

# Vorlesungs-Skript

## Numerische Finanzmathematik

Rüdiger Seydel



Universität zu Köln  
Mathematisches Institut

2015

Version 2015

Dieses Skriptum kann heruntergeladen werden von  
<http://www.mi.uni-koeln.de/~seydel/skript.html>

© R. Seydel, Köln 2007, 2009, 2011–2016

seydel@mi.uni-koeln.de  
[www.compfin.de](http://www.compfin.de)

**CCFinance**

Cologne Computational Finance Skriptum

# Vorwort

Dieses Skriptum spiegelt in wesentlichen Zügen den Inhalt meiner Vorlesung *Einführung in die Numerische Finanzmathematik* wider.

Die Themenauswahl entspricht einer Grundausbildung. Die Vorlesung und dieses Skript versuchen, die Thematik in großer Breite darzustellen, einschließlich stochastischer Grundlagen, numerischer Prinzipien und algorithmischer Ausführung, jeweils mit Hinweisen auf den Finanz-Hintergrund. Die Betonung liegt auf wichtigen Ideen. Das Skript bildet den Kern der Veranstaltung ab. Zwangsläufig geht die Kürze des einführenden Skriptums bei gleichzeitiger Breite des Stoffs auf Kosten der Tiefe.

Die Erfahrung des Dozenten mündet in Rat, zusätzlichen Erklärungen und Details, vielen Skizzen und wertenden Kommentaren. Solche Ergänzungen werden während der Vorlesung gegeben und finden sich nur selten im Skript. Deswegen ersetzt die Lektüre des Skriptums nicht den Besuch der Vorlesung.

Der Grad der Detaillierung des Skripts ist unterschiedlich. Die einer Kernkompetenz zugeordneten Methoden werden ausführlicher hergeleitet. In anderen Teilen des Skripts werden “nur” Ideen vorgestellt und Begriffe erläutert. Der Stil versucht, die Ideen und Methoden nicht zu formalistisch aufzubereiten. Wo immer möglich und sinnvoll, wird der natürlichen Reihenfolge *Herleitung* → *Resultat* entsprochen.

In diesem Skript werden zwar wichtige Finanzbegriffe erläutert, das Skript ersetzt aber keinesfalls eine gründliche Beschäftigung mit den finanzwirtschaftlichen Hintergründen. Wesentliche Elemente der Vorlesung mit einigen Herleitungen sind als Übungsaufgaben formuliert. Deswegen gehören die Übungen als wichtiger Bestandteil der Veranstaltung dazu. Sie sind in Englisch formuliert, um den Übergang zu dieser den Finanzbereich dominierenden Sprache zu erleichtern.

Diese Lehrveranstaltung wurde viele Jahre lang jährlich an der Universität zu Köln gelesen, vorher in Ulm, und nachher in Frankfurt und München. Die Inhalte variierten jeweils etwas. Erst 2007 wurde versucht, den Kern meiner einführenden Vorlesung auch als Skript zu fixieren. Die erste Version des Skripts basierte auf einer Vorlesungsmitschrift, sie wurde später mehrmals überarbeitet, insbesondere 2014. Dank für Korrekturlesen und für Verbesserungsvorschläge gebührt unter anderem Pascal Heider, Philipp Immenkötter, Christian Jonen, Markus Lücking und Frank Niekisch.

Köln, Juli 2014

Rüdiger Seydel

# Inhaltsverzeichnis

<b>Prolog: Derivate</b> .....	1
<b>Kapitel 1: Modellierung von Finanzoptionen</b> .....	7
1.1 Optionen .....	7
1.2 Mathematische Modelle .....	11
1.3 Binomial-Methode .....	14
1.4 Stochastische Prozesse .....	19
1.5 Stochastische Differentialgleichungen .....	23
1.6 Risikoneutrale Bewertung .....	30
<b>Kapitel 2: Berechnung von Zufallszahlen</b> .....	35
2.1 Gleichverteilte Zufallszahlen .....	35
2.2 Zufallszahlen mit anderen Verteilungen .....	40
2.3 Normalverteilte Zufallszahlen .....	44
2.4 Korrelierte Zufallszahlen .....	46
2.5 Zahlenfolgen niedriger Diskrepanz .....	48
<b>Kapitel 3: Monte-Carlo-Methoden</b> .....	51
3.1 Approximations-Fehler .....	52
3.2 Konstruktion von Integratoren für SDEs .....	54
3.3 Monte-Carlo-Methoden bei europäischen Optionen .....	58
3.4 Monte-Carlo-Methoden bei amerikanischen Optionen .....	63
<b>Kapitel 4: Finite Differenzen für Amerikanische Standard-Optionen</b> .....	69
4.1 Vorbereitungen .....	69
4.2 Grundlagen von Finite-Differenzen-Methoden .....	70
4.3 Crank-Nicolson-Methode .....	75
4.4 Randbedingungen .....	77
4.5 Vorzeitiges Ausüben — Freie Randwertprobleme .....	79
4.6 Komplementarität .....	83
4.7 Numerische Realisierung .....	87
<b>Ausblick</b> .....	92
<b>Übungen</b>	

# 0. (Prolog) Derivate

## A. Übersicht

Wesentliche Kapitalanlagen und Vermögenswerte (Assets) sind

Aktien	<i>equities, stocks</i>
Anleihen	<i>bonds</i>
Rohstoffe	<i>commodities</i> (Waren/handelbare Rohstoffe, z.B. Edel- u. Industriemetalle, Agrar-Produkte, ...)

Kassa-Markt (*spot market*): Kauf und Abwicklung sofort (“heute”)

### Termin-Geschäft

Vertrag heute ( $t = 0$ )

Abwicklung in der Zukunft (z.B. für  $t = T > 0$ ,  $T$  “Fälligkeit”, in Jahren)

Bezeichnung:  $S_t$  (oder  $S(t)$ ) sei der *Marktpreis* des Anlage-Objektes zum Zeitpunkt  $t$ . Das Anlage-Objekt heißt hier auch Asset, Basiswert oder *underlying*.

### Beispiele für Risiken

- 1.) Ein Aktionär will in einem halben Jahr ein Paket von Aktien verkaufen, deren heutiger Wert 50 Euro pro Stück beträgt. *Risiko: Crash!* Ziel/Wunsch: in 6 Monaten sollen die Aktien mindestens 50 Euro pro Stück einbringen! ( $T = 0.5$ )
- 2.) Ein Landwirt erwartet in drei Monaten (im Juli) eine Ernte von 800 Tonnen Mais. *Risiken:* Missernte, Preisverfall. Ziel/Wunsch: Er will diese Ernte heute schon verkaufen, und erst Ende Juli ( $T = 0.25$ ) liefern.  
Frage: Welcher Preis ist **heute** der “richtige”?
- 3.) Kaufe heute die Ladung eines Öltankers in Dollar, der erst in 2 Wochen in Rotterdam entladen werden kann und dann in Euro verkauft werden soll. Preis? *Risiken:* pünktliche Ankunft, Preisverfall, Währungsrisiko.
- 4.) Ein Unternehmen hat einen Kredit mit variablen Zinszahlungen aufgenommen, Zinsrate  $r_v(t)$ . *Risiko:* Zinszahlungen können steigen. Ziel: Absicherung

u.s.w. !

**Problem jeweils:** Der (über Gewinn/Verlust) entscheidende zukünftige Wert  $S_T$  ist heute ( $t = 0$ ) unbekannt. Man kennt heute lediglich  $S_0$  und die “gegenwärtige” Zinsrate, und eventuell weitere Daten von  $t \leq 0$ .

Aber:

Der Vertrag wird heute mit Angabe des zum Zeitpunkt  $T$  zu zahlenden Preises  $F$  abgeschlossen.

Frage z.B.:

Was ist der faire Preis  $F$ , so dass heute ein solches Termingeschäft zustande kommt?  
(Dies zieht weitere Fragen nach sich, z.B. wie sich “der Zins” entwickeln wird.)

Die meisten dieser Risiken lassen sich mit **Derivaten** absichern. (Das sind Finanzprodukte, die von den Basiswerten “abgeleitet” werden.) Drei große Klassen von Derivaten sind:

Futures/Forwards

Optionen

Swaps

“Zertifikate” bauen auf Optionen auf.

Praktische Frage: Bewertung solcher Derivate!?

“Financial Engineering”

Für die Bewertung von Futures/Forwards und Swaps reicht bei den Rechenmethoden üblicherweise relativ einfache Mathematik aus. Bei Optionen sind jedoch anspruchsvolle **numerische Näherungsverfahren** erforderlich.

**Vorlesung:** im Wesentlichen Numerik für Optionen.

Es werden benötigt:

Grundlagen der Stochastik

Grundlagen der Differentialgleichungen, auch elementare Partielle Dgl.

Numerik I (und II)

## Inhalt

Kapitel 1: Modellierung

Kapitel 2: Zufallszahlen

Kapitel 3: Monte Carlo Verfahren

Kapitel 4: Standardmethoden für Standardoptionen

(Kapitel 5: weitere Verfahren, insbesondere für exotische Optionen)

## B. Forwards und Arbitrage

### Definition:

Ein **Forward** ist ein Vertrag zwischen Vertragspartnern  $A$  und  $B$ , abgeschlossen “heute” zum Zeitpunkt  $t = 0$ , wonach zum Zeitpunkt  $T$  der Partner  $A$  eine bestimmte Menge eines Assets (heutiger Marktpreis  $S_0$ ) liefert und  $B$  dafür den Preis  $F$  zahlt.

### Behauptung

Falls für  $0 \leq t \leq T$  das Asset weder Kosten verursacht noch Erträge erzielt, und wenn  $r$  “die” Zinsrate einer risikofreien Anlage zur Laufzeit  $T$  ist, dann ist (unter geeigneten weiteren Voraussetzungen)

$$F = S_0 e^{rT}$$

der angemessene Preis.

Der Beweis von  $F = S_0 e^{rT}$  erfolgt mit dem No-Arbitrage-Prinzip.

**Definition: Arbitrage** = risikofreier Profit

oder konkreter:

Arbitrage ist die Existenz eines Portfolios, dessen Errichtung zu Beginn kein Kapital erfordert, und das bis zur Fälligkeit garantiert keinen Verlust macht, aber mit positiver Wahrscheinlichkeit einen Gewinn abwirft.

In einem (fiktiven) idealen Markt verbreitet sich Information so schnell, dass Arbitrage im Nu erkannt wird und wegen des Mechanismus "Angebot und Nachfrage" schnell verschwindet.

**General-Annahme: Es gibt keine Arbitrage!** (*no-arbitrage principle*)

weitere Annahmen:

- 1.) Zinses-Zins-Verzinsung (Faktor  $e^{rT}$ )
- 2.) Kreditzins = Guthabenzins

**Beweis von  $F = S_0 e^{rT}$ :**

1. Annahme  $F > S_0 e^{rT}$ .

Dann gibt es eine Arbitrage-Strategie wie folgt:

$t = 0$ : leihe Geldbetrag  $S_0$  zur Zinsrate  $r$ ,  
 kaufe den Handelsgegenstand zum Preis  $S_0$ ,  
 schließe Forward-Kontrakt ab, wonach für  $t = T$   
 das Asset zum Preis  $F$  verkauft wird.

$t = T$ : verkaufe gemäß Vertrag:  $+F$   
 zahle Kredit zurück:  $-S_0 e^{rT}$   
 Ergebnis  $F - S_0 e^{rT} > 0$ ,  
 d.h. risikofreier Gewinn, also Arbitrage!

Widerspruch zum No-Arbitrage-Prinzip.  $\Rightarrow F \leq S_0 e^{rT}$

2. Annahme  $F < S_0 e^{rT}$ .

$t = 0$ : Besitzer des Handelsgegenstandes verkaufen ihn:  $+S_0$   
 (Nicht-Besitzer: Leerverkauf)  
 lege  $S_0$  an zum Zinssatz  $r$ ,  
 und steige in Forward-Kontrakt ein, das Asset  
 zum Preis  $F$  zu kaufen.

$t = T$ : vom Vertrag:  $-F$   
 von Geldanlage:  $+S_0 e^{rT}$   
 Ergebnis  $S_0 e^{rT} - F > 0$  also Arbitrage

Widerspruch!  $\Rightarrow F \geq S_0 e^{rT}$

3. Zusammen:  $F = S_0 e^{rT}$

Bemerkung: Analog bei anderen Zinsmodellen. Bei zwei Zinssätzen:  $r_{\text{kredit}}$ ,  $r_{\text{guthaben}}$ , erhalte  $F \leq S_0 e^{r_{\text{kr}}T}$ ;  $F \geq S_0 e^{r_{\text{gut}}T}$ ; also:  $S_0 e^{r_{\text{gut}}T} \leq F \leq S_0 e^{r_{\text{kr}}T}$

## C. Zins

Es sei  $N(t)$  der Zeitwert einer Geld-Anlage. Bei

$$N(t + \Delta t) - N(t) = r(t, \Delta t) N(t) \Delta t$$

ist  $r(t, \Delta t)$  die Zinsrate zum Zeitpunkt  $t$  mit Anlagedauer  $\Delta t$ . Der **Diskontierungsfaktor** für einen Anlagezeitraum von  $t$  bis  $T$  ist

$$Z_r(t, T) := \frac{N(t)}{N(T)}$$

Für zeitstetige Betrachtung ( $t \in \mathbb{R}$ ) und integrierbare Momentan-Zinsrate  $r$  folgt mit  $\Delta t \rightarrow 0$ :

$$N(T) = N(t) \exp \int_t^T r(x) dx$$

Im Spezialfall  $r$  konstant bzgl.  $t$  (wird hier angenommen) folgt

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

**risikofreier Zins:** Wenn die Anleihe risikofrei ist, dann nennt man  $r$  “risikofreie” Zinsrate.

$Z$  wird aus Marktdaten gewonnen, speziell von Anleihen welche keinen Kupon auszahlen (*zero-bond*), daraus dann  $r(t, T - t)$ .

Für das Folgende sei verabredet:  $r$  meint immer den zum jeweiligen Anlagehorizont  $T - t$  passenden risikofreien Zins.

## LITERATUR

*deutschsprachiges Lehrbuch zu dieser Vorlesung:*

R. Seydel: Numerische Berechnung von Finanzderivaten. Zweite Auflage. Springer, 2016

*oder das Englisch-sprachige Buch*

R. Seydel: Tools for Computational Finance. 6th Ed., Springer, London 2017

*In diesem Buch ist der Inhalt der Vorlesung ausführlicher erklärt, versehen mit Literaturangaben und Ergänzungen, alles in der für dieses Gebiet entscheidenden Sprache Englisch.*

Jede Studentin und jeder Student sollte sich als Grundlage für eine einschlägige berufliche Arbeit weiteres Finanzwissen aneignen!

*Hintergrundwissen zu Derivaten:*

J.C. Hull: Options, Futures, and Other Derivates. Prentice Hall.

Weitere Spezial-Literatur zu einzelnen Themen wird in den jeweiligen Kapiteln angegeben.

Farbige Illustrationen werden in

*Topics in Computational Finance*

angeboten unter [www.compfin.de](http://www.compfin.de). Diese Illustrationen dienen der Ergänzung und der Vertiefung.

## D. Praxis

**(Standard-)Optionen** sind hoch-standardisierte Produkte, die an Börsen angeboten werden. Auf Grund der hohen Standardisierung (Underlying, Typ, Strike oder "Basispreis"  $K$ , Laufzeit  $T$ ) haben sie keine Wertpapierkennnummer (WKN). So ist etwa die Diskretisierungsstufe im Strike klar und einheitlich geregelt. Der Halter kann auch als Emittent auftreten (short).

**Optionsscheine** (*warrants*) dagegen werden von Banken und Unternehmen begeben, die Besitzer des Underlying sind. Die Optionsscheine können individuell gestaltet sein und sind Wertpapiere (mit WKN). Der Emittent bestimmt den Preis, gehandelt werden diese z.B. bei der Euwax, am Kassamarkt nur long Positionen [auch Zertifikate am Kassamarkt].

Was passiert, wenn eine Option ausgeübt wird? Wo kommen die Aktien her? Das wird alles vom *Market Maker* organisiert. Es werden Aktien genommen, die auf dem Markt gerade angeboten werden! Auswahl: zuerst wird mit einem Zufallsgenerator ein Broker ausgewählt. Bei dem gewählten Broker steht ein weiterer Zufallsgenerator, der dann einen Kunden mit passendem Aktien-Angebot auswählt. Es wird eher selten wirklich ausgeübt; angeblich nimmt der Market Maker rückverkaufte Optionen auch wieder vom Markt. Bei manchen risikoträchtigen Optionen und Optionsstrategien werden *Margins* verlangt.

Fälligkeit von Optionen immer am 3. Freitag im Monat. Deswegen wird nur der Monat angegeben. Die Quartalsmonate sind März, Juni, September, Dezember ( $\rightarrow$  Hexensabbat).

Bei Optionen werden *Kontrakte* gehandelt, d.h. im Allgemeinen eine Option auf mehrere Stücke des Underlyings. Bei den folgenden Betrachtungen und Berechnungen ist ein 1:1 Bezugsverhältnis vorausgesetzt.

An der Eurex sind Optionen auf Aktien immer vom amerikanischen Typ, solche auf Indices sind vom europäischen Typ.

Optionen auf Indices dotieren nicht auf Euro, sondern auf Punkte.

Unterschied DAX/Eurostoxx: Der DAX ist ein *performance* index, in dem auch Dividenden berücksichtigt werden. Dagegen ist der Eurostoxx ein *Preis*-Index (ohne Dividenden), ebenso der Dow Jones.

Es gibt vier Positionen: long call, short call, long put, short put.

long call	=	Kauf einer Kaufoption
short call	=	Verkauf einer Kaufoption
long put	=	Kauf einer Verkaufsoption
short put	=	Verkauf einer Verkaufsoption
		(am Kassamarkt gibt es solches Produkt: die Aktienanleihe)

**Innerer Wert:** call:  $(S - K)^+$ , put:  $(K - S)^+$  (für  $t \leq T$ )

(oft auch  $S - K$  bzw.  $K - S$ , d.h. kann negativ werden)

### Zeitwert

Preis der Option abzüglich |innerem Wert|.

*in the money*: innerer Wert positiv

*out of the money*: wenn kein innerer Wert

Der Zeitwertverlust ist besonders drastisch zum Ende der Laufzeit.

Die **Volatilität** ist hoch bei starken Kursbewegungen.

kurze Laufzeiten: etwa 1 – 2 Monate

“lange” Laufzeiten: 3 – 6 Monate, oder länger

### **Internet**

[www.boerse-frankfurt.com](http://www.boerse-frankfurt.com) (dort z.B. Aktien/Realtime)

[www.eurexchange.com](http://www.eurexchange.com) (dort z.B. /education/tools)

[www.boerse.de](http://www.boerse.de) (dort z.B. eurex)

[www.compfin.de](http://www.compfin.de) (zu Computational Finance)

[www.onvista.de](http://www.onvista.de) (aktuelle Daten)

[www.gs.de](http://www.gs.de) (aktuelle Daten)

# 1. Modellierung von Finanzoptionen

## 1.1 Optionen

### Definition (Option)

Eine *Option* ist das *Recht* (nicht aber die Verpflichtung) ein risikobehaftetes Asset zu einem spezifizierten Preis  $K$  bis zu einem Fälligkeitszeitpunkt  $T$  zu kaufen (“Call”) oder zu verkaufen (“Put”).

Die Option ist ein Vertrag über ein Geschäft, dessen Bedingungen vom Emittenten festgelegt werden (engl. *writer*). Der Erwerber der Option heißt *Halter*; er zahlt bei Erwerb der Option eine Prämie  $V$ .

Das Realisieren der Option, d.h. das Kaufen bzw. Verkaufen des *underlying* Asset (Basiswert) unter Verwendung des Options-Vertrages heißt ausüben (*exercise*). Der Basiswert kann auch ein Index sein.

**Frage:** Was ist der faire Preis  $V$ ?

Er hängt ab vom Preis  $K$ , vom Preis  $S_0$ , von  $T$  sowie von weiteren Marktdaten, wie dem Zinssatz  $r$  und der Volatilität  $\sigma$ .

### Begriffe

Die *Volatilität*  $\sigma$  ist ein Maß für die Schwankungsbreite des Assetpreises  $S_t$  und damit für das Risiko.

$T$ : Laufzeit der Option, *maturity date*, Fälligkeitsdatum

$K$ : Ausübungspreis, Basispreis, *strike*, *exercise price*

Eine *europäische* Option darf nur am Ende der Laufzeit ( $t = T$ ) ausgeübt werden, eine *amerikanische* Option zu jedem Zeitpunkt  $0 \leq t \leq T$ .

### Beispiel (Option auf den DAX)

Angenommenes Szenario: Ein Anleger besitzt “heute” (10.04.2007; DAX 7166 Punkte) ein Depot mit deutschen Standardaktien. Bei einem DAX-Wert von 7000 ist das Depot rund 100000 Euro wert. Der Anleger will sein Portfolio gegen Wertverlust absichern. (“Wertverlust” im Sinne  $\text{DAX} < 7000$ )

Der Markt bietet am 10.04.2007 laut FTD folgenden Put auf den Dax an:

Emittent:	(z.B.) BNP Paribas
Laufzeit:	20.06.2008
Ausübungs“preis” $K$ :	7000 Punkte
Preis $V$ :	3.26 Euro pro Put

Der Emittent zahlt pro DAX-Punkt unter 7000 einen Cent. (Falls der DAX-Wert etwa auf 6000 sinken würde, bekäme der Halter eines Puts bei Ausübung 10 Euro.) Zur

Sicherung seines Portfolios kauft der Anleger 1428 solcher Puts. (Warum 1428 ?) Bei dem Putpreis von 3.26 Euro kostet die Absicherung des Portfolios etwa 4700 Euro.

Zurück zur Frage: Was ist der Wert von  $V$ ?

Ein Aspekt ist leicht zu beantworten, nämlich der *Wert der Option zur Fälligkeit*: Dies ist die **Auszahlungsfunktion** (*payoff*).

### 1. Call in $t = T$

Der Halter der Option hat zwei Alternativen, wenn er das Asset erwerben will:

- (a) Er kauft es am Markt (an der Börse) und zahlt  $S_T$ , oder
- (b) er "kauft" es durch Ausüben der Option und zahlt  $K$ .

Je nach Marktlage sucht der rationale Halter die bessere Alternative aus.

1. Fall:  $S_T \leq K \Rightarrow$  Der Halter kauft am Markt, zahlt  $S_T$  und lässt Option verfallen. Dann ist die Option wertlos, also  $V = 0$ .

2. Fall:  $S_T > K \Rightarrow$  Der Halter "kauft" durch Ausüben und zahlt  $K$ . Er kann sofort am Markt zum Preis  $S_T$  verkaufen. Gewinn:  $S_T - K$ , also  $V = S_T - K$ .

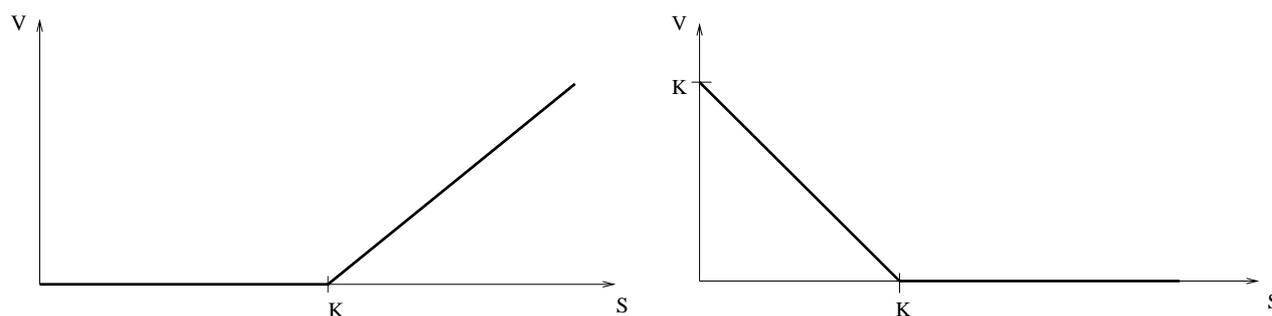
Somit gilt für den Wert des Calls zur Fälligkeit  $T$ :

$$\begin{aligned} V(S_T, T) &= \begin{cases} 0 & \text{falls } S_T \leq K \\ S_T - K & \text{falls } S_T > K \end{cases} \\ &= \max\{S_T - K, 0\} \quad =: (S_T - K)^+ \end{aligned}$$

### 2. Put in $t = T$

Hier führen analoge Überlegungen zu

$$\begin{aligned} V(S_T, T) &= \begin{cases} K - S_T & \text{falls } S_T \leq K \\ 0 & \text{falls } S_T > K \end{cases} \\ &= \max\{K - S_T, 0\} \quad =: (K - S_T)^+ \end{aligned}$$



Auszahlungsfunktion eines Calls (links) und eines Puts (rechts) in  $t = T$ .

Die gleiche Argumentation gilt auch für amerikanische Optionen für beliebiges  $t \leq T$ : Die Payoffs sind

$$\text{Put: } (K - S_t)^+$$

$$\text{Call: } (S_t - K)^+$$

Schwierig ist es, den Wert  $V$  für  $t < T$ , insbesondere für  $t = 0$ , zu bestimmen. Eine zentrale Rolle spielt hier das *No-Arbitrage-Prinzip*. Allein auf Grund dessen lassen sich elementare **Schranken für  $V$**  herleiten. Es folgen einige Beispiele.

Der Wert  $V(S, t)$  einer amerikanischen Option kann niemals kleiner als die Auszahlungsfunktion sein, denn (Beweis für Puts; Calls analog):

Offensichtlich gilt  $V \geq 0$  für alle  $S$ . *Annahme:* Es sei  $S < K$  und  $0 \leq V < K - S$ . Konstruiere nun die folgende Arbitrage-Möglichkeit:

Kaufe das Asset ( $-S$ ) und den Put ( $-V$ ), und übe den Put sofort aus: ( $+K$ ). Wegen  $K > S + V$  ergibt das einen risikolosen Gewinn  $K - S - V > 0$ . Widerspruch zum No-Arbitrage-Prinzip.

Es gilt also

$$V_{\text{Put}}^{\text{am}}(S, t) \geq (K - S)^+ \quad \forall S, t.$$

Analog:

$$V_{\text{Call}}^{\text{am}}(S, t) \geq (S - K)^+ \quad \forall S, t.$$

Zudem gilt

$$V^{\text{am}} \geq V^{\text{eu}},$$

weil eine amerikanische Option die Möglichkeiten einer europäischen Option beinhaltet.

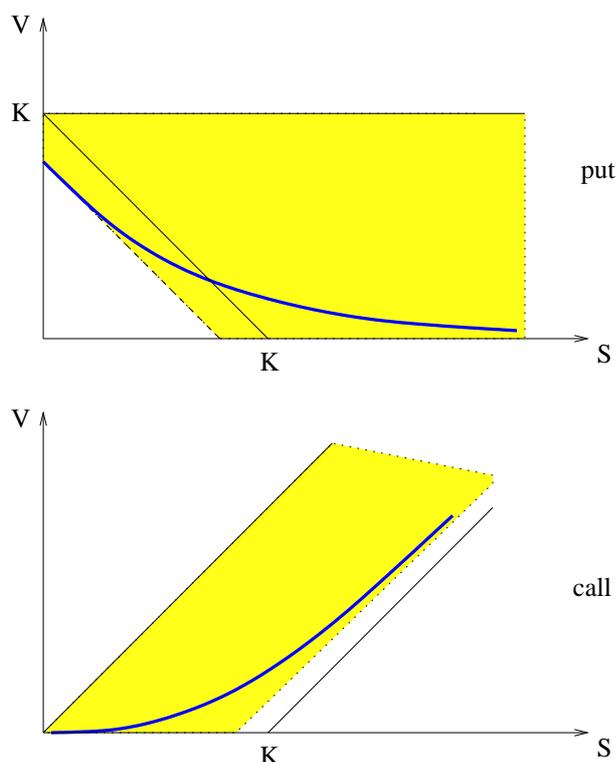
Speziell gilt für europäische Optionen auf ein während der Laufzeit dividendenloses Asset die **Put-Call-Parität**:

$$S + V_{\text{Put}} = V_{\text{Call}} + Ke^{-r(T-t)} \quad (\text{vgl. Übungsaufgaben})$$

Hieraus können weitere Folgerungen getroffen werden, wie z.B.

$$V_{\text{Put}}^{\text{eu}} \geq Ke^{-r(T-t)} - S.$$

Die folgende Grafik fasst diese a-priori-Schranken für europäische Optionen auf dividendenlose Assets zusammen (für  $r > 0$ ).



**Definition (historische Volatilität)**

Die historische Volatilität  $\sigma$  ist die Standardabweichung von  $S_t$ . Da sie mit Returns in Zeitintervallen  $\Delta t$  ermittelt wird, wird skaliert mit  $\frac{1}{\sqrt{\Delta t}}$  (Erklärung in Abschnitt 1.5).

In der Praxis ist  $\sigma$  nicht konstant, wird aber im klassischen Black-Scholes-Modell als konstant angenommen. Die Ermittlung von Markt-konsistenten Parameterwerten (wie  $\sigma$ ) ist i.A. anspruchsvoll. (“Kalibrierung”)

Hinweis zum Verständnis:

Bei jeder Option sind *drei Preise* gekoppelt: der Preis  $S_t$  des zugrunde liegenden Assets, der Strike-Preis  $K$  und der Preis  $V$  der Option.

**Definition**

Die Optionstypen mit den oben angegebenen Payoffs  $\Psi(S) := (K - S)^+$  oder  $\Psi(S) := (S - K)^+$  auf ein einzelnes Asset heißen *Standard-Optionen*, engl. *Vanilla-Options*. Es gibt noch viele andere Arten von Optionen mit anderen Modalitäten. Diese anderen Typen von Optionen heißen *exotisch*.

**Beispiele für exotische Optionen**

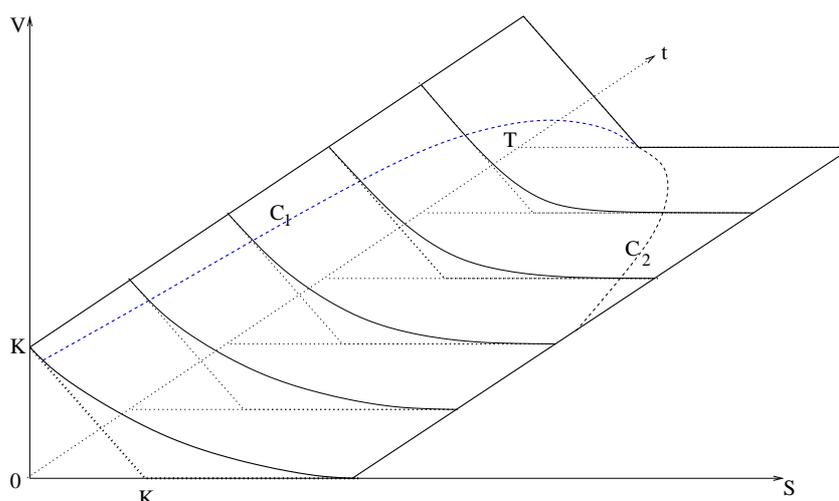
*Basket*: Das Underlying ist ein “Korb” bestehend aus *mehreren* Assets, z.B.  $\sum_{i=1}^m w_i S_i(t)$ , wobei  $S_i$  der Marktpreis des  $i$ -ten Asset ist,  $m > 1$ .

Optionen mit *anderen Payoffs*: z.B. *Binär-Put* mit

$$\text{Payoff} = \begin{cases} 0 & \text{falls } S_t > K \\ 1 & \text{falls } S_t \leq K. \end{cases}$$

*Pfadabhängigkeit*: Z.B. beim Payoff  $(\frac{1}{T} \int_0^T S(t) dt - K)^+$  geht der Mittelwert ein. Er hängt vom Pfad ab, d.h. vom Verlauf  $S(t)$  auf einem Intervall (*average price call*).

*Barrier*: Beispiel: Option wird wertlos, wenn  $S_t$  eine festgelegte “Barriere”  $B$  über- oder unterschreitet.

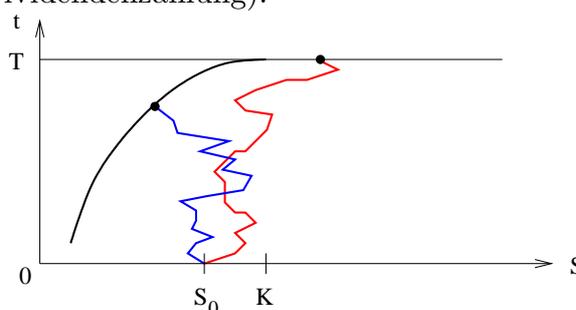


Optionspreissfläche eines amerikanischen Puts, schematisch

## Zur Geometrie der Optionen (Ausblick, Begründungen i.W. in Kapitel 4)

$V(S, t)$  kann als Fläche über dem Halbstreifen  $0 \leq t \leq T, S > 0$  interpretiert werden. Dieses  $V(S, t)$  heißt *Wertfunktion*. Die Fläche  $V(S, t)$  mündet bei amerikanischen Optionen (Illustration für einen Put) in der Ausübungskurve (*early-exercise curve*)  $C^1$ -glatt in die Ebene ein, welche durch den Payoff definiert ist.

**Bedeutung:** Wenn der Marktpreis  $S_t$  diese Kurve  $C_1$  erreicht, ist vorzeitiges/sofortiges Ausüben optimal: lege den Betrag  $K$  an zum Zinssatz  $r$ . Ansicht auf die Entscheidungsebene (Fall eines Puts, keine Dividendenzahlung):



Bei amerikanischen Call-Optionen *mit* Dividendenzahlung analoges Verhalten.

Die Ausübungskurve ist zunächst unbekannt und muss numerisch berechnet werden.

## 1.2 Mathematische Modelle

### A. Black-Scholes-Markt

Wir betrachten nun mathematische Modelle für die Entwicklung von  $S_t$ . Hierzu listen wir eine Reihe von Annahmen auf, welche i.W. auf Black, Scholes und Merton (1973, Nobel-Preis 1997) zurück gehen. Deren Annahmen führen zu einer partiellen Differentialgleichung, nämlich der berühmten **Black-Scholes-Gleichung**

$$\frac{\partial V}{\partial t} + \frac{1}{2} \sigma^2 S^2 \frac{\partial^2 V}{\partial S^2} + r S \frac{\partial V}{\partial S} - r V = 0$$

Diese ist ein Symbol für die klassische Theorie der Optionsmärkte. Jede Lösung  $V(S, t)$  einer Standard-Option muss für  $t = T$  der Endbedingung  $V(S, T) = \Psi(S)$  genügen, wobei  $\Psi$  den Payoff definiert.

### Annahmen des Modells

1. Es gibt keine Arbitragemöglichkeiten.
2. Der Markt ist "reibunglos". Das bedeutet: Gebühren und Steuern entfallen, Guthaben- und Kreditzinsen sind gleich. Zudem seien alle Werte beliebig oft teilbar (d.h.  $\in \mathbb{R}$ ). Ein Kauf oder Verkauf beeinflusst den Marktpreis nicht.
3. Der Marktpreis  $S_t$  folge einer geometrisch-brownschen Bewegung (Erklärung später).
4. Technische Annahmen:

$r$  und  $\sigma$  sind konstant für  $0 \leq t \leq T$ .

Es werden in  $0 \leq t \leq T$  keine Dividenden gezahlt.

(Einige dieser Annahmen können abgeschwächt werden.)

**Unter diesen Annahmen genügt die Wertfunktion  $V(S, t)$  einer europäischen Standard-Option der Black–Scholes-Gleichung.** Also ist ein möglicher numerischer Zugang für eine europäische Option derjenige, der die Black–Scholes-Gleichung löst. Es gibt eine analytische Lösung; sie ist am Ende dieses Kapitels aufgeführt (dort ist  $\delta$  eine kontinuierliche Dividendenrate).

Das obige Modell des Finanzmarktes ist der klassische Zugang, es gibt jedoch auch andere Marktmodelle.

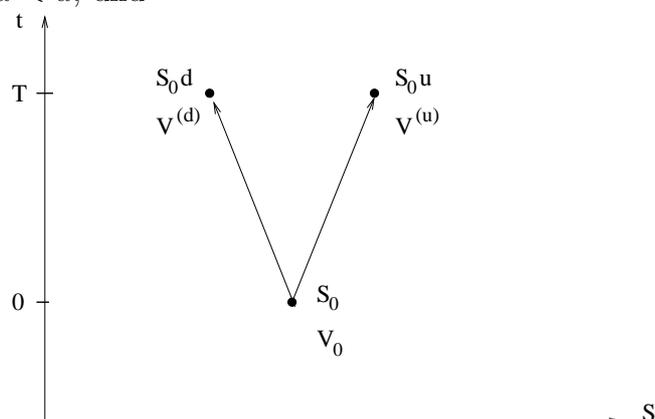
Das Modell mit seiner geometrisch-brownschen Bewegung (Abschnitt 1.5) ist *zeitstetig*, d.h.  $t \in \mathbb{R}$ . Es gibt auch *zeitdiskrete* Modelle des Marktes. Diese lassen nur eine diskrete Auswahl von Zeitpunkten zu. Solche zeitdiskreten Modelle betrachten wir hier i.W. nicht. Andere Marktmodelle sind zwar zeitstetig, verwenden aber die geometrisch-brownsche Bewegung nicht. Solche Modelle sind z.B. Sprungmodelle. (Spezial-Literatur zu zeitdiskreten Modellen z.B. [J. Kremer: Einführung in die diskrete Finanzmathematik]; zu Sprungmodellen z.B. [R. Cont, P. Tankov: Financial Modelling with Jump Processes])

### Numerische Aufgaben:

- Berechnung von  $V(S, t)$  insbesondere für  $t = 0$ , mit early-exercise Kurve bei amerikanischen Optionen,
- Berechnung der Sensitivitäten (“Griechen”), wie  $\frac{\partial V(S, 0)}{\partial S}$ ,
- Kalibrierung, d.h. Berechnung von Parametern der Modelle durch Anpassung an Marktdaten.

## B. Risikoneutrale Wahrscheinlichkeiten (Ein-Perioden-Modell)

Annahmen:  $0 < d < u$ , und



Es existieren nur zwei Zeitpunkte:  $0, T$ , und nur zwei mögliche zukünftige Aktienpreise  $S_0 d, S_0 u$ .  $V_0$  sei der (noch unbestimmte) Wert der Option “heute” für  $t = 0$ , und es sei  $S_0$  der heutige Wert der Aktie. Betrachte ein Portfolio mit zwei Posten:

1.  $\Delta$  viele Stücke der Aktie
2. eine emittierte (verkaufte) Option auf diese Aktie

Für den Wert  $\Pi_0$  des Portfolios zum Zeitpunkt 0 gilt

$$\Pi_0 = S_0 \Delta - V_0.$$

$\Delta$  ist noch offen. Zum Zeitpunkt  $T$  kann der Wert des Underlyings gestiegen (*up*) oder gefallen (*down*) sein. Also gilt für die Entwicklung des Portfolios in  $t = T$

$$\begin{aligned}\Pi^{(u)} &= S_0 u \Delta - V^{(u)} \\ \Pi^{(d)} &= S_0 d \Delta - V^{(d)}.\end{aligned}$$

$V^{(u)}$  und  $V^{(d)}$  sind durch den Payoff bekannt. Wähle nun  $\Delta$  so, dass das Portfolio in  $T$  risikolos ist, d.h. der Portfoliowert soll gleich sein, egal ob der Marktpreis “up” oder “down” geht,

$$\Pi^{(u)} = \Pi^{(d)} =: \Pi_T.$$

Somit gilt

$$\begin{aligned}S_0 \Delta (u - d) &= V^{(u)} - V^{(d)} \\ \text{oder } \Delta &= \frac{V^{(u)} - V^{(d)}}{S_0 u - S_0 d}.\end{aligned}$$

Dieser spezielle Wert von  $\Delta$  stellt das Portfolio risikofrei. Mit dem No-Arbitrage-Prinzip folgt nun:

Eine andere risikofreie Anlage muss den gleichen Wert haben, denn anderenfalls würden Arbitrageure die eine Anlage gegen die andere tauschen. Also:  $\Pi_T = \Pi_0 e^{rT}$

Nach elementarer Zwischenrechnung ergibt dies

$$V_0 = e^{-rT} \left( V^{(u)} q + V^{(d)} (1 - q) \right)$$

mit  $q := \frac{e^{rT} - d}{u - d}$ . Diese Formel hat die Struktur eines Erwartungswertes. Wenn  $0 < q < 1$  (benötige hierzu  $d < e^{rT} < u$ , eine Bedingung welche Arbitragefreiheit garantiert\*), dann wird durch dieses  $q$  eine Wahrscheinlichkeit  $Q$  induziert. Also gilt:

$$V_0 = e^{-rT} \mathbb{E}_Q[V_T]$$

[Denn in einem diskreten Wahrscheinlichkeitsraum mit Wahrscheinlichkeit  $P$  gilt

$$\mathbb{E}_P[X] = \sum_{i=1}^n x_i P(X = x_i),$$

wobei  $X$  eine Zufallsvariable ist.] Die oben definierte spezielle Wahrscheinlichkeit  $Q$  heißt **risikoneutrale Wahrscheinlichkeit**. Für  $S_0$  ergibt sich

$$\mathbb{E}_Q[S_T] = \underbrace{\frac{e^{rT} - d}{u - d}}_{=q} S_0 u + \underbrace{\frac{u - e^{rT}}{u - d}}_{=1-q} S_0 d = S_0 e^{rT}.$$

Also gilt auch

$$S_0 = e^{-rT} \mathbb{E}_Q[S_T].$$

---

\* Was sind die Arbitragestrategien für  $d \geq e^{rT}$  oder  $e^{rT} \geq u$  ?

**Zusammenfassung:**

Falls das Portfolio risikofrei ist (erreicht durch obigen speziellen Wert von  $\Delta$ ) und  $0 < q < 1$  mit  $q = \frac{e^{rT} - d}{u - d}$ , gibt es eine Wahrscheinlichkeit  $Q$ , so dass

$$V_0 = e^{-rT} \mathbf{E}_Q[V_T] \quad \text{und} \\ S_0 = e^{-rT} \mathbf{E}_Q[S_T].$$

Die Größe  $\Delta$  heißt *Delta*. Später wird sich zeigen, dass im zeitstetigen Fall  $\Delta = \frac{\partial V}{\partial S}$  gilt. Dies ist das erste und wichtigste Beispiel der “Greeks”, weitere sind  $\frac{\partial^2 V}{\partial S^2}$ ,  $\frac{\partial V}{\partial \sigma}$ , ... .  $\Delta$  ist der Schlüssel zum “*Delta-Hedgen*”, d.h. Risikobegrenzung oder Eliminierung des Risikos des Emittenten einer Option.

**Bemerkung**

Die Eigenschaft

$$e^{-rT} \mathbf{E}_Q[S_T] = S_0$$

für alle  $T$  ist die *Martingaleigenschaft* des mit dem risikofreien Zins  $r$  diskontierten Prozesses  $e^{-rt} S_t$  bezüglich des Wahrscheinlichkeitsmaßes  $Q$ .

### 1.3 Binomial-Methode

Wir möchten nun Optionen im zeitstetigen Fall bewerten. Für eine numerische Näherung muss in irgendeiner Weise diskretisiert werden. Unter den vielen möglichen Zugängen ist die Klasse der *Baum-Methoden* die einfachste und robusteste. Bei diesen Methoden ist das Diskretisierungsgitter *baumartig*. Der einfachste Fall einer Baum-Methode verwendet den Binomial-Baum. Das Black-Scholes-Modell ergibt sich hier im Grenzübergang mit Feinheit  $\rightarrow 0$ .

Definiere folgende äquidistante Zeit-Diskretisierung:

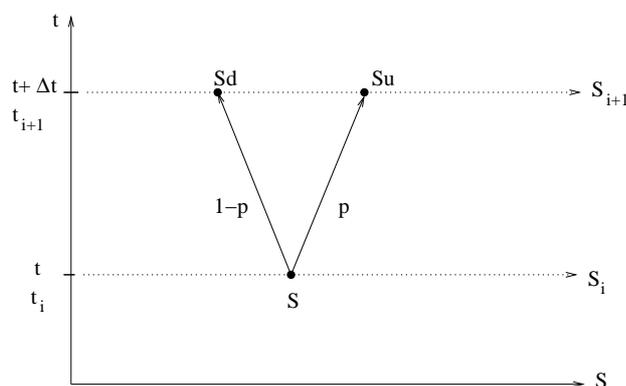
$M$ : Anzahl der Zeitschritte

$$\Delta t := \frac{T}{M}$$

$$t_i := i \cdot \Delta t, \quad i = 0, \dots, M$$

$$S_i := S(t_i)$$

Auf den  $S_i$ -Achsen werden im Folgenden diskrete  $S_{j,i}$ -Werte eingeführt.



## Annahmen

(Bi1) Der Marktpreis  $S$  über eine Periode  $\Delta t$  kann nur 2 Zustände annehmen:

$$Su \text{ oder } Sd \text{ mit } 0 < d < u.$$

(Bi2) Die Wahrscheinlichkeit der Bewegung “up” sei  $p$ , d.h.  $P(\text{up}) = p$ , mit  $0 < p < 1$ .

(Bi3) Die Erwartungswerte und Varianzen sollen identisch zu denjenigen des zeitstetigen Modells sein (für geometrisch-brownsche Bewegung  $S_t$  mit Wachstumsrate des risikofreien  $r$ ).

(Bi1) und (Bi2) definieren den Rahmen eines Binomialprozesses mit Wahrscheinlichkeit. Die Parameter  $u, d, p$  sind zunächst frei und sollen nun so bestimmt werden, dass sich die in (Bi3) gewünschten Eigenschaften ergeben.

## Bemerkungen

1.  $P$  ist die spezielle künstliche risikoneutrale Bewertung  $Q$ . Spezial-Literatur zu dem stochastischen Hintergrund: [Musielà&Rutkowski: Martingale Methods in Financial Modelling], [Shreve: Stochastic Calculus for Finance II (Continuous-time models)].
2. Für das zeitstetige Black–Scholes-Modell gilt bei risikoneutraler Bewertung (Zins  $r$ ; Beweis folgt unten in Abschnitt 1.5D):

$$\begin{aligned} E[S_t] &= S_0 e^{r(t-t_0)} \\ E[S_t^2] &= S_0^2 e^{(2r+\sigma^2)(t-t_0)} \end{aligned}$$

Setze hier  $S_i$  für  $S_0$ ,  $S_{i+1}$  für  $S_t$  und  $\Delta t$  für  $t - t_0$ .

3. Die Erwartungswerte sind bedingte Erwartungswerte, denn  $S(t_0)$  muss jeweils gleich dem Anfangswert sein.

**Folgerung für den Schritt  $i \rightarrow i + 1$ :**

$$\begin{aligned} E[S(t_{i+1}) | S(t_i) = S_i] &= S_i e^{r\Delta t} \\ \text{Var}[S(t_{i+1}) | S(t_i) = S_i] &= S_i^2 e^{2r\Delta t} (e^{\sigma^2\Delta t} - 1) \end{aligned}$$

Zunächst die Erwartungswerte: Im diskreten Modell gilt

$$E[S_{i+1}] = p S_i u + (1 - p) S_i d.$$

Durch Gleichsetzen der Erwartungswerte des diskreten und des stetigen Modells erhält man

$$e^{r\Delta t} = pu + (1 - p)d.$$

Dies ist die erste Gleichung für die 3 Unbekannten  $u, d, p$ . Es folgt

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

$0 < p < 1$  bedeutet

$$d < e^{r\Delta t} < u.$$

Dies muss gefordert werden, weil anderenfalls Arbitrage möglich ist.

Vergleich mit Abschnitt 1.2B zeigt: Dies  $p$  ist das  $q$  und repräsentiert die risikoneutrale Wahrscheinlichkeit.

Aus den Varianzen ergibt sich die zweite Gleichung

$$\begin{aligned}\text{Var}[S_{i+1}] &= \mathbb{E}[S_{i+1}^2] - (\mathbb{E}[S_{i+1}])^2 \\ &= p (S_i u)^2 + (1-p) (S_i d)^2 - S_i^2 (pu + (1-p)d)^2 \\ &\stackrel{!}{=} S_i^2 e^{2r\Delta t} (e^{\sigma^2 \Delta t} - 1).\end{aligned}$$

Hieraus erhält man

$$e^{2r\Delta t + \sigma^2 \Delta t} = pu^2 + (1-p)d^2.$$

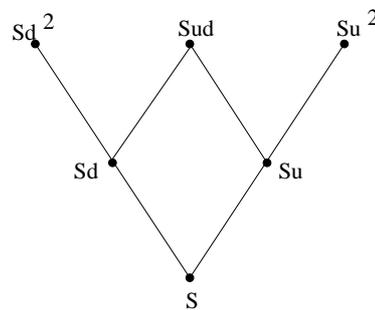
Eine dritte Gleichung kann willkürlich formuliert werden, etwa im Hinblick auf die Ökonomie der Rechnung, z.B.

$$u \cdot d = 1.$$

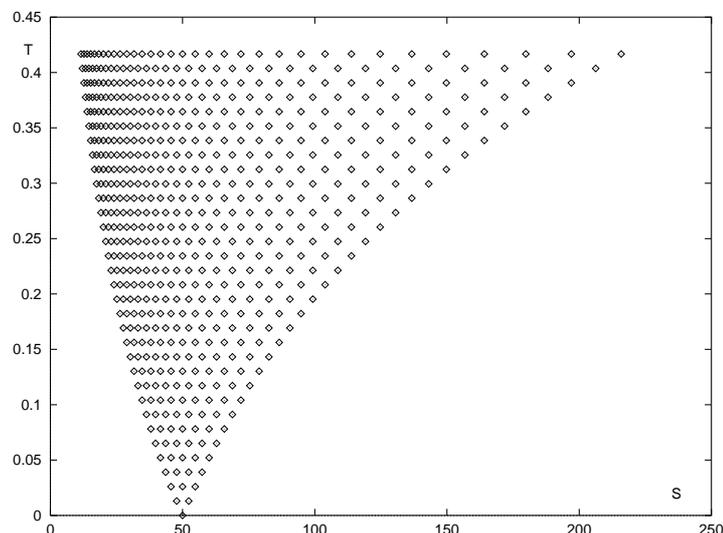
Als Lösung dieses nichtlinearen Gleichungssystems für  $u, d, p$  erhält man

$$\begin{aligned}\beta &:= \frac{1}{2}(e^{-r\Delta t} + e^{(r+\sigma^2)\Delta t}) \\ u &= \beta + \sqrt{\beta^2 - 1} \\ d &= 1/u = \beta - \sqrt{\beta^2 - 1} \\ p &= \frac{e^{r\Delta t} - d}{u - d}\end{aligned}$$

Der so mit  $ud = 1$  definierte einfache Baum ist "starr", da seine Parameter  $u, d, p$  nicht von  $K$  oder  $S_0$  abhängen. Der Baum ist rekombinierend.



Wegen  $S_{i+1} = \alpha S_i$ ,  $\alpha \in \{u, d\}$  liegt exponentielles Wachstum der "Äste" des Baumes vor.



Die  $S$ -Werte des Gitters sind also definiert durch

$$S_{j,i} := S_0 u^j d^{i-j}, \quad j = 0, \dots, i, \quad i = 1, \dots, M.$$

### Anwendung/Ausführung

Für  $t_M = T$  ist der Wert der Option durch die Auszahlungsfunktion  $\Psi(S) = (S - K)^+$  oder  $(K - S)^+$  bekannt:

$$V_{j,M} := \Psi(S_{j,M})$$

Aufbauend auf dem Prinzip der risikoneutralen Bewertung, welches oben für das Ein-Perioden-Modell hergeleitet wurde, gilt

$$V_i = e^{-r\Delta t} \mathbf{E}[V_{i+1}]$$

oder auf den Baum übertragen:

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

Diese zentrale Beziehung ist eine *Rekursion*, welche rückwärtslaufend von  $t_M$  anfangend die  $V$ -Bewertungen auf dem ganzen Baum berechnet, bis hin zu  $V_0 := V_{0,0}$ .

Im Falle der *amerikanischen* Option muss bei jeder Masche geprüft werden, ob vorzeitiges Ausüben sinnvoll ist. Der Halter der Option optimiert seine Position nach Vergleich von Payoff  $\Psi(S)$  und *Haltewert\**: er wählt den jeweils größeren Wert. Obige Rekursion muss also wie folgt modifiziert werden: (*cont* für continuation value = Wert, wenn nicht ausgeübt wird)

$$V_{j,i}^{\text{cont}} := e^{-r\Delta t} (pV_{j+1,i+1} + (1-p)V_{j,i+1})$$

Für europäische Optionen gilt  $V_{j,i} := V_{j,i}^{\text{cont}}$ . Für amerikanische Optionen gilt  $V_{j,i} := \max\{\Psi(S_{j,i}), V_{j,i}^{\text{cont}}\}$ , konkret:

$$\text{Call: } V_{j,i} := \max\{(S_{j,i} - K)^+, V_{j,i}^{\text{cont}}\}$$

$$\text{Put: } V_{j,i} := \max\{(K - S_{j,i})^+, V_{j,i}^{\text{cont}}\}$$

(Prinzip der Dynamischen Programmierung)

In der  $(S, t)$ -Ebene markieren die Knoten mit  $V_{j,i}^{\text{cont}} > \Psi(S_{j,i})$  den *Fortsetzungsbereich*, in dem das Halten der Option sinnvoll ist. Die komplementären Knoten markieren den *Stoppbereich*, hier ist Ausüben optimal. Wie die *early-exercise*-Kurve diese beiden Bereiche trennt, wird in Kapitel 4 genauer untersucht.

---

\* "Halten" im Sinne von *Behalten*, *Fortsetzen*.

**Algorithmus (Binomial-Methode, Grundversion)**

*Input:*  $r, \sigma, S = S_0, T, K$ , Put oder Call,  
europäisch oder amerikanisch,  $M$

*berechne:*  $\Delta t := T/M, u, d, p$  wie oben definiert  
 $S_{0,0} := S_0$   
 $S_{j,M} = S_{0,0} u^j d^{M-j}, j = 0, 1, \dots, M$   
(für amerikanische Optionen zusätzlich  $S_{j,i} = S_{0,0} u^j d^{i-j}$   
für  $0 < i < M, j = 0, 1, \dots, i$ )  
 $V_{j,M}$  aus dem Payoff  
 $V_{j,i}$  für  $i < M$  nach europäischer/amerikanischer Formel

*Output:*  $V_{0,0}$  ist Näherung für  $V(S_0, 0)$

**Vorteile der Methode**

- leicht zu implementieren
- robust
- Allgemeinere Typen von Optionen lassen sich relativ leicht realisieren.

**Nachteile der Methode**

- Genauigkeit ist eher bescheiden:  
Fehler  $\mathcal{O}(1/M) = \mathcal{O}(\Delta t)$ , d.h. lineare Konvergenz. Die Genauigkeit passt aber i.A. zur Praxis.
- Falls  $V_0$  für mehrere Werte von  $S_0$  benötigt wird, muss der Algorithmus entsprechend oft neu gestartet werden.

**Ergänzungen**

- Verbesserung zur Beseitigung von Oszillationen: Verallgemeinere  $ud = 1$  zu  $ud = \gamma$  mit  $\gamma$  so, dass für  $t = T$  ein Knoten auf den Strike  $K$  fällt. So ist der Baum flexibel, die Parameter  $u, d, p$  hängen von  $K$  und  $S_0$  ab, und die Genauigkeit wird verbessert: Extrapolation macht Sinn.
- diskrete Dividendenzahlung  $D$  zum Zeitpunkt  $t_D$ : schneide den Baum bei  $t_D$  und verschiebe die  $S$ -Werte um  $-D$ . Resultat: Auswertung an  $\tilde{S}_0 := S_0 - De^{-rt_D}$ . (Illustrationen in Topic 1 und 5 in den *Topics for CF*).
- Berechnung der Sensitivitäten (“Griechen”) mit Differenzenquotienten.

**Probleme**

Nicht nur bei Baummethoden: Im höher-dimensionalen Fall (z.B. Basket-Option mit 3 oder 4 Aktien) stellt sich die Frage, wie der Baum zu verallgemeinern ist.

In der Literatur wird obige Methode oft unter Cox-Ross-Rubinstein-Methode (CRR) aufgeführt.

Weiterentwicklungen: Trinomialmethode, “implizite Gitter” für variable  $\sigma(S, t)$ .

## 1.4 Stochastische Prozesse

Dieser Abschnitt dient als Grundlage für die zeitstetigen Modelle, wie z.B. dasjenige von Black, Scholes und Merton.

i.W. (geometrisch) brownische Bewegung

### Vorgeschichte

Brown (1827): engl. Botaniker, beschrieb Bewegung von Pollen.

Bachelier (1900): verwendete brownische Bewegung, um Aktienkurse zu modellieren.

Einstein (1905): Molekularbewegung

Wiener (1923): mathematisches Modell

seit 1940: Itô und andere

### Definition (Stochastischer Prozess)

Ein *Stochastischer Prozess* ist eine Familie von Zufallsvariablen  $X_t$  für  $t \geq 0$  oder  $0 \leq t \leq T$ .

D.h. jedem Zufallsexperiment ist eine Funktion  $X_t$  zugeordnet. Die resultierende Funktion  $X_t$  heißt *Pfad* oder *Trajektorie*.

### Definition (Wiener-Prozess / standard-brownische Bewegung)

$W_t$  (Bezeichn. auch  $W(t)$  oder  $W$  oder  $\{W_t\}_{t \geq 0}$ ) hat die Eigenschaften:

- (a)  $W_t$  ist ein *stetiger* stochastischer Prozess
- (b)  $W_0 = 0$
- (c)  $W_t \sim \mathcal{N}(0, t)$
- (d) Alle Inkremente  $\Delta W_t := W_{t+\Delta t} - W_t$  ( $\Delta t$  beliebig) auf nicht überlappenden  $t$ -Intervallen sind *unabhängig*.

(c) bedeutet:  $W_t$  ist normalverteilt mit  $E[W_t] = 0$  und  $\text{Var}[W_t] = E[W_t^2] = t$ .

### Bemerkungen

- 1) "Standard", weil als skalar angenommen,  $W_0 = 0$ , und kein Drift zugelassen ist.  
 $X_t = a + \mu t + W_t$  mit  $a, \mu \in \mathbb{R}$  wäre eine allgemeinere brownische Bewegung (mit Drift  $\mu$ ).
- 2) Folgerung (auch für  $W_0 = a$ ):

$$E[W_t - W_s] = 0, \quad \text{Var}[W_t - W_s] = t - s \quad \text{für } t > s.$$

- 3)  $W_t$  ist nirgendwo differenzierbar! Motivation:

$$\text{Var} \left[ \frac{\Delta W_t}{\Delta t} \right] = \frac{1}{(\Delta t)^2} \text{Var}[\Delta W_t] = \left( \frac{1}{\Delta t} \right)^2 \cdot \Delta t = \frac{1}{\Delta t}$$

würde für  $\Delta t \rightarrow 0$  gegen  $\infty$  gehen.

4) Ein Wiener-Prozess ist *selbst-ähnlich* im Sinne:

$$W_{\beta t} \stackrel{d}{=} \sqrt{\beta} W_t$$

(beide Seiten genügen der gleichen Verteilung). Allgemeiner gibt es fraktale Wiener-Prozesse mit

$$W_{\beta t} \stackrel{d}{=} \beta^H W_t,$$

für den normalen Wiener-Prozess ist  $H = \frac{1}{2}$ .  $H$  heißt *Hurst-Exponent*. Mandelbrot postulierte, dass Finanzmodelle fraktale Prozesse verwenden sollten.

### Zur Bedeutung

Der Wiener-Prozess ist Kern (“driving force”) wichtiger Finanzmodelle.

### Diskretisierung/Berechnung

Bisher haben wir  $W_t$  für zeitstetige Modelle ( $t \in \mathbb{R}$ ) betrachtet, und wollen nun  $W$  durch eine Diskretisierung approximieren. Es sei  $\Delta t > 0$  ein festes Zeit-Inkrement.

$$t_j := j \cdot \Delta t \quad \Rightarrow \quad W_{j\Delta t} = \sum_{k=1}^j (W_{k\Delta t} - W_{(k-1)\Delta t}) = \sum_{k=1}^j \Delta W_k$$

Die  $\Delta W_k$  sind unabhängig und haben wegen Bemerkung 2 die Eigenschaften

$$\mathbb{E}(\Delta W_k) = 0, \quad \text{Var}(\Delta W_k) = \Delta t.$$

Falls  $Z$  eine Zufallsvariable ist mit  $Z \sim \mathcal{N}(0, 1)$  [vgl. Kap.2], so gilt

$$Z\sqrt{\Delta t} \sim \mathcal{N}(0, \Delta t).$$

Deswegen ist

$$Z \cdot \sqrt{\Delta t} \quad \text{für } Z \sim \mathcal{N}(0, 1)$$

ein Modell für den diskreten Wiener Prozess der  $\Delta W_k$ .

### Algorithmus (Simulation eines Wiener-Prozesses)

*Start:*  $t_0 = 0, W_0 = 0$ ; wähle  $\Delta t$ .

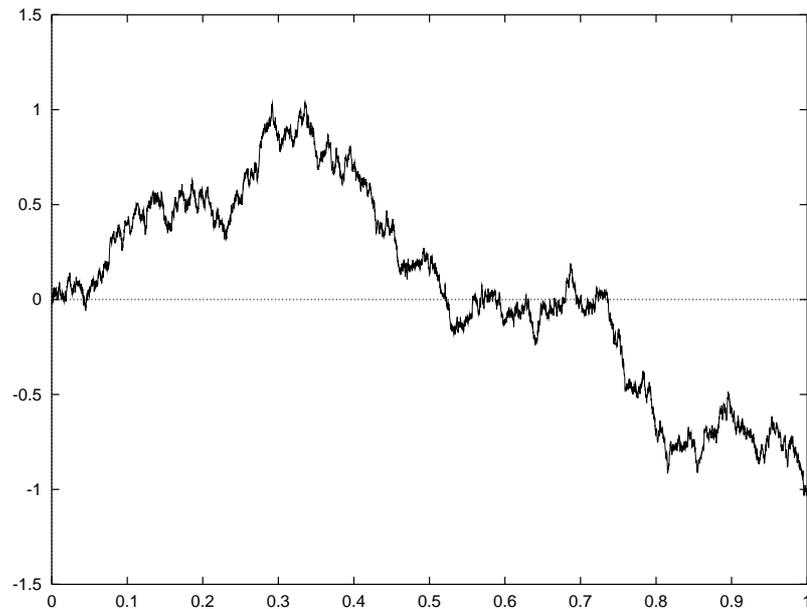
*Schleife*  $j = 1, 2, \dots$ :

$$t_j = t_{j-1} + \Delta t$$

ziehe  $Z \sim \mathcal{N}(0, 1)$  (numerisch)

$$W_j = W_{j-1} + Z\sqrt{\Delta t}$$

Dabei ist  $W_j$  eine Realisierung von  $W_t$  an der Stelle  $t_j$ . Die folgende Figur zeigt eine Simulation mit Zeit-Schrittweite  $\Delta t = 0.0002$ .



## Stochastisches Integral

### Motivation:

Ein Aktienkurs sei durch einen Wiener-Prozess  $W_t$  beschrieben. Es sei  $b(t)$  die Anzahl der Aktien im Portfolio zum Zeitpunkt  $t$ . Zur Vereinfachung nehmen wir an, dass es nur diskrete Handelszeitpunkte

$$0 = t_0 < t_1 < \dots < t_N = T$$

gibt. Somit ist  $b(t)$  stückweise konstant:

$$b(t) = b(t_{j-1}) \quad \text{für} \quad t_{j-1} \leq t < t_j \quad (*)$$

Es ergibt sich der Handelsgewinn

$$\sum_{j=1}^N b(t_{j-1})(W_{t_j} - W_{t_{j-1}}) \quad \text{für} \quad 0 \leq t \leq T.$$

Nun sei die Aktie beliebig handelbar ( $\rightarrow$  zeitstetiges Modell).

**Frage:** Konvergiert die Summe für  $N \rightarrow \infty$ ?

Für allgemeine  $b$  existiert das Integral

$$\int_0^T b(t) dW_t$$

nicht als Riemann-Stieltjes-Integral. (Hinreichend für Existenz ist Endlichkeit der ersten Variation von  $W_t$ .)

Zeige: Die erste Variation  $\sum_{j=1}^N |W_{t_j} - W_{t_{j-1}}|$  ist nicht beschränkt.

*Beweis:* Es gilt

$$\sum_{j=1}^N |W_{t_j} - W_{t_{j-1}}|^2 \leq \max_j (|W_{t_j} - W_{t_{j-1}}|) \sum_{j=1}^N |W_{t_j} - W_{t_{j-1}}|$$

für jede Zerlegung des Intervalls  $[0, T]$ . Nun  $\Delta t \rightarrow 0$ . Die zweite Variation ist beschränkt, bzw. konvergiert gegen ein  $c \neq 0$  (siehe Lemma unten). Wegen der Stetigkeit von  $W$  geht der erste Faktor der rechten Seite gegen 0 und damit der zweite Faktor (die erste Variation) gegen  $\infty$ .

Es bleibt noch zu untersuchen, was aus der zweiten Variation entsteht. Der verwendete Konvergenzbegriff ist die *Konvergenz im Mittel*, d.h.

$$\lim_{N \rightarrow \infty} \mathbb{E}[(X - X_N)^2] = 0,$$

bezeichnet als:  $X = \text{l.i.m.}_{N \rightarrow \infty} X_N$ .

Es bleibt zu zeigen:

### Lemma

Es sei  $t_0 = t_0^{(N)} < t_1^{(N)} < \dots < t_N^{(N)} = T$  eine Folge von Partitionen des Intervalls  $t_0 \leq t \leq T$  mit  $\delta_N := \max_{j=1}^N (t_j^{(N)} - t_{j-1}^{(N)})$ . Dann gilt:

$$\text{l.i.m.}_{\delta_N \rightarrow 0} \sum_{j=1}^N (W_{t_j^{(N)}} - W_{t_{j-1}^{(N)}})^2 = T - t_0$$

*Beweis:* Übungen

Hinweis: Im Beweis des Lemmas treten die folgenden Aussagen auf

$$\mathbb{E}[(\Delta W_t)^2 - \Delta t] = 0$$

$$\text{Var}[(\Delta W_t)^2 - \Delta t] = 2 \cdot (\Delta t)^2.$$

Dies bedeutet, dass die Zufallsvariable  $\Delta W_t^2$  sich im obigen probabilistischen Sinn “ähnlich verhält” wie  $\Delta t$ . Symbolische Abkürzung:

$$(dW_t)^2 = dt$$

(wichtig für Größenordnungsbetrachtungen)

Die Konstruktion eines (für unsere einschlägigen Integranden  $b$ ) existierenden stochastischen Integrals

$$\int_{t_0}^t b(s) dW_s$$

erfolgt aufbauend auf  $\int_{t_0}^t b(s) dW_s := \sum_{j=1}^N b(t_{j-1})(W_{t_j} - W_{t_{j-1}})$  für alle Stufenfunktionen  $b$  im Sinne von (\*).

Konstruktion für allgemeinere  $b$  über Konvergenz im Mittel von geeignet gegen  $b$  konvergierenden Stufenfunktionen, vgl. Literatur, z.B. [Øksendal: Stochastic Differential Equations], [Shreve: Stochastic Calculus].

## 1.5 Stochastische Differentialgleichungen

### A. Begriff

#### Definition (Diffusionsmodell)

Die *Integralgleichung* für einen stochastischen Prozess  $X_t$

$$X_t = X_{t_0} + \int_{t_0}^t a(X_s, s) ds + \int_{t_0}^t b(X_s, s) dW_s$$

heißt “Itô-stochastische Differentialgleichung”. Man benutzt die symbolische Schreibweise

$$dX_t = a(X_t, t) dt + b(X_t, t) dW_t$$

für die obige Integral-Gleichung. Lösungen dieser stochastischen DGL (d.h. obiger Integralgleichung) heißen *stochastische Diffusion*, oder *Itô-Prozess*. Der “Term”  $a(X_t, t)$  heißt *Drift*-Term, und  $b(X_t, t)$  heißt *Diffusion*.

#### Spezialfälle

- Den Standard-Wiener-Prozess erhält man mit  $X_t = W_t$ ,  $a = 0$ ,  $b = 1$ .
- Im deterministischen Fall gilt  $b = 0$ , d.h.  $\frac{dX_t}{dt} = a(X_t, t)$ .

#### Algorithmus (analog wie beim Wiener-Prozess)

Grundlage: diskrete Version

$$\Delta X_t = a(X_t, t) \Delta t + b(X_t, t) \Delta W_t$$

mit  $\Delta W$  und  $\Delta t$  aus Abschnitt 1.4. Es bezeichne  $y_j$  eine Näherung zu  $X_{t_j}$ .

*Start:*  $t_0, y_0 = X_0$ , wähle  $\Delta t$ .

*Schleife:*  $j = 0, 1, 2, \dots$

$$t_{j+1} = t_j + \Delta t$$

$$\Delta W = Z \sqrt{\Delta t} \text{ mit } Z \sim \mathcal{N}(0, 1)$$

$$y_{j+1} = y_j + a(y_j, t_j) \Delta t + b(y_j, t_j) \Delta W$$

(wird in Kapitel 3 vertieft: Ordnung “nur”  $\frac{1}{2}$ ; anschaulich klar wegen  $dW^2 = dt$ .)

## B. Anwendung auf den Aktienmarkt

**Modell** (GBM = Geometrisch-Brownsche Bewegung)

$$dS_t = \mu S_t dt + \sigma S_t dW_t$$

Dies ist eine Itô-stoch.-DGL mit  $a = \mu S_t$  und  $b = \sigma S_t$ . Sie ist *linear* solange  $\mu$  und  $\sigma$  nicht auch von  $S_t$  abhängen. Bei Black und Scholes sind  $\mu$  und  $\sigma$  konstant.

(Damit ist die “Lücke” GBM der Annahme 3 in Abschnitt 1.2 im mathematischen Modell des Marktes gefüllt.)

### Bedeutung

$\mu$  wird als “Wachstumsrate” und  $\sigma$  als Volatilität interpretiert. Für die relative Änderung gilt

$$\frac{dS_t}{S_t} = \mu dt + \sigma dW_t.$$

Wir treffen die Annahme, dass eine GBM mit konstantem  $\mu, \sigma$  vorliegt (entspricht dem Modell von Black–Scholes–Merton).

(Im Modell von Bachelier

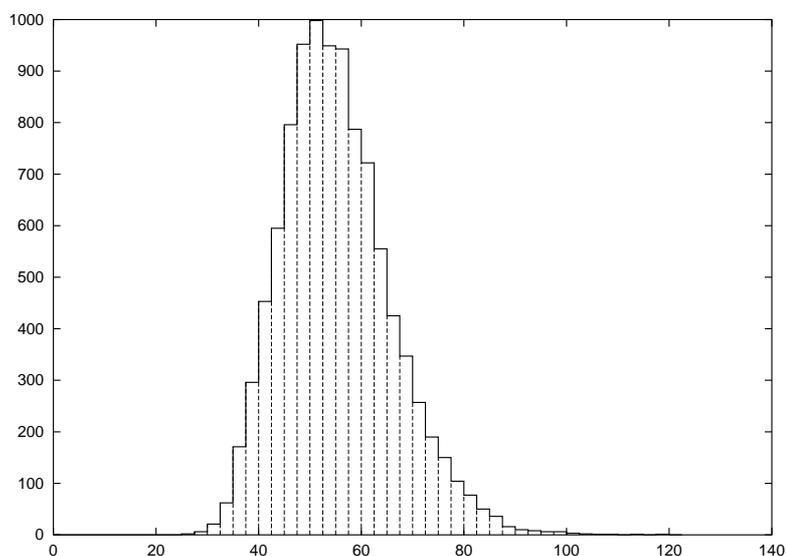
$$dS_t = \mu dt + \sigma dW_t$$

kann der Aktienkurs  $S_t$  negativ werden.)

### Empfehlung

Den Algorithmus implementieren (mit  $Z$  aus Kapitel 2) und ein konkretes Modell (z.B. mit  $S_0 = 50$ ,  $\mu = 0.1$ ,  $\sigma = 0.2$ ) mehrmals (etwa 10000 mal) bis  $t = 1$  integrieren und danach die erhaltenen Werte  $S_1$  in Klassen (Teilintervalle) einteilen und zählen.

→ Histogramm mit “log-normal” Verteilung (siehe Figur)



**Folgerung:**

Aus

$$\frac{\Delta S}{S} = \mu \Delta t + \sigma \Delta W$$

folgt für die Verteilung der  $\frac{\Delta S}{S}$ :

1) normalverteilt

2)  $E[\frac{\Delta S}{S}] = \mu \Delta t$

3)  $\text{Var}[\frac{\Delta S}{S}] = \sigma^2 \Delta t$

also:  $\frac{\Delta S}{S} \sim \mathcal{N}(\mu \Delta t, \sigma^2 \Delta t)$ Hieraus ergibt sich die folgende Methode, Volatilitäten  $\sigma$  empirisch zu ermitteln:

Die Daten  $\frac{\Delta S}{S}$  für eine Folge von Handelstagen seien  $R_i$  (*returns*) genannt, wobei  $R_{i+1}$  und  $R_i$  im Zeitabstand  $\Delta t$  gemessen sind. Die Returns mögen möglichst gut durch das GBM-Modell beschrieben sein.

$$\Rightarrow \sigma = \frac{1}{\sqrt{\Delta t}} * \text{Standardabweichung der } R_i$$

Dieser spezielle aus den Daten der Vergangenheit ermittelte Wert  $\sigma$  heißt *historische Volatilität* (zur *impliziten* Volatilität vgl. Übungen).

Der Prozess  $S_t$  unter GBM kann mit dem Algorithmus in 1.5A approximiert werden solange  $\Delta t > 0$  klein genug ist und  $S > 0$ .

**Andere Modelle**

GBM ist stetig und hat einige Eigenschaften (wie “dünne” Flanken der Dichtefunktion), die oft nicht überzeugen. Deswegen verwendet man auch andere stochastische Prozesse, z.B. Sprungprozesse oder Prozesse mit stochastische Volatilitäten. Vorläufig bleiben wir in der Klasse der Itô-DGL, d.h. von einem Wiener-Prozess angetriebene stetige Prozesse.

**Mean-Reversion** (oft für Zins-Modelle verwendet)

Hier sei  $R$  ein mittleres Zinsniveau. Betrachte die SDE (Stoch.DGL.) für den stochastischen Prozess  $r_t$ :

$$dr_t = \alpha(R - r_t) dt + \sigma^r r_t^\beta dW_t, \quad \alpha > 0$$

Das heißt

$$a(r_t, t) = \alpha(R - r_t) \quad \text{mean reversion Drift}$$

$$b(r_t, t) = \sigma^r r_t^\beta$$

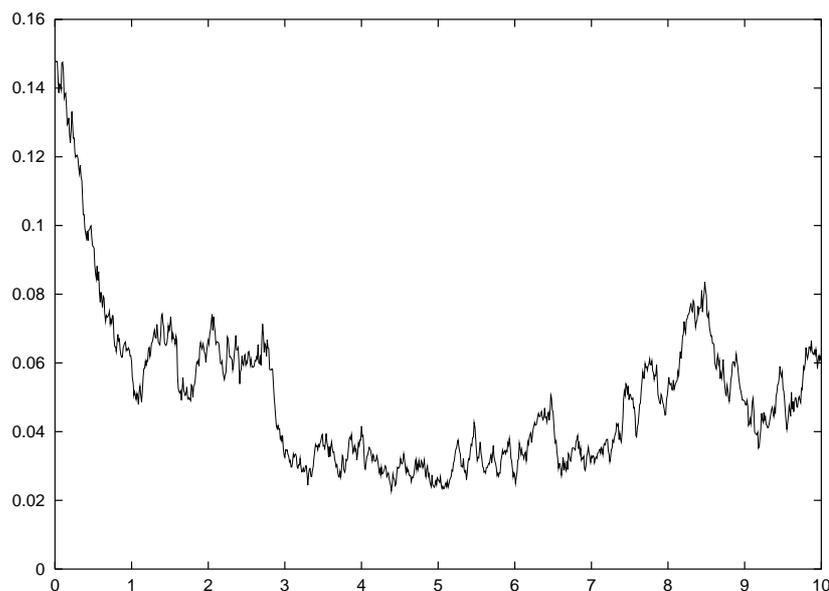
mit geeigneten Parametern  $R, \alpha, \sigma^r, \beta$  (durch Kalibrierung ermittelt). Dies hat folgende Wirkung für den Drift:

$r_t < R \Rightarrow$  Wachstum des deterministischen Anteils

$r_t > R \Rightarrow$  Abfall

Also verbleibt  $r_t$ , abgesehen von der Wirkung des stochastischen Terms, in der Größenordnung von  $R$ . Der stochastische Term wird  $r_t$  immer wieder von  $R$  wegtreiben, aber latent kehrt  $r_t$  zu  $R$  zurück.  $\alpha$  steuert die Geschwindigkeit der Rückkehr.

Für  $\beta = \frac{1}{2}$ , d.h.  $b(r_t, t) = \sigma^r \sqrt{r_t}$ , heißt das Modell CIR-Modell (Cox-Ingersoll-Ross-Modell).



Eine Simulation  $r_t$  des Cox-Ingersoll-Ross-Modells für  $R = 0.05$ ,  $\alpha = 1$ ,  $\beta = 0.5$ ,  $r_0 = 0.15$ ,  $\sigma^F = 0.1$ ,  $\Delta t = 0.01$

Nächste Erweiterung:

### Vektorwertige Prozesse

Es sei  $W_t = (W_t^{(1)}, \dots, W_t^{(m)})$  eine  $m$ -dimensionale brownische Bewegung. Definiere für  $i = 1, \dots, n$

$$X_t^{(i)} = X_{t_0}^{(i)} + \int_{t_0}^t a_i(X_s, s) ds + \sum_{k=1}^m \int_{t_0}^t b_{i,k}(X_s, s) dW_s^{(k)},$$

mit den Vektoren

$$X_t = \begin{pmatrix} X_t^{(1)} \\ \vdots \\ X_t^{(n)} \end{pmatrix}, \quad a(X_s, s) = \begin{pmatrix} a_1(X_s^{(1)}, \dots, X_s^{(n)}, s) \\ \vdots \\ a_n(X_s^{(1)}, \dots, X_s^{(n)}, s) \end{pmatrix}$$

und der Matrix

$$((b_{i,k}))_{i=1, \dots, n}^{k=1, \dots, m}$$

in welcher die Kovarianzen des Vektorprozesses stecken.

### Beispiel 1 Heston-Modell

$$\begin{aligned} dS_t &= \mu S_t dt + \sqrt{v_t} S_t dW^{(1)} \\ dv_t &= \kappa(\theta - v_t) dt + \sigma^{\text{vola}} \sqrt{v_t} dW^{(2)} \end{aligned}$$

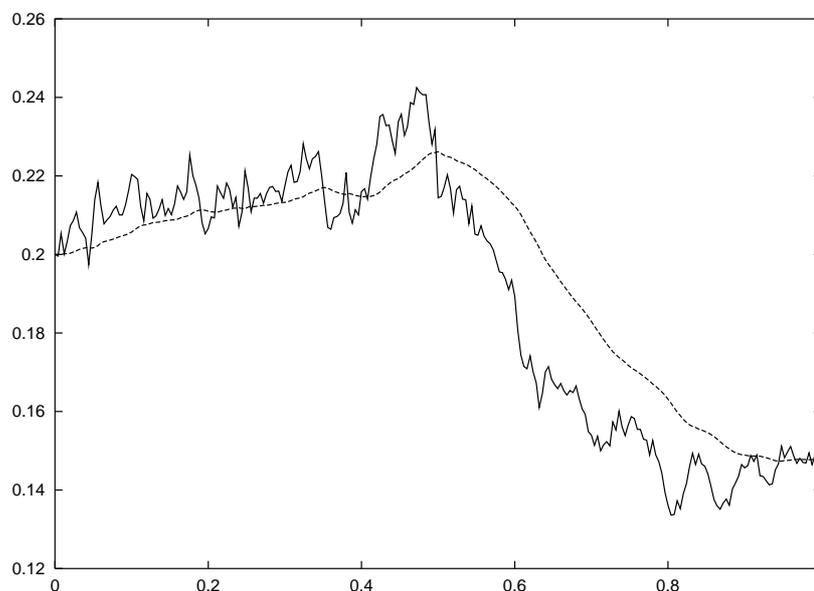
d.h. die **stochastische Volatilität**  $\sqrt{v_t}$  ist über einen Mean-Reversion-Prozess für die Varianz  $v_t$  definiert. Dieses Modell (mit  $n = 2$  und  $m = 2$ ) benutzt die Parameter  $\kappa, \theta, \sigma^{\text{vola}}$ , Korrelation  $\rho$  zwischen  $W^{(1)}$  und  $W^{(2)}$ , sowie der Anfangswert  $v_0$ , evtl.  $\mu$  (falls nicht  $\mu = r$ , siehe unten). D.h. es müssen circa 5 Parameter kalibriert werden. Das Heston-Modell wird viel verwendet.

**Beispiel 2 Volatilitäts-Tandem**

$$dS = \sigma S dW^{(1)}$$

$$d\sigma = -(\sigma - \zeta)dt + \alpha\sigma dW^{(2)}$$

$$d\zeta = \beta(\sigma - \zeta) dt$$



$\alpha = 0.3, \beta = 10$ ; gestrichelt:  $\zeta$

**Hinweis:** Es gibt auch noch den Begriff der *lokalen* Volatilität

$$\sigma = \sigma(t, S_t).$$

**C. Itô-Lemma**

**Motivation** (deterministischer Fall)

Es sei  $x(t)$  eine Funktion und  $y(t) := g(x(t), t)$ . Mit der Kettenregel folgt

$$\frac{d}{dt}g = \frac{\partial g}{\partial x} \cdot \frac{dx}{dt} + \frac{\partial g}{\partial t}.$$

Dies kann mit  $dx = a(x(t), t)dt$  wie folgt geschrieben werden

$$dg = \left( \frac{\partial g}{\partial x} a + \frac{\partial g}{\partial t} \right) dt$$

**Lemma (Itô)**

$X_t$  folge einem Itô-Prozess  $dX_t = a(X_t, t) dt + b(X_t, t) dW_t$  und es sei  $g(x, t) \in C^2$ . Dann gilt für  $Y_t := g(X_t, t)$  die SDGI

$$dY_t = \left( \frac{\partial g}{\partial x} a + \frac{\partial g}{\partial t} + \frac{1}{2} \frac{\partial^2 g}{\partial x^2} b^2 \right) dt + \frac{\partial g}{\partial x} b dW_t.$$

D.h.  $Y_t$  ist ein Itô-Prozess mit dem gleichen Wiener-Prozess wie der Ausgangsprozess.

**Beweis-Skizze:**

$$\left. \begin{array}{l} t \rightarrow t + \Delta t \\ X \rightarrow X + \Delta X \end{array} \right\} \rightarrow g(X + \Delta X, t + \Delta t) = Y + \Delta Y$$

Taylorentwicklung von  $g$  ergibt  $\Delta Y$ :

$$\Delta Y = \frac{\partial g}{\partial X} \cdot \Delta X + \frac{\partial g}{\partial t} \Delta t + \text{Terme quadratisch in } \Delta t, \Delta X$$

Nach Einsetzen von

$$\begin{aligned} \Delta X &= a \Delta t + b \Delta W \\ (\Delta X)^2 &= a^2 \Delta t^2 + b^2 \underbrace{\Delta W^2}_{=O(\Delta t)} + 2ab \Delta t \Delta W \end{aligned}$$

und Sortieren nach Potenzen von  $\Delta t, \Delta W$  erhalte

$$\Delta Y = \left( \frac{\partial g}{\partial X} a + \frac{\partial g}{\partial t} + \frac{1}{2} \frac{\partial^2 g}{\partial X^2} b^2 \right) \Delta t + b \frac{\partial g}{\partial X} \Delta W + \text{T.h.O.}$$

Ähnlich wie in Abschnitt 1.4 kann  $\Delta W$  als Summe geschrieben werden, und *Konvergenz im Mittel* wird angewendet (technisch aufwändig). Siehe [Øksendal].

**D. Anwendung auf das GBM-Modell**

Betrachte das GBM-Modell

$$dS = \mu S dt + \sigma S dW$$

mit  $\mu$  und  $\sigma$  konstant, d.h.  $X = S, a = \mu S, b = \sigma S$ .

- 1) Es sei  $V(S, t)$  eine glatte Funktion ( $\in C^2$ )

$$\Rightarrow dV = \left( \frac{\partial V}{\partial S} \mu S + \frac{\partial V}{\partial t} + \frac{1}{2} \sigma^2 S^2 \frac{\partial^2 V}{\partial S^2} \right) dt + \frac{\partial V}{\partial S} \sigma S dW$$

Hieraus folgt mit weiteren Überlegungen ( $\rightarrow$  später) die partielle DGL nach Black und Scholes für die Wertfunktion  $V(S, t)$  einer europäischen Standardoption.

- 2)  $Y_t := \log(S_t)$ , d.h.  $g(x) = \log x$

$$\begin{aligned} \Rightarrow \frac{\partial g}{\partial x} &= \frac{1}{x} \quad \text{und} \quad \frac{\partial^2 g}{\partial x^2} = -\frac{1}{x^2} \\ \Rightarrow d(\log S_t) &= \left( \mu - \frac{\sigma^2}{2} \right) dt + \sigma dW_t \end{aligned}$$

D.h. für die log-Preise  $Y_t = \log S_t$  gilt die einfachst-denkbare SDE, mit der elementaren Lösung:

$$\begin{aligned} Y_t &= Y_{t_0} + \left( \mu - \frac{\sigma^2}{2} \right) (t - t_0) + \sigma (W_t - W_{t_0}) \\ \Rightarrow \log S_t - \log S_{t_0} &= \log \frac{S_t}{S_{t_0}} = \left( \mu - \frac{\sigma^2}{2} \right) (t - t_0) + \sigma (W_t - W_{t_0}) \\ \Rightarrow S_t &= S_{t_0} \cdot \exp \left[ \left( \mu - \frac{\sigma^2}{2} \right) (t - t_0) + \sigma (W_t - W_{t_0}) \right] \end{aligned}$$

Es sei o.E.  $t_0 = 0$  und  $W_{t_0} = W_0 = 0$ , dann erhält man

$$S_t = S_0 \exp \left[ \left( \mu - \frac{\sigma^2}{2} \right) t + \sigma W_t \right]$$

Also ist  $S_t$  ist Exponentialfunktion einer brownischen Bewegung mit Drift.

**Folgerungen** für  $t_0 = 0$ :

- a)  $\log S_t$  ist normal-verteilt.
- b)  $E[\log S_t] = E[\log S_0] + (\mu - \frac{\sigma^2}{2})t + 0 = \log S_0 + (\mu - \frac{\sigma^2}{2})t$
- c)  $\text{Var}[\log S_t] = \text{Var}[\sigma W_t] = \sigma^2 t$
- a) – c) zusammengefasst ergibt

$$\log \frac{S_t}{S_0} \sim \mathcal{N} \left( \left( \mu - \frac{\sigma^2}{2} \right) t, \sigma^2 t \right)$$

- d) Hieraus ergibt sich die Dichtefunktion für  $Y = \log S$

$$\hat{f}(Y) = \hat{f}(\log S_t) = \frac{1}{\sigma \sqrt{2\pi t}} \exp \left[ -\frac{(\log(S_t/S_0) - (\mu - \sigma^2/2)t)^2}{2\sigma^2 t} \right].$$

Was ist nun die Dichtefunktion von  $S_t$ ? Diese kann durch eine Transformation der Verteilungsfunktion mittels des Transformationssatzes (Skript Abschnitt 2.2B) mit  $Y := \log S$  berechnet werden

$$\hat{f}(Y) dY = \underbrace{\hat{f}(\log S)}_{f(S_t)} \frac{1}{S} dS.$$

So erhält man als Dichte  $f$  der Verteilung des Aktienkurses  $S_t$

$$f(S_t, t; S_0, \mu, \sigma) := \frac{1}{S_t \sigma \sqrt{2\pi t}} \exp \left[ -\frac{(\log(S_t/S_0) - (\mu - \sigma^2/2)t)^2}{2\sigma^2 t} \right].$$

Dies ist die Dichte  $f_{\text{GBM}}$  der sogenannten *lognormal-Verteilung*. Sie beschreibt die Übergangswahrscheinlichkeit für den Übergang  $(S_0, 0) \rightarrow (S_t, t)$  unter GBM.

- e) Nun kann die letzte Lücke der Herleitung der Binomialmethode geschlossen werden; das dortige “zeitstetige” Modelle ist das hier betrachtete GBM-Modell (siehe Übungen):

$$E(S) = \int_0^{\infty} S f(\dots) dS = S_0 e^{\mu(t-t_0)}$$

$$E(S^2) = \int_0^{\infty} S^2 f(\dots) dS = S_0^2 e^{(\sigma^2 + 2\mu)(t-t_0)}$$

## 1.6 Risikoneutrale Bewertung

(Dieser Abschnitt *skizziert* wichtige Begriffe und Ideen. Für eine genauere Darstellung sei auf die Literatur Stochastischer Finanzmathematik verwiesen, wie z.B. [Musiela&Rutkowski: Martingale Methods in Financial Modelling])

**Wiederholung** (Ein-Perioden-Modell)

$$V_0 = e^{-rT} \mathbb{E}_Q[\Psi(S_T)]$$

wobei  $Q$  eine künstliche Wahrscheinlichkeit ist, siehe Abschnitt 1.2, und  $\Psi(S_T)$  den Payoff bezeichnet.

Für das **Modell mit kontinuierlicher Zeit** gilt formal die gleiche Beziehung. Aber  $Q$  and  $\mathbb{E}_Q$  sind hier anders. Es stellt sich heraus, dass  $f(S_t, t; S_0, r, \sigma)$  die Dichte von  $Q$  ist, mit  $\mu$  ersetzt durch  $r$ . Damit gilt

$$V_0 = e^{-rT} \int_0^\infty \Psi(S_T) \cdot f(S_T, T; S_0, r, \sigma) dS_T$$

für das GBM-basierte kontinuierliche Modell. Dies wird im Folgenden begründet. Von zentraler Bedeutung ist der

### Hauptsatz der Asset-Bewertung

Das Marktmodell ist arbitragefrei dann und nur dann wenn es ein Wahrscheinlichkeitsmaß  $Q$  gibt so dass die diskontierten Asset-Preise  $e^{-rt}S_t$  Martingale bzgl.  $Q$  sind.

### Vorbemerkung:

Stichprobenraum und  $\sigma$ -Algebra  $(\Omega, \mathcal{F})$ , die einem Wiener Prozess zugrunde liegen, werden nicht weiter spezifiziert. Die gewählte Wahrscheinlichkeit  $P$  ergänzt  $(\Omega, \mathcal{F})$  zum Wahrscheinlichkeitsraum  $(\Omega, \mathcal{F}, P)$ . Die Unabhängigkeit der Zuwächse des Wiener-Prozesses hängt von  $P$  ab, d.h.  $W$  kann Wiener-Prozess bzgl.  $P$  sein, ist dies aber nicht bezüglich einer anderen Wahrscheinlichkeit  $\hat{P}$ .

### Exkurs: Martingal

Ein Martingal  $M_t$  ist ein stochastischer Prozess mit der Eigenschaft

$$\mathbb{E}[M_t | \mathcal{F}_s] = M_s \quad \text{für alle } t, s \text{ mit } s \leq t,$$

wobei  $\mathcal{F}_s$  eine Filtration ist, d.h. eine Familie von  $\sigma$ -Algebren mit  $\mathcal{F}_s \subseteq \mathcal{F}_t \forall s \leq t$ . Eine solche Filtration ist ein Modell für den Informationsgehalt des Marktes.

$\mathbb{E}[M_t | \mathcal{F}_s]$  ist ein *bedingter Erwartungswert* und ist deutbar als Erwartungswert von  $M_t$  unter der Informationsmenge, die bis zum Zeitpunkt  $s$  zur Verfügung steht.

$M_t$  "Martingal" heißt also, dass  $M_s$  zum Zeitpunkt  $s$  die bestmögliche Prognose ist für  $t \geq s$ .

Martingal bzgl. einer Wahrscheinlichkeit  $Q$ :  $\mathbb{E}_Q[M_t | \mathcal{F}_s] = M_s$  für alle  $t, s$  mit  $s \leq t$ .

### Beispiele für Martingale

- 1) jeder Wiener-Prozess
- 2)  $W_t^2 - t$  für jeden Wiener-Prozess  $W$ .
- 3) notwendiges Kriterium für Martingale: Driftfreiheit.

Im Wesentlichen sind driftfreie Prozesse Martingale.

## Marktpreis des Risikos

$$\begin{aligned} dS &= \mu S dt + \sigma S dW \\ &= rS dt + (\mu - r)S dt + \sigma S dW \\ &= rS dt + \sigma S \left[ \frac{\mu - r}{\sigma} dt + dW \right] \end{aligned}$$

Der Anleger erwartet eine Rendite  $\mu > r$  als Ausgleich für das übernommene Risiko, welches durch  $\sigma$  repräsentiert wird.  $\mu - r$  ist eine Art von Kompensation für das Risiko (*excess return*).

$$\begin{aligned} \gamma &:= \frac{\mu - r}{\sigma} = \text{“Marktpreis des Risikos”} \\ &= \text{Kompensationsrate im Verhältnis zum Risiko} \end{aligned}$$

Also

$$dS = rS dt + \sigma S [\gamma dt + dW] \quad (*)$$

Bezüglich der Wahrscheinlichkeit  $\mathbb{P}$  repräsentiert der Term in eckigen Klammern keinen Wiener-Prozess, sondern eine brownische Bewegung mit Drift.

### Satz von Girsanov

Es sei  $W$  Wiener-Prozess bzgl.  $(\Omega, \mathcal{F}, \mathbb{P})$ . Falls  $\gamma$  geeignete Voraussetzungen erfüllt, dann gibt es eine Wahrscheinlichkeit  $\mathbb{Q}$  so dass

$$W_t^\gamma := W_t + \int_0^t \gamma ds$$

ein (standard) Wiener-Prozess ist unter  $\mathbb{Q}$ .

(Wahrscheinlichkeitstheorie:  $\mathbb{Q}$  ergibt sich aus dem Satz von Radon-Nikodym.  $\mathbb{Q}$  und  $\mathbb{P}$  sind äquivalent. Für konstante  $\gamma$  sind die Voraussetzungen von Girsanov erfüllt.)

### Anwendung

Einsetzen von  $dW^\gamma = dW + \gamma dt$  in  $(*)$  liefert

$$dS = rS dt + \sigma S dW^\gamma.$$

Dies ist ein Drift-Wechsel von  $\mu$  zu  $r$ ;  $\sigma$  bleibt unverändert. Das Verhalten von  $S_t$  unter der Wahrscheinlichkeit  $\mathbb{Q}$  ist definiert durch die Dichte  $f(\dots, r, \sigma)$ . Der Übergang von  $f(\dots, \mu, \sigma)$  zu  $f(\dots, r, \sigma)$  entspricht einer Justierung der Wahrscheinlichkeit von  $\mathbb{P}$  zu  $\mathbb{Q}$ . Dann ist das diskontierte  $e^{-rt} S_t$  driftfrei unter  $\mathbb{Q}$  und Martingal. Man nennt  $\mathbb{Q}$  das “risiko-neutrale Maß”.

### Handelsstrategie

Es sei  $X_t$  ein stochastischer Vektor-Prozess von Marktpreisen.  $b_t$  sei der Vektor der jeweils im Portfolio gehaltenen Stückzahlen. Folglich ist  $b_t^T X_t$  der Wert des jeweiligen Portfolios.

Beispiel

$$X_t := \begin{pmatrix} S_t \\ B_t \end{pmatrix},$$

wobei  $S_t$  Marktpreis der einer Option zugrunde liegenden Aktie ist und  $B_t$  der Wert einer risikofreien Anlage.

Bezeichnung: Es sei  $V_t$  die Zufallsvariable des Wertes einer europäischen Option.

**Annahmen:** (ähnlich auch für die Herleitung der Black–Scholes Gleichung)

- (1) Es existiere eine Strategie  $b_t$ , welche den Payoff der Option zum Zeitpunkt  $T$  repliziert, d.h.

$$b_T^* X_T = \text{Payoff}.$$

$b_t$  muss  $\mathcal{F}_t$ -messbar sein für alle  $t$ . (D.h. der Händler kann nicht in die Zukunft schauen. Beachte: Der Wert des Payoff ist Zufallsvariable.)

- (2) Das Portfolio ist gegenüber Geld-Zuflüssen und -Abflüssen abgeschlossen: Diese Eigenschaft heißt *selbstfinanzierend* und ist definiert als

$$d(b_t^* X_t) = b_t^* dX_t.$$

- (3) Der Markt ist arbitragefrei.

(1), (2), (3)  $\Rightarrow V_t = b_t^* X_t$  für  $0 \leq t \leq T$  (weil anderenfalls eine Arbitragemöglichkeit existieren würde.)

Wir betrachten eine europäische Option und einen Diskontierungsprozess  $Y_t$  mit der Eigenschaft, dass  $Y_t X_t$  ein Martingal ist. Man kann zeigen, dass dann auch  $Y_t b_t^* X_t$  ein Martingal ist (jeweils bzgl.  $\mathbb{Q}$ ).

**Folgerungen für Europäische Optionen** für  $t \leq T$

$$\begin{aligned} Y_t V_t = Y_t b_t^* X_t &= \mathbb{E}_{\mathbb{Q}}[Y_T b_T^* X_T \mid \mathcal{F}_t] && \text{(weil Martingal)} \\ &= \mathbb{E}_{\mathbb{Q}}[Y_T \cdot \text{Payoff} \mid \mathcal{F}_t] && \text{(Replikation)} \end{aligned}$$

Falls der Payoff eine Funktion  $\Psi$  von  $S_T$  ist (Vanilla-Option mit GBM), dann

$$= \mathbb{E}_{\mathbb{Q}}[Y_T \cdot \Psi(S_T)]$$

(weil  $W_T - W_t$  unabhängig von  $\mathcal{F}_t$ ). Mit Diskontierung  $Y_t = e^{-rt}$  folgt speziell für  $t = 0$

$$1 \cdot V_0 = \mathbb{E}_{\mathbb{Q}}[e^{-rT} \cdot \Psi(S_T)]$$

und damit

$$V_0 = e^{-rT} \int_0^\infty \Psi(S_T) \cdot f(S_T, T; S_0, r, \sigma) dS_T.$$

Wir nennen dies *risiko-neutrale Bewertung*.

**Literatur zur Stochastischen Finanzmathematik:** [Elliot & Kopp: Mathematics of Financial Markets], [Korn: Optionsbewertung und Portfolio-Optimierung], [Musiela & Rutkowski: Martingale Methods in Financial Modelling], [Shreve: Stochastic Calculus for Finance], u.a.

### Ausblick

Bisher wurden nur (Wert-)stetige Modelle für  $S_t$ , und ebenso  $W_t$ , betrachtet. Im Hinblick auf gelegentlich drastische Preisbewegungen ist die Stetigkeit eine starke Einschränkung. Abhilfe gibt uns hier eine *stochastische* Volatilität, oder die Einführung von Sprungprozessen.

Dieses Kapitel ist eine Kurzfassung von *Chapter 1* aus [R. Seydel: Tools for CF]. Letzteres enthält zusätzlich unter anderem Ausführungen zu Sprungprozessen und zur Kalibrierung.

## Ergänzungen zur Vorlesung

Die “**Griechen**” beschreiben die Sensitivitäten von  $V(S, t; \sigma, r)$  und sind wie folgt definiert:

$$\text{Delta} = \frac{\partial V}{\partial S}, \quad \text{gamma} = \frac{\partial^2 V}{\partial S^2}, \quad \text{theta} = \frac{\partial V}{\partial t}, \quad \text{vega} = \frac{\partial V}{\partial \sigma}, \quad \text{rho} = \frac{\partial V}{\partial r}$$

### Black–Scholes-Formel

For a European call the analytic solution of the Black–Scholes equation is

$$d_1 := \frac{\log \frac{S}{K} + \left(r - \delta + \frac{\sigma^2}{2}\right) (T - t)}{\sigma \sqrt{T - t}}$$

$$d_2 := d_1 - \sigma \sqrt{T - t} = \frac{\log \frac{S}{K} + \left(r - \delta - \frac{\sigma^2}{2}\right) (T - t)}{\sigma \sqrt{T - t}}$$

$$V_C(S, t) = S e^{-\delta(T-t)} F(d_1) - K e^{-r(T-t)} F(d_2),$$

where  $F$  denotes the standard normal cumulative distribution (compare Exercises), and  $\delta$  is a continuous dividend yield. The value  $V_P(S, t)$  of a put is obtained by applying the put-call parity

$$V_P = V_C - S e^{-\delta(T-t)} + K e^{-r(T-t)}$$

from which

$$V_P = -S e^{-\delta(T-t)} F(-d_1) + K e^{-r(T-t)} F(-d_2)$$

follows.

## 2. Berechnung von Zufallszahlen

### Definition (Stichprobe einer Verteilung)

Eine Folge von Zahlen heißt *Stichprobe (sample) von einer Verteilungsfunktion  $F$* , wenn die Zahlen unabhängige Realisierungen einer Zufallsvariablen mit diesem  $F$  sind.

### Beispiele

Wenn  $F$  diejenige der Gleichverteilung auf dem Intervall  $[0, 1]$  ist, so nennen wir die Stichprobe von  $F$  *gleichverteilt*. Bezeichnung:  $\sim \mathcal{U}[0, 1]$ .

Wenn  $F$  diejenige der Standardnormalverteilung ist, so nennen wir die Stichprobe von  $F$  *standardnormalverteilt*. Bezeichnung:  $\sim \mathcal{N}(0, 1)$ .

Die Grundlage ist die Berechnung von Zahlen  $\sim \mathcal{U}[0, 1]$ . Darauf baut die Berechnung von Zahlen  $\sim \mathcal{N}(0, 1)$  auf, und hierauf weitere, z.B.  $\sim \mathcal{N}(\mu, \Sigma)$ .

### 2.1 Gleichverteilte Zufallszahlen

#### A. Lineare Kongruenz-Generatoren

Wähle  $a, b, M \in \mathbb{N}$ ,  $a \neq 0$ ,  $a, b < M$ , und definiere für ein  $N_0 \in \mathbb{N}$  (“seed”) eine Folge von Zahlen durch

#### Algorithmus (Linearer Kongruenz-Generator)

Wähle  $N_0$ .  
Für  $i = 1, 2, \dots$  berechne  
 $N_i = (aN_{i-1} + b) \bmod M$

Definiere nun  $U_i \in [0, 1)$  durch

$$U_i = \frac{N_i}{M}.$$

Die Zahlen  $U_i$  werden als gleichverteilte Zufallszahlen verwendet.

Einfachste **Eigenschaften** der  $U_i$  bzw. der  $N_i$

- (a)  $N_i \in \{0, 1, \dots, M - 1\}$
- (b) Die Folge der  $N_i$  ist periodisch mit einer Periode  $p \leq M$ .  
(denn: Es gibt keine  $M + 1$  verschiedenen Zahlen  $N_i$ . So müssen zwei Zahlen aus  $\{N_0, \dots, N_M\}$  gleich sein, d.h.  $N_i = N_{i+p}$  mit  $p \leq M$ . Es folgt  $p$ -Periodizität.)

Literatur: [D. Knuth: The Art of Computer Programming, Volume 2]

Obige Zahlen  $U_i$  sind keine wirklichen Zufallszahlen, sondern deterministisch bestimmte, reproduzierbare Zahlen. Wir nennen sie *Pseudo-Zufallszahlen*, lassen aber im Folgenden den Präfix “Pseudo” weg. Das Ziel ist, die  $M, a, b$  so zu wählen, dass die Zahlen ein guter Ersatz echter Zufallszahlen sind.

### Beispiel

$$M = 244944, \quad a = 1597, \quad b = 51749$$

Weitere brauchbare Parameter  $a, b, M$  finden sich in [Press et al.: Numerical Recipes].

**Frage:** Was sind “gute” Zufallszahlen?

Praktische (und hypothetische) Antwort: Die Zahlen müssen möglichst “alle” Tests erfüllen.

**1. Ziel:** Verlange eine **große Periode**  $p$ , also  $M$  so groß wie möglich. Z.B. im Binärrechner  $M \approx 2^l$ , wobei  $l$  die Mantissenlänge ist. Geeignete  $a, b$  können mit zahlentheoretischen Methoden hergeleitet werden. [Knuth]

**2. Ziel:** Die Zahlen müssen wie gewünscht verteilt sein (Dichte  $f$ , Erwartungswert  $\mu$ , Varianz  $\sigma^2$ ). Hierzu: **Statistische Tests**. Berechne zunächst mit dem Algorithmus eine große Anzahl von  $U_i$ -Werten. Dann z.B.

- (a) Berechne von dieser Stichprobe den Mittelwert  $\hat{\mu}$  und die Varianz  $\hat{s}^2$ . Es muss gelten  $\hat{\mu} \approx \mu$  und  $\hat{s}^2 \approx \sigma^2$ .
- (b) Werte Korrelationen der  $U_i$  mit früheren  $U_{i-j}$  aus. Falls starke Korrelationen vorliegen (z.B. wenn es sehr wahrscheinlich ist, dass auf kleine  $U$ -Werte wiederum kleine Werte folgen), dann ist der Generator schlecht.
- (c) Schätzung der Dichtefunktion  $\hat{f}$  der Stichprobe. Es soll gelten:  $\hat{f} \approx f$ . Ein einfacher Prototyp eines solchen Tests ist: Teile  $[0, 1]$  in äquidistante Stücke

$$k\Delta U \leq U < (k+1)\Delta U,$$

wobei  $\Delta U$  die Länge der Teilintervalle bezeichnet. (Bei anderen Verteilungen wähle das Intervall so, dass alle  $U_i$  darin enthalten sind; definiere die Teilintervalle entsprechend.) Bei insgesamt  $j$  vielen Zahlen sei  $j_k$  die Anzahl der in das  $k$ -te Teilintervall fallenden Zahlen. Die Wahrscheinlichkeit, dass das  $k$ -te Teilintervall getroffen wird, ist  $\frac{j_k}{j}$ . Der Sollwert ist

$$\int_{k\Delta U}^{(k+1)\Delta U} f(x) dx \quad (f = 1 \text{ im Fall der Gleichverteilung}).$$

Eine Näherung des Integrals ist

$$\Delta U f(\bar{U}),$$

mit  $\bar{U}$  im  $k$ -ten Teilintervall. Also sollte für einen guten Generator gelten

$$\Delta U \hat{f}(\bar{U}) = \frac{j_k}{j} \stackrel{!}{=} \Delta U f(\bar{U}),$$

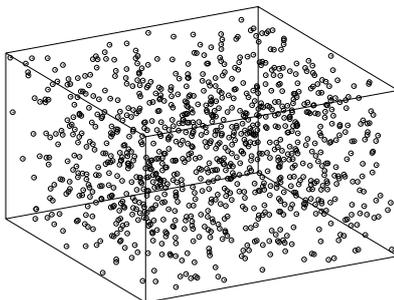
jedenfalls für kleine  $\Delta U$ . Man erhält die empirische Dichte auf dem  $k$ -ten Teilintervall

$$\hat{f} = \frac{j_k}{j\Delta U}.$$

**3. Ziel:** Man benötigt eine günstige **Gitterstruktur**. Bilde Vektoren als  $m$ -Tupel durch  $m$  aufeinanderfolgende Zahlen:

$$(U_i, U_{i+1}, \dots, U_{i+m-1})$$

Wenn  $U \sim \mathcal{U}[0, 1]$ , dann sollen diese Punkte den  $m$ -dimensionalen Einheits-Würfel möglichst gleichmäßig auffüllen. Die Folgen der Punkte/Vektoren liegen auf  $(m - 1)$ -dimensionalen Hyperebenen. Trivialer Fall:  $M$  Parallele Ebenen durch  $U = \frac{i}{M}$ ,  $i = 0, \dots, M - 1$  (für eine beliebige der Komponenten). Schlecht wäre es, wenn alle Punkte auf nur wenige Hyperebenen fallen. Dann würden in den “breiten” Bereichen zwischen den Hyperebenen *keine* Zufallszahlen liegen (schlechte Eigenschaft). Analysiere hierzu die “Gitterstruktur”, hier die Struktur der Hyperebenen, auf denen alle Punkte im  $[0, 1]^m$  “landen”.



*Analyse für  $m = 2$ :* Hyperebenen in der  $(U_{i-1}, U_i)$ -Ebene sind Geraden  $z_0 U_{i-1} + z_1 U_i = \lambda$ , für Parameter  $z_0, z_1, \lambda$ . Mit

$$\begin{aligned} N_i &= (aN_{i-1} + b) \bmod M \\ &= aN_{i-1} + b - kM \quad \text{für } kM \leq aN_{i-1} + b < (k+1)M \end{aligned}$$

folgt für beliebige Zahlen  $z_0, z_1$

$$\begin{aligned} z_0 N_{i-1} + z_1 N_i &= z_0 N_{i-1} + z_1 (aN_{i-1} + b - kM) \\ &= N_{i-1} (z_0 + az_1) + z_1 b - z_1 kM \\ &= M \underbrace{\left( N_{i-1} \frac{z_0 + az_1}{M} - z_1 k \right)}_{=: c=c(i)} + z_1 b \end{aligned}$$

Nach Division durch  $M$  erhält man

$$z_0 U_{i-1} + z_1 U_i = c + z_1 b M^{-1}.$$

Dies ist eine Geraden-Gleichung in der  $(U_{i-1}, U_i)$ -Ebene. Also liegt hier für festes  $z_0, z_1$  eine *Schar* von parallelen Geraden/Ebenen vor, parametrisiert durch  $c$ .

Frage: Gibt es eine Schar solcher Geraden/Ebenen, d.h. ein Tupel  $(z_0, z_1)$ , so dass nur *wenige* Geraden/Ebenen den Einheitswürfel schneiden? Die minimale Anzahl von parallelen Hyperebenen, die alle Punkte enthalten, beschreibt den schlimmsten Fall. Entscheidend ist die Kardinalität der  $c$ 's. Um den schlimmsten Fall zu finden, beschränken wir uns auf  $z_1, z_0 \in \mathbb{Z}$  und  $z_0 + az_1 \bmod M = 0$ . Dann ist  $c$  ganzzahlig, und

$$c = z_0 U_{i-1} + z_1 U_i - z_1 b M^{-1} \in \mathbb{Z}.$$

( $z_1 b M^{-1}$  ist hierbei eine konstante Parallelverschiebung, welche die Anzahl der Ebenen nicht beeinflusst.) Wie viele  $c$ 's gibt es? Für  $0 \leq U < 1$  erhält man eine Abschätzung für die  $c$ 's durch ein maximales Zahlenintervall  $I_c \subset \mathbb{Z}$ , so dass gilt

$$c \in I_c \Rightarrow \text{Gerade berührt oder schneidet den Einheitswürfel.}$$

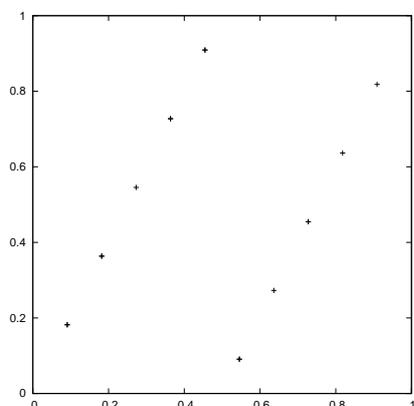
Die Mächtigkeit von  $I_c$  ist ein Maß dafür, wie eng der Abstand zwischen den parallelen Ebenen ist. Schlecht ist es, wenn die Mächtigkeit gering ist.

**Akademisches Beispiel**  $N_i = 2N_{i-1} \pmod{11}$  (also  $a = 2$ ,  $b = 0$ ,  $M = 11$ )

Die Kombination  $(z_0, z_1) = (-2, 1)$  löst  $z_0 + az_1 = 0 \pmod{M}$ . Dann gilt

$$-2U_{i-1} + U_i = c.$$

Aus  $0 \leq U < 1$  folgt  $-2 < c < 1$ , außerdem gilt  $c \in \mathbb{Z}$ . Es bleiben nur  $c = -1$  und  $c = 0$ . D.h. bei dieser Wahl von  $(z_0, z_1)$  fallen alle 10 Punkte in  $[0, 1)^2$  auf nur 2 Geraden. (0 wird nicht angenommen für  $N_0 \neq 11k$ ,  $k \in \mathbb{Z}$ .)



$(U_{i-1}, U_i)$ -Ebene.  
10 Punkte für  $N_0 \neq 11k$ ,  $k \in \mathbb{Z}$

**Beispiel**  $N_i = (1229N_{i-1} + 1) \pmod{2048}$

Die Bedingung  $z_0 + az_1 = 0 \pmod{M}$

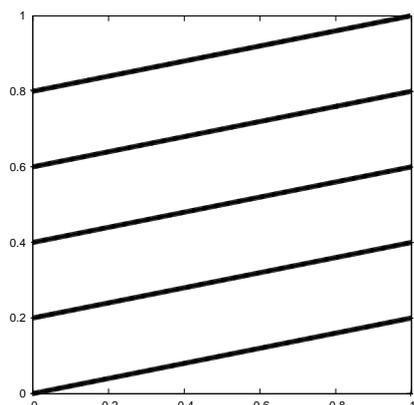
$$\frac{z_0 + 1229z_1}{2048} \in \mathbb{Z}$$

ist erfüllt durch  $z_0 = -1$ ,  $z_1 = 5$ , denn

$$-1 + 1229 \cdot 5 = 6144 = 3 \cdot 2048.$$

Wegen  $c = -U_{i-1} + 5U_i - \frac{5}{2048}$  gilt  $-1 - \frac{5}{2048} < c < 5 - \frac{5}{2048}$ . Alle Punkte in  $[0, 1)^2$  fallen auf 6 Geraden, da  $c \in \{-1, 0, 1, 2, 3, 4\}$ .

Der  $U_i$ -Abstand von zwei benachbarten Geraden ist  $\frac{1}{z_1} = \frac{1}{5}$ .



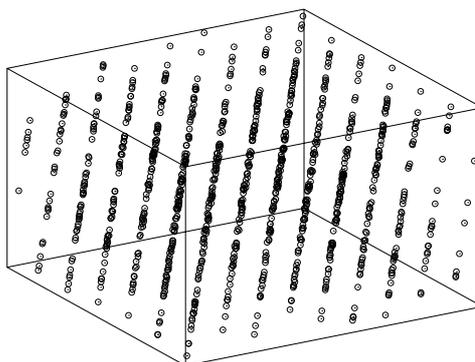
$(U_{i-1}, U_i)$ -Ebene.  
Die diskreten Punkte werden in dieser Grafik nicht getrennt. Die sechste Gerade besteht aus einem Punkt.

Die beiden obigen Beispiele zeigen für  $m = 2$  offensichtlich keine gleichmäßige Verteilung der  $(U_{i-1}, U_i)$ -Punkte. Die Zufallspunkte des folgende Beipiels sind in dieser Hinsicht gut verteilt. Aber es zeigt, dass sich eine gleichmäßige Verteilung für  $m = 2$  nicht auf größere Werte von  $m$  übertragen lässt.

### Beispiel (RANDU)

$$N_i = aN_{i-1} \bmod M, \quad \text{mit } a = 2^{16} + 3, \quad M = 2^{31}$$

Für  $m = 2$  zeigen Experimente, dass die Punkte  $(U_{i-1}, U_i)$  schön gleichmäßig auf dem Quadrat verteilt sind. Aber für  $m = 3$  zeigt die Analyse, dass die Zufallspunkte im Würfel  $[0, 1)^3$  auf nur 15 Ebenen liegen.



(Analyse für große  $m$  analog. Illustration in *Topic 14*)

## B. Fibonacci-Generatoren

Es gibt auch andere Klassen von Zufallszahlen-Generatoren, z.B. diejenige der **Fibonacci-Generatoren**. Ein Prototyp solcher Generatoren ist definiert durch

$$N_{i+1} := N_{i-\nu} - N_{i-\mu} \bmod M$$

für geeignete  $\mu, \nu$  (evtl. auch mit “+” oder mit mehr Termen). Literatur: [Knuth]

### Beispiel

$$U_i := U_{i-17} - U_{i-5},$$

falls  $U_i < 0$  setze  $U_i := U_i + 1.0$

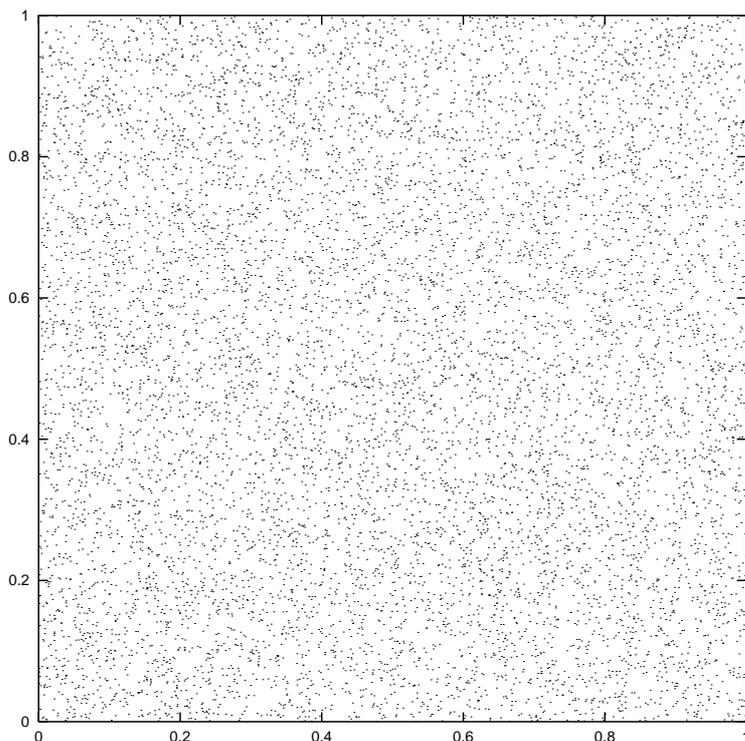
(einfaches Beispiel mit relativ guten Eigenschaften, jedoch gibt es Korrelationen.)

### Algorithmus (Schleife eines einfachen Fibonacci-Generators)

*Repeat:*  $\zeta := U_i - U_j$   
 falls  $\zeta < 0$ , setze  $\zeta := \zeta + 1$   
 $U_i := \zeta$   
 $i := i - 1$   
 $j := j - 1$   
 falls  $i = 0$ , setze  $i := 17$   
 falls  $j = 0$ , setze  $j := 17$

Initialisierung: Setze  $i = 17$ ,  $j = 5$ , und berechne  $U_1, \dots, U_{17}$  mit einem Kongruenz-Generator mit beispielsweise  $M = 714025$ ,  $a = 1366$ ,  $b = 150889$ .

**Professioneller Generator** zur Erzeugung von gleichverteilten Zufallszahlen: “Mersenne Twister” von Matsumoto, Nishimura, in: ACM Transactions on Modelling and Computer Simulations **8** (1998), S.3-30. Dieser Generator hat exzellente Eigenschaften, mit enorm großer Periode und gut verteilten Punkten auch für große  $m$ .



10000 Zufallszahlen  $(U_{i-1}, U_i)$ , mit Fibonacci-Generator erstellt

## 2.2 Zufallszahlen mit anderen Verteilungen

Die Erzeugung von z.B. normalverteilten Zufallszahlen baut auf gleichverteilten Zufallszahlen auf. Zur Erzeugung von Zufallszahlen mit vorgegebenen Verteilungen gibt es Inversions- und Transformations-Methoden, sowie Verwerfungsmethoden.

### A. Inversion

Es sei  $F(x) := P(X \leq x)$  Verteilungsfunktion, wobei  $X$  eine Zufallsvariable und  $P$  die mit  $F$  verbundene Wahrscheinlichkeit ist.

#### Satz (Inversion)

Angenommen  $U \sim \mathcal{U}[0, 1]$  und  $F$  sei stetige und streng monoton wachsende Verteilungsfunktion, dann ist  $X := F^{-1}(U)$  nach  $F$  verteilt.

*Beweis:*

$U \sim \mathcal{U}[0, 1]$  heißt  $P(U \leq \xi) = \xi$  für  $0 \leq \xi \leq 1$ . Also

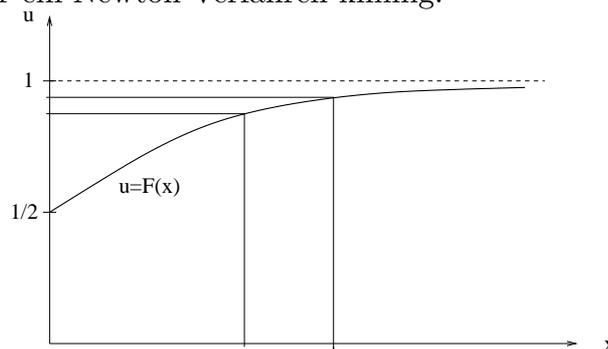
$$P(F^{-1}(U) \leq x) = P(U \leq F(x)) = F(x).$$

## Anwendung

Berechne  $u \sim \mathcal{U}[0, 1]$ . Werte  $F^{-1}(u)$  aus, diese Zahlen haben die gewünschte Verteilung. Die Inversion muss meist numerisch durchgeführt werden, weil  $F^{-1}$  i.A. nicht analytisch bekannt ist.

Zwei Varianten des numerischen Zugangs:

- (a)  $F(x) = u$  ist eine nichtlineare Gleichung für  $x$ , welche mit den Standardmethoden der Numerik (z.B. Newton-Verfahren) iterativ gelöst werden kann. Schwierigkeit bei der Normalverteilung (Figur): Wegen der großen Auswirkung von Störungen in  $u$  auf das zugehörige  $x$  in den Bereichen  $u \approx 0$ ,  $u \approx 1$  sind die Abbruchkriterien für ein Newton-Verfahren knifflig.



- (b) Konstruiere eine “einfache” Näherungsfunktion  $G$ , so dass  $G(u) \approx F^{-1}(u)$ , und werte dann  $x = G(u)$  aus. Man kann  $G$  so konstruieren, dass das asymptotische Verhalten, d.h. die Pole von  $G$ , gut wiedergegeben werden. Bei Ausnutzung der Symmetrie der Normalverteilung (zu  $(x, u) = (0, \frac{1}{2})$ ) braucht man nur das Polverhalten bei  $u = 1$  zu berücksichtigen. Hierzu Ansatz einer rationalen Funktion für  $G(u)$ , d.h. das Nennerpolynom hat  $u = 1$  als Nullstelle.

## B. Transformation

Zunächst der skalare Fall:  $X$  sei eine Zufallsvariable. Wie ist ein transformiertes  $h(X)$  verteilt?

### Satz (skalare Transformation)

Es sei  $X$  eine Zufallsvariable mit Dichtefunktion  $f$  und Verteilungsfunktion  $F$ . Weiter sei

$$h : S \rightarrow B$$

mit  $S, B \subseteq \mathbb{R}$ , wobei  $S$  der Träger von  $f$  ist, zudem sei  $h$  streng monoton.

- (a)  $Y := h(X)$  ist Zufallsvariable mit Verteilungsfunktion

$$\begin{aligned} F(h^{-1}(y)) & \quad \text{für monoton wachsendes } h \\ 1 - F(h^{-1}(y)) & \quad \text{für monoton fallendes } h \end{aligned}$$

- (b) Falls  $h^{-1}$  absolut stetig ist, dann ist für fast alle  $y$

$$f(h^{-1}(y)) \left| \frac{dh^{-1}(y)}{dy} \right|$$

die Dichte von  $h(X)$ .

*Beweis:* (schreibe auch  $F^X$  für  $F$ )

$$(a) F^Y(y) := \mathbf{P}(h(X) \leq y) =$$

falls  $h$  monoton wachsend:

$$= \mathbf{P}(X \leq h^{-1}(y)) = F^X(h^{-1}(y))$$

falls  $h$  monoton fallend:

$$= \mathbf{P}(X \geq h^{-1}(y)) = 1 - \mathbf{P}(X < h^{-1}(y)) = 1 - F^X(h^{-1}(y))$$

(b)  $h^{-1}$  absolut stetig  $\Rightarrow$  Dichte von  $Y = h(X)$  ist gleich der Ableitung der Verteilungsfunktion (fast überall). Die Auswertung von  $\frac{dF(h^{-1}(y))}{dy}$  liefert mit der Kettenregel das gewünschte Ergebnis; dabei ist zwischen wachendem und fallendem  $h$  zu unterscheiden.

### Anwendung

Beginne mit  $X \sim \mathcal{U}[0, 1]$  und  $f$  als Dichte der Gleichverteilung, also

$$f(x) = \begin{cases} 1 & \text{für } 0 \leq x \leq 1 \\ 0 & \text{sonst} \end{cases}$$

d.h.  $S = [0, 1]$ . Zu berechnen sind Zufallszahlen  $Y$  mit einer vorgegebenen Dichte  $g(y)$ . Benötige also eine Transformation  $h$ , so dass

$$f(h^{-1}(y)) \left| \frac{dh^{-1}(y)}{dy} \right| = 1 \left| \frac{dh^{-1}(y)}{dy} \right| \stackrel{!}{=} g(y).$$

Dann ist  $h(X)$  wie gewünscht verteilt.

### Beispiel (Exponentialverteilung)

Die Exponentialverteilung mit Parameter  $\lambda > 0$  hat die Dichte

$$g(y) = \begin{cases} \lambda e^{-\lambda y} & \text{für } y \geq 0 \\ 0 & \text{für } y < 0. \end{cases}$$

d.h.  $B$  sind die nichtnegativen reellen Zahlen. Transformation von  $[0, 1] \rightarrow B$  mit der monoton fallenden Funktion

$$y = h(x) := -\frac{1}{\lambda} \log x$$

und der Umkehrfunktion  $h^{-1}(y) = e^{-\lambda y}$  für  $y \geq 0$ . Für dieses  $h$  gilt

$$f(h^{-1}(y)) \left| \frac{dh^{-1}(y)}{dy} \right| = 1 \cdot |(-\lambda)e^{-\lambda y}| = \lambda e^{-\lambda y} = g(y).$$

Also ist  $h(X)$  exponentialverteilt, sofern  $X \sim \mathcal{U}[0, 1]$ .

### Anwendung

Berechne gleichverteilte  $U_1, U_2, \dots$

$$\Rightarrow -\frac{1}{\lambda} \log(U_1), \quad -\frac{1}{\lambda} \log(U_2), \quad \dots \text{ sind exponentialverteilt.}$$

(Die Abstände der Sprungzeiten bei Poisson-Prozessen sind exponentialverteilt.)

Versuch zur Normalverteilung: Suche  $h$  so, dass

$$1 \cdot \left| \frac{dh^{-1}(y)}{dy} \right| = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}y^2\right).$$

Dies ist eine Differentialgleichung für  $h^{-1}$  ohne analytische Lösungsmöglichkeit. Hier hilft die mehrdimensionale Version des Transformationssatzes.

**Satz (Transformationsatz im  $\mathbb{R}^n$ )**

Es sei  $X$  eine Zufallsvariable im  $\mathbb{R}^n$  mit Dichte  $f(x) > 0$  auf dem Träger  $S$ . Die Transformation  $h : S \rightarrow B$ ,  $S, B \subseteq \mathbb{R}^n$  sei umkehrbar und die Umkehrfunktion stetig differenzierbar auf  $B$ . Dann hat  $Y := h(X)$  die Dichte

$$f(h^{-1}(y)) \left| \frac{\partial(x_1, \dots, x_n)}{\partial(y_1, \dots, y_n)} \right|, \quad y \in B, \quad (2.7)$$

wobei  $\frac{\partial(x_1, \dots, x_n)}{\partial(y_1, \dots, y_n)}$  die Determinante der Jacobimatrix von  $h^{-1}(y)$  ist.

*Beweis:* siehe Theorem 4.2 in [L.Devroye: Non-Uniform Random Variate Generation (1986)]

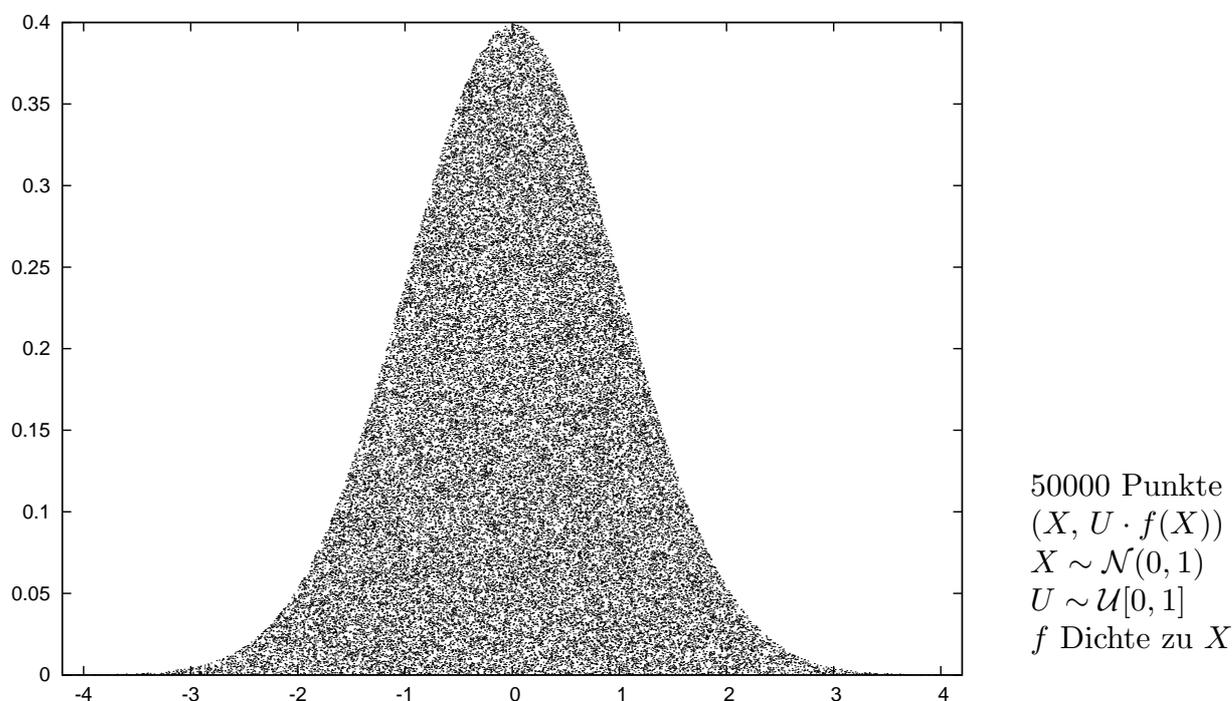
Die zweidimensionale Version wird in Abschnitt 2.3 angewendet zur Berechnung normalverteilter Zufallszahlen.

**C. Verwerfungsmethode**

Grundlage: Es sei  $f$  eine Dichtefunktion auf  $S \subset \mathbb{R}$  and  $\mathcal{A}_f$  die Fläche zwischen der  $x$ -Achse und dem Graphen von  $f$ . Weiter seien zwei Zufallsvariable  $U$  und  $X$  unabhängig voneinander, mit  $U \sim \mathcal{U}[0, 1]$  und  $X$  nach  $f$  verteilt. Dann sind die Punkte

$$(x, y) := (X, U \cdot f(X))$$

auf  $\mathcal{A}_f$  gleichverteilt. (Die Umkehrung gilt ebenso.) Dies ist in der Figur illustriert für die Normalverteilung. Schneidet man von der Fläche  $\mathcal{A}_f$  etwas ab, dann sind die verbleibenden Punkte immer noch gleichverteilt. Dies machen sich Verwerfungsmethoden zu Nutze.



Es sei  $g$  eine andere Dichte auf  $S$ , und für eine Konstante  $c \geq 1$  gelte

$$f(x) \leq c \cdot g(x) \quad \text{für alle } x \in S.$$

Die Funktion  $cg$  majorisiert  $f$ . Die Menge  $\mathcal{A}_f$  ist Teilfläche der Fläche  $\mathcal{A}_{cg}$  unterhalb des Graphen der Majorante  $cg$ . Der Verwerfungs-Algorithmus setzt voraus, dass nach  $g$  verteilte

Zufallszahlen  $x$  leicht berechnet werden können, und damit auch auf  $\mathcal{A}_{cg}$  gleichverteilte Punkte  $(x, u_{cg}(x))$ . Zu berechnen sind nach  $f$  verteilte Zufallszahlen. Die Punkte oberhalb  $\mathcal{A}_f$  (dort gilt  $u_{cg}(x) > f(x)$ ) werden verworfen; die  $x$ -Koordinaten der akzeptierten Punkte mit  $u_{cg}(x) \leq f(x)$  sind wie gewünscht verteilt.

*Repeat:*

$x :=$  Zufallszahl nach  $g$  verteilt,

$u :=$  Zufallszahl  $\sim \mathcal{U}[0, 1]$  unabhängig von  $x$

*until*  $u_{cg}(x) \leq f(x)$

*return:*  $x$

Beispiel zur Übung: Laplace-Dichte  $g(x) := \frac{1}{2} \exp(-|x|)$ ,  $f$  Dichte der Standard-Normalverteilung. Was ist  $c$ ?\*

## 2.3 Normalverteilte Zufallszahlen

In diesem Abschnitt wird der Transformationssatz im  $\mathbb{R}^2$  angewendet zur Berechnung von normalverteilten Zufallszahlen. Es sei darauf hingewiesen, dass Inversions- und Verwerfsmethoden Alternativen bieten.

### A. Methode von Box und Muller

$S := [0, 1]^2$ ,  $X$  gleichverteilt auf  $S$  mit Dichte  $f = 1$  auf  $S$ . Transformation  $h$ :

$$\begin{cases} y_1 = \sqrt{-2 \log x_1} \cos 2\pi x_2 =: h_1(x_1, x_2) \\ y_2 = \sqrt{-2 \log x_1} \sin 2\pi x_2 =: h_2(x_1, x_2) \end{cases}$$

Umkehrabbildung  $h^{-1}$ :

$$\begin{cases} x_1 = \exp \left\{ -\frac{1}{2}(y_1^2 + y_2^2) \right\} \\ x_2 = \frac{1}{2\pi} \arctan \frac{y_2}{y_1} \end{cases}$$

Für diese Transformation gilt

$$\begin{aligned} \frac{\partial(x_1, x_2)}{\partial(y_1, y_2)} &= \det \begin{pmatrix} \frac{\partial x_1}{\partial y_1} & \frac{\partial x_1}{\partial y_2} \\ \frac{\partial x_2}{\partial y_1} & \frac{\partial x_2}{\partial y_2} \end{pmatrix} = \\ &= -\frac{1}{2\pi} \exp \left\{ -\frac{1}{2}(y_1^2 + y_2^2) \right\}. \end{aligned}$$

Absolut genommen ist dies die Dichte einer zweidimensionalen Normalverteilung. Wegen

$$\left| \frac{\partial(x_1, x_2)}{\partial(y_1, y_2)} \right| = \left[ \frac{1}{\sqrt{2\pi}} \exp \left( -\frac{1}{2}y_1^2 \right) \right] \cdot \left[ \frac{1}{\sqrt{2\pi}} \exp \left( -\frac{1}{2}y_2^2 \right) \right]$$

---

\* farbig in Topic 3 in den *Topics for CF*

ist die zweidimensionale Dichte das Produkt der eindimensionalen Dichten der Standardnormalverteilung. Folgerung: Die beiden Komponenten  $y_1, y_2$  des Vektors  $Y$  sind unabhängig. [z.B. Hesse: Angewandte Wahrscheinlichkeitstheorie, S. 68]

### Anwendung

Wenn die beiden Komponenten  $x_1, x_2$  nach  $\sim \mathcal{U}[0, 1]$  verteilt sind, dann liefert die Transformation zwei unabhängige Zufallszahlen  $y_1, y_2 \sim \mathcal{N}(0, 1)$ .

### Algorithmus (Box-Muller)

- (1) generiere  $U_1 \sim \mathcal{U}[0, 1]$  und  $U_2 \sim \mathcal{U}[0, 1]$ .
- (2)  $\theta := 2\pi U_2$ ,  $\rho := \sqrt{-2 \log U_1}$
- (3)  $Z_1 := \rho \cos \theta$  ist  $\sim \mathcal{N}(0, 1)$   
(genauso wie  $Z_2 := \rho \sin \theta$ ).

## B. Variante von Marsaglia

Präpariere die Eingangsdaten  $x_1, x_2$  für die Box & Muller -Transformation derart, dass die trigonometrischen Funktionen entfallen. Mit  $U \sim \mathcal{U}[0, 1]$  erhält man  $V := 2U - 1 \sim \mathcal{U}[-1, 1]$ . Zwei solcher Zahlen  $V_1, V_2$  definieren einen Punkt im  $\mathbb{R}^2$ . Definiere

$$\mathcal{D} := \{(V_1, V_2) : V_1^2 + V_2^2 < 1\}.$$

Akzeptiere nur solche Tupel  $(U_1, U_2)$ , so dass  $(V_1, V_2) \in \mathcal{D}$ . Diese verbleibenden Punkte sind auf  $\mathcal{D}$  gleichverteilt. Transformation auf (Radius)<sup>2</sup> und normierten Winkel:

$$\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} \sqrt{V_1^2 + V_2^2} \\ \frac{1}{2\pi} \arg(V_1, V_2) \end{pmatrix}.$$

Diese  $(x_1, x_2)$  sind in  $S$  gleichverteilt ( $\rightarrow$  Übung) und werden als Eingabe für Box-Muller verwendet. Vorteilhaft gilt

$$\begin{aligned} \cos(2\pi x_2) &= \frac{V_1}{\sqrt{V_1^2 + V_2^2}} \\ \sin(2\pi x_2) &= \frac{V_2}{\sqrt{V_1^2 + V_2^2}} \end{aligned}$$

### Algorithmus (Polar-Methode)

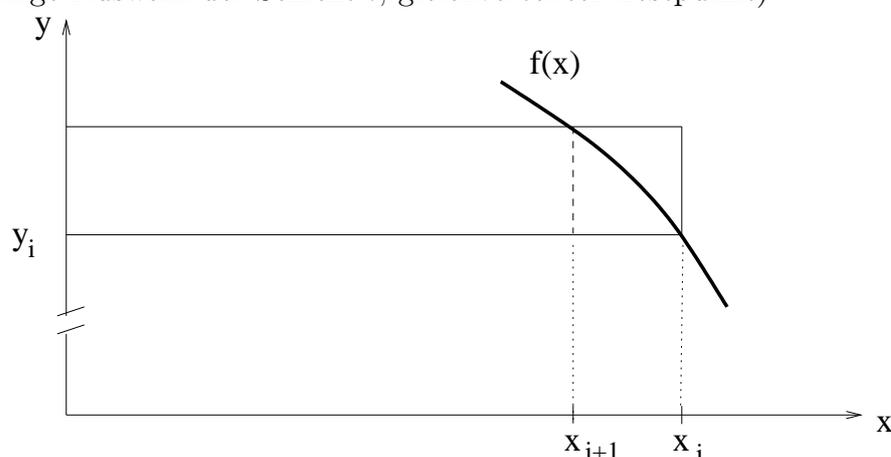
- (1) *Repeat:* generiere  $U_1, U_2 \sim \mathcal{U}[0, 1]$ ;  
 $V_1 := 2U_1 - 1$ ,  $V_2 := 2U_2 - 1$ ;  
*until*  $w := V_1^2 + V_2^2 < 1$ .
- (2)  $Z_1 := V_1 \sqrt{-2 \log(w)/w}$  ist  $\sim \mathcal{N}(0, 1)$   
(ebenso wie  $Z_2 := V_2 \sqrt{-2 \log(w)/w}$ ).

Die Wahrscheinlichkeit, dass  $w < 1$ , ist das Verhältnis der Flächen, also  $\frac{\pi}{4} \approx 0.785\dots$ . D.h. in 21% aller Ziehungen von  $(U_1, U_2)$  gilt  $w \geq 1$ , und die Ziehung wird verworfen. Aber die Einsparung der trigonometrischen Funktionen überwiegt den Nachteil, und Marsaglias Polar Methode ist effizienter als Box-Muller.

### C. Ziggurat Algorithmus

Ein besonders effizienter Algorithmus zur Erzeugung normalverteilter Zufallszahlen ist der Ziggurat-Algorithmus, der auf einer Verwerfungsmethode beruht.

Im Wesentlichen ist die Majorante der Gaussschen Dichte  $f$  eine Stufenfunktion größer oder gleich  $f$ . Die Konstruktion erfolgt mit  $N$  horizontalen Schichten: oben  $N - 1$  Rechtecke gleichen Flächeninhaltes, und unten ein Segment ebenfalls von dem gleichen Flächeninhalt, aber kein Rechteck, da die unendliche Flanke von  $f$  enthalten ist. Die Figur illustriert schematisch die Situation für  $x \geq 0$  und ein  $i$  mit  $0 < i < N - 1$ , wo das Rechteck aus zwei Teilen besteht. Das macht einen besonders effektiven Test auf Akzeptanz oder Verwerfung möglich (zufällige Auswahl der Schicht  $i$ ; gleichverteilter Testpunkt).



## 2.4 Korrelierte Zufallsvariable

Das Ziel sei die Berechnung eines normalverteilten Zufallsvektors  $X = (X_1, \dots, X_n)$  mit vorgegebenen

$$\mu = \mathbf{E}X = (\mathbf{E}X_1, \dots, \mathbf{E}X_n),$$

Kovarianzmatrix

$$\Sigma_{ij} = (\text{Cov}X)_{ij} := \mathbf{E}((X_i - \mu_i)(X_j - \mu_j)); \quad \sigma_i^2 = \Sigma_{ii}$$

und den Korrelationen

$$\rho_{ij} := \frac{\Sigma_{ij}}{\sigma_i \sigma_j}.$$

Für das Folgende sei angenommen, dass  $\Sigma$  symmetrisch und positiv definit ist. Es gilt die lineare Transformations-Eigenschaft für nichtsinguläre Matrizen  $A$ :

$$Z \sim \mathcal{N}(0, I) \Rightarrow AZ \sim \mathcal{N}(0, AA^t)$$

*Begründung:*

Erinnert sei an die Dichtefunktion  $f(x_1, \dots, x_n)$  von  $\mathcal{N}(\mu, \Sigma)$

$$f(x) = \frac{1}{(2\pi)^{n/2}} \frac{1}{(\det \Sigma)^{1/2}} \exp \left\{ -\frac{1}{2} (x - \mu)^t \Sigma^{-1} (x - \mu) \right\}.$$

Angenommen  $Z \sim \mathcal{N}(0, I)$  und  $x = Az$ ,  $A \in \mathbb{R}^{n \times n}$ , wobei  $z$  eine Realisierung von  $Z$  und  $I$  die Einheitsmatrix ist. Wir wenden den Transformationssatz für  $X = h(Z) := AZ$  an: Die Dichte von  $X$  ist

$$\begin{aligned} f(A^{-1}x) |\det(A^{-1})| &= \frac{1}{(2\pi)^{n/2}} \exp \left\{ -\frac{1}{2} (A^{-1}x)^t (A^{-1}x) \right\} \frac{1}{|\det(A)|} \\ &= \frac{1}{(2\pi)^{n/2}} \frac{1}{|\det(A)|} \exp \left\{ -\frac{1}{2} x^t (AA^t)^{-1} x \right\} \end{aligned}$$

für beliebige nichtsinguläre Matrizen  $A$ . Falls  $AA^t$  eine Faktorisierung von  $\Sigma$  ist, also  $\Sigma = AA^t$  und folglich  $|\det A| = (\det \Sigma)^{1/2}$ , dann gilt:

$$AZ \sim \mathcal{N}(0, \Sigma).$$

Folgerung:

$$\mu + AZ \sim \mathcal{N}(\mu, \Sigma).$$

Beispiel: Wähle die Cholesky-Zerlegung von  $\Sigma$ .

Alternative: Eine Vektoren-Analyse der Matrix  $\Sigma$  (*principal component analysis*) ermöglicht eine andere Zerlegung.

### Algorithmus (Korrelierte normalverteilte Zufallsvariablen)

- (1) Berechne zu  $\Sigma$  die Cholesky-Zerlegung  $AA^t = \Sigma$
- (2) Berechne  $Z \sim \mathcal{N}(0, I)$  komponentenweise  
mit  $Z_i \sim \mathcal{N}(0, 1)$  für  $i = 1, \dots, n$ , z.B. mit  
Marsaglias Polar-Methode
- (3)  $\mu + AZ$  hat die gewünschte Verteilung  $\sim \mathcal{N}(\mu, \Sigma)$

**Beispiel:** Wenn  $\Sigma = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}$  gefordert ist, dann wird dies durch

$\begin{pmatrix} \sigma_1 Z_1 \\ \sigma_2 \rho Z_1 + \sigma_2 \sqrt{1 - \rho^2} Z_2 \end{pmatrix}$  erfüllt.

## 2.5 Zahlenfolgen niedriger Diskrepanz

**Ziel:** Konstruktion von Punkten derart, dass die Verteilung ähnlich ist wie bei Zufallszahlen, aber deren Nachteile (Auftreten größerer Lücken oder Klumpen) vermieden werden. **Idee:** Es sei  $Q$  ein Quader in  $[0, 1]^m$ ,  $m \geq 1$ . Ziel einer **gleichmäßigen** (*equidistributed*) Verteilung der Punkte ist

$$\frac{\# \text{ der } x_i \in Q}{\# \text{ aller Punkte in } [0, 1]^m} \approx \frac{\text{vol}(Q)}{\text{vol}([0, 1]^m)}$$

für möglichst alle Quader  $Q$ . Die Konstruktion hat keinen Wahrscheinlichkeits-Hintergrund.

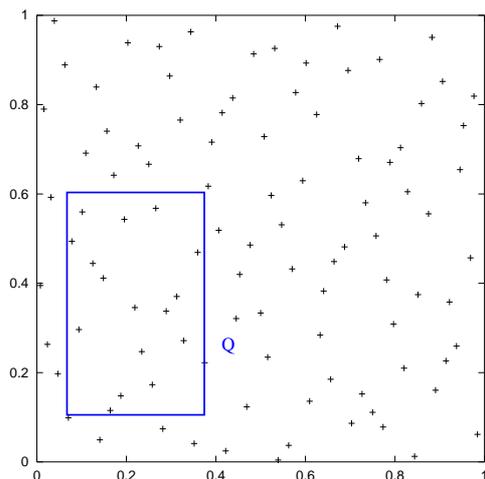


Illustration für  $m = 2$ ; Punkte in  $[0, 1]^2$

### Definition (Diskrepanz)

Die Diskrepanz einer Punktmenge  $\{x_1, \dots, x_N\}$  mit  $x_i \in [0, 1]^m$  ist

$$D_N := \sup_Q \left| \frac{\# \text{ der } x_i \in Q}{N} - \text{vol}(Q) \right|.$$

Das Ziel ist es, Zahlenfolgen (Punktfolgen) zu finden, deren Diskrepanz  $D_N$  für  $N \rightarrow \infty$  möglichst "schnell" gegen 0 geht. Zum Vergleich diene die Folge

$$\frac{1}{\sqrt{N}},$$

welche charakteristisch ist für den probabilistischen Fehler von Monte-Carlo-Methoden. Für *zufällig* verteilte Punkte ist die Diskrepanz von der ähnlicher Größenordnung

$$\sqrt{\frac{\log \log N}{N}}.$$

### Definition (Folge niedriger Diskrepanz)

Eine Folge von Punkten  $x_1, \dots, x_N \dots \in [0, 1]^m$  heißt Folge niedriger Diskrepanz, wenn es eine Konstante  $C_m$  gibt, so dass für alle  $N$  gilt:

$$D_N \leq C_m \frac{(\log N)^m}{N}$$

**Kommentar**

Der Nenner in  $\frac{1}{N}$  beschreibt eine relativ schnelle Abnahme von  $D_N$  mit der Anzahl der Punkte  $N$ , verglichen mit dem  $\frac{1}{\sqrt{N}}$  von Monte-Carlo-Verfahren.

Allerdings: bei der Abnahme gemäß “niedriger Diskrepanz” haben wir noch die Zunahme des Zählers  $(\log N)^m$ . Da  $\log N$  nur sehr bescheiden wächst, ist für niedrige Dimensionen  $m$  die Abnahme von  $D_N$  deutlich schneller als die Abnahme des Fehlers von Monte-Carlo-Verfahren (vgl. Tabelle).

Tabelle verschiedener Nullfolgen:

$N$	$\frac{1}{\sqrt{N}}$	$\sqrt{\frac{\log \log N}{N}}$	$\frac{\log N}{N}$	$\frac{(\log N)^2}{N}$	$\frac{(\log N)^3}{N}$
$10^1$	.31622777	.28879620	.23025851	.53018981	1.22080716
$10^2$	.10000000	.12357911	.04605170	.21207592	.97664572
$10^3$	.03162278	.04396186	.00690776	.04771708	.32961793
$10^4$	.01000000	.01490076	.00092103	.00848304	.07813166
$10^5$	.00316228	.00494315	.00011513	.00132547	.01526009
$10^6$	.00100000	.00162043	.00001382	.00019087	.00263694
$10^7$	.00031623	.00052725	.00000161	.00002598	.00041874
$10^8$	.00010000	.00017069	.00000018	.00000339	.00006251
$10^9$	.00003162	.00005506	.00000002	.00000043	.00000890

**Frage:** Gibt es Zahlenfolgen/Punktfolgen niedriger Diskrepanz?

**Beispiel:** ( $m = 1$ ) **Van der Corput-Folge**

$$\frac{1}{2}, \frac{1}{4}, \frac{3}{4}, \frac{1}{8}, \frac{5}{8}, \frac{3}{8}, \frac{7}{8}, \frac{1}{16}, \dots$$

Bildungsgesetz am Beispiel  $x_6 = \frac{3}{8}$ . Index 6 binär: 110.

radix-invertiert: .011 wird dezimal  $\frac{3}{8}$

**Definition (Radix-inverse Funktion)**

Für  $i = 1, 2, \dots$  sei

$$i = \sum_{k=0}^j d_k b^k$$

die Darstellung von  $i$  zur Basis  $b$  (ganzzahlig  $\geq 2$ ), mit  $d_k \in \{0, 1, \dots, b-1\}$ . Die *Radix-inverse (radical-inverse) Funktion* ist definiert durch

$$\phi_b(i) := \sum_{k=0}^j d_k b^{-k-1} .$$

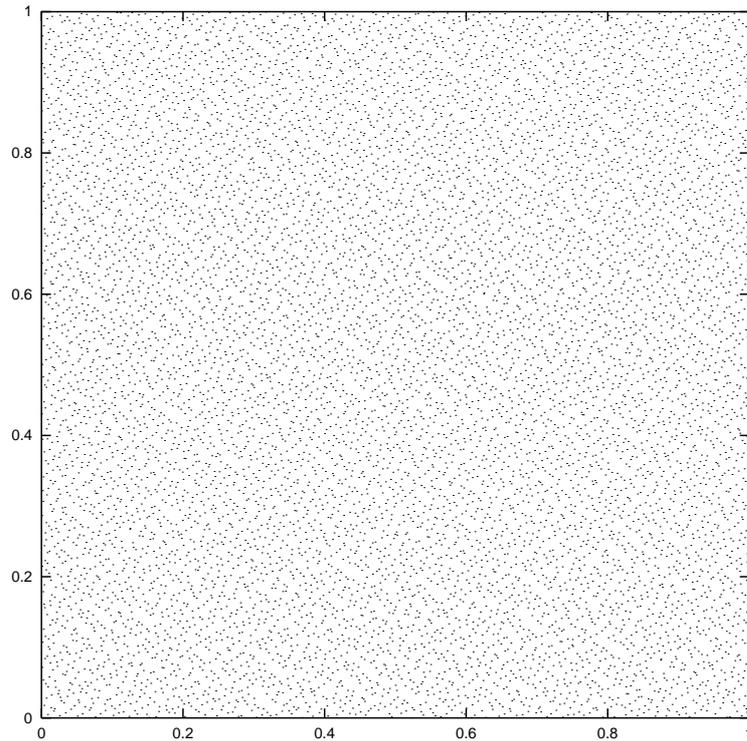
Ein eindimensionales Beispiel:  $x_i := \phi_2(i)$  ist die Van der Corput-Folge.

**Definition (Halton-Folge)**

Es seien  $p_1, \dots, p_m$  natürliche und paarweise teilerfremde Zahlen. Die *Halton-Folge* ist definiert als Folge der Vektoren

$$x_i := (\phi_{p_1}(i), \dots, \phi_{p_m}(i)), \quad i = 1, 2, \dots$$

Die Halton-Folge ist von niedriger Diskrepanz mit  $C_2 = 0.2602$  für  $m = 2$  und ist einfach zu generieren.



Die Abbildung zeigt die ersten 10000 Halton-Punkte mit  $m = 2$  und  $p_1 = 2, p_2 = 3$ .

### Weitere Folgen niedriger Diskrepanz:

- Faure-Folge
- Sobol-Folge
- Niederreiter-Folge
- Halton "leaped": Für große  $m$  hat die Halton-Folge gewisse Nachteile (Korrelationen).  
Abhilfe:

$$x_i := (\phi_{p_1}(li), \dots, \phi_{p_m}(li)), \quad i = 1, 2, \dots$$

für eine geeignete Primzahl  $l$  verschieden von den  $p_k$ , wie z.B.  $l = 409$ .

Die deterministischen Zahlenfolgen niedriger Diskrepanz heißen **Quasi-Zufallszahlen**. (Es sind keine Zufallszahlen!)

Literatur zu Quasi-Zufallszahlen: [H. Niederreiter: Random Number Generation and Quasi-Monte Carlo Methods (1992)]

### 3. Monte-Carlo-Methoden

Kapitel 1 hat die folgende Formel zur risiko-neutralen Bewertung von europäischen Optionen eingeführt:

$$V(S_0, 0) = e^{-rT} \mathbf{E}_Q [\Psi(S_T) \mid S_0] ,$$

wobei  $\Psi(S_T)$  den Payoff bezeichnet. Im Black-Scholes-Modell ergab sich daraus der Spezialfall

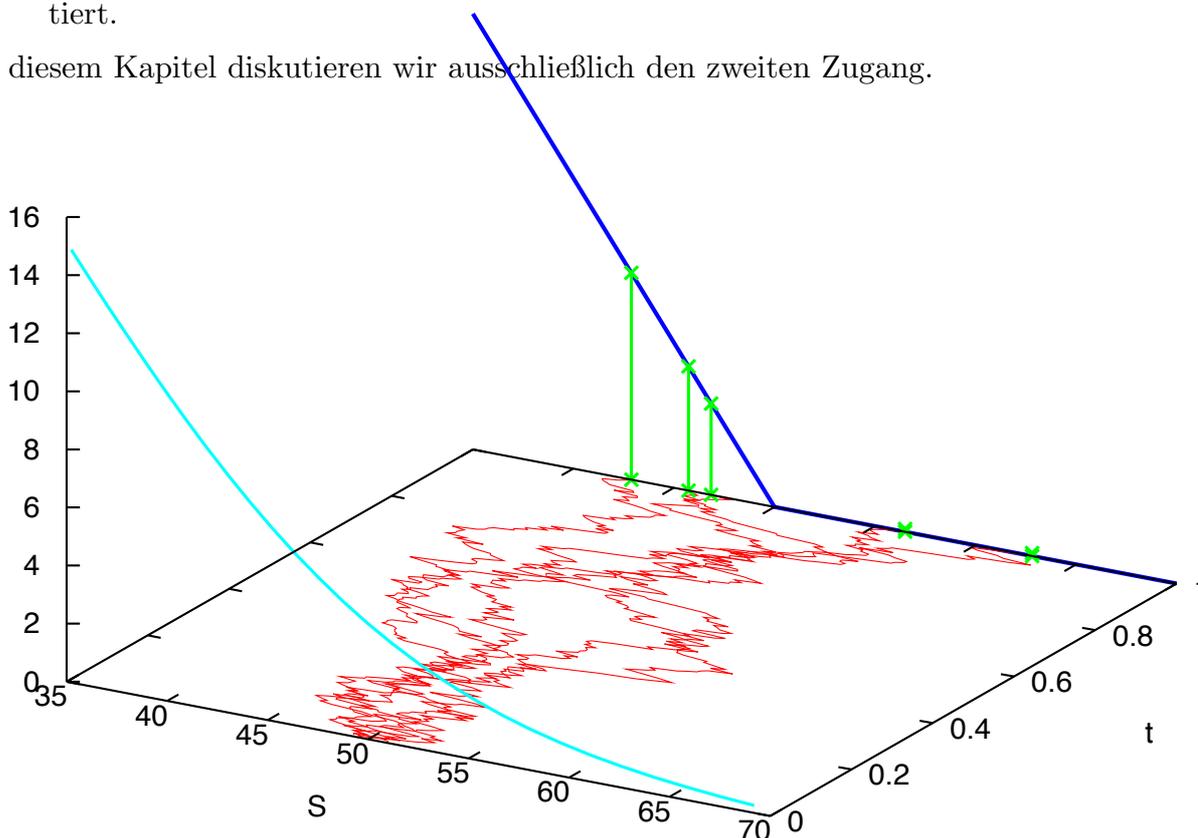
$$V(S_0, 0) = e^{-rT} \int_0^\infty \Psi(S_T) \cdot f_{\text{GBM}}(S_T, T; S_0, r, \sigma) dS_T . \quad (\text{Int})$$

(Übergangsdichte  $f_{\text{GBM}}$  vgl. Abschnitt 1.5D.) Die daraus resultierende PDGL des Black-Scholes-Modells wird in Kapitel 4 diskutiert und gelöst. Solche PDGLen sind für allgemeinere Modelle nicht immer bekannt, oder nicht einfach zu lösen. Deswegen leiten wir Monte-Carlo-Methoden her, die in jedem Fall benutzt werden können.

Es gibt zwei Zugänge, das obige Integral zu berechnen:

- 1) Das Integral (Int) wird mit numerischen Quadratur-Methoden berechnet.
- 2) Man wendet Monte-Carlo-Simulationen an. D.h. man "spielt" mit Hilfe von Zufallszahlen Asset-Pfade  $S_t$  durch unter den angenommenen (risiko-neutralen) Wahrscheinlichkeitsverteilungen, die zu dem jeweiligen Modell passen. Dies ist der Hauptteil der Simulation; anschließend wird der Mittelwert der Payoff-Werte gebildet und diskontiert.

In diesem Kapitel diskutieren wir ausschließlich den zweiten Zugang.



Fünf simulierte Börsenkurs-Pfade mit Payoff-Werten.

### Bezeichnungen (Wiederholung)

Das einfache skalare Modell mit Antrieb durch einen Wiener-Prozess ist

$$dX_t = a(X_t, t) dt + b(X_t, t) dW_t. \quad (\text{SDE})$$

Wir diskretisieren die Zeit  $t$  mit einem Gitter

$$\dots < t_{j-1} < t_j < t_{j+1} < \dots,$$

mit äquidistanter Gitterweite  $h$  bzw.  $\Delta t = t_{j+1} - t_j$ . Es sei  $y_j$  eine Approximation zu  $X_{t_j}$ , wobei  $y_0 := X_0$ .

Beispiel: Aus Kapitel 1 kennen wir die **Euler-Diskretisierung**

$$\begin{aligned} y_{j+1} &= y_j + a(y_j, t_j) \Delta t + b(y_j, t_j) \Delta W_j, \\ t_j &= j \Delta t, \\ \Delta W_j &= W_{t_{j+1}} - W_{t_j} = Z \sqrt{\Delta t} \\ \text{mit } Z &\sim \mathcal{N}(0, 1). \end{aligned} \quad (\text{Euler})$$

## 3.1 Approximations-Fehler

### Definition

Für einen gegebenen Pfad des Wiener-Prozesses  $W_t$  nennen wir eine Lösung  $X_t$  von (SDE) eine **starke Lösung**. Falls der Wiener-Prozess frei ist, so nennen wir  $X_t$  bzw.  $(X_t, W_t)$  eine **schwache Lösung**.

Bei starken Lösungen wird die numerische Diskretisierung mit dem *gleichen*  $W_t$  berechnet, welches für die SDE zugrunde liegt. Dies hat den Vorteil, dass die pfadweise Differenz  $X_t - y_t$  auf Konvergenzverhalten für  $h \rightarrow 0$  untersucht werden kann.

**Bezeichnung:** Schreibe  $y_t^h$  für eine numerisch mit Schrittweite  $h$  berechnete Näherung  $y$  an der Stelle  $t$ , insbesondere für  $t = T$ .

### Definition (absoluter Fehler)

Für eine starke Lösung  $X_t$  von (SDE) und eine Näherung  $y_t^h$  ist der absolute Fehler für  $t = T$

$$\varepsilon(h) := \mathbf{E} [ |X_T - y_T^h| ].$$

Dieser Wert kann bei GBM, bei der die analytische Lösung  $X_t$  bekannt ist, leicht empirisch ermittelt werden:

Setze hierzu in (SDE)  $a(X_t, t) = \alpha X_t$  und  $b(X_t, t) = \beta X_t$ . So erhält man mit der Lösungsformel aus Abschnitt 1.5D für gegebenes  $W_T$  den Wert

$$X_T = X_0 \exp \left[ \left( \alpha - \frac{\beta^2}{2} \right) T + \beta W_T \right].$$

Ein Schätzer für den Erwartungswert  $\varepsilon(h)$  ist der Mittelwert über eine große Anzahl von Auswertungen von  $|X_T - y_T^h|$ . Dies liefert beim Euler-Verfahren empirisch als Ergebnis

$$\varepsilon(h) = O(h^{\frac{1}{2}}),$$

eine geringe Genauigkeit im Vergleich zum  $O(h)$  des deterministischen Falls. Dieses Resultat ist plausibel weil  $\Delta W$  von der Ordnung  $O(\sqrt{h})$  ist (in Wahrscheinlichkeit), vgl. Abschnitt 1.4.

### Definition (starke Konvergenz)

$y_T^h$  konvergiert *stark* gegen  $X_T$  mit Ordnung  $\gamma > 0$ , wenn

$$\varepsilon(h) = \mathbb{E}[|X_T - y_T^h|] = O(h^\gamma).$$

Man sagt,  $y_T^h$  konvergiert *stark*, wenn

$$\lim_{h \rightarrow 0} \mathbb{E}[|X_T - y_T^h|] = 0.$$

### Beispiel

Wenn  $a$  und  $b$  globalen Lipschitzbedingungen und einer Beschränkung des Wachstums genügen, dann hat die Euler-Diskretisierung die Ordnung  $\gamma = \frac{1}{2}$ .

**Vorsicht:** Für etliche in den Finanzanwendungen wichtige Beispiele (wie CIR, Heston) gilt die globale Lipschitzbedingung *nicht*. Modifikationen vom Standard-Euler-Verfahren sind dann ratsam, auch um  $S_t \geq 0$  zu garantieren.

Was ist mit den schwachen Lösungen?

In vielen Fällen der Praxis sind die jeweiligen Pfade von  $X_t$  ohne Interesse. Stattdessen interessiert uns nur ein *Moment* von  $X_T$ , d.h. entweder  $\mathbb{E}[X_T]$  oder  $\text{Var}[X_T]$ , ohne Wert auf die *einzelnen* Werte von  $X_T$  zu legen. Bei Optionen interessiert  $\mathbb{E}[\Psi(X_T)]$ .

### Definition (schwache Konvergenz)

$y_T^h$  konvergiert *schwach* gegen  $X_T$  bzgl. einer Funktion  $g$  mit Ordnung  $\beta > 0$ , wenn

$$\mathbb{E}[g(X_T)] - \mathbb{E}[g(y_T^h)] = O(h^\beta),$$

und konvergiert schwach mit Ordnung  $\beta$ , wenn dies für alle Polynome  $g$  gilt.

### Beispiel

Wenn  $a$  und  $b$  viermal stetig differenzierbar sind, dann ist die Euler-Methode von schwacher Konvergenz mit Ordnung  $\beta = 1$ .

### Bedeutung von $g$

Wenn die Konvergenzordnung  $\beta$  für alle Polynome gilt, so folgt u.a. Konvergenz aller Momente.

*Beweis* (für die ersten beiden Momente):

(a) Für  $g(x) := x$  gilt

$$\mathbb{E}[X_T] - \mathbb{E}[y_T^h] = O(h^\beta),$$

d.h. Konvergenz der Mittel.

- (b) Falls die Konvergenzordnung zusätzlich für  $g(x) := x^2$  gilt (mit Bezeichnungen  $y := y_T^h$  und  $X := X_T$ ),

$$\begin{aligned} |\text{Var}[X_T] - \text{Var}[y_T^h]| &= |\mathbb{E}[X^2] - \mathbb{E}[y^2] - (\mathbb{E}[X])^2 + (\mathbb{E}[y])^2| \\ &\leq \underbrace{|\mathbb{E}[X^2] - \mathbb{E}[y^2]|}_{=O(h^\beta)} + \underbrace{|\mathbb{E}[X] + \mathbb{E}[y]|}_{\leq \text{const}} \cdot \underbrace{|\mathbb{E}[X] - \mathbb{E}[y]|}_{=O(h^\beta)}, \end{aligned}$$

d.h. Konvergenz der Varianz.

### Bemerkung

Starke Konvergenz impliziert schwache Konvergenz bzgl.  $g(x) = x$ .

Denn aus den Eigenschaften der Integration

$$|\mathbb{E}[X] - \mathbb{E}[Y]| = |\mathbb{E}[X - Y]| \leq \mathbb{E}[|X - Y|]$$

folgt

$$\mathbb{E}[|X - Y|] = O(h^\gamma) \implies \mathbb{E}[X] - \mathbb{E}[Y] = O(h^\gamma).$$

### Praktischer Vorteil der schwachen Konvergenz:

Die Inkremente  $\Delta W$  für die Berechnung von  $y^h$  können ersetzt werden durch andere Zufallsvariablen  $\widehat{\Delta W}$  mit übereinstimmenden ersten Momenten. Die schwache Konvergenzordnung bleibt erhalten.

### Beispiel

$\widehat{\Delta W} := \pm\sqrt{\Delta t}$ , wobei beide Vorzeichen jeweils mit Wahrscheinlichkeit 1/2 angenommen werden. (billiger zu approximieren als die Berechnung von  $Z \sim \mathcal{N}(0, 1)$ )

Folgerung: u.a.  $\mathbb{E}(\widehat{\Delta W}) = 0$  und  $\mathbb{E}((\widehat{\Delta W})^2) = \Delta t$  ( $\implies \text{Var}(\widehat{\Delta W}) = \Delta t$ )

Mit diesem  $\widehat{\Delta W}$  statt  $\Delta W$  ergibt sich das "vereinfachte Euler-Verfahren". Es ist schwach konvergent mit Ordnung 1.

## 3.2 Konstruktion von Integratoren für SDEs

Ein wichtiges Hilfsmittel zur Konstruktion von Integratoren für SDEs ist die Stochastische Taylorentwicklung.

### A. Stochastische Taylorentwicklung

(nach [P. Kloeden & E. Platen: Numerical Solution of SDEs])

Zur Motivation betrachten wir zunächst den deterministischen autonomen Fall

$$\frac{d}{dt}X_t = a(X_t).$$

Die Kettenregel für beliebiges  $f \in C^1(\mathbb{R})$  besagt

$$\begin{aligned}\frac{d}{dt}f(X_t) &= \frac{df(X)}{dX} \cdot \frac{dX}{dt} = \frac{df(X)}{dX}a(X_t) \\ &= a(X_t)\frac{d}{dX}f(X_t) =: Lf(X_t). \\ \implies f(X_t) &= f(X_{t_0}) + \int_{t_0}^t \underbrace{Lf(X_s)}_{=: \tilde{f}} ds\end{aligned}$$

Diese Version wird mit

$$\tilde{f}(X_s) := Lf(X_s)$$

in sich selbst eingesetzt:

$$\begin{aligned}f(X_t) &= f(X_{t_0}) + \int_{t_0}^t \left\{ \tilde{f}(X_{t_0}) + \int_{t_0}^s L\tilde{f}(X_z) dz \right\} ds \\ &= f(X_{t_0}) + \tilde{f}(X_{t_0}) \int_{t_0}^t ds + \int_{t_0}^t \int_{t_0}^s L\tilde{f}(X_z) dz ds \\ &= f(X_{t_0}) + Lf(X_{t_0})(t - t_0) + \int_{t_0}^t \int_{t_0}^s L^2 f(X_z) dz ds.\end{aligned}$$

Dies ist die Taylorentwicklung mit Restglied in Integralform, hier entwickelt bis einschließlich des linearen Terms; das Restglied ist ein Doppelintegral. Setzt man diesen Prozess fort, erhält man die deterministische Taylorentwicklung mit Restglied in Integralform. (Alle auftretenden Ableitungen mögen existieren.)

Nun der **stochastische Fall**, d.h. die Itô-Taylor-Entwicklung:

Das Itô-Lemma auf  $f(X)$  und die autonome SDE

$$dX_t = a(X_t) dt + b(X_t) dW_t$$

angewendet ergibt

$$df(X_t) = \underbrace{\left\{ a(X_t)\frac{\partial}{\partial x}f(X_t) + \frac{1}{2}(b(X_t))^2\frac{\partial^2}{\partial x^2}f(X_t) \right\}}_{=: L^0 f(X_t)} dt + \underbrace{b(X_t)\frac{\partial}{\partial x}f(X_t)}_{=: L^1 f(X_t)} dW_t,$$

d.h.

$$\boxed{f(X_t) = f(X_{t_0}) + \int_{t_0}^t L^0 f(X_s) ds + \int_{t_0}^t L^1 f(X_s) dW_s.} \quad (*)$$

Im Spezialfall  $f(x) = x$  enthält (\*) die Ausgangs-SDE

$$X_t = X_{t_0} + \int_{t_0}^t a(X_s) ds + \int_{t_0}^t b(X_s) dW_s.$$

Wende nun (\*) an für geeignete  $\tilde{f}$ , zunächst mit  $\tilde{f} := a$  und  $\tilde{f} := b$ , und erhalte

$$X_t = X_{t_0} + \int_{t_0}^t \left\{ a(X_{t_0}) + \int_{t_0}^s L^0 a(X_z) dz + \int_{t_0}^s L^1 a(X_z) dW_z \right\} ds \\ + \int_{t_0}^t \left\{ b(X_{t_0}) + \int_{t_0}^s L^0 b(X_z) dz + \int_{t_0}^s L^1 b(X_z) dW_z \right\} dW_s.$$

Dies schreiben wir

$$X_t = X_{t_0} + a(X_{t_0}) \int_{t_0}^t ds + b(X_{t_0}) \int_{t_0}^t dW_s + R,$$

mit

$$R = \int_{t_0}^t \int_{t_0}^s L^0 a(X_z) dz ds + \int_{t_0}^t \int_{t_0}^s L^1 a(X_z) dW_z ds \\ + \int_{t_0}^t \int_{t_0}^s L^0 b(X_z) dz dW_s + \int_{t_0}^t \int_{t_0}^s L^1 b(X_z) dW_z dW_s.$$

Die Integranden sind

$$L^0 a = aa' + \frac{1}{2} b^2 a'' \quad L^0 b = ab' + \frac{1}{2} b^2 b'' \\ L^1 a = ba' \quad L^1 b = bb'.$$

In analoger Weise können die Integranden der Doppelintegrale in  $R$  durch Anwendung von (\*) mit geeignetem  $\tilde{f}$  ersetzt werden. Dabei treten die Doppelintegrale

$$\underbrace{\int_{t_0}^t \int_{t_0}^s dz ds}_{=: \mathcal{I}(0,0) = \frac{1}{2} (\Delta t)^2}, \quad \underbrace{\int_{t_0}^t \int_{t_0}^s dW_z ds}_{=: \mathcal{I}(1,0)}, \quad \underbrace{\int_{t_0}^t \int_{t_0}^s dz dW_s}_{=: \mathcal{I}(0,1)}, \quad \underbrace{\int_{t_0}^t \int_{t_0}^s dW_z dW_s}_{=: \mathcal{I}(1,1)}$$

als Faktoren auf.  $\mathcal{I}(1,0)$ ,  $\mathcal{I}(0,1)$ ,  $\mathcal{I}(1,1)$  sind stochastische Variable. Mit einer Plausibilitätsbetrachtung (ersetze  $\Delta W_s := W_s - W_{t_0}$  durch den Erwartungswert  $\sqrt{s - t_0}$ ) erwarte, dass  $\mathcal{I}(1,1)$  das Integral von niedrigster Ordnung ist:  $O(\Delta t)$ . Wir beginnen mit dem Integral niedrigster Ordnung, also demjenigen mit dem Integranden  $\tilde{f} := L^1 b(X)$ . Aus (\*) folgt dann

$$\int_{t_0}^t \int_{t_0}^s L^1 b(X_z) dW_z dW_s = L^1 b(X_{t_0}) \int_{t_0}^t \int_{t_0}^s dW_z dW_s + \text{zwei Dreifachintegrale}$$

Damit gilt

$$R = \text{drei Doppelintegrale (s.o.)} + \underbrace{b(X_{t_0})b'(X_{t_0})}_{=: L^1 b(X_{t_0})} \mathcal{I}(1,1) + \text{zwei Dreifachintegrale}$$

Berechnung des Doppelintegrales  $\mathcal{I}(1,1)$ :

Es sei  $g(x) := x^2$  und  $X_t = W_t$ , d.h. SDE mit  $a = 0$  und  $b = 1$ . Mit dem Itô-Lemma erhält man

$$d(W_t^2) = \frac{1}{2} 2 dt + 2W_t dW_t = dt + 2W_t dW_t,$$

und daraus wiederum

$$\begin{aligned} \int_{t_0}^t \int_{t_0}^s dW_z dW_s &= \int_{t_0}^t (W_s - W_{t_0}) dW_s = \int_{t_0}^t W_s dW_s - W_{t_0} \int_{t_0}^t dW_s = \\ &= \int_{t_0}^t \frac{1}{2} [d(W_s^2) - ds] - W_{t_0}(W_t - W_{t_0}) = \\ &= \frac{1}{2}(W_t^2 - W_{t_0}^2) - \frac{1}{2}(t - t_0) - \frac{2}{2}W_{t_0}(W_t - W_{t_0}) = \frac{1}{2}(\Delta W_t)^2 - \frac{1}{2}\Delta t. \end{aligned}$$

Dies bestätigt die erwartete Ordnung  $O(\Delta t)$ .

Für die allgemeine Herleitung der stochastischen Taylorentwicklung setze die begonnene systematische Definition der nach obigem Prozess entstehenden Mehrfachintegrale fort, wie zum Beispiel  $\mathcal{I}(0, 0, 0), \dots$ . Dabei steht “0” für eine deterministische und eine “1” für eine stochastische Integration.

### Anwendung auf die Numerik

Durch Hinzufügen weiterer führender Terme der stochastischen Taylorentwicklung erhält man Integrationsverfahren höherer Ordnung.

#### Beispiel (Milstein-Verfahren)

$$y_{j+1} = y_j + a \Delta t + b \Delta W_j + \frac{1}{2}bb' \{(\Delta W_j)^2 - \Delta t\}$$

Die ersten Summanden entsprechen der Euler-Methode, und der letzte ist ein Korrekturterm, der die Liste der  $O(\Delta t)$ -Terme vervollständigt und ein Verfahren der (starken) Ordnung 1 entstehen lässt. (Nachprüfung z.B. empirisch) Die schwache Ordnung ist ebenfalls 1. Frage: Was bedeutet der Fall  $b' = 0$  für das Euler-Verfahren?

## B. Positivität

Wie schon in Abschnitt 3.1 erwähnt, will man bei Finanz-relevanten SDEs häufig sicherstellen, dass die numerischen Näherungen nicht negativ werden. Beispielhaft erläutern wir Möglichkeiten für den wichtigen CIR-Prozess, Baustein auch des Heston-Modells.

#### Beispiel CIR

$$dX_t = \kappa(\theta - X_t) dt + \sigma\sqrt{X_t} dW_t$$

mit  $\kappa, \theta, \sigma > 0$ ,  $X_0 = x_0 > 0$ . Positivität von  $X_t$  für alle  $t$  wird durch die “Feller-Bedingung”

$$\kappa\theta \geq \frac{1}{2}\sigma^2$$

erreicht, die ein genügend starkes Wachstum quantifiziert. — Hinweis:  $b(X) = \sigma\sqrt{X}$  erfüllt keine globale Lipschitz-Bedingung.

Euler-Schema:

$$y_{j+1} = y_j + \kappa(\theta - y_j)\Delta t + \sigma\sqrt{y_j} \Delta W_j$$

mit  $y_0 := x_0$ , solange  $y_j \geq 0$ . Die Wahrscheinlichkeit, dass  $y_{j+1}$  negativ wird, ist positiv. Wenn  $X$  einen Aktienkurs, einen Zinssatz, oder eine Varianz (beim Heston-Modell) repräsentiert, dann macht  $y < 0$  keinen Sinn.

### Varianten

Ersetze zum Beispiel  $\sqrt{y}$  durch  $\sqrt{|y|}$  oder durch  $\sqrt{y^+}$ , womit das Schema für alle  $y \in \mathbb{R}$  definiert ist. Eine andere Variante berechnet

$$y_{j+1} = |y_j + \kappa(\theta - y_j)\Delta t + \sigma\sqrt{y_j} \Delta W_j|.$$

Erfolgreich werden auch implizite Euler-Methoden angewendet, wie zum Beispiel das Drift-implizite Schema

$$y_{j+1} = y_j + a(y_{j+1})\Delta t + b(y_j) \Delta W_j.$$

Wendet man dieses Schema an auf die SDE für den Prozess  $\sqrt{X_t}$ , dann ergibt sich eine quadratische Gleichung für  $y_{j+1}$  mit eindeutiger positiver Lösung (Übung!). [A. Alfonsi (2005)]

## 3.3 Monte-Carlo-Methoden bei europäischen Optionen

Wir wollen nun den Wert

$$V(S_0, 0) = e^{-rT} \mathbf{E}_Q [\Psi(S_T) | S(0) = S_0]$$

einer europäischen Option berechnen, wobei  $\Psi$  den Payoff bezeichnet und  $Q$  ein risikofreies Wahrscheinlichkeitsmaß ist.

### A. Grundprinzip

Dieses Integral kann mit Monte-Carlo-Methoden approximiert werden. Dazu muss zunächst entschieden werden, welches Modell zugrunde gelegt wird (z.B. Heston- oder Black-Scholes-Modell). Wir betrachten hier insbesondere das Black-Scholes-Modell mit GBM

$$dS_t = S_t (r dt + \sigma dW_t).$$

Der Rest erfolgt analog zur

#### Monte-Carlo-Quadratur:

Zu approximieren sei das bestimmte Integral

$$\int_0^1 f(x) dx = \int_{-\infty}^{\infty} f(x) \mathbf{1}_{[0,1]} dx = \mathbf{E}(f),$$

wobei  $\mathbf{1}_{[0,1]}$  die Dichte der Gleichverteilung  $\mathcal{U}[0, 1]$  ist. Anwendung des Gesetzes der großen Zahlen ergibt

$$\frac{1}{N} \sum_{k=1}^N f(x_k) \longrightarrow \mathbf{E}(f) \quad \text{für } N \rightarrow \infty,$$

wobei  $x_k$  zufällig gezogene unabhängige und gleichverteilte Punkte im Definitionsbereich  $\mathcal{D} := [0, 1]$  sind. Also kann

$$\frac{1}{N} \sum_{k=1}^N f(x_k)$$

als Näherung zum Integral berechnet werden.

Für allgemeinere Definitionsbereiche  $\mathcal{D}$  gilt:

$$\int_{\mathcal{D}} f(x) dx \approx \frac{\text{Vol}(\mathcal{D})}{N} \sum_{k=1}^N f(x_k).$$

Der **Fehler** ist von probabilistischer Natur (vgl. Zentraler Grenzwertsatz), d.h. es gelten Aussagen wie:

Mit 95% Wahrscheinlichkeit liegt der wahre Wert des Integrals in dem “Konfidenzintervall”, welches durch die halbe Breite  $a\sigma/\sqrt{N}$  gegeben ist. Dabei ist  $\sigma$  die Standardabweichung, und  $a = 1.96$  (bei 95%).

Soweit das allgemeine MC-Prinzip, angewendet auf Quadratur.

**Frage:** Welche Struktur hat  $f(x)$  bei der Bewertung von Optionen unter GBM?

Verteilung: Die Dichte ist hier  $f_{\text{GBM}}$ , entsprechend müssen die  $x_k$  lognormal-verteilt sein.

### Algorithmus: Monte-Carlo-Verfahren bei europäischen Optionen

Führe  $N$  Simulationen des Aktienkurses entsprechend dem risikoneutralen Maß  $\mathbb{Q}$  aus, jeweils startend mit  $S_0$ , und erhalte so  $x_k := (S_T)_k$  für  $k = 1, \dots, N$ . Werte jeweils den Payoff  $\Psi(S_T)$  aus:

$$f(x_k) := \Psi((S_T)_k),$$

mit anschließender Mittelwertbildung und dann Diskontierung mit Faktor  $e^{-rT}$ . Der Erwartungswert ergibt den wahren Wert  $V$ , vorausgesetzt,  $f$  (also  $S_T$  und  $\Psi$ ) werden verzerrungsfrei berechnet.

(Illustration zu Beginn des Kapitels)

### Beispiele für den Payoff $\Psi$

1.) *Binary*- oder *Digital*-Option, z.B. Binär-Call:

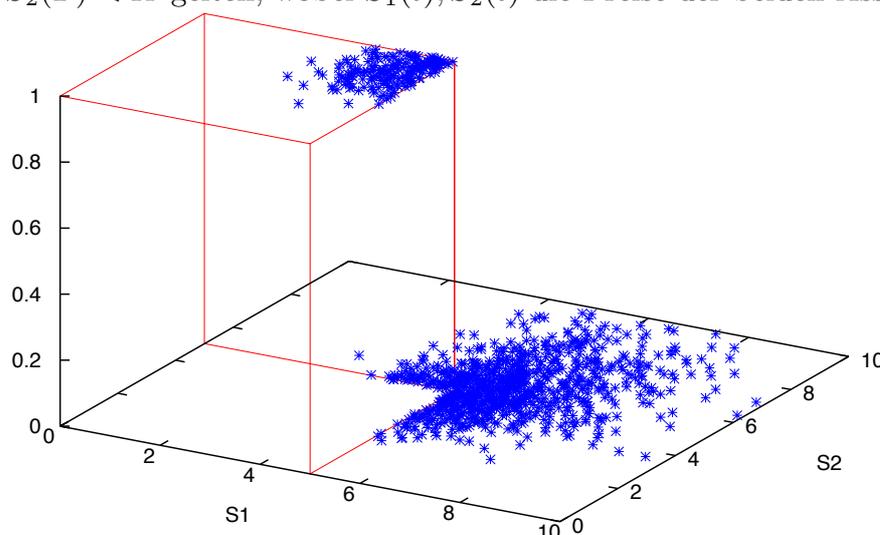
$$\Psi(S_T) = \mathbf{1}_{S_T > K} = \begin{cases} 1 & \text{falls } S_T > K \\ 0 & \text{sonst} \end{cases}$$

2.) *Barrier*-Option mit Barriere  $B$ , z.B. die *down-and-out* Call Option mit

$$\Psi(S) = \begin{cases} 0 & \text{falls } S_t \leq B \text{ für ein } t \text{ in } 0 \leq t \leq T \\ (S_T - K)^+ & \text{sonst} \end{cases}$$

Bei dieser pfadabhängigen exotischen Option ist der gesamte Kursverlauf  $S_t$  auf  $0 \leq t \leq T$  von Interesse. (sinnvoll nur für  $S_0 > B > K$ ; Illustration dreidimensional im  $(S, t, V)$ -Raum unter Berücksichtigung der Randbedingung entlang  $S = B$ )

- 3.) *two-asset cash-or-nothing Put*: Die Auszahlung ist 1 wenn die Ungleichungen  $S_1(T) < K_1$  und  $S_2(T) < K_2$  gelten, wobei  $S_1(t), S_2(t)$  die Preise der beiden Assets sind.



Hinweis: Viele analytische Lösungsformeln finden sich in [E.G. Haug: Option Pricing Formulas]; weitere Figuren in [www.compfin.de](http://www.compfin.de).

### Ausführung der Monte-Carlo-Methode

Beim GBM-Modell kann die analytische Lösungsformel

$$S_t = S_0 \exp \left\{ \left( r - \frac{1}{2} \sigma^2 \right) t + \sigma W_t \right\}$$

verwendet werden. Bei nicht-pfadabhängigen Optionen muss dann nur für  $t = T$  für jeden generierten Pfad  $(S_t)_i$  der Zufallsgenerator einmal benutzt werden, um  $W_T$  und damit  $S_T$  zu ermitteln. Alternativ, wichtig bei allgemeineren Modellen, bei denen keine analytische Formel existiert, muss (z.B. mit dem Euler-Verfahren) numerisch integriert werden. Das Monte-Carlo-Verfahren besitzt dann außer der äußeren Schleife ( $k = 1, \dots, N$ ) noch eine innere Schleife ( $j = 1, \dots, M$ ), wobei  $\Delta t = \frac{T}{M}$  die Schrittweite des Integrators ist. Wenn GBM angenommen wird, kann bei pfadabhängigen Optionen auch stückweise für jedes  $t_j$  die analytische Formel verwendet werden:

$$S_{t_{j+1}} = S_{t_j} \exp \left\{ \left( r - \frac{1}{2} \sigma^2 \right) \Delta t + \sigma \Delta W \right\}$$

mit  $\Delta W = \sqrt{\Delta t} Z$ ,  $Z \sim \mathcal{N}(0, 1)$ .

**Dimensionsabhängigkeit:** Monte-Carlo-Methoden arbeiten genauso im Fall höherer Dimensionen. Der Aufwand ist weitgehend unabhängig von der Dimension. Dies ist ein wichtiger Vorteil von Monte-Carlo-Methoden.

## B. Genauigkeit

a) Es sei

$$\hat{\mu} := \frac{1}{N} \sum_{k=1}^N f(x_k), \quad \mu \text{ der Erwartungswert, und}$$

$$\hat{s}^2 := \frac{1}{N-1} \sum_{k=1}^N (f(x_k) - \hat{\mu})^2.$$

Die Näherung  $\hat{\mu}$  verhält sich nach dem **Zentralen Grenzwertsatz** wie  $\mathcal{N}(\mu, \sigma^2)$ ,

$$P\left(\hat{\mu} - \mu \leq a \frac{\sigma}{\sqrt{N}}\right) = F(a),$$

mit Verteilungsfunktion  $F$ . In der Praxis wird  $\sigma^2$  durch die Näherung  $\hat{\sigma}^2$  ersetzt. Der Fehler verhält sich also wie  $\frac{\hat{\sigma}}{\sqrt{N}}$ . Um diesen statistischen Fehler verringern zu können, muss entweder der Zähler kleiner werden (*Varianzreduktion*) oder der Nenner größer, d.h. mehr Simulationen durchgeführt werden. Die zweite Möglichkeit ist sehr aufwändig, denn um z.B. eine weitere korrekte Dezimalstelle zu gewinnen, muss der Fehler um den Faktor  $\frac{1}{10}$  reduziert werden, woraus eine *Aufwandssteigerung um den Faktor*  $100 = \left(\frac{1}{10}\right)^{-2}$  folgt.

- b) Wenn die Berechnung von  $f(x_i)$  erwartungstreu ist, ist der obige Fehler der einzige. Ansonsten kommt ein weiterer Fehler hinzu, nämlich die **Verzerrung** (*bias*).

Es sei  $\hat{x}$  ein Schätzer für den zu schätzenden wahren Wert  $x$ , so ist die Verzerrung definiert als

$$\text{bias}(\hat{x}) := E[\hat{x}] - x.$$

## Beispiele

- 1.) Bei einer *Lookback*-Option tritt im Payoff der Term

$$x := E\left[\max_{0 \leq t \leq T} S_t\right]$$

auf. Eine Approximation hierzu ist

$$\hat{x} := \max_{0 \leq j \leq M} S_{t_j}.$$

Klar ist  $\hat{x} \leq x$ .  $\hat{x}$  wird  $x$  fast sicher unterschätzen, d.h.  $E[\hat{x}] < x$  fast sicher.  $\hat{x}$  liefert also nur verzerrte Ergebnisse, hier gilt  $\text{bias}(\hat{x}) \neq 0$ .

- 2.) Im Vergleich zur analytischen Lösung liefert auch das Euler-Verfahren verzerrte Ergebnisse. Bei GBM ist die Verwendung von

$$S_{t_{j+1}} = S_{t_j} \exp\left\{\left(r - \frac{1}{2}\sigma^2\right)\Delta t + \sigma \Delta W\right\}$$

erwartungstreu (unverzerrt), während der Eulerschritt

$$S_{t_{j+1}} = S_{t_j}(1 + r \Delta t + \sigma \Delta W)$$

verzerrt ist (*biased*).

Diese beiden Beispiele sind "asymptotisch unverzerrt", weil die Verzerrung mit  $M \rightarrow \infty$  verschwindet. Es stellt sich also die Frage, wo man beim Verringern der Fehler mehr Aufwand betreiben sollte: entweder mit

- Varianzreduktion,
- eine größere Anzahl  $N$  an Pfaden simulieren, oder
- bias verkleinern, d.h.  $M$  vergrößern ( $\Delta t$  verkleinern),

oder mit mehreren dieser Maßnahmen. Eine Vergrößerung von  $N$  und  $M$  sollte ausgewogen erfolgen. Der Gesamtfehler wird wiedergegeben durch den mittleren quadratischen Fehler:

**Definition (mittlerer quadratischer Fehler)**

$$\text{MSE}(\hat{x}) := \text{mean square error} := \mathbb{E} [(x - \hat{x})^2].$$

Eine einfache Rechnung zeigt:

$$\begin{aligned} \text{MSE}(\hat{x}) &= (\mathbb{E}[\hat{x}] - x)^2 + \mathbb{E} [(\hat{x} - \mathbb{E}(\hat{x}))^2] \\ &= (\text{bias}(\hat{x}))^2 + \text{Var}(\hat{x}). \end{aligned}$$

**C. Varianzreduktion**

Die einfachste (aber vielleicht am wenigsten wirksame) der vielen möglichen Methoden der Varianzreduktion ist die Methode der Antithetischen Variablen (*antithetic variates*). Diese basiert auf folgender Idee: Einerseits werden “normal” mit Zufallszahlen  $Z_1, Z_2, \dots$  Pfade  $S_t$  berechnet mit einer MC-Näherung, die wir hier mit  $\hat{V}$  bezeichnen. Wir verwenden jeweils parallel die Zahlen  $-Z_1, -Z_2, \dots$ , welche auch  $\sim \mathcal{N}(0, 1)$  sind, um gespiegelte “Zwillings-Pfade”  $S_t^-$  zu berechnen, aus welchen die Payoff-Werte  $\Psi(S_T^-)$  ermittelt werden. So erhalten wir einen *zweiten* Monte-Carlo-Wert  $V^-$ . Nach Konstruktion gilt  $\text{Var}(\hat{V}) = \text{Var}(V^-)$ . Der Aufwand zur Berechnung von  $V^-$  ist etwas geringer als der von  $\hat{V}$ , weil die  $Z$  wiederverwendet werden. Für den Mittelwert

$$V_{\text{AV}} := \frac{1}{2}(\hat{V} + V^-)$$

folgt

$$\begin{aligned} \text{Var}(V_{\text{AV}}) &= \frac{1}{4}\text{Var}(\hat{V} + V^-) \\ &= \frac{1}{4}(\text{Var}\hat{V} + \text{Var}V^- + 2\text{Cov}(\hat{V}, V^-)) \\ &= \frac{1}{2}\text{Var}\hat{V} + \frac{1}{2}\text{Cov}(\hat{V}, V^-) \end{aligned}$$

Wegen der anti-symmetrischen Konstruktion der Pfade erhofft man eine negative Korrelation im Sinn  $\text{Cov}(\hat{V}, V^-) < 0$ . Dies gilt falls der Output  $V$  monoton vom Input  $Z$  abhängt. Für  $\text{Cov}(\hat{V}, V^-) < 0$  folgt

$$\text{Var}(V_{\text{AV}}) < \frac{1}{2}\text{Var}(\hat{V}).$$

Dies wird erreicht mit maximal den doppelten Kosten. Eine entsprechende Fehlerreduktion um den Faktor  $\frac{1}{2}$  alleine über  $N$  erfordert wenigstens eine Vervierfachung des Aufwands.

**Beispiel GBM:** Der Index  $k$  in  $V_k$  sei die Nummer einer MC-Simulation,  $k = 1, \dots, N$ , für GBM. Für den Payoff  $\Psi$  ziehe  $Z_k \sim \mathcal{N}(0, 1)$  und berechne die Paare  $\hat{V}_k, V_k^-$  und die Antithetische Variable  $V_{\text{AV},k}$  wie folgt:

$$\begin{aligned} \hat{V}_k &= \Psi \left( S_0 \exp \left\{ \left( r - \frac{\sigma^2}{2} \right) T + \sigma \sqrt{T} Z_k \right\} \right) \\ V_k^- &= \Psi \left( S_0 \exp \left\{ \left( r - \frac{\sigma^2}{2} \right) T - \sigma \sqrt{T} Z_k \right\} \right) \\ V_{\text{AV},k} &= \frac{1}{2}(\hat{V}_k + V_k^-) \end{aligned}$$

Für jedes  $k$  sind  $\hat{V}_k$  und  $V_k^-$  abhängig, aber die Unabhängigkeit der  $Z_k \sim \mathcal{N}(0, 1)$  bewirkt die Unabhängigkeit der  $V_{AV,k}$  für  $k = 1, \dots, N$ . Monte Carlo kann angewendet werden, und der diskontierte Mittelwert (Faktor  $e^{-rT}$ ) approximiert  $V$ .

(Die nächst-komplexere Methode der Varianzreduktion ist die Methode *control variates*, vgl. Literatur.)

Hinweis: Für den eindimensionalen Fall von Vanilla-Optionen wurde MC nicht entwickelt. Die Möglichkeiten von MC entfalten sich bei exotischen Optionen, vor allem im höherdimensionalen Fall.

## 3.4 Monte-Carlo-Methoden bei amerikanischen Optionen

### A. Stoppzeiten

Beispiele für Entscheidungen an Finanzmärkten sind der Verkauf von Aktien oder das Ausüben von Optionen. Solche Entscheidungen können als “Stoppen” interpretiert werden; das Halten der Aktie oder der Option wird gestoppt. Den Zeitpunkt der Entscheidung nennt man “Stoppzeit”, hier mit  $\tau$  bezeichnet. Entscheidungen können nur auf Informationen gründen, die bis zum Entscheidungszeitpunkt vorliegen. Deswegen darf eine “Stoppzeit” **nicht voreilend** sein (*non-anticipating*); die Stoppregel darf nicht in die Zukunft schauen:

Für jedes  $t$  braucht man Gewissheit, ob die Entscheidung gefallen ist, d.h. ob  $\tau \leq t$  oder  $\tau > t$ .

Diese Charakterisierung einer Stoppzeit kann mit den Mitteln der Stochastik formal definiert werden, aufbauend auf dem zugrundeliegenden Prozess  $S_t$ :

Filtration  $\mathcal{F}_t$ : Ein stochastischer Prozess  $S_t$  heißt  $\mathcal{F}_t$ -adaptiert, wenn  $S_t$   $\mathcal{F}_t$ -messbar ist für alle  $t$ . Die natürliche Filtration  $\mathcal{F}_t^S$  ist die kleinste Sigma-Algebra über  $\{S_s \mid 0 \leq s \leq t\}$ , vereinigt mit den P-Nullmengen.  $S_t$  ist bezüglich  $\mathcal{F}_t^S$  adaptiert. Filtrationen repräsentieren die Informationsmenge zum Zeitpunkt  $t$ . Fordere also für die Menge  $\{\tau \leq t\}$  aller Entscheidungen bis  $t$

$$\{\tau \leq t\} \in \mathcal{F}_t .$$

Das ist die  $\mathcal{F}_t$ -Messbarkeit von  $\tau$ .

#### Definition (Stoppzeit)

Eine Stoppzeit  $\tau$  bezüglich einer Filtration  $\mathcal{F}_t$  ist eine Zufallsvariable mit Werten in  $[0, T] \cup \{\infty\}$ , welche für alle  $t$   $\mathcal{F}_t$ -messbar ist.

Das folgende Resultat von Bensoussan (1984) unterstreicht die Bedeutung von Stoppzeiten für die Bewertung amerikanischer Optionen: Es sei  $\Psi(S_t)$  ein Payoff, z.B.  $\Psi(S_t) = (K - S_t)^+$ , so gilt

$$V(S, 0) = \sup_{\substack{0 \leq \tau \leq T \\ \tau \text{ Stoppzeit}}} \mathbf{E}_Q [e^{-r\tau} \Psi(S_\tau) \mid S_0 = S] \quad (*)$$

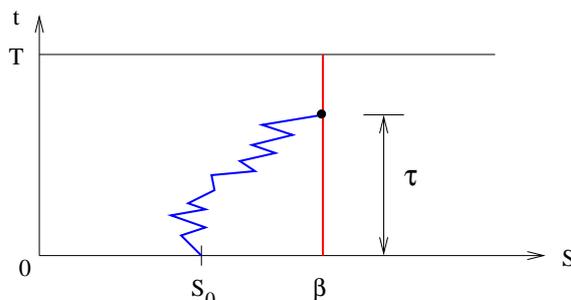
Dabei ist  $\tau$  eine Stoppzeit bezüglich einer natürlichen Filtration  $\mathcal{F}_t$  zu  $S_t$ .

## Beispiele für Stoppzeiten

1) Es sei

$$\tau := \inf \{ t > 0 \mid S_t \geq \beta \}$$

für gegebenes  $\beta > S_0$ . (engl. *hitting time*) Falls kein solches  $t$  existiert, setze  $\tau := \infty$ . Offensichtlich ist dieses  $\tau$  nicht vorgeifend. Denn das Setzen von  $\tau$  bedeutet das Setzen einer Marke, sobald  $S_t$  den Wert  $\beta$  erreicht. Dann kann für beliebiges  $t$  geprüft werden, ob die Entscheidung gefallen ist.\*



2) Definiere  $t^*$  als denjenigen Zeitpunkt, an dem  $\max_{0 \leq t \leq T} S_t$  erreicht wird. Dies ist keine Stoppzeit! Denn für ein beliebiges  $t$  kann nicht entschieden werden, ob  $t^* \leq t$  oder  $t^* > t$ .

3) Es sei

$$\tau := \min \{ t \leq T \mid (t, S_t) \in \text{Stopp-Bereich} \}$$

Dieses Beispiel ist ähnlich wie Beispiel 1, wenn der Zeitpunkt betrachtet wird, in dem  $S_t$  den Rand des Stopp-Bereiches trifft. Zu dieser *Ausübungs-Kurve* (*early-exercise curve*) vgl. Abschnitt 4.5 im folgenden Kapitel.

## B. Parametrische Methoden

In (\*) wird das Supremum über *alle* Stoppzeiten gebildet. Wir konstruieren nun eine *spezielle* Stoppzeit. Ähnlich wie in Beispiel 1 oder 3 definieren wir eine Kurve im  $(S, t)$ -Halbstreifen, welche eine Näherung zur *early-exercise* Kurve darstellen soll. Damit ist eine spezielle Stopp-Strategie  $\tilde{\tau}$  definiert durch das Treffen dieser Kurve. Wenn  $\beta$  einen Vektor von Parametern repräsentiert, der die Kurve definiert, dann hängt die Stoppregel und damit  $\tilde{\tau}$  von  $\beta$  ab. Die spezielle  $\beta$ -abhängige Stopp-Strategie  $\tilde{\tau}$  führt zu einer **unteren Schranke**

$$V^{\text{low}(\beta)}(S, 0) := \mathbf{E}_{\mathbf{Q}} [e^{-r\tilde{\tau}} \Psi(S_{\tilde{\tau}}) \mid S_0 = S] \leq V(S, 0).$$

**Anwendung:** Offensichtlich kann  $V(S, 0)$  über geeignete Stopp-Kurven näherungsweise als

$$\sup_{\beta} V^{\text{low}(\beta)}$$

charakterisiert werden, über eine optimale Stopp-Strategie. Diese Idee führt auf einen brauchbaren Ansatz:

Man konstruiere eine Kurve in Abhängigkeit von einem Parametervektor  $\beta$  derart, dass sie die *Ausübungs-Kurve* approximiert. Stopp-Strategie: Stoppe wenn der Pfad  $S_t$  die durch  $\beta$  definierte Kurve überschreitet. Für  $N$  solcher Pfade wertet man den Payoff

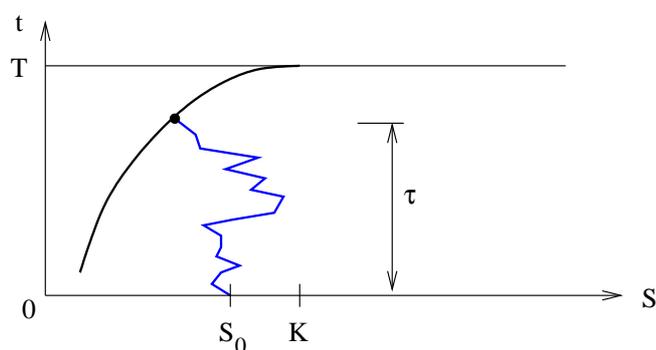
---

\* Für einen formalen Beweis, dass dieses  $\tau$  eine Stoppzeit ist, siehe z.B. [Hunt & Kennedy (2000)].

aus, und approximiert den Wert  $V^{\text{low}(\beta)}$  wie bei den “gewöhnlichen” Monte-Carlo-Methoden. Anschließend versuche, die untere Schranke  $V^{\text{low}(\beta)}$  durch Wiederholungen des Verfahrens mit “besseren” Parameter-Werten  $\beta$  zu maximieren.

**Beispiel** mit  $\beta \in \mathbb{R}^1$ :

Betrachte eine Parabel mit Scheitel in  $(S, t) = (K, T)$  mit nur einem Parameter  $\beta$  als Näherung der early-exercise-Kurve eines Puts (vgl. Figur). Man berechne Pfade (z.B. 10000 mal) bis der linke Ast der Parabel erreicht wird (ergibt  $\tilde{\tau}$ ) wie in Beispiel 1, oder  $t = T$ . Die MC-Methode liefert einen Wert für  $V^{\text{low}(\beta)}$ . Jeder einzelne Wert  $V^{\text{low}(\beta)}$  wird also mit einem Aufwand berechnet, der etwa dem MC-Verfahren bei einer europäischen Option entspricht. Anschließend Wiederholung für bessere  $\beta$ . — Diese Werte  $V^{\text{low}(\beta)}$  werden nicht gegen  $V(S, 0)$  konvergieren. Denn ein systematischer Fehler bleibt, weil die Ausübungs-Kurve mit einer Parabel nur schlecht modelliert werden kann.



Vollständig wird die Methode erst, wenn auch obere Schranken  $V^{\text{up}}$  konstruiert werden und eventuell sogar Konvergenz gegeben ist. Literatur: [P. Glasserman: Monte Carlo Methods in Financial Engineering (2004)]

## C. Regressionsmethoden

### Definition (Bermuda-Option)

Eine Bermuda-Option ist eine Option, die nur an einer endlichen Anzahl  $M$  von diskreten Zeitpunkten  $t_j$  ausgeübt werden kann.

Speziell für

$$\Delta t := \frac{T}{M}, \quad t_j := j \Delta t \quad (j = 0, \dots, M)$$

sei der Wert einer solchen Bermuda-Option mit  $V^{\text{Be}(M)}$  bezeichnet. Auf Grund der zusätzlichen Anzahl an Ausübungsmöglichkeiten gilt

$$V^{\text{Eu}} \leq V^{\text{Be}(M)} \leq V^{\text{Am}} .$$

Außerdem kann man zeigen, dass

$$\lim_{M \rightarrow \infty} V^{\text{Be}(M)} = V^{\text{Am}} .$$

Für geeignetes  $M$  wird  $V^{\text{Be}(M)}$  als Näherung von  $V^{\text{Am}}$  verwendet. Die lineare Konvergenz legt nahe, mit wenigen moderaten Werten von  $M$  zu arbeiten und Richardson-Extrapolation anzuwenden. So lassen sich die hohen Kosten von Monte-Carlo-Verfahren bei amerikanischen Optionen beschränken.

**Wiederholung** (Binomial-Methode, vgl. Abschnitt 1.3)

Der Wert einer amerikanischen Option wird dort rückwärts rekursiv berechnet: streifenweise europäische Optionen, mit Fortsetzungswert  $V^{\text{cont}}$  und für jedes  $t_j$

$$V^{\text{Am}} = \max \{ \Psi(S), V^{\text{cont}} \} .$$

Denn an jedem  $t_j$  entscheidet der Halter der Option, welche der beiden Möglichkeiten {Ausüben, Halten} die bessere ist.\*

Bei einer Bermuda-Option definieren wir den Fortsetzungswert an  $t_j$  analog:

$$C_j(x) := e^{-r\Delta t} \mathbb{E}_{\mathbb{Q}} [V(S_{t_{j+1}}, t_{j+1}) \mid S_{t_j} = x] .$$

Diese Funktionen  $C_j(x)$  müssen approximiert werden.

**Allgemeine Rekursion**

Setze  $V_M(x) \equiv \Psi(x)$ .

Für  $j = M - 1, \dots, 1$

konstruiere  $C_j(x)$  für  $x > 0$ ;

$V_j(x) := V(x, t_j) = \max \{ \Psi(x), C_j(x) \}$  für Gitterpunkte  $x$ .

$V_0 := V(S_{t_0}, t_0) = \max \{ \Psi(S_0), C_0(S_0) \}$ .

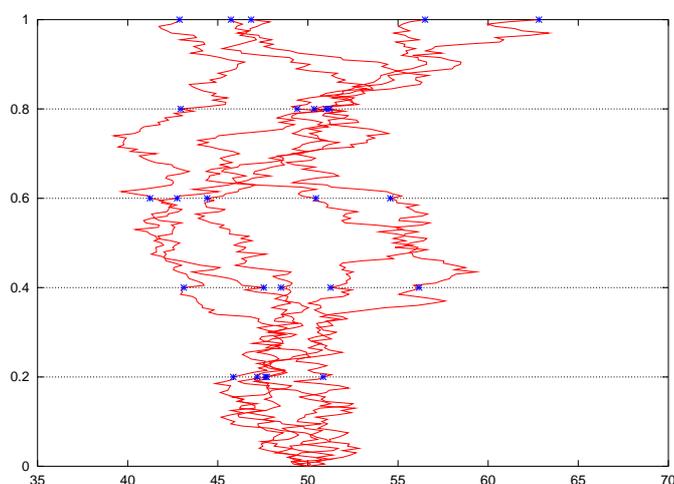
Spezielle  $x$  werden unten durch ein stochastisches Gitter definiert.

Um die Funktionen  $C_j(x)$  mit Monte Carlo zu berechnen, zieht man Informationen aus durch Simulation erzeugten Pfaden, und approximiert  $C_j(x)$  durch eine Regressionskurve  $\hat{C}_j(x)$ .

**Regression** (Grundprinzip)

(a) Simuliere  $N$  Pfade  $S_1(t), \dots, S_N(t)$ : Berechne und speichere die Werte

$$S_{j,k} := S_k(t_j), \quad j = 1, \dots, M, \quad k = 1, \dots, N .$$

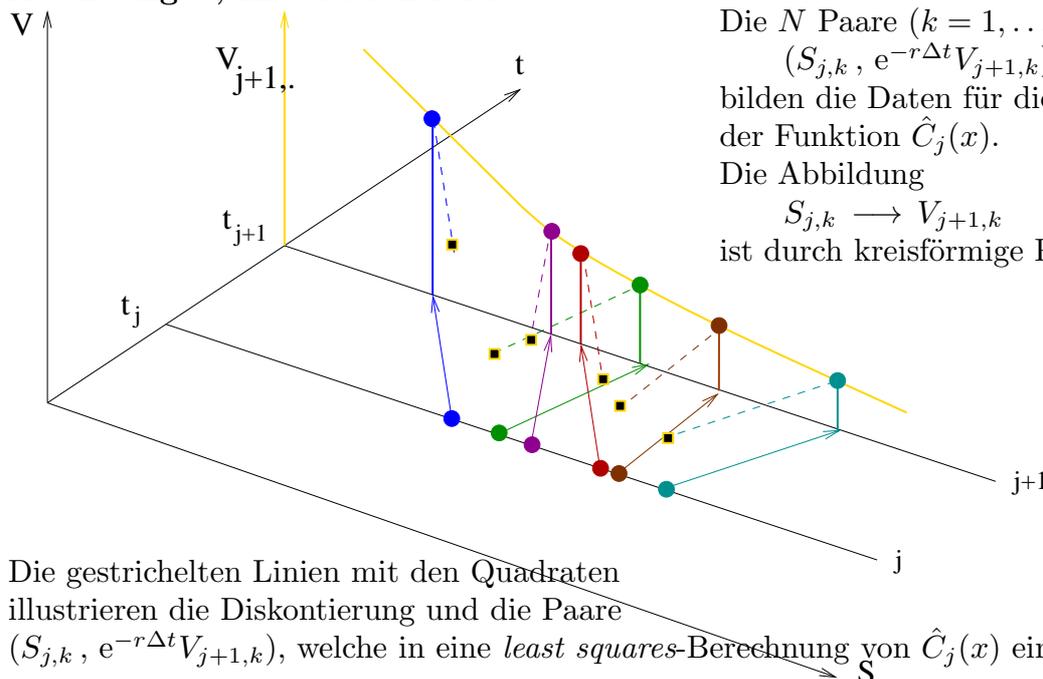


fünf Trajektorien und  
Punkte  $(S_{j,k}, t_j)$   
für  $j = 1, \dots, 5, k = 1, \dots, 5$

---

\* Prinzip der Dynamischen Programmierung

Abbildung C, Illustration für  $N = 6$



Die  $N$  Paare ( $k = 1, \dots, N$ )  
 $(S_{j,k}, e^{-r\Delta t}V_{j+1,k})$   
 bilden die Daten für die Berechnung  
 der Funktion  $\hat{C}_j(x)$ .  
 Die Abbildung  
 $S_{j,k} \rightarrow V_{j+1,k}$   
 ist durch kreisförmige Punkte illustriert.

Die gestrichelten Linien mit den Quadraten  
 illustrieren die Diskontierung und die Paare  
 $(S_{j,k}, e^{-r\Delta t}V_{j+1,k})$ , welche in eine *least squares*-Berechnung von  $\hat{C}_j(x)$  eingehen.

**Regression** (Fortsetzung)

- (b) Setze für  $j = M$ :  $V_{M,k} = \Psi(S_{M,k})$  für alle  $k$ .
- (c) Für  $j = M - 1, \dots, 1$ :  
 Approximiere  $C_j(x)$  mit geeigneten Basisfunktionen  $\phi_0, \dots, \phi_L$  (z.B. Monome)

$$C_j(x) \approx \sum_{l=0}^L a_l \phi_l(x) =: \hat{C}_j(x)$$

Bestimme hierzu mit der Methode der kleinsten Quadrate (“least squares”) über die  $N$  Punkte

$$(S_{j,k}, e^{-r\Delta t}V_{j+1,k}), \quad k = 1, \dots, N$$

die Koeffizienten  $a_0, \dots, a_L$  und damit  $\hat{C}_j$ .

Auswertung:

$$V_{j,k} := \max \left\{ \Psi(S_{j,k}), \hat{C}_j(S_{j,k}) \right\}.$$

- (d) Setze

$$V_0 := \max \left\{ \Psi(S_0), e^{-r\Delta t} \frac{1}{N} (V_{1,1} + \dots + V_{1,N}) \right\}.$$

Zu Schritt (d): Der eigentliche Algorithmus von (c) lässt sich für  $j = 0$  nicht durchführen, weil  $S_{0,k} = S_0$  für alle  $k$ . Deswegen die Mittelwertbildung in (d). Konvergenz des Algorithmus wurde bewiesen.

**Aufwand:** Aufwändig sind die Schritte (a) und (c). Der Aufwand bei Amerikanischen Optionen hängt in Schritt (c) von der Dimension ab, also von der Anzahl der zugrundeliegenden Assets. (Warum?) Wenn die Gesamtrechnzeit für die Bewertung einer amerikanischen Option beschränkt ist, dann beeinflusst diese Dimensionsabhängigkeit des Aufwands auch den

erreichbaren Fehler. Insofern ist der Fehler bei Monte-Carlo-Methoden *doch* dimensionsabhängig.

Auf dem Grundprinzip des obigen Regressions-Algorithmus baut der Algorithmus von Longstaff & Schwartz (2001) auf, sowie der noch effizientere von C. Jonen (2009). Insbesondere der Teil (c) des Algorithmus bietet Verbesserungspotential.

---

Bei Longstaff & Schwartz wird der Algorithmus wie folgt modifiziert:

Ein dynamisches Programmierprinzip für die optimalen Stoppzeiten wird eingearbeitet. Für jeden Pfad wird eine eigene optimale Stoppzeit  $\tau_k$  vorgesehen für  $k = 1, \dots, N$  (es genügt wegen  $\tau_k = k\Delta t$  die Speicherung des Index  $k$ ).

#### **Algorithmus**

Initialisierung:  $\tau_k := M$  für alle  $k$ .

Für jedes  $j = M - 1, \dots, 1$ :

Schleife über alle Pfade  $k = 1, \dots, N$ :

Falls  $\Psi(S_{j,k}) \geq \hat{C}_j(S_{j,k})$  dann setze  $\tau_k := j$ .

Andernfalls lasse  $\tau_k$  unverändert.

Dieser Algorithmus hat den Vorteil, dass über mehrere Zeitschichten hinweg gearbeitet werden kann. Dank einer Modifikation von C. Jonen [Intern.J. Computer Math. **86** (2009); Dissertation 2011] ist dies ein effizientes Verfahren.

Weitere Hinweise zur Regression und zur Berechnung von Sensitivitäten (Griechen), sowie auf Literatur finden sich in [R. Seydel: Tools for Computational Finance].

Vergleiche auch Topic 6 in den *Topics for CF*.

## 4. Finite Differenzen für Amerikanische Standard-Optionen

Betrachte Vanilla-Optionen mit Payoff  $\Psi(S)$ , dabei ist  $S$  der Preis eines *underlying* Assets. Modell: GBM. Zugelassen sei eine kontinuierliche Dividendenzahlung mit konstanter *Dividendenrate*  $\delta \geq 0$ , also

$$\frac{dS_t}{S_t} = (\mu - \delta) dt + \sigma dW_t.$$

### 4.1 Vorbereitungen

Die Black-Scholes-Gleichung für den Wert  $V(S, t)$  einer Standard-Option ist

$$\frac{\partial V}{\partial t} + \frac{1}{2} \sigma^2 S^2 \frac{\partial^2 V}{\partial S^2} + (r - \delta) S \frac{\partial V}{\partial S} - r V = 0.$$

Definitionsbereich:  $S > 0$ ,  $0 \leq t \leq T$ . Endbedingung:  $V(S, T) = \Psi(S)$ .

Diese partielle DGL kann direkt mit numerischen Methoden gelöst werden. Allerdings müssen in dieser Version kompliziertere Instabilitäten berücksichtigt werden. ( $\rightarrow$  später)

Mit Transformationen können einfachere Versionen aufgestellt werden, z.B. mit  $S = e^x$ .

**Annahme:**  $\sigma$ ,  $r$  und  $\delta$  seien konstant.

Dann kann die Black-Scholes-Gleichung auf einen besonders einfachen Typ von PDGL transformiert werden (vgl. Übungen). Mit  $t = T - \frac{2\tau}{\sigma^2}$ ,  $S = Ke^x$ ,  $q := \frac{2r}{\sigma^2}$ ,  $q_\delta := \frac{2(r-\delta)}{\sigma^2}$  und  $y(x, \tau)$  definiert durch

$$V(S, t) = V(Ke^x, T - \frac{2\tau}{\sigma^2}) =: v(x, \tau)$$

$$v(x, \tau) = K \exp \left\{ -\frac{1}{2}(q_\delta - 1)x - \left( \frac{1}{4}(q_\delta - 1)^2 + q \right) \tau \right\} y(x, \tau)$$

ergibt sich

$$\frac{\partial y}{\partial \tau} = \frac{\partial^2 y}{\partial x^2}$$

Dies ist eine parabolische Gleichung, welche eine analytische Lösung besitzt. Nach Rücktransformation ergibt sich die bekannte Black-Scholes-*Formel*. Somit benötigen wir keine weiteren numerischen Methoden für Standard-Optionen europäischen Typs.

Unser Ziel ist es, **amerikanische** Optionen zu bewerten.

Generische Anwendungen für die einfache PDGl  $y_\tau = y_{xx}$  sind die Wärmeleitung und die Diffusion in einem eindimensionalen Medium. Dies Problem ist sachgemäß gestellt (*well-posed*) als **Anfangswertproblem** mit wachsendem  $\tau$ . Wegen der Zeit-Transformation  $t \rightarrow \tau$  wird aus der Endbedingung für  $V(S, T) = \Psi(S)$  nun eine Anfangsbedingung für  $\tau = 0$ :

$$\begin{aligned} \text{Call : } y(x, 0) &= \max \left\{ e^{\frac{x}{2}(q_\delta+1)} - e^{\frac{x}{2}(q_\delta-1)}, 0 \right\} \\ \text{Put : } y(x, 0) &= \max \left\{ e^{\frac{x}{2}(q_\delta-1)} - e^{\frac{x}{2}(q_\delta+1)}, 0 \right\} \end{aligned}$$

Der Definitionsbereich für  $y$  ist

$$-\infty < x < +\infty, \quad 0 \leq \tau \leq \frac{1}{2}\sigma^2 T =: \tau_{\max}$$

d.h. ein *Streifen* in der  $(x, \tau)$ -Ebene. “Rand” bedeutet hier  $x \rightarrow -\infty$ ,  $x \rightarrow +\infty$ .

### Randwertproblem

Für numerische Zwecke muss der Streifen durch ein Rechteck ersetzt werden, mit

$$x_{\min} \leq x \leq x_{\max}.$$

Diese “Lokalisierung“ beschneidet den Einfluss der Anfangsbedingung. Um ein sachgemäß gestelltes Problem zu erhalten, müssen an den Rändern  $x_{\min}$  und  $x_{\max}$  zusätzliche Randbedingungen formuliert werden. Die Lösung  $w(x, \tau)$  des resultierenden **Randwertproblems** soll eine gute Näherung für  $y(x, \tau)$  sein, deswegen muss das Rechteck groß genug sein. Die Wahl von  $x_{\min}$  und  $x_{\max}$  muss auch berücksichtigen, dass der Bereich des finanziellen Interesses, also die  $x$ -Werte von  $S$  und  $K$ , eingeschlossen ist, also

$$x_{\min} < \min \left\{ 0, \log \frac{S_0}{K} \right\}, \quad \max \left\{ 0, \log \frac{S_0}{K} \right\} < x_{\max}.$$

Zur Vereinfachung denke man etwa an  $x_{\min} = -3$  und  $x_{\max} = 3$ . — Später werden wir sehen, dass für amerikanische Optionen aus der Partiiellen Differentialgleichung eine Ungleichung wird.

## 4.2 Grundlagen von Finite-Differenzen-Methoden (FDM)

### A. Differenzen-Approximationen

Es gilt

$$f'(x) = \frac{f(x+h) - f(x)}{h} - \frac{h}{2} f''(\xi) \quad \text{für ein } \xi \in (x, x+h) \text{ und } f \in C^2$$

Wir führen ein äquidistantes Gitter mit den Gitterpunkten  $x_i$

$$\dots < x_{i-1} < x_i < x_{i+1} < \dots$$

und  $h := x_{i+1} - x_i$  ein. [Im nicht äquidistanten Fall verwendet man eher Finite-Elemente-Methoden (FEM).] Wir arbeiten in diesem Kapitel ausschließlich mit äquidistanten Gittern.

Mit der Bezeichnung  $f_i = f(x_i)$  gilt analog:

$$\begin{aligned} f'(x_i) &= \frac{f_{i+1} - f_{i-1}}{2h} + O(h^2) && \text{für } f \in C^3 \\ f''(x_i) &= \frac{f_{i+1} - 2f_i + f_{i-1}}{h^2} + O(h^2) && \text{für } f \in C^4 \\ f'(x_i) &= \frac{-f_{i+2} + 4f_{i+1} - 3f_i}{2h} + O(h^2) && \text{für } f \in C^3 \end{aligned}$$

## B. Das Gitter

Für ein  $m \in \mathbb{N}$  und ein  $\nu_{\max} \in \mathbb{N}$  definiere

$$\begin{aligned} \Delta x &:= \frac{x_{\max} - x_{\min}}{m}, \quad x_i := x_{\min} + i \cdot \Delta x, \quad i = 0, 1, \dots, m; \\ \Delta \tau &:= \frac{\tau_{\max}}{\nu_{\max}}, \quad \tau_\nu := \nu \cdot \Delta \tau, \quad \nu = 0, \dots, \nu_{\max}. \end{aligned}$$

$y_{i,\nu} := y(x_i, \tau_\nu)$  ist Wert von  $y$  am ‘‘Knoten’’  $(x_i, \tau_\nu)$ .  $w_{i,\nu}$  ist eine Approximation von  $y_{i,\nu}$ . Über die Anfangsbedingung sind bekannt

$$w_{i,0} = y(x_i, 0).$$

## C. Explizite Methode

Ersetze

$$\begin{aligned} \frac{\partial y_{i,\nu}}{\partial \tau} &:= \frac{\partial y(x_i, \tau_\nu)}{\partial \tau} = \frac{y_{i,\nu+1} - y_{i,\nu}}{\Delta \tau} + O(\Delta \tau) \quad \text{und} \\ \frac{\partial^2 y_{i,\nu}}{\partial x^2} &= \frac{y_{i+1,\nu} - 2y_{i,\nu} + y_{i-1,\nu}}{\Delta x^2} + O(\Delta x^2) \end{aligned}$$

in der partiellen Differentialgleichung, lasse die  $O$ -Fehlerterme weg, ersetze  $y \rightarrow w$  und erhalte die Differenzengleichung

$$\frac{w_{i,\nu+1} - w_{i,\nu}}{\Delta \tau} = \frac{w_{i+1,\nu} - 2w_{i,\nu} + w_{i-1,\nu}}{\Delta x^2}.$$

Wenn für eine ‘‘Zeitschicht’’  $\nu$  alle Werte  $w$  berechnet sind, können explizit mit

$$w_{i,\nu+1} = w_{i,\nu} + \frac{\Delta \tau}{\Delta x^2} (w_{i+1,\nu} - 2w_{i,\nu} + w_{i-1,\nu})$$

die Werte für die Zeitschicht  $\nu + 1$  berechnet werden. Mit der Abkürzung

$$\lambda := \frac{\Delta \tau}{\Delta x^2}$$

erhält man

$$w_{i,\nu+1} = \lambda w_{i-1,\nu} + (1 - 2\lambda)w_{i,\nu} + \lambda w_{i+1,\nu}. \quad (*)$$

Beginne mit  $\nu = 0$ , weil dort die  $w_{i,0}$  durch die Anfangsbedingung der PDGI bekannt sind. D.h. für  $\nu = 1$  können alle  $w_{i,1}$  mit der expliziten Formel (\*) ausgerechnet werden und ebenso Zeitschicht-weise für wachsende  $\nu$  die weiteren Werte.

$\nu$  ist der äußere Schleifenindex und  $i$  der innere. Dies legt nahe, das Vorgehen mit Vektoren und Matrizen zu beschreiben:

$$w^{(\nu)} := (w_{1,\nu}, \dots, w_{m-1,\nu})^T \quad (\text{noch ohne Randkomponenten } w_{0,\nu} \text{ und } w_{m,\nu}),$$

also die Werte der  $\nu$ -ten Zeitschicht, und die  $(m-1) \times (m-1)$  Matrix

$$A := A_{\text{expl}} := \begin{pmatrix} 1-2\lambda & \lambda & 0 & \cdots & 0 \\ \lambda & 1-2\lambda & \ddots & \ddots & \vdots \\ 0 & \ddots & \ddots & \ddots & 0 \\ \vdots & \ddots & \ddots & \ddots & \lambda \\ 0 & \cdots & 0 & \lambda & 1-2\lambda \end{pmatrix}$$

Somit lautet die explizite Methode

$$w^{(\nu+1)} = Aw^{(\nu)} \quad \text{für } \nu = 0, 1, 2, \dots, \nu_{\text{max}} - 1$$

bis auf eine Korrektur wegen Randbedingungen für  $w_{0,\nu}$  und  $w_{m,\nu}$ . Zur Einsparung von Schreibarbeit setzen wir *vorläufig* die Randbedingungen auf null.

### Beispiel für Instabilität

$y_\tau = y_{xx}$  mit  $y(x, 0) = \sin \pi x$ ,  $x_0 = 0$ ,  $x_m = 1$  und Randbedingungen  $= 0$ .

Wir wollen eine Näherung zu  $y(x = 0.2, \tau = 0.5)$  berechnen und benutzen dafür das Gitter mit  $\Delta x = 0.1$ , d.h.  $m = 10$  und  $0.2 = x_2$ , und zwei verschiedene Ausführungen für  $\Delta \tau$ :

- $\Delta \tau = 0.0005 \Rightarrow \lambda = 0.05$  und  $0.5 = \tau_{1000}$   
ergibt  $w_{2,1000} = 0.00435$  (exakt wäre 0.004227)
- $\Delta \tau = 0.01 \Rightarrow \lambda = 1$  und  $0.5 = \tau_{50}$   
ergibt  $w_{2,50} = -1.5 \cdot 10^8$  (also Instabilität)

## D. Stabilität

Fehleranalyse von  $w^{(\nu+1)} = Aw^{(\nu)} + d^{(\nu)}$  (oben:  $d^{(\nu)} = 0$ )

### Bezeichnungen

$w^{(\nu)}$  seien die exakten Vektoren von  $w^{(\nu+1)} = Aw^{(\nu)} + d^{(\nu)}$ ,

$\bar{w}^{(\nu)}$  sei die im Rechner berechnete Version, also mit Rundungsfehlern,

$e^{(\nu)} := \bar{w}^{(\nu)} - w^{(\nu)}$ .

Bei exakter Rechnung ist der Vektor  $\bar{w}^{(\nu+1)} - A\bar{w}^{(\nu)} - d^{(\nu)}$  von Null verschieden; nennen wir ihn  $r^{(\nu+1)}$ , also

$$\bar{w}^{(\nu+1)} = A\bar{w}^{(\nu)} + d^{(\nu)} + r^{(\nu+1)}.$$

Der Vektor  $r^{(\nu+1)}$  repräsentiert die Rundungsfehler im  $\nu$ -ten Schritt.

Es genügt, die Entwicklung *eines* Fehlers zu analysieren. Entsprechend setze  $r^{(\nu)} = 0$  für  $\nu > 1$ , d.h. studiere die Fortpflanzung des Fehlers  $e^{(0)}$  im Laufe der weiteren Iterationen.

$$\begin{aligned} \Rightarrow \bar{w}^{(\nu+1)} &= A\bar{w}^{(\nu)} + d^{(\nu)} \quad (\nu > 1) \\ \Rightarrow Ae^{(\nu)} &= A\bar{w}^{(\nu)} - Aw^{(\nu)} = \bar{w}^{(\nu+1)} - w^{(\nu+1)} = e^{(\nu+1)} \\ \Rightarrow e^{(\nu)} &= A^\nu e^{(0)} \end{aligned}$$

Wir fordern für ein stabiles Verhalten  $A^\nu e^{(0)} \rightarrow 0$  für  $\nu \rightarrow \infty$ .

Bezeichnung: Spektralradius

$$\rho(A) := \max_k |\mu_k^A| \text{ mit } \mu^A \text{ ist Eigenwert von } A$$

### Lemma 1

$$\rho(A) < 1 \iff A^\nu z \rightarrow 0 \text{ für alle } z \text{ und } \nu \rightarrow \infty$$

*Beweis:* Vorlesungen oder Lehrbücher über Numerik.

$A$  ist hier von folgender Gestalt

$$A = I - \lambda \cdot \underbrace{\begin{pmatrix} 2 & -1 & & & 0 \\ -1 & \ddots & \ddots & & \\ & \ddots & \ddots & \ddots & \\ & & \ddots & \ddots & -1 \\ 0 & & & -1 & 2 \end{pmatrix}}_{=:G}.$$

$$\Rightarrow \mu^A = 1 - \lambda\mu^G.$$

### Lemma 2

$$\text{Für } G = \begin{pmatrix} \alpha & \beta & & & 0 \\ \gamma & \ddots & \ddots & & \\ & \ddots & \ddots & \ddots & \\ & & \ddots & \ddots & \beta \\ 0 & & & \gamma & \alpha \end{pmatrix} \in \mathbb{R}^{N \times N} \text{ gilt}$$

$$\mu_k^G = \alpha + 2\beta\sqrt{\frac{\gamma}{\beta}} \cos \frac{k\pi}{N+1}, \quad k = 1, \dots, N.$$

*Beweis:*

Mit den Eigenvektoren

$$v^{(k)} = \left( \sqrt{\frac{\gamma}{\beta}} \sin \frac{k\pi}{N+1}, \left( \sqrt{\frac{\gamma}{\beta}} \right)^2 \sin \frac{2k\pi}{N+1}, \dots, \left( \sqrt{\frac{\gamma}{\beta}} \right)^N \sin \frac{Nk\pi}{N+1} \right)^T$$

prüfe  $Gv^{(k)} = \mu^G v^{(k)}$ .

(Für  $\gamma = \beta$  hängen die Eigenvektoren nicht mehr von  $\alpha, \beta, \gamma$  ab.)

Die Anwendung von Lemma 2 direkt auf  $A$ , oder auf  $G$  mit  $N = m - 1$ ,  $\alpha = 2$  und  $\beta = \gamma = -1$  ergibt

$$\mu_k^G = 2 - 2 \cos \frac{k\pi}{m} = 4 \sin^2 \left( \frac{k\pi}{2m} \right)$$

$$\mu_k^A = 1 - 4\lambda \sin^2 \frac{k\pi}{2m}.$$

Mit Lemma 1:

$$\text{Stabilität} \iff \left| 1 - 4\lambda \sin^2 \frac{k\pi}{2m} \right| < 1, \quad k = 1, \dots, m-1$$

$$\iff \lambda > 0 \text{ und } -1 < 1 - 4\lambda \sin^2 \frac{k\pi}{2m}, \quad \text{bzw. } \frac{1}{2} > \lambda \sin^2 \frac{k\pi}{2m}$$

Für die zweite Ungleichung ist  $\lambda \leq 1/2$  hinreichend. Zusammengefasst:

Für  $0 < \lambda \leq \frac{1}{2}$  ist die explizite Methode  $w^{(\nu+1)} = Aw^{(\nu)}$  mit  $A = A_{\text{expl}}$  stabil.

Wegen  $\lambda = \frac{\Delta\tau}{\Delta x^2}$ , bzw.  $0 < \Delta\tau \leq \frac{1}{2}\Delta x^2$  sind  $\Delta\tau$  und  $\Delta x$  nicht unabhängig voneinander wählbar.

### Folgerung

Diese explizite Methode ist unbefriedigend.

## E. Eine implizite Methode

Der ‘‘Rückwärts’’-Differenzenquotient

$$\frac{\partial y_{i,\nu}}{\partial \tau} = \frac{y_{i,\nu} - y_{i,\nu-1}}{\Delta\tau} + O(\Delta\tau).$$

führt zu

$$-\lambda w_{i+1,\nu} + (1 + 2\lambda)w_{i,\nu} - \lambda w_{i-1,\nu} = w_{i,\nu-1}.$$

Dies ist ein System von gekoppelten linearen Gleichungen für die  $w_{i,\nu}$ . Mit

$$A := A_{\text{impl}} := \begin{pmatrix} 1 + 2\lambda & -\lambda & & 0 \\ -\lambda & \ddots & \ddots & \\ & \ddots & \ddots & -\lambda \\ 0 & & -\lambda & 1 + 2\lambda \end{pmatrix}$$

ist der Vektor  $w^{(\nu)}$  implizit definiert als Lösung von  $Aw^{(\nu)} = w^{(\nu-1)}$ , oder

$$Aw^{(\nu+1)} = w^{(\nu)}, \nu = 0, \dots, \nu_{\max} - 1.$$

(Wiederum haben wir die Randbedingungen vorläufig auf  $w_{0,\nu} = w_{m,\nu} = 0$  gesetzt.) Diese Methode heißt *Rückwärts-Differenzen-Methode* oder *backwards-time-centerspace (BTCS)* oder *fully implicit*.

### Stabilität:

Auf Grund der Lemmata ist das Verfahren *uneingeschränkt stabil*, und  $\Delta\tau$  und  $\Delta x$  sind unabhängig voneinander.

### Kosten des Verfahrens:

$\nu_{\max}O(m)$ , weil nur eine *LR*-Zerlegung von  $A$  für  $\nu = 0$  notwendig ist. Die Kosten sind minimal, weil  $A$  Tridiagonal-Gestalt besitzt. Für jedes  $\nu$  muss nur eine Schleife rückwärts durchlaufen werden, welche  $O(m)$  Operationen benötigt.

Ein Nachteil dieser Methode (und auch der expliziten Methode) ist die Genauigkeit: Der Fehler ist von der Ordnung

$$O(\Delta x^2) + O(\Delta\tau).$$

## 4.3 Crank-Nicolson-Methode

Gesucht ist eine stabile Methode der Genauigkeit  $O(\Delta\tau^2)$  für

$$\frac{\partial y}{\partial \tau} = \frac{\partial^2 y}{\partial x^2}.$$

Vorwärts für  $\nu$ :

$$\frac{w_{i,\nu+1} - w_{i,\nu}}{\Delta\tau} = \frac{w_{i+1,\nu} - 2w_{i,\nu} + w_{i-1,\nu}}{\Delta x^2}$$

und rückwärts für  $\nu + 1$ :

$$\frac{w_{i,\nu+1} - w_{i,\nu}}{\Delta\tau} = \frac{w_{i+1,\nu+1} - 2w_{i,\nu+1} + w_{i-1,\nu+1}}{\Delta x^2}$$

Addition/Mitteln der beiden Gleichungen ergibt

$$\frac{w_{i,\nu+1} - w_{i,\nu}}{\Delta\tau} = \frac{1}{2\Delta x^2} (w_{i+1,\nu} - 2w_{i,\nu} + w_{i-1,\nu} + w_{i+1,\nu+1} - 2w_{i,\nu+1} + w_{i-1,\nu+1}).$$

### Satz (Crank–Nicolson)

Es gelten die folgende Aussagen:

- 1.) Für  $y \in C^4$  ist die Methode von der Ordnung  $O(\Delta x^2) + O(\Delta\tau^2)$ .
- 2.) Für jedes  $\nu$  muss ein lineares Gleichungssystem in Tridiagonalgestalt gelöst werden.
- 3.) Für alle  $\Delta\tau > 0$  ist das Verfahren stabil.

*Beweis:*

zu 1.)

Mit der Abkürzung

$$\delta_{xx}w_{i,\nu} := \frac{w_{i+1,\nu} - 2w_{i,\nu} + w_{i-1,\nu}}{\Delta x^2}$$

gilt mit Taylor-Entwicklung für  $y \in C^4$

$$\delta_{xx}y_{i,\nu} = \frac{\partial^2}{\partial x^2}y_{i,\nu} + \frac{\Delta x^2}{12} \frac{\partial^4}{\partial x^4}y_{i,\nu} + O(\Delta x^4).$$

Für den lokalen Diskretisierungsfehler

$$\varepsilon := \frac{y_{i,\nu+1} - y_{i,\nu}}{\Delta \tau} - \frac{1}{2} (\delta_{xx}y_{i,\nu} + \delta_{xx}y_{i,\nu+1})$$

gilt dann

$$\varepsilon = O(\Delta x^2) + O(\Delta \tau^2).$$

zu 2.)

Mit  $\lambda := \frac{\Delta \tau}{\Delta x^2}$  erhält man Gleichungen der Form

$$\begin{aligned} -\frac{\lambda}{2}w_{i-1,\nu+1} + (1 + \lambda)w_{i,\nu+1} - \frac{\lambda}{2}w_{i+1,\nu+1} \\ = \frac{\lambda}{2}w_{i-1,\nu} + (1 - \lambda)w_{i,\nu} + \frac{\lambda}{2}w_{i+1,\nu}. \end{aligned}$$

Mit den vorläufigen Randbedingungen  $w_{0,\nu} = w_{m,\nu} = 0$  ergibt sich das Matrix-Vektor-System

$$Aw^{(\nu+1)} = Bw^{(\nu)}$$

wobei

$$A := \begin{pmatrix} 1 + \lambda & -\frac{\lambda}{2} & & 0 \\ -\frac{\lambda}{2} & \ddots & \ddots & \\ & \ddots & \ddots & -\frac{\lambda}{2} \\ 0 & & -\frac{\lambda}{2} & 1 + \lambda \end{pmatrix}, \quad B := \begin{pmatrix} 1 - \lambda & \frac{\lambda}{2} & & 0 \\ \frac{\lambda}{2} & \ddots & \ddots & \\ & \ddots & \ddots & \frac{\lambda}{2} \\ 0 & & \frac{\lambda}{2} & 1 - \lambda \end{pmatrix}.$$

Die Eigenwerte von  $A$  müssen nach dem Satz von Gerschgorin zwischen 1 und  $1 + 2\lambda$  liegen. Somit ist Null kein Eigenwert und  $A$  nichtsingulär, und das Gleichungssystem hat eine eindeutige Lösung.

zu 3.)

Setze

$$A = I + \frac{\lambda}{2}G \quad \text{mit} \quad G := \begin{pmatrix} 2 & -1 & & 0 \\ -1 & \ddots & \ddots & \\ & \ddots & \ddots & -1 \\ 0 & & -1 & 2 \end{pmatrix} \quad \text{und} \quad B = I - \frac{\lambda}{2}G.$$

So gilt

$$\begin{aligned} \underbrace{(2I + \lambda G)}_{=:C} w^{(\nu+1)} &= (2I - \lambda G) w^{(\nu)} \\ &= (4I - 2I - \lambda G) w^{(\nu)} \\ &= (4I - C) w^{(\nu)}, \end{aligned}$$

welches zur expliziten Form

$$w^{(\nu+1)} = (4C^{-1} - I) w^{(\nu)}$$

führt. Mit Lemma 1 lautet die Stabilitätsforderung

$$\left| \frac{4}{\mu_k^C} - 1 \right| < 1 \quad \text{für alle } k.$$

Die Eigenwerte  $\mu_k^C$  von  $C$  sind nach Abschnitt 4.2D

$$\mu_k^C = 2 + \lambda \mu_k^G = 2 + 4\lambda \sin^2 \frac{k\pi}{2m} > 2.$$

Das Verfahren ist also stabil für alle  $\lambda > 0$  ( $\Delta\tau > 0$ ).

### Algorithmus (Crank-Nicolson)

*Start:* Wähle  $m, \nu_{\max}$ ; berechne  $\Delta x, \Delta\tau$

$$w_i^{(0)} = y(x_i, 0) \quad (0 \leq i \leq m)$$

Berechne  $LR$ -Zerlegung (bzw.  $RL$ -Zerlegung) von  $A$

*Schleife:* für  $\nu = 0, 1, \dots, \nu_{\max} - 1$  :

Berechne  $c := Bw^{(\nu)} + 0$  (vorläufige Randbedingungen 0)

Löse  $Ax = c$  (mit Hilfe der  $LR/RL$ -Zerlegung)

$$w^{(\nu+1)} := x$$

## 4.4 Randbedingungen

Da Anfangsbedingungen nur noch für das Intervall  $x_{\min} \leq x \leq x_{\max}$  aktiv sind, werden Randbedingungen

$$y(x, \tau) \text{ für } x_{\min} \text{ und } x_{\max}, \quad \text{oder} \\ w_{0,\nu} \text{ und } w_{m,\nu} \text{ für } \nu = 1, \dots, \nu_{\max}$$

benötigt, damit das Problem sachgemäß gestellt ist. Diese Randbedingungen sind künstlich und haben Einfluss auf die Genauigkeit der numerischen Lösung.

In den  $S$ -Variablen folgt aus dem Modell der GBM für  $S_t$

$$\begin{aligned} S(0) = 0 &\Rightarrow S_t = 0 \text{ für alle } t \\ S(0) \rightarrow \infty &\Rightarrow S_t \text{ groß für alle } t \leq T \end{aligned}$$

woraus folgt dass für alle  $t$

$$\begin{aligned} V_C(S, t) &= 0 \quad \text{für } S = 0, \quad \text{und} \\ V_P(S, t) &\rightarrow 0 \quad \text{für } S \rightarrow \infty. \end{aligned}$$

Wir nutzen diese Randbedingungen als Approximation für  $S_{\min} \approx 0$  und große  $S_{\max}$ . Und ebenso in den transformierten  $x$ -Variablen, sofern  $-x_{\min} = -x_0$  und  $x_{\max} = x_m$  groß genug sind:

$$\begin{aligned} \text{Call: "links"} \quad (S = 0) \quad V &= 0, \quad \text{d.h. } w_{0,\nu} = 0 \text{ für alle } \nu \\ \text{Put: "rechts"} \quad (S \rightarrow \infty) \quad V &= 0, \quad \text{d.h. } w_{m,\nu} = 0 \text{ für alle } \nu \end{aligned}$$

Die bisherige Argumentation ist sowohl für europäische als auch für amerikanische Optionen gültig. Nun stellt sich die Frage, was man für die  $V$ -Werte auf der jeweiligen "anderen" Seite des Intervalls erhält.

Wir betrachten zuerst **europäische** Optionen.

Wende die *Put-Call-Parität*

$$V_P = V_C - Se^{-\delta(T-t)} + Ke^{-r(T-t)}$$

an, um die Randbedingungen auf der anderen Seite zu erhalten. Man erhält so

$$\begin{aligned} V_C(S, t) &= Se^{-\delta(T-t)} - Ke^{-r(T-t)} \quad \text{für } S \rightarrow \infty \\ V_P(S, t) &= Ke^{-r(T-t)} - Se^{-\delta(T-t)} \quad \text{für } S \rightarrow 0. \end{aligned}$$

Nach der Transformation  $(S, t) \rightarrow (x, \tau)$  ergibt sich folgendes asymptotisches Verhalten:

$$\begin{aligned} y(x, \tau) &= r_1(x, \tau) \text{ für } x \rightarrow -\infty, \quad y(x, \tau) = r_2(x, \tau) \text{ für } x \rightarrow \infty, \text{ mit} \\ \text{Call: } r_1(x, \tau) &:= 0, \quad r_2(x, \tau) := \exp\left(\frac{1}{2}(q_\delta + 1)x + \frac{1}{4}(q_\delta + 1)^2\tau\right) \\ \text{Put: } r_1(x, \tau) &:= \exp\left(\frac{1}{2}(q_\delta - 1)x + \frac{1}{4}(q_\delta - 1)^2\tau\right), \quad r_2(x, \tau) := 0 \end{aligned}$$

Für das endliche Intervall  $a := x_{\min} \leq x \leq x_{\max} =: b$  sind  $r_1, r_2$  die dominierenden Terme der Randbedingungen und können als Approximation verwendet werden. Setze also für die Randbedingungen

$$\begin{aligned} w_{0,\nu} &= r_1(a, \tau_\nu) \\ w_{m,\nu} &= r_2(b, \tau_\nu). \end{aligned}$$

Dies sind Randbedingungen vom Dirichlet-Typ. Sie führen im Crank-Nicolson-Schema auf die zusätzlichen Terme

$$\begin{aligned} -\frac{\lambda}{2} w_{0,\nu+1} &= -\frac{\lambda}{2} r_1(a, \tau_{\nu+1}) \\ \frac{\lambda}{2} w_{0,\nu} &= \frac{\lambda}{2} r_1(a, \tau_\nu) \end{aligned}$$

und

$$\begin{aligned} -\frac{\lambda}{2} w_{m,\nu+1} &= -\frac{\lambda}{2} r_2(b, \tau_{\nu+1}) \\ \frac{\lambda}{2} w_{m,\nu} &= \frac{\lambda}{2} r_2(b, \tau_{\nu}), \end{aligned}$$

welche durch den Vektor

$$d^{(\nu)} := \frac{\lambda}{2} \cdot \begin{pmatrix} r_1(a, \tau_{\nu+1}) + r_1(a, \tau_{\nu}) \\ 0 \\ \vdots \\ 0 \\ r_2(b, \tau_{\nu+1}) + r_2(b, \tau_{\nu}) \end{pmatrix} = \frac{\lambda}{2} \cdot \begin{pmatrix} w_{0,\nu} + w_{0,\nu+1} \\ 0 \\ \vdots \\ 0 \\ w_{m,\nu} + w_{m,\nu+1} \end{pmatrix}$$

dargestellt werden. Im Crank-Nicolson-Algorithmus ist die rechte Seite des Gleichungssystems nun

$$c := Bw^{(\nu)} + d^{(\nu)}$$

Die Matrix  $A$  bleibt unverändert.

## 4.5 Vorzeitiges Ausüben — Freie Randwertprobleme

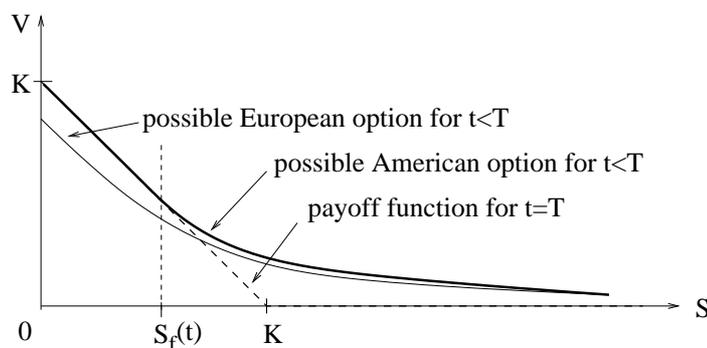
Im vorigen Abschnitt haben wir auch für amerikanische Optionen die Randbedingungen für  $S = 0$  beim Call und für  $S \rightarrow \infty$  beim Put festgestellt. Es fehlen noch die Randbedingungen auf der jeweils anderen Seite. Hierzu ist eine weitere Analyse der Lösungsstruktur amerikanischer Optionen notwendig.

**Grundlagen** (für den Put beim Black-Scholes Modell)

- $V(S, t)$  ist stetig,  $V \geq 0$ , und  $V \rightarrow 0$  für  $S \rightarrow \infty$ .
- Für  $r > 0$  und alle  $t < T$  gilt

$$V_P^{\text{Eu}}(0, t) = Ke^{-r(T-t)} < K,$$

woraus folgt: Es gibt ein  $S^* = S^{*\text{Eu}}(t)$  so dass  $V_P^{\text{Eu}}(S^*, t) = K - S^*$ .



- $V_P$  ist monoton fallend bzgl.  $S$ , und konvex. [R.C. Merton: Theory of Rational Option Pricing (1973)]

Hieraus folgt die Eindeutigkeit von  $S^*$ .

- $V_P^{\text{Am}} \geq (K - S)^+$  und  $V_P^{\text{Am}} \leq K$

**Behauptung 1**

Auch für den amerikanischen Put gibt es für  $r > 0$  ein  $S^* > 0$ , so dass  $V_P^{\text{Am}}(S^*, t) = K - S^*$ .

*Beweis:*

*Annahme:*  $V_P^{\text{Am}} > K - S$  für alle  $S > 0$ .

Somit gilt:

$$-V_P^{\text{Am}} + K - S < 0,$$

d.h. den Put auszuüben würde für alle  $S$  einen Verlust verursachen. Somit wäre *early exercise* nie sinnvoll, und es würde  $V_P^{\text{Am}} = V_P^{\text{Eu}}$  gelten. Es folgt

$$V_P^{\text{Eu}} = V_P^{\text{Am}} > K - S,$$

für alle  $S > 0$ , was im Widerspruch zur Folgerung aus b) steht.

Nun  $S^* := \max\{S \mid V^{\text{Am}}(S, t) = K - S\}$

Wegen c) und d) gilt  $V = K - S$  für alle  $S < S^*$ .

Es bleibt das Verhalten von  $V$  für  $S \geq S^*$  zu untersuchen.

**Behauptung 2**

(Voraussetzung: Für alle  $S < S^*$  gelte  $V = K - S$ )

Die rechtsseitige Ableitung der Funktion  $V_P^{\text{Am}}(S, t)$  an der Stelle  $S^*$  hat den Wert  $-1$ .

*Beweisskizze:*

Wir benutzen die Notation  $V := V_P^{\text{Am}}$ , und  $\frac{\partial V}{\partial S}$  meint die rechtsseitige Ableitung.

Für die rechtsseitige Ableitung kann  $\frac{\partial V(S^*, t)}{\partial S} < -1$  nicht sein, weil sonst für  $S > S^*$  die Eigenschaft d) verletzt wäre. Also gilt  $\frac{\partial V(S^*, t)}{\partial S} \geq -1$ .

*Annahme:*

$$\frac{\partial V(S^*, t)}{\partial S} > -1.$$

*Widerlegung:*

Setze ein Portfolio an:  $\Pi := V + S$ , also Anfangswert

$$\Pi^* := V + S^*.$$

(=  $K$ ; leihe hierfür von der Bank den Betrag  $K$ .) Für die unterstellte GBM

$$dS = rS dt + \sigma S dW$$

treffen wir die Annahme, dass  $dt$  so klein ist, dass  $\sqrt{dt} \gg dt$ . Ist  $dt$  klein genug, dann gilt (intuitiv)

$$dW > 0 \iff dS > 0.$$

Das Itô-Lemma ergibt

$$\begin{aligned} d\Pi &= (\dots)dt + \frac{\partial \Pi}{\partial S} \sigma S dW \\ &= O(dt) + \left( \frac{\partial V}{\partial S} + 1 \right) \sigma S dW. \end{aligned}$$

Für  $dS > 0$  ist dies für hinreichend kleine  $dt$  wegen der Annahme und  $dW > 0$  positiv. Für  $dS < 0$  gilt  $\Pi \equiv K$  und deswegen  $d\Pi = 0$ .

Insgesamt:  $E(d\Pi) > 0$  und von der Ordnung  $O(\sqrt{dt})$ . Verkaufe das Portfolio nach Ablauf von  $dt$  und erwarde folgende Bilanz

$$\begin{aligned}\Pi^* + E(d\Pi) - Ke^{r dt} &= V + S^* + E(d\Pi) - K(1 + O(dt)) \\ &= E(d\Pi) + O(dt)\end{aligned}$$

Dies ist positiv weil  $E(d\Pi)$  wegen der Kleinheit von  $dt$  dominiert. Widerspruch zum No-Arbitrage-Prinzip!

Andere Beweise verwenden das Argument, den Wert der Option zu maximieren. Dies kann für den Fall der *perpetual option* (eine Option ohne Fälligkeit) elementar durchgeführt werden. (Übungen)

Zusammenfassend bleibt im Fall eines amerikanischen Puts der Schluss, dass Folgendes gilt:

- (1)  $V(S^*, t) = K - S^*$  (Dirichlet Randbedingung)
- (2)  $\frac{\partial V(S^*, t)}{\partial S} = -1$  (Neumann Randbedingung)
- (3)  $V(S, t) > K - S$  für  $S > S^*$
- (4)  $V(S, t) = K - S$  für  $S \leq S^*$

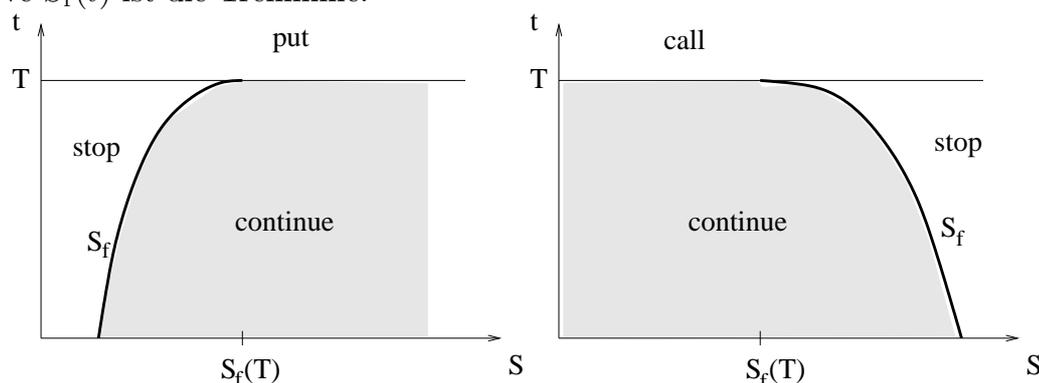
Die Eigenschaft (2) heißt *Kontaktbedingung* (engl. *high contact, smooth pasting*).

Da für jedes  $t$  ein derartiges  $S^*$  existiert, ist hierdurch eine Funktion definiert, welche wir mit  $S_f(t)$  bezeichnen. Das “f” steht für “freier Rand”, bzw. *free-boundary problem*.

$S_f(t)$  bewirkt eine Zweiteilung des Halbstreifens, nämlich

- 1.)  $S > S_f$  heißt Fortsetzungsbereich, auch “Haltebereich” (*continuation region*) des Puts.
- 2.)  $S \leq S_f$  heißt Stoppbereich (*stopping region*) des Puts.

Für Standard-Optionen ohne diskrete Dividende sind diese Bereiche einfach zusammenhängend. Die Kurve  $S_f(t)$  ist die Trennlinie.



### Begründung

- 1.) Falls  $V_P > (K - S)^+$ , so bedeutet vorzeitiges Ausüben  $-V + K - S < 0$ , also Verlust. Also übt man in diesem Bereich *nicht* aus: “continuation”, d.h. die Option wird gehalten.
- 2.) Wenn der Kurs die Kurve  $S_f(t)$  erreicht bzw. überschreitet zu  $S \leq S_f(t)$ , dann sollte man sofort vorzeitig ausüben (“stopping”), weil so der Geldbetrag  $K$ , der beim

Ausüben erhalten wird, angelegt werden kann und für  $r > 0$ ,  $t < T$  folgender Gewinn erzielt wird:

$$Ke^{r(T-t)} - K = K(e^{r(T-t)} - 1)$$

Bei Ausüben ist die Schlussbilanz  $Ke^{r(T-t)}$  größer als  $Se^{\delta(T-t)}$ , zumindest für  $r(T-t) < 1$ .

“Freies Randwertproblem” bedeutet:

Die Black-Scholes-Gleichung ist *nur* im Fortsetzungsbereich gültig, und nicht im Stopp-Bereich. D.h. ihr Definitionsbereich für den amerikanischen Put ist

$$S_f(t) < S < \infty.$$

Der linke Rand  $S_f(t)$  ist “frei”, insbesondere ist er zunächst unbekannt, und muss erst numerisch berechnet werden. Die zusätzliche Randbedingung, welche implizit die Lage von  $S_f(t)$  bestimmt, ist die Kontaktbedingung  $\frac{\partial V}{\partial S} = -1$ .

Die Eigenschaften für einen Call können analog hergeleitet werden.

**Wichtige Eigenschaften** der early-exercise-Kurve **im Falle eines Puts** nach dem Black-Scholes-Modell (kontinuierliche Dividendenrate  $\delta \geq 0$  ist möglich, aber diskrete Dividendenzahlung wird hier ausgeschlossen) sind folgende:

- 1.)  $S_f(t)$  ist für  $t < T$  stetig differenzierbar, und
- 2.)  $S_f(t)$  ist monoton wachsend.
- 3.)

$$S_f(T) := \lim_{\substack{t \rightarrow T \\ t < T}} S_f(t) = \begin{cases} K & \text{für } 0 \leq \delta \leq r \\ \frac{r}{\delta}K & \text{für } r < \delta \end{cases}$$

*Beweis* von 3.) (Bez.  $V = V_P^{\text{Am}}$ )

Wegen  $V^{\text{Am}} \geq V^{\text{Eu}}$  gilt  $S_f(t) \leq S^{*\text{Eu}}(t) < K$  für alle  $t < T$ . Also  $S_f(T) \leq K$ .

**Vorbemerkung:** Für  $V(S, t)$  gilt wegen  $V \geq \Psi$  und  $V(S, T) = \Psi(S)$  für  $t = T$

$$\frac{\partial V}{\partial t} \leq 0.$$

Als Vorbereitung für einige indirekte Beweise studieren wir, wie sich dieses Vorzeichen im Fall  $S_f(T) < S < K$  mit der Black-Scholes-Gleichung verträgt. Für  $t = T$  erhalten wir mit  $V(S, T) = K - S$

$$\frac{\partial V}{\partial t} + 0 - (r - \delta)S - rV = 0$$

$$\implies \frac{\partial V(S, T)}{\partial t} = rK - \delta S.$$

Für welche  $(\delta, r)$ -Kombinationen gilt nun  $\frac{\partial V(S, T)}{\partial t} \leq 0$  ?

**Fall:**  $\delta > r$

Hier gilt  $\frac{r}{\delta}K < K$ . Es gilt entweder  $S_f(T) = \frac{r}{\delta}K$  (also die Behauptung), oder es existiert eines von zwei offenen Teilintervallen (i)  $S_f(T) < \frac{r}{\delta}K$  und (ii)  $\frac{r}{\delta}K < S_f(T)$ .

- (i) Für  $S$  im Intervall  $S_f(T) < S < \frac{r}{\delta}K$  gilt  $\frac{\partial V(S,T)}{\partial t} = rK - \delta S > 0$ , ein Widerspruch zu

$$\frac{\partial V}{\partial t} \leq 0.$$

- (ii) Für jedes  $S$  im Intervall  $\frac{r}{\delta}K < S < S_f(T)$  gibt es ein kleines  $dt$ , so dass  $(S, T-dt)$  im Stoppbereich liegt. Hier gilt  $rK < \delta S$  und deswegen auch  $rKdt < \delta Sdt$ , oder

$$K(e^{rdt} - 1) < S(e^{\delta dt} - 1).$$

Das heißt, der Dividendenenertrag über einen Zeitraum  $dt$ , den man bei Fortsetzen der Option erhielte, ist größer als der Zinsertrag bei Ausüben. Somit ist vorzeitiges Ausüben nicht optimal, welches ein Widerspruch zur Bedeutung von  $S < S_f(t)$  ist.

Es muss im Fall  $\delta > r$  also  $S_f(T) = \frac{r}{\delta}K$  gelten.

**Fall:**  $\delta \leq r$

Annahme  $S_f(T) < K$ . Für  $S$  im Intervall  $S_f(T) < S < K$  erhalte einen Widerspruch aus

$$\underbrace{\frac{\partial V}{\partial t}}_{\leq 0} = \underbrace{rK - \delta S}_{> 0}.$$

**Hinweis:** Im Fall von *diskreten* Dividendenzahlungen gelten diese Aussagen so nicht mehr. Insbesondere ist  $S_f$  dann unstetig! Zum Beispiel ist bei einem amerikanischen Put ein vorzeitiges Ausüben in einem Zeitintervall vor dem Ausschüttungszeitpunkt sinnlos.

## 4.6 Komplementarität

Gesucht ist eine numerische Methode, welche die Unbekannte  $S_f$  nicht explizit verwendet.

### A. Ungleichung

Die Black-Scholes-Gleichung gilt bei amerikanischen Optionen nicht mehr im ganzen Halbstreifen  $0 \leq t \leq T$ ,  $0 < S$ , sondern sie gilt nur noch im Fortsetzungsbereich. Was gilt nun aber im Stoppbereich?

Bezeichnung

$$\mathcal{L}_{BS}(V) := \frac{1}{2}\sigma^2 S^2 \frac{\partial^2 V}{\partial S^2} + (r - \delta)S \frac{\partial V}{\partial S} - rV.$$

Damit lautet die Black-Scholes-Gleichung

$$\frac{\partial V}{\partial t} + \mathcal{L}_{BS}(V) = 0.$$

Für  $S < S_f$  gilt für den Put  $V \equiv \Psi$  und deshalb

$$V = K - S, \quad \frac{\partial V}{\partial S} = -1, \quad \frac{\partial^2 V}{\partial S^2} = 0, \quad \frac{\partial V}{\partial t} = 0,$$

also

$$\frac{\partial V}{\partial t} + \mathcal{L}_{\text{BS}}(V) = \delta S - rK.$$

Aus den oberen Schranken  $S < S_f(T)$  ergibt sich:

im Fall  $r < \delta$ :  $S < \frac{r}{\delta}K$ , also  $\delta S - rK < 0$

im Fall  $r \geq \delta$ :  $\delta S - rK \leq r(S - K) < 0$  weil  $S < K$

also:

$$\frac{\partial V}{\partial t} + \mathcal{L}_{\text{BS}}(V) < 0.$$

Für einen Call folgt das gleiche Ergebnis aus analogen Überlegungen.

**Folgerung:** Für alle  $S > 0$ ,  $0 \leq t \leq T$  gilt die partielle **Differential-Ungleichung**

$$\frac{\partial V}{\partial t} + \mathcal{L}_{\text{BS}}(V) \leq 0.$$

Dies gilt für den Put und für den Call.

### Übersicht

Put:	$V_{\text{P}}^{\text{Am}} = K - S$ für $S \leq S_f$
	$V_{\text{P}}^{\text{Am}}$ löst die BS-Gleichung für $S > S_f$
	Kontaktbedingung: $\frac{\partial V(S_f, t)}{\partial S} = -1$
Call:	$V_{\text{C}}^{\text{Am}} = S - K$ für $S \geq S_f$
	$V_{\text{C}}^{\text{Am}}$ löst die BS-Gleichung für $S < S_f$
	Kontaktbedingung: $\frac{\partial V(S_f, t)}{\partial S} = 1$

**Bemerkungen:** Die zweite Ableitung von  $V$  nach  $S$  ist an  $S_f$  unstetig. D.h.  $V$  ist zwar glatt im Inneren des Fortsetzungsgebietes, nicht jedoch auf dem gesamten Halbstreifen. Die Transformation von Abschnitt 4.1 bedeutet u.a.

$$\frac{\partial V}{\partial t} + \mathcal{L}_{\text{BS}}(V) = -\frac{\partial y}{\partial \tau} + \frac{\partial^2 y}{\partial x^2}.$$

## B. Formulierung mit Strafterm

Eine einheitliche Behandlung von  $\frac{\partial V}{\partial t} + \mathcal{L}_{\text{BS}}(V) \leq 0$  auf dem ganzen Halbstreifen ist möglich. Hierzu verlange für eine geeignete nicht-negative Funktion  $p(V)$

$$\frac{\partial V}{\partial t} + \mathcal{L}_{\text{BS}}(V) + p(V) = 0.$$

Der Strafterm (*penalty term*)  $p$  sollte 0 sein im Fortsetzungsbereich und positiv im Stoppbereich. Den Abstand zu  $S_f$  kennt man nicht, aber den Abstand  $V - \Psi$  von  $V$  zum Payoff  $\Psi$ , und der wird als Steuergröße verwendet. Ein Beispiel eines Strafterms ist

$$p(V) := \frac{\epsilon}{V - \Psi} \quad \text{für ein kleines } \epsilon > 0.$$

Es bezeichne  $V_\epsilon(S, t)$  die Lösung obiger mit dem Strafterm versehenen partiellen Differentialgleichung. Zwei extreme Fälle charakterisieren die Wirkungsweise des Strafterms, für  $(S, t)$  im Fortsetzungsbereich und im Stoppbereich:

- $V_\epsilon - \Psi \gg \epsilon$  impliziert  $p \approx 0$ . Dann reduziert sich die Gleichung i.W. zur Black–Scholes–Gleichung.
- $0 < V_\epsilon - \Psi \ll \epsilon$  impliziert einen großen Wert von  $p$ . Dann wird der BS-Anteil der Gleichung von  $p$  dominiert, die Black–Scholes Gleichung ist außer Kraft gesetzt, und  $V_\epsilon \approx \Psi$ .

Die diesen beiden Fällen entsprechenden Zweige der Lösung  $V_\epsilon$  seien hier “Fortsetzungszweig” ( $p \approx 0$ ) und “Stoppzweig” ( $V_\epsilon \approx \Psi$ ) genannt. Offensichtlich approximieren diese beiden Zweige die exakte Lösung  $V$  des Black–Scholes-Problems. Der Zwischenbereich mit  $V_\epsilon - \Psi = O(\epsilon)$  kennzeichnet eine Grenzschicht zwischen Fortsetzungsbereich und Stoppbereich, eine Grenzschicht um  $S_f$ . Hier kann  $V_\epsilon$  als eine Verbindung zwischen der BS-Fläche und der Payoff-Ebene angesehen werden.<sup>a</sup>

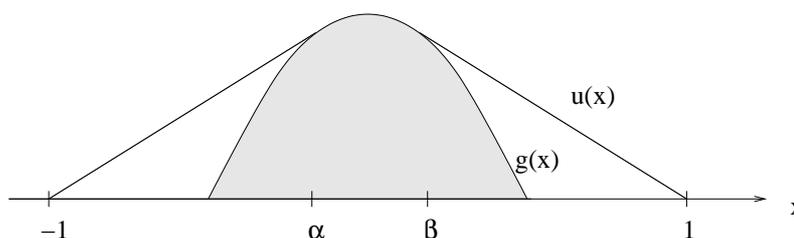
**Bemerkung:**  $p$  und die resultierende PDE sind nichtlinear in  $V$ . Eine Implementierung muss darauf achten, dass  $V_\epsilon \leq \Psi$  vermieden wird; nicht jede Wahl von  $\epsilon$  und  $\Delta t$  funktioniert.

Im Folgenden suchen wir nach einer eleganteren Lösung, die auf die eindimensionale Black-Scholes-Situation zugeschnitten ist.

### C. Einfaches Hindernisproblem (Exkurs)

Wir betrachten ein “Hindernis”  $g$  mit den Eigenschaften<sup>b</sup>

$$g > 0 \text{ für ein Teilintervall von } -1 < x < 1, g \in C^2, g'' \leq 0, g(-1) < 0 \text{ und } g(1) < 0.$$



Charakterisierung einer Funktion  $u \in C^1$  mit minimaler Länge mit  $u(-1) = u(1) = 0$  und  $u \geq g$  für das  $g$  der Figur: Es gibt  $\alpha, \beta$  so dass

$$\begin{aligned} -1 < x < \alpha &: u'' = 0 \quad (u > g) \\ \alpha < x < \beta &: u = g \quad (u'' = g'' \leq 0) \\ \beta < x < 1 &: u'' = 0 \quad (u > g) \end{aligned}$$

Darin drückt sich eine Komplementarität aus im Sinne von

$$\begin{aligned} \text{falls } u - g > 0, & \text{ dann } u'' = 0, \\ \text{falls } u - g = 0, & \text{ dann } u'' \leq 0. \end{aligned}$$

In der Sprache der Optionen heißt dies

$$\begin{aligned} \text{falls } V - \Psi > 0, & \text{ dann } \dot{V} + \mathcal{L}_{\text{BS}}(V) = 0, \\ \text{falls } V - \Psi = 0, & \text{ dann } \dot{V} + \mathcal{L}_{\text{BS}}(V) \leq 0. \end{aligned}$$

<sup>a</sup> Illustrationen finden sich in Topic 9 auf der Homepage [www.compfin.de](http://www.compfin.de); dort ist auch die Payoff-Funktion  $p$  gezeigt.

<sup>b</sup> Dieser Abschnitt und weitere Teile dieses Kapitels lehnen sich an [P. Wilmott, J. Dewynne, S. Howison: The Mathematics of Financial Derivatives] an.

Für das einfache Hindernisproblem lautet eine äquivalente Formulierung:

$$\begin{aligned} &\text{Finde eine Funktion } u, \text{ so dass} \\ &u''(u - g) = 0, \quad -u'' \geq 0, \quad u - g \geq 0, \quad (*) \\ &u(-1) = u(1) = 0, \quad u \in \mathcal{C}^1[-1, 1]. \end{aligned}$$

$\alpha$  und  $\beta$  treten hier nicht explizit auf. Durch die numerische Lösung werden die Werte von  $\alpha$  und  $\beta$  im Nachhinein ablesbar sein. Das Problem (\*) ist ein **lineares Komplementaritätsproblem** (LCP); es hat die Form

$$\mathcal{A} \cdot \mathcal{B} = 0, \quad \mathcal{A} \geq 0, \quad \mathcal{B} \geq 0, \quad \text{für geeignete } \mathcal{A}, \mathcal{B}.$$

Für eine **numerische Lösung** des einfachen Hindernisproblems verwende das Gitter

$$x_i = -1 + i\Delta x, \quad \Delta x := \frac{2}{m}, \quad g_i := g(x_i).$$

Dies ergibt die diskrete Form des Hindernisproblems

$$\left. \begin{aligned} (w_{i-1} - 2w_i + w_{i+1})(w_i - g_i) &= 0, \\ -w_{i-1} + 2w_i - w_{i+1} &\geq 0, \\ w_i - g_i &\geq 0 \end{aligned} \right\} \text{ für } 0 < i < m$$

mit  $w_0 = w_m = 0$ . Passend dazu definieren wir

$$G := \begin{pmatrix} 2 & -1 & & 0 \\ -1 & \ddots & \ddots & \\ & \ddots & \ddots & -1 \\ 0 & & -1 & 2 \end{pmatrix} \quad \text{und} \quad w := \begin{pmatrix} w_1 \\ \vdots \\ w_{m-1} \end{pmatrix}, \quad g := \begin{pmatrix} g_1 \\ \vdots \\ g_{m-1} \end{pmatrix}.$$

Damit lautet das lineare Komplementaritätsproblem in Vektornotation

$$\left. \begin{aligned} (w - g)^T G w &= 0 \\ G w &\geq 0 \\ w - g &\geq 0 \end{aligned} \right\}$$

(Ende des Exkurses)

## D. Komplementarität für das Black-Scholes-Problem

Speziell für den amerikanischen Put wird nun das transformierte Problem

$$\frac{\partial y}{\partial \tau} = \frac{\partial^2 y}{\partial x^2} \quad \text{solange} \quad V_P^{\text{am}} > (K - S)^+$$

als LCP formuliert. Dazu müssen wir auch die Nebenbedingungen transformieren.

$$\begin{aligned} V_P^{\text{Am}}(S, t) &\geq (K - S)^+ \\ &= K \max\{1 - e^x, 0\} \end{aligned}$$

führt mit  $q = \frac{2r}{\sigma^2}$  auf (hier speziell für  $\delta = 0$ ):

$$\begin{aligned} y(x, \tau) &\geq \exp\left\{\frac{1}{2}(q-1)x + \frac{1}{4}(q+1)^2\tau\right\} \max\{1 - e^x, 0\} \\ &= \exp\left\{\frac{1}{4}(q+1)^2\tau\right\} \max\{(1 - e^x)e^{\frac{1}{2}(q-1)x}, 0\} \\ &= \exp\left\{\frac{1}{4}(q+1)^2\tau\right\} \max\{e^{\frac{1}{2}(q-1)x} - e^{\frac{1}{2}(q+1)x}, 0\} \\ &=: g(x, \tau). \end{aligned}$$

Es stellt sich heraus, dass mit diesem  $g$  auch die Rand- und Anfangsbedingungen ausdrückbar sind:

$$\begin{aligned} y(x, 0) &= g(x, 0) \\ y(x_{\min}, \tau) &= g(x_{\min}, \tau) \\ y(x_{\max}, \tau) &= g(x_{\max}, \tau) \end{aligned}$$

Damit lautet das lineare Komplementaritätsproblem für den amerikanischen Put

$$\left. \begin{aligned} \left(\frac{\partial y}{\partial \tau} - \frac{\partial^2 y}{\partial x^2}\right) (y - g) &= 0 \\ \frac{\partial y}{\partial \tau} - \frac{\partial^2 y}{\partial x^2} &\geq 0 \\ y - g &\geq 0 \end{aligned} \right\}$$

mit Rand- und Anfangsbedingungen von oben.

Diese Formulierung ist identisch mit derjenigen des allgemeineren Falls  $\delta \neq 0$ , bis auf eine Anpassung von  $g$ :

$$q = \frac{2r}{\sigma^2}, \quad q_\delta = \frac{2(r - \delta)}{\sigma^2},$$

$$\text{Put: } (r > 0) \quad g(x, \tau) := \exp\left\{\frac{1}{4}((q_\delta - 1)^2 + 4q)\tau\right\} \max\{e^{\frac{1}{2}(q_\delta - 1)x} - e^{\frac{1}{2}(q_\delta + 1)x}, 0\}$$

$$\text{Call: } (\delta > 0) \quad g(x, \tau) := \exp\left\{\frac{1}{4}((q_\delta - 1)^2 + 4q)\tau\right\} \max\{e^{\frac{1}{2}(q_\delta + 1)x} - e^{\frac{1}{2}(q_\delta - 1)x}, 0\}$$

**Bemerkung:** Für einen amerikanischen Call ohne Dividendenzahlung und  $r > 0$ ,  $t < T$  gilt  $V_C^{\text{Am}} = V_C^{\text{Eur}}$ . Beweis für  $\delta = 0$ :

$$V_C^{\text{Am}} \geq V_C^{\text{Eur}} \geq S - Ke^{-r(T-t)} > S - K$$

Also  $V_C^{\text{Am}} > \text{Payoff}$ ; kein vorzeitiges Ausüben.

## 4.7 Numerische Realisierung

### A. Diskretisierung

Wir diskretisieren nun das LCP-Problem mit dem Gitter vom Beginn des Kapitels:

$$\begin{aligned} \frac{w_{i,\nu+1} - w_{i,\nu}}{\Delta\tau} &= \theta \frac{w_{i+1,\nu+1} - 2w_{i,\nu+1} + w_{i-1,\nu+1}}{\Delta x^2} + \\ &\quad (1 - \theta) \frac{w_{i+1,\nu} - 2w_{i,\nu} + w_{i-1,\nu}}{\Delta x^2} \end{aligned}$$

(“Theta-Methode”), mit  $\theta = 0$  (explizites Verfahren),  $\theta = \frac{1}{2}$  (Crank–Nicolson) oder  $\theta = 1$  (BTCS-Verfahren). Mit der Abkürzung  $\lambda := \frac{\Delta\tau}{(\Delta x)^2}$  wird aus  $\frac{\partial y}{\partial \tau} - \frac{\partial^2 y}{\partial x^2} \geq 0$  nun

$$\begin{aligned} w_{i,\nu+1} - \lambda\theta(w_{i+1,\nu+1} - 2w_{i,\nu+1} + w_{i-1,\nu+1}) \\ - w_{i,\nu} - \lambda(1-\theta)(w_{i+1,\nu} - 2w_{i,\nu} + w_{i-1,\nu}) \geq 0. \end{aligned} \quad (*)$$

Passend dazu definieren wir für die Terme der  $\nu$ -ten Zeitschicht

$$b_{i,\nu} := w_{i,\nu} + \lambda(1-\theta)(w_{i+1,\nu} - 2w_{i,\nu} + w_{i-1,\nu}), \text{ für } i = 2, \dots, m-2.$$

Die Randbedingungen sind in  $b_{1,\nu}$  und  $b_{m-1,\nu}$  enthalten:

$$\begin{aligned} b_{1,\nu} &:= w_{1,\nu} + \lambda(1-\theta)(w_{2,\nu} - 2w_{1,\nu} + g_{0,\nu}) + \lambda\theta g_{0,\nu+1} \\ b_{m-1,\nu} &:= w_{m-1,\nu} + \lambda(1-\theta)(g_{m,\nu} - 2w_{m-1,\nu} + w_{m-2,\nu}) + \lambda\theta g_{m,\nu+1}, \end{aligned}$$

wobei

$$g_{i,\nu} := g(x_i, \tau_\nu) \quad (0 \leq i \leq m, 0 \leq \nu \leq \nu_{\max}).$$

Definiere nun

$$b^{(\nu)} := (b_{1,\nu}, \dots, b_{m-1,\nu})^t$$

und  $w^{(\nu)}, g^{(\nu)}$  analog. Mit der Matrix

$$A := \begin{pmatrix} 1 + 2\lambda\theta & -\lambda\theta & & 0 \\ -\lambda\theta & \ddots & \ddots & \\ & \ddots & \ddots & -\lambda\theta \\ 0 & & -\lambda\theta & 1 + 2\lambda\theta \end{pmatrix} \in \mathbb{R}^{(m-1) \times (m-1)}$$

lautet das Problem (\*)

$$Aw^{(\nu+1)} \geq b^{(\nu)} \quad \text{für alle } \nu.$$

Aus  $y - g \geq 0$  wird

$$w^{(\nu)} \geq g^{(\nu)}$$

und schließlich wird aus  $\left(\frac{\partial y}{\partial \tau} - \frac{\partial^2 y}{\partial x^2}\right)(y - g) = 0$

$$\left(Aw^{(\nu+1)} - b^{(\nu)}\right)^t \left(w^{(\nu+1)} - g^{(\nu+1)}\right) = 0.$$

Wir erhalten so den folgenden Makro-Algorithmus:

### Algorithmus

Für  $\nu = 0, 1, \dots, \nu_{\max} - 1$ :

berechne  $g := g^{(\nu+1)}$ ,  $b := b^{(\nu)}$ , s.o.

berechne  $w$  als Lösung von

$$Aw - b \geq 0, \quad w \geq g, \quad (Aw - b)^t(w - g) = 0. \quad (**)$$

setze  $w^{(\nu+1)} := w$

Für jede Zeitschicht  $\nu$  ist in (\*\*) ein LCP zu lösen, das wir im Folgenden umformen und analysieren.

### Behauptung

Mit der Transformation  $x := w - g$  und  $y := Aw - b$  ist (\*\*) äquivalent zu:

$$\begin{array}{l} \text{Berechne Vektoren } x, y, \text{ so dass für } \hat{b} := b - Ag \text{ Folgendes gilt:} \\ Ax - y = \hat{b}, \quad x \geq 0, \quad y \geq 0, \quad x^T y = 0 \end{array} \quad (***)$$

Beachte, dass  $x, y$  hier allgemeine Vektoren  $\in \mathbb{R}^{m-1}$  sind!

*Beweis:* Anwendung der Transformation.

### Lemma

Die Aufgabe (\*\*\*) hat eine eindeutige Lösung.

*Beweis:*

1.) Definiere

$$G(x) := \frac{1}{2} x^T A x - \hat{b}^T x$$

mit  $A$  von oben, d.h.

$$A = I + \lambda\theta \begin{pmatrix} 2 & -1 & & 0 \\ -1 & \ddots & \ddots & \\ & \ddots & \ddots & -1 \\ 0 & & -1 & 2 \end{pmatrix}$$

Nach Lemma 2 aus Abschnitt 4.2D erhalten wir die Eigenwerte der Matrix  $A$  als

$$\mu_k^A = 1 + \lambda\theta 4 \sin^2(k\pi/2m) \geq 1, \quad k = 1, \dots, m-1.$$

Somit ist die symmetrische Matrix  $A$  positiv definit. Durch Ableiten von  $G$  nach  $x$  erhalte

$$G_x = Ax - \hat{b}, \quad G_{xx} = A,$$

d.h.  $A$  ist die Hesse-Matrix von  $G$ . Da  $G \in C^2$  ist, gilt:

$G$  ist strikt konvex.  $\iff$  Die Hesse-Matrix von  $G$  ist positiv definit.

Also ist die oben definierte quadratische Form  $G$  strikt konvex und hat somit auf jeder konvexen Teilmenge des  $\mathbb{R}^N$  ein eindeutiges Minimum,  $N = m-1$ . Man beachte, dass durch  $x \geq 0$  eine konvexe Menge definiert ist.

2.) Verwende den Satz von Karush, Kuhn, Tucker (KKT):

Für das Minimum  $x^*$  einer konvexen Funktion  $G$  unter den Nebenbedingungen  $H_i(x) \leq d_i$  für  $i = 1, \dots, N$ , wobei  $H_i$  konvex sind, gilt: Es gibt  $y_i \geq 0$  so dass

$$\frac{\partial G(x^*)}{\partial x_j} + y_1 \frac{\partial H_1(x^*)}{\partial x_j} + \dots + y_N \frac{\partial H_N(x^*)}{\partial x_j} = 0 \quad (j = 1, \dots, N)$$

mit  $H_i(x^*) \leq d_i$  und  $y_i(H_i(x^*) - d_i) = 0$  für  $i = 1, \dots, N$ .

(Literatur: z.B. [Stoer & Witzgall], [Strang], u.v.a.)

Anwendung mit  $N = m-1$ : Die Ungleichung  $x \geq 0$  oder  $x_i \geq 0$  für alle  $i = 1, \dots, m-1$  impliziert  $d_i = 0$  und  $H_i(x) = -x_i$ . Aus dem Satz von KKT folgt die Existenz von  $y \geq 0$  mit

$$\frac{\partial G}{\partial x_j} + y_j \frac{\partial H_j}{\partial x_j} + 0 = \frac{\partial G}{\partial x_j} - y_j = 0$$

und  $y_i x_i = 0$  für alle  $i$ . Für unser spezielles  $G$  gilt:

$$G_x = Ax - \hat{b} \implies Ax - \hat{b} - y = 0$$

D.h. (\*\*\*) resultiert aus dem KKT-Satz und hat deswegen eine eindeutige Lösung. Dies gilt dann auch für (\*\*).

## B. Numerische Lösung

Die Aufgabe (\*\*) kann schnell **direkt** gelöst werden. Brennan & Schwartz schlagen zur Lösung von (\*\*) folgendes Vorgehen vor:

Löse  $Aw = b$  *komponentenweise* unter der NB  $w \geq g$

Dieses Prinzip alleine reicht noch nicht: Die Implementierung ist der Schlüssel. Sie beruht auf dem Gaußschen Algorithmus, der in einer ersten Phase  $Aw = b$  zu einem äquivalenten System  $\tilde{A}w = \tilde{b}$  umformt, so dass  $\tilde{A}$  eine Dreiecksmatrix (hier Bidiagonalmatrix) ist. Nach Berechnung von  $\tilde{A}, \tilde{b}$  lässt sich in einer zweiten Phase obiges Prinzip mit einer Schleife lösen. Wenn  $\tilde{A}$  ein oberes Dreieck ist, dann erfolgt die abschließende Schleife zur Lösung von  $\tilde{A}w = \tilde{b}$  *rückwärts*. Wenn  $\tilde{A}$  ein unteres Dreieck ist, dann erfolgt diese Schleife *vorwärts*. Wenn im  $i$ -ten Schritt der Schleife  $\tilde{w}_i$  die Komponente der Lösung von  $\tilde{A}w = \tilde{b}$  ist, dann scheint mit  $w_i := \max\{\tilde{w}_i, g_i\}$  eine Realisierung gefunden zu sein.

Aber  $w$  hängt davon ab, wie die Schleife durchlaufen wird. Nur eine der beiden Möglichkeiten funktioniert. Eine Implementierung muss sicherstellen, dass die für die jeweilige Option charakteristische Ordnung erreicht wird. Für den **Put** heißt das:

Es sei  $i_f$  der Index des Knotens  $S_i$ , der am nächsten am Kontaktpunkt liegt\*, also da wo  $V$  den Payoff berührt. Genauer, in den transformierten Variablen,

für  $i \leq i_f$  gilt  $w_i = g_i$  und

für  $i_f < i \leq m$  gilt  $w_i > g_i$ .

Diese Struktur ist durch die Theorie festgelegt, jedoch ist  $i_f$  unbekannt. Bei einem Put ist der Startpunkt  $w_1 = g_1$ , und die  $w_i := \max\{\tilde{w}_i, g_i\}$ -Schleife erfolgt *vorwärts*. Deswegen muss  $\tilde{A}$  ein unteres Dreieck sein. Also ist für einen Put eine *RL*-Zerlegung von  $A$  notwendig. Dies erzeugt eine untere Dreiecksmatrix  $\tilde{A} := L$  und  $\tilde{b}$  aus  $R\tilde{b} = b$ . Zu Beginn der Schleife werden die ersten Komponenten  $w_i = g_i$  sein, bis zu dem ersten Index, an dem  $w_i > g_i$  gilt. Dies ergibt den Index  $i_f$ .

---

\* Das setzt ein genügend großes Intervall voraus,  $S_1 < S_f$ .

**Algorithmus (Put)**

1. Phase: Berechne  $RL$ -Zerlegung von  $A$ . Dann setze  $\tilde{A} = L$  und berechne  $\tilde{b}$  (Rückwärts-Schleife).
2. Phase: Vorwärts-Schleife; starte mit  $i = 1$ . Berechne die jeweils nächste Komponente von  $\tilde{A}w = \tilde{b}$ ; sie sei  $\tilde{w}_i$ . Setze  $w_i := \max\{\tilde{w}_i, g_i\}$ .

Der **Aufwand** ist gering (Lösung eines linearen Systems mit Tridiagonalmatrix). Es kann gezeigt werden, dass obiges Vorgehen für die Standard-Option mit der hier vorliegenden Matrix  $A$  funktioniert.

Für den **Call** verfährt man genau umgekehrt: Der Startpunkt ist  $w_m = g_m$ , und die zweite Phase wird rückwärts durchlaufen. Damit dies funktioniert, wird in der ersten Phase durch die "traditionelle"  $LR$ -Zerlegung von  $A$  ein oberes Dreieck  $\tilde{A} = R$  erzeugt, und  $\tilde{b}$  ergibt sich aus  $L\tilde{b} = b$  in einer Vorwärts-Schleife.

**Bemerkung zur Genauigkeit:**

Da  $V(S, t)$  nicht zweimal stetig differenzierbar bzgl.  $S$  an  $S_f(t)$  ist, hat dies Auswirkungen auf die Genauigkeit. Der Konvergenzbeweis von Crank-Nicolson fordert sogar  $y \in C^4$ . Trotz der fehlenden Glätte ist der Crank-Nicolson-Ansatz hier von ausreichender Genauigkeit, Oszillationen werden relativ schnell gedämpft. Die fehlende Glattheit des Payoff ist gravierender. Sogar Extrapolation funktioniert ziemlich gut, obwohl die (hinreichenden) Voraussetzungen nicht erfüllt sind.

## Ausblick

Das beendet die Einführung in Grundlagen der numerischen Finanzmathematik. Ein wichtiger Bestandteil der Veranstaltung sind die Übungen, die zu dem Skript dazugehören (von der Homepage herunterzuladen), insbesondere die Programmieraufgaben.

Wird der Stoff dieses Skripts in einem Semester erklärt, dann bleibt am Ende des Semesters noch Zeit für Ergänzungen oder Vertiefungen. (Bei mir sind es typischerweise zwei Wochen.) Dieser weitere Stoff ist hier im Skript nicht ausgeführt, weil er in jedem Kurs anders sein wird, abhängig von den jeweiligen Vorkenntnissen und Interessen. Für eine lehrbuchartige Aufbereitung weiteren Materials siehe [Seydel: Tools for Computational Finance. 6.Auflage (2017)], hieraus auch die folgenden Abschnittsnummern. Mögliche Themen sind zum Beispiel:

- als weitere analytische Methode die Integralgleichung nach Kim (§4.8.4)
- *upwind*-Schemata und ihre Bedeutung (§6.4 – 6.5)
- eine Penalty-Methode im zweidimensionalen Fall (§6.7)
- Fallstudien, wie z.B. die zweidimensionale Baummethode von Übungsaufgabe 6.2
- Sprung-Diffusionsprozesse (§1.9, §7.3)

Wenn eine solche Veranstaltung sogar zweisemestrig gelesen oder durch ein Seminar ergänzt werden kann, so bieten sich an

- Finite Elemente (Kapitel 5)
- Nichtlineare Black–Scholes-Probleme (§7.1 – 7.2)