

Numerische Methoden der Finanzmathematik

Lars Grüne
Mathematisches Institut
Fakultät für Mathematik und Physik
Universität Bayreuth
95440 Bayreuth
lars.gruene@uni-bayreuth.de
www.math.uni-bayreuth.de/~lgruene/

Vorlesungsskript
2. Auflage
Wintersemester 2010/2011

Vorwort

Dieses Skript ist im Rahmen einer gleichnamigen Vorlesung entstanden, die ich im Wintersemester 2010/2011 an der Universität Bayreuth gehalten habe. Es ist die zweite Auflage eines Skriptes, das zuerst im Sommersemester 2009 erstellt wurde, für diese zweite Auflage aber grundlegend überarbeitet und erweitert wurde.

Die einzelnen Kapitel des Skriptes wurden auf Basis der im Literaturverzeichnis angegebenen Lehrbücher und Monographien erstellt. Wie immer möchte ich mich bei allen aufmerksamen StudentInnen bedanken, die viele kleinere Fehler gefunden haben, die in dieser Version korrigiert werden konnten.

Eine elektronische Version dieses Skripts sowie die zu dieser Vorlesung gehörigen Übungsaufgaben finden sich im WWW unter dem Link “Lehrveranstaltungen” auf der Seite <http://www.math.uni-bayreuth.de/~lgruene/>.

Bayreuth, Februar 2011

LARS GRÜNE

Inhaltsverzeichnis

Vorwort	i
1 Einführung und Grundlagen	1
1.1 Optionen	2
1.2 Das Marktmodell	3
2 Optionen und ihre Bewertung	7
2.1 Risikoneutrale Bewertung	8
2.2 Grundalgorithmus der Optionsbewertung	11
2.3 Analytische Aussagen über Optionswerte	12
2.4 Folgerungen aus der risikoneutralen Bewertung	16
3 Die Binomialmethode	19
3.1 Zeitdiskrete stochastische dynamische Systeme	19
3.2 Das Binomialmodell eines Aktienkurses	20
3.3 Optionswertberechnung mit dem Binomialmodell	21
3.4 Die Parameter im Binomialmodell	23
4 Stochastische Differentialgleichungen	25
4.1 Der Wiener Prozess	25
4.2 Das Itô-Integral	27
4.2.1 Das Itô-Lemma	30
4.2.2 Die geometrische Brownsche Bewegung	32
4.3 Ein anspruchsvolleres Kursmodell	34
5 Die Monte-Carlo Methode	37
5.1 Grundidee	37
5.2 Konvergenzgeschwindigkeit	39
5.3 Antithetische Zufallsvariablen	40

6	Numerische Lösung deterministischer DGLen	43
6.1	Definition	43
6.1.1	Anfangswertprobleme	44
6.1.2	Grafische Darstellung der Lösungen	45
6.2	Allgemeine Theorie der Einschrittverfahren	47
6.2.1	Diskrete Approximationen	47
6.2.2	Erste einfache Einschrittverfahren	48
6.2.3	Konvergenztheorie	50
6.3	Taylor-Verfahren	55
6.3.1	Definition	55
6.3.2	Eigenschaften	57
6.4	Ableitungsfreie Verfahren	58
6.4.1	Konsistenz beliebiger Einschrittverfahren	58
6.4.2	Explizite Runge-Kutta-Verfahren	61
7	Numerische Lösung stochastischer DGLen	65
7.1	Stochastische Konvergenz- und Approximationsbegriffe	65
7.2	Approximationen des Wiener-Prozesses	68
7.3	Das stochastische Euler-Verfahren	71
7.4	Anwendung auf die Optionsbewertung	73
7.5	Die Itô-Taylor-Entwicklung	75
7.6	Itô-Taylor-Verfahren	79
7.7	Spezielle Verfahren	82
7.8	Das Binomialmodell als schwache Approximation	86
8	Die Black-Scholes Gleichung	89
9	Hedging und die “Griechen”	95
9.1	Zeitdiskretes Hedging	96
9.2	Die “Griechen”	100
9.3	Monte-Carlo für das Δ	101
10	Schätzung der Volatilität	105
10.1	Historische Volatilität	105
10.2	Implizite Volatilität	110

11 Einführung in die Finite Differenzen-Methode	115
11.1 Black-Scholes- und Wärmeleitungsgleichung	116
11.2 Grundidee der Finiten Differenzen	117
11.3 Lösung der Finiten Differenzgleichungen	120
11.4 Konsistenz, Stabilität und Konvergenz	122
12 Finite Differenzen für die Black-Scholes Gleichung	127
12.1 Wahl der numerischen Randbedingungen	127
12.2 Direkte Diskretisierung der Black-Scholes-Gleichung	129
12.3 Finite Differenzen und Binomialmethode	131
Literaturverzeichnis	134
Index	136

Kapitel 1

Einführung und Grundlagen

Die Finanzmathematik ist ein relativ junges und im letzten Jahrzehnt stark gewachsenes Teilgebiet der angewandten Mathematik. Aufgrund der ständig steigenden Komplexität der weltweit gehandelten Finanzinstrumente sind diese — wenn überhaupt — oft nur noch mit mathematischen Methoden wirklich zu verstehen.

Der Bereich der Finanzmathematik, den wir hier näher betrachten wollen, beschäftigt sich weniger mit Börsenspekulationen (obwohl das nicht immer streng zu trennen ist, vgl. Abschnitt 1.1) sondern vielmehr mit Finanzinstrumenten, die vor allem zur Absicherung von tatsächlichen Geschäften dienen, wie z.B. zur Absicherung gegen Wechselkursschwankungen, die für international agierende Firmen zum Alltagsgeschäft gehört. Als Modellproblem betrachten wir dazu in dieser Vorlesung die Bewertung europäischer und amerikanischer Optionen. Diese Problemklasse hat den Vorteil, dass die mathematische Theorie und die zugehörige Numerik inzwischen relativ gut verstanden und ausgearbeitet ist. Zudem lassen sich die Methoden auf komplexere Finanzinstrumente erweitern, wenn man die Grundprinzipien einmal verstanden hat. Als ergänzende Literatur empfehle ich dabei die Bücher

- [1] M. Günther und A. Jüngel, Finanzderivate mit MATLAB. Vieweg, 2003
- [2] D.J. Higham, An introduction to financial option valuation. Mathematics, stochastics and computation. Cambridge University Press, 2004
- [4] R. Seydel, Tools for computational finance. Springer, 2009 (4. Auflage)¹,

die auch die Grundlage dieses Skripts bilden.

Im Rahmen der immer noch nicht überstandenen globalen Finanzkrise stellt sich natürlich die Frage, inwieweit gerade die in dieser Vorlesung behandelten Modelle und Finanzprodukte sowie die zugehörigen numerischen Methoden Ursache der derzeitigen Probleme sind. Da ich kein Experte für Finanzmärkte bin, möchte ich hier keine Spekulationen anstellen und nur einige offensichtliche Zusammenhänge — speziell in den folgenden beiden Abschnitten — aufzeigen. Was die derzeitige Krise für die Zukunft der Finanzmathematik bedeutet,

¹Von [4] gibt es auch eine — allerdings ältere und weniger umfangreiche — deutsche Version: R. Seydel, Einführung in die numerische Berechnung von Finanz-Derivaten, Springer, 2000.

ist aus meiner Sicht derzeit noch nicht absehbar. Unstrittig erscheint mir aber, dass das Verständnis komplexer Finanzinstrumente und ihres mathematischen Hintergrundes gerade in der Krise wichtig sind.

1.1 Optionen

Die Optionsbewertung ist eine der einfachsten Aufgabenstellungen der Finanzmathematik. Trotzdem ist sie alles andere als trivial, zudem können komplexere finanzmathematische Probleme oft mit einer Verallgemeinerung der Methoden für Optionen gelöst werden. Daher werden wir sie in dieser Vorlesung als beispielhafte finanzmathematische Anwendung betrachten. Wir definieren nun zunächst, was eine Option überhaupt ist und erläutern kurz das Problem der Optionsbewertung. In den nächsten Kapiteln werden wir dann Schritt für Schritt verschiedene Methoden zur numerischen Lösung des Problems kennenlernen.

Ganz allgemein ist eine Option ein Vertrag, der dem Inhaber die Möglichkeit (aber nicht die Verpflichtung) gibt, einen *Basiswert* (z.B. ein Aktienpaket oder einen festgelegten Betrag einer Fremdwährung) zu einem vorher vereinbarten *Ausübungspreis* K zu kaufen (*Call-Option*) oder zu verkaufen (*Put-Option*). Optionen und andere ähnliche Finanzinstrumente werden *Derivate* genannt, da sich ihr Wert von dem Basiswert ableitet.

Wir betrachten dabei sowohl die *europäischen Optionen* als auch die *amerikanischen Optionen*. Bei den europäischen Optionen muss der Kauf (bzw. Verkauf) zu einem von vornherein festgelegten *Ausübungszeitpunkt* T stattfinden. Ist dieser Ausübungszeitpunkt T erreicht, so kann der Inhaber die Option ausüben, also den Basiswert zu dem vorher festgelegten Ausübungspreis kaufen (bzw. verkaufen), oder er kann die Option verfallen lassen, den Basiswert also nicht kaufen (bzw. verkaufen).

Bei der amerikanischen Option kann die Ausübung zu jedem beliebigen Zeitpunkt $t \in [0, T]$ durchgeführt werden, wobei T ein vertraglich festgelegter Zeitpunkt ist. Der Startzeitpunkt der Option wird hier per Definition immer als 0 festgelegt, was durch einfache Verschiebung des tatsächlichen zeitlichen Arguments natürlich o.B.d.A. möglich ist. In der Praxis wird der tatsächliche Kauf oder Verkauf dabei übrigens häufig gar nicht ausgeführt, sondern der Differenzbetrag zwischen Ausübungspreis und Marktpreis ausgezahlt.

Optionen können vielfältig eingesetzt werden. Eine offensichtliche Anwendung ist die Absicherung gegenüber Währungsschwankungen: Erwartet z.B. ein europäisches Unternehmen in T Monaten eine Zahlung von 1 Mio. US-Dollar, so könnte es für diesen Zeitpunkt eine Put-Option auf den Verkauf von 1 Mio. US-Dollar für K Euro erwerben und hätte sich damit einen garantierten Wechselkurs gesichert. Liegt der Wert der 1 Mio. US-Dollar dann zum Ausübungszeitpunkt T unter dem Wert K Euro, so würde die Option ausgeübt um die garantierten Einnahmen in Euro zu realisieren. Liegt der Wert oberhalb der K Euro, so würde die Option nicht ausgeübt und das Geld am Markt umgetauscht. Analog kann man sich z.B. gegen schwankende Rohstoffpreise absichern.

Optionen eignen sich aber auch hervorragend zur Spekulation. Vermutet ein Spekulant, dass der Wert eines Aktienpakets eines Unternehmens vom heutigen Kurs K_1 bis zum Zeitpunkt T auf den Wert $K_2 < K_1$ fällt, so kann er eine Put-Option mit Ausübungspreis $K = K_1$ erwerben. Geht die Spekulation auf, so kann er zum Ausübungszeitpunkt das

Aktienpaket für den günstigeren Preis K_2 an der Börse kaufen und durch Ausübung der Option für den Preis $K = K_1$ teurer verkaufen.

Klar ist nun, dass der Ausgeber (oder Emittent) der Option (typischerweise eine Bank oder eine Versicherung) eine solche Option nicht umsonst ausgeben wird sondern einen Preis verlangen wird, der das mit der Option behaftete Risiko widerspiegelt. Das Problem der *Optionsbewertung* ist nun: Welchen Wert besitzt diese Option selbst zu einem Zeitpunkt $t < T$? Aus diesem Wert kann dann die Gebühr (typischerweise natürlich noch mit einem Aufschlag) errechnet werden. Der so ermittelte Wert der Option wird auch *fairer Preis* genannt, denn in der Praxis werden auch viele Optionen selbst wieder an der Börse gehandelt, so dass der tatsächliche Preis der Option wiederum durch den Markt bestimmt wird. Trotzdem ist die Berechnung des fairen Preises zur Einschätzung des Wertes einer Option auf jeden Fall theoretisch wichtig, im Falle spezialisierter Optionen, für die kein großer Markt besteht, aber auch ganz praktisch.

Ein eng verwandtes Problem ist die Frage, wie sich der Ausgeber der Option nun selbst wieder gegen das Risiko absichert. Je besser das gelingt, desto günstiger kann der Preis kalkuliert werden. Wir werden hier das Prinzip der *risikoneutralen Bewertung* zu Grunde legen. Dieses Prinzip hat seine Begründung in einer aus der Black-Scholes Theorie stammenden *Hedging-Strategie* (Absicherungsstrategie) und wird in späteren Kapiteln ausführlich betrachtet. Anschaulich lässt sich das Prinzip allerdings bereits an einem sehr einfachen Kursmodell erläutern, was wir in Abschnitt 2.1 machen werden.

1.2 Das Marktmodell

Um den Finanzmarkt mathematisch modellieren zu können, müssen wir Annahmen treffen, die sich in mathematische Aussagen umsetzen lassen. Hierzu existieren eine ganze Reihe sehr ausgefeilter Modelle; hier werden wir uns aber auf die folgenden einfachen Grundannahmen beschränken, die für unsere Zwecke ausreichen:

- Zinsen werden kontinuierlich mit dem jährlichen Zinssatz r berechnet: für eine Anlage von x Euro für t Jahre ergibt sich der Zins zu $e^{rt}x$.
- Geld kann jederzeit zu dem zeitlich unveränderlichen Zinssatz r über einen beliebigen Zeitraum festverzinslich angelegt werden.
- Kredite können jederzeit zu dem gleichen Zinssatz r für einen beliebigen Zeitraum aufgenommen werden.
- Der Markt ist *arbitragefrei*, d.h. es gibt keine risikofreie Anlage, mit der ein Betrag x so angelegt werden kann, dass er nach einer Laufzeit T eine Auszahlung $y > e^{rT}x$ liefert.

Einige dieser Annahmen sind lediglich vereinfachend und können (allerdings mit z.T. nicht unerheblichem mathematischem Aufwand) gelockert werden, z.B. die Annahme, dass der Zinssatz r nicht mit der Zeit schwankt oder die Annahme, dass der Zinssatz für Anlagen und Kredite gleich ist. Andere Annahmen, wie z.B. die Verfügbarkeit von Krediten oder die Arbitragefreiheit sind wesentlich für die im Folgenden entwickelten Methoden. Beide

sind durchaus problematisch: gerade in einer Finanzkrise ist die Annahme, dass Kredite in beliebigem Umfang verfügbar sind, sicherlich nicht gerechtfertigt und möglicherweise ist diese falsche Annahme eine der Ursachen für das Ausmaß der derzeitigen Krise. Auch die Arbitragefreiheit gilt in der Praxis nicht in dieser Reinform. Arbitragefreiheit verbietet z.B., dass Produkte auf verschiedenen Finanzmärkten zu verschiedenen Preisen gehandelt werden, da man sonst durch Kauf eines Produktes auf dem “billigeren” und Verkauf auf dem “teureren” Markt einen risikofreien Gewinn erzielen könnte. Tatsächlich gibt es aber solche Preisunterschiede, allerdings sind diese oft nur klein oder nur von kurzer Dauer, so dass wir die Annahme näherungsweise als erfüllt ansehen können.

Um zu zeigen, wie man aus diesen Annahmen präzise mathematische Aussagen ableiten kann, betrachten wir die folgende Frage der *Abzinsung* (oder *Diskontierung*), die später für die Optionsbewertung eine wichtige Rolle spielen wird: Angenommen, wir kennen den Wert $P(T) > 0$ eines Produktes P (Rohstoff, Anlage, Währung etc.) zu einem zukünftigen Zeitpunkt T . Was ist dann der Wert $P(t)$ dieses Produktes zu einem früheren Zeitpunkt $t < T$? Als konkretes Beispiel könnte $P(T)$ die vertraglich gesicherte Auszahlung einer Anlage zu einem festgelegten zukünftigen Termin T sein und die Frage wäre, welchen Betrag $P(t)$ wir sinnvollerweise zahlen sollten, wenn wir diese Anlage zum heutigen Zeitpunkt t einer anderen Person abkaufen wollten.

Um den Preis $P(t)$ zu ermitteln, vergleichen wir den erzielbaren Gewinn mit dem Gewinn aus einer festverzinslichen Anlage.

Nehmen wir an, wir besitzen das Produkt P und würden es zum Zeitpunkt t verkaufen und den Erlös $P(t)$ festverzinslich anlegen. Auf diese Weise erhielten wir gemäß unserer Annahmen zum Zeitpunkt T gerade $B(T) = e^{r(T-t)}P(t)$. Hätten wir P gehalten, so erhielten wir zum Zeitpunkt T gerade $P(T)$. Gälte für den Preis $P(t)$ nun

$$P(t) > e^{r(t-T)}P(T),$$

so hätten wir zum Zeitpunkt T einen risikofreien Gewinn von

$$B(T) - P(T) = e^{r(T-t)}P(t) - P(T) > e^{r(T-t)}e^{r(t-T)}P(T) - P(T) = 0$$

erzielt, was nach der Annahme der Arbitragefreiheit nicht möglich ist.

Andererseits könnten wir zum Zeitpunkt t einen Kredit der Höhe $P(t)$ aufnehmen und dafür das Produkt zum Preis $P(t)$ erwerben. Zum Zeitpunkt T erhalten wir dann vertragsgemäß $P(T)$ und zahlen davon den Kredit mit Zinsen, also $B(T) = e^{r(T-t)}P(t)$ zurück. Gälte für den Preis $P(t)$ nun

$$P(t) < e^{r(t-T)}P(T),$$

so hätten wir zum Zeitpunkt T einen risikofreien Gewinn von

$$P(T) - B(T) = P(T) - e^{r(T-t)}P(t) > P(T) - e^{r(T-t)}e^{r(t-T)}P(T) = 0$$

erzielt, was nach der Annahme der Arbitragefreiheit wiederum nicht möglich ist. Wir haben damit den folgenden Satz bewiesen.

Satz 1.1 Der aus den Marktannahmen folgende faire Preis des Produktes P zum Zeitpunkt $t < T$ beträgt

$$P(t) = e^{r(t-T)}P(T).$$

Bemerkung 1.2 Der Satz lässt sich wie folgt verallgemeinern (Übungsaufgabe): Wenn der Wert von P zum Zeitpunkt T garantiert die Ungleichung $P(T) \geq P_0$ für ein $P_0 \in \mathbb{R}$ erfüllt, so folgt $P(t) \geq e^{r(t-T)}P_0$. Analog gilt dies für “ \leq ”. Beachte, dass $P(T)$ hier negativ sein kann. P wäre in diesem Fall eine Anlage, bei der zum Zeitpunkt T eine Zahlungsverpflichtung eintritt, z.B. ein Kredit mit Fälligkeit zum Zeitpunkt T . \square

Kapitel 2

Optionen und ihre Bewertung

Wie bereits in Abschnitt 1.1 beschrieben, ist eine Option ein Vertrag, der dem Inhaber die Möglichkeit (aber nicht die Verpflichtung) gibt, einen *Basiswert* (z.B. ein Aktienpaket oder einen festgelegten Betrag einer Fremdwahrung) zu einem vorher vereinbarten *Ausungspreis* K zu kaufen (*Call-Option*) oder zu verkaufen (*Put-Option*). Optionen und andere hnliche Finanzinstrumente werden *Derivate* genannt, da sich ihr Wert von dem Basiswert ableitet.

Wir betrachten dabei sowohl die *europaischen Optionen* als auch die *amerikanischen Optionen*. Bei den europaischen Optionen muss der Kauf (bzw. Verkauf) zu einem von vornherein festgelegten *Ausungszeitpunkt* T stattfinden. Ist dieser Ausungszeitpunkt T erreicht, so kann der Inhaber die Option ausen, also den Basiswert zu dem vorher festgelegten Ausungspreis kaufen (bzw. verkaufen), oder er kann die Option verfallen lassen, den Basiswert also nicht kaufen (bzw. verkaufen).

Bei der amerikanischen Option kann die Ausung zu jedem beliebigen Zeitpunkt $t \in [0, T]$ durchgefurt werden, wobei T ein vertraglich festgelegter Zeitpunkt ist. Der Startzeitpunkt der Option wird hier per Definition immer als 0 festgelegt, was durch einfache Verschiebung des Zeitachse o.B.d.A. moglich ist.

Im Folgenden verwenden wir die folgenden Bezeichnungen:

$S(t)$: Basiswert zur Zeit t mit $S(t) \geq 0$ fur alle $t \in [0, T]$

$V(t, S)$: Preis der Option zur Zeit t abhangig vom aktuellen Basiswert S

$K > 0$: vereinbarter Ausungspreis

$T > 0$: vereinbarte Laufzeit

Falls wir genauer spezifizieren mochten, um welche Option es sich handelt, verwenden wir die Bezeichnungen V_C und V_P fur Put bzw. Call oder noch genauer V_{Ce} , V_{Pe} , V_{Ca} oder V_{Pa} fur den europaischen Call, europaischen Put, amerikanischen Call und amerikanischen Put.

2.1 Risikoneutrale Bewertung

Eines der wesentlichen Konzepte der Optionsbewertung, das durch die Arbeiten von Black, Scholes und Merton eingeführt wurde, ist das Prinzip der *risikoneutralen Bewertung*.

Dieses Prinzip besagt, dass der Wert einer Option $V(t, S(t))$ zum Zeitpunkt $t < T$ analog zu der Formel aus Satz 1.1 als

$$V(t, S(t)) = e^{r(t-T)} E(V(T, S(T)))$$

berechnet wird, obwohl der Optionswert $V(T, S(T))$ zum Zeitpunkt $t < T$ *nicht* sicher bekannt ist sondern nur sein Erwartungswert.

Warum sollte man das so machen? Der Grund dafür ist, dass der sogenannte *faire Wert* der Option gar nicht über den zufälligen zukünftigen Wert der Option definiert wird sondern vielmehr über das Kapital, das die Ausgeberin der Option (typischerweise die Bank) aufwenden muss, um sich gegen Verluste aus dem Optionsgeschäft abzusichern.

Dieses Vorgehen lässt sich am Einfachsten an einem Beispiel erläutern, bei dem wir eine Aktie durch ein sehr einfaches stochastisches Modell beschreiben. Wir betrachten dazu eine Aktie $S(t)$ mit bekanntem Wert $S(0)$. Für den zukünftigen Zeitpunkt $T > 0$ ist der Wert $S(T)$ unbekannt, kann aber wie folgt stochastisch beschrieben werden:

Mit Wahrscheinlichkeit $p \in [0, 1]$ steigt die Aktie auf den Wert $S(T) = S^u > S(0)$ und mit Wahrscheinlichkeit $1 - p$ fällt die Aktie auf den Wert $S(T) = S^d < S(0)$ (u steht für “up” und d für “down”).

Desweiteren stehe auf unserem einfachen Finanzmarkt eine festverzinsliche Anleihe $B(t)$ zur Verfügung, deren Wert gemäß der Annahmen aus Kapitel 1 mit dem festen Zinssatz wächst, also $B(T) = e^{rT} B(0)$.

Betrachten wir nun eine europäische Call-Option V_C auf S mit Ausübungszeitpunkt T und Ausübungspreis K . Zum Zeitpunkt T ist der Wert der Option nun gerade durch die Höhe der Ersparnis bestimmt, die man durch die Option erhält. Falls $S(T) > K$ ist, ist dies gerade $S(T) - K$. Falls $S(T) < K$ ist würde man die Aktie über die Option teurer als am Markt erwerben. Folglich würde man die Option nicht ausüben und hat folglich die Ersparnis 0. Insgesamt erhält man also

$$V(T, S(T)) = \max\{S(T) - K, 0\}$$

und den Erwartungswert

$$E(V(T, S(T))) = p \max\{S^u - K, 0\} + (1 - p) \max\{S^d - K, 0\}.$$

Dieser Wert ist nun gerade der erwartete Verlust der Ausgeberin der Option zum Zeitpunkt T . Die Bank kann also aus dem Kursmodell den erwarteten Verlust, nicht aber den tatsächlichen Verlust berechnen. Auf Basis des Erwartungswertes kann die Bank also nicht ermitteln, was genau sie zu Zeitpunkt T an den Halter der Option zahlen muss.

Um dieses Risiko zu vermeiden, kann die Bank nun zum Zeitpunkt $t = 0$ ein Portfolio π zusammenstellen, das zum Zeitpunkt T genau den Wert $V(T, S(T))$ besitzt. Dass dies tatsächlich geht, jedenfalls wenn das stochastische Aktienmodell stimmt, zeigt die folgende Rechnung:

Das Portfolio bestehe aus Δ Anteilen der Aktie und β Anteilen der festverzinslichen Anleihe. Es besitzt zum Zeitpunkt $t = 0$ also den Wert

$$\pi(0) = \Delta S(0) + \beta B(0).$$

Falls die Aktie steigt, besitzt das Portfolio zum Zeitpunkt T den Wert

$$\pi(T) = \Delta S(T) + \beta B(T) = \Delta S^u + \beta e^{rT} B(0)$$

und falls die Aktie fällt den Wert

$$\pi(T) = \Delta S(T) + \beta B(T) = \Delta S^d + \beta e^{rT} B(0).$$

Damit das Portfolio zum Zeitpunkt T in beiden Fällen gerade den Wert $V(T, S(T))$ annimmt, müssen also die Gleichungen

$$\begin{aligned} \Delta S^u + \beta e^{rT} B(0) &= \max\{S^u - K, 0\} \\ \Delta S^d + \beta e^{rT} B(0) &= \max\{S^d - K, 0\} \end{aligned}$$

gelten. Dies ist ein lineares Gleichungssystem mit den Unbekannten Δ und β , das wegen $S^u \neq S^d$ eindeutig lösbar ist. Die Lösung ist

$$\begin{aligned} \Delta &= \frac{\max\{S^d - K, 0\} - \max\{S^u - K, 0\}}{S^d - S^u} \\ \beta &= \frac{\max\{S^d - K, 0\} S^u - \max\{S^u - K, 0\} S^d}{e^{rT} B(0) (S^u - S^d)}. \end{aligned}$$

Stellt die Bank also das Portfolio zum Zeitpunkt $t = 0$ mit diesen Anteilen zusammen, so besitzt das Portfolio zum Zeitpunkt $t = T$ gerade den Wert der Option, egal ob die Aktie steigt oder fällt. Die Bank kann das Portfolio also zum Zeitpunkt T verkaufen, dadurch ihre Kosten aus dem Optionsgeschäft bestreiten und hat so kein Risiko mehr.

Dieses Vorgehen nennt sich *Hedging* und bildet die Basis für die Definition des fairen Optionspreises $V(0, S(0))$. Dieser ist nämlich gerade als der Wert des Portfolios π mit den gerade ermittelten Parametern Δ und β zur Zeit $t = 0$ definiert, also

$$V(0, S(0)) = \pi(0).$$

Die Vorstellung dabei ist, dass die Bank die Optionsgebühr zum Zeitpunkt $t = 0$ in das Portfolio investiert und ihr möglicher Verlust aus dem Optionsgeschäft dann zum Zeitpunkt $t = T$ durch den Wert des Portfolios abgedeckt ist. Das dafür konstruierte Portfolio π heißt *replizierendes (oder nachbildendes) Portfolio*, weil es den Wert der Option nachbildet, egal wie sich der Aktienkurs entwickelt.

Was hat dies nun mit der eingangs erwähnten risikoneutralen Bewertung zu tun? Schaut man sich die obige Herleitung an, so stellt man fest, dass der faire Optionspreis durch die Lösung eines linearen Gleichungssystems gegeben ist. Möchte man nun komplexere stochastische Modelle zur Modellierung des Kursverlaufs $S(t)$ verwenden — was unbedingt anzuraten ist, da das obige einfache Modell die Realität sicherlich nur sehr unzutreffend widerspiegelt — so kann man sich leicht vorstellen, dass die Berechnung von $\pi(0)$ auf die

oben beschriebene Weise sehr kompliziert wird. Insbesondere werden wir in den nächsten Kapiteln Modelle kennenlernen, in denen $S(T)$ mehr als zwei oder sogar unendlich viele verschiedene Werte annehmen kann. In diesem Fall ist der Ansatz über das lineare Gleichungssystem nicht praktikabel. Mit der risikoneutralen Bewertung kann man die explizite Berechnung des replizierenden Portfolios π nun vermeiden.

Dazu verwenden wir, dass π die Gleichung

$$E(\pi(T)) = E(V(T, S(T)))$$

erfüllt, denn wenn $\pi(T)$ in jedem Fall mit $V(T, S(T))$ übereinstimmt, muss das auch für den Erwartungswert gelten. Es gilt nun wegen der Linearität des Erwartungswertes, der Annahme an $S(T)$ und der Tatsache, dass $B(T)$ nicht vom Zufall abhängt

$$E(\pi(T)) = E(\Delta S(T) + \beta B(T)) = \Delta E(S(T)) + \beta B(T) = \Delta(pS^u + (1-p)S^d) + \beta e^{rT} B(0).$$

Zudem gilt, dass der faire Optionswert überhaupt nicht von den Wahrscheinlichkeiten p und $1-p$ abhängt, denn weder im Ansatz für π noch in dem Gleichungssystem für Δ und β tauchen diese Werte auf. Wir können also p so wählen, dass sich der Ausdruck für $E(\pi(T))$ “schön” vereinfacht ohne dass dies den Optionswert beeinflusst. Mit der Wahl

$$p = \frac{e^{rT} S(0) - S^d}{S^u - S^d}$$

erhalten wir $pS^u + (1-p)S^d = e^{rT} S(0)$ und damit

$$E(\pi(T)) = \Delta e^{rT} S(0) + \beta e^{rT} B(0) = e^{rT} \pi(0).$$

Dies ergibt

$$V(0, S(0)) = \pi(0) = e^{-rT} E(\pi(T)) = e^{-rT} E(V(T, S(T))).$$

Dies ist genau die eingangs erwähnte risikoneutrale Bewertungsformel für $t = 0$. Zu beachten ist hierbei, dass die Formel für p nur für $S_d \leq e^{rT} S(0) \leq S^u$ eine Wahrscheinlichkeit in $[0, 1]$ liefert. Diese Ungleichungen folgen aber — mit ein klein wenig Nachdenken — aus Bemerkung 1.2 und sind damit unter der Annahme der Arbitragefreiheit automatisch erfüllt.

Bemerkung 2.1 Beachte, dass β in dem obigen Rechenbeispiel negativ sein kann, was z.B. für $S^d \leq K$ leicht zu sehen ist. Dies bedeutet, dass das Portfolio einen Kredit enthält, der zum Zeitpunkt $t = 0$ aufgenommen wird und zum Zeitpunkt $t = T$ zurückgezahlt wird. Bei komplizierteren Hedging-Strategien kann es auch passieren, dass Δ negativ wird, was einem sogenannten Leerverkauf der Aktie S entspricht. Das Hedging funktioniert in diesem Fall nur, wenn Leerverkäufe erlaubt sind. Wir kommen im übernächsten Abschnitt noch einmal auf die Leerverkäufe zurück. \square

Die hier für das einfache Aktienmodell durchgeführte Herleitung zeigt, dass das Hedging fester Bestandteil der risikoneutralen Bewertung ist. Insbesondere ist bei der Optionsbewertung nicht nur der Wert der Option, sondern auch die Werte Δ und β im replizierenden Portfolio wichtig, die bei komplexeren Kursmodellen auch zeitabhängig sein können. Wir werden in den folgenden Kapitel trotzdem zunächst nur das Problem der Optionsbewertung — also die Berechnung von $V(t, S(t))$ — betrachten. Auf das Hedging kommen wir dann in Kapitel 9 zurück, nachdem wir die Black-Scholes Theorie eingeführt haben.

2.2 Grundalgorithmus der Optionsbewertung

Die Grundidee der risikoneutralen Bewertung europäischer Optionen, die in allen im Verlauf dieser Vorlesung behandelten numerischen Methoden gleich ist, besteht aus den folgenden drei Schritten:

Algorithmus 2.2 (Grundalgorithmus der Optionsbewertung)

- (1) Bestimme eine Formel für den Wert $V(T, S)$ zum Laufzeitende (in Abhängigkeit vom Kurs $S = S(T)$ am Laufzeitende)
- (2) Bestimme ausgehend vom Basiswert $S(t)$ zur Zeit $t < T$ mit Hilfe eines stochastischen Kursmodells die Zufallsvariable $S(T) = S(T, \omega)$
- (3) Berechne den Optionswert als den abgezinnten Erwartungswert

$$V(t, S(t)) = e^{r(t-T)} E(V(T, S(T)))$$

□

Schritt (3) ist dabei gerade das im vorhergehenden Abschnitt motivierte Prinzip der *risikoneutralen Bewertung*. Wenngleich wir dieses Prinzip nur für ein sehr einfaches Kursmodell hergeleitet haben, wird es auch für komplexere Modelle verwendet. Wir werden im Rahmen der Black-Scholes-Theorie sehen, dass das Prinzip auch für das dort verwendete Kursmodell gerechtfertigt ist. Der Name “risikoneutral” ergibt sich daraus, dass man dem in der Preisentwicklung enthaltenen Risiko neutral gegenübersteht und den Erwartungswert wie den sicheren Wert in Satz 1.1 behandelt. Im Gegensatz dazu stehen die *risikoaverse Bewertung*, in der man zum Ausgleich des Risikos einen kleineren Faktor als $e^{r(t-T)}$ verwendet, und die *risikofreudige Bewertung*, in der man einen größeren Faktor benutzt. Tatsächlich hat diese Bewertungsmethode Auswirkungen auf das Kursmodell in (2), die wir in Abschnitt 2.4 betrachten.

Um von Algorithmus 2.2 zu tatsächlich implementierbaren numerischen Algorithmen zu kommen, werden wir im Folgenden die einzelnen Schritte genauer untersuchen: Schritt (1) werden wir im folgenden Abschnitt behandeln. Für Schritt (2) werden wir in den folgenden Kapiteln verschiedene stochastische Modelle kennen lernen. Die numerische Umsetzung von Schritt (3) hängt dann wesentlich von dem in (2) verwendeten Modell ab. Je nach Modell kann der Erwartungswert entweder direkt (vgl. Kapitel 3), durch stochastische Simulation (vgl. Kapitel 5), oder durch analytische Umformung in eine partielle Differentialgleichung mit anschließender Diskretisierung (vgl. Kapitel 8 und 11) numerisch berechnet werden.

Für amerikanische Optionen muss Schritt (3) modifiziert werden, denn es muss zusätzlich für jeden Zeitpunkt zwischen t und T geprüft werden, ob eine vorzeitige Ausübung vorteilhaft ist. Dies funktioniert nicht für jedes der in dieser Vorlesung vorgestellten Verfahren. Aber auch wenn es prinzipiell funktioniert, ist eine Überprüfung für *jeden* Zeitpunkt dabei numerisch aus offensichtlichen Gründen nicht möglich. Tatsächlich wird diese Überprüfung nur zu diskreten Zeitpunkten durchgeführt. Welche Zeitpunkte dafür verwendet werden und wie dies im Detail implementiert wird, hängt dabei von der konkreten numerischen Ausgestaltung von Schritt (3) ab.

2.3 Analytische Aussagen über Optionswerte

In diesem Abschnitt berechnen wir zunächst die in Schritt (1) der Bewertung benötigten Formeln für den Optionswert am Laufzeitende T . Danach leiten wir daraus auf analytischem Wege einige Folgerungen für die Optionswerte zu beliebigen Zeiten $t \in [0, T]$ her.

Der Wert der Option $V(T, S)$ am Laufzeitende T ist bei Kenntnis des Basiswertes $S = S(T)$ leicht zu berechnen: Sei K der festgelegte Ausübungspreis einer Call-Option. Falls $S > K$ ist, so erzielt man durch Ausübung der Option offenbar den Gewinn $V(T, S) = S - K > 0$. Falls $S \leq K$, so würde man durch Ausübung der Option den Basiswert zu einem höheren Preis als dem aktuellen Marktpreis kaufen, folglich würde man die Option nicht ausüben, der Wert ist also $V(T, S) = 0$. Zusammen erhält man so den Wert

$$V_C(T, S) = \max\{S - K, 0\} =: (S - K)^+. \quad (2.1)$$

Gerade umgekehrt verhält es sich bei einer Put-Option. Hier wird man die Option nur dann ausüben, wenn der Marktpreis S unterhalb des Ausübungspreises K liegt. Man erhält so

$$V_P(T, S) = \max\{K - S, 0\} =: (K - S)^+. \quad (2.2)$$

Beachte, dass es am Laufzeitende keinen Unterschied zwischen europäischen und amerikanischen Optionen gibt.

Zusammen mit den Annahmen an unser Marktmodell können wir aus diesen einfachen Formeln Aussagen über den Zusammenhang der einzelnen Optionen sowie Abschätzungen für ihren Wert ableiten. In den Beweisen der folgenden Sätze spielen *Portfolios* eine wichtige Rolle. Ein spezielles Portfolio haben wir in Abschnitt 2.1 bereits kennengelernt. Allgemein ist ein Portfolio

$$\pi = \alpha_1 P_1 + \alpha_2 P_2 + \dots + \alpha_n P_n$$

eine Summe verschiedener Finanzanlagen P_i mit Gewichtungen $\alpha_i \in \mathbb{R}$ (in den folgenden Beweisen sind die α_i stets gleich $+1$ oder -1 und konstant in der Zeit, können aber allgemein beliebig und auch zeitvariant gewählt werden). Wichtig ist dabei, dass die Gewichtungen auch negativ sein können. Eine negativ gewichtete Anlage P_i kann in der Praxis durchaus eine sinnvolle Interpretation haben: Ist P_i z.B. eine Option, so ist die Halterin der negativen Option gerade die Emittentin der Option, die sich verpflichtet, den Basiswert vom Halter der Option (falls von diesem gewünscht) zum festgesetzten Ausübungspreis zu kaufen (Put) oder zu verkaufen (Call). In den folgenden Beweisen dieses Abschnitts gehen tatsächlich nur Optionen negativ in die Portfolios ein, so dass wir immer diese Interpretation verwenden können.

Prinzipiell ist es aber in der Finanzmathematik möglich, dass jede beliebige Anlage P_i mit negativem Gewicht $\alpha_i < 0$ in ein Portfolio eingehen kann — man nennt dies einen *Leerverkauf*. In der Praxis entspricht dies dem Verkauf von $-\alpha_i$ Anteilen von P_i beim Kauf des Portfolios und den Kauf von $-\alpha_i$ Anteilen von P_i beim Verkauf des Portfolios, natürlich zum jeweils gültigen Preis. Mit Leerverkäufen von Aktien lässt sich also Gewinn machen, wenn die Aktie fällt — ein Grund dafür, warum Leerverkäufe in der Finanzkrise in Verruf geraten sind und in einigen Ländern für manche Aktien sogar zeitweise verboten wurden. Voraussetzung für Leerverkäufe ist natürlich, dass sich auf dem Markt ein Partner

für das entsprechende Geschäft findet. Einige Hinweise zur technischen Abwicklung solcher Leerverkäufe finden sich z.B. auf Wikipedia.

Wir betrachten nun zunächst europäische Optionen.

Satz 2.3 Für alle $t \in [0, T]$ gilt

$$S(t) + V_{Pe}(t, S(t)) - V_{Ce}(t, S(t)) = Ke^{r(t-T)}.$$

Diese Gleichung wird auch als *Put-Call-Parität* bezeichnet und erlaubt es, den Wert des europäischen Put aus dem europäischen Call (und umgekehrt) zu berechnen.

Beweis: Wir betrachten ein Portfolio π , das aus dem Basiswert, einem europäischen Put und einem negativen europäischen Call besteht. Der Wert dieses Portfolios zur Zeit t beträgt

$$\pi(t) = S(t) + V_{Pe}(t, S(t)) - V_{Ce}(t, S(t))$$

und zur Zeit T gerade

$$\pi(T) = S(T) + (K - S(T))^+ - (S(T) - K)^+ = K.$$

Wir erhalten also zur Zeit T die garantierte Auszahlung K , weswegen der Wert des Portfolios zur Zeit t gemäß Satz 1.1 gerade $e^{r(t-T)}K$ ist, womit die Behauptung folgt.

Satz 2.4 Für alle $t \in [0, T]$ gelten die Ungleichungen

$$(i) (S(t) - Ke^{r(t-T)})^+ \leq V_{Ce}(t, S(t)) \leq S(t)$$

$$(ii) (Ke^{r(t-T)} - S(t))^+ \leq V_{Pe}(t, S(t)) \leq Ke^{r(t-T)}$$

Beweis: (i) Wir beweisen die Ungleichungen einzeln:

“ $0 \leq V_{Ce}(t, S(t))$ ”: Da sicher die Ungleichung $V_{Ce}(T, S(T)) \geq 0$ gilt, folgt die Aussage aus Bemerkung 1.2 mit $P(t) = V_{Ce}(t, S(t))$ und $P_0 = 0$.

“ $S(t) - Ke^{r(t-T)} \leq V_{Ce}(t, S(t))$ ”: Der Wert des Portfolios

$$\pi(t) = S(t) - V_{Ce}(t, S(t))$$

erfüllt zum Zeitpunkt T sicher die Ungleichung

$$\pi(T) = S(T) - (S(T) - K)^+ \leq K.$$

Aus Bemerkung 1.2 mit $P(t) = \pi(t)$ und $P_0 = K$ folgt daher

$$\pi(t) \leq e^{r(t-T)}K,$$

woraus die Ungleichung sofort folgt.

“ $V_{Ce}(t, S(t)) \leq S(t)$ ”: Das Portfolio

$$\pi(t) = S(t) - V_{Ce}(t, S(t))$$

erfüllt zum Zeitpunkt T sicher die Ungleichung

$$\pi(T) = S(T) - (S(T) - K)^+ \geq 0.$$

Die Ungleichung folgt daher aus Bemerkung 1.2 mit $P(t) = \pi(t)$ und $P_0 = 0$.

(ii) Folgt aus (i) mit Satz 2.3. □

Als nächstes betrachten wir die Beziehung zwischen den europäischen und amerikanischen Optionen. Klar ist, dass die beiden Ungleichungen

$$V_{Ca}(t, S) \geq V_{Ce}(t, S) \quad \text{und} \quad V_{Pa}(t, S) \geq V_{Pe}(t, S) \quad (2.3)$$

gelten, denn da die amerikanischen Optionen genau wie die europäischen Optionen zum Zeitpunkt T ausgeübt werden können, kann mindestens der gleiche Gewinn erzielt werden.

Durch die mögliche vorzeitige Ausübung könnte allerdings eventuell ein höherer Gewinn erzielt werden, der dann zu einem strikt größeren Wert führen würde. Der folgende Satz zeigt, dass die mögliche frühere Ausübung für den Call allerdings keinen solchen Vorteil bringt.

Satz 2.5 Der durch die vorzeitige Ausübung der amerikanischen Call-Option zu einem Zeitpunkt $t < T$ erzielbare Ertrag $(S(t) - K)^+$ ist immer kleiner oder gleich dem Wert der Option, d.h.

$$(S(t) - K)^+ \leq V_{Ca}(t, S(t)), \quad (2.4)$$

wobei die Ungleichung strikt ist, falls $(S(t) - K)^+ > 0$ gilt. Insbesondere ist eine vorzeitige Ausübung der Option also niemals sinnvoll und es folgt

$$V_{Ce}(t, S(t)) = V_{Ca}(t, S(t)).$$

Beweis: Angenommen, wir üben die amerikanische Option vorzeitig zum Zeitpunkt $t < T$ aus. Dann erhalten wir den Betrag $(S(t) - K)^+$. Aus (2.3) und Satz 2.4(i) folgt dann die Ungleichung

$$V_{Ca}(t, S(t)) \geq V_{Ce}(t, S(t)) \geq (S(t) - Ke^{r(t-T)})^+ \geq (S(t) - K)^+,$$

weil ja $e^{r(t-T)} < 1$ gilt wegen $t < T$. Die letzte Ungleichung dieser Kette ist dabei strikt, wenn $(S(t) - K)^+ > 0$ ist. Damit folgt (2.4).

Damit ist es also stets optimal, die Option entweder zum Zeitpunkt $t = T$ oder gar nicht auszuüben. Da dies genau die Möglichkeiten der europäischen Option sind, müssen die Werte übereinstimmen. □

Wir beenden den Abschnitt mit Ungleichungen für amerikanische Optionen, die analog zu Satz 2.3 und Satz 2.4(ii) sind.

Satz 2.6 Für alle $t \in [0, T]$ gelten die Ungleichungen

$$(i) \quad Ke^{r(t-T)} \leq S(t) + V_{Pa}(t, S(t)) - V_{Ca}(t, S(t)) \leq K$$

$$(ii) \quad (Ke^{r(t-T)} - S(t))^+ \leq V_{Pa}(t, S(t)) \leq K$$

Beweis: (i) Wir zeigen die beiden Ungleichungen separat.

“ $Ke^{r(t-T)} \leq S(t) + V_{Pa}(t, S(t)) - V_{Ca}(t, S(t))$ ”: Durch Anwendung von Satz 2.5, Ungleichung (2.3) und Satz 2.3 folgt

$$\begin{aligned} S(t) + V_{Pa}(t, S(t)) - V_{Ca}(t, S(t)) &= S(t) + V_{Pa}(t, S(t)) - V_{Ce}(t, S(t)) \\ &\geq S(t) + V_{Pe}(t, S(t)) - V_{Ce}(t, S(t)) \\ &= Ke^{r(t-T)}, \end{aligned}$$

also die gewünschte Ungleichung.

“ $S(t) + V_{Pa}(t, S(t)) - V_{Ca}(t, S(t)) \leq K$ ”:

Betrachte das Portfolio

$$\pi(t) = S(t) + V_{Pa}(t, S(t)) - V_{Ca}(t, S(t)).$$

Für $V_{Ca}(t, S(t))$ gilt wegen (2.5) und Satz 2.4(i) die Ungleichung $V_{Ca}(t, S(t)) \geq (S(t) - Ke^{r(t-T)})^+$ und damit

$$\pi(t) \leq S(t) + V_{Pa}(t, S(t)) - (S(t) - Ke^{r(t-T)})^+.$$

Bezeichne nun mit $t^* \in [t, T]$ den Ausübungszeitpunkt des amerikanischen Put.

Im Fall $K \geq S(t^*)$ erhalten wir dann

$$\pi(t^*) \leq S(t^*) + K - S(t^*) - (S(t^*) - Ke^{r(t^*-T)})^+ \leq K$$

und für $K \leq S(t^*)$ ergibt sich

$$\pi(t^*) \leq S(t^*) + 0 - (S(t^*) - Ke^{r(t^*-T)})^+ = Ke^{r(t^*-T)}.$$

Wegen $Ke^{r(t^*-T)} \leq K$ erhalten wir also die sichere obere Schranke

$$\pi(t^*) \leq K.$$

Damit folgt aus Bemerkung 1.2

$$\pi(t) \leq e^{r(t-t^*)} K \leq K$$

und damit die Behauptung.

(ii) Aus (i) folgt

$$Ke^{r(t-T)} - S(t) + V_{Ca}(t, S(t)) \leq V_{Pa}(t, S(t)) \leq K - S(t) + V_{Ca}(t, S(t)).$$

Damit und mit Satz 2.5 und Satz 2.4(i) folgt

$$\begin{aligned} V_{Pa}(t, S(t)) &\geq Ke^{r(t-T)} - S(t) + V_{Ca}(t, S(t)) \\ &= Ke^{r(t-T)} - S(t) + V_{Ce}(t, S(t)) \\ &\geq Ke^{r(t-T)} - S(t) + (S(t) - Ke^{r(t-T)})^+ = (Ke^{r(t-T)} - S(t))^+, \end{aligned}$$

also die erste Ungleichung in (ii). Mit den gleichen Sätzen folgt zudem

$$\begin{aligned} V_{Pa}(t, S(t)) &\leq V_{Pa}(t, S(t)) \leq K - S(t) + V_{Ca}(t, S(t)) \\ &= K - S(t) + V_{Ce}(t, S(t)) \leq K - S(t) + S(t) = K, \end{aligned}$$

also die zweite Ungleichung in (ii). □

2.4 Folgerungen aus der risikoneutralen Bewertung

Wir haben die risikoneutrale Bewertung in Abschnitt 2.1 durch eine geschickte Wahl der Wahrscheinlichkeiten p und $1 - p$ hergeleitet. Eine Konsequenz dieser Wahl ist, dass für den Erwartungswert des Kurses $S(T)$ die Gleichung

$$E(S(T)) = e^{rT} S(0)$$

gilt. Die Aktie verhält sich also im Mittel genau wie eine festverzinsliche Anleihe.

Nun könnte man natürlich auf die Idee kommen, in Algorithmus 2.2 ein Kursmodell zu verwenden, dessen Erwartungswert nicht wie eine festverzinsliche Anleihe wächst. Wir werden jetzt aber zeigen, dass dies einen Widerspruch zu den gerade bewiesenen Sätzen liefert. Betrachtet man die Beweise im vorhergehenden Abschnitt genauer, so stellt man zunächst fest, dass wir die risikoneutrale Bewertung dort gar nicht verwendet haben, denn alle Aussagen wurden aus sicheren Portfoliowerten zur Endzeit ermittelt. Die Sätze gelten also auch, wenn wir den Faktor $e^{r(t-T)}$ in Algorithmus 2.2(3) durch einen beliebigen anderen Faktor ersetzen.

Legen wir aber eine risikoneutrale Bewertung zu Grunde, so erhalten wir aus der ersten Ungleichung aus Satz (2.4)(i) und Algorithmus 2.2(3)

$$(S(t) - Ke^{r(t-T)})^+ \leq V_{Ce}(t, S(t)) = e^{r(t-T)} E(V_{Ce}(T, S(T))) = e^{r(t-T)} E((S(T) - K)^+).$$

Beachte dabei, dass beide Seiten der Ungleichung stetig in K sind und für alle $K > 0$ gelten sollen. Damit müssen die Ungleichungen also auch für $K = 0$ gelten, woraus wegen $S(t) \geq 0$ folgt

$$S(t) \leq e^{r(t-T)} E(S(T)).$$

Mit der zweiten Ungleichung aus Satz (2.4)(i) erhalten wir

$$e^{r(t-T)} E((S(T) - K)^+) = V_{Ce}(t, S(t)) \leq S(t)$$

und damit für $K = 0$

$$e^{r(t-T)} E(S(T)) \leq S(t).$$

Zusammen ergibt sich also

$$e^{r(t-T)} E(S(T)) = S(t) \quad \text{oder äquivalent} \quad E(S(T)) = e^{r(T-t)} S(t). \quad (2.5)$$

Die risikoneutrale Bewertung der Option verlangt also, dass auch auf $S(t)$ das Prinzip der risikoneutralen Bewertung angewendet werden muss. Oder, äquivalent formuliert, der Erwartungswert $E(S(T))$ des Basiswertes muss sich wie eine risikofreie Anleihe verhalten, d.h. mit dem gleichen Zinssatz wachsen.

Dies erscheint zunächst sehr einschränkend, denn man möchte ja eigentlich in der Lage sein, Basiswerte mit unterschiedlichen durchschnittlichen Wachstumsraten modellieren zu können. Insbesondere würden wir ja eher erwarten, dass eine Aktie, deren Kursverlauf einem gewissen Risiko unterworfen ist, im Mittel eine höhere Rendite als eine festverzinsliche Anleihe bringen sollte, gewissermaßen als Ausgleich für das eingegangene Risiko.

Der Grund dafür ist, dass wir die Wahrscheinlichkeiten p und $1 - p$ in Abschnitt 2.1 verändert haben, um die risikoneutrale Bewertungsformel zu erhalten. Das Aktienkursmodell zur Berechnung von $S(t)$ in Algorithmus 2.2 ist also eigentlich gar kein Modell eines “richtigen” Aktienkurses, sondern ein Rechenmodell, das zwar einige stochastische Charakteristika des zu Grunde liegenden Kursverlaufs besitzt — wie z.B. die Varianz, die in der Finanzmathematik zumeist *Volatilität* genannt wird — manche aber nicht — wie eben gerade den Erwartungswert.

Steigt man etwas tiefer in die stochastische Finanzmathematik ein, so sieht man, dass man dies auch eleganter interpretieren kann: man kann $S(t)$ durchaus als “richtigen” Aktienkurs verstehen, dessen Entwicklung aber mit einem modifizierten Wahrscheinlichkeitsmaß — dem sogenannten risikoneutralen Maß — gemessen wird, das gerade so gewählt wird, dass bezüglich dieses Maßes Gleichung (2.5) gilt. Für unsere numerischen Zwecke ist diese Interpretation aber nicht unbedingt notwendig. Wir können $S(t)$ in Algorithmus 2.2 einfach als ein geeignetes Rechenmodell verstehen.

Kapitel 3

Die Binomialmethode

In diesem Abschnitt beschreiben wir eine erste Methode zur Berechnung des fairen Optionspreises. Der Basiswert $S(t)$ wird dabei meist als Aktienpaket interpretiert und wir spezifizieren nun zunächst das stochastische Modell, mit dem wir aus dem derzeitigen Kurs $S(t)$ die möglichen zukünftigen Kurse $S(t')$, $t' > t$, berechnen können.

3.1 Zeitdiskrete stochastische dynamische Systeme

Das hier verwendete Modell ist zeitdiskret mit Zeitachse \mathbb{Z} , d.h. wir berechnen die möglichen Kurswerte für die zukünftigen Zeiten $S(t+1)$, $S(t+2)$, \dots , $S(T)$. Formal abstrakt ist das Modell dabei von der folgenden Form.

Definition 3.1 Ein zeitdiskretes stochastisches System ist gegeben durch eine Abbildung

$$f : \mathbb{R}^n \times \mathbb{R}^m \rightarrow \mathbb{R}^n, (S, Z) \mapsto f(S, Z).$$

Hierbei ist $S \in \mathbb{R}^n$ der Zustand und $Z \in \mathbb{R}^m$ der stochastische Einfluss. Für einen Anfangswert $S_0 \in \mathbb{R}^n$ und eine Folge von Zufallsvariablen Z_0, Z_1, \dots definieren wir die *Lösung* $S(t) = S(t, S_0)$ des Systems für $t \geq 0$ induktiv mittels

$$S(0) = S_0, \quad S(t+1) = f(S(t), Z_t). \tag{3.1}$$

□

Bemerkung 3.2 (i) Wir verwenden im Folgenden oft die Kurzschreibweise $S(t)$ und führen das zusätzliche Argument S_0 nur dann explizit auf, wenn ansonsten Uneindeutigkeiten entstünden.

(ii) Die Lösung $S(t)$ ist zu jeder Zeit $t > t_0$ eine Zufallsvariable, da sie von den Zufallsvariablen Z_{t_0}, \dots, Z_{t-1} abhängt. Für jeden Anfangswert S_0 ist $S(t, S_0)$ ein sogenannter *stochastischer Prozess*, eine zeitabhängige Zufallsvariable.

(iii) Wenn wir Realisierungen $Z_t(\omega)$ der Zufallsvariablen Z_t betrachten, erhalten wir eine Realisierung $S(t, \omega) = S(t, S_0, \omega)$. Diese Realisierung ist nun eine "normale" Funktion $t \mapsto S(t, S_0, \omega)$ von \mathbb{N}_0 nach \mathbb{R}^n und wird *Pfad* von $S(t)$ genannt.

(iv) Falls Z_0, \dots, Z_{t-1} und Z_t unabhängige Zufallsvariablen sind, sind auch $S(t)$ und Z_t unabhängig, da $S(t)$ nur von Z_0, \dots, Z_{t-1} abhängt.

(v) Die genaue Skalierung der Zeiteinheit “ t ” ist variabel. Die Zeit $t = 1$ könnte ein Jahr, einen Monat oder einen Tag bedeuten. Wichtig ist aber, dass die Parameter des Modells an die Zeitskala angepasst werden. Für den jährlichen Zinssatz r bedeutet dies z.B., dass sich die Zinsen im Intervall $[0, T]$ bei einer Zeitskala in Jahren zu $e^{rT} - 1$, bei einer Zeitskala in Tagen hingegen zu $e^{\frac{r}{360}T} - 1$ ergeben (wenn das Jahr wie üblich in 12 Monate mit je 30 Tagen eingeteilt wird). Um die Notation einfach zu halten, verwenden wir im Folgenden stets das Symbol “ r ” für den Zinssatz, der zu der verwendeten Zeiteinheit passt, so dass die Zinsen in einer Zeiteinheit immer als $e^r - 1$ und die Vermehrung des Kapitals immer als e^r geschrieben werden können. \square

3.2 Das Binomialmodell eines Aktienkurses

Das Modell, das der Binomialmethode seinen Namen gibt, ist das sogenannte *Binomialmodell* eines Aktienkurses. Ausgehend von dem Anfangskurs S_0 kann der Kurs dabei in jedem Zeitschritt entweder mit Wahrscheinlichkeit p um einen Faktor $\alpha_u > 1$ steigen (“up”) oder mit Wahrscheinlichkeit $1 - p$ um einen Faktor $\alpha_d < 1$ fallen (“down”). Hierbei soll die Bewegung im Zeitschritt $t + 1$ jeweils stochastisch unabhängig von der Bewegung im vorhergehenden Zeitschritt t sein.

Zur Modellierung als zeitdiskretes stochastisches System betrachten wir daher unabhängige Zufallsvariablen $Z_t : \Omega \rightarrow \{\alpha_u, \alpha_d\}$ mit

$$\mathbb{P}_{Z_t}(\{\alpha_u\}) = p \quad \text{und} \quad \mathbb{P}_{Z_t}(\{\alpha_d\}) = 1 - p$$

und definieren

$$f(S, Z) = S \cdot Z.$$

Die Lösung ergibt sich damit für $t \geq 1$ zu

$$S(t) = Z_{t-1}Z_{t-2} \dots Z_0 S_0.$$

Abbildung 3.1 zeigt einige Beispielpfade dieses Modells für $\alpha_u = 1.03$, $\alpha_d = 0.98$, $p = 1/2$ und $S_0 = 1$.

Das Binomialmodell hat eine wesentliche Eigenschaft, die es für die numerische Optionsbewertung günstig macht: Die Menge der möglichen zukünftigen Zustände $S(t)$ nimmt für wachsendes t relativ langsam zu:

Auf den ersten Blick sieht es so aus, als wenn sich die Anzahl der möglichen Zustände in jedem Schritt verdoppelt. Beginnen wir in $t = 0$ mit S_0 , so erhalten wir in $t = 1$ die zwei möglichen Zustände $\{\alpha_u S_0, \alpha_d S_0\}$, in $t = 2$ die vier Zustände $\{\alpha_u^2 S_0, \alpha_d \alpha_u S_0, \alpha_u \alpha_d S_0, \alpha_d^2 S_0\}$ usw. Da nun aber offensichtlich $\alpha_d \alpha_u = \alpha_u \alpha_d$ gilt, fallen hier zwei der Zustände zusammen. Ähnliches passiert in allen nachfolgenden Zeitschritten — man sagt, das Modell ist *rekombinierbar*. Daher lässt sich per Induktion leicht beweisen, dass es zum Zeitpunkt t gerade $t + 1$ verschiedene Zustände gibt. Abbildung 3.2 illustriert die entstehende Baumstruktur (mit den Wahrscheinlichkeiten jedes Teilpfades), die auch *Binomialbaum* genannt wird.

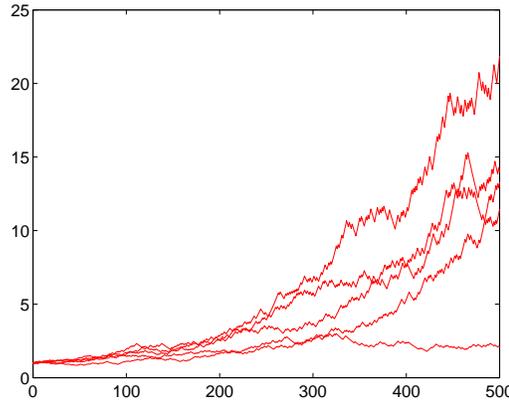


Abbildung 3.1: Beispielpfade des Binomialmodells

Die Bedeutung der Parameter p , α_u und α_d betrachten wir im übernächsten Abschnitt noch etwas genauer. Zunächst wollen wir aber erläutern, wie man mit Hilfe dieses Binomialmodells den Wert einer Option berechnet.

3.3 Optionswertberechnung mit dem Binomialmodell

Mit Hilfe des Binomialmodells können wir nun den Schritt (3) von Algorithmus 2.2 numerisch umsetzen. Wir beginnen mit europäischen Optionen und stellen dazu eine Methode vor, die iterativ vorgeht. Diese lässt sich danach leicht auf amerikanische Optionen verallgemeinern.

Wir bezeichnen die möglichen Zustände des Binomialmodells mit $S_k(t)$, $k = 0, \dots, t$, wobei wir aufsteigend nummerieren. Dann gilt für die $S(t)$ -Werte aus unserem Modell gerade $S(0) = S_0 = S_0(0)$ und

$$\begin{aligned} \mathbb{P}_{S(t+1)}\left(S_{k+1}(t+1) \mid S(t) = S_k(t)\right) &= p, \\ \mathbb{P}_{S(t+1)}\left(S_k(t+1) \mid S(t) = S_k(t)\right) &= 1 - p. \end{aligned}$$

Damit erhalten wir das folgende Lemma.

Lemma 3.3 Wenn der Basiswert $S(t)$ einer europäischen Option sich gemäß dem Binomialmodell verhält, so gilt für alle $t \in [0, T - 1]$ für den Wert $V(t, S)$ bei risikoneutraler Bewertung die Gleichung

$$V(t, S_k(t)) = e^{-r} \left(pV(t+1, S_{k+1}(t+1)) + (1-p)V(t+1, S_k(t+1)) \right). \quad (3.2)$$

Beweis: Bei risikoneutraler Bewertung gilt

$$V(t, S^1(t)) = e^{r(t-T)} E\left(V(T, S(T)) \mid S(t) = S^1(t)\right)$$

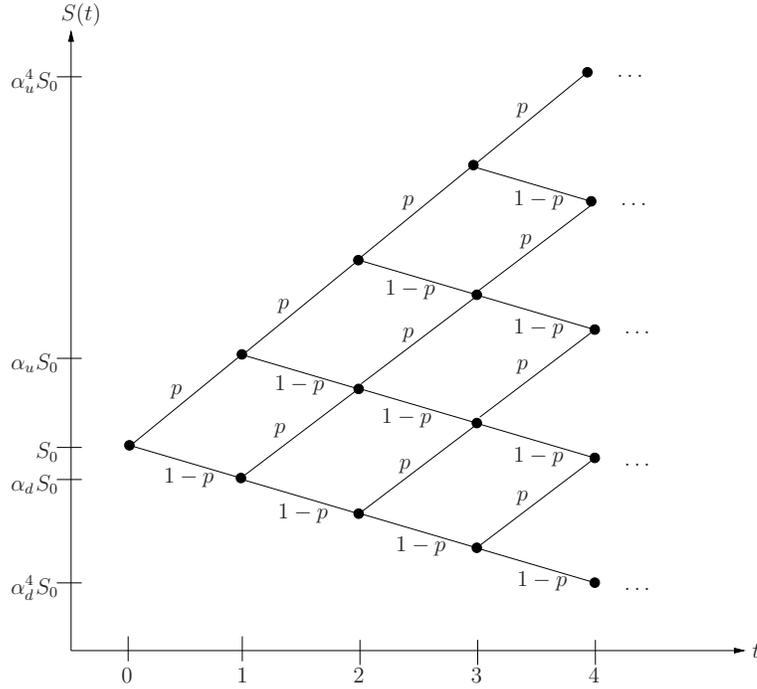


Abbildung 3.2: Binomialbaum: Mögliche Zustände im Binomialmodell

und

$$V(t+1, S^2(t+1)) = e^{r(t+1-T)} E\left(V(T, S(T)) \mid S(t+1) = S^2(t+1)\right),$$

wobei S^1 , S^2 und S dem Binomialmodell folgen. Aus der zweiten Gleichung folgt dabei durch Anwendung des bedingten Erwartungswertes $E(\cdot \mid S^2(t) = S^1(t))$ die Gleichung

$$\begin{aligned} & E\left(V(t+1, S^2(t+1)) \mid S^2(t) = S^1(t)\right) \\ &= e^{r(t+1-T)} E\left(E\left(V(T, S(T)) \mid S(t+1) = S^2(t+1)\right) \mid S^2(t) = S^1(t)\right) \\ &= e^{r(t+1-T)} E\left(V(T, S(T)) \mid S(t) = S^1(t)\right) \end{aligned}$$

und damit

$$V(t, S^1(t)) = e^{-r} E\left(V(t+1, S^2(t+1)) \mid S^2(t) = S^1(t)\right).$$

Nach den obigen Wahrscheinlichkeiten $\mathbb{P}_{S(t+1)}$ gilt für $S^2(t) = S^1(t) = S_k(t)$ gerade $S^2(t+1) = S_k(t+1)$ mit Wahrscheinlichkeit $1-p$ und $S^2(t+1) = S_{k+1}(t+1)$ mit Wahrscheinlichkeit p . Damit folgt die angegebene Formel. \square

Mit Hilfe der Formel (3.2) können wir nun den Optionswert $V(0, S_0)$ wie folgt berechnen.

Algorithmus 3.4 (Iterative Bewertung einer europäischen Option mit der Binomialmethode)

Eingabe: $S_0 \in \mathbb{R}^+$, $K \in \mathbb{R}^+$, $T \in \mathbb{N}$, $p \in (0, 1)$, $\alpha_u > \alpha_d \in \mathbb{R}_+$

- (1) Für i von 0 bis T
 - setze $S_i = S_0 \alpha_u^i \alpha_d^{T-i}$
 - setze $V_i := (K - S_i)^+$ für einen Put bzw. $V_i := (S_i - K)^+$ für einen Call
 Ende der Schleife
- (2) Für t von $T - 1$ bis 0
 - Für i von 0 bis t
 - setze $V_i := e^{-r}(pV_{i+1} + (1-p)V_i)$
 Ende der Schleifen

Ausgabe: $V_0 = V(0, S_0)$ □

Dieser Algorithmus lässt sich nun leicht auf amerikanische Put-Optionen verallgemeinern. Hierzu maximieren wir in jedem Zeitschritt über den bei der vorzeitigen Ausübung erzielbaren und den durch Halten der Option für (mindestens) einen weiteren Zeitschritt erzielbaren Wert. Beachte, dass wir hier nicht kontinuierlich prüfen, ob die vorzeitige Ausübung sinnvoll ist sondern nur zu den durch das Binomialmodell vorgegebenen diskreten Zeitpunkten.

Algorithmus 3.5 (Iterative Bewertung einer amerikanischen Put-Option mit der Binomialmethode)

Eingabe: $S_0 \in \mathbb{R}^+$, $K \in \mathbb{R}^+$, $T \in \mathbb{N}$, $p \in (0, 1)$, $\alpha_u > \alpha_d \in \mathbb{R}_+$

- (1) Für i von 0 bis T
 - setze $S_i = S_0 \alpha_u^i \alpha_d^{T-i}$
 - setze $V_i := (K - S_i)^+$
 Ende der Schleife
- (2) Für t von $T - 1$ bis 0
 - Für i von 0 bis t
 - setze $S_i := S_0 \alpha_u^i \alpha_d^{t-i}$
 - setze $V_i := \max\{e^{-r}(pV_{i+1} + (1-p)V_i), K - S_i\}$
 Ende der Schleifen

Ausgabe: $V_0 = V_{Pa}(0, S_0)$ □

3.4 Die Parameter im Binomialmodell

Um eine reale Kursentwicklung brauchbar stochastisch zu simulieren, müssen die Parameter p , α_u und α_d geeignet angepasst werden. Das absolute Minimum an stochastischen Werten, das ein Aktienkursmodell widerspiegeln sollte, sind der Mittelwert des Kurses und seine Varianz, im Zusammenhang mit Kursen oft Volatilität genannt. Da wir es bei den Aktienkursmodellen mit stochastischen Prozessen zu tun haben, sind diese Werte hier nicht wie bei einer einfachen Zufallsvariablen statisch sondern abhängig von der Zeit und vom Anfangswert, d.h. wir erhalten

$$E(S(t, S_0)) \quad \text{und} \quad Var(S(t, S_0))$$

abhängig von t und S_0 . Aus Formel (2.5) folgt dabei, dass für unsere risikoneutrale Bewertung die Gleichung

$$E(S(t+1)) = e^r S(t) \quad (3.3)$$

gelten muss.

Im Binomialmodell gibt es nun gerade die drei Parameter p , α_u und α_d , die zunächst so angepasst werden, dass diese Bedingung erfüllt ist. Wegen

$$E(S(t+1)) = (p\alpha_u + (1-p)\alpha_d)S(t)$$

folgt aus (3.3) die Gleichung

$$p\alpha_u + (1-p)\alpha_d = e^r.$$

Dadurch ist einer der drei Parameter bereits eindeutig bestimmt, wenn die beiden anderen Parameter festgelegt sind. Die beiden anderen Parameter können nun auf unterschiedliche Weisen berechnet werden. Zum einen können sie so gewählt werden, dass die Varianz (und gegebenenfalls weitere stochastische Kenngrößen) des Modells mit den vergangenen Kursdaten gut übereinstimmen — man nennt dies *historische Parameter*. Die Hoffnung ist natürlich, dass die aus der Vergangenheit ermittelten Werte einen halbwegs zuverlässigen stochastischen Blick in die Zukunft erlauben. Eine andere Methode der Parameterbestimmung besteht darin, die Parameter so zu bestimmen, dass die ermittelten fairen Preise mit den beobachteten Marktpreisen für die Optionen übereinstimmen — man spricht von *impliziten Parametern*. Auf das Problem der Parameterbestimmung kommen wir später genauer zurück.

Oft wird das Binomialmodell als Näherung der geometrischen Brownschen Bewegung angesehen, die dem Black-Scholes-Modell zu Grunde liegt. Wie die Parameter in diesem Fall gewählt werden müssen, werden wir in Abschnitt 4.2.2 genauer besprechen.

Kapitel 4

Stochastische Differentialgleichungen

Um eine stochastische Differentialgleichung zu formulieren, müssen wir zunächst erklären, auf welche Weise der “Zufall” in unsere Differentialgleichung eingeht.

Vereinfacht kann man eine stochastische Differentialgleichung zunächst als eine “übliche” gewöhnliche Differentialgleichung mit der speziellen Struktur

$$\dot{x}(t) = f(t, x(t)) := a(t, x(t)) + b(t, x(t))g(t) \quad (4.1)$$

schreiben¹. Hierbei sind $a, b : \mathbb{R} \times \mathbb{R}^n \rightarrow \mathbb{R}^n$ gewöhnliche Funktionen, der Zufall kommt durch die Funktion g ins Spiel, die nämlich keine “feste” Funktion ist, sondern eine Funktion, die abhängig vom Zufall verschiedene Gestalt annehmen kann. Für jedes mögliche g erhalten wir für einen festen Anfangswert weiterhin eine eindeutige Lösung. Da diese aber nun von g abhängt und g zufällig ist, ist in diesem Sinne auch die Lösungstrajektorie zufällig.

Natürlich ist g nicht “irgendwie” vom Zufall abhängig, sondern genügt gewissen statistischen Regeln. Um diese zu definieren, werden wir zuerst einen grundlegenden stochastischen Prozess einführen. Dabei spielt die Gauß- oder Normal-Verteilung eine wichtige Rolle. Wir erinnern hier daran, dass eine Zufallsvariable X normalverteilt mit Erwartungswert μ und Varianz σ^2 ist (Notation: $X \sim N(\mu, \sigma^2)$), falls sie die Dichtefunktion

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

besitzt.

4.1 Der Wiener Prozess

Zur Definition der stochastischen Differentialgleichung benötigen wir nun einen Begriff für zufällige Funktionen. Eine “zufällige Funktion” ist dabei nichts anderes als eine Funktion $X : \mathbb{R} \times \Omega \rightarrow \mathbb{R}$, die für jedes feste t gerade eine Zufallsvariable $X(t, \cdot) : \Omega \rightarrow \mathbb{R}$ ist,

¹Dies ist eine vorläufige Schreibweise, siehe (4.5) für die mathematisch gebräuchliche Formulierung.

wobei für alle t der gleiche Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, \mathbb{P})$ zu Grunde liegt. Solch eine Funktion nennt man *stochastischen Prozess*. Beachte, dass wir keinerlei Annahmen über den Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, \mathbb{P})$ voraussetzen; wichtig ist nur, dass alle Zufallsvariablen X auf dem selben Raum definiert sind. Für jedes feste $\omega \in \Omega$ ist die Realisierung $X(\cdot, \omega) : \mathbb{R} \rightarrow \mathbb{R}$ eine “normale” reelle Funktion. Wie bereits im vorhergehenden Kapitel erwähnt, nennt man die Realisierung bei stochastischen Prozessen *Pfad*.

Der stochastische Prozess, den wir hier näher betrachten wollen und aus dem das obige g später hervorgehen wird, ist der sogenannte *Wiener Prozess*, der üblicherweise mit W bezeichnet wird und für $t \geq 0$ definiert ist. Die Definition des Wiener Prozesses ergibt sich aus der von N. Wiener² eingeführten mathematischen Beschreibung der *Brownschen Bewegung*, die in der Physik die zufällige Bewegung eines auf einer Wasseroberfläche schwimmenden Teilchens beschreibt. Formal verlangt man die folgenden Bedingungen:

- (i) $W(t)$ ist eine Gauß-verteilte Zufallsvariable mit $E(W(t)) = 0$ und $\text{Var}(W(t)) = t$
- (ii) Für $t_1 \geq t_0 \geq 0$ sind die *Inkmente* $W(t_1) - W(t_0)$ Gauß-verteilte Zufallsvariablen mit $E(W(t_1) - W(t_0)) = 0$ und $\text{Var}(W(t_1) - W(t_0)) = t_1 - t_0$
- (iii) Für $s_1 \geq s_0 \geq t_1 \geq t_0 \geq 0$ sind die Inkmente $W(t_1) - W(t_0)$ und $W(s_1) - W(s_0)$ unabhängige Zufallsvariablen.

Beachte, dass wir oben die Gauß-Verteilung nur für $\sigma > 0$ definiert haben. Hier erhalten wir für $t = 0$ die Bedingung $\text{Var}(W(0)) = 0$, womit einfach $\mathbb{P}_{W(0)}(\mu) = 1$ gemeint ist. Mit anderen Worten nimmt die Zufallsvariable $W(0)$ also fast sicher ihren Erwartungswert μ (hier $\mu = 0$) an.

Ein Pfad $W(t, \omega)$ von W ist also gerade eine der vielen möglichen zufälligen Funktionen, die (in ihrer Gesamtheit) den obigen Bedingungen genügen. Für diese Pfade kann man beweisen, dass sie fast sicher stetig in t sind (zur Erinnerung: dies bedeutet, dass das Ereignis $A = \{\omega \in \Omega \mid X(t, \omega) \text{ ist stetig in } t\}$ die Wahrscheinlichkeit $\mathbb{P}(A) = 1$ besitzt) aber auch fast sicher nirgends differenzierbar. Abbildung 4.1 zeigt drei verschiedene Pfade des Wiener Prozesses.

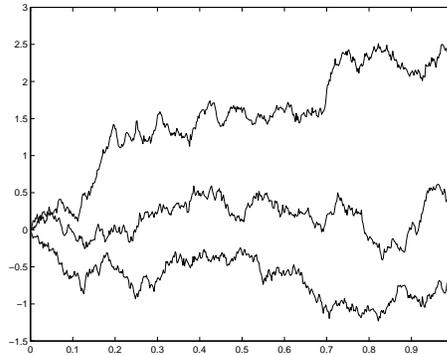


Abbildung 4.1: Verschiedene Pfade des Wiener Prozesses

²US-amerikanischer Mathematiker, 1894–1964

Die Bedingung (i) besagt gerade, dass die Streuung der Werte der Pfade $W(t, \omega)$ um so größer wird, je größer t wird. Der Mittelwert ist dabei zu jedem Zeitpunkt gleich 0. Die Bedingung (ii) bedeutet, dass der stochastische Prozess $\widetilde{W}(t) = W(t + t_0) - W(t_0)$ wieder ein Wiener Prozess ist. Wenn wir also ein “Endstück” eines Pfades abschneiden und dessen Anfangspunkt in den Nullpunkt verschieben, so erhalten wir wieder einen Pfad des Wiener Prozesses. Diese zwei Eigenschaften kann man anhand der Grafik gut nachvollziehen. Bedingung (iii) schließlich besagt, dass man aus der Kenntnis eines Teilstücks eines Pfades auf einem Teilintervall $[t_0, t_1]$ keinerlei Vorhersagen über den Verlauf auf einem anderen Teilintervall $[s_0, s_1]$ mit $s_0 \geq t_1$ machen kann, wenn man nur die Informationen über die Anfangswerte durch Abziehen der Werte $W(s_0)$ und $W(t_0)$ “löscht”. Der Pfad kann sich also zu jedem Zeitpunkt mit exakt der gleichen Wahrscheinlichkeit nach oben oder unten bewegen, egal welchen Verlauf er bis zu diesem Zeitpunkt genommen hat.

4.2 Das Itô-Integral

Betrachtet man die Pfade in Abbildung 4.1, so kann man durchaus eine vage Ähnlichkeit mit den Kursdiagrammen z.B. von Aktienkursen erkennen. Selbst für eine sehr grobe Modellierung von solchen Kurswerten ist der Wiener Prozess aber doch noch zu einfach, denn es gibt keine Parameter, an denen man “drehen” könnte, um den modellierten Verlauf einem echten Kursverlauf anzupassen: Der Wiener Prozess besitzt noch zu wenig Struktur.

Er ist aber bestens geeignet, um als “Zutat” in der Definition stochastischer Differentialgleichungen eine wichtige Rolle zu spielen; vereinfachend gesagt soll die zufällige Funktion g in (4.1) gerade die Ableitung des Wiener Prozesses sein. Um die dabei auftretenden mathematischen Probleme und ihre Lösung zu erläutern, wollen wir hier zunächst die denkbar einfachste stochastische Differentialgleichung betrachten und dann zur allgemeinen Formulierung übergehen.

Zunächst einige Anmerkungen zur Notation: Da die Lösung einer SDG über den eingehenden Wiener Prozess wieder eine zufällige Funktion — also ein stochastischer Prozess — ist, verwenden wir hier für den gesuchten unbekannt Prozess die groß geschriebene Bezeichnung $X(t)$, bzw. mit Anfangswert $X_0 \in \mathbb{R}^n$ und Anfangszeit $t_0 \in \mathbb{R}$ die Schreibweise $X(t; t_0, X_0)$. Wir erlauben hierbei, dass X vektorwertig, also aus dem \mathbb{R}^n ist, was einfach bedeutet, dass $X = (X_1, X_2, \dots, X_n)^T$ ist, wobei die X_i reellwertige stochastische Prozesse im oben eingeführten Sinne sind. Zu jedem Pfad $W(t, \omega)$ des eingehenden Wiener Prozesses gehört dann ein Lösungspfad des X -Prozesses, den wir mit $X(t; t_0, X_0, \omega)$ bezeichnen.

Die technische Hauptschwierigkeit in der mathematischen Formulierung stochastischer Differentialgleichungen zeigt sich nun bereits bei einer scheinbar trivialen Aufgabe, nämlich dem Problem, eine stochastische Differentialgleichung aufzustellen, deren Lösung gerade der Wiener Prozess ist. Scheinbar trivial ist die Aufgabe deswegen, weil wir ja den Wiener Prozess als gegeben voraussetzen und in der Formulierung verwenden dürfen, weswegen es nahe liegt, einfach die Differentialgleichung

$$\frac{d}{dt}X(t) = \frac{d}{dt}W(t) \quad (4.2)$$

mit Anfangsbedingung $X_0 = W(0)$ zur Anfangszeit $t_0 = 0$ zu verwenden. Das Problem ist jetzt aber: Was verstehen wir unter “ $\frac{d}{dt}W(t)$ ”? Man würde vielleicht versuchen, die

Ableitung pfadweise aufzufassen, d.h., wir berechnen die Ableitung für jeden Pfad $W(t, \omega)$. Nur ist ein typischer Pfad $W(t, \omega)$, wie oben erwähnt, nirgends differenzierbar.

Wie bei einer deterministischen gewöhnlichen Differentialgleichung, können wir nun Gleichung (4.2) zunächst in Integralform

$$X(t) = X_0 + \int_0^t \frac{d}{d\tau} W(\tau) d\tau$$

schreiben. Jetzt könnten wir formal integrieren, was uns aber bei der Frage “was ist $\frac{d}{dt} W(t)$?” nicht weiter bringt. Für das obige sogenannte *stochastische Integral* hat sich in der Literatur die kürzere Schreibweise

$$\int_0^t dW_\tau$$

eingebürgert, die wir hier übernehmen wollen. Dies zeigt die Richtung auf, die wir zur Lösung unseres Problems einschlagen wollen: Anstatt die Ableitung $\frac{d}{dt} W(t)$ zu betrachten, werden wir versuchen, diesem stochastischen Integral eine mathematische Definition zu geben, die

- (i) wohldefiniert ist, obwohl $\frac{d}{dt} W(t)$ nicht existiert
- (ii) das gewünschte Ergebnis, nämlich $X(t) = W(t)$, liefert
- (iii) sich auf allgemeinere Integrale der Form

$$I(F) := \int_{t_0}^{t_1} F(t) dW_t \tag{4.3}$$

verallgemeinern lässt, damit wir auch kompliziertere SDGs formulieren können. Hierbei ist F wiederum ein stochastischer Prozess, der auf dem selben Wahrscheinlichkeitsraum wie W definiert ist.

Wir wollen dieses Konzept nun für Integrale der Form (4.3) angeben. Die hier vorgestellte Lösung geht auf Kiyosi Itô³ zurück und wurde in den 1940er Jahren entwickelt. Die Idee besteht darin, das Integral (4.3) für jedes Paar von Pfaden $F(t, \omega)$ und $W(t, \omega)$ durch den Limes einer geeigneten Summe zu approximieren. Wir wählen dazu ein $N \in \mathbb{N}$ und eine Folge von Zeiten $\tau_i^{(N)}$, $i = 0, 1, \dots, N$ mit

$$t_0 = \tau_0^{(N)} < \tau_1^{(N)} < \dots < \tau_N^{(N)} = t_1$$

und definieren für jedes $\omega \in \Omega$

$$I^{(N)}(F)(\omega) := \sum_{i=0}^{N-1} F(\tau_i^{(N)}, \omega) (W(\tau_{i+1}^{(N)}, \omega) - W(\tau_i^{(N)}, \omega)).$$

³japanischer Mathematiker, 1915–2008

Das Integral (4.3) wird nun über den Limes dieser Summe definiert. Betrachte eine Familie von Folgen $\tau_i^{(N)}$ für $N \in \mathbb{N}$ mit $\lim_{N \rightarrow \infty} \delta(N) = 0$, wobei $\delta(N) := \max_{i=1, \dots, N} \tau_i^{(N)} - \tau_{i-1}^{(N)}$ ist. Dann definieren wir

$$I(F) := \lim_{N \rightarrow \infty} I^{(N)}(F). \quad (4.4)$$

Diese Definition wirft zunächst eine Reihe von Fragen auf, denn da die Pfade des Wiener Prozesses sehr unangenehme Funktionen sein können, ist nicht garantiert, dass dieser Limes für jedes ω überhaupt existiert. Tatsächlich lag der Haupttrick von Itô darin, zu definieren, was der Limes in (4.4) eigentlich bedeuten soll. Dieser Limes ist nämlich *nicht* pfadweise zu verstehen (in dem Sinne, dass wir $\lim_{N \rightarrow \infty} I^{(N)}(F)(\omega)$ für jedes $\omega \in \Omega$ bilden), sondern man muss die Werte $I^{(N)}(F)$ ebenso wie das Integral $I(F)$ wieder als Zufallsvariablen $I^{(N)}(F) : \Omega \rightarrow \mathbb{R}$ bzw. $I(F) : \Omega \rightarrow \mathbb{R}$ auffassen. Für Zufallsvariablen gibt es verschiedene Konvergenzbegriffe und der hier geeignete ist der Begriff der *Quadrat-Mittel-Konvergenz*, der wie folgt definiert ist: Eine Folge von Zufallsvariablen $X_N : \Omega \rightarrow \mathbb{R}$ konvergiert im Quadrat-Mittel-Sinne gegen eine Zufallsvariable $X : \Omega \rightarrow \mathbb{R}$, falls

$$\lim_{N \rightarrow \infty} E(|X_N - X|^2) = 0$$

gilt. Mit diesem Konvergenzbegriff kann man zeigen, dass die Folge $I^{(N)}(F)$ (unter geeigneten Bedingungen an F) tatsächlich konvergiert und (4.4) also wohldefiniert ist. Das resultierende Integral wird *Itô-Integral* genannt und es besitzt tatsächlich die oben aufgeführten gewünschten Eigenschaften (i)–(iii).

Mit Hilfe des Itô-Integrals können wir die informelle Schreibweise (4.1) mathematisch präzise formulieren. Statt der üblichen Differentialgleichungsschreibweise schreibt man *Itô-stochastische Differentialgleichungen* nämlich als

$$dX(t) = a(t, X(t))dt + b(t, X(t))dW_t. \quad (4.5)$$

Dies ist nur eine symbolische Schreibweise; was mit (4.5) tatsächlich gemeint ist, ist die längere Integralschreibweise

$$X(t) = X(t_0) + \int_{t_0}^t a(\tau, X(\tau))d\tau + \int_{t_0}^t b(\tau, X(\tau))dW_\tau,$$

bei der das zweite Integral gerade das Itô-Integral ist. Dies liefert eine mathematisch fundierte und brauchbare Definition stochastischer Differentialgleichungen. Falls $b(t, x) \equiv 0$ ist, also kein stochastischer Anteil vorhanden ist, reduziert sich (4.5) auf

$$X(t) = X(t_0) + \int_{t_0}^t a(\tau, X(\tau))dt \iff \frac{d}{dt}X(t) = a(t, X(t)),$$

also auf die wohlbekannte deterministische gewöhnliche Differentialgleichung. Der deterministische Anteil der Gleichung $a(t, X(t))$ wird oft *Drift* genannt, der stochastische Anteil $b(t, X(t))$ oft als *Diffusion* bezeichnet.

Natürlich lässt sich (4.5) in vielfacher Hinsicht erweitern, z.B. kann man statt nur einem W mehrere unabhängige Wiener Prozesse W^1, \dots, W^m eingehen lassen, was zur Gleichung

$$dX(t) = a(t, X(t))dt + \sum_{j=1}^m b_j(t, X(t))dW_t^j$$

führt.

Für das Itô-Integral lassen sich eine Reihe von Eigenschaften beweisen; auf die meisten können wir hier nicht näher eingehen. In Lemma 4.2 werden wir eine wichtige Eigenschaft kennen lernen, eine weitere Eigenschaft ist die Tatsache, dass

$$E \left(\int_{t_0}^{t_1} F(t) dW_t \right) = 0 \quad (4.6)$$

ist, falls die Zufallsvariablen $F(t)$ und $W(s) - W(t)$ unabhängig sind für $s > t \geq 0$ (dies gilt immer, falls F die Lösung einer SDG bzgl. W ist). Eigenschaft (4.6) gilt zunächst für die Approximationen $I^{(N)}(F)$, weil nach Annahme an W und wegen der Unabhängigkeit⁴ die Gleichung

$$E(F(\tau_i)(W(\tau_{i+1}) - W(\tau_i))) = E(F(\tau_i)) \underbrace{E(W(\tau_{i+1}) - W(\tau_i))}_{=0} = 0 \quad (4.8)$$

gilt, und überträgt sich durch die Limesbildung auf $I(F)$.

Bemerkung 4.1 Es sollte hier erwähnt werden, dass es eine weitere sinnvolle stochastische Integraldefinition gibt, die auf R. Stratonovich⁵ zurück geht. Das Stratonovich-Integral

$$\int_{t_0}^{t_1} F(t) \circ dW_t$$

wird über eine ähnliche Limes-Bildung wie das Itô-Integral definiert und liefert ebenfalls eine mathematisch fundierte Definition stochastischer DGLs. Die beiden Integrale unterscheiden sich allerdings in den Rechenregeln ebenso wie in der Form der Lösungen. Die zugehörigen *Stratonovich-SDGs* werden in der Form

$$dX(t) = a(t, X(t))dt + b(t, X(t)) \circ dW_t$$

geschrieben. Aus Zeitgründen können wir auf diese zweite Definition und auf die Gemeinsamkeiten und Unterschiede zum Itô-Integral hier nicht näher eingehen. \square

4.2.1 Das Itô-Lemma

Zur Bewertung von Optionen, aber auch, um ganz allgemein mit dem Itô-Integral zu rechnen, benötigt man geeignete Rechenregeln. Falls $x(t)$ eine Lösung einer eindimensionalen deterministischen DGL $\dot{x}(t) = f(t, x(t))$ ist, so rechnet man für eine beliebige Funktion $g : \mathbb{R} \times \mathbb{R} \rightarrow \mathbb{R}$ mittels der Kettenregel leicht nach, dass die Funktion $y(t) = g(t, x(t))$ die Gleichung

$$\dot{y}(t) = \frac{\partial g}{\partial t}(t, x(t)) + \frac{\partial g}{\partial x}(t, x(t))f(t, x(t))$$

⁴Für unabhängige Zufallsvariablen X_1, \dots, X_k gilt die Gleichung

$$E(X_1 \cdots X_k) = E(X_1) \cdots E(X_k). \quad (4.7)$$

⁵russischer Mathematiker, 1930–1997

erfüllt.

Das Itô-Lemma liefert nun eine Verallgemeinerung dieser Gleichung für die Lösung Itô-stochastischer DGL, und wird daher oft auch als stochastische Verallgemeinerung der Kettenregel bezeichnet.

Lemma 4.2 Sei $g : \mathbb{R} \times \mathbb{R} \rightarrow \mathbb{R}$ eine zwei mal stetig differenzierbare Funktion und sei $X(t)$ die Lösung einer reellwertigen Itô-SDG vom Typ (4.5). Dann erfüllt $Y(t) = g(t, X(t))$ die Gleichung

$$\begin{aligned} dY(t) &= \left(\frac{\partial g}{\partial t}(t, X(t)) + \frac{\partial g}{\partial x}(t, X(t))a(t, X(t)) + \frac{1}{2} \frac{\partial^2 g}{\partial x^2} b(t, X(t))^2 \right) dt \\ &\quad + \frac{\partial g}{\partial x}(t, X(t))b(t, X(t))dW_t, \end{aligned}$$

wobei W hier gerade der Wiener Prozess aus der SDG ist, die $X(t)$ erfüllt. Diese Formel wird auch *Itô-Formel* genannt.

Beweisidee: Wir werden hier keinen vollständigen Beweis betrachten, wollen aber begründen, warum die (im Vergleich zur deterministischen Formel ungewöhnliche) zweite Ableitung von g nach x hier auftritt. Wir betrachten dazu zwei Zeitpunkte t und $t + \Delta t$. Dann gilt $X(t + \Delta t) = X(t) + \Delta X(t)$ mit

$$\begin{aligned} \Delta X(t) &= \int_t^{t+\Delta t} a(s, X(s))ds + \int_t^{t+\Delta t} b(s, X(s))dW_s \\ &\approx a(t, X(t))\Delta t + b(t, X(t))\Delta W_t, \end{aligned} \tag{4.9}$$

wobei $\Delta W_t = W(t + \Delta t) - W(t)$ ist. Die Approximation (4.9) folgt hierbei aus der Limes-Definition des Itô-Integrals, wenn wir $\tau_{i+1} - \tau_i = \Delta t$ setzen. Wir betrachten nun die Größe

$$\Delta Y := g(t + \Delta t, X(t) + \Delta X(t)) - g(t, X(t)).$$

Aus der Taylor-Entwicklung von g folgt

$$\begin{aligned} \Delta Y &\approx \frac{\partial g}{\partial x} \Delta X + \frac{\partial g}{\partial t} \Delta t + \frac{1}{2} \frac{\partial^2 g}{\partial x^2} (\Delta X)^2 \\ &\approx \frac{\partial g}{\partial x} \left(a(t, X(t))\Delta t + b(t, X(t))\Delta W_t \right) + \frac{\partial g}{\partial t} \Delta t \\ &\quad + \frac{1}{2} \frac{\partial^2 g}{\partial x^2} \left(a(t, X(t))^2 (\Delta t)^2 + 2a(t, X(t))b(t, X(t))\Delta t \Delta W_t + (b(t, X(t))\Delta W_t)^2 \right), \end{aligned}$$

wobei alle Ableitungen in $(t, X(t))$ ausgewertet werden. Um aus dieser Approximation die ‘‘Ableitung’’

$$dY(t) = \lim_{\Delta t \rightarrow 0} \Delta Y$$

zu berechnen, müssen wir alle Terme der Ordnung $O(\Delta t)$ berücksichtigen. Wären alle Größen deterministisch, so blieben hier gerade die Terme mit den ersten Ableitungen von

g stehen. Im stochastischen Fall ist aber auch $(\Delta W_t)^2$ ein Term der Ordnung $O(\Delta t)$, denn es gilt

$$E((\Delta W_t)^2) = E(\underbrace{(E(\Delta W_t) - \Delta W_t)}_{=0})^2 = \text{Var}(\Delta W_t) = \text{Var}(W(t + \Delta t) - W(t)) = \Delta t.$$

Der Term

$$\frac{1}{2} \frac{\partial^2 g}{\partial x^2} b(t, X(t))^2 (\Delta W_t)^2$$

muss also mit berücksichtigt werden. Führt man nun den Grenzübergang für $\Delta t \rightarrow 0$ im richtigen stochastischen Sinne durch, so erhält man gerade die behauptete Formel. \square

Bemerkung 4.3 Es existiert auch eine Version des Itô-Lemmas für mehrdimensionale SDGs. In diesem Fall gehen — durch die im Beweis verwendete Taylor-Entwicklung für g — alle gemischten zweiten Ableitungen der Form $\frac{\partial^2 g}{\partial x_i \partial x_j}$ in die Gleichung ein. \square

4.2.2 Die geometrische Brownsche Bewegung

Zum Abschluss dieses Kapitels stellen wir hier ein einfaches aber sehr verbreitetes SDG-Modell zur Modellierung von Kursverläufen vor. Wie in den vergangenen Kapiteln schreiben wir hier nun wieder $S(t)$ statt $X(t)$.

Wie bereits in Abschnitt 3.4 erwähnt, möchte man in dieser Modellierung mindestens den Mittelwert und die Varianz des Kurses bestimmen. In der geometrischen Brownschen Bewegung stehen dazu die zwei Parameter μ und σ zur Verfügung, wobei die *Rücklaufquote* oder auch *Rendite* μ das mittlere Wachstum oder den “Trend” des Kurses modelliert und die *Volatilität* σ die Varianz, die hier von der aktuellen Höhe des Kurses abhängt.

Das zugehörige SDG-Modell ist gegeben durch

$$dS(t) = \mu S(t) dt + \sigma S(t) dW_t. \quad (4.10)$$

Trotz seiner Einfachheit ist dieses Modell Grundlage vieler finanzmathematischer Anwendungen in der Modellierung von Kursverläufen, auch wenn es heutzutage immer mehr durch komplexere Modelle ersetzt wird.

Das Schöne an dieser einfachen Gleichung ist die Tatsache, dass man die Lösungen noch per Hand angeben kann. Mit Hilfe des Itô-Lemmas rechnet man nach, dass

$$S(t; S_0) = S_0 \exp \left(\left(\mu - \frac{1}{2} \sigma^2 \right) t + \sigma W(t) \right) \quad (4.11)$$

die Lösung von (4.10) ist: Wir betrachten die SDG $dZ(t) = dW(t)$ (also $a \equiv 0$ und $b \equiv 1$) und wenden das Itô-Lemma an auf $Y(t) = g(t, Z(t))$ mit

$$g(t, x) = S_0 \exp \left(\left(\mu - \frac{1}{2} \sigma^2 \right) t + \sigma x \right).$$

Wegen

$$\frac{\partial g}{\partial t}(t, x) = g(t, x)\left(\mu - \frac{1}{2}\sigma^2\right), \quad \frac{\partial g}{\partial x}(t, x) = g(t, x)\sigma \quad \text{und} \quad \frac{\partial^2 g}{\partial x^2}(t, x) = g(t, x)\sigma^2$$

und $Y(t) = g(t, Z(t)) = S(t)$ (weil $Z(t) = W(t)$ ist) folgt aus dem Itô-Lemma die SDG

$$dS(t) = dY(t) = \left(S(t)\left(\mu - \frac{1}{2}\sigma^2\right) + \frac{1}{2}\sigma^2 S(t) \right) dt + S(t)\sigma dW_t = \mu S(t)dt + \sigma S(t)dW_t,$$

was zeigt, dass (4.11) tatsächlich eine Lösung der SDG (4.10) ist, wegen der Eindeutigkeit der Lösung also die einzige.

Für $\sigma = 0$ erhalten wir aus (4.10) gerade die bekannte lineare DGL $\dot{x}(t) = \mu x(t)$ und ihre zugehörige Lösung $x(t) = x(0)e^{\mu t}$. Durch diese Gleichung ist der Erwartungswert $E(S(t; S_0))$ bestimmt. Für die Lösungen von (4.10) gilt nämlich

$$\begin{aligned} E(S(t; S_0)) &= E(S(0)) + E\left(\int_0^t \mu S(\tau; S_0) d\tau\right) + \underbrace{E\left(\int_0^t \sigma S(\tau; S_0) dW_\tau\right)}_{=0, \text{ wegen (4.6)}} \\ &= E(S(0)) + \int_0^t \mu E(S(\tau; S_0)) d\tau. \end{aligned}$$

Die Funktion $e(t) = E(S(t, x_0))$ erfüllt also gerade die lineare DGL $\dot{e}(t) = \mu e(t)$ mit Anfangswert $e(0) = E(S_0) = S_0$, weswegen $E(S(t, x_0)) = S_0 e^{\mu t}$ ist.

Ebenso wie den Erwartungswert kann man auch die Varianz der Lösungen von (4.10) explizit berechnen: Es gilt

$$\begin{aligned} \text{Var}(S(t)) &= E\left(S_0^2 \exp(2((\mu - \frac{1}{2}\sigma^2)t + \sigma W(t))) - S_0^2 e^{2\mu t}\right) \\ &= E\left(S_0^2 e^{2\mu t} (\exp(-\sigma^2 t + 2\sigma W(t)) - 1)\right) \\ &= S_0^2 e^{2\mu t} (E(\exp(-\sigma^2 t + 2\sigma W(t))) - 1). \end{aligned}$$

Der Ausdruck unter dem Erwartungswert lässt sich nun schreiben als

$$\exp(-\sigma^2 t + 2\sigma W(t)) = \tilde{S}_0 \exp\left(\left(\tilde{\mu} - \frac{1}{2}\tilde{\sigma}^2\right)t + \tilde{\sigma}W(t)\right)$$

mit $\tilde{S}_0 = 1$, $\tilde{\mu} = \sigma^2$ und $\tilde{\sigma} = 2\sigma$. Er ist also gerade wieder eine geometrische Brownsche Bewegung, für die folglich

$$E\left(\tilde{S}_0 \exp\left(\left(\tilde{\mu} - \frac{1}{2}\tilde{\sigma}^2\right)t + \tilde{\sigma}W(t)\right)\right) = \tilde{S}_0 e^{\tilde{\mu}t} = e^{\sigma^2 t}$$

gilt. Also erhalten wir für die Varianz

$$\text{Var}(S(t)) = S_0^2 e^{2\mu t} (e^{\sigma^2 t} - 1).$$

Um die geometrische Brownsche Bewegung in der risikoneutralen Optionsbewertung verwenden zu können, muss gemäß (2.5) die Bedingung

$$E(S(T)) = e^{rT} S(0)$$

gelten, woraus sofort $\mu = r$ folgt. Es bleibt also nur ein einziger Parameter übrig, nämlich die Volatilität σ . Wie bereits in Abschnitt 3.4 erläutert, wird dieser Parameter entweder aus vergangenen Kursverläufen historisch oder aus Marktwerten implizit geschätzt, genauer werden wir das später noch erläutern. An diese Stelle soll noch einmal betont werden, dass natürlich auch die Kursmodellierung mit stochastischen Differentialgleichungen keine Vorhersagen über den Kurs zulässt; sie erlaubt nur eine stochastische Modellierung des zukünftigen Verlaufes auf Basis geschätzter charakteristischer Werte.

Das im letzten Kapitel besprochene Binomialmodell wird oft als Approximation der geometrischen Brownschen Bewegung interpretiert. Warum dies gerechtfertigt ist, können wir im Detail erst dann verstehen, wenn wir uns mit der Numerik stochastischer Differentialgleichungen beschäftigt haben. Der Zusammenhang zwischen den Parametern in den beiden Modellen wird allerdings bereits auf dem aktuellen Übungsblatt untersucht.

4.3 Ein anspruchsvolleres Kursmodell

Um zu zeigen, dass man mit stochastischen Differentialgleichungen auch viel anspruchsvollere Modelle als nur die geometrische Brownsche Bewegung aufstellen kann, wollen wir hier noch kurz ein mehrdimensionales Kursmodell angeben. Dieses ist gegeben durch den dreidimensionalen Prozess $X(t) = (S(t), \sigma(t), \zeta(t))^T$, der durch die Gleichungen

$$\begin{aligned} dS(t) &= \mu S(t)dt + \sigma(t)S(t)dW_t^1 \\ d\sigma(t) &= -(\sigma(t) - \zeta(t))dt + \alpha\sigma(t)dW_t^2 \\ d\zeta(t) &= \beta(\sigma(t) - \zeta(t))dt \end{aligned}$$

mit Parametern $\mu \in \mathbb{R}$, $\alpha > 0$ und $\beta \geq 0$ gegeben ist. In der ersten Komponente erkennen wir (4.10) wieder, allerdings ist σ nun kein fester Parameter sondern ebenfalls ein stochastischer Prozess, was die Realität sicherlich deutlich angemessener widerspiegelt. Um zu verstehen, wie sich σ hier in der Zeit verändert, ist es sinnvoll, zunächst die Gleichung für ζ anzuschauen. Für jeden Pfad $\sigma(t, \omega)$ ist dies eine einfache inhomogene lineare DGL, für die man die exakte Lösung als

$$\zeta(t) = e^{-\beta t} \zeta(0) + \int_0^t e^{-\beta(t-s)} \sigma(s, \omega) ds$$

angeben kann. Der Wert ζ kann also als (durch $e^{-\beta(t-s)}$ gewichteter) Mittelwert für σ aufgefasst werden. Die Gleichung für σ bewirkt dann, dass die Volatilität σ ihrem Mittelwert folgt. Wäre dies eine deterministische Gleichung, so wäre das Verhalten langweilig da (für $\beta > 0$) sowohl ζ als auch σ gegen Null konvergieren würden. Da die Gleichung für σ aber durch einen zweiten Wiener Prozess W^2 wiederum stochastisch gestört wird, ergibt sich ein durchaus interessantes Verhalten der Gleichung.

Zur Erstellung noch realistischerer Kursmodelle geht man in der aktuellen Forschung inzwischen über Itô-SDGen hinaus. Diese sind nämlich nur schlecht geeignet, um plötzliche Kurssprünge modellieren zu können. Zu diesem Zweck erweitert man die Gleichungen um Sprungterme, die durch Poisson-Prozesse modelliert werden. Die Optionsbewertung auf Basis solcher Modelle wäre wiederum ein schönes Diplomarbeitsthema.

Kapitel 5

Die Monte-Carlo Methode

In diesem Kapitel werden wir einen Algorithmus zur Bewertung von Optionen kennen lernen, der auf stochastischen Simulationen beruht. Im Gegensatz zur Binomialmethode hat er den großen Vorteil, auch für komplexere Kursmodelle zu funktionieren. Es gibt aber auch Nachteile: die Methode konvergiert relativ langsam und sie eignet sich nicht für amerikanische Optionen (jedenfalls nicht für den amerikanischen Put).

5.1 Grundidee

Die Monte-Carlo Methode dient dazu, den Erwartungswert $\mu = E(X)$ einer Zufallsvariablen X zu schätzen. Sie beruht auf der wiederholten Durchführung von Zufallsexperimenten, d.h. auf simulierten Realisierungen von Zufallsvariablen X . In der Praxis geschieht dies durch den Zufallszahlengenerator im Computer. Dieser simuliert die Durchführung voneinander unabhängiger Zufallsexperimente. Je nach Programmiersprache bzw. Programmierumgebung stehen dabei verschiedene Verteilungen zur Verfügung. Die für uns oft wichtigen normalverteilten Zufallsvariablen werden in MATLAB z.B. mit der Anweisung `randn` erzeugt.

Das wiederholte Ziehen von Zufallszahlen x_1, x_2, x_3, \dots kann dabei auf zwei Arten stochastisch interpretiert werden: Zum einen kann man die x_i als verschiedene unabhängige Realisierungen ein und derselben Zufallsvariablen X interpretieren, also $x_i = X(\omega_i)$, zum anderen kann man die x_i als Realisierungen unabhängiger und identisch verteilter Zufallsvariablen $X_i \sim X$, also $x_i = X_i(\omega)$ auffassen. Beide Sichtweisen sind äquivalent und wir werden beide im Folgenden verwenden.

Die Funktion der Monte-Carlo Methode beruht auf dem zentralen Grenzwertsatz der Stochastik. Dieser Satz gilt für unabhängige und identisch verteilte Zufallsvariablen X_1, X_2, \dots mit Mittelwert $\mu = E(X_i)$ und Varianz $\sigma^2 = \text{Var}(X_i)$. Definiert man die neuen Zufallsvariablen

$$Y_n := \sum_{i=1}^n X_i,$$

so besagt der Satz, dass Y_n für große n approximativ $N(n\mu, n\sigma^2)$ -verteilt ist, also normal-

verteilt mit Mittelwert $n\mu$ und Varianz $n\sigma^2$. Genauer gilt die Konvergenz

$$\lim_{n \rightarrow \infty} P\left(\frac{Y_n - n\mu}{\sigma\sqrt{n}} \leq x\right) = F(x) \quad (5.1)$$

punktweise für alle $x \in \mathbb{R}$, wobei

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

die Verteilungsfunktion der Standard-Normalverteilung ist. Ist das (zentrierte) dritte Moment $E((X_i - \mu)^3)$ endlich, so gilt darüberhinaus nach dem Satz von Berry-Esseen

$$\left| P\left(\frac{Y_n - n\mu}{\sigma\sqrt{n}} \leq x\right) - F(x) \right| = O\left(\frac{1}{\sqrt{n}}\right)$$

und die Konvergenz ist gleichmäßig in x .

Geben wir uns nun ein $\varepsilon > 0$ vor, so folgt

$$P\left(\frac{Y_n}{n} - \mu \leq \varepsilon\right) = P(Y_n - n\mu \leq n\varepsilon) = P\left(\frac{Y_n - n\mu}{\sigma\sqrt{n}} \leq \frac{\sqrt{n}\varepsilon}{\sigma}\right) \approx F\left(\frac{\sqrt{n}\varepsilon}{\sigma}\right) \rightarrow 1$$

und

$$P\left(\frac{Y_n}{n} - \mu \leq -\varepsilon\right) = P(Y_n - n\mu \leq -n\varepsilon) = P\left(\frac{Y_n - n\mu}{\sigma\sqrt{n}} \leq -\frac{\sqrt{n}\varepsilon}{\sigma}\right) \approx F\left(-\frac{\sqrt{n}\varepsilon}{\sigma}\right) \rightarrow 0$$

für $n \rightarrow \infty$. Damit folgt

$$P\left(\left|\frac{Y_n}{n} - \mu\right| \leq \varepsilon\right) \approx F\left(\frac{\sqrt{n}\varepsilon}{\sigma}\right) - F\left(-\frac{\sqrt{n}\varepsilon}{\sigma}\right) \rightarrow 1 \quad (5.2)$$

für $n \rightarrow \infty$. Betrachten wir nun Realisierungen $X_i(\omega)$ der Zufallsvariablen, oder — äquivalent — verschiedene unabhängige Realisierungen $X(\omega_i)$ unserer ursprünglichen Zufallsvariablen $X \sim X_i$, so gilt für große n mit hoher Wahrscheinlichkeit die Approximation

$$\frac{1}{n} \sum_{i=1}^n X(\omega_i) \approx \mu = E(X).$$

Dies ist die einfache Grundidee der Monte-Carlo-Simulation: Um den Erwartungswert $E(X)$ einer Zufallsvariablen zu schätzen, führen wir n (numerisch simulierte) Zufallsexperimente $X(\omega_1), \dots, X(\omega_n)$ durch und berechnen den Mittelwert.

Für eine Monte-Carlo-Simulation auf Basis der geometrischen Brownschen Bewegung (4.10) spezifizieren wir die Schritte (2) und (3) des Grundalgorithmus 2.2 für europäische Optionen daher wie folgt:

- (2) Setze $S(T) = S_0 \exp\left((\mu - \frac{1}{2}\sigma^2)T + \sigma W(T)\right)$ mit $W(T) \sim N(0, T)$

- (3) Simuliere n Kurswerte $S(T, \omega_1), \dots, S(T, \omega_n)$ numerisch und approximiere den Optionswert als den abgezinnten Mittelwert

$$V(t, S(t)) = e^{r(t-T)} \frac{1}{n} \sum_{i=1}^n V(T, S(T, \omega_i)).$$

Zur Berechnung von $S(T, \omega_1), \dots, S(T, \omega_n)$ werden dabei mittels des Zufallsgenerators die n $N(0, T)$ -verteilten Zufallszahlen $W(T, \omega_1), \dots, W(T, \omega_n)$ erzeugt und in die Formel in (2) eingesetzt. Die Formel für $V(T, S)$ in (3) ergibt sich gemäß (2.1) oder (2.2), je nachdem ob ein Call oder Put bewertet werden soll.

Der große Vorteil der Monte-Carlo Methode ist es, dass in Schritt (2) des Algorithmus im Prinzip jedes numerisch simulierbare Kursmodell eingesetzt werden kann. Dieses muss dabei gar keine geschlossene Lösungsformel zulassen, da man statt dessen numerische Lösungsverfahren verwenden kann. Näheres dazu behandeln wir in Kapitel 7.

Bemerkung 5.1 Im Gegensatz zur Binomialmethode lässt sich die Monte-Carlo Methode leider nicht auf amerikanische Put-Optionen verallgemeinern. Der Grund dafür ist der folgende: Um zu einem Zeitpunkt t auf Basis des Kurses $S(t)$ zu entscheiden, ob die vorzeitige Ausübung einen Vorteil bietet (und damit zu einem höheren Wert führt), muss man den Erwartungswert des Optionswertes über die von $S(t)$ ausgehenden zukünftigen Kurspfade kennen. In der Binomialmethode sind diese Werte vorhanden, in der Monte-Carlo-Methode hingegen nicht, weil hier nur einzelne Pfade $S(t)$, $t \in [0, T]$, aber keine Bäume betrachtet werden.

Zwar ließe sich die Methode so variieren, dass entlang jedes Pfades an einer Auswahl diskreter Zeitpunkte der Vorteil der vorzeitigen Ausübung überprüft wird. Dies liefert aber nicht den richtigen Wert (sondern im Allgemeinen einen deutlich größeren), da die Entscheidung dann nur auf Basis eines einzigen zukünftigen Pfades und nicht auf Basis des Erwartungswertes getroffen wird. Dies wäre auch in der Praxis nicht umsetzbar, da dies einer Ausübungsentscheidung mit "Blick in die Zukunft" entsprechen würde. \square

5.2 Konvergenzgeschwindigkeit

Numerische Tests (Übungsaufgabe) zeigen, dass die Monte-Carlo Methode in ihrer einfachen Form für wachsende Anzahl von Simulationen n relativ langsam konvergiert. Wir wollen die Konvergenzgeschwindigkeit hier zunächst theoretisch untersuchen um dann eine Strategie zu entwickeln, mit der man schnellere Konvergenz erhält.

Da die Monte-Carlo Methode nur im stochastischem Sinne konvergiert, müssen wir dazu zunächst ein geeignetes Maß finden. Dies ist das folgende:

Definition 5.2 Gegeben sei $p \in [0, 1]$, $n \in \mathbb{N}$. Ein Intervall der Form $I = [\mu - \varepsilon, \mu + \varepsilon]$ heißt p -Konfidenzintervall der Monte-Carlo Methode, wenn

$$P\left(\frac{1}{n} \sum_{i=1}^n X_i \in I\right) = p$$

ist. \square

Halten wir nun p fest, so können wir untersuchen, wie groß diese Konfidenzintervalle abhängig von n sind. Hierfür gilt der folgende Satz:

Satz 5.3 Gegeben sei $p \in (0, 1)$. Dann existiert $k > 0$ und eine Folge $p_n \rightarrow p$, so dass die p_n -Konfidenzintervalle I_n von der Form

$$I_n = \left[\mu - \frac{k\sigma}{\sqrt{n}}, \mu + \frac{k\sigma}{\sqrt{n}} \right]$$

sind.

Beweis: Wähle $x \in \mathbb{R}$ so, dass für die Verteilungsfunktion der Standard-Normalverteilung

$$F(x) - F(-x) = p$$

gilt. Auf Grund der Konvergenz (5.1) existiert für jedes $x \in \mathbb{R}$ eine Folge $\delta_n(x) \rightarrow 0$ mit

$$\left| P \left(\frac{Y_n - n\mu}{\sigma\sqrt{n}} \leq x \right) - F(x) \right| = \delta_n(x).$$

Analog zur Herleitung von (5.2) erhalten wir

$$P \left(\left| \frac{Y_n - n\mu}{\sigma\sqrt{n}} \right| \leq x \right) = F(x) - F(-x) + \delta_n(x) - \delta_n(-x) = p + \delta_n(x) - \delta_n(-x) =: p_n.$$

Sicherlich gilt mit dieser Definition $p_n \rightarrow p$. Zudem gilt

$$\left| \frac{Y_n - n\mu}{\sigma\sqrt{n}} \right| \leq x \Leftrightarrow \frac{Y_n}{n} \in \left[\mu - \frac{x\sigma}{\sqrt{n}}, \mu + \frac{x\sigma}{\sqrt{n}} \right].$$

Die Behauptung folgt also mit $k = x$. □

Wichtig an diesem Satz ist weniger die Folge $p_n \rightarrow p$, sondern die Aussage darüber, wie schnell die Konfidenzintervalle in Abhängigkeit von n schrumpfen. Tatsächlich hängt die Breite des Intervalls von \sqrt{n} ab, was bedeutet, dass man hundert mal so viele Simulationen benötigt um im Ergebnis eine korrekte Nachkommastelle hinzuzugewinnen. Dies erklärt die langsame Konvergenz der Methode.

Die Analyse zeigt aber auch, dass die Konfidenzintervalle kleiner werden, wenn die Varianz σ^2 der Zufallsvariablen X kleiner wird. Dies ist ein wichtiger Ansatzpunkt zur Erhöhung der Konvergenzgeschwindigkeit. Im folgenden Abschnitt betrachten wir daher eine Methode zur Reduktion der Varianz von X .

5.3 Antithetische Zufallsvariablen

Wir betrachten nun eine Monte-Carlo Simulation mit Zufallsvariable $X = f(Z)$, wobei $f : \mathbb{R} \rightarrow \mathbb{R}$ eine gegebene Funktion ist und $Z \sim N(0, 1)$ eine standard-normalverteilte Zufallsvariable ist.

Dies ist genau die Form unseres Optionsbewertungsproblems aus den obigen Schritten (2) und (3) mit

$$f(Z) = \left(S_0 \exp \left(\left(\mu - \frac{1}{2} \sigma^2 \right) T + \sigma \sqrt{T} Z \right) - K \right)^+ \quad (5.3)$$

(für einen Call, alles Folgende geht analog auch für den Put), wobei wir ausgenutzt haben, dass sich die Zufallsvariable $W(T) \sim N(0, T)$ auch als $\sqrt{T}Z$ mit $Z \sim N(0, 1)$ schreiben lässt.

Die Idee ist nun, die Abbildung f durch eine neue Zufallsvariable \hat{f} zu ersetzen, so dass $E(\hat{f}(Z)) = E(f(Z))$ und $\text{Var}(\hat{f}(Z)) < \text{Var}(f(Z))$ gilt.

Dazu verwenden wir die sogenannte *antithetische* Zufallsvariable zu Z , die gerade durch $-Z$ gegeben ist. Wir verwenden diese, in dem wir $f(Z)$ durch

$$\hat{f}(Z) := \frac{f(Z) + f(-Z)}{2}$$

ersetzen. In der Monte-Carlo Simulation führt dies zum Ersetzen von

$$\frac{1}{n} \sum_{i=1}^n f(Z(\omega_i)) \quad \text{durch} \quad \frac{1}{n} \sum_{i=1}^n \frac{f(Z(\omega_i)) + f(-Z(\omega_i))}{2}.$$

Jede gezogene Zufallszahl $Z(\omega_i)$ geht also zweimal — einmal positiv und einmal negativ — in f und damit in die Summe ein.

Der folgende Satz zeigt, dass wir damit unser Ziel der Varianzreduktion bei gleichbleibendem Erwartungswert erreichen.

Satz 5.4 Betrachte die Monte-Carlo Methode zur Optionsbewertung einer europäischen Call- oder Put-Option mit der geometrischen Brownschen Bewegung als Kursmodell. Dann gelten für die Zufallsvariable

$$\hat{f}(Z) = \frac{f(Z) + f(-Z)}{2}$$

die (Un-)Gleichungen

$$E(\hat{f}(Z)) = E(f(Z)) \quad \text{und} \quad \text{Var}(\hat{f}(Z)) \leq \text{Var}(f(Z))/2.$$

Beweis: Es gilt

$$E \left(\frac{f(Z) + f(-Z)}{2} \right) = \frac{E(f(Z)) + E(f(-Z))}{2} = \frac{E(f(Z)) + E(f(Z))}{2} = E(f(Z))$$

wegen $-Z \sim N(0, 1) \Leftrightarrow Z \sim N(0, 1)$. Dies zeigt die Gleichheit der Erwartungswerte.

Zur Untersuchung der Varianz verwenden wir die Formel

$$\text{Var} \left(\frac{f(Z) + f(-Z)}{2} \right) = \frac{1}{2} \left(\text{Var}(f(Z)) + \text{Cov}(f(Z), f(-Z)) \right), \quad (5.4)$$

die mit der üblichen Kovarianz

$$\text{Cov}(X, Y) = E[(X - E(X))(Y - E(Y))] = E[XY] - E[X]E[Y]$$

aus der Gleichung

$$\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y) + 2\text{Cov}(X, Y)$$

und der Identität $\text{Var}(f(Z)) = \text{Var}(f(-Z))$ folgt.

Wir beweisen zunächst das folgende Hilfsresultat: Es seien $a, b : \mathbb{R} \rightarrow \mathbb{R}$ monotone Abbildungen mit umgekehrter Monotonie (also eine wachsend und eine fallend) Dann gilt für jede beliebige Zufallsvariable X die Ungleichung

$$\text{Cov}(a(X), b(X)) \leq 0. \quad (5.5)$$

Zum Beweis von (5.5) betrachten wir eine weitere beliebige Zufallsvariable Y . Dann gilt wegen der Monotonie von a und b fast sicher

$$(a(X) - a(Y))(b(X) - b(Y)) \leq 0$$

und daher auch $E[(a(X) - a(Y))(b(X) - b(Y))] \leq 0$. Damit erhalten wir

$$\begin{aligned} 0 &\geq E[(a(X) - a(Y))(b(X) - b(Y))] \\ &= E[a(X)b(X)] - E[a(X)b(Y)] - E[a(Y)b(X)] + E[a(Y)b(Y)]. \end{aligned}$$

Wählen wir nun $X \sim Y$ und Y unabhängig von X , so folgt

$$\begin{aligned} 0 &\geq E[a(X)b(X)] - E[a(X)b(Y)] - E[a(Y)b(X)] + E[a(Y)b(Y)] \\ &= 2E[a(X)b(X)] - 2E[a(X)]E[b(X)] \\ &= 2\text{Cov}(a(X), b(X)) \end{aligned}$$

und damit (5.5).

Unsere Funktion f aus (5.3) ist monoton wachsend (für einen Put wäre sie monoton fallend), weswegen die Funktion $x \mapsto f(-x)$ monoton fallend ist (für einen Put wäre sie monoton wachsend). Also können wir (5.5) mit $a(x) = f(x)$ und $b(x) = f(-x)$ anwenden und es folgt

$$\text{Cov}(f(Z), f(-Z)) \leq 0,$$

was eingesetzt in (5.4) die Behauptung zeigt. \square

In der Praxis beobachtet man oft, dass die Verwendung der antithetischen Variablen einen deutlich größeren Vorteil bringt, als es die bloße Halbierung der Varianz vermuten lässt.

Kapitel 6

Numerische Lösung deterministischer Differentialgleichungen

In der finanzmathematischen Modellierung von Kursverläufen werden stochastische Differentialgleichungen benutzt. Sofern man diese nicht — wie bei der geometrischen Brown'schen Bewegung — analytisch lösen kann, müssen numerische Methoden eingesetzt werden. Bevor wir uns im nächsten Kapitel mit solchen numerischen Methoden für stochastische Differentialgleichungen beschäftigen, werden wir zuerst numerische Methoden für gewöhnliche deterministische Differentialgleichungen betrachten. Wengleich die Beweise im stochastischen Fall anders verlaufen, sind die wesentlichen Ideen zur Konstruktion numerischer Verfahren im deterministischen Fall ähnlich und liefern daher eine gute Einführung in das Thema.

6.1 Definition

Eine gewöhnliche Differentialgleichung setzt die Ableitung einer Funktion $x : \mathbb{R} \rightarrow \mathbb{R}^n$ nach ihrem (eindimensionalen) Argument mit der Funktion selbst in Beziehung. Formal beschreibt dies die folgende Definition.

Definition 6.1 Ein *gewöhnliche Differentialgleichung* (DGL) im \mathbb{R}^n , $n \in \mathbb{N}$, ist gegeben durch die Gleichung

$$\frac{d}{dt}x(t) = f(t, x(t)), \quad (6.1)$$

wobei $f : D \rightarrow \mathbb{R}^n$ eine stetige Funktion ist und *Vektorfeld* genannt wird, deren Definitionsbereich D eine offene Teilmenge von $\mathbb{R} \times \mathbb{R}^n$ ist.

Eine *Lösung* von (6.1) ist eine stetig differenzierbare Funktion $x : \mathbb{R} \rightarrow \mathbb{R}^n$, die (6.1) erfüllt. \square

Einige Anmerkungen zur Notation bzw. Sprechweise:

- Die unabhängige Variable t werden wir üblicherweise als Zeit interpretieren, obwohl (abhängig vom modellierten Sachverhalt) gelegentlich auch andere Interpretationen möglich sind.
- Statt $\frac{d}{dt}x(t)$ schreiben wir oft kurz $\dot{x}(t)$.
- Die Lösungsfunktion $x(t)$ nennen wir auch *Lösungskurve* oder (*Lösungs-*)*Trajektorie*.
- Falls das Vektorfeld f nicht von t abhängt, also $\dot{x}(t) = f(x(t))$ ist, nennen wir die Differentialgleichung *autonom*.

6.1.1 Anfangswertprobleme

Eine gewöhnliche Differentialgleichung besitzt im Allgemeinen unendlich viele Lösungen. Als Beispiel betrachte die (sehr einfache) eindimensionale DGL mit $f(x, t) = x$, also

$$\dot{x}(t) = x(t)$$

mit $x(t) \in \mathbb{R}$. Betrachte die Funktion $x(t) = Ce^t$ mit beliebigem $C \in \mathbb{R}$. Dann gilt

$$\dot{x}(t) = \frac{d}{dt}Ce^t = Ce^t = x(t).$$

Für jedes feste C löst Ce^t die obige DGL, es gibt also unendlich viele Lösungen.

Um *eindeutige* Lösungen zu erhalten, müssen wir eine weitere Bedingung festlegen. Dies geschieht in der folgenden Definition.

Definition 6.2 Ein *Anfangswertproblem* für die gewöhnliche Differentialgleichung (6.1) besteht darin, zu gegebenem $t_0 \in \mathbb{R}$ und $x_0 \in \mathbb{R}^n$ eine Lösungsfunktion $x(t)$ zu finden, die (6.1) erfüllt und für die darüberhinaus die Gleichung

$$x(t_0) = x_0 \tag{6.2}$$

gilt. □

Notation und Sprechweisen:

- Für die Lösung $x(t)$, die (6.1) und (6.2) erfüllt, schreiben wir $x(t; t_0, x_0)$. Im Spezialfall $t_0 = 0$ werden wir oft kurz $x(t; x_0)$ schreiben.
- Die Zeit $t_0 \in \mathbb{R}$ bezeichnen wir als *Anfangszeit*, den Wert $x_0 \in \mathbb{R}^n$ als *Anfangswert*. Das Paar (t_0, x_0) bezeichnen wir als *Anfangsbedingung*, ebenso nennen wir die Gleichung (6.2) *Anfangsbedingung*.

Bemerkung 6.3 Eine stetig differenzierbare Funktion $x : I \rightarrow \mathbb{R}^n$ löst das Anfangswertproblem (6.1), (6.2) für ein $t_0 \in I$ und ein $x_0 \in \mathbb{R}^n$ genau dann, wenn sie für alle $t \in I$ die *Integralgleichung*

$$x(t) = x_0 + \int_{t_0}^t f(\tau, x(\tau))d\tau \tag{6.3}$$

erfüllt. Dies folgt sofort durch Integrieren von (6.1) bzgl. t bzw. durch Differenzieren von (6.3) nach t unter Verwendung des Hauptsatzes der Differential- und Integralrechnung. Beachte dabei, dass eine stetige Funktion x , die (6.3) erfüllt, „automatisch“ stetig differenzierbar ist, da aus der Stetigkeit von $t \mapsto x(t)$ und der daraus folgenden Stetigkeit von $t \mapsto f(t, x(t))$ die stetige Differenzierbarkeit der rechten Seite in (6.3) und damit wegen der Gleichheit auch für x selbst folgt. \square

Um nun tatsächlich eine eindeutige Lösung zu einem gegebenen Anfangswert zu erhalten, brauchen wir eine weitere Bedingung.

Definition 6.4 Eine gewöhnliche Differentialgleichung (6.1) für ein stetiges $f : D \rightarrow \mathbb{R}^n$ mit $D \subseteq \mathbb{R} \times \mathbb{R}^n$ offen erfüllt die Lipschitzbedingung im zweiten Argument, wenn das folgende gilt: Für jede kompakte Teilmenge $K \subset D$ existiert eine Konstante $L > 0$, so dass die Ungleichung

$$\|f(t, x) - f(t, y)\| \leq L\|x - y\|$$

gilt für alle $t \in \mathbb{R}$ und $x, y \in \mathbb{R}^n$ mit $(t, x), (t, y) \in K$. \square

Unter dieser Bedingung kann man beweisen, dass für jede Anfangsbedingung $(t_0, x_0) \in D$ genau eine Lösung $x(t; t_0, x_0)$ des Anfangswertproblems (6.1), (6.2) existiert. Diese ist definiert für alle t aus einem offenen *maximalen Existenzintervall* $I_{t_0, x_0} \subseteq \mathbb{R}$ mit $t_0 \in I_{t_0, x_0}$.

Ein Beweis dieses Satzes findet sich z.B. in meinem Skript “Numerische Methoden für gewöhnliche Differentialgleichungen”.

Wir werden im Folgenden immer annehmen, dass die obige Annahme erfüllt ist, auch ohne dies explizit zu erwähnen. Auch werden wir oft Mengen der Form $[t_1, t_2] \times K$ mit $K \subset \mathbb{R}^n$ betrachten, bei denen wir — ebenfalls ohne dies immer explizit zu erwähnen — annehmen, dass alle Lösungen $x(t; t_0, x_0)$ mit $x_0 \in K$ für alle $t_0, t \in [t_1, t_2]$ existieren.

Eine einfache Konsequenz aus der Eindeutigkeit ist die sogenannte *Kozykluseigenschaft* der Lösungen, die für $(t_0, x_0) \in D$ und zwei Zeiten $t_1, t \in \mathbb{R}$ gegeben ist durch

$$x(t; t_0, x_0) = x(t; t_1, x(t_1; t_0, x_0)), \quad (6.4)$$

vorausgesetzt natürlich, dass alle hier auftretenden Lösungen zu den angegebenen Zeiten auch existieren. Zum Beweis rechnet man nach, dass der linke Ausdruck in (6.4) das Anfangswertproblem (6.1), (6.2) zur Anfangsbedingung $(t_1, x(t_1; t_0, x_0))$ löst. Da der rechte dies ebenfalls tut, müssen beide übereinstimmen.

Unter der obigen Lipschitz-Bedingung ist die Lösungsabbildung $x(t; t_0, x_0)$ zudem stetig in all ihren Variablen, also in t, t_0 und x_0 .

6.1.2 Grafische Darstellung der Lösungen

Zur grafischen Darstellung von Lösungen verwenden wir zwei verschiedene Methoden, die wir hier an der zweidimensionalen DGL

$$\dot{x}(t) = \begin{pmatrix} -0.1 & 1 \\ -1 & -0.1 \end{pmatrix} x(t)$$

mit $x(t) = (x_1(t), x_2(t))^T$ und Anfangsbedingung $x(0) = (1, 1)^T$ illustrieren wollen. Da jede Lösung einer Differentialgleichung eine Funktion von \mathbb{R} nach \mathbb{R}^n darstellt, kann man die Graphen der einzelnen Komponenten $x_i(t)$ der Lösung in Abhängigkeit von t darstellen. Für die obige DGL ist dies in Abbildung 6.1 dargestellt. Die durchgezogene Linie zeigt $x_1(t)$ während die gestrichelte Linie $x_2(t)$ darstellt.

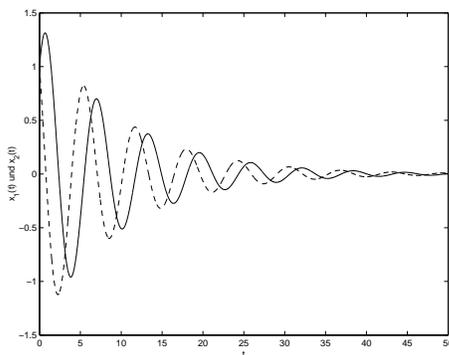


Abbildung 6.1: Darstellung von $x(t)$ mittels Graphen ($x_1(t)$ durchgezogen, $x_2(t)$ gestrichelt)

Eine alternative Darstellung, die speziell für zwei- und dreidimensionale Differentialgleichungen geeignet ist, ergibt sich, wenn man statt der Funktionsgraphen der Komponenten x_i die Kurve $\{x(t) \mid t \in [0, T]\} \subset \mathbb{R}^n$ darstellt. Hier geht in der Grafik die Information über die Zeit (sowohl über die Anfangszeit t_0 als auch über die laufende Zeit t) verloren. Letzteres kann zumindest teilweise durch das Anbringen von Pfeilen, die die Zeitrichtung symbolisieren, ausgeglichen werden. Ein Beispiel für diese Darstellung zeigt Abbildung 6.2.

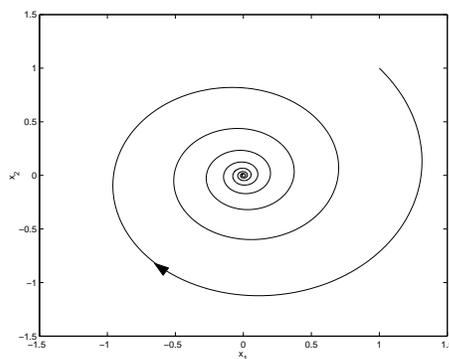


Abbildung 6.2: Darstellung von $x(t)$ als Kurve

Am Computer kann man die Darstellung als Kurve mit einer Animation verbinden, so dass man die Information über den zeitlichen Ablauf der Lösung über die Animation wieder zurück erhält. Ein MATLAB M-File, das sowohl die Abbildungen 6.1 und 6.2 sowie eine animierte Version von Abbildung 6.2 erstellt, findet sich auf der Homepage zur Vorle-

sung “Numerische Methoden für gewöhnliche Differentialgleichungen”¹ unter dem Namen “darstellung.m”.

Für autonome Differentialgleichungen ist der Verlust der Anfangszeit in der Grafik nicht weiter schlimm, da die Lösungen nicht wirklich von der Anfangszeit abhängen: man rechnet leicht nach, dass hier für die Anfangszeiten t_0 und $t_0 + t_1$ die Beziehung

$$x(t; t_0 + t_1, x_0) = x(t - t_1; t_0, x_0) \quad (6.5)$$

gilt. Die Lösung verschiebt sich also auf der t -Achse, verändert sich aber ansonsten nicht. Insbesondere ist die in Abbildung 6.2 dargestellte Kurve für autonome DGL für alle Anfangszeiten gleich.

6.2 Allgemeine Theorie der Einschnittverfahren

In diesem Kapitel werden wir eine wichtige Klasse von Verfahren zur Lösung gewöhnlicher Differentialgleichungen einführen und analysieren, die *Einschnittverfahren*.

6.2.1 Diskrete Approximationen

In der Numerik gewöhnlicher Differentialgleichungen wollen wir eine Approximation an die Lösungsfunktion $x(t; t_0, x_0)$ für $t \in [t_0, T]$ berechnen (wir nehmen hier immer an, dass die Lösungen auf den angegebenen Intervallen existieren). In der folgenden Definition definieren wir die Art von Approximationen, die wir betrachten wollen und einen Begriff der Konvergenzordnung.

Definition 6.5 (i) Eine Menge $\mathcal{T} = \{t_0, t_1, \dots, t_N\}$ von Zeiten mit $t_0 < t_1 < \dots < t_N = T$ heißt *Gitter* auf dem Intervall $[t_0, T]$. Die Werte

$$h_i = t_{i+1} - t_i$$

heißen *Schrittweiten*, der Wert

$$\bar{h} = \max_{i=0, \dots, N-1} h_i$$

heißt *maximale Schrittweite*. Im Fall *äquidistanter Schrittweiten* $h_0 = h_1 = \dots = h_{N-1}$ schreiben wir zumeist h statt h_i .

(ii) Eine Funktion $\tilde{x} : \mathcal{T} \rightarrow \mathbb{R}^n$ heißt *Gitterfunktion*.

(iii) Eine Familie von Gitterfunktionen \tilde{x}_j , $j \in \mathbb{N}$, auf Gittern \mathcal{T}_j auf dem Intervall $[t_0, T] \subset I_{t_0, x_0}$ mit maximalen Schrittweiten \bar{h}_j heißt (*diskrete*) *Approximation* der Lösung $x(t; t_0, x_0)$ von (6.1), falls

$$\max_{t_i \in \mathcal{T}_j} \|\tilde{x}_j(t_i) - x(t_i; t_0, x_0)\| \rightarrow 0$$

¹<http://www.uni-bayreuth.de/departments/math/~lgruene/numerik05/>

für $\bar{h}_j \rightarrow 0$. Eine von der Anfangsbedingung (t_0, x_0) abhängige Familie von Gitterfunktionen $\tilde{x}_j(\cdot; t_0, x_0)$, $j \in \mathbb{N}$, hat die *Konvergenzordnung* $p > 0$, falls für jede kompakte Menge $K \subset D$ und alle $T > 0$ mit $[t_0, T] \subset I_{t_0, x_0}$ für alle $(t_0, x_0) \in K$ ein $C > 0$ existiert, so dass

$$\max_{t_i \in \mathcal{T}_j} \|\tilde{x}_j(t_i; t_0, x_0) - x(t_i; t_0, x_0)\| \leq C\bar{h}_j^p$$

gilt für alle $(t_0, x_0) \in K$ und alle hinreichend feinen Gitter \mathcal{T}_j auf $[t_0, T]$. In diesem Fall schreiben wir auch $\tilde{x}_j(t_i; t_0, x_0) = x(t_i; t_0, x_0) + O(\bar{h}_j^p)$. \square

Bemerkung 6.6 Wir haben in der Einführung in die Numerik verschiedene Methoden kennen gelernt, mit denen man Funktionen numerisch darstellen kann, z.B. Polynom- oder Splineinterpolation. Jede Gitterfunktion gemäß Definition 6.5 kann natürlich mit diesen Methoden zu einer “echten” Funktion erweitert werden. \square

Ein Einschrittverfahren ist nun gegeben durch eine numerisch auswertbare Funktion Φ , mittels derer wir eine Gitterfunktion zu einem gegebenen Gitter berechnen können. Formal ist dies wie folgt definiert.

Definition 6.7 Ein *Einschrittverfahren* ist gegeben durch eine stetige Abbildung

$$\Phi : \mathbb{R} \times \mathbb{R}^n \times \mathbb{R} \rightarrow \mathbb{R}^n,$$

mit der zu jedem Gitter \mathcal{T} und jedem Anfangswert x_0 mittels

$$\tilde{x}(t_0) = x_0, \quad \tilde{x}(t_{i+1}) = \Phi(t_i, \tilde{x}(t_i), h_i) \text{ für } i = 0, 1, \dots, N-1$$

rekursiv eine Gitterfunktion definiert werden kann.

Wenn die so erzeugten Gitterfunktionen die Bedingung aus Definition 6.5 (iii) erfüllen, so nennen wir das Einschrittverfahren *konvergent* bzw. *konvergent mit Konvergenzordnung* p . \square

Der Name *Einschrittverfahren* ergibt sich dabei aus der Tatsache, dass der Wert $\tilde{x}(t_{i+1})$ nur aus dem direkten Vorgängerwert $\tilde{x}(t_i)$ berechnet wird. Wir werden später auch *Mehrschrittverfahren* kennen lernen, bei denen $\tilde{x}(t_{i+1})$ aus $\tilde{x}(t_{i-k}), \tilde{x}(t_{i-k+1}), \dots, \tilde{x}(t_i)$ berechnet wird.

6.2.2 Erste einfache Einschrittverfahren

Bevor wir in die Konvergenztheorie einsteigen und mathematisch untersuchen, welche Bedingungen Φ erfüllen muss, damit die erzeugte Gitterfunktion eine Approximation darstellt, wollen wir in diesem Abschnitt zwei Einschrittverfahren heuristisch betrachten.

Die Idee der Verfahren erschließt sich am einfachsten über die Integralgleichung (6.3). Die exakte Lösung erfüllt ja gerade

$$x(t_{i+1}) = x(t_i) + \int_{t_i}^{t_{i+1}} f(\tau, x(\tau)) d\tau.$$

Die Idee ist nun, das Integral durch einen Ausdruck zu ersetzen, der numerisch berechenbar ist, wenn wir $x(\tau)$ für $\tau > t_i$ nicht kennen. Die einfachste Approximation ist die Rechteck-Regel (oder Newton-Cotes Formel mit $n = 0$, die wir in der Einführung in die Numerik wegen ihrer Einfachheit gar nicht betrachtet haben)

$$\int_{t_i}^{t_{i+1}} f(\tau, x(\tau)) d\tau \approx (t_{i+1} - t_i) f(t_i, x(t_i)) = h_i f(t_i, x(t_i)). \quad (6.6)$$

Setzen wir also

$$\Phi(t, x, h) = x + h f(t, x), \quad (6.7)$$

so gilt

$$\tilde{x}(t_{i+1}) = \Phi(t_i, \tilde{x}(t_i), h_i) = \tilde{x}(t_i) + h_i f(t_i, \tilde{x}(t_i))$$

und wenn wir $\tilde{x}(t_i) \approx x(t_i)$ annehmen, so können wir fortfahren

$$\dots \approx x(t_i) + h_i f(t_i, x(t_i)) \approx x(t_i) + \int_{t_i}^{t_{i+1}} f(\tau, x(\tau)) d\tau.$$

Da $\tilde{x}(t_0) = x_0 = x(t_0)$ ist, kann man damit rekursiv zeigen, dass $\tilde{x}(t_{i+1})$ eine Approximation von $x(t_{i+1})$ ist. Wir werden dies im nächsten Abschnitt mathematisch präzisieren.

Das durch (6.7) gegebene Verfahren ist das einfachste Einschrittverfahren und heißt *Euler'sche Polygonzugmethode* oder einfach *Euler-Verfahren*. Es hat eine einfache geometrische Interpretation: In jedem Punkt $\tilde{x}(t_i)$ berechnen wir die Steigung der exakten Lösung durch diesen Punkt (das ist gerade $f(t_i, \tilde{x}(t_i))$) und folgen der dadurch definierten Geraden bis zum nächsten Zeitschritt. Das Prinzip ist in Abbildung 6.3 grafisch dargestellt.

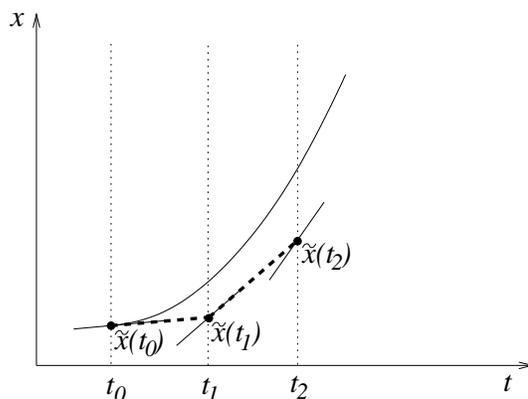


Abbildung 6.3: Grafische Veranschaulichung des Euler-Verfahrens

Das Euler-Verfahren liefert nur eine recht grobe Approximation der Lösung. Bessere Verfahren kann man erhalten, wenn man statt (6.6) eine genauere Approximation verwendet. Eine bessere Möglichkeit ist z.B.

$$\int_{t_i}^{t_{i+1}} f(\tau, x(\tau)) d\tau \approx \frac{h_i}{2} \left(f(t_i, x(t_i)) + f(t_{i+1}, x(t_i) + h_i f(t_i, x(t_i))) \right). \quad (6.8)$$

Dies ist nichts anderes als die Trapez-Regel (oder Newton-Cotes Formel mit $n = 1$), bei der wir den unbekanntem Wert $x(t_{i+1})$ durch die Euler-Approximation $x(t_{i+1}) \approx x(t_i) + h_i f(t_i, x(t_i))$ ersetzen. Das daraus resultierende Verfahren ist gegeben durch

$$\Phi(t, x, h) = x + \frac{h}{2} \left(f(t, x) + f\left(t + h, x + hf(t, x)\right) \right)$$

und heißt *Heun-Verfahren*. Es ist tatsächlich schon deutlich besser als das Euler-Verfahren.

Man kann sich leicht vorstellen, dass weitere bessere Verfahren sehr komplizierte Formeln benötigen. Wir werden deshalb später einen Formalismus kennen lernen, mit dem man auch sehr komplizierte Verfahren einfach aufschreiben und implementieren kann.

Ein Grundalgorithmus zur Approximation einer Lösung $x(t; t_0, x_0)$ auf $[t_0, T]$ mittels eines Einschrittverfahrens Φ lässt sich nun leicht angeben. Wir beschränken uns hierbei zunächst auf Gitter mit konstanter Schrittweite, also $h_i = h$ für alle $i = 0, 1, 2, \dots, N$, wobei wir N als Parameter vorgeben.

Algorithmus 6.8 (Lösung eines Anfangswertproblems mit Einschrittverfahren)

Eingabe: Anfangsbedingung (t_0, x_0) , Endzeit T , Schrittzahl N , Einschrittverfahren Φ

(1) Setze $h := (T - t_0)/N$, $\tilde{x}_0 = x_0$

(2) Berechne $t_{i+1} = t_i + h$, $\tilde{x}_{i+1} := \Phi(t_i, \tilde{x}_i, h)$ für $i = 0, \dots, N - 1$.

Ausgabe: Werte der Gitterfunktion $\tilde{x}(t_i) = \tilde{x}_i$ in t_0, \dots, t_N □

6.2.3 Konvergenztheorie

Die Grundidee der Konvergenztheorie für numerische Methoden für Differentialgleichungen liegt in einem geschickten Trick, mit dem verschiedene Fehlerquellen separiert werden können. Wir schreiben hier kurz $x(t) = x(t; t_0, x_0)$. Um nun den Fehler

$$\|\tilde{x}(t_i) - x(t_i)\| = \|\Phi(t_{i-1}, \tilde{x}(t_{i-1}), h_{i-1}) - x(t_i)\|$$

abzuschätzen, schieben wir mittels der Dreiecksungleichung die Hilfsgröße

$$\Phi(t_{i-1}, x(t_{i-1}), h_{i-1})$$

ein. Wir erhalten so mit (6.4) die Abschätzung

$$\begin{aligned} \|\tilde{x}(t_i) - x(t_i)\| &\leq \|\Phi(t_{i-1}, \tilde{x}(t_{i-1}), h_{i-1}) - \Phi(t_{i-1}, x(t_{i-1}), h_{i-1})\| \\ &\quad + \|\Phi(t_{i-1}, x(t_{i-1}), h_{i-1}) - x(t_i)\| \\ &= \|\Phi(t_{i-1}, \tilde{x}(t_{i-1}), h_{i-1}) - \Phi(t_{i-1}, x(t_{i-1}), h_{i-1})\| \\ &\quad + \|\Phi(t_{i-1}, x(t_{i-1}), h_{i-1}) - x(t_i; t_{i-1}, x_{i-1})\| \end{aligned}$$

Statt also direkt den Fehler zur Zeit t_i abzuschätzen, betrachten wir getrennt die zwei Terme

- (a) $\|\Phi(t_{i-1}, \tilde{x}(t_{i-1}), h_{i-1}) - \Phi(t_{i-1}, x(t_{i-1}), h_{i-1})\|$, also die Auswirkung des Fehlers bis zur Zeit t_{i-1} in Φ
- (b) $\|\Phi(t_{i-1}, x(t_{i-1}), h_{i-1}) - x(t_i; t_{i-1}, x_{i-1})\|$, also den lokalen Fehler beim Schritt von $x(t_{i-1})$ nach $x(t_i)$

Die folgende Definition gibt die benötigten Eigenschaften an Φ an, mit denen diese Fehler abgeschätzt werden können.

Definition 6.9 (i) Ein Einschrittverfahren erfüllt die *Lipschitzbedingung* (oder *Stabilitätsbedingung*), falls für jede kompakte Menge $K \subset D$ des Definitionsbereiches der Differentialgleichung ein $L > 0$ existiert, so dass für alle Paare $(t_0, x_1), (t_0, x_2) \in K$ und alle hinreichend kleinen $h > 0$ die Abschätzung

$$\|\Phi(t_0, x_1, h) - \Phi(t_0, x_2, h)\| \leq (1 + Lh)\|x_1 - x_2\| \quad (6.9)$$

gilt.

(ii) Ein Einschrittverfahren Φ heißt *konsistent*, falls für jede kompakte Menge $K \subset D$ des Definitionsbereiches der Differentialgleichung eine Funktion $\varepsilon(h)$ mit $\lim_{h \rightarrow 0} \varepsilon(h) = 0$ existiert, so dass für alle $(t_0, x_0) \in K$ und alle hinreichend kleinen $h > 0$ die Ungleichung

$$\|\Phi(t_0, x_0, h) - x(t_0 + h; t_0, x_0)\| \leq h\varepsilon(h) \quad (6.10)$$

gilt. O.B.d.A. nehmen wir dabei an, dass $\varepsilon(h)$ monoton ist, ansonsten können wir $\varepsilon(h)$ durch $\sup_{h \in [0, h]} \varepsilon(h)$ ersetzen.

Das Verfahren hat die *Konsistenzordnung* $p > 0$, falls für jede kompakte Menge $K \subset D$ ein $E > 0$ existiert, so dass $\varepsilon(h) = Eh^p$ gewählt werden kann. In diesem Fall schreiben wir auch $\Phi(t_0, x_0, h) = x(t_0 + h; t_0, x_0) + O(h^{p+1})$. \square

Offenbar garantiert (6.9), dass der Fehlerterm (a) nicht zu groß wird, während (6.10) dazu dient, den Term (b) abzuschätzen. Der formale Beweis folgt in Satz 6.11. Bevor wir diesen formulieren, wollen wir uns noch überlegen, ob die im vorherigen Abschnitt definierten Verfahren diese Bedingungen erfüllen.

Man rechnet leicht nach, dass das Euler- und das Heun-Verfahren die Lipschitzbedingung erfüllen. Die Konsistenzbedingung (6.10) ist allerdings nicht so leicht nachzuprüfen, da sie mit Hilfe der (unbekannten) Lösungen $x(t; t_0, x_0)$ formuliert ist. Das folgende Lemma stellt eine alternative und leichter nachprüfbare Formulierung der Bedingung vor.

Lemma 6.10 Gegeben sei ein Einschrittverfahren Φ der Form

$$\Phi(t, x, h) = x + h\varphi(t, x, h)$$

mit einer stetigen Funktion $\varphi : \mathbb{R} \times \mathbb{R}^n \times \mathbb{R} \rightarrow \mathbb{R}^n$. Dann ist das Verfahren genau dann konsistent, falls für alle $(t, x) \in D$ die Bedingung

$$\varphi(t, x, 0) = f(t, x) \quad (6.11)$$

gilt.

Beweis: Wir schreiben wieder kurz $x(t) = x(t; t_0, x_0)$. Es gilt

$$\begin{aligned}
& \frac{\Phi(t_0, x_0, h) - x(t_0 + h)}{h} \\
&= \frac{1}{h} \left(\Phi(t_0, x_0, h) - x_0 - \int_{t_0}^{t_0+h} f(\tau, x(\tau)) d\tau \right) \\
&= \frac{1}{h} \left(\Phi(t_0, x_0, h) - x_0 - \int_{t_0}^{t_0+h} f(t_0, x_0) d\tau + \int_{t_0}^{t_0+h} f(t_0, x_0) d\tau - \int_{t_0}^{t_0+h} f(\tau, x(\tau)) d\tau \right) \\
&= \frac{1}{h} \left(h\varphi(t_0, x_0, h) - \int_{t_0}^{t_0+h} f(t_0, x_0) d\tau \right) + \frac{1}{h} \left(\int_{t_0}^{t_0+h} f(t_0, x_0) - f(\tau, x(\tau)) d\tau \right) \\
&= \varphi(t_0, x_0, h) - f(t_0, x_0) + \frac{1}{h} \left(\int_{t_0}^{t_0+h} f(t_0, x_0) - f(\tau, x(\tau)) d\tau \right)
\end{aligned}$$

Sei nun $K \subset D$ gegeben. Die Funktion $f(t_0 + s, x(t_0 + s; t_0, x_0))$ ist stetig in s , t_0 und x_0 , also gleichmäßig stetig für $(s, t_0, x_0) \in [0, h] \times K$ für hinreichend kleines $h > 0$ (so klein, dass die Lösungen $x(t_0 + s; t_0, x_0)$ für $s \in [0, h]$ existieren), da diese Menge kompakt ist. Also existiert eine Funktion $\varepsilon_1(h) \rightarrow 0$ mit

$$\|f(\tau, x(\tau)) - f(t_0, x(t_0))\| \leq \varepsilon_1(h)$$

für $\tau = t_0 + s \in [t_0, t_0 + h]$ und damit

$$\frac{1}{h} \left\| \int_{t_0}^{t_0+h} f(t_0, x_0) - f(\tau, x(\tau)) d\tau \right\| \leq \frac{1}{h} \int_{t_0}^{t_0+h} \|f(t_0, x_0) - f(\tau, x(\tau))\| d\tau \leq \varepsilon_1(h). \quad (6.12)$$

Wir nehmen nun an, dass (6.11) gilt. Ebenfalls wegen gleichmäßiger Stetigkeit und wegen (6.11) existiert eine Funktion $\varepsilon_2(h) \rightarrow 0$ mit

$$\|\varphi(t_0, x_0, h) - f(t_0, x_0)\| \leq \varepsilon_2(h).$$

Damit folgt

$$\frac{\|\Phi(t_0, x_0, h) - x(t_0 + h)\|}{h} \leq \varepsilon_2(h) + \varepsilon_1(h),$$

also (6.10) mit $\varepsilon(h) = \varepsilon_1(h) + \varepsilon_2(h)$.

Gelte umgekehrt (6.10). Sei $(x, t) \in D$ gegeben und sei $[t_1, t_2]$ und K so gewählt, dass $(x, t) \in [t_1, t_2] \times K$ gilt. Wiederum mit (6.12) folgt

$$\|\varphi(t_0, x_0, h) - f(t_0, x_0)\| \leq \varepsilon(h) + \varepsilon_1(h),$$

also

$$\lim_{h \rightarrow 0} \|\varphi(t_0, x_0, h) - f(t_0, x_0)\| = 0$$

und damit (6.11) wegen der Stetigkeit von φ . \square

Mit Hilfe der Bedingung (6.11) prüft man leicht nach, dass das Euler- und das Heun-Verfahren konsistent sind. Die Konsistenzordnung kann man aus (6.11) allerdings nicht ableiten, da die Abschätzung von $\varepsilon(h)$ mittels $\varepsilon_1(h)$ und $\varepsilon_2(h)$ dafür zu grob ist, denn falls $f \neq 0$ ist, gilt $\varepsilon_1(h) \geq O(h)$, so dass man maximal die Konsistenzordnung $p = 1$ nachweisen könnte. Wir werden später sehen, wie man die Konsistenzordnung berechnen kann.

Wir kommen nun zu unserem ersten wichtigen Satz, der besagt, dass Lipschitzbedingung und Konsistenz tatsächlich ausreichend für die Konvergenz sind.

Satz 6.11 Betrachte ein Einschrittverfahren Φ , das die Lipschitzbedingung erfüllt und konsistent ist. Dann ist das Verfahren konvergent. Falls das Verfahren dabei die Konsistenzordnung p besitzt, so besitzt es auch die Konvergenzordnung p .

Beweis: Wir müssen die Eigenschaft aus Definition 6.5(iii) nachprüfen. Sei dazu eine kompakte Menge $K \subset D$ und ein $T > 0$ mit $[t_0, T] \subset I_{t_0, x_0}$ für alle $(t_0, x_0) \in K$ gegeben. Die Menge

$$K_1 := \{(t, x(t; t_0, x_0)) \mid (t_0, x_0) \in K, t \in [t_0, T]\}$$

ist dann ebenfalls kompakt, da x stetig in allen Variablen ist und Bilder kompakter Mengen unter stetigen Funktionen wieder kompakt sind. Wir wählen ein $\delta > 0$ und betrachten die kompakte Menge

$$K_2 := \bigcup_{(t,x) \in K_1} \{t\} \times \overline{B}_\delta(x).$$

Diese Menge kompakt ist, folgt aus der Beschränktheit und der Abgeschlossenheit; beide Eigenschaften folgen mit einer kleinen Rechnung aus der Kompaktheit von K_1 .

Die Menge K_2 ist genau die Menge aller Punkte (t, x) , deren x -Komponente einen Abstand $\leq \delta$ von einer Lösung $x(t; t_0, x_0)$ mit $x_0 \in K$ hat. Für hinreichend kleines $\delta > 0$ ist K_2 Teilmenge des Definitionsbereiches D von f , da D offen ist und $K_1 \subset D$ gilt. Das betrachtete Einschrittverfahren ist deswegen konsistent auf K_2 mit einer Funktion $\varepsilon(h)$, wobei $\varepsilon(h) = Eh^p$ im Falle der Konsistenzordnung p ist. Ebenfalls erfüllt Φ auf K_2 die Lipschitzbedingung mit einer Konstanten $L > 0$.

Wir beweisen die Konvergenz nun zunächst unter der folgenden Annahme, deren Gültigkeit wir später beweisen werden:

Für alle hinreichend feinen Gitter \mathcal{T} und alle Anfangsbedingungen $(t_0, x_0) \in K$ gilt für die gemäß Definition 6.7 erzeugte Gitterfunktion \tilde{x} (6.13) die Beziehung $(t_i, \tilde{x}(t_i)) \in K_2$ für alle $t_i \in \mathcal{T}$.

Zum Beweis der Konvergenz wählen wir eine Anfangsbedingung $(t_0, x_0) \in K$ und schreiben wieder kurz $x(t) = x(t; t_0, x_0)$. Mit \tilde{x} bezeichnen wir die zugehörige numerisch approximierende Gitterfunktion und mit

$$e(t_i) := \|\tilde{x}(t_i) - x(t_i)\|$$

bezeichnen wir den Fehler zur Zeit $t_i \in \mathcal{T}$. Dann gilt nach den Vorüberlegungen am Anfang dieses Abschnitts

$$\begin{aligned} e(t_i) &= \|\tilde{x}(t_i) - x(t_i)\| \leq \|\Phi(t_{i-1}, \tilde{x}(t_{i-1}), h_{i-1}) - \Phi(t_{i-1}, x(t_{i-1}), h_{i-1})\| \\ &\quad + \|\Phi(t_{i-1}, x(t_{i-1}), h_{i-1}) - x(t_i)\| \\ &= \|\Phi(t_{i-1}, \tilde{x}(t_{i-1}), h_{i-1}) - \Phi(t_{i-1}, x(t_{i-1}), h_{i-1})\| \\ &\quad + \|\Phi(t_{i-1}, x(t_{i-1}), h_{i-1}) - x(t_i; t_{i-1}, x(t_{i-1}))\| \\ &\leq (1 + Lh_{i-1})\|\tilde{x}(t_{i-1}) - x(t_{i-1})\| + h_{i-1}\varepsilon(h_{i-1}) \\ &= (1 + Lh_{i-1})e(t_{i-1}) + h_{i-1}\varepsilon(h_{i-1}) \end{aligned}$$

wobei wir im vorletzten Schritt die Lipschitzbedingung und die Konsistenz sowie die Tatsache, dass $(t_{i-1}, \tilde{x}(t_{i-1})) \in K_2$ liegt, ausgenutzt haben. Wir erhalten also für den Fehler $e(t_i)$ die rekursive Gleichung

$$e(t_i) \leq (1 + Lh_{i-1})e(t_{i-1}) + h_{i-1}\varepsilon(h_{i-1})$$

gemeinsam mit der “Anfangsbedingung” $e(t_0) = 0$, da $\tilde{x}(t_0) = x_0 = x(t_0)$ ist.

Mittels Induktion zeigen wir nun, dass daraus die Abschätzung

$$e(t_i) \leq \varepsilon(\bar{h}) \frac{1}{L} (\exp(L(t_i - t_0)) - 1)$$

folgt. Für $i = 0$ ist die Abschätzung klar. Für $i - 1 \rightarrow i$ verwenden wir

$$\exp(Lh_i) = 1 + Lh_i + \frac{L^2 h_i^2}{2} + \dots \geq 1 + Lh_i$$

und erhalten damit mit der Induktionsannahme

$$\begin{aligned} e(t_i) &\leq (1 + Lh_{i-1})e(t_{i-1}) + h_{i-1}\varepsilon(h_{i-1}) \\ &\leq (1 + Lh_{i-1})\varepsilon(\bar{h}) \frac{1}{L} (\exp(L(t_{i-1} - t_0)) - 1) + h_{i-1} \underbrace{\varepsilon(h_{i-1})}_{\leq \varepsilon(\bar{h})} \\ &= \varepsilon(\bar{h}) \frac{1}{L} \left(h_{i-1}L + (1 + Lh_{i-1})(\exp(L(t_{i-1} - t_0)) - 1) \right) \\ &= \varepsilon(\bar{h}) \frac{1}{L} \left(h_{i-1}L + (1 + Lh_{i-1}) \exp(L(t_{i-1} - t_0)) - 1 - Lh_{i-1} \right) \\ &= \varepsilon(\bar{h}) \frac{1}{L} \left((1 + Lh_{i-1}) \exp(L(t_{i-1} - t_0)) - 1 \right) \\ &\leq \varepsilon(\bar{h}) \frac{1}{L} \left(\exp(Lh_{i-1}) \exp(L(t_{i-1} - t_0)) - 1 \right) \\ &= \varepsilon(\bar{h}) \frac{1}{L} (\exp(L(t_i - t_0)) - 1). \end{aligned}$$

Damit folgt die Konvergenz und im Falle von $\varepsilon(\bar{h}) \leq E\bar{h}^p$ auch die Konvergenzordnung mit $C = E(\exp(L(T - t_0)) - 1)/L$.

Es bleibt zu zeigen, dass unsere oben gemachte Annahme (6.13) tatsächlich erfüllt ist. Wir zeigen, dass (6.13) für alle Gitter \mathcal{T} gilt, deren maximale Schrittweite \bar{h} die Ungleichung

$$\varepsilon(\bar{h}) \leq \frac{\delta L}{\exp(L(T - t_0)) - 1}$$

erfüllt. Wir betrachten dazu eine Lösung \tilde{x} mit Anfangswert $x_0 \in K$ und beweisen die Annahme per Induktion. Für $\tilde{x}(t_0)$ ist wegen $\tilde{x}(t_0) = x_0$ nichts zu zeigen. Für den Induktionsschritt $i - 1 \rightarrow i$ sei $(t_k, \tilde{x}(t_k)) \in K_2$ für $k = 0, 1, \dots, i - 1$. Wir müssen zeigen, dass $(t_i, \tilde{x}(t_i)) \in K_2$ liegt. Beachte, dass die oben gezeigte Abschätzung

$$e(t_i) \leq \varepsilon(\bar{h}) \frac{1}{L} (\exp(L(T - t_0)) - 1)$$

bereits gilt, falls $(t_k, \tilde{x}(t_k)) \in K_2$ liegt für $k = 0, 1, \dots, i - 1$. Mit der Wahl von h folgt damit $e(t_i) \leq \delta$, also

$$\|\tilde{x}(t_i) - x(t_i)\| \leq \delta.$$

Da $(t_i, x(t_i)) \in K_1$ liegt, folgt $(t_i, \tilde{x}(t_i)) \in \{t_i\} \times \overline{B}_\delta(x(t_i)) \subset K_2$, also die gewünschte Beziehung. \square

Bemerkung 6.12 (i) Schematisch dargestellt besagt Satz 6.11 das Folgende:

$$\begin{array}{ll} \text{Lipschitzbedingung} + \text{Konsistenz} & \Rightarrow \text{Konvergenz} \\ \text{Lipschitzbedingung} + \text{Konsistenzordnung } p & \Rightarrow \text{Konvergenzordnung } p \end{array}$$

(ii) Die Schranke für $e(T)$ wächst — sogar sehr schnell — wenn die Intervallgröße $T - t_0$ wächst. Insbesondere lassen sich mit dieser Abschätzung keinerlei Aussagen über das Langzeitverhalten numerischer Lösungen machen, z.B. über Grenzwerte $\tilde{x}(t_i)$ für $t_i \rightarrow \infty$. Tatsächlich kann es passieren, dass der “numerische Grenzwert” von $\tilde{x}(t_i)$ für $t_i \rightarrow \infty$ für beliebig feine Gitter \mathcal{T} weit von dem tatsächlichen Grenzwert der exakten Lösung $x(t)$ entfernt ist. Wir werden später genauer auf dieses Problem eingehen.

(iii) Der Konsistenzfehler $\varepsilon(h)h$ wird auch als *lokaler Fehler* bezeichnet, während der im Beweis abgeschätzte Fehler $e(t)$ als *globaler Fehler* bezeichnet wird. Im Falle der Konsistenzordnung p gilt $\varepsilon(h)h = O(h^{p+1})$ und $e(t) = O(h^p)$. Man “verliert” also eine Ordnung beim Übergang vom lokalen zum globalen Fehler. Dies lässt sich anschaulich wie folgt erklären: Bis zur Zeit t muss man (bei äquidistantem Gitter) gerade ca. $N(t) = (t - t_0)/h$ Schritte machen, weswegen sich $N(t)$ lokale Fehler aufsummieren, was zu dem globalen Fehler $O(h^{p+1})N(t) = O(h^{p+1})/h = O(h^p)$ führt. \square

6.3 Taylor-Verfahren

Wir werden in diesem Kapitel eine spezielle Klasse von Einschrittverfahren einführen, die in der numerischen Praxis zwar eher selten verwendet werden (wir werden später sehen, wieso), für das Verständnis der weiteren Einschrittverfahren aber sehr nützlich sind.

6.3.1 Definition

Die Taylor-Verfahren haben ihren Namen von der zu Grunde liegenden Taylor-Formel und gehen in direkter Weise aus diesen hervor. Allerdings wird die Taylor-Formel in zunächst etwas ungewohnt erscheinender Weise angewendet: Wir verwenden den Differentialoperator L_f^i , $i \in \mathbb{N}$, der für (hinreichend oft differenzierbare) Funktionen $f, g : D \rightarrow \mathbb{R}^n$ mit $D \subseteq \mathbb{R} \times \mathbb{R}^n$ mittels

$$L_f^0 g(t, x) := g(t, x), \quad L_f^1 g(t, x) := \frac{\partial g}{\partial t}(t, x) + \frac{\partial g}{\partial x}(t, x)f(t, x), \quad L_f^{i+1} g(t, x) = L_f^1 L_f^i g(t, x)$$

definiert ist. Beachte, dass $L_f^i g$ wieder eine Funktion von D nach \mathbb{R}^n ist. Der folgende Satz stellt die hier benötigte Version der Taylor-Formel vor.

Satz 6.13 Gegeben sei eine Differentialgleichung (6.1) mit p -mal stetig differenzierbarem Vektorfeld f . Sei $x(t) = x(t; t_0, x_0)$ eine Lösung dieser Differentialgleichung. Dann gilt

$$x(t) = x_0 + \sum_{i=1}^p \frac{(t-t_0)^i}{i!} L_f^{i-1} f(t_0, x_0) + O((t-t_0)^{p+1}),$$

wobei das O -Symbol im Sinne von Definition 6.5(iii) verwendet wird.

Beweis: Aus der Theorie der gewöhnlichen Differentialgleichungen ist bekannt, dass die Lösung $x(t)$ unter der vorausgesetzten Differenzierbarkeitsbedingung an f $p+1$ -mal stetig differenzierbar nach t ist. Nach der aus der Analysis bekannten Taylor-Formel für Funktionen von \mathbb{R} nach \mathbb{R}^n gilt demnach

$$x(t) = x_0 + \sum_{i=1}^p \frac{(t-t_0)^i}{i!} \frac{d^i x}{dt^i}(t_0) + O((t-t_0)^{p+1}).$$

Zum Beweis des Satzes werden wir nun nachweisen, dass

$$\frac{d^i x}{dt^i}(t) = L_f^{i-1} f(t, x(t)) \quad (6.14)$$

ist, denn dann folgt die Behauptung aus

$$\frac{d^i x}{dt^i}(t_0) = L_f^{i-1} f(t_0, x(t_0)) = L_f^{i-1} f(t_0, x_0).$$

Wir zeigen (6.14) per Induktion über i . Für $i = 1$ gilt

$$\frac{dx}{dt}(t_0) = f(t_0, x(t_0)) = f(t_0, x_0) = L_f^0 f(t_0, x_0).$$

Für $i \rightarrow i+1$ beachte, dass für je zwei differenzierbare Funktionen $g : D \rightarrow \mathbb{R}^n$ und $x : \mathbb{R} \rightarrow \mathbb{R}^n$ die Gleichung

$$\frac{d}{dt} g(t, x(t)) = \frac{\partial g}{\partial t}(t, x(t)) + \frac{\partial g}{\partial x}(t, x(t)) \frac{d}{dt} x(t)$$

gilt (man nennt dies auch die *totale Ableitung* von g entlang der Funktion $x(t)$). Mit $g(t, x) = L_f^{i-1} f(t, x)$ gilt damit

$$\begin{aligned} \frac{d^{i+1} x}{dt^{i+1}}(t) &= \frac{d}{dt} \frac{d^i x}{dt^i}(t) = \frac{d}{dt} L_f^{i-1} f(t, x(t)) = \frac{d}{dt} g(t, x(t)) \\ &= \frac{\partial g}{\partial t}(t, x(t)) + \frac{\partial g}{\partial x}(t, x(t)) \frac{d}{dt} x(t) \\ &= \frac{\partial g}{\partial t}(t, x(t)) + \frac{\partial g}{\partial x}(t, x(t)) f(t, x(t)) \\ &= L_f^1 g(t, x(t)) = L_f^1 L_f^{i-1} f(t, x(t)) = L_f^i f(t, x(t)), \end{aligned}$$

also gerade (6.14). □

Die Idee der Taylor-Verfahren ist nun denkbar einfach: Wir verwenden die Taylor-Formel und lassen den Restterm weg.

Definition 6.14 Das *Taylor-Verfahren der Ordnung* $p \in \mathbb{N}$ ist gegeben durch

$$\Phi(t, x, h) = x + \sum_{i=1}^p \frac{h^i}{i!} L_f^{i-1} f(t, x).$$

□

6.3.2 Eigenschaften

Der folgende Satz gibt die wesentlichen Eigenschaften der Taylor-Verfahren an.

Satz 6.15 Gegeben sei eine Differentialgleichung mit p -mal stetig differenzierbarem Vektorfeld $f : D \rightarrow \mathbb{R}^n$. Dann erfüllt das Taylor-Verfahren der Ordnung p die Lipschitzbedingung und ist konsistent mit Konsistenzordnung p .

Beweis: Wir zeigen zunächst die Lipschitzbedingung. Beachte, dass in der Formulierung der Taylor-Verfahren partielle Ableitungen von f bis zur Ordnung $p-1$ auftreten. Jede der auftretenden Funktionen $L_f^{i-1} f$ ist also ein weiteres mal stetig differenzierbar, woraus (mit dem Mittelwertsatz der Differentialrechnung) folgt, dass für jede kompakte Menge $K \subset D$ Lipschitz-Konstanten $L_i > 0$ existieren, so dass $L_f^{i-1} f$ Lipschitz in x mit dieser Konstante ist. Für die Funktion Φ gilt also für alle $h \leq 1$ die Abschätzung

$$\begin{aligned} \|\Phi(t, x_1, h) - \Phi(t, x_2, h)\| &\leq \|x_1 - x_2\| + \sum_{i=1}^p \frac{h^i}{i!} L_i \|x_1 - x_2\| \\ &\leq \|x_1 - x_2\| + \sum_{i=1}^p h L_i \|x_1 - x_2\| = (1 + Lh) \|x_1 - x_2\| \end{aligned}$$

mit

$$L = \sum_{i=1}^p L_i.$$

Dies ist gerade die gewünschte Lipschitz-Bedingung.

Die Konsistenz sowie die behauptete Konsistenzordnung folgt direkt aus Satz 6.13. □

Bemerkung 6.16 Wenn alle auftretenden Ableitungen auf ganz D beschränkt sind, so sind auch die Konstanten in den Lipschitz- und Konsistenzabschätzungen unabhängig von K gültig, man erhält also globale Fehlerabschätzungen. □

Beachte, dass das Taylor-Verfahren der Ordnung $p = 1$ durch

$$\Phi(t, x, h) = x + h L_f^0 f(t, x) = x + h f(t, x).$$

gegeben ist, also gerade das Euler-Verfahren ist. Dies führt sofort zu dem folgenden Korollar.

Korollar 6.17 Falls f einmal stetig differenzierbar ist, so ist das Euler-Verfahren konsistent mit Konsistenzordnung $p = 1$.

Beweis: Das Taylor-Verfahren der Ordnung $p = 1$ ist gerade das Euler-Verfahren, das also nach Satz 6.15 die Konsistenzordnung $p = 1$ besitzt. \square

Bemerkung 6.18 Mit einem direkten Beweis kann man die Konsistenzordnung $p = 1$ für das Euler-Verfahren auch beweisen, wenn f nur Lipschitz-stetig (in x und t) ist. Die Beweisidee geht wie folgt: Zunächst zeigt man, dass $\|x(t+h) - x(t)\| \leq C_1|h|$ für ein $C_1 > 0$ und alle hinreichend kleinen h ist; dies verwendet man dann, um

$$\int_t^{t+h} \|f(\tau, x(\tau)) - f(t, x(t))\| d\tau \leq C_2 h^2$$

für ein $C_2 > 0$ zu beweisen. Damit kann man schließlich die Konsistenzordnung zeigen. \square

Das Euler-Verfahren ist das einzige Taylor-Verfahren, bei dem keine Ableitungen des Vektorfeldes f auftreten. Das Auftreten der Ableitungen ist tatsächlich der Hauptgrund dafür, dass Taylor-Verfahren in der Praxis eher selten verwendet werden, da man dort Verfahren bevorzugt, die ohne explizite Verwendung der Ableitung funktionieren (auch wenn symbolische Mathematikprogramme wie z.B. MAPLE heutzutage zur automatischen Berechnung der benötigten Ableitungen verwendet werden können). Trotzdem gibt es Spezialanwendungen, in denen Taylor-Verfahren verwendet werden: Für hochgenaue Numerik, bei der Verfahren sehr hoher Ordnung ($p \geq 15$) benötigt werden, sind Taylor-Verfahren nützlich, da sie systematisch für beliebige Konsistenzordnungen hergeleitet werden können und die auftretenden Konstanten (in der Lipschitzbedingung und der Konsistenzabschätzung) durch genaue Analyse der Ableitungen und Restterme exakt abgeschätzt werden können.

6.4 Ableitungsfreie Verfahren

Um ein Taylor-Verfahren aufzustellen, muss man die Ableitungen des Vektorfeldes explizit berechnen, was im Allgemeinen recht aufwändige Berechnungen erfordert. Wir haben aber bereits in der Einleitung zwei Verfahren kennen gelernt, in denen man überhaupt keine Ableitung braucht, nämlich das Euler- und das Heun-Verfahren. Weitere Verfahren werden wir im übernächsten Abschnitt kennen lernen. Zunächst untersuchen wir aber, wie man die Konsistenzordnung beliebiger Einschrittverfahren analysieren kann.

6.4.1 Konsistenz beliebiger Einschrittverfahren

Die Konsistenzanalyse beliebiger Schemata führt man durch, indem man die Abbildung Φ mit den Koeffizienten der Taylor-Verfahren bzw. der Taylor-Entwicklung aus Satz 6.13 vergleicht. Genauer gilt der folgende Satz.

Satz 6.19 Sei $f : D \rightarrow \mathbb{R}^n$ p -mal stetig differenzierbar. Gegeben sei ein Einschrittverfahren $\Phi : \mathbb{R} \times \mathbb{R}^n \times \mathbb{R}$, das $p + 1$ -mal stetig differenzierbar ist. Dann besitzt Φ genau dann die Konsistenzordnung $p \in \mathbb{N}$, wenn die Bedingungen

$$\Phi(t, x, 0) = x \quad \text{und} \quad \frac{\partial^i \Phi}{\partial h^i}(t, x, 0) = L_f^{i-1} f(t, x) \quad \text{für } i = 1, \dots, p \quad (6.15)$$

für alle $(t, x) \in D$ gelten.

Beweis: Es bezeichne $\Phi_{T,p}$ das Taylor-Verfahren der Ordnung p . Die Taylor-Entwicklung von Φ nach der Variablen h in $h = 0$ ist gegeben durch

$$\Phi(t, x, h) = \Phi(t, x, 0) + \sum_{i=1}^p \frac{h^i}{i!} \frac{\partial^i \Phi}{\partial h^i}(t, x, 0) + O(h^{p+1}).$$

Sei nun (6.15) erfüllt. Dann liefert der Koeffizientenvergleich mit $\Phi_{T,p}$

$$\Phi(t, x, h) = \Phi_{T,p}(t, x, h) + O(h^{p+1})$$

Aus Satz (6.15) folgt daher

$$x(t+h; t, x) = \Phi_{T,p}(t, x, h) + O(h^{p+1}) = \Phi(t, x, h) + O(h^{p+1}),$$

was die Konsistenz zeigt.

Falls (6.15) nicht erfüllt ist, so gibt es $(t, x) \in D$, so dass entweder $\Phi(t, x, 0) \neq x$ gilt (in diesem Fall setzen wir $i^* = 0$) oder

$$\frac{\partial^{i^*} \Phi}{\partial h^{i^*}}(t, x, 0) \neq L_f^{i^*-1} f(t, x)$$

für ein $i^* \in \{1, \dots, p\}$ gilt. Wenn wir i^* minimal mit dieser Eigenschaft wählen, so folgt aus dem Koeffizientenvergleich mit $\Phi_{T,p}$, dass ein $C > 0$ existiert, so dass für alle hinreichend kleinen $h > 0$ die Ungleichung

$$\|\Phi(t, x, h) - \Phi_{T,p}(t, x, h)\| > Ch^{i^*}$$

gilt. Mit Satz (6.15) erhalten wir daher

$$\|x(t+h, t, x) - \Phi(t, x, h)\| > Ch^{i^*} - O(h^{p+1}) > \tilde{C}h^{i^*}$$

für geeignetes $0 < \tilde{C} < C$ und alle hinreichend kleinen $h > 0$, was der Konsistenz widerspricht. Also folgt die behauptete Äquivalenz. \square

Mit diesem Satz können wir die Konsistenzordnung beliebiger Einschrittverfahren überprüfen. Beachte, dass die Aussage über die Ordnung nur stimmt, wenn das Vektorfeld f hinreichend oft differenzierbar ist. Verfahren mit hoher Konsistenzordnung verlieren diese typischerweise, wenn das Vektorfeld der zu lösenden DGL nicht die nötige Differenzierbarkeit besitzt!

Ein wesentlicher Nachteil dieses Satzes ist, dass die Ausdrücke $L_f^i f(t, x)$ für große i sehr umfangreich und kompliziert werden. Hier können — wie bereits erwähnt — symbolische Mathematikprogramme wie MAPLE bei den Rechnungen helfen. Das folgende MAPLE Programm berechnet die Ableitungen $L_f^i f(t, x)$ für $i = 0, \dots, p$. (Vor der Ausführung muss der Variablen `p` natürlich ein Wert zugewiesen werden.)

```

> L[0]:=f(t,x);
> for i from 1 to p do
>   L[i] := simplify(diff(L[i-1],t) + diff(L[i-1],x)*f(t,x));
> od;

```

Die Ausgabe für $p:=3$ ist

$$L_0 := f(t, x)$$

$$L_1 := \left(\frac{\partial}{\partial t} f(t, x)\right) + \left(\frac{\partial}{\partial x} f(t, x)\right) f(t, x)$$

$$L_2 := \left(\frac{\partial^2}{\partial t^2} f(t, x)\right) + 2\left(\frac{\partial^2}{\partial x \partial t} f(t, x)\right) f(t, x) + \left(\frac{\partial}{\partial x} f(t, x)\right) \left(\frac{\partial}{\partial t} f(t, x)\right) \\ + \left(\frac{\partial^2}{\partial x^2} f(t, x)\right) f(t, x)^2 + f(t, x) \left(\frac{\partial}{\partial x} f(t, x)\right)^2$$

$$L_3 := \left(\frac{\partial^3}{\partial t^3} f(t, x)\right) + 3\left(\frac{\partial^3}{\partial x \partial t^2} f(t, x)\right) f(t, x) + 3\left(\frac{\partial^2}{\partial x \partial t} f(t, x)\right) \left(\frac{\partial}{\partial t} f(t, x)\right) \\ + \left(\frac{\partial}{\partial x} f(t, x)\right) \left(\frac{\partial^2}{\partial t^2} f(t, x)\right) + 3\left(\frac{\partial^3}{\partial x^2 \partial t} f(t, x)\right) f(t, x)^2 \\ + 3\left(\frac{\partial^2}{\partial x^2} f(t, x)\right) f(t, x) \left(\frac{\partial}{\partial t} f(t, x)\right) + \left(\frac{\partial}{\partial t} f(t, x)\right) \left(\frac{\partial}{\partial x} f(t, x)\right)^2 \\ + 5f(t, x) \left(\frac{\partial}{\partial x} f(t, x)\right) \left(\frac{\partial^2}{\partial x \partial t} f(t, x)\right) + \left(\frac{\partial^3}{\partial x^3} f(t, x)\right) f(t, x)^3 \\ + 4\left(\frac{\partial^2}{\partial x^2} f(t, x)\right) f(t, x)^2 \left(\frac{\partial}{\partial x} f(t, x)\right) + f(t, x) \left(\frac{\partial}{\partial x} f(t, x)\right)^3$$

Diese Ausdrücke gelten für den skalaren Fall $x \in \mathbb{R}$, für höhere Dimensionen muss das MAPLE-Programm erweitert werden.

Bemerkung 6.20 Man sieht, dass die Ausdrücke tatsächlich sehr unübersichtlich werden; ebenso ist das natürlich bei den entsprechenden Termen der Einschrittverfahren. Eine Hilfe hierfür bietet ein Formalismus, der von dem neuseeländischen Mathematiker J.C. Butcher in den 1960er Jahren entwickelt wurde, und bei dem die auftretenden Ableitungen mittels einer grafischen Repräsentierung in einer Baumstruktur übersichtlich strukturiert werden. \square

Satz 6.19 kann auf verschiedene Art und Weise angewendet werden. Eine Möglichkeit ist, jede Ableitung in einem Taylor-Verfahren durch einen Differenzenquotienten zu ersetzen, der die entsprechende Ableitung mit hinreichend hoher Genauigkeit approximiert. Als Beispiel dafür betrachten wir das Taylor-Verfahren zweiter Ordnung, das gegeben ist durch

$$\Phi_{T,2}(t, x, h) = x + hL_f^0 f(t, x) + \frac{h^2}{2} L_f^1 f(t, x) = x + hf(x, u) + \frac{h^2}{2} \frac{\partial f}{\partial t} f(t, x) + \frac{h^2}{2} \frac{\partial f}{\partial x}(t, x) f(t, x).$$

Mit der Approximation

$$\begin{aligned} \frac{\partial f}{\partial t} f(t, x) + \frac{\partial f}{\partial x}(t, x) f(t, x) &= \left. \frac{d}{ds} \right|_{s=0} f(t + s, x + s f(t, x)) \\ &= \frac{f(t + h, x + h f(t, x)) - f(t, x)}{h} + O(h) \end{aligned}$$

erhalten wir

$$\Phi_{T_2}(t, x, h) = x + h f(t, x) + \frac{h^2}{2} \left(\frac{f(t + h, x + h f(t, x)) - f(t, x)}{h} \right) + O(h^3).$$

Damit folgt, dass

$$\begin{aligned} \Phi(t, x, h) &= x + h f(t, x) + \frac{h^2}{2} \left(\frac{f(t + h, x + h f(t, x)) - f(t, x)}{h} \right) \\ &= x + \frac{h}{2} (f(t, x) + f(t + h, x + h f(t, x))) \end{aligned}$$

den Satz 6.19 für $p = 2$ erfüllt. Das so erhaltene Einschrittverfahren Φ ist tatsächlich nichts anderes als das bereits bekannte Heun-Verfahren.

6.4.2 Explizite Runge-Kutta-Verfahren

Das gerade skizzierte Vorgehen führt bei höheren Konsistenzordnungen schnell an seine praktischen Grenzen, weil man Ableitungen hoher Ordnung geeignet approximieren muss. Hier ist ein anderes Vorgehen effizienter: wir formulieren zunächst eine ganze Klasse ableitungsfreier Verfahren, die von einer Reihe von Parameter abhängen und verwenden Satz 6.19 dann, um die Parameter zu bestimmen.

Die Klasse von Verfahren, die wir hier betrachten, sind die sogenannten expliziten Runge-Kutta-Verfahren, deren Idee wir am Heun-Verfahren erläutern. Bei der Konstruktion des Heun-Verfahrens haben wir das Euler-Verfahren verwendet, um einen Schätzwert für den unbekanntes Wert $x(t_{i+1})$ zu erhalten. Es liegt nun nahe, diese Methode systematisch rekursiv anzuwenden, um zu Verfahren höherer Konsistenzordnung zu gelangen. Genau dies ist die Grundidee der Runge-Kutta-Verfahren.

Um die dabei entstehenden Verfahren übersichtlich zu schreiben, benötigen wir einen geeigneten Formalismus. Wir erläutern diesen am Beispiel des Heun-Verfahrens

$$\Phi(t, x, h) = x + \frac{h}{2} \left(f(t, x) + f(t + h, x + h f(t, x)) \right).$$

Wir schreiben dieses nun als

$$\begin{aligned} k_1 &= f(t, x) \\ k_2 &= f(t + h, x + h k_1) \\ \Phi(t, x, h) &= x + h \left(\frac{1}{2} k_1 + \frac{1}{2} k_2 \right) \end{aligned}$$

Was zunächst vielleicht komplizierter als die geschlossene Formel aussieht, erweist sich als sehr günstige Schreibweise, wenn man weitere k_i -Terme hinzufügen will. Dies ist gerade die Schreibweise der expliziten Runge-Kutta-Verfahren.

Definition 6.21 Ein s -stufiges explizites Runge-Kutta-Verfahren ist gegeben durch

$$k_i = f \left(t + c_i h, x + h \sum_{j=1}^{i-1} a_{ij} k_j \right) \quad \text{für } i = 1, \dots, s$$

$$\Phi(t, x, h) = x + h \sum_{i=1}^s b_i k_i.$$

Den Wert $k_i = k_i(t, x, h)$ bezeichnen wir dabei als i -te Stufe des Verfahrens. □

Die Koeffizienten eines Runge-Kutta-Verfahrens können wir mittels

$$b = \begin{pmatrix} b_1 \\ b_2 \\ b_3 \\ \vdots \\ b_s \end{pmatrix} \in \mathbb{R}^s, \quad c = \begin{pmatrix} c_1 \\ c_2 \\ c_3 \\ \vdots \\ c_s \end{pmatrix} \in \mathbb{R}^s, \quad \mathcal{A} = \begin{pmatrix} 0 & & & & & \\ a_{21} & 0 & & & & \\ a_{31} & a_{32} & 0 & & & \\ \vdots & \vdots & \ddots & \ddots & & \\ a_{s1} & \cdots & \cdots & a_{s,s-1} & 0 & \end{pmatrix} \in \mathbb{R}^{s \times s}$$

kompakt schreiben. Konkrete Verfahren werden meist in Form des Butcher-Tableaus (oder Butcher-Schemas)

$$\begin{array}{c|ccc} c_1 & & & \\ c_2 & a_{21} & & \\ c_3 & a_{31} & a_{32} & \\ \vdots & \vdots & \vdots & \ddots \\ c_s & a_{s1} & a_{s2} & \cdots & a_{s,s-1} \\ \hline & b_1 & b_2 & \cdots & b_{s-1} & b_s \end{array}$$

geschrieben, das wiederum auf J.C. Butcher zurückgeht.

Einfache Beispiele solcher Verfahren sind das Euler-Verfahren ($s = 1$), das Heun-Verfahren ($s = 2$) und das sogenannte *klassische Runge-Kutta-Verfahren* ($s = 4$), das von C. Runge² und M. Kutta³ entwickelt wurde, und dem die ganze Verfahrensklasse ihren Namen verdankt. Diese Verfahren sind (von links nach rechts) gegeben durch die Butcher-Tableaus

$$\begin{array}{c|c} 0 & \\ \hline & 1 \end{array} \quad \begin{array}{c|cc} 0 & & \\ 1 & 1 & \\ \hline & \frac{1}{2} & \frac{1}{2} \end{array} \quad \begin{array}{c|ccc} 0 & & & \\ \frac{1}{2} & \frac{1}{2} & & \\ \frac{1}{2} & 0 & \frac{1}{2} & \\ 1 & 0 & 0 & 1 \\ \hline & \frac{1}{6} & \frac{2}{6} & \frac{2}{6} & \frac{1}{6} \end{array}$$

²deutscher Mathematiker, 1856–1927

³deutscher Mathematiker und Ingenieur, 1867–1944

Beachte, dass das Euler-Verfahren sowohl das einfachste Runge-Kutta-Verfahren als auch das einfachste Taylor-Verfahren ist; es ist das einzige Verfahren, das in beiden Klassen liegt, da alle Runge-Kutta-Verfahren per Definition ohne Ableitungen von f auskommen, was gegenüber den Taylor-Verfahren einen großen Vorteil darstellt.

Es ist in diesem Zusammenhang interessant, den Aufwand des Heun-Verfahrens und des Taylor-Verfahrens der Ordnung 2 z.B. für $x \in \mathbb{R}$ zu vergleichen, die ja die gleiche Konsistenzordnung besitzen. Beim Taylor-Verfahren der Ordnung 2 müssen in jedem Schritt $L_f^0 f(t, x) = f(t, x)$ und

$$L_f^1 f(t, x) = \frac{\partial f}{\partial t}(t, x) + \frac{\partial f}{\partial x}(t, x)f(t, x)$$

ausgewertet werden, also 3 Funktionsauswertungen; beim Heun Verfahren müssen $k_1 = f(t, x)$ und $f(t + h, x + hk_1)$, also 2 Funktionen ausgewertet werden. Der Aufwand ist folglich nur $2/3$ so groß. Dieser geringere Aufwand, der bei höherer Konsistenzordnung noch deutlicher ausfällt, ist typisch für Runge-Kutta-Verfahren, ein weiterer Vorteil gegenüber den Taylor-Verfahren.

Beachte, dass Runge-Kutta-Verfahren immer die Lipschitz-Bedingung erfüllen, wenn das Vektorfeld f Lipschitz-stetig im Sinne von Definition 6.4 ist: Mittels Induktion sieht man leicht, dass jede Stufe k_i Lipschitz-stetig ist. Damit gilt dies auch für ihre Summe, weswegen Φ die gewünschte Bedingung erfüllt.

Satz 6.19 kann nun verwendet werden, um Bedingungen an die Koeffizienten des Verfahrens herzuleiten, unter denen Konsistenz und die Erfüllung einer gegebenen Konsistenzordnung gewährleistet ist. Details zur Herleitung der folgenden Bedingungen finden sich in meinem Skript "Numerische Methoden für gewöhnliche Differentialgleichungen".

Es gilt: Ein explizites Runge-Kutta-Verfahren ist genau dann konsistent, wenn die Bedingung

$$\sum_{i=1}^s b_i = 1$$

erfüllt ist. Gilt darüber hinaus die Bedingung

$$c_i = \sum_{j=1}^{i-1} a_{ij},$$

so folgt für alle Vektorfelder $f \in C^p(D, \mathbb{R}^n)$:

- (i) Das Verfahren besitzt genau dann die Konsistenzordnung $p = 1$, wenn die Gleichung

$$\sum_i b_i = 1$$

gilt.

- (ii) Es besitzt genau dann die Konsistenzordnung $p = 2$, wenn zusätzlich zu (i) die Gleichung

$$\sum_i b_i c_i = 1/2$$

gilt.

- (iii) Es besitzt genau dann die Konsistenzordnung $p = 3$, wenn zusätzlich zu (i), (ii) die Gleichungen

$$\sum_i b_i c_i^2 = 1/3, \quad \sum_{ij} b_i a_{ij} c_j = 1/6$$

gelten.

- (iv) Es besitzt genau dann die Konsistenzordnung $p = 4$, wenn zusätzlich zu (i)–(iii) die Gleichungen

$$\begin{aligned} \sum_i b_i c_i^3 &= 1/4, & \sum_{ij} b_i a_{ij} c_i c_j &= 1/8 \\ \sum_{ij} b_i a_{ij} c_j^2 &= 1/12, & \sum_{ijk} b_i a_{ij} a_{jk} c_k &= 1/24 \end{aligned}$$

gelten.

Hierbei laufen die Summations-Indizes in den Grenzen $i = 1, \dots, s$, $j = 1, \dots, i - 1$ und $k = 1, \dots, j - 1$. Diese Gleichungen an die Koeffizienten werden *Bedingungsgleichungen* genannt. Wie komplex das Problem des Aufstellens der Bedingungsgleichungen für große p wird, zeigt die folgende Tabelle, die die Anzahl der Gleichungen für gegebenes p angibt.

Konsistenzordnung p	1	2	3	4	5	6	7	8	9	10	20
Anzahl Bedingungsgl'en	1	2	4	8	17	37	85	200	486	1205	20247374

Nicht nur das Aufstellen, auch das Lösen dieser (nichtlinearen!) Gleichungssysteme wird ziemlich komplex. Hier kommt wieder das in Bemerkung 6.20 bereits erwähnte grafische Verfahren von Butcher ins Spiel. Mit diesem Verfahren können die einzelnen Terme der L_f^i -Ableitungen ebenso wie die Terme der Ableitungen von Φ mittels einer Baumstruktur grafisch dargestellt werden. Dieses Verfahren erlaubt eine Einsicht in die Struktur dieser riesigen nichtlinearen Gleichungssysteme, womit es gelungen ist, die Gleichungen bis $p = 10$ (ohne Computerhilfe) zu lösen. Eine wichtige Rolle spielt dabei natürlich die Stufenzahl s der betrachteten Verfahren. Insbesondere ist hierbei wichtig, wie viele Stufen s man zur Realisierung einer gegebenen Konsistenzordnung p benötigt. Die folgende Tabelle gibt die ebenfalls durch Butcher (in den Jahren 1964–1985) berechneten bekannten minimalen Schranken an.

Konsistenzordnung p	1	2	3	4	5	6	7	8	≥ 9
minimale Stufenzahl s	1	2	3	4	6	7	9	11	$\geq p + 3$

Der Eintrag für $p \geq 9$ bedeutet nicht, dass für jedes $p \geq 9$ ein Verfahren mit $s = p + 3$ Stufen bekannt ist, sondern dass es kein Verfahren mit weniger Stufen geben kann. Für $p = 10$ wurde 1978 von E. Hairer ein Verfahren mit $s = 17$ Stufen angegeben, das sich im Guinness-Buch der Rekorde findet. Möglichst wenig Stufen zu verwenden ist allerdings nicht das einzige Qualitätsmerkmal für Runge-Kutta-Verfahren, oftmals spielen andere Kriterien eine wichtigere Rolle. Details dazu finden sich ebenfalls in meinem Skript "Numerische Methoden für gewöhnliche Differentialgleichungen".

Kapitel 7

Numerische Lösung stochastischer Differentialgleichungen

In diesem Kapitel geben wir eine Einführung in die numerische Lösung stochastischer Itô-Differentialgleichungen (4.5). Aus Zeitgründen werden wir dabei einige Aussagen nicht beweisen können und statt dessen auf die Literatur verweisen, speziell auf das Buch von Kloeden und Platen [3].

Ziel des Kapitels ist zum einen das Bereitstellen von Algorithmen, mit denen Approximationen $\tilde{X}(T) \approx X(T, X_0)$ der Lösung von (4.5) zum Anfangswert X_0 berechnet werden können. Dies werden wiederum Einschrittverfahren sein, die wir als Verallgemeinerung der Einschrittverfahren für deterministische Differentialgleichungen aus dem vorhergehenden Kapitel auffassen können. Wir beginnen dabei mit dem stochastischen Euler-Verfahren und zeigen, wie dieses in der Optionsbewertung verwendet werden kann. Danach betrachten wir das stochastische Analogon der Taylor-Entwicklung aus Satz 6.13, das analog zum deterministischen Fall zum einen zur Entwicklung von stochastischen Taylor-Verfahren verwendet werden kann und zum anderen zur Analyse der Konvergenzordnung ableitungsfreier Verfahren. Wir werden dabei sehen, dass die Konstruktion von Verfahren höherer Ordnung $p \geq 2$ für stochastische Differentialgleichungen sehr viel schwieriger ist als im deterministischen Fall. Zum Abschluss werden wir dann noch zeigen, dass — und in welchem Sinne — das in der Binomialmethode verwendete Modell eine numerische Approximation der geometrischen Brown'schen Bewegung darstellt.

Bevor wir konkrete Verfahren betrachten, wollen wir im folgenden ersten Abschnitt aber klären, was man für stochastische Differentialgleichungen unter einer numerischen Approximation überhaupt versteht.

7.1 Stochastische Konvergenz- und Approximationsbegriffe

Bei der numerischen Lösung stochastischer Differentialgleichungen kann man im Allgemeinen nicht erwarten, dass man für jeden möglichen Pfad $X(T, X_0, \omega)$ eine gute Approximation $\tilde{X}(T, \omega)$ berechnen kann. Dies liegt daran, dass manche Pfade für die numerische

Approximation sehr unangenehme Eigenschaften haben können, z.B. können sie ihre Richtung innerhalb eines kurzen Zeitintervalls sehr schnell und stark ändern, was sich mit einem Einschrittverfahren nur durch Verwendung eines sehr feinen Gitters reproduzieren lässt. Eine gute Approximation jedes einzelnen Pfades ist aber nicht unbedingt notwendig: Für die allermeisten Anwendungen reicht es aus, wenn die berechneten Approximationen im Mittel hinreichend gut sind und da Pfade mit numerisch unangenehmen Eigenschaften mit relativ geringer Wahrscheinlichkeit auftreten, kann man dies auch mit nicht zu feinen Gittern erreichen.

Es gibt allerdings verschiedene Möglichkeiten, wie man dieses “im Mittel” definieren kann und je nach dem, was man berechnen möchte, benötigt man verschiedene Konvergenzbegriffe. Tatsächlich gibt es in der stochastischen Numerik eine Vielzahl von Konvergenzbegriffen, von denen wir hier nur zwei herausheben wollen, die für unsere Zwecke besonders wichtig sind.

Wir betrachten dabei einen stochastischen Prozess X auf dem Intervall $[0, T]$. Dieser soll durch eine Folge numerischer Approximationen \tilde{X}_j auf Zeitgittern der Form $\mathcal{T}_j = \{0, \bar{h}_j, 2\bar{h}_j, \dots, N_j\bar{h}_j\}$ mit konstanten maximalen Schrittweiten $\bar{h}_j = T/N_j$ (also $N_j\bar{h}_j = T$) approximiert werden mit $N_j \rightarrow \infty$ für $j \rightarrow \infty$. Die folgende Definition bildet das stochastische Gegenstück der Konvergenz einer diskreten Approximation, vgl. Definition 6.5. Im Gegensatz zu den deterministischen Gleichungen beschränken wir uns hier darauf, die Approximation zum Zeitpunkt T zu betrachten, d.h. nicht wie dort an allen Gitterpunkten in \mathcal{T}_j . Wir schreiben dazu kurz $X(T) = X(T, \omega)$ und $\tilde{X}_j(T) = \tilde{X}_j(T, \omega)$.

Definition 7.1 (i) Die Folge \tilde{X}_j von stochastischen Prozessen heißt *starke Approximation* für X zur Zeit T bzgl. einer Funktion $g : \mathbb{R}^n \rightarrow \mathbb{R}^m$, wenn die Bedingung

$$\lim_{j \rightarrow \infty} E(\|g(X(T)) - g(\tilde{X}_j(T))\|) = 0$$

gilt. Sie heißt *starke Approximation mit Ordnung $\gamma > 0$* , falls für alle $j \geq j_0$ zusätzlich die Abschätzung

$$E(\|g(X(T)) - g(\tilde{X}_j(T))\|) \leq Ch_j^\gamma$$

für ein $C > 0$ gilt.

(ii) Die Folge \tilde{X}_j von stochastischen Prozessen heißt *schwache Approximation* für X zur Zeit T bzgl. einer Funktion $g : \mathbb{R}^n \rightarrow \mathbb{R}^m$, wenn die Bedingung

$$\lim_{j \rightarrow \infty} \|E(g(X(T))) - E(g(\tilde{X}_j(T)))\| = 0$$

gilt. Sie heißt *schwache Approximation mit Ordnung $\beta > 0$* , falls für alle $j \geq j_0$ zusätzlich die Abschätzung

$$\|E(g(X(T))) - E(g(\tilde{X}_j(T)))\| \leq Ch_j^\beta$$

für ein $C > 0$ gilt. □

Eine starke Approximation liefert also eine Approximation der Pfade, bei denen zwar u.U. nicht jeder Pfad gut approximiert wird, der Fehler aber zumindest im Mittel klein werden muss. Dies ist die natürliche Verallgemeinerung der deterministischen Approximation.

Bei einer schwachen Approximation hingegen (die sich, wie wir sehen werden, mit deutlich geringerem numerischen Aufwand berechnen lässt), können die einzelnen Pfade ganz anders aussehen, wichtig ist hier nur, dass die bezüglich g ermittelten statistischen Eigenschaften gleich sind.

Kurz gesagt können wir festhalten: Eine starke Approximation liefert eine *Approximation der Pfade* von X während eine schwache Approximation eine *Approximation der statistischen Eigenschaften* von X liefert, jeweils gemessen bzgl. g .

Beachte, dass wir die Konstante C in der Abschätzung der Approximationsordnung hier nicht wie bei den deterministischen Gleichungen für eine kompakten Menge von Anfangsbedingungen sondern nur für eine feste Lösung definiert haben. Ebenso betrachten wir nur äquidistante Gitter. Beides hat vor allem den Grund, die Konstruktion und Analyse der Verfahren technisch etwas zu vereinfachen.

Das folgende Beispiel zeigt typische Anwendungen starker und schwacher Approximationen.

Beispiel 7.2 (i) Wenn wir den Erwartungswert $E(X(T))$ numerisch berechnen wollen, genügt eine schwache Approximation bzgl. $g(x) = x$, denn dafür gilt

$$\lim_{j \rightarrow \infty} |E(X(T)) - E(\tilde{X}_j(T))| = 0,$$

und damit

$$\lim_{j \rightarrow \infty} E(\tilde{X}_j(T)) = E(X(T)).$$

(ii) Wenn wir die Varianz $\text{Var}(X(T))$ numerisch berechnen wollen, genügt eine schwache Approximation bzgl. $g(x) = x$ und $g(x) = x^2$.

(iii) Für die Optionsbewertung genügt eine schwache Approximation bzgl. $g(x) = (x - K)^+$ bzw. $g(x) = (K - x)^+$.

(iv) Für kompliziertere Berechnungen reichen schwache Approximationen allerdings i.A. nicht mehr aus. Wenn wir z.B. für einen Wert $c \in \mathbb{R}$ die “minimale Überschreitungszeit”

$$t(\omega) := \inf\{t \geq 0 \mid X(t, \omega) \geq c\}$$

für ein vorgegebenes $c \in \mathbb{R}$ definieren, so erhalten wir eine weitere Zufallsvariable $t(\omega)$. Dies könnte z.B. die Überschreitung eines vorgegebenen Kursniveaus sein, was zur Bewertung von Barriere-Optionen wichtig ist. Bei diesen Optionstypen hängt der Wert der Option davon ab, ob der Wert des Kurses bis zum Laufzeitende ein vorgegebenes Kursniveau über- oder unterschreitet. Ziel einer numerischen Berechnung könnte es nun sein, die mittlere Zeit $E\{t(\omega)\}$ zu ermitteln. Da sich diese Größe nicht als $E\{g(X(T))\}$ schreiben lässt, genügt hier eine schwache Approximation nicht mehr. □

Dass die starke Approximation tatsächlich eine stärkere Eigenschaft als die schwache Approximation ist, zeigt das folgende Lemma, in dem wir uns der Einfachheit halber auf reellwertige g beschränken.

Lemma 7.3 Wenn $\tilde{X}_j(T)$ eine starke Approximation von $X(T)$ bzgl. $g : \mathbb{R}^n \rightarrow \mathbb{R}$ ist, so ist $\tilde{X}_j(T)$ auch eine schwache Approximation bzgl. g . Hierbei bleibt die Konvergenzordnung erhalten, d.h. es gilt $\beta \geq \gamma$.

Beweis: Da der Erwartungswert monoton ist, folgt

$$E(g(X(T)) - g(\tilde{X}(T))) \leq E(|g(X(T)) - g(\tilde{X}(T))|)$$

und

$$E(g(\tilde{X}(T)) - g(X(T))) \leq E(|g(\tilde{X}(T)) - g(X(T))|) = E(|g(X(T)) - g(\tilde{X}(T))|).$$

Damit folgt

$$|E(g(\tilde{X}(T)) - g(X(T)))| \leq E(|g(X(T)) - g(\tilde{X}(T))|)$$

und damit die Behauptung. \square

Eine nützliche Eigenschaft für starke Approximationen gibt das folgende Lemma.

Lemma 7.4 Es sei $h : \mathbb{R}^m \rightarrow \mathbb{R}^l$ eine global Lipschitz stetige Funktion. Falls \tilde{X}_i eine starke Approximation mit einer gegebenen Ordnung für eine Funktion $g : \mathbb{R}^n \rightarrow \mathbb{R}^m$ ist, so ist \tilde{X}_i auch eine starke Approximation mit einer gegebenen Ordnung für die Funktion $h \circ g$.

Beweis: Es sei $L > 0$ die globale Lipschitz-Konstante von h . Für den Ausdruck, der die starke Konvergenz beschreibt, gilt dann

$$E(\|h \circ g(X(T)) - h \circ g(\tilde{X}_i(T))\|) \leq E(L\|g(X(T)) - g(\tilde{X}_i(T))\|) = LE(\|g(X(T)) - g(\tilde{X}_i(T))\|),$$

weswegen sich sowohl die starke Konvergenz als auch die dazugehörige Ordnung von g auf $h \circ g$ übertragen. \square

Leider gilt diese praktische Eigenschaft für die schwache Approximation nicht.

7.2 Approximationen des Wiener-Prozesses

Bevor wir uns der Lösung stochastischer Differentialgleichungen widmen, wollen wir zuerst untersuchen, wie wir den Wiener-Prozess approximieren können. Diese sind für die im Folgenden entwickelten numerischen Verfahren wichtig: Da der Wiener-Prozess in die Definition der stochastischen Differentialgleichungen eingeht, werden diese approximativen Wiener-Prozesse in den numerischen Schemata benötigt.

Algorithmus 7.5 (Starke Approximation des Wiener-Prozesses durch Gitterfunktion)

Gegeben: Schrittweite h , Gitter $\mathcal{T} = \{t_0, \dots, t_N\}$ mit $t_i = ih$

Gesucht: M approximative Pfade $\tilde{W}(t_i, \omega_1), \dots, \tilde{W}(t_i, \omega_M)$ des Wiener Prozesses auf \mathcal{T} .

(1) Für $j = 1, \dots, M$:

(2a) Erzeuge unabhängige $N(0, h)$ -verteilte Zufallszahlen $\Delta W_i(\omega_j)$ für $i = 0, \dots, N - 1$

(2b) Erzeuge Gitterfunktionen $\tilde{W}(t_i, \omega_j)$ mittels der Rekursion

$$\tilde{W}(t_0, \omega_j) = 0, \quad \tilde{W}(t_{i+1}, \omega_j) = \tilde{W}(t_i, \omega_j) + \Delta W_i(\omega_j), \quad i = 0, \dots, N - 1$$

(3) Ende der j -Schleife \square

Man kann beweisen, dass dieser approximative Wiener-Prozess die Bedingungen des exakten Wiener-Prozesses an allen Gitterpunkten t_i erfüllt ((i) und (ii) sind Übungsaufgabe, (iii) folgt aus der Unabhängigkeit der Zufallszahlen in Algorithmus 7.5). Daher kann man jedem Pfad $\widetilde{W}(\cdot, \omega)$ des approximierten Wiener-Prozesses gerade einen Pfad des exakten Prozesses $W(\cdot, \omega)$ mit $\widetilde{W}(t_i, \omega) = W(t_i, \omega)$ für jedes $t_i \in \mathcal{T}$ zuordnen, so dass die Approximation für jedes $g : \mathbb{R} \rightarrow \mathbb{R}$ und jeden Gitterpunkt $t_i \in \mathcal{T}$ die Bedingung

$$E\{|g(\widetilde{W}(t_i, \omega)) - g(W(t_i, \omega))|\} = 0$$

erfüllt. Es handelt sich also um eine “besonders starke” Form der starken Konvergenz.

Wir wollen nun untersuchen, wie man eine schwache Approximation des Wiener-Prozesses erhalten kann — in der Hoffnung, dass sich der Algorithmus dabei vereinfacht. Der folgende Algorithmus zeigt, wie dies geht. Wir benötigen eine spezielle Form von gleichverteilten diskreten Zufallsvariablen, nämlich die durch

$$X(\Omega) = \{x_1, x_2\}, \quad \mathbb{P}_X(\{x_1\}) = \mathbb{P}_X(\{x_2\}) = \frac{1}{2}$$

definierte *zweipunktverteilte Zufallsvariable* mit den zwei Werten $\{x_1, x_2\}$.

Algorithmus 7.6 (Schwache Approximation des Wiener-Prozesses durch Gitterfunktion)

Gegeben: Schrittweite h , Gitter $\mathcal{T} = \{t_0, \dots, t_N\}$ mit $t_i = ih$

Gesucht: M schwach approximierende Pfade $\widetilde{W}(T, \omega_1), \dots, \widetilde{W}(T, \omega_M)$ des Wiener Prozesses.

(1) Für $j = 1, \dots, M$:

(2a) Erzeuge unabhängige zweipunktverteilte Zufallszahlen $\Delta W_i(\omega_j)$ für $i = 0, \dots, N - 1$ mit $x_1 = -\sqrt{h}$, $x_2 = \sqrt{h}$.

(2b) Erzeuge Gitterfunktionen $\widetilde{W}(t_i, \omega_j)$ mittels der Rekursion

$$\widetilde{W}(t_0, \omega_j) = 0, \quad \widetilde{W}(t_{i+1}, \omega_j) = \widetilde{W}(t_i, \omega_j) + \Delta W_i(\omega_j), i = 0, \dots, N - 1$$

(3) Ende der j -Schleife □

Der große Unterschied in der Konstruktion besteht darin, dass wir hier in jedem Schritt nur endlich viele (nämlich gerade 2) Möglichkeiten für $\Delta W_i(\omega_j)$ haben. Die Anzahl der möglichen Pfade ist demnach endlich, weswegen man hier tatsächlich alle möglichen Pfade mitsamt der Wahrscheinlichkeiten ihres Auftretens berechnet.

Abbildung 7.1 zeigt einen Pfad des Wienerprozesses (durchgezogen) sowie seine schwache Approximation (gestrichelt). Um vergleichbare Pfade zu erhalten, wurden die Zufallszahlen in den beiden Algorithmen so gekoppelt, dass sie jeweils gleiches Vorzeichen besitzen. Man sieht, dass die Pfade erheblich voneinander abweichen können.

Dass dieser Algorithmus trotzdem die richtigen statistischen Eigenschaften besitzt, zeigt das folgende Lemma.

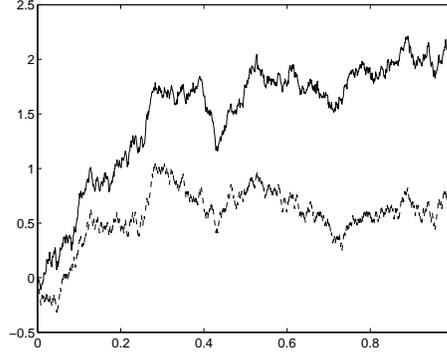


Abbildung 7.1: Schwache Approximation (- -) eines Pfades eines Wiener-Prozesses (—)

Lemma 7.7 Der Algorithmus 7.6 liefert eine schwache Approximation für den Wiener-Prozess bzgl. $g(x) = x$ und $g(x) = x^2$ für jeden Gitterpunkt $t_i \in \mathcal{T}$.

Beweis: Für die zweipunktverteilten Zufallszahlen ΔW_i berechnet man

$$E\{\Delta W_i(\cdot)\} = -\sqrt{h}/2 + \sqrt{h}/2 = 0 \text{ und } E\{\Delta W_i(\cdot)^2\} = h/2 + h/2 = h.$$

Aus der Definition folgt

$$\widetilde{W}(t_i, \omega) = \sum_{k=0}^{i-1} \Delta W_k(\omega).$$

Also ergibt sich

$$E\{\widetilde{W}(t_i, \cdot)\} = \sum_{k=0}^{i-1} E\{\Delta W_k(\cdot)\} = 0 = E\{W(t_i, \cdot)\}$$

und, indem wir die Unabhängigkeit der Zufallszahlen und (4.7) ausnutzen, auch

$$\begin{aligned} E\{\widetilde{W}(t_i, \cdot)^2\} &= E\left\{\left(\sum_{k=0}^{i-1} \Delta W_k(\cdot)\right)^2\right\} \\ &= E\left\{\sum_{k=0}^{i-1} \Delta W_k(\cdot)^2 + \sum_{k \neq j} \Delta W_k(\cdot) \Delta W_j(\cdot)\right\} \\ &= \sum_{k=1}^i \underbrace{E\{\Delta W_k(\cdot)^2\}}_{=h} + \sum_{k \neq j} \underbrace{E\{\Delta W_k(\cdot)\} E\{\Delta W_j(\cdot)\}}_{=0} \\ &= hi = t_i = E\{W(t_i, \cdot)^2\}, \end{aligned}$$

denn es gilt

$$E\{W(t_i, \cdot)^2\} = E\{\widetilde{W}(t_i, \cdot)^2\} - \underbrace{E\{\widetilde{W}(t_i, \cdot)\}^2}_{=0} = \text{Var}(W(t_i, \cdot)) = t_i.$$

□

7.3 Das stochastische Euler-Verfahren

Das stochastische Euler-Verfahren lässt sich genau wie sein deterministisches Gegenstück heuristisch herleiten. Wir betrachten dazu vektorwertige stochastische Differentialgleichungen, deren Lösungen von mehreren Wiener Prozessen W^1, \dots, W^m abhängen. Diese sind von der Form

$$dX(t) = a(t, X(t))dt + \sum_{j=1}^m b_j(t, X(t))dW_t^j. \quad (7.1)$$

Hierbei ist $X(t, \omega) \in \mathbb{R}^n$,

$$a = (a_1, \dots, a_n)^T : \mathbb{R} \times \mathbb{R}^n \rightarrow \mathbb{R}^n$$

mit $a_k : \mathbb{R} \times \mathbb{R}^n \rightarrow \mathbb{R}$ für $k = 1, \dots, n$ und

$$b_j = (b_{j,1}, \dots, b_{j,n})^T : \mathbb{R} \times \mathbb{R}^n \rightarrow \mathbb{R}^n$$

mit $b_{j,k} : \mathbb{R} \times \mathbb{R}^n \rightarrow \mathbb{R}$ für $j = 1, \dots, m$ und $k = 1, \dots, n$. Wir erinnern daran, dass (7.1) eine Kurzschreibweise ist für die Integralgleichung

$$X(t) = X(t_0) + \int_{t_0}^t a(\tau, X(\tau))d\tau + \sum_{j=1}^m \int_{t_0}^t b_j(\tau, X(\tau))dW_\tau^j. \quad (7.2)$$

Für die Lösung zur Zeit $t + h$ gilt also

$$X(t+h) = X(t) + \int_t^{t+h} a(\tau, X(\tau))d\tau + \sum_{j=1}^m \int_t^{t+h} b_j(\tau, X(\tau))dW_\tau^j,$$

für die mit $a(\tau, X(\tau)) \approx a(t, X(t))$, $b_j(\tau, X(\tau)) \approx b_j(t, X(t))$ und der Definition des Itô-Integrals (vgl. auch (4.9)) die Approximation

$$\approx X(t) + ha(t, X(t)) + \sum_{j=1}^m \Delta W^j(t) b_j(t, X(t))$$

mit $\Delta W^j(t) = W^j(t+h) - W^j(t)$ gilt.

Dies führt auf das stochastische Euler-Verfahren

$$\Phi(t, X, h, W, \omega) = X(\omega) + ha(t, X(\omega)) + \sum_{j=1}^m \Delta W^j(t, \omega) b_j(t, X(\omega)), \quad (7.3)$$

das auch als Euler-Maruyama-Verfahren bezeichnet wird. Hierbei schreiben wir kurz $W = (W^1, \dots, W^m)$. Für die geometrische Brownsche Bewegung (4.10) (hier gilt $m = 1$) ergibt sich die Verfahrensfunktion Φ gerade zu

$$\Phi(t, X, h, W, \omega) = X(\omega) + h\mu X(\omega) + \Delta W(t, \omega)\sigma X(\omega). \quad (7.4)$$

Dies ist das einfachste Beispiel eines *stochastischen Einschnittverfahrens*, deren Anwendung wir als Algorithmus formulieren wollen.

Algorithmus 7.8 (Lösung einer SDG mit stoch. Einschritt-Verfahren)

Gegeben: Schrittweite h , Gitter $\mathcal{T} = \{t_0, \dots, t_N\}$ mit $t_i = ih$

Gesucht: M approximative Pfade $\tilde{X}(\cdot, \omega_1), \dots, \tilde{X}(\cdot, \omega_M)$ für die SDG (7.1) auf \mathcal{T} mit $X(t_0, \omega_j) = x_0$.

(0) Erzeuge für jeden Wiener-Prozess W^j , $j = 1, \dots, m$, je M approximative Pfade $\tilde{W}^j(\cdot, \omega_k)$, $k = 1, \dots, M$ auf \mathcal{T}

(1) Für $k = 1, \dots, M$:

(2) Erzeuge Gitterfunktionen $\tilde{X}(t_i, \omega_k)$ mittels der Rekursion $\tilde{X}(t_0, \omega_j) = x_0$,

$$\tilde{X}(t_{i+1}, \omega_k) = \Phi(t_i, \tilde{X}(t_i, \omega_k), h, \tilde{W}^j, \omega_k)$$

für $i = 0, \dots, N - 1$

(3) Ende der k -Schleife □

Natürlich kann man die Implementierung noch optimieren, statt z.B. die Pfade $\tilde{W}^j(\cdot, \omega_k)$ im Voraus zu berechnen, kann man die im Euler-Verfahren benötigten Zufallszahlen ΔW^j auch direkt an der Stelle per Zufallsgenerator erzeugen, an der man sie benötigt. Der Algorithmus 7.8 soll in erster Linie den prinzipiellen Ablauf einer solchen Simulation verdeutlichen.

Die Konvergenzeigenschaften des stochastischen Euler-Verfahrens fasst der folgende Satz zusammen. Hierbei verwenden wir den Raum C_P^q . Dies ist die Menge aller Funktionen, die q -mal stetig differenzierbar ist und deren partielle Ableitungen bis zum Grad q in der Norm durch ein Polynom P beschränkt sind.

Satz 7.9 Betrachte eine stochastische Differentialgleichung (7.1) und das stochastische Euler-Verfahren Φ , das gemäß Algorithmus 7.8 auf einem Intervall $[0, T]$ angewendet wird.

(i) Angenommen, es gelten für geeignete Konstanten K_i und alle $t, s \in [0, T]$, $x, y \in \mathbb{R}^n$ die folgenden Ungleichungen

$$\|a(t, x) - a(t, y)\| + \sum_{j=1}^m \|b_j(t, x) - b_j(t, y)\| \leq K_1 \|x - y\| \quad (7.5)$$

$$\|a(t, x)\| + \sum_{j=1}^m \|b_j(t, x)\| \leq K_2(1 + \|x\|) \quad (7.6)$$

$$\|a(t, x) - a(s, x)\| + \sum_{j=1}^m \|b_j(t, x) - b_j(s, x)\| \leq K_3(1 + \|x\|)|t - s|^{1/2}. \quad (7.7)$$

Darüberhinaus seien die approximativen Pfade des Wiener-Prozesses in Algorithmus 7.8 die starken Approximationen des Wiener-Prozesses aus Algorithmus 7.5. Dann liefert ist das stochastische Euler-Verfahren eine starke Approximation der Ordnung $\gamma = 1/2$ bzgl. $g(x) = x$.

(ii) Angenommen, es gelten die Ungleichungen aus (i), zudem seien a und b_1, \dots, b_m unabhängig von t und C_P^4 in x . Darüberhinaus seien die approximativen Pfade des Wiener-Prozesses in Algorithmus 7.8 die schwachen Approximationen des Wiener-Prozesses aus

Algorithmus 7.6. Dann liefert ist das stochastische Euler-Verfahren eine schwache Approximation der Ordnung $\beta = 1$ bzgl. jeder C^4_P -Funktion g , wobei die Konstante C in Definition 7.1 von g abhängt.

Beweis: Folgt aus Satz 7.20, vgl. Abschnitt 7.7. \square

Beachte, dass ein wesentlicher Unterschied zwischen (i) und (ii) in der Tatsache liegt, dass in (ii) nur eine schwache Approximation des Wiener-Prozesses verwendet wird.

Ebenso wie bei deterministischen Differentialgleichungen gilt, dass das Euler-Schema relativ langsam konvergiert und dass man bessere Ergebnisse für Schemata höherer Ordnung erhält. Wir kommen später darauf zurück.

7.4 Anwendung auf die Optionsbewertung

Aus Satz 7.9(i), Lemma 7.3 und Lemma 7.4 sowie der Lipschitz-Stetigkeit der Funktionen $g(x) = (x - K)^+$ bzw. $g(x) = (K - x)^+$ folgt sofort, dass der für die Optionsbewertung benötigte Erwartungswert durch eine starke Approximation korrekt approximiert wird.

Da wir für die Optionsbewertung aber nur den Erwartungswert $E(g(\tilde{X}(T)))$ für $g(x) = (x - K)^+$ oder $g(x) = (K - x)^+$ benötigen, reicht es im Prinzip aus, hier eine schwache Approximation zu verwenden. Das Problem bei der Anwendung von Satz 7.9(ii) liegt aber darin, dass diese Funktionen g nicht C^4 sind. Da sie einen Knick an der Stelle $x = K$ besitzen, sind sie nämlich noch nicht einmal einfach differenzierbar.

Wir wollen nun beweisen, dass wir für diese nichtdifferenzierbaren g trotzdem Konvergenz erhalten, d.h. dass

$$\lim_{h \rightarrow 0} |E(g(X(T))) - E(g(\tilde{X}(T)))| = 0$$

gilt, wobei wir hier Schrittweiten der Form $h = T/N$ betrachten, damit $t_N = T$ gilt. Dazu führen wir C^4 -Hilfsfunktionen g_ε für $\varepsilon > 0$ ein, die sich nur wenig von g unterscheidet, und für die wir Konvergenz erhalten. Diese verwenden wir allerdings nicht im Algorithmus — dort benutzen wir die “richtigen” g — sondern nur in der Konvergenzanalyse. Wir machen dies beispielhaft für $g(x) = (x - K)^+$, also für einen Call.

Die approximierende Funktion g_ε erhalten wir, indem wir für $\varepsilon > 0$ eine C^4 -Funktion $h_\varepsilon : [K - \varepsilon, K + \varepsilon] \rightarrow \mathbb{R}$ wählen mit den Eigenschaften

$$h_\varepsilon^{(k)}(K - \varepsilon) = 0, k = 0, 1, 2, 3, 4,$$

$$h_\varepsilon(K + \varepsilon) = \varepsilon, h'_\varepsilon(K + \varepsilon) = 1, h_\varepsilon^{(k)}(K + \varepsilon) = 0, k = 2, 3, 4$$

wobei $h_\varepsilon^{(k)}$ die k -te Ableitung bezeichnet. Zudem verlangen wir $h_\varepsilon(x) \in [0, \varepsilon]$ für $x \in [K - \varepsilon, K + \varepsilon]$. Eine solche Funktion existiert und kann z.B. durch Hermite-Interpolation erzeugt werden. Damit definieren wir

$$\tilde{g}_\varepsilon(x) = \begin{cases} g(x), & x \notin [K - \varepsilon, K + \varepsilon] \\ h_\varepsilon(x), & x \in [K - \varepsilon, K + \varepsilon] \end{cases}$$

Abbildung 7.2 zeigt die Originalfunktion g (dicke grüne Linie) und die mit Hermite-Interpolation erzeugte Funktion g_ε (dünne schwarze Linie) für $\varepsilon = 0.2$.

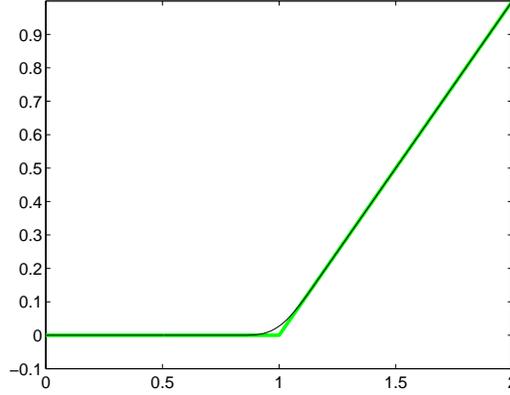


Abbildung 7.2: Die Funktionen $g(x) = (x - K)^+$ (dicke grüne Linie) und $g_\varepsilon(x)$ (dünne schwarze Linie) für $K = 1$ und $\varepsilon = 0.2$

Nun schätzen wir den Fehler mit der Dreiecksungleichung ab

$$\begin{aligned} |E(g(X(T))) - E(g(\tilde{X}(T)))| &\leq |E(g(X(T)) - E(g_\varepsilon(X(T))))| \\ &\quad + |E(g_\varepsilon(X(T))) - E(g_\varepsilon(\tilde{X}(T)))| \\ &\quad + |E(g_\varepsilon(\tilde{X}(T))) - E(g(\tilde{X}(T)))| \end{aligned}$$

Nun gilt einerseits nach Konstruktion $|g_\varepsilon(x) - g(x)| \leq \varepsilon$ für alle x , andererseits unterscheiden sich die beiden Funktionen nur auf einem Intervall der Länge 2ε . Da die Wahrscheinlichkeit, dass $X(T)$ bzw. $\tilde{X}(T)$ in diesem Intervall liegen proportional zu ε ist, folgt

$$|E(g(X(T)) - E(g_\varepsilon(X(T))))| \leq K\varepsilon^2 \quad \text{und} \quad |E(g(\tilde{X}(T))) - E(g_\varepsilon(\tilde{X}(T)))| \leq K\varepsilon^2$$

für eine geeignete Konstante $K > 0$. Den mittleren Term können wir mit Satz 7.9(ii) abschätzen, wobei die Konstante in der Konvergenz aber von ε abhängt, d.h. es gilt

$$|E(g_\varepsilon(X(T))) - E(g_\varepsilon(\tilde{X}(T)))| \leq C_\varepsilon h.$$

Insgesamt erhalten wir also

$$|E(g(X(T))) - E(g(\tilde{X}(T)))| \leq 2K\varepsilon^2 + C_\varepsilon h.$$

Die Konstante C_ε kann dabei für $\varepsilon \rightarrow 0$ möglicherweise gegen unendlich streben. Wir können nun aber zu gegebenem hinreichend kleinem $h > 0$ ein $\varepsilon(h) > 0$ so wählen, dass $C_\varepsilon \leq h^{-1/2}$ gilt. Wählen wir dieses $\varepsilon(h)$ minimal mit dieser Eigenschaft, so folgt $\varepsilon(h) \rightarrow 0$ für $h \rightarrow 0$. Damit erhalten wir

$$|E(g(X(T))) - E(g(\tilde{X}(T)))| \leq 2K\varepsilon(h)^2 + h^{1/2} \rightarrow 0$$

für $h \rightarrow 0$ und damit die gewünschte Konvergenz.

Beachte, dass wir mit diesem Beweis keine Aussage über die Konvergenzordnung treffen können. Dazu müssten wir wissen, wie schnell C_ε für $\varepsilon \rightarrow 0$ wächst, was im Prinzip berechenbar ist, aber nur mit sehr hohem Aufwand, weswegen wir uns hier mit der Konvergenzaussage ohne Abschätzung der Ordnung zufrieden geben.

7.5 Die Itô-Taylor-Entwicklung

Bei den deterministischen Einschrittverfahren haben wir in Satz 6.19 gesehen, dass man die Konsistenz und damit letztendlich die Konvergenz analysieren kann, indem man die Ableitungen der Verfahren mit den Koeffizienten der Taylor-Entwicklung aus Satz 6.13 vergleicht. Ebenso kann diese Entwicklung verwendet werden, um die Taylor-Verfahren als spezielle Einschrittverfahren zu definieren. Um dies für stochastische Differentialgleichungen auf ähnliche Weise machen zu können, wollen wir nun eine analoge Taylor-Entwicklung der Lösungen für Itô-SDGs herleiten. Wir betrachten dazu wieder die vektorwertige Itô-SDG (7.1) und die zugehörige Integralgleichung (7.2). Um zu zeigen, dass die hier vorgestellte Theorie tatsächlich eine Erweiterung der deterministischen Theorie ist, verwenden wir, dass sich (7.1) im Spezialfall $m = 0$ (mit der Konvention $\sum_{j=1}^0 \cdots = 0$) und $a(t, x) = f(t, x)$ gerade auf die deterministische gewöhnliche DGL (6.1) reduziert.

Ziel dieses Abschnitts ist es, für eine gegebene Funktion $g : \mathbb{R} \times \mathbb{R}^n \rightarrow \mathbb{R}^q$ eine Taylor-Entwicklung für $g(t, X(t))$ in $t = t_0$ anzugeben. Dazu benötigen wir einige Definitionen.

Definition 7.10 Ein Multi-Index ist ein Zeilenvektor $\alpha = (j_1, j_2, \dots, j_l)$, mit Einträgen $j_i \in \{0, 1, \dots, m\}$ für $i = 1, \dots, l$. Mit $l(\alpha) = l$ bezeichnen wir die Länge des Multi-Indexes und mit \odot bezeichnen wir den Multi-Index der Länge Null, d.h. $l(\odot) = 0$.

Die Menge aller Multi-Indexes bezeichnen wir mit

$$\mathcal{M}_m = \left\{ (j_1, j_2, \dots, j_l) : j_i \in \{0, \dots, m\}, i \in \{1, \dots, l\}, l \in \{1, 2, \dots\} \right\} \cup \{\odot\}.$$

Für jeden Multi-Index $\alpha = (j_1, j_2, \dots, j_l) \in \mathcal{M}_m$ mit $l(\alpha) \geq 1$ bezeichnen wir mit $-\alpha$ und $\alpha-$ den Multi-Index aus \mathcal{M}_m , den wir durch Entfernen der ersten bzw. letzten Komponente von α erhalten, also

$$-\alpha = (j_2, \dots, j_l) \quad \alpha- = (j_1, \dots, j_{l-1}).$$

Mit $n(\alpha)$ bezeichnen wir die Anzahl der Einträge eines Multi-Indexes $\alpha \in \mathcal{M}_m$, die gleich 0 sind. \square

Mit Hilfe dieser Multi-Indexes können wir die folgenden Integrale definieren.

Definition 7.11 Für einen Multi-Index $\alpha = (j_1, j_2, \dots, j_l) \in \mathcal{M}_m$, einen Itô-integrierbaren vektorwertigen stochastischen Prozess $g(t)$ auf $[t_0, T]$ und eine Zeit $t \in [t_0, T]$ definieren wir das mehrfache Itô-Integral $I_\alpha[g(\cdot)]_{t_0, t}$ rekursiv als

$$I_\alpha[g(\cdot)]_{t_0, t} := \begin{cases} g(t), & \text{falls } l = 0 \\ \int_{t_0}^t I_{\alpha-}[g(\cdot)]_{t_0, s} ds, & \text{falls } l \geq 1 \text{ und } j_l = 0 \\ \int_{t_0}^t I_{\alpha-}[g(\cdot)]_{t_0, s} dW_s^{j_l}, & \text{falls } l \geq 1 \text{ und } j_l \in \{1, \dots, m\}. \end{cases}$$

\square

Beispiel 7.12 Einige Spezialfälle dieser rekursiven Definition sind

$$\begin{aligned} I_{\odot}[g(\cdot)]_{t_0,t} &= g(t), & I_{(0)}[g(\cdot)]_{t_0,t} &= \int_{t_0}^t g(s) ds, \\ I_{(1)}[g(\cdot)]_{t_0,t} &= \int_{t_0}^t g(s) dW_s^1, \\ I_{(0,1)}[g(\cdot)]_{t_0,0,t} &= \int_{t_0}^t \int_{t_0}^{s_2} g(s_1) ds_1 dW_{s_2}^1, \\ I_{(0,2,1)}[g(\cdot)]_{t_0,t} &= \int_{t_0}^t \int_{t_0}^{s_3} \int_{t_0}^{s_2} g(s_1) ds_1 dW_{s_2}^2 dW_{s_3}^1. \end{aligned}$$

□

Im Folgenden werden einige Eigenschaften dieser Integrale gebraucht. Für konstante Funktionen $g(t) \equiv g_0$ gilt

$$I_{\alpha}[g(\cdot)]_{t_0,t} = I_{\alpha}[g_0]_{t_0,t} = g_0 I_{\alpha}[1]_{t_0,t}.$$

Hierbei kann g_0 auch ein Vektor sein, die Multiplikation von g_0 mit der skalaren Größe $I_{\alpha}[1]_{t_0,t}$ ist dann komponentenweise zu verstehen. Zur Vereinfachung der Notation schreiben wir oft kurz $I_{\alpha,t_0,t}$ oder I_{α} statt $I_{\alpha}[1]_{t_0,t}$.

Für $\alpha = (0, \dots, 0)$ mit $l(\alpha) = i$ ergibt sich

$$I_{\alpha}[g(\cdot)]_{t_0,t} = \int_{t_0}^t \int_{t_0}^{s_i} \dots \int_{t_0}^{s_2} g(s_1) ds_1 ds_2 \dots ds_i \quad (7.8)$$

woraus für $g(t) \equiv g_0$ folgt

$$I_{\alpha}[g_0]_{t_0,t} = g_0 I_{\alpha,t_0,t} = g_0 \frac{(t-t_0)^i}{i!}. \quad (7.9)$$

Für deterministische und auf $[t_0, t]$ beschränkte Funktionen $g : \mathbb{R} \rightarrow \mathbb{R}^n$ folgt daraus die Ungleichung

$$\|I_{\alpha}[g(\cdot)]_{t_0,t}\| \leq \sup_{t \in [t_0,t]} \|g(t)\| I_{\alpha,t_0,t} \leq \sup_{t \in [t_0,t]} \|g(t)\| \frac{(t-t_0)^i}{i!}. \quad (7.10)$$

Definition 7.13 Für einen Multi-Index $\alpha = (j_1, \dots, j_l) \in \mathcal{M}_m$ und eine hinreichend oft differenzierbare Funktion $g : [t_0, T] \times \mathbb{R}^n \rightarrow \mathbb{R}^q$ definieren wir die Mehrfachableitung von g_{α} bzgl. α rekursiv als

$$g_{\alpha} = \begin{cases} g, & \text{falls } l = 0 \\ L_{j_1} g_{-\alpha}, & \text{falls } l \geq 1. \end{cases} \quad (7.11)$$

Hierbei sind die Differentialoperatoren L_j definiert mittels

$$\begin{aligned} L_0 g(t, x) &= \frac{\partial}{\partial t} g(t, x) + \frac{\partial}{\partial x} g(t, x) a(t, x) + \frac{1}{2} \sum_{k,l=1}^n \sum_{j=1}^m b_{j,k}(t, x) b_{j,l}(t, x) \frac{\partial^2}{\partial x^k \partial x^l} g(t, x) \\ L_j g(t, x) &= \frac{\partial}{\partial x} g(t, x) b_j(t, x), \quad j = 1, \dots, m. \end{aligned}$$

□

Beachte, dass auch a und b_1, \dots, b_m in (7.1) hinreichend oft differenzierbar sein müssen, damit diese Mehrfachableitung wohldefiniert ist.

Diese Definition verallgemeinert die Definition von L_f^i aus Abschnitt 6.3: Für $m = 0$, $a(t, x) = f(t, x)$ und einen Multi-Index $\alpha = (0, \dots, 0)$ mit $l(\alpha) = i$ erhalten wir gerade

$$g_\alpha(t, x) = L_f^i g(t, x). \quad (7.12)$$

Für den Spezialfall $g(t, x) = x$ erhalten wir wegen

$$L_f^1 g(t, x) = \frac{\partial}{\partial t} g(t, x) + \frac{\partial}{\partial x} g(t, x) f(t, x) = f(t, x)$$

für $l(\alpha) = i \geq 1$

$$g_\alpha(t, x) = L_f^i g(t, x) = L_f^{i-1} L_f^1 g(t, x) = L_f^{i-1} f(t, x). \quad (7.13)$$

Beispiel 7.14 Im eindimensionalen Fall $n = 1$ ergibt sich für die Identität $g(t, x) \equiv x$ und Index-Einträge $j_1, j_2 \geq 1$

$$g_{(0)} = a, \quad g_{(j_1)} = b_{j_1}, \quad g_{(0,0)} = \dot{a} + a'a + \frac{1}{2} \sum_{j=1}^m (b_j)^2 a'',$$

$$g_{(0,j_1)} = \dot{b}_{j_1} + b'_{j_1} a + \frac{1}{2} \sum_{j=1}^m (b_j)^2 b''_{j_1}, \quad g_{(j_1,0)} = a' b_{j_1}, \quad g_{(j_1,j_2)} = b'_{j_2} b_{j_1},$$

wobei der Strich ' die Ableitung bzgl. x und der Punkt die Ableitung bzgl. t bezeichnet und die Argumente (t, x) bei allen Funktionen weggelassen wurden. □

Um aus diesen Ableitungen eine Taylor-Entwicklung zu konstruieren, brauchen wir noch eine letzte Definition.

Definition 7.15 Eine endliche Teilmenge $\mathcal{H} \subset \mathcal{M}_m$ heißt *hierarchische Menge*, wenn $\mathcal{H} \neq \emptyset$ ist und

$$-\alpha \in \mathcal{H} \quad \text{für jedes } \alpha \in \mathcal{H} \setminus \{\odot\}$$

gilt, wobei \odot wie üblich den Multi-Index der Länge 0 bezeichnet.

Für eine gegebene hierarchische Menge \mathcal{H} ist die zugehörige *Restmenge* $\mathcal{R}(\mathcal{H})$ definiert als

$$\mathcal{R}(\mathcal{H}) = \{\alpha \in \mathcal{M}_m \setminus \mathcal{H} \mid -\alpha \in \mathcal{H}\}.$$

□

Mit anderen Worten ist eine Menge gerade dann eine hierarchische Menge, wenn für jeden Multi-Index $\alpha \in \mathcal{H}$ auch der Multi-Index $-\alpha$ mit entfernter erster Komponente in \mathcal{H} liegt.

Die Restmenge besteht aus gerade den Multi-Indizes, die durch Anfügen eines neuen ersten Eintrags an ein Element aus \mathcal{H} entstehen und nicht bereits in \mathcal{H} liegen.

Mit dieser Notation können wir nun eine Itô-Taylor Entwicklung formulieren.

Satz 7.16 Sei $g : \mathbb{R}_0^+ \times \mathbb{R}^n \rightarrow \mathbb{R}^q$ and sei $\mathcal{H} \subset \mathcal{M}_m$ eine hierarchische Menge mit Restmenge $\mathcal{R}(\mathcal{H})$ und sei $X(t)$ eine Lösung von (7.1) auf $[t_0, T]$. Dann gilt die Itô-Taylor-Entwicklung bzgl. \mathcal{H}

$$g(t, X(t)) = \sum_{\alpha \in \mathcal{H}} g_\alpha(t_0, X(t_0)) I_{\alpha, t_0, t} + \sum_{\alpha \in \mathcal{R}(\mathcal{H})} I_\alpha [g_\alpha(\cdot, X(\cdot))]_{t_0, t} \quad (7.14)$$

für alle $t \in [t_0, T]$, vorausgesetzt dass alle auftretenden Ableitungen von g , a , b_1, \dots, b_m und alle auftretenden Mehrfach-Itô-Integrale existieren.

Beweis: Der Beweis folgt durch iterative Anwendung des (mehrdimensionalen) Itô-Lemmas, für Details siehe Kloeden/Platen [3, Theorem 5.5.1]. \square

Beachte den Unterschied zwischen den beiden Summen in (7.14): In der ersten Summe sind die Integranden konstant, da g_α zur festen Zeit t_0 ausgewertet wird. In der zweiten Summe hingegen sind die Integranden Funktionen.

Wir illustrieren diese Definition an Hand von drei Beispielen. Das erste Beispiel zeigt dabei insbesondere, dass (7.14) tatsächlich eine Erweiterung der deterministischen Taylor-Entwicklung aus Satz 6.13 ist.

Beispiel 7.17 (i) Im Fall $m = 0$ und $a(t, x) = f(t, x)$ wird (7.1) zur gewöhnlichen Differentialgleichung $\dot{x}(t) = f(t, x(t))$ mit Lösung $X(t) = x(t, t_0, x_0)$. Wegen $m = 0$ besteht \mathcal{M}_m nur aus Multi-Indizes der Form $\alpha = \odot$ und $\alpha = (0, \dots, 0)$. Folglich ist jede hierarchische Menge von der Form

$$\mathcal{H} = \{\alpha = (0, \dots, 0) \mid l(\alpha) \leq l^*\} \cup \{\odot\}$$

für ein $l^* \geq 0$ und die zugehörige Restmenge ist

$$\mathcal{R}(\mathcal{H}) = \{\alpha = (0, \dots, 0) \mid l(\alpha) = l^* + 1\}.$$

Für $g(t, x) = x$ ergibt sich (7.14) wegen (7.13) und (7.8)–(7.10) damit zu

$$\begin{aligned} X(t) &= g(t, X(t)) \\ &= \sum_{\alpha \in \mathcal{H}} g_\alpha(t_0, X(t_0)) I_{\alpha, t_0, t} + \sum_{\alpha \in \mathcal{R}(\mathcal{H})} I_\alpha [g_\alpha(\cdot, X(\cdot))]_{t_0, t} \\ &= \sum_{i=0}^{l^*} L_f^i g(t_0, X(t_0)) \underbrace{I_{(0, \dots, 0)}_{t_0, t}}_{i \text{ Einträge}} + I_{\underbrace{(0, \dots, 0)}_{l^*+1 \text{ Einträge}}} \left[L_f^{l^*+1} g(\cdot, X(\cdot)) \right]_{t_0, t} \\ &= X(t_0) + \sum_{i=1}^{l^*} L_f^{i-1} f(t_0, X(t_0)) \frac{(t-t_0)^i}{i!} + \underbrace{I_{\underbrace{(0, \dots, 0)}_{l^*+1 \text{ Einträge}}} \left[L_f^{l^*+1} g(\cdot, X(\cdot)) \right]_{t_0, t}}_{\|\cdot\| \leq C(t-t_0)^{l^*+1}} \end{aligned}$$

vorausgesetzt, dass $L_f^{l^*+1} g(\tau, X(\tau))$ beschränkt ist für $\tau \in [t_0, t]$. Für $X(t) = x(t, t_0, x_0)$ ist dies gerade die Aussage von Satz 6.13. Dies liefert die folgende Interpretation der beiden Summen in (7.14): die erste Summe stellt die Taylor-Approximation dar während die zweite Summe das Restglied ist.

(ii) Für $\mathcal{H} = \{\odot\}$ und $\mathcal{R}(\{\odot\}) = \{(0), \dots, (m)\}$ erhalten wir

$$\begin{aligned} g(t, X(t)) &= g_{\odot}(t_0, X(t_0)) I_{\odot, t_0, t} + \sum_{\alpha \in \mathcal{R}(\{\odot\})} I_{\alpha} [g_{\alpha}(\cdot, X(\cdot))]_{t_0, t} \\ &= g(t_0, X(t_0)) + \int_{t_0}^t L_0 g(s, X(s)) ds + \sum_{j=1}^m \int_{t_0}^t L_j g(s, X(s)) dW_s^j. \end{aligned}$$

Schreiben wir diese Integralgleichung in der üblichen Itô-SDG Kurzschreibweise, so erhalten wir

$$d(g(t, X(t))) = L_0 g(t, X(t)) dt + \sum_{j=1}^m L_j g(t, X(t)) dW_t^j.$$

Für $g(x) = x$ ist dies gerade die Gleichung (7.1) selbst. Im skalaren Fall $m = n = 1$ und für allgemeine $g : \mathbb{R} \times \mathbb{R} \rightarrow \mathbb{R}$ ist dies gerade das Itô-Lemma 4.2.

(iii) Im skalaren Fall $n = m = 1$ mit $g(t, x) \equiv x$, $a(t, x) = a(x)$, $b_1(t, x) = b(x)$, $t_0 = 0$,

$$\mathcal{H} = \{\alpha \in \mathcal{M}_1 : l(\alpha) \leq 2\} \quad \text{und} \quad \mathcal{R}(\mathcal{H}) = \{\alpha \in \mathcal{M}_1 : l(\alpha) = 3\}$$

ergibt sich

$$\begin{aligned} X(t) &= X(t_0) + a(x)I_{(0)} + b(x)I_{(1)} + a(x)a'(x)I_{(0,0)} \\ &\quad + a(x)b'(x)I_{(0,1)} + b(x)a'(x)I_{(1,0)} \\ &\quad + b(x)b'(x)I_{(1,1)} + R_3(t, t_0), \end{aligned}$$

wobei die Integrale jeweils auf dem Intervall $[t_0, t]$ zu verstehen sind, der Strich ' die Ableitung bzgl. x bezeichnet und $R_3(t, t_0)$ eine Kurzschreibweise für die zweite Summe in 7.14 ist. \square

7.6 Itô-Taylor-Verfahren

Satz 7.16 liefert nun die Basis für die folgenden Itô-Taylor-Schemata. Wie im deterministischen Fall erhalten wir diese Schemata, indem wir die Restterme einfach weglassen lassen. Wie man die hierarchischen Mengen in der folgenden Definition sinnvollerweise wählt und wie man die auftretenden Integralterme und die darin enthaltenen Pfade der Wiener-Prozesse numerisch auswertet, betrachten wir anschließend.

Definition 7.18 Für eine gegebene hierarchische Menge \mathcal{H} und $x = (x_1, \dots, x_n)^T \in \mathbb{R}^n$ definieren wir das *Itô-Taylor-Verfahren* als

$$\Phi(t, X, h, W, \omega) = X(\omega) + \sum_{\alpha \in \mathcal{H} \setminus \{\odot\}} g_{\alpha}(t, X(\omega)) I_{\alpha, t, t+h}(\omega) \quad (7.15)$$

mit $g(t, x) = x$, wobei in $W = (W^1, \dots, W^m)$ die m Wiener-Prozesse zusammengefasst sind, welche in die Integralterme $I_{\alpha, t, t+h}$ eingehen. \square

Um solch ein Schema gemäß Algorithmus 7.8 implementieren zu können, müssen wir

- die hierarchische Menge \mathcal{H} angeben
- die auftretenden Ableitungen g_α berechnen
- eine numerische Formel $\tilde{I}_{\alpha,t,t+h}$ für die auftretenden Mehrfachintegrale $I_{\alpha,t,t+h}$ herleiten. Diese basiert dabei auf geeigneten Approximationen \tilde{W}^j der eingehenden Wiener-Prozesse.

Bevor wir einen allgemeinen Satz formulieren, der uns sagt, wie \mathcal{H} für eine gewünschte Ordnung gewählt werden muss und welche Bedingung $\tilde{I}_{\alpha,t,t+h}$ erfüllen muss, schauen wir uns einen einfachen Spezialfall an.

Beispiel 7.19 Wir betrachten die Menge $\mathcal{H} = \{\odot, (0), \dots, (m)\}$. Wegen $g_{(0)}(t, x) = a(t, x)$ und $g_{(k)}(t, x) = b_k(t, x)$ für $k = 1, \dots, m$ erhalten wir mit der Definition der Mehrfachintegrale

$$\begin{aligned} \Phi(t, X, h, W, \omega) &= X(\omega) + a(t, X(\omega))I_{(0),t,t+h}(\omega) + \sum_{j=1}^m b_j(t, X(\omega))I_{(j),t,t+h}(\omega) \\ &= X(\omega) + a(t, X(\omega))h + \sum_{j=1}^m b_j(t, X(\omega))\Delta W^j(t, \omega) \end{aligned}$$

mit $\Delta W^j(t, \omega) = W^j(t+h, \omega) - W^j(t, \omega)$. Dies ist gerade das stochastische Euler-Verfahren bei dem wir die auftretenden Integrale approximieren können durch $\tilde{I}_{(0),t,t+h} = h$ und $\tilde{I}_{(j),t,t+h} = \Delta \tilde{W}^j(t)$ mit den numerischen Approximationen \tilde{W}^j der Wiener-Prozesse aus Abschnitt 7.2. \square

Für größere hierarchische Mengen ist die Berechnung der auftretenden Mehrfachintegrale nicht mehr so leicht. Dies ist ein Hauptgrund dafür, dass Verfahren höherer Ordnung für stochastische Differentialgleichungen schwer zu implementieren sind. Bevor wir dieses Problem etwas näher betrachten, geben wir zunächst den bereits angekündigten allgemeinen Konvergenzsatz für Itô-Taylor-Verfahren.

Satz 7.20 Betrachte eine stochastische Differentialgleichung (7.1) und ein stochastisches Einschrittverfahren Φ , das gemäß Algorithmus 7.8 angewendet wird.

(i) Es sei Φ das Itô-Taylor-Verfahren zur hierarchischen Menge

$$\mathcal{A}_\gamma := \left\{ \alpha \in \mathcal{M}_m \mid l(\alpha) + n(\alpha) \leq 2\gamma \text{ oder } l(\alpha) = n(\alpha) = \gamma + \frac{1}{2} \right\}$$

für $\gamma = 0.5, 1, 1.5, 2, \dots$. Es seien die folgenden Bedingungen für die in Φ auftretenden Funktionen g_α , Konstanten $K_1, K_2, K_3 > 0$ und alle $t \in [0, T]$ und $x, y \in \mathbb{R}^n$ erfüllt.

- $\|g_\alpha(t, x) - g_\alpha(t, y)\| \leq K_1 \|x - y\|$ für alle $\alpha \in \mathcal{A}_\gamma$

- alle Mehrfachintegrale $I_\alpha[g_\alpha(t, x)]_{0, T}$ existieren und es gelte $g_{-\alpha} \in C^2$ für alle $\alpha \in \mathcal{A}_\gamma \cup \mathcal{R}(\mathcal{A}_\gamma)$
- $\|g_\alpha(t, x)\| \leq K_2(1 + \|x\|)$ für alle $\alpha \in \mathcal{A}_\gamma \cup \mathcal{R}(\mathcal{A}_\gamma)$
- die Approximationen der Mehrfachintegrale erfüllen

$$E \left(\left| \tilde{I}_{\alpha, t, t+h} - I_{\alpha, t, t+h} \right|^2 \right) \leq K_3 h^{2\gamma+1} \quad (7.16)$$

für alle $\alpha \in \mathcal{A}_\gamma \setminus \{\odot\}$ und alle $t \in \mathcal{T}$.

Dann liefert das Itô-Taylor-Verfahren eine starke Approximation der Ordnung γ bzgl. $g(x) = x$.

(ii) Es sei Φ das Itô-Taylor-Verfahren zur hierarchischen Menge

$$\Gamma_\beta := \{\alpha \in \mathcal{M}_m \mid l(\alpha) \leq \beta\}$$

für $\beta = 1, 2, 3, \dots$ Es seien die folgenden Bedingungen für die in Φ auftretenden Funktionen g_α , Konstanten $K_1, K_2 > 0$ und alle $x \in \mathbb{R}^n$ erfüllt.

- a und b_1, \dots, b_m seien unabhängig von t , global Lipschitz und aus $C_P^{2(\beta+1)}$
- $\|g_\alpha(x)\| \leq K_1(1 + \|x\|)$ für alle $\alpha \in \Gamma_\beta$
- die Approximationen der Mehrfachintegrale erfüllen

$$\left| E \left(\left(\sum_{\alpha \in \Gamma_\beta \setminus \{\odot\}} \tilde{I}_{\alpha, t, t+h} \right)^l - \left(\sum_{\alpha \in \Gamma_\beta \setminus \{\odot\}} I_{\alpha, t, t+h} \right)^l \right) \right| \leq K_2 h^{\beta+1} \quad (7.17)$$

für alle $l = 1, \dots, 2\beta + 1$ und alle $t \in \mathcal{T}$.

Dann liefert das Itô-Taylor-Verfahren eine schwache Approximation der Ordnung β bzgl. jeder $C_P^{2(\beta+1)}$ -Funktion g , wobei die Konstante C in Definition 7.1 von g abhängt.

Beweisskizze: (i) Im starken Fall läuft der Beweis ähnlich wie im deterministischen Fall, vgl. den Beweis von Satz 6.11. An Stelle der Fehlerfunktion $e(t)$ tritt dabei die Funktion

$$Z(t) = E \left(\sup_{s \in [0, t]} \|\tilde{X}(s) - X(s)\|^2 \right),$$

zu deren Definition die zunächst nur für t aus dem Gitter \mathcal{T} definierte Funktion \tilde{X} in geeigneter Weise auf dem ganzen Intervall $[0, T]$ fortgesetzt wird. Für Z erhält man dann durch Abschätzen der Restterme der Taylor-Entwicklung, Ausnutzung von (7.16) und Verwendung des Gronwall-Lemmas (das an Stelle der Induktion im Beweis von Satz 6.11 tritt) die Ungleichung

$$Z(t) \leq \underbrace{K(1 + |X(0)|^2)}_{=: C^2} h^{2\gamma}.$$

Aus der sogenannten Jensenschen Ungleichung $E(Y)^2 \leq E(Y^2)$ (hier für den Spezialfall $x \mapsto x^2$), die für beliebige reelle Zufallsvariablen gilt, folgt dann

$$E\left(\sup_{s \in [0, t]} \|\tilde{X}(s) - X(s)\|\right) \leq \sqrt{Z(t)} \leq Ch^\gamma,$$

woraus die starke Konvergenz mit $t = T$ folgt. Für Details siehe [3, Theorem 10.6.3 und Corollary 10.6.5].

(ii) Der Beweis im schwachen Fall wird indirekter geführt. Hier können wir uns auf reellwertige Funktionen g beschränken; die Abschätzung im vektorwertigen Fall folgt dann durch komponentenweise Anwendung des Erwartungswertes. Um die gewünschte Ungleichung

$$|E(g(\tilde{X}(T)) - g(X(T)))| \leq Ch^\beta$$

herzuleiten, wird ausgenutzt, dass $E(g(X(T)))$ sich als Lösung $u(0, X(0))$ der partiellen Differentialgleichung $L_0 u(t, x) = 0$ mit Endbedingung $u(T, x) = g(x)$ schreiben lässt (hier wird die Differenzierbarkeit von g benötigt). Dann wird nachgewiesen, dass die Funktion $g(\tilde{X}(t))$ für die numerische Approximation \tilde{X} diese partielle Differentialgleichung approximativ löst, wobei der Fehler, der durch Analyse der Restterme der Taylor-Entwicklung und (7.17) abgeschätzt wird, gerade von der gesuchten Ordnung ist. Details finden sich in [3, Theorem 14.5.1 und 14.5.2]. \square

7.7 Spezielle Verfahren

Satz 7.20 gibt im Prinzip alle Bedingungen an, die bei der Konstruktion der numerischen Verfahren beachtet werden müssen. Die Bedingungen an die numerischen Mehrfachintegrale $\tilde{I}_{\alpha, t, t+h}$ sind allerdings implizit und geben zunächst keinen Hinweis darauf, wie man diese Ausdrücke denn nun in einer praktischen Implementierung berechnet. Tatsächlich ist dies der wesentliche Hinderungsgrund für die Verwendung von Verfahren höherer Ordnung. Für einige Spezialfälle lassen sich aber geeignete Approximationen angeben. Einige Beispiele werden wir im Folgenden betrachten. Danach werden wir noch kurz Verfahren ansprechen, in denen keine Ableitungen der Form g_α berechnet werden müssen. Für Details zu den angegebenen Aussagen siehe [3, Abschnitte 10.2, 10.3, 14.1 und 14.2].

Das stochastische Euler-Verfahren

In Beispiel 7.19 haben wir bereits gesehen, dass wir für die hierarchische Menge $\mathcal{H} = \{\odot, (0), \dots, (m)\}$ gerade das stochastische Euler-Verfahren erhalten. Diese Menge ist gerade gleich $\mathcal{A}_{0.5}$ im Fall (i) und gleich Γ_1 im Fall (ii) von Satz 7.20. Um den Satz anwenden zu können (und damit den Beweis von Satz 7.9 zu erhalten) müssen wir noch nachweisen, dass die im Satz 7.9 implizit verwendeten Approximationen der Mehrfachintegrale die Bedingungen von Satz 7.20 erfüllen. Dies sind gemäß Beispiel 7.19 gerade

$$\tilde{I}_{(0), t, t+h} = h \quad \text{und} \quad \tilde{I}_{(j), t, t+h} = \Delta \tilde{W}^j(t).$$

Die zugehörigen exakten Mehrfachintegrale lauten

$$I_{(0),t,t+h} = h \quad \text{und} \quad I_{(j),t,t+h} = \Delta W^j(t).$$

Im starken Fall (i) ist die Bedingung (7.16) für $\alpha = (0)$ damit unabhängig von der Konstruktion der \widetilde{W}^j erfüllt. Für $\alpha = (j)$ mit $j \geq 1$ ergibt sich (7.16) zu

$$E \left(\left| \Delta \widetilde{W}^j(t) - \Delta W^j(t) \right|^2 \right) \leq K_3 h^2.$$

Dies ist nach den Überlegungen nach Algorithmus 7.5 für die starken Approximationen des Wiener-Prozesses aus diesem Algorithmus aber gerade erfüllt, sogar mit $K_3 = 0$.

Im schwachen Fall (ii) nutzt man aus, dass alle in (7.17) auftretenden Mehrfachintegrale stochastisch unabhängig sind. Daher kann der Erwartungswert in die gemischten Terme in den potenzierten Summen hineingezogen werden und es reicht aus, die Ungleichungen

$$\left| E \left((\widetilde{I}_{\alpha,t,t+h}^l - (I_{\alpha,t,t+h}^l)^l) \right) \right| \leq K_2 h^2$$

für alle $\alpha \in \Gamma_\beta \setminus \{\odot\}$ und $l = 1, 2, 3$ zu zeigen. Für $\alpha = (0)$ folgt dies wie oben, für $\alpha = (j)$ mit $j \geq 1$ muss

$$\left| E \left(\widetilde{W}^j(t)^l - W^j(t)^l \right) \right| \leq K_2 h^2$$

für $l = 1, 2, 3$ und die schwache Approximation des Wiener-Prozesses 7.6 gelten. Aus der Konstruktion dieser Approximation folgt, dass die Inkremente \widetilde{W}^j gerade die zweipunktverteilten Zufallsvariablen ΔW_i sind. Daher gilt

$$E(\Delta \widetilde{W}^j(t)^1) = 0, \quad E(\Delta \widetilde{W}^j(t)^2) = h \quad \text{und} \quad E(\Delta \widetilde{W}^j(t)^3) = 0.$$

Dies sind gerade die gleichen Werte wie bei den Inkrementen des exakten Wiener-Prozesses $\Delta W^j(t)$, da diese $N(0, h)$ -verteilt sind und für $N(0, \sigma^2)$ -verteilte Zufallsvariablen X gilt $E(X) = 0$, $E(X^2) = \sigma^2$ und $E(X^3) = 0$. Die Bedingung (7.17) ist daher für alle $l = 1, 2, 3$ mit $K_2 = 0$ erfüllt.

Das Milstein-Verfahren

Das Milstein-Verfahren ist das Itô-Taylor Verfahren mit starker Konvergenzordnung $\gamma = 1$. Die zugehörige hierarchische Menge ist gegeben durch

$$\mathcal{A}_1 = \{\odot, (0), \dots, (m), (1, 1), \dots, (1, m), (2, 1), \dots, (m, m)\}.$$

Im Vergleich zum Euler-Verfahren kommen hier m^2 weitere Terme dazu. Im Fall $m = 1$ lautet das Verfahren wegen

$$L_{j_1} b_{j_2}(t, x) = \frac{\partial}{\partial x} b_{j_2}(t, x) b_{j_1}(t, x)$$

explizit ausgeschrieben

$$\begin{aligned} \Phi(t, X, h, W, \omega) &= X(\omega) + a(t, X(\omega)) \widetilde{I}_{(0),t,t+h} + b_1(t, X(\omega)) \widetilde{I}_{(1),t,t+h} \\ &\quad + \frac{\partial}{\partial x} b_1(t, X(\omega)) b_1(t, X(\omega)) \widetilde{I}_{(1,1),t,t+h}. \end{aligned}$$

Hier kann man beweisen, dass (7.16) gilt, wenn die Mehrfachintegrale als

$$\tilde{I}_{(0),t,t+h} = h, \quad \tilde{I}_{(1),t,t+h} = \Delta \tilde{W}^1(t) \quad \text{und} \quad \tilde{I}_{(1,1),t,t+h} = \frac{1}{2}(\Delta \tilde{W}^1(t))^2 - h$$

gewählt werden mit der starken Approximation \tilde{W}^1 aus Algorithmus 7.5.

Für $m \geq 2$ lautet das Milstein-Verfahren

$$\begin{aligned} \Phi(t, X, h, W, \omega) &= X(\omega) + a(t, X(\omega))\tilde{I}_{(0),t,t+h} + \sum_{j=1}^m b_j(t, X(\omega))\tilde{I}_{(j),t,t+h} \\ &\quad + \sum_{j_1, j_2=1}^m \frac{\partial}{\partial x} b_{j_2}(t, X(\omega))b_{j_1}(t, X(\omega))\tilde{I}_{(j_1, j_2),t,t+h}. \end{aligned}$$

Hier gilt (7.16) für die Mehrfachintegrale für $\alpha = (j_1, j_2)$ mit $j_1 = j_2 = j \geq 1$ analog zum Fall $j = 1$ mit

$$\tilde{I}_{(j,j),t,t+h} = \frac{1}{2}(\Delta \tilde{W}^j(t))^2 - h.$$

Ein Problem stellt allerdings die Berechnung der Mehrfachintegrale $\tilde{I}_{(j_1, j_2),t,t+h}$ für $j_1 \neq j_2$ dar. Für diese gibt es keine Möglichkeit, sie aus den starken Approximationen \tilde{W}^{j_1} , \tilde{W}^{j_2} aus Algorithmus 7.5 zu berechnen. Es gibt zwar Approximationsmethoden für diese Integrale, vgl. [3, Abschnitt 10.3], diese sind aber sehr aufwändig und verlangen die Erzeugung vieler weiterer Zufallszahlen.

Unter gewissen strukturellen Annahmen an (7.1) fallen die zu diesen Integralen gehörigen Koeffizienten im Milstein-Verfahren weg. Dies ist z.B. der Fall wenn $n = m$ ist und die Bedingungen $b_{j,k} \equiv 0$ und $\partial b_{j,j} / \partial x_k \equiv 0$ für $k \neq j$ gelten. Diesen Fall nennt man *diagonales Rauschen*.

Ebenso gilt für manche Gleichungen die Identität $L_{j_1} b_{j_2} = L_{j_2} b_{j_1}$. In diesem Fall kann man je zwei Summanden im Milstein-Verfahren in der Form

$$L_{j_1} b_{j_2} \tilde{I}_{(j_1, j_2),t,t+h} + L_{j_2} b_{j_1} \tilde{I}_{(j_2, j_1),t,t+h} = L_{j_1} b_{j_2} (\tilde{I}_{(j_1, j_2),t,t+h} + \tilde{I}_{(j_2, j_1),t,t+h})$$

zusammenfassen. Die dadurch entstehende Summe von Mehrfachintegralen kann man dann wieder mittels der starken Approximation der Wiener-Prozesses ausdrücken, und zwar als

$$\tilde{I}_{(j_1, j_2),t,t+h} + \tilde{I}_{(j_2, j_1),t,t+h} = \Delta \tilde{W}^{j_1}(t) \Delta \tilde{W}^{j_2}(t).$$

Auf das Aktienkursmodell aus Abschnitt 4.3 ist aber leider keiner dieser Tricks anwendbar.

Das Itô-Taylor-Verfahren der schwachen Ordnung 2

Für dieses Verfahren benötigen wir die hierarchische Menge

$$\Gamma_2 = \{\odot, (0), \dots, (m), (0, 0), (0, 1), \dots, (0, m), (1, 0), \dots, (m, m)\}.$$

Um das Verfahren explizit aufzuschreiben ist es am einfachsten, die einzelnen Komponenten (Φ_1, \dots, Φ_n) von Φ anzugeben. Zur Abkürzung lassen wir zudem das Argument

$X(\omega)$ in allen auftretenden Funktionen sowie die Indizes $t, t+h$ bei den approximierten Mehrfachintegralen weg. Dann gilt

$$\begin{aligned}\Phi_k(t, X, h, W, \omega) &= X_k(\omega) + a_k \tilde{I}_{(0)} + L_0 a_k \tilde{I}_{(0,0)} \\ &+ \sum_{j=1}^m \left(b_{j,k} \tilde{I}_{(j)} + L_0 b_{j,k} \tilde{I}_{(0,j)} + L_j a_k \tilde{I}_{(j,0)} \right) \\ &+ \sum_{j_1, j_2=1}^m L_{j_1} b_{j_2, k} \tilde{I}_{(j_1, j_2)}\end{aligned}$$

mit den in Definition 7.13 definierten Operatoren L_j .

Im Fall $m = 1$ benötigen wir hier nur Approximationen der Mehrfachintegrale für einen Wiener-Prozess. Eine Analyse der Bedingung (7.17) zeigt hier, dass die schwache Approximation des Wiener-Prozesses aus Algorithmus 7.6 hier aber nicht genau genug ist, damit (7.17) für $\beta = 2$ gilt. Es gibt aber eine leichte Abhilfe: ersetzen wir die zweipunktverteilte Zufallsvariable ΔW_i mit den Werten $x_1 = -\sqrt{h}$, $x_2 = \sqrt{h}$ in Algorithmus 7.6 durch eine dreipunktverteilte Zufallsvariable mit den Werten $x_1 = -\sqrt{3h}$, $x_2 = 0$ und $x_3 = \sqrt{3h}$ und den Wahrscheinlichkeiten

$$\mathbb{P}_{\Delta W_i}(\{-\sqrt{3h}\}) = \mathbb{P}_{\Delta W_i}(\{\sqrt{3h}\}) = \frac{1}{6} \quad \text{und} \quad \mathbb{P}_{\Delta W_i}(\{0\}) = \frac{2}{3},$$

so gilt (7.17), wenn wir die Approximationen

$$\tilde{I}_{(0)} = h, \quad \tilde{I}_{(1)} = \Delta \tilde{W}, \quad \tilde{I}_{(0,0)} = \frac{h^2}{2}, \quad \tilde{I}_{(0,1)} = \tilde{I}_{(1,0)} = \frac{1}{2} h \Delta \tilde{W}, \quad \tilde{I}_{(1,1)} = \frac{1}{2} ((\Delta \tilde{W})^2 - h)$$

verwenden. Im Fall $m \geq 2$ können $I_{(j)}$, $I_{(0,j)}$, $I_{(j,0)}$ und $I_{(j,j)}$ analog mittels Approximationen $\Delta \tilde{W}^j$ berechnet werden, welche wiederum aus Algorithmus 7.6 mit der oben angegebenen dreipunktverteilten Zufallsvariablen berechnet werden. Die Approximation von I_{j_1, j_2} für $j_1 \neq j_2$, die beim Milstein-Verfahren im dort benötigten starken Fall große Probleme bereitet, ist im hier benötigten schwachen Fall einfacher, allerdings benötigt man auch hier über die im Algorithmus 7.6 verwendeten Zufallszahlen hinaus weitere Zufallszahlen. Genauer erzeugt man in jedem Schritt des Verfahrens für $j_1 = 1, \dots, m$ weitere zweipunktverteilte Zufallszahlen V_{j_1, j_2} mit $x_1 = -h$, $x_2 = h$ und

$$\mathbb{P}_{V_{j_1, j_2}}(\{-h\}) = \mathbb{P}_{V_{j_1, j_2}}(\{h\}) = \frac{1}{2}$$

für $j_2 = 1, \dots, j_1 - 1$ und setzt darüberhinaus $V_{j_1, j_1} := -h$ und $V_{j_1, j_2} := -V_{j_2, j_1}$ für $j_2 = j_1 + 1, \dots, m$. Mit Hilfe dieser Zahlen definiert man dann die Approximationen

$$\tilde{I}_{(j_1, j_2)} = \frac{1}{2} (\Delta \tilde{W}^{j_1} \Delta \tilde{W}^{j_2} + V_{j_1, j_2}).$$

Für diese Wahl von \tilde{I} kann man dann beweisen, dass (7.17) für $\beta = 2$ gilt. Insbesondere ist dieses Verfahren damit auf das Aktienkursmodell aus Abschnitt 4.3 anwendbar.

Ein ableitungsfreies Schema (in der Vorlesung nicht behandelt)

Ganz analog zur Herleitung des Heun-Verfahrens am Ende von Abschnitt 6.4.1 kann man die Ableitungen in den bisher betrachteten Schemata durch Differenzenquotienten passender Ordnung ersetzen, um stochastische Einschrittverfahren höherer Ordnung zu erhalten, in denen keine Ableitungen der Funktionen a und b_j verwendet werden.

Dies kann man für starke und schwache Verfahren machen; wir betrachten hier beispielhaft nur das schwache Itô-Taylor-Verfahren der Ordnung 2 im Spezialfall $m = 1$ mit a und $b = b_1$ unabhängig von t und ersetzen die auftretenden approximierten Mehrfachintegrale \tilde{I}_α gleich durch die oben verwendeten Approximationen. Mit etwas Rechnung ergibt sich dann das Verfahren

$$\begin{aligned}\Phi_k(t, X, h, \tilde{W}, \omega) &= X(\omega) + \frac{h}{2}(a(Y_1) + a(X(\omega))) \\ &\quad + \frac{1}{4}(b(Y_2) + b(Y_3) + 2b(X(\omega))\Delta\tilde{W}) \\ &\quad + \frac{1}{4\sqrt{h}}(b(Y_2) - b(Y_3))\left((\Delta\tilde{W})^2 - h\right)\end{aligned}$$

mit

$$\begin{aligned}Y_1 &= X(\omega) + ha(X(\omega)) + b(X(\omega))\Delta\tilde{W} \\ Y_2 &= X(\omega) + ha(X(\omega)) + b(X(\omega))\sqrt{h} \\ Y_3 &= X(\omega) + ha(X(\omega)) - b(X(\omega))\sqrt{h}.\end{aligned}$$

Im deterministischen Fall $b \equiv 0$ ist dies nichts anderes als das Heun-Verfahren. Für viele weitere Verfahren dieser Art siehe [3, Kapitel 11 und 15].

7.8 Das Binomialmodell als schwache Approximation der geometrischen Brownschen Bewegung

In Abschnitt 3.2 haben wir das Binomialmodell für einen Aktienkurs eingeführt. Bezeichnen wir die Kursvariable in diesem Modell mit S_B , so ist das Modell durch die Iterationsvorschrift

$$S_B(t+1) = S_B(t)Z_t$$

mit der zweipunktverteilten Zufallsvariablen $Z_t : \Omega \rightarrow \{\alpha_u, \alpha_d\}$ mit

$$\mathbb{P}_{Z_t}(\{\alpha_u\}) = p \quad \text{und} \quad \mathbb{P}_{Z_t}(\{\alpha_d\}) = 1 - p$$

gegeben ist. Wir setzen hier $p = 1/2$.

Betrachten wir nun die Euler-Approximation (7.4) der geometrischen Brownschen Bewegung und verwenden in Algorithmus 7.8 eine schwache Approximation, so ergibt sich die Iterationsvorschrift zu

$$\tilde{X}(t_{i+1}) = \tilde{X}(t_i) + hr\tilde{X}(t_i) + \Delta\tilde{W}(t_i)\sigma\tilde{X}(t_i),$$

wobei $\Delta\widetilde{W}(t_i)$ eine zweipunktverteilte Zufallsvariable mit Wert $\pm\sqrt{h}$ ist. Beachte, dass hier, wie in der risikoneutralen Optionsbewertung benötigt, $\mu = r$ gesetzt wurde.

Betrachten wir also das Binomialmodell mit

$$\alpha_d = 1 + hr - \sqrt{h}\sigma \quad \text{und} \quad \alpha_u = 1 + hr + \sqrt{h}\sigma, \quad (7.18)$$

so stimmen die beiden Iterationsvorschriften (bei gleichen Anfangswerten) überein, d.h. es gilt

$$S_B(i) \sim \widetilde{X}(t_i).$$

Mit dieser Wahl von α_d und α_u erhält man für $h \rightarrow 0$ aus der Binomialmethode also gerade den Optionswert auf Basis der geometrischen Brownschen Bewegung, der — wie wir im kommenden Kapitel sehen werden — gerade der Wert des Black-Scholes Modells ist. Beachte, dass ein Zeitschritt “1” im Binomialmodell hier gerade einem Zeitschritt “ h ” im Euler-Verfahren entspricht. Soll also der Optionswert zum Zeitpunkt T mit $h = T/N$ mit dem Binomialmodell mit α_u und α_d aus (7.18) berechnet werden, so müssen wir die Binomialmethode mit N Schritten anwenden. Zudem muss die Zinsrate r in der Binomialmethode durch hr ersetzt werden.

Die Wahl (7.18) ist allerdings nicht die einzig mögliche. In Aufgabe 10 (3. Übungsblatt) haben wir gesehen, dass Erwartungswert und Varianz des Binomialmodells mit den Werten der geometrischen Brownschen Bewegung übereinstimmen, wenn wir

$$\alpha_d = e^{hr} - \sqrt{e^{2hr}(e^{h\sigma^2} - 1)} \quad \text{und} \quad \alpha_u = e^{hr} + \sqrt{e^{2hr}(e^{h\sigma^2} - 1)} \quad (7.19)$$

wählen (diese Werte sind bereits auf einen Zeitschritt der Länge h umgerechnet, vgl. Aufgabe 12(ii) vom 4. Übungsblatt). Auch mit dieser Wahl der Parameter ist das Binomialmodell eine schwache Approximation der geometrischen Brownschen Bewegung, allerdings eine, die nicht mit dem Euler-Verfahren übereinstimmt. Dies folgt aus der Tatsache, dass man mit Hilfe der Taylor-Entwicklung nachrechnen kann, dass die Erwartungswerte $E(g(S_B(h)))$ für die Parameter aus (7.18) und (7.19) für beliebige C_P^4 Funktionen g nur um $O(h^2)$ voneinander abweichen, was aufsummiert über das Verfahren eine Abweichung von $O(h)$ ergibt, die für $h \rightarrow 0$ gegen Null strebt.

Kapitel 8

Die Black-Scholes Gleichung

Fisher Black¹ und Myron Scholes² sowie Robert Merton³ haben die nach den ersten beiden Autoren benannte partielle Differentialgleichung gleichzeitig in der ersten Hälfte der 1970er Jahre entwickelt. Merton und Scholes bekamen für ihre Arbeiten auf dem Gebiet der Finanzmathematik 1997 den Nobel-Preis für Wirtschaftswissenschaften (Black starb bereits 1995).

Wir wollen hier sowohl die Herleitung dieser Gleichung als auch ihre explizite Lösung angeben. Ziel ist die Berechnung der Funktion $V(t, S)$, die hier als Funktion in den zwei reellen Variablen S und t aufgefasst wird, also $V : \mathbb{R}_0^t \times [0, T] \rightarrow \mathbb{R}$. Wir erinnern daran, dass der Wert $V(T, S)$ dieser Funktion durch (2.1) oder (2.2) gegeben ist.

Wir fassen zunächst die Grundannahmen für die Black-Scholes Gleichung zusammen:

- (a) Es gibt keine Transaktionskosten und Gebühren, ebenso hat das eigene Kaufverhalten keinen Einfluss auf den Markt.
- (b) Der Markt ist Arbitrage-frei.
- (c) Der Kursverlauf ist durch (4.10) beschrieben.
- (d) Die risikofreie Zinsrate r ist konstant und im betrachteten Zeitraum werden keine Dividenden gezahlt.
- (e) Es wird eine risikoneutrale Optionsbewertung durchgeführt.

Beachte, dass wir diese Annahmen bereits bisher — explizit oder implizit — gemacht haben. Annahme (e) werden wir hier ähnlich wie in Abschnitt 2.1 verwenden, allerdings werden wir Portfolios betrachten, die im Zeitraum $[0, T]$ beliebig oft kostenneutral umgeschichtet werden können. Viele dieser Annahmen können abgeschwächt werden, führen dann aber auf (zum Teil deutlich) kompliziertere Gleichungen.

Die Idee der Black-Scholes Gleichung besteht darin, eine Gleichung für V herleiten, die nicht explizit von den Lösungen $S(t)$ von (4.10) abhängt. Diese Herleitung werden wir nun

¹US-amerikanischer Wirtschaftswissenschaftler, 1938–1995

²kanadischer Wirtschaftswissenschaftler, *1941

³US-amerikanischer Mathematiker und Wirtschaftswissenschaftler, *1944

durchführen wozu wir annehmen, dass $V(t, S)$ eine für $t < T$ zweimal stetig differenzierbare Funktion in t und S ist.

Unter dieser Annahme können wir das Itô-Lemma 4.2 auf $V(t, S(t))$ anwenden. Um die Schreibweise zu verkürzen, lassen wir die Argumente der Funktionen im Folgenden oft weg. Damit erhalten wir

$$dV = \left(\mu S \frac{\partial V}{\partial S} + \frac{\partial V}{\partial t} + \frac{1}{2} \sigma^2 S^2 \frac{\partial^2 V}{\partial S^2} \right) dt + \sigma S \frac{\partial V}{\partial S} dW. \quad (8.1)$$

$V(t, S)$ fassen wir hier als deterministische Funktion in den unabhängigen Variablen S und t auf, W und S hingegen sind stochastische Prozesse.

Nun kommt die risikoneutrale Bewertung ins Spiel: Dazu stellt der Emittent der Option ein Portfolio Π zusammen, das aus einer ausgegebenen (also negativen) Option sowie zeitlich veränderlichen Anteilen des Basiswertes sowie einer festverzinslichen Anleihe besteht, also

$$\Pi(t) = -V(t, S(t)) + \Delta(t)S(t) + \beta(t)B(t). \quad (8.2)$$

Ziel der risikoneutralen Bewertung ist es nun, den Preis $V(t, S)$ so zu bestimmen, dass sich dieses Portfolio bei geeigneter Wahl von $\Delta(t)$ und $\beta(t)$ risikofrei entwickelt, also – gemäß der Arbitragefreiheit – wie eine festverzinsliche Anleihe verhält. Durch Anlegen des Portfolios zum Zeitpunkt $t = 0$ (wobei der vom Optionskäufer zu zahlende Preis $V(0, S(0))$ investiert wird) und Auflösen zur Zeit $t = T$ (wobei dem Kunden dann der Wert $V(T, S(T))$ ausgezahlt wird) kann die Bank also das mit der Ausgabe der Option verbundene Risiko komplett “weghedgen”.

Dieses Portfolio soll dabei selbstfinanzierend sein, d.h. es soll während der Laufzeit nur umgeschichtet werden, es darf aber kein Geld hineingesteckt oder herausgenommen werden. Was dies formal bedeutet, macht man sich am Besten für stückweise konstante Werte $\Delta(t)$ und $\beta(t)$ klar. Nehmen wir an, es gilt $\Delta|_{[t_0, t_1)} = \Delta_0$, $\Delta|_{[t_1, t_2)} = \Delta_1$, $\beta|_{[t_0, t_1)} = \beta_0$ und $\beta|_{[t_1, t_2)} = \beta_1$ für Zeiten $t_0 < t_1 < t_2$. Zum Zeitpunkt t_1 wird das Portfolio also umgeschichtet.

Dann beträgt der Wert des Portfolios unmittelbar vor der Umschichtung

$$\Pi_v(t_1) = -V(t_1, S(t_1)) + \Delta_0 S(t_1) + \beta_0 B(t_1)$$

und nach der Umschichtung

$$\Pi_n(t_1) = -V(t_1, S(t_1)) + \Delta_1 S(t_1) + \beta_1 B(t_1).$$

Das Portfolio ist nun gerade dann selbstfinanzierend, wenn diese beiden Werte übereinstimmen, also wenn

$$\Delta_0 S(t_1) + \beta_0 B(t_1) = \Delta_1 S(t_1) + \beta_1 B(t_1)$$

gilt. Diese Bedingung können wir auch anders ausdrücken: Betrachten wir den Wert

$$\Pi(t_2) = -V(t_2, S(t_2)) + \Delta_1 S(t_2) + \beta_1 B(t_2)$$

so ist das Portfolio genau dann selbstfinanzierend, wenn

$$\begin{aligned} \Pi(t_2) &= \Pi(t_0) + \Delta_0 S(t_1) + \beta_0 B(t_1) - \Delta_0 S(t_0) - \beta_0 B(t_0) + V(t_0, S(t_0)) \\ &\quad + \Delta_1 S(t_2) + \beta_1 B(t_2) - \Delta_1 S(t_1) - \beta_1 B(t_1) - V(t_2, S(t_2)) \end{aligned}$$

gilt. Für $p + 1$ Zeitpunkte t_0, \dots, t_p ergibt sich die Bedingung dann zu

$$\begin{aligned} \Pi(t_p) &= \Pi(t_0) + V(t_0, S(t_0)) - V(t_p, S(t_p)) \\ &\quad + \sum_{k=0}^{p-1} \Delta(t_k)(S(t_{k+1}) - S(t_k)) + \sum_{k=0}^{p-1} \beta(t_k)(B(t_{k+1}) - B(t_k)). \end{aligned}$$

Für kontinuierliche Umschichtung erhalten wir durch Grenzübergang $t_{k+1} - t_k \rightarrow 0$ daraus die Bedingung

$$\Pi(t) = \Pi(t_0) + V(t_0, S(t_0)) - V(t, S(t)) + \int_{t_0}^t \Delta(\tau) dS(\tau) + \int_{t_0}^t \beta(\tau) dB(\tau),$$

wobei die Integrale im Itô-Sinne mit S bzw. B an Stelle von W zu verstehen sind. Schreiben wir dies wieder in unserer symbolischen SDG-Schreibweise und nutzen aus, dass $dB = rBdt$ gilt, so erhalten wir die Gleichung

$$d\Pi(t) = -dV(t, S(t)) + \Delta(t)dS(t) + \beta(t)rB(t)dt. \quad (8.3)$$

Setzen wir die Gleichungen (4.10) und (8.1) in diese Gleichung ein, so erhalten wir

$$d\Pi = - \left(\mu S \left(\frac{\partial V}{\partial S} - \Delta \right) + \frac{\partial V}{\partial t} + \frac{1}{2} \sigma^2 S^2 \frac{\partial^2 V}{\partial S^2} - \beta r B \right) dt - \sigma S \left(\frac{\partial V}{\partial S} - \Delta \right) dW.$$

Damit sich das Portfolio risikofrei entwickeln kann, müssen zunächst die stochastischen Terme in $d\Pi$ eliminiert werden. Dies erreichen wir, indem wir

$$\Delta(t) = \frac{\partial V}{\partial S}(t, S(t))$$

setzen. Damit erhalten wir

$$d\Pi = - \left(\frac{\partial V}{\partial t} + \frac{1}{2} \sigma^2 S^2 \frac{\partial^2 V}{\partial S^2} - \beta r B \right) dt \quad (8.4)$$

also eine deterministische Differentialgleichung.

Schließlich verwenden wir die Arbitrage-Freiheit, um V zu bestimmen. Wir suchen einen Wert von V , für den sich Π risikofrei entwickelt. Wenn ein solcher Wert existiert, muss sich Π nach der Annahme der Arbitrage-Freiheit wie eine festverzinsliche Anleihe entwickeln, es muss also $\Pi(t) = e^{rt}\Pi(0)$ gelten. Daraus folgt die Differentialgleichung

$$d\Pi = r\Pi dt,$$

woraus mit (8.2) und $\Delta(t) = \frac{\partial V}{\partial S}(t, S(t))$ die Gleichung

$$d\Pi = r\Pi dt = \left(-rV + rS \frac{\partial V}{\partial S} + r\beta B \right) dt, \quad (8.5)$$

folgt.

Nun haben wir zwei Gleichungen für $d\Pi$, nämlich gerade (8.4) und (8.5). Gleichsetzen dieser Gleichungen liefert

$$\frac{\partial V}{\partial t} + \frac{1}{2}\sigma^2 S^2 \frac{\partial^2 V}{\partial S^2} - rV + rS \frac{\partial V}{\partial S} = 0. \quad (8.6)$$

Dies ist gerade die Black-Scholes Gleichung.

Die Black-Scholes Gleichung (8.6) ist also eine partielle Differentialgleichung, die von V erfüllt wird. Die Gleichung (8.6) hat (genau wie eine gewöhnliche Differentialgleichung ohne Anfangsbedingung) unendlich viele Lösungen, wenn wir aber geeignete *Randwerte* $V(T, S)$, für ein $T > 0$ und alle $S \in [0, \infty)$ festlegen, so erhalten wir eine eindeutige Lösung. Dies ist hier einfach, da sich die Randwerte in natürlicher Weise aus den Optionswerten zum Fälligkeitszeitpunkt $t = T$ ergeben.

Tatsächlich ist Gleichung (8.6) explizit lösbar. Für eine europäische Call-Option gilt

$$V_{Ce}(t, S) = SP(a) - Ke^{-r(T-t)}P(b), \quad (8.7)$$

während sich für die europäische Put-Option

$$V_{Pe}(t, S) = S(P(a) - 1) - Ke^{-r(T-t)}(P(b) - 1)$$

ergibt. Hierbei sind in beiden Fällen

$$a = \frac{\log \frac{S}{K} + \left(r + \frac{\sigma^2}{2}\right)(T-t)}{\sigma\sqrt{T-t}}, \quad b = a - \sigma\sqrt{T-t}$$

und die Funktion P ist die Verteilungsfunktion der Standard-Normalverteilung gegeben durch

$$P(y) = \int_{-\infty}^y \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right) dx.$$

Abbildung 8.1 zeigt Lösungen für den europäischen Call mit $r = 0.06$, $K = 1$, $T = 1$ und $\sigma = 0.3$ bzw. $\sigma = 1$.

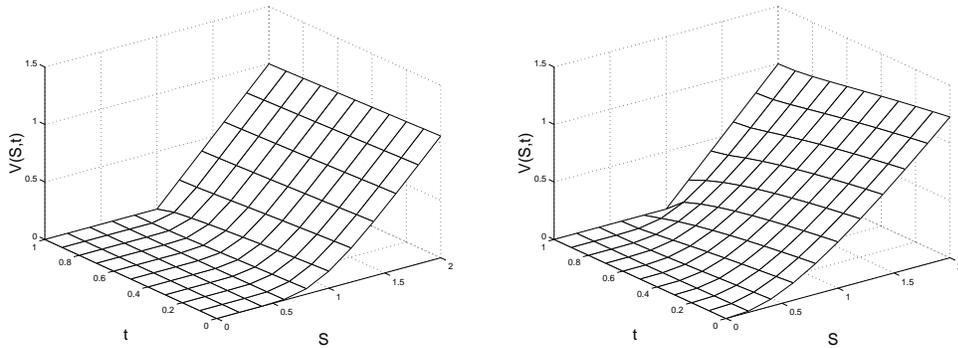


Abbildung 8.1: Lösung der Black-Scholes Gleichung für den europäischen Call mit $r = 0.06$, $K = 1$, $T = 1$ und $\sigma = 0.3$ (links) bzw. $\sigma = 1$ (rechts)

Das hier verwendete Portfolio hat eine enge Beziehung zu dem Portfolio π aus Abschnitt 2.1. Wählen wir $\beta(0)$ so, dass $\Pi(0) = 0$ gilt, so hat das Portfolio Π für alle Zeiten $t \in [0, T]$ den Wert $\Pi(t) = 0$, da es sich ja risikoneutral entwickelt. Folglich hat das Portfolio $\Delta(t)S(t) + \beta(t)$ gerade den Wert $V(t, S(t))$ und bildet damit den Optionswert nach — es ist also wieder ein replizierendes Portfolio.

Wenn man die vereinfachenden Annahmen abschwächt oder wenn man amerikanische an Stelle von europäischen Optionen betrachtet, so erhält man deutlich kompliziertere partielle Differentialgleichungen von ähnlichem Typ wie (8.6), die dann nur noch numerisch gelöst werden können.

Bemerkung 8.1 (i) Die Annahme der risikoneutralen Bewertung wurde hier in die Modellannahme umgesetzt, dass der Emittent der Option sein Risiko durch das Vorhalten des selbstfinanzierenden Portfolios Π ausschalten kann. Diese Annahme ist wesentlich für den resultierenden Optionspreis. Sie bedeutet aber auch, dass der Emittent dazu sein Portfolio stets geeignet umschichten muss; tut er dies nicht, bleibt ein Risiko erhalten. Dieses Ausschalten des Risikos durch Umschichten nennt man “Hedging”. Dies bedeutet insbesondere, dass die Black-Scholes Gleichung tatsächlich viel mehr als nur den Optionspreis liefert: Sie liefert auch eine Regel für die risikoneutrale Anlagestrategie (auch “Hedging-Strategie” genannt) des Emittenten. Wir werden auf das Hedging im folgenden Kapitel näher eingehen.

(ii) Beachte, dass wir hier weder die Abzinsungsformel aus Algorithmus 2.2(3) noch die Wachstumsbedingung (2.5) an S verwendet haben. Insbesondere gilt der durch die Black-Scholes Gleichung ermittelte Wert tatsächlich für Basiswerte mit beliebigem Trend μ , da der Effekt des Trends durch das Hedging eliminiert wird.

Tatsächlich folgt die Abzinsungsformel in Algorithmus 2.2(3) (und damit auch (2.5)) analog zur Herleitung in Abschnitt 2.1 aus der hier rein ökonomisch interpretierten Risikoneutralität, vgl. auch [2, Chapter 12]. \square

Kapitel 9

Hedging und die “Griechen”

Die Black-Scholes Gleichung liefert nicht nur eine — im Fall europäischer Optionen sogar explizit lösbare — Gleichung für den Optionswert. Aus den Annahmen in ihrer Herleitung ergibt sich darüberhinaus eine Regel, wie sich eine Emittentin gegen das durch die Ausgabe der Option entstehende Risiko absichern kann. Dies geschieht nämlich gerade durch das in der Herleitung konstruierte Portfolio mit den aus der Herleitung folgenden Umschichtungen Δ und β . Dieses Vorgehen nennt man *Hedging*.

Kontrolltheoretisch kann man Δ und β als Kontrollfunktionen betrachten, mit denen man das Portfolio Π steuert. In diesem Sinne ist die Black-Scholes Strategie

$$\Delta = \Delta(t, S(t)) = \frac{\partial V}{\partial S}(t, S(t))$$

dann ein sogenanntes (zeitabhängiges) Feedback, d.h. der gewünschte aktuelle Anteil des Basiswertes S im Portfolio hängt von der aktuellen Zeit t und vom aktuellen Kurs $S(t)$ ab. Um dies in der Praxis (oder hier zumindest in numerischen Simulationen) umzusetzen, müssen wir $\Delta(t, S)$ als Funktion in t und S berechnen. Wir machen dies ausführlich für den europäischen Call. Hierzu benötigen wir zwei Hilfsresultate. Zum einen gilt offenbar

$$P'(x) = \frac{1}{\sqrt{2\pi}} \exp\left(\frac{-x^2}{2}\right)$$

und zum anderen gilt die Gleichung

$$SP'(a) - e^{-r(T-t)}KP'(b) = 0, \tag{9.1}$$

was in einer Übungsaufgabe nachgerechnet wird. Leiten wir nun (8.7) nach S ab so erhalten wir

$$\begin{aligned} \Delta(t, S) &= P(a) + SP'(a)\frac{\partial a}{\partial S} - Ke^{-r(T-t)}P'(b)\frac{\partial b}{\partial S} \\ &= P(a) + \frac{P'(a)}{\sigma\sqrt{T-t}} - Ke^{-r(T-t)}\frac{P'(b)}{S\sigma\sqrt{T-t}}. \end{aligned}$$

Wegen (9.1) heben sich die beiden hinteren Terme gerade gegeneinander auf, woraus $\Delta(t, S) = P(a)$ folgt. Eine ähnliche Rechnung für den europäischen Put liefert $\Delta(t, S) = P(a) - 1$, woraus insgesamt

$$\Delta(t, S(t)) = P(a) \text{ (europ. Call)} \quad \text{und} \quad \Delta(t, S(t)) = P(a) - 1 \text{ (europ. Put)} \tag{9.2}$$

folgt.

Die zweite Größe β ergibt sich dann aus der Selbstfinanzierungs-Bedingung und könnte daraus errechnet werden. Für die praktische Implementierung des Hedging werden wir im Folgenden aber etwas anders vorgehen, so dass wir die exakte Formel für β gar nicht benötigen.

9.1 Zeitdiskretes Hedging

Das kontinuierliche Umschichten des Portfolios Π ist in der Praxis nicht zu bewerkstelligen. Dies würde nämlich bedeuten, die Gewichte Δ und β zu jedem Zeitpunkt $t \in [0, T]$ an die aktuellen Kurse anzupassen.

Realistisch ist es hingegen nur möglich, die Gewichte zu festen diskreten Zeitpunkten $t_i \in [0, T]$ anzupassen, d.h. das Portfolio zu diesen Zeiten umzuschichten. Wir betrachten hier den einfachsten Fall, nämlich dass das Portfolio alle τ Zeiteinheiten umgeschichtet wird, d.h. zu den Zeiten $t_i = i\tau$, $i = 1, \dots, N$, wobei $\tau > 0$ so gewählt ist, dass $N\tau = T$ gilt.

Das solchermaßen zeitdiskret gehedgte Portfolio ergibt sich damit für $t \in [t_i, t_{i+1}]$ zu

$$\Pi(t) = -V(t, S(t)) + \Delta(t_i, S(t_i))S(t) + \beta(t_i)B(t).$$

In der Sprache der Kontroll- bzw. Regelungstheorie werden Δ und β als sogenanntes *Abtastfeedback* implementiert. Zu jedem Zeitpunkt t_i , $i = 1, \dots, N - 1$ werden die Anteile dann umgeschichtet, indem $\Delta(t_{i-1}, S(t_{i-1}))$ durch $\Delta(t_i, S(t_i))$ und $\beta(t_{i-1})$ durch $\beta(t_i)$ ersetzt wird. Das $\Delta(t_i, S(t_i))$ wird dabei gemäß der Zeit t_i und dem Kurs $S(t_i)$ gemäß (9.2) berechnet und das $\beta(t_i)$ wird so gewählt, dass die Umschichtung selbstfinanzierend, also kostenneutral erfolgt.

Wir berechnen nun eine Formel für dieses $\beta(t_i)$. Zum Zeitpunkt t_i besitzt das Portfolio vor der Umschichtung den Wert

$$\Pi(t_i) = -V(t_i, S(t_i)) + \Delta(t_{i-1}, S(t_{i-1}))S(t_i) + \beta(t_{i-1})B(t_i). \quad (9.3)$$

Zum Zeitpunkt t_i wird nun der Anteil des Basiswertes auf $\Delta(t_i, S(t_i))S(t_i)$ verändert. Das Portfolio ist bei dieser Umschichtung genau dann selbstfinanzierend, wenn sich der Wert des Portfolios dabei nicht ändert. Mit $\Pi(t_i)$ aus (9.3) muss also nach der Umschichtung

$$\Pi(t_i) = -V(t_i, S(t_i)) + \Delta(t_i, S(t_i))S(t_i) + \beta(t_i)B(t_i) \quad (9.4)$$

gelten. Aus (9.4) lässt sich $\beta(t_i)$ nun eindeutig bestimmen. Es gilt

$$\beta(t_i) = \frac{\Pi(t_i) + V(t_i, S(t_i)) - \Delta(t_i, S(t_i))S(t_i)}{B(t_i)}.$$

Dies führt zu dem folgenden Algorithmus, in dem ein mittels der Black-Scholes Formeln gehedgtes Portfolio numerisch simuliert wird.

Algorithmus 9.1 (Simulation des Black-Scholes Portfolios mit zeitdiskretem Hedging)

Gegeben: $r, \sigma, T, S_0, B_0, \beta_0 > 0$ und ein Zeitgitter $\mathcal{T} = \{t_0, \dots, t_N\}$ mit $t_i = i\tau$, $\tau > 0$ und $t_N = T$.

- (1) Berechne Kurswerte $S(t_i)$, $i = 0, \dots, N$ mit $S(t_0) = S_0$ mittels der geometrischen Brownschen Bewegung auf Basis eines Pfades der starken Approximation des Wiener-Prozesses auf \mathcal{T}
- (2) Berechne $V(t_i, S(t_i))$, $\Delta(t_i, S(t_i))$ und $B(t_i) = e^{rt_i} B_0$, $i = 0, \dots, N$
- (3) Berechne das Portfolio induktiv mittels $\beta(0) := \beta_0$,

$$\Pi(t_0) := -V(t_0, S(t_0)) + \Delta(t_0, S(t_0))S(t_0) + \beta(t_0)B(t_0)$$

und

$$\Pi(t_i) := -V(t_i, S(t_i)) + \Delta(t_{i-1}, S(t_{i-1}))S(t_i) + \beta(t_{i-1})B(t_i),$$

$$\beta(t_i) := \frac{\Pi(t_i) + V(t_i, S(t_i)) - \Delta(t_i, S(t_i))S(t_i)}{B(t_i)}$$

für $i = 1, \dots, N$

□

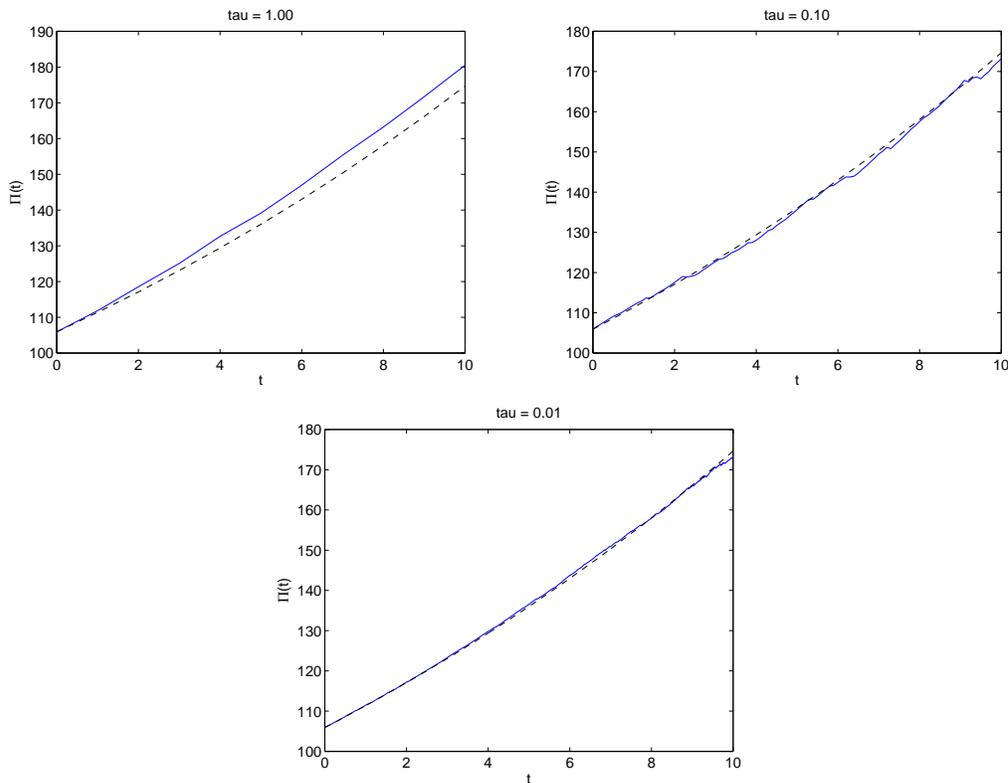


Abbildung 9.1: Simulierte Entwicklung des Portfolios Π bei diskretem Hedging mit $\tau = 1, 0.1, 0.01$ (blau durchgezogen). Zum Vergleich ist die theoretisch erwartete Entwicklung $e^{rt}\Pi(0)$ schwarz gestrichelt eingezeichnet.

Abbildung 9.1 zeigt Ergebnisse dieser Simulation für einen europäischen Call mit den Parametern $S_0 = B_0 = 80$, $T = 10$, $r = 0.05$, $\sigma = 0.3$, $K = 100$, $\beta_0 = 1$ sowie $\tau = 1, 0.1$ und

0.01. Wie man sieht, erfüllt das Ergebnis voll und ganz die Erwartungen: Das Portfolio folgt fast genau der deterministischen Entwicklung $e^{rt}\Pi(0)$, die zum Vergleich gestrichelt eingezeichnet ist, und zwar um so genauer je kleiner τ ist.

Ein interessantes Phänomen zeigt sich, wenn man den Anteil Δ des Basiswertes im Portfolio in Abhängigkeit von t plottet, vgl. Abbildung 9.2.

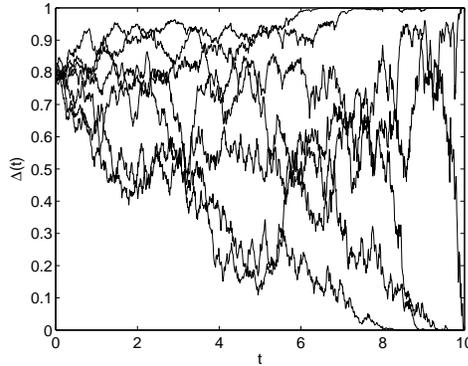


Abbildung 9.2: Verlauf von $\Delta(t)$ für verschiedene simulierte Entwicklungen des Portfolios Π mit $\tau = 0.01$

Betrachtet man den Wert $\Delta(T)$ am Laufzeitende, so stellt man fest, dass hier nur die Werte 0 oder 1 angenommen werden, d.h. entweder befindet sich am Laufzeitende genau ein Anteil des Basiswertes im Portfolio oder gar kein Anteil. Dass dies kein Zufall ist, sieht man in Abbildung 9.3, in der Δ (für die gleichen Parameterwerte wie oben) in Abhängigkeit von t und S dargestellt ist.

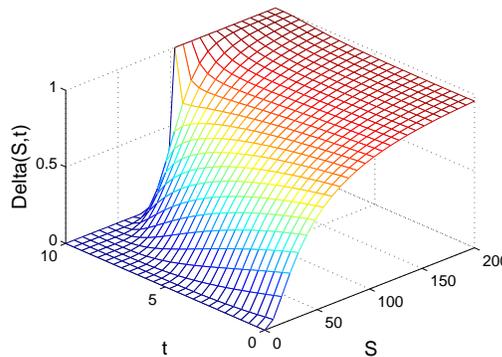


Abbildung 9.3: Δ in Abhängigkeit von S und t

Zum Zeitpunkt $t = T = 10$ nimmt Δ hier scheinbar nur zwei unterschiedliche Werte an: Im Falle $S < K = 100$ gilt $\Delta(t, S) = 0$ und im Fall $S > K = 100$ gilt $\Delta(t, S) = 1$. Dies kann man aus der Formel für Δ auch explizit errechnen: Für $S > K$ gilt $a \rightarrow \infty$ für $t \nearrow T$ und damit $P(a) \rightarrow 1$ während für $S < K$ gilt $a \rightarrow -\infty$ für $t \nearrow T$ und damit $P(a) \rightarrow 0$ folgt

(die Schreibweise $t \nearrow T$ bedeutet hier “ $t \rightarrow T$ und $t < T$ ”). Im Grenzfall $S = K$ gilt $a = 0$ und damit $P(a) \rightarrow 1/2$ für $t \nearrow T$; tatsächlich gibt es also noch einen dritten Wert, der allerdings in stochastischen Simulationen wie in Abbildung 9.2 praktisch nie erreicht wird, weil er nur in einem Punkt angenommen wird. Für die drei Fälle $S(T) > K$, $S(T) = K$ und $S(T) < K$ haben sich in der Finanzmathematik die Begriffe “im Geld”, “am Geld” und “aus dem Geld” eingebürgert. Zudem sieht man an der Formel ebenso wie in Abbildung 9.3, dass Δ hier immer im Intervall $[0, 1]$ liegt.

Für den europäischen Put kann man eine ähnliche Analyse machen. Hier erhält man, dass Δ stets im Intervall $[-1, 0]$ liegt (das Portfolio also in der Regel Leerverkäufe von S enthält) und dass $\Delta(T, S(T)) = 0$ für $S(T) > K$, $\Delta(T, S(T)) = -1$ für $S(T) < K$, und $\Delta(T, S(T)) = -1/2$ für $S(T) = K$ gilt.

Eine ganz wesentliche Annahme dafür, dass das Hedging so gut funktioniert, ist die Korrektheit (im stochastischen Sinne) des zu Grunde liegenden Kursmodells. Die Hedging-Strategie ist nämlich darauf angewiesen, dass sich der Kurs tatsächlich gemäß des Modells verhält, da sie durch die Herleitung aus den Modellannahmen implizit auf einer stochastischen Vorhersage des zukünftigen Verhaltens basiert.

Verhält sich der Kurs nicht entsprechend den Vorhersagen, so bricht die deterministische Entwicklung des Portfolios Π zusammen. Als Beispiel betrachten wir noch einmal die obige Simulation, in der der simulierte Kurswert mittels der Skalierung $S(t_i) \rightarrow S(t_i)/2$ für $t_i \geq 5$ verändert wurde. Dies entspricht einem “Kurssturz” zur Zeit $t = 5$, bei dem sich der Kurs halbiert. Nach dem Sturz verläuft der Kurs wieder gemäß der geometrischen Brownschen Bewegung, allerdings nun auf niedrigerem Niveau. Abbildung 9.4 zeigt, was dabei mit dem Portfolio Π passiert: der Wert bricht zur Zeit $T/2$ deutlich ein. Er wächst danach (wenn das Modell wieder stimmt) zwar wieder deterministisch, allerdings auf niedrigerem Niveau. Die Auswirkungen dieses Verhaltens auf die Bilanzen von Banken, Fondsgesellschaften etc. in Zeiten der Finanzkrise brauchen sicherlich nicht weiter kommentiert werden.

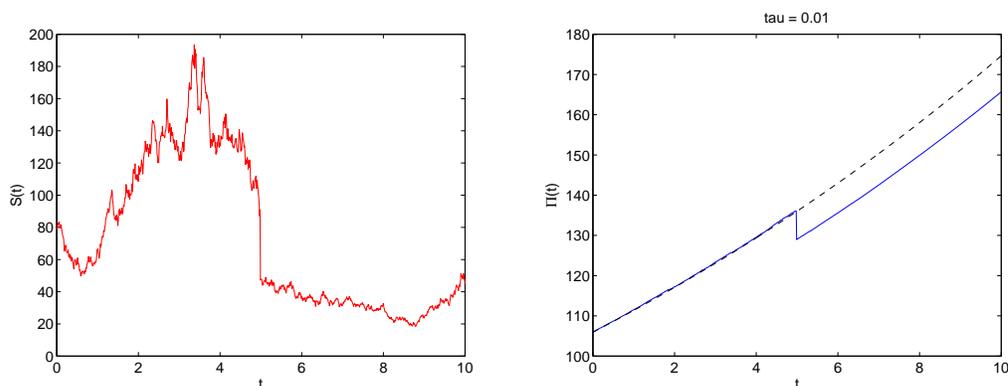


Abbildung 9.4: Simulierte Entwicklung des Kurses S (links) und des Portfolios Π (rechts) bei diskretem Hedging mit $\tau = 0.01$ und simuliertem “Kurssturz” zur Zeit $t = 5$ (blau durchgezogen). Zum Vergleich ist die theoretisch erwartete Entwicklung $e^{rt}\Pi(0)$ schwarz gestrichelt eingezeichnet.

Mit diesem Kurssturz hängt das genaue Verhalten von Π nun wieder vom Zufall ab. Je nach

dem sonstigen Verlauf des Kurses $S(t_i)$ wirkt sich der Sturz mehr oder weniger schlimm aus. Dieses Phänomen soll — ebenso wie die Frage was bei einem unerwarteten Kursanstieg oder einem vorübergehenden Kurssturz passiert — auf dem aktuellen Übungsblatt numerisch untersucht werden.

9.2 Die "Griechen"

Die Ableitung $\Delta = \frac{\partial V}{\partial S}$ spielt offenbar eine bedeutende Rolle beim Hedging. Neben diesem "Delta" werden auch alle anderen Ableitungen des Optionswerts mit griechischen Buchstaben versehen. Es bezeichnen

$$\begin{aligned}\Gamma &= \frac{\partial^2 V}{\partial S^2} \quad (\text{"Gamma"}) & \rho &= \frac{\partial V}{\partial r} \quad (\text{"Rho"}) \\ \Theta &= \frac{\partial V}{\partial t} \quad (\text{"Theta"}) & \text{vega} &= \frac{\partial V}{\partial \sigma} \quad (\text{"Vega"}).\end{aligned}$$

Im Englischen werden diese Werte als "Greeks", also "Griechen" bezeichnet, im deutschen existiert dafür auch der Begriff "Kennzahlen" oder "Sensitivitäten".

Neben dem für das praktische Hedging besonders wichtigen Δ geben auch die anderen "Griechen" gewisse Einsichten. Wir berechnen hier beispielhaft Formeln für Γ und ρ für den europäischen Call.

Zur Berechnung von Γ müssen wir Δ noch einmal ableiten. Dies ergibt

$$\Gamma = \frac{\partial \Delta}{\partial S} = \frac{\partial P(a)}{\partial S} = \frac{P'(a)}{S\sigma\sqrt{T-t}}.$$

Für ρ erhalten wir

$$\begin{aligned}\rho &= SP'(a)\frac{\partial a}{\partial r} + (T-t)Ke^{-r(T-t)}P(b) - Ke^{-r(T-t)}P'(b)\frac{\partial b}{\partial r} \\ &= SP'(a)\frac{T-t}{\sigma\sqrt{T-t}} + (T-t)Ke^{-r(T-t)}P(a) - Ke^{-r(T-t)}P'(b)\frac{T-t}{\sigma\sqrt{T-t}}\end{aligned}$$

Wegen (9.1) heben sich der erste und der dritte Term gerade gegeneinander auf, woraus

$$\rho = (T-t)Ke^{-r(T-t)}P(a)$$

folgt. Ähnliche Rechnungen zeigen

$$\Theta = \frac{-S\sigma}{2\sqrt{T-t}}P'(a) - rKe^{-r(T-t)}P(b)$$

und

$$\text{vega} = S\sqrt{T-t}P'(a).$$

Aus diesen Formeln sieht man z.B. die folgenden Ungleichungen und die daraus folgenden Monotonieeigenschaften des Optionswertes für $t < T$ (hier jeweils für den europäischen Call):

- es gilt $\rho > 0$, d.h. der Optionswert steigt, wenn die risikofreie Zinsrate steigt. Dies kann anschaulich wie folgt erklärt werden. Wird r erhöht, hat dies in den Formeln den gleichen Effekt als wenn K verringert wird (genauer ist eine Erhöhung von r auf $\bar{r} > r$ äquivalent zur Verringerung von K auf $\bar{K} = Ke^{-(\bar{r}-r)(T-t)}$). Ein niedrigerer Ausübungspreis erhöht aber den Wert der Call-Option, da der Basiswert günstiger erworben werden kann.
- es gilt $\Theta < 0$, d.h. der Optionswert fällt mit fortschreitender Zeit t . Diese Eigenschaft kann tatsächlich unabhängig von dem verwendeten Kursmodell mit Arbitrageargumenten ähnlich zu denen in Abschnitt 2.3 hergeleitet werden, siehe Higham [2, Abschnitt 10.3 und Übungsaufgabe 10.2].
- es gilt $vega > 0$, d.h. der Wert der Option steigt, wenn die Volatilität σ steigt. Dies kann man wie folgt erklären: bei größerer Volatilität überdecken die Werte von $S(T)$ ein größeres Intervall. Werte, die näher an Null liegen, liefern dabei aber keine kleineren Optionswerte, da für $S(T) \leq K$ in jedem Fall $V(T, S(T)) = (S(T) - K)^+ = 0$ gilt. Größere Werte von $S(T)$ ergeben größere Werte, weswegen der Wert der Option zum Ausübungszeitpunkt T und damit auch für kleinere Zeiten $t \in [0, T)$ steigt.
- es gilt $\Gamma > 0$, d.h. der Anteil Δ des Basiswertes im Portfolio wird um so größer je höher der Preis des Basiswertes ist. Eine anschauliche Erklärung dafür ist, dass die Schwankung des Anteils $-V(t, S(t))$ im Portfolio um so größer ist, je größer der Wert der Option ist. Damit sich das Portfolio insgesamt risikofrei entwickelt, muss dies durch einen höheren Anteil des Basiswertes im Portfolio ausgeglichen werden.

9.3 Monte-Carlo für das Δ

Wenn wir komplexere Kursmodelle als die geometrische Brownsche Bewegung verwenden, ist die Monte-Carlo Simulation oft die einzige Methode, mit denen eine Optionsbewertung numerisch durchgeführt werden kann. Ebenso gibt es eine Reihe sogenannter pfadabhängiger Optionsarten, für die die Monte-Carlo Methode gut geeignet ist. Ein Beispiel ist die sogenannte asiatische Option, bei der die Auszahlung (im Falle eines Call) am Ende der Laufzeit durch

$$V(T, S(\cdot)) = \max\left(\frac{1}{T} \int_0^T S(\tau) d\tau - E, 0\right)$$

gegeben ist. Der Wert der Option hängt also vom Durchschnittswert und nicht nur vom Endwert des Basiskurses im Intervall $[0, T]$ ab.

Approximiert man das Integral durch eine Quadraturformel (z.B. durch die zusammengesetzte Trapez-Regel) mit Stützstellen $t_i = hi$, $i = 0, \dots, N$, $hN = T$, und erweitert man die Monte-Carlo-Methode so, dass nicht nur wie in Kapitel 5 beschrieben der Wert $S(T)$ sondern die Werte $S(t_i)$, $i = 0, \dots, N$ berechnet werden, so kann man den Wert jedes Pfades errechnen. Bezeichnen wir den mit $e^{r(t-T)}$ abdiskontierten Wert des i -ten simulierten Pfades zur Zeit t und Anfangskurswert $S > 0$ mit $V(t, S, \omega_i)$ (der einfacheren Notation wegen lassen wir hier und im Folgenden die sonst üblichen Schlangen über den Ausdrücken, die die numerische Approximation kennzeichnen, weg), so kann man den Monte-Carlo-Wert

der Option als

$$V_{Mc}(t, S) = \frac{1}{n} \sum_{i=1}^n V(t, S, \omega_i)$$

schreiben.

Unabhängig davon, welche Option man bewertet und welches Kursmodell man verwendet, benötigt man auch bei der Verwendung von Monte-Carlo-Methoden nun das Delta $\Delta = \frac{\partial V}{\partial S}$, um die Hedging-Strategie zu bestimmen¹. Zur Berechnung von Δ verwenden wir die aus der Taylor-Entwicklung stammende Approximation

$$\Delta(t, S) = \frac{\partial V}{\partial S}(t, S) = \frac{V(t, S + \Delta S) - V(t, S)}{\Delta S} + O(\Delta S)$$

für kleine $\Delta S \in \mathbb{R}$.

Eine naheliegende Anwendung dieser Formel wäre nun, das Δ mittels

$$\Delta(t, S) \approx \frac{V_{Mc}(t, S + \Delta S) - V_{Mc}(t, S)}{\Delta S} \quad (9.5)$$

zu approximieren. Da der Fehler in V_{Mc} von der Ordnung $O(\sqrt{n})$ ist und durch ΔS geteilt wird, ist der Fehler der Approximation (9.5) von der Ordnung

$$O(\sqrt{n}/\Delta S) + O(\Delta S).$$

Der sowieso schon recht große Fehler der Monte-Carlo-Simulation verstärkt sich also durch das Bilden des Differenzenquotienten noch weiter.

Obwohl sich dieses grundsätzliche Problem nicht vollständig beheben lässt, kann man eine bessere Approximation erhalten, wenn man das Δ pfadweise berechnet. Dazu setzt man zunächst

$$\Delta(t, S, \omega_i) := \frac{V(t, S + \Delta S, \omega_i) - V(t, S, \omega_i)}{\Delta S},$$

wobei gleiches ω_i bedeutet, dass wir die gleichen Pfade des Wiener-Prozesses verwenden. Dann setzen wir

$$\Delta_{Mc}(t, S) := \frac{1}{n} \sum_{i=1}^n \Delta(t, S, \omega_i). \quad (9.6)$$

Der Grund dafür, warum diese zweite Variante bessere Werte liefert, liegt in der niedrigeren Varianz der aufsummierten Zufallszahlen. Dazu erinnern wir uns an Satz 5.3, der besagt dass die p_n -Konfidenzintervalle I_n der Monte-Carlo-Methode von der Form

$$I_n = \left[\mu - \frac{k\sigma}{\sqrt{n}}, \mu + \frac{k\sigma}{\sqrt{n}} \right]$$

sind, wobei σ die Varianz der Zufallsvariablen in der Monte-Carlo-Methode beschreibt.

¹Tatsächlich ist a priori natürlich nicht klar, ob die aus der Black-Scholes-Theorie abgeleitete Formel $\Delta = \partial V / \partial S$ korrekt ist, wenn sich für das betrachtete Problem gar keine Black-Scholes-Gleichung aufstellen lässt. Wir gehen hier der Einfachheit halber davon aus, dass dies der Fall ist.

Dass die Varianz in (9.6) geringer ist als in (9.5), kann man an Hand der bereits in Abschnitt 5.3 verwendeten Formel

$$\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y) + 2\text{Cov}(X, Y) \quad (9.7)$$

sehen. Dazu schreiben wir (9.5) als

$$\Delta(t, S) \approx \frac{1}{n\Delta S} \sum_{i=1}^n (V(t, S + \Delta S, \omega_i) - V(t, S, \tilde{\omega}_i)), \quad (9.8)$$

wobei $V(t, \cdot, \omega)$ den numerisch ermittelten Optionswert entlang des durch ω parametrisierten Pfades bezeichnet. Die verschiedenen Ereignisse ω_i und $\tilde{\omega}_i$ deuten dabei an, dass der erste Ausdruck der Differenz unabhängig von dem zweiten Ausdruck der Differenz berechnet wird, d.h. für unabhängig voneinander simulierte Pfade des Wiener-Prozesses. Die Methode (9.6) schreiben wir als

$$\Delta(t, S) \approx \frac{1}{n\Delta S} \sum_{i=1}^n (V(t, S, \omega_i) - V(t, S, \omega_i)). \quad (9.9)$$

Hier taucht nun in beiden Ausdrücken das gleiche ω_i auf, weil beide Ausdrücke für den gleichen Pfad des Wiener-Prozesses berechnet werden.

Im Fall (9.8) setzen wir nun $X = V(t, S + \Delta S, \cdot)$ und $Y = -\bar{V}_{Mc}(t, S, \cdot)$. Da $\bar{V}_{Mc}(t, S + \Delta S, \cdot)$ und $\bar{V}_{Mc}(t, S, \cdot)$ unabhängig voneinander simuliert werden, sind X und Y unabhängig und es folgt $\text{Cov}(X, Y) = 0$ und damit

$$\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y).$$

Im Fall (9.9) setzen wir $X = V_{Mc}(t, S + \Delta S, \cdot)$ und $Y = -V_{Mc}(t, S, \cdot)$. Beachte, dass $\text{Var}(X)$ und $\text{Var}(Y)$ die gleichen Werte wie im Fall (9.8) annehmen, da es sich um Zufallsvariablen mit identischer Verteilung handelt. Im Gegensatz zu (9.9) beruhen diese Zufallsvariablen für gleiches nun aber auf dem jeweils gleichen Pfad des Wiener-Prozesses, weswegen (da ΔS ja klein ist und die Ausdrücke stetig in S sind) $V_{Mc}(t, S + \Delta S, \omega_i) \approx V_{Mc}(t, S, \omega_i)$ und damit

$$X(\omega_i) \approx -Y(\omega_i)$$

gilt. Daraus folgt $\text{Cov}(X, Y) < 0$ und damit

$$\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y) + 2\text{Cov}(X, Y) < \text{Var}(X) + \text{Var}(Y),$$

d.h. die Varianz im Fall (9.9) ist geringer und die Monte-Carlo-Simulation sollte daher bessere Werte liefern. Dass diese nicht nur ein bisschen sondern deutlich besser sind, zeigt die Simulation in Abbildung 9.5 (die Daten stammen aus dem Seminarvortrag von Thomas Höllbacher aus dem Sommersemester 2010).

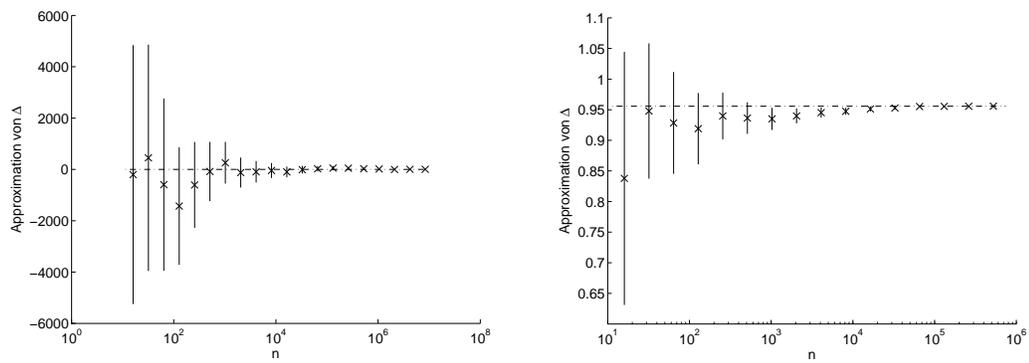


Abbildung 9.5: Vergleich der Monte-Carlo-Ergebnisse für (9.5) (links) und (9.6) (rechts). Die Balken zeigen die (aus den empirisch berechneten Varianzen ermittelten) 95%-Konfidenzintervalle. Beachte die unterschiedlichen Skalen in beiden Grafiken.

Kapitel 10

Schätzung der Volatilität

Um den Wert einer Option und/oder die zugehörige Hedging-Strategie für reale Wertpapiere mit den bisher vorgestellten Methoden zu berechnen, muss man die Parameter in den Kursmodellen bestimmen. Je nach dem, welches Modell man verwendet, sind dies Parameter unterschiedlicher Anzahl und Art. Wir werden dies Problem hier beispielhaft für das Black-Scholes-Modell betrachten, in dem das zu Grunde liegende Kursmodell die geometrische Brownsche Bewegung. Das Modell ist gegeben durch

$$dS(t) = \mu S(t)dt + \sigma S(t)dW_t,$$

vgl. (4.10) mit den Lösungen

$$S(t; S_0) = S_0 \exp\left(\left(\mu - \frac{1}{2}\sigma^2\right)t + \sigma W(t)\right),$$

siehe (4.11). Erwartungswert und Varianz der Lösung sind gegeben durch

$$E(S(t, x_0)) = S_0 e^{\mu t} \quad \text{und} \quad \text{Var}(S(t)) = S_0^2 e^{2\mu t} (e^{\sigma^2 t} - 1).$$

Da wir in der Optionsbewertung $\mu = r$ setzen und der risikofreie Zinssatz r bekannt ist, muss für dieses Modell tatsächlich nur ein Parameter bestimmt werden, nämlich die Volatilität σ . Konzeptionell lassen sich aber zumindest einige der Methoden aus diesem Kapitel auch auf anspruchsvollere Kursmodelle übertragen, in denen mehrere Parameter bestimmt werden müssen.

10.1 Historische Volatilität

Die Methode der historischen Volatilität entspricht dem statistischen Ansatz, den man z.B. auch in den Naturwissenschaften machen würde, wenn man Parameter aus Messwerten schätzen möchte. Die “Messwerte” sind hier die Aktienkurse $S(t_i)$, die man zu gegebenen Zeiten t_i aus den Kurstabellen der Börse ablesen kann. Hierbei nehmen wir an, dass die Daten zu äquidistanten Zeitpunkten vorliegen, d.h. dass $t_i - t_{i-1} = \Delta t$ für alle i gilt. Die Idee ist, dass man die Volatilität σ aus dem vergangenen Kursverlauf, d.h. aus historischen Daten schätzt — natürlich in der Hoffnung, dass sich σ in der Zukunft nicht sehr stark

ändert. In wie weit diese Hoffnung berechtigt ist, hängt natürlich stark vom betrachteten Kurs, dem Marktumfeld, den ökonomischen Rahmbedingungen etc. ab und ist ohne Zweifel ein wesentlicher Unsicherheitsfaktor der Methode.

Die Grundannahme im Black-Scholes-Modell ist ja bekanntlich, dass sich diese Kurse gemäß der geometrischen Brownschen Bewegung verhalten. Dies bedeutet, dass

$$S(t_i) = S_0 \exp \left(\left(\mu - \frac{1}{2} \sigma^2 \right) t_i + \sigma W(t_i) \right)$$

gilt. Interessant für unsere Zwecke ist nun — wie bereits mehrfach erwähnt — die Volatilität $\sigma > 0$. Um diese zu schätzen, definieren wir die neuen Zufallsvariablen

$$U_i := \log \frac{S(t_i)}{S(t_{i-1})}.$$

Setzen wir $\Delta t := t_i - t_{i-1}$ (aus der obigen Annahme an die t_i folgt, dass Δt tatsächlich unabhängig von i ist) und $\Delta W_i = W(t_i) - W(t_{i-1})$ und nehmen wir an, dass sich die $S(t_i)$ gemäß der geometrischen Brownschen Bewegung verhalten, so gilt wegen

$$U_i = \log \frac{S(t_i)}{S(t_{i-1})} = \log S(t_i) - \log S(t_{i-1}) \quad (10.1)$$

die Gleichung

$$U_i = \left(\mu - \frac{1}{2} \sigma^2 \right) t_i + \sigma W(t_i) - \left(\mu - \frac{1}{2} \sigma^2 \right) t_{i-1} - \sigma W(t_{i-1}) = \left(\mu - \frac{1}{2} \sigma^2 \right) \Delta t + \sigma \Delta W_i.$$

Da ΔW_i als Inkrement des Wiener-Prozesses normalverteilt mit Erwartungswert 0 und Varianz Δt ist, kann U_i folglich als Realisierung einer normalverteilten Zufallsvariablen U mit Erwartungswert $\mathbb{E}[U] = (\mu - \sigma^2/2)\Delta t$ und Varianz $\text{Var}(U) = \mathbb{E}[(U_i - \mathbb{E}(U_i))^2] = \sigma^2 \Delta t$, d.h. $U \sim N((\mu - \sigma^2/2)\Delta t, \sigma^2 \Delta t)$.

Die Werte U_i können also als Zufallszahlen gemäß dieser Verteilung aufgefasst werden; folglich können $\mathbb{E}[U]$ und $\text{Var}(U)$ mit der Monte-Carlo-Methode geschätzt werden. Dazu nehmen wir an, dass $t = t_n$ die aktuelle Zeit ist und wir die $M + 1$ letzten Werte $S(t_{n-M}), \dots, S(t_n)$ kennen.

Dann können wir

$$\begin{aligned} \mu_M &:= \frac{1}{M} \sum_{i=1}^M U_{n+1-i} \\ \sigma_M^2 &:= \frac{1}{M-1} \sum_{i=1}^M (U_{n+1-i} - \mu_M)^2 \end{aligned}$$

setzen und erhalten so Schätzwerte $\mu_M \approx \mathbb{E}[U_i]$ und $\sigma_M^2 \approx \text{Var}(U_i)$. Auf den ersten Blick erscheint der Nenner “ $M - 1$ ” in der Formel für σ_M^2 ungewöhnlich, denn auch hier summieren wir ja M Werte auf und müssten demnach nach dem üblichen Monte-Carlo-Prinzip durch M teilen. Der Grund für die Wahl “ $M - 1$ ” liegt darin, dass es erstrebenswert ist, dass σ_M^2 ein sogenannter *erwartungstreuer Schätzer* für $\text{Var}(U_i)$ ist. Dies bedeutet, dass

$\mathbb{E}[\sigma_M^2] = \text{Var}(U_i)$ gilt, d.h. dass bei wiederholter Ausführung der Schätzung (mit identisch verteilten Zufallszahlen U_i) im Mittel gerade der korrekte Wert herauskommt. Dies ist aber gerade dann der Fall, wenn der Nenner als $M - 1$ und nicht als M gewählt wird (Details siehe Übungsaufgabe).

Wegen $\text{Var}(U_i) = \sigma^2 \Delta t$ ist nun

$$\sigma^* = \frac{\sigma_M}{\sqrt{\Delta t}}$$

eine Schätzung für die Volatilität. Eine kleine Rechnung zeigt, dass dies auch direkt in Abhängigkeit von den U_i als

$$\sigma^* = \sqrt{\frac{1}{\Delta t} \left(\frac{1}{M-1} \sum_{i=1}^M U_{i+1-i}^2 - \frac{1}{M(M-1)} \left(\sum_{i=1}^M U_{n+1-i} \right)^2 \right)} \quad (10.2)$$

geschrieben werden kann (Übungsaufgabe).

Wie genau ist nun diese Schätzung? Um einen Eindruck davon zu bekommen, betrachten wir ein vereinfachtes Problem und untersuchen eine Monte-Carlo-Schätzung für die Varianz $\text{Var}(U)$. Dabei vernachlässigen wir in unserer Analyse die Tatsache, dass in der Formel für σ_M^2 der geschätzte Erwartungswert μ_M und nicht der exakte Erwartungswert $\mathbb{E}[U_i]$ verwendet wird. Ebenso verwenden wir den üblichen Nenner M (statt des erwartungstreuen $M - 1$); dieser Unterschied ist für große M aber vernachlässigbar. Wir ersetzen die Formel für σ_M^2 also durch die Monte-Carlo-Approximation

$$\text{Var}(U) \approx \frac{1}{M} \sum_{i=1}^M (U_i - \mathbb{E}[U])^2 =: \tilde{\sigma}_M^2$$

mit $U \sim N((\mu - \sigma^2/2)\Delta t, \sigma^2 \Delta t)$ und untersuchen den Fehler der Schätzung $\sigma^* = \tilde{\sigma}_M / \sqrt{\Delta t}$. Indem wir $\hat{U} = U / \sqrt{\Delta t}$ und $\hat{U}_i = U_i / \sqrt{\Delta t}$ setzen, erhalten wir

$$(\sigma^*)^2 = \frac{1}{M} \sum_{i=1}^M (\hat{U}_i - \mathbb{E}[\hat{U}])^2$$

mit $\hat{U} \sim N((\mu - \sigma^2/2)\sqrt{\Delta t}, \sigma^2)$. Aus Satz 5.3 folgt nun, dass das 95%-Konfidenzintervall der Methode approximativ durch

$$\left[(\sigma^*)^2 - \frac{1.96v}{\sqrt{M}}, (\sigma^*)^2 + \frac{1.96v}{\sqrt{M}} \right]$$

gegeben ist, wobei v^2 die Varianz der Zufallsvariablen $(\hat{U} - \mathbb{E}[\hat{U}])^2$ ist. Schreiben wir $Y = \hat{U} - \mathbb{E}[\hat{U}]$ so gilt $Y \sim N(0, \sigma^2)$ und damit

$$\text{Var}((\hat{U} - \mathbb{E}[\hat{U}])^2) = \text{Var}(Y^2) = \mathbb{E}[Y^4] - \mathbb{E}[Y^2]^2 = 3\sigma^4 - \sigma^4 = 2\sigma^4.$$

Das approximative 95%-Konfidenzintervall ergibt sich damit zu

$$\left[(\sigma^*)^2 - \frac{1.96\sqrt{2}\sigma^2}{\sqrt{M}}, (\sigma^*)^2 + \frac{1.96\sqrt{2}\sigma^2}{\sqrt{M}} \right].$$

Für hinreichend große M gilt unter Verwendung der Taylor-Approximation $\sqrt{1 \pm \varepsilon} \approx 1 \pm \varepsilon/2$ dann

$$\sqrt{(\sigma^*)^2 \pm \frac{1.96\sqrt{2}\sigma^2}{\sqrt{M}}} \approx \sqrt{(\sigma^*)^2 \pm \frac{1.96\sqrt{2}(\sigma^*)^2}{\sqrt{M}}} = \sigma^* \sqrt{1 \pm \frac{1.96\sqrt{2}}{\sqrt{M}}} \approx \sigma^* \pm \frac{1.96\sigma^*}{\sqrt{2M}},$$

womit wir das approximative (und numerisch aus den Daten berechenbare) 95%-Konfidenzintervall

$$\left[\sigma^* - \frac{1.96\sigma^*}{\sqrt{2M}}, \sigma^* + \frac{1.96\sigma^*}{\sqrt{2M}} \right]$$

für σ^* erhalten. Wie bei Monte-Carlo-Approximationen üblich, ist der Fehler also von der Ordnung $O(1/\sqrt{M})$.

Im Vergleich zu den bisher betrachteten Monte-Carlo-Simulationen mit Zufallszahlen gibt es bei den hier verwendeten realen Daten U_i aber einen grundlegenden Unterschied: Wenn wir die Genauigkeit der Approximation durch Wahl eines größeren M erhöhen wollen, können wir nicht einfach mehr Zufallszahlen ziehen sondern benötigen mehr Daten! Hierzu gibt es im Prinzip zwei Möglichkeiten:

- (1) Halte Δt fest und verwende weiter zurückliegende Daten
- (2) Halte den Zeitraum der Datenerhebung $M\Delta t$ fest und verringere Δt , d.h. den zeitlichen Abstand zwischen zwei Daten

Beide Ansätze sind nicht wirklich zufriedenstellend: Methode (1) führt zur Verwendung sehr alter Daten, die in der Regel für die Bestimmung der zukünftigen Volatilität nur recht begrenzte Aussagekraft haben. Methode (2) stößt schnell an praktische Grenzen, weil Kurswerte mit sehr geringem zeitlichen Abstand (z.B. wegen Verzögerungen in der Preisbildung an der Börse) aus statistischer Sicht keine brauchbaren Informationen liefern. Eine gebräuchliche Faustregel in der Optionsbewertung ist, historische Daten aus einem Zeitraum der Länge der Laufzeit der Option zu wählen, es gibt aber auch Untersuchungen, die nahelegen, dass die Verwendung älterer Daten zu besseren Ergebnissen führen kann. Insgesamt benötigt die Wahl des richtigen Zeitraums und der richtigen Anzahl der Daten viel praktische Erfahrung und lässt sich nicht allein mit rigorosen mathematischen Methoden begründen.

In der Praxis wird die oben angegebene Formel für σ^* noch weiter vereinfacht. Grundlage dafür ist die folgende Umstellung der oben angegebenen Formel für μ_M , in der wir die Identität (10.1) verwenden:

$$\begin{aligned} \mu_M &= \frac{1}{M} \sum_{i=1}^M (\log S(t_{n+1-i}) - \log S(t_{n-i})) \\ &= \frac{1}{M} \log S(t_n) - \log S(t_{n-M}) = \log \frac{S(t_n)}{S(t_{n-M})}. \end{aligned}$$

Der Mittelwert μ_M hängt also nur vom ersten und letzten Kurswert ab. Wegen

$$\log S(t_n) - \log S(t_{n-M}) = \left(\mu - \frac{1}{2}\sigma^2 \right) \Delta t M + \sigma(W(t_n) - W(t_{n-M}))$$

folgt $\mu_M \sim N((\mu - \sigma^2/2)\Delta t, \sigma^2\Delta t/M)$. Der Mittelwert μ_M ist also eine normalverteilte Zufallsvariable mit kleinem Mittelwert und kleiner Varianz und wird daher in der Praxis üblicherweise auf 0 gesetzt. Damit vereinfacht sich (10.2) zu

$$\sigma^* = \sqrt{\frac{1}{\Delta t} \frac{1}{M-1} \sum_{i=1}^M U_{n+1-i}^2}. \quad (10.3)$$

Dies ist die Formel, die in der Praxis oft verwendet wird und die auch in der Programmieraufgabe auf dem aktuellen Übungsblatt verwendet werden soll.

Zum Abschluss betrachten wir noch eine Variante von (10.3). Wir hatten oben bereits besprochen, dass die Verwendung von weit zurück liegenden Daten problematisch sein kann, weil diese u.U. nicht viel über die aktuelle Volatilität aussagen. Andererseits würden wir aber gerne viele Daten verwenden, um die Zuverlässigkeit der Schätzung zu erhöhen. Ein Kompromiss, mit dem man beide Aspekte berücksichtigen kann, ist die Verwendung exponentiell gewichteter Monte-Carlo-Schätzer, kurz EWMA von “exponentially weighted moving averages” genannt. Diese Variante kann für den vereinfachten erwartungstreuen Schätzer (10.3) oder für die “echte” Monte-Carlo-Approximation, d.h. für (10.3) mit Nenner M an Stelle von $M-1$ angewendet werden; wir betrachten hier die zweite Variante, weil diese technisch etwas einfacher aufzuschreiben ist. Für diese Variante kann (10.3) umgeschrieben werden zu

$$\Delta t(\sigma^*)^2 = \frac{1}{M} \sum_{i=1}^M U_{n+1-i}^2,$$

so dass $\Delta t(\sigma^*)^2$ als Approximation des Erwartungswertes $\mathbb{E}[U^2]$ aufgefasst werden kann. In dieser Formel kann man nun Gewichte $\alpha_i > 0$ mit $\sum_{i=1}^M \alpha_i = 1$ einführen und sie zu

$$\Delta t(\sigma^*)^2 = \sum_{i=1}^M \alpha_i U_{n+1-i}^2$$

verallgemeinern. Typischerweise wählt man $\alpha_1 > \alpha_2 > \dots > \alpha_M > 0$ um aktuellere Daten stärker zu gewichten. Oft werden die Gewichte dabei rekursiv mittels $\alpha_{i+1} = \omega \alpha_i$ für ein $\omega \in (0, 1)$ gewählt. Einsetzen dieser Gleichung liefert

$$\Delta t(\sigma^*)^2 = \frac{\sum_{i=1}^M \omega^i U_{n+1-i}^2}{\sum_{i=1}^M \omega^i},$$

wobei die Skalierung im Nenner die Summenbedingung $\sum_{i=1}^M \alpha_i = 1$ sicher stellt. Ein typischer Wert hier ist $\omega = 0.94$. Bei dieser rekursiven Wahl der α_i kann auch die Volatilität rekursiv berechnet werden. Sei dazu $\Delta t(\sigma_n^*)^2$ der Schätzwert zur Zeit t_n , der aus den Daten U_{n+1-M}, \dots, U_n berechnet wurde. Zur Zeit t_{n+1} ist die neue Schätzung $\Delta t(\sigma_{n+1}^*)^2$ zur Zeit t_{n+1} basierend auf den Daten $U_{n+1-M}, \dots, U_{n+1}$ dann gegeben durch

$$\Delta t(\sigma_{n+1}^*)^2 = \omega \Delta t(\sigma_n^*)^2 + (1 - \omega) U_{n+1}^2.$$

Die Schätzungen können auf diese Weise fortlaufend aktualisiert werden, ohne dass stets die komplette Summe neu berechnet werden muss.

10.2 Implizite Volatilität

Das zweite Verfahren zur Schätzung von σ , das wir betrachten wollen, verwendet eine vollständig andere Herangehensweise. Bei der Methode der impliziten Volatilität nutzt man aus, dass Optionen an der Börse gehandelt werden und sich durch Angebot und Nachfrage ein Marktpreis $V_M(t, S)$ der Option einstellt. Im Gegensatz zur historischen Volatilität, die rein auf der Vergangenheit der Kurse beruht, spiegelt der Marktpreis der Option auch die Erwartungen der Marktteilnehmer über die zukünftige Kursentwicklung wider. Diese Erwartungen müssen natürlich nicht mit der dann tatsächlich eintretenden Kursentwicklung übereinstimmen; ist man aber gewillt anzunehmen, dass die Marktteilnehmer ihre Einschätzungen auf Basis aller verfügbaren Informationen und mit genügend ökonomischem Sachverstand treffen, so besteht zumindest die Hoffnung, dass die Marktpreise der Optionen realistische Annahmen über die zukünftige Kursentwicklung widerspiegeln.

Die Idee der impliziten Volatilität ist nun, die Volatilität σ so zu bestimmen, dass der Black-Scholes-Preis $V(t, S)$ für diese Wahl der Volatilität gerade mit dem Marktpreis $V_M(t, S)$ übereinstimmt. Im Gegensatz zu dem Ansatz der historischen Volatilität erfordert dieser Ansatz keinerlei statistische Methoden; statt dessen stoßen wir auf das klassische Problem der numerischen Lösung einer nichtlinearen Gleichung. Insbesondere werden wir dabei sehen, dass wir hier über die aus der Einführung in die Numerik bekannte Konvergenzanalyse hinaus aus der speziellen Struktur des Problems schärfere Aussagen ableiten können.

Zur besseren Notation nehmen wir die Volatilität als weiteren Parameter in den Black-Scholes-Optionspreis auf und schreiben diesen als $V(t, S, \sigma)$. Gesucht ist dann also $\sigma > 0$, so dass die Gleichung

$$V(t, S, \sigma) = V_M(t, S)$$

gilt. Hierbei müssen die sonstigen Parameter der Optionen (wie z.B. Ausübungszeit T und Ausübungspreis K) natürlich übereinstimmen. Definieren wir die Funktion $f : \mathbb{R}_0^+ \rightarrow \mathbb{R}$

$$f(\sigma) := V(t, S, \sigma) - V_M(t, S), \quad (10.4)$$

so ist die Berechnung der impliziten Volatilität σ^* äquivalent zur Berechnung der Nullstelle $f(\sigma^*) = 0$.

Für dieses Problem haben wir in der Einführung in die Numerik verschiedene Algorithmen kennen gelernt. Eine Möglichkeit ist das Bisektionsverfahren, das den Vorteil besitzt, dass es stets konvergiert, wenn die Funktion f stetig ist (was hier der Fall ist) und wenn wir Anfangswerte σ_a, σ_b finden können, so dass $f(\sigma_a)$ und $f(\sigma_b)$ unterschiedliches Vorzeichen besitzen. Nachteil dieses Verfahrens ist, dass es relativ langsam konvergiert; in jedem Iterationsschritt halbiert sich der Fehler gerade, d.h. das Verfahren ist linear konvergent.

Viel schneller ist das Newton-Verfahren, bei dem ausgehend von einem Startwert σ_0 iterativ neue Werte σ_{i+1} gemäß der Iteration

$$\sigma_{i+1} = \sigma_i - \frac{f(\sigma_i)}{f'(\sigma_i)}$$

erzeugt werden. Diese Iteration wird dann so lange durchgeführt, bis $|\sigma_{i+1} - \sigma_i| < tol$ für eine vorgegebene Toleranz tol ist. Der Konvergenzsatz für dieses Verfahren besagt nun das Folgende: Wenn f zwei mal stetig differenzierbar ist, die Ableitung $f'(\sigma) \neq 0$ ist in einer

Umgebung von σ^* und der Startwert σ_0 hinreichend nahe an der Nullstelle σ^* liegt, so konvergiert die Folge σ_i quadratisch gegen σ^* .

Die quadratische Konvergenz des Newton-Verfahrens ist tatsächlich viel schneller als die lineare Konvergenz des Bisektionsverfahrens. Während die Anzahl der korrekten Nachkommastellen der Näherungen σ_i bei der linearen Konvergenz proportional zur Anzahl der Iterationen i ist, verdoppelt sich die Anzahl der korrekten Nachkommastellen bei der quadratischen Konvergenz in jedem Schritt. Man überlegt sich leicht (z.B. durch Ausprobieren an einem Beispiel), dass dies sehr viel schneller ist.

Aus Sicht der Konvergenzgeschwindigkeit ist das Newton-Verfahren also deutlich vorzuziehen, es besitzt aber auch Nachteile. Die Funktion f muss zwei mal stetig differenzierbar sein, die Ableitung f' muss bekannt sein und darf nicht gleich Null sein und es muss eine gute Startschätzung σ_0 bekannt sein, für die man Konvergenz des Verfahrens sicher stellen kann. Wir werden diese Aspekte im Folgenden — beispielhaft für den europäischen Call — im Detail untersuchen. Dazu erinnern wir an die Formeln

$$V(t, S, \sigma) = SP(a) - Ke^{-r(T-t)}P(b)$$

mit

$$a = \frac{\log \frac{S}{K} + \left(r + \frac{\sigma^2}{2}\right)(T-t)}{\sigma\sqrt{T-t}}, \quad b = a - \sigma\sqrt{T-t} = \frac{\log \frac{S}{K} + \left(r - \frac{\sigma^2}{2}\right)(T-t)}{\sigma\sqrt{T-t}}$$

und

$$P(y) = \int_{-\infty}^y \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right) dx,$$

vgl. Kapitel 8.

Da $V_M(t, S)$ eine Konstante ist, ist die Ableitung $f'(\sigma)$ von f aus (10.4) gerade

$$f'(\sigma) = \frac{\partial V}{\partial \sigma}(t, S, \sigma) = \text{vega} = S\sqrt{T-t}P'(a),$$

vgl. Abschnitt 9.2. Wie bereits dort beobachtet, ist $f'(\sigma) = \text{vega}$ für $\sigma > 0$ positiv, also insbesondere ungleich Null. Da vega zudem stetig differenzierbar nach σ ist, und $f''(\sigma) = \partial \text{vega} / \partial \sigma$ ist, ist f zwei mal stetig differenzierbar. Falls also eine positive Lösung σ^* des Nullstellenproblems $f(\sigma^*) = 0$ existiert, ist das Newton-Verfahren anwendbar und die quadratische Konvergenz des Verfahrens für einen geeigneten Startwert σ_0 sicher gestellt.

Die schwierigere Frage ist nun die Bestimmung eines geeigneten Startwerts, für den man sicher stellen kann, dass das Verfahren auch gegen σ^* konvergiert. Um diese Frage zu beantworten, brauchen wir noch einige Vorüberlegungen.

Zunächst untersuchen wir, für welche Werte von $V_M(t, S)$ eine Lösung σ^* existiert und zeigen, dass diese dann eindeutig ist. Dazu betrachten wir zunächst, wie der Wert $V(t, S, \sigma)$ asymptotisch von σ abhängt.

Für $\sigma \rightarrow \infty$ divergiert a gegen unendlich, weswegen $P(a) \rightarrow 1$ gilt. Da der von σ abhängige Summand in a gerade (nach Kürzen) gleich $\sigma\sqrt{T-t}/2$ ist, konvergiert $b = a - \sigma\sqrt{T-t}$ gegen 0 für $\sigma \rightarrow \infty$, also $P(b) \rightarrow 0$. Damit ergibt sich aus der Formel für $V(t, S, \sigma)$ die Konvergenz

$$\lim_{\sigma \rightarrow \infty} V(t, S, \sigma) = S.$$

Für $\sigma \rightarrow 0$ mit $\sigma > 0$ (was wir im Folgenden kurz als $\sigma \searrow 0$ schreiben) unterscheiden wir drei verschiedene Fälle:

- 1) $S - Ke^{-r(T-t)} > 0$: In diesem Fall gilt $\log S/K + r(T-t) > 0$ woraus $a \rightarrow \infty$ und $b \rightarrow \infty$ für $\sigma \searrow 0$ und damit $V(t, S, \sigma) \rightarrow S - Ke^{-r(T-t)}$ folgt
- 2) $S - Ke^{-r(T-t)} < 0$: Hier gilt $\log S/K + r(T-t) < 0$ und damit $a \rightarrow -\infty$ und $b \rightarrow -\infty$. Daraus folgt $V(t, S, \sigma) \rightarrow 0$.
- 3) $S - Ke^{-r(T-t)} = 0$: In diesem Fall erhalten wir $\log S/K + r(T-t) = 0$ und somit $a \rightarrow 0$ und $b \rightarrow 0$. Daraus folgt $V(t, S, \sigma) \rightarrow (S - Ke^{-r(T-t)})/2 = 0$.

Zusammenfassend folgt aus den drei Fällen die Konvergenz

$$\lim_{\sigma \searrow 0} V(t, S, \sigma) = (S - Ke^{-r(T-t)})^+.$$

Da $vega = \partial V / \partial \sigma$ positiv ist, ist $V(t, S, \sigma)$ in σ streng monoton wachsend. Aus den gerade berechneten Grenzwerten folgt damit

$$V(t, S, \sigma) \in \left((S - Ke^{-r(T-t)})^+, S \right).$$

Diese Rechnungen zeigen, dass es für jeden Marktpreis $V_M(t, S)$ aus diesem Intervall genau ein $\sigma^* > 0$ mit $f(\sigma^*) = 0$ gibt.

Zur Analyse des Newton-Verfahrens benötigen wir weitere Informationen über die zweite Ableitung $f''(\sigma)$ (warum, wird in Kürze klar werden). Für diese gilt

$$f''(\sigma) = \frac{\partial veega}{\partial \sigma} = -\frac{S\sqrt{T-t}}{\sqrt{2\pi}} e^{-a^2/2} a \frac{\partial a}{\partial \sigma}.$$

Aus den Formeln für a und b folgt

$$\begin{aligned} \frac{\partial a}{\partial \sigma} &= -\frac{\log(S/K) + r(T-t)}{\sigma^2 \sqrt{T-t}} + \frac{1}{2} \sqrt{T-t} \\ &= -\left[\frac{\log(S/K) + (r - (\sigma^2/2))(T-t)}{\sigma^2 \sqrt{T-t}} \right] = -\frac{b}{\sigma}. \end{aligned}$$

Einsetzen in die Formel für f'' liefert dann

$$f''(\sigma) = \frac{S\sqrt{T-t}}{\sqrt{2\pi}} e^{-a^2/2} \frac{ab}{\sigma} = S\sqrt{T-t} P'(a) \frac{ab}{\sigma} = \frac{ab}{\sigma} f'(\sigma). \quad (10.5)$$

Da der Integrand von P maximal wird für $x = 0$, wird $f'(\sigma) = veega$ genau dann maximal in σ , wenn $a = 0$ gilt. Auflösen der Gleichung $a = 0$ nach σ zeigt, dass dies gerade für

$$\hat{\sigma} = \sqrt{2 \left| \frac{\log S/K + r(T-t)}{T-t} \right|} \quad (10.6)$$

der Fall ist. Unter Ausnutzung der Tatsache, dass das Einsetzen von $\hat{\sigma}$ in a gerade $a = 0$ liefert, zeigt man dann, dass aus (10.5) die Gleichung

$$f''(\sigma) = \frac{T-t}{4\sigma^3}(\hat{\sigma}^4 - \sigma^4)f'(\sigma) \quad (10.7)$$

folgt (Übungsaufgabe). Wegen $f'(\sigma) > 0$ und $(T-t)/4\sigma^3 > 0$ folgt daher

$$f''(\sigma) \begin{cases} > 0, & \sigma \in (0, \hat{\sigma}) \\ < 0, & \sigma \in (\hat{\sigma}, \infty). \end{cases}$$

Dies bedeutet, dass f konvex ist für $\sigma \in (0, \hat{\sigma})$ und konkav für $\sigma \in (\hat{\sigma}, \infty)$.

Was bedeutet dies nun für das Newton-Verfahren? Dazu analysieren wir die Entwicklung des Fehlers $\sigma_i - \sigma^*$ des Newton-Verfahrens mit Hilfe des Mittelwertsatzes der Differentialrechnung:

$$\sigma_{i+1} - \sigma^* = \sigma_i - \sigma^* - \frac{f(\sigma_i) - f(\sigma^*)}{f'(\sigma_i)} = \sigma_i - \sigma^* - \frac{(\sigma_i - \sigma^*)f'(\xi_i)}{f'(\sigma_i)} \quad (10.8)$$

mit $\xi_i \in (\sigma_i, \sigma^*)$ oder $\xi_i \in (\sigma^*, \sigma_i)$, wobei wir im ersten Schritt $f(\sigma^*) = 0$ ausgenutzt haben. Daraus folgt

$$\frac{\sigma_{i+1} - \sigma^*}{\sigma_i - \sigma^*} = 1 - \frac{f'(\xi_i)}{f'(\sigma_i)}. \quad (10.9)$$

Wir wählen nun für das Newton-Verfahren den Startwert $\sigma_0 = \hat{\sigma}$ aus (10.6) und beweisen per Induktion über i für alle i mit $f(\sigma_i) \neq 0$ und $f(\sigma_{i+1}) \neq 0$ (andernfalls hätten wir die exakte Nullstelle gefunden und würden das Verfahren abbrechen) die Ungleichungen

$$(a) \sigma_i > 0, \quad (b) 0 < f'(\xi_i) < f'(\sigma_i), \quad (c) 0 < \frac{\sigma_{i+1} - \sigma^*}{\sigma_i - \sigma^*} < 1, \quad (d) 0 < \frac{\sigma_i - \sigma^*}{\sigma_0 - \sigma^*} \leq 1.$$

Beachte dazu, dass aus der Bedingung $f(\sigma_{i+1}) \neq 0$ die Ungleichung $f'(\xi_i) \neq f'(\sigma_i)$ folgt, denn aus $f'(\xi_i) = f'(\sigma_i)$ folgt mit (10.8)

$$\sigma_{i+1} - \sigma^* = \sigma_i - \sigma^* - \frac{(\sigma_i - \sigma^*)f'(\xi_i)}{f'(\sigma_i)} = 0 \Rightarrow \sigma_{i+1} - \sigma^* = 0 \Rightarrow f(\sigma_{i+1}) = 0.$$

Zudem folgt wegen der Eindeutigkeit der Nullstelle $\sigma_i \neq \sigma^*$ und $\sigma_{i+1} \neq \sigma^*$.

Induktionsanfang $i = 0$: Ungleichung (a) folgt sofort wegen $\sigma_0 = \hat{\sigma} > 0$. Aus der Tatsache, dass $\hat{\sigma}$ die Funktion f' maximiert und der Ungleichung $f(\xi_i) \neq f'(\sigma_i)$ folgt Ungleichung (b). Beachte, dass ξ_0 zwischen σ^* und σ_0 liegt und damit positiv ist; daher ist $f'(\xi_0)$ definiert und positiv. Daraus und aus Ungleichung (b) folgt Ungleichung (c) wegen (10.9). Ungleichung (d) folgt schließlich, weil für $i = 0$ Zähler und Nenner identisch sind.

Induktionsschritt $i \rightarrow i + 1$: als Induktionsannahme nehmen wir an, dass (a)–(d) für i gilt. Wir betrachten zunächst den Fall $\sigma_0 < \sigma^*$. In diesem Fall folgt aus (d) die Ungleichung $\sigma_0 \leq \sigma_i \leq \sigma^*$ und aus (c) die Ungleichung $\sigma_i < \sigma_{i+1} < \sigma^*$. Damit folgt aus Ungleichung (a) für i die Ungleichung (a) für $i + 1$, zudem folgt $\sigma_{i+1} > \sigma_0$. Da $f''(\sigma) < 0$ ist für $\sigma \in (\sigma_0, \infty)$ ist f' streng monoton fallend in diesem Intervall. Da ξ_{i+1} echt zwischen σ_{i+1} und σ^* liegt, gilt also $\xi_{i+1} > \sigma_{i+1} > \sigma_0$ und damit $f'(\xi_{i+1}) > 0$ und $f'(\xi_{i+1}) < f'(\sigma_{i+1})$, also (b) für $i + 1$.

Ungleichung (c) für $i+1$ folgt damit aus (10.9). Ungleichung (d) für $i+1$ folgt schließlich aus der Kombination von (c) und (d) für i . Im Fall $\sigma_0 > \sigma^*$ verläuft die Argumentation genau so mit umgekehrten Ungleichungen; hier nutzt man aus, dass $f''(\sigma) > 0$ ist für $\sigma \in (0, \sigma_0)$.

Ungleichung (c) zeigt, dass der Fehler $\sigma_i - \sigma^*$ entweder für alle i negativ ist und monoton wächst (falls $\sigma_0 < \sigma^*$) oder für alle i positiv ist und monoton fällt (falls $\sigma_0 > \sigma^*$). Damit konvergiert $\sigma_i - \sigma^*$ gegen einen Grenzwert s und σ_i konvergiert gegen einen Grenzwert $\tilde{\sigma} > 0$. Für $i \rightarrow \infty$ folgt daher aus (10.8) und der Stetigkeit von f und f'

$$\frac{f(\tilde{\sigma}) - f(\sigma^*)}{f'(\tilde{\sigma})} = \lim_{i \rightarrow \infty} \frac{f(\sigma_i) - f(\sigma^*)}{f'(\sigma_i)} = \lim_{i \rightarrow \infty} (\sigma_{i+1} - \sigma^*) - (\sigma_i - \sigma^*) = s - s = 0.$$

Daraus folgt aber $f(\tilde{\sigma}) = 0$ und wegen der Eindeutigkeit der Nullstelle also $\tilde{\sigma} = \sigma^*$. Für den Startwert $\sigma_0 = \hat{\sigma}$ aus (10.6) konvergiert das Newton-Verfahren also stets gegen den gesuchten Wert und da wir bereits gesehen haben, dass die Bedingungen für lokal quadratische Konvergenz erfüllt sind, erhalten wir diese sobald sich die Folge σ_i in der Umgebung von σ^* befindet, in der die lokale quadratische Konvergenz gilt.

Kapitel 11

Einführung in die Finite Differenzen-Methode

Während die Black-Scholes Gleichung für europäische Optionen analytisch gelöst werden kann, ist dies für viele weitere Optionsarten nicht möglich.

Wenn andere Optionen als europäische Calls oder Puts betrachtet werden, kann es sein, dass die zugehörige Black-Scholes-Gleichung entsprechend verändert werden muss. Dies ist z.B. der Fall für die bereits erwähnten asiatischen Optionen oder für sogenannte Basket-Optionen, in denen die Option von mehreren Basiswerten abhängt, vgl. Günther und Jüngel, Abschnitt 6.1 bzw. Abschnitt 6.5. Im Falle amerikanischer Optionen ändert sich dabei sogar der Typ der zu lösenden Gleichung, vgl. Günther und Jüngel, Kapitel 7.

Es kann aber auch sein, dass sich die Black-Scholes Gleichung selbst gar nicht ändert, sondern dass nur andere Randbedingungen eingesetzt werden müssen. Dies ist z.B. der Fall für die sogenannten (zugegebenermaßen relativ exotischen) Power-Optionen, in denen die Auszahlungsfunktion z.B. (im Falle eines sogenannten europäischen capped symmetrischen Power-Calls) durch $V(T, S) = \min\{L, ((S - K)^+)^p\}$ für Parameter $L, p > 0$ gegeben ist. Hier bleibt die Gleichung bestehen und es ändern sich nur die Randbedingungen. Ähnlich ist dies bei sogenannten Barriere-Optionen, bei denen die Auszahlung nur dann fällig wird, wenn der Basiswert während der Laufzeit eine gegebene Schranke B unterschreitet (down-and-in) oder nicht unterschreitet (down-and-out), die sich im europäischen Fall allerdings wieder analytisch lösen lassen.

In diesem Kapitel werden wir ein numerisches Verfahren kennenlernen, mit dem sich die Black-Scholes-Gleichung numerisch lösen lässt, die sogenannte Finite Differenzen Methode. Allerdings werden wir das Verfahren nicht direkt für diese Gleichung sondern für eine einfachere partielle Differentialgleichung, die sogenannte Wärmeleitungsgleichung einführen. Ein Grund dafür ist, dass diese Gleichung viel einfacher strukturiert ist, weswegen sich das Verfahren daran leichter erklären lässt.

11.1 Black-Scholes- und Wärmeleitungsgleichung

Der zweite Grund für die Betrachtung der Wärmeleitungsgleichung ist, dass wir diese in gewissem Sinne “o.B.d.A.” verwenden können, da sich die Black-Scholes-Gleichung durch eine Variablentransformation in diese Gleichung umformen lässt. Diese Transformation wollen wir zunächst erläutern.

Wir verwenden in diesem Kapitel für Ableitungen die übliche Notation aus Theorie und Numerik partieller Differentialgleichungen, nämlich z.B.

$$\frac{\partial V}{\partial t} = V_t \quad \text{oder} \quad \frac{\partial^2 V}{\partial S^2} = V_{SS}.$$

Die Black-Scholes-Gleichung (8.6) schreibt sich damit als

$$V_t(t, S) + \frac{1}{2}\sigma^2 S^2 V_{SS}(t, S) - rV + rSV_S(t, S) = 0,$$

wobei $V : [0, T] \times \mathbb{R}_{\geq 0} \rightarrow \mathbb{R}$ die unbekannte Funktion ist, welche die Randbedingung (wir betrachten hier beispielhaft einen Call)

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

zur Zeit $t = T$ für alle $S \in [0, \infty)$ erfüllt.

Die Wärmeleitungsgleichung hingegen lautet einfach

$$u_t(t, x) - u_{xx}(t, x) = 0 \tag{11.1}$$

für die unbekannte Funktion $u(t, x)$ mit $u : \mathbb{R}_{\geq 0} \times \mathbb{R} \rightarrow \mathbb{R}$. Die Randbedingung ist hierbei typischerweise zur Zeit $T = 0$ als $u(0, x) = u_0(x)$ für alle $x \in \mathbb{R}$ und eine vorgebene Funktion $u_0 : \mathbb{R} \rightarrow \mathbb{R}$. Die Gleichung beschreibt die Temperatur $u(t, x)$ zur Zeit t im Punkt x in einem (näherungsweise) eindimensionalen Stab unendlicher Länge.

Die Transformation der Black-Scholes-Gleichung in die Wärmeleitungsgleichung erfolgt nun in zwei Schritten. Zunächst definieren wir die Variablentransformationen

$$x = \ln(S/K) \Leftrightarrow S = Ke^x, \quad \tau = \sigma^2(T - t)/2 \Leftrightarrow t = T - 2\tau/\sigma^2$$

und setzen

$$v(\tau, x) := V(t, S)/K.$$

Beachte, dass diese Transformation das Kursintervall $S \in (0, \infty)$ auf das Intervall $x \in (-\infty, \infty)$ abbildet und die Zeit $t = T$ auf $\tau = 0$, so dass insbesondere die Randbedingung für V in $t = T$ nun eine Randbedingung für v in $\tau = 0$ wird.

Man rechnet nach, dass v die Gleichung

$$v_\tau(\tau, x) - v_{xx}(\tau, x) + (1 - \kappa)v_x(\tau, x) + \kappa v(\tau, x) = 0$$

für alle $x \in \mathbb{R}$ und alle $\tau \in [0, T_0]$ mit $\kappa = 2r/\sigma^2$ und $T_0 = \sigma^2 T/2$ erfüllt. Zudem gilt die (Anfangs-)Randbedingung

$$v(0, x) = (e^x - 1)^+.$$

Setzen wir nun

$$u(\tau, x) = \exp\left(\frac{1}{2}(\kappa - 1)x + \frac{1}{4}(\kappa + 1)^2\tau\right)v(\tau, x),$$

so rechnet man nach, dass u die Wärmeleitungsgleichung

$$u_\tau(\tau, x) - u_{xx}(\tau, x) = 0$$

mit der Randbedingung

$$u(0, x) = u_0(x) = \left(e^{(\kappa+1)x/2} - e^{(\kappa-1)x/2}\right)^+$$

erfüllt.

Die Rücktransformation lautet

$$V(t, S) = K \exp\left(-\frac{1}{2}(\kappa - 1)x - \frac{1}{4}(\kappa + 1)^2\tau\right)u(\tau, x)$$

mit den oben definierten Beziehungen zwischen t und τ bzw. S und x .

Wenn wir also die Wärmeleitungsgleichung numerisch mit der transformierten Black-Scholes Randbedingung zur Zeit $\tau = 0$ auf dem Zeitintervall $[0, T_0]$ lösen, so können wir durch Rücktransformation die (approximative) Lösung der Black-Scholes-Gleichung erhalten.

11.2 Grundidee der Finiten Differenzen

Wir beschreiben nun die Methode der finiten Differenzen für die Wärmeleitungsgleichung (11.1). Dabei bezeichnen wir die Zeitvariable wieder mit t (statt τ) und das Zeitintervall mit $[0, T]$ (statt $[0, T_0]$).

Die Lösung $u(t, x)$ der Wärmeleitungsgleichung (11.1) ist durch Ihre Ableitungen an unendlich vielen Punkten (t, x) charakterisiert. Die Methode der finiten Differenzen besteht nun darin, diese unendlich vielen Ableitungen durch eine endliche (finite) Anzahl von Differenzenquotienten (Differenzen) zu ersetzen und so eine approximative Lösung der Gleichung zu erhalten.

Die verwendeten Differenzenquotienten sind dabei für $h, s > 0$

$$u_{xx}(t, x) \approx \frac{1}{h^2}(u(t, x + h) - 2u(t, x) + u(t, x - h))$$

für die 2. Ableitung nach x und entweder die Vorwärtsdifferenz

$$u_t(t, x) \approx \frac{1}{s}(u(t + s, x) - u(t, x))$$

oder die Rückwärtsdifferenz

$$u_t(t, x) \approx \frac{1}{s}(u(t, x) - u(t - s, x))$$

für die 1. Ableitung nach t .

Hierfür gilt das folgende Lemma. Darin (und allgemein in diesem Kapitel) verwenden wir das Landau-Symbol O im folgenden Sinne: Für einen beliebigen reellen Ausdruck $y(t, x, z)$ mit $t, x, z \in \mathbb{R}$, $t \geq 0$, $z > 0$, schreiben wir $y(t, x) = O(z^p)$, falls für jede kompakte Menge $K \subset [0, \infty) \times \mathbb{R}$ eine Konstante $C_K > 0$ existiert mit $|y(t, x, z)| \leq C_K z^p$ für alle hinreichend kleinen z und alle $(t, x) \in K$.

Lemma 11.1 Für $u \in C^4$ und $h, s > 0$ gilt

$$u_{xx}(t, x) = \frac{1}{h^2}(u(t, x+h) - 2u(t, x) + u(t, x-h)) + O(h^2),$$

$$u_t(t, x) = \frac{1}{s}(u(t+s, x) - u(t, x)) + O(s)$$

und

$$u_t(t, x) = \frac{1}{s}(u(t, x) - u(t-s, x)) + O(s).$$

Beweis: Nach der Taylor-Entwicklung gelten die beiden Gleichungen

$$u(t, x+h) = u(t, x) + u_x(t, x)h + u_{xx}(t, x)\frac{h^2}{2} + u_{xxx}(t, x)\frac{h^3}{6} + O(h^4)$$

$$u(t, x-h) = u(t, x) - u_x(t, x)h + u_{xx}(t, x)\frac{h^2}{2} - u_{xxx}(t, x)\frac{h^3}{6} + O(h^4).$$

Addition der Gleichungen und Division durch h^2 ergibt

$$\frac{1}{h^2}u(t, x+h) + \frac{1}{h^2}u(t, x-h) = \frac{1}{h^2}2u(t, x) + u_{xx}(t, x) + \frac{1}{h^2}O(h^4)$$

und wegen $O(h^4)/h^2 = O(h^2)$ folgt die erste Gleichung.

Die zweite und dritte Gleichung folgen direkt aus der Taylor-Entwicklung in t . Die geforderte Existenz der Konstanten C_K folgt dabei in allen Fällen aus der Stetigkeit der Ableitungen, wodurch die entsprechenden Faktoren in den Taylor-Restgliedern stetig sind und damit auf kompakten Mengen beschränkt sind. \square

Wir erhalten damit das folgende Korollar.

Korollar 11.2 Für die Lösung $u(t, x)$ der Wärmeleitungsgleichung gilt

$$\frac{1}{s}(u(t+s, x) - u(t, x)) - \frac{1}{h^2}(u(t, x+h) - 2u(t, x) + u(t, x-h)) = O(s+h^2) \quad (11.2)$$

und

$$\frac{1}{s}(u(t, x) - u(t-s, x)) - \frac{1}{h^2}(u(t, x+h) - 2u(t, x) + u(t, x-h)) = O(s+h^2) \quad (11.3)$$

sowie für jedes $\theta \in [0, 1]$

$$\begin{aligned} & \frac{1}{s}(u(t+s, x) - u(t, x)) \\ & - \frac{1-\theta}{h^2}(u(t, x+h) - 2u(t, x) + u(t, x-h)) \\ & - \frac{\theta}{h^2}(u(t+s, x+h) - 2u(t+s, x) + u(t+s, x-h)) = O(s+h^2). \end{aligned} \quad (11.4)$$

Beweis: Die Gleichungen (11.2) und (11.3) folgen sofort durch Einsetzen der Gleichungen aus Lemma 11.1 in Gleichung (11.1) und $O(h^2) + O(s) = O(h^2 + s)$. Zum Beweis von Gleichung (11.4) betrachten wir (11.3) für $t + s$ an Stelle von t , also

$$\frac{1}{s}(u(t+s, x) - u(t, x)) - \frac{1}{h^2}(u(t+s, x+h) - 2u(t+s, x) + u(t+s, x-h)) = O(s+h^2).$$

Multiplizieren wir diese Gleichung mit θ und addieren (11.2) multipliziert mit $1 - \theta$, so erhalten wir gerade (11.4). \square

Für $\theta = 0$ ist (11.4) gerade (11.2), für $\theta = 1$ gerade (bis auf Verschiebung der Zeitvariablen t) (11.3). Wir werden später sehen, dass man für jedes $\theta \in [0, 1]$ aus (11.4) sinnvolle numerische Schemata ableiten kann, allerdings mit unterschiedlichen Eigenschaften.

Um nun zu einem implementierbaren Schema zu kommen, müssen wir zunächst das Definitionsgebiet $[0, T] \times \mathbb{R}$ von u auf eine kompakte Menge einschränken. Wir definieren dazu das Rechengebiet $\Omega = [0, T] \times [-a, a]$ mit $a > 0$. Weil das Intervall der x -Werte nun nicht mehr unendlich ist, müssen wir an den neuen Rändern $[0, T] \times \{a\}$ und $[0, T] \times \{-a\}$ Randbedingungen festlegen, um eine eindeutige Lösung zu erhalten. Wir setzen

$$u(t, -a) = g_1(t) \quad \text{und} \quad u(t, a) = g_2(t) \quad \text{für } t \in [0, T].$$

Für die physikalische Interpretation der Gleichung bedeutet diese sogenannten *Dirichlet-Randbedingungen*, dass der eindimensionale Stab nun nicht mehr unendlich lang ist, sondern die Länge $2a$ besitzt. Die Funktionen g_1 und g_2 beschreiben dabei gerade die Temperatur an den freien Endpunkten $-a$ und a des Stabes. Wie die Funktionen $g_1(t)$ und $g_2(t)$ im Fall der transformierten Black-Scholes-Gleichung sinnvoll gewählt werden, betrachten wir in Abschnitt 12.1.

Auf dem Gebiet Ω definieren wir nun ein regelmäßiges Gitter aus $(M+1) \cdot (N+1)$ Punkten (t_j, x_i) , $j = 0, \dots, M$, $i = 0, \dots, N$. Dazu setzen wir $s := T/M$ und $h := 2a/N$ und definieren

$$t_j := sj, \quad j = 0, \dots, M \quad \text{und} \quad x_i := -a + hi, \quad i = 0, \dots, N.$$

Beachte, dass damit $t_{j+1} = t_j + s$ und $x_{i+1} = x_i + h$ sowie $x_{i-1} = x_i - h$ gilt.

Bezeichnen wir nun die Werte der exakten Lösung u in den Gitterpunkten kurz mit

$$u_i^j := u(t_j, x_i),$$

so können wir (11.4) für $t = t_j$ und $x = x_i$ schreiben als

$$\frac{1}{s}(u_i^{j+1} - u_i^j) - \frac{1-\theta}{h^2}(u_{i+1}^j - 2u_i^j + u_{i-1}^j) - \frac{\theta}{h^2}(u_{i+1}^{j+1} - 2u_i^{j+1} + u_{i-1}^{j+1}) = O(s+h^2). \quad (11.5)$$

Die Finite Differenzenmethode beruht nun darauf, für gegebenes $\theta \in [0, 1]$ Werte $w_i^j \in \mathbb{R}$ zu berechnen, indem diese Gleichungen mit rechter Seite $= 0$ gelöst werden, wobei die entsprechenden Randbedingungen berücksichtigt werden.

Wir suchen also also Werte $w_i^j \in \mathbb{R}$, $i = 0, \dots, N$, $j = 0, \dots, M$, welche die Gleichungen

$$\frac{1}{s}(w_i^{j+1} - w_i^j) - \frac{1-\theta}{h^2}(w_{i+1}^j - 2w_i^j + w_{i-1}^j) - \frac{\theta}{h^2}(w_{i+1}^{j+1} - 2w_i^{j+1} + w_{i-1}^{j+1}) = 0 \quad (11.6)$$

für $i = 1, \dots, N - 1$ und $j = 0, \dots, M - 1$ sowie die Randbedingungen

$$w_i^0 = u_0(x_i), \quad i = 0, \dots, N \quad \text{und} \quad w_0^j = g_1(t_j), \quad w_N^j = g_2(t_j), \quad j = 0, \dots, M \quad (11.7)$$

erfüllen. Die Hoffnung ist dabei natürlich, dass die Werte Approximationen $w_i^j \approx u_i^j$ der exakten Funktionswerte $u_i^j = u(t_j, x_i)$ in den Gitterpunkten darstellen. Wir werden später beweisen, dass dies unter geeigneten Voraussetzungen auch tatsächlich so ist.

11.3 Lösung der Finiten Differenzgleichungen

Wie löst man nun das Gleichungssystem (11.6), (11.7) numerisch? Um (11.6) dafür noch etwas zu vereinfachen, multiplizieren wir die Gleichung mit s und setzen $\alpha := s/h^2$. Mit Umstellen der Terme (alle Terme mit $j + 1$ nach links, den Rest nach rechts) ist (11.6) dann äquivalent zu

$$-\alpha\theta w_{i+1}^{j+1} + (2\alpha\theta + 1)w_i^{j+1} - \alpha\theta w_{i-1}^{j+1} = \alpha(1 - \theta)w_{i+1}^j + (1 - 2\alpha(1 - \theta))w_i^j + \alpha(1 - \theta)w_{i-1}^j.$$

Diese Gleichung schreiben wir nun rekursiv in Matrixform und setzen die Randwerte gleich ein. Dies ergibt die implizite rekursive Vorschrift

$$Aw^{j+1} = Bw^j + d^j, \quad j = 0, \dots, M - 1 \quad (11.8)$$

mit den Unbekannten

$$w^j = (w_1^j, \dots, w_{N-1}^j)^T,$$

der Matrix

$$A = \begin{pmatrix} 2\alpha\theta + 1 & -\alpha\theta & 0 & \cdots & 0 \\ -\alpha\theta & 2\alpha\theta + 1 & -\alpha\theta & & 0 \\ 0 & \ddots & \ddots & \ddots & 0 \\ \vdots & & & & 0 \\ 0 & \cdots & 0 & -\alpha\theta & 2\alpha\theta + 1 \end{pmatrix},$$

die wir kurz auch als $A = \text{diag}(-\alpha\theta, 2\alpha\theta + 1, \alpha\theta)$ schreiben, der Matrix

$$B = \text{diag}(\alpha(1 - \theta), 1 - 2\alpha(1 - \theta), \alpha(1 - \theta))$$

und dem Vektor

$$d^j = \begin{pmatrix} \alpha(1 - \theta)g_1(t_j) + \alpha\theta g_1(t_{j+1}) \\ 0 \\ \vdots \\ 0 \\ \alpha(1 - \theta)g_2(t_j) + \alpha\theta g_2(t_{j+1}) \end{pmatrix},$$

der sich aus den Randbedingungen für w_0^j und w_N^j in (11.7) ergibt. Da durch (11.7) zudem die Werte w^0 eindeutig bestimmt sind, können wir (11.8) damit rekursiv für $j = 0, \dots, M - 1$ lösen, indem wir in jedem Schritt das lineare Gleichungssystem mit der angegebenen Matrix A numerisch lösen, vorausgesetzt A ist invertierbar, was wir am Ende dieses Abschnitts untersuchen.

Formal könnte man die Iteration (11.8) dann auch explizit als

$$w^{j+1} = A^{-1}Bw^j + A^{-1}d^j$$

schreiben, aus der Einführung in die Numerik ist aber bekannt, dass die explizite Verwendung von A^{-1} numerisch weniger effizient als die Lösung des zugehörigen linearen Gleichungssystems (11.8) ist.

Für $\theta = 0$ nennt man das Verfahren *Vorwärts-Differenzenverfahren*, für $\theta = 1$ *Rückwärts-Differenzenverfahren*. Für $\theta \in (0, 1)$ erhält man *gemischte Vorwärts-Rückwärts-Differenzenverfahren*, von denen besonders das Verfahren für $\theta = 1/2$ wichtig ist, das sogenannte *Crank-Nicolson-Verfahren*. Warum das so ist, sehen wir später.

Für das Vorwärts-Differenzenverfahren mit $\theta = 0$ wird die Lösung von (11.8) besonders einfach, da A dann die Einheitsmatrix ist. Wir werden aber im nächsten Abschnitt sehen, dass dies aus anderen Gründen keine besonders gute Wahl ist.

Um die Lösbarkeit von (11.8) und allgemein die Existenz eindeutiger Lösungen w^j zu garantieren, reicht es aus, die Invertierbarkeit von A zu untersuchen. Dazu müssen wir nachweisen, dass A keinen Eigenwert $\lambda = 0$ besitzt. Dies kann man leicht mit dem folgenden Satz aus der linearen Algebra beweisen.

Satz 11.3 (Satz von Gerschgorin) Für alle Eigenwerte $\lambda \in \mathbb{C}$ einer Matrix $A \in \mathbb{R}^{n \times n}$ mit Einträgen a_{ij} gilt

$$\lambda \in \bigcup_{i=1}^n \left\{ z \in \mathbb{C} \mid |z - a_{ii}| \leq \sum_{\substack{j=1 \\ j \neq i}}^n |a_{ij}| \right\}.$$

Beweis: Sei $x = (x_1, \dots, x_n)^T \neq 0$ ein Eigenvektor von A zu λ und sei x_i der betragsmäßig größte Eintrag von x . Wegen

$$\lambda x_i = (Ax)_i = \sum_{j=1}^n a_{ij} x_j$$

folgt mit Division durch $x_i \neq 0$

$$|\lambda - a_{ii}| = \left| \sum_{j=1}^n a_{ij} \frac{x_j}{x_i} - a_{ii} \right| = \left| \sum_{\substack{j=1 \\ j \neq i}}^n a_{ij} \frac{x_j}{x_i} \right| \leq \sum_{\substack{j=1 \\ j \neq i}}^n |a_{ij}| \left| \frac{x_j}{x_i} \right| \leq \sum_{\substack{j=1 \\ j \neq i}}^n |a_{ij}|.$$

Daraus folgt die Behauptung. □

Korollar 11.4 Für alle Eigenwerte λ der Matrix A aus (11.8) gilt $|\lambda| \geq 1$ für alle $\alpha > 0$ und alle $\theta \in [0, 1]$. Insbesondere ist A invertierbar.

Beweis: Sei λ ein Eigenwert von A . Nach Satz 11.3 gilt dann für ein $i = 1, \dots, N - 1$

$$|\lambda - a_{ii}| \leq \sum_{\substack{j=1 \\ j \neq i}}^{N-1} |a_{ij}| \leq 2\alpha\theta.$$

Damit folgt mit der umgekehrten Dreiecksungleichung und wegen $a_{ii} = 2\alpha\theta + 1 > 0$

$$2\alpha\theta \geq |\lambda - a_{ii}| \geq 2\alpha\theta + 1 - |\lambda|,$$

woraus die Behauptung folgt. \square

11.4 Konsistenz, Stabilität und Konvergenz

Für die Konvergenzanalyse des Finiten Differenzschemas benötigen wir zwei Eigenschaften: Konsistenz und Stabilität¹. Die Konsistenzbedingung stellt dabei sicher, dass die Lösung in einem Schritt von (11.8) nur leicht von der exakten Lösung abweicht, die Stabilitätsbedingung garantiert, dass sich diese kleinen Abweichungen in der Iteration in (11.8) nicht zu sehr großen Fehlern aufschaukeln. Wir werden diese beiden Begriffe nun genau definieren und beweisen, dass daraus tatsächlich die Konvergenz — in geeignetem Sinne — folgt.

Bevor wir die Konsistenz definieren, erinnern wir an den Konsistenzbegriff für Einschrittverfahren für gewöhnliche Differentialgleichungen aus Kapitel 6. Dort war ein Verfahren $\Phi(t, x, h)$ konsistent mit Konsistenzordnung p , wenn für alle hinreichend kleinen $h > 0$ die Ungleichung

$$\|\Phi(t_0, x_0, h) - x(t_0 + h; t_0, x_0)\| \leq Eh^{p+1} \quad (11.9)$$

gilt, vgl. Definition 6.9. Wir haben diese Ungleichung dort als Vergleich eines Schritts des numerischen Verfahrens mit der exakten Lösung interpretiert. Man kann die Konsistenzbedingung aber auch anders interpretieren: Betrachten wir die numerische Lösung $\tilde{x}(t_i)$, so erfüllt diese (bei konstanter Schrittweite $h > 0$) gerade

$$\tilde{x}(t_{i+1}) = \Phi(t_i, \tilde{x}(t_i), h).$$

Dies kann man äquivalent auch als

$$\|\Phi(t_i, \tilde{x}(t_i), h) - \tilde{x}(t_{i+1})\| = 0 \quad (11.10)$$

schreiben. Ersetzen wir nun $\tilde{x}(t_i)$ durch die exakte Lösung $x(t_i)$ und nutzen aus, dass $x(t_i + h; t_i, x(t_i)) = x(t_{i+1})$ ist, so erhalten wir nach Ersetzen von \tilde{x} durch x in (11.10) für ein konsistentes Verfahren aus (11.9) gerade

$$\|\Phi(t_i, x(t_i), h) - x(t_{i+1})\| \leq Eh^{p+1}.$$

Der Konsistenzfehler kann also alternativ auch als der Fehler betrachtet werden, den man beim Einsetzen der exakten Lösung in das numerische Schema erhält. Konsistenz heißt

¹Teilnehmer/innen der Vorlesung “Numerik Gewöhnlicher Differentialgleichungen” kennen diese Begriffe und das folgende Vorgehen bereits aus der Konvergenzanalyse von linearen Mehrschrittverfahren für gewöhnliche Differentialgleichungen.

dann, dass die exakte Lösung die numerische Iterationsvorschrift approximativ (d.h. hier bis auf den Fehlerterm Eh^{p+1}) erfüllt.

Dies ist genau die Definition der Konsistenz, die wir für das Finite Differenzenschema verwenden: Wir verlangen, dass die exakte Lösung der Gleichung (11.1) die Iterationsvorschrift (11.8) approximativ erfüllt.

Definition 11.5 Das Finite Differenzenschema (11.8) heißt konsistent mit Ordnung $a > 0$ in der Zeit und Ordnung $b > 0$ im Raum, falls für die mittels $u^j = (u(t_j, x_1), \dots, u(t_j, x_{N-1}))^T$ definierten Vektoren mit den exakten Lösungswerten die Ungleichung

$$\|Au^{j+1} - Bu^j - d^j\|_\infty \leq Cs(s^a + h^b) \quad (11.11)$$

für eine Konstante $C > 0$ und alle $j = 0, \dots, M - 1$ erfüllen. \square

Beachte, dass die Bedingung einen zusätzlichen Faktor s aber keinen Faktor h in der Abschätzung verlangt. Dies liegt daran, dass über die t -Variable $M = T/s$ Schritte iteriert wird, wodurch — ganz analog zu den numerischen Schemata für gewöhnliche Differentialgleichungen — in der Konvergenzanalyse eine Potenz “verlorenght”. Da über x nicht iteriert wird, ist eine Multiplikation mit der Ortsschrittweite h nicht nötig.

Der folgende Satz zeigt, dass das Schema konsistent ist.

Satz 11.6 Die Lösung der Wärmeleitungsgleichung erfülle $u \in C^4$. Dann ist das Schema (11.8) für alle $\theta \in [0, 1]$ konsistent mit $a = 1$ und $b = 2$. Im Falle $\theta = 1/2$ erhöht sich die erste Konsistenzordnung auf $a = 2$.

Beweis: Die allgemeine Aussage folgt sofort aus (11.5) wenn wir beachten, dass die Komponenten der rechten Seite von (11.11) gerade durch Multiplikation von (11.5) mit s entstehen, weswegen wir nun $O(s(s + h^2))$ erhalten. Mit $K = \Omega$ und der vor Lemma 11.1 definierten Gleichmäßigkeit der Konstanten in den O -Termen folgt die Existenz von $C = C_K$.

Die spezielle Aussage für $\theta = 1/2$ folgt, wenn wir zusätzlich zu den Abschätzungen aus Lemma 11.1 noch die zusätzlichen Taylor-Entwicklungen

$$\frac{1}{s}(u_i^{j+1} - u_i^j) = u_t(t_j, x_i) + \frac{s}{2}u_{tt}(t_j, x_i) + O(s^2),$$

$$\frac{1}{s}(u_{xx}(t_{j+1}, x_i) - u_{xx}(t_j, x_i)) = u_{xxt}(t_j, x_i) + O(s)$$

sowie

$$\begin{aligned} & \frac{1-\theta}{h^2}(u_{i+1}^j - 2u_i^j + u_{i-1}^j) - \frac{\theta}{h^2}(u_{i+1}^{j+1} - 2u_i^{j+1} + u_{i-1}^{j+1}) \\ &= (1-\theta)u_{xx}(t_j, x_i) + \theta u_{xx}(t_{j+1}, x_i) + O(h^2) \\ &= u_{xx}(t_j, x_i) + \theta(u_{xx}(t_{j+1}, x_i) - u_{xx}(t_j, x_i)) + O(h^2) \\ &= u_{xx}(t_j, x_i) + \theta s u_{xxt}(t_j, x_i) + O(s^2 + h^2) \end{aligned}$$

betrachten. Damit lässt sich (11.5) verbessern zu

$$\begin{aligned} & \frac{1}{s}(u_i^{j+1} - u_i^j) - \frac{1-\theta}{h^2}(u_{i+1}^j - 2u_i^j + u_{i-1}^j) - \frac{\theta}{h^2}(u_{i+1}^{j+1} - 2u_i^{j+1} + u_{i-1}^{j+1}) \\ &= \frac{s}{2}(u_{tt}(t_j, x_i) - 2\theta u_{xxt}(t_j, x_t)) + O(s^2 + h^2). \end{aligned}$$

Wegen $u_{tt} - u_{xxt} = (u_t - u_{xx})_t = 0$ (denn u löst ja gerade (11.1), also $u_t - u_{xx} = 0$) fällt der erste Summand für $\theta = 1/2$ weg. Wie oben erhalten wir durch Multiplikation dieser Gleichung mit s gerade die Komponenten von (11.11) und damit die Behauptung. \square

Im Sinne der Konsistenz ist folglich $\theta = 1/2$ — also das Crank-Nicolson-Verfahren — am Besten.

Die zweite Bedingung, die wir zur Herleitung der Konvergenz benötigen, ist die sogenannte Stabilität. Wir erläutern zunächst anschaulich, warum wir eine zweite Bedingung benötigen:

Bezeichnen wir mit $e^j = Au^{j+1} - Bu^j - d^j$ den lokalen Konsistenzfehler und mit $\hat{e}^j = u^j - w^j$ den globalen Fehler der numerischen Lösung, so gilt für den Fehler die Rekursion

$$A\hat{e}^{j+1} = Au^{j+1} - Aw^{j+1} = Bu^j + d^j + e^j - Bw^j - d^j = B\hat{e}^j + e^j,$$

was wir auch als

$$\hat{e}^{j+1} = A^{-1}B\hat{e}^j + A^{-1}e^j$$

schreiben können. Schreiben wir kurz $\tilde{A} = A^{-1}B$, $\tilde{e}^j = A^{-1}e^j$ und verwenden $\hat{e}_0 = 0$ (da die Randbedingungen in $t_0 = 0$ für u und w ja identisch sind), so können wir per Induktion leicht die explizite Formel

$$\hat{e}^j = \sum_{k=0}^{j-1} \tilde{A}^{j-k-1} \tilde{e}^k \quad (11.12)$$

herleiten. Für immer kleiner werdende s und h passieren nun zwei Dinge: Die Anzahl der Schritte j wird immer größer und die Matrix A^{-1} , die den Vektor \tilde{e}^k definiert, verändert sich. Das zweite Problem untersuchen wir später im Beweis von Satz 11.9, das erste betrachten wir jetzt.

Da wir nicht davon ausgehen können, dass sich die einzelnen Summanden günstig gegeneinander aufheben, muss nun, damit \hat{e}^j klein wird, jeder der Summanden klein sein. Für jedes beliebige $k \in \mathbb{N}$ mit $k < j$ muss also der Term $\tilde{A}^{j-k-1} \tilde{e}^k$ klein sein. Da wir für immer kleinere Zeitschritte s auch bei festem Zeitintervall $[0, T]$ immer mehr Schritte und damit immer größere j erhalten, können wir keine von s unabhängige obere Schranke für $j - k - 1$ angeben; ebenso haben wir keine Informationen über \tilde{e}^k . Wir können zwar annehmen, dass dieser Fehler klein ist, er wird für $s \rightarrow 0$ aber nicht beliebig klein, da ja auch die räumliche Diskretisierung zu diesem Fehler beiträgt.

Wir brauchen also eine Eigenschaft der Matrix \tilde{A} , die sicherstellt, dass $\tilde{A}^l x$ für beliebige Vektoren x mit kleiner Norm und beliebigem $l \in \mathbb{N}$ ebenfalls klein ist.

Dies liefert uns Satz 9.10 aus der Numerik gewöhnlicher Differentialgleichungen:

Satz 11.7 Für eine Matrix $A \in \mathbb{R}^{n \times n}$ existiert genau dann eine Konstante $C > 0$, so dass

$$\|A^l x\| \leq C \|x\|$$

gilt für alle $l \in \mathbb{N}$ und alle $x \in \mathbb{R}^n$, wenn alle Eigenwerte λ von A die Ungleichung $|\lambda| \leq 1$ erfüllen und alle Eigenwerte mit $|\lambda| = 1$ halbeinfach sind (d.h. dass λ eine einfache Nullstelle des Minimalpolynoms ist).

Der folgende Satz wendet dies Kriterium auf die Matrix $A^{-1}B$ an.

Satz 11.8 Die Matrix $A^{-1}B$ erfüllt die Stabilitätsbedingung aus Satz 11.7, wenn die Ungleichungen

$$\begin{aligned} 0 < \alpha \leq \frac{1}{2 - 4\theta}, & \quad \text{falls } 0 \leq \theta < \frac{1}{2} \\ 0 < \alpha, & \quad \text{falls } \frac{1}{2} \leq \theta \leq 1 \end{aligned}$$

gelten. Für die euklidische Norm $\|\cdot\|_2$ kann die Konstante C in Satz 11.7 sogar als $C = 1$ gewählt werden.

Beweis: Wegen

$$B = \frac{1}{\theta} \text{Id} - \frac{1 - \theta}{\theta} A$$

gilt

$$A^{-1}B = \frac{1}{\theta} A^{-1} - \frac{1 - \theta}{\theta} \text{Id}.$$

Jeder Eigenwert von $A^{-1}B$ ist also von der Form $\lambda = \frac{1}{\theta\mu} - \frac{1-\theta}{\theta}$, wobei μ ein Eigenwert von A ist. Wir können nun A schreiben als $A = \text{Id} + \alpha\theta G$ mit $G = \text{diag}(-1, 2, -1)$. Für G kann man nachrechnen, dass die Eigenwerte gerade durch

$$4 \sin^2 \left(\frac{k\pi}{2N} \right), \quad k = 1, \dots, N - 1$$

gegeben sind. Die Eigenwerte von A lauten daher

$$\mu_k = 1 + 4\alpha\theta \sin^2 \left(\frac{k\pi}{2N} \right), \quad k = 1, \dots, N - 1$$

und die von $A^{-1}B$ folglich

$$\lambda_k = \frac{1}{\theta} \frac{1}{1 + 4\alpha\theta \sin^2(k\pi/2N)} - \frac{1 - \theta}{\theta} = 1 - \frac{4\alpha \sin^2(k\pi/2N)}{1 + 4\alpha\theta \sin^2(k\pi/2N)}.$$

Weil $A^{-1}B$ symmetrisch und damit diagonalisierbar ist, sind alle Eigenwerte halbeinfach. Es reicht also, $|\lambda_k| \leq 1$ nachzuweisen.

Wegen $\alpha > 0$ ist $|\lambda_k| \leq 1$ äquivalent zu

$$\frac{4\alpha \sin^2(k\pi/2N)}{1 + 4\alpha\theta \sin^2(k\pi/2N)} \leq 2 \Leftrightarrow (2 - 4\theta)\alpha \sin^2(k\pi/2N) \leq 1.$$

Für $\theta \geq 1/2$ ist diese Ungleichung immer erfüllt, für $\theta \in [0, 1/2)$ wegen $\sin^2(k\pi/2N) \leq 1$ für alle $\alpha \leq 1/(2 - 4\theta)$.

Dass die Konstante als $C = 1$ gewählt werden kann, folgt wiederum aus der Symmetrie von $A^{-1}B$. Für symmetrische Matrizen A ist die induzierte Matrixnorm $\|A\|_2$ gerade gleich dem maximalen Betrag der Eigenwerte (dies folgt aus der Darstellung der 2-Norm für Matrizen über den Spektralradius, vgl. Satz 2.5 aus der Einführung in die Numerik). Damit folgt für alle $x \in \mathbb{R}^n$ und alle $j \in \mathbb{N}$ die Ungleichung $\|A^j x\|_2 \leq \|A\|_2^j \|x\|_2 \leq \|x\|_2$, also $C = 1$. \square

Im Sinne der Stabilität sind also die Verfahren mit $\theta \geq 1/2$ besonders gut, weil sie für alle Kombinationen aus Zeit- und Raumschrittweite s und h stabil sind. Man sagt, diese Verfahren sind *unbedingt stabil*.

Nun können wir schließlich die Konvergenz beweisen.

Satz 11.9 Unter den Voraussetzungen der Sätze 11.6 und 11.8 existiert eine Konstante $C_0 > 0$, so dass für alle $\theta \in [0, 1]$ und alle hinreichend kleinen $s, h > 0$ gilt

$$\max_{j=1, \dots, M} \|w^j - u^j\|_{l_2} \leq C_0(s + h^2).$$

Im Falle $\theta = 1/2$ gilt diese Abschätzung sogar mit $C_0(s^2 + h^2)$.

Hierbei ist die *diskrete l_2 -Norm* $\|\cdot\|_{l_2}$ definiert durch

$$\|x\|_{l_2} = \|x\|_2 / (N - 1).$$

Beweis: Betrachte die Fehlergleichung (11.12), die sich durch einfache Ummummerierung der Indizes schreiben lässt als

$$\hat{e}^j = \sum_{k=0}^{j-1} \tilde{A}^k \tilde{e}^{j-k-1}$$

mit $\tilde{A} = A^{-1}B$, $\tilde{e}^j = A^{-1}e^j$ und $\hat{e}^j = u^j - w^j$. Wir müssen zeigen, dass $\|\hat{e}^j\|_{l_2} \leq C_0(s + h^2)$ bzw. $\leq C_0(s^2 + h^2)$ gilt für alle $j = 1, \dots, M$.

Für die lokalen Fehler e^j gilt nach Satz 11.6 und der Definition der diskreten l_2 -Norm die Ungleichung $\|e^j\|_{l_2} \leq \|e^j\|_\infty \leq Cs(s^p + h^2)$ mit $p = 1$ für allgemeines θ und $p = 2$ für $\theta = 1/2$. Nach Korollar 11.4 gilt $|\lambda| \geq 1$ für alle Eigenwerte λ von A . Also gilt $|\lambda| \leq 1$ für alle Eigenwerte von A^{-1} und weil A und damit auch A^{-1} symmetrisch sind folgt $\|A^{-1}\|_2 \leq 1$ (vgl. auch das Ende des Beweises von Satz 11.8). Damit folgt $\|\tilde{e}^j\|_{l_2} \leq Cs(s^p + h^2)$.

Nach Satz 11.8 gilt für die Summanden von \hat{e}^j nun $\|\tilde{A}^k \tilde{e}^{j-k-1}\|_{l_2} \leq \|\tilde{e}^{j-k-1}\|_{l_2}$. Damit folgt

$$\|\hat{e}^j\|_{l_2} \leq \sum_{k=0}^{j-1} \|\tilde{A}^k \tilde{e}^{j-k-1}\|_{l_2} \leq jCs(s^p + h^2).$$

Wegen $j \leq M \leq T/s$ folgt

$$\|\hat{e}^j\|_{l_2} \leq \frac{T}{s} Cs(s^p + h^2) = TC(s^p + h^2).$$

Damit ergibt sich die Behauptung mit $C_0 = TC$. \square

Dieser Satz zeigt sofort, warum das Crank-Nicolson-Verfahren, also $\theta = 1/2$, das bevorzugte Verfahren ist: Es hat die höchste Konvergenzordnung und ist unbedingt — d.h. für alle Kombinationen von s und h bzw. von N und M — stabil.

Kapitel 12

Finite Differenzen für die Black-Scholes Gleichung

12.1 Wahl der numerischen Randbedingungen

Im letzten Kapitel haben wir die Finite Differenzen-Methode für die Wärmeleitungsgleichung

$$u_t - u_{xx} = 0$$

betrachtet. Um nun die Black-Scholes-Gleichung mit dieser Methode zu lösen, können wir diese wie in Abschnitt 11.1 beschrieben in die Wärmeleitungsgleichung transformieren, die Wärmeleitungsgleichung mit den transformierten Randbedingungen lösen und die so erhaltene numerische Lösung mit der am Ende von Abschnitt 11.1 angegebenen Formel zurücktransformieren.

Wenn wir so vorgehen, müssen wir berücksichtigen, dass wir den Bereich $[0, \infty)$ der räumlichen Variablen S mittels $x = \ln(S/K)$ auf das Gebiet $(-\infty, \infty)$ für x transformieren, das wir für die numerische Approximation auf ein endliches Intervall $[-a, a]$ einschränken müssen. Dadurch muss an jedem der Ränder $[0, T_0] \times \{-a\}$ und $[0, T_0] \times \{a\}$ eine Randbedingung eingesetzt werden. Da die exakte Lösung nicht als bekannt vorausgesetzt werden soll (denn ansonsten bräuchten wir ja keine numerische Lösung), müssen hier Näherungswerte für die (transformierte) Lösung der Black-Scholes-Gleichung eingesetzt werden.

In den ursprünglichen Koordinaten entsprechen diese Ränder gerade

$$[0, T] \times \{Ke^{-a}\} \quad \text{und} \quad [0, T] \times \{Ke^a\}.$$

Der “untere” Rand liegt also für große a nahe an der Null, der “obere” für große a weit weg von der Null. Für die Wahl geeigneter Randbedingungen für europäische Optionen betrachten wir die a priori Abschätzungen aus Satz 2.4:

- (i) $(S - Ke^{r(t-T)})^+ \leq V_{Ce}(t, S) \leq S$
- (ii) $(Ke^{r(t-T)} - S)^+ \leq V_{Pe}(t, S) \leq Ke^{r(t-T)}.$

Für $S \rightarrow 0$ folgt hieraus $V_{Ce}(t, S) \rightarrow 0$ und $V_{Pe}(t, S) \rightarrow Ke^{r(t-T)}$. Für große S gilt $V_{Ce}(t, S) \approx S$ und $V_{Pe}(t, S) \approx 0$, wobei “ \approx ” hier im Sinne des relativen Fehlers zu verstehen

ist, d.h. es gilt

$$\frac{V_{Ce}(t, S) - S}{S} \rightarrow 0 \quad \text{und} \quad \frac{V_{Pe}(t, S) - 0}{S} \rightarrow 0$$

für $S \rightarrow \infty$. Dies führt zu den Randbedingungen

$$V_{Ce}(t, Ke^{-a}) = 0, \quad V_{Ce}(t, Ke^a) = Ke^a$$

und

$$V_{Pe}(t, Ke^{-a}) = Ke^{r(t-T)}, \quad V_{Pe}(t, Ke^a) = 0.$$

Beachte, dass wir durch die nur approximativ gültigen Randwerte hier einen Fehler erzeugen, der sich in der numerischen Lösung auswirkt. Man kann allerdings beweisen, dass der Fehler für festes S und t und immer größer werdendes a immer kleiner wird.

Für die numerische Lösung der Wärmeleitungsgleichung müssen wir diese Randbedingung nun noch entsprechend transformieren. Es gilt dann mit den Bezeichnungen des vorhergehenden Kapitels

$$g_1(t) = 0, \quad g_2(t) = \exp\left(-\frac{1}{2}(\kappa - 1)a + \frac{1}{4}(\kappa + 1)^2 t\right) e^a$$

für den europäischen Call und

$$g_1(t) = \exp\left(-\frac{1}{2}(\kappa - 1)a + \frac{1}{4}(\kappa + 1)^2 t\right) e^{r(-2t/\sigma^2)}, \quad g_2(t) = 0$$

für den europäischen Put.

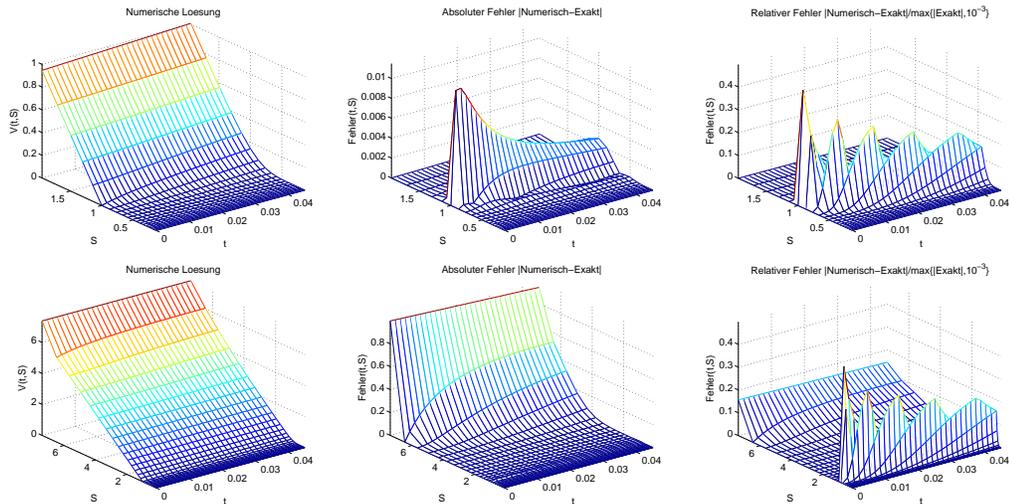


Abbildung 12.1: Numerische Ergebnisse und Fehler auf verschiedenen Gebieten

In Abbildung 12.1 sind die numerischen Lösungen sowie die — durch den Vergleich mit der exakten Lösung erhaltenen — Fehler für den Call abgebildet. Gerechnet wurde hier mit $a = 2$, also für $S \in [e^{-a}, e^a] \approx [0.1353, 7.3891]$. Das Gebiet wurde mit 30 Intervallen je Richtung diskretisiert und die Lösung mit dem Crank-Nicolson-Verfahren berechnet.

In der oberen Zeile sind die Lösungen auf $[0, T] \times [e^{-a}, 2]$ dargestellt. Man sieht, dass sich die nicht ganz korrekten Randwerte hier kaum auswirken. Anders ist dies, wenn die Lösung auf dem ganzen Rechengebiet $[0, T] \times [e^{-a}, e^a]$ dargestellt wird, was in der unteren Zeile des Bildes gemacht wurde. Hier zeigt sich, dass sich der Fehler deutlich stärker auswirkt.

Dass der Fehler im Inneren des Rechengebietes besonders in der Nähe des “Knicks” im Ausübungspreis groß ist, ist natürlich kein Zufall. Hier ist die Lösung am Rand nicht differenzierbar und auch wenn die Lösung im Inneren des Gebietes differenzierbar ist, besitzt sie doch große Ableitungen, was große Restglieder in der Taylor-Entwicklung und damit große numerische Fehler nach sich zieht.

12.2 Direkte Diskretisierung der Black-Scholes-Gleichung

Die numerische Lösung der Black-Scholes-Gleichung über die Transformation in die Wärmeleitungsgleichung ist nur ein möglicher Weg, um diese Gleichung numerisch zu lösen. Alternativ können wir die Gleichung nämlich auch direkt diskretisieren und die Transformation damit vermeiden.

Betrachten wir die Black-Scholes-Gleichung

$$V_t(t, S) + \frac{1}{2}\sigma^2 S^2 V_{SS}(t, S) - rV + rSV_S(t, S) = 0,$$

so haben wir bereits Differenzenquotienten für die Ableitungen V_t und V_{SS} eingeführt. Zusätzlich tauchen hier noch die Terme V und V_S auf. Für den ersten brauchen wir keinen Differenzenquotienten, weil es sich nicht um eine Ableitung handelt. Für den zweiten haben wir verschiedene Möglichkeiten: Wir könnten wieder Vorwärts- oder Rückwärtsdifferenzen verwenden, also

$$V_S(t, S) \approx \frac{1}{h}(V(t, S+h) - V(t, S)) \quad \text{oder} \quad V_S(t, S) \approx \frac{1}{h}(V(t, S) - V(t, S-h)).$$

Es gibt aber auch die Möglichkeit, den sogenannten *zentralen Differenzenquotienten*

$$V_S(t, S) \approx \frac{1}{2h}(V(t, S+h) - V(t, S-h))$$

zu verwenden. Aus der Taylor-Entwicklung folgt

$$V(t, S+h) = V(t, S) + hV_S(t, S) + \frac{h^2}{2}V_{SS}(t, S) + O(h^3)$$

sowie

$$V(t, S-h) = V(t, S) - hV_S(t, S) + \frac{h^2}{2}V_{SS}(t, S) + O(h^3).$$

Subtraktion der zweiten Gleichung von der ersten liefert

$$V(t, S+h) - V(t, S-h) = 2hV_S(t, S) + O(h^3),$$

woraus durch Division durch h und Umstellen der Terme

$$V_S(t, S) = \frac{1}{2h}(V(t, S+h) - V(t, S-h)) + O(h^2)$$

folgt. Der zentrale Differenzenquotient ist also eine Approximation zweiter Ordnung und liefert daher eine bessere Approximation als der Vorwärts- oder Rückwärts-Quotient, vgl. Lemma 11.1.

Das Finite Differenzschema für die Black-Scholes-Gleichung erhalten wir nun, indem wir die Ableitungen in der Gleichung durch die entsprechenden Differenzenquotienten ersetzen. Betrachten wir wieder ein Gitter mit Gitterpunkten (t_j, S_i) und schreiben kurz $V_i^j = V(t_j, S_i)$, so erhalten wir mit Vorwärtsdifferenzen in der Zeit

$$\frac{1}{s}(V_i^{j+1} - V_i^j) + \frac{\sigma^2 S_i^2}{2h^2}(V_{i+1}^j - 2V_i^j + V_{i-1}^j) - rV_i^j + \frac{rS_i}{2h}(V_{i+1}^j - V_{i-1}^j) = O(s + h^2)$$

und mit Rückwärtsdifferenzen in der Zeit (mit verschobenem Zeitargument)

$$\frac{1}{s}(V_i^{j+1} - V_i^j) + \frac{\sigma^2 S_i^2}{2h^2}(V_{i+1}^{j+1} - 2V_i^{j+1} + V_{i-1}^{j+1}) - rV_i^{j+1} + \frac{rS_i}{2h}(V_{i+1}^{j+1} - V_{i-1}^{j+1}) = O(s + h^2).$$

Mit $1 - \theta$ bzw. θ gewichtet und aufsummiert ergibt sich

$$\begin{aligned} & \frac{1}{s}(V_i^{j+1} - V_i^j) \\ & + (1 - \theta) \left(\frac{\sigma^2 S_i^2}{2h^2}(V_{i+1}^j - 2V_i^j + V_{i-1}^j) - rV_i^j + \frac{rS_i}{2h}(V_{i+1}^j - V_{i-1}^j) \right) \\ & + \theta \left(\frac{\sigma^2 S_i^2}{2h^2}(V_{i+1}^{j+1} - 2V_i^{j+1} + V_{i-1}^{j+1}) - rV_i^{j+1} + \frac{rS_i}{2h}(V_{i+1}^{j+1} - V_{i-1}^{j+1}) \right) = O(s + h^2). \end{aligned}$$

Wie im vorhergehenden Kapitel erhalten wir nun Approximationen $W_i^j \approx V_i^j$, indem wir den $O(s + h^2)$ -Term auf der rechten Seite dieser Gleichung vernachlässigen, also gleich Null setzen, und das entstehende Gleichungssystem

$$\begin{aligned} & \frac{1}{s}(W_i^{j+1} - W_i^j) \\ & + (1 - \theta) \left(\frac{\sigma^2 S_i^2}{2h^2}(W_{i+1}^j - 2W_i^j + W_{i-1}^j) - rW_i^j + \frac{rS_i}{2h}(W_{i+1}^j - W_{i-1}^j) \right) \\ & + \theta \left(\frac{\sigma^2 S_i^2}{2h^2}(W_{i+1}^{j+1} - 2W_i^{j+1} + W_{i-1}^{j+1}) - rW_i^{j+1} + \frac{rS_i}{2h}(W_{i+1}^{j+1} - W_{i-1}^{j+1}) \right) = 0 \end{aligned}$$

iterativ lösen. Auch hier liefert das Crank-Nicolson-Verfahren mit $\theta = 1/2$ die beste Approximation. Weil die Black-Scholes-Gleichung rückwärts von der Endbedingung zur Zeit $t_M = T$ aus berechnet werden muss, muss die Gleichung dafür nach W^j aufgelöst und in Vektorform geschrieben werden (während wir bei der Wärmeleitungsgleichung nach w^{j+1} aufgelöst hatten). Beachte, dass dadurch das Schema mit $\theta = 1$ explizit wird (bei der Wärmeleitungsgleichung war dies für $\theta = 0$ der Fall), während in allen Schemata für $\theta \in [0, 1)$ in jedem Schritt ein Gleichungssystem gelöst werden muss.

Da das räumliche Gebiet $S \in [0, \infty)$ hier nur einseitig unbeschränkt ist, müssen wir dabei nur auf einer Seite "abschneiden", d.h. auf $[0, L]$ für ein $L > 0$ rechnen und den Randwert $V(t, L)$ analog zum vorhergehenden Abschnitt approximativ einsetzen. Die Details der vollständigen Herleitung der zu lösenden Matrix-Gleichung wollen wir hier aus Zeitgründen nicht betrachten.

12.3 Finite Differenzen und Binomialmethode

Wir haben diese Vorlesung in Abschnitt 3.2 mit der Binomialmethode begonnen und in Abschnitt 7.8 gezeigt, dass diese Methode als schwache Approximation der geometrischen Brownschen Bewegung verstanden werden kann. Zum Abschluss der Vorlesung zeigen wir nun, dass die Binomialmethode auch mit der Finiten Differenzenmethode zusammenhängt.

Wir betrachten dazu das Finite-Differenzenschema angewendet auf die Hilfsfunktion $v(t, x) = V(t, S)$ mit $x = \ln(S)$. Ähnlich wie in Abschnitt 11.1 rechnet man nach, dass diese Funktion die Gleichung

$$v_t + \frac{1}{2}\sigma^2 v_{xx} + (r - \frac{1}{2}\sigma^2)v_x - rv = 0$$

erfüllt. Setzen wir nun $w(t, x) = e^{-rt}v(t, x)$, so rechnet man nach, dass w die Gleichung

$$w_t + \frac{1}{2}\sigma^2 w_{xx} + (r - \frac{1}{2}\sigma^2)w_x = 0$$

erfüllt. Ersetzen wir nun die Ableitungen wie im vorherigen Abschnitt durch die Differenzenquotienten und verwenden die Rückwärtsdifferenzen in der Zeit, so erhalten wir Approximationen $\tilde{w}_i^j \approx w(t_j, x_i)$ mittels

$$\frac{1}{s}(\tilde{w}_i^{j+1} - \tilde{w}_i^j) + \frac{1}{2h^2}\sigma^2(\tilde{w}_{i+1}^{j+1} - 2\tilde{w}_i^{j+1} + \tilde{w}_{i-1}^{j+1}) + \frac{r - \sigma^2/2}{2h}(\tilde{w}_{i+1}^{j+1} - \tilde{w}_{i-1}^{j+1}) = 0.$$

Aufgelöst nach \tilde{w}_i^j ergibt sich

$$\begin{aligned} \tilde{w}_i^j &= \tilde{w}_i^{j+1} + \frac{s}{2h^2}\sigma^2(\tilde{w}_{i+1}^{j+1} - 2\tilde{w}_i^{j+1} + \tilde{w}_{i-1}^{j+1}) + \frac{s(r - \sigma^2/2)}{2h}(\tilde{w}_{i+1}^{j+1} - \tilde{w}_{i-1}^{j+1}) \\ &= \left(1 - \frac{s\sigma^2}{h^2}\right)\tilde{w}_i^{j+1} + \left(\frac{s\sigma^2}{2h^2} + \frac{s(r - \sigma^2/2)}{2h}\right)\tilde{w}_{i+1}^{j+1} + \left(\frac{s\sigma^2}{2h^2} - \frac{s(r - \sigma^2/2)}{2h}\right)\tilde{w}_{i-1}^{j+1}. \end{aligned}$$

Mit der Wahl $h = \sigma\sqrt{s}$ folgt

$$\begin{aligned} \tilde{w}_i^j &= \left(\frac{s\sigma^2}{2h^2} + \frac{s(r - \sigma^2/2)}{2h}\right)\tilde{w}_{i+1}^{j+1} + \left(\frac{s\sigma^2}{2h^2} - \frac{s(r - \sigma^2/2)}{2h}\right)\tilde{w}_{i-1}^{j+1} \\ &= \left(\frac{1}{2} + \frac{\sqrt{s}(r - \sigma^2/2)}{2\sigma}\right)\tilde{w}_{i+1}^{j+1} + \left(\frac{1}{2} - \frac{\sqrt{s}(r - \sigma^2/2)}{2\sigma}\right)\tilde{w}_{i-1}^{j+1} \\ &= p\tilde{w}_{i+1}^{j+1} + (1-p)\tilde{w}_{i-1}^{j+1} \end{aligned}$$

mit

$$p = \frac{1}{2} + \frac{\sqrt{s}(r - \sigma^2/2)}{2\sigma}.$$

Für $\tilde{V}_i^j \approx V(t_j, S_i)$ erhalten wir wegen $V(t_j, S_i) = e^{rt}w(t_j, x_i)$ mit $S_i = e^{x_i}$ die Vorschrift

$$\tilde{V}_i^j = e^{-rs} \left(p\tilde{V}_{i+1}^{j+1} + (1-p)\tilde{V}_{i-1}^{j+1} \right).$$

Dies ist genau die Iterationsvorschrift der Binomialmethode. Beachte, dass wir hier kein Binomialmodell mit $p = 1/2$ herausbekommen.

Zur Berechnung der zugehörigen Dynamik des Binomialmodells

$$f(S, Z) = S \cdot Z \quad \text{mit} \quad \mathbb{P}_Z(\alpha_u) = p \quad \text{und} \quad \mathbb{P}_Z(\alpha_d) = 1 - p,$$

müssen wir noch die Werte α_u und α_d bestimmen. Da sich die Werte für \tilde{V}_i^j aus V_{i+1}^{j+1} und V_{i-1}^{j+1} ergeben, muss $f(S_i, Z) = S_{i+1}$ mit Wahrscheinlichkeit p und $f(S_i, Z) = S_{i-1}$ mit Wahrscheinlichkeit $1 - p$ gelten. Wegen $x_i = hi$ gilt

$$S_i = e^{hi} = e^{\sigma i \sqrt{s}}$$

und damit

$$S_{i+1} = e^{\sigma \sqrt{s}} S_i \quad \text{und} \quad S_{i-1} = e^{-\sigma \sqrt{s}} S_i.$$

Folglich gilt $\alpha_u = e^{\sigma \sqrt{s}}$ und $\alpha_d = e^{-\sigma \sqrt{s}}$. Für den Erwartungswert dieses Modells gilt

$$\begin{aligned} E(f(S, Z)) &= pe^{\sigma \sqrt{s}} S + (1-p)e^{-\sigma \sqrt{s}} S \\ &= \underbrace{p\left(1 + \sigma \sqrt{s} + \frac{\sigma^2 s}{2}\right) S + (1-p)\left(1 - \sigma \sqrt{s} + \frac{\sigma^2 s}{2}\right) S}_{=(1+sr)S} + O(s^{3/2}) \\ &= (1 + sr)S + O(s^{3/2}) = e^{sr} S + O(s^{3/2}), \end{aligned}$$

d.h. bis auf einen Fehlerterm der Ordnung $s^{3/2}$ von der für die risikoneutralen Bewertung benötigten Form.

Der wesentliche Unterschied der Binomialmethode zur Finiten Differenzenmethode besteht darin, dass die Binomialmethode stets nur die Werte in den Knoten (t_j, S_i) berechnet, die zur Berechnung des gewünschten Wertes in $(0, S_{i_0})$ unbedingt benötigt werden, also diejenigen, die in dem entsprechenden Binomialbaum liegen. Im Gegensatz dazu liefert die Finite Differenzenmethode Werte für alle vorgegebenen Gitterknoten. Abbildung 12.2 veranschaulicht diese "Einbettung" des Binomialbaums in das Finite Differenzengitter schematisch.

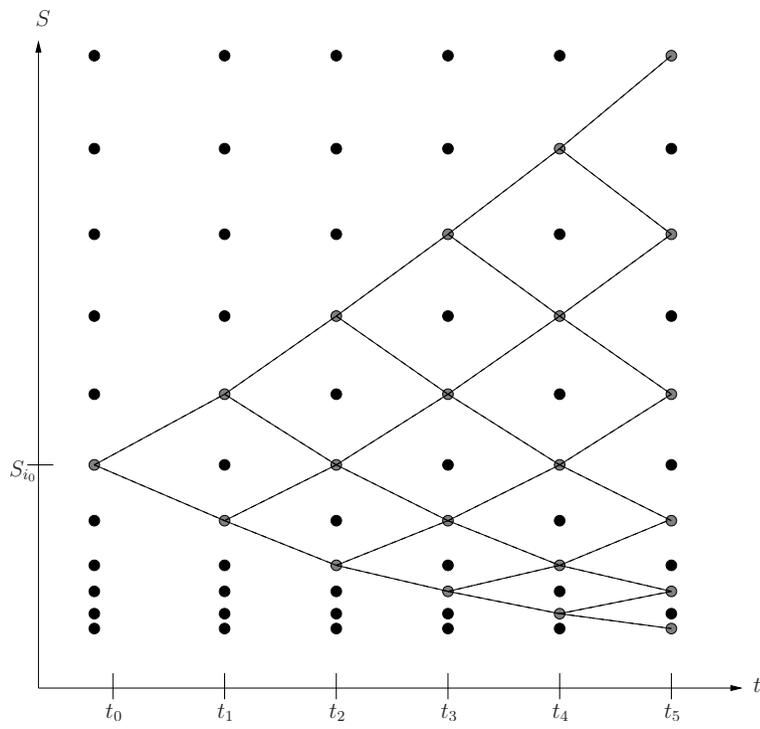


Abbildung 12.2: Binomialbaum und Finite Differenzengitter

Literaturverzeichnis

- [1] M. GÜNTHER AND A. JÜNGEL, *Finanzderivate mit MATLAB®*, Vieweg-Verlag, Braunschweig, 2003.
- [2] D. J. HIGHAM, *An introduction to financial option valuation*, Cambridge University Press, Cambridge, 2004. Mathematics, stochastics and computation.
- [3] P. E. KLOEDEN AND E. PLATEN, *Numerical Solution of Stochastic Differential Equations*, Springer-Verlag, Heidelberg, 1992. (3rd revised and updated printing, 1999).
- [4] R. U. SEYDEL, *Tools for computational finance*, Universitext, Springer-Verlag, Berlin, fourth ed., 2009.

Index

- Ableitung
 - mehrfach, 76
- Abtastfeedback, 96
- Abzinsung, 4
- Anfangsbedingung, 44
- Anfangswert, 44
- Anfangswertproblem, 44
- Anfangszeit, 44
- antithetische Zufallsvariable, 41
- Approximation, 47
 - schwach, 66
 - stark, 66
- Arbitrage, 3
- Arbitragefreiheit, 3
- Ausübungspreis, 2, 7
- Ausübungszeitpunkt, 2, 7
- autonome Differentialgleichung, 44, 47

- Basiswert, 2, 7
- Bedingungsgleichungen, 64
- Binomialmodell, 20
 - als schwache Approximation, 86
 - Parameter, 23
 - und Finite Differenzen, 131
- Binomialbaum, 20
- Black-Scholes Gleichung, 92
 - und Wärmeleitungsgleichung, 116
- Brownsche Bewegung, 26
 - geometrisch, 32
- Butcher-Tableau, 62

- Call-Option, 2, 7
- Crank-Nicolson-Verfahren, 121, 126

- Derivat, 7
- Differentialgleichung
 - gewöhnlich, 43
 - stochastisch, 25
- Differenzenquotient, 117, 129

- Diffusion, 29
- Dirichlet-Randbedingungen, 119
- Diskontierung, 4
- diskrete l_2 -Norm, 126
- diskrete Approximation, 47
- Drift, 29

- Einschrittverfahren, 48
 - Grundalgorithmus, 50
 - Konvergenzsatz, 52
 - schematisch, 55
 - stochastisch, 71
 - Grundalgorithmus, 71
 - Konvergenzeigenschaften, 80
- Euler-Maruyama, *siehe* Euler-Verfahren, stochastisch
- Euler-Verfahren, 49, 62
 - stochastisch, 71, 82
 - Konvergenzeigenschaften, 72
- Eulersche Polygonzugmethode, 49

- fairer Preis, 3
- Feedback, 95
- Fehler
 - global, 55
 - lokal, 55
- Finite Differenzen, 117
 - und Binomialmethode, 131

- geometrische Brownsche Bewegung, 32
- Gerschgorin
 - Satz von, 121
- gewöhnliche Differentialgleichung, 43
- Gitter, 47
- Gitterfunktion, 47
- globaler Fehler, 55
- grafische Darstellung
 - als Graph, 46
 - als Kurve, 46

- Griechen, 100
- Hedging, 3, 9, 93, 95
 - Simulation, 96
 - zeitdiskret, 96
- Heun-Verfahren, 50, 62
- Hierarchische Menge, 77
- historische Parameter, 24
- implizite Parameter, 24
- Inkrement des Wiener Prozesses, 26
- Itô-Integral, 29
 - mehrfach, 75
- Itô-Lemma, 31
- Kennzahlen, 100
- Konfidenzintervall, 39
- Konsistenz, 51, 123
 - einfache Bedingung, 51
- Konsistenzanalyse, 58
- Konsistenzordnung, 51
- Konvergenz, 126
- Konvergenzordnung, 48
- Kozykluseigenschaft, 45
- Kursverlaufsmodel, 32
- Lösungskurve, 44
- Lösungstrajektorie, 44
- Leerverkauf, 10, 12
- Lipschitzbedingung, 51
- lokaler Fehler, 55
- Maple, 59
- Marktmodell, 3
- Mehrfachableitung, 76
- Milstein-Verfahren, 83
- Monte-Carlo-Methode, 37, 101
 - Algorithmus, 38
 - Konvergenzgeschwindigkeit, 39
 - Schätzer, 106
- Multi-Index, 75
- Newton-Verfahren, 110
 - Startwert, 111
- Normalverteilung, 25
 - Verteilungsfunktion, 38
- Option, 2, 7
 - amerikanisch, 2, 7
 - asiatisch, 101
 - Barriere, 115
 - Call, 2, 7
 - europäisch, 2, 7
 - Power, 115
 - Put, 2, 7
- Optionsbewertung, 3
 - fair, 8
 - Grundalgorithmus, 11
 - risikoneutral, 3, 8, 11, 16, 89, 93
- Pfad, 19, 26
- Portfolio, 8, 12
 - Black-Scholes, 90
 - replizierend, 9, 93
 - selbstfinanzierend, 90
- Put-Call-Parität, 13
- Put-Option, 2, 7
- Randbedingung
 - numerisch, 127
- Randbedingungen
 - Dirichlet, 119
- Randwert, 92
- rekombinierbar, 20
- Restmenge, 77
- risikoavers, 11
- risikofreudig, 11
- risikoneutral, *siehe* Optionsbewertung, risikoneutral
- risikoneutrales Maß, 17
- Rücklaufquote, 32
- Rückwärts-Differenzenverfahren, 121
- Runge-Kutta-Verfahren
 - explizit, 62
 - klassisch, 62
- Schätzer
 - erwartungstreu, 106
 - EWMA, 109
 - Monte-Carlo, 106
- Schrittweite, 47
- schwache Approximation, 66
- selbstfinanzierend, 90, 96
- Stabilität, 124
 - unbedingt, 126

- Stabilitätsbedingung, 51, 125
- starke Approximation, 66
- stochastisches Integral, 28
- stochastische Differentialgleichung, 25
 - Itô, 29
 - Stratonovich, 30
- stochastischer Prozess, 19, 26
- stochastisches dynamisches System, 19

- Taylor-Entwicklung, 55
 - stochastisch, 77
- Taylor-Verfahren, 56
 - stochastisch, 79
 - schwache Ordnung 2, 84
- Trajektorie, 44
- Trapez-Regel, 50

- Varianzreduktion, 40
- Vektorfeld, 43
- Volatilität, 17, 23, 32
 - historisch, 105
 - implizit, 110
- Vorwärts-Differenzenverfahren, 121

- Wärmeleitungsgleichung, 116
- Wiener Prozess, 26
 - schwache Approximation, 69
 - starke Approximation, 68

- Zufallsvariable
 - antithetisch, 41
- Zufallszahlen, 37