sonst} \end{cases}$$

für $i = 1, \dots, n_\ell$ mit den an die aktuelle Kontaktmenge angepassten Räumen

$$S_\ell^\nu := \{v \in S_\ell : v_i = 0 \text{ für } i \in K_\ell^\bullet(\bar{u}_J^\nu), i = 1, \dots, n_\ell\} \subset S_\ell.$$

Damit kann nun der Abschneideoperator $T_\ell : S_\ell \rightarrow S_J^\nu$ als

$$(3.31) \quad T_\ell := t_J \circ p_{J-1}^J \circ \dots \circ t_{\ell+1} \circ p_\ell^{\ell+1}$$

definiert werden. Hierbei bezeichnet $p_\ell^{\ell+1}$ den kanonischen Interpolationsoperator von S_ℓ nach $S_{\ell+1}$ (Gittertransferoperatoren für B-Splines sind im nächsten Abschnitt dargestellt).

Zur Veranschaulichung ist in Abbildung 3.2 die Basisfunktion $N_{1,2,\Delta_{J-2}} \in S_{J-2}$ und die abgeschnittenen Basisfunktion

$$T_{J-1} N_{1,2,\Delta_{J-2}} := t_{J-1} \circ p_{J-2}^{J-1} \circ N_{1,2,\Delta_{J-2}} \in S_{J-1}^\nu$$

und

$$T_{J-2} N_{1,2,\Delta_{J-2}} := t_{J-1} \circ p_{J-2}^{J-1} \circ T_{J-1} N_{1,2,\Delta_{J-2}} \in S_J^\nu$$

für den Fall von zwei Kontaktpunkten an den Stellen $x = 4$ und $x = 7$ dargestellt. Dabei ist das feinste Gitter $\Delta_J \subset \mathbb{Z}$ und das gröbste Gitter $\Delta_{J-2} \subset 4\mathbb{Z}$.

Bemerkung 3.2.10 Die Definitionen der Kontaktmengen aus (3.30) und der Abschneideoperatoren aus (3.31) entsprechen den Definitionen aus [Ko1] für den Fall $k = 2$.

Die abgeschnittene Version ist nun dadurch definiert, dass die an die Kontaktmenge angepasste Menge

$$(3.32) \quad \tilde{M}_c^\nu := \{T_\ell N_{i,k,\Delta_\ell} \text{ mit } i = 1, \dots, n_\ell, \ell = 0, \dots, J-1\}$$

an Grobgitter-Suchrichtungen zur Bestimmung der Grobgitterkorrekturen verwendet wird.

Die reduzierte Menge an Suchrichtungen (s. (3.27)) sei mit \tilde{M}° bezeichnet. Da offensichtlich $M^\circ \subset \tilde{M}^\circ$ gilt, können bei der abgeschnittenen Version verbesserte Konvergenzraten in der asymptotischen Phase erwartet werden. Weil variable Suchrichtungen $\{M^\nu\}_{\nu \in \mathbb{N}}$ von Anfang an mit berücksichtigt worden sind, lassen sich die Ergebnisse von Satz 3.2.3, Lemma 3.2.6 und Satz 3.2.9 direkt auf die abgeschnittene Version übertragen.

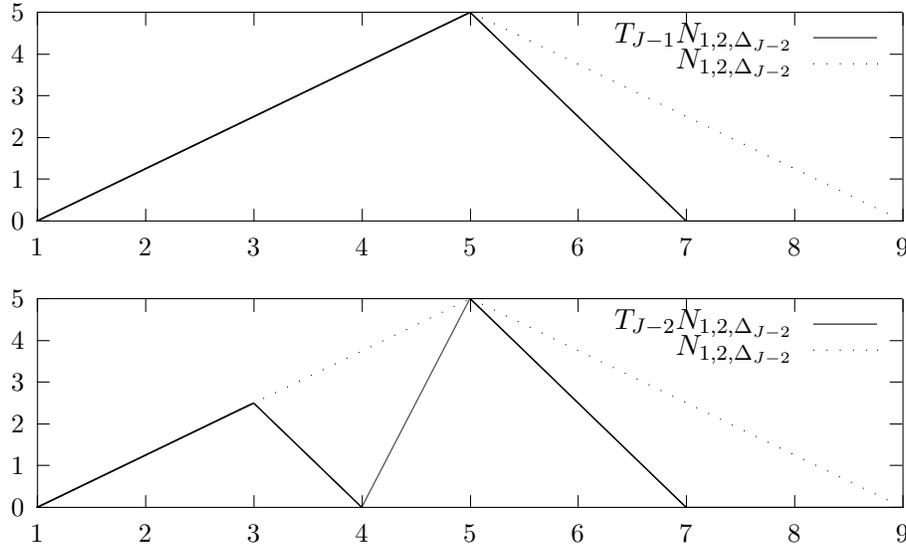


Abbildung 3.2: Hutfunktion $N_{1,2,\Delta_{J-2}}$ und abgeschnittene Basisfunktion $T_{J-1}N_{1,2,\Delta_{J-2}}$ (links) bzw. $T_{J-2}N_{1,2,\Delta_{J-2}}$ (rechts) bei den beiden Kontaktpunkten $x = 4, 7$.

3.3 Monotone Mehrgitterverfahren (MMG)

Nach den vorhergehenden Abschnitten kann das *monotone Mehrgitterverfahren (MMG)* aus Algorithmus 3.1.4 als ein approximatives erweitertes Relaxationsverfahren mit der konstanten Menge an Suchrichtungen $M := \Sigma_J \cup M_c$ mit M_c aus (3.29) aufgefasst werden. Auf diese Weise wurde die globale Konvergenz und die asymptotische Reduktion der Verfahren zu einer linearen Relaxation nachgewiesen. Insbesondere wurden über diese Betrachtungsweise die Forderungen nach Monotonie und Quasioptimalität der lokalen Hindernisse motiviert.

In diesem Abschnitt werden zunächst konkrete Hinweise zur Implementierung der monotonen Mehrgitterverfahren aus Algorithmus 3.1.4 und der abgeschnittenen Version gegeben. Anschließend wird ein endgültiger Konvergenzsatz und asymptotische Fehlerabschätzungen für die Verfahren geliefert.

3.3.1 Implementierung des Verfahrens

Um die beschriebenen B-Spline-basierten Mehrgitterverfahren implementieren zu können, werden zunächst geeignete Gittertransferoperatoren für B-Splines der Ordnung k benötigt. Die Unterteilungsformel aus (A.10) liefert durch

$$(3.33) \quad p(N_{i,\Delta_{\ell-1}}) := \sum_{j=0}^k 2^{1-k} \binom{k}{j} N_{i+j,\Delta_{\ell}}$$

einen kanonischen Prolongationsoperator $p = p_{\ell-1}^{\ell}$ von $S_{\ell-1}$ nach S_{ℓ} für B-Splinefunktionen $N_{i,\Delta_{\ell-1}}$ der Ordnung k . [H2] folgend, wird als gewichteter Restriktionsoperator $r = r_{\ell}^{\ell-1}$ die Adjungierte p^* von p bezüglich des vom Gitter Δ_{ℓ} abhängigen Skalarproduktes

$$\langle u_{\ell}, v_{\ell} \rangle := \sum_{x \in \Delta_{\ell}} h_{\ell} u_{\ell}(x) v_{\ell}(x) \quad \text{für } u_{\ell}, v_{\ell} \in S_{\ell}$$

gewählt. In unserem Fall $h_{\ell-1} = 2h_{\ell}$ rechnet man leicht nach, dass $p^* = \frac{1}{2}p^T$ gilt.

Auf diese Weise wird der Defekt, die Massen- und die Steifigkeitsmatrix auf gröbere Gitter transportiert. Der Operator kann allerdings nicht für die Konstruktion der monotonen und quasioptimalen lokalen Hindernisse $\psi^{\nu,j}$ verwendet werden, da er die Voraussetzungen aus Satz 3.3.1 nicht erfüllt.

Anstelle dessen werden zunächst monotone und quasioptimale Approximationen $\Psi_\ell \in S_\ell$ für $\ell < J$ über eine geeignete, sukzessive Restriktion der exakten Hindernisfunktion

$$\Psi_J := g - \bar{u}'_J \in S_J$$

konstruiert. Die Konstruktion wird ausführlich in Kapitel 4 beschrieben. Besitzt die Funktion $\Psi_\ell \in S_\ell$ bei Entwicklung in der Basis $\Sigma_\ell \subset S_\ell$ die Darstellung

$$\Psi_\ell = \sum_{i=1}^{n_\ell} c_i N_{i,\Delta_{n_\ell}},$$

dann können für $\mu^{\nu,j} \in M$ die lokalen Hindernisse durch

$$\psi^{\nu,j} := c_i N_{i,\Delta_{n_\ell}} \in \text{span}\{\mu^{\nu,j}\}$$

definiert werden. Dabei sei erinnert, dass es nach Konstruktion der Menge M für alle $j = 1, \dots, m$ ein $\ell \in \{1, \dots, J\}$ und ein $i \in \{1, \dots, n_\ell\}$ gibt, so dass $\mu^{\nu,j} = N_{i,\Delta_{n_\ell}}$ gilt. Dass sich auf diese Weise wirklich monotone und quasioptimale lokale Hindernisse im Sinne der Definitionen 3.2.2 bzw. 3.2.8 ergeben, wird ebenfalls in Kapitel 4 in Lemma 4.2.7 gezeigt.

Unter Verwendung der oben beschriebenen Gittertransfer- und Restriktionsoperatoren, kann das monotone Mehrgitterverfahren jetzt als Modifikation eines klassischen linearen Mehrgitterverfahrens implementiert werden.

Der Glätter, das klassische Gauss-Seidel-Verfahren, muss um den Projektionsschritt zum projektiven Gauss-Seidel-Verfahren ergänzt werden. Massen-, Steifigkeitsmatrix und Defekt werden mit gewichteten Restriktionen bestimmt. Die Korrektur wird kanonisch mit Interpolation auf das feinste Gitter zurücktransportiert.

Für die abgeschnittene Version (TrMMG) des monotonen Mehrgitterverfahrens müssen außerdem folgende Modifikation durchgeführt werden.

Beiträge von Entwicklungskoeffizienten mit Index $i \in K_J^\bullet(\bar{u}'_J)$ werden bei der Berechnung des Defekts und der Steifigkeitsmatrix zu Null gesetzt. Die Entwicklungskoeffizienten der Hindernisfunktion werden für $i \in K_J^\bullet(\bar{u}'_J)$ auf plus (bei einem oberen Hindernis) bzw. minus (bei einem unteren Hindernis) unendlich gesetzt, so dass sie keinen restriktiven Einfluss mehr auf die Bestimmung der Korrekturen haben. Dies ist in Abbildung 4.8 bzw. 4.9 in Abschnitt 4.4.3 für den Fall stückweise linearer bzw. quadratischer Ansatzfunktionen veranschaulicht. Korrekturen werden auf jedem Level ℓ nur zu Entwicklungskoeffizienten $(u'_\ell)_i$ addiert, für die $i \notin K_J^\bullet(\bar{u}'_J)$ gilt.

3.3.2 Asymptotische Fehlerabschätzungen

Für das monotone Mehrgitterverfahren und die abgeschnittene Version lassen sich ohne Annahmen an die Regularität des freien Randes oder an die Koeffizienten in der Bilinearform $a(\cdot, \cdot)$ folgende asymptotische Fehlerabschätzung für Gebiete $\Omega \subset \mathbb{R}^2$ angeben.

Satz 3.3.1 *Das diskrete Minimierungsproblem aus (3.10) sei nicht-degeneriert, und es seien monotone und quasioptimale Hindernisse $\{\psi^{\nu,j}\}_{\nu \in \mathbb{N}}$ gegeben. Dann konvergiert das durch die Menge $M := \Sigma_J \cup M_c$ mit M_c aus (3.29) induzierte monotone Mehrgitterverfahren als auch*

die durch die Menge $\tilde{M}^\nu := \Sigma_J \cup \tilde{M}_c^\nu$ mit \tilde{M}_c^ν aus (3.32) induzierte abgeschnittene Version für alle Startiterierten $u_J^0 \in S_J$ gegen die Lösung u_J von (3.10) und es gilt in der Energienorm $\|v\| := \sqrt{a(v,v)}$ für ein hinreichend großes $\nu_0 \geq 0$ die Fehlerabschätzung

$$(3.34) \quad \|u_J - u_J^{\nu+1}\| \leq \left(1 - \frac{c}{(J+1)^4}\right) \|u_J - u_J^\nu\| \quad \text{für alle } \nu \geq \nu_0$$

für das monotone Mehrgitterverfahren und

$$(3.35) \quad \|u_J - u_J^{\nu+1}\| \leq \left(1 - \frac{c}{(J+1)^6}\right) \|u_J - u_J^\nu\| \quad \text{für alle } \nu \geq \nu_0$$

für die abgeschnittene Version. Die Konstante $0 \leq c < 1$ hängt nur von der Elliptizität von $a(\cdot, \cdot)$ und vom Gitter Δ_1 ab.

Beweisskizze: Die Mengen M, \tilde{M}^ν definieren reguläre Folgen an Suchrichtungen. Damit folgt die globale Konvergenz des monotonen Mehrgitterverfahrens und der abgeschnittenen Version aus Satz 3.2.9. Außerdem folgt aus dem Satz, dass sich die Verfahren asymptotisch zur linearen multiplikativen Schwarz-Methode zur Lösung des reduzierten linearen Problems aus (3.28) reduzieren, die durch die Zerlegung

$$S_J^\circ = \sum_{\mu \in M^\circ} \text{span } \mu$$

induziert wird. Für $v \in S_J^\circ$ gilt also die Darstellung $v = \sum_{\mu \in M^\circ} v_\mu$ mit $v_\mu \in \text{span } \mu$. Der Arbeit [GO] folgend wird für $v \in S_J^\circ$ die Norm

$$\|v\|^2 := \inf \left\{ \sum_{\mu \in M^\circ} \|v_\mu\|^2 : v_\mu \in \text{span } \mu, v = \sum_{\mu \in M^\circ} v_\mu \in S_J^\circ \right\}$$

eingeführt. Resultate aus [KY] liefern dann die Abschätzung

$$\|v\|^2 \lesssim (J+1)^2 \|v\|^2$$

nach oben und aus der verschärften Cauchy-Schwarz Ungleichung folgt die Abschätzung

$$\|v\|^2 \lesssim \|v\|^2$$

nach unten. Die beiden Abschätzungen führen schließlich nach Satz 4 aus [GO] zu den behaupteten Fehlerabschätzungen aus (3.34) und (3.35). \square

Bemerkung 3.3.2 Die Fehlerschranken sind nicht gitterweitenunabhängig. Wegen $J \leq |\log h_J|$ wächst die Schranke aber nur logarithmisch und damit nur sehr mäßig. Ob sich der logarithmische Faktor in den Fehlerabschätzungen vermeiden lässt, ist noch eine offene Frage.

Bemerkung 3.3.3 Numerische Tests zeigen, dass die Abschätzungen den schlechtesten Fall beschreiben. Insbesondere ist die Konvergenz der abgeschnittenen Version TrMMG in numerischen Tests immer schneller als die des Standardverfahrens MMG (s. [Ko2] und Kapitel 6). Man stellt sogar fest, dass die abgeschnittene Version bei restringierten Minimierungsproblemen die gleiche asymptotische Effizienz erreicht wie bei unrestringierten Minimierungsproblemen (s. [Ko2] und Kapitel 6).

Kapitel 4

Konstruktion monotoner und quasioptimaler Grogitterapproximationen

Um die Konstruktion der monotonen Mehrgitterverfahren aus Kapitel 3 abzuschließen, werden noch monotone und quasioptimale Restriktionsoperatoren für die Hindernisfunktion benötigt. Die Konstruktion solcher Grogitterapproximationen an die Hindernisfunktion ist der zentrale Inhalt dieses Kapitels.

Der Übersichtlichkeit halber werden nur zwei Gitter betrachtet. Die Verallgemeinerung auf mehrere Gitter ist aber offensichtlich. Zu vorgegebener Ordnung k und zu vorgegebenem $n \in \mathbb{N}$ sei $T := \{\theta_i\}_{i=1, \dots, n+k}$ eine äquidistante erweiterte Knotenfolge (s. Anhang A.1)

$$(4.1) \quad \theta_1 = \dots = \theta_k = a < \theta_{k+1} < \dots < \theta_n < b = \theta_{n+1} = \dots = \theta_{n+k}$$

in $I := [a, b] \subset \mathbb{R}$ mit Gitterweite $H = \theta_{n+1} - \theta_n = \dots = \theta_{k+1} - \theta_k$. Weiter sei $\Delta := \{\tilde{\theta}_i\}_{i=1, \dots, \tilde{n}+k}$ die verfeinerte Knotenfolge

$$(4.2) \quad \tilde{\theta}_1 = \dots = \tilde{\theta}_k = a < \tilde{\theta}_{k+1} < \dots < \tilde{\theta}_{\tilde{n}} < b = \tilde{\theta}_{\tilde{n}+1} = \dots = \tilde{\theta}_{\tilde{n}+k}$$

mit Gitterweite $h = \frac{1}{2}H$ und $\tilde{\theta}_{k+2j} = \theta_{k+j}$ und $\tilde{\theta}_{k+2j-1} = \frac{1}{2}(\theta_{k+j-1} + \theta_{k+j})$ für $j = 1, \dots, n-k$. Dann ist

$$(4.3) \quad \tilde{n} = 2n + 1 - k.$$

Bezüglich der Knotenfolgen Δ bzw. T sei $\mathcal{N}_{k,\Delta}$ bzw. $\mathcal{N}_{k,T}$ der in Anhang A.1 definierte Raum der B-Splines der Ordnung k . Weiter sei $\tilde{S} \in \mathcal{N}_{k,\Delta}$ eine auf dem feinen Gitter Δ definierte Hindernisfunktion mit Koeffizienten $\tilde{\mathbf{c}} = (\tilde{c}_1, \dots, \tilde{c}_{\tilde{n}})^T$. In Hinblick auf Bemerkung 3.16 wird $\tilde{c}_i \geq 0$ für $i = 1, \dots, \tilde{n}$ gefordert. Schließlich sei $S \in \mathcal{N}_{k,T}$ eine Splinefunktion auf dem groben Gitter T mit den Entwicklungskoeffizienten $\mathbf{c} = (c_1, \dots, c_n)^T$. Es gelten also die Darstellungen

$$(4.4) \quad \tilde{S}(\theta) = \sum_{i=1}^{\tilde{n}} \tilde{c}_i N_{i,k,\Delta}(\theta), \quad S(\theta) = \sum_{i=1}^n c_i N_{i,k,T}(\theta).$$

Abkürzend wird auch die Schreibweise als Skalarprodukt

$$\tilde{S} = \tilde{\mathbf{c}}^T \mathbf{N}_{k,\Delta}, \quad S = \mathbf{c}^T \mathbf{N}_{k,T}$$

verwendet. Hierbei wird die Menge der Basisfunktionen $\{N_{i,k,T} : i = 1, \dots, n\}$ als ein geordneter Vektor $\mathbf{N}_{k,T} \in \mathbb{R}^n$ aufgefasst.

4.1 Monotone Grobgitterapproximationen

Definition 4.1.1 (Monotone Approximation) Die Funktion $f : I \rightarrow \mathbb{R}$ heißt monotone (oder einseitige) untere Approximation an die Funktion $g : I \rightarrow \mathbb{R}$, falls $f(x) \leq g(x)$ für alle $x \in I$ gilt.

Es gibt eine Fülle an Literatur (z.B. [Bu, DV, LP, Mv, Pi]) insbesondere aus der Approximationstheorie, die sich mit monotonen Approximationen an Funktionen beschäftigt. Meist wird eine bestimmte Approximationsgenauigkeit ϵ vorgegeben und die Anzahl n der Freiheitsgrade so gewählt, dass diese Genauigkeit erreicht werden kann.

Die Fragestellung in diesem Abschnitt ist dagegen anders. Hier soll zu einer vorgegebenen Anzahl n an Freiheitsgraden eine B-Splinefunktion $S \in \mathcal{N}_{k,T}$ gefunden werden, die eine "möglichst gute" monotone Approximation an eine gegebene Hindernisfunktion $\tilde{S} \in \mathcal{N}_{k,\Delta}$ liefert. Die Funktion \tilde{S} ist ebenfalls ein B-Spline der Ordnung k , aber mit einer höheren Anzahl $\tilde{n} > n$ an Freiheitsgraden. Eine solche Approximation wird im folgenden als *monotone Grobgitterapproximation* bezeichnet.

Definition 4.1.2 (Monotone Grobgitterapproximation) Es sei T die Knotenfolge aus (4.1) und Δ die verfeinerte Knotenfolge aus (4.2). Die Funktion $S \in \mathcal{N}_{k,T}$ heißt monotone untere Grobgitterapproximation an $\tilde{S} \in \mathcal{N}_{k,\Delta}$, falls $S(x) \leq \tilde{S}(x)$ für alle $x \in I$ gilt.

Wie in Kapitel 3 dargestellt wurde, sind solche Approximationen ein wichtiger Baustein von monotonen Mehrgitterverfahren zur Lösung von freien Randwertproblemen. Ihr Einsatz ist aber auch in anderen Anwendungen denkbar, z.B. zur Komplexitätsreduktion von Geometrien (s. [LP]). Wichtig ist, dass bei Anwendung im Rahmen eines Mehrgitterverfahrens der Aufwand zur Konstruktion einer monotonen Grobgitterapproximation $O(n)$ Operationen, mit n der Anzahl der Freiheitsgrade des Splines S bzgl. des groben Gitters, nicht überschreiten darf. Um eine möglichst effiziente Grobgitterkorrektur zu erhalten, sollte die Grobgitterfunktion die Hindernisfunktion "möglichst gut" approximieren. Die Fragestellung dieses Kapitels ist damit die folgende:

Konstruiere mit $O(n)$ Operationen zu einer gegebenen Hindernisfunktion $\tilde{S} \in \mathcal{N}_{k,\Delta}$ eine monotone untere bzw. obere Grobgitterapproximation $S \in \mathcal{N}_{k,T}$, welche die Funktion \tilde{S} im Hinblick auf eine noch zu spezifizierende Zielfunktion möglichst gut approximiert.

Bemerkung 4.1.3 Für Hutfunktionen sind in [Ma, Kol] Restriktionsoperatoren angegeben, die obige Forderungen erfüllen. Für B-Splines höherer Ordnung gibt es solche Operatoren bisher noch nicht. Die Konstruktion wird in diesem Kapitel diskutiert.

Ist $S \in \mathcal{N}_{2,T}$ eine stetige, stückweise lineare Funktion, θ_i und θ_{i+1} zwei aufeinanderfolgende Stützstellen und ist $x \in [\theta_i, \theta_{i+1}]$, dann gilt

$$\min \{S(\theta_i), S(\theta_{i+1})\} \leq S(x) \leq \max \{S(\theta_i), S(\theta_{i+1})\}.$$

Für stückweise Funktionen $S \in \mathcal{N}_{k,T}$ der Ordnung $k \geq 3$ ist dies im Allgemeinen aber nicht mehr richtig. Es reicht folglich nicht mehr aus, nur Funktionswerte an Stützstellen zu betrachten. Daraus resultiert eine Hauptschwierigkeit bei der Konstruktion monotoner Grobgitterapproximationen und der wesentliche Unterschied zum stückweise linearen Fall. Unter der Ausnutzung der Eigenschaften von B-Splines wird hier eine Konstruktion vorgeschlagen, die wie in Lemma 2.6.5 über die Entwicklungskoeffizienten der B-Splinefunktion erfolgt. Für Hutfunktionen reduziert sie sich auf die Konstruktionen in [Ma, Kol].

Ordnet man alle Summanden nach den Basisfunktionen $N_{i,k,\Delta}$ und führt einen Indexwechsel durch erhält man weiter

$$\begin{aligned} S(x) &= \sum_{\substack{i=1 \\ i \text{ ungerade}}}^{\tilde{n}} (a_{k-1} c_{(i+1)/2} + a_{k-3} c_{(i+3)/2} + \dots + a_1 c_{(i+k-1)/2}) N_{i,k,\Delta}(x) \\ &+ \sum_{\substack{i=2 \\ i \text{ gerade}}}^{\tilde{n}-1} (a_k c_{i/2} + a_{k-2} c_{(i+2)/2} + \dots + a_0 c_{(i+k)/2}) N_{i,k,\Delta}(x). \end{aligned}$$

Dabei sind alle c_j mit $j < 1$ oder $j > n$ wie Null zu behandeln. Die Darstellung der Differenz $\tilde{S}(x) - S(x)$ in der Basis von $\mathcal{N}_{k,\Delta}$ lautet dann

$$(4.7) \quad \tilde{S}(x) - S(x) = \sum_{i=1}^{\tilde{n}} d_i N_{i,k,\Delta}(x)$$

mit den Koeffizienten

$$d_i := \begin{cases} \tilde{c}_i - (a_{k-1} c_{(i+1)/2} + a_{k-3} c_{(i+3)/2} + \dots + a_1 c_{(i+k-1)/2}), & \text{falls } i \text{ ungerade ist} \\ \tilde{c}_i - (a_k c_{i/2} + a_{k-2} c_{(i+2)/2} + \dots + a_0 c_{(i+k)/2}), & \text{falls } i \text{ gerade ist.} \end{cases}$$

Wegen $N_{i,k,\Delta}(x) \geq 0$ für alle $x \in I$, folgt wie in Lemma 2.6.5 aus $d_i \geq 0$ für alle $i = 1, \dots, \tilde{n}$, dass $\tilde{S}(x) - S(x) \geq 0$ für alle $x \in I$ ist. Damit ist S monotone untere Grobgitterapproximation an \tilde{S} . Schreibt man die Bedingungen $d_i \geq 0$ für $i = 1, \dots, \tilde{n}$ in Matrixform, erhält man das Ungleichungssystem aus (4.5).

Ist k ungerade verläuft der Beweis analog. Man erhält

$$d_i := \begin{cases} \tilde{c}_i - (a_{k-1} c_{(i+1)/2} + a_{k-3} c_{(i+3)/2} + \dots + a_0 c_{(i+k)/2}), & \text{falls } i \text{ ungerade ist} \\ \tilde{c}_i - (a_k c_{i/2} + a_{k-2} c_{(i+2)/2} + \dots + a_1 c_{(i+k-1)/2}), & \text{falls } i \text{ gerade ist.} \end{cases}$$

Fasst man die beiden Fälle zusammen, erhält man für beliebige Ordnungen $k \geq 1$ die Koeffizienten d_i für $i = 1, \dots, \tilde{n}$ mit

$$(4.8) \quad d_i := \begin{cases} \tilde{c}_i - (a_{k-1} c_{(i+1)/2} + a_{k-3} c_{(i+3)/2} + \dots + a_0 c_{\lfloor (i+k)/2 \rfloor}), & \text{falls } i \text{ ungerade ist} \\ \tilde{c}_i - (a_k c_{i/2} + a_{k-2} c_{(i+2)/2} + \dots + a_1 c_{\lfloor (i+k)/2 \rfloor}), & \text{falls } i \text{ gerade ist.} \end{cases}$$

Hierbei bezeichnet $\lfloor x \rfloor$ für $x \in \mathbb{R}$ die größte ganze Zahl kleiner oder gleich x . □

Zur Illustration sind die Matrizen A_k für einige Spezialfälle explizit aufgelistet.

Beispiel 4.1.5 Für stetige, stückweise lineare ($k = 2$), C^1 -glatte, stückweise quadratische ($k = 3$) und C^2 -glatte, stückweise kubische ($k = 4$) Splines erhält man die Matrizen

Beweis: Nach (A.11) sind die Zeilensummen der Matrix A_k immer gleich eins. Daher erfüllt der Vektor $\mathbf{q} := (q_1, \dots, q_n)^T$ aus (4.11) offensichtlich das System $A_k \mathbf{q} \leq \tilde{\mathbf{c}}$. Damit folgt die Behauptung direkt aus Satz 4.1.4. \square

Für den Fall $k = 2$ bzw. $k = 3$ ist eine untere bzw. obere Hindernisfunktion und die monotone und quasioptimale Approximation nach Proposition 4.2.1 in Abbildung 4.1 bzw. 4.2 dargestellt. Man erkennt, dass die Approximationen in bestimmten Bereichen deutlich verbesserungsfähig sind. Für den glatten Fall sind in der rechten Abbildung zusätzlich die Kontrollpolynome der Funktionen abgebildet. Hier sieht man, dass für die Kontrollpolynome die gleiche Monotonie gilt, wie für die Funktionen. Im Unterschied dazu können in Abschnitt 4.4.3 über ein Optimierungsverfahren monotone untere Grobgitterapproximationen konstruiert werden, deren Kontrollpolynome zum Teil oberhalb des Kontrollpolynoms der Hindernisfunktion verlaufen.

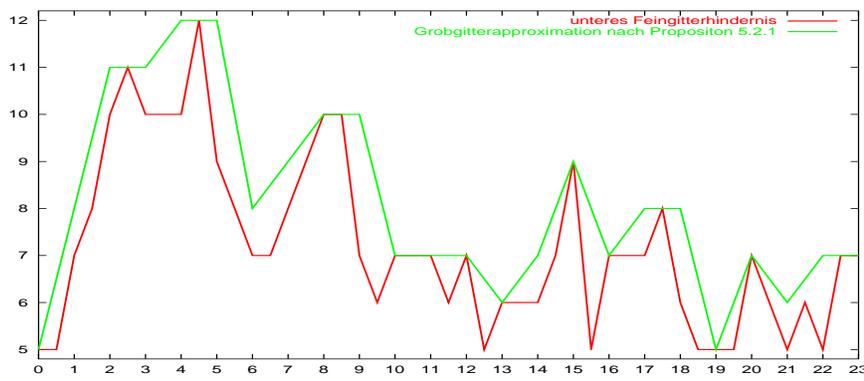


Abbildung 4.1: Unteres stückweise lineares Hindernis auf dem Gitter $[0, 23] \cap 1/2\mathbb{Z}$ und Grobgitterapproximation L_2 auf dem Gitter $[0, 23] \cap \mathbb{Z}$.

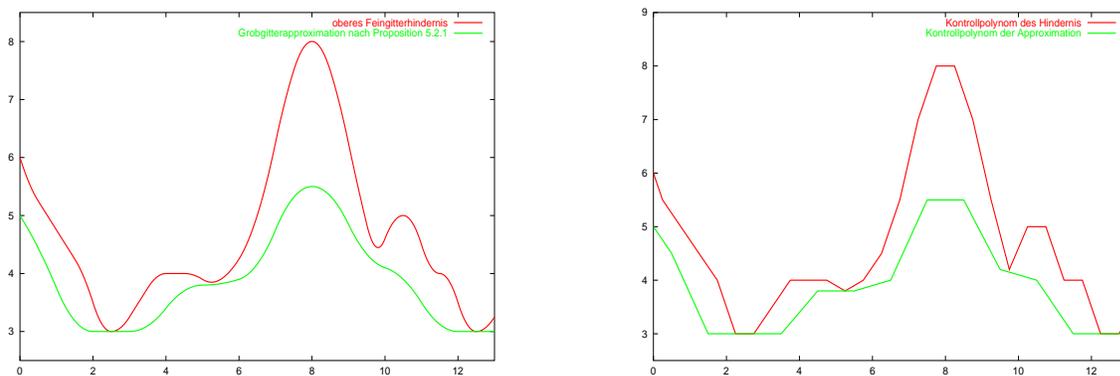


Abbildung 4.2: Links: C^1 -glatter, stückweise quadratischer Spline auf dem Gitter $[0, 13] \cap \mathbb{Z}/2$ und Grobgitterapproximation L_3 auf dem Gitter $[0, 13] \cap \mathbb{Z}$. Rechts: Zugehörige Kontrollpolynome.

Bemerkung 4.2.2 Für $k = 2$ entspricht der in Proposition 4.2.1 definierte Restriktionsoperator $\hat{r} : \mathcal{N}_{2,\Delta} \rightarrow \mathcal{N}_{2,T}$, $\tilde{S} \mapsto L_2$ dem Restriktionsoperator aus [Ma].

Es sei weiter an die Definition der Menge M^ν an Suchrichtungen aus (3.12), an die Definition der reduzierten Menge M° an Suchrichtungen aus (3.27) und an die Definition der Kontaktmenge $K_\ell^\bullet(u_j^\nu)$ der Näherung u_j^ν auf dem Level ℓ aus (3.30) erinnert. Man beachte, dass es für jedes $\mu^{\nu,j} \in M^\nu$, ein $\ell \in \{1, \dots, J\}$ und ein $i \in \{1, \dots, n_\ell\}$ gibt, so dass $\mu^{\nu,j} = N_{i,k,\Delta_{n_\ell}}$ gilt.

Lemma 4.2.7 *Ist $\Psi_\ell \in \mathcal{N}_{k,\Delta_\ell}$ eine monotone und quasioptimale Grobgitterapproximation an die Hindernisfunktion $\Psi_{\ell+1} \in \mathcal{N}_{k,\Delta_{\ell+1}}$ im Sinne der Definition 4.1.2 bzw. 4.2.5 und $\mu^{\nu,j} \in M^\nu$ für $j = j(i, \ell) \in \{1, \dots, m^\nu\}$ eine Suchrichtung, dann definieren*

$$(4.13) \quad \psi^{\nu,j} := c_i N_{i,k,\Delta_{n_\ell}} \in \text{span}\{\mu^{\nu,j}\}$$

monotone und quasioptimale Hindernisse im Sinne der Definition 3.2.2 bzw. 3.2.8.

Beweis: Aus der Annahme der Monotonie der Grobgitterapproximation folgt nach Definition der lokalen Hindernisse $\psi^{\nu,j}$, dass $D^{\nu,j} \subset D^{*,\nu,j}$ gilt. Nach Definition 3.2.2 sind die lokalen Hindernisse dann monoton.

Um auch Quasioptimalität nachzuweisen, gelte $K_{\ell+1}^\bullet(u_j^\nu) = K_{\ell+1}^\bullet(u_J)$ auf dem Level $\ell+1$ für alle $\nu > \nu_0$ und es sei $\mu^{\nu,j} \in M^\circ$. Dann gibt es ein $\ell \in \{1, \dots, J\}$ und ein $i \in \{1, \dots, n_\ell\}$, so dass

$$\mu^{\nu,j} = N_{i,k,\Delta_{n_\ell}}$$

gilt. Bei Entwicklung in der B-Spline-Basis $\mathcal{N}_{k,\Delta_{\ell+1}}$ gilt nach (A.10) die Darstellung

$$\mu^{\nu,j} = \sum_{l=2i-k}^{2i} a_{l-2i+k} N_{i,k,\Delta_{n_{\ell+1}}}$$

mit den Verfeinerungskoeffizienten $a_0, \dots, a_k > 0$ (*). Wegen $\mu^{\nu,j} \in M^\circ$ und der Darstellung aus (4.12) von $\Psi_{\ell+1}$ folgt hieraus

$$\tilde{c}_{2i-k}, \dots, \tilde{c}_{2i} > 0. \quad (**)$$

Denn wäre $\tilde{c}_l = 0$ für ein $l \in \{2i-k, \dots, 2i\}$, dann ist nach Definition der Kontaktmenge

$$l \in K_{\ell+1}^\bullet(u_j^\nu) = K_{\ell+1}^\bullet(u_J)$$

für alle $\nu > \nu_0$. Wegen $\mu^{\nu,j} \in M^\circ$ wäre dann aber $a_{l-2i+k} = 0$ im Widerspruch zu (*). Aus (**) und der Definition von q_i erhält man schließlich, dass $q_i > 0$ gilt. Aus der Annahme der Quasioptimalität $c_i \geq q_i$ folgt damit

$$\psi^{\nu,j} := c_i N_{i,k,\Delta_{n_\ell}} \geq q_i N_{i,k,\Delta_{n_\ell}} > 0.$$

Nach Definition 3.2.8 ist das lokale Hindernis $\psi^{\nu,j}$ dann quasioptimal und wegen $\mu^{\nu,j} = N_{i,k,\Delta_{n_\ell}}$ gilt $\psi^{\nu,j} \in \text{span}\{\mu^{\nu,j}\}$. \square

4.3 Formulierung als lineare Optimierungsaufgabe

Die im vorherigen Abschnitt konstruierten monotonen und quasioptimalen Grobgitterapproximationen sollen weiter optimiert werden. Dazu soll das Problem, eine monotone und quasioptimale Grobgitterapproximation $S = \mathbf{c}^T \mathbf{N}_{k,T}$ an einen Spline $\tilde{S} = \tilde{\mathbf{c}}^T \mathbf{N}_{k,\Delta}$ zu finden, in die Standardform einer linearen Optimierungsaufgabe gebracht werden.

Lineare Optimierungsaufgabe

$$(4.14) \quad \begin{aligned} &\text{Minimiere die Zielfunktion} && F(\mathbf{c}) = \xi^T \mathbf{c} + \eta \\ &\text{unter den Nebenbedingungen} && A_k \mathbf{c} \leq \tilde{\mathbf{c}} \text{ und } \mathbf{c} \geq \mathbf{q}. \end{aligned}$$

Dabei sind $\mathbf{q} \in \mathbb{R}^n$, $\xi \in \mathbb{R}^n$, $A_k \in \mathbb{R}^{\tilde{n} \times n}$, $\tilde{\mathbf{c}} \in \mathbb{R}^m$ und $\eta \in \mathbb{R}$ gegeben, und es gilt $\tilde{n} > n$.

Bemerkung 4.3.1 Die Nebenbedingung $A_k \mathbf{c} \leq \tilde{\mathbf{c}}$ ist die Bedingung aus Satz 4.1.4 und sichert die Monotonie der Approximation, d.h. $\tilde{S}(x) \geq S(x)$ für alle $x \in I$. Die Nebenbedingung $\mathbf{c} \geq \mathbf{q}$ beinhaltet die unteren Schranken aus (4.11) und sichert die Quasioptimalität.

Um die Güte verschiedener monotoner und quasioptimaler Approximationen vergleichen zu können, wird nun die lineare Zielfunktion F definiert. Sie sei so gewählt, dass die Summe der Abstände von Grob- und Feingitterspline an allen Gitterpunkten $\theta \in T$ des groben Gitters minimal ist, d.h.

$$(4.15) \quad F(\mathbf{c}) := \sum_{\theta \in T} |\tilde{S}(\theta) - S(\theta)|$$

soll minimal sein.

Lemma 4.3.2 Die Zielfunktion F aus (4.15) definiert eine lineare Funktion $\mathbb{R}^n \rightarrow \mathbb{R}$ in der Form

$$F(\mathbf{c}) = \xi^T \mathbf{c} + \eta.$$

Dabei gilt für ungerades k :

$$(i) \quad \eta = \frac{1}{2} \sum_{i=1}^{\tilde{n}} \tilde{c}_i, \quad \xi_i = \begin{cases} -(a_{k-2i+1} + \dots + a_k)/2 & \text{für } i = 1, \dots, \lfloor k/2 \rfloor, \\ -(a_0 + \dots + a_k)/2 = -1 & \text{für } i = \lfloor k/2 \rfloor + 1, \dots, n - \lfloor k/2 \rfloor, \\ -(a_0 + \dots + a_{2n-2i+1})/2 & \text{für } i = n - \lfloor k/2 \rfloor + 1, \dots, n \end{cases}$$

und für gerades k :

$$(ii) \quad \eta = \mathbf{s}_k \tilde{\mathbf{c}}, \quad \xi_i = \begin{cases} -(\gamma_k a_{k-2i+1} + \dots + \beta_k a_k) & \text{für } i = 1, \dots, \lfloor k/2 \rfloor, \\ -(\beta_k a_0 + \gamma_k a_1 + \dots + \beta_k a_k) & \text{für } i = \lfloor k/2 \rfloor + 1, \dots, n - \lfloor k/2 \rfloor, \\ -(\beta_k a_0 + \dots + \gamma_k a_{2n-2i+1}) & \text{für } i = n - \lfloor k/2 \rfloor + 1, \dots, n. \end{cases}$$

Dabei ist $\mathbf{s}_k := (\beta_k, \gamma_k, \beta_k, \dots)^T \in \mathbb{R}^{\tilde{n}}$ mit β_k, γ_k aus Lemma A.1.3. Die Verfeinerungskoeffizienten a_j sind in (4.6) definiert.

Beweis: Wegen der Nebenbedingung $A_k \mathbf{c} \leq \tilde{\mathbf{c}}$ ist $\tilde{S}(x) \geq S(x)$ für alle $x \in I$ und damit $|\tilde{S}(x) - S(x)| = \tilde{S}(x) - S(x)$. Die Differenzen $\tilde{S}(x) - S(x)$ wurden bereits in (4.7) berechnet. Damit gilt für $k \in \mathbb{N}$:

$$\begin{aligned} F(\mathbf{c}) &= \sum_{\theta \in T} (\tilde{S}(\theta) - S(\theta)) \\ &= \sum_{\theta \in T} \sum_{i=1}^{\tilde{n}} d_i N_{i,k,\Delta}(\theta) \\ &= \sum_{i=1}^{\tilde{n}} d_i \sum_{\theta \in T} N_{i,k,\Delta}(\theta). \end{aligned}$$

Wegen $d_i = \tilde{c}_i - A_k c_i$ und mit der Abkürzung $(s_k)_i := \sum_{\theta \in T} N_{i,k,\Delta}(\theta)$ erhält man weiter

$$(4.16) \quad F(\mathbf{c}) = \sum_{i=1}^{\tilde{n}} (s_k)_i (\tilde{c}_i - A_k c_i).$$

Für ungerades k gilt $(s_k)_i = \frac{1}{2}$ unabhängig von i nach Lemma A.1.3. Damit müssen nur noch die verbliebenen Terme in (4.16) summiert und nach c_i geordnet werden. Mit ξ_i wie

oben über die Verfeinerungskoeffizienten der Matrix A_k definiert und unter Berücksichtigung des Faktors $\frac{1}{2}$ folgt schließlich

$$F(\mathbf{c}) = \frac{1}{2} \sum_{i=1}^{\tilde{n}} \tilde{c}_i - \sum_{i=1}^n \xi_i c_i.$$

Wegen A.11 gilt für alle Koeffizienten ξ_i weit genug entfernt vom Rand, d.h. für $i = \lfloor k/2 \rfloor + 1, \dots, n - \lfloor k/2 \rfloor$, dass

$$\xi_i := -\frac{1}{2}(a_0 + \dots + a_k) = -1$$

gilt. Damit ist (i) gezeigt.

Für gerades k sind je nach der Position des B-Splines $N_{i,k,\Delta}$ zwei Fälle zu unterscheiden. Je nach Fall gilt nach Lemma A.1.3 $(s_k)_i = \beta_k$ bzw. $(s_k)_i = \gamma_k$. Für $i = 1$ gehört der Knoten $\theta_{1+k/2}$, der genau in der Mitte des Träger von $N_{1,k,\Delta}$ liegt, zum groben Gitter T . Daher gilt $(s_k)_1 = \beta_k$ (s. Lemma A.1.3). Da sich die beiden Fälle jeweils abwechseln, ergibt sich insgesamt $\mathbf{s}_k := (\beta_k, \gamma_k, \beta_k, \dots)^T$. Durch Aufsummieren und Ordnen folgt damit aus (4.16) die Behauptung (ii), d.h.

$$F(\mathbf{c}) = \mathbf{s}_k \tilde{\mathbf{c}} - \sum_{i=1}^n \xi_i c_i.$$

Die durch β_k und γ_k gewichteten Koeffizienten ξ_i sind hierbei wie oben definiert. \square

Für $k = 2$ ist z.B. $\beta_2 = 1$ und $\gamma_2 = 0$ nach Lemma A.1.3 und man erhält

$$(4.17) \quad F(\mathbf{c}) = \sum_{i=1}^n (\tilde{c}_{2i-1} - c_i).$$

Für $k = 3$ ergibt sich

$$(4.18) \quad F(\mathbf{c}) = \sum_{i=1}^{\tilde{n}} \tilde{c}_i - \frac{1}{2} c_1 - \sum_{i=2}^{n-1} c_i - \frac{1}{2} c_n.$$

Man beachte, dass für Ordnungen $k > 2$ nach Lemma 4.3.2 Entwicklungskoeffizienten c_i nahe am Rand, d.h. für $i \leq \lfloor k/2 \rfloor$ oder $i > n - \lfloor k/2 \rfloor$, weniger stark in der Optimierung gewichtet werden als Entwicklungskoeffizienten im Inneren des Gebiets.

4.4 Lösung der linearen Optimierungsaufgabe

Im vorhergehenden Abschnitt wurde das Problem, eine monotone und quasioptimale Grobgitterapproximation zu konstruieren, für welche die Summe der Abstände von Fein- zu Grobgitterspline an Grobgitterpunkten minimal ist, in Form der linearen Optimierungsaufgabe (4.14) formuliert. In diesem Abschnitt werden drei verschiedene Möglichkeiten zur Lösung angegeben. Besonders nahe liegt eine Lösung mit dem *Simplex-Algorithmus*. In einem optimalen Mehrgitterverfahren kann dieser jedoch wegen des zu hohem Aufwandes nicht eingesetzt werden. Daher werden zwei weitere Lösungsmöglichkeiten dargestellt. Dies ist zum einen eine *direkte Lösung mit Fourier-Motzkin Elimination* für den Fall $k = 2$ und zum anderen eine *näherungsweise Lösung mit dem neuen OGK-Verfahren* für allgemeines k .

4.4.1 Lösung mit dem Simplex-Algorithmus

Die lineare Optimierungsaufgabe (4.14) kann durch einen einfachen *Simplex-Algorithmus* gelöst werden. Für mehr Informationen über das Simplex-Verfahren siehe z.B. [HH, Sj].

Wegen Proposition 4.2.1 kann der Punkt $\mathbf{q} \in \mathbb{R}^n$ aus (4.11), der die untere Schranke angibt, hierbei als Startecke gewählt werden. Eine separate Konstruktion der Startecke ist also nicht erforderlich.

Um beim Simplex-Verfahren von einer Ecke zu einer Nachbarecke zu kommen, beträgt der Aufwand $2n(m+1)+1$ Multiplikationen und $n(m+1)$ Additionen. Die Anzahl der Ecken ist durch $\binom{m}{n}$ beschränkt. In der Praxis beobachtet man, dass man meist mit $\frac{3}{2}(m-n)$ Schritten auskommt. In jedem Fall liegt die Komplexität aber oberhalb von $O(n)$ Operationen. Für eine Lösung der linearen Optimierungsaufgabe im Rahmen eines Mehrgitteralgorithmus müssen daher alternative Verfahren verwendet werden.

4.4.2 Direkte Lösung für Hutfunktionen mit Fourier-Motzkin-Elimination

Für stetige, stückweise lineare Splines ist eine rekursiv definierte direkte Lösung der linearen Optimierungsaufgabe (4.14) in Abhängigkeit der Entwicklungskoeffizienten $\tilde{\mathbf{c}}$ bestimmbar. Die Nummerierung der groben und feinen Gitterpunkte im Fall $k=2$ ist aus Abbildung 4.3 ersichtlich.

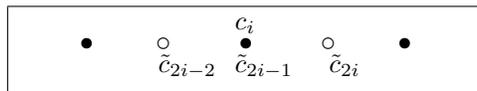


Abbildung 4.3: Nummerierung der groben Gitterpunkte c_i und der feinen Gitterpunkte \tilde{c}_j im Fall $k=2$.

Lemma 4.4.1 (Direkte Lösung für Hutfunktionen) Für $k=2$ und gegebenem $\tilde{\mathbf{c}} \in \mathbb{R}^{\tilde{n}}$ ist die Lösung der linearen Optimierungsaufgabe aus (4.14) rekursiv durch

$$\begin{aligned}
 c_n &:= \min\{\tilde{c}_{2n-1}, 2\tilde{c}_{2n-2} - q_{n-1}\} \\
 (4.19) \quad c_i &:= \min\{2\tilde{c}_{2i-2} - q_{i-1}, \tilde{c}_{2i-1}, 2\tilde{c}_{2i} - c_{i+1}\} \quad \text{für } i = n-1, \dots, 2 \text{ und} \\
 c_1 &:= \min\{\tilde{c}_1, 2\tilde{c}_2 - c_2\}
 \end{aligned}$$

gegeben. Dabei ist

$$(4.20) \quad q_i = \min\{\tilde{c}_{2i-2}, \tilde{c}_{2i-1}, \tilde{c}_{2i}\} \quad \text{für } i = 1, \dots, n \quad (\text{vgl. hierzu (4.11)}).$$

Insbesondere ist der B-Spline $S = \mathbf{c}^T \mathbf{N}_{k,T}$ eine quasioptimale und monotone Grobgitterapproximation an den B-Spline $\tilde{S} = \tilde{\mathbf{c}}^T \mathbf{N}_{2,\Delta}$.

Beweis: Der Beweis erfolgt mit der sogenannten Fourier-Motzkin-Elimination. Dies ist eine Technik, mit der für lineare Ungleichungssysteme der Form $A\mathbf{c} \leq \mathbf{b}$ mit der Matrix $A \in \mathbb{R}^{\tilde{n},n}$ und $\tilde{n} \geq n$ der zulässige Lösungsbereich der Koeffizienten c_i bestimmt werden kann. Insbesondere lässt sich feststellen, ob das System überhaupt eine Lösung besitzt. Eine genaue Beschreibung findet sich z.B. in [Sj]. Zunächst wird jede Ungleichungen so mit einem positiven Faktor multipliziert, dass die erste Spalte nur $+1$, -1 und 0 Einträge enthält. Wendet man die Methode auf die Ungleichungen

$$B_2 \mathbf{c} \leq \mathbf{e}$$

aus Beispiel 4.2.3 an, muss also die dritte Ungleichung mit zwei multipliziert werden. Man erhält

$$\begin{pmatrix} -1 & 0 & \cdots \\ 1 & 0 & \cdots \\ 1 & 1 & \cdots \\ 0 & -1 & \cdots \\ \vdots & \vdots & \end{pmatrix} \begin{pmatrix} c_1 \\ c_2 \\ c_3 \\ c_4 \\ \vdots \end{pmatrix} \leq \begin{pmatrix} -q_1 \\ \tilde{c}_1 \\ 2\tilde{c}_2 \\ -q_2 \\ \vdots \end{pmatrix}.$$

Hieran liest man ab, dass

$$(4.21) \quad q_1 \leq c_1 \leq \min\{\tilde{c}_1, 2\tilde{c}_2 - c_2\}$$

gelten muss, damit die ersten drei Ungleichungen erfüllt sind. Da in (4.21) bereits alle Informationen der ersten Spalte des Systems enthalten sind, kann im nächsten Schritt das um die erste Spalte reduzierte System betrachtet werden. Allerdings muss, um die Gültigkeit von (4.21) sicherzustellen, die Ungleichung

$$c_2 \leq 2\tilde{c}_2 - q_1$$

als neue Zeile in das System mit aufgenommen werden. Die andere Ungleichung $q_1 \leq \tilde{c}_1$ ist wegen (4.20) automatisch erfüllt. Im zweiten Schritt wird somit das System

$$B_2^{(1)} \mathbf{c}^{(1)} \leq \mathbf{e}^{(1)}$$

mit der modifizierten Matrix $B_2^{(1)}$ der entsprechend modifizierten rechten Seite $\mathbf{e}^{(1)}$ und dem um die erste Komponente reduzierten Vektor $\mathbf{c}^{(1)} := (c_2, \dots, c_n)^T \in \mathbb{R}^{n-1}$ betrachtet, d.h.

$$\begin{pmatrix} 1 & 0 & \cdots \\ -1 & 0 & \cdots \\ 1 & 0 & \cdots \\ 1 & 1 & \cdots \\ \vdots & \vdots & \end{pmatrix} \begin{pmatrix} c_2 \\ c_3 \\ c_4 \\ c_5 \\ \vdots \end{pmatrix} \leq \begin{pmatrix} 2\tilde{c}_2 - q_1 \\ -q_2 \\ \tilde{c}_3 \\ 2\tilde{c}_4 \\ \vdots \end{pmatrix}.$$

Hieran liest man die Bedingungen ab, die c_2 in Abhängigkeit von c_3, \dots, c_n erfüllen muss. Auf diese Weise fährt man fort, bis nur noch das System

$$B_2^{(n-1)} \mathbf{c}^{(n-1)} \leq \mathbf{e}^{(n-1)}$$

mit $\mathbf{c}^{(n-1)} := c_n \in \mathbb{R}$ übrig ist, an dem die Bedingungen für den letzten Koeffizienten c_n ablesbar sind. Für das System $B_2 \mathbf{c} \leq \mathbf{e}$ erhält man auf diese Weise schließlich

$$q_n \leq c_n \leq \min\{\tilde{c}_{2n-1}, 2\tilde{c}_{2n-2} - q_{n-1}\},$$

$$q_i \leq c_i \leq \min\{2\tilde{c}_{2i-2} - q_{i-1}, \tilde{c}_{2i-1}, 2\tilde{c}_{2i} - c_{i+1}\} \quad \text{für } i = n-1, \dots, 2 \text{ und}$$

$$q_1 \leq c_1 \leq \min\{\tilde{c}_1, 2\tilde{c}_2 - c_2\}.$$

Wegen (4.20) gilt $q_n \leq \min\{\tilde{c}_{2n-1}, 2\tilde{c}_{2n-2} - q_{n-1}\}$.

Um die Zielfunktion $F(\mathbf{c})$ aus (4.17) zu minimieren, die sich im Fall $k = 2$ aus Lemma 4.3.2 ergibt, müssen die Entwicklungskoeffizienten c_i jeweils möglichst groß gewählt werden. Dies führt zu der Wahl in (4.19). \square

Bemerkung 4.4.2 Der durch Lemma 4.4.1 definierte Restriktionsoperator $r : \mathcal{N}_{2,\Delta} \rightarrow \mathcal{N}_{2,T}$, $\tilde{S} \mapsto S$ entspricht dem mit geometrischen Überlegungen hergeleiteten Restriktionsoperator aus [Ko1]. Er ist eine Verbesserung des Restriktionsoperators \hat{r} aus Bemerkung 4.2.2 [Ma], da

$$r(\tilde{S}) \geq \hat{r}(\tilde{S}) \quad \text{für alle } \tilde{S} \in \mathcal{N}_{2,\Delta}$$

gilt.

In der Abbildung 4.4 ist im Intervall $[0, 23]$ eine stetige, stückweise lineare, untere Hindernisfunktion und die nach Lemma 4.4.1 konstruierte optimale Grobgitterapproximation in Vergleich zu der Grobgitterapproximation aus Proposition 4.2.1 dargestellt. Die Verbesserung ist in den Bereichen $[0, 2]$, $[5, 8]$, $[13, 15]$ und $[20, 22]$ deutlich erkennbar.

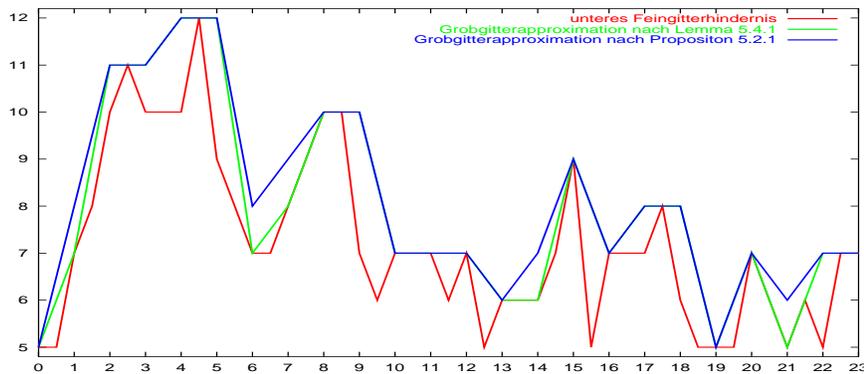


Abbildung 4.4: Stetige, stückweise lineare untere Hindernisfunktion auf dem Feingitter $[0, 23] \cap \mathbb{Z}/2$, Grobgitterapproximation nach Lemma 4.4.1 und Grobgitterapproximation nach Proposition 4.2.1 auf dem Grobgitter $[0, 23] \cap \mathbb{Z}$.

Bemerkung 4.4.3 Die Fourier–Motzkin–Elimination ist nur für kleine Matrizen mit wenigen Nicht–Nulleinträgen geeignet (s. [Sj]). Für Ordnungen $k > 2$ führt – bedingt durch die gestiegene Bandbreite der Matrix A_k – ein analoges Vorgehen auf eine mit n stetig wachsende Anzahl an unteren und oberen Schranken, die die Entwicklungskoeffizienten c_i erfüllen und die in jedem weiteren Eliminationsschritt mit berücksichtigt werden müssen. Für allgemeines $k \in \mathbb{N}$ und großem $n \in \mathbb{N}$ ist es daher zur Lösung der linearen Optimierungsaufgabe (4.14) nicht geeignet. Anstelle dessen wird im folgenden Abschnitt eine näherungsweise Lösung konstruiert.

4.4.3 Näherungsweise Lösung für B–Splines höherer Ordnung mit dem Optimierten Grobgitterhindernis–Konstruktionsalgorithmus (OGK)

Die lineare Optimierungsaufgabe aus (4.14) ist zwar prinzipiell mit dem Simplex–Algorithmus für beliebige $k \in \mathbb{N}$ lösbar. Im Rahmen eines Mehrgitterverfahrens zerstört die Anwendung des Simplex–Algorithmus aber sofort die optimale Komplexität von $O(n)$ Operationen. Das folgende Vorgehen nutzt dagegen die spezielle Struktur der Matrix A_k aus (4.5) aus. Dadurch ist es möglich in optimaler Komplexität eine zulässige Lösung des Ungleichungssystems aus (4.5) zu konstruieren, welche die untere Schranke L_k aus Proposition 4.2.1 erfüllt. Die numerischen Ergebnisse zeigen, dass die Lösung des Verfahrens nahe an der durch die lineare Optimierungsaufgabe definierten optimalen Lösung liegt.

Betrachte dazu das lineare Ungleichungssystem, zu vorgegebenem $\mathbf{b} \in \mathbb{R}^m$ ein $\mathbf{c} \in \mathbb{R}^n$ mit

$$(4.22) \quad A\mathbf{c} \leq \mathbf{b}$$

Ausgehend von (4.23) kann jetzt rekursiv eine zulässige Lösung des Ungleichungssystems bestimmt werden.

Lemma 4.4.5 *Es seien $A \in \mathbb{R}^{m \times n}$ und $\mathbf{b} \in \mathbb{R}^m$ wie in (4.22) gegeben und es sei \tilde{b}_i wie in (4.24). Dann erfüllen die rekursiv definierten Koeffizienten*

$$(4.25) \quad c_i := \min \left\{ \frac{\tilde{b}_{2i-1}}{a_{2i-1,i}}, \frac{\tilde{b}_{2i}}{a_{2i,i}} \right\} \quad \text{für } i = 1, \dots, n$$

das Ungleichungssystem $A\mathbf{c} \leq \mathbf{b}$. Dabei sind nur die Terme mit $a_{j,i} \neq 0$ zu betrachten.

Beweis: Die Behauptung folgt sofort aus der Darstellung (4.23). □

Bemerkung 4.4.6 Im allgemeinen weist die mit (4.25) berechnete Lösung Oszillationen im folgenden Sinn auf. Bei bereits bestimmten Koeffizienten c_1, \dots, c_{i-1} wird der Koeffizient c_i nach (4.25) genau so groß gewählt, dass die $(2i-1)$ -te und die $(2i)$ -te Ungleichung gerade noch erfüllt sind. Je größer aber c_i ist, desto kleiner wird der Wert \tilde{b}_j für $j > 2i$ der rechten Seite. Der folgende Koeffizient c_{i+1} kann dann eventuell nur noch sehr klein sein, unter Umständen muss er sogar negativ gewählt werden.

Um solche Oszillationen in der Lösung zu verhindern, soll bei der Bestimmung von c_i in (4.25) zusätzlich

$$(4.26) \quad \mathbf{c} \geq \mathbf{q}$$

für einen gegebenen Vektor $\mathbf{q} \in \mathbb{R}^n$ gefordert werden. Damit darf in der Rekursion jeder Koeffizient c_i höchstens so groß gewählt werden, dass alle folgenden Koeffizienten c_j mit $j > i$ die untere Schranke $c_j \geq q_j$ noch erfüllen können, ohne die $(2j-1)$ -te und die $2j$ -te Ungleichung in (4.23) zu verletzen. Es ist klar, dass es für zu groß gewählte untere Schranken \mathbf{q} keinen Vektor \mathbf{c} mehr geben kann, der $A\mathbf{c} \leq \mathbf{b}$ und $\mathbf{c} \geq \mathbf{q}$ erfüllt. Wir erinnern allerdings, dass bereits in (4.20) eine geeignete Schranke \mathbf{q} angegeben wurde. Nach Proposition 4.2.1 ist in diesem Fall auch die Lösbarkeit des Systems sichergestellt.

Die Forderung $c_j \geq q_j$, die $(2j-1)$ -te Ungleichung und die Definition von \tilde{b}_i in (4.24) implizieren

$$a_{2j-1,j} q_j \leq a_{2j-1,j} c_j \leq \tilde{b}_{2j-1} = b_{2j-1} - \sum_{\ell=1}^{j-1} a_{2j-1,\ell} c_\ell.$$

Hieraus folgt mit erneuter Anwendung von $c_j \geq q_j$ für alle $j > i$ die Ungleichung

$$a_{2j-1,i} c_i \leq b_{2j-1} - \sum_{\ell=1}^{i-1} a_{2j-1,\ell} c_\ell - \sum_{\ell=i+1}^j a_{2j-1,\ell} q_\ell =: \hat{b}_{2j-1,i}.$$

Falls $a_{2j-1,i} \neq 0$ ist, ergibt sich also für den Koeffizienten c_i die zusätzliche Bedingung

$$c_i \leq \hat{b}_{2j-1,i} / a_{2j-1,i}.$$

Mit $c_j \geq q_j$, der $2j$ -ten Ungleichung und der Definition von \tilde{b}_i folgt analog, dass

$$a_{2j,i} c_i \leq b_{2j} - \sum_{\ell=1}^{i-1} a_{2j,\ell} c_\ell - \sum_{\ell=i+1}^j a_{2j,\ell} q_\ell =: \hat{b}_{2j,i}$$

und damit

$$c_i \leq \hat{b}_{2j,i}/a_{2j,i}$$

für alle $j > i$ gelten muss.

Zusammen mit Lemma 4.4.5 ist damit der folgende Satz zur näherungsweise Lösung der linearen Optimierungsaufgabe aus (4.14) bewiesen.

Satz 4.4.7 (Optimierter Grobgitterhindernis-Konstruktionsalgorithmus (OGK)) *Es seien $A \in \mathbb{R}^{m \times n}$ und $\mathbf{b} \in \mathbb{R}^m$ wie in (4.22) gegeben, \tilde{b}_i wie in (4.24) und*

$$(4.27) \quad \hat{b}_{j,i} := b_j - \sum_{\ell=1}^{i-1} a_{j,\ell} c_\ell - \sum_{\ell=i+1}^{\lceil j/2 \rceil} a_{j,\ell} c_\ell \quad \text{für } i = 1, \dots, n, \text{ und } j > i.$$

Die Vektor \mathbf{c} sei rekursiv durch

$$(4.28) \quad c_i := \min \left\{ \frac{\tilde{b}_{2i-1}}{a_{2i-1,i}}, \frac{\tilde{b}_{2i}}{a_{2i,i}}, \frac{\hat{b}_{2i+1,i}}{a_{2i+1,i}}, \frac{\hat{b}_{2i+2,i}}{a_{2i+2,i}}, \dots, \frac{\hat{b}_{m,i}}{a_{m,i}} \right\} \quad \text{für } i = 1, \dots, n$$

definiert, wobei nur die Terme zu betrachten sind, bei denen $a_{j,i} \neq 0$ ist. Dann erfüllt der Vektor $\mathbf{c} \in \mathbb{R}^n$ das Ungleichungssystem

$$(4.29) \quad A\mathbf{c} \leq \mathbf{b}, \quad \mathbf{c} \geq \mathbf{q},$$

falls die Lösungsmenge von (4.29) nicht leer ist.

Für das speziellere Problem, eine monotone und quasioptimale Grobgitterapproximation zu konstruieren, ist wegen Proposition 4.2.1 die Existenz einer Lösung des Systems (4.29) gesichert. Damit ist auch das folgende Korollar bewiesen.

Korollar 4.4.8 *Es sei speziell \mathbf{q} aus (4.20), A_k aus (4.5), und \mathbf{b} wie in Bemerkung 4.4.4. Weiter sei der B-Spline $\tilde{S} \in \mathcal{N}_{k,\Delta}$ mit $\tilde{S} = \tilde{\mathbf{c}}^T \mathbf{N}_{k,\Delta}$ gegeben. Dann definiert der B-Spline $S = \mathbf{c}^T \mathbf{N}_{k,T} \in \mathcal{N}_{k,T}$ mit den Koeffizienten c_i aus (4.28) eine monotone und quasioptimale Grobgitterapproximation an den B-Spline $\tilde{S} \in \mathcal{N}_{k,\Delta}$.*

Bemerkung 4.4.9 Bei der Lösung des ursprünglichen Problems $A_k \mathbf{c} \leq \tilde{\mathbf{c}}$ aus (4.5) sind bei der Bestimmung des Minimums in (4.28) wegen der geringen Bandbreite der Matrizen A_k nur sehr wenige Terme zu berechnen, wie das folgende Beispiel für den Fall stückweise quadratischer und kubischer B-Splines zeigt.

Beispiel 4.4.10 Betrachte das Problem $A_3 \mathbf{c} \leq \tilde{\mathbf{c}}$ mit $\mathbf{c} \geq \mathbf{q}$ aus (4.14) zur Konstruktion einer monotonen und quasioptimalen Grobgitterapproximation einer quadratischen Hindernisfunktion. Vergleiche hierzu die Definition der Matrix A_k in (4.5) und Bemerkung 4.4.4. Dann müssen in (4.28) nur $\hat{b}_{j,2i+1}$ und $\hat{b}_{j,2i+2}$ berechnet werden. Alle anderen Terme \hat{b}_j mit $j > 2i + 2$ entfallen wegen $a_{j,i} = 0$ für $j > 2i + 2$. Entsprechend vereinfacht sich (4.28) wesentlich auf

$$(4.30) \quad c_i = \min \left\{ \begin{array}{l} 4\tilde{c}_{2i-3} - 3c_{i-1}, \\ \frac{4}{3}\tilde{c}_{2i-2} - \frac{1}{3}c_{i-1}, \\ \frac{4}{3}\tilde{c}_{2i-1} - \frac{1}{3}q_{i+1}, \\ 4\tilde{c}_{2i} - 3q_{i+1} \end{array} \right\}.$$

Für kubische B-Splines, d.h. $k = 4$ erhält man

$$(4.31) \quad c_i := \min \left\{ \begin{array}{l} 8\tilde{c}_{2i-4} - c_{i-2} - 6c_{i-1}, \\ 2\tilde{c}_{2i-3} - c_{i-1}, \\ \frac{4}{3}\tilde{c}_{2i-2} - \frac{1}{6}c_{i-1} - \frac{1}{6}q_{i+1}, \\ 2\tilde{c}_{2i-1} - q_{i+1} \\ 8\tilde{c}_{2i} - q_{i+2} - 6q_{i+1} \end{array} \right\}.$$

Für jeden Koeffizienten c_i sind also lediglich vier Terme zu berechnen. Dass dies auch für allgemeines k gilt, wird durch die Struktur der Matrix des äquivalenten Problems (4.23) illustriert.

Zum Aufwand des Verfahrens

Für die rekursive Bestimmung der Koeffizienten c_i nach dem nicht optimierten Verfahren (4.25) benötigt man $m + 2n$ Multiplikationen und $2m$ Additionen.

Für die rekursive Bestimmung der Koeffizienten c_i nach dem optimierten Verfahren OGK benötigt man $3N$ Multiplikationen und $N + m$ Additionen.

Dabei bezeichnet N die Anzahl der Nicht-Nullen in der Matrix A . Für das ursprüngliche Problem aus (4.5) ist $m = 2n + 1 - k$ wegen (4.3) und für die Anzahl der Nicht-Nullen in der Matrix A_k gilt offensichtlich $N \leq (k + 1)n$, so dass der Aufwand nach oben durch $3(k + 1)n$ Multiplikationen und $(k + 1)n + (2n + 1 - k)$ Additionen, also durch

$$(4k + 6)n + 1 - k \text{ Operationen}$$

beschränkt ist.

Satz 4.4.11 *Für festes $k \in \mathbb{N}$ liegt der Aufwand des nicht optimierten Verfahrens (4.25), als auch der Aufwand des optimierten Verfahrens OGK in der Größenordnung von $O(n)$ Operationen.*

Numerische Beispiele

Das Verfahren OGK ist in dem Sinne stabil, dass keine Oszillationen in der Lösung auftauchen und sichergestellt ist, dass bei der Durchführung nicht durch Null geteilt wird. Für den Fall, dass sich die Hindernisfunktion exakt auf dem groben Gitter darstellen lässt (vgl. hierzu die Bedingung aus (4.10)), liefert das Verfahren immer diese optimale Lösung. Dies ist dadurch begründet, dass in diesem Fall alle Ungleichungen mit gleich erfüllt werden können. Mit der Funktion L_k aus Proposition 4.2.1 kann eine gute untere Schranke für die Lösung des Verfahrens direkt angegeben werden.

Die nachfolgenden numerischen Ergebnisse zeigen, dass diese nochmal deutlich durch die mit dem OGK-Algorithmus verbundenen Optimierung verbessert werden kann. In der linken Abbildung 4.5 definiert die oberste Linie einen C^1 -glatten, quadratischen oberen Hindernis-B-Spline zur Knotenfolge $\Delta := [0, 13] \cap \mathbb{Z}/2$. Die unterste Linie stellt die untere Schranke $L_3 \in \mathcal{N}_{3,T}$ nach Proposition 4.2.1 dar, die auf der größeren Knotenfolge $T := [0, 13] \cap \mathbb{Z}$ definiert ist. Die mit dem OGK-Algorithmus bestimmte Approximation verläuft zwischen diesen Linien. Die Verbesserung im Vergleich zum Spline L_3 ist deutlich erkennbar.

Sehr glatte Bereiche des Hindernisses können erwartungsgemäß gut approximiert werden. Hochfrequenter Schwankungen des Hindernisses können dagegen nur bedingt approximiert werden. Dies ist z.B. im Bereich $[10, 12]$ erkennbar.

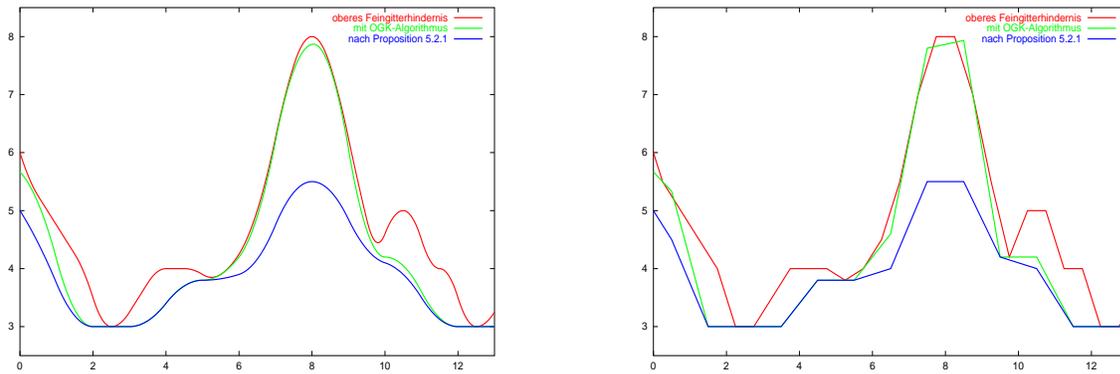


Abbildung 4.5: Links: C^1 – glatte, quadratische obere Hindernisfunktion auf dem Feingitter $\Delta := [0, 13] \cap \mathbb{Z}/2$ mit OGK-optimierter quadratischer Restriktion und der unteren Schranke L_3 , die beide auf dem Grobgitter $T := [0, 13] \cap \mathbb{Z}$ definiert sind. Rechts: Zugehörige Kontrollpolynome.

In der rechten Abbildung 4.5 sind die zugehörigen Kontrollpolynome dargestellt. Es fällt auf, dass das Kontrollpolynom der Approximation teilweise oberhalb des Kontrollpolynoms der Hindernisfunktion verläuft, obwohl die OGK-Approximation echt unterhalb des Hindernisses liegt. Dies zeigt, dass das OGK-Verfahren aus Satz 4.4.7 allen alternativen Verfahren überlegen ist, in denen monotone Approximationen über eine monotone Restriktion des Kontrollpolynoms der Hindernisfunktion bestimmt werden (vgl. hierzu auch Bemerkung 4.2.4).

In der linken Abbildung 4.6 wird die nach Satz 4.4.7 bestimmte C^1 – glatte, stückweise quadratische Restriktion mit der optimalen Restriktion verglichen, die mit dem Simplex-Algorithmus ermittelt wurde. In der rechten Abbildung wird sie außerdem mit einer optimalen stetigen, stückweise linearen Restriktion verglichen.

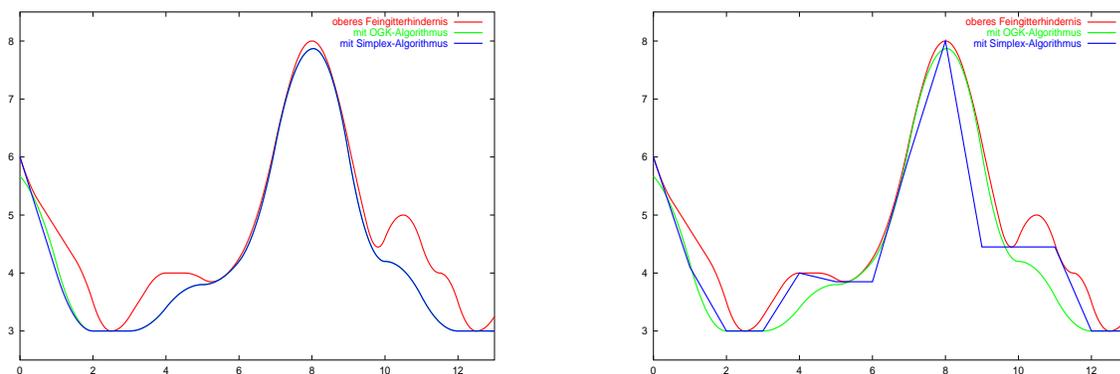


Abbildung 4.6: Vergleich der mit dem OGK-Algorithmus bestimmten monotonen und quasi-optimalen C^1 – glatten Grobgitterapproximation mit der optimalen Approximation nach dem Simplex-Algorithmus (links) und mit einer Approximation durch einen stetigen, stückweise linearen Spline (rechts).

Für die Hindernisfunktion aus Abbildung 4.6 ergeben sich die folgenden Werte der zu minimierenden Zielfunktion F aus (4.15), je nachdem mit welchem Verfahren die Grobgitterapproxi-

mation bestimmt wird.

mit Simplex-Algorithmus	$F(\mathbf{c}) = 4.56$
mit OGK-Algorithmus	$F(\mathbf{c}) = 4.83$
mit stückweise linearen Splines	$F(\mathbf{c}) = 5.05$
nach Proposition 4.2.1	$F(\mathbf{c}) = 11.45$

Man erkennt an den Werten und in der linken Abbildung 4.6, dass die mit dem OGK-Algorithmus bestimmte C^1 -glatte, stückweise quadratische Approximation mit einem Zielwert $F(\mathbf{c}) = 4.83$ der optimalen mit dem Simplex-Algorithmus (Zielwert $F(\mathbf{c}) = 4.56$) bestimmten Approximation bis auf kleinerer Randeffekte entspricht. Den numerischen Ergebnissen nach scheint dies in vielen Fällen der Fall zu sein. In der rechten Abbildung ist der Vergleich mit einer stetigen, stückweise linearen Approximation dargestellt. Diese liefert den höheren Zielwert von $F(\mathbf{c}) = 5.05$. Für die Approximation nach Proposition 4.2.1 erhält man den Wert $F(\mathbf{c}) = 11.45$. Man beachte bei den Ergebnissen, dass der OGK-Algorithmus und das Verfahren aus Proposition 4.2.1 im Unterschied zum Simplex-Verfahren nach Satz 4.4.11 nur optimale $O(n)$ Operationen benötigt. Die guten Ergebnisse spiegeln sich auch in den Konvergenzgeschwindigkeiten der Mehrgitterverfahren wieder, die in Kapitel 6 diskutiert werden.

Bemerkung 4.4.12 Es wurde auch die Verwendung von mehrfachen Stützstellen in der Knotenfolge untersucht. Die dadurch erhöhte Anzahl an Freiheitsgraden führt zu einer höheren Flexibilität in der Wahl der Grobgitterapproximation. Mit geringen Modifikationen des beschriebenen Lösungsverfahrens können auf diese Weise Grobgitterapproximationen konstruiert werden, die sich an die Glattheit der Hindernisfunktion anpassen. In Abbildung 4.7 ist ein stetiger, quadratischer B-Spline \tilde{S} mit doppelten Stützstellen auf dem Feingitter $\Delta := [0, 14] \cap \frac{1}{2}\mathbb{Z}$ dargestellt, der im ersten Bereich für $x \in [0, 9]$ glatt und im zweiten für $x \in [9, 14]$ stückweise linear ist. Die untere Grobgitterapproximation S passt sich automatisch an diese Struktur an, um die vorgegebene Zielfunktion minimieren zu können. Insbesondere können a priori bekannte, unterschiedliche Glattheitsstrukturen analog wie oben behandelt werden. Im Rahmen eines Mehrgitterverfahrens hat sich die Verwendung von Funktionen mit mehrfachen Stützstellen jedoch nicht angeboten und wird daher in dieser Arbeit auch nicht weiter verfolgt.

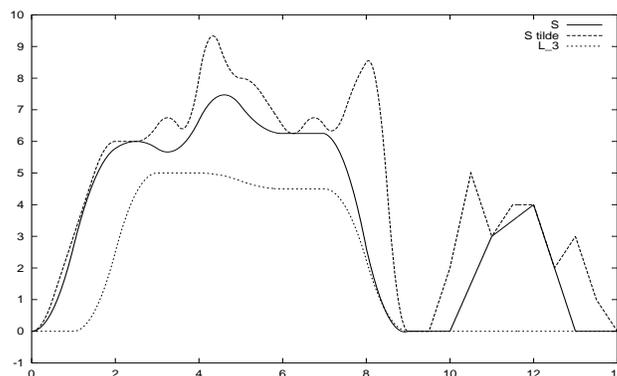


Abbildung 4.7: Stetige, quadratische obere Hindernisfunktion \tilde{S} mit doppelten Stützstellen auf dem Feingitter $\Delta := [0, 14] \cap \mathbb{Z}/2$ mit quadratischer Restriktion S und der unteren Schranke L_3 jeweils auf dem Grobgitter $T := [0, 14] \cap \mathbb{Z}$.

Bemerkung 4.4.13 Bei der abgeschnittenen Version des monotonen Mehrgitterverfahrens aus Abschnitt 3.2.4 muss das Verfahren aus Satz 4.4.7 bzw. aus Lemma 4.4.1 leicht modifiziert werden. Die Entwicklungskoeffizienten \tilde{c}_i der Hindernisfunktion \tilde{S} mit Kontakt, d.h. für die $i \in K_j^\bullet(\bar{u}_j')$ (s. Definition 3.2.4) gilt, werden auf plus (bei einem oberen Hindernis) bzw. minus (bei einem unteren Hindernis) unendlich gesetzt, so dass sie keinen restriktiven Einfluss mehr auf die Bestimmung der Korrekturen haben. Die entsprechend modifizierten Restriktionen sind für stetige, stückweise lineare bzw. glatte, stückweise quadratische Hindernisse in den Abbildungen 4.8 bzw. 4.9 in Abschnitt 3.3 dargestellt.

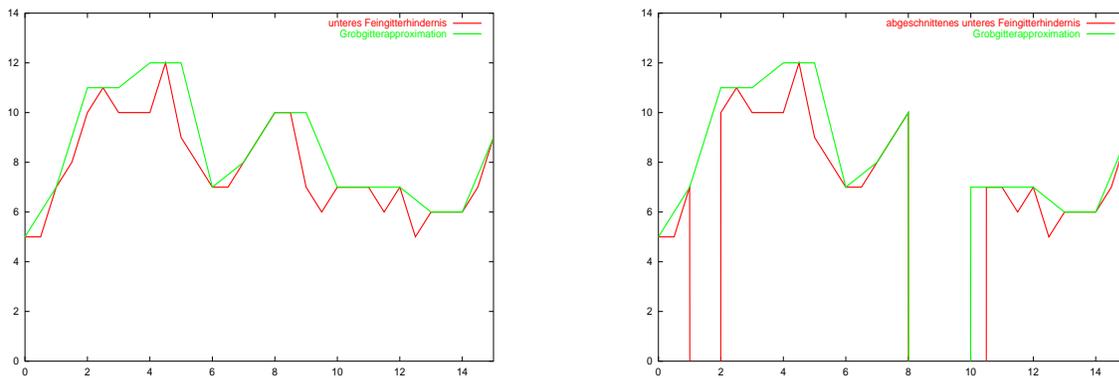


Abbildung 4.8: Links: Stetige, stückweise lineare Hindernisfunktion auf dem feinen Gitter $[0, 15] \cap \mathbb{Z}/2$ und stetige, stückweise lineare Approximation auf dem groben Gitter $[0, 15] \cap \mathbb{Z}$. Rechts: Abgeschnittene Hindernisfunktion und Restriktion bei den Kontaktpunkten $x = 1.5, 8.5, 9.0, 9.5, 10.0$.

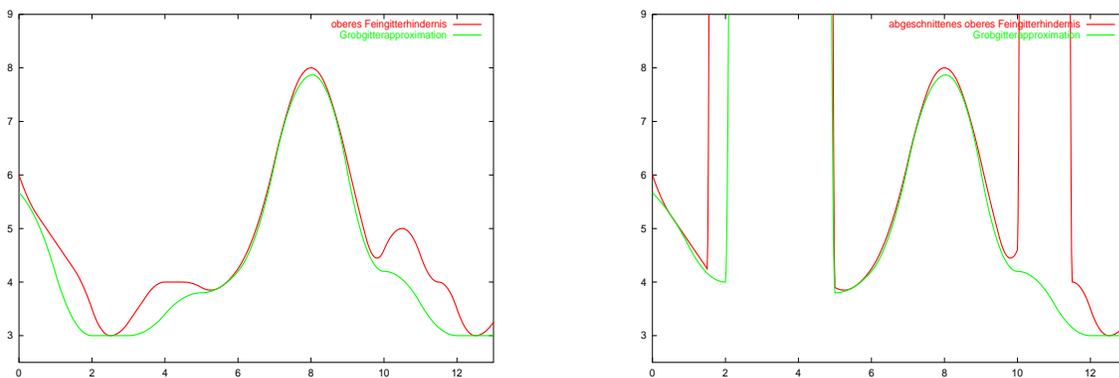


Abbildung 4.9: Links: C^1 -glatte, stückweise quadratische *obere* Hindernisfunktion auf dem feinen Gitter $[0, 13] \cap \mathbb{Z}/2$ und glatte, stückweise quadratische Restriktion auf dem groben Gitter $[0, 13] \cap \mathbb{Z}$. Rechts: Abgeschnittene Hindernisfunktion und Restriktion bei den Kontaktpunkten $x = 2.5, 3.0, 3.5, 4.0, 4.5, 11.0$.

Kapitel 5

Verallgemeinerung auf mehrere Dimensionen

In diesem Kapitel wird die Verallgemeinerbarkeit der bisherigen Ergebnisse auf höhere Dimensionen diskutiert. In Abschnitt 5.3 wird gezeigt, dass die Optionspreistheorie eine Reihe wichtiger Anwendungen bietet, die auf mehrdimensionale Probleme führen. Für solche Probleme muss das monotone Mehrgitterverfahren aus den Kapiteln 3 und 4 auf höhere Dimensionen verallgemeinert werden. Für Hutfunktionen ist das Verfahren für zwei Raumdimensionen in [Ko1] beschrieben.

In Abschnitt 5.2 wird ausgeführt, dass über einen Tensorproduktansatz eine Verallgemeinerung der in Kapitel 4 für den eindimensionalen Fall konstruierten Restriktionsoperatoren für B-Splines beliebiger Ordnung auf Quadergebieten möglich ist.

Das Ungleichungssystem

$$A_k^d \mathbf{c} \leq \tilde{\mathbf{c}}$$

im d -dimensionalen Fall wird über Tensorproduktbildung der Matrizen A_k aus (4.5) aufgestellt. Der B-Spline

$$S = \mathbf{c}^T \mathbf{N}_{k,T}^d$$

bildet dann eine d -dimensionale monotone Restriktion an die Hindernisfunktion

$$\tilde{S} = \tilde{\mathbf{c}}^T \mathbf{N}_{k,\Delta}^d.$$

Hierbei ist Δ bzw. T ein feines bzw. grobes Tensorprodukt-Gitter im \mathbb{R}^d . Mit Hilfe der Fourier-Motzkin-Elimination kann damit der optimale monotone Restriktionsoperator für stetige, stückweise lineare Splines angegeben werden. Bei Verwendung von B-Splines höherer Ordnung kann die Lösung wie im eindimensionalen Fall nicht mehr direkt bestimmt werden. Hier ist eine Verallgemeinerung des OGK-Algorithmus aus Abschnitt 4.4.3 notwendig. Für den Fall $d = 2$ und $k = 3$ wird auf diese Weise eine zweidimensionale, C^1 -glatte, monotone und quasioptimale Grobgitterapproximation konstruiert.

5.1 Tensorprodukt-B-Splines

Zu Intervallgrenzen $\mathbf{a}, \mathbf{b} \in \mathbb{R}^d$ sei

$$\Omega := \prod_{\ell=1}^d (a_\ell, b_\ell) \subset \mathbb{R}^d$$

ein Quader im \mathbb{R}^d .

Außerdem seien für $\ell = 1, \dots, d$ in $[a_\ell, b_\ell]$ erweiterte Knotenfolgen

$$(5.1) \quad \theta_{\ell,1} = \dots = \theta_{\ell,k} = a_\ell < \theta_{\ell,k+1} < \dots < \theta_{\ell,n} < b_\ell = \theta_{\ell,n+1} = \dots = \theta_{\ell,n+k}$$

gegeben und auf diesen eindimensionale B-Splines $N_{j_\ell,k}$ für $j = 1, \dots, n+k$ definiert (s. Anhang A.1). Die Menge aller Knotenpunkte definiert dann ein Gitter

$$T := \{\theta_{\ell,j} : \ell = 1, \dots, d, j = 1, \dots, n+k\}$$

in $\Omega \subset \mathbb{R}^d$.

Definition 5.1.1 (Tensorprodukt-B-Splines) Der i -te, d -dimensionale Tensorprodukt-B-Spline der Ordnung k zum Gitter $T \subset \Omega$ ist als

$$(5.2) \quad N_{i,k,T}^d(x) = N_{i,k}^d(x) := \prod_{\ell=1}^d N_{i_\ell,k}(x)$$

definiert. Dabei ist $i := (i_1, \dots, i_d)$ ein Multiindex.

Ein zweidimensionaler Tensorprodukt-B-Splines der Ordnung $k = 4$ ist in Abbildung 5.1 dargestellt.

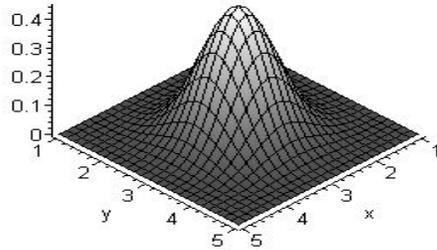


Abbildung 5.1: Zweidimensionaler Tensorprodukt-B-Splines der Ordnung $k = 4$.

Ein Tensorprodukt B-Spline besitzt wie im eindimensionalen Fall die Eigenschaften:

- (i) $\text{supp } N_{i,k}^d \subseteq \prod_{\ell=1}^d [\theta_{\ell,i}, \theta_{\ell,i+k}]$ (lokaler Träger)
- (ii) $N_{i,k}^d(x) \geq 0$ für alle $x \in \Omega$ (nicht negativ)
- (iii) $\sum_i N_{i,k}^d(x) = 1$ für alle $x \in \Omega$ (Zerlegung der Eins)
- (iv) $N_{i,k}^d \in C^{k-2}(\Omega)$ (Differenzierbarkeit)

Die Menge

$$\Sigma_k^d := \left\{ N_{i,k}^d(x) : i = (i_1, \dots, i_d), i_\ell = 1, \dots, n+k \right\}$$

bildet eine Basis für den Raum $\mathcal{N}_{k,T}^d$ der Splines der Ordnung k in Ω . Die Auswertung von Tensorprodukt-B-Splines und ihren Ableitungen kann analog zu den Neville-artigen Schemata aus (A.6) und (A.5) erfolgen. Ebenso lassen sich die Unterteilungs- und Knotenverfeinerungsstrategien aus (A.9) und (A.10) auf beliebige Dimensionen übertragen.

Auf Basis der Tensorprodukt-B-Splines lassen sich nun Splinefunktionen $S : \Omega \rightarrow \mathbb{R}$ der Ordnung k in der Form

$$S(\theta) = \sum_{i=(i_1, \dots, i_d)} c_i N_{i,k,T}^d(\theta)$$

mit $i_j = 1, \dots, n+k$ für $j = 1, \dots, d$ mit den Entwicklungskoeffizienten $\mathbf{c} \in \mathbb{R}^{n \times d}$ darstellen. Abkürzend wird die Schreibweise als Skalarprodukt

$$S = \mathbf{c}^T \mathbf{N}_{k,T}^d$$

verwendet.

5.2 Monotone und quasioptimale Grobgitterapproximationen

Für den Rest dieses Kapitels sei T das Gitter aus (5.1) in $\Omega \subset \mathbb{R}^d$ mit äquidistanter Gitterweite H und $\Delta \supset T$ ein verfeinertes Gitter mit Gitterweite $h := H/2$.

Als Verallgemeinerung der Definitionen 4.1.2 und 4.2.5 aus Kapitel 4 ergeben sich für d Raumdimensionen die folgenden Definitionen.

Definition 5.2.1 (Monotone Grobgitterapproximation in \mathbb{R}^d) Die Funktion $S \in \mathcal{N}_{k,T}^d$ heißt monotone untere Grobgitterapproximation an $\tilde{S} \in \mathcal{N}_{k,\Delta}^d$, falls $S(x) \leq \tilde{S}(x)$ für alle $x \in \Omega$ gilt.

Definition 5.2.2 (Quasioptimale Grobgitterapproximation in \mathbb{R}^d) Eine monotone untere Grobgitterapproximation $S \in \mathcal{N}_{k,T}^d$ mit $S = \mathbf{c}^T \mathbf{N}_{k,T}^d$ an den Spline $\tilde{S} \in \mathcal{N}_{k,\Delta}^d$ mit $\tilde{S} = \tilde{\mathbf{c}}^T \mathbf{N}_{k,\Delta}^d$ heißt quasioptimal, falls $\mathbf{c} \geq \mathbf{q}$ gilt. Dabei ist $\mathbf{q} \in \mathbb{R}^{n \times d}$ ein Vektor mit den Einträgen

$$(5.3) \quad q_i := \min \{ \tilde{c}_l : l = (i_1 \pm r_1, \dots, i_d \pm r_d), r_j \in \{0, \dots, k-1\}, j = 1, \dots, d \}$$

mit $i := (i_1, \dots, i_d)$.

Bemerkung 5.2.3 Für den Fall der Ordnung $k = 2$ entspricht Definition 5.2.2 der Definition aus [Ko1]. Für den Fall der Dimension $d = 1$ entspricht sie Definition 4.2.5.

Monotone und quasioptimale Grobgitterapproximationen S sollen analog zum eindimensionalen Fall über Tensorproduktbildung bestimmt werden.

Definition 5.2.4 (Tensorprodukt) Das Tensorprodukt $A \otimes B$ zweier Matrizen $A, B \in \mathbb{R}^{\tilde{n} \times n}$ ist definiert durch

$$(A \otimes B)_{i,j,k,l} := A_{i,k} B_{j,l} \quad \text{für } i, j = 1, \dots, \tilde{n} \text{ und } k, l = 1, \dots, n.$$

Insbesondere gilt $A \otimes B \in \mathbb{R}^{\tilde{n} \times \tilde{n} \times n \times n}$.

Analog zu Satz 4.1.4 aus Kapitel 4 liefert der folgende Satz 5.2.6 in d Dimensionen die Basis zur Konstruktion einer monotonen Grobgitterapproximation. Er beruht auf der Positivität, den Verfeinerungseigenschaften von B-Splines und dem folgenden Lemma.

Lemma 5.2.5 Erfüllen die Matrizen $A, B \in \mathbb{R}^{\tilde{n} \times n}$ und die Vektoren $\mathbf{c}, \mathbf{d} \in \mathbb{R}^n$ mit $\mathbf{c}, \mathbf{d} \geq 0$ und $\tilde{\mathbf{c}}, \tilde{\mathbf{d}} \in \mathbb{R}^{\tilde{n}}$ mit $\tilde{\mathbf{c}}, \tilde{\mathbf{d}} \geq 0$ die Ungleichungssysteme

$$A \mathbf{c} \leq \tilde{\mathbf{c}} \quad \text{und} \quad B \mathbf{d} \leq \tilde{\mathbf{d}},$$

dann folgt für die Tensorprodukte

$$(A \otimes B) (\mathbf{c} \otimes \mathbf{d}) \leq \tilde{\mathbf{c}} \otimes \tilde{\mathbf{d}}.$$

Beweis: Einsetzen der Definition der Tensorprodukte und die Voraussetzung führen zu

$$\begin{aligned}
((A \otimes B)(\mathbf{c} \otimes \mathbf{d}))_{i,j} &= \sum_{k,l=1}^n A_{i,k} B_{j,l} c_k d_l \\
&= \sum_{k=1}^n A_{i,k} c_k \sum_{l=1}^n B_{j,l} d_l \\
&= (\mathbf{A}\mathbf{c})_i (\mathbf{B}\mathbf{d})_j \\
&\leq \tilde{c}_i \tilde{d}_j.
\end{aligned}$$

□

Satz 5.2.6 (Monotone Grobgitterapproximationen im \mathbb{R}^d) *Es sei $\tilde{S} \in \mathcal{N}_{k,\Delta}^d$ ein Tensorprodukt-B-Spline mit der Darstellung $\tilde{S} = \tilde{\mathbf{c}}^T \mathbf{N}_{k,\Delta}^d$. Dann ist $S \in \mathcal{N}_{k,T}$ mit $S = \mathbf{c}^T \mathbf{N}_{k,T}^d$ monotone untere Grobgitterapproximation an \tilde{S} , falls das Ungleichungssystem*

$$(5.4) \quad A_k^d \mathbf{c} \leq \tilde{\mathbf{c}}$$

erfüllt ist. Hierbei ist

$$A_k^d := A_k \otimes \dots \otimes A_k$$

mit der Matrix A_k aus (4.5).

Proposition 5.2.7 *Der Spline $L_k^d \in \mathcal{N}_{k,T}^d$ mit $L_k^d := \mathbf{q}^T \mathbf{N}_{k,T}^d$ und den Entwicklungskoeffizienten \mathbf{q} aus (5.3) ist monotone und quasioptimale untere Grobgitterapproximation an den Spline $\tilde{S} \in \mathcal{N}_{k,\Delta}$ mit $\tilde{S} = \tilde{\mathbf{c}}^T \mathbf{N}_{k,\Delta}$.*

Beweis: Der Beweis folgt wie in Proposition 4.2.1 direkt aus Satz 5.2.6 und Definition 5.2.2. □

Bemerkung 5.2.8 Die Restriktion L_k^d entspricht für den speziellen Fall $k = 2$ und $d = 2$ der Restriktion aus [Ma].

Optimale Elemente aus der Menge der monotonen und quasioptimalen Grobgitterapproximationen sind analog zu Abschnitt 4.3 dadurch charakterisiert, dass sie die Summe der Abstände zum Hindernis an allen Grobgitterpunkten minimieren.

Direkte Lösung für Hutfunktionen

Die Bestimmung einer optimalen Grobgitterapproximation für den Fall $k = 2$ aus Lemma 4.4.1 kann auf beliebig viele Dimensionen erweitert werden.

Lemma 5.2.9 (Direkte Lösung für Hutfunktionen in \mathbb{R}^2) *Es sei $\tilde{S} \in \mathcal{N}_{2,\Delta}^2$ eine Hutfunktion mit der Darstellung $\tilde{S} = \tilde{\mathbf{c}}^T \mathbf{N}_{2,\Delta}^2$. Dann ist die monotone und quasioptimale untere Grobgitterapproximation, welche die Abstände zum Hindernis \tilde{S} an allen Grobgitterpunkten minimiert, durch $S = \mathbf{c}^T \mathbf{N}_{2,T}^d \in \mathcal{N}_{2,T}^d$ und den Entwicklungskoeffizienten*

$$\begin{aligned}
c_{n,n} &:= q_{n,n}, \\
(5.5) \quad c_{i,j} &:= \min \left\{ \begin{array}{ll} 4\tilde{c}_{2i-2,2j} - q_{i-1,j} - c_{i+1,j+1} - c_{i,j+1}, & 2\tilde{c}_{2i-1,2j} - c_{i,j+1}, \\ 2\tilde{c}_{2i-2,2j-1} - q_{i-1,j}, & \tilde{c}_{2i-1,2j-1}, \\ 4\tilde{c}_{2i-2,2j-2} - q_{i-1,j} - q_{i-1,j-1} - q_{i,j-1}, & 2\tilde{c}_{2i-1,2j-2} - q_{i,j-1}, \\ & 4\tilde{c}_{2i,2j} - c_{i,j+1} - c_{i+1,j+1} - c_{i+1,j} \\ & 2\tilde{c}_{2i,2j-1} - c_{i+1,j} \\ & 4\tilde{c}_{2i,2j-2} - q_{i,j-1} - q_{i+1,j-1} - c_{i+1,j} \end{array} \right\}
\end{aligned}$$

für $i, j = 1, \dots, n$ gegeben. Dabei ist

$$q_{i,j} := \min \left\{ \begin{array}{lll} \tilde{c}_{2i-2,2j}, & \tilde{c}_{2i-1,2j}, & \tilde{c}_{2i,2j} \\ \tilde{c}_{2i-2,2j-1}, & \tilde{c}_{2i-1,2j-1}, & \tilde{c}_{2i,2j-1} \\ \tilde{c}_{2i-2,2j-2}, & \tilde{c}_{2i-1,2j-2}, & \tilde{c}_{2i,2j-2} \end{array} \right\}.$$

Beweis: Der Beweis verläuft mit Fourier–Motzkin–Elimination wie in Lemma 4.4.1. Dafür müssen zunächst die $2n$ Bedingungen

$$-\mathbf{c} \leq -\mathbf{q}$$

mit dem Vektor \mathbf{q} aus (5.3) in das System

$$A_2^2 \mathbf{c} \leq \tilde{\mathbf{c}}$$

aus Satz 5.2.6 analog zu Beispiel 4.2.3 integriert werden. Auf das resultierende Ungleichungssystem

$$B_2^2 \mathbf{c} \leq \mathbf{b}$$

wird dann die Fourier–Motzkin–Elimination angewandt. \square

In Abbildung 5.2 ist für den zweidimensionalen Fall eine obere Hindernis–Hutfunktion aus $\mathcal{N}_{2,\Delta}^2$ und die optimale Grobgitterapproximation aus $\mathcal{N}_{2,T}^2$ dargestellt, die sich mit den Koeffizienten aus Lemma 5.2.9 ergibt. Betrachtet man nur den Rand, dann sind die eindimensionalen Approximationen aus Abschnitt 4.4.2 wiederzuerkennen. Die Nummerierung der groben und feinen Gitterpunkte für den Fall $k = d = 2$ ist in Abbildung 5.3 veranschaulicht.

Bemerkung 5.2.10 Für den speziellen Fall $k = 2$ und $d = 2$ entsprechen die Approximationen S aus Lemma 5.2.9 den Approximationen aus [Ko1]. Dort wird allerdings auch der allgemeine Fall von Nicht–Tensorproduktgittern behandelt.

Verallgemeinerungen auf beliebige Dimensionen d sind analog möglich. In drei Dimensionen ist das Minimum in (5.5) über 27 Terme zu bilden, in d Dimensionen allgemein über 3^d Terme. Die Formel in (5.5) kann leicht hergeleitet werden, wenn die drei Vorfaktoren für den eindimensionalen Fall bekannt sind. Für $d = 1$ (s. Lemma 4.4.1), $d = 2$ (s. Lemma 5.2.9) und $d = 3$ sind die Vorfaktoren in Abbildung 5.4 dargestellt.

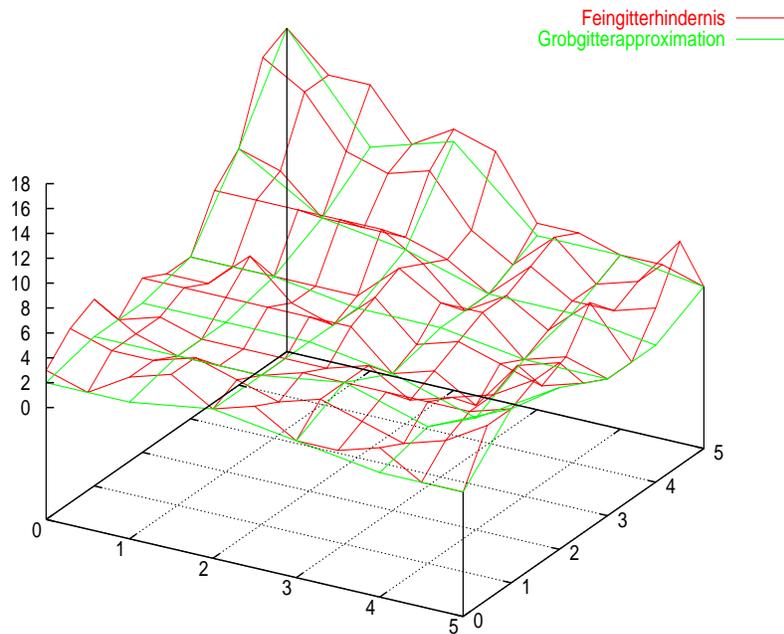


Abbildung 5.2: Stetige, stückweise lineare obere Hindernisfunktion mit monotoner und quasiop-
timaler Grobitterapproximation nach Lemma 5.2.9.

Näherungsweise Lösung für B-Splines der Ordnung $k \geq 3$

Für B-Splines der Ordnung $k \geq 3$ kann bei großem n die Lösung wie im eindimensionalen Fall nicht mehr mit Fourier–Motzkin–Elimination bestimmt werden. Anstelle dessen ist eine Verallgemeinerung des Näherungsverfahrens aus Abschnitt 4.4.3 notwendig. Die Vorgehensweise soll für den Fall $k = 3$ und $d = 2$ skizziert werden.

Beispiel 5.2.11 Zunächst wird das lineare Ungleichungssystem

$$A_3^2 \mathbf{c} \leq \tilde{\mathbf{c}}$$

für glatte, zweidimensionale quadratische B-Splines mit der Matrix

$$A_3^2 := A_3 \otimes A_3 \in \mathbb{R}^{\tilde{n} \times \tilde{n} \times n \times n}$$

aufgestellt. Dabei ist A_3 die Matrix aus (4.5) mit $k = 3$. Bei geeigneter Anordnung der Koeffizienten erhält man das System

für $i, j = 2, \dots, n$. Diese Bedingungen sichern die Monotonie der Approximation. Ist zusätzlich Quasioptimalität gefordert, ergeben sich in Analogie zu Satz 4.4.7 aus Abschnitt 4.4.3 die zusätzlichen Bedingungen

$$c_{1,j} \geq q_{1,j} \quad \text{für } j = 1, \dots, n$$

und

$$c_{i,j} \leq \min \left\{ \begin{array}{l} \frac{1}{3} \tilde{c}_{2i-3,2j-1} - 3c_{i-1,j} - 3c_{i-1,j+1} - \frac{1}{3} q_{i,j+1}, \\ \tilde{c}_{2i-3,2j} - 3c_{i-1,j} - 9c_{i-1,j+1} - 3q_{i,j+1}, \\ \frac{1}{9} \tilde{c}_{2i-2,2j-1} - \frac{1}{3} c_{i-1,j} - \frac{1}{9} c_{i-1,j+1} - \frac{1}{3} q_{i,j+1}, \\ \frac{1}{3} \tilde{c}_{2i-2,2j} - \frac{1}{3} c_{i-1,j} - c_{i-1,j+1} - 3q_{i,j+1} \end{array} \right\} =: m_2(i, j)$$

für $i = 2, \dots, n, j = 1, \dots, n$. An dieser Stelle wird die Wahl

$$c_{1,j} := q_{1,j} \quad \text{für } j = 1, \dots, n$$

$$c_{i,j} := \min \{m_1(i, j), m_2(i, j)\} \quad \text{für } i = 2, \dots, n, j = 1, \dots, n$$

vorgeschlagen. Damit ist eine C^1 -glatte, monotone und quasioptimale Groggitterapproximation $S = \mathbf{c}^T \mathbf{N}_{3,T}^2$ an den Spline $\tilde{S} = \tilde{\mathbf{c}}^T \mathbf{N}_{3,\Delta}^2$ definiert. Die Frage, ob bessere Approximationen an \tilde{S} möglich sind, und die Herleitung allgemeiner Formeln für beliebige Dimensionen d und Ordnungen k würde den Umfang dieser Arbeit sprengen und soll hier nicht weiter verfolgt werden.

5.3 Mehrdimensionale Optionspreistheorie

Wie in Abschnitt 1.2.3 hergeleitet wurde, führt die Bewertung Amerikanischer Standard-Optionen auf ein eindimensionales instationäres freies Randwertproblem. Im folgenden sollen Beispiele für Finanzverträge vorgestellt werden, die auf freie Randwertprobleme mit mehr als nur einer Raumdimension führen. Analog zur Bewertung Amerikanischer Standard-Optionen kann jeweils eine linear komplementäre Formulierung hergeleitet werden, in der der freie Rand nicht mehr explizit auftritt und auf die eine entsprechend mehrdimensionale Verallgemeinerung des Finite-Elemente-Verfahrens aus Kapitel 2 und des Mehrgitterverfahrens aus Kapitel 3 angewendet werden kann. Die folgenden Abschnitte sind den Arbeiten [TR], [Oo] und [WHD1] entnommen.

5.3.1 Amerikanische Basket-Optionen

Eine *Basket-Option* ist eine Option, deren Preis von mehreren zugrundeliegenden Wertpapieren abhängt. Für $i = 1, \dots, d$ bezeichnet im folgenden S_i den Preis des i -ten Wertpapiers mit Volatilität σ_i und der Korrelation ρ_{ij} mit dem j -ten Wertpapier. Zusammen ergibt sich der Vektor $\mathbf{S} := (S_1, \dots, S_d)^T$.

Unter den Voraussetzungen aus Abschnitt 1.2.1 erfüllt der Preis $V = V(\mathbf{S}, t)$ einer Europäischen Basket Verkaufsoption die d -dimensionale Black-Scholes Gleichung (s. [TR])

$$\mathcal{L}V := \frac{\partial}{\partial t} V + \frac{1}{2} \sum_{i=1}^d \sum_{j=1}^d \rho_{ij} \sigma_i \sigma_j S_i S_j \frac{\partial^2}{\partial S_i \partial S_j} V + \sum_{i=1}^d r S_i \frac{\partial}{\partial S_i} V - rV = 0 \quad \text{für } S_i \geq 0, 0 \leq t < T$$

mit der Auszahlungsfunktion

$$h(\mathbf{S}) := \left(K - \sum_{i=1}^d \alpha_i S_i \right)^+$$

für gegebene Konstanten $\alpha_i \in \mathbb{R}^+$.

Die Bewertung Amerikanischer Basket-Optionen führt dann auf ein instationäres freies Randwertproblem in d Raumdimensionen zu obiger Differentialgleichung mit der Bedingung, dass V und $\nabla V := (\partial V/\partial S_1, \dots, \partial V/\partial S_d)$ stetig über den freien Rand verlaufen. Die Möglichkeit die Option jederzeit ausüben zu können, impliziert

$$V(\mathbf{S}, t) \geq h(\mathbf{S})$$

für alle \mathbf{S} und alle t . Außerdem gilt

$$\mathcal{L}V \leq 0.$$

Damit erhält man analog zu Abschnitt 2.2.1 die lineare komplementäre Formulierung

$$(\mathcal{L}V)(V - h) = 0,$$

$$\mathcal{L}V \leq 0, \quad V - h \geq 0$$

des freien Randwertproblems mit der Endbedingung $V(\mathbf{S}, T) = h(\mathbf{S})$. Die Randbedingungen lauten $V(\mathbf{S}, T) = 0$ für $S_i \rightarrow \infty$ und $V(\mathbf{S}, T) = b_i(\mathbf{S}, T)$ für $S_i = 0$ für $i = 1, \dots, d$. Dabei bezeichnet b_i die Lösung des entsprechenden $(d - 1)$ -dimensionalen Black-Scholes Problem (s. [TR]).

5.3.2 Amerikanische Optionen mit stochastischer Volatilität

Im Black-Scholes Modell (s. Abschnitt 1.2) ist die Aktienkursdynamik $S(t)$ modelliert durch die stochastische Differentialgleichung

$$dS(t) = \mu S(t) dt + \sqrt{y} S(t) dB_1(t)$$

mit konstanter Drift $\mu \in \mathbb{R}$ und konstanter Volatilität $\sigma = \sqrt{y} \in \mathbb{R}^+$. In der Praxis ist die Volatilität jedoch weder konstant noch für längere Zeiträume vorhersagbar. Sie muss daher aus historischen Daten geschätzt oder aus bekannten Optionswerten zurückgerechnet werden. Allgemeiner ist es, die Volatilität ebenfalls als einen stochastischen Prozess $\sigma(t) := \sqrt{y(t)}$ aufzufassen. Üblicherweise wird er dann durch die stochastische Differentialgleichung

$$dy(t) = \alpha(\beta - y(t))dt + \gamma\sqrt{y(t)} dB_2(t)$$

mit den Konstanten α, β und $\gamma \in \mathbb{R}^+$ modelliert. $\mathbf{B}(t) = (B_1(t), B_2(t))$ bezeichnet dabei eine zweidimensionale standardisierte Brownsche Bewegung mit Korrelations-Koeffizient $\rho \in [-1, 1]$. Für eine Definition der mehrdimensionalen Brownschen Bewegung siehe z.B. [Ba].

Der Optionswert $V = V(S, y, t)$ hängt jetzt zusätzlich von einer dritten Variablen y ab. Er erfüllt (s. [BR]) die Konvektions-Diffusions-Reaktionsgleichung

$$\mathcal{L}V := \frac{\partial V}{\partial t} + \frac{1}{2} \left(S^2 y \frac{\partial^2 V}{\partial S^2} + \frac{\partial^2 V}{\partial S \partial y} + \gamma^2 y \frac{\partial V}{\partial y^2} \right) + rS \frac{\partial V}{\partial S} + (\alpha(\beta - y) - \lambda\gamma\sqrt{y}) + \frac{\partial V}{\partial y} - rV = 0$$

in $\Omega := \mathbb{R}^+ \times \mathbb{R}^+ \times [0, T)$. Hierbei bezeichnet $\lambda = \lambda(S, y, t)$ die sogenannte "Market Price of Risk" Funktion. Sie wird oft als konstant gleich Null angenommen (s. [Oo], [CP]). Für eine Verkaufsoption gelten die Dirichlet Randbedingungen

$$V(0, y, t) = K, \quad \text{und} \quad V(S, 0, t) = (K - S)^+ =: \mathcal{H}(S)$$

und die Neumann Randbedingungen

$$\frac{\partial V}{\partial S} = 0 \text{ für } S \rightarrow \infty, \quad \text{und} \quad \frac{\partial V}{\partial y} = 0 \text{ für } y \rightarrow \infty.$$

Amerikanische Optionen mit stochastischer Volatilität führen somit zu einem instationären zwei-dimensionalen freien Randwertproblem. Analog zu Abschnitt 2.2.2 lässt es sich in ein lineares komplementäres Problem der Form

$$\begin{aligned} (\mathcal{L}V) \cdot (V - h) &= 0, \\ \mathcal{L}V &\leq 0, \quad V - h \geq 0 \end{aligned}$$

mit den obigen Randbedingungen umschreiben.

Ausgehend von dieser Formulierung wird das Problem in den Arbeiten [CP] und [Oo] numerisch mittels Finite-Differenzen Diskretisierung behandelt.

5.3.3 Pfadabhängige Amerikanische Optionen

Die Auszahlung \mathcal{H} zum Fälligkeitszeitpunkt T einer pfadabhängigen Option ist nicht wie bisher nur vom Aktienkurs $S(T)$ zum Zeitpunkt T abhängig, sondern allgemeiner von allen Kursen $S(t)$ mit $0 \leq t \leq T$.

Ein Beispiel sind *Asiatische Optionen*. Hier ist die Auszahlungsfunktion \mathcal{H} abhängig vom durchschnittlichen Kursverlauf.

Definition 5.3.1 (Asiatische Kaufoptionen) Für eine Asiatische Kaufoption mit *arithmetischem Durchschnitt* gilt

$$\mathcal{H}(S, I) := (I - K)^+ \quad \text{mit } I := \frac{1}{T} \int_0^T S(t) dt$$

und bei *geometrischem Durchschnitt*

$$\mathcal{H}(S, I) := (e^I - K)^+ \quad \text{mit } I := \frac{1}{T} \int_0^T \log S(t) dt.$$

Pfadabhängige Optionen lassen sich in einem einheitlichen Rahmen zusammenfassen (s. [WHD1], Abschnitt 9.2). Sei hierzu für eine gegebene Funktion $f = f(S(t), t)$ die neue Variable

$$I := \int_0^T f(S, t) dt$$

eingeführt. Im Falle Asiatischer Optionen gilt $f(S, t) = \frac{1}{T} S(t)$ bei arithmetischem Durchschnitt bzw. $f(S, t) = \frac{1}{T} \log S(t)$ bei geometrischem Durchschnitt. Da nach den Annahmen des Black-Scholes Modells die Kursvergangenheit unabhängig vom aktuellen Aktienkurs ist, sind S, I und t drei unabhängige Variablen, von denen der Optionswert $V = V(S, I, t)$ abhängt.

Mit dem Lemma von Itô und Arbitrage-Argumenten folgt ähnlich wie in Abschnitt 1.2.2, dass der Wert einer pfadabhängigen Europäischen Option die partielle Differentialgleichung

$$\mathcal{L}V := \frac{\partial}{\partial t} V + \frac{\sigma^2}{2} S^2 \frac{\partial^2}{\partial S^2} V + f(S, t) \frac{\partial}{\partial I} V + rS \frac{\partial}{\partial S} V - rV = 0$$

mit der Endbedingung

$$V(S, I, T) = h(S, I)$$

erfüllt. Die genaue Form der Auszahlungsfunktion und der Randbedingungen hängt hierbei von der jeweiligen Art der Option ab. Sie sind für die wichtigsten pfadabhängigen Optionsformen in [WHD1] aufgeführt. Der Wert einer pfadabhängigen Europäischen Option erfüllt somit ein instationäres Konvektions-Diffusions-Reaktions-Randwertproblem in zwei Raumdimensionen.

Bei einer pfadabhängigen Amerikanischen Option führt die Möglichkeit, die Option vorzeitig ausüben zu können, zu der Nebenbedingung

$$V(S, I, t) \geq h(S, I) \quad \text{für alle } 0 \leq t \leq T$$

und damit auf ein instationäres freies Randwertproblem in zwei Raumdimensionen mit der Bedingung, dass V und $\partial V/\partial S$ auch über den freien Rand hinweg stetig sind. Es kann in linear komplementärer Form formuliert werden als

$$\begin{aligned} (\mathcal{L}V) \cdot (V - h) &= 0, \\ \mathcal{L}V &\leq 0, \quad V - h \geq 0. \end{aligned}$$

Für pfadabhängige Europäische Optionen lassen sich in einigen Fällen analytische Lösungen angeben. In einigen anderen ist eine Dimensionsreduktion von zwei Dimensionen auf eine möglich. Ansonsten muss auf numerische Verfahren zurückgegriffen werden. Dies ist generell der Fall, wenn es sich um pfadabhängige Amerikanische Optionen handelt.

Kapitel 6

Numerische Ergebnisse

In den vorhergehenden Kapiteln wurde ein auf B-Splines beliebiger Glattheit basierendes Finite-Elemente-Verfahren und verschiedene Varianten monotoner Mehrgitterverfahren zur numerischen Lösung des zeitabhängigen freien Randwertproblems 2.1.1 vorgestellt, welches bei der Bewertung von Amerikanischen Optionen auftritt. Die Konvergenzeigenschaften der Verfahren sollen in diesem Kapitel untersucht und verglichen werden.

Die Korrektheit der Algorithmen wird zunächst anhand des unrestringierten Randwertproblems 2.1.2 überprüft, für das eine analytische Lösung bekannt ist. Der Lösungsalgorithmus ändert sich dabei im Vergleich zum restringierten Problem nur insofern, dass das Hindernis so gesetzt wird, dass es keinen restriktiven Einfluss mehr auf die Lösung hat. Um auch die Ergebnisse für das schwierigere freie Randwertproblems 2.1.1 mit einem unabhängigen Verfahren vergleichen zu können, wurde zusätzlich das Binomialverfahren aus Unterabschnitt 1.4.1 implementiert.

Im ersten Abschnitt werden die Lösungen der beiden Randwertprobleme graphisch dargestellt. Der Verlauf der Lösungen und der Verlauf des freien Randes wird in Hinblick auf die theoretischen Aussagen aus Kapitel 1 diskutiert. In Abschnitt 6.2 werden dann die numerischen Ergebnisse bei der Bestimmung der Lösung und bei der Bestimmung der Ableitungen der Lösung für Ansatzfunktionen verschiedener Glattheit aufgeführt. Der Einfluss unterschiedlicher Parameter auf die Lösung ist in Abschnitt 6.3 dargestellt. In Abschnitt 6.4 wird das Konvergenzverhalten der verschiedenen Varianten des monotonen Mehrgitterverfahrens aus Kapitel 3 untersucht. Insbesondere werden die Iterationszahlen, die benötigt werden bis die Lösung bis auf Maschinengenauigkeit bestimmt ist, und die Iterationszahlen, die benötigt werden, bis die Kontaktmenge der Lösung richtig identifiziert ist, verglichen. Abschließend werden in Abschnitt 6.5 die numerischen Ergebnisse zu den asymptotischen Mehrgitter-Konvergenzraten in Abhängigkeit von der Anzahl der Unbekannten und in Abhängigkeit von der Anzahl der Glättungsschritte dargestellt.

Sämtliche Algorithmen sind auf einem Intel Pentium 4 CPU mit 2,66 Gigahertz in der Programmiersprache C implementiert worden. Der Quellcode umfasst etwa 2500 Zeilen.

6.1 Europäische und Amerikanische Optionspreise

Wie in Abschnitt 1.2 hergeleitet wurde, beschreibt die Lösung des unrestringierten Randwertproblems 2.1.2 den Wert einer Europäischen Kaufoption und die Lösung des freien Randwertproblems 2.1.1 den Wert einer Amerikanischen Verkaufsoption. Mit N bzw. M wird wie bisher die Anzahl der Gitterpunkte in der Zeit bzw. im Ort bezeichnet. Für die Darstellungen in diesem Abschnitt wurde generell $M = N = 64$ gewählt. Wenn es nicht explizit anders erwähnt wird, werden die Parameter aus Tabelle 6.1 verwendet.

Ausübungspreis	$K = 10$
Fälligkeitszeitpunkt	$T = 1$
Volatilität	$\sigma = 0.6$
Dividende	$d = 0$
Zinssatz	$r = 2.5\%$

Tabelle 6.1: Gewählte Parameter für die Randwertprobleme 2.1.1 und 2.1.2.

Der Preis $V(S, t)$ einer Europäischen Kaufoption ist in Abhängigkeit vom Aktienkurs S und der Zeit t in der linken Abbildung 6.1 dargestellt. Über die Put–Call–Parität aus Satz 1.1.11 kann aus dem Preis einer Europäischen Kaufoption der Preis der entsprechenden Verkaufsoption bestimmt werden. Dieser ist in der rechten Abbildung 6.1 dargestellt. Für $t = 1$ erkennt die Randdaten $V(S, T) = (S - K)^+ = \mathcal{H}_c(S)$ bzw. $V(S, T) = (K - S)^+ = \mathcal{H}_p(S)$ mit der Nicht-Differentierbarkeitsstelle am Punkt $(10, 1)$.

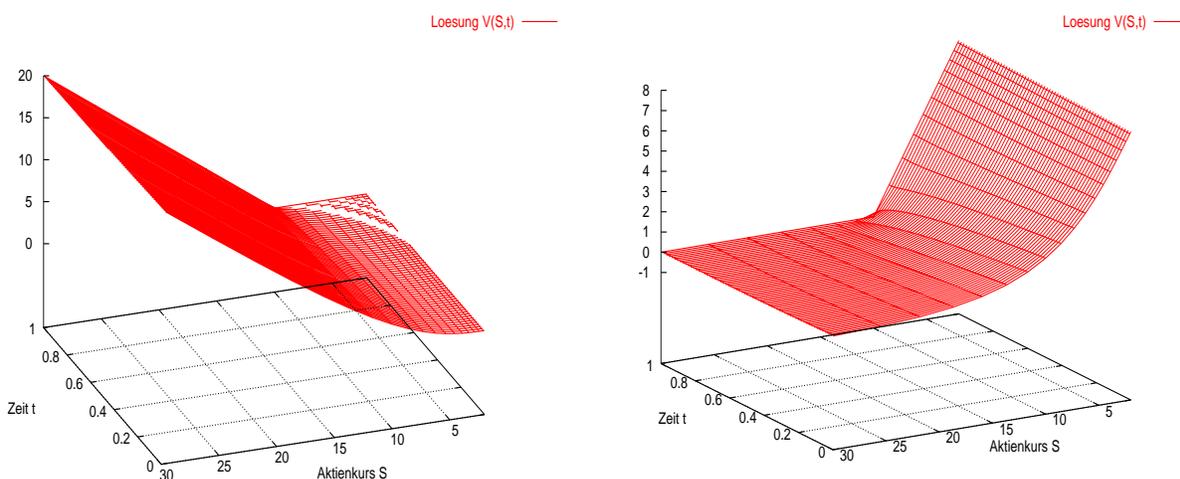


Abbildung 6.1: Preis einer Europäischen Kaufoption (links) und einer Europäischen Verkaufsoption (rechts).

Solange keine Dividendenzahlungen auftreten, entspricht der Preis einer Amerikanischen Kaufoption nach Satz 1.1.9 immer dem einer Europäischen. Für Verkaufsoptionen gilt dies aber nicht, wie in Satz 1.1.10 gezeigt wurde. Der Preis einer Amerikanischen Verkaufsoption entspricht der Lösung $V(S, t)$ des freien Randwertproblems 2.1.1. Er ist zusammen mit der zu berücksichtigenden Hindernisfunktion \mathcal{H} in der Abbildung 6.2 dargestellt. In der linken Abbildung wurden die Parametern aus Tabelle 6.1 gewählt. Die in der rechten Abbildung dargestellte Lösung ergibt sich, wenn man eine etwas höhere Volatilität $\sigma = 0.8$ des Aktienkursprozesses annimmt.

Man erkennt, dass die Lösung V immer oberhalb der Hindernisfunktion $\mathcal{H} = \mathcal{H}(S, t)$ liegt. Dies bestätigt die theoretische Aussage aus (1.2). Die Hindernisfunktion \mathcal{H} beschreibt den inneren Wert einer Amerikanischen Option. Der Wert einer Amerikanischen Option ist also immer mindestens so hoch wie ihr innerer Wert. Zum Fälligkeitszeitpunkt T entspricht der innere Wert der Auszahlungsfunktion.

Speziell für den Zeitpunkt $t = 0$ sind in Abbildung 6.3 die unterschiedlichen Verläufe der Lösun-

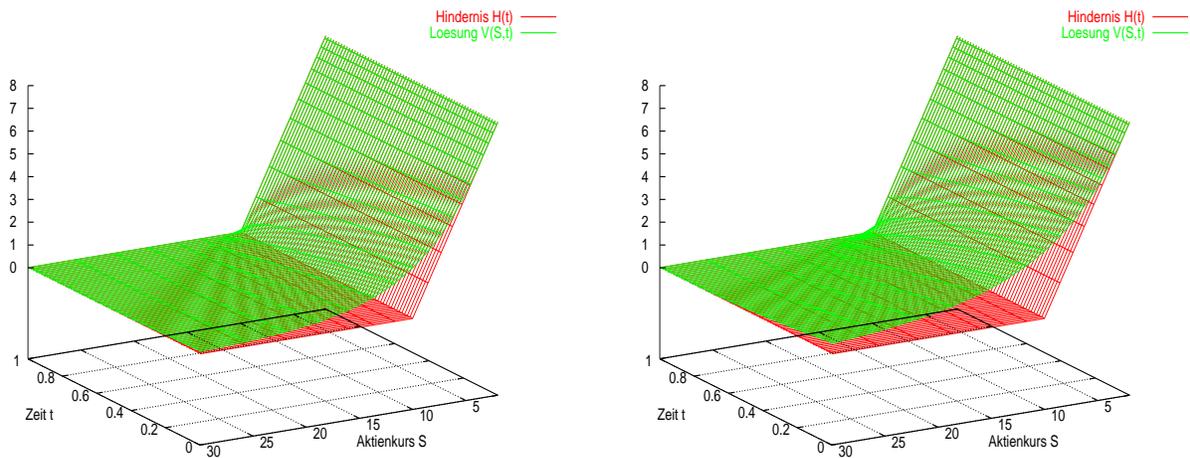


Abbildung 6.2: Preis $V(S, t)$ einer Amerikanischen Verkaufsoption mit Volatilität $\sigma = 0.6$ (links) und $\sigma = 0.8$ (rechts).

gen des unrestringierten Randwertproblems 2.1.2 und des restringierten Randwertproblems 2.1.1 in Vergleich mit der Auszahlungsfunktion $\mathcal{H}(S) = (K - S)^+$ dargestellt. Um den Unterschied möglichst deutlich zu machen, wurde abweichend von den Parametern aus Tabelle 6.1 hier die längere Laufzeit $T = 10$ gewählt. Man erkennt gut die restriktive Wirkung der Hindernisfunktion im Falle Amerikanischer Optionen. Zum Zeitpunkt $t = T$ entsprechen die Lösungen beider Probleme der Hindernisfunktion \mathcal{H} .

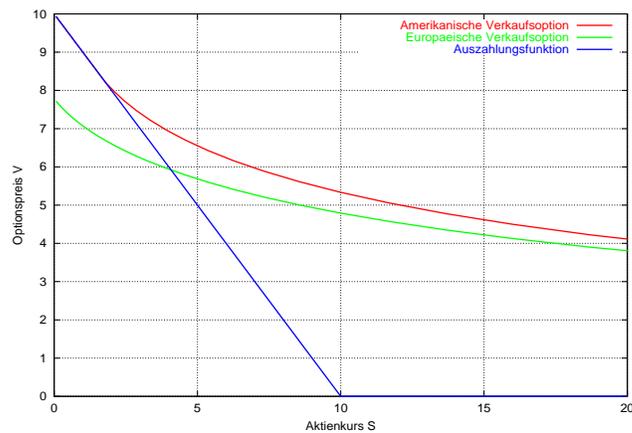


Abbildung 6.3: Preis einer Europäischen und einer Amerikanischen Verkaufsoption zum Zeitpunkt $t = 0$.

Es sei erinnert, dass in Lemma 2.2.7 bewiesen wurde, dass die Lösung des transformierten Problems 2.2.6 nicht nur homogene Randdaten hat, sondern sogar in einer größeren Umgebung der Ränder verschwindet. Dies wird in Abbildung 6.4 bestätigt. Dort ist die bereits zurücktransformierte Lösung des auf homogene Randdaten reduzierten Problems dargestellt. Man erkennt,

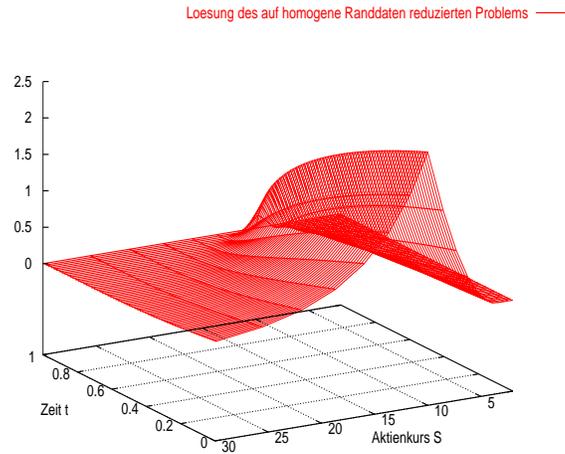


Abbildung 6.4: Lösung des auf homogene Randbedingungen reduzierten Problems.

dass die Lösung in den Bereichen $S < 4$ und $S > 28$ für beliebiges t verschwindet. Dies hat den großen Vorteil, dass die Modellierung der Randdaten einfach durch Nullsetzen der Rand B-Splines ohne Verlust an Approximationsgüte auch für B-Splines höherer Ordnung möglich ist.

Der freie Rand

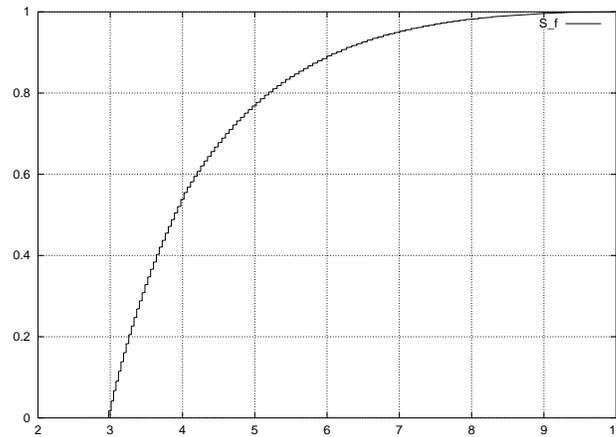
Der zeitabhängige freie Rand S_f , der in Problem 2.1.1 auftritt, und den optimalen Ausübungszeitpunkt einer Amerikanischen Option beschreibt, ist für die Parameter aus Tabelle 6.1 in Abbildung 6.5 dargestellt. Auf der x -Achse ist dort der Aktienkurs S , auf der y -Achse die Zeit t aufgetragen. Man erkennt die Abbildung 1.2 wieder.

Im Bereich links von S_f gilt $V(S, t) = \mathcal{H}(S)$ und Ausüben der Option ist optimal. Dagegen ist rechts von S_f die Black-Scholes Gleichung

$$\frac{\partial}{\partial t} V + \frac{\sigma^2}{2} S^2 \frac{\partial^2}{\partial S^2} V + rS \frac{\partial}{\partial S} V - rV = 0$$

erfüllt und es ist günstiger die Option zu halten. Man sieht in Übereinstimmung mit den theoretischen Aussagen aus Abschnitt 1.2.3, dass der freie Rand $S_f(t)$ eine stetige, monoton wachsende Funktion der Zeit ist und, dass $S_f(t) = K$ für $t \rightarrow T$ gilt.

Der freie Rand ist auch in der Abbildung 6.2 und speziell für $t = 0$ in der Abbildung 6.3 erkennbar. Er wird durch die Punkte gebildet, an denen (von links nach rechts gesehen) die Lösung V zum ersten Mal echt oberhalb der Hindernisfunktion \mathcal{H} liegt. Ab dort erfüllt die Lösung dann die Black-Scholes Gleichung. In den Abbildungen erkennt man gut den glatten Verlauf Amerikanischer Optionspreise auch über den freien Rand hinweg, solange $t < T$ gilt. Es sei erinnert, dass an nur C^1 -Glattheit vorliegt. Die Bestimmung der Ableitungen wird im nächsten Abschnitt behandelt.

Abbildung 6.5: Freier Rand S_f einer Amerikanischen Verkaufsoption.

Die Berechnung der rechten Seite

Zur Berechnung der rechten Seite des in jedem Zeitschritt zu lösenden Ungleichungssystems 2.3.4 bzw. Gleichungssystems 2.3.5 wurden alle drei Möglichkeiten aus Abschnitt 2.5.1 implementiert. Es wurden in allen drei Fällen keine signifikanten Unterschiede in der Qualität der Lösung festgestellt. In Tabelle 6.2 sind die Abweichungen der Finiten-Elemente Lösung $V(10, 0)$ des freien Randwertproblems 2.1.1 zu der Lösung, die ein Binomialverfahren mit $n = 256$ Teilintervallen liefert, angegeben. Die numerische Integration wurde hierbei mit dem Quasi-Monte-Carlo Verfahren und 5000 Stützstellen durchgeführt.

Bestimmung der rechten Seite	Abweichung
direkt nach Lemma 2.5.2	0,0028
mit Interpolation nach (2.40)	0,0038
mit numerischer Integration	0,0035

Tabelle 6.2: Abweichungen der Finite-Elemente Lösung $V(10, 0)$ bei unterschiedlicher Berechnung der rechten Seite zum Vergleichswert eines Binomialverfahrens mit $n = 256$ Teilintervallen.

6.2 Einfluss der Ordnung der Ansatzfunktionen auf die Qualität der Lösung

In diesem Abschnitt wird untersucht, in wieweit sich die Verwendung glatter Ansatzfunktionen in der Qualität der Lösung widerspiegelt. Insbesondere bei der Bestimmung der Ableitungen der Lösung werden Vorteile von der Wahl glatter Funktionen erwartet, da in diesem Fall ein direktes Ableiten der Ansatzfunktionen möglich ist. Zunächst werden die numerischen Ergebnisse für das Problem 2.1.2 behandelt und mit der analytischen Lösung verglichen.

Die Konvergenzaussagen aus Abschnitt 2.4.1 lassen bei Verwendung von Hutfunktionen und bei Verwendung des Crank-Nicholson Verfahrens ($\theta = \frac{1}{2}$) quadratische Konvergenz und bei Verwendung eines rein expliziten ($\theta = 0$) oder eines rein impliziten Euler Verfahrens ($\theta = 1$) lineare Konvergenz erwarten. Dies wird numerisch in Abbildung 6.6 bestätigt. Dort ist in logarithmischer Skala die Anzahl M der Unbekannten im Ort und der ℓ^2 -Fehler zum Zeitpunkt

$t = 0$ aufgetragen. Es wurden jeweils $N = M$ Zeitschritte gewählt. Wird die Zeitschrittweite $\Delta\tau$ im Vergleich zur Ortsschrittweite h zu gross gewählt, genauer $\Delta\tau \geq \frac{1}{2}h^2$, dann ist eine explizite Zeitdiskretisierung ($\theta < \frac{1}{2}$) nicht mehr stabil.

Wegen der schnelleren Konvergenz ist bei den folgenden Tests ausschließlich das Crank–Nicholson Verfahren verwendet worden. Mit dieser einfachen Art der Zeitdiskretisierung lässt sich auch mit Ansatzfunktionen höherer Ordnung im Ort nicht mehr als quadratische Konvergenz erreichen. Allerdings kann die Zeitschrittweite $\Delta\tau$ im Vergleich zur Ortsschrittweite h größer gewählt werden, ohne die quadratische Konvergenz zu verlieren, und zwar in der Ordnung $O(h^{2/k})$ bei Ansatzfunktionen der Ordnung $k \geq 2$. Abbildung 6.6 bestätigt die quadratische Konvergenz und zeigt für den Fall $O(\Delta\tau) = O(h)$, dass sich die Konstanten mit steigender Glattheit der Basisfunktionen leicht verbessern.

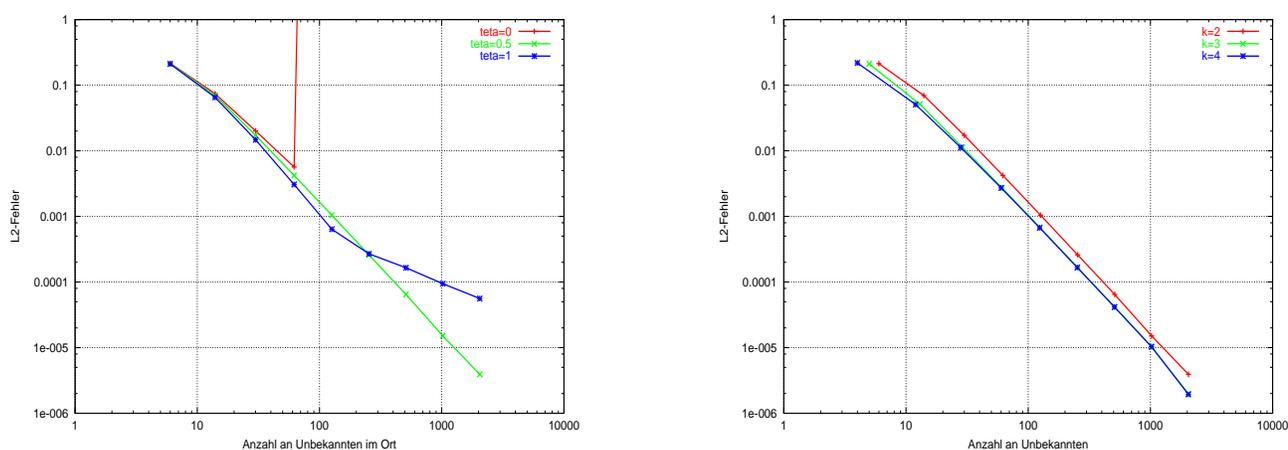


Abbildung 6.6: Links: Konvergenzraten für das Crank–Nicholson Verfahren ($\theta = \frac{1}{2}$), für das explizite ($\theta = 0$) und das implizite ($\theta = 1$) Eulerverfahren. Rechts: Konvergenzraten für stetige, stückweise lineare ($k = 2$), C^1 -glatte, quadratische ($k = 3$) und C^2 -glatte, kubische ($k = 4$) B-Splines.

Die Bestimmung der Ableitungen

Im folgenden werden die numerischen Ergebnisse bei der Bestimmung der Ableitungen der Lösung dargestellt. Dazu wird auf Ergebnisse und die Notation aus Abschnitt 1.3 zurückgegriffen. Dort wurde erläutert, dass sich im Rahmen der Optionsbewertung aus den Ableitungen die wichtigen Risikokennziffern Delta, Gamma und Theta ergeben. Der Verlauf von

$$\text{Delta} := \frac{\partial V(S, 0)}{\partial S} \quad \text{und} \quad \text{Gamma} := \frac{\partial^2 V(S, 0)}{\partial S^2},$$

also der ersten und zweiten Ableitung der Lösung sind in der Abbildung 6.7 für eine Amerikanische Verkaufsoption dargestellt. In der linken Abbildung ist gut erkennbar, dass die Nicht-Differenzierbarkeitsstelle in den Anfangsdaten \mathcal{H} am Punkt (K, T) zu einer Sprungstelle in der ersten Ableitung führt. Für $t < T$ sind Optionspreise C^1 -glatt. In der rechten Abbildung sieht man in Übereinstimmung mit den theoretischen Resultaten, dass die zweite Ableitung eine Sprungstelle an freien Randpunkten aufweist. Der Sprung ist umso größer, je größer t ist.

In Abbildung 6.8 sind für stetige, stückweise lineare und C^1 -glatte, stückweise quadratische Ansatzfunktionen die Konvergenzraten bei der Bestimmung der ersten Ableitung und die Ver-

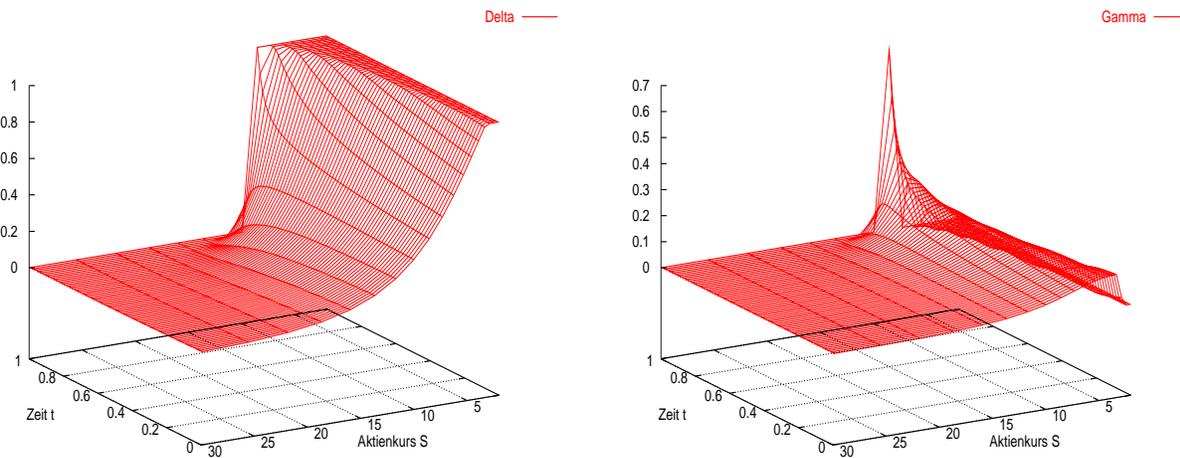


Abbildung 6.7: Delta := $\partial V/\partial S$ (links) und Gamma := $\partial^2 V/\partial S^2$ (rechts) einer Amerikanischen Verkaufsoption mit den Parametern aus Tabelle 6.1.

teilung des punktweisen Fehlers

$$\left| \frac{\partial V(S, 0)}{\partial S} - N(d_1) \right|$$

mit $N = M = 275$ dargestellt. Links ist in logarithmischer Skala der ℓ^2 -Fehler zum Zeitpunkt $t = 0$ in Abhängigkeit von der Anzahl der Unbekannten und rechts der punktweise Fehler in Abhängigkeit vom Aktienkurs S aufgetragen. Für Hutfunktionen ($k = 2$) können die Ableitungen nur durch numerische Differentiation bestimmt werden. Für Ordnungen $k \geq 3$ wurde die Formel aus Lemma 2.5.4 verwendet. Die Ableitungen werden von glatten Funktionen erwartungsgemäß wesentlich genauer approximiert als von stückweise linearen. Mit B-Splines der Ordnungen $k \geq 3$ wird auf diese Weise die $(k - 2)$ -te Ableitung der Lösung mit quadratischer Konvergenz bestimmt.

Analog sind in Abbildung 6.9 die Konvergenzraten für stetige, stückweise lineare und C^2 -glatte, kubische Ansatzfunktionen bei der Bestimmung der zweiten Ableitung und die Verteilung des punktweisen Fehlers

$$\left| \frac{\partial^2 V(S, 0)}{\partial S^2} - \frac{N'(d_1)}{S \sigma \sqrt{T}} \right|$$

der gleichen Option mit $N = M = 275$ abgebildet. Bemerkenswert ist, dass mit numerischer Differentiation der Hutfunktionen immer noch die Konvergenzordnung $O(h)$ erreicht wird. Die theoretischen Resultate aus Kapitel 2.4.2 implizieren nur $O(1)$. Es liegt hier eine Superkonvergenzeigenschaft vor. Für C^2 -glatte stückweise kubische Funktionen ($k = 4$) wurde die zweite Ableitung direkt mit der Formel aus Lemma 2.5.4 bestimmt und es wird quadratische Konvergenz erreicht.

In den Tabellen 7.3 – 7.5 sind alle Ergebnisse für B-Splines der Ordnungen $k = 2, 3, 4$ noch einmal zusammengefasst. Die exakte Lösung ist mit V die Näherungslösung mit V_h bezeichnet. Die Anzahl an verwendeten Gitterpunkten ist in der ersten Spalte angegeben. Jeweils in der 2., 6. und 7. Spalte sind die ℓ^2 -Fehler der Lösung, der ersten Ableitung und der zweiten Ableitung dargestellt. Der maximale punktweise Fehler findet sich in der 3. Spalte. In der 4.

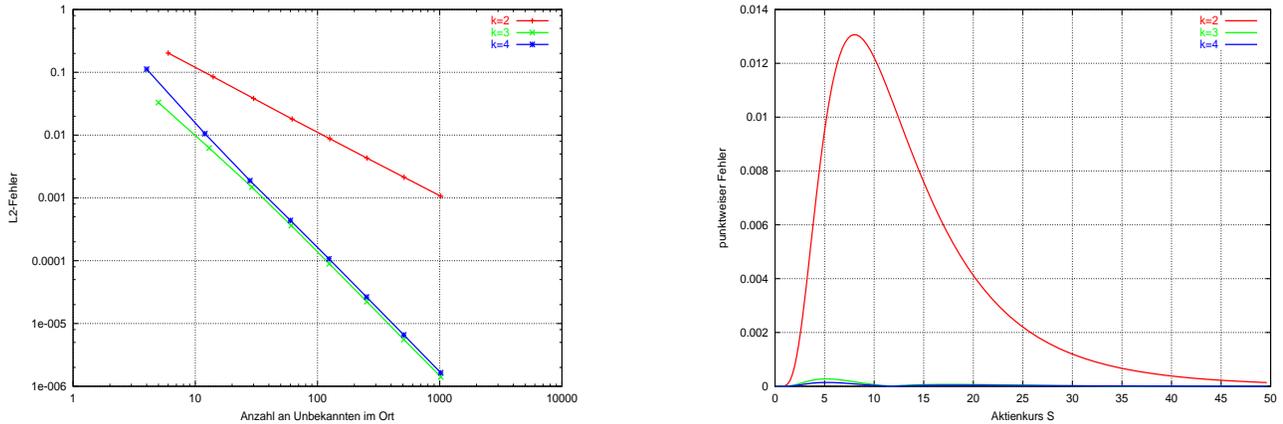


Abbildung 6.8: Konvergenzraten (links) und Fehlerverteilung (rechts) bei der Bestimmung der ersten Ableitung.

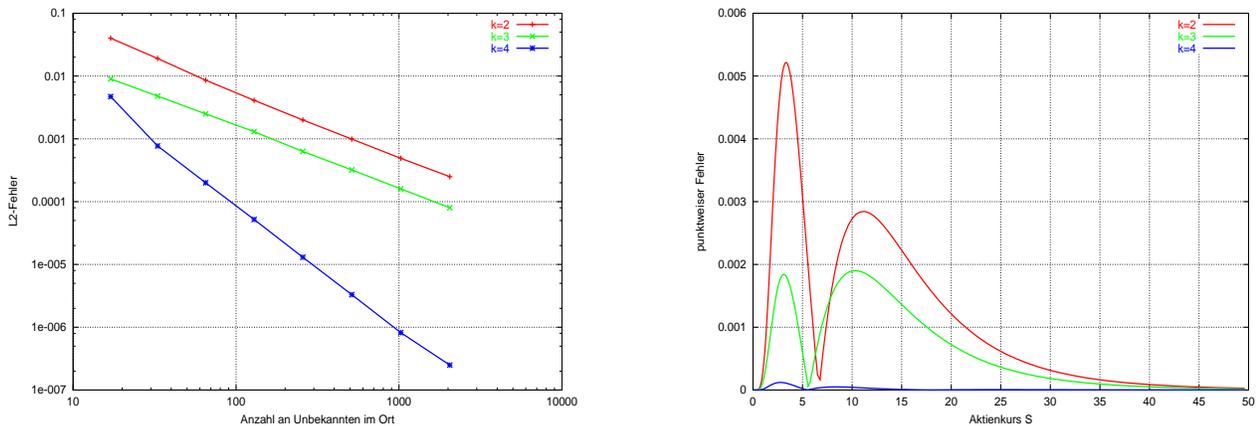


Abbildung 6.9: Konvergenzraten (links) und Fehlerverteilung (rechts) bei der Bestimmung der zweiten Ableitung.

und 5. Spalte ist der Wert S bzw. t angegeben, an dem der maximale Fehler gemessen wurde. In der letzten Spalte ist schließlich die Näherungslösung am Punkt $(K, 0)$ angegeben. Der exakte Preis $V(K, 0) = 2,4545$ entspricht dem fairen Preis einer Europäischen Option zum Zeitpunkt $t = 0$, falls ein Ausübungspreis K in gleicher Höhe des aktuellen Aktienkurses S vereinbart wird. Zusammenfassend beobachtet man bei Inspektion der Werte in Hinblick auf die Ordnung der Ansatzfunktionen in allen Segmenten geringere Fehler bei steigender Glattheit der Basisfunktionen.

# Gitterpunkte	$\ V_h - V\ _{\ell^2}$	$\ V_h - V\ _{\infty}$ bei (S, t)	$\ \Delta_h - \Delta\ _{\ell^2}$	$\ \Gamma_h - \Gamma\ _{\ell^2}$	$V_h(K, 0)$		
9 × 9	1.6e-001	3.5e-001	17.43	0.22	1.5e-001	9.6e-002	4.6763
17 × 17	5.6e-002	1.9e-001	13.42	0.71	7.4e-002	4.0e-002	3.1262
33 × 33	1.5e-002	9.3e-002	11.64	0.91	3.6e-002	1.9e-002	2.6404
65 × 65	4.0e-003	4.6e-002	10.80	0.98	1.7e-002	8.6e-003	2.5032
129 × 129	1.0e-003	2.4e-002	10.40	0.99	8.6e-003	4.1e-003	2.4669
257 × 257	2.6e-004	1.3e-002	10.20	1.00	4.3e-003	2.0e-003	2.4577
513 × 513	6.4e-005	8.7e-003	10.10	1.00	2.1e-003	9.9e-004	2.4553
1025 × 1025	1.5e-005	7.0e-003	10.05	1.00	1.1e-003	4.9e-004	2.4547
2049 × 2049	3.9e-006	6.1e-003	10.02	1.00	5.3e-004	2.5e-004	2.4546
exakt	—	—	—	—	—	—	2.4545

Tabelle 7.3: Ergebnisse für stetige, stückweise lineare Elemente ($k = 2$).

# Gitterpunkte	$\ V_h - V\ _{\ell^2}$	$\ V_h - V\ _{\infty}$ bei (S, t)	$\ \Delta_h - \Delta\ _{\ell^2}$	$\ \Gamma_h - \Gamma\ _{\ell^2}$	$V_h(K, 0)$		
9 × 9	1.5e-001	3.1e-001	17.43	0.22	2.1e-002	4.5e-002	1.9544
17 × 17	3.8e-002	1.5e-001	13.42	0.76	4.9e-003	9.0e-003	2.2934
33 × 33	1.0e-002	7.6e-002	11.64	0.94	1.3e-003	4.8e-003	2.4092
65 × 65	2.6e-003	3.8e-002	10.80	0.98	3.4e-004	2.5e-003	2.4425
129 × 129	6.5e-004	1.8e-002	10.40	0.99	8.7e-005	1.3e-003	2.4514
257 × 257	1.6e-004	8.9e-003	10.20	1.00	2.2e-005	6.3e-004	2.4537
513 × 513	4.1e-005	6.1e-003	10.10	1.00	5.5e-006	3.2e-004	2.4543
1025 × 1025	1.0e-005	5.6e-003	10.05	1.00	1.4e-006	1.6e-004	2.4545
2049 × 2049	5.5e-007	5.4e-003	10.02	1.00	6.9e-007	8.0e-005	2.4545
exakt	—	—	—	—	—	—	2.4545

Tabelle 7.4: Ergebnisse für C^1 - glatte, stückweise quadratische Elemente ($k = 3$).

# Gitterpunkte	$\ V_h - V\ _{\ell^2}$	$\ V_h - V\ _{\infty}$ bei (S, t)	$\ \Delta_h - \Delta\ _{\ell^2}$	$\ \Gamma_h - \Gamma\ _{\ell^2}$	$V_h(K, 0)$		
9 × 9	1.3e-001	2.7e-001	17.43	0.33	4.8e-002	4.1e-001	1.9648
17 × 17	3.8e-002	1.5e-001	13.42	0.76	7.3e-003	4.7e-003	2.2889
33 × 33	1.0e-002	7.4e-002	11.64	0.94	1.7e-003	7.7e-004	2.4089
65 × 65	2.6e-003	3.7e-002	10.80	0.98	4.1e-004	2.0e-004	2.4425
129 × 129	6.5e-004	1.8e-002	10.40	0.99	1.0e-004	5.2e-005	2.4514
257 × 257	1.6e-004	8.7e-003	10.20	1.00	2.6e-005	1.3e-005	2.4537
513 × 513	4.1e-005	6.0e-003	10.10	1.00	6.5e-006	3.3e-006	2.4543
1025 × 1025	1.0e-005	5.6e-003	10.05	1.00	1.6e-006	8.2e-007	2.4545
2049 × 2049	5.5e-007	5.4e-003	10.02	1.00	3.0e-007	2.5e-007	2.4545
exakt	—	—	—	—	—	—	2.4545

Tabelle 7.5: Ergebnisse für C^2 - glatte, stückweise kubische Elemente ($k = 4$).

6.3 Einfluss der Parameter

In Tabelle 7.6 ist der Einfluss der Parameter auf die Lösungen V^{am} bzw. V^{eur} der Randwertprobleme 2.1.1 bzw. 2.1.2 untersucht. In der ersten Zeile sind die Ergebnisse mit den Parametern aus Tabelle 6.1 aufgeführt. In den nachfolgenden Zeilen wurde jeweils der in der ersten

Spalte angegebene Parameter variiert. Im Finite-Elemente-Ansatz sind die Optionswerte mit $M \times N = 257 \times 257$ Gitterpunkten berechnet worden. Zur Validierung der Ergebnisse des Finite-Elemente-Verfahrens sind in der letzten Spalte die Amerikanischen Optionswerte angegeben, die ein Binomialverfahren mit $n = 256$ Teilintervallen liefert.

Parameter	$V^{\text{eur}}(K, 0)$	$V^{\text{am}}(K, 0)$		
		Finite-Elem.	Binomialverf.	Abweichung
Tabelle 6.1	2.2076	2.2312	2.2299	0.0013
$r = 10\%$	1.8004	1.9232	1.9226	0.006
$\sigma = 0.4$	1.4451	1.4681	1.4669	0.0012
$T = 0.5$	1.6081	1.6181	1.6167	0.0014
$T = 0.1$	0.7424	0.7441	0.7432	0.009
$S = 9$	2.61	2.6381	2.6401	0.002
$S = 11$	1.8711	1.8899	1.8910	0.0011

Tabelle 7.6: Einfluss der Parameter auf Europäische und Amerikanische Optionswerte.

Im Übereinstimmung mit den theoretischen Aussagen aus Kapitel 1 erkennt man, dass der Preis einer Verkaufsoption fallend im Aktienkurs S , im Zinssatz r , in der Volatilität σ und steigend in der Laufzeit T ist. Amerikanische Optionen sind generell teurer als Europäische Optionen. Mit abnehmender Laufzeiten $T \rightarrow 0$ nähern sich die Werte von Amerikanischen und Europäischen Optionen aneinander an.

6.4 Konvergenzverhalten des PSOR- und der Mehrgitterverfahren

Im folgenden wird jeweils nur noch ein Zeitschritt des Finite-Elemente-Verfahrens betrachtet, um die Wirkung des Verfahrens zu analysieren. In Abbildung 6.10 sind für $M = 256$ zunächst die Iterationsfehler des projektiven Gauss-Seidel-Verfahrens für Finite-Elemente-Funktionen unterschiedlicher Ordnung verglichen. Der Einfluss der Ordnung auf die Güte des Verfahrens ist deutlich sichtbar.

Für Hutfunktionen ist die Anzahl n der Entwicklungskoeffizienten gleich der Anzahl M der Stützstellen. Allgemein gilt für B-Splines der Ordnung k

$$n = M + k - 2.$$

Bei Verwendung mehrerer Gitter mit Gitterweite $h_\ell = 2^{-\ell+1} h_1$ auf dem Level ℓ beträgt die Anzahl der Entwicklungskoeffizienten $n_{\ell,k}$ in Abhängigkeit vom Level ℓ und der Ordnung k

$$n_{\ell,k} = k 2^{\ell-1} + k - 1.$$

Hierbei bezeichnet h_1 die Gitterweite des größten Gitters. Für $\ell = 1, \dots, 10$ und $k = 2, 3, 4$ sind die Werte $n_{\ell,k}$ in Tabelle 7.7 aufgelistet.

Level ℓ	1	2	3	4	5	6	7	8	9	10
$n_{\ell,2}$	3	5	9	17	33	65	129	257	513	1025
$n_{\ell,3}$	5	8	14	26	50	98	194	386	770	1538
$n_{\ell,4}$	7	11	19	35	67	131	259	515	1027	2051

Tabelle 7.7: Anzahl an Freiheitsgraden in Abhängigkeit vom Level ℓ und der Ordnung k .

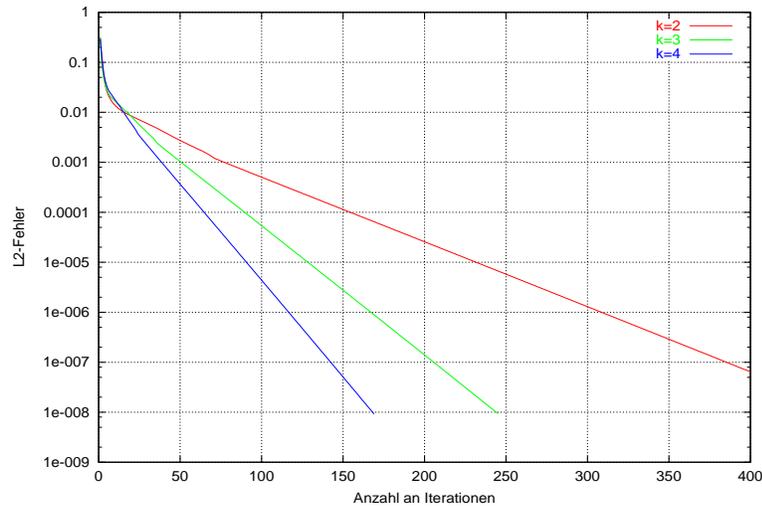


Abbildung 6.10: Vergleich der PSOR-Iterationsfehler eines Zeitschrittes für $k = 2$, $k = 3$ und $k = 4$.

In den Kapiteln 3 und 4 wurden die folgenden Verfahren vorgestellt.

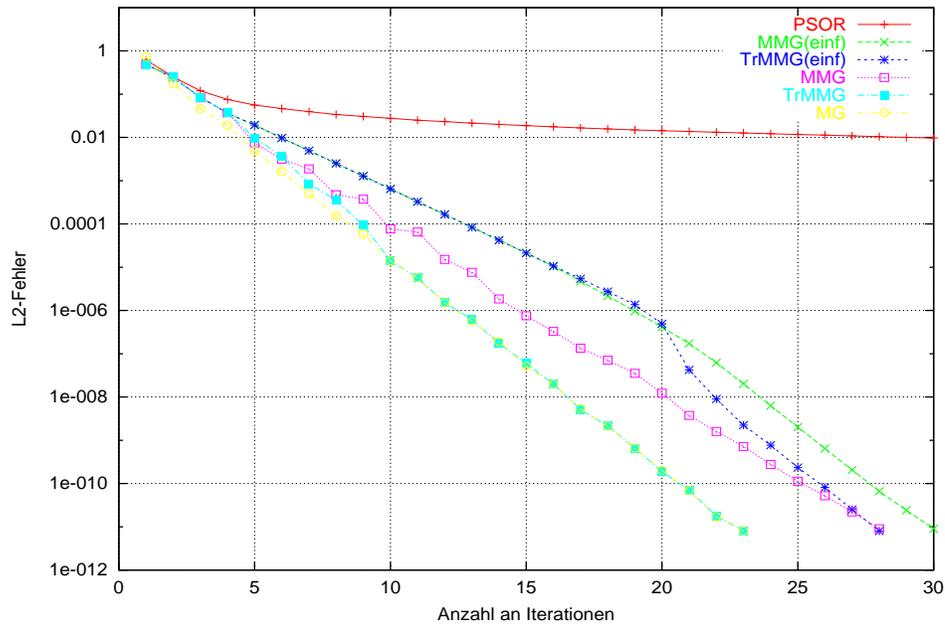
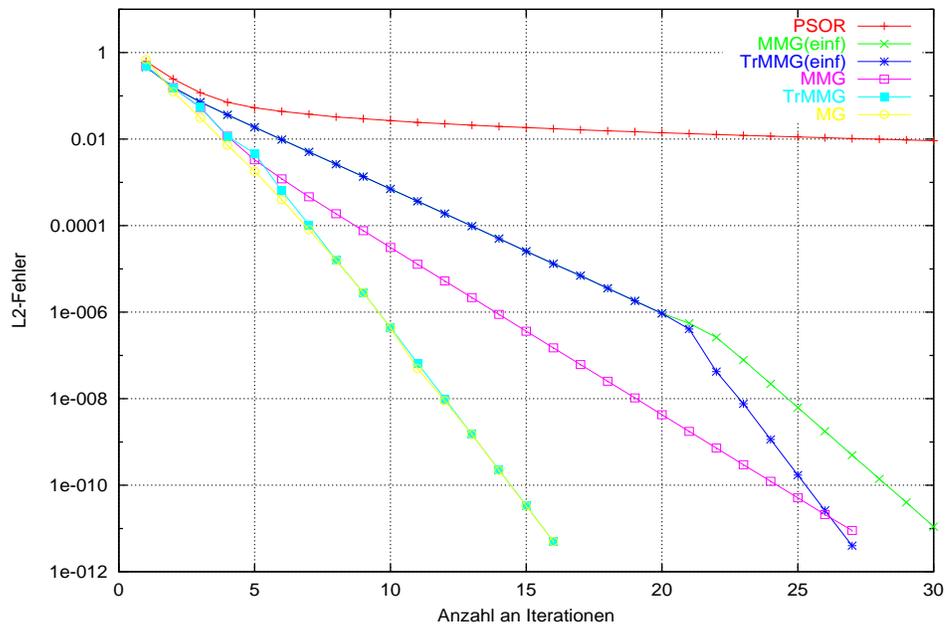
- PSOR: Projektives Gauss-Seidel-Verfahren
- MMG: Monotones Mehrgitterverfahren mit optimierter Approximation des Hindernisses nach Lemma 4.4.1 bzw. Satz 4.4.7
- TrMMG: Abgeschnittene Version des monotonen Mehrgitterverfahrens mit optimierter Approximation des Hindernisses nach Lemma 4.4.1 bzw. Satz 4.4.7
- MMG (einf): Monotones Mehrgitterverfahren mit einfacher Approximation des Hindernisses nach Proposition 4.2.1
- TrMMG (einf): Abgeschnittene Version des monotonen Mehrgitterverfahrens mit einfacher Approximation des Hindernisses nach Proposition 4.2.1
- MG: Lineares Mehrgitterverfahren angewandt auf das unrestringierte Problem.

Im folgenden werden die unterschiedlichen Konvergenzverhalten der Verfahren verglichen. Um den Einfluss der Ordnung k auf das Konvergenzverhalten zu untersuchen, wurde systematisch der Fall $k = 3$ mit dem Fall $k = 2$ verglichen. Für den Fall $k > 3$ erwartet man entsprechende Ergebnisse. Hierfür wurde mit den Parametern aus Tabelle 6.1 auf dem Level $J = 7$ ein Zeitschritt des instationären freien Randwertproblems betrachtet und für diesen das resultierende elliptische Problem bis auf Maschinengenauigkeit gelöst. Bei den Mehrgitterverfahren wird auf jedem Level nur ein Glättungsschritt durchgeführt. Es wurde eine zufällige Anfangsnäherung verwendet. Um sicherzustellen, dass nicht zu früh abgebrochen wird, wurde unabhängig vom Diskretisierungsfehler das Stoppkriterium

$$\|u_J^{\nu+1} - u_J^\nu\|_\infty \leq 10^{-12}$$

verwendet. Dabei bezeichnet u_J^ν die ν -te Iterierte auf dem feinsten Level J .

Die Ergebnisse für den Fall $k = 2$ und $k = 3$ sind in den Abbildungen 6.11 und 6.12 und in den Tabellen 7.8 und 7.9 zusammengefasst. In den Abbildungen sind für das Level $J = 7$ die Iterationsfehler sämtlicher Verfahren bis Erreichen der Maschinengenauigkeit dargestellt.

Abbildung 6.11: Vergleich der Iterationsfehler für Hutfunktionen ($k = 2$)Abbildung 6.12: Vergleich der Iterationsfehler für C^1 -glatte Ansatzfunktionen ($k = 3$)

In der Tabellen 7.8 wird für $k = 2$ und für $k = 3$ in der dritten Spalte die Anzahl ν_0 an Iterationen verglichen, die benötigt wird, um die Menge $K^\bullet(u_J)$ der Kontaktpunkte (s. Definition 3.2.4) richtig zu bestimmen. In der nächsten Spalte $\#$ It. wird festgehalten, wieviele Iterationen benötigt werden, um das Problem bis auf Maschinengenauigkeit zu lösen.

Um den unterschiedlichen Aufwand der Verfahren zu vergleichen, wird auf die Definition einer Workunit (WU) aus [BC] zurückgegriffen. Danach bezeichnet $WU = WU_J$ den Aufwand für einen Iterationsschritt des projektiven Gauss–Seidel–Verfahrens auf dem feinsten Level J . Der Aufwand WU_ℓ eines Iterationsschritt auf dem Level $\ell \leq J$ beträgt dann

$$WU_\ell = 2^{J-\ell} WU_J.$$

Der Aufwand in WU – Einheiten, den die Verfahren bis Erreichen des Stoppkriteriums benötigen, steht in der Spalte $\#$ WU. Die Ergebnisse sind jeweils für ein und für zwei Glättungsschritte auf jedem Mehrgitter–Level aufgeführt.

	Verfahren	1 Glättungsschritt			2 Glättungsschritte		
		ν_0	$\#$ It.	$\#$ WU	ν_0	$\#$ It.	$\#$ WU
$k = 2$	PSOR	134	403	403	-	-	-
	MMG (einf)	6	30	59.06	5	21	82.69
	TrMMG (einf)	7	28	55.13	5	17	66.94
	MMG	7	28	55.13	5	14	55.13
	TrMMG	7	23	45.28	5	13	51.19
$k = 3$	PSOR	103	447	447	-	-	-
	MMG (einf)	5	31	61.03	4	20	78.75
	TrMMG (einf)	6	27	53.16	4	17	66.94
	MMG	5	27	53.16	4	14	55.13
	TrMMG	5	16	31.5	4	11	43.31

Tabelle 7.8: Anzahl an Iterationen, die zur Identifikation der Kontaktmenge bzw. zur Lösung bis auf Maschinengenauigkeit benötigt werden und Aufwand in WU –Einheiten.

Man sieht an den Ergebnissen, dass es im Hinblick auf Aufwand und Genauigkeit bereits reicht nur ein bis zwei Glättungsschritte auf jedem Verfeinerungslevel bei den Mehrgitterverfahren durchzuführen. Der Aufwand ist bei allen Mehrgitterverfahren im Vergleich zum projektiven Gauss–Seidel–Verfahren deutlich reduziert. Die abgeschnittenen Versionen TrMMG und TrMMG (einf), in der das Hindernis an den aktuellen freien Rand angepasst wird, konvergieren in allen Fällen schneller als die Standardversionen MMG bzw. MMG (einf). Weiter erkennt man, dass die Mehrgitterverfahren mit einer optimierten Approximation des Hindernisses nach Lemma 4.4.1 bzw. Satz 4.4.7 schneller konvergieren als die Mehrgitterverfahren mit der einfachen Approximation nach Proposition 4.2.1. Dies spricht für die Qualität der mit dem OGK–Algorithmus in Kapitel 4 konstruierten näherungsweise monotonen und quasioptimalen Restriktionen. Die Kontaktmenge wird von allen Mehrgitterverfahren innerhalb weniger Iterationen richtig identifiziert.

Bei Inspektion der Werte in Hinblick auf die Ordnung k erkennt man, dass alle Verfahren die Kontaktmenge mit glatten Ansatzfunktionen schneller identifizieren. Außerdem konvergieren die Verfahren (mit Ausnahme von MMG (einf)) etwas schneller und weisen einen etwas geringeren Aufwand auf. Dabei ist zu beachten, dass mit dem Level $J = 7$ für $k = 2$ weniger Freiheitsgrade verbunden sind als für $k = 3$, wie aus Tabelle 7.7 ersichtlich ist. Daher werden die Werte für $k = 3$ zusätzlich mit den Werten für $k = 2$ auf dem Level $J = 8$ aus Tabelle 7.9 verglichen. Hierbei stellt man in allen Segmenten zum Teil sogar deutlich bessere Werte bei glatten Ansatzfunktionen fest.

Verfahren	1 Glättungsschritt			2 Glättungsschritte		
	ν_0	# It.	# WU	ν_0	# It.	# WU
PSOR	641	1962	1962	-	-	-
MMG (einf)	10	46	91.28	6	27	107.16
TrMMG (einf)	10	36	71.44	7	23	91.28
MMG	9	28	55.56	5	24	95.25
TrMMG	9	23	45.64	5	14	55.56

Tabelle 7.9: Ergebnisse für $k = 2$ auf dem Level $J = 8$

Zur Einbettung in die Zeitdiskretisierung ist zu bemerken, dass bei der Lösung des instationären Problems die durchschnittliche Anzahl an benötigten Iterationen pro Zeitschritt deutlich niedriger ausfällt, da in diesem Fall mit der Lösung des vorhergehenden Zeitschrittes gute Startnäherungen zu Verfügung stehen. In diesem Fall kann man davon ausgehen, dass die asymptotische Phase das Konvergenzverhalten der Mehrgitterverfahren dominieren wird. Die asymptotischen Mehrgitter-Konvergenzraten werden abschließend im folgenden Abschnitt diskutiert.

6.5 Mehrgitter-Konvergenzraten

Die Konvergenzrate ρ_ℓ eines Mehrgitterverfahrens mit $\ell + 1$ Ebenen ist gegeben durch

$$\|u_\ell^{\nu+1} - u_\ell\| \leq \rho_\ell \|u_\ell^\nu - u_\ell\|.$$

Hierbei bezeichnet $u_\ell \in S_\ell$ die exakte Lösung und $u_\ell^\nu \in S_\ell$ die Näherungslösung im ν -ten Iterationsschritt. Mehrgitterkonvergenz liegt dann vor, wenn ρ_ℓ durch eine Konstante $\rho_\infty < 1$ beschränkt ist, die unabhängig von der Gitterweite ist.

Die asymptotischen Konvergenzraten sollen für den V-Zyklus der abgeschnittenen Version des monotonen Mehrgitterverfahrens TrMMG mit $\ell + 1$ Ebenen nach

$$\rho_\ell \approx \frac{\|u_\ell^{\nu^*+1} - u_\ell^{\nu^*}\|_{\ell^2}}{\|u_\ell^{\nu^*} - u_\ell^{\nu^*-1}\|_{\ell^2}}$$

geschätzt werden. Dabei wird ν^* so gewählt, dass

$$\|u_\ell^{\nu^*+1} - u_\ell^{\nu^*}\|_{\ell^2} \leq 10^{-12}.$$

In der Abbildungen 6.13 sind die Ergebnisse links für stetige, stückweise lineare und rechts für C^1 -glatte, stückweise quadratische Ansatzfunktionen in Abhängigkeit von der Anzahl $n_{\ell,k}$ an Unbekannten dargestellt. Zusätzlich sind in der Abbildung 6.14 die asymptotischen Konvergenzraten auf dem Level $J = 10$ in Abhängigkeit von der Anzahl ν an Glättungsschritten aufgeführt. Die Ergebnisse entsprechen den bekanntermaßen guten Konvergenzraten von Standard-Mehrgitterverfahren. Bei nur einem Glättungsschritt pro Verfeinerungslevel erhält man $\rho_\infty \approx 0.31$ für $k = 2$ und $\rho_\infty \approx 0.27$ für $k = 3$. Eine Erhöhung der Anzahl der Glättungsschritte, damit auch des Aufwandes, führt erwartungsgemäß zu niedrigeren Konvergenzraten. Bei mehr als fünf Glättungsschritten verbessern sich die Konvergenzraten kaum noch und liegen etwa bei $\rho_\infty \approx 0.035$.

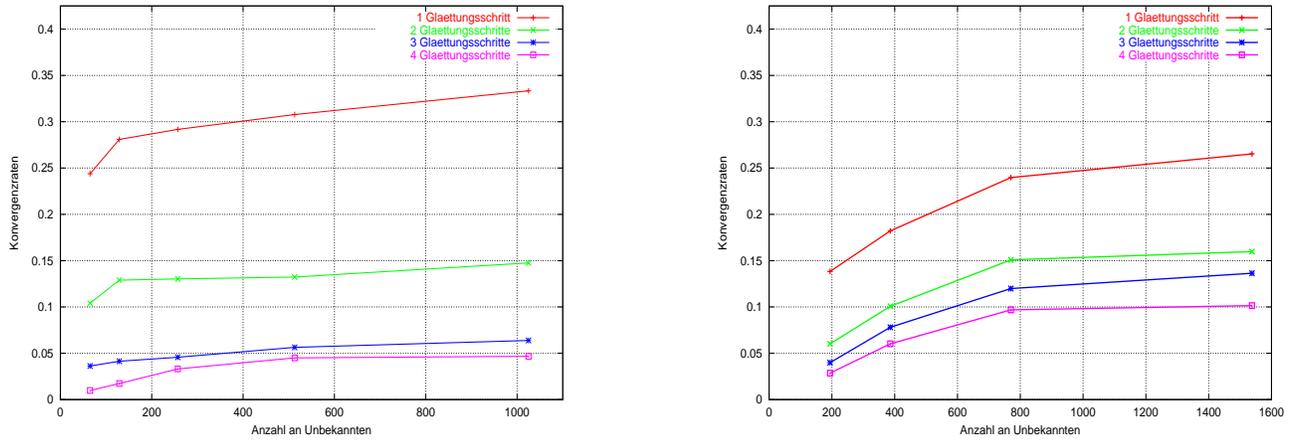


Abbildung 6.13: Konvergenzraten für stetige, stückweise lineare (links) und C^1 – glatte, stückweise quadratische Funktionen (rechts) in Abhängigkeit von der Anzahl M der Unbekannten.

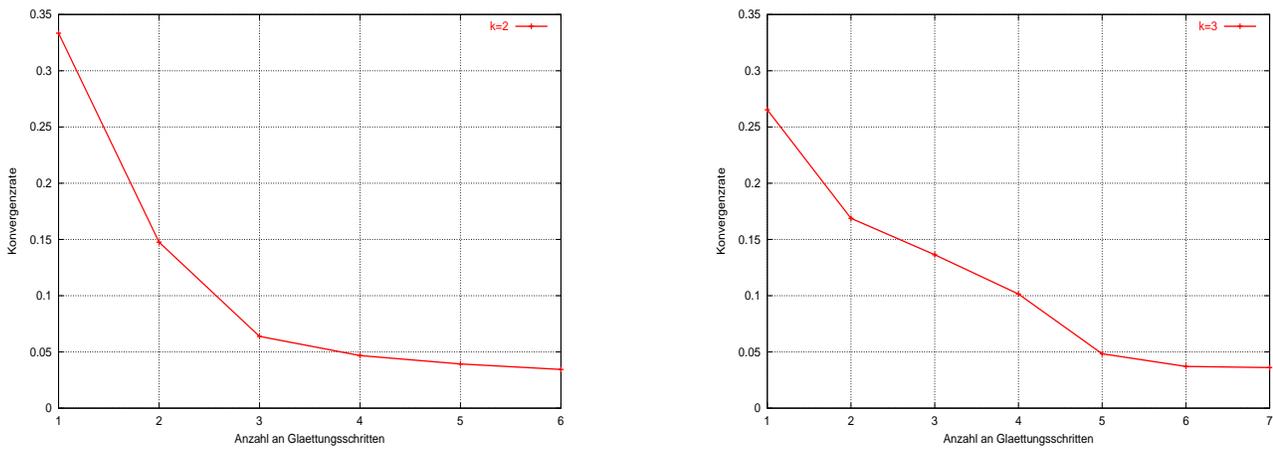


Abbildung 6.14: Konvergenzraten für stetige, stückweise lineare (links) und C^1 – glatte, stückweise quadratische Funktionen (rechts) in Abhängigkeit von der Anzahl der Glättungsschritte.

Zusammenfassung

In dieser Arbeit wurde ein robustes und optimales B-Spline-basiertes Mehrgitterverfahren zur Bestimmung Amerikanischer Optionspreise und ihrer Risikokennziffern vorgestellt. Dafür wurde die Konstruktion und die Theorie der monotonen Mehrgitterverfahren aus [Ko1] erstmalig auf B-Spline-Ansatzfunktionen beliebiger Glattheit übertragen. Desweiteren wurde die Idee der monotonen Mehrgitterverfahren hier zum ersten Mal auf die Bewertung Amerikanischer Optionen angewandt.

Die wesentliche Aufgabe bestand darin, in optimaler Komplexität eine monotone und quasioptimale Grobgitterapproximation an die Hindernisfunktion zu konstruieren. Die Konstruktion wurde auf die Lösung einer linearen Optimierungsaufgabe zurückgeführt. Ausgehend davon konnte für Hutfunktionen über Fourier-Motzkin-Elimination eine direkte Lösung bestimmt werden. Dabei wurden die Approximationen aus [Ma, Ko1] wiederentdeckt. Für B-Splines beliebiger Glattheit wurde mit dem optimierten Grobgitterhindernis-Konstruktionsalgorithmus (OGK) ein neues Verfahren vorgestellt, welches in optimaler Komplexität eine monotone und quasioptimale näherungsweise Lösung der linearen Optimierungsaufgabe liefert.

Die hohe Qualität der Lösung wurde in verschiedenen numerischen Tests bestätigt. Insbesondere konvergieren mit den OGK-Approximationen ausgestattete monotone Mehrgitterverfahren schneller als solche mit nicht optimierten Approximationen. Bei der abgeschnittenen Version der monotonen Mehrgitterverfahren wurden asymptotisch die gleichen, bekanntermaßen guten Konvergenzraten linearer Mehrgitterverfahren festgestellt. Durch direktes Ableiten genügend glatter B-Spline-Ansatzfunktionen konnten die Risikokennziffern Europäischer und Amerikanischer Optionen, die durch die Ableitungen der Lösung einer Variationsgleichung bzw. Ungleichung gegeben sind, wesentlich exakter bestimmt werden als es für Hutfunktionen der Fall ist.

Wie die Ergebnisse über einen Tensorproduktansatz auf beliebige Dimensionen verallgemeinerbar sind, wurde in Kapitel 5 dargestellt. In diesem Zusammenhang ist dort auch bereits ein Ausblick auf mögliche mehrdimensionale Anwendungen aus der Optionspreistheorie für das dargestellte Mehrgitterverfahren gegeben. Eine robuste und effiziente Lösung solcher Probleme existiert bisher höchstens für Einzelfälle.

Anhang A

Mathematische Grundlagen

A.1 B-Splines

Seit den Arbeiten [Sb] und [Bo] kommt B-Spline-Techniken aufgrund ihrer guten Approximationseigenschaften eine große Bedeutung bei einer Vielzahl von Anwendungen zu. Die Definition und die wichtigsten Eigenschaften von B-Splines sollen in diesem Abschnitt zusammengestellt werden.

Für $k \in \mathbb{N}$ und $n \in \mathbb{N}$ sei $T := \{\theta_i\}_{i=1, \dots, n+k}$ eine Knotenfolge im Intervall $I := [a, b]$ der Form

$$(A.1) \quad \theta_1 = \dots = \theta_k = a < \theta_{k+1} < \dots < \theta_n < b = \theta_{n+1} = \dots = \theta_{n+k}.$$

Die Knotenfolge T wird als *erweiterte Knotenfolge* bezeichnet. Wir beschränken uns auf den Spezialfall nicht zusammenfallender Knoten (vgl. aber Bemerkung 4.4.12).

Definition A.1.1 (B-Splines) Die B-Splines $N_{i,k}$ der Ordnung k sind für $i = 1, \dots, n+k$ und für $x \in I$ rekursiv durch

$$(A.2) \quad N_{i,1}(x) = \begin{cases} 1, & \text{falls } x \in [\theta_i, \theta_{i+1}) \\ 0, & \text{sonst} \end{cases}$$
$$N_{i,k}(x) = \frac{x - \theta_i}{\theta_{i+k-1} - \theta_i} N_{i,k-1}(x) + \frac{\theta_{i+k} - x}{\theta_{i+k} - \theta_{i+1}} N_{i+1,k-1}(x)$$

definiert.

Für die Ordnungen $k = 2, 3, 4$ sind die B-Splines $N_{i,k}$ in Abbildung A.1 dargestellt.

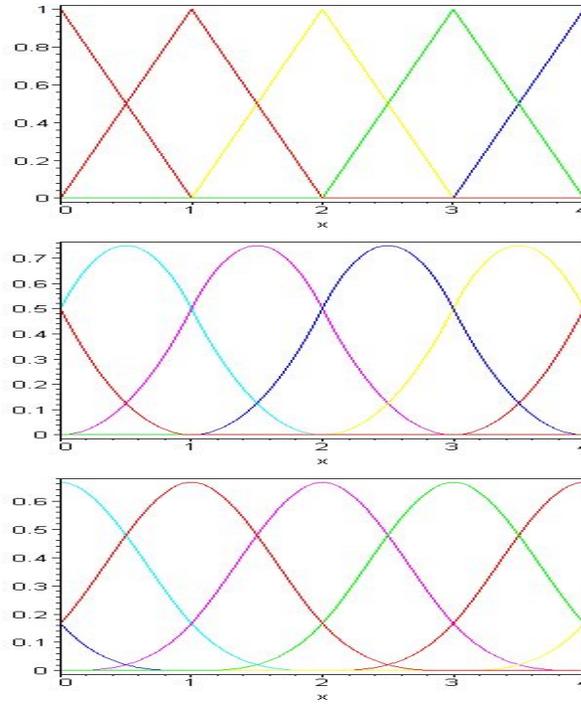
Die folgenden Eigenschaften von B-Splines werden in dieser Arbeit benötigt und sollen hier ohne Beweis zusammengestellt werden. Alle Beweise findet man in [Bo].

- (i) $\text{supp } N_{i,k} \subseteq [\theta_i, \theta_{i+k}]$ (lokaler Träger)
- (ii) $N_{i,k}(x) \geq 0$ für alle $x \in I$ (nicht negativ)
- (iii) $\sum_i N_{i,k}(x) = 1$ für alle $x \in I$ (Zerlegung der Eins)
- (iv) $N_{i,k} \in C^{k-2}(I)$ (Differenzierbarkeit)

Auf der Basis der B-Splines lassen sich nun im Intervall $I = [a, b]$ Splinefunktionen in der Form

$$(A.3) \quad S(x) = \sum_{i=1}^n c_i N_{i,k}(x)$$

darstellen.

Abbildung A.1: Basis-B-Splines der Ordnungen $k = 2, 3, 4$.

Satz A.1.2 Die B-Splines $\{N_{i,k}\}_{i=1,\dots,n}$ zur erweiterten Knotenfolge T bilden eine Basis für den Raum $\mathcal{N}_{k,T}$ der Splines der Ordnung k . Die B-Spline-Basis ist lokal unabhängig, d.h. gilt

$$\sum_{i=1}^n c_i N_{i,k}(x) = 0 \text{ für } x \in (c, d) \subseteq I$$

dann folgt

$$c_i = 0, \text{ falls } (c, d) \cap (\theta_i, \theta_{i+k}) \neq \emptyset.$$

Die B-Spline-Basis ist unconditionell stabil, d.h. es gilt für ein $c > 0$

$$c \max_{i=1,\dots,n} |c_i| \leq \left\| \sum_{i=1}^n c_i N_{i,k}(x) \right\|_{\infty} \leq \max_{i=1,\dots,n} |c_i| \quad \text{für alle } x \in I.$$

Die Ableitung von B-Splines lässt sich rekursiv als

$$(A.4) \quad N'_{i,k}(x) = (k-1) \left(\frac{N_{i,k-1}(x)}{\theta_{i+k-1} - \theta_i} - \frac{N_{i+1,k-1}(x)}{\theta_{i+k} - \theta_{i+1}} \right)$$

darstellen und für die ℓ -te Ableitung von Splinefunktionen S mit der Darstellung aus (A.3) gilt die Rekursionsformel

$$(A.5) \quad S^{(\ell)}(x) = (k-1) \cdot \dots \cdot (k-\ell) \sum_{i=\ell+1}^n c_i^{(\ell)} N_{i,k-\ell}(x)$$

mit

$$c_i^{(\ell)} := \begin{cases} c_i & \text{für } \ell = 0 \\ \frac{c_i^{(\ell-1)} - c_{i-1}^{(\ell-1)}}{\theta_{i+k-\ell} - \theta_i} & \text{für } \ell > 0. \end{cases}$$

Eine stabile und effiziente Auswertung von B-Splines kann mit dem Neville-artige Schema

$$(A.6) \quad S(x) = \sum_{i=\ell+1}^n c_i^{[\ell]}(x) N_{i, k-\ell}(x)$$

mit

$$c_i^{[\ell]}(x) := \begin{cases} c_i & \text{für } \ell = 0 \\ \frac{x - \theta_i}{\theta_{i+k-\ell} - \theta_i} c_i^{[\ell-1]}(x) + \frac{\theta_{i+k-\ell} - x}{\theta_{i+k-\ell} - \theta_i} c_{i-1}^{[\ell-1]}(x) & \text{für } \ell > 0 \\ 0 & \text{für } \theta_{i+k-\ell} = \theta_i \end{cases}$$

erfolgen. B-Splines besitzen gute Approximationseigenschaften an stetige Funktionen. Für $f \in C^k(I)$ gilt mit $h := \max_i |\theta_{i+1} - \theta_i|$ die Abschätzung

$$(A.7) \quad \inf_{S \in \mathcal{N}_{k, T_h}} \|f - S\|_\infty \leq c h^k \|f^{(k)}\|_\infty.$$

Ist $T := \{\theta_i\}_{i=1, \dots, n+k}$ eine erweiterte Knotenfolge wie in (A.1) mit zusätzlich äquidistantem Knotenabstand H , dann gilt

$$(A.8) \quad N_{i, k, T}(x) = N_{0, k, T}(x - iH) =: N_{k, T}(x - iH).$$

Ist weiter Δ eine verfeinerte Knotenfolge mit äquidistantem Knotenabstand $h = 1/p$, $p \in \mathbb{N}$, zur Knotenfolge T mit äquidistantem Knotenabstand $H = 1$, dann gilt die Skalierungsbeziehung

$$(A.9) \quad N_{k, \Delta}(x) = N_{k, T}\left(\frac{x}{h} - i\right).$$

Umgekehrt lässt sich jeder Grobgitterspline $N_{k, T}$ als Linearkombination von $(k+1)$ Feingittersplines $N_{k, \Delta}$ darstellen. Speziell für $h = 1/2$ erhält man die Darstellung

$$(A.10) \quad N_{k, T}(x) = \sum_{i=0}^k a_i N_{k, \Delta}\left(x - \frac{i}{2}\right)$$

mit den Verfeinerungskoeffizienten $a_i := 2^{1-k} \binom{k}{i}$.

Im Fall $k = 2$ wird also die Hutfunktion $N_{2, T}$ durch drei mit den Koeffizienten $a_0 = \frac{1}{2}$, $a_1 = 1$ und $a_2 = \frac{1}{2}$ gewichteten Hutfunktionen $N_{2, \Delta}$ wie in Abbildung A.2 dargestellt.

Abbildung A.2: Darstellung der Hutfunktion $N_{2, T}$ auf \mathbb{Z} durch drei Hutfunktionen auf $\mathbb{Z}/2$.

Ähnliche Beziehungen zwischen Grob- und Feingitter B-Splines gelten auch für nicht äquidistante Knotenfolgen mit $T \subset \Delta$.

Für die Verfeinerungskoeffizienten a_i gilt

$$\sum_{\substack{i=0 \\ i \text{ ungerade}}}^k a_i = \sum_{\substack{i=0 \\ i \text{ gerade}}}^k a_i = 1$$

und damit

$$(A.11) \quad \sum_{i=0}^k a_i = 2^{1-k} \sum_{i=0}^k \binom{i}{k} = 2.$$

In Abschnitt 4.3 wird der Wert benötigt, der sich ergibt, wenn ein B-Splines $N_{i, k, \Delta}$ an genau jedem zweiten Knotenpunkt $\theta \in \Delta$ ausgewertet wird und diese Werte aufaddiert werden.

Lemma A.1.3 *Es sei $N_{i,k,\Delta}$ der i -te Basis-B-Spline zur Knotenfolge Δ mit Gitterweite h und Träger $[\theta_i, \theta_{i+k}]$. Ist T eine gröbere Knotenfolge mit Gitterweite $H = 2h$ und $T \subset \Delta$, dann gilt*

$$\begin{aligned} (i) \quad & \sum_{\theta \in \Delta} N_{i,k,\Delta}(\theta) = 1, \quad \text{für } k \in \mathbb{N}, \\ (ii) \quad & \sum_{\theta \in T} N_{i,k,\Delta}(\theta) = \frac{1}{2}, \quad \text{für ungerades } k, \\ (iii) \quad & \sum_{\theta \in T} N_{i,k,\Delta}(\theta) = \begin{cases} \beta_k, & \text{falls } \theta_{i+k/2} \in T \\ \gamma_k, & \text{falls } \theta_{i+k/2} \notin T \end{cases} \quad \text{für gerades } k. \end{aligned}$$

Für $k = 2, 4, 6, 8$ ergeben sich für β_k bzw. γ_k die Werte

k	2	4	6	8
β_k	1	$\frac{2}{3}$	$\frac{17}{30}$	$\frac{166}{315}$
γ_k	0	$\frac{1}{3}$	$\frac{13}{30}$	$\frac{149}{315}$

Beweis: Da sich die Aussage von Lemma A.1.3 nicht in der zitierten Literatur über B-Splines findet, wird der Beweis hier durchgeführt.

Die Behauptung (i) ist für $k = 1$ offensichtlich richtig. Für $k \geq 1$ folgt sie induktiv wegen

$$\begin{aligned} \sum_{\theta \in \Delta} N_{i,k+1,\Delta}(\theta) &= \sum_{j=1}^k N_{i,k+1,\Delta}(\theta_{i+j}) \\ &= \sum_{j=1}^k \frac{j}{k} \frac{h}{h} N_{i,k,\Delta}(\theta_{i+j}) + \frac{(k+1-j)h}{kh} N_{i+1,k,\Delta}(\theta_{i+j}) \\ &= \sum_{j=1}^{k-1} \frac{j}{k} N_{i,k,\Delta}(\theta_{i+j}) + \sum_{j=2}^k \frac{k-(j-1)}{k} N_{i,k,\Delta}(\theta_{i+j-1}) \\ &= \sum_{j=1}^{k-1} \frac{j+k-j}{k} N_{i,k,\Delta}(\theta_{i+j}) \\ &= \sum_{\theta \in \Delta} N_{i,k,\Delta}(\theta). \end{aligned}$$

Dabei ist $\text{supp } N_{i,k,\Delta} \subseteq [\theta_i, \theta_{i+k}]$ und die Rekursion aus (A.2) verwendet worden. Falls nur über $\theta \in T$ summiert wird und falls k ungerade ist, ergibt sich aus Symmetriegründen direkt (ii). Für gerades k sind die beiden Fälle $\theta_{i+k/2} \in T$ und $\theta_{i+k/2} \notin T$ zu unterscheiden. Durch Auswerten des B-Splines und Aufaddieren ergeben sich je nach Fall die Werte β_k bzw. γ_k , die für $k = 2, 4, 6, 8$ im Lemma angegeben sind. \square

A.2 Sobolevräume und Bilinearformen

Finite-Elemente-Verfahren basieren auf der Variationsformulierung elliptischer Differentialgleichungen. Unter bestimmten Voraussetzungen an die Bilinearform lassen sich *Variationsgleichungen* äquivalent als unrestringierte Minimierungsprobleme und *Variationsungleichungen* als Minimierungsprobleme über konvexe Menge formulieren. Die Minimierungsprobleme besitzen nach Satz A.3.3 von Lax-Milgram Lösungen in Funktionenräumen, die als *Sobolevräume* bezeichnet werden.

Sobolevräume

Im folgenden sei Ω eine offene Teilmenge des \mathbb{R}^d . Es bezeichne $C_0^\infty(\Omega)$ den Raum der beliebig oft differenzierbaren Funktionen $v : \Omega \rightarrow \mathbb{R}$ mit kompaktem Träger in Ω und $L^2(\Omega)$ den Raum der über Ω quadrat-integrierbaren Funktionen $v : \Omega \rightarrow \mathbb{R}$. Mit dem Skalarprodukt

$$\langle u, v \rangle := \int_{\Omega} u(x)v(x)dx$$

ist $L^2(\Omega)$ ein Hilbertraum. Die induzierte L^2 -Norm wird mit $\|\cdot\|_0$ bezeichnet.

Definition A.2.1 (Schwache Ableitung) Eine Funktion $u \in L^2(\Omega)$ besitzt in $L^2(\Omega)$ die schwache Ableitung $\partial^\alpha u$, falls $\partial^\alpha u \in L^2(\Omega)$ und

$$\langle v, \partial^\alpha u \rangle = (-1)^{|\alpha|} \langle \partial^\alpha v, u \rangle \quad \text{für alle } v \in C_0^\infty(\Omega).$$

$\alpha := (\alpha_1, \dots, \alpha_d)$ ist hierbei ein Multiindex.

Definition A.2.2 (Sobolevräume) Für $m \in \mathbb{N}$ bezeichnet $H^m(\Omega)$ die Menge aller Funktionen $u \in L^2(\Omega)$, die schwache Ableitungen $\partial^\alpha u$ für alle $|\alpha| \leq m$ besitzen. Die Räume $H^m(\Omega)$ werden als Sobolevräume bezeichnet.

Mit dem Skalarprodukt

$$\langle u, v \rangle_m := \sum_{|\alpha| \leq m} \langle \partial^\alpha u, \partial^\alpha v \rangle$$

ist der Sobolevraum $H^m(\Omega)$ ein Hilbertraum.

Das Skalarprodukt induziert die *Sobolevnorm*

$$\|v\|_{m,\Omega} = \|v\|_m := \sqrt{\langle v, v \rangle_m} = \sqrt{\sum_{|\alpha| \leq m} \|\partial^\alpha v\|_0^2}.$$

Daneben betrachtet man die *Seminormen*

$$|v|_{m,\Omega} = |v|_m := \sqrt{\sum_{|\alpha|=m} \|\partial^\alpha v\|_0^2}.$$

Ist v ein stückweises Polynom auf einer Zerlegung $\{T_j\}_j$ von Ω , dann sind die höheren Sobolevnormen nicht erklärt. Genauer gilt der folgende Satz.

Satz A.2.3 Es sei $k \geq 1$ und Ω beschränkt. Eine stückweise beliebig oft differenzierbare Funktion $v : \bar{\Omega} \rightarrow \mathbb{R}$ gehört genau dann zu $H^k(\Omega)$, wenn $v \in C^{k-1}(\bar{\Omega})$ gilt.

Beweis: Der Beweis erfolgt durch Anwendung der Greenschen Formel in jedem Bereich, in dem v beliebig oft differenzierbar ist (s. [Br], S. 59). \square

Als Ersatz wählt man *gitterabhängige Sobolevnormen*

$$\|v\|_{m,h} := \sqrt{\sum_j \|v\|_{m,T_j}^2}.$$

Für $v \in H^m(\Omega)$ gilt offensichtlich $\|v\|_{m,h} = \|v\|_{m,\Omega}$.

Einen weiteren Zusammenhang zwischen den Sobolevräumen $H^m(\Omega)$ und dem Räumen $C^k(\bar{\Omega})$ der k -mal stetig differenzierbaren Funktionen liefert der Einbettungssatz von Sobolev- in Hölderräume. Da er in dieser Arbeit nicht für allgemeine Hölderräume benötigt wird, wird er nur für die Räume $C^k(\bar{\Omega})$ formuliert.

Satz A.2.4 (Einbettung von Sobolev- in Hölderräume) *Es sei $\Omega \subset \mathbb{R}^d$ offen und beschränkt mit Lipschitz-Rand, $m \geq 1$, sowie $k \geq 0$. Dann gilt $H^m(\Omega) \subset C^k(\bar{\Omega})$, falls*

$$m > k + \frac{d}{2}.$$

Beweis: Der Beweis beruht auf dem Einbettungssatz für Hölderräume (s. [Alt], S.307), dem Einbettungssatz für Sobolevräume (s. [Alt], S.314) und dem Satz von Morrey (s. [Alt], S.317). In dieser Form ist er in [Alt], S.319 durchgeführt. \square

Für beliebiges $\epsilon > 0$ gelten also die Einbettungen

$$\begin{aligned} H^{1+\epsilon}(\Omega) &\subset C^0(\bar{\Omega}), \quad H^{2+\epsilon}(\Omega) \subset C^1(\bar{\Omega}) \quad \text{und} \quad H^{3+\epsilon}(\Omega) \subset C^2(\bar{\Omega}), \quad \text{falls } d = 2, \\ H^{0.5+\epsilon}(\Omega) &\subset C^0(\bar{\Omega}), \quad H^{1.5+\epsilon}(\Omega) \subset C^1(\bar{\Omega}) \quad \text{und} \quad H^{2.5+\epsilon}(\Omega) \subset C^2(\bar{\Omega}), \quad \text{falls } d = 1. \end{aligned}$$

Es bezeichne weiter $H_0^m(\Omega)$ den Raum aller Funktionen $v \in H^m(\Omega)$ mit kompaktem Träger in Ω . Man kann zeigen, dass $H_0^m(\Omega)$ die Vervollständigung von $C_0^\infty(\Omega)$ bezüglich der Sobolevnorm $\|\cdot\|_m$ ist. Der Dualraum von $H_0^m(\Omega)$ sei mit $H^{-m}(\Omega)$ bezeichnet.

Bilinearformen

Es sei H ein Hilbertraum mit der Norm $\|\cdot\|$. Eine Abbildung $a : H \times H \rightarrow \mathbb{R}$ heißt *Bilinearform*, wenn für alle $u, v \in H$ die Abbildungen $v \mapsto a(u, v)$ und $u \mapsto a(u, v)$ linear sind.

Definition A.2.5 *Eine Bilinearform $a : H \times H \rightarrow \mathbb{R}$ heißt*

- symmetrisch, falls $a(u, v) = a(v, u)$ für alle $u, v \in H$,
- positiv definit in H , falls $a(u, u) > 0$ für alle $u \in H, u \neq 0$,
- stetig, wenn es ein $C > 0$ gibt, so dass

$$|a(u, v)| \leq C \|u\| \|v\| \quad \text{für alle } u, v \in H,$$

- H -elliptisch oder koerziv, wenn es ein $\alpha > 0$ gibt, so dass

$$a(u, u) \geq \alpha \|u\|^2 \quad \text{für alle } u \in H.$$

Lemma A.2.6 *Es sei Ω ein Gebiet im \mathbb{R}^d und $c_1, c_2 \in \mathbb{R}^+$. Dann ist die Bilinearform*

$$a(u, v) := \int_{\Omega} \left(c_1 u v + c_2 \frac{\partial u}{\partial x} \frac{\partial v}{\partial x} \right) dx$$

stetig, symmetrisch und positiv definit in $H^1(\Omega)$. Ist $c_2 > 0$, dann ist $a(\cdot, \cdot)$ außerdem $H_0^1(\Omega)$ -elliptisch.

Beweis: Die Symmetrie von $a(\cdot, \cdot)$ ist offensichtlich. Die Stetigkeit folgt wegen

$$\begin{aligned} |a(u, v)| &\leq \int_{\Omega} c_1 |u v| + c_2 \left| \frac{\partial u}{\partial x} \frac{\partial v}{\partial x} \right| dx \\ &\leq c_1 \|u\|_0 \|v\|_0 + c_2 \|u\|_1 \|v\|_1 \\ &\leq \max\{c_1, c_2\} \|u\|_1 \|v\|_1 \end{aligned}$$

direkt aus der Cauchy–Schwarz Ungleichung. Wegen $c_1, c_2 \geq 0$ ist außerdem

$$a(v, v) = \int_{\Omega} (c_1 |v|^2 + c_2 |\partial v / \partial x|^2) dx \geq c_2 |v|_1^2 \geq 0$$

für alle $v \in H^1(\Omega)$. Speziell für $v \in H_0^1(\Omega)$ folgt insbesondere mit der Poincaré–Friedrichs–Ungleichung mit der Konstanten $\gamma_{\Omega} > 0$ (s. [Br], S. 29) wegen

$$a(v, v) \geq c_2 |v|_1^2 \geq c_2 \gamma_{\Omega} \|v\|^2$$

die $H_0^1(\Omega)$ –Elliptizität der Bilinearform. Der Beweis ist [Hb], S. 679 entnommen. \square

A.3 Variationsungleichungen und Minimierungsprobleme

In diesem Abschnitt sind einige Grundlagen über Variationsungleichungen zusammengestellt. Während Variationsgleichungen äquivalent zu unrestringierten Minimierungsproblemen sind, lassen sich Variationsungleichungen als Minimierungsprobleme über einer konvexen Menge auffassen.

In Regularitätsfragen unterscheiden sie sich dadurch, dass die Lösung von Variationsgleichungen im Allgemeinen um so glatter ist, je glatter die Daten inklusive Gebiet sind, während bei Variationsungleichungen im Allgemeinen eine bestimmte Glattheitsschranke der Lösung auch bei beliebig glatten Daten nicht überschritten wird.

Für mehr Informationen über Variationsungleichungen sei auf die Bücher [EO], [KS] oder [GLT] verwiesen.

Formulierung als Minimierungsproblem

Es gilt folgender wohlbekannter Zusammenhang zwischen Variationsgleichungen und Minimierungsproblemen.

Satz A.3.1 *Es sei $a : V \times V \rightarrow \mathbb{R}$ eine symmetrische und auf V positiv definite Bilinearform, $f : V \rightarrow \mathbb{R}$ ein lineares Funktional und V ein linearer Raum. Dann ist die Variationsgleichung*

$$\text{finde } u \in V : a(u, v) = f(v) \text{ für alle } v \in V \quad (*)$$

äquivalent zum Minimierungsproblem

$$\text{finde } u \in V : \mathcal{J}(u) \leq \mathcal{J}(v) \text{ für alle } v \in V$$

mit dem quadratischen Funktional $\mathcal{J}(v) := \frac{1}{2}a(v, v) - f(v)$.

Beweis: Für $t \in \mathbb{R}$ rechnet man mit der Definition von \mathcal{J} leicht die Identität

$$(A.12) \quad \mathcal{J}(u + tv) = \mathcal{J}(u) + t[a(u, v) - f(v)] + \frac{1}{2}t^2 a(v, v)$$

nach. Ist $u \in V$ Lösung der Variationsgleichung folgt aus (A.12) mit $t = 1$ und Einsetzen von (*), dass

$$\mathcal{J}(u + v) = \mathcal{J}(u) + \frac{1}{2}a(v, v) > \mathcal{J}(u) \quad \text{für alle } v \in V, v \neq 0$$

gilt. Ist umgekehrt $u \in V$ Lösung des Minimierungsproblems, dann muss die Ableitung der Funktion $t \mapsto \mathcal{J}(u + tv)$ bei $t = 0$ verschwinden. Nach (A.12) beträgt die Ableitung $a(u, v) - f(v)$, also folgt (*). Der Beweis ist [Br], S.34 entnommen. \square

Für Variationsungleichungen lässt sich die folgende analoge Aussage treffen.

Satz A.3.2 *Es sei $a : V \times V \rightarrow \mathbb{R}$ eine symmetrische und auf V positiv definite Bilinearform, $f : V \rightarrow \mathbb{R}$ ein lineares Funktional und \mathcal{K} eine konvexe Teilmenge eines linearen Raumes V . Dann ist die Variationsungleichung*

$$\text{finde } u \in \mathcal{K} : a(u, v - u) \geq f(v - u) \text{ für alle } v \in \mathcal{K} \quad (**)$$

äquivalent zum Minimierungsproblem

$$\text{finde } u \in \mathcal{K} : \mathcal{J}(u) \leq \mathcal{J}(v) \text{ für alle } v \in \mathcal{K}$$

mit dem quadratischen Funktional $\mathcal{J}(v) := \frac{1}{2}a(v, v) - f(v)$.

Beweis: Angenommen, u ist Lösung der Variationsungleichung (**). Dann rechnet man leicht

$$\mathcal{J}(u) - \mathcal{J}(v) = f(v - u) - a(u, v - u) - \frac{1}{2}a(u - v, u - v) \leq 0 \text{ für alle } v \in \mathcal{K}$$

nach. Die Funktion u ist also auch Lösung des Minimierungsproblems.

Sei nun umgekehrt $u \in \mathcal{K}$ Lösung des Minimierungsproblems. Da \mathcal{K} konvex ist, gilt $(1 - \lambda)u + \lambda v \in \mathcal{K}$ für jedes $v \in \mathcal{K}$ und $\lambda \in (0, 1]$. Aus $\mathcal{J}(u) \leq \mathcal{J}((1 - \lambda)u + \lambda v)$ erhält man nach Einsetzen der Definition von \mathcal{J} , dass

$$0 \leq \lambda[a(u, v - u) - f(v - u)] + \frac{1}{2}\lambda^2 a(v - u, v - u)$$

für alle $v \in \mathcal{K}$ und $\lambda \in (0, 1]$ gilt. Nach Dividieren durch λ folgt mit $\lambda \rightarrow 0$ die Variationsungleichung

$$a(u, v - u) \geq f(v - u) \text{ für alle } v \in \mathcal{K}.$$

Der Beweis ist [EO], S.101 entnommen. □

Existenz und Eindeutigkeit der Lösung

In der Formulierung als Minimierungsproblem folgt die Existenz und Eindeutigkeit einer Lösung für Variationsgleichungen und –ungleichungen in Hilberträumen mit dem Satz von Lax–Milgram.

Satz A.3.3 (Lax–Milgram) *Es sei \mathcal{K} eine abgeschlossene, konvexe Menge in einem Hilbertraum H und $a : H \times H \rightarrow \mathbb{R}$ eine H – elliptische und stetige Bilinearform. Dann besitzt das Minimierungsproblem*

$$\text{finde } u \in \mathcal{K} : \mathcal{J}(u) \leq \mathcal{J}(v) \text{ für alle } v \in \mathcal{K}$$

mit dem quadratischen Funktional $\mathcal{J}(v) := \frac{1}{2}a(v, v) - f(v)$ für jedes stetige Funktional $f : H \rightarrow \mathbb{R}$ eine eindeutige Lösung $u \in \mathcal{K}$.

Beweis: \mathcal{J} ist nach unten beschränkt, da aus der Elliptizität von a , der Stetigkeit von f und nach Ergänzen zur Binomischen Formel

$$\begin{aligned} \mathcal{J}(v) &\geq \frac{1}{2}\alpha\|v\|^2 - \|f\|\|v\| \\ &= \frac{1}{2\alpha}(\alpha\|v\| - \|f\|)^2 - \frac{\|f\|^2}{2\alpha} \\ &\geq -\frac{\|f\|^2}{2\alpha} \end{aligned}$$

folgt. Sei $\{v_n\}_{n \in \mathbb{N}}$ eine Minimalfolge in \mathcal{K} , so dass

$$(A.13) \quad \lim_{n \rightarrow \infty} \mathcal{J}(v_n) \rightarrow \inf\{\mathcal{J}(v) : v \in \mathcal{K}\} =: c.$$

Mit der Elliptizität der Bilinearform $a(\cdot, \cdot)$ rechnet man leicht

$$\alpha\|v_n - v_m\|^2 \leq 4\mathcal{J}(v_n) + 4\mathcal{J}(v_m) - 8\mathcal{J}\left(\frac{v_n + v_m}{2}\right)$$

nach. Da \mathcal{K} konvex ist, gilt $\mathcal{J}\left(\frac{v_n + v_m}{2}\right) \geq c$ und damit folgt wegen (A.13), dass $\|v_n - v_m\| \rightarrow 0$ für $n, m \rightarrow \infty$. Also ist v_n Cauchy–Folge in H und es existiert $u := \lim_{n \rightarrow \infty} v_n$. Da \mathcal{K} abgeschlossen ist, gilt $u \in \mathcal{K}$. Die Stetigkeit von \mathcal{J} impliziert schließlich

$$\mathcal{J}(u) = \lim_{n \rightarrow \infty} \mathcal{J}(v_n) = \inf\{\mathcal{J}(v) : v \in \mathcal{K}\},$$

also die Existenz einer Lösung des Minimierungsproblems. Da der Grenzwert einer Cauchyfolge eindeutig bestimmt ist, ist die Lösung auch eindeutig. Der Beweis ist [Br], S. 37 entnommen. \square

A.4 Hindernisprobleme und lineare komplementäre Probleme

Eine spezielle Klasse freier Randwertprobleme bilden die sogenannten *Hindernisprobleme*. Hier darf die Lösung eine gegebene Hindernisfunktion nicht über– bzw. unterschreiten. Der a priori unbekannte freie Rand des Problems, wird dann durch die Punkte gebildet, an denen die Lösung die Hindernisfunktion tangential berührt.

Im folgenden werden Hindernisprobleme und lineare komplementäre Probleme definiert und es wird gezeigt, dass sich Hindernisprobleme in linear komplementäre Probleme umformulieren lassen. Betrachte zur Veranschaulichung hierbei auch das einfache Beispiel 2.2.1 zu Beginn von Kapitel 2.

Definition A.4.1 (Hindernisproblem) Ein Hindernisproblem bezeichnet eine Variationsungleichung

$$(A.14) \quad \text{finde } u \in \mathcal{K} : a(u, v - u) \geq f(v - u) \quad \text{für alle } v \in \mathcal{K}$$

mit einer konvexen Menge \mathcal{K} der speziellen Form

$$(A.15) \quad \mathcal{K} := \{v \in H_0^1(\Omega) : v(x) \geq g(x) \quad \text{für alle } x \in \Omega\}$$

gilt.

Die Funktion g beschreibt also eine (untere) Hindernisfunktion für die Lösung. Analog sind auch obere Hindernisse oder beidseitige Hindernisse möglich. Damit das Hindernisproblem A.4.1 wohldefiniert ist, muss $g(x) < 0$ wegen $\mathcal{K} \subset H_0^1(\Omega)$ für alle $x \in \Gamma$ gelten, wobei Γ den Rand des Gebietes Ω bezeichnet.

Im folgenden wird gezeigt, dass sich Hindernisprobleme als lineare komplementäre Probleme formulieren lassen. Dazu sei $u \in H_0^1(\Omega)$ die Lösung eines Hindernisproblems mit einer stetigen und elliptischen Bilinearform $a(\cdot, \cdot)$ und mit $f \in L^2(\Omega)$. Es sei $u \in H^2(\Omega)$ angenommen. Weiter sei $A : H_0^1(\Omega) \rightarrow H^{-1}$ der Riesz-Operator, der $\langle Au, v \rangle = a(u, v)$ für alle $v \in H$ erfüllt.

Definition A.4.2 (Kontaktmenge) Die Menge $K^\bullet(u) := \{x \in \Omega : u(x) = g(x)\} \subset \Omega$ wird als Kontaktmenge der Funktion u bezeichnet.

Lemma A.4.3 Erfüllt $u \in H_0^1(\Omega)$ das Hindernisproblem A.4.1, dann gilt $u \geq g$ und $Au = f$ fast überall in der Menge $K^\circ(u) := \Omega \setminus K^\bullet(u)$.

Beweis: Die erste Behauptung folgt sofort wegen $u \in \mathcal{K}$. Ist $x \in K^\circ(u)$, dann existiert eine Kugel $B_r(x)$ mit Radius r und Mittelpunkt x und eine Funktion $\eta \in C_0^\infty(B_r(x))$ mit $\eta(x) \geq 0$ für $x \in B_{r/2}(x)$, so dass

$$u \geq g + \eta \quad \text{für alle } x \in B_r(x).$$

Für alle $\xi \in C_0^\infty(B_{r/2}(x))$ gibt es daher ein $\epsilon > 0$, so dass

$$u + \epsilon\xi \geq g + \frac{1}{2}\eta \quad \text{für alle } x \in B_{r/2}(x).$$

Also ist $v = u + \epsilon\xi \in \mathcal{K}$. Einsetzen in die Variationsungleichung aus (A.14) liefert

$$a(u, \xi) \geq f(\xi) \quad \text{für alle } x \in B_{r/2}(x).$$

Da dies analog auch für $-\xi$ gilt, folgt mit der Testfunktion $v = u - \epsilon\xi \in \mathcal{K}$, dass umgekehrt

$$a(u, \xi) \leq f(\xi) \quad \text{für alle } x \in B_{r/2}(x)$$

gilt und damit $Au = f$ in $K^\circ(u)$. Der Beweis ist [KS] entnommen. \square

Lemma A.4.4 Erfüllt $u \in H_0^1(\Omega)$ das Hindernisproblem A.4.1, dann gilt $u = g$ und $Au \geq f$ fast überall in der Menge $K^\bullet(u)$.

Beweis: Nach Definition der Kontaktmenge gilt $u = g$ fast überall in $K^\bullet(u)$. Es sei $\varphi \in C_0^\infty(K^\bullet(u))$ eine nicht negative Testfunktion mit $u + \varphi \geq g$. Einsetzen in die Variationsungleichung aus (A.14) liefert

$$\langle Au - f, \varphi \rangle \geq 0 \quad \text{für alle } \varphi \geq 0, \varphi \in C_0^\infty(K^\bullet(u)).$$

Damit folgt $Au \geq f$ fast überall in der Kontaktmenge $K^\bullet(u)$. Der Beweis ist [KS] entnommen. \square

Definition A.4.5 (Lineares komplementäres Problem) *Es seien $\mathcal{A}, \mathcal{B} : H_0^1(\Omega) \rightarrow H_0^1(\Omega)$ zwei lineare Operatoren. Ein lineares komplementäres Problem bezeichnet ein Problem der Form*

$$\text{finde } u \in H_0^1(\Omega) : \mathcal{A}u \cdot \mathcal{B}u = 0, \mathcal{A}u \geq 0, \mathcal{B}u \geq 0.$$

Mit Lemma A.4.3 und Lemma A.4.4 lässt sich das Hindernisproblem A.4.1 als ein lineares komplementäres Problem formulieren.

Lemma A.4.6 *Ist $u \in H^2(\Omega)$ die Lösung des Hindernisproblems A.4.1, dann erfüllt u das lineare komplementäre Problem*

$$(Au - f) \cdot (u - g) = 0,$$

$$Au - f \geq 0, \quad u - g \geq 0$$

fast überall in Ω .

Beweis: Die Behauptung folgt sofort aus Lemma A.4.3 und Lemma A.4.4. \square

Die Umkehrung von Lemma A.4.6 ist auch richtig. Dies wurde bereits in Abschnitt 2.2.3 gezeigt. Dort erfolgt die Herleitung der variationellen Formulierung A.4.1 aus der linear komplementären Formulierung des Problems über partielle Integration.

Notation

Zur besseren Unterscheidung sind Vektoren generell fettgedruckt dargestellt. Bezeichnet $u : \Omega \rightarrow \mathbb{R}$ eine Funktion, dann wird der Koeffizientenvektor von u bei Entwicklung in einer Basis der Länge n also fettgedruckt mit $\mathbf{u} := (u_1, \dots, u_n)^T$ bezeichnet.

Ungleichungen $\mathbf{a} \leq \mathbf{b}$ von Vektoren $\mathbf{a}, \mathbf{b} \in \mathbb{R}^n$ sind generell komponentenweise als $a_i \leq b_i$ für alle $i = 1, \dots, n$ aufzufassen.

Allgemeine Bezeichnungen

$\Omega \subset \mathbb{R}^d$	polygonales Gebiet im \mathbb{R}^d
$L^2(\Omega)$	Raum der über Ω quadrat-integrierbaren Funktionen
$\langle \cdot, \cdot \rangle$	L^2 -Skalarprodukt
$\ \cdot \ _0$	induzierte L^2 -Norm
ℓ^2	Raum der quadrat-integrierbaren Folgen mit der Norm $\ \cdot \ _{\ell^2}$
$H^m(\Omega)$	Sobolevraum von L^2 -Funktionen mit quadrat-integrierbaren schwachen Ableitungen bis zur Ordnung m
$H_0^m(\Omega)$	Unterraum von $H^m(\Omega)$ mit verallgemeinerten Nullrandbedingungen
$\ \cdot \ _m$	Sobolevnorm der Ordnung m
$ \cdot _m$	Sobolev-Seminorm der Ordnung m
$C^k(\Omega)$	Menge der Funktionen mit stetigen Ableitungen der Ordnung k
$C_0^k(\Omega)$	Unterraum von $C^k(\Omega)$ mit kompaktem Träger
$\ \cdot \ _\infty$	Supremumsnorm
$S_h, S_J \subset H_0^1(\Omega)$	endlich-dimensionale Räume stückweiser Polynome mit Gitterweite $h = 2^{-J}$
$\Sigma_J \subset S_J$	Basis von S_J mit Kardinalität n_J
$I_h : C(\bar{\Omega}) \rightarrow S_h$	Interpolationsoperator
\mathcal{N}_{k, T_h}	Raum der Splines der Ordnung k bzgl. der erweiterten Knotenfolge T_h
$N_{i,k} \in \mathcal{N}_{k, T_h}$	i -ter Basis-B-Spline der Ordnung k
\mathcal{L}, \mathcal{W}	lineare Differentialoperatoren
\mathcal{N}	nichtlinearer Differentialoperator
$\text{supp } f(x)$	Träger der Funktion f
$a \wedge b$	$\min\{a, b\}$
a^+	$\max\{a, 0\}$
$[x]$	kleinste ganze Zahl größer oder gleich $x \in \mathbb{R}$
$\lfloor x \rfloor$	größte ganze Zahl kleiner oder gleich $x \in \mathbb{R}$
$x \lesssim y$	$x \leq c y$ mit einer gitterweiten-unabhängigen Konstanten $c \in \mathbb{R}^+$

Bezeichnungen in Zusammenhang mit Optionen

$T \in \mathbb{R}^+$	Fälligkeitszeitpunkt der Option
$K \in \mathbb{R}^+$	Ausübungspreis der Option
$S(t) \in \mathbb{R}^+$	Preis des Basiswertes in Abhängigkeit vom Zeitpunkt $t \in \mathbb{R}^+$
$S_f(t) \in \mathbb{R}^+$	optimaler Ausübungszeitpunkt einer Amerikanischen Option
$\mathcal{H}(S) \in \mathbb{R}^+$	Auszahlungsfunktion der Option in Abhängigkeit vom Preis S des Basiswertes
$V(S, t) \in \mathbb{R}^+$	Wert der Option zum Zeitpunkt $t \in \mathbb{R}^+$ in Abhängigkeit vom Preis $S \in \mathbb{R}^+$ des Basiswertes
$r \in \mathbb{R}^+$	konstanter Jahreszinssatz für eine sichere Anlage
$d \in \mathbb{R}^+$	Höhe der stetigen Dividende des Wertpapiers
$\mu \in \mathbb{R}$	Drift des stochastischen Prozesses $S(t)$
$\sigma \in \mathbb{R}^+$	Volatilität des stochastischen Prozesses $S(t)$
$B(t) \in \mathbb{R}$	eindimensionale standardisierte Brownsche Bewegung

Bezeichnungen in Zusammenhang mit Variationsungleichungen

$\mathcal{J} : H^1(\Omega) \rightarrow \mathbb{R}$	quadratisches Variationsfunktional
$a : H^1(\Omega) \times H^1(\Omega) \rightarrow \mathbb{R}$	stetige, symmetrische und in H^1 positiv definite Bilinearform
$\ell, f : H^1(\Omega) \rightarrow \mathbb{R}$	lineare Funktionale
$\mathcal{K} \subset H^1(\Omega)$	konvexer Lösungsraum
$\mathcal{K}_h \subset S_h, \mathcal{K}_J \subset S_J$	diskrete konvexe Lösungsräume
$g \in H^1(\Omega)$	obere bzw. untere Hindernisfunktion mit Koeffizienten g_i
$K^\bullet(v) := \{x \in \Omega : v(x) = g(x)\}$	Kontaktmenge der Funktion $v \in H^1(\Omega)$
$K_J^\bullet(v) := \{i \in \{1, \dots, n_J\} : v_i = g_i\}$	diskrete Kontaktmenge für die Entwicklungskoeffizienten der Funktion $v \in S_J$
$K_J^\circ(v) := \{i \in \{1, \dots, n_J\} : v_i \neq g_i\}$	diskrete Menge der Koeffizienten ohne Kontakt

Bezeichnungen in Zusammenhang mit Finite-Elementen

$\nu \in \mathbb{N}$	(Zeit-) Iterationszähler
$\theta \in [0, 1]$	Parameter der Zeitdiskretisierung ($\theta = 0$: explizites Eulerverfahren, $\theta = 1/2$: Crank-Nicholson Verfahren, $\theta = 1$: implizites Eulerverfahren)
$\Delta\tau$	Zeitschrittweite
$N + 1 \in \mathbb{N}$	Anzahl der Zeitschritte der Diskretisierung
$M \in \mathbb{N}$	Anzahl der Ortsschritte der Diskretisierung
$y : \mathbb{R} \times \mathbb{R}^+ \rightarrow \mathbb{R}$	Lösung der transformierten Randwertprobleme mit homogenen Randdaten
$y^\nu \in H_0^1(\mathbb{R})$	semidiskrete Lösung im ν -ten Zeitschritt
$y_h^\nu \in S_h$	vollständig diskrete Lösung im ν -ten Zeitschritt
$A, B \in \mathbb{R}^{n \times n}$	Steifigkeits- bzw. Massenmatrix
$\mathbf{b} \in \mathbb{R}^n$	diskrete rechte Seite
$\mathbf{w} \in \mathbb{R}^n$	Koeffizientenvektor der diskreten Lösung y_h^ν

Bezeichnungen in Zusammenhang mit Relaxationsverfahren

$\nu \in \mathbb{N}$	Iterationszähler
$j \in \mathbb{N}$	Zwischeniterationszähler
$M^\nu \subset S_J$	Menge der Suchrichtungen des Relaxationsverfahrens im ν -ten Iterationsschritt
$m^\nu \in \mathbb{N}$	Kardinalität von M^ν
$M_c^\nu \subset M^\nu$	Menge der Grobgitterfunktionen in M^ν
$\mu^{\nu,j} \in M^\nu$	j -te Suchrichtung des Relaxationsverfahrens im ν -ten Iterationsschritt
$D^{\nu,j}, D^{*,\nu,j} \subset \text{span}\{\mu^{\nu,j}\}$	eindimensionale Lösungsräume lokaler Teilprobleme
$v^{\nu,j} \in D^{\nu,j}, v^{*,\nu,j} \in D^{*,\nu,j}$	approximative bzw. exakte Korrektur in Richtung von $\mu^{\nu,j}$
$\psi^{\nu,j} \in \text{span}\{\mu^{\nu,j}\}$	lokale approximierte Hindernisfunktionen
S_J°	reduzierter Teilraum von S_J
M°	reduzierte Menge an Suchrichtungen
$S_\ell^\nu \subset S_\ell$	an die Position der Kontaktmenge im ν -ten Iterationsschritt angepasste endlich-dimensionale Räume
$T_\ell : S_\ell \rightarrow S_\ell^\nu$	Abschneideoperator zum Anpassen der Suchrichtungen an die Position des freien Randes im ν -ten Iterationsschritt

Bezeichnungen in Zusammenhang mit Mehrgitterverfahren

$\ell = 0, \dots, J$	Mehrgitterlevel
J	feinstes Level
$S_\ell \subset S_J \subset H_0^1(\Omega)$	endlich-dimensionale Räume mit Dimension n_ℓ zum Gitter Δ_ℓ
Δ_ℓ	Gitter mit Gitterweite h_ℓ
$\Sigma_\ell \subset S_\ell$	Basis von S_ℓ der Länge n_ℓ
$u^{\nu,\ell,m}$	Näherung im ν -ten Zykel auf Ebene ℓ im m -ten Teilschritt
$\gamma \in \mathbb{N}$	$\gamma = 1$: V-Zyklus, $\gamma = 2$: W-Zyklus
η_1, η_2	Anzahl der Vor- bzw. Nachglättungsschritte
\mathcal{S}	Glättungsoperator
\mathcal{P}	Projektionsoperator
$r_{\ell+1}^\ell : S_{\ell+1} \rightarrow S_\ell$	Restriktionsoperator für den Defekt
$p_\ell^{\ell+1} : S_\ell \rightarrow S_{\ell+1}$	Prolongationsoperator für die Korrektur

Verwendete Abkürzungen

PSOR	Projektives Gauss-Seidel-Verfahren
MMG	Monotones Mehrgitterverfahren mit optimaler Approximation des Hindernisses ([Ko1])
TrMMG	Abgeschnittene Version des monotonen Mehrgitterverfahrens mit optimaler Approximation des Hindernisses ([Ko1])
MMG (einf)	Monotones Mehrgitterverfahren mit einfacher Approximation des Hindernisses ([Ma])
TrMMG (einf)	Abgeschnittene Version des monotonen Mehrgitterverfahrens mit einfacher Approximation des Hindernisses
MG	Mehrgitterverfahren für lineare partielle Differentialgleichungen
FAS	Mehrgitterverfahren für nichtlineare partielle Differentialgleichungen ([B])
PFAS	Projektive Variante des FAS Verfahrens für lineare komplementäre Probleme ([BC])
PDPFAS	Variante des PFAS Verfahrens mit schnellerer Konvergenz ([BC])
OGK	Optimierter Grobgitterhindernis-Konstruktionsalgorithmus aus Satz 4.4.7

Literaturverzeichnis

- [Alt] Alt, H.-W.: *Lineare Funktionalanalysis*. Springer, 3.Auflage, 1999.
- [B] Brandt, A.: *Multi-level adaptive technique (MLAT) for fast numerical solution to boundary value problems*. in [CT], 1973.
- [Ba] Bauer, H.: *Wahrscheinlichkeitstheorie*. deGruyter, 5.Auflage, 2001.
- [Bc] Baiocchi, C.: *Estimation d'error dans L^∞ pour les inequations a obstacle*. Lecture Notes in Math. 606, 27-34, 1977.
- [BC] Brandt, A.; Cryer, C.W.: *Multigrid algorithms for the solution of linear complementary problems arising from free boundary problems*. SIAM J. Sci. Stat. Comput. 4, 655-684, 1983.
- [BG] Broadi, M.; Glasserman, P.: *Pricing American-style securities using simulation*. J. Econom. Dynam. Control 21, 1323-1352, 1997.
- [BHR] Brezzi, F.; Hager, W.; Raviart P.A.: *Error estimates for the finite element solution of variational inequalities*. Numer. Math. 28, 431-443, 1977.
- [BjS] Bjerksund, P.; Stensland, G.: *Closed-form approximation of American options*. Scandinavian J. Management 9, 87-99, 1993.
- [Bk] Black, F.: *The pricing of commodity contracts*. J. Financial Economics 3, 167-179, 1976.
- [Bl] Barles, G.: *Convergence of numerical schemes for degenerate parabolic equations arising in finance theory*. In [RT], 1-21, 1997.
- [BMR] Barndorff-Nielsen, O.E.; Mikosch, T.; Resnick, S.: *Lévy Processes: Theorie and Applications*. Birkhäuser, 2001.
- [Bn] Breen, R.: *The accelerated binomial option pricing model*. J. Financial Quant. Anal. 26, 153-164, 1991.
- [Bo] de Boor, C.: *A Practical Guide to Splines*. Springer, 1978.
- [Br] Braess, D.: *Finite Elemente*. Springer, 3.Auflage, 2003.
- [BR] Ball, C.A.; Roma, A.: *Stochastic volatility option pricing*. J. Financial Quant. Anal. 29, 589-607, 1994.
- [BS] Black, F.; Scholes, M.: *The pricing of options and corporate liabilities*. J. Political Economy 81, 637-654, 1973.
- [BS1] Brennan, M.J; Schwartz, E.S.: *The valuation of American put options*. J. Finance 32, 449-462, 1978.
- [BS2] Brennan, M.J; Schwartz, E.S.: *Finite difference methods and jump processes arising in the pricing of contingent claims*. J. Financial Quant. Anal. 13, 462-474, 1978.
- [Bu] Buck, R.C.: *Applications of duality in approximation theory*. In Approximation of Functions (Proc. Sympos. General Motors Res. Lab.), 27-42, Elsevier Publ. Co., 1965.
- [Bv] Bachvalov, N.S.: *On the convergence of a relaxation method with natural constraints on the elliptic operator*. USSR Comp. Math. and Math. Phys. 6, 101-135, 1966.
- [BW] Barone-Adesi, G.; Whaley, R.: *Efficient analytic approximation of American option values*. J. Finance 42, 301-320, 1987.
- [By] Boyle, P.P.: *A lattice framework for option pricing with two state variables*. J. Financial Quant. Anal. 23, 1-12, 1988.

- [Ch] Chan, T.: *Pricing contingent claims on stocks driven by Lévy processes*. Annals of Appl. Prob. 9, 504-528, 1999.
- [CP] Clarke, N.; Parrott, K.: *Multigrid for American option pricing with stochastic volatility*. Appl. Math. Finance 6, 177-195, 1999.
- [Cr] Cryer, C.W.: *The solution of a quadratic programming problem using systematic overrelaxation*. SIAM J. Control 9, 385-392, 1971.
- [CRR] Cox, J.C.; Ross, S.A.; Rubinstein, M.: *Option pricing: a simplified approach*. J. Financial Economics 7, 229-263, 1979.
- [CT] Cabannes, H.; Temam, R.: *Proceedings of the third international conference on numerical methods in fluid mechanics*. Lecture Notes in Phys. 18, Springer, 1973.
- [DH] Dempster, M.; Hutton, J.P.: *Pricing American stock options by linear programming*. Math. Finance 9, 229-254, 1999.
- [DKU] Dahmen, W.; Kunoth, A.; Urban, K.: *Biorthogonal spline wavelets on the interval – stability and moment conditions*. Appl. and Comp. Harmon. Anal. 6, 132-196, 1999.
- [DM] Dahmen, W.; Micchelli, C.A.: *Banded matrices with banded inverses, II: Locally finite decomposition of spline spaces*. Constr. Appr. 9, 263-281, 1993.
- [Du] Duffie, D.: *Dynamic Asset Pricing Theory*. Princeton University Press, 1996.
- [DV] DeVore, R.: *One-sided approximation of functions*. J. Approximation Theory 1, 11-25, 1968.
- [Eb] Eberlein, E.: *Application of generalized hyperbolic Lévy motions to finance*. in [BMR], 319-337, 2001.
- [EO] Elliott, C.M.; Ockendon, J.K.: *Weak and Variational Methods for Moving Boundary Problems*. Pitman, 1982.
- [Ev] Evans, L.C.: *Partial Differential Equations*. Amer. Math. Soc., 1980.
- [F1] Fedorenko, R.P.: *A relaxation method for solving elliptic difference equations*. USSR Comp. Math. and Math. Phys. 1, 1092-1096, 1961.
- [F2] Fedorenko, R.P.: *The speed of convergence of one iterative process*. USSR Comp. Math. and Math. Phys. 4, 227-235, 1964.
- [FG] Figlewski, S.; Gao, B.: *The adaptive mesh model: a new approach to efficient option pricing*. J. Financial Economics 53, 313-351, 1999.
- [FKV] Fasshauer, G.E.; Khaliq, A.Q., Voss, D.A.: *Using Meshfree Approximation for Multi-Asset American option problems*. North-Holland, 1981.
- [FPS] Fouque, J.-P.; Papanicolaou, G.; Sircar, K.: *Derivatives in Financial Markets with Stochastic Volatility*. Cambridge University Press, 2000.
- [FV] Forsyth, P.; Vetzal, K.: *Quadratic convergence of a penalty method for valuing American options*. SIAM J. Sci. Comp. 23, 2096-2123, 2002.
- [Ge] Geske, R.: *A note on an analytic valuation formula for unprotected American Options on stocks with known dividends*. J. Financial Economics 7, 375-380, 1979.
- [GG] Gerstner, T.; Griebel, M.: *Praktikumsausarbeitung Computational Finance*. Inst. Angew. Math. Uni. Bonn, SS 2003.
- [GLT] Glowinski, R.; Lions, J.L.; Trémolières, R.: *Numerical Analysis of Variational Inequalities*. North-Holland, 1981.
- [GO] Griebel, M.; Oswald, P.: *On the abstract theory of additive and multiplicative Schwarz algorithms*. Numer.Math. 70, 163-180, 1995.
- [GS] Geske, R., Shastri, K.: *Valuation by approximation: a comparison of alternative option valuation techniques*. J. Financial Quant. Anal. 20, 45-71, 1985.
- [H1] Hackbusch, W.: *A fast iterative method solving Poisson's equation in a general region*. In Burlisch, R.; Griegorieff, R.D.; Schröder, J.: *Numerical treatment of differential equations*. Lecture Notes in Math. 631, Springer, 1978.
- [H2] Hackbusch, W.: *Multi-Grid Methods and Applications*. Springer, 1985.

- [Hb] Hanke-Bourgeois, M.: *Grundlagen der Numerischen Mathematik und des Wissenschaftlichen Rechnens*. Teubner, 2002.
- [HH] Hämmerlin, G.; Hoffmann, K.H.: *Numerische Mathematik*. Springer, 1994.
- [HIW] Hull, J.; White, A.: *The use of the control variate technique in option pricing*. J. Financial Quant. Anal. 23, 237-251, 1988.
- [HM] Hackbusch, W.; Mittelmann, H.D.: *On multigrid methods for variational inequalities*. Numer. Math. 42, 65-76, 1983.
- [Ho] Hoppe, R.H.W.: *Multigrid algorithms for variational inequalities*. SIAM J. Numer. Anal. 24, 1046-1065, 1987.
- [Hö] Höllig, K.: *Finite-Element Methods with B-Splines*. SIAM, Frontiers in Appl. Math., 2003.
- [Hu] Hull, J.: *Options, Futures and other Derivative Securities*. Prentice Hall, Upper Saddle River, 2000.
- [HW] Han, H.; Wu, X.: *A fast numerical method for the Black-Scholes equation of American Options*. SIAM J. Numer. Anal. 41, 2081-2095, 2003.
- [JLL] Jaillet, P.; Lamberton, D.; Lapeyre, B.: *Variational inequalities and the pricing of American options*. Acta Appl. Math. 21, 263-289, 1990.
- [K] Kunoth, A.: *Mathematische Methoden in der Geometrischen Datenverarbeitung (CAGD)*. Vorlesungsskript, FU Berlin, 1999.
- [Ko1] Kornhuber, R.: *Monotone multigrid methods for elliptic variational inequalities I*. Numer. Math. 69, 167-184, 1994.
- [Ko2] Kornhuber, R.: *Adaptive Monotone Multigrid Methods for Nonlinear Variational Problems*. Teubner, 1997.
- [Kr] Krause, R.: *Monotone multigrid methods for Signorini's problem with friction*. Dissertation, FU Berlin, 2001.
- [KS] Kinderlehrer, D.; Stampacchia, G.: *An Introduction to Variational Inequalities and their Applications*. Academic Press, 1980.
- [Kw] Kwok, Y.-K.: *Mathematical Models of Financial Derivatives*. Springer, 1998.
- [KY] Kornhuber, R.; Yserentant, H.: *Multilevel methods for elliptic problems on domains not resolved by the coarse grid*. Contemp. Math. 180, 49-60, 1994.
- [LL] Lamberton, D.; Lapeyre, B.: *Introduction to Stochastic Calculus Applied to Finance*. Chapman & Hall, 1999.
- [LM] Lions, P.L.; Mercier, B.: *Approximation numérique des équations de Hamilton-Jacobi-Bellman*. RAIRO Anal. Numér. 14, 369-393, 1980.
- [LP] Lutterkort, D.; Peters, J.: *Smooth paths in a polygonal channel*. ACM Symposium on Comput. Geometry, 1999.
- [Ma] Mandel, J.: *A multilevel iterative method for symmetric, positive definite linear complementarity problems*. Appl. Math. Optimization 11, 77-95, 1984.
- [Mc] McKean, H.P.: *Appendix: a free boundary problem for the heat equation arising from a problem in Mathematical Economics*. Industrial Management Review 6, 32-39, 1965.
- [MCC] Madan, D.-B.; Carr, P.; Chang, E.: *The Variance Gamma Process and Option Pricing*. European Finance Review 2, 79-105, 1998.
- [Me] Meyer, G.-H.: *On pricing american and asian options with PDE methods*. Acta Math. Univ. Comen. New Ser. 70, 153-165, 2001.
- [Mi] MacMillan, L.W.: *Analytic approximation for the American put option*. Advances Futures Options Res. 1, 119-139, 1986.
- [MPS] Matache, A.-M.; Petersdorff, T.; Schwab, C.: *Fast deterministic pricing of options on levy driven assets*. erscheint in Math. Modeling and Numer. Anal., 2002.
- [Mv] Matveev, E.R.: *On a super one-sided spline approximation of functions of several variables*. Izvestiya VUZ. Matematika 32, 49-54, 1988.

- [N] Nitsche, J.A.: L^∞ convergence of Finite Element approximations. Lecture Notes in Math. 606, 261-274, 1977.
- [Oo] Oosterlee, C.-W.: On multigrid for linear complementarity problems with application to American-style options. Electron. Trans. Numer. Anal. 15, 165-185, 2003.
- [PFV] Pooley, D.-M.; Forsyth, A.; Vetzal, K.-R.: Numerical convergence properties of option pricing PDEs with uncertain volatility. J. Numer. Anal. 23, 241-267, 2000.
- [Pi] Pinkus, A.M.: On L^1 - Approximation. Cambridge University Press, 1989.
- [Ra] Raible, S.: Levy processes in Finance: Theory, Numerics, and Empirical Facts. Doktorarbeit Albert-Ludwigs-Uni. Freiburg i.Br., 2000.
- [Ro] Roll, R.: An analytic valuation formula for unprotected American Call options on stocks with known dividends. J. Financial Economics 5, 251-258, 1977.
- [RT] Rogers, L.-C.; Taylay, D.: Numerical Methods in Finance. Cambridge, University Press, 1997.
- [Sa] Sandmann, K.: Einführung in die Stochastik der Finanzmärkte. Springer, 1999.
- [Sb] Schoenberg I.J.: Contributions to the problem of approximation of equidistant data by analytic functions. Quart. Appl. Math. 4, 45-99, 1946.
- [Se] Seydel, R.: Tools for Computational Finance. Springer, 2002.
- [Si] Sircar, K.-R.; Papanicolaou, G.: Stochastic volatility, smile and asymptotics. Applied Mathematical Finance 6, 107-145, 1999.
- [Sj] Schrijver, A.: Theory of Linear and Integer Programming. John Wiley and Sons, 1986.
- [Tai] Tai, X.-C.: Rate of convergence for some constraint decomposition methods for nonlinear variational inequalities. Numer. Math. 93, 755-786, 2003.
- [Th] Thomée, V.: Galerkin Finite Element Methods for Parabolic Problems. Springer, 1997.
- [TR] Tavella, D.; Randall, C.: Pricing Financial Instruments. John Wiley and Sons, 2000.
- [Wh] Whaley, R.: On the valuation of American Call options on stocks with known dividends. J. Financial Economics 9, 207-211, 1981.
- [WHD1] Wilmott, P.; Howison S.; Dewynne, J.: Option Pricing. Oxford, Financial Press, 1993.
- [WHD2] Wilmott, P.; Howison S.; Dewynne, J.: The Mathematics of Financial Derivatives. Cambridge, University Press, 1995.
- [Zh1] Zhang, X.-L.: Multilevel Schwarz Methods. Numer. Math. 63, 521-539, 1992.
- [Zh2] Zhang, X.-L.: Valuation of American options in a jump-diffusion model. in [RT], 93-111, 1997.