

Offen im Denken

UNIVERSITÄT DUISBURG-ESSEN

FAKULTÄT FÜR MATHEMATIK

LEHRSTUHL FÜR ANGEWANDTE STOCHASTIK

Generative Zeitreihenmodellierung mittels Schrödinger-Brücke mit Kernel-Regression

eingereicht von:
Stephan Fischer
Arnoldstraße 19
41751 Viersen

Erstprüfer:
Prof. Dr. Denis Belomestny
Zweitprüfer:
Prof. Dr. Mikhail Urusov

Wintersemester 2024/2025

Zusammenfassung

Diese Arbeit untersucht generative Modelle für Zeitreihen aus der Perspektive des Schrödinger-Brücken-Problems, einem klassischen Problem in stochastischen Prozessen und im optimalen Transport. Ein besonderer Schwerpunkt der Forschung liegt auf der Verbindung zwischen theoretischen Grundlagen und der praktischen Umsetzung von Schrödinger-Brücken-Methoden. Ziel ist die Generierung von Zeitreihen, die sowohl vorgegebene statistische Eigenschaften als auch Randbedingungen erfüllen. Die Arbeiten umfassen eine detaillierte Analyse stochastischer Differentialgleichungen, die Entwicklung numerischer Verfahren zur effizienten Lösung von Schrödinger-Brücken-Problemen und deren Anwendung zur Erstellung realistischer generativer Modelle. Die Ergebnisse werden durch umfangreiche numerische Experimente, Visualisierungen und Simulationen veranschaulicht, wobei Finanzzeitreihen, aber auch Bilder als beispielhafte Anwendung dienen. Die Arbeit zeigt, wie die Kombination von stochastischer Steuerung und modernen Berechnungsmethoden genutzt werden kann, um komplexe zeitliche Dynamik präzise und flexibel zu modellieren.

Abstract

This thesis investigates generative models for time series from the perspective of the Schrödinger bridge problem, a classical problem in stochastic processes and optimal transport. A particular focus of the research is on the connection between theoretical foundations and the practical implementation of Schrödinger bridge methods, with the aim of generating time series that fulfill both specified statistical properties and boundary conditions. The research includes a thorough study of stochastic differential equations, the formulation of numerical methods for efficiently solving Schrödinger bridge problems, and their utilization in generating realistic generative models. The findings are substantiated by extensive numerical experiments, visualizations and simulations, with financial time series and images serving as illustrative examples. The study demonstrates the efficacy of integrating stochastic control and current computational methods in modelling complex temporal dynamics with precision and adaptability.

Inhaltsverzeichnis

Abbildungsverzeichnis	ii
Tabellenverzeichnis	iv
1 Einleitung	1
2 Grundlagen	2
2.1 Wahrscheinlichkeitstheorie	2
2.2 Stochastische Prozesse	4
2.3 Das Itô-Integral	12
2.4 Die Itô-Formel	14
2.5 Stochastische Differentialgleichungen	15
3 Zeitreihenanalyse	23
3.1 Definition von Zeitreihen	23
3.2 Modellierung von Zeitreihen	25
3.3 Modellwahl und Parameter-Schätzung	29
4 Schrödinger-Brücken-Problem	34
4.1 Historischer Hintergrund	34
4.2 Einleitung	35
4.3 Problemformulierung	37
4.4 Steuerungen und Kostenfunktion	40
4.5 Lösung des SBP	42
5 SBP für Zeitreihen	59
5.1 Motivation	59
5.2 Problemformulierung	59
5.3 Stochastisches Kontrollproblem	60
5.4 Lösung für SBTS	61
5.5 Generative Modelle	69
5.6 SBTS-Algorithmus	76
5.7 Numerische Experimente	77
5.8 Abgrenzung zu verwandten Methoden	87
6 Fazit und Ausblick	90
Literaturverzeichnis	91

Abbildungsverzeichnis

1	Dichtefunktion einer Zufallsvariable $X \sim \mathcal{N}(0, 1)$	4
2	Beispielhafte Pfade eines zeitdiskreten Random Walks.	6
3	Beispiele von Pfaden einer Brownschen Bewegung $\{B_t\}_{t \geq 0}$ in \mathbb{R} . Die Pfade sind fast sicher stetig, aber nirgendwo differenzierbar.	8
4	Beispiele von Pfaden einer Brownschen Brücke $\{B_t\}_{t \in [0, T]}$, jeweils mit $B_0 = 0$ und $B_T = 0$	11
5	Approximation der quadratischen Variation einer Brownschen Bewegung. Die Summe der quadratischen Inkremente nähert sich t an.	12
6	Realisationen eines AR(1)-Prozesses mit $\phi = 0,8$ und $\epsilon_t \sim \mathcal{N}(0, 1)$. Linke Grafik: Zeitreihe; Rechte Grafik: Autokorrelationsfunktion.	27
7	Volatilitätsdynamik und Autokorrelationsfunktion (ACF) der simulierten GARCH(1,1)-Renditen. Oben: Simulierte Renditen; Unten: Geschätzte bedingte Varianz σ_t^2 und ACF der simulierten Renditen.	28
8	Oben: ARMA(1,1)-Zeitreihe (simuliert und mit geschätzten Parametern). Unten: Residuen $\hat{\epsilon}_t$	33
9	Histogramm der simulierten Endwerte bei $T = 20$ des ungesteuerten Prozesses und die theoretische Dichte einer Normalverteilung $\mathcal{N}(20, 21)$	47
10	Simulationsergebnisse zur SDGL $dX_t = dt + dB_t$ mit Startwert $x_0 = 3$, $T = 20$ und die Endverteilung ist eine Standardnormalverteilung.	49
11	Simulationsergebnisse zur SDGL $dX_t = dt + dB_t$ mit Startwert $X_0 = 3$ und Endzeit $T = 5$	51
12	Simulationsergebnisse zur SDGL $dX_t = AX_t dt + dB_t$ mit Startwert $X_0 = 0$ und Endzeit $T = 2$	55
13	Simulationsergebnisse zur SDGL $dX_t = dt + dB_t$ bei normalverteilter Startverteilung $X_0 \sim \mathcal{N}(2, 1^2)$	57
14	Histogramm der Daten und KDE mit Bandbreiten $h = 0.2$ und $h = 0.5$	72
15	Schätzung der Regressionsfunktion mithilfe des Nadaraya-Watson-Schätzers mit zwei Bandbreiten $h = 0.3$ und $h = 0.7$	73
16	Vergleich von X_{t_3} zwischen echten Daten in pink und simulierten Schrödinger-Brücken in blau.	78
17	Gemeinsamer Verteilungsplot der Zufallsvariablen X_{t_1} , X_{t_2} und X_{t_3} für echte Daten und simulierte Schrödinger-Brücken. Die Diagonale zeigt die marginalen Verteilungen, während die oberen und unteren Dreiecke die bivariaten Dichtekonturen darstellen.	78
18	Empirische kumulative Verteilungsfunktionen (ECDF) für X_{t_1} , X_{t_2} und X_{t_3} bei echten Daten und simulierten Schrödinger-Brücken. Die obere Zeile zeigt die Ergebnisse für X_{t_1} und X_{t_2} , während die untere Zeile die Ergebnisse für X_{t_3} zusammenfasst.	79
19	Boxplots für X_{t_1} , X_{t_2} und X_{t_3} bei echten Daten und simulierten Schrödinger-Brücken. Die Boxplots zeigen die zentrale Tendenz und Streuung der Daten im Vergleich zu den simulierten Schrödinger-Brücken.	80
20	Vergleich von GARCH-Zeitreihen (links) und simulierten Schrödinger-Brücken (rechts) für vier zufällig ausgewählte Pfade.	81

21	Zufällig ausgewählte Pfade der Microsoft-Aktie (links) und simulierte Pfade (rechts).	82
22	Kovarianzmatrizen der Originaldaten und simulierten Daten	83
23	Verteilung der quadratischen Variation für Microsoft-Daten und Schrödinger-Brücken-Simulationen.	84
24	Verteilung der Renditen der Microsoft-Daten, simulierten Schrödinger-Brücken-Daten und einer Normalverteilung.	85
25	Vergleich zwischen zufällig ausgewählten Bildern aus dem MNIST-Fashion-Datensatz (links) und simulierten Bildern (rechts).	86
26	Vergleich zwischen zufällig ausgewählten Originalbildern (links) und simulierten Bildern (rechts) des CIFAR-10-Datensatzes.	86
27	Zeitliche Entwicklung des gesteuerten Prozesses bei der Generierung von RGB-Bildern. Die Abbildung zeigt den Übergang von einer uniformen Startverteilung zu den finalen simulierten Bildern.	87

Tabellenverzeichnis

1	Vergleich zwischen wahren und geschätzten Parametern im ARMA(1,1)-Modell; Maximum-Likelihood-Schätzung.	31
2	Ergebnisse der statistischen Tests zur Validierung des ARMA(1,1)-Modells.	32
3	Quantile und p-Werte der echten Daten und simulierten Schrödinger-Brücken.	80

1 Einleitung

Die stochastische Analysis stellt ein fundamentales und vielseitiges Teilgebiet der Mathematik dar, welches sich mit der Modellierung und Analyse zufälliger Prozesse befasst. Diese Prozesse zeigen sich in einer Vielzahl von Anwendungsbereichen, darunter Physik, Biologie, Finanzmärkte sowie Ingenieurwesen. Die Analyse solcher Prozesse ermöglicht nicht nur die Gewinnung tieferer Einblicke in die zugrunde liegenden Dynamiken, sondern auch die Entwicklung von Verfahren, die in unsicheren Umgebungen die Vorhersage von Ereignissen sowie die Durchführung von Optimierungen ermöglichen. Ein besonderer Fokus liegt dabei auf der Analyse von Zeitreihen, d. h. Daten, die in zeitlicher Abfolge vorliegen und sich dynamisch entwickeln können. Die Zeitreihenanalyse stellt eine komplexe Herausforderung dar, da sie die Entwicklung präziser Modelle erfordert, die zugleich flexibel genug sind, um komplexe Dynamiken und Unsicherheiten zu erfassen. Die gängigen Verfahren, wie autoregressive Modelle, GARCH-Modelle oder ARIMA-Modelle, erzielen in der Tat robuste Resultate. Allerdings sind diese durch Annahmen wie Linearität oder Stationarität teilweise begrenzt. In diesem Kontext eröffnet das Schrödinger-Brücken-Problem, ein Konzept aus der statistischen Mechanik, neue Perspektiven. Es verbindet die Theorie des optimalen Transports mit stochastischer Steuerung und ermöglicht die Interpolation zwischen Wahrscheinlichkeitsverteilungen unter dynamischen Randbedingungen. Diese Methode erweist sich insbesondere bei nicht-stationären Zeitreihen als nützlich, da sie Übergänge zwischen unterschiedlichen Verteilungen dynamisch abbilden kann.

Ein exemplarisches Anwendungsbeispiel der Schrödinger-Brücke findet sich in der Finanzwelt, wo Zeitreihen wie Aktienkurse häufig plötzlichen Schwankungen oder Veränderungen unterliegen. Herkömmliche Modelle können sich in diesem Kontext als unpassend erweisen, da sie auf stationäre Prozesse ausgelegt sind. Die Schrödinger-Brücke hingegen erlaubt eine stochastisch gesteuerte Interpolation zwischen verschiedenen Verteilungen, die unterschiedlichen Marktphasen entsprechen. Dynamische Modelle, welche plötzliche Veränderungen wie Marktcrashes und Phasen der Erholung realistisch abbilden, können auf diese Weise entwickelt werden.

Die vorliegende Arbeit untersucht, inwieweit die Schrödinger-Brücke eine geeignete Methode zur Modellierung und Generierung komplexer Zeitreihen darstellt. Der Schwerpunkt liegt dabei auf der Herleitung eines effizienten numerischen Verfahrens für die praktische Anwendung. Die Arbeit beinhaltet die Einführung grundlegender Konzepte der stochastischen Differentialgleichungen, die das Verständnis der vorgestellten Methoden unterstützen. Die numerischen Algorithmen zur Approximation der Schrödinger-Brücke werden in Python implementiert und die Ergebnisse durch aussagekräftige Visualisierungen präsentiert.

2 Grundlagen

Die folgenden Grundlagen bilden das Fundament für das Verständnis der stochastischen Analysis im Kontext der Schrödinger-Brücke. Insbesondere sind Begriffe wie *Wahrscheinlichkeitsraum*, *stochastische Prozesse* und *Martingale* für den formalen Aufbau von Zeitreihenmodellen und deren stochastischer Interpolation essenziell. In klassischen Zeitreihenmodellen wie ARIMA oder GARCH wird oft Stationarität vorausgesetzt. Die Schrödinger-Brücke ermöglicht hingegen eine *nicht-stationäre* Modellierung und nutzt steuerbare Diffusionsprozesse, deren Grundlagen im Folgenden dargestellt werden. Um Redundanz zu vermeiden, werden nur die für diese Arbeit zentralen Konzepte zusammenfassend vorgestellt. Weiterführende Details finden sich in [12] und [16].

2.1 Wahrscheinlichkeitstheorie

Die Wahrscheinlichkeitstheorie liefert das Fundament der stochastischen Analysis und damit auch der Schrödinger-Brücke. Dabei werden *Zufallsexperimente* formalisiert, um Wahrscheinlichkeiten zu quantifizieren. Im Folgenden bezeichnet das Tripel $(\Omega, \mathcal{F}, \mathbb{P})$ immer einen Wahrscheinlichkeitsraum.

Definition 2.1.1 (Zufallsvariable)

Eine **Zufallsvariable** ist eine messbare Abbildung $X : (\Omega, \mathcal{F}, \mathbb{P}) \rightarrow (\mathbb{R}, \mathcal{B}(\mathbb{R}))$. Hierbei ist $\mathcal{B}(\mathbb{R})$ die Borel- σ -Algebra auf \mathbb{R} .

Definition 2.1.2 (Verteilung und Dichtefunktion)

Die **Verteilung** einer Zufallsvariable X ist das Maß μ_X auf $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$, definiert durch

$$\mu_X(B) := \mathbb{P}(X^{-1}(B)).$$

Ist μ_X absolut stetig, so besitzt sie eine Dichtefunktion $f_X(x)$, und es gilt $\mu_X(B) = \int_B f_X(x) dx$.

Definition 2.1.3 (Erwartungswert, Momente, Varianz)

Sei $(\Omega, \mathcal{F}, \mathbb{P})$ ein Wahrscheinlichkeitsraum, und sei $X : \Omega \rightarrow \mathbb{R}$ eine reelle Zufallsvariable.

(i) X heißt **integrierbar**, wenn $X \in L^1(\mathbb{P})$, das heißt, wenn

$$\int_{\Omega} |X| d\mathbb{P} < \infty.$$

In diesem Fall ist der **Erwartungswert** von X definiert als

$$\mathbb{E}[X] := \int_{\Omega} X d\mathbb{P}.$$

Falls X entweder in seinem positiven oder negativen Teil integrierbar ist, schreiben wir ebenfalls $\mathbb{E}[X] := \int_{\Omega} X d\mathbb{P}$. Ist $\mathbb{E}[X] = 0$, so heißt X **zentriert**.

(ii) Für $n \in \mathbb{N}$ und $X \in L^n(\mathbb{P})$ sind die n -ten **Momente** und **absoluten Momente** definiert als

$$m_k := \mathbb{E}[X^k], \quad M_k := \mathbb{E}[|X|^k], \quad \text{für } k = 1, \dots, n.$$

(iii) X heißt **quadratintegrierbar**, wenn $X \in L^2(\mathbb{P})$, das heißt, wenn

$$\int_{\Omega} X^2 d\mathbb{P} < \infty.$$

In diesem Fall ist die **Varianz** von X definiert als

$$\text{Var}[X] := \mathbb{E}[X^2] - (\mathbb{E}[X])^2.$$

Die Zahl $\sigma := \sqrt{\text{Var}[X]}$ wird als **Standardabweichung** von X bezeichnet. Falls $\mathbb{E}[X^2] = \infty$, setzen wir formal $\text{Var}[X] = \infty$.

(iv) Für $X, Y \in L^2(\mathbb{P})$ ist die **Kovarianz** von X und Y definiert als

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

Die Kovarianz kann äquivalent geschrieben werden als

$$\text{Cov}[X, Y] = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y].$$

Die Zufallsvariablen X und Y heißen **unkorreliert**, wenn $\text{Cov}[X, Y] = 0$.

Beispiel 2.1.1 (Normalverteilte Zufallsvariable und ihre Eigenschaften.)

Betrachten wir eine normalverteilte Zufallsvariable $X \sim \mathcal{N}(\mu, \sigma^2)$ mit Erwartungswert $\mu \in \mathbb{R}$ und Varianz $\sigma^2 > 0$. Die Verteilung von X ist durch die Dichtefunktion

$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right), \quad x \in \mathbb{R},$$

gegeben und ist für $\mu = 0$, $\sigma = 1$ in Abbildung 1 zu sehen.

Die **Verteilung** von X ist ein Maß μ_X auf $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$, das durch

$$\mu_X(B) = \int_B f_X(x) dx, \quad \forall B \in \mathcal{B}(\mathbb{R}),$$

definiert ist. Der **Erwartungswert** von X ergibt sich durch Integration mit dieser Dichte:

$$\mathbb{E}[X] = \int_{\mathbb{R}} x f_X(x) dx = \mu.$$

Die Eigenschaft $\mathbb{E}[X] = \mu$ entspricht der Interpretation von μ als Schwerpunkt der Verteilung.

Die **Varianz** von X ist definiert als

$$\text{Var}[X] = \mathbb{E}[X^2] - (\mathbb{E}[X])^2.$$

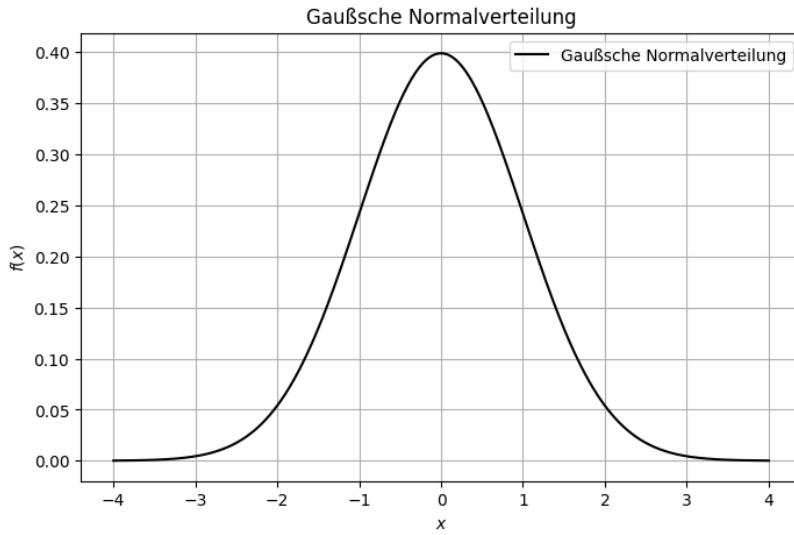


Abbildung 1: Dichtefunktion einer Zufallsvariable $X \sim \mathcal{N}(0, 1)$.

Für die Normalverteilung gilt

$$\mathbb{E}[X^2] = \mu^2 + \sigma^2, \quad \text{daher folgt } \text{Var}[X] = \sigma^2.$$

Normalverteilte Zufallsvariablen treten häufig in stochastischen Prozessen auf, insbesondere bei Schrödinger-Brücken-Problemen, da sie natürliche Modelle für Drift- und Diffusionsprozesse darstellen. Die hier erläuterten Eigenschaften, insbesondere der Erwartungswert und die Varianz, spielen eine zentrale Rolle bei der Modellierung und Analyse solcher Prozesse.

Die Jensen-Ungleichung ist ein zentrales Resultat der Wahrscheinlichkeitstheorie, das die Konvexität von Funktionen mit Erwartungswerten verknüpft.

Satz 2.1.1 (Jensen-Ungleichung für \mathbb{R}^n)

Sei X eine Zufallsvariable mit Werten in \mathbb{R}^n , definiert auf und sei $f : \mathbb{R}^n \rightarrow \mathbb{R}$ eine konvexe Funktion, für die $\mathbb{E}[|f(X)|] < \infty$ gilt. Weiterhin sei X integrierbar. Dann gilt:

$$f(\mathbb{E}[X]) \leq \mathbb{E}[f(X)].$$

2.2 Stochastische Prozesse

Stochastische Prozesse modellieren dynamische Systeme, bei denen der Zustand sich zufällig im Laufe der Zeit ändert. Viele Schlüsselideen – insbesondere in der Schrödinger-Brücke – basieren darauf, unterschiedliche Verteilungen zu verschiedenen Zeitpunkten zu verknüpfen. Im Folgenden wird größtenteils auf die Darstellung in REVUZ UND YOR [16] [16] zurückgegriffen, um die grundlegenden Konzepte und Eigenschaften stochastischer Prozesse zu entwickeln und soll lediglich als Überblick dienen.

Definition 2.2.1 (Stochastischer Prozess)

Ein \mathbb{R}^d -wertiger **stochastischer Prozess** $\{X_t\}_{t \in T}$ ist eine Familie von Zufallsvariablen $X_t : \Omega \rightarrow \mathbb{R}^d$, indexiert durch $t \in T$, wobei $T \subseteq \mathbb{R}_+$ typischerweise einen Zeitbereich darstellt.

Ein stochastischer Prozess $\{X_t\}_{t \in T}$ heißt **stetig**, wenn die Abbildung $t \mapsto X_t(\omega)$ für jedes $\omega \in \Omega$ stetig ist.

Die endlichen dimensionalen Verteilungen eines stochastischen Prozesses sind eine wichtige Eigenschaft, da sie die Wahrscheinlichkeitsstruktur eines Prozesses vollständig beschreiben. Insbesondere spielen sie eine zentrale Rolle bei der Analyse und Modellierung solcher Prozesse.

Definition 2.2.2 (Verteilung eines stochastischen Prozesses)

Die **endlichen dimensionalen Verteilungen** eines stochastischen Prozesses $\{X_t\}_{t \in T}$ sind die gemeinsamen Verteilungen der Zufallsvektoren $(X_{t_1}, X_{t_2}, \dots, X_{t_n})$ für beliebige $t_1, t_2, \dots, t_n \in T$. Diese Verteilungen charakterisieren den Prozess vollständig, da sie alle Wahrscheinlichkeitsstrukturen des Prozesses beschreiben.

Die Begriffe der endlichen dimensionalen Verteilungen und der Stationarität sind essenziell für die Untersuchung stochastischer Prozesse. Sie bilden eine Grundlage für die spätere Diskussion von speziellen Prozessen wie der Brownschen Bewegung oder Markov-Prozessen, die in dieser Arbeit von Bedeutung sind.

Definition 2.2.3 (Stationarität eines stochastischen Prozesses)

Ein stochastischer Prozess $\{X_t\}_{t \in T}$ heißt **stationär**, wenn seine endlichen dimensionalen Verteilungen invariant unter Zeitschiebungen sind. Das bedeutet, für alle $h \geq 0$ und alle $t_1, t_2, \dots, t_n \in T$ gilt

$$(X_{t_1}, X_{t_2}, \dots, X_{t_n}) \stackrel{d}{=} (X_{t_1+h}, X_{t_2+h}, \dots, X_{t_n+h}),$$

wobei $\stackrel{d}{=}$ die Gleichheit der Verteilungen bezeichnet. Stationarität impliziert, dass die statistischen Eigenschaften des Prozesses zeitlich unveränderlich sind.

Beispiel 2.2.1 (Zeitdiskreter Random Walk auf \mathbb{R})

Ein **Random Walk** auf \mathbb{R} ist ein stochastischer Prozess $\{S_n\}_{n \in \mathbb{N}_0}$, definiert durch

$$S_0 = 0, \quad S_n = \sum_{k=1}^n X_k, \quad n \in \mathbb{N},$$

wobei $\{X_k\}_{k \in \mathbb{N}}$ eine Familie unabhängiger, identisch verteilter Zufallsvariablen auf einem Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, \mathbb{P})$ ist.

Der Random Walk ist nicht stationär, da die Varianz von S_n mit n wächst und somit von der Zeit n abhängt.

Abbildung 2 zeigt fünf verschiedene Pfade eines zeitdiskreten Random Walks, wobei $X_k \sim \mathcal{N}(0, 1)$.

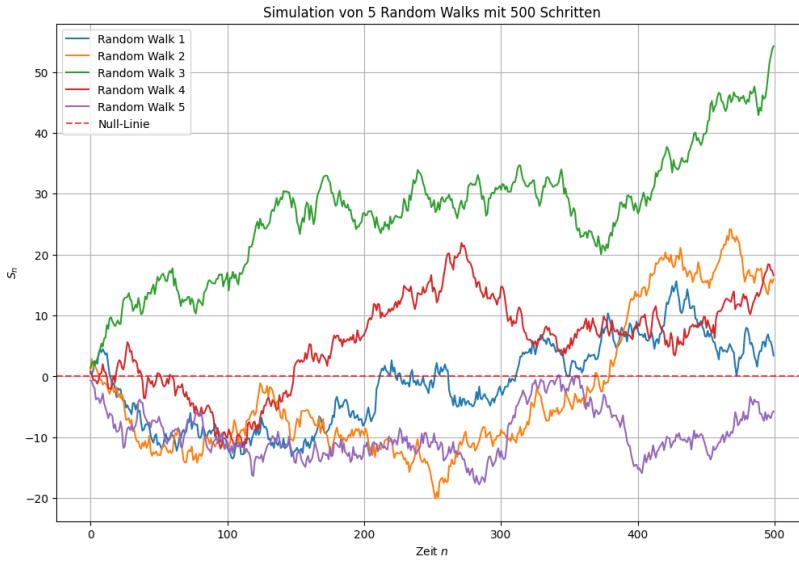


Abbildung 2: Beispielhafte Pfade eines zeitdiskreten Random Walks.

In der stochastischen Analysis spielen **Filtrationen** eine zentrale Rolle, da sie die zeitliche Entwicklung der verfügbaren Information modellieren. Eine Filtration beschreibt, welche Ereignisse zu einem bestimmten Zeitpunkt bekannt sind und wie sich diese Information im Laufe der Zeit erweitert.

Definition 2.2.4 (Adaptivität, Filtration)

Eine **Filtration** $\{\mathcal{F}_t\}_{t \geq 0}$ ist eine aufsteigende Familie von σ -Algebren. Ein Prozess $\{X_t\}$ heißt **adaptiert**, wenn X_t für jedes t \mathcal{F}_t -messbar ist.

In der stochastischen Analysis finden Stoppzeiten häufig Anwendung bei der formalen Beschreibung von Ereignissen, wie beispielsweise dem erstmaligen Überschreiten einer Grenze oder dem Verlassen eines bestimmten Bereichs.

Definition 2.2.5 (Stoppzeit)

Sei $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t \geq 0}, \mathbb{P})$ ein filtrierter Wahrscheinlichkeitsraum. Eine Abbildung $\tau : \Omega \rightarrow [0, \infty]$ heißt **Stoppzeit** relativ zur Filtration $\{\mathcal{F}_t\}$, wenn für jedes $t \geq 0$ gilt:

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

das heißt, das Ereignis, dass die Stoppzeit τ bis zur Zeit t eintritt, ist \mathcal{F}_t -messbar.

Beispiel 2.2.2 (Erstmaliges Überschreiten einer Schranke)

Betrachten wir eine Brownsche Bewegung $\{B_t\}_{t \geq 0}$ und definieren die Stoppzeit

$$\tau_a = \inf\{t \geq 0 : B_t \geq a\},$$

wobei $a > 0$. Dann ist τ_a eine Stoppzeit bezüglich der natürlichen Filtration von $\{B_t\}$.

Definition 2.2.6 (Martingal)

Ein Prozess $\{M_t\}_{t \geq 0}$ heißt **Martingal**, falls $\mathbb{E}[|M_t|] < \infty$, M_t adaptiert ist und

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

Ein Prozess $\{M_t\}_{t \geq 0}$ heißt **Submartingal**, falls gilt:

$$\mathbb{E}[M_t | \mathcal{F}_s] \geq M_s \quad \text{fast sicher.}$$

Entsprechend heißt $\{M_t\}_{t \geq 0}$ **Supermartingal**, wenn gilt:

$$\mathbb{E}[M_t | \mathcal{F}_s] \leq M_s \quad \text{fast sicher.}$$

Martingale eignen sich hervorragend, um dynamische Systeme zu beschreiben, bei denen zukünftige Entwicklungen ausschließlich auf der aktuell bekannten Information basieren. Sie sind unverzichtbar in der stochastischen Analysis und bilden die Grundlage für viele weiterführende Konzepte wie die Theorie der stochastischen Differentialgleichungen und die Bewertung von Finanzinstrumenten.

Das Optional Sampling Theorem spielt eine zentrale Rolle in der Martingaltheorie. Es trifft eine Aussage darüber, wie sich der Erwartungswert eines Martingals an einer zufälligen, von der Entwicklung des Prozesses abhängigen Stopzeit verhält. Unter geeigneten Regularitätsbedingungen bleibt der erwartete Wert des Martingals erhalten, selbst wenn man an zufälligen Zeitpunkten stoppen darf.

Satz 2.2.1 (Optional Sampling Theorem)

Sei $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t \geq 0}, \mathbb{P})$ ein filtrierter Wahrscheinlichkeitsraum und $\{X_t\}_{t \geq 0}$ ein $\{\mathcal{F}_t\}$ -adaptierter stochastischer Prozess. Sei τ eine $\{\mathcal{F}_t\}$ -Stopzeit.

Wenn $\{X_t\}_{t \geq 0}$ ein Martingal oder Submartingal ist, dann gilt unter einer der folgenden Bedingungen:

- (i) τ ist fast sicher beschränkt, d. h. es existiert ein konstantes $C > 0$ mit $\tau(\omega) \leq C$ für fast alle $\omega \in \Omega$.
- (ii) $\{X_t\}_{t \geq 0}$ ist ein gleichgradig integrierbares Martingal oder Submartingal, und es gilt $\mathbb{E}[\tau] < \infty$.
- (iii) Die Familie $\{X_\sigma : \sigma \text{ ist eine beschränkte Stopzeit}\}$ ist gleichgradig integrierbar, und wiederum $\mathbb{E}[\tau] < \infty$.

Dann gilt

$$\mathbb{E}[X_\tau] \geq \mathbb{E}[X_0] \quad \text{falls } \{X_t\} \text{ ein Submartingal ist,}$$

bzw.

$$\mathbb{E}[X_\tau] = \mathbb{E}[X_0] \quad \text{falls } \{X_t\} \text{ ein Martingal ist.}$$

Die **Brownsche Bewegung** $\{B_t\}$ ist eines der zentralen Beispiele für einen Markov-Prozess und spielt in der Schrödinger-Brücke eine Schlüsselrolle als Referenzprozess.

Definition 2.2.7 (Brownsche Bewegung / Wiener-Prozess)

Ein \mathbb{R}^d -wertiger Prozess $\{B_t\}_{t \geq 0}$ heißt **Brownsche Bewegung** (oder **Wiener-Prozess**), wenn:

- (i) $B_0 = 0$ fast sicher,
- (ii) $B_{t_i} - B_{t_{i-1}}$ sind unabhängig für disjunkte Intervalle,
- (iii) $B_t - B_s \sim \mathcal{N}(0, (t-s)I_d)$ für $0 \leq s < t$,
- (iv) $t \mapsto B_t(\omega)$ ist fast sicher stetig.

Die Brownsche Bewegung ist durch ihre zufälligen, aber stetigen Pfade gekennzeichnet. Abbildung 3 zeigt beispielhaft drei Pfade einer Brownschen Bewegung $\{B_t\}_{t \geq 0}$ in \mathbb{R} .

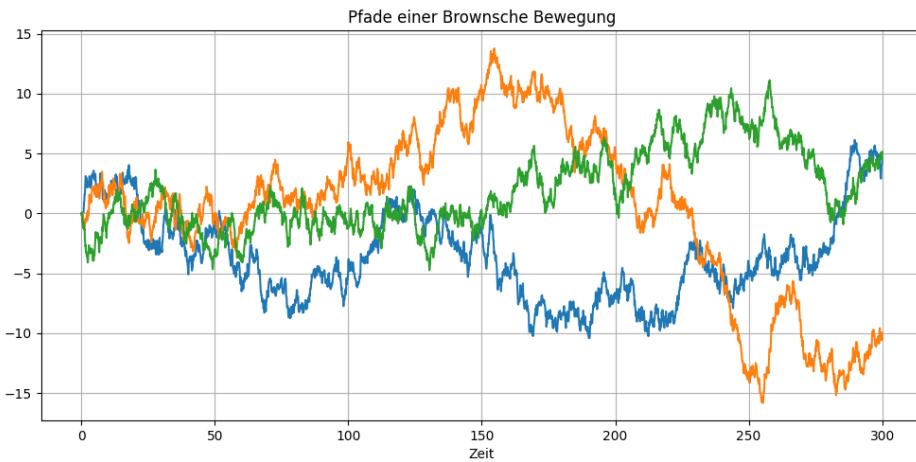


Abbildung 3: Beispiele von Pfaden einer Brownschen Bewegung $\{B_t\}_{t \geq 0}$ in \mathbb{R} . Die Pfade sind fast sicher stetig, aber nirgendwo differenzierbar.

Markov-Eigenschaft. Die **Markov-Eigenschaft** eines Prozesses $\{X_t\}$ besagt, dass die Zukunft nur vom aktuellen Zustand, nicht aber von der Vorgeschichte abhängt:

$$\mathbb{P}(X_{t+h} \in B \mid \mathcal{F}_t) = \mathbb{P}(X_{t+h} \in B \mid X_t),$$

wobei \mathcal{F}_t die bis zur Zeit t verfügbare Information bezeichnet.

Definition 2.2.8 ($\{\mathcal{F}_t\}$ -Brownsche Bewegung)

Ein \mathbb{R}^d -wertiger stochastischer Prozess $\{B_t\}_{t \geq 0}$ heißt **$\{\mathcal{F}_t\}$ -Brownsche Bewegung**, falls

- (i) $\{B_t\}$ eine (Standard-)Brownsche Bewegung ist,
- (ii) $\{B_t\}$ an die Filtration $\{\mathcal{F}_t\}$ adaptiert ist,
- (iii) für $0 \leq s < t$ gilt: $B_t - B_s$ ist unabhängig von \mathcal{F}_s .

Diese Definition stärkt die Unabhängigkeit der Inkremente, indem sie eine Anpassung an die Filtration $\{\mathcal{F}_t\}$ sicherstellt. Im Folgenden wird nicht explizit auf die Adaptiertheit eingegangen, sofern klar ist, dass auf einem filtrierten Wahrscheinlichkeitsraum gearbeitet wird.

Wird die Brownsche Bewegung $\{B_t\}$ wie in Definition 2.2.8 an eine Filtration $\{\mathcal{F}_t\}$ angepasst, so erfüllt sie automatisch die **Markov-Eigenschaft**. Für einen Zeitpunkt $t \geq 0$ und eine künftige Zeit $s > 0$ hängt B_{t+s} nur vom aktuellen Zustand B_t ab und nicht von der gesamten Vergangenheit $\{B_u : u \leq t\}$. Dies folgt aus der Unabhängigkeit der Zuwächse ($B_{t+s} - B_t$ ist unabhängig von \mathcal{F}_t) und daraus, dass die Verteilung der Zuwächse allein durch die Zeitdifferenz s bestimmt wird:

$$B_{t+s} - B_t \sim \mathcal{N}(0, s I_d).$$

Somit hängt die zukünftige Entwicklung nur vom gegenwärtigen Zustand B_t ab.

Die $\{\mathcal{F}_t\}$ -Brownsche Bewegung ist außerdem ein **Martingal** bezüglich der natürlichen Filtration (oder jeder größeren geeigneten Filtration). Sei $\mathcal{F}_t := \sigma(\{B_u : 0 \leq u \leq t\})$ die natürliche Filtration. Für $0 \leq s \leq t$ gilt:

$$\mathbb{E}[B_t \mid \mathcal{F}_s] = \mathbb{E}[B_s + (B_t - B_s) \mid \mathcal{F}_s].$$

Da $(B_t - B_s)$ unabhängig von \mathcal{F}_s ist und einen Erwartungswert von 0 besitzt, folgt

$$\mathbb{E}[B_t \mid \mathcal{F}_s] = B_s.$$

Somit ist $\{B_t\}$ ein Martingal.

Fast alle Pfade der Brownschen Bewegung $\{B_t\}$ sind **überall nicht-differenzierbar**. Darüber hinaus ist sie **fast sicher unbeschränkt**, das heißt, für jedes $M > 0$ existiert (mit Wahrscheinlichkeit 1) eine Zeit $t > 0$, so dass $\|B_t\| > M$. Diese Eigenschaften verdeutlichen, weshalb klassische Konzepte wie *Deterministische Ableitung* hier nicht greifen und zugleich, warum Maßänderungen (etwa via Girsanov) so mächtig sind.

Die *Schrödinger-Brücke* lässt sich als Steuerungsproblem für Markov-Prozesse auffassen, bei dem zwei Randverteilungen (Anfangs- und Endverteilung) optimal verbunden werden sollen. Konkret wird ein *Diffusionsprozess* gesucht, der zur Zeit 0 eine vorgegebene Startverteilung besitzt und zur Zeit T eine bestimmte Zielverteilung einnimmt. Der Satz von Girsanov 2.5.2 ermöglicht dabei, eine *standardisierte* Brownsche Bewegung durch Hinzufügen eines Driftterms in einen *driftbehafteten* Diffusionsprozess zu verwandeln, der genau diese vorgegebenen Randverteilungen erfüllt. Dies bildet den Kern des Schrödinger-Brücken-Problems.

Satz 2.2.2 (Übergangsfunktion)

Sei $\{X_t\}_{t \geq 0}$ ein Markov-Prozess. Dann existiert eine Funktion $p(t, x, B)$, die für jeden Startzustand $x \in \mathbb{R}^d$, jede Zeit $t \geq 0$ und jede messbare Menge $B \in \mathcal{B}(\mathbb{R}^d)$ definiert ist durch

$$p(t, x, B) = \mathbb{P}(X_t \in B \mid X_0 = x).$$

Falls $p(t, x, B)$ eine Dichte $p(t, x, y)$ bezüglich des Lebesgue-Maßes besitzt, gilt:

$$\mathbb{P}(X_t \in B \mid X_0 = x) = \int_B p(t, x, y) dy.$$

Um in einem Schrödinger-Brücken-Setting den Zeitverlauf der Verteilung von $\{X_t\}$ zu steuern, spielen **Übergangsfunktionen** wie $p(t, x, B)$ eine zentrale Rolle. Sie erlauben, die Verteilung zu jedem Zeitpunkt t exakt zu beschreiben, sodass Anfangs- und Endverteilungen (bei $t = 0$ bzw. $t = T$) gezielt erreicht werden können.

Definition 2.2.9 (Homogener Markov-Prozess)

Ein Markov-Prozess $\{X_t\}$ heißt **homogen**, wenn seine Übergangsfunktion $p(t, x, B)$ nur von der Zeitdifferenz $t - s$ abhängt. Es existiert also eine Funktion $p_h(x, B)$ derart, dass

$$p(t, x, B) = p_h(x, B), \quad \text{wobei } h = t - s.$$

Satz 2.2.3 (Stationarität der Übergangsfunktion)

Sei $\{X_t\}$ ein homogener Markov-Prozess. Dann hängt die Übergangsfunktion $p(t, x, B)$ nur von der Zeitdifferenz $t - s$ ab. Es existiert also $p_h(x, B)$ mit

$$p(t, x, B) = p_h(x, B) \quad \text{für alle } t \geq s.$$

Falls eine Übergangsdichte $p_h(x, y)$ existiert, gilt:

$$p(t, x, B) = \int_B p_h(x, y) dy.$$

Satz 2.2.4 (Zeitentwicklung der Verteilung)

Sei $\{X_t\}$ ein Markov-Prozess mit Übergangsfunktion $p(t, x, B)$ und Startverteilung $\mu(dx)$. Dann ist die Verteilung des Prozesses zur Zeit t gegeben durch:

$$\mathbb{P}(X_t \in B) = \int_{\mathbb{R}^d} p(t, x, B) \mu(dx), \quad B \in \mathcal{B}(\mathbb{R}^d).$$

Falls zudem eine Übergangsdichte $p(t, x, y)$ existiert, folgt

$$\mathbb{P}(X_t \in B) = \int_B \int_{\mathbb{R}^d} p(t, x, y) \mu(dx) dy.$$

Beispiel 2.2.3 (Brownsche Bewegung als Markov-Prozess)

Die **Brownsche Bewegung** $\{B_t\}$ erfüllt die Markov-Eigenschaft durch ihre unabhängigen Zuwächse. Für $t \geq 0$, $h > 0$ und ein Ereignis $B \in \mathcal{B}(\mathbb{R}^d)$ gilt:

$$\mathbb{P}(B_{t+h} \in B \mid \mathcal{F}_t) = \mathbb{P}(B_{t+h} \in B \mid B_t).$$

Die Übergangsdichte ist

$$p(t, x, y) = \frac{1}{(2\pi t)^{d/2}} \exp\left(-\frac{\|y-x\|^2}{2t}\right), \quad t > 0.$$

Dies beschreibt die Wahrscheinlichkeit, nach Zeit t von x nach y zu gehen.

Da diese Dichte nur von $t - s$ abhängt, ist die Brownsche Bewegung homogen. Startet sie mit einer Verteilung μ , ergibt sich die Verteilung zur Zeit t als

$$\mathbb{P}(B_t \in B) = \int_{\mathbb{R}^d} \int_B p(t, x, y) \mu(dx) dy.$$

Definition 2.2.10 (Brownsche Brücke)

Eine **Brownsche Brücke** $\{B_t^{\text{BB}}\}_{t \in [0, T]}$ ist ein stochastischer Prozess, der aus einer Standard-Brownschen Bewegung $\{W_t\}$ konstruiert wird und die Randbedingungen

$$B_0^{\text{BB}} = 0 \quad \text{und} \quad B_T^{\text{BB}} = 0 \quad \text{fast sicher}$$

erfüllt. Konkret setzt man

$$B_t^{\text{BB}} = W_t - \frac{t}{T} W_T, \quad t \in [0, T].$$

Im Gegensatz zur "freien" Brownschen Bewegung, die keine Randbedingungen besitzt, muss die Brownsche Brücke bei $t = 0$ und $t = T$ jeweils 0 sein. Abbildung 4 illustriert einige Pfade dieses Prozesses.

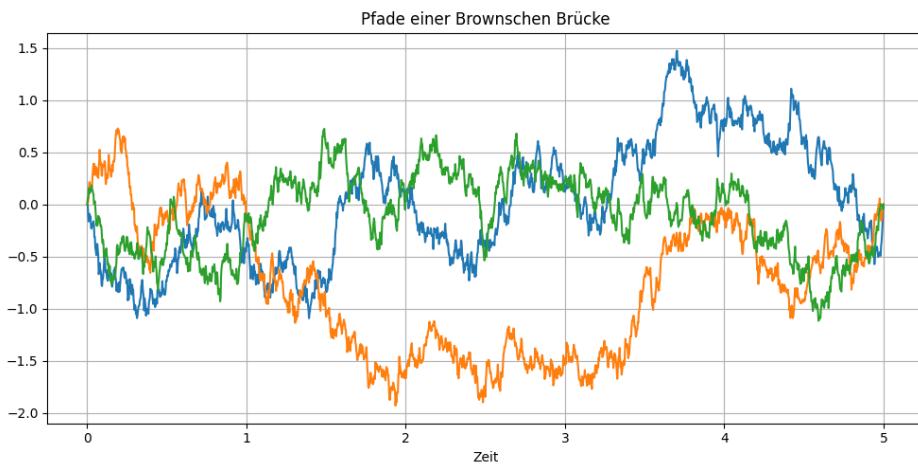


Abbildung 4: Beispiele von Pfaden einer Brownschen Brücke $\{B_t\}_{t \in [0, T]}$, jeweils mit $B_0 = 0$ und $B_T = 0$.

Ein wichtiges Konzept in der stochastischen Analysis ist die **quadratische Variation**, die insbesondere für die Brownsche Bewegung eine zentrale Rolle spielt.

Definition 2.2.11 (Quadratische Variation)

Sei $\{X_t\}_{t \geq 0}$ ein stetiger Prozess. Die **quadratische Variation** $\{\langle X, X \rangle_t\}_{t \geq 0}$ ist definiert durch

$$\langle X, X \rangle_t := \lim_{\|\Pi\| \rightarrow 0} \sum_{i=1}^n (X_{t_i} - X_{t_{i-1}})^2,$$

wobei $\Pi = \{0 = t_0 < \dots < t_n = t\}$ eine Zerlegung von $[0, t]$ ist.

Satz 2.2.5 (Quadratische Variation der Brownschen Bewegung)

Sei $\{B_t\}$ eine Brownsche Bewegung. Dann gilt

$$\langle B, B \rangle_t = t \quad \text{fast sicher für alle } t \geq 0.$$

Beispiel 2.2.4 (Numerische Illustration der quadratischen Variation)

In Abbildung 5 wird die quadratische Variation einer Brownschen Bewegung für verschiedene Zerlegungen approximiert. Die Summe der quadratischen Inkremente konvergiert gegen die Zeit t .

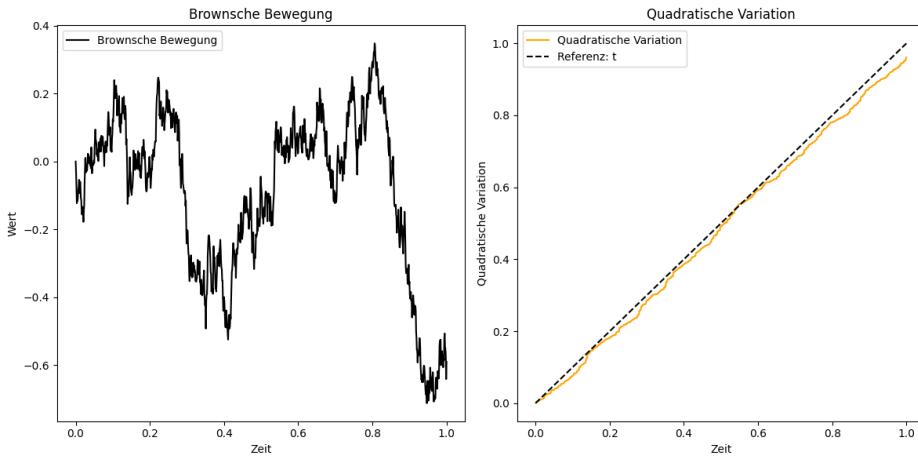


Abbildung 5: Approximation der quadratischen Variation einer Brownschen Bewegung. Die Summe der quadratischen Inkremente nähert sich t an.

2.3 Das Itô-Integral

Ein zentraler Baustein der stochastischen Analysis ist das Itô-Integral, welches die Integration über stochastische Prozesse wie die Brownsche Bewegung erlaubt. Im Gegensatz zu klassischen Integralen, die auf der Differenzierbarkeit oder der Riemann-Summen-Konstruktion basieren, verwendet das Itô-Integral eine spezielle Definition, die die Eigenschaften der Brownschen Bewegung und ihrer quadratischen Variation berücksichtigt.

Definition 2.3.1 (Itô-Integral)

Sei $\{B_t\}_{t \geq 0}$ eine Brownsche Bewegung auf einem filtrierten Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t \geq 0}, \mathbb{P})$. Ein stochastischer Prozess $\{H_t\}_{t \geq 0}$ heißt **Itô-integrierbar** auf dem Intervall $[0, T]$, wenn er $\{\mathcal{F}_t\}$ -adaptiert und messbar ist und die Bedingung

$$\mathbb{E} \left[\int_0^T H_t^2 dt \right] < \infty$$

erfüllt.

Das **Itô-Integral** von H_t bezüglich $\{B_t\}$ ist definiert als der $L^2(\Omega)$ -Grenzwert der Summen

$$\int_0^T H_t dB_t := \lim_{|\Pi| \rightarrow 0} \sum_{i=0}^{n-1} H_{t_i} (W_{t_{i+1}} - W_{t_i}),$$

wobei $\Pi = \{0 = t_0 < t_1 < \dots < t_n = T\}$ eine Zerlegung des Intervalls $[0, T]$ ist und die Grenzwertbildung über die Feinheit der Zerlegung $|\Pi| = \max_i(t_{i+1} - t_i)$ erfolgt.

Die Existenz dieses Grenzwerts wird durch die Isometrie-Eigenschaft des Itô-Integrals und die L^2 -Integrierbarkeit von H_t garantiert.

Das Itô-Integral erfüllt eine Reihe fundamentaler Eigenschaften, die es von klassischen Integralen unterscheiden:

1. **Linearität:** Für Itô-integrierbare Prozesse $\{H_t\}$ und $\{G_t\}$ sowie Skalare $\alpha, \beta \in \mathbb{R}$ gilt:

$$\int_0^T (\alpha H_t + \beta G_t) dB_t = \alpha \int_0^T H_t dB_t + \beta \int_0^T G_t dB_t.$$

2. **Isometrie (Itô-Isometrie):** Für jeden Itô-integrierbaren Prozess $\{H_t\}$ gilt:

$$\mathbb{E} \left[\left(\int_0^T H_t dB_t \right)^2 \right] = \mathbb{E} \left[\int_0^T H_t^2 dt \right].$$

3. **Martingaleigenschaft:** Das Itô-Integral $\left\{ \int_0^t H_s dW_s \right\}_{t \geq 0}$ ist ein $\{\mathcal{F}_t\}$ -Martingal.

4. **Stetigkeit:** Das Itô-Integral ist ein stetiger Prozess in t mit fast sicher stetigen Pfaden, d. h.,

$$t \mapsto \int_0^t H_s dW_s \quad \text{ist fast sicher stetig für alle } t \in [0, T].$$

Eine wichtige Eigenschaft des Itô-Integrals ist seine Verbindung zur quadratischen Variation. Sei $\{H_t\}_{t \geq 0}$ ein Itô-integrierbarer Prozess. Dann besitzt der Prozess $\left\{ \int_0^t H_s dW_s \right\}_{t \geq 0}$ die quadratische Variation

$$\left\langle \int_0^\cdot H_s dW_s \right\rangle_t = \int_0^t H_s^2 ds.$$

Dies folgt direkt aus der Itô-Formel und der Eigenschaft der Brownschen Bewegung, dass $\langle W \rangle_t = t$.

Beispiel 2.3.1 (Itô-Integral einer konstanten Funktion)

Betrachten wir den Prozess $H_t = c$, wobei $c \in \mathbb{R}$ eine Konstante ist. Da H_t $\{\mathcal{F}_t\}$ -adaptiert und quadratintegrierbar ist, ist er Itô-integrierbar. Das Itô-Integral ist in diesem Fall gegeben durch:

$$\int_0^T c dB_t = cB_T.$$

Dies illustriert, dass das Itô-Integral für konstante Prozesse einfach der skalierte Wert der Brownschen Bewegung zum Zeitpunkt T ist.

Obwohl hier das Itô-Integral bezüglich der Brownschen Bewegung definiert wurde, kann das Konzept auf allgemeine Semimartingale erweitert werden. Sei X_t ein Semimartingal, dann kann das Itô-Integral $\int_0^T H_t dX_t$ unter geeigneten Bedingungen ähnlich konstruiert werden.

2.4 Die Itô-Formel

Die Itô Formel erweitert die klassische Kettenregel auf stochastische Prozesse und berücksichtigt sowohl deterministische als auch stochastische Dynamiken.

Satz 2.4.1 (Itô-Formel für zeit- und ortsabhängige Funktionen)

Sei $\{X_t\}_{t \geq 0}$ ein n -dimensionaler Itô-Prozess, der die SDGL

$$dX_t = \mu(X_t, t) dt + \sigma(X_t, t) dB_t$$

erfüllt, wobei $\mu: [0, T] \times \mathbb{R}^n \rightarrow \mathbb{R}^n$ und $\sigma: [0, T] \times \mathbb{R}^n \rightarrow \mathbb{R}^{n \times d}$ hinreichend regulär sind, und $\{B_t\}$ eine d -dimensionale Brownsche Bewegung ist. Sei ferner

$$f: [0, T] \times \mathbb{R}^n \rightarrow \mathbb{R}$$

eine Funktion aus der Klasse $C^{1,2}([0, T] \times \mathbb{R}^n)$. Dann gilt für den Prozess $Y_t := f(t, X_t)$ die folgende Differentialgleichung:

$$\begin{aligned} df(t, X_t) &= \frac{\partial f}{\partial t}(t, X_t) dt + \nabla_x f(t, X_t)^\top \mu(X_t, t) dt \\ &\quad + \frac{1}{2} \text{tr}\left(\sigma(X_t, t)^\top D_x^2 f(t, X_t) \sigma(X_t, t)\right) dt + \nabla_x f(t, X_t)^\top \sigma(X_t, t) dB_t. \end{aligned}$$

wobei $\nabla_x f$ der Gradient und $D_x^2 f$ die Hesse-Matrix von f in den Raumvariablen bezeichnet.

Die Itô Formel beschreibt, wie sich $f(t, X_t)$, eine Funktion von Zeit t und einem n -dimensionalen stochastischen Prozess X_t , infinitesimal ändert. Sie setzt sich aus einem deterministischen Term der die zeitliche Änderung, die Drift und den Einfluss der Varianz beschreibt und einem stochastischen Term der die zufällige Komponente der Dynamik erfasst.

Beispiel 2.4.1 (Brownsche Bewegung in \mathbb{R}^d)

Sei $\{B_t\}_{t \geq 0}$ eine \mathbb{R}^d -wertige Brownsche Bewegung, also ein Prozess mit unabhängigen Komponenten $\{B_t^i\}_{t \geq 0}$, $i = 1, \dots, d$, und $\langle W^i, W^j \rangle_t = \delta_{ij} t$. Für $f \in C^2(\mathbb{R}^d)$ gilt:

$$f(B_t) = f(W_0) + \sum_{i=1}^d \int_0^t \frac{\partial f}{\partial x_i}(W_s) dW_s^i + \frac{1}{2} \sum_{i=1}^d \int_0^t \frac{\partial^2 f}{\partial x_i^2}(W_s) ds,$$

da die Brownsche Bewegung keine quadratische Kovariation hat ($\langle W^i, W^j \rangle_t = 0$ für $i \neq j$).

Beispiel 2.4.2

Betrachte die eindimensionale stochastische Differentialgleichung

$$dX_t = \mu dt + \sigma dB_t,$$

wobei $\mu, \sigma \in \mathbb{R}$ Konstanten und $\{B_t\}$ eine Brownsche Bewegung sind.
Sei ferner $f: \mathbb{R} \rightarrow \mathbb{R}$ durch $f(x) = x^2$ gegeben. Dann gelten

$$f'(x) = 2x \quad \text{und} \quad f''(x) = 2.$$

Wendet man die Itô-Formel 2.4.1 auf $f(X_t)$ an, so ergibt sich

$$X_t^2 = X_0^2 + \int_0^t 2X_s \mu \, ds + \int_0^t 2X_s \sigma \, dB_s + \int_0^t \sigma^2 \, ds.$$

2.5 Stochastische Differentialgleichungen

Stochastische Differentialgleichungen (SDGL) beschreiben die Dynamik stochastischer Prozesse, die sowohl deterministischen als auch zufälligen Einflüssen unterliegen. Sie spielen in zahlreichen Anwendungen – insbesondere in der Physik, Biologie und Finanzmathematik eine zentrale Rolle (vgl. [22]).

Die Motivation für stochastische Differentialgleichungen ergibt sich aus der Notwendigkeit, dynamische Systeme zu modellieren, die durch Zufallseinflüsse gestört werden. Klassische (deterministische) Differentialgleichungen beschreiben Systeme, in denen der zukünftige Zustand ausschließlich vom aktuellen Zustand abhängt. In vielen realen Anwendungen ist es jedoch erforderlich, zufällige Störungen explizit zu berücksichtigen. SDGLs erweitern daher die klassischen Differentialgleichungen, indem sie *stochastische Prozesse* – insbesondere die Brownsche Bewegung – als treibende Kräfte einbeziehen.

Eine typische SDGL in \mathbb{R}^n hat die Form

$$dX_t = \mu(X_t, t) dt + \sigma(X_t, t) dB_t,$$

wobei

$$\mu: \mathbb{R}^n \times [0, T] \rightarrow \mathbb{R}^n$$

die **Driftfunktion** ist, welche die deterministische Tendenz des Systems beschreibt, und

$$\sigma: \mathbb{R}^n \times [0, T] \rightarrow \mathbb{R}^{n \times d}$$

die **Diffusionsmatrix**, welche die Intensität der zufälligen Einflüsse angibt. Schließlich ist $\{B_t\}_{t \geq 0}$ eine d -dimensionale Brownsche Bewegung.

Bevor Lösungen von SDGLs untersucht werden, ist zu präzisieren, was unter einer Lösung verstanden wird.

Definition 2.5.1 (Lösung einer SDGL)

Sei $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t \geq 0}, \mathbb{P})$ ein filtrierter Wahrscheinlichkeitsraum und $\{B_t\}_{t \geq 0}$ eine d -dimensionale $\{\mathcal{F}_t\}$ -Brownsche Bewegung. Eine \mathbb{R}^n -wertige, $\{\mathcal{F}_t\}$ -adaptierte, stetige Prozessfamilie $\{X_t\}_{t \geq 0}$ heißt eine (**starke**) Lösung der stochastischen Differentialgleichung

$$dX_t = \mu(X_t, t) dt + \sigma(X_t, t) dB_t, \quad X_0 = x_0,$$

wenn für alle $t \in [0, T]$ fast sicher gilt:

$$X_t = x_0 + \int_0^t \mu(X_s, s) ds + \int_0^t \sigma(X_s, s) dW_s.$$

Bemerkung 2.5.1

Die **starke Lösung** bezieht sich auf eine Lösung bezüglich eines vorgegebenen Wahrscheinlichkeitsraums und einer festen Brownschen Bewegung. Im Gegensatz dazu wird von einer **schwachen Lösung** gesprochen, wenn lediglich gefordert ist, dass irgendein Wahrscheinlichkeitsraum gefunden wird, auf dem eine geeignete Brown'sche Bewegung und ein Prozess $\{X_t\}$ existieren, welche die SDGL erfüllen.

Die Frage nach Existenz und Eindeutigkeit von Lösungen einer SDGL ist zentral in der stochastischen Analysis. Unter bestimmten Bedingungen an die Koeffizientenfunktionen μ und σ lässt sich zeigen, dass eine eindeutige starke Lösung existiert.

Satz 2.5.1 (Existenz und Eindeutigkeit der starken Lösung)

Seien

$$\mu: \mathbb{R}^n \times [0, T] \rightarrow \mathbb{R}^n \quad \text{und} \quad \sigma: \mathbb{R}^n \times [0, T] \rightarrow \mathbb{R}^{n \times d}$$

messbare Funktionen. Angenommen, es existieren Konstanten $K > 0$ und $L > 0$, so dass für alle $x, y \in \mathbb{R}^n$ und alle $t \in [0, T]$ gilt:

(i) (Lipschitz-Bedingung)

$$\|\mu(x, t) - \mu(y, t)\| + \|\sigma(x, t) - \sigma(y, t)\| \leq L \|x - y\|.$$

(ii) (Lineares Wachstum)

$$\|\mu(x, t)\|^2 + \|\sigma(x, t)\|^2 \leq K(1 + \|x\|^2).$$

Dann existiert eine eindeutige $\{\mathcal{F}_t\}$ -adaptierte, stetige starke Lösung $\{X_t\}_{t \geq 0}$ der obigen SDGL.

Mehrere spezielle SDGLs haben in der Praxis besondere Bedeutung.

Beispiel 2.5.1 (Geometrische Brownsche Bewegung (GBM))

Die Geometrische Brownsche Bewegung ist in der Finanzmathematik von zentraler Bedeutung, insbesondere im Black–Scholes-Modell. Sie wird durch die stochastische Differentialgleichung

$$dS_t = \mu S_t dt + \sigma S_t dB_t \quad (S_0 > 0, \mu, \sigma \in \mathbb{R})$$

beschrieben. Zur expliziten Lösung betrachtet man üblicherweise den Logarithmus $\ln S_t$ und wendet die Itô-Formel an.

Setze

$$Y_t := \ln(S_t).$$

Dann gilt nach Itô-Formel

$$dY_t = d(\ln(S_t)) = \frac{1}{S_t} dS_t - \frac{1}{2} \frac{1}{S_t^2} \sigma^2 S_t^2 dt.$$

Eingesetzt in die SDE von S_t führt das zu:

$$dY_t = \frac{1}{S_t} \left(\mu S_t dt + \sigma S_t dB_t \right) - \frac{1}{2} \sigma^2 dt = \left(\mu - \frac{\sigma^2}{2} \right) dt + \sigma dB_t.$$

Somit ist Y_t eine Brownsche Bewegung mit Drift $\mu - \frac{\sigma^2}{2}$ und Diffusion σ .

Da

$$Y_t = Y_0 + \int_0^t \left(\mu - \frac{\sigma^2}{2} \right) ds + \int_0^t \sigma dB_s,$$

und $Y_0 = \ln(S_0)$, folgt:

$$Y_t = \ln(S_0) + \left(\mu - \frac{\sigma^2}{2} \right) t + \sigma B_t.$$

Somit gilt

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

Hieraus erkennt man, dass S_t lognormalverteilt ist, mit $\mathbb{E}[S_t] = S_0 e^{\mu t}$ und $\text{Var}(S_t) = S_0^2 e^{2\mu t} (e^{\sigma^2 t} - 1)$.

Beispiel 2.5.2 (Ornstein–Uhlenbeck-Prozess)

Der Ornstein–Uhlenbeck-Prozess (OU-Prozess) ist ein stochastischer Prozess, der eine "Rückstellkraft" in Richtung eines Gleichgewichtszustandes besitzt. Diese Eigenschaft führt zu einem charakteristischen exponentiellen Abklingen von Abweichungen vom Ruhpunkt. Der OU-Prozess wird häufig in der Physik (z. B. zur Modellierung von Geschwindigkeiten bei Teilchenbewegungen), in der Finanzmathematik (als "Mean-Reverting"-Modelle) und in anderen Bereichen eingesetzt.

Der OU-Prozess $\{X_t\}$ erfüllt die stochastische Differentialgleichung

$$dX_t = -AX_t dt + \sigma dB_t, \quad A > 0, \sigma > 0.$$

Hierbei ist $A > 0$ die Rückstell- oder Mean-Reversion-Rate, $\sigma > 0$ die "Rausch-" oder Diffusionsstärke und B_t eine Brownsche Bewegung.

Es gilt

$$e^{\theta t} dX_t = -\theta e^{\theta t} X_t dt + \sigma e^{\theta t} dB_t.$$

Andererseits gilt nach Itô-Formel:

$$d(e^{\theta t} X_t) = e^{\theta t} dX_t + \theta e^{\theta t} X_t dt.$$

In die erste Gleichung eingesetzt, vereinfacht sich das $-\theta e^{\theta t} X_t dt$ mit dem $\theta e^{\theta t} X_t dt$ zu einem reinen Differentialausdruck:

$$d(e^{\theta t} X_t) = \sigma e^{\theta t} dB_t.$$

Integration über $[0, t]$:

$$e^{\theta t} X_t - e^{\theta \cdot 0} X_0 = \sigma \int_0^t e^{\theta s} dB_s.$$

Schließlich erhält man

$$X_t = e^{-\theta t} X_0 + \sigma e^{-\theta t} \int_0^t e^{\theta s} dB_s$$

und das wird mit Ausklammern zu

$$X_t = e^{-\theta t} X_0 + \sigma \int_0^t e^{-\theta(t-s)} dB_s.$$

Also ist die Lösung der Ornstein–Uhlenbeck-SDGL

$$X_t = x_0 e^{-\theta t} + \sigma \int_0^t e^{-\theta(t-s)} dB_s,$$

wobei x_0 den Startwert bezeichnet.

Das Girsanov-Theorem ist ein fundamentaler Satz der stochastischen Analysis und beschreibt die Änderung des Wahrscheinlichkeitsmaßes, unter der eine Brownsche Bewegung eine Drift erhält. Dieser Satz ist insbesondere für die Theorie stochastischer Differentialgleichungen zentral und findet Anwendung in Gebieten wie der Finanzmathematik oder in Schrödinger-Brücken-Problemen.

Satz 2.5.2 (Satz von Girsanov)

Sei $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t \geq 0}, \mathbb{P})$ ein filtrierter Wahrscheinlichkeitsraum, auf dem $\{B_t\}_{t \geq 0}$ eine $\{\mathcal{F}_t\}$ -Brownsche Bewegung unter \mathbb{P} ist. Sei $\{\theta_t\}_{t \geq 0}$ ein $\{\mathcal{F}_t\}$ -adaptierter Prozess mit

$$\int_0^T \theta_t^2 dt < \infty \quad \text{fast sicher, für jedes } T > 0.$$

Definiere das stochastische Exponential Λ_T durch

$$\Lambda_T = \exp\left(-\int_0^T \theta_t dB_t - \frac{1}{2} \int_0^T \theta_t^2 dt\right),$$

und setze für \mathcal{F}_T das neue Wahrscheinlichkeitsmaß \mathbb{Q} via

$$\frac{d\mathbb{Q}}{d\mathbb{P}} = \Lambda_T.$$

Dann wird unter \mathbb{Q} der Prozess $\{\tilde{B}_t\}_{t \geq 0}$, definiert durch

$$\tilde{B}_t = B_t + \int_0^t \theta_s ds,$$

zu einer $\{\mathcal{F}_t\}$ -Brownschen Bewegung.

Bemerkung 2.5.2

Das Girsanov-Theorem ermöglicht den Wechsel zwischen verschiedenen Wahrscheinlichkeitsmaßen, unter denen die gleiche Brownsche Bewegung unterschiedliche Drifteigenschaften aufweist. Dies ist besonders hilfreich bei der Behandlung von Schrödinger-Brücken-Problemen, wo die Maßänderung genutzt wird, um optimale Pfade zu konstruieren.

Darüber hinaus ist die zugehörige Dichte

$$Z(\omega) = \frac{d\mathbb{Q}}{d\mathbb{P}}(\omega)$$

strikt positiv sein, wodurch Wahrscheinlichkeiten und Erwartungswerte zwischen \mathbb{Q} und \mathbb{P} transformiert werden können. Die Pfadstruktur stochastischer Prozesse – darunter Stetigkeit, Sprünge oder die Adaptiertheit an eine Filtration – bleibt unter beiden Maßen unverändert. Auch Martingaleigenschaften bleiben weitgehend erhalten, solange $Z_t := \frac{d\mathbb{Q}}{d\mathbb{P}}|_{\mathcal{F}_t}$ hinreichend glatt ist. Diese Eigenschaften sind in der Finanzmathematik (etwa bei Derivatebewertung) und in Schrödinger-Brücken-Problemen gleichermaßen bedeutsam. Für die Existenz eines Radon-Nikodým-Derivats bei der Maßänderung müssen bestimmte Bedingungen erfüllt sein, welche sicherstellen, dass die Gesamtwahrscheinlichkeit erhalten bleibt, d. h.

$$\mathbb{E}_{\mathbb{P}}\left[\frac{d\mathbb{Q}}{d\mathbb{P}}\right] = 1.$$

Darüber hinaus wird häufig die Novikov-Bedingung verwendet, die eine hinreichende Voraussetzung zur Existenz des Exponentialmartingals ist:

$$\mathbb{E}_{\mathbb{P}}\left[\exp\left(\frac{1}{2} \int_0^T \|\theta_t\|^2 dt + \int_0^T \theta_t dW_t\right)\right] < \infty.$$

Diese Bedingung garantiert, dass das stochastische Exponential

$$\Lambda_T = \exp\left(-\int_0^T \theta_t dW_t - \frac{1}{2} \int_0^T \|\theta_t\|^2 dt\right)$$

wohldefiniert bleibt.

In der Theorie der Schrödinger-Brücken nimmt genau diese Maß-Transformation eine zentrale Stellung ein. Ausgehend vom Pfadraum wird zunächst eine standardisierte Brownsche Bewegung betrachtet und anschließend mithilfe der Girsanov-Transformation ein neues Wahrscheinlichkeitsmaß ermittelt. Unter diesem neuen Maß erhält der Prozess einen Drift und es werden vorgegebene Randverteilungen erfüllt.

Beispiel 2.5.3 (Finanzmathematik)

Betrachten wir einen Aktienkurs $\{S_t\}$, der durch die stochastische Differentialgleichung

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

beschrieben wird. Hierbei ist $\{B_t\}_{t \geq 0}$ eine Brownsche Bewegung unter dem realen Wahrscheinlichkeitsmaß \mathbb{P} , μ die erwartete Rendite und σ die Volatilität. Wechselt man nun

zu einem risikoneutralen Maß \mathbb{Q} , so wird μ durch den risikofreien Zinssatz r ersetzt. Nach dem Girsanov-Theorem ändert sich die Brownsche Bewegung zu

$$\tilde{B}_t = B_t + \frac{\mu - r}{\sigma} t,$$

wobei die Dichte

$$Z_T = \exp\left(-\int_0^T \frac{\mu-r}{\sigma} dB_t - \frac{1}{2} \int_0^T \left(\frac{\mu-r}{\sigma}\right)^2 dt\right)$$

das neue Wahrscheinlichkeitsmaß \mathbb{Q} definiert. Unter \mathbb{Q} kann man die obige Gleichung dann als

$$dS_t = r S_t dt + \sigma S_t d\tilde{B}_t$$

schreiben. Die Pfadstruktur des Prozesses $\{B_t\}$ bleibt dabei erhalten, nur die Drift ändert sich. Dieses Vorgehen ist in der Finanzmathematik wesentlich, um etwa den fairen Preis von Derivaten unter dem risikoneutralen Maß \mathbb{Q} zu bestimmen.

2.5.1 Fokker-Planck-Gleichung

Ein zentrales Ziel in der stochastischen Analysis besteht darin, die Verteilung eines stochastischen Prozesses im Zeitverlauf zu beschreiben. Für Markovsche Diffusionsprozesse – wie solche, die durch $dX_t = \mu(X_t, t)dt + \sigma(X_t, t)dB_t$ angetrieben werden – ergibt sich eine partielle Differentialgleichung für die Wahrscheinlichkeitsdichte. Dieser Zusammenhang zeigt sich bereits im Spezialfall einer reinen Brownschen Bewegung $\{B_t\}$ ($dX_t = dB_t$), bei der die Verteilung sich gemäß der Wärmeleitungsgleichung entwickelt.

Es wird also zunächst die eindimensionale SDGL

$$dX_t = dB_t, \quad X_0 = 0 \in \mathbb{R},$$

betrachtet, wobei $\{B_t\}_{t \geq 0}$ eine Brownsche Bewegung ist. Im Folgenden wird hergeleitet, wie sich die Wahrscheinlichkeitsdichte $p(x, t)$ von X_t zeitlich entwickelt.

Herleitung im Spezialfall. Sei $f: \mathbb{R} \rightarrow \mathbb{R}$ zweimal stetig differenzierbar ($f \in C^2(\mathbb{R})$). Nach Itô gilt für einen Itô-Prozess X_t mit Dynamik $dX_t = dB_t$:

$$df(X_t) = \frac{\partial f}{\partial x}(X_t) dB_t + \frac{1}{2} \frac{\partial^2 f}{\partial x^2}(X_t) (dB_t)^2.$$

Da $(dB_t)^2 = dt$ (quadratische Variation), ergibt sich

$$df(X_t) = \frac{\partial f}{\partial x}(X_t) dB_t + \frac{1}{2} \frac{\partial^2 f}{\partial x^2}(X_t) dt.$$

Der Erwartungswert beider Seiten liefert

$$\mathbb{E}[df(X_t)] = \frac{1}{2} \mathbb{E}\left[\frac{\partial^2 f}{\partial x^2}(X_t)\right] dt,$$

da der stochastische Integralterm mit Drift $\partial f / \partial x$ unter hinreichender Glattheit von f erwartungstreu verschwindet. Falls X_t eine Dichte $p(x, t)$ besitzt, gilt

$$\mathbb{E}[f(X_t)] = \int_{-\infty}^{\infty} f(x) p(x, t) dx.$$

Eine Ableitung nach t und partielle Integration (vorausgesetzt, $p(x, t)$ und seine Ableitungen gehen an den Rändern hinreichend schnell gegen null) führen zur partiellen Differentialgleichung

$$\frac{\partial p}{\partial t}(x, t) = \frac{1}{2} \frac{\partial^2 p}{\partial x^2}(x, t),$$

die als *Wärmeleitungsgleichung* (Heat Equation) bezeichnet wird. Dieser Spezialfall beschreibt die Ausbreitung der Wahrscheinlichkeitsdichte eines Prozesses mit $dX_t = dB_t$, was einer Normalverteilung mit wachsender Varianz entspricht.

Allgemeiner Fall (Fokker-Planck-Gleichung). Eine analoge Argumentation lässt sich auf die Gleichung

$$dX_t = \mu(X_t, t) dt + \sigma(X_t, t) dB_t$$

übertragen (vgl. [22] oder [11]). Dann ergibt sich die **Fokker-Planck-Gleichung** (auch Kolmogorov-Vorwärtsgleichung):

$$\frac{\partial p}{\partial t}(x, t) = - \sum_{i=1}^n \frac{\partial}{\partial x_i} [\mu_i(x, t) p(x, t)] + \frac{1}{2} \sum_{i,j=1}^n \frac{\partial^2}{\partial x_i \partial x_j} [(\sigma \sigma^\top)_{ij}(x, t) p(x, t)].$$

Diese partielle Differentialgleichung beschreibt die zeitliche Entwicklung der Wahrscheinlichkeitsdichte eines \mathbb{R}^n -wertigen Itô-Prozesses mit Driftfunktion μ und Diffusionskoeffizient σ .

Bemerkung 2.5.3 (Bezug zur Schrödinger-Brücke)

Das Schrödinger-Brücken-Problem lässt sich als ein Randwertproblem für die Fokker-Planck-Gleichung interpretieren: Gesucht ist eine geeignete Driftfunktion, welche die von $\{X_t\}$ eingenommene Verteilung zur Zeit $t = 0$ in eine vorgegebene Zielverteilung zur Zeit $t = T$ überführt. Die Fokker-Planck-Gleichung erfasst dabei genau die Entwicklung der Wahrscheinlichkeitsdichte im Verlauf der Zeit. Damit wird deutlich, dass sie in der Theorie der Schrödinger-Brücken eine zentrale Rolle als "Herzstück" des Problems spielt.

In diesem Kapitel wurden die zentralen Konzepte der stochastischen Analysis vorgestellt, die für das Verständnis der Schrödinger-Brücke unverzichtbar sind. Insbesondere wurden grundlegende Begriffe wie Zufallsvariablen, Verteilungen und Momente eingeführt, woran sich die Diskussion über stochastische Prozesse, Filtrationen und Martingale anschloss. Die Brownsche Bewegung diente dabei mehrfach als Schlüsselbeispiel, das ihre Rolle als Referenzprozess verdeutlicht. Anschließend wurden das Itô-Integral und die Itô-Formel behandelt, die den Grundstein für stochastische Differentialgleichungen (SDGL) legen. Die Fokker-Planck-Gleichung bildet schließlich das Bindeglied zur Wahrscheinlichkeitsdichteentwicklung, welche im

Zentrum des Schrödinger-Brücken-Problems steht.

Im Hinblick auf die *Schrödinger-Brücke* hat sich vor allem gezeigt, dass stochastische Prozesse mittels einer *Maßänderung* (Girsanov-Theorem) zielgerichtet manipuliert werden können. Durch geeignete Driftterme lassen sich vorgegebene Anfangs- und Endverteilungen erreichen, was die Grundlage für eine nicht-stationäre Modellierung bildet.

3 Zeitreihenanalyse

Die Zeitreihenanalyse befasst sich mit der statistischen Analyse von Daten, die in zeitlicher Reihenfolge erfasst wurden. Eine *Zeitreihe* ist eine Folge von Beobachtungen $\{X_t\}_{t \in T}$, wobei T ein diskretes oder kontinuierliches Zeitgitter ist. Typische Anwendungsbereiche umfassen Finanzmärkte, Wettervorhersagen, medizinische Daten und viele weitere Felder [3, 4].

3.1 Definition von Zeitreihen

Eine Zeitreihe ist definiert als $\{X_t\}_{t \in T}$, wobei T entweder diskret ($T = \{t_1, t_2, \dots, t_N\}$) oder kontinuierlich ($T = [0, T]$) sein kann. Die Beobachtungen X_t können entweder eindimensional oder mehrdimensional sein, abhängig von der Anwendung.

Die Zeitreihenanalyse hat mehrere zentrale Ziele, die für viele Anwendungsbereiche entscheidend sind. Ein wesentliches Ziel ist die *Modellierung*, welche darauf abzielt, Modelle zu entwickeln, die die zugrunde liegende Dynamik einer Zeitreihe angemessen beschreiben. Die erstellten Modelle dienen folglich zum besseren Verständnis hinsichtlich der Struktur und der zugrunde liegenden Prozesse der Daten.

Ein weiteres Hauptziel ist die *Prognose*, die es ermöglicht, zukünftige Werte einer Zeitreihe basierend auf historischen Daten vorherzusagen. Dabei finden Methoden Anwendung, die sowohl die Dynamik der Zeitreihe als auch potenzielle externe Einflüsse berücksichtigen, um präzise Prognosen zu erstellen.

Des Weiteren umfasst die Zeitreihenanalyse die Untersuchung der *statistischen Eigenschaften* der Daten. Dies beinhaltet die Analyse von Trends, Saisonalitäten, Autokorrelationen sowie Spektraldichten, um tiefgreifendere Einblicke in die Struktur der Zeitreihe zu gewinnen. Die Identifikation solcher Eigenschaften spielt eine zentrale Rolle bei der Entwicklung und Validierung von Modellen sowie bei der Interpretation der Ergebnisse.

Ein weiteres grundlegendes Konzept in der Zeitreihenanalyse ist die *Stationarität*. Eine Zeitreihe wird als stationär bezeichnet, wenn ihre statistischen Eigenschaften, wie der Mittelwert, die Varianz und die Autokorrelationsstruktur, zeitlich konstant sind. Stationarität ist von zentraler Bedeutung, da viele Modelle in der Zeitreihenanalyse, wie z. B. autoregressive Modelle (AR) oder Moving-Average-Modelle (MA), die Annahme der Stationarität erfordern.

Definition 3.1.1 (Strikte Stationarität)

Eine Zeitreihe $\{X_t\}$ ist strikt stationär, wenn die gemeinsamen Verteilungen für alle k und für alle Zeitpunkte t_1, t_2, \dots, t_k gleich den Verteilungen für $t_1 + \tau, t_2 + \tau, \dots, t_k + \tau$ sind, für jede Verschiebung τ .

Eine weitere Betrachtung gilt der schwachen Stationarität von Zeitreihen.

Definition 3.1.2 (Schwache Stationarität)

Eine Zeitreihe $\{X_t\}$ ist schwach stationär, wenn

- $\mathbb{E}[X_t] = \mu$ für alle t .
- $\text{Var}(X_t) = \sigma^2$ für alle t .
- $\text{Cov}(X_t, X_{t+h})$ hängt nur von dem Lag h und nicht von der Zeit t ab.

Die Begriffe *strikte Stationarität* und *schwache Stationarität* beschreiben unterschiedliche Grade von "Zeitinvarianz" in einer Zeitreihe. Während die strikte Stationarität verlangt, dass *sämtliche* Verteilungen einer Zeitreihe unter Zeitverschiebung τ identisch bleiben, setzt die schwache Stationarität lediglich auf konstante Erwartung und Varianz, sowie die von t unabhängige Struktur der Kovarianz.

Da die strikte Stationarität alle Momente einschließt, impliziert sie unter hinreichenden Bedingungen (z. B. Existenz bestimmter Momente) auch die schwache Stationarität. Umgekehrt garantiert die schwache Stationarität nicht zwangsläufig auch die strikte Stationarität.

Viele statistische Modelle, wie zum Beispiel ARMA-Modelle, setzen *stationäre* Zeitreihen voraus. Eine solche Stationarität erleichtert die Modellierung und Analyse, da die Verteilungseigenschaften einer Zeitreihe sich nicht mit der Zeit verändern. Eine zentrale Größe bei der Untersuchung dieser Zeitreihen ist die *Autokorrelation*, welche die Korrelation einer Reihe $\{X_t\}$ mit sich selbst nach einer Verzögerung (Lag) um h Zeitpunkte misst.

Definition 3.1.3 (Autokorrelationsfunktion (ACF))

Die Autokorrelationsfunktion einer Zeitreihe $\{X_t\}$ ist definiert als

$$\rho(h) = \frac{\text{Cov}(X_t, X_{t+h})}{\text{Var}(X_t)}, \quad h = 0, 1, 2, \dots$$

Die *Spektraldichte* einer Zeitreihe gibt Auskunft darüber, wie sich die Varianz der Zeitreihe auf unterschiedliche Frequenzen verteilt. Sie ermöglicht insbesondere, periodische oder saisonale Einflüsse zu identifizieren und zu quantifizieren.

Definition 3.1.4 (Spektraldichte)

Die Spektraldichte $f(\lambda)$ einer Zeitreihe $\{X_t\}$ ist die Fourier-Transformierte ihrer Autokorrelationsfunktion:

$$f(\lambda) = \frac{1}{2\pi} \sum_{h=-\infty}^{\infty} \rho(h) e^{-i\lambda h},$$

wobei λ die Frequenz bezeichnet.

Auf diese Weise kann man den *energetischen* Anteil (d. h. die Varianz) der Zeitreihe in unterschiedlichen Frequenzbereichen untersuchen und damit saisonale oder periodische Muster besser verstehen. Durch ACF und Spektraldichte gemeinsam werden häufig wertvolle Hinweise für die Auswahl eines geeigneten Zeitreihenmodells (z. B. ARIMA oder saisonale Erweiterungen) gewonnen.

3.2 Modellierung von Zeitreihen

Die Modellierung von Zeitreihen kann auf klassischen statistischen Methoden oder auf modernen stochastischen Prozessen basieren. In diesem Abschnitt werden zunächst die grundlegenden Prozesse der Autoregression (AR) und des Moving Average (MA) vorgestellt, bevor diese in kombinierten Modellen wie ARMA und ARIMA Verwendung finden.

3.2.1 AR-Modelle

Ein *autoregressives Modell* der Ordnung p (AR(p)) beschreibt die Zeitreihe $\{X_t\}$ als eine lineare Kombination ihrer eigenen vorherigen Werte und eines Fehlerterms. Mathematisch wird dies wie folgt ausgedrückt

$$X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \dots + \phi_p X_{t-p} + \epsilon_t,$$

wobei $\phi_1, \phi_2, \dots, \phi_p$ die Parameter des AR-Modells sind und ϵ_t der Fehlerterm zum Zeitpunkt t ist, der als $\epsilon_t \sim \mathcal{N}(0, \sigma^2)$ angenommen wird.

Die aktuelle Beobachtung X_t hängt *linear* von den vorherigen p Werten ab. Für das AR(p)-Modell ist die *Stationaritätsbedingung* gegeben durch die Eigenschaft, dass alle Wurzeln des charakteristischen Polynoms außerhalb des Einheitskreises liegen

$$\phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p = 0,$$

wobei B der Backshift-Operator ist ($BX_t = X_{t-1}$).

3.2.2 MA-Modelle

Ein *Moving Average Modell* der Ordnung q (MA(q)) beschreibt die Zeitreihe $\{X_t\}$ als eine lineare Kombination der aktuellen und vorherigen Fehlerterme. Mathematisch wird dies wie folgt ausgedrückt:

$$X_t = \epsilon_t + \theta_1 \epsilon_{t-1} + \theta_2 \epsilon_{t-2} + \dots + \theta_q \epsilon_{t-q},$$

wobei $\theta_1, \theta_2, \dots, \theta_q$ die Parameter des MA-Modells sind und ϵ_t der Fehlerterm zum Zeitpunkt t ist, der als $\epsilon_t \sim \mathcal{N}(0, \sigma^2)$ angenommen wird.

Die aktuelle Beobachtung X_t hängt linear von den aktuellen und den letzten q Fehlertermen ab. MA(q)-Modelle haben eine Autokorrelationsfunktion (ACF), die nach der Lag q abrupt auf null abfällt.

3.2.3 ARMA-Modelle

Das *Autoregressive Moving Average Modell* der Ordnung p und q (ARMA(p, q)) kombiniert sowohl autoregressive als auch gleitende Mittelwertkomponenten. Mathematisch wird dies wie folgt ausgedrückt

$$X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \dots + \phi_p X_{t-p} + \epsilon_t + \theta_1 \epsilon_{t-1} + \theta_2 \epsilon_{t-2} + \dots + \theta_q \epsilon_{t-q},$$

wobei $\phi_1, \phi_2, \dots, \phi_p$ die autoregressiven Parameter sind, $\theta_1, \theta_2, \dots, \theta_q$ die gleitenden Mittelwertparameter sind und $\epsilon_t \sim \mathcal{N}(0, \sigma^2)$ der Fehlerterm ist.

ARMA-Modelle können eine Vielzahl von Zeitreihenmustern durch die Kombination von AR- und MA-Komponenten erfassen. Für die Stationarität und Invertierbarkeit des ARMA-Modells müssen sowohl die AR- als auch die MA-Komponenten die entsprechenden Bedingungen erfüllen.

3.2.4 ARIMA-Modelle

Das *Autoregressive Integrated Moving Average Modell* (ARIMA(p, d, q)) erweitert das ARMA-Modell um eine Integrationskomponente d , die Differenzierung zur Handhabung nicht-stationärer Zeitreihen beinhaltet. Mathematisch wird dies wie folgt ausgedrückt

$$\phi(B)(1 - B)^d X_t = \theta(B)\epsilon_t,$$

wobei, $(1 - B)^d$ der Differenzierungsoperator ist und $\phi(B)$ und $\theta(B)$ die AR- bzw. MA-Polynome sind.

Durch die Differenzierung können Trends und Saisonalitäten in der Zeitreihe entfernt werden, um Stationarität zu erreichen. ARIMA-Modelle sind besonders nützlich für Zeitreihen, die nicht stationär sind, da sie eine differenzierte Version der Daten modellieren.

Klassische Zeitreihenmodelle finden breite Anwendung in verschiedenen Bereichen, in denen Daten über Zeiträume hinweg analysiert werden. Ein prominentes Beispiel sind Finanzzeitreihen, wie etwa Aktienkurse, Wechselkurse oder Zinsraten. Diese Daten zeichnen sich häufig durch komplexe Trends und Volatilitätsmuster aus. Zur Analyse solcher Zeitreihen werden häufig ARIMA-Modelle eingesetzt, da sie in der Lage sind, sowohl Trends als auch Autokorrelationen und zufällige Schwankungen zu erfassen. Für die Beschreibung der dynamischen Volatilität in Finanzzeitreihen werden ARIMA-Modelle oft durch spezialisierte Ansätze wie GARCH-Modelle ergänzt.

Ein weiteres Anwendungsfeld sind Wetterdaten, beispielsweise Temperaturmessungen oder Niederschlagsmengen. Diese Zeitreihen weisen typischerweise ausgeprägte saisonale Muster auf, die periodische Schwankungen innerhalb eines Jahres darstellen. Für solche Daten eignen sich SARIMA-Modelle (Seasonal ARIMA) besonders gut, da sie auch saisonale Effekte explizit berücksichtigen können. Saisonale Modelle enthalten zusätzlich saisonale Operatoren, $\Phi(B^s)$ und $\Theta(B^s)$, um periodische Muster explizit zu modellieren [3]. Diese Modelle sind nicht nur für die Vorhersage kurzfristiger klimatischer Bedingungen hilfreich, sondern auch für die Analyse langfristiger klimatischer Trends.

Durch den Einsatz spezialisierter Modelle wie ARIMA können somit sowohl komplexe Finanzzeitreihen als auch saisonal beeinflusste Wetterdaten beschrieben und analysiert werden.

3.2.5 GARCH-Modelle

Die *Generalized Autoregressive Conditional Heteroskedasticity (GARCH)*-Modelle [2,6] erfassen zeitliche Abhängigkeiten in der bedingten Varianz von $\{X_t\}$.

Im GARCH(p, q)-Modell besteht die Zeitreihe meist aus einer *Mean Equation* (oft ARMA) und einer *Variance Equation*

$$\sigma_t^2 = \alpha_0 + \alpha_1 \epsilon_{t-1}^2 + \dots + \alpha_q \epsilon_{t-q}^2 + \beta_1 \sigma_{t-1}^2 + \dots + \beta_p \sigma_{t-p}^2.$$

Die Bedingung $\sum_{i=1}^q \alpha_i + \sum_{j=1}^p \beta_j < 1$ gewährleistet (schwache) Stationarität. Insbesondere im Finanzbereich sind GARCH-Modelle sehr beliebt, weil sie Volatilitätsclustering ("Phasen hoher Volatilität folgen auf hohe Volatilität") abbilden.

Beispiel 3.2.1

Wir betrachten ein einfaches Autoregressionsmodell erster Ordnung (AR(1))

$$X_t = \phi X_{t-1} + \epsilon_t,$$

mit $\phi = 0,8$ und $\epsilon_t \sim \mathcal{N}(0, 1)$.

Die Simulation des AR(1)-Prozesses erfolgt durch die Generierung einer Folge von X_t -Werten, die gemäß der obigen Gleichung bestimmt sind. Der Ausgangspunkt ist dabei $X_0 = \epsilon_0$ und die Iteration erfolgt über die Zeitpunkte. Eine Realisierung ist in Abbildung 6 zu sehen, sowie die zugehörige Autokorrelationsfunktion.

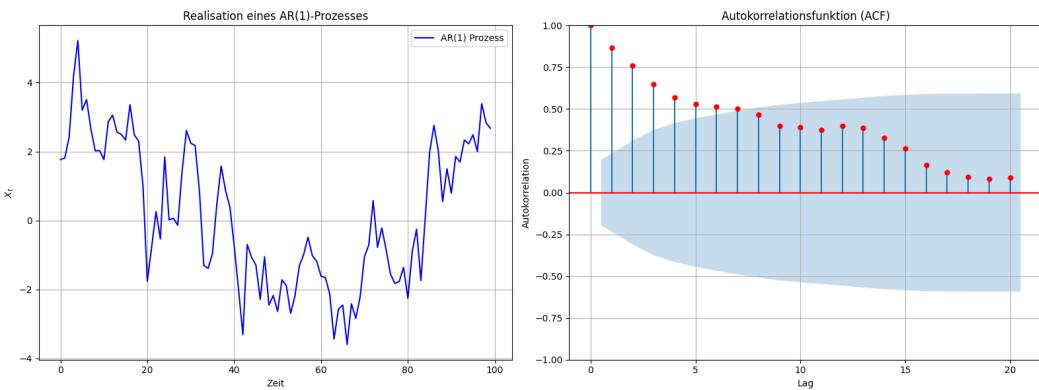


Abbildung 6: Realisationen eines AR(1)-Prozesses mit $\phi = 0,8$ und $\epsilon_t \sim \mathcal{N}(0, 1)$. Linke Grafik: Zeitreihe; Rechte Grafik: Autokorrelationsfunktion.

Beispiel 3.2.2

Angenommen, es werden die täglichen Renditen einer Aktie R_t mittels eines GARCH(1,1)-Modells modelliert.

$$R_t = \mu + \epsilon_t,$$

$$\epsilon_t = \sigma_t Z_t, \quad Z_t \sim \mathcal{N}(0, 1),$$

$$\sigma_t^2 = \alpha_0 + \alpha_1 \epsilon_{t-1}^2 + \beta_1 \sigma_{t-1}^2.$$

Hierbei beschreibt μ den durchschnittlichen Renditenwert, σ_t^2 die bedingte Varianz zum Zeitpunkt t , und Z_t ist eine standardnormalverteilte Zufallsvariable.

Wir betrachten das GARCH-Modell erster Ordnung ($GARCH(1,1)$)

$$\sigma_t^2 = \alpha_0 + \alpha_1 \epsilon_{t-1}^2 + \beta_1 \sigma_{t-1}^2,$$

wobei σ_t^2 die bedingte Varianz zum Zeitpunkt t ist, $\alpha_0 > 0$ der konstante Term ist, $\alpha_1 \geq 0$ den ARCH-Term darstellt, der die kurzfristigen Schwankungen erfasst, $\beta_1 \geq 0$ den GARCH-Term darstellt, der die langfristigen Schwankungen erfasst und $\epsilon_t \sim \mathcal{N}(0, \sigma_t^2)$ der Fehlerterm ist.

Dieses Modell ermöglicht es, die Volatilität der Zeitreihe dynamisch zu modellieren, indem es vergangene Fehlerterme und vergangene Varianzen berücksichtigt. Eine Realisation des $GARCH(1,1)$ -Prozesses sowie die zugehörige Autokorrelationsfunktion der simulierten Renditen sind in Abbildung 7 zu sehen.

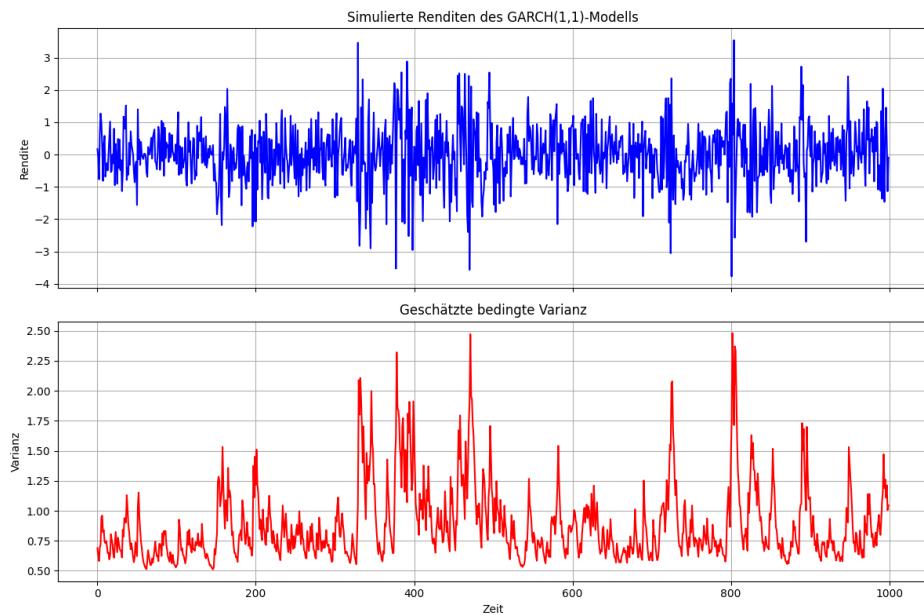


Abbildung 7: Volatilitätsdynamik und Autokorrelationsfunktion (ACF) der simulierten GARCH(1,1)-Renditen. Oben: Simulierte Renditen; Unten: Geschätzte bedingte Varianz σ_t^2 und ACF der simulierten Renditen.

3.3 Modellwahl und Parameter-Schätzung

Im Folgenden wird ein allgemeiner Ansatz zur *Modellauswahl* und *Parameterschätzung* beschrieben. Die Prinzipien gelten insbesondere für *lineare* und *nichtlineare* Zeitreihenstrukturen, unter Einbeziehung von Regularisierung, saisonalen Komponenten oder exogenen Variablen.

Explorative Analyse und Modellauswahl

Zu Beginn einer Zeitreihenanalyse empfiehlt sich eine gründliche *Datenbereinigung* und *Vorverarbeitung*, bei der etwaige Ausreißer, fehlende Werte oder abrupte Strukturbrüche identifiziert und behandelt werden. Zeigen sich starke Nichtstationaritäten, können Verfahren wie die *Differenzierung* oder eine *Transformation* (etwa nach Box-Cox) notwendig sein, um die Daten für stationäre Modelle tauglich zu machen.

Im nächsten Schritt wird ein *Test auf Stationarität* durchgeführt, beispielsweise mithilfe des Augmented-Dickey–Fuller-Tests (ADF) oder des KPSS-Tests. Viele klassische Zeitreihenmodelle wie ARMA und ARIMA setzen stationäre Daten voraus; dementsprechend entfernt man vorhandene Trend- oder Saisoneffekte (sofern vorhanden), bevor man das Modell anpasst.

Ein nützliches Werkzeug für die Bestimmung der Modellordnung sind *ACF-* und *PACF-Diagramme*. Die *Autokorrelationsfunktion* (ACF) und die *Partielle Autokorrelationsfunktion* (PACF) ermöglichen eine erste Einschätzung darüber, welche autoregressiven (AR) bzw. MA(gleitenden) Komponenten ein Modell haben könnte. Bei stark saisonalen Daten (z. B. Temperaturreihen) betrachtet man zusätzlich eine *Saisonaler-ACF* (SACF), um Hinweise auf periodische Strukturen zu erkennen.

Nach der Vorauswahl eines oder mehrerer plausibler Modelle erfolgt deren *quantitative Vergleich* mithilfe von *Informationskriterien* wie dem Akaike-Information-Kriterium (AIC), dem Bayes-Information-Kriterium (BIC) oder dem Hannan–Quinn-Kriterium. Ein Modell mit einem möglichst kleinen Informationskriterium gilt dabei als besonders guter Kompromiss zwischen Modellgüte und Sparsamkeit (Ockhams Rasiermesser). Diese Kriterien können zudem um spezifische Komponenten (z. B. saisonale Anteile, GARCH-Effekte) erweitert werden, falls die Daten entsprechende Strukturen nahelegen.

Parameterschätzung

Eine *präzise Parameterschätzung* bestimmt maßgeblich die Modellgüte und die Zuverlässigkeit von Prognosen. Man unterscheidet grundsätzlich *parametrische Methoden*, die konkrete Verteilungsannahmen voraussetzen (z. B. Normalverteilung der Fehler), und *nichtparametrische Ansätze*, die mithilfe von Kernel- oder Wavelet-Methoden auf vorgegebene Verteilungsformen verzichten. Für lineare Zeitreihenmodelle (AR, MA, ARMA) kommen häufig folgende Verfahren zum Einsatz:

Die *Methode der kleinsten Quadrate* (OLS) minimiert die Summe der quadrierten Fehler

$$\hat{\theta}_{\text{OLS}} = \arg \min_{\theta} \sum_{t=1}^T (Y_t - \hat{Y}_t(\theta))^2.$$

Bei AR(1)-Prozessen ist der OLS-Schätzer analytisch bestimmbar. Für komplexere Strukturen (z. B. AR(p) mit großem p) wird häufig auf numerische Optimierung oder Regularisierungsverfahren (LASSO, Ridge) zurückgegriffen, um Overfitting zu vermeiden.

Der *Maximum-Likelihood-Schätzer* (MLE) maximiert die Likelihood unter Annahme einer bestimmten *Fehlerverteilung* (oft Normalverteilung)

$$\hat{\theta}_{\text{MLE}} = \arg \max_{\theta} \prod_{t=1}^T p(Y_t | Y_{t-1}, \dots, Y_1; \theta).$$

Insbesondere nichtlineare Modelle (z. B. GARCH oder ARIMA-Modelle) werden üblicherweise per MLE geschätzt.

Der Schätzer basierend auf den Yule-Walker-Gleichungen bei einem reinen AR(p)-Modell ermöglicht eine effiziente Bestimmung der Parameter. Dazu ist zunächst die Schätzung der Autokorrelationsfunktionen aus den Daten erforderlich, woraufhin ein lineares System gelöst werden kann.

Modellvalidierung

Nach der Schätzung des Zeitreihenmodells folgt die *Validierung*, mit der die Güte und Zuverlässigkeit des angepassten Modells überprüft werden. Ein zentrales Instrument hierfür ist die *Residuenanalyse*: Die Residuen $\hat{e}_t = Y_t - \hat{Y}_t$ sollten keine Autokorrelation mehr aufweisen — sind sie noch korreliert, so deutet das auf eine unzureichende Modellierung hin. Zur formalen Überprüfung kann etwa der Ljung–Box-Test herangezogen werden. Abweichungen von der Normalverteilungsannahme lassen sich in einem *Normal-QQ-Plot* erkennen.

Ein weiteres Kriterium ist die *Prognosegenauigkeit*, die *Out-of-Sample* validiert wird. Man teilt etwa den Beobachtungszeitraum in ein Trainings- und ein Testfenster und beurteilt die Vorhersagen anhand von Kennzahlen wie dem RMSE (Root Mean Squared Error) oder dem MAE (Mean Absolute Error). In Finanzmarktanwendungen oder anderen dynamischen Umgebungen wird das Verfahren oft in *Rolling Windows* wiederholt, um eine zeitliche Stabilität der Prognosegüte zu gewährleisten.

Im *Vergleich mit Alternativmodellen* (z. B. ARIMA(p, d, q), SARIMA bei saisonalen Komponenten, GARCH bei varianzdynamischen Effekten) kommen erneut *Informationskriterien* (AIC, BIC) oder Cross-Validation-Techniken zum Einsatz, um verschiedene Modellklassen quantitativ gegeneinander abzuwägen. Schließlich wird die *Stabilität* des Modells untersucht, indem man konstante Parameter über die Zeit fordert (z. B. CUSUM-Test). Signifikante Änderungen der geschätzten Modellparamete-

ter können auf Strukturbrüche oder eine deutliche Zeitvariabilität im Datenprozess hinweisen.

Im Folgenden werden die *Modellauswahl*, die *Parameterschätzung* und die *Validierung* anhand einer simulierten Zeitreihe veranschaulicht. Ziel ist es, zu zeigen, wie ein geeignetes ARMA(p, q)-Modell identifiziert und die Modellparameter geschätzt werden können. Besonderes Augenmerk wird auf die Validierung der Modellgüte durch Residuenanalyse und statistische Tests gelegt. Die Schritte umfassen die Modellauswahl auf Basis explorativer Verfahren, die Parameterschätzung mittels Maximum-Likelihood und schließlich die Validierung anhand der Residuen. Die Ergebnisse werden im Kontext der ARMA(1,1)-Schätzung erläutert.

Beispiel 3.3.1

Eine ARMA(1,1)-Zeitreihe wird mit den wahren Parametern

$$\phi_{\text{true}} = 0.6, \quad \theta_{\text{true}} = 0.3, \quad \sigma_{\text{true}}^2 = 1.0$$

simuliert. Der zugrunde liegende Prozess lautet

$$Y_t = \phi Y_{t-1} + \epsilon_t + \theta \epsilon_{t-1}, \quad t = 1, \dots, 1000,$$

wobei ϵ_t weißes Rauschen mit Varianz σ^2 ist. Eine Realisation $\{Y_t\}$ der Länge $T = 1000$ wird erzeugt und anschließend durch ein ARMA(1,1)-Modell mithilfe der Maximum-Likelihood-Methode geschätzt.

Tabelle 1 fasst die wahren Parameter, die geschätzten Werte sowie deren jeweilige Abweichungen zusammen.

Parameter	Wahrer Wert	Geschätzter Wert	Abweichung
ϕ	0.6	0.6111	0.0111
θ	0.3	0.3165	0.0165
σ^2	1.0	0.9996	0.0004

Tabelle 1: Vergleich zwischen wahren und geschätzten Parametern im ARMA(1,1)-Modell; Maximum-Likelihood-Schätzung.

Aus Tabelle 1 wird deutlich, dass die geschätzten Parameter sehr nah an den wahren Werten liegen. Die Abweichungen betragen lediglich $|\phi - \hat{\phi}| \approx 0.011$, $|\theta - \hat{\theta}| \approx 0.017$ und $|\sigma^2 - \hat{\sigma}^2| \approx 0.0004$. Dies unterstreicht die Konsistenz der Maximum-Likelihood-Schätzung bei einer ausreichend langen Zeitreihe.

Nach der Schätzung des ARMA(1,1)-Modells werden die Residuen analysiert, um die Modellanpassung zu validieren. Tabelle 2 fasst die Ergebnisse der durchgeführten Tests zusammen.

Die Testergebnisse bestätigen die gute Anpassung des ARMA(1,1)-Modells. Der Ljung-Box-Test (Q) weist eine Teststatistik von 0.00 und einen p -Wert von 0.95 auf. Dies zeigt,

Test/Kennzahl	Wert
Ljung-Box (L1) (Q)	0.00
Prob(Q)	0.95
Jarque-Bera (JB)	0.17
Prob(JB)	0.92
Heteroskedasticity (H)	1.00
Prob(H) (two-sided)	0.98
Skew	-0.03
Kurtosis	2.98

Tabelle 2: Ergebnisse der statistischen Tests zur Validierung des ARMA(1,1)-Modells.

dass keine signifikanten Autokorrelationen in den Residuen vorliegen, was darauf hindeutet, dass die Abhängigkeiten durch das Modell hinreichend erklärt werden.

Der Jarque-Bera-Test ergibt eine Teststatistik von 0.17 und einen p-Wert von 0.92. Dies deutet darauf hin, dass die Residuen keine signifikante Abweichung von der Normalverteilung zeigen. Da die Annahme der Normalverteilung für viele statistische Modelle wichtig ist, unterstützen diese Ergebnisse die Validität des Modells.

Die Analyse der Heteroskedastizität ergibt eine Teststatistik von 1.00 und einen p-Wert von 0.98. Dies weist darauf hin, dass die Varianz der Residuen als konstant angenommen werden kann, was ebenfalls ein Zeichen für eine gute Modellanpassung ist.

Die Kennzahlen zur Schiefe -0.03 und Kurtosis (Wölbung) 2.98 liegen nahe an den Werten einer idealen Normalverteilung (Schiefe = 0, Kurtosis = 3). Die Residuen sind daher nahezu symmetrisch und weisen keine ungewöhnlichen Extremwerte auf.

Zusammenfassend zeigen die Testergebnisse, dass die Residuen des geschätzten ARMA(1,1)-Modells die theoretischen Annahmen erfüllen, was auch zu erwarten war, da es sich um ein künstliches Beispiel handelt. Das Modell erklärt die zugrunde liegende Zeitreihe adäquat und liefert valide Ergebnisse, die durch die statistischen Tests gestützt werden. Abbildung 8 verdeutlicht zusätzlich in der oberen Grafik den Verlauf der simulierten Zeitreihe mit den geschätzten ARMA(1,1)-Parametern. Die untere Grafik illustriert die Residuen \hat{e}_t . Deren Zufallscharakter ohne auffällige Muster oder Korrelationen deutet auf eine gute Modellanpassung hin.

Die in diesem Kapitel vorgestellten Methoden und Ergebnisse verdeutlichen die Bedeutung einer fundierten Modellauswahl und -validierung für Zeitreihen aller Art – stationär oder nicht-stationär. Für komplexe Anwendungsfelder, etwa Finanzzeitreihen oder ausgeprägte Saisonalitäten, empfiehlt sich der Einsatz erweiterter Modellansätze wie ARIMA-, SARIMA- oder GARCH-Modelle, um die zeitliche Dynamik angemessen abzubilden. Dieser strukturierte Ansatz bietet eine solide Grundlage für den praktischen Einsatz von Zeitreihenmodellen in unterschiedlichsten Disziplinen.

Dennoch basiert die klassische Zeitreihenanalyse, wie sie etwa mit ARIMA- oder

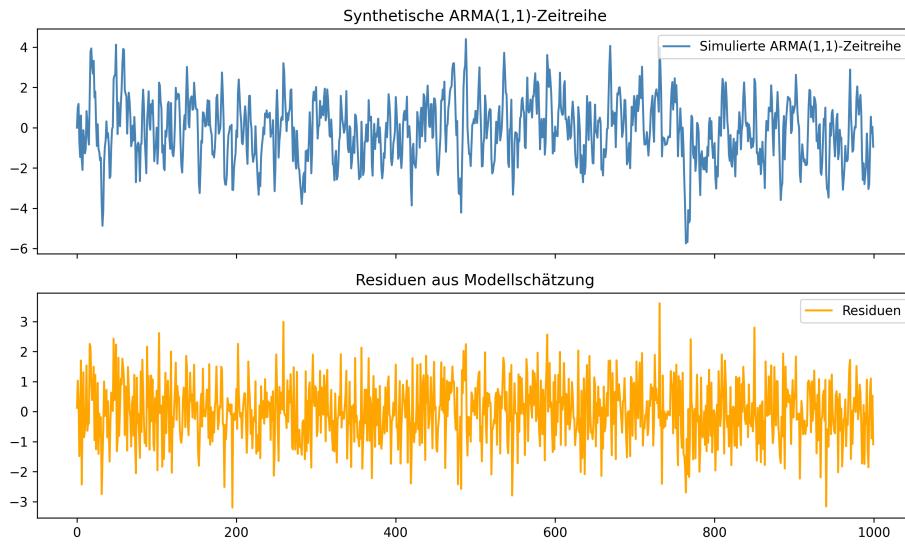


Abbildung 8: Oben: ARMA(1,1)-Zeitreihe (simuliert und mit geschätzten Parametern). Unten: Residuen $\hat{\epsilon}_t$.

GARCH-Modellen realisiert wird, häufig auf der Annahme (nahezu) stationärer oder linearer Strukturen. Tatsächlich weisen reale Zeitreihen oft nichtlineare Effekte oder signifikante Ausreißer auf, sodass klassische Modelle nur eine begrenzte Abbildung dieser Phänomene ermöglichen. Das später folgende Kapitel 5 stellt daher das *Schrödinger-Problem* für Zeitreihen vor. Diese Methode erlaubt es, aus vorhandenen Zeitreihen neue Pfade zu generieren, die die statistischen Eigenschaften der Ursprungsdaten bewahren.

Eine zentrale Stärke der Schrödinger-Brücke liegt in ihrer Fähigkeit, fließende Übergänge zwischen verschiedenen Verteilungen zu modellieren. Anders als klassische Modelle, bei denen Eigenschaften wie Varianz oder Drift zeitlich fest vorgegeben sind, ermöglicht sie eine stochastische Interpolation zwischen unterschiedlichen Verteilungen. Dies ist vor allem für Zeitreihen vorteilhaft, die durch nichtstationäre Phasen oder abrupte Dynamikwechsel charakterisiert sind.

Zudem erlaubt die Schrödinger-Brücke eine *kontrollierte Pfadkonstruktion*. Durch Girsanov-Transformation oder optimalen Transport im Wahrscheinlichkeitsraum lassen sich gezielt Anfangs- und Endverteilungen verbinden, auch in hochdimensionalen Szenarien. Hierdurch wird ein Bereich eröffnet, in dem sich komplexe Zeitreihen, deren Modellierung mit herkömmlichen Ansätzen nur schwer realisierbar ist, auf nachvollziehbare Weise generieren lassen.

4 Schrödinger-Brücken-Problem

Die Theorie der Schrödinger-Brücken stellt eine Verbindung zwischen stochastischen Prozessen, der Variationsrechnung und der statistischen Mechanik her. Im Kern geht es um die Frage, wie sich ein System zwischen zwei vorgegebenen Zustandsverteilungen probabilistisch entwickelt, wobei eine optimale stochastische Dynamik gesucht wird. Diese Dynamik minimiert die Abweichung von einem gegebenen Referenzprozess und respektiert gleichzeitig die beobachteten Anfangs- und Endverteilungen. Damit bietet das Schrödinger-Brücken-Problem eine mathematische Grundlage für Anwendungen in Bereichen wie der Bildverarbeitung, Quantenmechanik und dem maschinellen Lernen.

4.1 Historischer Hintergrund

Das *Schrödinger-Brücken-Problem* wurde erstmals von ERWIN SCHRÖDINGER im Jahr 1931 im Kontext der statistischen Mechanik in [18] formuliert. Schrödinger untersuchte die wahrscheinlichste Entwicklung eines Systems von Teilchen, das zwischen zwei beobachteten Dichteverteilungen evolviert, unter der Annahme, dass die Teilchen zufälligen thermischen Fluktuationen unterliegen. Dieses Problem verbindet Konzepte aus der Thermodynamik, Quantenmechanik und Wahrscheinlichkeitstheorie.

Die grundlegende Motivation des Schrödinger-Brücken-Problems liegt darin, eine stochastische Dynamik zu finden, die die beobachteten Anfangs- und Endzustände reproduziert, während die Entropie des Systems minimiert wird. Damit stellt es eine Variation des klassischen Variationsprinzips dar, bei der die Entropie anstelle der Energie minimiert wird.

Formalisiert beschäftigt sich das Problem mit der folgenden Fragestellung: Gegeben seien zwei Wahrscheinlichkeitsmaße μ_0 und μ_T auf \mathbb{R}^n sowie ein Markov-Prozess mit Übergangsdichte $q(s, y, t, x)$ für $0 < s < t < T$. Die klassische Interpretation ist, dass die Markov-Dynamik, ausgehend von einer Startverteilung μ_0 , zu einer Endverteilung μ_T führt. Beobachtet man jedoch eine tatsächliche Endverteilung, die von dieser Markov-Dynamik abweicht, stellt sich die Frage, welches alternative Übergangsdichte $\tilde{q}(s, y, t, x)$ die Anfangs- und Endverteilung μ_0 und μ_T erfüllt und der ursprünglichen Dynamik q möglichst nahekommt. Dieses Problem wird als *Schrödinger-Brücken-Problem (SBP)* bezeichnet.

Das SBP führt zu *reziproken Prozessen*, einer Klasse von stochastischen Prozessen, die zwischen Markov- und allgemeinen nicht-Markov-Prozessen liegen. Während jeder Markov-Prozess reziprok ist, gilt die Umkehrung im Allgemeinen nicht. Die Theorie reziproker Prozesse wurde umfassend untersucht, siehe etwa [13] für eine Übersicht. Innerhalb dieser Klasse liefert das SBP eine Methode, aus zwei vorgegebenen Randverteilungen μ_0 und μ_T einen Prozess zu konstruieren, dessen Zwischenverteilungen zu Zeiten $0 < s < t < T$ die Bedingungen erfüllen und der in einem geeigneten Sinn

die minimale Abweichung von der Referenzdynamik q aufweist.

Ein zentraler Beitrag zur Theorie des Schrödinger-Brücken-Problems stammt von DAI PRA [15], der das Problem als ein stochastisches Kontrollproblem formulierte. Dabei wird die Transformation eines Ausgangsdiffusionsprozesses zu einem Zielprozess, der die vorgegebenen Randbedingungen erfüllt, als Steuerungsproblem mit minimaler Energie interpretiert. Dieser Ansatz ermöglicht unter geeigneten Annahmen eine explizite oder zumindest gut charakterisierbare Lösung des Schrödinger-Brücken-Problems.

4.2 Einleitung

Ein reeller n -dimensionaler Diffusionsprozess X_t auf dem Intervall $[0, T]$ wird durch die SDGL

$$dX_t = \mu(X_t, t) dt + \sigma(X_t, t) dB_t, \quad (4.1)$$

beschrieben, wobei $\{B_t\}_{t \geq 0}$ im Allgemeinen eine n -dimensionale Brownsche Bewegung ist und μ, σ die Drift- bzw. Diffusionskomponenten darstellen. Es wird zusätzlich angenommen, dass die elliptische Bedingung

$$\xi^\top (\sigma \sigma^\top)(x, t) \xi \geq \eta \|\xi\|^2, \quad \eta > 0,$$

erfüllt ist. Dies gewährleistet, dass die Diffusionsmatrix $a(x, t) := \sigma(x, t) \sigma(x, t)^\top$ symmetrisch, positiv definit und von unten durch η beschränkt ist. Unter geeigneten Regularitätsbedingungen besitzt die obige SDGL eine eindeutig definierte schwache Lösung (siehe Annahmen 4.2.1 und 4.2.2).

Der Generator des Diffusionsprozesses X_t wird durch den Operator L definiert, der $f : \mathbb{R}^n \times [0, T] \rightarrow \mathbb{R}$ mit $f \in C^{2,1}(\mathbb{R}^n \times [0, T])$ wie folgt wirkt

$$Lf(x, t) = \sum_{i=1}^n \mu_i(x, t) \frac{\partial f}{\partial x_i}(x, t) + \frac{1}{2} \sum_{i,j=1}^n (\sigma \sigma^\top)_{ij}(x, t) \frac{\partial^2 f}{\partial x_i \partial x_j}(x, t).$$

Dieser Generator bildet die Grundlage für die Herleitung parabolischer partieller Differentialgleichungen, die die zeitliche Entwicklung von Wahrscheinlichkeitsdichten beschreiben (vgl. Fokker–Planck-Gleichung). Im Spezialfall einer reinen Brownschen Bewegung, bei der $\mu(x, t) = 0$ und $\sigma(x, t)$ die Einheitsmatrix ist, reduziert sich der Generator auf

$$Lf(x, t) = \frac{1}{2} \Delta f(x, t),$$

wobei Δ der Laplace-Operator ist. In diesem Fall führt die zugehörige Wahrscheinlichkeitsdichte $p(x, t)$ zur Wärmeleitungsgleichung (vgl. [18]), die die zeitliche Entwicklung einer sich ausbreitenden Normalverteilung beschreibt.

Für den allgemeinen Fall eines stochastischen Prozesses mit Drift- und Diffusionstermen, wie in (4.1) beschrieben, ergibt sich die *Fokker–Planck-Gleichung* (auch *Kolmogorov–Vorwärtsgleichung* genannt). Diese beschreibt die zeitliche Entwicklung der Wahrscheinlichkeitsdichte $p(x, t)$ unter Einfluss von Drift μ und Diffusion σ . Im Folgenden sind die zu treffenden Annahmen aufgeführt.

Annahme 4.2.1 (Annahme A1)

Für jede quadratintegrierbare Zufallsvariable X_0 , die unabhängig von der Brownschen Bewegung B_t ist, besitzt die SDGL (4.1) eine eindeutig definierte schwache Lösung auf dem Intervall $[0, T]$ mit Anfangswert $X_0 = x_0$.

Annahme 4.2.2 (Annahme A2)

Sei $q(s, y, t, x)$ die Übergangsdichte des Prozesses X_t . Es wird angenommen, dass $q(s, y, t, x)$ die parabolische partielle Differentialgleichung (PDGL)

$$\frac{\partial}{\partial t} f(t, x) + \mathcal{L}f(t, x) = 0$$

löst, wobei \mathcal{L} der Generator des Diffusionsprozesses ist. Diese PDGL beschreibt analog zur Fokker-Planck-Gleichung die zeitliche Entwicklung von Wahrscheinlichkeitsdichten.

Die Übergangsdichte $q(s, y, t, x)$ sei in zweimal stetig differenzierbar in y und stetig differenzierbar in s . Ferner gilt: Für jede positiv messbare Funktion g mit

$$\int_{\mathbb{R}^n} q(0, x, T, z) g(z) dz < +\infty,$$

ist die Funktion

$$h(x, t) := \int_{\mathbb{R}^n} q(t, x, T, z) g(z) dz$$

in $C^{2,1}(\mathbb{R}^n \times [0, T])$ enthalten und erfüllt

$$\frac{\partial}{\partial t} h(x, t) + \mathcal{L}h(x, t) = 0.$$

Proposition 4.2.1

Die folgenden Bedingungen gewährleisten die Gültigkeit von Annahme (4.2.1) und Annahme (4.2.2):

- (i) Die Koeffizienten $\mu(X, t)$ sind beschränkt, stetig und erfüllen Hölder-Bedingungen bezüglich X .
- (ii) Die Koeffizienten $\sigma(X, t)$ sind in $C^1(\mathbb{R}^n \times [0, T])$. Außerdem sind $\sigma(X, t)$ und deren Ableitungen $\frac{\partial \sigma}{\partial x_k}$ beschränkt und erfüllen Hölder-Bedingungen sowohl bezüglich des Ortes als auch der Zeitvariablen.
- (iii) Die Diffusionsmatrix $a(X_t, t) = \sigma(X_t, t) \sigma(X_t, t)^\top$ erfüllt für alle $X \in \mathbb{R}^n$, $t \in [0, T]$ die elliptische Bedingung

$$\xi^\top a(X_t, t) \xi \geq \eta \|\xi\|^2 \quad \text{für alle } \xi \in \mathbb{R}^n,$$

wobei $\eta > 0$ eine Konstante ist.

4.3 Problemformulierung

Wie in der Arbeit von DAI PRA [15] gezeigt, kann das Schrödinger-Brücken-Problem als Steuerungsproblem mit minimaler Energie interpretiert werden. Dabei wird anstelle einer rein passiven Betrachtung der Diffusion (4.1) ein Steuerungsterm u_t eingeführt. Die gesteuerte Dynamik ergibt sich dann zu

$$dX_t^u = (\mu(X_t^u, t) + u_t) dt + \sigma(X_t^u, t) dB_t. \quad (4.2)$$

Das Ziel ist es, den Prozess so zu steuern, dass die Randverteilungen μ_0 und μ_T erfüllt werden, während eine Energie- oder Kostenfunktion

$$J(u) = \mathbb{E} \left[\int_0^T \|u_t\|^2 dt \right]$$

minimiert wird. Der optimale Steuerungsterm u_t^* charakterisiert die Schrödinger-Brücke und entspricht der Lösung eines Variationsproblems.

Ein bekannter Zusammenhang zwischen *Driftanpassung* und *Maßänderung* lässt sich im Kontext des Schrödinger-Brücken-Problems durch JAMISON [10] herleiten. Im Wesentlichen zeigt dieser Zusammenhang, dass das Hinzufügen eines Drifts $\nabla \ln h$ zu einer Diffusion dem Ändern des Wahrscheinlichkeitsmaßes durch Multiplikation mit $h(X_t, t)$ (Girsanov-Transformation) entspricht.

Satz 4.3.1 (Driftmodifikation)

Sei X_t eine schwache Lösung der SDGL (4.1) und $h(x, t)$ eine strikt positive Lösung der zugehörigen parabolischen Gleichung (etwa der Wärmeleitungsgleichung), die in $C^{2,1}(\mathbb{R}^n \times [0, T])$ liegt. Es gelte zudem, dass $\mathbb{E}[h(X_t, t)] < \infty$ und $\mathbb{E}[h(X_t, t) | X_s = x] = h(x, s)$ fast sicher für jedes $0 \leq s < t < T$. Dann besitzt die modifizierte SDGL

$$dX_t = (\mu(X_t, t) + a(X_t, t) \nabla \ln h(X_t, t)) dt + \sigma(X_t, t) dB_t \quad (4.3)$$

eine schwache Lösung auf dem Intervall $[0, T - \varepsilon]$ für jedes $\varepsilon > 0$. Darüber hinaus lässt sich die Übergangsdichte eines Prozesses mit dieser Dynamik durch

$$q^h(s, y, t, x) = \frac{q(s, y, t, x) h(x, t)}{h(y, s)} \quad (4.4)$$

bestimmen, wobei $q(s, y, t, x)$ die Übergangsdichte des ursprünglichen Prozesses ist.

Beweis.

Es soll gezeigt werden, dass die Funktion $h(X_t, t)$, die die parabolische Differentialgleichung

$$\frac{\partial h}{\partial t} + \mu \cdot \nabla h + \frac{1}{2} \text{Tr}(\sigma \sigma^\top \nabla^2 h) = 0,$$

mit $\sigma \sigma^\top = a$ erfüllt, die Martingaleigenschaft besitzt und die Übergangsdichte des neuen Prozesses korrekt bestimmt.

Die Dynamik von $h(X_t, t)$ wird mithilfe der Itô-Formel bestimmt. Da X_t durch die stochastische Differentialgleichung (4.1) gegeben ist, ergibt sich für $h(X_t, t)$

$$dh(X_t, t) = \frac{\partial h}{\partial t} dt + (\nabla h)^\top dX_t + \frac{1}{2} \text{Tr}(\sigma \sigma^\top \nabla^2 h) dt.$$

Setzt man $dX_t = \mu(X_t, t) dt + \sigma(X_t, t) dB_t$ ein, folgt

$$dh(X_t, t) = \left[\frac{\partial h}{\partial t} + \mu \cdot \nabla h + \frac{1}{2} \text{Tr}(\sigma \sigma^\top \nabla^2 h) \right] dt + (\nabla h)^\top \sigma dB_t.$$

Da $h(X_t, t)$ die parabolische PDGL erfüllt, verschwindet der Driftterm. Es bleibt

$$dh(X_t, t) = (\nabla h)^\top \sigma dB_t.$$

Der Prozess $Z_t = \frac{h(X_t, t)}{h(X_0, 0)}$ ist daher ein Martingal bezüglich der Filtration $\mathcal{F}_t = \sigma(X_s : 0 \leq s \leq t)$. Da $\mathbb{E}[Z_T] = Z_0 = 1$, kann durch $Z_{T-\varepsilon}$ für ein festes $\varepsilon > 0$ eine neue Wahrscheinlichkeitsverteilung \mathbb{Q} definiert werden

$$\frac{d\mathbb{Q}}{d\mathbb{P}} = Z_{T-\varepsilon}.$$

Für $h(X_t, t)f(X_t)$ gilt wiederum nach der Itô-Formel

$$d(h(X_t, t)f(X_t)) = \left[\frac{\partial h}{\partial t} + \mathcal{L}h \right] f(X_t) dt + h(X_t, t) \mathcal{L}f(X_t) dt + h(X_t, t) \nabla f \cdot \sigma dB_t.$$

Da $h(X_t, t)$ die parabolische PDGL erfüllt, verschwindet der Term $\frac{\partial h}{\partial t} + \mathcal{L}h$. Es folgt

$$d(h(X_t, t)f(X_t)) = h(X_t, t) \mathcal{L}f(X_t) dt + h(X_t, t) \nabla f \cdot \sigma dB_t.$$

Für $0 \leq s < t \leq T - \varepsilon$ und $f \in C_0^\infty(\mathbb{R}^n)$ ergibt sich unter dem neuen Maß \mathbb{Q} daher

$$\mathbb{E}_{\mathbb{Q}}[f(X_t) | X_s = x] - f(x)$$

$$\begin{aligned} &= \frac{1}{h(x, s)} \mathbb{E}[h(X_t, t)f(X_t) - h(x, s)f(x) | X_s = x] \\ &= \frac{1}{h(x, s)} \mathbb{E} \left[\int_s^t (h \mathcal{L}f + a \nabla h \cdot \nabla f)(X_\tau, \tau) d\tau | X_s = x \right] \\ &= \frac{1}{h(x, s)} \mathbb{E} \left[\int_s^t h(X_\tau, \tau) (b(X_\tau, \tau) + a \nabla \ln h(X_\tau, \tau)) \cdot \nabla f(X_\tau, \tau) d\tau | X_s = x \right] \\ &= \frac{1}{h(x, s)} \mathbb{E} \left[\int_s^t h(X_\tau, \tau) \left((b + a \nabla \ln h) \cdot \nabla f + \frac{1}{2} \sum_{i,j=1}^n a_{ij} \frac{\partial^2 f}{\partial x_i \partial x_j} \right) (X_\tau, \tau) d\tau | X_s = x \right] \\ &= \mathbb{E} \left[\int_s^t \left((b + a \nabla \ln h) \cdot \nabla f + \frac{1}{2} \sum_{i,j=1}^n a_{ij} \frac{\partial^2 f}{\partial x_i \partial x_j} \right) (X_\tau, \tau) d\tau | X_s = x \right]. \end{aligned}$$

Das bedeutet, dass die Verteilung des Prozesses X_t , wie sie unter dem geänderten Maß \mathbb{Q} definiert ist, das Problem für den modifizierten Generator

$$\mathcal{L}^h f(x, t) = \left(b(x, t) + a(x, t) \nabla \ln h(x, t) \right) \cdot \nabla f(x, t) + \frac{1}{2} \sum_{i,j=1}^n a_{ij}(x, t) \frac{\partial^2 f}{\partial x_i \partial x_j}(x, t)$$

löst.

Also besitzt die SDGL die ermittelte schwache Lösung auf $[0, T - \varepsilon]$.

Um die Übergangsdichte zu überprüfen, wird angenommen, dass $h(X_t, t) = g(X(T))$ für eine messbare Funktion g . Dies führt zu

$$\frac{d\mathbb{Q}}{d\mathbb{P}} = g(X(T)).$$

Unter Verwendung der Definition der bedingten Erwartung ergibt sich

$$\mathbb{E}_{\mathbb{Q}}[f(X_t) - f(x)] = \mathbb{E}_{\mathbb{Q}} \left[\int_s^t \left((b + a \nabla \ln h) \cdot \nabla f + \frac{1}{2} \sum_{i,j} a_{ij} \frac{\partial^2 f}{\partial x_i \partial x_j} \right)(X_\tau, \tau) d\tau \mid X_s = x \right].$$

Da f kompakten Träger hat und $X_t \rightarrow X_T$, wenn $t \rightarrow T$ fast sicher, kann man das Intervall mit dem Satz über die dominierte Konvergenz auf $[0, T]$ erweitern. Um (4.4) zu zeigen, muss überprüft werden, dass gilt

$$\mathbb{E}^{\mathbb{Q}}[f(X_t) \mid X_s = x] = \int_{\mathbb{R}^n} q^h(s, y, t, x) f(y) dy \quad \text{für alle } f \in C_0^\infty(\mathbb{R}^n).$$

Tatsächlich ergibt sich für alle $f \in C_0^\infty(\mathbb{R}^n)$

$$\mathbb{E}_{\mathbb{Q}}[f(X_t) \mid X_s = x] = \frac{1}{h(x, s)} \mathbb{E}[h(X_t, t) f(X_t) \mid X_s = x],$$

wobei die rechte Seite berechnet werden kann als

$$\frac{1}{h(x, s)} \int_{\mathbb{R}^n} q(s, y, t, x) h(y, t) f(y) dy = \int_{\mathbb{R}^n} q^h(s, y, t, x) f(y) dy.$$

Somit ist die Gleichung gezeigt, wobei $q^h(s, y, t, x)$ durch

$$q^h(s, y, t, x) = \frac{q(s, y, t, x) h(x, t)}{h(y, s)}$$

gegeben ist.

Dies bestätigt die Gültigkeit der Übergangsdichte unter \mathbb{Q} . \square

4.4 Zulässige Steuerungen und Kostenfunktion

Betrachtet wird ein gesteuerter Diffusionsprozess X_t^u , der die SDGL (4.2) in schwacher Form löst. Hierbei bezeichnet u_t die Steuerung.

Definition 4.4.1 (Zulässige Steuerung)

Eine Steuerung u_t wird als zulässig (admissible) bezeichnet, wenn sie die folgenden Bedingungen erfüllt.

- (i) u_t ist $\mathcal{F}_t^{X^u}$ -messbar, wobei $\mathcal{F}_t^{X^u}$ die natürliche Filtration des gesteuerten Prozesses X_t^u ist.
- (ii) Die Differentialgleichung (4.2) besitzt auf dem Intervall $[0, T]$ eine schwache Lösung.
- (iii) Es gilt die Integrabilitätsbedingung:

$$\mathbb{E} \left[\int_0^T \|u_t\|^2 dt \right] < \infty.$$

Gegeben seien zwei Wahrscheinlichkeitsmaße μ_0 und μ_T auf \mathbb{R}^n , die die Anfangs- bzw. Endverteilung des Prozesses beschreiben. Das stochastische Kontrollproblem kann nun wie folgt formuliert werden.

Schrödinger-Brücken-Problem (SBP): Finde eine zulässige Steuerung u_t^* , sodass

$$\begin{cases} \text{Minimiere über } u \in \mathcal{U}, \quad J(u) = \mathbb{E}_P \left[\frac{1}{2} \int_0^T \|u_t\|^2 dt \right], \\ \text{unter der Bedingung } X_0^{u^*} \sim \mu_0 \quad \text{und} \quad X_T^{u^*} \sim \mu_T, \end{cases}$$

wobei \mathcal{U} die Menge aller zulässigen Steuerungen ist.

Dieses Problem ist eine Version des von SCHRÖDINGER [18] formulierten Problems der stochastischen Steuerung. Die Lösung wird mithilfe der logarithmischen Transformation, die in den Arbeiten von FLEMING [7] eingeführt wurde, behandelt. Dabei werden die technischen Voraussetzungen aus Annahmen (4.2.1) und (4.2.2) als erfüllt vorausgesetzt.

Definition 4.4.2 (Zeitentwicklungsoperator für Maße)

Der Operator S_t beschreibt die Entwicklung von Wahrscheinlichkeitsmaßen unter einem Markov-Prozess und wirkt auf der Menge der σ -endlichen Maße auf \mathbb{R}^n . Für ein Maß μ auf \mathbb{R}^n ist $S_t \mu$ definiert durch

$$\frac{d(S_t \mu)}{d\lambda}(x) = \int_{\mathbb{R}^n} q(0, y, t, x) \mu(dy), \quad (4.5)$$

wobei $q(0, y, t, x)$ den Markov-Kern des (ungesteuerten) Prozesses bezeichnet. Die Notation $\frac{d(S_t \mu)}{d\lambda}$ steht für die Radon-Nikodým-Ableitung von $S_t \mu$ bezüglich des Lebesgue-Maßes λ .

Der Operator S_t beschreibt demnach die Verteilung des Prozesses zur Zeit t , wenn μ die Startverteilung des Prozesses ist.

Beispiel 4.4.1 (Zeitentwicklungsoperator S_t)

Um diesen Operator zu verdeutlichen, wird ein eindimensionaler Markov-Prozess $\{X_t\}_{0 \leq t \leq T}$ betrachtet, der eine reine Brownsche Bewegung ohne Drift ist. Ausgehend von einer Startverteilung μ auf \mathbb{R} , sei $X_0 \sim \mu$.

Die Übergangsdichte einer solchen Brownschen Bewegung lautet

$$q(0, y, t, x) = \frac{1}{\sqrt{2\pi t}} \exp\left(-\frac{(x-y)^2}{2t}\right).$$

Gemäß Definition (4.4.2) bestimmt sich dann $S_t\mu$, die Verteilung von X_t , durch

$$(S_t\mu)(B) = \int_{\mathbb{R}} \left(\int_B q(0, y, t, x) \lambda(dx) \right) \mu(dy),$$

wobei $B \subseteq \mathbb{R}$ eine Borelmenge ist und λ das Lebesgue-Maß bezeichnet.

Falls die Startverteilung ein Dirac-Maß δ_0 in $y = 0$ ist, gilt

$$(S_t\delta_0)(B) = \int_B \frac{1}{\sqrt{2\pi t}} \exp\left(-\frac{x^2}{2t}\right) \lambda(dx),$$

d. h. $X_t \sim \mathcal{N}(0, t)$. In diesem Fall ist $\frac{d(S_t\delta_0)}{d\lambda}(x) = \frac{1}{\sqrt{2\pi t}} \exp(-x^2/(2t))$. So zeigt sich anschaulich, dass S_t das Maß δ_0 in eine normalverteilte Dichte zur Zeit t transformiert.

Definition 4.4.3 (Relative Entropie / Kullback-Leibler-Divergenz)

Seien μ und ν zwei Wahrscheinlichkeitsmaße auf demselben messbaren Raum, mit $\mu \ll \nu$. Die relative Entropie (auch Kullback-Leibler-Divergenz) von μ relativ zu ν ist definiert durch

$$\mathcal{H}(\mu | \nu) = \begin{cases} \int \ln\left(\frac{d\mu}{d\nu}\right) d\mu, & \text{falls } \mu \ll \nu, \\ +\infty, & \text{sonst.} \end{cases} \quad (4.6)$$

Bemerkung 4.4.1

Die Kullback-Leibler-Divergenz $\mathcal{H}(\mu | \nu)$ ist im Allgemeinen keine Metrik, da sie weder Symmetrie ($\mathcal{H}(\mu | \nu) \neq \mathcal{H}(\nu | \mu)$) noch die Dreiecksungleichung erfüllt. Dennoch besitzt die Kullback-Leibler-Divergenz fundamentale Bedeutung in der Zeitreihenanalyse und Modellbewertung.

In der Zeitreihenanalyse wird die relative Entropie verwendet, um die Information zu quantifizieren, die ein Modell über die zugrunde liegende Wahrscheinlichkeitsstruktur liefert. Insbesondere wurde von AKAIKE in [1] gezeigt, dass die Kullback-Leibler-Divergenz eine Schlüsselrolle in der Entwicklung des Akaike-Informationskriteriums (AIC) spielt, das zur Modellselektion verwendet wird.

Beispiel 4.4.2 (KL-Divergenz zwischen eindimensionalen Normalverteilungen)
Es seien in der eindimensionalen Situation

$$\mu = \mathcal{N}(m_1, \sigma_1^2) \quad \text{und} \quad \nu = \mathcal{N}(m_2, \sigma_2^2).$$

Diese beiden Wahrscheinlichkeitsmaße besitzen die Dichten

$$f_\mu(x) = \frac{1}{\sqrt{2\pi}\sigma_1} \exp\left(-\frac{(x-m_1)^2}{2\sigma_1^2}\right) \quad \text{bzw.} \quad f_\nu(x) = \frac{1}{\sqrt{2\pi}\sigma_2} \exp\left(-\frac{(x-m_2)^2}{2\sigma_2^2}\right).$$

Der zugehörige Dichtequotient lautet

$$\frac{d\mu}{d\nu}(x) = \frac{f_\mu(x)}{f_\nu(x)} = \frac{\sigma_2}{\sigma_1} \exp\left(-\frac{(x-m_1)^2}{2\sigma_1^2} + \frac{(x-m_2)^2}{2\sigma_2^2}\right).$$

Der Logarithmus davon vereinfacht sich zu

$$\ln\left(\frac{d\mu}{d\nu}(x)\right) = \ln\left(\frac{\sigma_2}{\sigma_1}\right) + \frac{(x-m_2)^2}{2\sigma_2^2} - \frac{(x-m_1)^2}{2\sigma_1^2}.$$

Gemäß (4.6) ist

$$\mathcal{H}(\mu | \nu) = \int_{-\infty}^{+\infty} \ln\left(\frac{d\mu}{d\nu}(x)\right) d\mu(x).$$

Da $d\mu(x) = f_\mu(x)dx$, entsteht ein Gauß-Integral in x und man erhält das bekannte Resultat

$$\mathcal{H}\left(\mathcal{N}(m_1, \sigma_1^2) \middle| \mathcal{N}(m_2, \sigma_2^2)\right) = \ln\left(\frac{\sigma_2}{\sigma_1}\right) + \frac{\sigma_1^2 + (m_2 - m_1)^2}{2\sigma_2^2} - \frac{1}{2}.$$

4.5 Lösung des SBP

Im Folgenden wird davon ausgegangen, dass X_t eine schwache Lösung der SDGL

$$dX_t = \mu dt + \sigma dB_t, \tag{4.7}$$

ist, mit der Anfangsbedingung $X_0 = x_0$ (wobei $x_0 \sim \mu_0$ verteilt ist und $\mathbb{E}[|X_0|^2] < \infty$). Hierbei ist $\{B_t\}_{t \geq 0}$ eine Brownsche Bewegung, $\mu, \sigma \in \mathbb{R}$ und $\sigma > 0$.

Satz 4.5.1 (Lösung von Problem SBP bei deterministischem Anfangswert)

Sei μ_0 das Dirac-Maß, das in einem Punkt $x_0 \in \mathbb{R}^n$ konzentriert ist, und es gelte $\mathcal{H}(\mu_T | S_T \mu_0) < \infty$, wobei $\mathcal{H}(\cdot | \cdot)$ die Kullback-Leibler-Divergenz bezeichnet. Definiere die Funktion $h(x, t)$ durch

$$h(x, t) = \int_{\mathbb{R}^n} q(t, x, T, z) \frac{d\mu_T}{dS_T \mu_0}(z) dz, \tag{4.8}$$

wobei $q(t, x, T, z)$ der Markov-Kern des ungesteuerten Prozesses ist. Dann löst die Steuerung

$$u^*(x, t) = a(x, t) \nabla \ln h(x, t)$$

das Problem SBP, und der zugehörige Kostenwert ist gegeben durch

$$J(u^*) = \mathcal{H}(\mu_T | S_T \mu_0).$$

Beweis.

Es ist zu zeigen, dass die Funktion $h(x, t)$, wie in (4.8) definiert, die erforderlichen Eigenschaften erfüllt. Zunächst gilt

$$h(x_0, 0) = \int_{\mathbb{R}^n} q(0, x_0, T, z) \frac{d\mu_T}{dS_T \mu_0}(z) dz = \int_{\mathbb{R}^n} q(0, x_0, T, z) \frac{d\mu_T}{q(0, x_0, T, z)}(z) dz = 1,$$

da die Übergangsdichte $q(0, x_0, T, z)$ und die Maßänderung zusammen die Verteilung μ_T erzeugen. Aus Annahme (4.2.2) folgt, dass $h(x, t) \in C^{2,1}(\mathbb{R}^n \times [0, T])$ liegt und die parabolische Gleichung

$$\frac{\partial h}{\partial t} + \mathcal{L}h = 0, \quad h(x, T) = \frac{d\mu_T}{dS_T \mu_0}(x)$$

erfüllt, wobei \mathcal{L} der Generator des Diffusionsprozesses ist.

Da $h(x, t)$ die Differentialgleichung löst, ist $h(X_t, t)$ ein Martingal, und es gilt $\mathbb{E}[h(X_t, t)] = 1$. Nach Satz (4.3.1) folgt, dass die Steuerung

$$u^*(x, t) = \sigma^2 \nabla \ln h(x, t)$$

eine schwache Lösung der stochastischen Differentialgleichung

$$dX_t = (\mu + \sigma^2 \nabla \ln h(X_t, t)) dt + \sigma dB_t \quad (4.9)$$

auf dem Intervall $[0, T]$ erzeugt, sodass $X_T \sim \mu_T$.

Sei nun X_t^* eine schwache Lösung von (4.9) und seien $\mathbb{P}_X, \mathbb{P}_{X^*}$ zwei durch X_t und X_t^* induzierte Maße. Aus der Maßänderung in Satz (4.3.1) ergibt sich, dass

$$\frac{d\mathbb{P}^*}{d\mathbb{P}}(\xi(\cdot)) = h(\xi(T), T),$$

wobei \mathbb{P}^* das unter der optimalen Steuerung u^* induzierte Maß bezeichnet. Es gilt

$$\mathbb{E}_{\mathbb{P}^*}[\ln h(X_t, t)] = \mathbb{E}_{\mathbb{P}}[h(X_t, t) \ln h(X_t, t)] \leq \mathbb{E}_{\mathbb{P}}[h(X_T, T) \ln h(X_T, T)] = \mathcal{H}(\mu_T | S_T \mu_0).$$

Da die Funktion $\phi(x) = x \ln x$ konvex und beschränkt ist, folgt, dass $h(X_t, t) \ln h(X_t, t)$ ein Submartingal ist.

Definiere die folgende Folge von Stoppzeiten

$$\tau_n = \inf\{s : |X(s)| > n\} \quad \text{und} \quad \tau_n(\omega) = T, \quad \text{falls } |X_t(\omega)| \leq n \text{ für alle } 0 \leq t \leq T.$$

Für $t \leq T$ und mit Anwendung der Itô-Formel ergibt sich

$$\mathbb{E}[h(X_{t \wedge \tau_n}, t \wedge \tau_n) \ln h(X_{t \wedge \tau_n}, t \wedge \tau_n)] = \mathbb{E}\left[\int_0^{t \wedge \tau_n} \frac{1}{2} \|u^*(X_s, s)\|^2 h(X_s, s) ds\right].$$

Nach dem Optional Sampling Theorem gilt für Submartingale

$$\mathbb{E}[h(X_{t \wedge \tau_n}, t \wedge \tau_n) \ln h(X_{t \wedge \tau_n}, t \wedge \tau_n)] \leq \mathbb{E}[h(X_t, t) \ln h(X_t, t)],$$

wobei $\mathbb{E}[h(X_t, t) \ln h(X_t, t)] < \infty$ nach Annahme.

Für den Grenzübergang $t \wedge \tau_n \uparrow t$ für $n \rightarrow \infty$ gilt

$$\mathbb{E} \left[\int_0^t \frac{1}{2} \|u^*(X_s, s)\|^2 h(X_s, s) ds \right] \leq \mathbb{E}[h(X_t, t) \ln h(X_t, t)].$$

Eine analoge Begründung liefert

$$\mathbb{E} \left[\int_0^t \frac{1}{2} \|u^*(X_s, s)\|^2 h(X_s, s) ds \right] \geq \mathbb{E}[h(X_t, t) \ln h(X_t, t)].$$

Also gilt insgesamt

$$\mathbb{E} \left[\int_0^t \frac{1}{2} \|u^*(X_s, s)\|^2 h(X_s, s) ds \right] = \mathbb{E}[h(X_t, t) \ln h(X_t, t)],$$

Für den Grenzübergang $t \rightarrow T$ wird erneut Fatous Lemma angewendet, zusammen mit der Submartingaleigenschaft von $h(X_t, t) \ln h(X_t, t)$. Daraus folgt

$$\mathcal{H}(\mu_T | S_T \mu_0) = \mathbb{E}[h(X_T, T) \ln h(X_T, T)] = \mathbb{E} \left[\int_0^T \frac{1}{2} \|u^*(X_s, s)\|^2 h(X_s, s) ds \right].$$

Im nächsten Schritt ist zu zeigen, dass für jede zulässige Steuerung u_t gilt:

$$\mathbb{E} \left[\int_0^T \frac{1}{2} \|u_t\|^2 dt \right] \geq \mathcal{H}(\mu_T | S_T \mu_0).$$

Hierzu wird die Girsanov-Transformation verwendet. Sei u_t eine beliebige zulässige Steuerung. Dann folgt mit der Definition von $h(X_T, T)$ und der Jensen-Ungleichung

$$\begin{aligned} 1 &= \mathbb{E}[h(X_T, T)] \\ &= \mathbb{E} \left[h(X_T^u, T) \exp \left(- \int_0^T (\sigma^{-1}(X_t^u, t) u_t)^\top dB_t - \frac{1}{2} \int_0^T \|u_t\|^2 dt \right) \right] \\ &\geq \exp \left(\mathbb{E} \left[\ln h(X_T^u, T) - \int_0^T (\sigma^{-1}(X_t^u, t) u_t)^\top dB_t - \frac{1}{2} \int_0^T \|u_t\|^2 dt \right] \right) \\ &= \exp \left(\mathbb{E}[\ln h(X_T^u, T)] - \frac{1}{2} \mathbb{E} \left[\int_0^T \|u_t\|^2 dt \right] \right) \\ &= \exp \left(\mathcal{H}(\mu_T | S_T \mu_0) - \frac{1}{2} \mathbb{E} \left[\int_0^T \|u_t\|^2 dt \right] \right), \end{aligned}$$

da $X_T^u \sim \mu_T$.

Daraus folgt

$$\mathbb{E} \left[\int_0^T \frac{1}{2} \|u_t\|^2 dt \right] \geq \mathcal{H}(\mu_T | S_T \mu_0).$$

□

Bemerkung 4.5.1

Die Energie $J(u) = \frac{1}{2} \mathbb{E} \left[\int_0^T \|u_t\|^2 dt \right]$ besitzt eine interessante Interpretation im Kontext der Kullback-Leibler-Divergenz. Sei P_x das Maß, das durch den ungesteuerten Prozess X_t auf $C([0, T])$ induziert wird, und P_u das entsprechende Maß für den gesteuerten Prozess X_t^u . Mithilfe der Girsanov-Transformation ergibt sich für die Kullback-Leibler-Divergenz

$$\mathcal{H}(P_u | P_x) = \int \ln \frac{dP_u}{dP_x} dP_u = -\mathbb{E} \left[\int_0^T \sigma^{-1}(X_t^u, t) u_t dB_t - \frac{1}{2} \int_0^T \|u_t\|^2 dt \right].$$

Da das Itô-Integral aufgrund der Martingaleigenschaft verschwindet, reduziert sich dieser Ausdruck auf

$$\mathcal{H}(P_u | P_x) = J(u).$$

Somit erfüllt eine optimale Steuerung die Gleichung

$$\mathcal{H}(P_u | P_x) = \mathcal{H}(\mu_T | S_T \mu_0),$$

das heißt, die globale Kullback-Leibler-Divergenz entspricht der Kullback-Leibler-Divergenz zwischen den finalen Dichten. Diese Beziehung zeigt die zentrale Rolle der relativen Entropie sowohl für die Steuerungskosten als auch für die Maßänderung.

Um ein tieferes Verständnis für die im vorangegangenen Abschnitt behandelten gesteuerten Prozesse zu erlangen, ist es hilfreich, zunächst solche Prozesse zu betrachten, die keiner Steuerung unterliegen. Auf diese Weise wird deutlich, wie sich ein System allein aufgrund seiner stochastischen Dynamik entwickelt, ohne zusätzliche Eingriffe durch einen Steuerterm.

Die folgende Ausführung betrachtet daher einen einfachen *ungesteuerten* Prozess und zeigt, wie dessen Verteilungseigenschaften im zeitlichen Verlauf sind. Die gewonnenen Erkenntnisse bilden anschließend die Grundlage für das Verständnis gesteuerter Prozesse, bei denen ein zusätzlicher Driftterm eingesetzt wird, um den Verlauf der Prozesspfade in Richtung einer gewünschten Verteilung zu lenken.

Beispiel 4.5.1 (Ungesteuerter Prozess)

Betrachtet wird die eindimensionale SDGL

$$dX_t = \sigma dt + \mu dB_t \quad X_0 = x_0,$$

wobei $\sigma, \mu \in \mathbb{R}$ Konstanten sind und $\{B_t\}_{t \geq 0}$ eine Standard-Brownsche Bewegung bezeichnet. Falls die Zufallsvariable X_0 normalverteilt ist mit Mittelwert μ_0 und Varianz σ_0^2 , also

$$X_0 \sim \mathcal{N}(\mu_0, \sigma_0^2),$$

folgen aus den bekannten Eigenschaften linearer Transformationen normalverteilter Zufallsvariablen unmittelbar

$$\mathbb{E}[X_t] = \mathbb{E}[X_0] + \sigma t + \underbrace{\mu \mathbb{E}[B_t]}_{=0} = \mu_0 + \sigma t,$$

und

$$\text{Var}(X_t) = \text{Var}(X_0) + \mu^2 \text{Var}(B_t) = \sigma_0^2 + \mu^2 t.$$

Wegen der Stabilität unter Summation bleibt die Verteilung normalverteilt, das heißt

$$X_t \sim \mathcal{N}(\mu_0 + \sigma t, \sigma_0^2 + \mu^2 t).$$

Die obige Verteilungscharakterisierung kann auch über die Übergangsdichte $q(s, x, t, y)$ des Markov-Prozesses hergeleitet werden. Insbesondere gilt für $0 \leq s < t$

$$X_t = X_s + \sigma(t-s) + \mu(B_t - B_s),$$

so dass sich eine Normalverteilung mit dem Mittelwert $x + \sigma(t-s)$ und der Varianz $\mu^2(t-s)$ ergibt, wenn $X_s = x$ vorgegeben ist. Die entsprechende eindimensionale Übergangsdichte ist

$$q(s, x, t, y) = \frac{1}{\sqrt{2\pi\sigma^2(t-s)}} \exp\left(-\frac{(y-x-\mu(t-s))^2}{2\sigma^2(t-s)}\right).$$

Falls die Startverteilung von X_0 durch die Dichte $\mu_0(x)$ beschrieben wird, ergibt sich die Verteilung von X_t durch Faltung von μ_0 mit der Übergangsdichte

$$q(0, x; t, y) = \frac{1}{\sqrt{2\pi\sigma^2 t}} \exp\left(-\frac{(y-x-\mu t)^2}{2\sigma^2 t}\right).$$

Somit erhält man

$$\int_{\mathbb{R}} q(0, x, t, y) \mu_0(x) dx$$

als Endverteilung von X_t .

Sei nun $\sigma = 1$ und $\mu = 1$, sodass

$$dX_t = dt + dB_t$$

Wird als Startverteilung $X_0 \sim \mathcal{N}(0, 1)$ gewählt, lässt sich analytisch mit der Übergangsdichte

$$q(0, x; t, y) = \frac{1}{\sqrt{2\pi t}} \exp\left(-\frac{(y-x-t)^2}{2t}\right),$$

und der Startdichte $\mu_0(x) = \frac{1}{\sqrt{2\pi}} \exp(-x^2/2)$ die Endverteilung als

$$\int_{\mathbb{R}} q(0, x; t, y) \mu_0(x) dx = \frac{1}{\sqrt{2\pi(1+t)}} \exp\left(-\frac{(y-t)^2}{2(1+t)}\right),$$

bestimmen, was erneut $\mathcal{N}(t, 1+t)$ ergibt. Das Resultat kann entweder durch die in dieser Arbeit angewandte Methode der quadratischen Ergänzung oder alternativ mittels

Fourier-Transformationen hergeleitet werden. Es ist bekannt, dass die Faltung zweier Gauß-Funktionen erneut eine Gauß-artige Funktion ergibt. Diese Eigenschaft gestattet es, in den später analysierten Beispielen unmittelbar zu folgern, dass die optimale Steuerung linear in x sein muss.

Abbildung 9 veranschaulicht das Histogramm der simulierten Endwerte des Prozesses bei $T = 20$. Die empirische Verteilung zeigt eine hohe Übereinstimmung mit der theoretischen Dichte $\mathcal{N}(20, 21)$. Dies unterstreicht, dass die ungesteuerte Dynamik mit konstanter Drift und Diffusionskoeffizient bei einer Startverteilung $X_0 \sim \mathcal{N}(0, 1)$ in die Verteilung $\mathcal{N}(t, 1 + t)$ übergeht.

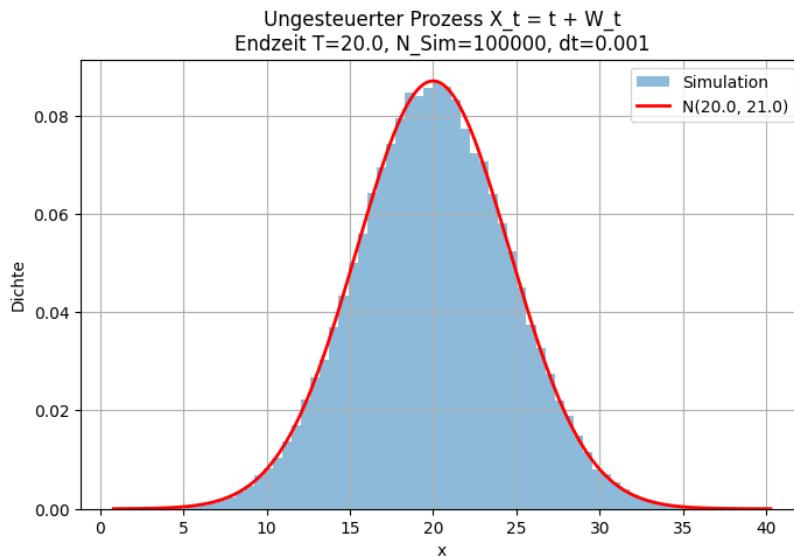


Abbildung 9: Histogramm der simulierten Endwerte bei $T = 20$ des ungesteuerten Prozesses und die theoretische Dichte einer Normalverteilung $\mathcal{N}(20, 21)$.

Beispiel 4.5.2 (Dirac-Startbedingung zur Standardnormalverteilung)

Es wird der stochastische Prozess

$$dX_t = dt + dB_t, \quad X_0 = x_0 \in \mathbb{R},$$

betrachtet, wobei $\{B_t\}_{t \geq 0}$ eine Brownsche Bewegung ist, dessen Übergangsdichte

$$q(y, t, x, s) = \frac{1}{\sqrt{2\pi(t-s)}} \exp\left(-\frac{[x - (y - (t-s))]^2}{2(t-s)}\right),$$

für $t > s$ ist.

Die Anfangsbedingung $X_0 = x_0$ stellt sicher, dass die Verteilung des Prozesses zum Zeitpunkt $t = 0$ eine Dirac-Delta-Verteilung in x_0 ist.

Im Weiteren wird die Funktion

$$h(x, t) = \int_{\mathbb{R}} q(t, x, T, z) \frac{\pi(z)}{q(0, 0, T, z)} dz$$

betrachtet, wobei $\pi(x)$ die Dichtefunktion der Normalverteilung

$$\pi(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right)$$

bezeichnet. Ausgehend von dieser Definition ergibt sich der Ausdruck

$$\frac{\pi(z)}{q(0, x_0, T, z)} = \sqrt{T} \exp\left(-\frac{z^2}{2} + \frac{[z - (x_0 + T)]^2}{2T}\right).$$

Somit lässt sich $h(x, t)$ als ein gefaltetes Integral von $q(t, x, T, z)$ mit dem Faktor $\pi(z)/q(0, x_0, T, z)$ interpretieren. Eingesetzt liefert das

$$h(x, t) = \underbrace{\int \frac{1}{\sqrt{2\pi(T-t)}} \exp\left(-\frac{[z - (x + (T-t))]^2}{2(T-t)}\right) dz}_{q(t, x, T, z)} \times \underbrace{\sqrt{T} \exp\left(-\frac{z^2}{2} + \frac{(z-T)^2}{2T}\right)}_{\pi(z)/q(0, 0, T, z)} dz.$$

Mit Hinblick auf die Berechnung der optimalen Steuerung wird $h(x, t)$ jetzt nur noch bis auf Proportionalität betrachtet. Das liefert

$$h(x, t) \propto \int \exp\left(-\frac{[z - (x + (T-t))]^2}{2(T-t)}\right) \exp\left(-\frac{z^2}{2} + \frac{(z-T)^2}{2T}\right) dz.$$

Durch eine detaillierte Herleitung mittels quadratischer Ergänzung ergibt sich für $h(x, t)$ die Darstellungsform

$$h(x, t) \propto \exp\left(C - \frac{B^2}{4A}\right) \sqrt{\frac{\pi}{-A}},$$

wobei die Konstanten A, B und C durch

$$A = -\frac{1}{2(T-t)} - \frac{1}{2} + \frac{1}{2T} = -\frac{T^2 - Tt + t}{2T(T-t)},$$

$$B = \frac{x + (T-t)}{T-t} - \frac{x_0 + T}{T} = \frac{T(x - x_0) + tx_0}{T(T-t)},$$

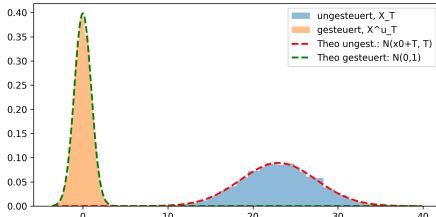
$$C = -\frac{(x + (T-t))^2}{2(T-t)} + \frac{(x_0 + T)^2}{2T} = -\frac{x^2}{2(T-t)} - x - \frac{(T-t)}{2} + \frac{x_0^2}{2T} + x_0 + \frac{T}{2},$$

definiert sind. Insbesondere zeigt sich anhand dieser quadratischen Ergänzung, wie sich $h(x, t)$ bis auf einen konstanten Vorfaktor als Exponentialfunktion eines Quadrates in x ausdrücken lässt, also wieder Gauß-artig ist. Daraus folgt direkt

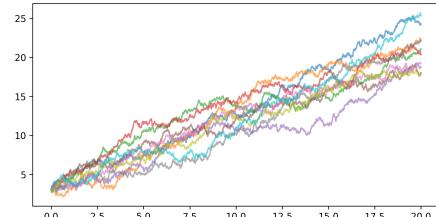
$$u^*(x, t) = \frac{d}{dx} \left(C - \frac{B^2}{4A} \right).$$

Einsetzen liefert dann

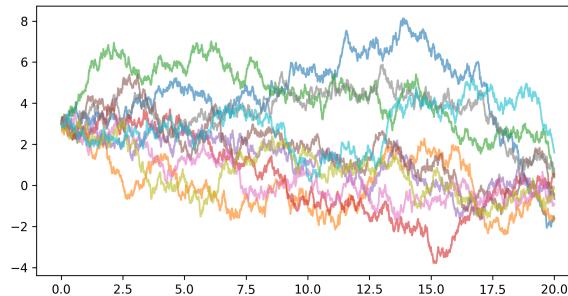
$$u^*(x, t) = -\frac{x + (T-t)}{T-t} - \frac{1}{2A(T-t)} \left[\frac{x + (T-t)}{T-t} - \frac{x_0 + T}{T} \right].$$



(a) Histogramm der Endverteilung



(b) Ungesteuerte Pfade



(c) Gesteuerte Pfade

Abbildung 10: Simulationsergebnisse zur SDGL $dX_t = dt + dB_t$ mit Startwert $x_0 = 3$, $T = 20$ und die Endverteilung ist eine Standardnormalverteilung.

In Abbildung 10a ist das Histogramm der Endverteilung für den ungestörten Fall dargestellt. Im Vergleich zum gesteuerten Fall wird die theoretische Dichtekurve einer Standardnormalverteilung dargestellt. der empirisch gemessenen Verteilung überlagert, um den Steuerungseffekt zu veranschaulichen.

In Abbildung 10b sind mehrere ungestörte Beispieldpfade ausgehend von $x_0 = 3$ zu erkennen, die das Wachstum des Prozesses im Zeitverlauf verdeutlichen. Im Gegensatz dazu zeigt Abbildung 10c wie derselbe Startwert mithilfe der Steuerung zu einer Standardnormalverteilung gelenkt wird.

Beispiel 4.5.3 (Allgemeine Normalverteilung als Zielverteilung)

Es wird das gleiche Setting wie in Beispiel 4.5.2 betrachtet, wobei jedoch die gewünschte Endverteilung nicht standardnormalverteilt ist, sondern eine Normalverteilung mit beliebigem Mittelwert μ und Varianz σ^2 . Die zugehörige Dichtefunktion lautet

$$\pi(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right).$$

Die Herleitung der zugehörigen Steuerung verläuft analog, wobei sich in den Exponenten der quadratischen Ergänzung nun die Parameter μ und σ^2 widerspiegeln. Auf diese Weise entsteht ein im Drift modifizierter Prozess, der in der Endverteilung an $\mathcal{N}(\mu, \sigma^2)$ angepasst wird.

Im betrachteten Setting ist erneut

$$h(x, t) \propto \int_{-\infty}^{\infty} \exp\left(-\frac{[z - (x + T - t)]^2}{2(T-t)} - \frac{(z - \mu)^2}{2\sigma^2} + \frac{[z - (x_0 + T)]^2}{2T}\right) dz.$$

Nach quadratischer Ergänzung ergibt sich bis auf Proportionalität

$$h(x, t) \propto \exp\left(C - \frac{B^2}{4A}\right) \int_{-\infty}^{\infty} \exp\left(A\left(z + \frac{B}{2A}\right)^2\right) dz,$$

wobei die Konstanten durch

$$\begin{aligned} A &= -\frac{1}{2(T-t)} - \frac{1}{2\sigma^2} + \frac{1}{2T}, \\ B &= \frac{x+T-t}{T-t} + \frac{\mu}{\sigma^2} - \frac{x_0+T}{T}, \\ C &= -\frac{(x+T-t)^2}{2(T-t)} - \frac{\mu^2}{2\sigma^2} + \frac{(x_0+T)^2}{2T} \end{aligned}$$

definiert sind. Insgesamt resultiert damit für die optimale Steuerung

$$u^*(x, t) = -\frac{x+(T-t)}{T-t} - \frac{1}{2(T-t)A} \left[\frac{x+(T-t)}{T-t} + \frac{\mu}{\sigma^2} - \frac{x_0+T}{T} \right].$$

Abbildung 11a zeigt das Histogramm der Endverteilung für den ungesteuerten bzw. gesteuerten Prozess. Theoretisch ergibt sich für den ungesteuerten Fall eine Verteilung $\mathcal{N}(x_0 + T, T)$, also $\mathcal{N}(8, 5)$ bei $x_0 = 3$ und $T = 5$. Im gesteuerten Fall wurde hier eine $\mathcal{N}(2, 1.5^2)$ -Endverteilung angestrebt.

Die ungestörten Pfade in Abbildung 11b verdeutlichen das Wachstum des Prozesses von $x_0 = 3$ bis zum Zeitpunkt $T = 5$. Die gesteuerten Pfade in Abbildung 11c streuen deutlich weniger und liegen näher an der gewünschten $\mathcal{N}(2, 1.5^2)$ -Verteilung.

Satz 4.5.2 (Lösung des Problems SBP für allgemeine Anfangsbedingungen)
Sei μ_0 ein Wahrscheinlichkeitsmaß auf \mathbb{R}^n , für das gilt:

$$\int_{\mathbb{R}^n} x^2 d\mu_0 < \infty, \quad \mathcal{H}(\mu_T | S_T \nu_0) < \infty \quad \text{und} \quad \int_{\mathbb{R}^n} \left(\frac{d\mu_0}{d\nu_0} \right) d\mu_0 < +\infty.$$

Sei außerdem die Funktion $h(x, t)$ definiert durch

$$h(x, t) = \int_{\mathbb{R}^n} q(t, x, T, z) \rho_T(z) dz, \tag{4.10}$$

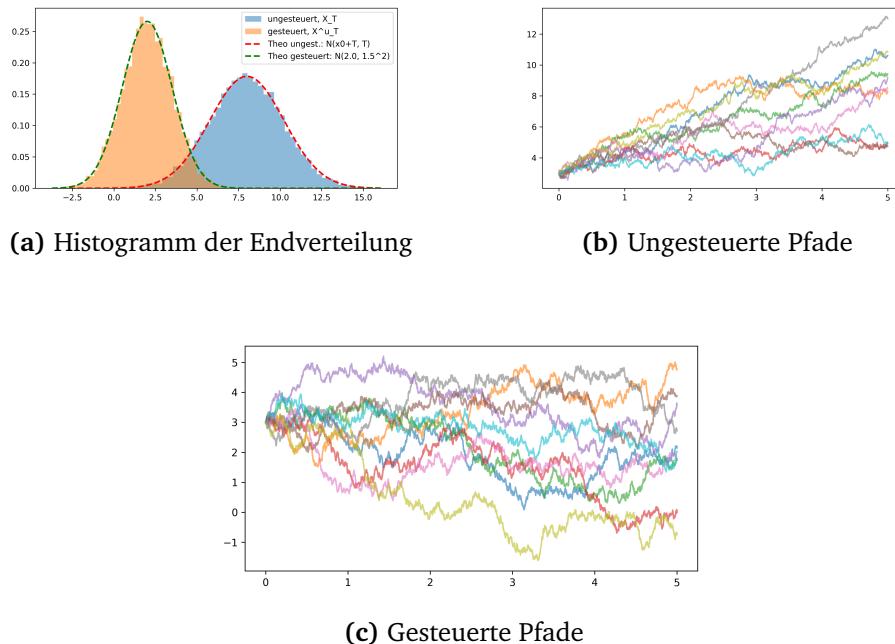


Abbildung 11: Simulationsergebnisse zur SDGL $dX_t = dt + dB_t$ mit Startwert $X_0 = 3$ und Endzeit $T = 5$.

wobei $q(t, x, T, z)$ der Übergangskern des zugrundeliegenden Markov-Prozesses und ρ_T die Dichte des Maßes ν_T ist.

Dann ist die Steuerung

$$u^*(x,t) = a(x,t) \nabla \ln h(x,t)$$

eine Lösung des Problems SBP, und der zugehörige Kostenwert ist gegeben durch

$$\mathbb{E} \left[\frac{1}{2} \int_0^T \|u^*(t)\|_{a^{-1}}^2 dt \right] = \mathcal{H}(\mu_T | S_T \nu_0) - \mathcal{H}(\mu_0 | \nu_0). \quad (4.11)$$

Beweis.

Zunächst ist zu zeigen, dass $h(X_t, t) = \mathbb{E}[\rho_T(X_T)|X_t]$. Nach Gleichung 4.10 ist dies erfüllt, wenn gilt $\mathbb{E}_{\rho_T}[X_T] < +\infty$. Tatsächlich ergibt sich

$$\begin{aligned} \mathbb{E}_{\rho_T}[X_T] &= \int \rho_T(x) dS_T \mu_0(x) \\ &= \int \rho_T(x) \left(\int q(0, y, T, x) d\mu_0(y) \right) dx \\ &= \int \left(\int q(0, y, T, x) \rho_T(x) dx \right) d\mu_0(y) = \int \frac{d\mu_0}{d\nu_0} d\mu_0 < +\infty. \end{aligned}$$

Nach Satz 4.3.1 besitzt (4.9) eine schwache Lösung in $[0, T]$. Der Beweis des Satzes wird analog zu Satz 4.5.1 geführt, wobei das Radon-Nikodym-Derivat berücksichtigt wird

$$\frac{d\mathbb{P}_{X^*}}{d\mathbb{P}_X}(\xi_{\cdot}) = \frac{\rho_T(\xi_T)}{h(\xi_0, 0)}.$$

Es ergibt sich für die bedingten Erwartungen

$$\begin{aligned}\mathbb{E}[\ln h^*(X_t, t)] &= \mathbb{E}\left[\frac{1}{h(X_0, 0)} \mathbb{E}[h(X_t, t) \ln h(X_t, t) | X_0]\right] \\ &\leq \mathbb{E}\left\{\frac{\rho_T(X_T)}{h(X_0, 0)} \ln h(X_T, T)\right\} \\ &= \mathbb{E}[\ln h(X_T, T)] = \mathcal{H}(\mu_T | S_T \nu_0).\end{aligned}$$

Analog zum Beweis von Satz 4.3.1 ergibt sich

$$\begin{aligned}\mathbb{E}\left\{\frac{1}{h(X_0, 0)} \mathbb{E}[h(X_{t \wedge \tau_n}, t \wedge \tau_n) \ln h(X_{t \wedge \tau_n}, t \wedge \tau_n) | X_0]\right\} - \mathbb{E}[\ln h(X_0, 0)] \\ = \mathbb{E}\left\{\frac{1}{h(X_0, 0)} \int_0^{t \wedge \tau_n} \frac{1}{2} \|u_s^*\|_{a^{-1}}^2 h(X_s, s) ds\right\}.\end{aligned}$$

Da $\mathbb{E}[\ln h(X_0, 0)] = \int \ln \frac{d\mu_0}{d\nu_0} d\mu_0 = \mathcal{H}(\mu_0 | \nu_0) < +\infty$, folgt aus der Jensen-Ungleichung

$$0 \leq \int \ln \frac{d\mu_0}{d\nu_0} d\mu_0 \leq \int \frac{d\mu_0}{d\nu_0} d\mu_0 < +\infty.$$

Mit Grenzübergang, wie in Satz 4.5.1, ergibt sich

$$\mathbb{E} \frac{1}{2} \int_0^T \|u_t^*\|_{a^{-1}}^2 dt = \mathcal{H}(\mu_T | S_T \nu_0) - \mathcal{H}(\mu_0 | \nu_0).$$

Da $\mathbb{E} \frac{1}{2} \int_0^T \|u_t\|_{a^{-1}}^2 dt \geq \mathcal{H}(\mu_T | S_T \nu_0) - \mathcal{H}(\mu_0 | \nu_0)$ für alle zulässigen Steuerungen u gilt, folgt die Gleichheit durch den Satz von Girsanov. \square

Beispiel 4.5.4 (Ornstein-Uhlenbeck-Prozess nach Dai Pra [15])

Es wird der eindimensionale Ornstein-Uhlenbeck-Prozess

$$dX_t = AX_t dt + dB_t$$

betrachtet, wobei $A \neq 0$, als Anfangsbedingung $X_0 = 0$ gilt und $\{B_t\}_{t \geq 0}$ eine Brownsche Bewegung ist. Die Endverteilung soll eine Standardnormalverteilung sein.

In Dai Pra [15] wird dieser Prozess im Rahmen des Schrödinger-Brücken-Problems untersucht.

Für $t \geq 0$ gilt

$$X_t = e^{At} X_0 + \int_0^t e^{A(t-s)} dB_s = \int_0^t e^{A(t-s)} dB_s$$

Für zwei Zeitpunkte $0 \leq t < T$ gilt

$$X_T = e^{A(T-t)} X_t + \int_t^T e^{A(T-s)} dB_s.$$

Der stochastische Anteil $\int_t^T e^{A(T-s)} dB_s$ ist normalverteilt mit Mittel 0 und Varianz

$$\sigma_{T-t}^2 = \int_t^T (e^{A(T-s)})^2 ds = \int_0^{T-t} e^{2Au} du,$$

wobei die Substitution $u = T - s$ erfolgte. Daraus folgt

$$\sigma_{T-t}^2 = \begin{cases} \frac{1}{2A} (e^{2A(T-t)} - 1), & \text{falls } A \neq 0, \\ T-t, & \text{falls } A = 0 \text{ (kein Drift).} \end{cases}$$

Damit lautet die Übergangsdichte

$$q(t, x; T, z) = \frac{1}{\sqrt{2\pi\sigma_{T-t}^2}} \exp\left(-\frac{1}{2\sigma_{T-t}^2} [z - e^{A(T-t)} x]^2\right),$$

$$\text{mit } \sigma_t^2 = \frac{1}{2A} (e^{2At} - 1).$$

Definiere wieder

$$h(x, t) = \int_{-\infty}^{\infty} q(t, x; T, z) \frac{\pi(z)}{q(0, 0; T, z)} dz.$$

Hierbei ist

$$q(0, 0; T, z) = \frac{1}{\sqrt{2\pi\sigma_T^2}} \exp\left(-\frac{1}{2\sigma_T^2} [z - e^{AT} \cdot 0]^2\right) = \frac{1}{\sqrt{2\pi\sigma_T^2}} \exp\left(-\frac{z^2}{2\sigma_T^2}\right).$$

Es gilt

$$\frac{\pi(z)}{q(0, 0; T, z)} = \frac{\frac{1}{\sqrt{2\pi}} \exp(-z^2/2)}{\frac{1}{\sqrt{2\pi\sigma_T^2}} \exp\left(-\frac{z^2}{2\sigma_T^2}\right)} = \sqrt{\sigma_T^2} \exp\left(-\frac{z^2}{2} + \frac{z^2}{2\sigma_T^2}\right).$$

Setzt man dies ein, erhält man

$$h(x, t) = \underbrace{\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma_{T-t}^2}} \exp\left(-\frac{[z - e^{A(T-t)} x]^2}{2\sigma_{T-t}^2}\right) dz}_{q(t, x; T, z)} \underbrace{\sqrt{\sigma_T^2} \exp\left(-\frac{z^2}{2} + \frac{z^2}{2\sigma_T^2}\right)}_{\frac{\pi(z)}{q(0, 0; T, z)}} dz.$$

Bis auf Proportionalität, kann man einen konstanten Faktor wie $\sqrt{\sigma_T^2}$ vernachlässigen

$$h(x, t) \propto \int_{-\infty}^{\infty} \exp\left(-\frac{[z - e^{A(T-t)} x]^2}{2\sigma_{T-t}^2}\right) \exp\left(-\frac{z^2}{2} + \frac{z^2}{2\sigma_T^2}\right) dz.$$

Dies kann man analog zu vorherigen Beispielen mittels quadratischer Ergänzung weiter vereinfachen und es folgt

$$h(x, t) \propto \exp\left(\gamma - \frac{\beta^2}{4\alpha}\right) \sqrt{\frac{\pi}{-\alpha}} \quad (\text{für } \alpha < 0),$$

wo

$$\alpha = -\frac{1}{2\sigma_{T-t}^2} + \frac{1}{2\sigma_T^2} - \frac{1}{2}, \quad \beta = \frac{e^{A(T-t)}x}{\sigma_{T-t}^2}, \quad \gamma = -\frac{e^{2A(T-t)}x^2}{2\sigma_{T-t}^2}.$$

Nach bilden des Logarithmus und ableiten nach x ergibt sich

$$u^*(x, t) = \frac{d}{dx} \ln h(x, t) = \frac{e^{2A(T-t)}[(1 - \sigma_T)x]}{e^{2A(T-t)}\sigma_t + \sigma_{T-t}\sigma T}$$

Abbildung 12a zeigt das Histogramm der Endverteilungen für den ungesteuerten bzw. gesteuerten Ornstein–Uhlenbeck-Prozess. Für den ungesteuerten Fall ergibt sich empirisch $\mathbb{E}[X_T] \approx 0.010$ und $\text{Var}(X_T) \approx 2.074$, während theoretisch $X_T \sim \mathcal{N}(0, 2.020)$ angenommen wird. Bei der gesteuerten Variante zeigen die Simulationen $\mathbb{E}[X_T] \approx 9.990$ und $\text{Var}(X_T) \approx 1.991$, was sehr gut zur angestrebten Verteilung $X_T \sim \mathcal{N}(10, 2)$ passt.

Die ungesteuerten Pfade in Abbildung 12b verdeutlichen die typische Entwicklung des Ornstein–Uhlenbeck-Prozesses von seinem jeweiligen Startwert bis zum Endzeitpunkt T . Die gesteuerten Pfade in Abbildung 12c streuen deutlich weniger, da sie durch die Steuerung u^* in Richtung der gewünschten Normalverteilung $\mathcal{N}(10, 2)$ korrigiert werden und so näher an der theoretischen Zielverteilung bleiben.

Bemerkung 4.5.2 (Normalverteilte Startbedingung)

Im vorangehenden Beispiel wurde eine Dirac-Startbedingung δ_{x_0} (deterministisch in x_0) verwendet, so dass

$$\frac{\pi(z)}{q(0, x_0; T, z)}$$

in das Integral $\int_{-\infty}^{\infty} q(t, x; T, z) \frac{\pi(z)}{q(0, x_0; T, z)} dz$ einfließt.

Ist dagegen zum Zeitpunkt $t = 0$ nicht ein fester Zustand x_0 vorgegeben, sondern eine beliebige Startverteilung ν_0 (z. B. eine Normalverteilung $\mathcal{N}(\mu_0, \sigma_0^2)$), so tritt anstelle von $q(0, x_0; T, z)$ die Faltung

$$q(0, \nu_0; T, z) = \int q(0, x_0; T, z) \nu_0(dx_0).$$

$h(x, t)$ erhält dann in analog in der Form

$$h(x, t) = \int q(t, x; T, z) \frac{\pi(z)}{q(0, \nu_0; T, z)} dz,$$

wobei $q(0, \nu_0; T, z)$ genau jene Faltung aus $q(0, x_0; T, z)$ und ν_0 ist.

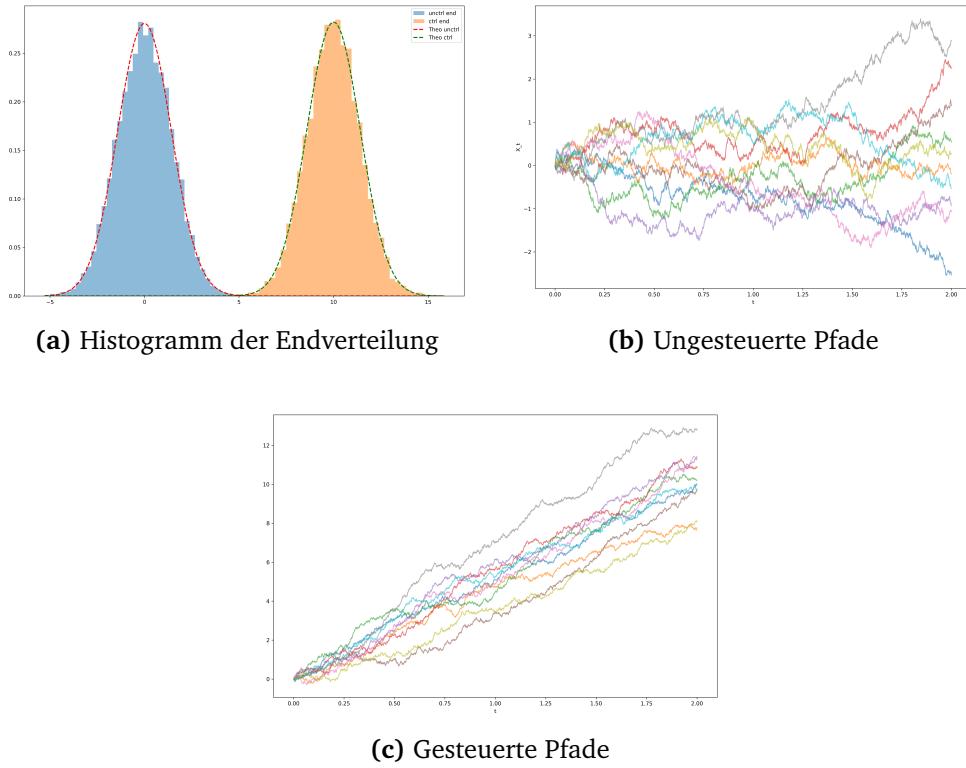


Abbildung 12: Simulationsergebnisse zur SDGL $dX_t = AX_t dt + dB_t$ mit Startwert $X_0 = 0$ und Endzeit $T = 2$.

Falls der Prozess wie oben definiert ist und die Startverteilung ν_0 eine Normalverteilung $\mathcal{N}(m_0, \sigma_0^2)$ ist, so ergibt sich

$$q(0, \nu_0; T, z) = \int_{\mathbb{R}} \left[\frac{1}{\sqrt{2\pi T}} \exp\left(-\frac{[z-(x_0+T)]^2}{2T}\right) \right] \left[\frac{1}{\sqrt{2\pi \sigma_0^2}} \exp\left(-\frac{(x_0-m_0)^2}{2\sigma_0^2}\right) \right] dx_0.$$

In einer Vielzahl dieser Berechnungen, insbesondere unter der Voraussetzung, dass Normalverteilungen und exponentielle Faktoren eine Rolle spielen, kann die Identität

$$C_2(a) \int_{\mathbb{R}} \varphi_{\mu_1, \sigma_1}(y-x) \exp(ax - C_1(a)y) \varphi_{\mu_2, \sigma_2}(x) dx = \varphi_{\mu_1 + \mu_2, \sqrt{\sigma_1^2 + \sigma_2^2}}(y)$$

nutzen. Dabei bezeichnet $\varphi_{\mu, \sigma}$ die Gaußsche Dichte mit Mittelwert μ und Standardabweichung σ ; $C_1(a), C_2(a)$ sind geeignete Korrektur-Faktoren in a und y , sodass am Ende eine reine Gauß-Funktion (in y) übrig bleibt.

Anstelle der manuellen Durchführung einer ausführlichen quadratischen Ergänzung (oder Fourier-Transformation) kann die obige Formel direkt angewendet werden. Es wird sofort ersichtlich, dass

$$q(0, \nu_0; T, z) = \varphi_{m_0 + T, \sqrt{\sigma_0^2 + T}}(z).$$

Beispiel 4.5.5 (Normalverteilte Anfangs- und Zielverteilung)

Es gilt erneut die SDGL

$$dX_t = dt + dB_t.$$

Außerdem sei

$$X_0 \sim \mathcal{N}(\mu_0, \sigma_0^2).$$

Es soll wieder eine $\mathcal{N}(\mu_1, \sigma_1^2)$ Endverteilung erreicht werden. Aus Beispiel 4.5.1 ergibt sich die Übergangsdichte und es gilt

$$q(0, x_0; T, z) = \frac{1}{\sqrt{2\pi T}} \exp\left(-\frac{[z - (x_0 + T)]^2}{2T}\right).$$

Daraus ergibt sich

$$\begin{aligned} q(0, X_0; T, z) &= \int_{-\infty}^{\infty} q(0, x_0; T, z) \nu_0(dx_0) \\ &= \int_{-\infty}^{\infty} \left[\frac{1}{\sqrt{2\pi T}} \exp\left(-\frac{(z - (x_0 + T))^2}{2T}\right) \right] \times \left[\frac{1}{\sqrt{2\pi \sigma_0^2}} \exp\left(-\frac{(x_0 - \mu_0)^2}{2\sigma_0^2}\right) \right] dx_0. \end{aligned}$$

Das ist wieder einer Faltung von zwei Gauß-artigen Faktoren und es ergibt sich $X_T \sim \mathcal{N}(\mu_0 + T, \sigma_0^2 + T)$, d. h.

$$q(0, X_0; T, z) = \frac{1}{\sqrt{2\pi(\sigma_0^2 + T)}} \exp\left(-\frac{[z - (\mu_0 + T)]^2}{2(\sigma_0^2 + T)}\right).$$

Daraus folgt wieder der Quotient

$$\frac{\pi(z)}{q(0, X_0; T, z)} = \sqrt{\frac{\sigma_0^2 + T}{\sigma_1^2}} \exp\left(-\frac{(z - \mu_1)^2}{2\sigma_1^2} + \frac{[z - (\mu_0 + T)]^2}{2(\sigma_0^2 + T)}\right).$$

Einsetzen liefert auch hier wieder eine ähnliche Form für

$$h(x, t) \propto \int_{-\infty}^{\infty} \exp\left(-\frac{[z - (x + (T-t))]^2}{2(T-t)} - \frac{(z - \mu_1)^2}{2\sigma_1^2} + \frac{[z - (\mu_0 + T)]^2}{2(\sigma_0^2 + T)}\right) dz.$$

Mit quadratischer Ergänzung ergeben sich wieder A, B und C

$$A = -\frac{1}{2(T-t)} - \frac{1}{2\sigma_1^2} + \frac{1}{2(\sigma_0^2 + T)}.$$

$$B = +\frac{x + (T-t)}{T-t} + \frac{\mu_1}{\sigma_1^2} - \frac{\mu_0 + T}{\sigma_0^2 + T}.$$

$$C = -\frac{[x + (T-t)]^2}{2(T-t)} - \frac{\mu_1^2}{2\sigma_1^2} + \frac{(\mu_0 + T)^2}{2(\sigma_0^2 + T)}.$$

Analog zu den vorherigen Beispielen folgt nun wieder

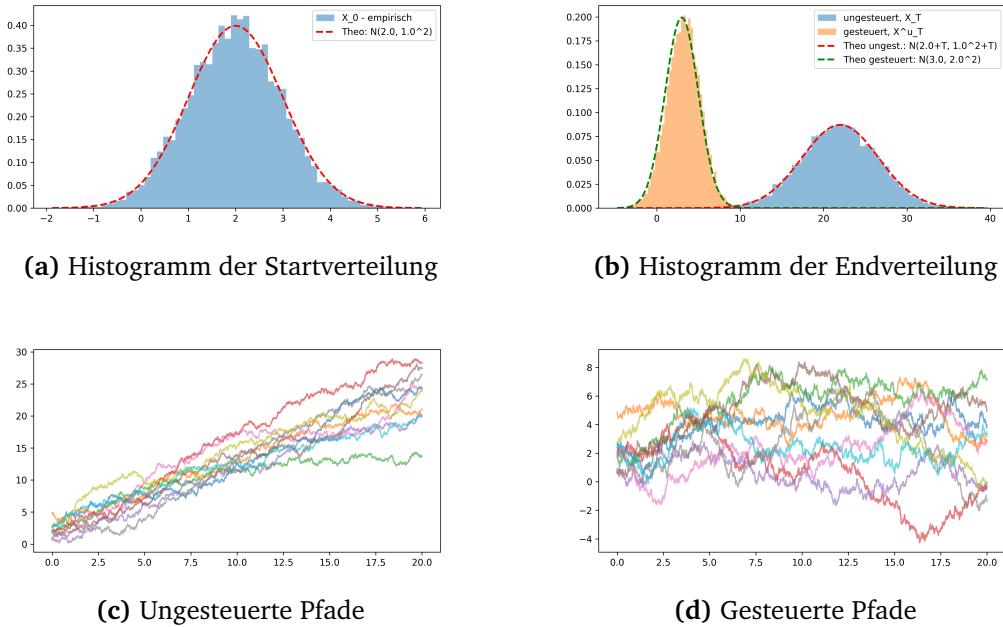


Abbildung 13: Simulationsergebnisse zur SDGL $dX_t = dt + dB_t$ bei normalverteilter Startverteilung $X_0 \sim \mathcal{N}(2, 1^2)$.

$$u^*(x, t) = -\frac{x + (T - t)}{T - t} - \frac{1}{2(T - t)A} \left[\frac{x + (T - t)}{T - t} + \frac{\mu_1}{\sigma_1^2} - \frac{\mu_0 + T}{\sigma_0^2 + T} \right].$$

Abbildung 13a illustriert die empirische Startverteilung, die theoretisch $\mathcal{N}(2, 1)$ ist. Die Endverteilung ohne Steuerung ist in Abbildung 13b zu sehen und sollte theoretisch $\mathcal{N}(22, 20)$ entsprechen. Empirisch ergibt sich dafür $\mathbb{E}[X_T] \approx 21.969$ und $\text{Var}(X_T) \approx 21.455$, was in guter Übereinstimmung liegt. Abbildung 13c verdeutlicht mehrere realisierte Pfade des ungesteuerten Prozesses, der im Mittel wächst und streut. In Abbildung 13d wird eine Steuerung angewandt, um den Prozess an die Zielverteilung $\mathcal{N}(3, 2^2)$ heranzuführen. Die empirischen Kennzahlen für diesen gesteuerten Prozess liegen bei $\mathbb{E}[X_T^u] \approx 3.171$ und $\text{Var}(X_T^u) \approx 3.881$, was die Annäherung an $\mathcal{N}(3, 4)$ belegt.

Bemerkung 4.5.3 (Zeitoptimierung per Kullback-Leibler-Divergenz)

Eine interessante Fragestellung betrifft die Wahl des Endzeitpunkts T . Betrachtet man die Kullback-Leibler-Divergenz

$$\mathcal{H}(\pi(x) | q(0, X_0; T, x))$$

zwischen einer gewünschten Endverteilung $\pi(x)$ und der ungesteuerten Übergangsdichte $q(0, X_0; T, x)$, so lässt sich diese Größe als Funktion von T auffassen. Durch

$$T^* = \arg \min_T \mathcal{H}(\pi(x) | q(0, X_0; T, x))$$

erhält man den optimalen Zeitpunkt T^* , an dem die ungesteuerte Entwicklung des Prozesses am nächsten an π liegt, gemessen in Bezug auf die relative Entropie. In Anwendungen kann dies von Bedeutung sein, um abzuschätzen, wann sich ein System ohne

Eingriff "spontan" in einer Verteilung nahe π befindet, sodass der Steuerungsaufwand beim Übergang zu π minimal wird.

In Beispiel 4.5.5 führt das zu

$$\mathcal{H}\left(\mathcal{N}(\mu, \sigma^2) \mid \mathcal{N}(\mu_0 + T, \sigma_0^2 + T)\right) = \underbrace{\ln\left(\sqrt{\frac{\sigma_0^2 + T}{\sigma^2}}\right) + \frac{\sigma^2 + ((\mu_0 + T) - \mu)^2}{2(\sigma_0^2 + T)} - \frac{1}{2}}_{=: \Psi_T}.$$

Dann ist die Ableitung

$$\Psi'_T = \frac{1}{2(\sigma_0^2 + T)} + \frac{2((\mu_0 - \mu) + T)(\sigma_0^2 + T) - [\sigma^2 + ((\mu_0 - \mu) + T)^2]}{2(\sigma_0^2 + T)^2}$$

Für $\Psi'_T = 0$ folgt mit quadratischer Ergänzung

$$T^* = \frac{-\left(1 + 2\sigma_0^2\right) + \sqrt{\Delta}}{2},$$

wobei

$$\Delta = \left(1 + 2\sigma_0^2\right)^2 - 4\left[\sigma_0^2(1 + 2(\mu_0 - \mu)) - \sigma^2 - (\mu_0 - \mu)^2\right].$$

Daher existiert ein eindeutiges $T^* \in [0, \infty)$, bei dem Ψ_T ein globales Minimum annimmt.

Der Wert T^* ist der Zeitpunkt, zu dem die ungesteuerte Normalverteilung $\mathcal{N}(\mu_0 + T, \sigma_0^2 + T)$ der Zielverteilung $\mathcal{N}(\mu, \sigma^2)$ am nächsten kommt, gemessen an der Kullback-Leibler-Divergenz. Das Minimum kann aber auch schon früher erreicht werden, denn falls z.B. $\sigma_0^2 = \sigma^2$ und $\mu_0 = \mu$ gelten, ist schon bei $t = 0$ die Endverteilung erreicht.

Dieses Kapitel hat gezeigt, wie das Schrödinger-Brücken-Problem (SBP) eine stochastische Dynamik durch minimale Eingriffe (gemessen an der Energie bzw. KL-Divergenz) von einem Referenzprozess zu einem Zielprozess transformieren kann. Über die logarithmische Transformation wird das SBP dabei auf ein entropieminimierendes Steuerungsproblem zurückgeführt.

Insbesondere illustrieren die vorgestellten Beispiele, von konstanten Drifts bis hin zum Ornstein–Uhlenbeck-Prozess, wie sich durch die Wahl eines geeigneten $\nabla \ln h$ die Pfade hin zu einer gewünschten Endverteilung steuern lassen. Die Kullback-Leibler-Divergenz spielt dabei die Rolle eines Maßes für die "Nähe" zwischen gesteuertem und ungesteuertem Prozess.

Im folgenden Kapitel werden diese Konzepte auf *Zeitreihen* übertragen. Dies erweitert den SBP-Ansatz auf praktische Anwendungen, bei denen die stochastische Struktur ganzer Datenreihen beeinflusst wird, um analog zu den hier behandelten Beispielen gewisse Anfangs- und Endverteilungen zu erfüllen.

5 SBP für Zeitreihen

5.1 Motivation

Zeitreihen werden in vielen gesellschaftlichen und industriellen Kontexten verwendet, von Finanzmärkten über Wettervorhersagen bis hin zu medizinischen Prognosen. In der Finanzindustrie beispielsweise werden Zeitreihenmodelle verwendet, um Szenarien für Marktstresstests zu simulieren, Risiken zu messen und dynamische Hedging-Strategien zu entwickeln. Die Modellierung solcher Prozesse ist jedoch eine anspruchsvolle Aufgabe, die häufig eine komplexe Kalibrierung und den Umgang mit Modellunsicherheiten erfordert. Diese Herausforderungen haben in den letzten Jahren das Interesse an der Generierung synthetischer Zeitreihen geweckt, um datengetriebene Ansätze in Bereichen wie dem Risikomanagement zu ermöglichen.

HAMDOUCHE ET AL. [9] erweitern diesen Ansatz durch die Einführung eines generativen Modells für Zeitreihen, das auf der Schrödinger-Brückentheorie basiert (SBTS). Die Methode beschreibt die entropische Interpolation zwischen einem Referenzmaß im Pfadraum und der Verteilung der Zeitreihendaten. Ein zentrales Merkmal des Ansatzes ist die dynamische Anpassung der Schrödinger-Brücke, die sowohl die Zeitabhängigkeit als auch die Pfadabhängigkeit der Driftfunktion berücksichtigt.

Im Gegensatz zu bestehenden Ansätzen, die sich primär auf statische Verteilungen oder marginale Abhängigkeiten konzentrieren, zielt der Ansatz von HAMDOUCHE darauf ab, die zeitlichen und korrelativen Strukturen von Zeitreihen präzise abzubilden. Neben theoretischen Beiträgen bietet die Methode auch praktische Vorteile, da die Simulation synthetischer Zeitreihen durch effiziente numerische Algorithmen ermöglicht wird. Anwendungen reichen von Spielzeugmodellen wie autoregressiven Prozessen bis hin zur Generierung realer Datensätze für Aufgaben wie Deep Hedging.

Dieses Kapitel widmet sich der detaillierten Darstellung des von HAMDOUCHE ET AL. vorgestellten Modells, wobei insbesondere der Bezug zu klassischen Schrödinger-Brücken-Problemen hergestellt und die zentralen mathematischen Ergebnisse in den Kontext der Zeitreihenmodellierung eingebettet werden. Ziel ist es, die Methodik und ihre Anwendungen schrittweise zu verstehen und die theoretischen Grundlagen mit numerischen Beispielen zu illustrieren. Dabei werden die Ergebnisse aus Hamdouce et al. wiedergegeben und genauer ausgeführt.

5.2 Problemformulierung

Eine Zeitreihe, die die Entwicklung eines Prozesses in \mathbb{R}^d beschreibt, wie beispielsweise eines Aktienkurses, eines Schadensmodells oder von Audiodaten, lässt sich durch eine Verteilung μ charakterisieren. Diese Verteilung gehört zur Menge der Wahrscheinlichkeitsmaße $\mathcal{P}((\mathbb{R}^d)^N)$, welche auf $(\mathbb{R}^d)^N$ definiert ist. Die Beobachtungspunkte dieser Zeitreihe sind diskrete Zeitmomente $T = \{t_1, t_2, \dots, t_N\}$, die sich

innerhalb des Intervalls $(0, \infty)$ befinden, wobei t_N den zeitlichen Endpunkt der Beobachtung markiert.

Das Ziel besteht darin, ein probabilistisches Modell zu entwickeln, das Stichproben generieren kann, welche der Zielverteilung μ möglichst nahekommen. Um dieses Problem zu lösen, wird auf einen dynamischen Ansatz zurückgegriffen, der auf der Schrödinger-Brücke basiert, wie in verschiedenen Arbeiten zu stochastischen Prozessen beschrieben (siehe z. B. [5, 9]). Der vorliegende Ansatz integriert die Prinzipien der stochastischen Optimierung mit Anforderungen an die Konsistenz der Daten.

Der zugrundeliegende Prozess ist definiert auf dem Raum $\Omega = C([0, T]; \mathbb{R}^d)$, der die Menge aller stetigen Funktionen von $[0, T]$ nach \mathbb{R}^d darstellt. Der Prozess $X = \{X_t\}_{t \in [0, T]}$ wird dabei als der kanonische Prozess angesehen, welcher bei $X_0 = 0$ startet. Die natürliche Filtration dieses Prozesses wird durch $\mathcal{F} = \{\mathcal{F}_t\}_{t \in [0, T]}$ bezeichnet.

Es soll ein Wahrscheinlichkeitsmaß $\mathbb{P} \in \mathcal{P}(\Omega)$ gefunden werden, das zum Wiener-Maß \mathbb{W} im Sinne der relativen Entropie möglichst nahe liegt und gleichzeitig die beobachteten Daten reproduziert. Dieses Optimierungsproblem wird durch Minimierung der Kullback-Leibler-Divergenz (relative Entropie) unter Berücksichtigung der beobachteten Randbedingungen gelöst. Mathematisch kann dies wie folgt formuliert werden:

$$\mathbb{P}^* \in \operatorname{argmin}_{\mathbb{P} \in \mathcal{P}_T^\mu(\Omega)} \mathcal{H}(P|W), \quad (5.1)$$

wobei $\mathcal{P}_T^\mu(\Omega)$ die Menge der Wahrscheinlichkeitsmaße $P \in \mathcal{P}(\Omega)$ mit gemeinsamer Verteilung μ zu den Zeitpunkten $T = (t_1, \dots, t_N)$ bezeichnet, d. h. $P \circ (X_{t_1}, \dots, X_{t_N})^{-1} = \mu$. Die relative Entropie $\mathcal{H}(P|W)$ zwischen den Maßen P und W ist gegeben in Definition 4.6.

Mit der Definition kann alternativ $\mathcal{H}(\mathbb{P}|\mathbb{W}) = \mathbb{E}_{\mathbb{P}} \left[\ln \frac{d\mathbb{P}}{d\mathbb{W}} \right] = \mathbb{E}_{\mathbb{W}} \left[\frac{d\mathbb{P}}{d\mathbb{W}} \ln \frac{d\mathbb{P}}{d\mathbb{W}} \right]$ geschrieben werden, falls $\mathbb{P} \ll \mathbb{W}$ gilt.

Im Gegensatz zur klassischen Schrödinger-Brücke, die sich mit der Interpolation zwischen Anfangs- und Endverteilungen beschäftigt, erweitert das SBTS-Problem diesen Ansatz, indem es zusätzlich die zeitliche Abhängigkeit und Zwischenverteilungen berücksichtigt. Die gesuchte Lösung in (5.1) liefert eine entropische Interpolation der Zeitreihenverteilung.

5.3 Stochastisches Kontrollproblem

Das Optimierungsproblem (5.1) kann in den Kontext der stochastischen Kontrolle überführt werden, ähnlich wie dies für die klassische Schrödinger-Brücke in [15] dargelegt wurde. Gegeben ein Wahrscheinlichkeitsmaß $\mathbb{P} \in \mathcal{P}(\Omega)$, für das die relative Entropie $\mathcal{H}(\mathbb{P}|\mathbb{W})$ endlich ist, lässt sich mithilfe des Satz von Girsanovs ein \mathcal{F} -adaptierter Prozess α_t definieren. Dieser Prozess ist \mathbb{R}^d -wertig, besitzt eine endli-

che Energie und erfüllt die notwendigen Bedingungen, sodass folgende Darstellung möglich wird:

$$\ln \frac{d\mathbb{P}}{d\mathbb{W}} = \int_0^T \alpha_t dX_t - \frac{1}{2} \int_0^T |\alpha_t|^2 dt,$$

wobei $X_t - \int_0^t \alpha_s ds$, $0 \leq t \leq T$, eine Brownsche Bewegung unter \mathbb{P} ist. Die relative Entropie wird dann gegeben durch $\mathcal{H}(\mathbb{P}|\mathbb{W}) = \mathbb{E}_P \left[\frac{1}{2} \int_0^T |\alpha_t|^2 dt \right]$. Damit lässt sich (5.1) äquivalent formulieren als

Schrödinger-Brücken-Problem für Zeitreihen (SBTS):

$$\begin{cases} \text{Minimiere über } \alpha \in \mathcal{A}, & J(\alpha) = \mathbb{E}_P \left[\frac{1}{2} \int_0^T |\alpha_t|^2 dt \right], \\ \text{unter der Bedingung } dX_t = \alpha_t dt + dB_t, & X_0 = 0, \quad (X_{t_1}, \dots, X_{t_N}) \sim \mu, \end{cases}$$

wobei \mathcal{A} die Menge der \mathbb{R}^d -wertigen, \mathcal{F} -adaptierten Prozesse mit $\mathbb{E}_P \left[\int_0^T |\alpha_t|^2 dt \right] < \infty$ ist. Das Infimum des Problems wird durch

$$V_{\text{SBTS}} := \inf_{\alpha \in \mathcal{A}_T^\mu} J(\alpha)$$

bezeichnet, wobei \mathcal{A}_T^μ die Menge der Steuerungen α ist, die $(X_{t_1}, \dots, X_{t_N}) \sim \mu$ erfüllen. Dieses stochastische Kontrollproblem wird im Folgenden eingehend untersucht, um die Existenz und Eigenschaften einer optimalen Steuerung α^* zu analysieren.

5.4 Lösung für SBTs

Im Folgenden wird, in Anlehnung an das klassische SBP, angenommen, dass die Zielverteilung μ auf $(\mathbb{R}^d)^N$ absolut stetig ist (d. h. eine Dichte in Bezug auf das Lebesgue-Maß besitzt). Zur besseren Lesbarkeit wird diese Dichte als $\mu(x_1, \dots, x_N)$ bezeichnet. Als Modell für den zeitdiskreten Pfad einer Brownschen Bewegung auf $[0, T]$ wird die Verteilung von $\{B_{t_1}, \dots, B_{t_N}\}$ notiert, die man mit μ_T^B bezeichnet. Letztere besitzt die gaußsche Dichte

$$\mu_T^B(x_1, \dots, x_N) = \prod_{i=0}^{N-1} \frac{1}{\sqrt{2\pi(t_{i+1} - t_i)}} \exp\left(-\frac{|x_{i+1} - x_i|^2}{2(t_{i+1} - t_i)}\right), \quad (5.2)$$

wobei $(x_1, \dots, x_N) \in (\mathbb{R}^d)^N$ und die Konventionen $t_0 = 0$ sowie $x_0 = 0$ gelten.

Man nehme an, dass das Zielmaß μ im Sinne $\mu \ll \mu_T^B$ (d. h. μ ist absolut stetig bzgl. μ_T^B) und die relative Entropie $\mathcal{H}(\mu|\mu_T^B)$ endlich ist, d. h.

$$\mathcal{H}(\mu|\mu_T^B) = \int \ln\left(\frac{\mu}{\mu_T^B}\right) d\mu < \infty.$$

Das folgende Resultat beschreibt die Lösung des zugehörigen SBP für Zeitreihen (SPTS) (5.3).

Satz 5.4.1 (Lösung des SBP für Zeitreihen (SPTS))

Ein Diffusionsprozess $X_t = \int_0^t \alpha_s^* ds + B_t$, $0 \leq t \leq T$, mit der Drift α^* definiert als

$$\alpha_t^* = a^*(t, X_t; \{X_{t_i}\}_{t_i \leq t}), \quad 0 \leq t < T,$$

wobei $a^*(t, x; \mathbf{x}_i)$ für $t \in [t_i, t_{i+1})$ und $\mathbf{x}_i = (x_1, \dots, x_i) \in (\mathbb{R}^d)^i$ und $\mathbf{X}_{t_i} = (X_{t_1}, \dots, X_{t_N})$ durch

$$a^*(t, x; \mathbf{x}_i) = \nabla_x \ln \mathbb{E}_{\mathbb{W}} \left[\frac{\mu}{\mu_T^B} (X_{t_1}, \dots, X_{t_N}) \middle| \mathbf{X}_{t_i} = \mathbf{x}_i, X_t = x \right]$$

definiert ist, induziert ein Wahrscheinlichkeitsmaß \mathbb{P}^* gegeben durch

$$d\mathbb{P}^* = \frac{\mu}{\mu_T^B} (X_{t_1}, \dots, X_{t_N}) d\mathbb{W},$$

welches das Schrödinger-Brücken-Zeitreihenproblem löst. Zudem gilt:

$$V_{SPTS} = \mathcal{H}(\mathbb{P}^* | \mathbb{W}) = \mathcal{H}(\mu | \mu_T^B).$$

Beweis.

Da μ ein Wahrscheinlichkeitsmaß ist, das absolut stetig bezüglich μ_T^B ist, existiert gemäß des Satz von Radon-Nikodym eine Dichtefunktion $\frac{\mu}{\mu_T^B}$, die das Verhältnis der beiden Maße beschreibt. Für jede messbare Menge $A \subseteq (\mathbb{R}^d)^N$ gilt:

$$\mu(A) = \int_A \frac{\mu}{\mu_T^B} (x_1, \dots, x_N) d\mu_T^B(x_1, \dots, x_N).$$

Insbesondere folgt aus der Normierung von μ , dass

$$\mu((\mathbb{R}^d)^N) = \int_{(\mathbb{R}^d)^N} \frac{\mu}{\mu_T^B} (x_1, \dots, x_N) d\mu_T^B(x_1, \dots, x_N) = 1.$$

Da der Erwartungswert unter dem Wiener-Maß \mathbb{W} durch Integration mit μ_T^B definiert ist, kann geschrieben werden

$$\mathbb{E}_{\mathbb{W}} \left[\frac{\mu}{\mu_T^B} (X_{t_1}, \dots, X_{t_N}) \right] = \int_{(\mathbb{R}^d)^N} \frac{\mu}{\mu_T^B} (x_1, \dots, x_N) d\mu_T^B(x_1, \dots, x_N).$$

Aus der oben gezeigten Normierung von μ folgt unmittelbar, dass dieser Erwartungswert gleich 1 ist.

Auf Basis dieser Eigenschaft kann ein Wahrscheinlichkeitsmaß $\mathbb{P}^* \ll \mathbb{W}$ definiert werden, dessen Prozess durch

$$Z_t = \mathbb{E}_{\mathbb{W}} \left[\frac{d\mathbb{P}^*}{d\mathbb{W}} \middle| \mathcal{F}_t \right] = \mathbb{E}_{\mathbb{W}} \left[\frac{\mu}{\mu_T^B} (X_{t_1}, \dots, X_{t_N}) \middle| \mathcal{F}_t \right], \quad 0 \leq t \leq T,$$

gegeben ist.

Aufgrund der Markov-Eigenschaften und der Gaußschen Struktur der Brownschen Bewegung kann der Prozess Z_t für $t \in [t_i, t_{i+1})$, $i = 0, \dots, N - 1$, durch eine Funktion der Form $h_i(t, X_t; \mathbf{X}_{t_i})$ beschrieben werden. Dabei hängt Z_t von der aktuellen Position X_t und der bisherigen Historie $\mathbf{X}_{t_i} = (X_{t_1}, \dots, X_{t_i})$ ab. Diese Abhängigkeit folgt aus der Markov-Eigenschaft, nach der die zukünftigen Werte $(X_{t_{i+1}}, \dots, X_{t_N})$ unabhängig von der Vergangenheit sind, wenn der aktuelle Zustand X_t bekannt ist.

Für einen Pfad $\mathbf{x}_i = (x_1, \dots, x_i) \in (\mathbb{R}^d)^i$ wird $h_i(t, x; \mathbf{x}_i)$ auf $[t_i, t_{i+1}) \times \mathbb{R}^d$ definiert durch:

$$h_i(t, x; \mathbf{x}_i) = \mathbb{E}_{Y \sim N(0, I_d)} \left[\frac{\mu}{\mu_T^B} \left(\mathbf{x}_i, x + \sqrt{t_{i+1} - t} Y, \dots, x + \sqrt{t_N - t} Y \right) \right].$$

Hierbei beschreibt der Zufallsvektor $Y \sim N(0, I_d)$ die zukünftigen inkrementellen Änderungen der Brownschen Bewegung, welche durch ihre Gaußsche Struktur vollständig charakterisiert sind. Insbesondere wird X_{t_j} für $j \geq i + 1$ bedingt durch X_t und Y dargestellt als:

$$X_{t_j} = x + \sqrt{t_j - t} Y.$$

Die obige Darstellung von $h_i(t, x; \mathbf{x}_i)$ ergibt sich direkt aus der Definition von Z_t und der Möglichkeit, die Integration über zukünftige Pfade $(X_{t_{i+1}}, \dots, X_{t_N})$ durch eine Integration über Y zu ersetzen.

Die Funktion h_i beschreibt somit die bedingte Erwartung der relativen Dichte $\frac{\mu}{\mu_T^B}$ über die zukünftigen Pfade, gegeben den aktuellen Zustand X_t und die bisherige Historie \mathbf{X}_{t_i} . Ihre Struktur ermöglicht es, Z_t rekursiv über die Intervalle $[t_i, t_{i+1})$ zu berechnen, indem zukünftige Zustände iterativ integriert werden.

Zudem gilt nach der bedingten Erwartungsformel

$$h_i(t, x; \mathbf{x}_i) = \mathbb{E}_{\mathbb{W}} \left[h_{i+1}(t_{i+1}, X_{t_{i+1}}; \mathbf{x}_i, X_{t_{i+1}}) \middle| X_t = x \right].$$

Diese Rekursion ergibt sich aus der Struktur des Prozesses Z_t und der Tatsache, dass h_i die bedingte Erwartung der Dichte $\frac{\mu}{\mu_T^B}$ beschreibt. Die zukünftigen Werte des Prozesses $X_{t_{i+1}}, \dots, X_{t_N}$ hängen nur von X_t und den vergangenen Werten \mathbf{X}_{t_i} ab. Aus der Definition von h_{i+1} folgt, dass sich h_i durch die Integration von h_{i+1} über die möglichen Zustände von $X_{t_{i+1}}$ berechnen lässt.

Bedingt auf den aktuellen Zustand $X_t = x$ kann man die Verteilung von $X_{t_{i+1}}$ als multivariate Normalverteilung beschreiben

$$X_{t_{i+1}} \mid X_t = x \sim \mathcal{N}(x, (t_{i+1} - t)I_d).$$

Damit lässt sich die bedingte Erwartung wie folgt explizit formulieren.

$$h_i(t, x; \mathbf{x}_i) = \int_{\mathbb{R}^d} h_{i+1}(t_{i+1}, x'; \mathbf{x}_i, x') p_{t, t_{i+1}}(x, x') dx',$$

wobei $p_{t,t_{i+1}}(x, x')$ die Übergangsdichte der Brownschen Bewegung von x nach x' zwischen den Zeitpunkten t und t_{i+1} beschreibt

$$p_{t,t_{i+1}}(x, x') = \frac{1}{(2\pi(t_{i+1} - t))^{d/2}} \exp\left(-\frac{|x' - x|^2}{2(t_{i+1} - t)}\right).$$

Diese Rekursion zeigt, dass h_i auf $[t_i, t_{i+1}]$ durch die Werte von h_{i+1} an t_{i+1} und die Übergangsdichte $p_{t,t_{i+1}}$ bestimmt wird. Sie erlaubt eine iterative Berechnung, beginnend bei der terminalen Bedingung

$$h_N(t_N, x; \mathbf{x}_{N-1}, x) = \frac{\mu}{\mu_T^B}(\mathbf{x}_{N-1}, x).$$

Die Kombination dieser Rekursion mit der Wärmeleitungsgleichung, die h_i auf jedem Teilintervall $[t_i, t_{i+1}]$ erfüllt, stellt sicher, dass h_i aufeinanderfolgende Intervalle konsistent miteinander verbindet. Die Wärmeleitungsgleichung für h_i lautet

$$\frac{\partial h_i}{\partial t} + \frac{1}{2} \Delta_x h_i = 0,$$

wobei Δ_x der Laplace-Operator ist. Die Randbedingung

$$h_i(t_{i+1}, x; \mathbf{x}_i) = h_{i+1}(t_{i+1}, x; \mathbf{x}_i, x)$$

stellt sicher, dass h_i und h_{i+1} an den Übergangszeitpunkten t_{i+1} nahtlos ineinander übergehen. Dies ermöglicht eine Definition des Prozesses Z_t über das gesamte Intervall $[0, T]$, wobei die terminale Bedingung $h_N(t_N, x; \mathbf{x}_{N-1}, x) = \frac{\mu}{\mu_T^B}(x_1, \dots, x_{N-1}, x)$ verwendet wird. Daraus folgt, dass $(t, x) \mapsto h_i(t, x; \mathbf{x}_i)$ auf $[t_i, t_{i+1}] \times \mathbb{R}^d$ die Wärmeleitungsgleichung löst

$$\frac{\partial h_i(\cdot; \mathbf{x}_i)}{\partial t} + \frac{1}{2} \Delta_x h_i(\cdot; \mathbf{x}_i) = 0,$$

mit der Randbedingung $h_i(t_{i+1}, x; \mathbf{x}_i) = h_{i+1}(t_{i+1}, x; \mathbf{x}_i, x)$.

Angewandt auf den Prozess Z_t unter dem Wiener-Maß \mathbb{W} ergibt die Itô-Formel

$$dZ_t = \nabla_x h_i(t, X_t; X_{t_i}) dX_t = Z_t \nabla_x \ln h_i(t, X_t; X_{t_i}) dX_t, \quad t_i \leq t < t_{i+1},$$

für $i = 0, \dots, N-1$. Damit definiert man den Prozess α^* durch $\alpha_t^* = \nabla_x \ln h_i(t, X_t; X_{t_i})$ für $t \in [t_i, t_{i+1}]$. Damit ergibt sich

$$\frac{d\mathbb{P}^*}{d\mathbb{W}} = \exp\left(\int_0^T \alpha_t^* \cdot dX_t - \frac{1}{2} \int_0^T |\alpha_t^*|^2 dt\right).$$

Nach dem Satz von Girsanov ist $X_t - \int_0^t \alpha_s^* ds$ eine Brownsche Bewegung unter \mathbb{P}^* . Zudem gilt nach der Definition von \mathbb{P}^* und der Bayes-Formel für jede beschränkte, messbare Funktion φ

$$\begin{aligned}
\mathbb{E}_{\mathbb{P}^*}[\varphi(X_{t_1}, \dots, X_{t_N})] &= \mathbb{E}_W \left[\frac{\mu}{\mu_T^B}(X_{t_1}, \dots, X_{t_N}) \varphi(X_{t_1}, \dots, X_{t_N}) \right] \\
&= \int \frac{\mu}{\mu_T^B}(x_1, \dots, x_N) \varphi(x_1, \dots, x_N) \mu_T^B(x_1, \dots, x_N) dx_1, \dots, dx_N \\
&= \int \varphi(x_1, \dots, x_N) \mu(x_1, \dots, x_N) dx_1 \dots dx_N.
\end{aligned}$$

Somit folgt, dass $(X_{t_1}, \dots, X_{t_N}) \stackrel{\mathbb{P}^*}{\sim} \mu$ und dass $\alpha^* \in \mathcal{A}_T^\mu$ folgt aus $J(\alpha^*) = \mathbb{E}_{\mathbb{P}^*}[\int_0^T \frac{1}{2} |\alpha_t^*|^2 dt] = \mathbb{E}_{\mathbb{P}^*}[\ln \frac{d\mathbb{P}^*}{d\mathbb{W}}] = \mathcal{H}(\mathbb{P}^* | \mathbb{W}) = \mathcal{H}(\mu | \mu_T^B) < \infty$. Es bleibt zu zeigen, dass für jedes $\alpha \in \mathcal{A}_T^\mu$, das einem Wahrscheinlichkeitsmaß $\mathbb{P} \ll \mathbb{W}$ mit einer Dichte, die durch (5.2) gegeben ist, zugeordnet werden kann, gilt

$$J(\alpha) \geq H(\mu | \mu_W^T). \quad (5.3)$$

Um dies zu zeigen, wird wieder die Bayes-Formel verwendet, sowie die Eigenschaft, dass $\tilde{B}_t = X_t - \int_0^t \alpha_s ds$ unter dem Wahrscheinlichkeitsmaß \mathbb{P} eine Brownsche Bewegung ist, wie es durch das Satz von Girsanov garantiert wird. Daraus ergibt sich

$$1 = \mathbb{E}_W \left[\frac{\mu}{\mu_T^B}(X_{t_1}, \dots, X_{t_N}) \right] = \mathbb{E}_{\mathbb{P}} \left[\exp \left(\ln \frac{\mu}{\mu_T^B}(X_{t_1}, \dots, X_{t_N}) - \int_0^T \alpha_t \cdot dB_t - \frac{1}{2} \int_0^T |\alpha_t|^2 dt \right) \right].$$

Die erste Gleichheit ergibt sich aus der Definition der relativen Dichte $\frac{\mu}{\mu_T^B}$ und der Normierung des Wiener-Maßes \mathbb{W} . Die zweite folgt aus der Definition des Wahrscheinlichkeitsmaßes \mathbb{P} und der Dichte gemäß (5.2).

Anschließend wird die Jensen-Ungleichung benutzt. Da die Exponentialfunktion konvex ist, gilt

$$1 \geq \exp \left(\mathbb{E}_{\mathbb{P}} \left[\ln \frac{\mu}{\mu_T^B}(X_{t_1}, \dots, X_{t_N}) - \int_0^T \alpha_t dB_t - \frac{1}{2} \int_0^T |\alpha_t|^2 dt \right] \right).$$

Weil $(X_{t_1}, \dots, X_{t_N}) \stackrel{\mathbb{P}}{\sim} \mu$ gilt, reduziert sich der erste Term im Exponenten auf die relative Entropie $\mathcal{H}(\mu | \mu_T^B)$, und der dritte Term entspricht $J(\alpha)$. Somit gilt

$$1 \geq \exp(H(\mu | \mu_W^T) - J(\alpha)).$$

Durch Anwendung des Logarithmus auf beide Seiten folgt

$$0 \geq H(\mu | \mu_T^B) - J(\alpha),$$

oder äquivalent

$$J(\alpha) \geq H(\mu | \mu_T^B).$$

Damit ist die Ungleichung (5.3) bewiesen. \square

Bemerkung 5.4.1

Der optimale Drift des SBTS-Prozesses ist von dem gesamten Pfad des Prozesses abhängig. Zu einem Zeitpunkt t wird die Drift nicht nur durch den aktuellen Zustand X_t , sondern auch durch die vergangenen Zustände $X_{t_1}, \dots, X_{t_{\eta(t)}}$ beeinflusst, wobei $\eta(t)$ der letzte bekannte Zeitpunkt vor t ist. Der Prozess kann durch folgende SDGL beschrieben werden

$$dX_t = a^*(t, X_t; \mathbf{X}_{\eta(t)}) dt + dB_t, \quad 0 \leq t \leq T, \quad X_0 = 0.$$

Die Driftfunktion a^* ist für jedes Intervall $[t_i, t_{i+1})$ durch die bedingte Erwartung der relativen Dichte $\rho := \frac{\mu}{\mu_T^B}$ charakterisiert. Sie lässt sich über die Funktion h_i darstellen für $t \in [t_i, t_{i+1})$, welche den vergangenen Pfad berücksichtigt

$$a^*(t, x; x_i) = \frac{\nabla_x h_i(t, x; \mathbf{x}_i)}{h_i(t, x; \mathbf{x}_i)}, \quad x \in \mathbb{R}^d, \quad x_i \in (\mathbb{R}^d)^i.$$

Dabei ist h_i definiert als

$$h_i(t, x; x_i) = \mathbb{E}_{Y \sim N(0, I_d)} [\rho(x_i, x + \sqrt{t_{i+1} - t} Y, \dots, x + \sqrt{t_N - t} Y)].$$

Die Funktion h_i beschreibt, wie die Zielverteilung μ unter der relativen Dichte ρ über zukünftige Pfade des Prozesses integriert wird. Die Zukunft wird dabei über die Gaußsche Struktur der Brownschen Bewegung modelliert, sodass die Zufallsvariable $Y \sim N(0, I_d)$ die inkrementellen Änderungen des Prozesses beschreibt. Die Drift a^* hängt daher nicht nur lokal von t und X_t , sondern auch global von dem gesamten Pfad bis zur Zeit t ab.

Diese Darstellung der Driftfunktion erlaubt es, die pfadabhängige Struktur des Prozesses mathematisch präzise zu beschreiben und bildet die Grundlage für numerische Algorithmen zur Berechnung.

Proposition 5.4.1

Für $i = 0, \dots, N-1$, $t \in [t_i, t_{i+1})$, $\mathbf{x}_i = (x_1, \dots, x_i) \in (\mathbb{R}^d)^i$ und $x \in \mathbb{R}^d$ wird die Driftfunktion $a^*(t, x; \mathbf{x}_i)$ gegeben durch

$$a^*(t, x; \mathbf{x}_i) = \frac{1}{t_{i+1} - t} \frac{\mathbb{E}_\mu \left[(X_{t_{i+1}} - x) \exp \left(-\frac{|X_{t_{i+1}} - x|^2}{2(t_{i+1} - t)} - \frac{|X_{t_{i+1}} - x_i|^2}{2(t_{i+1} - t_i)} \right) \middle| \mathbf{X}_{t_i} = \mathbf{x}_i \right]}{\mathbb{E}_\mu \left[\exp \left(-\frac{|X_{t_{i+1}} - x|^2}{2(t_{i+1} - t)} - \frac{|X_{t_{i+1}} - x_i|^2}{2(t_{i+1} - t_i)} \right) \middle| \mathbf{X}_{t_i} = \mathbf{x}_i \right]}. \quad (5.4)$$

Beweis.

Es sei $t \in [t_i, t_{i+1})$ und $\mathbf{x}_i = (x_1, \dots, x_i)$ für ein $i \in \{0, 1, \dots, N-1\}$ fixiert. Die bedingte Erwartung unter dem Wiener-Maß \mathbb{W} ist gegeben durch

$$\mathbb{E}_{\mathbb{W}} \left[\frac{\mu}{\mu_T^B} (X_{t_1}, \dots, X_{t_N}) \middle| \mathbf{X}_{t_i} = \mathbf{x}_i, X_t = x \right].$$

Die Dichte des Wiener-Prozesses μ_T^B in einer Zerlegung des Intervalls $[0, T]$ ist durch folgende Formel gegeben

$$\mu_T^B(x_1, \dots, x_N) = \prod_{j=0}^{N-1} \frac{1}{\sqrt{2\pi(t_{j+1} - t_j)}} \exp\left(-\frac{|x_{j+1} - x_j|^2}{2(t_{j+1} - t_j)}\right),$$

wie in (5.2) definiert. Diese Dichte beschreibt die Übergangswahrscheinlichkeiten des Wiener-Prozesses zwischen den Zeitpunkten t_j und t_{j+1} .

Bei der Bedingung auf $X_{t_i} = \mathbf{x}_i$ wird der Anteil dieser Pfade quasi herausgefiltert. Dies führt zur Konstante

$$C = \frac{\mu(\mathbf{x}_i)}{\mu_T^B(\mathbf{x}_i)} = \frac{\mu(x_1, \dots, x_i)}{\prod_{j=0}^{i-1} \frac{1}{\sqrt{2\pi\Delta t_j}} \exp\left(-\frac{|x_{j+1} - x_j|^2}{2\Delta t_j}\right)},$$

wobei $\Delta t_j = t_{j+1} - t_j$. Diese Konstante berücksichtigt die historischen Pfade bis zum Zeitpunkt t_i , ist jedoch unabhängig von den zukünftigen Zuständen (x_{i+1}, \dots, x_N) . Eingesetzt in die Definition der bedingten Erwartung liefert

$$\mathbb{E}_{\mathbb{W}}\left[\frac{\mu}{\mu_T^B}(X_{t_1}, \dots, X_{t_N}) | \mathbf{X}_{t_i} = \mathbf{x}_i, X_t = x\right] = C \int_{\mathbb{R}^d} \cdots \int_{\mathbb{R}^d} \rho(X_{t_1}, \dots, X_{t_N}) dX_{t_{i+1}} \cdots dX_{t_N}.$$

Da die Übergangswahrscheinlichkeit des Wiener-Prozesses von X_t nach $X_{t_{i+1}}$ durch

$$p(X_{t_{i+1}} | X_t = x) = \frac{1}{(2\pi(t_{i+1} - t))^{d/2}} \exp\left(-\frac{|X_{t_{i+1}} - x|^2}{2(t_{i+1} - t)}\right)$$

beschrieben wird, ergibt sich

$$\mathbb{E}_{\mathbb{W}}\left[\frac{\mu}{\mu_T^B} | \mathbf{X}_{t_i} = \mathbf{x}_i, X_t = x\right] = C \cdot F_i(t, X_{t_i} x, X_{t_{i+1}}),$$

wobei $F_i(t, x_i, x, x_{i+1})$ definiert ist als

$$F_i(t, x_i, x, x_{i+1}) = \exp\left(-\frac{|x_{i+1} - x|^2}{2(t_{i+1} - t)} - \frac{|x_{i+1} - x_i|^2}{2(t_{i+1} - t_i)}\right).$$

Unter dem Zielmaß μ wird die bedingte Erwartung zu

$$\mathbb{E}_{\mu}\left[F_i(t, x_i, x, X_{t_{i+1}}) | \mathbf{X}_{t_i} = \mathbf{x}_i, X_t = x\right].$$

Die Konstante C kürzt sich im finalen Ausdruck der Driftfunktion heraus, sodass gilt

$$a^*(t, x; \mathbf{x}_i) = \frac{1}{t_{i+1} - t} \frac{\mathbb{E}_{\mu}\left[(X_{t_{i+1}} - x) F_i(t, x, X_{t_{i+1}}) | \mathbf{X}_{t_i} = \mathbf{x}_i\right]}{\mathbb{E}_{\mu}\left[F_i(t, x, X_{t_{i+1}}) | \mathbf{X}_{t_i} = \mathbf{x}_i\right]}.$$

□

Bemerkung 5.4.2

Die Zielverteilung μ kann in manchen Fällen durch eine Markov-Kette beschrieben werden. Eine solche Verteilung lässt sich durch Übergangskerne v_i auf \mathbb{R}^d in der Form

$$\mu(dx_1, \dots, dx_N) = \prod_{i=1}^{N-1} v_i(x_i, dx_{i+1})$$

darstellen. In dieser Konstellation wird die bedingte Erwartung in (5.4) erheblich vereinfacht: Sie hängt nicht mehr von der gesamten Historie der Pfade X_{t_1}, \dots, X_{t_t} ab, sondern ausschließlich vom aktuellen Zustand X_t . Das bedeutet, dass die Dynamik in diesem Fall nur über den direkt vorhergehenden Wert gesteuert wird, was die essentielle Eigenschaft eines Markov-Prozesses widerspiegelt.

5.4.1 Erweiterung: Soft-constrained Schrödinger Bridge (SSB)

Das Soft-constrained Schrödinger Bridge (SSB) Problem wurde von JHANVI GARG, XIANYANG ZHANG und QUAN ZHOU in [8] vorgestellt. Es stellt eine Verallgemeinerung des klassischen Schrödinger-Brücken-Problems dar, indem die harte Nebenbedingung, dass die Randverteilung mit der Zielverteilung μ_T übereinstimmen muss, durch eine weiche Einschränkung ersetzt wird. Dies wird durch die Einführung eines Regularisierungsparameters $\beta > 0$ erreicht, der die Abweichung von der Zielverteilung quantifiziert.

Das SSB-Problem wird wie folgt formuliert

$$V = \inf_{u \in \mathcal{U}} J_\beta(u), \quad (5.5)$$

wobei die Zielfunktion $J_\beta(u)$ gegeben ist durch:

$$J_\beta(u) = \beta \cdot \mathcal{H}(\text{Law}(X_T^u) | \mu_T) + \mathbb{E} \left[\int_0^T \frac{|u_t|^2}{2\sigma^2} dt \right]. \quad (5.6)$$

Hierbei bezeichnet $\text{Law}(X_T^u)$ die Verteilung des Prozesses X_T^u am Endzeitpunkt T , der unter dem Einfluss der Steuerung u gemäß der stochastischen Differentialgleichung

$$dX_t = u_t dt + \sigma dB_t, \quad X_0 = x_0,$$

entsteht. Der Parameter $\beta > 0$ kontrolliert den Einfluss der Kullback-Leibler-Divergenz, die durch $\mathcal{H}(\cdot | \cdot)$ zwischen $\text{Law}(X_T^u)$ und der Zielverteilung μ_T ausgedrückt wird.

Das SSB ersetzt somit die harte Nebenbedingung des klassischen Schrödinger-Brücken-Problems, dass $\text{Law}(X_T^u) = \mu_T$ gelten muss, durch eine Regularisierung, die Abweichungen erlaubt und durch β parametrisiert wird. Für $\beta \rightarrow \infty$ wird die weiche Einschränkung zunehmend restriktiv, bis die Zielverteilung exakt erfüllt wird, und das SSB geht in das klassische Schrödinger-Brücken-Problem über. Dieser Übergang wird mathematisch durch den Grenzwert

$$\lim_{\beta \rightarrow \infty} [\beta \cdot \mathcal{H}(\text{Law}(X_T^u) | \mu_T)] = \begin{cases} 0, & \text{wenn } \text{Law}(X_T^u) = \mu_T, \\ \infty, & \text{sonst,} \end{cases}$$

beschrieben.

Das SBTS kann als spezieller Grenzfall des SSB interpretiert werden, bei dem $\beta \rightarrow \infty$ gilt. Im SBTS wird die Zielfunktion auf die Minimierung der Energie der Steuerung reduziert, während die Zielverteilung exakt reproduziert werden muss

$$V = \inf_{u \in \mathcal{U}} \mathbb{E} \left[\int_0^T \frac{|u_t|^2}{2\sigma^2} dt \right], \quad (5.7)$$

unter der Nebenbedingung, dass $\text{Law}(X_T^u) = \mu_T$. Dieser Zusammenhang zeigt, warum das Schrödinger-Brücken-Problem für Zeitreihen als Grenzfall des SSB betrachtet werden kann.

Der Übergang wird formell durch den Ausdruck

$$\lim_{\beta \rightarrow \infty} \mathcal{H}(\text{Law}(X_T^u) | \mu_T) = \begin{cases} 0, & \text{wenn } \text{Law}(X_T^u) = \mu_T, \\ \infty, & \text{sonst,} \end{cases}$$

beschrieben. Somit entsteht bei $\beta \rightarrow \infty$ ein striktes Schrödinger-Brücken-Problem mit harten Randbedingungen.

Die Soft-constrained Schrödinger Bridge bietet signifikante Vorteile bei numerischen Berechnungen, da sie Regularisierung erlaubt und somit numerisch stabiler ist. Für Anwendungen in hochdimensionalen Zeitreihenmodellen oder bei unvollständigen Zielverteilungen ist dieser Ansatz besonders nützlich. Der Übergang zu $\beta \rightarrow \infty$ führt zu einer optimalen Steuerung u_t^* , die mit der des SBTS-Problems übereinstimmt, wie in [8] nachgewiesen wird. Dies eröffnet neue Möglichkeiten für die Anwendung generativer Modelle in der Finanzmathematik und der stochastischen Modellierung. Die Soft-constrained Schrödinger Bridge erweitert die klassischen Schrödinger-Brücken-Probleme auf elegante Weise und bietet zusätzliche Flexibilität durch Regularisierung.

5.5 Generative Modelle

Generative Modelle sind eine Klasse maschineller Lernverfahren, deren Ziel es ist, Datenproben zu erzeugen, die einer gegebenen Zielverteilung μ folgen. Im Kontext der SBP ermöglichen diese Modelle die Simulation des Diffusionsprozesses, indem sie sowohl die zugrundeliegenden stochastischen Prozesse als auch die Driftterme a_i^* approximieren. Die Driftterme beschreiben die Dynamik des Prozesses und spielen eine zentrale Rolle bei der Konstruktion von Proben, die die zeitlichen und statistischen Eigenschaften der Zielverteilung präzise abbilden.

Ein Ansatz, um die Driftterme a_i^* zu approximieren, basiert auf der Schätzung bedingter Erwartungswerte unter der Zielverteilung μ . Diese bedingten Erwartungswerte können auf verschiedene Weisen geschätzt werden. Zu den gängigsten Methoden zählen die Kernel-Schätzung und die Approximation durch neuronale Netze wie LSTM-Modelle, die insbesondere für pfadabhängige Prozesse geeignet sind.

5.5.1 Kernel-Schätzung der Drift

Die Kernel-Schätzung ist eine Methode zur Annäherung von Dichten und bedingten Erwartungswerten basierend auf einer gegebenen Menge von Datenproben. Diese Technik ist besonders nützlich, wenn die Driftterme a_i^* explizit von bedingten Erwartungswerten abhängen, die direkt aus Daten geschätzt werden können.

Sei $\{X_k\}_{k=1}^n \subset \mathbb{R}^d$ eine Menge unabhängiger und identisch verteilter Datenpunkte, die aus einer Verteilung f gezogen wurden. Die Kernel-Dichteschätzung von f ist gegeben durch

$$\hat{f}_h(x) = \frac{1}{nh^d} \sum_{k=1}^n K\left(\frac{x - X_k}{h}\right),$$

wobei $K : \mathbb{R}^d \rightarrow \mathbb{R}$ ein Kernel ist, der typischerweise eine symmetrische und positiv definierte Funktion darstellt, und $h > 0$ die Bandbreite des Kernels bezeichnet.

Unter geeigneten Regularitätsbedingungen konvergiert die Kernel-Schätzung $\hat{f}_h(x)$ fast sicher gegen $f(x)$, wenn $n \rightarrow \infty$ und $h \rightarrow 0$ mit $nh^d \rightarrow \infty$.

Ein typischer Kernel ist der Gauß-Kernel, definiert durch

$$K(x) = \frac{1}{(2\pi)^{d/2}} \exp\left(-\frac{|x|^2}{2}\right).$$

Sei $\{X_k\}_{k=1}^n \subset \mathbb{R}$ eine Menge von Datenpunkten. Für einen Punkt $x \in \mathbb{R}$ ist die Kernel-Schätzung der Dichte an x mit Bandbreite $h > 0$ gegeben durch

$$\hat{f}_h(x) = \frac{1}{nh} \sum_{k=1}^n \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{(x - X_k)^2}{2h^2}\right).$$

Neben dem Gauß-Kernel existieren jedoch eine Vielzahl anderer Kernel-Funktionen, die sich in der Praxis bewährt haben. Diese alternativen Kernel unterscheiden sich in ihrer Form und den Eigenschaften, die sie auf die Dichteschätzung übertragen.

Beispielhafte Kernel:

1. **Quadratischer Kernel:** Der quadratische Kernel ist durch folgende Funktion definiert

$$K(x) = \begin{cases} (H^2 - x^2)^2 & \text{falls } |x| < H, \\ 0 & \text{sonst.} \end{cases}$$

Dieser Kernel hat einen kompakten Träger und ist bevorzugt in Anwendungen, in denen lokale Glättung gewünscht wird. Die Bandbreite H steuert die Breite des Einflussbereichs.

2. **Gauß-Kernel:** Der Gauß-Kernel, wie oben beschrieben, hat die Form:

$$K(x) = \frac{1}{\sqrt{2\pi}H^2} \exp\left(-\frac{x^2}{2H^2}\right),$$

wobei H die Bandbreite steuert. Er hat einen unbeschränkten Träger und glättet die Dichte über den gesamten Wertebereich der Daten.

3. Laplace-Kernel: Der Laplace-Kernel ist gegeben durch

$$K(x) = \frac{1}{2H} \exp\left(-\frac{|x|}{H}\right).$$

Dieser Kernel ist besonders geeignet für Daten mit exponentiellen Verteilungen und bietet eine robustere Schätzung bei Ausreißern.

4. Polynom-Kernel: Der Polynom-Kernel der Ordnung p wird definiert durch

$$K(x) = \left(1 + \frac{x}{H}\right)^p.$$

Hierbei steuert H die Skalierung, und der Grad p des Polynoms bestimmt die Glättungseigenschaften. Höhere Grade führen zu stärkeren Glättungseffekten.

Für alle genannten Kernel kann die Dichteschätzung durch die allgemeine Formel

$$\hat{f}_h(x) = \frac{1}{nh} \sum_{k=1}^n K\left(\frac{x-X_k}{h}\right)$$

berechnet werden, wobei $K(x)$ die jeweilige Kernel-Funktion darstellt. Die Wahl des Kernels und der Bandbreite h hängt von den spezifischen Eigenschaften des Datensatzes und dem gewünschten Glättungsgrad ab. Der Gauß-Kernel wird häufig als Standard gewählt, während der Laplace- oder quadratische Kernel in Szenarien mit Ausreißern oder begrenztem Datenbereich bevorzugt werden können.

In der Praxis kann der Gauß-Kernel verwendet werden, um eine glatte Dichte zu approximieren, während der quadratische Kernel in Schrödinger-Brücken-Problemen bei der Approximation von Drifttermen nützlich ist. Die polynomiale Variante eignet sich insbesondere für Daten mit stark nichtlinearer Struktur, da sie die Flexibilität bietet, zusätzliche Glättung oder Nichtlinearität in die Schätzung einzuführen.

Beispiel 5.5.1 (Kernel-Dichteschätzung)

In diesem Beispiel wird die Kernel-Dichteschätzung (KDE) verwendet, um eine Wahrscheinlichkeitsdichte auf Basis eines Datensatzes zu approximieren. Die Implementierung erfolgt in Python mithilfe der Bibliothek scikit-learn. Zusätzlich wird die Dichteschätzung mit der wahren Dichte verglichen, um die Genauigkeit der Methode zu illustrieren.

Zunächst wird ein Datensatz simuliert, der aus zwei Normalverteilungen besteht. Diese Verteilungen werden gemischt, um eine multimodale Verteilung zu erzeugen. Die Kernel-Dichteschätzung erfolgt mithilfe des Moduls KernelDensity aus scikit-learn. Ein Gauß-Kernel wird verwendet, um die Dichte zu schätzen, und die Bandbreite steuert die Glättung. Zur Visualisierung wird die geschätzte Dichtekurve

gemeinsam mit einem Histogramm der Daten und der wahren Dichte (falls bekannt) geplottet.

Die Bandbreite des Kernels ist ein zentraler Parameter, der die Glättung der Dichteschätzung beeinflusst. Eine kleine Bandbreite führt zu einer stark angepassten, aber möglicherweise rauschenden Schätzung, während eine große Bandbreite die Daten stärker glättet, aber eventuell wichtige Strukturen verwischt.

Die geschätzte Dichtekurve zeigt eine glatte Annäherung an die zugrunde liegende Datenverteilung. Durch die Wahl eines geeigneten Bandbreitenparameters lässt sich ein guter Kompromiss zwischen Glättung und Anpassung erzielen. Im Vergleich zur wahren Dichte zeigt sich, dass die Kernel-Dichteschätzung die multimodale Struktur der Daten gut erfasst.

Abbildung 14 zeigt ein Histogramm der simulierten Daten, die geschätzte Dichtekurve (KDE) mit zwei unterschiedlichen Bandbreiten ($h = 0.5$ und $h = 0.2$) sowie die wahre Dichte der zugrunde liegenden Verteilung. Die kleinere Bandbreite ($h = 0.2$) führt zu einer stärkeren Anpassung an die Daten, während die größere Bandbreite ($h = 0.5$) eine geglättete Schätzung liefert.

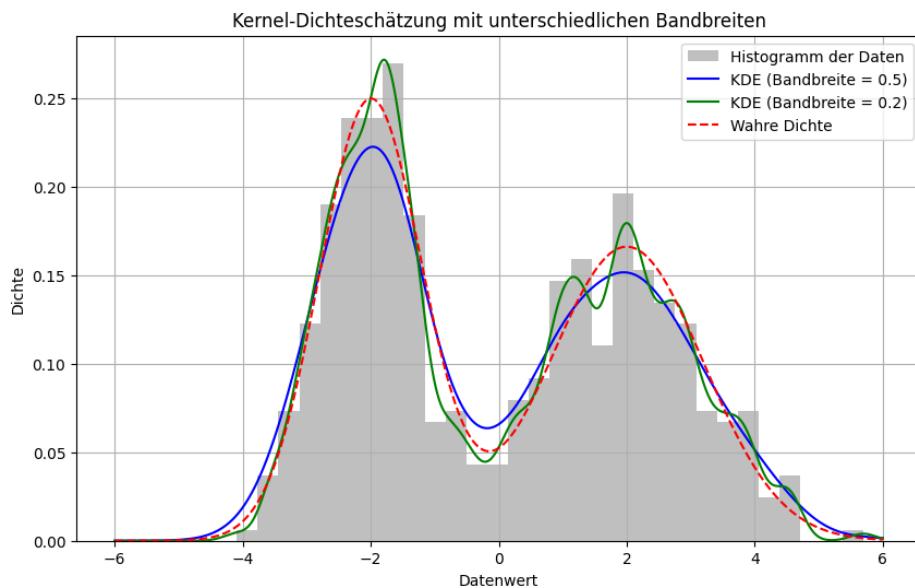


Abbildung 14: Histogramm der Daten, Kernel-Dichteschätzung (KDE) mit Bandbreiten $h = 0.2$ und $h = 0.5$, sowie wahre Dichte.

Der Nadaraya-Watson-Schätzer ist eine nicht-parametrische Methode zur Schätzung der bedingten Erwartung $\mathbb{E}[Y|X = x]$. Im Gegensatz zur Kernel-Dichteschätzung, die die Wahrscheinlichkeitsdichte einer Verteilung approximiert, fokussiert sich der Nadaraya-Watson-Schätzer auf die Modellierung der Beziehung zwischen einer EingabevARIABLE X und einer Zielvariablen Y . Dabei werden die Zielwerte Y_i der Datenpunkte (X_i, Y_i) mit Gewichten versehen, die durch die Nähe der Eingabewerte

X_i zum Punkt x bestimmt werden. Diese Gewichtung erfolgt durch eine Kernel-Funktion, wie den Gauß-Kernel. Der Schätzer wird häufig in nicht-parametrischen Regressionsproblemen eingesetzt und bietet eine datengestützte Alternative zu klassischen Regressionsmethoden.

$$\hat{f}(x) = \frac{\sum_{i=1}^n K_h(x - x_i)y_i}{\sum_{i=1}^n K_h(x - x_i)},$$

wobei $K_h(x)$ eine Kernel-Funktion mit Bandbreite $h > 0$ ist, typischerweise der Gauß-Kernel

$$K_h(x) = \frac{1}{\sqrt{2\pi}h} \exp\left(-\frac{x^2}{2h^2}\right).$$

Der Nadaraya-Watson-Schätzer gewichtet die Zielwerte y_i in Abhängigkeit von der Nähe der Eingabewerte x_i zum Schätzpunkt x . Die Bandbreite h steuert die Glättung der Schätzung.

Beispiel 5.5.2 (Nadaraya-Watson-Schätzer)

Um den Schätzer zu veranschaulichen, werden die Daten aus einem nicht-linearen Zusammenhang $Y = \sin(X) + \varepsilon$ simuliert, wobei $X \sim \mathcal{U}[-\pi, \pi]$ und $\varepsilon \sim \mathcal{N}(0, 0.1^2)$. Anschließend wird die Regressionsfunktion geschätzt. Abbildung 15 visualisiert den Nadaraya-Watson-Schätzer für die Bandbreiten $h = 0.3$ und $h = 0.7$ gegeben der wahren Funktion Y .

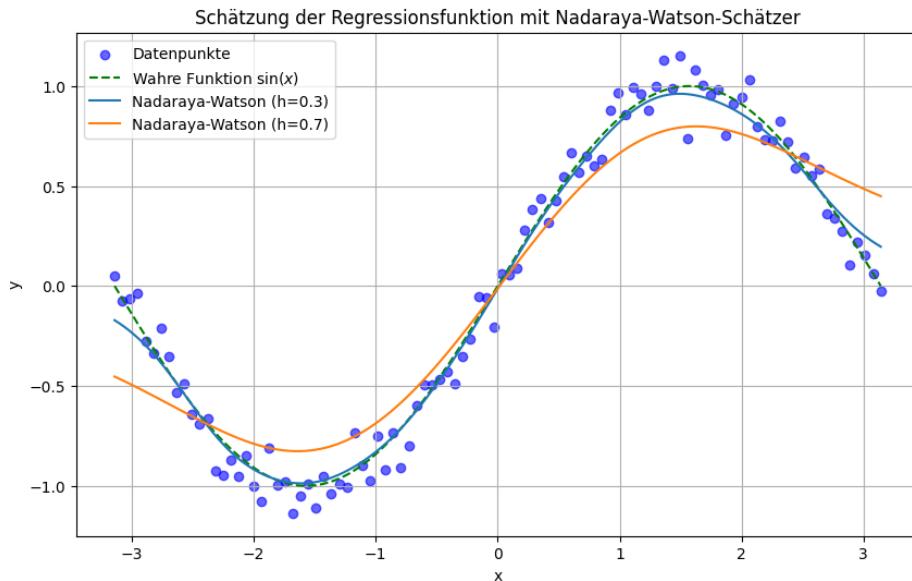


Abbildung 15: Schätzung der Regressionsfunktion mithilfe des Nadaraya-Watson-Schätzers mit zwei Bandbreiten $h = 0.3$ und $h = 0.7$.

Im Kontext der Schrödinger-Brücke für Zeitreihen wird der Nadaraya-Watson Schätzer verwendet, um die bedingten Erwartungswerte zu approximieren, die in der Formel für die Driftterme a_i^* auftreten. Der Ansatz ermöglicht es, diese Driften auf Grundlage von Datenproben aus der Zielverteilung μ zu schätzen, ohne auf explizite analytische

Formeln zurückgreifen zu müssen.

Die Nadaraya-Watson-Schätzung der Driftfunktion basiert auf den gegebenen Datenproben $(\mathbf{X}_{t_i}^{(m)})_{i=1,\dots,N}$ für $m = 1, \dots, M$ aus μ gezogen und ist wie folgt definiert

$$\hat{a}(t, x; \mathbf{x}_i) = \frac{1}{(t_{i+1} - t)} \frac{\sum_{m=1}^M \left(X_{t_{i+1}}^{(m)} - x \right) F_i(t, X_{t_i}^{(m)}, x, X_{t_{i+1}}^{(m)}) \prod_{j=1}^i K_h(x_j - X_{t_j}^{(m)})}{\sum_{m=1}^M F_i(t, X_{t_i}^{(m)}, x, X_{t_{i+1}}^{(m)}) \prod_{j=1}^i K_h(x_j - X_{t_j}^{(m)})}.$$

wobei $t \in [t_i, t_{i+1}]$, $x_i \in (\mathbb{R}^d)^i$, $x \in \mathbb{R}^d$, und K_h ein Kernel mit Bandbreite $h > 0$ ist und

$$F_i(t, x_i, x, x_{i+1}) = \exp\left(-\frac{|x_{i+1} - x|^2}{2(t_{i+1} - t)} - \frac{|x_{i+1} - x_i|^2}{2(t_{i+1} - t_i)}\right).$$

Der verwendete Kernel K_h ist eine nicht-negative, reellwertige, integrierbare und symmetrische Funktion. Einige Kernel wurden bereits oben eingeführt.

Die Schätzfunktion $\hat{a}(t, x; \mathbf{x}_i)$ kombiniert Informationen aus den Datenproben mithilfe von Gewichtungsfaktoren, die durch den Kernel und die Funktion F_i definiert werden. Die Funktion F_i ist eine glättende Gewichtung, die die bedingte Struktur der Drift innerhalb der Schrödinger-Brücke berücksichtigt.

Dieser Ansatz erlaubt eine datengetriebene Approximation der Drift, die sowohl flexibel als auch effizient ist, insbesondere in Fällen, in denen die analytische Form der Drift unbekannt oder schwer zugänglich ist. Durch die Nutzung von Kernel-Methoden wird eine hohe Anpassungsfähigkeit an die zugrunde liegenden Verteilungen gewährleistet, was die Anwendung des Schrödinger-Brücken-Problems für Zeitreihen auf realistische Szenarien erleichtert.

5.5.2 Schätzung des Drift mit neuronalen Netzen

Eine alternative Möglichkeit besteht darin, die Driftterme a_i^* , insbesondere für pfa-dabhängige Prozesse, durch neuronale Netzwerke zu approximieren. In WANG ET AL. [21] wird eine Methode gezeigt, welche die Approximation des RN-Derivats $\rho = \frac{\mu}{\mu_T^B}$ mit Hilfe eines neuronalen Netzes kombiniert. Diese Methode erlaubt es, sowohl die Nicht-Markov-Struktur der Prozesse als auch deren zeitliche Abhängigkeiten effizient zu modellieren.

Die Idee basiert darauf, die logarithmierte Dichte $r_\theta(\mathbf{X})$ zu approximieren, wobei r_θ durch ein neuronales Netz mit Parametern θ dargestellt wird. Die Optimierung dieser Parameter erfolgt durch Minimierung einer logistischen Verlustfunktion (loss function)

$$L_{\text{logistic}}(r) = \mathbb{E}_\mu [\ln(1 + \exp(-r(\mathbf{X})))] + \mathbb{E}_{\mu_T^B} [\ln(1 + \exp(r(\mathbf{X})))].$$

Gegeben Datenstichproben $(\mathbf{X}_{t_i}^{(m)})_{i=1,\dots,N}$ für $m = 1, \dots, M$ aus μ gezogen und zusätzliche Stichproben $\{Y_t^{(m)}\}_{m=1}^M$ aus μ_T^B gezogen, wird ρ durch

$$\hat{\rho}(x) = \exp(r_{\hat{\theta}}(x))$$

ermöglicht. Die Optimierung erfolgt durch

$$\theta \mapsto \frac{1}{M} \sum_{m=1}^M \left[\ln \left(1 + \exp(-r(\mathbf{X}^{(m)})) \right) + \ln \left(1 + \exp(r(\mathbf{Y}^{(m)})) \right) \right].$$

Die Schätzung der Driftfunktion $a^*(t, x; \mathbf{x}_i)$ erfolgt anschließend durch die Anwendung der folgenden Formel

$$a^*(t, x; \mathbf{x}_i) = \frac{\mathbb{E}_{Y \sim \mathcal{N}(0, I_d)} \left[\nabla_x \rho(x_i, x + \sqrt{t_{i+1} - t} Y, \dots, x + \sqrt{t_N - t} Y) \right]}{\mathbb{E}_{Y \sim \mathcal{N}(0, I_d)} \left[\rho(x_i, x + \sqrt{t_{i+1} - t} Y, \dots, x + \sqrt{t_N - t} Y) \right]},$$

wobei das Radon-Nikodým Derivat ρ und ihre Gradienten durch das neuronale Netz geschätzt werden. Diese Methode kombiniert die Flexibilität neuronaler Netzwerke mit Monte-Carlo-Simulationen, um sowohl ρ als auch die Drift zu approximieren.

Die Verwendung neuronaler Netzwerke ermöglicht es, hochdimensionale und komplexe Abhängigkeiten in der Zielverteilung μ zu erfassen, die mit klassischen Methoden schwer zugänglich sind. Gleichzeitig erlaubt der Ansatz durch die Nutzung von Monte-Carlo-Techniken eine präzise Approximation der Erwartungswerte. Ein Nachteil dieser Methode ist der hohe Rechenaufwand, der sowohl für das Training des neuronalen Netzes als auch für die Monte-Carlo-Simulation zur Schätzung der Driftterme erforderlich ist [21].

5.5.3 Simulation der Schrödinger-Brücken-Zeitreihe

Sei eine SDGL der Form

$$dX_t = f(X_t) dt + g(X_t) dB_t,$$

gegeben, wobei $\{B_t\}_{t \geq 0}$ eine Brownsche Bewegung ist. Das Euler-Maruyama-Schema approximiert die Lösung durch

$$X_{t_{i+1}} = X_{t_i} + f(X_{t_i}) \Delta t + g(X_{t_i}) \sqrt{\Delta t} Z_i,$$

wobei $Z_i \sim \mathcal{N}(0, I_d)$ standardnormalverteilte Zufallsvariablen sind.

Für die SBTs wird das Euler-Maruyama-Schema angewendet, wobei $f(X_t)$ durch den approximierten Driftterm $\hat{a}_i(X_t)$ ersetzt wird und $\sigma = 1$ gesetzt wird

$$X_{t_{i+1}} = X_{t_i} + \hat{a}_i(X_{t_i}) \Delta t + \sqrt{\Delta t} Z_i.$$

Hierbei ist \hat{a}_i die approximierte Driftfunktion, die durch Nadaraya-Watson-Schätzer oder neuronale Netzwerke ermittelt wird.

5.6 SBTS-Algorithmus

Ausgehend von der geschätzten Driftfunktion \hat{a} , die die Pfadabhängigkeit berücksichtigt, kann die SDGL durch ein Euler-Maruyama-Schema simuliert werden. Zur feinen Diskretisierung zwischen zwei Beobachtungszeitpunkten t_i und t_{i+1} wird das Intervall in N^π gleichmäßige Zeitschritte unterteilt. Die diskreten Zeitpunkte werden dabei durch $t_{k,i}^\pi = t_i + \frac{k}{N^\pi}$ mit $k = 0, \dots, N^\pi - 1$ definiert. Der folgende Algorithmus beschreibt die Simulation der Schrödinger-Brücken-Zeitreihe.

Eingabe: Gegebene Zeitreihen $(X_t^{(m)})_{m=1,\dots,M}$ an den Zeitpunkten t_1, \dots, t_N

Ausgabe: Simulierte Zeitreihe x_1, \dots, x_N

Initialisierung: Setze $x_0 = 0$;

for $i = 0$ bis $N - 1$ **do**

Initialisiere den Zustand $y_0 = x_i$;

for $k = 0$ bis $N^\pi - 1$ **do**

Berechne $\hat{a}(t_k^\pi, y_k; \mathbf{x}_i)$ (z. B. mit Kernel-Schätzer);

Ziehe $\varepsilon_k \sim \mathcal{N}(0, 1)$;

Aktualisiere:

$$y_{k+1} = y_k + \frac{1}{N^\pi} \hat{a}(t_{k,i}^\pi, y_k; \mathbf{x}_i) + \frac{1}{\sqrt{N^\pi}} \varepsilon_k$$

end

Setze $x_{i+1} = y_{N^\pi}$;

end

return x_1, \dots, x_N

Algorithmus 1: Simulation der Schrödinger-Brücken-Zeitreihe

Algorithmus 1 zeigt die systematische Anwendung des Euler-Maruyama-Schemas zur Simulation von Schrödinger-Brücken-Zeitreihen. Durch die Kombination aus Approximation des Drifts und stochastischem Anteil wird eine flexible Modellierung von Prozessen ermöglicht, die sowohl deterministische als auch zufällige Komponenten enthalten. Die feine Diskretisierung des Zeitintervalls erlaubt es, die dynamischen Eigenschaften der Schrödinger-Brücke präzise abzubilden und die Zielverteilung schrittweise zu rekonstruieren.

Die Verwendung von Nadaraya-Watson-Schätzern zur Bestimmung des optimalen Drifts bietet eine flexible und datengetriebene Methode, um komplexe Verteilungen zu modellieren. Gleichzeitig bleibt der Algorithmus numerisch effizient und anpassbar an verschiedene Szenarien, indem er die Eigenschaften der SDGL nutzt. Die strukturierte Iteration über die Zeitgitter und die stochastischen Terme ermöglichen eine robuste Simulation, die insbesondere in Anwendungen wie der Finanzmathematik oder der Zeitreihenanalyse von großer Bedeutung ist.

5.7 Numerische Experimente

In diesem Kapitel werden numerische Experimente zur Konstruktion und Simulation von SBTS für eindimensionale Systeme vorgestellt. Die Ergebnisse basieren auf Implementierungen, die mit der Programmiersprache Python und der Jupyter-Notebook-Umgebung entwickelt wurden. Ziel ist es, die praktische Anwendung der theoretisch formulierten Schrödinger-Brücke auf einfache, eindimensionale Szenarien zu demonstrieren. Dazu werden zunächst die zugrunde liegenden Modelle und Randbedingungen definiert, bevor die numerischen Ergebnisse ausführlich diskutiert und visualisiert werden.

Im ersten Beispiel wird die Simulation einer Schrödinger-Brücke anhand einer synthetischen Zeitreihe illustriert. Dies dient als Einstieg in die Methodik, da ARIMA-Modelle als gut verstandene Systeme eine ideale Grundlage für die Demonstration bilden.

Beispiel 5.7.1 (Simulation einer Schrödinger-Brücke basierend auf Zeitreihen)
Eine Zeitreihe wird mithilfe eines ARIMA(2,0,1)-Modells erzeugt, wobei Datensätze der Größe $M = 1000$ verwendet werden. Das Modell folgt der Struktur

$$X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \theta_1 \varepsilon_{t-1} + \varepsilon_t,$$

wobei ϕ_1, ϕ_2 die autoregressiven Koeffizienten und θ_1 der MA-Koeffizient sind. Das Rauschen ε_t folgt einer Normalverteilung $N(0, \sigma^2)$. Für die Simulation der Schrödinger-Brücke wurden die folgenden Parameter verwendet. Die Anzahl der Zeitintervalle wurde auf $N = 3$ gesetzt, die Anzahl der Proben auf $M = 1000$. Die Schrittweite wurde mit $\Delta t = \frac{7}{252}$ festgelegt, während die Bandbreite des Kernels $H = 0.05$ beträgt. Für die zeitliche Diskretisierung wurden $N^\pi = 100$ Schritte pro Zeitschritt gewählt, und insgesamt wurden 1000 Schrödinger-Brücken simuliert. Die Driftfunktion wurde mit einem Kernel geschätzt, wobei das Euler-Verfahren zur Simulation genutzt wurde.

Die zeitliche Entwicklung der Schrödinger-Brücke für den letzten Zeitschritt X_{t_3} wird in Abbildung 16 dargestellt. Dabei zeigt sich eine hohe Übereinstimmung zwischen den simulierten Schrödinger-Brücken und den echten Daten, wodurch die Fähigkeit der Methode zur zuverlässigen Modellierung zeitlicher Abhängigkeiten demonstriert wird.

Zusätzlich zur zeitlichen Entwicklung wird in Abbildung 17 ein gemeinsamer Verteilungsplot der Zufallsvariablen X_{t_1}, X_{t_2} und X_{t_3} dargestellt. Die Abbildung zeigt sowohl die marginalen Verteilungen (entlang der Diagonalen) als auch die bivariaten Verteilungen (in den oberen und unteren Dreiecken). Die Konturdiagramme illustrieren die Dichteschätzungen der Verteilungen für die echten Daten und die simulierten Schrödinger-Brücken. Diese Darstellung verdeutlicht, wie gut die simulierten Schrödinger-Brücken die Abhängigkeiten zwischen den Variablen und deren Verteilungseigenschaften nachbilden.

Die empirischen kumulativen Verteilungsfunktionen (ECDFs) für jedes X_{t_i} wurden in Abbildung 18 dargestellt. Diese zeigen, dass die simulierten Daten die Verteilungseigen-

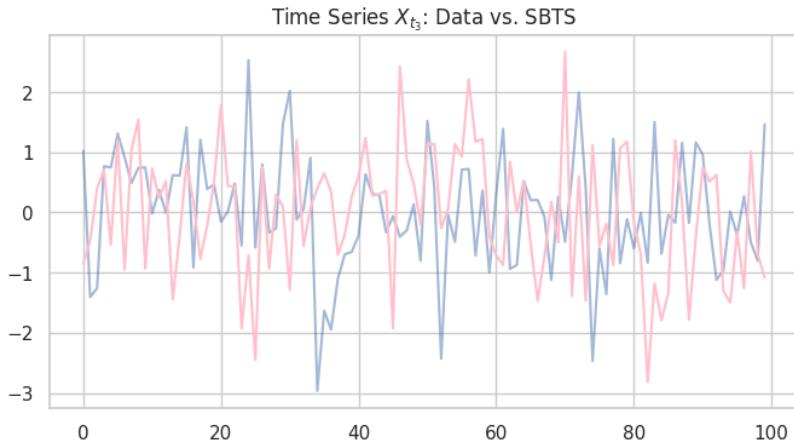


Abbildung 16: Vergleich von X_{t_3} zwischen echten Daten in pink und simulierten Schrödinger-Brücken in blau.

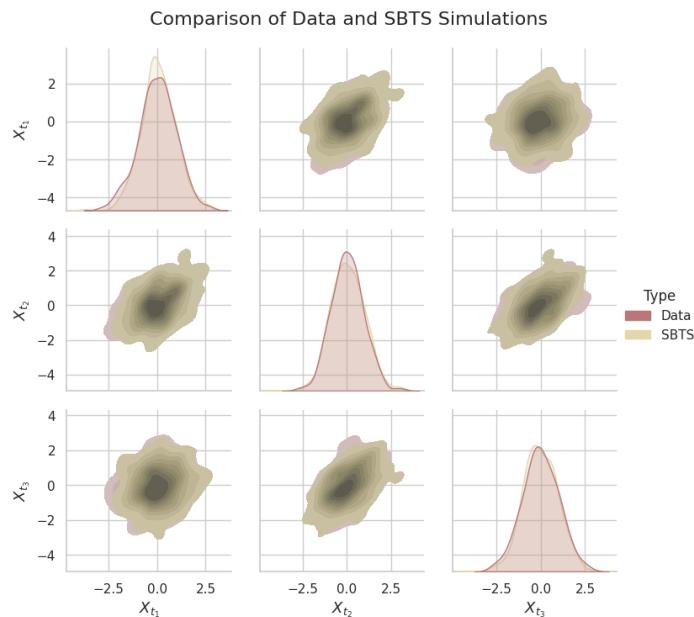


Abbildung 17: Gemeinsamer Verteilungsplot der Zufallsvariablen X_{t_1} , X_{t_2} und X_{t_3} für echte Daten und simulierte Schrödinger-Brücken. Die Diagonale zeigt die marginalen Verteilungen, während die oberen und unteren Dreiecke die bivariaten Dichtekonturen darstellen.

schaften der echten Daten konsistent abbilden.

Darüber hinaus zeigen die in Abbildung 19 dargestellten Boxplots, dass die simulierten Daten die zentralen Tendenzen und die Streuung der echten Daten gut nachbilden. Die Ergebnisse bestätigen die Fähigkeit der Schrödinger-Brücke, sowohl die zeitlichen als auch die verteilungsmäßigen Strukturen der Daten zuverlässig zu simulieren.

Quantitative Analysen in Tabelle 3 zeigen, dass die Quantile der simulierten Daten in hohem Maße mit denen der echten Daten übereinstimmen. Die p-Werte verdeutlichen,

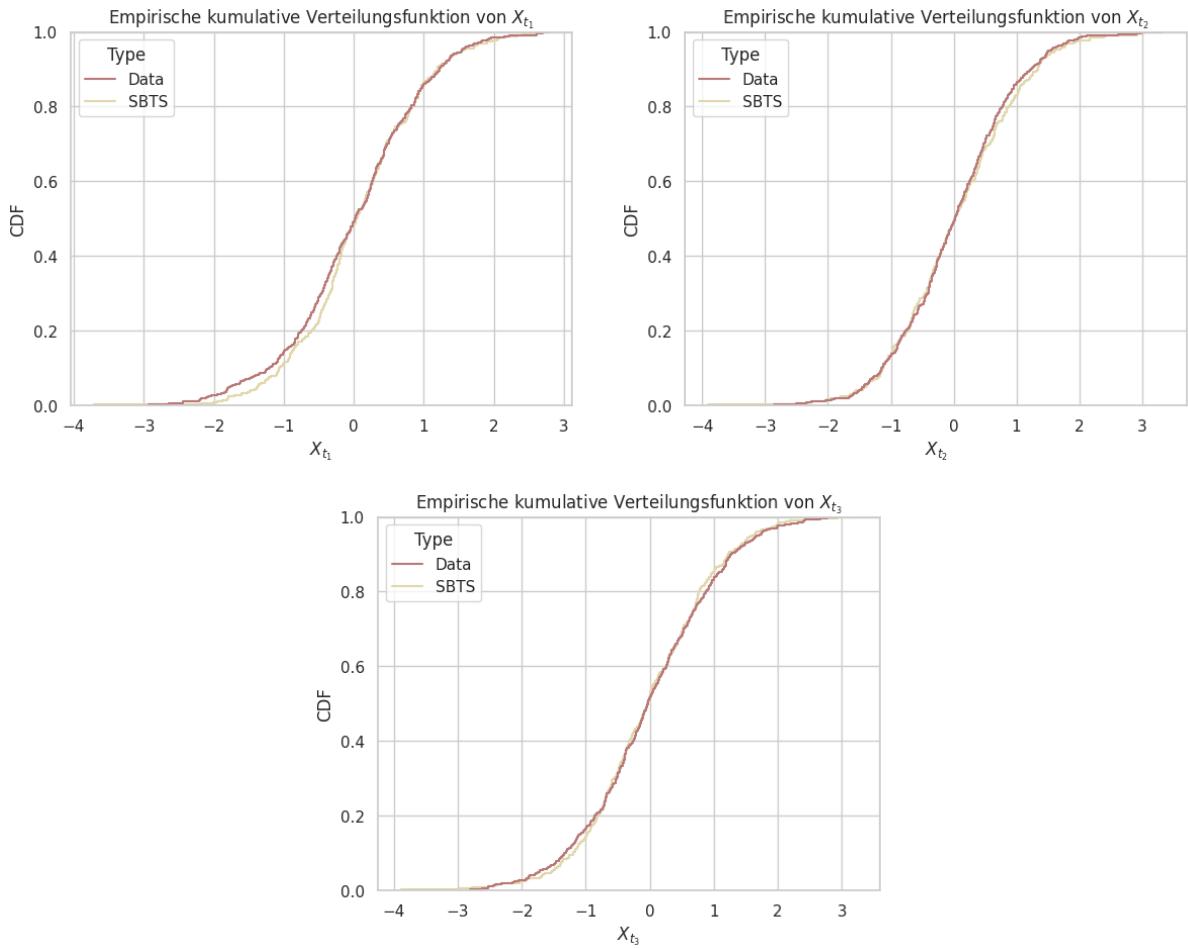


Abbildung 18: Empirische kumulative Verteilungsfunktionen (ECDF) für X_{t_1} , X_{t_2} und X_{t_3} bei echten Daten und simulierten Schrödinger-Brücken. Die obere Zeile zeigt die Ergebnisse für X_{t_1} und X_{t_2} , während die untere Zeile die Ergebnisse für X_{t_3} zusammenfasst.

dass keine signifikanten Unterschiede zwischen den Verteilungen bestehen.

Die Korrelationsmatrizen für die echten und simulierten Daten sowie ihre Differenzen werden im Folgenden dargestellt:

$$\text{Datenkorrelation: } \begin{bmatrix} 1.000 & 0.458 & 0.184 \\ 0.458 & 1.000 & 0.489 \\ 0.184 & 0.489 & 1.000 \end{bmatrix},$$

$$\text{SB-Korrelation: } \begin{bmatrix} 1.000 & 0.454 & 0.197 \\ 0.454 & 1.000 & 0.502 \\ 0.197 & 0.502 & 1.000 \end{bmatrix},$$

$$\text{Differenz: } \begin{bmatrix} 0.000 & 0.005 & -0.013 \\ 0.005 & 0.000 & -0.013 \\ -0.013 & -0.013 & 0.000 \end{bmatrix}.$$

Die Ergebnisse zeigen, dass die Schrödinger-Brücke für Zeitreihen sowohl die zeitlichen Entwicklungen der Zeitreihe X_t als auch deren Verteilungs- und Abhängigkeitsstrukturen

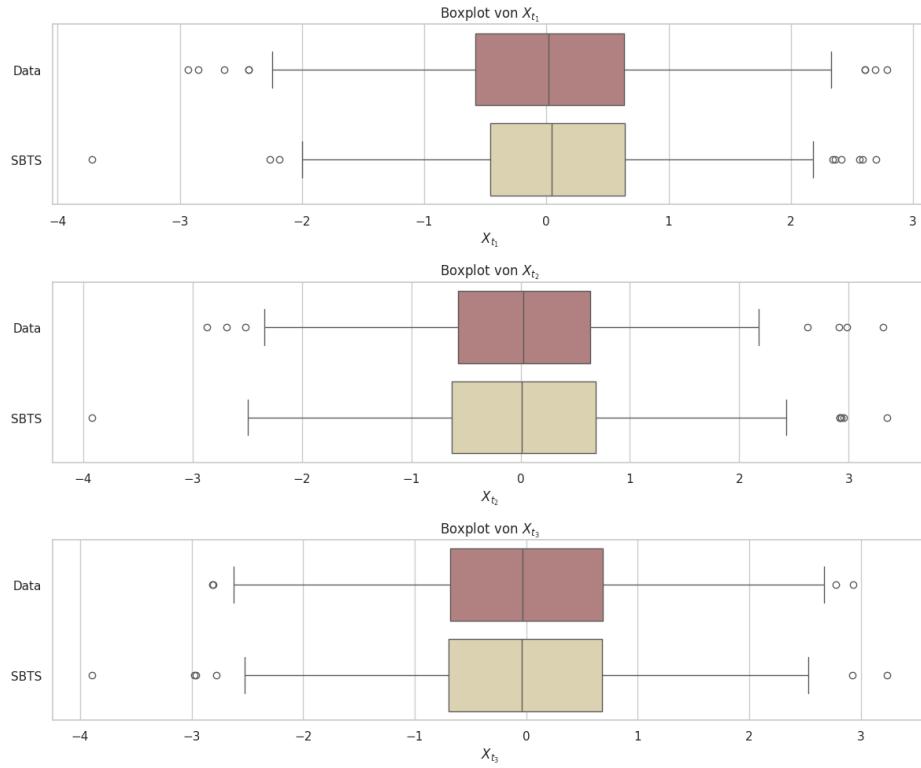


Abbildung 19: Boxplots für X_{t_1} , X_{t_2} und X_{t_3} bei echten Daten und simulierten Schrödinger-Brücken. Die Boxplots zeigen die zentrale Tendenz und Streuung der Daten im Vergleich zu den simulierten Schrödinger-Brücken.

<i>p</i> -Wert	5%-Quantil (Daten)	5%-Quantil (SB)	95%-Quantil (Daten)	95%-Quantil (SB)
0.313	-1.7316	-1.451	1.6105	1.4834
0.9136	-1.5979	-1.4217	1.6103	1.6082
0.9542	-1.6542	-1.5350	1.6527	1.6952

Tabelle 3: Quantile und *p*-Werte der echten Daten und simulierten Schrödinger-Brücken.

zuverlässig simulieren kann. Dies macht sie zu einer leistungsfähigen Methode für die Modellierung dynamischer Systeme.

Im zweiten Beispiel wird die Komplexität erhöht, indem ein GARCH-Modell verwendet wird. GARCH-Zeitreihen repräsentieren reale Finanzzeitreihen besser, insbesondere durch ihre Fähigkeit, Volatilität zu modellieren.

Beispiel 5.7.2 (SBTS basierend auf GARCH-Zeitreihen)

In diesem Abschnitt wird ein GARCH-Modell (Generalized Autoregressive Conditional Heteroskedasticity) verwendet, um Zeitreihen zu generieren. Das GARCH-Modell wird häufig genutzt, um die Volatilität von Finanzzeitreihen zu modellieren. Für die Simulation wurde ein GARCH(1,2)-Modell gewählt, das durch die folgenden Gleichungen beschrieben wird

$$X_{t+1} = \sigma_{t+1} \varepsilon_{t+1}, \quad \sigma_{t+1}^2 = \alpha_0 + \alpha_1 X_t^2 + \alpha_2 X_{t-1}^2,$$

wobei ε_t unabhängige und identisch normalverteilte Rauschkomponenten mit einem Mittelwert von 0 und einer Standardabweichung von 0.1 darstellen. Die Parameter des Modells wurden wie folgt festgelegt

$$\alpha_0 = 2, \quad \alpha_1 = -0.4, \quad \alpha_2 = 0.1.$$

Die Größe der generierten Zeitreihen beträgt $N = 60$. Um eine stationäre Verteilung zu erreichen, wurden zunächst 50 Schritte durchlaufen, bevor die eigentliche Zeitreihe erzeugt wurde. Für die Generierung der SBTS wurden folgende Parameter verwendet

- Anzahl der Zeitreihen: $M = 1000$,
- Anzahl der Schritte für die Euler-Maruyama-Diskretisierung: $N^\pi = 100$,
- Bandbreite für die Kernel-Schätzung: $h = 0.2$,
- Zeitschrittweite: $\Delta t = \frac{7}{252}$.

Die Schrödinger-Brücke wurde mit einem Kernel-Schätzer für die Driftfunktion simuliert. Abbildung 20 zeigt die Pfade von vier zufällig ausgewählten Zeitreihen aus dem GARCH-Modell im Vergleich zu den entsprechenden simulierten Schrödinger-Brücken. Es zeigt sich eine hohe Übereinstimmung zwischen den simulierten Werten und den echten Daten, was die Effektivität der Methode zur Modellierung zeitlicher Abhängigkeiten unterstreicht.

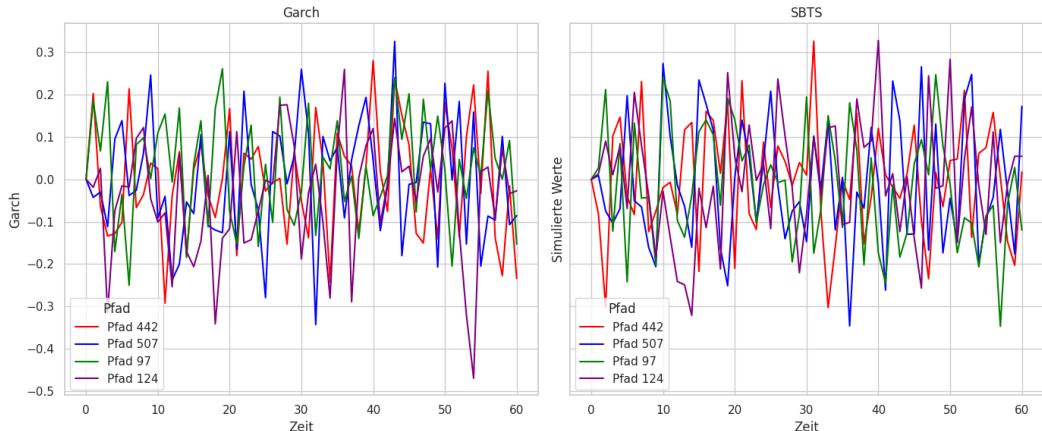


Abbildung 20: Vergleich von GARCH-Zeitreihen (links) und simulierten Schrödinger-Brücken (rechts) für vier zufällig ausgewählte Pfade.

Die in Abbildung 20 dargestellten Pfade illustrieren, wie das SBTS-Modell dynamische und volatile Prozesse wie GARCH-Zeitreihen präzise approximieren kann. Diese Eigenschaft macht die Methode zu einem wertvollen Werkzeug für die Analyse von Finanzzeitreihen, insbesondere für Anwendungen, bei denen Unsicherheiten und Volatilitäten eine entscheidende Rolle spielen.

Die Simulation mit GARCH-Zeitreihen im vorgehenden Beispiel bildet eine Brücke zur Modellierung realer Daten wie den logarithmischen Renditen von Microsoft-Aktien. Letztere enthalten ähnliche Eigenschaften wie autoregressive Abhängigkeiten und Heteroskedastizität, was die Anwendung der SPTS auf Finanzzeitreihen besonders relevant macht.

Beispiel 5.7.3 (Generierung von Microsoft-Aktienzeitreihen)

In diesem Beispiel werden historische Daten der Microsoft-Aktie verwendet, die über die Yahoo Finance API heruntergeladen wurden. Der betrachtete Zeitraum umfasst Januar 2010 bis Januar 2020. Basierend auf den logarithmischen Renditen der Aktie wird die Schrödinger-Brücke simuliert, um die zugrunde liegende Dynamik der Zeitreihe zu modellieren. Die Fenstergröße für die Zeitreihe beträgt $N = 60$, wobei die Anzahl der Proben $M = 1000$, die Bandbreite $H = 0.05$, die Anzahl der Schritte pro Zeitschritt $N^\pi = 100$ und die Zeitschrittweite $\Delta t = \frac{1}{252}$ (Handelstage).

Abbildung 21 zeigt zufällig ausgewählte Pfade der logarithmischen Renditen der Microsoft-Aktie im Vergleich zu simulierten Paden, die mithilfe der Schrödinger-Brücke generiert wurden. Die zufällig ausgewählten Pfade der simulierten Daten zeigen eine ähnlich ausgeprägte Volatilität wie die echten Microsoft-Daten. Schwankungen und Trends sind in beiden Fällen sichtbar, was darauf hinweist, dass die Schrödinger-Brücke für Zeitreihen in der Lage ist, die kurzfristigen Dynamiken und Trends der Microsoft-Daten nachzubilden.

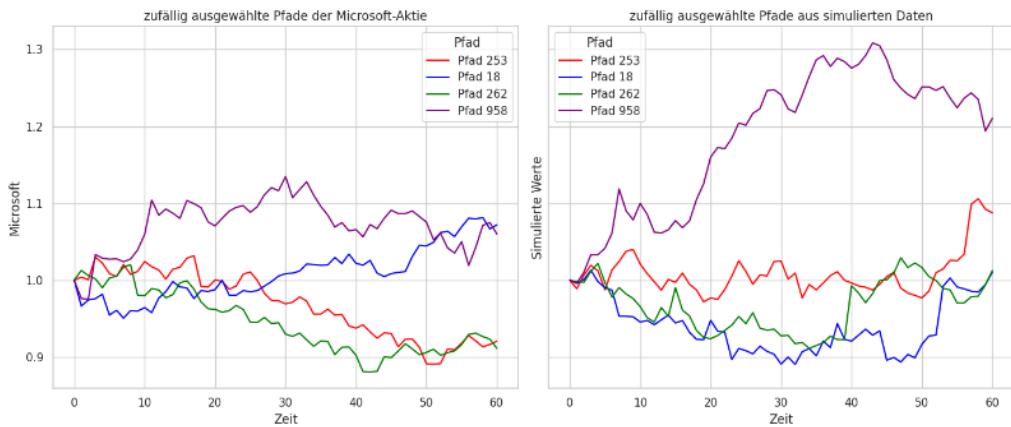


Abbildung 21: Zufällig ausgewählte Pfade der Microsoft-Aktie (links) und simulierte Pfade (rechts).

Die Kovarianzmatrizen der Originaldaten und der simulierten Schrödinger-Brücken-Daten werden in Abbildung 22 dargestellt. Auffällig ist, dass die höchste Kovarianz bei den simulierten Daten (0.012) größer ist als bei den Originaldaten (0.008). Dies zeigt, dass die SPTS möglicherweise stärkere Abhängigkeiten zwischen den Variablen modelliert, als sie tatsächlich in den Originaldaten existieren. Dennoch bleibt die Struktur der Kovarianzmatrizen vergleichbar, was die Fähigkeit der Schrödinger-Brücke unterstreicht, die Abhängigkeitsstruktur der Microsoft-Zeitreihen in großem Maße zu rekon-

struieren.

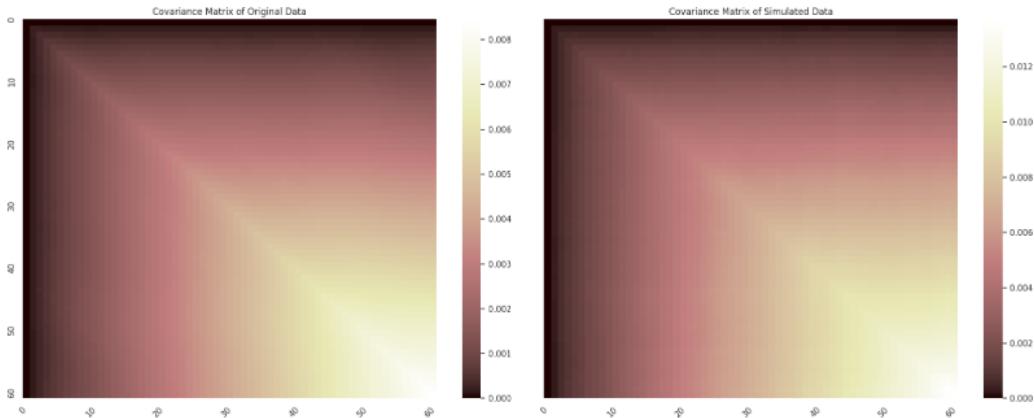


Abbildung 22: Kovarianzmatrizen der Originaldaten (links) und simulierten Daten (rechts).

Die quadratische Variation, welche die aggregierte Volatilität der Zeitreihe misst, wird in Abbildung 23 dargestellt. Die Verteilung zeigt, dass die quadratische Variation der SBTS-Daten eine signifikante Ähnlichkeit zu den echten Microsoft-Daten aufweist. Dennoch zeigt sich, dass die SBTS-Daten tendenziell eine etwas breitere Verteilung mit einer höheren Dichte bei größeren Werten haben. Dies deutet darauf hin, dass die Schrödinger-Brücke für Zeitreihen eine leicht geglättete Darstellung der tatsächlichen Volatilität erzeugt, während sie gleichzeitig Variationen in einem breiteren Bereich ermöglicht. Die wesentlichen Merkmale der Originaldaten bleiben dabei weitgehend erhalten.

Abbildung 24 zeigt schließlich die Verteilung der Renditen der Microsoft-Daten, der simulierten Schrödinger-Brücken-Daten und einer Normalverteilung. Die Verteilung der simulierten Daten ähnelt zwar den Microsoft-Daten, weicht jedoch bei extremen Renditen ab, was auf die Tendenz der Schrödinger-Brücke hinweist, die extremen Schwankungen der echten Daten nicht vollständig abzubilden. Gleichzeitig wird die Abweichung der Microsoft-Daten von einer Normalverteilung durch die Schrödinger-Brücke gut reproduziert.

Zusammenfassend zeigt dieses Beispiel, dass die Schrödinger-Brücke für Zeitreihen die Dynamik und Verteilungseigenschaften der Microsoft-Zeitreihen weitgehend zuverlässig modelliert. Die Kovarianzmatrix der simulierten Daten weist jedoch leichte Abhängigkeiten auf, was möglicherweise eine Überanpassung an die simulierten Datenstrukturen darstellt. Dennoch bleibt die SBTS eine leistungsstarke Methode zur Modellierung der zeitlichen Dynamiken und Verteilungseigenschaften solcher Finanzzeitreihen.

Nach der Anwendung der Schrödinger-Brücke auf eindimensionale Finanzzeitreihen zeigt dieses Beispiel eine Erweiterung auf hochdimensionale Daten in Form von Bildern. Der MNIST-Fashion-Datensatz dient hier als einfacheres Szenario, um die Methodik in einer neuen Domäne zu validieren.

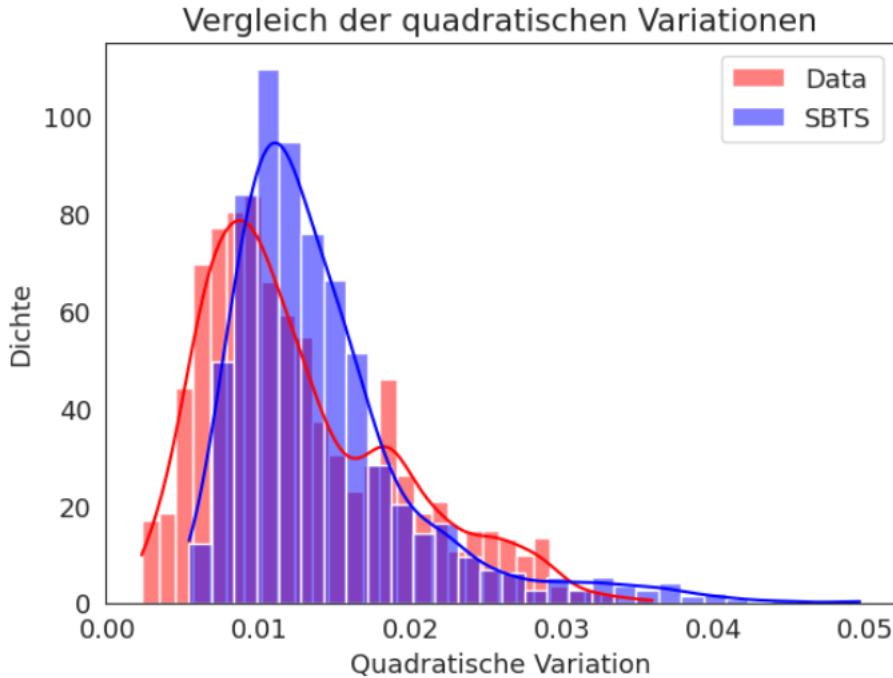


Abbildung 23: Verteilung der quadratischen Variation für Microsoft-Daten und Schrödinger-Brücken-Simulationen.

Beispiel 5.7.4 (Generierung von Bildern mit SBTS-Modell)

In diesem Beispiel wird der MNIST-Fashion-Datensatz verwendet, um mit dem SBTS-Modell neue Bilder zu generieren. Der MNIST-Fashion-Datensatz besteht aus 28×28 -Graustufenbildern von Kleidungsstücken wie T-Shirts, Schuhen oder Taschen.

Die Pixelwerte der Bilder wurden normalisiert, um numerische Probleme bei der Berechnung zu vermeiden. Die Normalisierung erfolgt, indem die Pixelwerte durch 255 geteilt und anschließend um 0.05 verschoben werden. Um das SBTS-Modell anwenden zu können, werden die Bilder in flache Vektoren umgewandelt. Die Dimension eines Bildes ist also $784 = 28 \times 28$.

Der SBTS-Algorithmus wird mit den normalisierten und transformierten Bildern initialisiert. Anschließend wird das Modell verwendet, um neue Bilder zu generieren. Die Simulation erfolgt, indem die SBTS zwischen einer uniformen Verteilung (als Startpunkt) und der Zielverteilung der Bilder aus dem Datensatz erstellt wird.

Abbildung 25 zeigt zufällig ausgewählte originale Bilder des MNIST-Fashion-Datensatzes im Vergleich zu simulierten Bildern, die mithilfe der Schrödinger-Brücke für Zeitreihen erzeugt wurden. Die Simulation zeigt, dass die Schrödinger-Brücke in der Lage ist, die Verteilung der Bilder realitätsnah zu rekonstruieren, wobei kleine Unterschiede in den Details sichtbar bleiben, wie das erhöhte Rauschen anstatt eines klaren schwarzen Hintergrundes.

Die Ergebnisse zeigen, dass das SBTS-Modell für Bilder eine leistungsfähige Methode

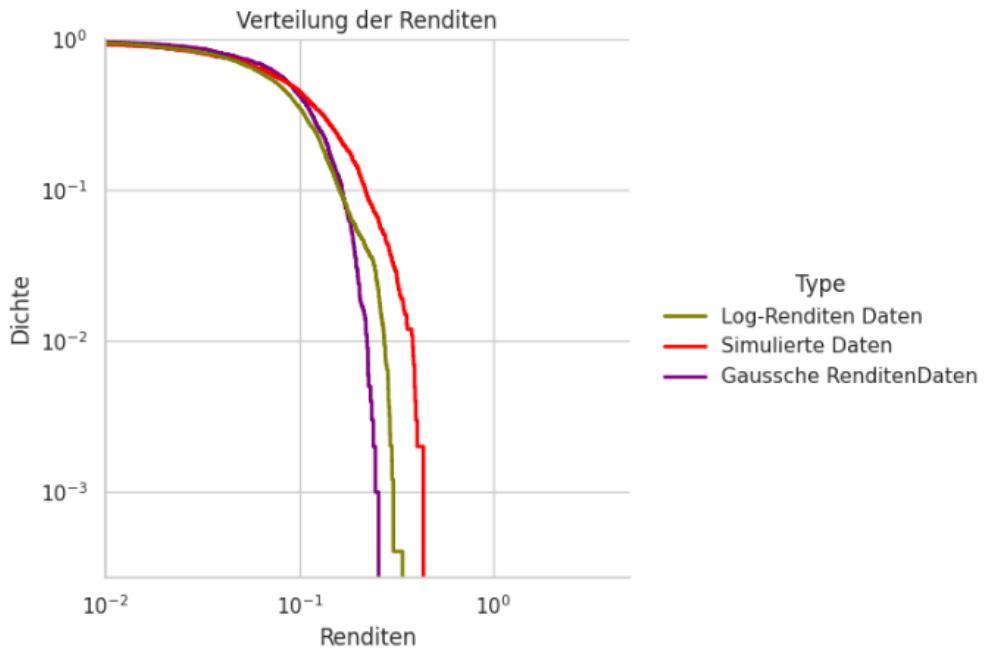


Abbildung 24: Verteilung der Renditen der Microsoft-Daten, simulierten Schrödinger-Brücken-Daten und einer Normalverteilung.

darstellt, um komplexe Verteilungen in hochdimensionalen Daten zu modellieren. Die generierten Bilder zeigen ähnliche Strukturen und Muster wie die Originalbilder, jedoch mit leicht geglätteten oder vereinfachten Details. Die Abweichungen sind auf die Bandbreite des Kernels und die Dimension der Daten zurückzuführen, was eine interessante Herausforderung für zukünftige Arbeiten darstellt. Insgesamt bestätigt die Visualisierung, dass die Schrödinger-Brücke für Zeitreihen erfolgreich hochdimensionale Bilddaten simulieren kann.

Nach der erfolgreichen Anwendung auf Graustufenbilder werden im nächsten Beispiel RGB-Bilddaten untersucht, um die Fähigkeit der Schrödinger-Brücke bei noch komplexeren Datensätzen zu validieren.

Beispiel 5.7.5 (Generierung von RGB-Bildern mit CIFAR-10-Daten)

In diesem Beispiel wird der CIFAR-10-Datensatz verwendet, der aus farbigen 32×32 -Bildern in drei Kanälen (RGB) besteht. Aufgrund der hohen Dimension der Daten werden die Bilder auf 16×16 -Pixel verkleinert, um numerische und zeitliche Einschränkungen zu berücksichtigen. Trotzdem hat jeder Datenpunkt die Dimension $16 \times 16 \times 3$, da es sich um farbige Bilder handelt.

Die Bilder werden normalisiert, indem ihre Pixelwerte durch 255 geteilt und anschließend um 0.05 verschoben werden, um numerische Instabilitäten zu vermeiden. Die resultierenden Bilder werden in flache Vektoren umgewandelt, die vom SBTs-Modell verarbeitet werden können. Insgesamt werden $M = 1000$ Bilder verwendet.

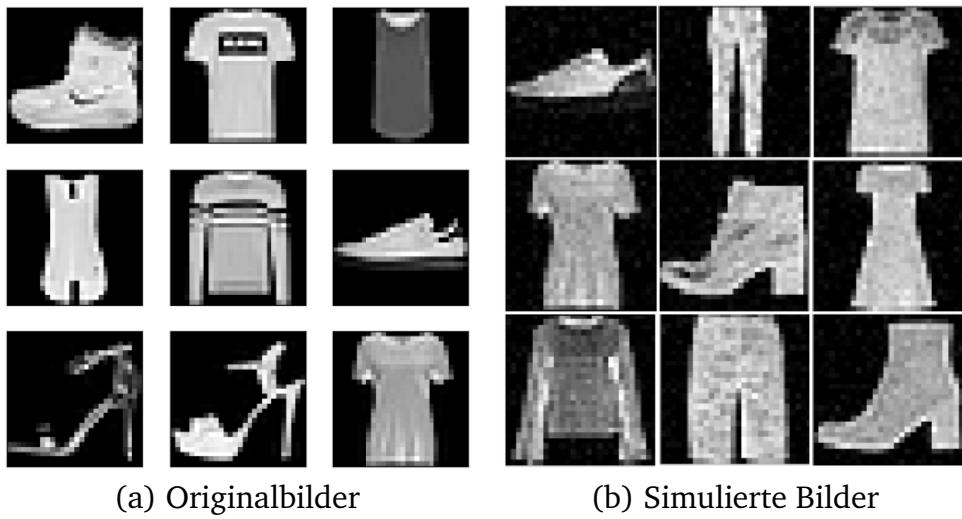


Abbildung 25: Vergleich zwischen zufällig ausgewählten Bildern aus dem MNIST-Fashion-Datensatz (links) und simulierten Bildern (rechts).

Abbildung 26 zeigt zufällig ausgewählte Originalbilder des CIFAR-10-Datensatzes im Vergleich zu simulierten Bildern, die mithilfe der Schrödinger-Brücke für Zeitreihen erzeugt wurden. Die Originalbilder weisen charakteristische Merkmale der CIFAR-10-Klassen wie Tiere und Fahrzeuge auf, während die simulierten Bilder eine reduzierte, aber dennoch erkennbare Struktur zeigen.

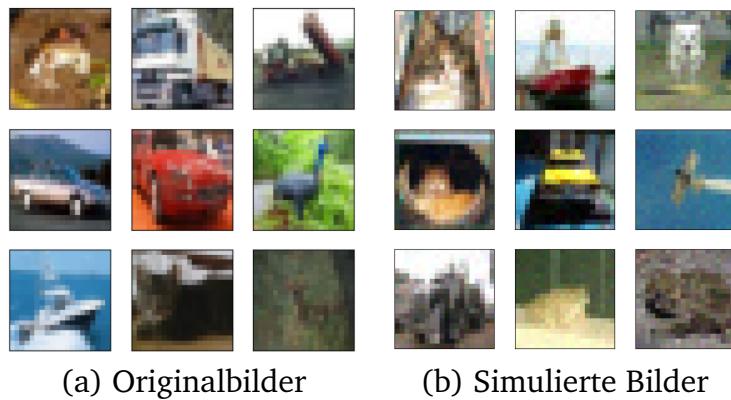


Abbildung 26: Vergleich zwischen zufällig ausgewählten Originalbildern (links) und simulierten Bildern (rechts) des CIFAR-10-Datensatzes.

Zusätzlich zur Darstellung der finalen Ergebnisse bietet Abbildung 27 eine Illustration der zeitlichen Entwicklung des gesteuerten Prozesses. Dabei wird sichtbar, wie die Schrödinger-Brücke für Zeitreihen schrittweise aus einer uniformen Startverteilung die Zielbilder erzeugt. Der Übergang veranschaulicht die iterative Annäherung des Modells an die Verteilung der Originalbilder und zeigt den Fortschritt bei der Rekonstruktion der Bildstruktur.

Die Visualisierung zeigt, dass das SBTS-Modell in der Lage ist, grundlegende Struk-

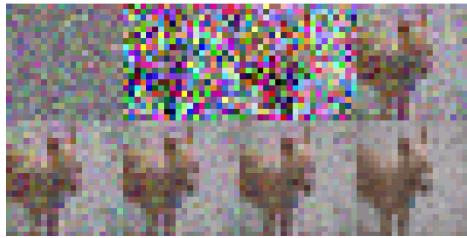


Abbildung 27: Zeitliche Entwicklung des gesteuerten Prozesses bei der Generierung von RGB-Bildern. Die Abbildung zeigt den Übergang von einer uniformen Startverteilung zu den finalen simulierten Bildern.

turmerkmale der CIFAR-10-Daten zu erfassen. Jedoch weisen die simulierten Bilder stärkere Glättungen und weniger Details auf, was auf die begrenzte Bandbreite des Kernels und die reduzierte Pixelauflösung zurückzuführen ist. Dieses Beispiel illustriert, dass die Schrödinger-Brücke auch bei RGB-Bildern ein leistungsfähiges Werkzeug zur Modellierung hochdimensionaler Verteilungen darstellt, obwohl die Qualität der Rekonstruktion stark von der Wahl der Hyperparameter und der Datenvorverarbeitung abhängt.

5.8 Abgrenzung zu verwandten Methoden

Das in diesem Kapitel vorgestellte SBTS lässt sich in den breiteren Kontext moderner generativer Modelle einordnen, welche in den letzten Jahren insbesondere im Bereich des maschinellen Lernens stark an Bedeutung gewonnen haben. In diesem Zusammenhang sind vor allem *Score-Based Diffusion Models* (auch als *DDPMs* oder *Score-Matching-Ansätze* bekannt) sowie *Normalizing Flows* zu nennen, die in vielerlei Hinsicht ähnliche Ideen wie die Schrödinger-Brücke aufgreifen, aber andere mathematische oder algorithmische Schwerpunkte setzen. Dieser Abschnitt gibt einen kurzen Überblick über zentrale Gemeinsamkeiten und Unterschiede zwischen SBTS und diesen verwandten Verfahren, um den innovativen Charakter der Schrödinger-Brücken-Methodik zu verdeutlichen.

Score-Based Diffusion Models (siehe z. B. [19, 20]) beruhen in ihrer Grundidee darauf, dass man die Daten schrittweise durch Hinzufügen von Rauschen in eine nahezu "uniforme" oder "weiß-rauschige" Verteilung überführt (Vorwärts-Diffusion) und anschließend mithilfe eines gelernten *Scores* (Gradient des Logarithmus der Wahrscheinlichkeitsdichte) wieder einen Pfad zurück in die Datenverteilung findet (Rückwärts-Diffusion). Eine spezielle Ausprägung dieses Ansatzes sind *Denoising Diffusion Probabilistic Models* (DDPM), bei denen die Rückwärts-Diffusion diskretisiert und über eine Reihe einzelner Schritte trainiert wird.

Parallelen zur Schrödinger-Brücke zeigen sich insbesondere in der Vorwärts-Rückwärts-Struktur. Auch hier kommt man von einer einfachen Referenzverteilung \mathbb{W} (dem Wiener Maß) durch eine kontrollierte Drift zur Zielverteilung μ . Beide Methoden arbeiten zudem mit *Forward-Backward-SDGLs*, die man – je nach Perspektive – als eine Art *generatives stochastisches Interpolationsproblem* auffassen kann. Ein wesent-

licher Unterschied besteht jedoch in der *Trainingsphilosophie*: Score-Based Models lernen explizit die *Score-Funktion* (also den Gradienten $\nabla_x \ln p_t(x)$) über sämtliche Zeitpunkte t , während bei SBTS direkt die zugehörige Drift $\alpha^*(t, x)$ ermittelt wird, welche aus einer *Entropie-Minimierung* (bzw. KL-Divergenz-Minimierung) folgt. Insbesondere betont das Schrödinger-Brücken-Framework die Minimierung der *relativen Entropie* gegenüber einem Referenzprozess und ist mathematisch stark in der stochastischen Kontrolltheorie verankert.

Normalizing Flows (siehe z. B. [14, 17]) verfolgen einen anderen Ansatz. Sie definieren eine deterministisch-stochastische Transformation (oft als diffeomorphes Mapping) von einer einfachen Verteilung (häufig einer Normalverteilung in \mathbb{R}^d) zu einer Zielverteilung, wobei die *Jacobi-Determinante* dieser Transformation in geschlossener Form bestimmbar bleibt. Dadurch lässt sich die Likelihood (bzw. Dichte) der Zieldaten explizit auswerten, was ein effizientes *Maximum-Likelihood-Training* ermöglicht.

Im Vergleich zur Schrödinger-Brücke unterscheidet sich die generative Perspektive von Normalizing Flows darin, dass hier durchgehend *deterministische Flüsse* (mit invertiblen Transformationen) genutzt werden, wohingegen SBTS *stochastische Pfade* konstruieren muss, um die entropische Minimierung via Girsanov-Transformation zu erfüllen. Zwar kann man Flows in einer zeitkontinuierlichen Variante (Continuous Normalizing Flows) ebenfalls als *stochastische Flüsse* interpretieren, doch bleibt das Hauptaugenmerk meist auf deterministischen ODEs. Aus mathematischer Perspektive führt die Schrödinger-Brücke darüber hinaus *Pfadanpassungen* ein, statt nur Zustandsverteilungen in \mathbb{R}^d . Dennoch besteht eine gewisse Nähe, da beide Methoden (SBTS und Flows) von einer einfachen Verteilung ausgehen und schrittweise in Richtung der Zielverteilung transformieren.

Alle drei Methoden – SBTS, Score-Based Diffusion Models und Normalizing Flows – verfolgen das Ziel, *synthetische Daten zu generieren*, die einer (hochdimensionalen) Zielverteilung möglichst nahekommen. Dabei nutzen sie jeweils eine Referenzverteilung (z. B. Gaußsche Normalverteilung oder Wiener-Prozess) sowie eine *gezielte Transformation*, um zwischen Referenz und Zielverteilung zu interpolieren. In der Praxis werden oft *Neural-Network-Approximationen* verwendet, um benötigte Funktionen (Score-Funktion, Drift oder Flow-Parameter) aus Daten zu schätzen.

Der wichtigste konzeptionelle Kontrast liegt im *stochastischen Ansatz* und *Trainingsmethodik*. SBTS stützt sich stark auf die Minimierung der *relativen Entropie* (Kullback-Leibler-Divergenz) unter Nebenbedingungen und agiert im Pfadraum, was es prädestiniert für *Zeitreihenmodelle* oder generell pfadabhängige Phänomene. *Diffusionsmodelle* (Score-Based) extrahieren und lernen die *lokalen Gradienten der Dichte* (Scores) in einem Forward-Backward-Szenario, was zwar ebenfalls stochastische Diffusionen erfordert, aber eher auf ein *Pixel- oder Feature-basiertes Rauschen* (bei Bildern) bzw. Punkt-zu-Punkt-Denoising setzt. *Normalizing Flows* setzen häufig auf *deterministische, invertierbare Transformationen*, deren Wahrscheinlichkeitsdichte man via Jacobi-Determinante

explizit kontrolliert, was in vielen Fällen die *Likelihood-basierte Optimierung* vereinfacht, bei komplexen Zeitreihen oder Pfadproblemen aber nicht immer direkt anwendbar ist.

Aus diesem Vergleich ergibt sich, dass *Schrödinger-Brücken-Methoden* eine interessante Alternative zu den verbreiteten Diffusionsmodellen und Normalizing Flows darstellen. Insbesondere bei *Zeitreihen* oder *pfadabhängigen* Anwendungen bietet SPTS durch seine stochastische Kontrollinterpretation einen sehr natürlichen Zugang. Gleichzeitig greifen *Score-Based Ansätze* denselben mathematischen Kern (Forward-Backward-SDGLs) auf, nutzen jedoch eine andere Lernstrategie, die sich besonders gut bei hochdimensionalen Bilddatensätzen bewährt hat.

Es lässt sich festhalten, dass das SPTS nicht allein, sondern im Kontext anderer moderner generativer Methoden betrachtet werden sollte. Diese *Abgrenzung* hebt sowohl die *Stärken* der Schrödinger-Brücke hervor, stochastische Kontrolle und Entropieminimierung, als auch ihre *Herausforderungen* im Vergleich zu Diffusionsmodellen und Flows. Damit wird der innovative und eigenständige Charakter der SPTS-Methodik für zeitabhängige Daten hervorgehoben und zugleich Anknüpfungspunkte für den fachlichen Austausch und die Weiterentwicklung moderner generativer Verfahren aufgezeigt.

Die in diesem Kapitel präsentierten Resultate bilden die Grundlage für den Einsatz des Schrödinger-Brücken-Ansatzes in der Zeitreihenanalyse. Im anschließenden Fazit- und Ausblick-Kapitel werden die wesentlichen Erkenntnisse zusammengefasst und potenzielle Weiterentwicklungen sowie offene Forschungsfragen erörtert. Ziel ist es, einen umfassenden Überblick über die Potenziale und Limitationen des SPTS-Ansatzes zu vermitteln und den Weg für weitere Forschungsarbeiten zu bereiten.

6 Fazit und Ausblick

Die vorliegende Arbeit hat die Anwendung des Schrödinger-Brücken-Problems (SBP) auf die generative Modellierung von Zeitreihen untersucht. Diese Methoden, die stochastische Steuerung und numerische Optimierungsverfahren verbinden, bieten innovative Möglichkeiten zur Modellierung und Simulation komplexer dynamischer Systeme. Ein zentraler Beitrag dieser Arbeit war die theoretische und praktische Untersuchung des Schrödinger-Brücken-Problems sowie des Schrödinger-Brücken-Problems für Zeitreihen (SBTS). Die effiziente Lösung der SBTS erlaubte es, stochastische Prozesse unter Einhaltung von Randbedingungen und statistischen Eigenschaften zu modellieren. Dabei bietet die vorgestellte Methode den Vorteil, dass sie ohne das Training großer neuronaler Netzwerke auskommt und direkt mit den vorhandenen Datenstichproben arbeitet. Dies reduziert nicht nur den Rechenaufwand, sondern ermöglicht auch eine einfachere und direktere Implementierung.

Die numerischen Experimente zeigten jedoch Herausforderungen in sehr hohen Dimensionen. Die berechneten Gewichte neigten dazu, exponentiell zu wachsen, was zu numerischen Instabilitäten führte. Eine mögliche Lösung könnte in der Regularisierung der Optimierungsprobleme oder in der Verwendung von skalierbaren Varianten wie dem Soft Constrained Schrödinger Bridge (SSB) liegen, der bereits kurz in dieser Arbeit eingeführt wurde.

Das Konzept des SSB bietet vielversprechende Perspektiven für zukünftige Forschung. Insbesondere könnte untersucht werden, wie SSB-Methoden genutzt werden können, um numerische Stabilität in hochdimensionalen Szenarien zu gewährleisten. Eine Python-Implementierung des SSB würde die Anwendung und Weiterentwicklung dieser Ansätze erleichtern und bietet ein wichtiges Ziel für zukünftige Studien.

Ein weiteres wichtiges Ergebnis der Arbeit war, dass die Simulation generativer Daten in vielen Fällen zu realistischen Ergebnissen führte, jedoch nicht ohne Einschränkungen blieb. Beispielsweise zeigten die numerischen Experimente leichte Unterschiede zwischen simulierten und echten Finanzzeitreihen, was auf Optimierungspotenziale bei der Modellerstellung hinweist. Hingegen verlief die Generierung synthetischer Bilder, wie etwa mit dem MNIST- oder CIFAR-10-Datensatz, erfolgreich, solange die Dimensionen nicht zu hoch wurden. Diese Ergebnisse unterstreichen die Notwendigkeit von Methoden zur Reduktion von Dimensionen oder speziell angepassten Algorithmen für hochdimensionale Daten.

Zusammenfassend hat diese Arbeit gezeigt, dass die Schrödinger-Brücke eine vielseitige und mächtige Methode darstellt, die sowohl theoretische Tiefe als auch praktische Relevanz besitzt. Die Erweiterung der hier entwickelten Algorithmen um SSB-Ansätze und deren Anwendung auf hochdimensionale Probleme bietet eine spannende Möglichkeit, die generative Modellierung weiterzuentwickeln. Darüber hinaus könnte die Kombination mit maschinellem Lernen eine zusätzliche Ebene der Effizienz und Skalierbarkeit schaffen, insbesondere bei der Modellierung komplexer zeitlicher Dynamiken.

Literatur

- [1] Hirotugu Akaike. Canonical correlation analysis of time series and the use of an information criterion. In Raman K. Mehra and Dimitri G. Lainiotis, editors, *System Identification Advances and Case Studies*, volume 126 of *Mathematics in Science and Engineering*, pages 27–96. Elsevier, 1976. 41
- [2] Tim Bollerslev. Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, 31(3):307–327, 1986. 27
- [3] George E. P. Box, Gwilym M. Jenkins, Gregory C. Reinsel, and Greta M. Ljung. *Time Series Analysis: Forecasting and Control*. Wiley, Hoboken, NJ, 5th edition, 2015. 23, 26
- [4] Peter J. Brockwell and Richard A. Davis. *Introduction to Time Series and Forecasting*. Springer Texts in Statistics. Springer, Cham, Switzerland, 3rd edition, 2016. 23
- [5] Yongxin Chen, Tryphon Georgiou, and Michele Pavon. On the relation between optimal transport and schrödinger bridges: A stochastic control viewpoint, 2014. 60
- [6] Robert F. Engle. Autoregressive conditional heteroscedasticity with estimates of the variance of united kingdom inflation. *Econometrica*, 50(4):987–1007, 1982. 27
- [7] Wendell Fleming. Exit probabilities and stochastic control theory. *Applied Mathematics and Optimization*, 4:329–346, 1978. 40
- [8] Jhanvi Garg, Xianyang Zhang, and Quan Zhou. Soft-constrained schrödinger bridge: A stochastic control approach, 2024. 68, 69
- [9] Mohamed Hamdouche, Pierre Henry-Labordere, and Huyêñ Pham. Generative modeling for time series via schrödinger bridge, 2023. 59, 60
- [10] Benton Jamison. The markov processes of schrödinger. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, 32(4):323–331, 1975. 37
- [11] Ioannis Karatzas and Steven E. Shreve. *Brownian Motion and Stochastic Calculus*, volume 113 of *Graduate Texts in Mathematics*. Springer, New York, 1991. 21
- [12] Achim Klenke. *Wahrscheinlichkeitstheorie: Eine Einführung in die mathematischen Grundlagen*. Springer, Berlin, Heidelberg, 5. auflage edition, 2020. 2
- [13] Christian Léonard. A survey of the schrödinger problem and some of its connections with optimal transport, 2013. 34

- [14] George Papamakarios, Eric Nalisnick, Danilo Jimenez Rezende, Shakir Mohamed, and Balaji Lakshminarayanan. Normalizing flows for probabilistic modeling and inference. *Journal of Machine Learning Research*, 22(57):1–64, 2021. 88
- [15] Paolo Dai Pra. A stochastic control approach to reciprocal diffusion processes. *Applied Mathematics and Optimization*, 23(1):313–329, 1991. 35, 37, 52, 60
- [16] Daniel Revuz and Marc Yor. *Continuous martingales and Brownian motion*. Number 293 in Grundlehren der mathematischen Wissenschaften. Springer, Berlin [u.a.], 3. ed edition, 1999. 2, 4
- [17] Danilo Jimenez Rezende and Shakir Mohamed. Variational inference with normalizing flows. In *International Conference on Machine Learning (ICML)*, 2015. 88
- [18] Erwin Schrödinger. Über die umkehrung der naturgesetze. *Sitzungsberichte der Preußischen Akademie der Wissenschaften, Physikalisch-mathematische Klasse*, pages 144–153, 1931. 34, 35, 40
- [19] Yang Song and Stefano Ermon. Generative modeling by estimating gradients of the data distribution. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2019. 87
- [20] Yang Song, Jascha Sohl-Dickstein, Diederik P. Kingma, A. Kumar, Stefano Ermon, and Ben Poole. Score-based generative modeling through stochastic differential equations. In *International Conference on Learning Representations (ICLR)*, 2021. 87
- [21] Gefei Wang, Yuling Jiao, Qian Xu, Yang Wang, and Can Yang. Deep generative learning via schrödinger bridge, 2021. 74, 75
- [22] Bernt Øksendal. *Stochastic Differential Equations: An Introduction with Applications*. Universitext. Springer Berlin, Heidelberg, 6th edition, 2003. Springer Book Archive, Copyright Springer-Verlag Berlin Heidelberg 2003. 15, 21

Eidesstattliche Erklärung

Der Verfasser erklärt an Eides statt, dass er die vorliegende Arbeit selbstständig, ohne fremde Hilfe und ohne Benutzung anderer als die angegebenen Hilfsmittel angefertigt hat.

Die aus fremden Quellen, einschließlich elektronischer Quellen, direkt oder indirekt übernommenen Gedanken sind ausnahmslos als solche kenntlich gemacht.

Die Arbeit ist in gleicher oder ähnlicher Form oder auszugsweise im Rahmen einer anderen Prüfung noch nicht vorgelegt worden.

Ort, Datum

Unterschrift