

Analyse stochastischer Prozesse

Eine iterative Methode zur Bestimmung von Drift und Diffusion

als Diplomarbeit vorgelegt von
David Kleinhans

Juni 2005



Westfälische Wilhelms-Universität Münster

Die Sorge um die Menschen und ihr Schicksal muß stets das Hauptinteresse allen technischen Strebens bilden, die großen ungelösten Fragen der Organisation von Arbeit und der Güterverteilung, damit die Erzeugnisse unseres Geistes dem Menschengeschlecht zum Segen gereichen und nicht zum Fluche. Vergeßt dies nie über Euren Zeichnungen und Gleichungen.

ALBERT EINSTEIN, Februar 1931

Inhaltsverzeichnis

Einleitung	1
1 Einführung in die Theorie stochastischer Prozesse	3
1.1 Mittelwerte und bedingte Verteilungen	3
1.1.1 Motivation	3
1.1.2 Zeitabhängige Verteilungen	4
1.1.3 Verbundwahrscheinlichkeiten	5
1.1.4 Bayes'sches Theorem	5
1.2 Grundlagen stochastischer Prozesse	6
1.2.1 Markow-Prozesse	6
1.2.2 Chapman-Kolmogorov-Gleichung	7
1.2.3 Langevin-Gleichungen	7
1.2.4 Interpretation von Stieltjes-Integralen nach Itô und Stratonovich	8
1.2.5 Langevin-Gleichungen mit endlichem Zeitinkrement	9
1.2.6 Fokker-Planck-Gleichung	10
1.2.7 Stationäre Lösung der Fokker-Planck-Gleichung	11
2 Schätzen von Drift und Diffusion	13
2.1 Statistik von Zeitreihen	13
2.2 Schätzen der Koeffizienten	15
2.3 Beispiele	16
2.3.1 Physiologische und pathologische Fluktuation der Herzfrequenz	16
2.3.2 Analyse von Straßenverkehrs-Daten	18
2.3.3 Nachweis des Parkinson-Syndroms	19
3 Fehler beim Schätzen von Drift und Diffusion	23
3.1 Schätzung von $D^{(1)}$ und $D^{(2)}$ für endliche τ	23
3.1.1 Berechnung des ersten bedingten Momentes	24
3.1.2 Berechnung des zweiten bedingten Momentes	25
3.1.3 Einfluss auf die in Kapitel 2 beschriebene Methode zur Datenanalyse	26
3.1.4 Beispiel: Ornstein-Uhlenbeck-Prozess	27
3.2 Einfluss einer endlichen Stichprobe auf die Schätzungen	27
3.3 Durch Diskretisierung des Ortsraums zu erwartende Fehler	31
3.4 Einfluss von Messrauschen	33
3.4.1 Berechnung des ersten Momentes	34
3.4.2 Berechnung des zweiten Momentes	35
3.4.3 Beispiel: Ornstein-Uhlenbeck-Prozess	36

4 Iteratives Verfahren zur Bestimmung von Drift und Diffusion	41
4.1 Allgemeines Verfahren	41
4.2 Eigenschaften des Kullback-Maßes	42
4.3 Vereinfachungen für einige Spezialfälle	43
4.3.1 Eindimensionale Systeme	43
4.3.2 Potential-Systeme	44
5 Beispiele	45
5.1 Numerische Grundlagen	45
5.1.1 Integration von Langevin-Gleichungen	45
5.1.2 Numerische Integration der Fokker-Planck-Gleichung	46
5.1.3 Numerische Ableitung von Daten	47
5.1.4 Monte-Carlo-Techniken zur Integration	48
5.2 Eine Dimension, additives Rauschen	49
5.2.1 Analyse der Zeitreihe	54
5.2.2 Numerische Umsetzung	55
5.3 Eine Dimension, multiplikatives Rauschen	56
5.3.1 Analyse der Zeitreihe	57
5.3.2 Numerische Umsetzung	58
5.4 Zweidimensionales System mit additivem Rauschen	60
5.4.1 Diskussion des Potentials	62
5.4.2 Numerische Umsetzung	66
5.4.3 Analyse der Zeitreihe	67
6 Zusammenfassung	69
A Ausgegliederte Berechnungen	71
A.1 Berechnung von Integralen wie (3.10c) durch partielle Integration	71
A.2 Auswirkung der Diskretisierung: Auswertung von (3.22)	74
B Kommentierter Programmcode	75
Literaturverzeichnis	81

Einleitung

Mitte des 19. Jahrhunderts schien das Ende der grundlegenden Naturwissenschaften aus Sicht mancher Forscher schon fast besiegelt. Es herrschte die Ansicht, dass die aus heutiger Sicht „klassische Physik“ die Natur hinreichend genau beschrieb. Die Erklärung einiger aus Experimenten bekannter Ergebnisse wollte jedoch nicht gelingen.

Wenige Jahre später änderte sich vieles: Planck und Einstein begründeten mit der Beschreibung der Hohlraumstrahlung [1] und der Erklärung des lichtelektrischen Effektes [2] die Quantenphysik. Mit der Erkenntnis, dass zum Verständnis der offen gebliebenen Fragen ein vollkommen neuer Formalismus benötigt wurde, der das klassische Bild in sich zusammenbrechen ließ, taten sich eine große Anzahl neuer Fragen auf. Sollte der so anschauliche, vollkommene Determinismus einem Modell weichen müssen, das auf den kleinsten Längenskalen nur „Wahrscheinlichkeiten“ kennt? Die Frage nach der Existenz kleinster Einheiten hatte Philosophen schon seit jeher beschäftigt und es bestand bis zu diesem Zeitpunkt weitgehend die Vorstellung, dass jeder physikalischen Prozess vollkommen kontinuierlich ablaufen müsse, wie man es von makroskopischen Prozessen gewohnt war. Nun zeigte sich, dass sich manche Phänomene nur erklären ließen, wenn man mutig genug war, eine solche Ansicht zu verwerfen.

Durch Einführung eines neuen Konzepts hat sich auch das Verständnis *komplexer Systeme* verändert. Komplexe Systeme treten in der Natur – in allen Disziplinen der Wissenschaften – auf und sind charakterisiert durch die hohe Anzahl an Freiheitsgraden, die ihren Ursprung in der hohen Anzahl gekoppelter Untersysteme hat. So kann man z.B. turbulente Strömung [3], Transportprozesse granularer Materie [4], ökonomische Prozesse [5] sowie fast alle Prozesse, die sich im menschlichen Körper abspielen [6], als „komplex“ bezeichnen. Prinzipiell ist eine Lösung der Entwicklungsgleichungen der einzelnen Freiheitsgrade in einigen Fällen möglich. Auf der anderen Seite ist man üblicherweise mit zwei Problemen konfrontiert: Die genauen Anfangsbedingungen sowie die mikroskopischen Wechselwirkungen sind im Allgemeinen nicht hinreichend genau bekannt.

Findet ein Energie-Fluss durch das System statt, handelt es sich also um ein System fernab vom thermodynamischen Gleichgewicht, so können Effekte der *Strukturbildung* und *Selbstorganisation* auftreten. In diesem Fall zeigt das Kollektiv der einzelnen, gekoppelten Systeme ein Verhalten, das auf mikroskopischer Ebene – im isolierten Zustand – nicht zu vermuten wäre. H. Haken hat für das Verständnis dieses Phasenübergangs in den 60er Jahren das Konzept der *Ordnungsparameter* [7] eingeführt: In der Nähe von Instabilitäten kann man die Freiheitsgrade des Systems in Moden entwickeln, von denen einige instabil, die meisten aber stabil sind. Stabile Moden folgen – so fand Haken heraus – den instabilen Moden sehr schnell, sie werden *versklavt*. Man kann daher die instabilen Moden als Ordnungsparameter des Systems betrachten und die stabilen Moden adiabatisch aus den Bewegungsgleichungen eliminieren. Die Dynamik eines komplexen Systems wird – auf makroskopischer Ebene und endlichen Zeitskalen – nun durch die Dynamik einiger weniger Ordnungsparameter vollständig beschrieben. Die mikroskopischen Freiheitsgrade des Systems

tauchen in diesen Bewegungsgleichungen in Form von Fluktionskräften auf. Daher handelt es sich um stochastische Differentialgleichungen. Weil das Konzept, das ich hier kurz umrissen habe, so umfassend ist und sich die Anwendung über die Physik hinaus auch auf die anderen Naturwissenschaften bis hin zu Medizin, Soziologie und Ökonomie erstreckt, spielt die Analyse stochastischer Prozesse eine wichtige Rolle.

Durch Einführung und Weiterentwicklung von Computern haben sich die zur Verfügung stehenden Mittel in den letzten Jahrzehnten verändert; die Möglichkeiten, die der Wissenschaft dadurch zur Verfügung stehen, werden stets erweitert. Mittlerweile erstreckt sich der Anwendungsbereich neben der Durchführung aufwändiger Berechnungen auch auf die Datenerfassung. Es ist somit möglich, große Datenmengen zu erheben und zu bearbeiten. Welchen Einfluss hat das auf die Analyse komplexer Systeme? Es ist zu vermuten, dass die Lösung der mikroskopischen Bewegungsgleichungen für reale Systeme (mit $\sim 10^{23}$ Teilchen) auch in Zukunft nicht möglich sein wird. Wir können aber nach den Ordnungsparametern komplexer Systeme suchen und diese studieren. Dann ist es das Ziel, die Dynamik der Ordnungsparameter aus den Daten zu extrahieren und zu verstehen [8].

Es existiert eine Vielzahl von Verfahren, die einen Rückschluss von gemessenen Daten auf die Dynamik erlauben und sich im Aufwand der Durchführung und in den Anforderungen an die Daten stark voneinander unterscheiden. Vor einigen Jahren ist von Friedrich et al. gezeigt worden, dass sich bei sogenannten *Markow-Prozessen*, die sich durch die Angabe von Drift- und Diffusionsfunktionen beschreiben lassen, die Dynamik mit Hilfe eines einfachen Verfahrens direkt – d.h. ohne dass im Voraus Annahmen über die Struktur der Funktionen gemacht werden müssen – aus den Daten gewinnen lässt [9]. Es ist bekannt, dass diese Methode nur eingeschränkt funktioniert, wenn die Abtastrate der Daten, die analysiert werden sollen, nicht hoch genug ist. Wie in diesem Fall vorzugehen ist, wird seitdem diskutiert [10, 11, 12]. Diese Diskussion bildet den Ausgangspunkt für die vorliegende Arbeit.

Nach einer kurzen Einführung in die mathematischen Grundlagen wird in Kapitel 2 das Verfahren zur direkten Schätzung von Drift- und Diffusionsfunktionen vorgestellt. Anhand einiger Beispiele werden Anwendungsgebiete solcher Methoden demonstriert. Im darauf folgenden Kapitel werden Probleme, die bei der Schätzung der Funktionen auftreten, wie z.B. die bei niedrigen Samplingfrequenzen auftretenden Abweichungen, diskutiert. Ferner wird untersucht, wie sich bei der Messung der Daten entstandenes Rauschen auf die Ergebnisse auswirkt. Weil das ursprüngliche Verfahren in diesem Fall nicht mehr funktioniert, wird nach einer Modifikation gesucht. In Kapitel 4 wird dann ein neues, allgemeines Verfahren zur quantitativen Gewinnung von Drift und Diffusion [13] eingeführt, das auf der Minimierung des Kullback-Abstandes zweier Wahrscheinlichkeitsverteilungen beruht und die Unzulänglichkeiten der direkten Schätzung umgeht. Insbesondere ist dieses neue Verfahren nicht auf hohe Samplingfrequenzen beschränkt. Schließlich werden einige Beispiele behandelt, die mit Hilfe dieses Verfahrens untersucht worden sind. Dabei wird auf eine für weitere Arbeiten auf diesem Gebiet hilfreiche Dokumentation der numerischen Implementierung des Verfahrens Wert gelegt.

Kapitel 1

Einführung in die Theorie stochastischer Prozesse

Begonnen sei mit einer Einführung in die Grundlagen der Wahrscheinlichkeitstheorie und der stochastischen Prozesse, die für das weitere Verständnis dieser Arbeit notwendig sind. Dabei werde ich mich weitgehend an den Standardwerken von Risken [14] und Gardiner [15] sowie der im Wintersemester 2004/05 von Friedrich und Frank am hiesigen Institut gehaltenen Vorlesung [16] orientieren, die eine problemorientierte Einführung in die Thematik bieten.

Auf eine maßtheoretische Fundierung des Wahrscheinlichkeitsbegriffes wird verzichtet. Diese findet sich zum Beispiel bei Paul & Baschnagel [17], Dingel & Rost [18] sowie in mathematischer Einführungsliteratur. Speziellere Ausführungen zu einigen Themen sind den Büchern von Haken [7] (Grundlagen der Synergetik, Ordnungsparameter), Coffey [19] (Langevin-Gleichungen) und Klöden & Platen [20] (Numerik stochastischer Differentialgleichungen) zu entnehmen.

1.1 Mittelwerte und bedingte Verteilungen

1.1.1 Motivation

Ausgangspunkt sei ein Experiment, das ein Ergebnis ξ liefert, dessen Ausgang nicht vorherzusagen sei. ξ nennt man dann eine *Zufallsvariable*. Wird das Experiment N mal wiederholt, so haben wir eine Reihe von Realisierungen ξ_1, \dots, ξ_N unserer Zufallsvariablen.

Diese Ergebnisse ξ können zunächst alles darstellen: Zahlen, Vektoren, Meinungen, Farben, \dots . Wir haben im alltäglichen Leben gelernt, Zufallsvariablen zu akzeptieren und mit ihnen zu arbeiten: Wir wissen, dass uns die Mittelung über eine große Zahl solcher Variablen häufig wichtige Informationen gibt, die wir einzelnen Werten nicht entnehmen können.

Der einfachste Mittelwert ist

$$\langle \xi \rangle_{\xi} = \frac{1}{N} (\xi_1 + \xi_2 + \dots + \xi_N) \quad , \quad (1.1)$$

allgemein gilt für Funktionen h von Zufallsvariablen

$$\langle h(\xi) \rangle_{\xi} := \frac{1}{N} [h(\xi_1) + h(\xi_2) + \dots + h(\xi_N)] \quad . \quad (1.2)$$

Sei zunächst $\xi \in \mathbb{R}$, die Ausdrücke für $\xi \in \mathbb{R}^n$ folgen analog. Mit Hilfe der Stufen-

funktion

$$\Theta(x) = \begin{cases} 1 & x > 0 \\ \frac{1}{2} & \text{für } x = 0 \\ 0 & x < 0 \end{cases} \quad (1.3)$$

lässt sich eine Wahrscheinlichkeitsverteilung F definieren:

$$F(x) := \langle \Theta(x - \xi) \rangle_\xi \quad (1.4)$$

Aus (1.3) ergeben sich für F folgende Eigenschaften:

$$0 \leq F(x) \leq 1 \quad , \quad F(x \rightarrow -\infty) = 0 \quad \text{sowie} \quad F(x \rightarrow +\infty) = 1 \quad . \quad (1.5)$$

Nun kann die Wahrscheinlichkeitsdichte (PDF) $f_\xi(x)$ als Ableitung von F nach x eingeführt werden:

$$\begin{aligned} f_\xi(x) &:= \frac{\partial}{\partial x} F(x) = \frac{\partial}{\partial x} \langle \Theta(x - \xi) \rangle \\ &= \left\langle \frac{\partial}{\partial x} \Theta(x - \xi) \right\rangle = \langle \delta(x - \xi) \rangle \quad . \end{aligned} \quad (1.6)$$

Aufgrund der Monotonie von F ist f_ξ positiv, ferner ist die Normierung $\int_{-\infty}^{\infty} dx f_\xi(x) = 1$ wegen (1.5) stets erfüllt. Mit Hilfe von f_ξ können nun sämtliche Mittelwerte von Funktionen h der Zufallsvariablen ξ berechnet werden:

$$\begin{aligned} \langle h(\xi) \rangle_\xi &= \left\langle \int dx h(x) \delta(x - \xi) \right\rangle_\xi = \int dx h(x) \langle \delta(x - \xi) \rangle_\xi \\ &= \int dx h(x) f_\xi(x) \end{aligned} \quad (1.7)$$

Ein Beispiel für eine wichtige Verteilungsfunktion, die im Folgenden benötigt wird, ist die *Gauß'sche Normalverteilung*, kurz *Gauß-Verteilung*,

$$f_\sigma(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{x^2}{2\sigma^2}} \quad . \quad (1.8)$$

Sie ergibt sich als PDF der Summe sehr vieler, von einander unabhängiger Zufallsvariablen mit Mittelwert 0. Der Mittelwert von x (das erste Moment) über diese PDF verschwindet, weil f_σ in eine gerade Funktion in x ist. Für das zweite Moment ergibt sich:

$$\langle x^2 \rangle = \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{\infty} dx x^2 e^{-\frac{x^2}{2\sigma^2}} = \sigma^2 \quad . \quad (1.9)$$

σ^2 wird als Varianz der Verteilung bezeichnet.

1.1.2 Zeitabhängige Verteilungen

Zufallsvariablen lassen sich auch als Funktion der Zeit betrachten, zur Bildung von Mittelwerten sollen aber nur zur gleichen Zeit gemessene Variablen herangezogen werden. Eine Menge von zur gleichen Zeit gemessenen Zufallsvariablen bezeichnet man als *Ensemble*. Sei $\xi_i(t)$ der Wert der Variablen i zur Zeit t . Das *Ensemble-Mittel* zur Zeit t ist dann gegeben als:

$$f_{\xi(t)}(x, t) = \langle \delta(x - \xi(t)) \rangle = \frac{1}{N} \sum_{i=1}^N \xi_i(t) \quad . \quad (1.10)$$

Für zeitabhängige Zufallsvariablen werden in der Regel auch die PDF's explizit zeitabhängig.

1.1.3 Verbundwahrscheinlichkeiten

Zur Analyse zeitabhängiger Prozesse ist die Einführung von *Verbundwahrscheinlichkeiten* (JPDF's) hilfreich. So bezeichnet

$$f_{\xi(t_1), \dots, \xi(t_N)}^N(x_1, t_1; x_2, t_2; \dots; x_N, t_N) dx_1 \cdot \dots \cdot dx_N \quad (1.11)$$

mit

$$f_{\xi(t_1), \dots, \xi(t_N)}^N(x_1, t_1; x_2, t_2; \dots; x_N, t_N) := \langle \delta(x_1 - \xi(t_1)) \cdot \dots \cdot \delta(x_N - \xi(t_N)) \rangle$$

und $t_1 \geq t_2 \geq \dots \geq t_N$ die Wahrscheinlichkeit für eine Folge von Messungen, bei denen für alle i die Zufallsvariable zur Zeit t_i jeweils im Intervall $[x_i, x_i + dx_i]$ liegt. Durch Integrationen über x_i lassen sich einzelne Teile solcher Verbundwahrscheinlichkeiten eliminieren:

$$\begin{aligned} & \int dx_N f_{\xi(t_1), \dots, \xi(t_N)}^N(x_1, t_1; \dots; x_{N-1}, t_{N-1}; x_N, t_N) \\ &= f_{\xi(t_1), \dots, \xi(t_{N-1})}^{N-1}(x_1, t_1; \dots; x_{N-1}, t_{N-1}) \end{aligned} \quad (1.12)$$

Ebenso wichtig ist die Frage nach einer Wahrscheinlichkeits für $\xi(t_1), \dots, \xi(t_L)$ unter der Voraussetzung, dass die Variable zu den Zeiten t_L, \dots, t_N bekannt ist. Verteilungen, die solche Informationen liefern, werden als *bedingte Verteilungen* (CPDF's) bezeichnet. Die jeweilige Konditionierung wird durch einen vertikalen Strich im Argument gekennzeichnet. Für die bedingte Verteilung p muss in jedem Fall gelten:

$$\begin{aligned} & f_{\xi(t_1), \dots, \xi(t_N)}^N(x_1, t_1; \dots; x_N, t_N) \\ &= p_{\xi(t_1)}^{1,N}(x_1, t_1 | x_2, t_2; \dots; x_N, t_N) f_{\xi(t_2), \dots, \xi(t_N)}^{N-1}(x_2, t_2; \dots; x_N, t_N) \end{aligned} \quad (1.13)$$

Allgemein sind bedingte Verteilungen demnach durch

$$\begin{aligned} & p_{\xi(t_1), \dots, \xi(t_L)}^{L,N}(x_1, t_1; \dots; x_L, t_L | x_{L+1}, t_{L+1}; \dots; x_N, t_N) \\ &:= \frac{f_{\xi(t_1), \dots, \xi(t_N)}^N(x_1, t_1; \dots; x_N, t_N)}{f_{\xi(t_{L+1}), \dots, \xi(t_N)}^{N-L}(x_{L+1}, t_{L+1}; \dots; x_N, t_N)} \end{aligned} \quad (1.14)$$

gegeben.

Bislang wurden die Variablen, über die zu mitteln war, und die Verteilungen, über die diese Mittelung jeweils zu geschehen hatte, mit Hilfe der Indizes ξ_i , N und L gekennzeichnet. Aus Gründen der Übersichtlichkeit fallen diese Indizes von nun an weg. Aus (1.14) wird dann in der verkürzten Notation:

$$p(x_1, t_1; \dots; x_L, t_L | x_{L+1}, t_{L+1}; \dots; x_N, t_N) = \frac{f(x_1, t_1; \dots; x_N, t_N)}{f(x_{L+1}, t_{L+1}; \dots; x_N, t_N)} \quad (1.15)$$

1.1.4 Bayes'sches Theorem

Lassen wir die Zeitordnung der Messungen für einen kurzen Moment außer Acht, so lässt sich die JPDF zweier Ereignisse x und y schreiben als:

$$f(x; y) = p(x|y)f(y) = p(y|x)f(x) \quad . \quad (1.16)$$

Daraus ergibt sich der Zusammenhang, der als *Bayes'sches Theorem* bekannt geworden ist:

$$p(x|y) = \frac{p(y|x)f(x)}{f(y)} \quad (1.17)$$

Man bezeichnet hier $f(x)dx$ als *a priori*- und $p(x|y)dx$ als *a posteriori*-Wahrscheinlichkeit für das Auftreten von $x' \in [x, x + dx]$. Unter Annahme einer gewissen Ausgangs-Verteilung von x und dem Einfluss von x auf die Verteilung von y lässt sich so die Verteilung von x unter gegebenem y bestimmen. Auf diesem Verfahren, das sich iterativ durchführen lässt, basieren die sogenannten *Maximum-Likelihood-Methoden* [21] zur Bestimmung von Parametern eines gegebenen Prozesses.

Die bislang eingeführten Zusammenhänge gelten auch für Fälle, in denen die Zufallsvariablen vektorwertigen Charakter haben.

1.2 Grundlagen stochastischer Prozesse

Ein *stochastischer Prozess* ist die Entwicklung einer Zufallsvariablen $\xi(t) \in \mathbb{R}^n$ mit der Zeit. $\xi(t)$ wird in diesem Zusammenhang auch als *stochastische Variable* bezeichnet.

Beim einfachsten stochastischen Prozess, den man sich vorstellen kann, sind die einzelnen Messungen zu verschiedenen Zeiten *unkorreliert*, also *statistisch unabhängig*. In diesem Fall haben die vergangenen Messungen keinen Einfluss auf den Ausgang einer späteren Messung. Damit vereinfachen sich die bedingten PDF's zu einfachen Einpunkt-PDF's:

$$\begin{aligned} & f(\mathbf{x}_1, t_1; \dots; \mathbf{x}_N, t_N) \\ &= p(\mathbf{x}_1, t_1 | \mathbf{x}_2, t_2; \dots; \mathbf{x}_N, t_N) \dots p(\mathbf{x}_{N-1}, t_{N-1} | \mathbf{x}_N, t_N) f(\mathbf{x}_N, t_N) \\ &= f(\mathbf{x}_1, t_1) \dots f(\mathbf{x}_N, t_N) \end{aligned} \quad (1.18)$$

Die Mehrpunkt-Verbundwahrscheinlichkeiten faktorisieren also für unkorrelierte Zufallsereignisse.

Solche Prozesse sind jedoch unphysikalisch, da die Dynamik in der Regel zumindest stetig in der Zeit sein muss. Stetigkeit jedoch setzt – schon im streng mathematischen Sinne – eine endliche Korrelation von in kurzem Abstand gemessenen Werten voraus.

1.2.1 Markow-Prozesse

Wir betrachten daher den einfachstmöglichen Prozess, der ein gewisses *Gedächtnis* berücksichtigt. Zerfällt jede JPDF in ein Produkt aus Zweipunkt-CPDF's, so wird ein Prozess als *Markow-Prozess* bezeichnet. In diesem Falle gilt also:

$$\begin{aligned} & f(\mathbf{x}_1, t_1; \dots; \mathbf{x}_N, t_N) \\ &= p(\mathbf{x}_1, t_1 | \mathbf{x}_2, t_2; \dots; \mathbf{x}_N, t_N) \dots p(\mathbf{x}_{N-1}, t_{N-1} | \mathbf{x}_N, t_N) f(\mathbf{x}_N, t_N) \quad (1.19) \\ &= p(\mathbf{x}_1, t_1 | \mathbf{x}_2, t_2) p(\mathbf{x}_2, t_2 | \mathbf{x}_3, t_3) \dots p(\mathbf{x}_{N-1}, t_{N-1} | \mathbf{x}_N, t_N) f(\mathbf{x}_N, t_N) \end{aligned}$$

Die Angabe eines einzelnen Zustandes der Zufallsvariablen in der Vergangenheit bestimmt also, bei gegebenen Zweipunkt-CPDF's, die Statistik für spätere Zeiten vollständig. Es besteht die Vermutung, dass man alle Prozesse, die in der Natur auftreten, näherungsweise als Markow-Prozesse schreiben kann, wenn man die Dimension des Zustandsvektors nur weit genug erhöht. Auf das Problem der *Einbettung* wird in Kapitel 2 eingegangen.

1.2.2 Chapman-Kolmogorov-Gleichung

Aus der allgemeinen Gültigkeit von (1.12) folgt für eine beliebige bedingte Wahrscheinlichkeitsverteilung

$$\begin{aligned} p(\mathbf{x}_1, t_3 | \mathbf{x}_3, t_3) &= \int d^n x_2 \, p(\mathbf{x}_1, t_1; \mathbf{x}_2, t_2 | \mathbf{x}_3, t_3) \\ &= \int d^n x_2 \, p(\mathbf{x}_1, t_1 | \mathbf{x}_2, t_2; \mathbf{x}_3, t_3) p(\mathbf{x}_2, t_2 | \mathbf{x}_3, t_3) \quad . \end{aligned} \quad (1.20)$$

Für Markow-Prozesse vereinfacht sich dieser Ausdruck zu:

$$p(\mathbf{x}_1, t_3 | \mathbf{x}_3, t_3) = \int d^n x_2 \, p(\mathbf{x}_1, t_1 | \mathbf{x}_2, t_2) p(\mathbf{x}_2, t_2 | \mathbf{x}_3, t_3) \quad (1.21)$$

Diese Gleichung ist als *Chapman-Kolmogorov-Gleichung* bekannt. Sie ermöglicht die Entwicklung der Dynamik von Wahrscheinlichkeitsdichten durch eine beliebige Anzahl elementarer Prozesse und führt diese so auf Pfad-Integrale zurück.

1.2.3 Langevin-Gleichungen

Differentialgleichungen erster Ordnung in der Zeit, in denen eine stochastische, fluktuierende Kraft auftritt, fallen in der Klasse der Langevin-Gleichungen. Ihre Verwendung geht zurück auf Paul Langevin, der im Jahre 1908 mittels solcher Gleichungen den Diffusionsprozess der *Brown'schen Bewegung* erklären konnte [22]. Differentialgleichungen erster Ordnung beschreiben die zeitliche Entwicklung eines Parameters unter einer Anfangsbedingung. Für einzelne Teilchen gehen sie aus der Newtonschen Bewegungsgleichung als Gleichung für die Geschwindigkeit unter Einwirkung einer Stokes'schen Reibungskraft hervor. Sie lassen sich auch als Bewegungsgleichung eines überdämpften Teilchens verstehen.

Für eine stochastische Variable ξ lautet die allgemeine, nichtlineare Langevin-Gleichung:

$$\dot{\xi}_i = h_i(\xi, t) + \sum_j g_{ij}(\xi, t) \Gamma_j(t) \quad (1.22)$$

Dabei soll die Langevin-Kraft $\Gamma(t)$ eine gaußverteilte, deltakorrelierte Zufallsvariable mit verschwindenden Mittelwert sein. Aufgrund der konstanten spektralen Dichte nennt man Γ in Anlehnung an das Spektrum des Lichtes auch *weißes Rauschen*. Es soll also gelten:

$$\langle \Gamma(t) \rangle = 0 \quad \text{und} \quad \langle \Gamma_i(t) \Gamma_j(t') \rangle = 2 \delta_{ij} \delta(t - t') \quad (1.23)$$

Die Langevin-Kraft ist im Folgenden auf eine Amplitude von $\langle \Gamma_i(t) \Gamma_i(t) \rangle = 2$ normiert¹. Die Eigenschaften der Langevin-Kraft sind eng mit den Markow-Eigenschaften des Prozesses verknüpft. Langevin-Gleichungen beschreiben die Dynamik der stochastischen Variablen ξ . Mit ihrer Hilfe können z.B. Teilchenbahnen eines sich unter stochastischen Einflüssen bewegenden Teilchens simuliert werden. Formal lassen sie sich durch

$$\begin{aligned} \xi_i(t) &= \xi_i(t_0) + \int_{t_0}^t dt' \left(h_i(\xi, t') + \sum_j g_{ij}(\xi, t') \Gamma_j(t') \right) \\ &= \xi_i(t_0) + \int_{t_0}^t dt' h_i(\xi, t') + \sum_j \int_{t_0}^t dt' g_{ij}(\xi, t') \Gamma_j(t') \end{aligned} \quad (1.24)$$

¹ Die Normierung der Kraft wird in der Literatur unterschiedlich gehandhabt. Gelegentlich ist auch eine Normierung auf „1“ anzutreffen.

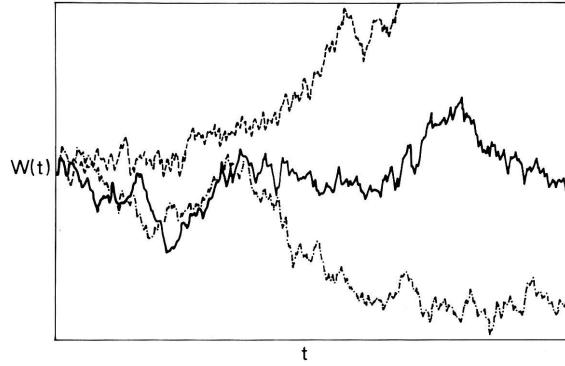


Abbildung 1.1: Drei simulierte Pfade eines Wiener Prozesses mit gleicher Anfangsbedingung. (Entnommen aus: [15])

lösen. Falls die Matrixelemente g_{ij} konstant sind, bezeichnet man das Rauschen als *additiv*, wenn sie Funktionen von ξ sind, als *multiplikativ*.

Die Auswertung des zweiten Integrals in (1.24) gestaltet sich für Rauschen, das nicht additiv ist, schwierig. Die Art und Weise, in der der Integrand $g_{ij}(\xi, t)$ auszuwerten ist, bedarf einer geeigneten Interpretation. In der Natur haben alle Prozesse eine endliche Korrelation $\langle \xi(t)\xi(t') \rangle \sim \frac{\varepsilon}{2}e^{-\varepsilon|t-t'|}$. Die Deltakorrelation, die hier als Idealisierung eingeführt worden ist, lässt sich nun als Grenzfunktion dieser Korrelationsfunktion für $\varepsilon \rightarrow 0$ verstehen. Es zeigt sich jedoch, dass bei Interpretation des Integrals im Riemann'schen Sinne die Stelle, an der dieser Grenzübergang vollzogen wird, kritisch ist und das Ergebnis qualitativ beeinflusst. Wir benötigen deshalb einen neuen Formalismus, um obiges Integral auswerten zu können.

1.2.4 Interpretation von Stieltjes-Integralen nach Itô und Stratonovich

Eine eindimensionale Langevin-Gleichung für ξ mit $h \equiv 0$ und $g \equiv 1$ beschreibt die Trajektorie eines Teilchens, das nur durch die Langevin-Kraft getrieben wird. Ein solcher reiner Diffusionsprozess wird als *Wiener Prozess* bezeichnet. In Abb. 1.1 sind drei Trajektorien eines solchen Prozesses skizziert. Für das Inkrement $dW(t)$ eines Wiener Prozesses $W(t)$ gilt:

$$dW(t) = W(t + dt) - W(t) = \Gamma(t)dt \quad . \quad (1.25)$$

Für einen allgemeinen stochastischen Prozess, der durch die Langevin-Gleichung (1.22) beschrieben sei, können wir damit den zweiten Summanden der Gleichung (1.24) durch ein sogenanntes *Stieltjes-Integral* darstellen. Dazu gehen wir von einer Integration über t zur Integration über einen Wiener Prozesses $W_j(t)$:

$$\int_{t_0}^t dt' g_{ij}(\xi, t') \Gamma_j(t') = \int_{W(t_0)}^{W(t)} g_{ij}(\xi(t'), t') dW_j(t') \quad (1.26)$$

Unterteilen wir das Intervall $[t_0, t]$ in n Intervalle $[t_{i-1}, t_i]$ gleicher Größe mit $t_0 < t_1 < \dots < t_n = t$, so ist dieses Integral definiert als der Grenzwert der Partialsummen

$$S_n = \sum_{i=1}^n g_{ij}(\xi(\tau_i), \tau_i) [W_j(t_i) - W_j(t_{i-1})] \quad (1.27)$$

für $n \rightarrow \infty$. Die Stellen, an denen die Funktion g_{ij} auszuwerten ist, müssen innerhalb des i -ten Intervalls liegen, es muss also $t_{i-1} \leq \tau_i < t_i$ gelten.

Die Trajektorien von $W_j(t)$ sind nicht differenzierbar, deswegen müssen die Inkremente für kleine Zeitunterschiede dt nicht mit der Ordnung von mindestens dt verschwinden. Das führt dazu, dass der Grenzwert der Partialsummen in diesem Fall maßgeblich von der Lage der Stellen τ_i abhängig ist. Seien die τ_i durch

$$\tau_i = \alpha t_i + (1 - \alpha)t_{i-1} \quad \text{mit } 0 \leq \alpha \leq 1 \quad (1.28)$$

festgelegt, so gibt es zwei in der statistischen Physik übliche Interpretationen für die Partialsummen und damit für die Auswertung von Stieltjes-Integralen:

$$\begin{aligned} \alpha = 0 & \quad \text{Interpretation nach Itô} \\ \alpha = \frac{1}{2} & \quad \text{Interpretation nach Stratonovich} \end{aligned}$$

Der Unterschied der beiden Interpretationen ist deutlich: Itô wertet $g_{ij}(\xi(\tau), \tau)$ in den Partialsummen jeweils zu Beginn eines Intervalls, Stratonovich hingegen in der Mitte aus. Keine der beiden ist „richtig“ oder „falsch“, maßgeblich für die Entscheidung zur Verwendung einer der beiden ist eine Abwägung von Vor- und Nachteilen.

Im Folgenden wird die Interpretation von Itô verwendet. Das hat den Nachteil, dass wir uns „neue Regeln für Integration und Differentiation“ [14] angewöhnen müssen. Auf der anderen Seite ist der Zusammenhang von \mathbf{h} und $(g_{ij})_{ij}$ zu den Koeffizienten der *Fokker-Planck-Gleichung*, die ich im Abschnitt 1.2.6 einführe, einfacher. Ferner bietet Itôs Interpretation Vorteile bei der Durchführung mathematischer Beweise.

1.2.5 Langevin-Gleichungen mit endlichem Zeitinkrement

Langevin-Gleichungen lassen sich zur numerischen Simulation von Zufallsvariablen verwenden. Dazu ist es notwendig, dass man sie für endliche Zeitinkremente $dt \rightarrow \tau$ formuliert. Die Funktion \mathbf{h} kann mit Hilfe des *Euler-Verfahrens* integriert werden. Für die Matrixelemente g_{ij} , die über das Inkrement $dW_j(t)$ integriert werden müssen, stellt sich heraus, dass der Term niedrigster Ordnung proportional zu $\sqrt{\tau}$ sein muss [15]. Ausgehend von (1.24) erhalten wir für $\tau \ll 1$:

$$\begin{aligned} \xi_i(t + \tau) &= \xi_i(t) + \tau h_i(\xi(t), t) + \sum_j \int_t^{t+\tau} dt' g_{ij}(\xi, t') dW_j(t') \\ &= \xi_i(t) + \tau h_i(\xi(t), t) + \sum_j \sqrt{\tau} g_{ij}(\xi, t) \Gamma_j(t) \end{aligned} \quad (1.29)$$

Sei τ nun konstant, also $t_i = t_0 + i\tau$. Wir benötigen für die Simulation von ξ zu diskreten Zeitpunkten Vektoren $\Gamma(t_i)$ von Zufallszahlen, die den Bedingungen (1.23) entsprechen. Üblicherweise liefern die gängigen Zufallszahlen-„Generatoren“ von Computern Zahlen r_i im Intervall $[0, 1]$.

Nach dem *Zentralen Grenzwertsatz* der Statistik strebt die Wahrscheinlichkeitsverteilung der Summe vieler unabhängiger Zufallsvariablen gegen eine *Normalverteilung* (1.8). Wir erhalten also durch Summation der vom Computer generierten Zufallszahlen gaußverteilte – und damit für die Simulation der obigen Gleichungen geeignete – Zufallszahlen. Sei r_i eine vom Computer generierte Folge von Zufallszahlen, so ist die Normierung berücksichtigt, wenn wir unsere $\Gamma_j(t_i)$ durch Summation von N Folgenglieder für hinreichend große N als

$$\Gamma_j(t_i) = \sqrt{\frac{24}{N}} \sum_{i=1}^N \left(r_i - \frac{1}{2} \right) \quad (1.30)$$

berechnen. Diese Methode ist zur Erzeugung einer großen Menge von Zufallszahlen, wie sie für die numerische Integration von Langevin-Gleichungen benötigt werden, jedoch nicht effizient.

Wesentlich schneller ist der *Box-Müller-Algorithmus*, der z.B. in [20] beschrieben ist. Nach diesem lassen sich aus zwei vom Computer generierten Zufallszahlen r_i und r_{i+1} jeweils zwei gaußverteilte Zufallsvariablen

$$\Gamma_j(t_i) = 2\sqrt{-\ln(r_1)} \cos(2\pi r_2) \quad (1.31a)$$

$$\Gamma_j(t_{i+1}) = 2\sqrt{-\ln(r_1)} \sin(2\pi r_2) \quad (1.31b)$$

erzeugen. Dieser Algorithmus findet in den im Rahmen dieser Arbeit durchgeführten numerischen Simulationen Verwendung.

1.2.6 Fokker-Planck-Gleichung

Einstein hat schon 1905 das Rätsel der Brownschen Bewegung ohne die Verwendung von Zufallkräften untersucht [23]. Es ist ihm gelungen, die Molekularbewegung mit Hilfe von statistischen Methoden auf makroskopischer Ebene zu beschreiben. Diese Vorgehensweise ist etwas eleganter, weil man dem Problem der Interpretation der Wirkung der stochastischen Kräfte, das in 1.2.4 diskutiert wurde, aus dem Wege gehen kann. Eine zur Langevin-Gleichung äquivalente Beschreibung gelingt auf Basis der Fokker-Planck-Gleichung, die nun eingeführt wird.

Wichtiger Ausgangspunkt für die Herleitung einer Differentialgleichung für die Wahrscheinlichkeitsdichte ist die Chapman-Kolmogorov-Gleichung (1.21). Ausgehend davon lässt sich durch Motivation einer sogenannten „Differentialen Chapman-Kolmogorov-Gleichung“ auf der Ebene von bedingten Wahrscheinlichkeitsverteilungen mit Hilfe einer Taylor-Entwicklung auf die Fokker-Planck-Gleichung kommen, wie es in Gardiner [15] durchgeführt ist. Wir folgen stattdessen dem von Risken [14] beschriebenen Weg.

Risken führt zunächst die bedingten Momente

$$M^{(n)}(\mathbf{x}, t|\tau) := \langle [\xi(t + \tau) - \xi(t)]^n | \xi(t) = \mathbf{x} \rangle \quad (1.32a)$$

$$= \int d^n x_1 \, [(\mathbf{x}_1 - \mathbf{x})]^n p(\mathbf{x}_1, t + \tau | \mathbf{x}, t) \quad (1.32b)$$

ein. Die Potenz ist dabei als Tensorprodukt zu interpretieren, M^n ist damit ein Tensor n -ter Stufe. Die Konditionierung in (1.32) soll bedeuten, dass für die Berechnung des Mittelwertes nur die Paare des Ensembles berücksichtigt werden, für die $\xi(t) = \mathbf{x}$ erfüllt ist. Mit Hilfe der Momente lassen sie die *Kramers-Moyal-Koeffizienten* $D^{(n)}$ definieren:

$$D^{(n)}(\mathbf{x}, t) := \frac{1}{n!} \lim_{\tau \rightarrow 0} \frac{1}{\tau} M^{(n)}(\mathbf{x}, t|\tau) \quad . \quad (1.33)$$

Die Konvergenz der Grenzwerte sei hier zunächst vorausgesetzt. Mit Hilfe der Kramers-Moyal-Koeffizienten lässt sich die zeitliche Entwicklung von PDF's durch eine Differentialgleichung erster Ordnung in der Zeit beschreiben:

$$\frac{\partial}{\partial t} f(\mathbf{x}, t) = \left\{ \sum_{n=1}^{\infty} (-\nabla)^n D^{(n)}(\mathbf{x}, t) \right\} f(\mathbf{x}, t) \quad (1.34)$$

Für Langevin-Prozesse lassen sich die Momente und damit die Kramers-Moyal-Koeffizienten bestimmen. Unter Ausnutzung der Markow-Eigenschaften gilt:

$$\mathbf{D}^{(1)}(\mathbf{x}, t) = \mathbf{h}(\mathbf{x}, t) \quad (1.35a)$$

$$D_{ij}^{(2)}(\mathbf{x}, t) = \sum_k g_{ik}(\mathbf{x}, t) g_{jk}(\mathbf{x}, t) \quad (1.35b)$$

$$D^{(i)}(\mathbf{x}, t) = 0 \quad \forall i > 2 \quad (1.35c)$$

Wir erhalten in diesem Falle aus der Kramers-Moyal-Entwicklung die Fokker-Planck-Gleichung² (FP-Gleichung):

$$\frac{\partial}{\partial t} f(\mathbf{x}, t) = \left(- \sum_i \frac{\partial}{\partial x_i} D_i^{(1)}(\mathbf{x}, t) + \sum_{ij} \frac{\partial^2}{\partial x_i \partial x_j} D_{ij}^{(2)}(\mathbf{x}, t) \right) f(\mathbf{x}, t) \quad (1.36)$$

Die FP-Gleichung lässt in der gleichen Form auch für bedingte Wahrscheinlichkeitsdichten formulieren:

$$\frac{\partial}{\partial t} p(\mathbf{x}, t | \mathbf{x}', t') = \left(- \sum_i \frac{\partial}{\partial x_i} D_i^{(1)}(\mathbf{x}, t) + \sum_{ij} \frac{\partial^2}{\partial x_i \partial x_j} D_{ij}^{(2)}(\mathbf{x}, t) \right) p(\mathbf{x}, t | \mathbf{x}', t') \quad (1.37)$$

Der Koeffizient $D^{(1)}(\mathbf{x}, t)$ wird nun üblicherweise als *Drift*, $D^{(2)}(\mathbf{x}, t)$ als *Diffusion* bezeichnet.

Mit der FP-Gleichung steht uns nun eine „Bewegungsgleichung für die PDF's“ zur Verfügung. Drift- und Diffusionsfunktionen, die in dieser Gleichung als Koeffizienten auftreten, charakterisieren stochastische Prozesse mit Markow-Eigenschaften vollständig und stehen in direktem Zusammenhang zu den Koeffizienten der Langevin-Gleichungen.

1.2.7 Stationäre Lösung der Fokker-Planck-Gleichung

Durch die Einführung eines Wahrscheinlichkeitsdichtestromes $\mathbf{j}(\mathbf{x}, t)$ mit

$$j_i(\mathbf{x}, t) = \left(D_i^{(1)}(\mathbf{x}, t) - \sum_j \frac{\partial}{\partial x_j} D_{ij}^{(2)}(\mathbf{x}, t) \right) f(\mathbf{x}, t) \quad (1.38)$$

können wir die FP-Gleichung wie folgt schreiben:

$$\frac{\partial}{\partial t} f(\mathbf{x}, t) + \nabla \cdot \mathbf{j}(\mathbf{x}, t) = 0 \quad (1.39)$$

Diese Gleichung stellt eine Kontinuitätsgleichung für die Wahrscheinlichkeitsdichte dar.

Falls $f(\mathbf{x}, t)$ nicht explizit von der Zeit abhängt, heißt $f_s(\mathbf{x}) := f(\mathbf{x}, t)$ stationär. Für diesen Fall folgt aus der Kontinuitätsgleichung:

$$\nabla \cdot \mathbf{j}(\mathbf{x}, t) = \left(- \sum_i \frac{\partial}{\partial x_i} D_i^{(1)}(\mathbf{x}) + \sum_{ij} \frac{\partial^2}{\partial x_i \partial x_j} D_{ij}^{(2)}(\mathbf{x}) \right) f_s(\mathbf{x}) = 0 \quad (1.40)$$

Für stochastische Prozesse in einer Dimension, bei denen die PDF natürlichen Randbedingungen unterliegen, lässt sich die stationäre Verteilung analytisch berechnen. In diesem Fall muss die PDF für große $|x|$ verschwinden, es kann deshalb kein Strom durch das System fließen. Integration von (1.40) liefert dann

$$-D^{(1)}(x) f_s(x) + \frac{\partial}{\partial x} D^{(2)}(x) f_s(x) = \mathcal{C} \quad (1.41)$$

mit der Integrationskonstante $\mathcal{C} = 0$. Es folgt:

$$f_s(x) = \frac{\mathcal{N}}{D^{(2)}(x)} e^{\int^x dx' \frac{D^{(1)}(x')}{D^{(2)}(x')}} \quad (1.42)$$

²Die Definition der Fokker-Planck-Gleichung ist in der Literatur nicht einheitlich. Gelegentlich findet sich ein Faktor „ $\frac{1}{2}$ “ vor dem Diffusionskoeffizienten, der dann im Wesentlichen zu einer veränderten (nämlich um einen Faktor „ 2 “ größeren) Diffusion führt.

\mathcal{N} ist an dieser Stelle eine geeignete Normierungskonstante. Für $\xi \in \mathbb{R}^n$ ist die analytische Berechnung im Allgemeinen nicht möglich, da die Wahrscheinlichkeitsströme üblicherweise nicht verschwinden.

Wir haben nun das mathematische Rüstzeug, um mit der Analyse von Daten stochastischer Prozesse zu beginnen. Weiterführende Erläuterungen sowie tiefergehende Grundlagen sind der zu Beginn dieses Kapitels aufgeführten Literatur zu entnehmen.

Kapitel 2

Schätzen von Drift und Diffusion

Ich habe in der Einleitung erläutert, dass stochastische Differentialgleichungen wie (1.22) als Entwicklungsgleichungen von Ordnungsparametern bei der Beschreibung komplexer Systeme eine entscheidende Rolle spielen. Nun beschäftige ich mich mit der Frage, wie wir Daten solcher Prozesse auswerten können. Dazu müssen die im letzten Kapitel eingeführten stochastischen Grundlagen zunächst an die Eigenschaften gemessener Datensätze angepasst werden.

2.1 Statistik von Zeitreihen

Experimentelle Daten stehen in der Regel in Form von (mehrdimensionalen) Zeitreihen zur Verfügung. Nur selten werden größere Ensembles zeitgleich gemessen. Die einzelnen Messungen erfolgen in vielen Fällen äquidistant in der Zeit, also mit konstantem Zeitinkrement τ . Je nach Quelle der Daten können die Messungen unterbrechungsfrei (z.B. medizinische Daten) oder in mehreren Stücken, zwischen denen die Messung unterbrochen wurde (z.B. Notierungen von Börsenkursen, astronomische Daten), vorliegen.

Sei zunächst eine eindimensionale Zeitreihe $x_i := x(t_i)$ mit $0 < i \leq N$ gegeben, die zu festen Zeiten $t_i = t_0 + i\tau$ gemessen sei. Im Allgemeinen sind nun die gemessenen x_i (stochastische) Funktionen einer dynamischen Variablen q :

$$x_i = f_M(q_i, q_{i-1}, \dots, q_0, i) + \Gamma_M(i) \quad . \quad (2.1)$$

Uns interessiert die interne Variable q , die z.B. einen Ordnungsparameter eines Systems darstellen kann. f_M ist eine sogenannte (deterministische) *Messfunktion*, Γ_M soll eine stochastische Funktion mit Mittelwert 0 sein und z.B. Messrauschen darstellen.

Es ist immer wünschenswert, dass die Daten keine Artefakte von speziellen Eigenschaften der Messapparatur enthalten. Darunter fallen z.B. zeitlicher Drift, additive oder multiplikative Konstanten und Hystereseffekte, die nicht die Dynamik des zu messenden Prozesses widerspiegeln. Solche Effekte werden durch die Messfunktion f_M berücksichtigt und bereiten, wenn sie nicht bekannt sind, Probleme, die der Beschreibung des Systems durch lineare stochastische Differentialgleichungen im Wege stehen können. Kantz beschreibt in [24] ein Verfahren, mit Hilfe dessen die Linearität der Messfunktion überprüft werden kann. Ich gehe von nun an davon aus, dass die Daten nicht durch eine Messfunktion verfälscht sind, verwende also $f_M(q_i) = q_i$.

Messrauschen hingegen tritt bei fast allen Messungen, schon allein bei der Digitalisierung der Messwerte, auf. Es sollte, wenn sich das Messgerät um den Wert

q_i linear verhält, gleichmäßig um 0 verteilt und gaußverteilt sein. Die Stärke des Messrauschen ist in diesem Fall durch die Varianz σ^2 der Verteilung bestimmt. Es ist immer wünschenswert, mit möglichst geringem Messrauschen zu arbeiten. Auf der anderen Seite eignen sich bei der Analyse stochastischer Prozesse Methoden zur nachträglichen Reduktion des Rauschens durch Mittelung über viele Werte oder Tiefpass-Filter, die das Spektrum modifizieren, nur bedingt. Sie können die Ergebnisse, die man für die Diffusionsfunktionen erhält, qualitativ beeinflussen. Messrauschen muss dehalb, wenn es eine gewisse Grenze überschreitet, bei der Analyse von Prozessen berücksichtigt werden. Untersuchen wir zunächst Daten, die kein oder nur sehr wenig Messrauschen enthalten und arbeiten mit $\sigma = 0$. Der Einfluss von σ auf das im Folgenden beschriebene Verfahren wird in Abschnitt 3.4 diskutiert.

Die Statistik, mit der bislang gearbeitet wurde, beruht auf der Mittelung über Ensemble. Falls alle Übergangswahrscheinlichkeiten zwischen Zuständen mit festem Zeitinkrement zeitunabhängig sind, so nennt man eine Zeitreihe *stationär*. Sind sie ferner unabhängig vom Startwert $x(t_0)$, so heißt die Zeitreihe *x ergodisch*. Für ergodische Zeitreihen sind Ensemble-Mittelwerte äquivalent zur Mittelung über die Zeit, die Drift- und Diffusionskoeffizienten in den stochastischen Differentialgleichungen werden zeitunabhängig. Eine weitergehende Diskussion der Begriffe der *Stationarität* und *Ergodizität* sowie die Beschreibung von Verfahren, mit denen sich Zeitreihen auf diese Eigenschaften hin überprüfen lassen, finden sich in [24].

Für den Mittelwert der Funktion h über eine stationäre Zeitreihe gilt damit:

$$\langle h[x(t)] \rangle = \frac{1}{T} \int_T dt h[x(t)] \quad (2.2a)$$

$$\langle h[x(t + \tau)|x(t) = x] \rangle = \frac{\int dt h[x(t + \tau)] \delta(x(t), x)}{\int dt \delta(x(t), x)} \quad (2.2b)$$

Weil wir es nun nicht mehr mit kontinuierlichen Größen zu tun haben, gehen die Integrale über die Zeit in Summen über:

$$\langle h[x(t)] \rangle = \frac{1}{N} \sum_{i=1}^N h[x_i] \quad (2.3a)$$

$$\langle h[x(t + \tau)|x(t) = x] \rangle = \frac{\sum_i h[x_{i+1}] \delta_{x,x_i}}{\sum_i \delta_{x,x_i}} \quad (2.3b)$$

Des Weiteren benötigen wir die im vorangegangenen Kapitel eingeführten PDF's. Diese werden numerisch wie folgt behandelt: zusätzlich zum diskreten Zeitschritt muss nun auch der Ortsraum diskretisiert – in sogenannte *Bins* eingeteilt – werden.

Üblich ist die Verwendung äquidistanter Bins. Dazu wird das Intervall $[x_a, x_b]$, das die zu diskretisierenden Werte enthält, in Intervalle $[x_a + (i-1)\Delta x, x_a + i\Delta x]$ mit $1 < i \leq k$ und $\Delta x = \frac{x_b - x_a}{k}$ eingeteilt. Das Kronecker-Delta in (2.3) kann dann wie folgt interpretiert werden:

$$\delta_{x_i, x} = \begin{cases} 1 & \text{falls } \exists j \text{ mit } x(i), x \in [x_a + (j-1)\Delta x, x_a + j\Delta x] \\ 0 & \text{sonst} \end{cases} \quad (2.4)$$

Nachteil der Verwendung äquidistanter Bins ist die unterschiedliche Dichte an Messwerten pro Intervall, die die Größe statistischer Fehler bestimmt. Es gibt auch andere Methoden zum Diskretisieren des Ortsraumes, die z.B. in mathematischer Literatur untersucht werden und die dieses Problem umgehen. Die Art und Weise, wie dieses *Binning* durchgeführt wird, kann großen Einfluss auf die späteren Ergebnisse haben. Wir verwenden im Folgenden eine äquidistante Diskretisierung, weil diese sowohl analytisch als auch numerisch gut durchzuführen ist. Die Fehler, die dabei zu erwarten sind, werden in Kapitel 3 im Detail diskutiert.

2.2 Schätzen der Koeffizienten

Drift- und Diffusionsfunktionen, wie sie in der FP-Gleichung auftauchen, sind über die Momente gemäß (1.33) definiert. Sind die Momente analytisch gegeben, so lassen sich die Funktionen berechnen als:

$$\mathbf{D}^{(1)}(\mathbf{x}) = \lim_{\tau \rightarrow 0} \frac{1}{\tau} \langle \mathbf{x}(t + \tau) - \mathbf{x}(t) | \mathbf{x}(t) = \mathbf{x} \rangle \quad (2.5a)$$

$$D_{ij}^{(2)}(\mathbf{x}) = \lim_{\tau \rightarrow 0} \frac{1}{2\tau} \langle [\mathbf{x}(t + \tau) - \mathbf{x}(t)]_i [\mathbf{q}(t + \tau) - \mathbf{x}(t)]_j | \mathbf{x}(t) = \mathbf{x} \rangle \quad . \quad (2.5b)$$

Diese Ausdrücke sollen nun für Daten stochastischer Prozesse, die in Form von Zeitreihen vorliegen mögen, bestmöglich ausgewertet werden. Problematisch gestaltet sich dabei der Grenzübergang $\tau \rightarrow 0$. Friedrich [9] und Siegert [25] haben kürzlich gezeigt, dass sich auch für endliche τ wertvolle Schätzungen für Drift- und Diffusionsfunktionen ergeben können. Stellen wir einmal Probleme, die dabei auftreten können, zusammen:

1. Empirische Datenreihen, die in Form von Zeitreihen vorliegen sollen, besitzen eine endliche Samplingfrequenz. Es gibt deswegen in jedem Fall ein minimales τ , für das obige Ausdrücke ausgewertet werden können. Es wird sich in Kapitel 3 zeigen, dass schon die endliche Menge an Daten, die zur Analyse zur Verfügung steht, die Qualität der Ergebnisse im Grenzwert $\tau \rightarrow 0$ drastisch vermindert.
2. Ist die Messung am stochastischen Prozess von Messrauschen Γ_M überlagert, so divergieren die Grenzwerte (2.5). Wie es zu diesem Verhalten kommt, wird ebenfalls in Kapitel 3 diskutiert.
3. Auch das Binning der Daten, das zur Berechnung der Momente erforderlich ist, hat einen Einfluss auf das Ergebnis. Wir werden in Kapitel 3 sehen, dass sowohl zu kleine als auch zu große Bins Fehler verursachen.
4. Die Länge der Mantisse bei Daten, die in digitaler Form vorliegen, ist beschränkt. Dadurch kann bei Subtraktion zweier annähernd gleich großer Werte der relative Fehler sehr groß werden. Dieses Problem ist numerischer Natur und muss bei der Berechnung der Momente für kleine Zeitintervalle beachtet werden. Insbesondere bereitet daher die Auswertung von Daten, die nur mit einer geringen Auflösung digitalisiert sind, Probleme.
5. Ferner kann der Grenzübergang Probleme bereiten, wenn es sich (unter den gewählten Variablen) nicht um einen Markow-Prozess handelt. In solchen Fällen ist zu prüfen, ob durch Hinzunahme weiterer (Delay-)Variablen der Prozess Markow'sch wird und die Grenzübergänge ein besseres Verhalten zeigen (siehe auch [24]).

Ob eine Einbettung die Markow-Eigenschaften erfüllt, ist in jedem Fall zunächst zu untersuchen. Hierzu ist zu prüfen, ob sich einfach bedingte von mehrfach bedingten Wahrscheinlichkeitsdichten unterscheiden. Diese Prüfung ist z.B. in [26] erläutert. Zunächst sollen einige Fälle vorgestellt werden, die mit Hilfe der Schätzungen (2.5) erfolgreich untersucht worden sind und in denen sich der Grenzübergang zufriedenstellend vollziehen ließ. In diesen Fällen kann der Wert des Ausdrucks, der für $\tau \rightarrow 0$ betrachtet werden soll, für einige endliche τ bis hin zum minimalen τ berechnet und schließlich für $\tau = 0$ extrapoliert werden. Danach liegen Drift- und Diffusions-Funktionen zunächst wieder diskret vor. Sie können dann, wenn gewünscht, in Funktionen entwickelt werden. Die Güte der erhaltenen Funktionen lässt sich überprüfen, indem man die Eigenschaften von simulierten Datenreihen oder simulierten Zweipunkt-JPDF's mit den Eigenschaften der ursprünglichen Datenreihe bzw. den daraus extrahierten JPDF's vergleicht.

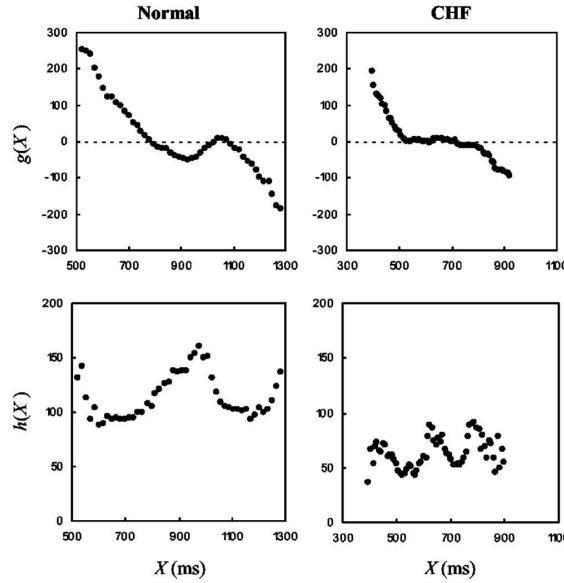


Abbildung 2.1: Typischer, aus Messdaten gewonnener Verlauf von Drift- (oben, hier $g(x)$ genannt) und Diffusions-Funktionen (unten, $D^{(2)}(x) = h^2(x)$) gesunder (links) sowie unter CHF leidender Personen (rechts). (Entnommen aus: [27])

2.3 Beispiele

Das soeben beschriebene Verfahren der direkten Schätzung von Drift und Diffusion hat bislang in vielen Bereichen Verwendung gefunden. In Tabelle 2.1 sind einige Veröffentlichungen aufgeführt, in denen die Methode verwendet oder diskutiert wird. Beachtenswert dabei ist insbesondere, in welch unterschiedlichen Bereichen derartige Verfahren benötigt werden. Einige Beispiele seien im Folgenden vorgestellt.

2.3.1 Physiologische und pathologische Fluktuation der Herzfrequenz

Tom Kuusela untersucht in seiner 2004 erschienenen Arbeit [27] Daten von Patienten, die unter kongestiver Herzinsuffizienz (CHF) leiden. Er vergleicht diese mit denen von gesunden Patienten und stellt einen qualitativen Unterschied fest.

Kuusela lagen von jeweils 15 gesunden und 15 an CHF erkrankten Personen Aufzeichnungen der Länge des R-R-Intervalls¹ über 22 bis 24 Stunden vor. Der Mittelwert der jeweiligen Messungen liegt immer in der Größenordnung von 1 s, die Datenreihen haben also einen Umfang von ca. 80000 – 100000 Punkten. Aus den Daten sind nun nach (2.5) Drift und Diffusion bestimmt worden. Abb. 2.1 zeigt den typischen Verlauf der Daten eines gesunden und eines erkrankten Patienten. Es ist deutlich ein qualitativer Unterschied zu erkennen:

$D^{(1)}$: Bei der gesunden Testperson sind deutlich drei Nulldurchgänge von $D^{(1)}$ zu erkennen. Diese bilden einen instabilen an der Stelle $x_u = 1050$ ms sowie zwei stabile Fixpunkte der Dynamik an den Stellen $x_{s,1} = 800$ ms sowie

¹Das R-R-Intervall bezeichnet den Abstand zweier aufeinanderfolgender Pulsschläge. Den einzelnen Ereignisse während einer Periode sind verschiedene Symbole zugewiesen, „R“ legt den zur Messung zu verwendenden Trigger fest.

Medizinische Anwendungen

2000	Friedrich et al.	Parkinson-Daten	[9]
2001	Sieger	Epilepsie-Daten	[25]
2004	Kuusela et al.	Schwankungen der Herzfrequenz	[27]

Sogenannte „Life Sciences“

2000	Friedrich et al.	Statistik auf Devisen-Märkten	[5]
2000	Gradišek et al.	Metall-Schneiden	[28]
2002	Kriso et al.	Straßenverkehrs-Daten	[29]
2003	Böttcher et al.	Statistik von Wind-Böen	[3]
2003	Jafari et al.	Analyse und Manipulation rauher Oberflächen	[30]
2003	Sura et al.	Wind im südlichen Ozean	[31]
2003	Waechter et al.	Oberflächen-Rauhigkeit von Fahrbahnen	[26]

Physikalische Anwendungen

1998	Sieger et al.	Synchronisation	[32]
1999	Lück et al.	Beschreibung turbulenter Felder	[33]
2001	Frank et al.	Optische Bistabilität von VCSELs	[34]
2003	Bödeker et al.	Dynamik von Gasentladungssystemen	[35]

Analyse von mathematischen Modellsystemen

1998	Sieger et al.	Versch. bistabile Systeme	[32]
2000	Friedrich et al.	Shinriki-Oszillatör	[9]
2000	Gradišek et al.	Lorentz-System	[28]

Probleme: endliche Samplingrate τ

2001	Ragwitz et al.	Korrekturen für endliche Zeiten	[11]
2002	Friedrich et al.	Korrekturen der Korrekturen	[10]
2002	Ragwitz et al.	Stellungnahme	[12]

Probleme: Messrauschen

2003	Siefert et al.	Analyse von Dynamischem- und Messrauschen	[37]
	Böttcher et al.	Analyse bei Messrauschen	[38]

Tabelle 2.1: Bisherige Anwendungen der und Veröffentlichungen zur in 2.2 beschriebenen Methode

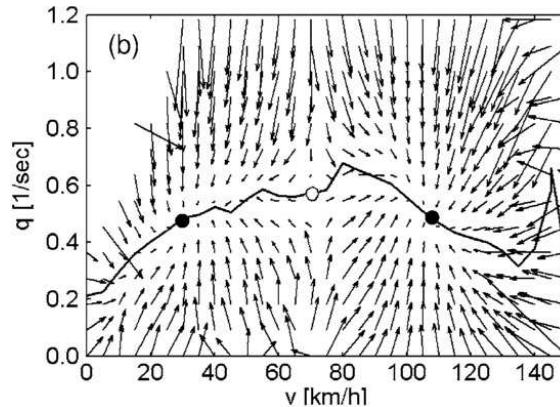


Abbildung 2.2: Deterministische Dynamik des Verkehrs auf einem Autobahnabschnitt nahe Köln als Funktion der Geschwindigkeit v der Fahrzeuge und des Verkehrsflusses q . (Entnommen aus: [29])

$x_{s,2} = 1100$ ms. Die von Null verschiedenen Steigungen in der Nähe der Fixpunkte treiben das System bei kleinen Abweichung immer wieder in die stabilen Fixpunkte zurück. Diese drei Fixpunkte vereinen sich bei erkrankten Personen zu einer breiten Zone, in der kein R-R-Wert hinreichend stabil ist.

$D^{(2)}$: Bei den Diffusionsfunktionen ist kein so deutlicher, qualitativer Unterschied auszumachen. Die Fluktuationen in der Herzfrequenz scheinen jedoch bei gesunden Probanden wesentlich stärker zu sein als bei erkrankten.

Es ist deutlich, dass sich die Drift-Funktion des R-R-Intervalls bei den erkrankten Personen qualitativ verändert ist. Aus physikalischer Sicht hat das System einen Phasenübergang vollzogen. Man erkennt einen Mechanismus, der aufgrund von Selbstorganisationseffekten bei gesunden Menschen zu einer Stabilisierung der Herzfrequenz führt. Bei CHF-Patienten hingegen scheint dieser Mechanismus nicht ordnungsgemäß zu funktionieren.

2.3.2 Analyse von Straßenverkehrs-Daten

In der Arbeit von Kriso et al. [29] ist die Dynamik des Verkehrs auf einem dreispurigen Autobahnabschnitt nahe Köln-Nord untersucht. Aus durch in die Fahrbahnen eingelassene Induktions-Schleifen gewonnenen Daten ist für jedes Fahrzeug unter anderem die Spur, auf der das Fahrzeug die Messstelle passiert, ermittelt und die Geschwindigkeit, Länge und der Abstand zum vorherfahrenden Fahrzeug berechnet worden. Kriso hat für seine Analyse aus diesen Daten die individuellen Geschwindigkeiten v_i und Verkehrs-Dichten q_i berechnet und untersucht.

Anders als bei Verkehrssimulationen üblich, ist nun die Entwicklung von v und q mit der Zeit als stochastischer Prozess betrachtet worden. Man kann sich vorstellen, dass der Verkehrsfluss auf der Autobahn gewissen deterministischen Gesetzmäßigkeiten folgt, auf der anderen Seite aber auch stochastische Fluktuationen eine Rolle spielen. Kriso hat ein System von zwei gekoppelten, Langevin-artigen Gleichungen

formuliert, das die Dynamik der beiden untersuchten Parameter beschreiben soll,

$$v_{i+1} = v_i + h_v(v_i, q_i) + \sum_{n=1}^2 g_{v,n}(v_i, q_i) \Gamma_{ni} \quad (2.6a)$$

$$q_{i+1} = q_i + h_q(v_i, q_i) + \sum_{n=1}^2 g_{q,n}(v_i, q_i) \Gamma_{ni} . \quad (2.6b)$$

Die $\Gamma_{n,i}$ stellen Rauschen mit einer normalen Verteilungsfunktion und folgenden Eigenschaften dar:

$$\langle \Gamma_{ni} \rangle = 0 \quad \text{und} \quad \langle \Gamma_{ni} \Gamma_{mk} \rangle = 2\delta_{mn}\delta_{ik} \quad \forall i, k, m, n \quad (2.7)$$

Die Funktionen h und g lassen sich nun, ähnlich den Gleichungen (2.5), durch Berechnung und Auswertung der bedingten Momente abschätzen. Die Faktoren τ bzw. $\sqrt{\tau}$, die an dieser Stelle fehlen, sind bei der von Kriso verwendeten Notation schon in die Funktionen h und g mit eingegangen. Ob die qualitativen Aussagen, die wir so erhalten, die wahre Situation widerspiegeln, ist später zu prüfen.

Abb. 2.2 zeigt den Drift (die Funktion h), die Kriso unter Berücksichtigung der Daten aller drei Spuren erhalten hat. Der Drift scheint eine charakteristische Struktur aufzuweisen: Es zeigt sich ein instabiler Fixpunkt der Dynamik bei einer Geschwindigkeit von ca. 80 km/h, rechts und links davon liegen bei 30 km/h und ungefähr 110 km/h zwei stabile Fixpunkte. Die einzelnen Fixpunkte lassen sich nun identifizieren:

- Der Fixpunkt von 30 km/h, in den das System bei Geschwindigkeiten unterhalb von 80 km/h getrieben wird, entspricht einem *Stau* auf der Straße.
- Oberhalb von 80 km/h relaxiert das System in den zweiten stabilen Fixpunkt bei etwa 110 km/h. Es herrscht nun *freie Fahrt*. Aus dem Teilstück der Autobahn, auf dem die Daten erhoben wurden, war die Höchsgeschwindigkeit auf 100 km/h begrenzt. Diese Beschränkung ist für den rücktreibenden Drift bei höheren Geschwindigkeiten verantwortlich.

Ferner ist dem Graphen zu entnehmen, dass die Stabilisierung des Verkehrsflusses auf ca. 0.5/s wesentlich stärker ist als die der Geschwindigkeit. Die Ergebnisse stimmen mit denen aus Experimenten und Simulationen zur Verkehrsdynamik weitgehend überein (siehe Referenzen in [29]). Der Vorteil dieser Art der Datenauswertung liegt in dem sehr geringen Rechenaufwand, mit dem Daten in Echtzeit analysiert werden können und in der übersichtlichen und qualitativ leicht auszuwertenden grafischen Darstellung.

2.3.3 Nachweis des Parkinson-Syndroms

In einem letzten Beispiel werden wieder medizinische Daten untersucht. Friedrich hat Hand-Tremor Messungen analysiert [9] und die Daten gesunder Probanden mit *physiologischem Tremor* (PT) mit denen von Patienten, die unter sogenanntem *essentiellen Tremor* (ET) bzw. unter der *Parkinson-Krankheit* (PD) leiden, verglichen.

Ausgangspunkt sind die mittels eines kleinen piezoresistiven Beschleunigungsaufnehmers gemessenen Daten mit einer Samplingfrequenz von 800 Hz. Die jeweiligen Zeitreihen umfassen 24000 Punkte und sind auf Varianz 1 normiert worden. In der Arbeit sind die Daten in ein zweidimensionales System aus Delay-Koordinaten eingebettet. In dieser Einbettung werden die Daten als stochastisches System betrachtet und Drift und Diffusion bestimmt. Es zeigt sich, dass sich den einzelnen

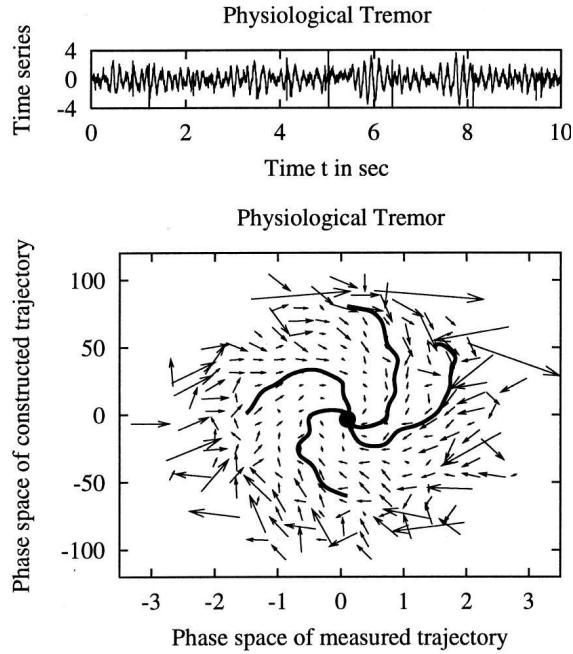


Abbildung 2.3: Hand Tremor eines gesunden Probanden und das aus dieser Zeitreihe für zweidimensionale Delay-Koordinaten gewonnene Drift-Vektorfeld. Im Zentrum liegt der deutlich sichtbare, stabile Fixpunkt der Dynamik. (Entnommen aus: [9])

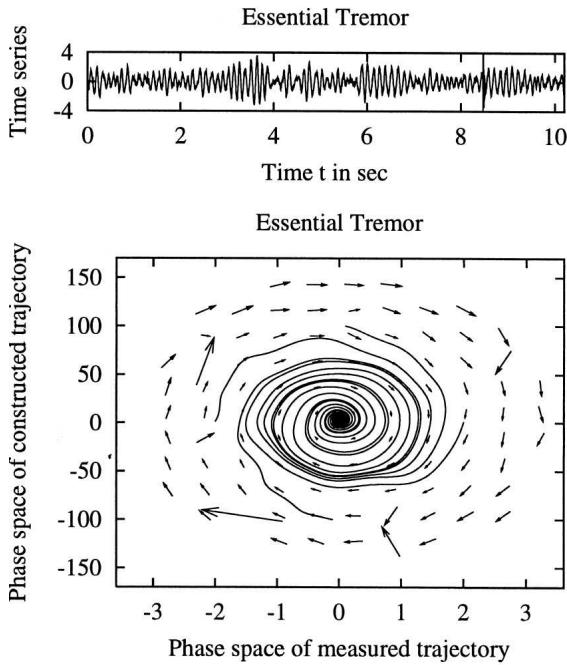


Abbildung 2.4: Hand Tremor eines ET-Probanden und das aus dieser Zeitreihe für zweidimensionale Delay-Koordinaten gewonnene Drift-Vektorfeld. Die Relaxationszeit in den stabilen Fixpunkt ist erheblich erhöht. (Entnommen aus: [9])

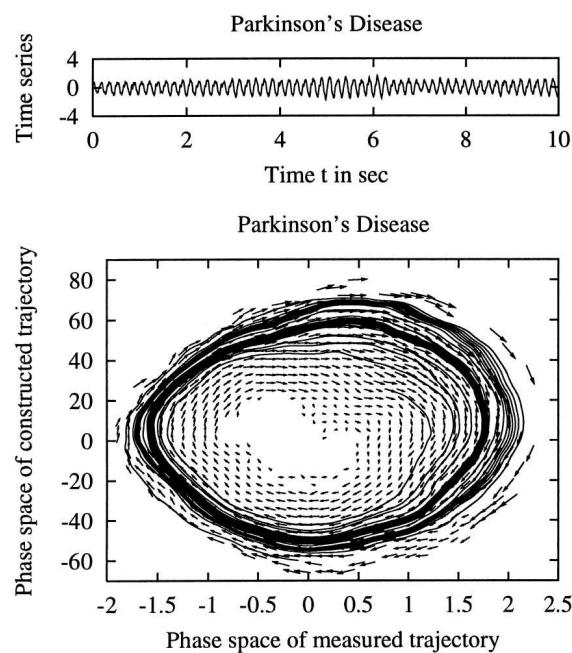


Abbildung 2.5: Hand Tremor eines PD-Probanden und das aus dieser Zeitreihe für zweidimensionale Delay-Koordinaten gewonnene Drift-Vektorfeld. Es hat eine Bifurkation von einem stabilen Fixpunkt in einen instabilen Fixpunkt und einen stabilen Grenz-Zyklus mit charakteristischer Umlauf-Frequenz stattgefunden. (Entnommen aus: [9])

Patientengruppen charakteristische Strukturen des Drift-Vektorfeldes zuordnen lassen. Es lässt sich ein Phasenübergang im System feststellen.

In den Abb. 2.3 bis 2.5 sind die ursprünglichen Daten sowie diese Vektorfelder für die drei Gruppen skizziert.

1. In Abb. 2.3 sind der Tremor einer gesunden Person und der zugehörige, aus den Daten extrahierte Drift aufgetragen. Beim Betrachten der Datenreihe fällt auf, dass sowohl Amplituden als auch Frequenz mit der Zeit stark schwanken. Dieses würde sich in einer schnell abfallenden Autokorrelationsfunktion (siehe z.B. [39]) äußern. Beim Drift sehen wir einen stabilen Fixpunkt im Ursprung, in den alle Beispieltrajektorien, zumindest wenn – wie hier – die Diffusion vernachlässigt wird, schnell relaxieren. Die gesunde Person ist also in der Lage, den Finger selbstständig zu stabilisieren. Die stets gegenwärtigen Fluktuationen jedoch führen dazu, dass sich das System nie vollkommen in Ruhe befindet.
2. Abb. 2.4 zeigt die gleiche Analyse, durchgeführt an Daten von Patienten, die unter sogenanntem *essentiellen Tremor* leiden. Den Daten kann man entnehmen, dass die Amplituden der Schwingungen etwas stabiler geworden ist, und sich die Frequenz wesentlich stabilisiert hat. Es treten jedoch Phasensprünge auf. Daraus folgt eine schwach abfallende Autokorrelationsfunktion der Messwerte. Es gibt noch immer einen attraktiven Fixpunkt im Ursprung. Auf diesen laufen die Beispieltrajektorien auf Spiralen zu und benötigen dafür einen wesentlich längeren Weg. Wählt man nun eine Diffusion > 0 so wird der Ursprung in der Regel nicht erreicht und das System führt Schwingungen wechselnder Amplitude aus. Bei Patienten, die unter essentiellem Tremor leiden, ist die Stabilisierung also so stark verlangsamt, dass sie in der Regel in Gegenwart natürlicher Fluktuationen nicht mehr gelingt. Eine solche Schwingung bezeichnet man auch als *Relaxations-Oszillation*.
3. Patienten, die unter der *Parkinson-Krankheit* leiden, können ihre Glieder nicht mehr stabilisieren (Abb. 2.5). Hier liegt die Ursache nicht in den Fluktuationen, die Zittern verursachen, sondern darin, dass nach einer Bifurkation des dynamischen Systems der ursprüngliche Fixpunkt instabil geworden ist. Das Zittern wird nun durch eine periodische Anregung verursacht, die sich bei der gewählten Einbettung in dem stabilen Grenzzyklus, der im Vektorfeld des Drifts zu erkennen ist, zeigt und zur hohen Stabilität von Frequenz und Amplitude führt. Die Autokorrelation der Daten ist in diesem Fall von großer Reichweite.

Kapitel 3

Fehler beim Schätzen von Drift und Diffusion

Im vorangangenen Kapitel habe ich bereits erwähnt, dass bei der Anwendung von (2.5) auf gemessene Daten Probleme auftreten. Die Auswirkungen der einzelnen Fehler auf die geschätzten Funktionen für Drift und Diffusion möchte ich nun etwas genauer betrachten.

Zunächst werden in Abschnitt 3.1 die Auswirkungen endlicher Zeitinkremente auf die Konvergenz und die Fehler der in letzten Kapitel eingeführten Schätzungen (2.5) untersucht. Neben der Konvergenz gegen die exakten Parameter für $\tau \rightarrow 0$ ergibt sich dabei eine Abschätzung für die Qualität der Ausdrücke für endliche τ . Des Weiteren wird gezeigt, wie man Fehler statistischer Natur, die aufgrund einer nur endlichen Anzahl von Datenpunkten, sowie Fehler, die durch die Diskretisierung des Ortsraums – des schon eingeführten „Binnings“ – entstehen, abschätzen kann.

Ferner werden in Abschnitt 3.4 Prozesse betrachtet, bei denen das Messrauschen Γ_M nicht verschwindet. Hier störe ich auf ein Verfahren, das sich auf einfache stochastische Prozesse anwenden lässt und, selbst bei stark verrauschten Messwerten, eine Analyse ermöglicht, mittels derer die Stärke des externe Rauschens und die Parameter der Dynamik des Systems bestimmt werden können.

Es sind nun Daten von Prozessen zu untersuchen, die bereits bei eindimensionaler Einbettung Markow-Eigenschaften zeigen. Als Beispiel wird jeweils ein eindimensionaler Ornstein-Uhlenbeck-Prozess betrachtet.

3.1 Schätzung von $D^{(1)}$ und $D^{(2)}$ für endliche τ

Mit Hilfe einer Methode, die von Friedrich in [10] vorgeschlagen wird, soll die Abhängigkeit der bedingten Momente vom gewählten Zeitinkrement τ untersucht werden. Die ersten beiden bedingten Momente lassen sich dazu in eine Taylor-Reihe entwickeln. In der Literatur findet sich eine solche Entwicklung gelegentlich unter dem Namen „Itô-Taylor-Entwicklung“ [38, 20]. Statt dessen wird nun die FP-Gleichung für kleine Zeitinkremente entwickelt. Damit lässt sich der in Abschnitt 1.2.4 diskutierten Interpretation stochastischer Differentialgleichungen aus dem Wege gehen.

Zunächst definieren wir den in der FP-Gleichung auftretenden linearen Differentialoperator als Fokker-Planck-Operator \hat{L} . Es gilt dann:

$$\frac{\partial}{\partial t} p(x, t|x_0, t_0) = \hat{L} p(x, t|x_0, t_0) \quad \text{mit} \quad \hat{L} = \left(-\frac{\partial}{\partial x} D^{(1)}(x, t) + \frac{\partial^2}{\partial x^2} D^{(2)}(x, t) \right) \quad (3.1)$$

Eine formale Lösung dieser Gleichung ist

$$p(x, t|x_0, t_0) = e^{\hat{L}(t-t_0)} p(x, t_0|x_0, t_0) = e^{\hat{L}(t-t_0)} \delta(x - x_0) . \quad (3.2)$$

Sei nun $t - t_0 =: \tau \ll 1$ das Zeitinkrement, für das wir die bedingten Verteilungen untersuchen wollen. Wir können die Exponentialfunktion dann nach τ entwickeln und erhalten:

$$p(x, t_0 + \tau|x_0, t_0) = \sum_{k=0}^{\infty} \frac{(\hat{L}\tau)^k}{k!} \delta(x - x_0) \quad (3.3)$$

Bei der Entwicklung der Momente sollen Terme bis zur dritten Ordnung in τ betrachtet werden. Dazu benötigen wir die ersten vier Reihenglieder der obigen Entwicklung:

$$p(x, t_0 + \tau|x_0, t_0) \approx \left(1 + \hat{L}\tau + \frac{1}{2}\hat{L}^2\tau^2 + \frac{1}{6}\hat{L}^3\tau^3 + \mathcal{O}(\tau^4) \right) \delta(x - x_0) \quad (3.4)$$

Hier treten Potenzen des Differentialoperators \hat{L} auf. Die höheren Potenzen von \hat{L} ergeben sich als Permutationen der Summanden des Differentialoperators. Weil zur Berechnung der Momente partiell integriert werden muss, werden die Potenzen des Operators zunächst notiert. Auf eine Testfunktion $f(x)$ wirken diese wie folgt¹:

$$\hat{L}^2 f(x) = \left[\frac{\partial}{\partial x} D^{(1)}(x, t) \frac{\partial}{\partial x} D^{(1)}(x, t) - \frac{\partial}{\partial x} D^{(1)}(x, t) \frac{\partial^2}{\partial x^2} D^{(2)}(x, t) \right. \quad (3.5a)$$

$$\left. - \frac{\partial^2}{\partial x^2} D^{(2)}(x, t) \frac{\partial}{\partial x} D^{(1)}(x, t) + \frac{\partial^2}{\partial x^2} D^{(2)}(x, t) \frac{\partial^2}{\partial x^2} D^{(2)}(x, t) \right] f(x)$$

$$\hat{L}^3 f(x) = \left[- \frac{\partial}{\partial x} D^{(1)}(x, t) \frac{\partial}{\partial x} D^{(1)}(x, t) \frac{\partial}{\partial x} D^{(1)}(x, t) \right. \quad (3.5b)$$

$$+ \frac{\partial}{\partial x} D^{(1)}(x, t) \frac{\partial}{\partial x} D^{(1)}(x, t) \frac{\partial^2}{\partial x^2} D^{(2)}(x, t)$$

$$+ \frac{\partial}{\partial x} D^{(1)}(x, t) \frac{\partial^2}{\partial x^2} D^{(2)}(x, t) \frac{\partial}{\partial x} D^{(1)}(x, t)$$

$$- \frac{\partial}{\partial x} D^{(1)}(x, t) \frac{\partial^2}{\partial x^2} D^{(2)}(x, t) \frac{\partial^2}{\partial x^2} D^{(2)}(x, t)$$

$$+ \frac{\partial^2}{\partial x^2} D^{(2)}(x, t) \frac{\partial}{\partial x} D^{(1)}(x, t) \frac{\partial}{\partial x} D^{(1)}(x, t)$$

$$- \frac{\partial^2}{\partial x^2} D^{(2)}(x, t) \frac{\partial}{\partial x} D^{(1)}(x, t) \frac{\partial^2}{\partial x^2} D^{(2)}(x, t)$$

$$- \frac{\partial^2}{\partial x^2} D^{(2)}(x, t) \frac{\partial^2}{\partial x^2} D^{(2)}(x, t) \frac{\partial}{\partial x} D^{(1)}(x, t)$$

$$+ \frac{\partial^2}{\partial x^2} D^{(2)}(x, t) \frac{\partial^2}{\partial x^2} D^{(2)}(x, t) \frac{\partial^2}{\partial x^2} D^{(2)}(x, t) \right] f(x)$$

3.1.1 Berechnung des ersten bedingten Momentes

Nach (1.32) kann das erste bedingte Moment aus der bedingten Wahrscheinlichkeitsdichte berechnet werden als

$$\langle x(t_0 + \tau) - x(t_0) | x(t_0) = x_0 \rangle = \int dx (x - x_0) p(x, t_0 + \tau|x_0, t_0) . \quad (3.6)$$

¹Die Differentialoperatoren sollen hier jeweils auf alle Faktoren, die rechts von ihnen stehen, wirken.

Für $\tau \ll 1$ kann die CPDF nach (3.3) entwickelt werden. Sortieren wir nach Potenzen von τ , so sind folgende Terme zu berechnen:

$$(3.6) \quad = \quad \int dx (x - x_0) \delta(x - x_0) \quad (3.7a)$$

$$+ \tau \int dx (x - x_0) \hat{L} \delta(x - x_0) \quad (3.7b)$$

$$+ \frac{\tau^2}{2} \int dx (x - x_0) \hat{L}^2 \delta(x - x_0) \quad (3.7c)$$

$$+ \frac{\tau^3}{6} \int dx (x - x_0) \hat{L}^3 \delta(x - x_0) \quad (3.7d)$$

$$+ \mathcal{O}(\tau^4)$$

Mittels partieller Integration lassen sich die einzelnen Integrale – nach Einsetzen der Potenzen des Fokker-Planck-Operators aus (3.5) – lösen. Im Anhang A.1 wird am Beispiel von (3.10c) gezeigt, wie diese Integrationen durchzuführen sind. Weil x im Integral nur in erster Ordnung auftritt, vereinfacht sich die Integration bei dem Integral, das \hat{L}^3 enthält: Alle Integrale mit einem Differentialoperator, der mit $D^{(2)}$ beginnt, ergeben keinen Beitrag. Die Summanden aus den Zeilen a bis d von (3.7), die jeweils die Korrektur des ersten Momentes in einer Ordnung von τ angeben, sind dann²:

$$(a) \quad = \quad 0 \quad (3.8a)$$

$$(b) \quad = \quad \tau D^{(1)}(x_0, t_0) \quad (3.8b)$$

$$(c) \quad = \quad \frac{\tau^2}{2} \left(D^{(1)} D_x^{(1)} + D^{(2)} D_{x^2}^{(1)} \right) \quad (3.8c)$$

$$(d) \quad = \quad \frac{\tau^3}{6} \left[D^{(1)} D^{(1)} D_{x^2}^{(1)} + D^{(1)} D_x^{(1)} D_x^{(1)} + 3D^{(2)} D_x^{(1)} D_{x^2}^{(1)} \right. \\ \left. + D^{(2)} D^{(1)} D_{x^3}^{(1)} + D^{(1)} D_x^{(2)} D_{x^2}^{(1)} + D^{(1)} D^{(2)} D_{x^3}^{(1)} \right. \\ \left. + D^{(2)} D_{x^2}^{(2)} D_{x^2}^{(1)} + 2D^{(2)} D_x^{(2)} D_{x^3}^{(1)} + D^{(2)} D^{(2)} D_{x^4}^{(1)} \right] \quad (3.8d)$$

3.1.2 Berechnung des zweiten bedingten Momentes

Für das zweite Moment lässt sich entsprechend verfahren. Hier ist nun eine Entwicklung von

$$\left\langle [x(t_0 + \tau) - x(t_0)]^2 | x(t_0) = x_0 \right\rangle = \int dx (x - x_0)^2 p(x, t_0 + \tau | x_0, t_0) \quad . \quad (3.9)$$

gesucht.

Wenn wir wieder nach Potenzen von \hat{L} ordnen, erhalten wir folgenden Ausdruck für Terme bis zur dritten Ordnung in τ :

²Drift und Diffusion sind jeweils – wie z.B. in (3.8b) – Funktionen von (x_0, t_0) . Zur weiteren Erhöhung der Übersichtlichkeit werden Ableitungen der Ordnung n nach dem Argument x_0 durch das Subscript x^n gekennzeichnet.

$$(3.9) = \int dx (x - x_0)^2 \delta(x - x_0) \quad (3.10a)$$

$$+ \tau \int dx (x - x_0)^2 \hat{L} \delta(x - x_0) \quad (3.10b)$$

$$+ \frac{\tau^2}{2} \int dx (x - x_0)^2 \hat{L}^2 \delta(x - x_0) \quad (3.10c)$$

$$+ \frac{\tau^3}{6} \int dx (x - x_0)^2 \hat{L}^3 \delta(x - x_0) \quad (3.10d)$$

$$+ \mathcal{O}(\tau^4)$$

Die einzelnen Zeilen tragen nun wie folgt zum Gesamtmoment bei:

$$(a) = 0 \quad (3.11a)$$

$$(b) = 2\tau D^{(2)}(x_0, t_0) \quad (3.11b)$$

$$(c) = \tau^2 \left[D^{(1)} D^{(1)} + 2D^{(2)} D_x^{(1)} + D^{(1)} D_x^{(2)} + D^{(2)} D_{x^2}^{(2)} \right] \quad (3.11c)$$

$$(d) = \frac{\tau^3}{3} \left[3D^{(1)} D^{(1)} D_x^{(1)} + 4D^{(2)} D_x^{(1)} D_x^{(1)} + 4D^{(2)} D^{(1)} D_{x^2}^{(1)} + 2D^{(1)} D_x^{(2)} D_x^{(1)} + 3D^{(1)} D^{(2)} D_{x^2}^{(1)} + 2D^{(2)} D_{x^2}^{(2)} D_x^{(1)} + 6D^{(2)} D_x^{(2)} D_{x^2}^{(1)} + 4D^{(2)} D^{(2)} D_{x^3}^{(1)} + D^{(1)} D_x^{(1)} D_x^{(2)} + D^{(1)} D^{(1)} D_{x^2}^{(2)} + D^{(2)} D_{x^2}^{(1)} D_x^{(2)} + 2D^{(2)} D_x^{(1)} D_{x^2}^{(2)} + D^{(2)} D^{(1)} D_{x^3}^{(2)} + D^{(1)} D_x^{(2)} D_{x^2}^{(2)} + D^{(1)} D^{(2)} D_{x^3}^{(2)} + D^{(2)} D_{x^2}^{(2)} D_{x^2}^{(2)} + 2D^{(2)} D_x^{(2)} D_{x^3}^{(2)} + D^{(2)} D^{(2)} D_{x^4}^{(2)} \right] \quad (3.11d)$$

3.1.3 Einfluss auf die in Kapitel 2 beschriebene Methode zur Datenanalyse

Mit Hilfe dieser Entwicklungen können wir nun die Genauigkeit der Schätzung unserer Funktionen aus empirischen Daten abschätzen. Sei τ dafür fest. Eine geeignete Wahl ist z.B. das durch die Samplingfrequenz der Daten festgelegte kleinstmögliche τ . Nehmen wir an, wir wüssten, dass der auszuwertenden Zeitreihe der Drift $D^{(1)}(x, t)$ und die Diffusion $D^{(2)}(x, t)$ zugrunde liegen. Welche Funktionen³ $D_E^{(1)}(x, t, \tau)$, $D_E^{(2)}(x, t, \tau)$ können wir für endliche τ erhalten, wenn keine weiteren Fehlerquellen einfließen?

Um dieses abschätzen zu können, setzen wir die Entwicklungen für endliches τ in (2.5) ein:

$$D_E^{(1)}(x_0, t_0, \tau) = \frac{1}{\tau} \langle x(t_0 + \tau) - x(t_0) | x(t_0) = x_0 \rangle \quad (3.12a)$$

$$= D^{(1)}(x_0, t_0) + \frac{\tau}{2} \left[D^{(1)} D_x^{(1)} + D^{(2)} D_{x^2}^{(1)} \right] + \frac{\tau^2}{6} \left[D^{(1)} D^{(1)} D_{x^2}^{(1)} + D^{(1)} D_x^{(1)} D_x^{(1)} - 3D^{(2)} D_x^{(1)} D_{x^2}^{(1)} - D^{(2)} D^{(1)} D_{x^3}^{(1)} - D^{(1)} D_x^{(2)} D_{x^2}^{(1)} - D^{(1)} D^{(2)} D_{x^3}^{(1)} + D^{(2)} D_{x^2}^{(2)} D_{x^2}^{(1)} + 2D^{(2)} D_x^{(2)} D_{x^3}^{(1)} + D^{(2)} D^{(2)} D_{x^4}^{(1)} \right] + \mathcal{O}(\tau^3) \quad (3.12b)$$

³Der Index „E“ kennzeichnet, dass es sich bei diesen Funktionen um aus den Daten gewonnene Schätzungen handelt, die von den nicht indizierten analytischen Funktionen zu unterscheiden sind.

$$D_E^{(2)}(x_0, t_0, \tau) = \frac{1}{2\tau} \langle [x(t_0 + \tau) - x(t_0)]^2 | x(t_0) = x_0 \rangle \quad (3.13a)$$

$$\begin{aligned} &= D^{(2)}(x_0, t_0) + \frac{\tau}{2} \left[D^{(1)}D^{(1)} + 2D^{(2)}D_x^{(1)} + D^{(1)}D_x^{(2)} + D^{(2)}D_{x^2}^{(2)} \right] \\ &\quad + \frac{\tau^2}{6} \left[3D^{(1)}D_x^{(1)} + 3D^{(2)}D_x^{(1)}D_x^{(1)} + D^{(2)}D^{(1)}D_{x^2}^{(1)} \right. \\ &\quad + 2D^{(1)}D_x^{(2)}D_x^{(1)} + 3D^{(1)}D^{(2)}D_{x^2}^{(1)} + 2D^{(2)}D_{x^2}^{(2)}D_x^{(1)} \\ &\quad + 6D^{(2)}D_x^{(2)}D_{x^2}^{(1)} + 4D^{(2)}D^{(2)}D_{x^3}^{(1)} + D^{(1)}D_x^{(1)}D_x^{(2)} + D^{(1)}D^{(1)}D_{x^2}^{(2)} \\ &\quad + D^{(2)}D_{x^2}^{(1)}D_x^{(2)} + 2D^{(2)}D_x^{(1)}D_{x^2}^{(2)} + D^{(2)}D^{(1)}D_{x^3}^{(2)} + D^{(1)}D_x^{(2)}D_{x^2}^{(2)} \\ &\quad \left. + D^{(1)}D^{(2)}D_{x^3}^{(2)} + D^{(2)}D_{x^2}^{(2)}D_{x^2}^{(2)} + D^{(2)}D_x^{(2)}D_{x^3}^{(2)} + D^{(2)}D^{(2)}D_{x^4}^{(2)} \right] \\ &\quad + \mathcal{O}(\tau^3) \end{aligned} \quad (3.13b)$$

Wir sehen, dass das Verfahren für $\tau \rightarrow 0$ die analytischen Funktionen liefert. Für endliche τ hingegen zeigen sich Abweichungen, die wir nun anhand eines einfachen Beispiels untersuchen wollen.

3.1.4 Beispiel: Ornstein-Uhlenbeck-Prozess

Zu analysieren seien nun die Daten eines stationären Ornstein-Uhlenbeck-Prozesses mit $D^{(1)}(x) = -\gamma x$ und $D^{(2)}(x) = Q$. In diesem Fall verschwindet ein Großteil der Ableitungen in (3.12b) und (3.13b) und die Abweichungen für endliches τ ergeben sich als:

$$D_E^{(1)}(x_0) = -\gamma x_0 + \frac{\tau}{2} \gamma^2 x_0 - \frac{\tau^2}{6} \gamma^3 x_0 + \mathcal{O}(\tau^3) \quad (3.14a)$$

$$D_E^{(2)}(x_0) = Q + \frac{\tau}{2} [\gamma^2 x_0^2 - 2Q\gamma x_0] + \frac{\tau^2}{6} [3\gamma^3 x_0^2 + 3Q\gamma^2] + \mathcal{O}(\tau^3) \quad (3.14b)$$

Wir sehen, dass der dominante Fehler in $D^{(2)}$, der für endliches τ entsteht, für $x_0 \gg Q$ quadratisch in x_0 wächst. Tatsächlich zeigt sich dieses Verhalten bei der Auswertung von Zeitreihen häufig. Auf der anderen Seite lässt sich durch rekursives Anwenden der Formeln (3.12b) und (3.13b) auf die erste Schätzung verbunden mit einer Veränderung von τ (z.B. zum nächstgelegenen Verfügbaren) abschätzen, welche Qualität die Schätzung hat und ob eine höhere Genauigkeit durch das Zeit-inkrement τ oder durch andere, die Schätzung beeinflussende Faktoren verhindert wird.

Wir können anhand der hier hergeleiteten Zusammenhänge für den einfachen Ornstein-Uhlenbeck-Prozess verstehen, dass es eine charakteristische Samplingfrequenz gibt, die die Güte der Schätzung beeinflusst. Diese ist proportional zur Dämpfung $f_C \sim \gamma$. Die genauere Untersuchung von Ornstein-Uhlenbeck-Prozessen zeigt, dass die Stärke γ der rücktreibenden Kraft die Geschwindigkeit der Dynamik bestimmt. Die Korrelation zweier Messwerte sinkt deshalb mit steigendem γ . Dieses ist der Grund für die Notwendigkeit der Anpassung der Samplingfrequenz an γ .

3.2 Einfluss einer endlichen Stichprobe auf die Schätzungen

Bislang sind wir davon ausgegangen, dass die Größe der Stichprobe, auf der wir unsere Statistik aufbauen, das Ergebnis nicht beeinflusst. Bei der Analyse empirischer Daten ist die Menge der zur Verfügung stehenden Daten aber immer beschränkt.

Bei Daten in Form von Zeitreihen stationärer Prozesse ist dieser Umstand nicht zu vernachlässigen: Wird der Ortsraum in äquidistante Bins eingeteilt, so variiert die Anzahl der Datenpaare pro Bin stark. Sie ist abhängig von der stationären Verteilung und verschwindet für Prozesse mit natürlichen Randbedingungen für große Entfernung vom Zentrum. Die Anzahl Datenpaare, die zur Bildung der Mittelwerte der Momente verwendet werden können, ist deshalb an einigen Stellen sehr gering.

Seien die Zeiten \tilde{t}_i so gewählt, dass $x(\tilde{t}_i)$ für alle $1 \leq i \leq N$ im Bin um x_0 liegt. Für diese endliche Stichprobe berechnen sich Abschätzungen für Drift und Diffusion dann als

$$D_E^{(1)}(x_0, \tau) = \frac{1}{N} \sum_{i=1}^N \frac{1}{\tau} \{x(\tilde{t}_i + \tau) - x(\tilde{t}_i)\} \quad (3.15a)$$

$$D_E^{(2)}(x_0, \tau) = \frac{1}{2N} \sum_{i=1}^N \frac{1}{\tau} \{x(\tilde{t}_i + \tau) - x(\tilde{t}_i)\}^2 \quad (3.15b)$$

im Grenzübergang $\tau \rightarrow 0$. Für die Abschätzungen von statistischen Fehlern wird die Varianz der einzelnen Summanden, die zur Mittelung beitragen, benötigt. Untersuchen wir zunächst den Drift an der Stelle x_0 für festes τ . Die Varianz σ^2 ergibt sich als:

$$\begin{aligned} \sigma^2 &= \left\langle \left[\frac{1}{\tau} \{x(t + \tau) - x(t)\} - \left\langle \frac{1}{\tau} \{x(t + \tau) - x(t)\} \right\rangle \right]^2 \middle| t \in \{\tilde{t}_1, \dots, \tilde{t}_N\} \right\rangle \\ &= \frac{1}{N} \sum_{i=1}^N \left[\frac{1}{\tau} \{x(\tilde{t}_i + \tau) - x(\tilde{t}_i)\} - D_E^{(1)}(x_0, \tau) \right]^2 \end{aligned} \quad (3.16a)$$

$$\begin{aligned} &= \frac{1}{\tau} \frac{1}{N} \sum_{i=1}^N \frac{1}{\tau} \{x(\tilde{t}_i + \tau) - x(\tilde{t}_i)\}^2 + \left[D_E^{(1)}(x_0, \tau) \right]^2 \frac{1}{N} \sum_{i=1}^N 1 \\ &\quad - 2D_E^{(1)}(x_0, \tau) \frac{1}{N} \sum_{i=1}^N \frac{1}{\tau} \{x(\tilde{t}_i + \tau) - x(\tilde{t}_i)\} \end{aligned} \quad (3.16b)$$

$$= \frac{2}{\tau} D_E^{(2)}(x_0, \tau) - \left[D_E^{(1)}(x_0, \tau) \right]^2 \quad (3.16c)$$

Wir haben die Varianz nun durch die aus den Daten gewonnenen Drift- und Diffusionsfunktionen ausgedrückt. Da die einzelnen Werte statistisch unabhängig sind, skaliert die Varianz bei der Mittelung über N Beiträge nach dem *Gauß'schen Fehlerfortpflanzungsgesetz* mit $\frac{1}{N}$. Ein wichtiges Maß für die Abschätzung des Fehlers einer Messung ist die Wurzel aus der Varianz oder ein Vielfaches davon, die von derselben Dimension ist wie das Messergebnis. Für den Fehler, der für den Drift zu erwarten ist, gilt nun:

$$E \left[D_E^{(1)}(x_0, \tau) \right] = \sqrt{\frac{2}{\tau} \frac{D_E^{(2)}(x_0, \tau)}{N} - \frac{\left[D_E^{(1)}(x_0, \tau) \right]^2}{N}} \quad (3.17)$$

Die Anzahl der Datenpunkte im Intervall kann nun durch die stationäre Verteilung ausgedrückt werden. Liegt eine Zeitreihe mit insgesamt N_0 Messungen vor und beträgt die Größe unserer Bins Δx , so können wir etwa $N(x_0) = N_0 f(x_0) \Delta x$ Messungen für die Bildung des Mittelwertes um die Stelle x_0 heranziehen. Es ergibt sich:

$$E \left[D_E^{(1)}(x_0, \tau) \right] = \sqrt{\frac{2}{\tau} \frac{D_E^{(2)}(x_0, \tau)}{N_0 f(x_0) \Delta x} - \frac{\left[D_E^{(1)}(x_0, \tau) \right]^2}{N_0 f(x_0) \Delta x}} \quad (3.18)$$

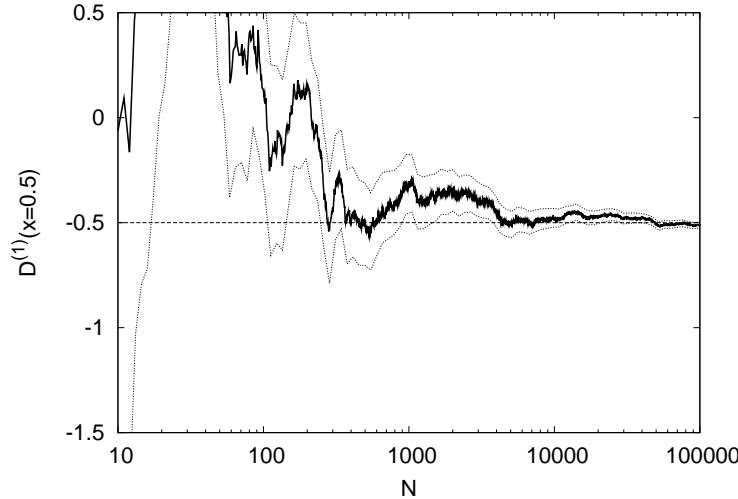


Abbildung 3.1: Abhängigkeit des Fehlers in $D_E^{(1)}(x = 0.5, \tau = 0.01)$ von der Anzahl der Messwerte im Bin am Beispiel eines Ornstein-Uhlenbeck-Prozesses. Die durchgezogene Linie markiert den Drift an der Stelle $x = 0.5$, der aus Mittelung über ein Ensemble von N Messungen zustande kommt, der gestrichelt eingetragene „Schlauch“ darum den Fehlerbereich, wie er nach (3.16a) aus den Daten gewonnen wird. Der Fehler zeigt den zu erwartenden Abfall $\sim N^{-1/2}$ und die Schätzung konvergiert für hohe N gegen den analytischen Werte (gestrichelte Linie).

Auf dieselbe Art und Weise lässt sich für die Diffusion nach (3.15b) verfahren. Es tritt dabei das vierte bedingte Moment $M_E^{(4)}(x_0, \tau)$ auf. Für den Fehler in der Diffusionsfunktion erhalten wir:

$$E \left[D_E^{(2)}(x_0, \tau) \right] = \sqrt{ \frac{1}{4\tau^2} \frac{M_E^{(4)}(x_0, \tau)}{N_0 f(x_0) \Delta x} - \frac{\left[D_E^{(2)}(x_0, \tau) \right]^2}{N_0 f(x_0) \Delta x} } \quad (3.19)$$

Es ist zunächst zu bemerken, dass beide Fehler für eine große Anzahl an Messungen mit $1/\sqrt{N_0}$ verschwinden. Für endliche Stichproben hingegen ist die Größe des Fehlerbereiches anhängig von

- der Besetzung der Bins. Für schwach besetzte Bins wird nur über eine geringe Anzahl an Ereignissen gemittelt, das Ergebnis ist deshalb nicht stabil. Hier lässt sich durch die Berücksichtigung weiterer Messwerte oder die Verbreiterung der einzelnen Bins Abhilfe schaffen. Der Einfluss von N_0 und Δx auf die Qualität der Schätzungen ist bekannt. Häufig überdecken die Fehlerbalken an den Rändern des Messbereiches den Datenbereich völlig, diese Stellen bleiben dann für die Schätzung von Drift und Diffusion in der Regel unberücksichtigt.
- dem Zeitinkrement. Es fällt auf, dass der Fehler im Drift bei endlichen Stichproben im Grenzübergang $\tau \rightarrow 0$ mit $1/\sqrt{\tau}$ divergiert, wenn Drift und Diffusion konvergentes Verhalten zeigen. Wir sehen nun, dass die Qualität der Schätzungen (2.5) für sehr kleine τ drastisch abnimmt. Für die Diffusion hingegen tritt dieses Problem nicht auf. Wir haben angenommen, dass es sich um Daten eines Markow-Prozesses handelt. In diesem Fall sind die Momente n -ter Ordnung mindestens von der Ordnung $n/2$ in τ . Der Fehler in der Diffusionsfunktion divergiert daher im Grenzübergang $\tau \rightarrow 0$ im Allgemeinen nicht.

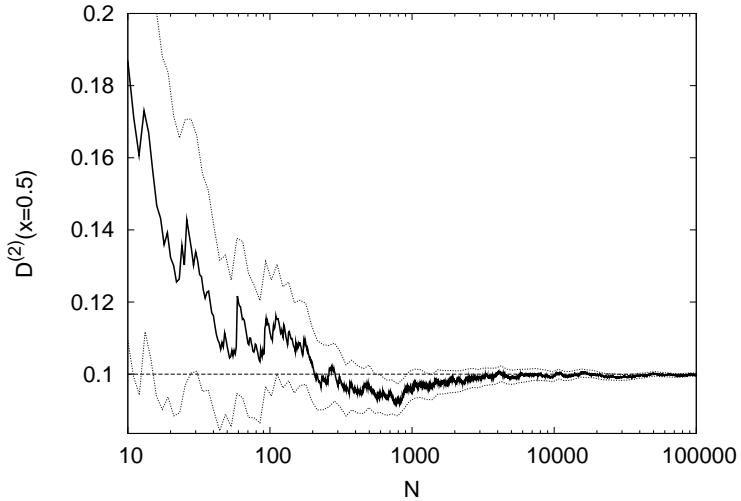


Abbildung 3.2: Abhängigkeit des Fehlers in $D_E^{(2)}(x = 0.5, \tau = 0.01)$ von der Anzahl der Messwerte im Bin am Beispiel eines Ornstein-Uhlenbeck-Prozesses. Die durchgezogene Linie markiert den Drift an der Stelle $x = 0.5$, der aus Mittelung über ein Ensemble von N Messungen zustande kommt, der gestrichelt eingetragene „Schlauch“ darum den Fehlerbereich nach (3.19). Die Konvergenz gegen den analytischen Wert (gestrichelte Linie) ist wesentlich schneller als beim Drift.

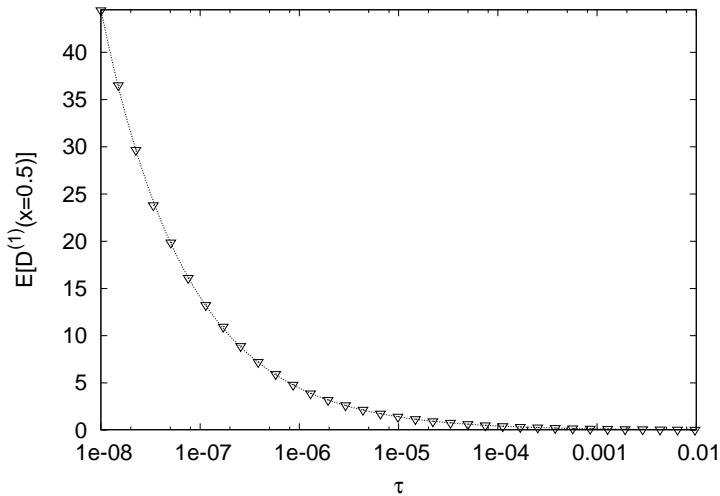


Abbildung 3.3: Verhalten des Fehlers für $\tau \rightarrow 0$ bei Stichproben endlicher Größe. Untersucht wurde ein Ensemble von $N = 10000$ Werten eines Ornstein-Uhlenbeck-Prozesses. Die gestrichelte Linie ist ein Fit der Daten durch die Funktion $f(x) = 0.00445/\sqrt{x}$. Der Fehler zeigt das divergente Verhalten, das nach (3.18) zu erwarten ist.

Eine Minimierung der durch die endliche Stichprobengröße bedingten Fehler ist möglich, wenn sowohl τ als auch Δx nicht zu klein gewählt werden.

Diese beiden unterschiedlichen Fehler, die auftreten können, sind am Beispiel eines numerisch simulierten Ornstein-Uhlenbeck-Prozesses mit $D^{(1)}(x) = -x$ und $D^{(2)}(x) = 0.1$ untersucht, die Ergebnisse sind in den Abb. 3.1 bis 3.3 grafisch dargestellt. Für den ersten Fall wurden dazu Ensembles verschiedener Größe N mit $x(t_0) = 0.5$ für das Zeitinkrement $\tau = 0.01$ erzeugt. Aus diesen können nach (3.15) Schätzungen für Drift und Diffusion an der Stelle 0.5 berechnet werden. Die Entwicklungen der Schätzungen mit steigendem N und die Fehler, die bei der Bildung der Mittelwerte entstehen, sind in den Abb. 3.1 und 3.2 dargestellt. Es ist deutlich zu sehen, dass die Schätzungen für dieses Zeitinkrement gutartig sind, weil sie für hohe N in beiden Fällen gegen die zur Simulation verwendeten Werte konvergieren. Eine genauere Analyse der Fehlers liefert ab ca. 200 Werten eine deutliche $N^{-1/2}$ -Abhängigkeit. Der Fehler in der Diffusionsfunktion ist für den hier verwendeten Zeitschritt wesentlich kleiner, eine sichere Schätzung ist in diesem Fall ab etwa 1000 Datenpaaren möglich.

Weiterhin wurde an einem Ensemble mit $N = 10000$ die Abhängigkeit der Schätzung des Drifts vom Zeitinkrement untersucht. Es ist Abb. 3.3 zu entnehmen, dass der Fehler dieser Schätzung mit $\tau^{-1/2}$ wächst. Die Divergenz für $\tau \rightarrow 0$ ist nicht auf numerische Artifakte wie eine endliche Länge der Mantisse zurückzuführen, sondern hat ihren Ursprung in (3.17). Für kleine τ dominiert der erste Summand unter der Wurzel. Das Einsetzen der analytischen Diffusion und der Größe des Ensembles liefert:

$$E \left[D_E^{(1)}(x_0 = 0.5, \tau) \right] = \sqrt{\frac{2}{\tau} \frac{D_E^{(2)}(x_0 = 0.5, \tau)}{N}} \approx \frac{0.004472}{\sqrt{\tau}} . \quad (3.20)$$

Dieses Ergebnis ist in sehr guter Übereinstimmung mit dem Fit, der an die Datenpunkte angepasst ist.

3.3 Durch Diskretisierung des Ortsraums zu erwartende Fehler

Wir haben gesehen, dass aufgrund der Verkleinerung der Stichprobe die Fehler für kleine Δx ansteigen. Auf der anderen Seite verursachen natürlich grobe Diskretisierungen Fehler, die ihre Ursache in einer Mittelung über einen zu großen Bereich haben. Diese Fehler sollen nun abgeschätzt und – für ein einfaches Beispiel – analytisch berechnet werden.

Dieses Problem wird wie folgt untersucht: Wir gehen wieder davon aus, dass die Stichprobe, auf der die Mittelwerte der Momente berechnet werden, sehr groß ist und die Momente für einzelne Werte damit exakt extrahiert werden können. Sei der Ortsraum nun so diskretisiert, dass für die Berechnung des bedingten Moments an der Stelle x_0 alle Paare $x(t + \tau), x(t)$ mit $x(t) \in [x_0 - \Delta x/2, x_0 + \Delta x/2]$ berücksichtigt werden. Als Schätzung für die bedingten Momente erhält man dann die Mittelung der Momente über den Bin, gewichtet mit der stationären Verteilung f des Prozesses:

$$M_E^{(n)}(x_0, \tau, \Delta x) = \frac{\int_{x_0 - \Delta x/2}^{x_0 + \Delta x/2} dx M^{(n)}(x, \tau) f(x)}{\int_{x_0 - \Delta x/2}^{x_0 + \Delta x/2} dx f(x)} \quad (3.21)$$

Die allgemeinen Überlegungen sollen an dieser Stelle nicht weiter vorangetrieben werden. Statt dessen untersuchen wir einen eindimensionalen Ornstein-Uhlenbeck-Prozess, für den sich das Integral weiter auswerten lässt. Sei $\tau \ll 1$, sodass Terme höherer Ordnung in τ verschwinden. Dann lässt sich der Drift $D_E^{(1)}$, den man nach Diskretisierung der Daten erwarten kann, abschätzen als:

$$D_E^{(1)}(x_0, \Delta x) \approx \frac{\int_{x_0-\Delta x/2}^{x_0+\Delta x/2} dx D^{(1)}(x) f(x)}{\int_{x_0-\Delta x/2}^{x_0+\Delta x/2} dx f(x)} \quad (3.22)$$

Der Ornstein-Uhlenbeck-Prozess ist charakterisiert durch:

$$\begin{aligned} D^{(1)}(x) &= -\gamma x \quad \text{mit} \quad \gamma > 0 \\ D^{(2)}(x) &= Q \\ f(x) &= \sqrt{\frac{\gamma}{2\pi Q}} e^{-\frac{\gamma}{2Q}x^2} \end{aligned} \quad (3.23)$$

Die stationäre Verteilung ergibt sich dabei, bei Verwendung natürlicher Randbedingungen, aus (1.42).

Die Wahrscheinlichkeitsdichte, über die im Nenner integriert wird, lässt sich um x_0 entwickeln. Weil die ungeraden Glieder der Taylor-Entwicklung bei der Integration verschwinden, verbleibt nur ein Fehler in dritter Ordnung von Δx . Dieser sowie die Fehler höherer Ordnung sind proportional zur zweiten bzw. zu höheren Ableitungen von $f(x_0)$ an der Stelle x_0 :

$$\int_{x_0-\Delta x/2}^{x_0+\Delta x/2} dx f(x) = f(x_0)\Delta x + \mathcal{O}(\Delta x^3) \quad (3.24)$$

Im Zähler ist eine deratige Entwicklung nicht nötig. Das Integral lässt sich für den hier betrachteten Prozess analytisch auswerten. Die Berechnung des Integrals ist im Anhang A.2 ausgeführt. Man erhält schließlich für den über einen Bin gemittelten Drift:

$$D_E^{(1)}(x_0, \Delta x) \approx \frac{2}{\Delta x} Q e^{-\frac{\gamma}{2Q}(\frac{\Delta x}{2})^2} \sinh\left(-\frac{\gamma}{Q}x_0 \frac{\Delta x}{2}\right) . \quad (3.25)$$

Sei nun $\Delta x \ll 1$. Wir können die Funktionen im Zähler nach ihrem Argument entwickeln, wie es im Anhang durchgeführt ist. Setzen wir diese Entwicklung für die Bestimmung des relativen Fehlers ein, so ergibt sich dieser als:

$$\frac{D_E^{(1)}(x_0, \Delta x)}{D^{(1)}(x_0)} = 1 + \frac{\gamma}{2Q} \left[\frac{\gamma x_0^2}{3Q} - 1 \right] \left(\frac{\Delta x}{2} \right)^2 + \mathcal{O}(\Delta x^4) \quad (3.26)$$

Diese Ausdrücke lassen sich nun interpretieren. Betrachten wir zunächst (3.22) für große Δx : Im Grenzübergang $\Delta x \rightarrow \infty$ liefert das Integral im Zähler keinen Beitrag und der Nenner ergibt 1. Der Drift, der sich aus der Schätzung ergibt, verschwindet also. Dieses Verhalten ist konsistent mit den natürlichen Randbedingungen, die wir hier vorausgesetzt haben: Würde der Drift bei Mittelung über das gesamte System nicht verschwinden, so gäbe es einen konstanten Fluss durch das System und die stationäre Verteilung würde fernab vom Zentrum nicht verschwinden. Für große Δx wird der Drift (betragsmäßig) also grundsätzlich unterschätzt.

Ferner lässt sich anhand (3.26) der relative Fehler bei sehr feiner Unterteilung untersuchen. Man sieht, dass das Verhältnis von Abschätzung zum ursprünglichen Drift für $\Delta x \rightarrow 0$ gegen 1 konvergiert, wie es schon Ausdruck (3.21) erwarten lässt. Die Korrekturen in zweiter Ordnung von Δx können, abhängig von γ , x_0 und Q , positiv oder negativ ausfallen.

3.4 Einfluss von Messrauschen

Es wurde bereits auf Probleme bei der Abschätzung der dynamischen Funktionen aus den Daten, deren Qualität durch Messrauschen beeinträchtigt ist, hingewiesen. Siefert hat eine Methode vorgestellt, nach der sich bei leicht verrauschten Daten die dynamischen Funktionen als Ableitung der bedingten Momente nach dem Zeitintervall ergeben [37]. Diese Methode ist jedoch auf sehr kleine Rausch-Amplituden beschränkt.

Das Problem der Separation von gemessenen Daten in Messrauschen und einen dynamischen Prozess ist schon seit längerer Zeit von großem Interesse. Hierzu eignen sich die sogenannten „Maximum Likelihood-Verfahren“, die sich Bayes’scher Methoden bedienen, um aus einem gegebenen Prozess den wahrscheinlichsten Satz an ursprünglichen Parametern zu extrahieren. Beispiele für aktuelle Arbeiten auf diesem Gebiet sind z.B. [40, 41, 42] sowie die in diesen Veröffentlichungen aufgeführten Referenzen. Für eine gewisse Klasse von Problemen sind mit diesen Methoden bei einer relativ geringen Anzahl von Datenpunkten gute Ergebnisse zu erzielen. Alle diese Methoden sind jedoch beschränkt auf additives Rauschen. Außerdem machen sie hochdimensionale Integrationen notwendig, weil zur Anwendung des Bayes’schen Theorems und Berechnung der bedingten Wahrscheinlichkeit eines Satzes von Parametern über sämtliche Datenpunkte integriert werden muss.

Frank Böttcher untersucht in [43, 38] das Verhalten der Momente für kleine τ bei verrauschten Daten. Es ergibt sich ein Verfahren, das geeignet ist, auch bei stark verrauschten Daten das externe Rauschen vom internen, dynamischen Rauschen zu trennen und die Stärke des Rauschens und die Parameter des internen Prozesses zu bestimmen. Dieses Verfahren basiert ebenfalls auf der Auswertung der bedingten Momente. Ich habe zu der mathematischen Fundierung dieses Verfahrens beigetragen und möchte es deshalb an dieser Stelle diskutieren und anhand eines Beispiels demonstrieren.

Wir betrachten jetzt Daten, die durch (2.1) mit $\Gamma_M \neq 0$ beschrieben werden. Es gelte weiterhin $f_M(q) = q$. In der Regel handelt es sich beim Messrauschen um ein Signal, das nicht mit dem Zustand q des Systems korreliert ist und kein Gedächtnis besitzt. Idealisiert lässt es sich deshalb durch gaußverteiltes, weißes Rauschen modellieren. Ein Maß für die Stärke des Rauschens ist die Varianz σ^2 der zugehörigen Verteilungsfunktion. Habe Γ_M hier diese Eigenschaften, so gilt für die Wahrscheinlichkeit der Messung eines Wertes $x' \in [x, x + dx]$, unter der Bedingung, dass das System im Zustand q ist:

$$p_\sigma(x|q)dx = \sqrt{\frac{1}{2\pi\sigma^2}} e^{-\frac{(x-q)^2}{2\sigma^2}} dx \quad . \quad (3.27)$$

Wir haben gesehen, dass für Markow-Prozesse Zweipunkt-PDF's eine entscheidende Rolle spielen. Bei unseren von Messrauschen überlagerten Prozessen lassen sich diese Zweipunkt-Übergangswahrscheinlichkeiten mit Hilfe von (1.21) als Integral über das Produkt bedingter Wahrscheinlichkeiten schreiben. Sei $p_\tau(q|q_0)$ die bedingte Übergangswahrscheinlichkeit des unverrauschten Prozesses für ein gewisses, festes τ . Dann gilt für das n -te bedingte Moment der gemessenen Daten x :

$$\begin{aligned} M^{(n)}(x_0, \tau) &= \langle [x(t + \tau) - x(t)]^n | x(t) = x_0 \rangle \\ &= \int dq_0 \int dq \int dx (x - x_0)^n p_\sigma(x|q) p_\tau(q|q_0) \bar{p}_\sigma(q_0|x_0) \end{aligned} \quad (3.28)$$

Diese Formulierung ähnelt der Lösung von Problemen mit Hilfe von Pfadintegralen: Für die Analyse werden die internen, verborgenen Variablen q_0 und q eingeführt. Die auf diese Weise in die Berechnung der Momente eingeführte CPDF p_τ charakterisiert den internen Prozess, der üblicherweise untersucht werden soll. Weil a priori alle

internen Realisierungen mit den gemessenen, verrauschten Werten verträglich sind, muss jeweils über den gesamten Raum integriert werden. Sie werden aber mit der Wahrscheinlichkeit, dass sie unter gegebenen Messwerten x auftreten, gewichtet. Dafür sorgen die bedingten Verteilungen p_σ .

Zu beachten ist, dass der letzte Faktor \bar{p}_σ des Produktes der CPDF's gewissermaßen das zu (3.27) inverse Problem darstellt. Gesucht ist hier die Wahrscheinlichkeit, dass unter gegebenem x_0 die interne Variable q_0 einen bestimmten Wert annimmt. An dieser Stelle muss auf das in Abschnitt 1.1.4 eingeführte Bayes'sche Theorem zurückgegriffen werden, mit Hilfe dessen sich \bar{p}_σ aus (3.27) berechnen lässt. Seien die verschiedenen stationären Verteilungen von x und q durch Indizes σ bzw. τ gekennzeichnet, dann lässt sich die inverse CPDF berechnen als:

$$\bar{p}_\sigma(q_0|x_0) = \frac{p_\sigma(x_0|q_0)f_q(q_0)}{f_x(x_0)} . \quad (3.29)$$

Wir werden diese Berechnung später für einen einfachen stochastischen Prozess durchführen. Die stationäre Verteilung, die im Nenner auftaucht, ist als Faltungsintegral zu berechnen:

$$f_x(x_0) = \int dq_0 p_\sigma(x_0|q_0)f_q(q_0) . \quad (3.30)$$

Wir berechnen nun die ersten beiden Momente des verrauschten Prozesses und werden dabei auf die Momente des verborgenen Prozesses p_τ stoßen. Für $\tau \ll 1$ lassen sich die Entwicklungen (3.7) und (3.10) verwenden und alle Terme $\sim \tau^2$ vernachlässigen. Die Funktionen $D^{(n)}$ beziehen sich auf den verborgenen, dynamischen Prozess.

3.4.1 Berechnung des ersten Momentes

Die Berechnung des ersten Momentes soll für festes Zeitinkrement $\tau \ll 1$ durchgeführt werden. Wir integrieren nun der Reihe nach über die Variablen x , q und q_0 :

$$\begin{aligned} M_E^{(1)}(x_0, \tau) &= \int dq_0 \int dq \int dx (x - x_0) p_\sigma(x|q) p_\tau(q|q_0) \bar{p}_\sigma(q_0|x_0) \\ &= \int dq_0 \int dq p_\tau(q|q_0) \bar{p}_\sigma(q_0|x_0) \int dx (x - q + q - x_0) p_\sigma(x|q) \\ &= \int dq_0 \int dq p_\tau(q|q_0) \bar{p}_\sigma(q_0|x_0) \left\{ \int dx (q - x_0) p_\sigma(x|q) + \int dx (x - q) p_\sigma(x|q) \right\} \end{aligned} \quad (3.31a)$$

Die beiden Integrale, die geklammert sind, können nun ausgewertet werden: Im ersten taucht die Integrationsvariable nicht mehr auf. Integration über die Verteilung liefert aufgrund der Normierung der PDF 1. Es verbleibt also der Faktor $q - x_0$. Das zweite Integral ist das erste Moment einer Gaußfunktion und verschwindet. Wir können nun fortfahren:

$$\begin{aligned} \dots &= \int dq_0 \bar{p}_\sigma(q_0|x_0) \int dq (q - q_0 + q_0 - x_0) p_\tau(q|q_0) \\ &= \int dq_0 \bar{p}_\sigma(q_0|x_0) \left\{ \int dq (q_0 - x_0) p_\tau(q|q_0) + \int dq (q - q_0) p_\tau(q|q_0) \right\} \end{aligned} \quad (3.31b)$$

Auch hier liefert die Integration im ersten der beiden geklammerten Integrale nur den Beitrag $q_0 - q$. Das zweite Integral ist das erste Moment der internen Prozesses. Das einsetzen von (3.7) für $\tau \ll 1$ liefert:

$$\begin{aligned}
 \dots &= \int dq_0 (q_0 - x_0) \bar{p}_\sigma(q_0|x_0) + \tau \int dq_0 D^{(1)}(q_0) \bar{p}_\sigma(q_0|x_0) + \mathcal{O}(\tau^2) \quad (3.31c) \\
 &= \frac{\int dq_0 (q_0 - x_0) p_\sigma(x_0|q_0) f_q(q_0)}{\int dq_0 p_\sigma(x_0|q_0) f_q(q_0)} \\
 &\quad + \tau \frac{\int dq_0 D^{(1)}(q_0) p_\sigma(x_0|q_0) f_q(q_0)}{\int dq_0 p_\sigma(x_0|q_0) f_q(q_0)} + \mathcal{O}(\tau^2) \\
 &= \gamma_1(x_0) + \tau \frac{\int dq_0 D^{(1)}(q_0) p_\sigma(x_0|q_0) f_q(q_0)}{\int dq_0 p_\sigma(x_0|q_0) f_q(q_0)} + \mathcal{O}(\tau^2) \quad (3.31d)
 \end{aligned}$$

Das Ergebnis, das wir hier erhalten haben, ist ein sehr wichtiges Resultat, weil es erklärt, warum die gewöhnliche Abschätzung des Drifts bei verrauschten Daten nicht mehr funktioniert. Es taucht hier ein Term auf, der unabhängig von τ ist, den wir $\gamma_1(x_0)$ nennen wollen. Das widerspricht bisherigen Annahmen z.B. in [37]. Im Grenzübergang $\tau \rightarrow 0$ divergiert der Ausdruck $\frac{1}{\tau} M_E^{(1)}$ und liefert damit keine vernünftige Abschätzung mehr für den Drift. Der hier entwickelte Zusammenhang lässt sich nutzen, wenn man das Verhalten der Momente als Funktion von τ untersucht. Aus dem Graphen lassen sich das Offset γ_1 sowie die Steigung für $\tau \ll 1$, die dem zweiten Summanden von Ausdruck (3.31d) entspricht, extrahieren. Daraus kann man dann – zumindest für einfache Prozesse, für die obige Integrale analytisch zu lösen sind – Informationen über Drift und Rauschen aus dem Graphen gewinnen. Für den Grenzwert kleinen Messrauschen erhält man – wie in [37] vorgeschlagen – den Drift direkt aus der Ableitung des ersten Momentes nach dem Zeitinkrement für $\tau \ll 1$.

3.4.2 Berechnung des zweiten Momentes

Für das zweite Moment ist die Existenz eines Offsets und damit das divergente Verhalten der Abschätzung der Diffusion bekannt. Man weiß, dass das Offset für schwaches Messrauschen $2\sigma^2$ beträgt. Wir untersuchen nun auch das zweite Moment unseres verrauschten Prozesses:

$$\begin{aligned}
 M_E^{(2)}(x_0, \tau) &= \int dq_0 \int dx (x - x_0)^2 p_\sigma(x|q) p_\tau(q|q_0) \bar{p}_\sigma(q_0|x_0) \\
 &= \int dq_0 \int dq p_\tau(q|q_0) \bar{p}_\sigma(q_0|x_0) \int dx (x - x_0)^2 p_\sigma(x|q) \\
 &= \int dq_0 \int dq p_\tau(q|q_0) \bar{p}_\sigma(q_0|x_0) \left\{ \int dx (x - q)^2 p_\sigma(x|q) \right. \\
 &\quad \left. + 2(q - x_0) \int dx (x - q) p_\sigma(x|q) + (q - x_0)^2 \int dx p_\sigma(x|q) \right\} \quad (3.32a)
 \end{aligned}$$

Von den drei Termen, die in der Klammer stehen, entspricht der erste dem zweiten Moment über die Gaußverteilung f_σ und liefert den Beitrag σ^2 . Der zweite Summand, das erste Moment der Gaußverteilung, verschwindet wie schon in der vorangegangenen Rechnung. Die dritte Integration über die normierte Verteilung

liefert 1. Wir können also fortfahren:

$$\begin{aligned} \dots &= \int dq_0 \bar{p}_\sigma(q_0|x_0) \int dq \left[\sigma^2 + (q - x_0)^2 \right] p_\tau(q|q_0) \\ &= \int dq_0 \bar{p}_\sigma(q_0|x_0) \int dq \left\{ \int dx (q - q_0)^2 p_\tau(q|q_0) \right. \\ &\quad \left. + 2(q_0 - x_0) \int dx (q - q_0) p_\tau(q|q_0) + \left[\sigma^2 + (q_0 - x_0)^2 \right] \int dx p_\tau(q|q_0) \right\} \end{aligned} \quad (3.32b)$$

Die ersten beiden Integrale in der Klammer sind gerade das zweite und das erste Moment des internen Prozesses. Wie schon bei der letzten Rechnung setzen wir nun die niedrigsten Ordnungen der Entwicklungen (3.7) und (3.10) ein. Die Integration im letzten Summanden liefert wieder 1. Damit erhalten wir:

$$\begin{aligned} \dots &= \tau \int dq_0 \left[D^{(2)}(q_0) + 2(q_0 - x_0) D^{(1)}(q_0) \right] \bar{p}_\sigma(q_0|x_0) \\ &\quad + \int dq_0 \left[\sigma^2 + (q_0 - x_0)^2 \right] \bar{p}_\sigma(q_0|x_0) + \mathcal{O}(\tau^2) \end{aligned} \quad (3.32c)$$

$$\begin{aligned} &= \tau \frac{\int dq_0 \left[D^{(2)}(q_0) + 2(q_0 - x_0) D^{(1)}(q_0) \right] \bar{p}_\sigma(x_0|q_0) f_q(q_0)}{\int dq_0 p_\sigma(x_0|q_0) f_q(q_0)} \\ &\quad + \frac{\int dq_0 (q_0 - x_0)^2 \bar{p}_\sigma(q_0|x_0)}{\int dq_0 p_\sigma(x_0|q_0) f_q(q_0)} + \sigma^2 + \mathcal{O}(\tau^2) \\ &= \gamma_2(x_0) + \tau \frac{\int dq_0 \left[D^{(2)}(q_0) + 2(q_0 - x_0) D^{(1)}(q_0) \right] \bar{p}_\sigma(x_0|q_0) f_q(q_0)}{\int dq_0 p_\sigma(x_0|q_0) f_q(q_0)} + \mathcal{O}(\tau^2) \end{aligned} \quad (3.32d)$$

Auch hier tritt ein konstanter Term auf, der $\gamma_2(x_0)$ genannt werden soll. Dieser ist abhängig von x_0 und geht nur im Grenzwert verschwindenden Messrauschen in das von Siefert angegebene konstante Offset von $2\sigma^2$ über. Wie schon beim ersten Moment diskutiert, können wir den konstanten sowie den zu τ proportionalen Teil aus der Untersuchung der τ -Abhängigkeit des Momentes für kleine τ gewinnen.

Die Integrale, die in den Ausdrücken (3.31) und (3.32) auftauchen, sehen zunächst einmal sehr unanschaulich aus. Um sie zur Analyse von Daten verwenden zu können, muss in der Regel vorausgesetzt werden, dass diesen Daten einfache stochastische Prozesse zugrunde liegen die durch wenige Parameter beschrieben werden können. Für diese Modelle können obige Integrale gelöst werden und lassen sich dann zur Charakterisierung der Prozesse verwenden. Als Beispiel wird wiederum ein eindimensionaler Ornstein-Uhlenbeck-Prozess betrachtet, den Frank Böttcher mit dieser Methode untersucht hat.

3.4.3 Beispiel: Ornstein-Uhlenbeck-Prozess

Der Prozess sei wie in (3.23) charakterisiert und es liege eine gemessene Zeitreihe $x(t)$ vor, die mit Messrauschen der Varianz σ^2 behaftet sei. Böttcher hat die Offsets γ_i für diesen Prozess berechnet und

$$\gamma_1(x_0, \sigma) = -\frac{\sigma^2}{\sigma^2 + \left(\frac{Q}{2\gamma}\right)^2} x_0^2 \quad (3.33a)$$

$$\gamma_2(x_0, \sigma) = \frac{\sigma^2 \left(\frac{Q}{2\gamma}\right)^4 + \sigma^4 \left(\frac{Q}{2\gamma}\right)^2}{\left[\sigma^2 + \left(\frac{Q}{2\gamma}\right)^2\right]^2} + \frac{\sigma^4}{\left[\sigma^2 + \left(\frac{Q}{2\gamma}\right)^2\right]^2} x_0^2 + \sigma^2 \quad (3.33b)$$

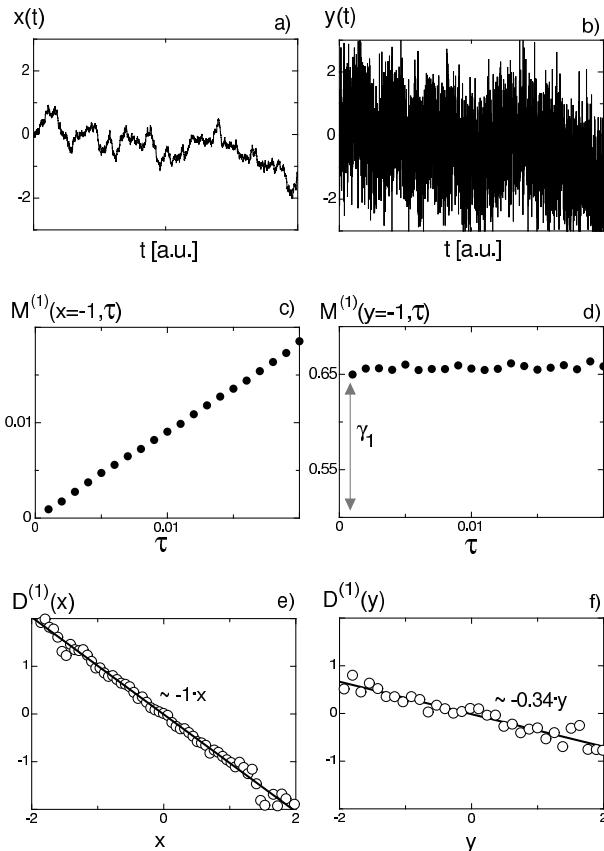


Abbildung 3.4: Analyse synthetischer Daten eines Ornstein-Uhlenbeck-Prozesses mit $\gamma = Q = 1$. Den Plots auf der rechten Seite wurde Messrauschen mit Varianz $\sigma^2 = 1$ hinzugefügt. Zu sehen sind ein Ausschnitt aus der Zeitreihe (oben), das erste Moment aufgetragen gegen das Zeitinkrement τ (mittlere Zeile) sowie die aus der Steigung des mittleren Graphen gewonnene Funktion $D^{(1)}$ (unten). (Entnommen aus: [38])

erhalten. Für $\sigma \ll 1$ zeigen die Offsets also das von Siefert [37] beschriebene Verhalten: Der Drift kann weiterhin aus den Daten bestimmt werden, weil γ_1 verschwindet, für das zweite Moment ergibt sich ein Offset von $2\sigma^2$.

Die beiden Terme $\sim \tau$ im ersten und zweiten Moment lassen sich für unseren Prozess berechnen. Einsetzen von $D^{(1)}$ in (3.31c) liefert für den ersten Summanden:

$$\begin{aligned} & \tau \int dq_0 D^{(1)}(q_0) \bar{p}_\sigma(q_0|x_0) \\ &= -\tau\gamma \left[x_0 \int dq_0 \bar{p}_\sigma(q_0|x_0) + \int dq_0 (q_0 - x_0) \bar{p}_\sigma(q_0|x_0) \right] \\ &= -\tau\gamma [x_0 + \gamma_1(x_0, \sigma)] \quad . \end{aligned} \quad (3.34)$$

Ebenso lässt sich auch der zu τ proportionale Term in (3.32c) auswerten:

$$\begin{aligned} & \tau \int dq_0 \left[D^{(2)}(q_0) - 2(q_0 - x_0) D^{(1)}(q_0) \right] \bar{p}_\sigma(q_0|x_0) \\ &= \tau \int dq_0 [Q - 2\gamma(q_0 - x_0) q_0] \bar{p}_\sigma(q_0|x_0) \\ &= \tau Q - 2\gamma \int dq_0 \left[(q_0 - x_0)^2 + x_0 (q_0 - x_0) \right] \bar{p}_\sigma(q_0|x_0) \\ &= \tau [Q - 2\gamma(\gamma_2(x_0, \sigma) - \sigma^2 + x_0 \gamma_1(x_0, \sigma))] \end{aligned} \quad (3.35)$$

Betrachten wir den Fall kleiner σ , so stellen wir fest, dass in diesem Falle die Steigung des ersten Momentes gegen den Drift konvergiert. In diesem Falle funktioniert also das von Siefert et al. [37] vorgeschlagene Verfahren. Für größere τ wird der Drift nach Sieferts Methode aber unterschätzt. Beim zweiten Moment ist das übliche Verfahren zur Schätzung wieder nur für sehr geringes Messrauschen exakt, andernfalls ergeben sich die berechneten Korrekturen.

Zur Schätzung des Drifts aus einer stark verrauschten Zeitreihe eines Ornstein-Uhlenbeck-Prozesses ist nun wie folgt vorzugehen:

- Die bedingten Momente müssen für ein festes x_0 für verschiedene Zeitintervalle τ berechnet und aufgetragen werden. Die Steigung der Momente sollte für genügend kleines τ linear sein.
- Aus dem Graphen kann nun der Wert $\gamma_1(x_0, \sigma)$ als Ordinate für $\tau \equiv 0$ extrapoliert werden. Weiterhin ist die Steigung $m(x_0)$ des Graphen für kleine τ zu bestimmen.
- Der Parameter γ des Drifts ergibt sich als

$$\gamma(x_0) = -\frac{m(x_0)}{x_0 + \gamma_1(x_0, \sigma)} \quad . \quad (3.36)$$

Dieses Procedere kann für verschiedene x_0 wiederholt und die Qualität unserer Schätzung auf diese Weise erhöht werden.

In Abb. 3.4 ist das Verfahren für einen Ornstein-Uhlenbeck-Prozess demonstriert. Die Analyse beruht auf einer synthetisch generierten Zeitreihe mit den Parametern $\gamma = Q = 1$. Ein Ausschnitt dieser Zeitreihe ist in Abb. 3.4a zu sehen. Diese Daten sind zusätzlich mit einem Messrauschen der Varianz $\sigma^2 = 1$ überlagert worden und in Abb. 3.4b abgebildet. Es ist deutlich, dass das Rauschen den internen Prozess für das blosse Auge unkenntlich macht und eine Analyse der Daten erschwert. Nun werden diese beiden Reihen, die ursprüngliche (linke Seite) sowie die verrauschte (rechte Seite), an der Stelle $x_0 = -1$ analysiert. Die direkte Schätzung des Drifts aus der Ableitung der Momente nach der Zeit liefert für die unverrauschte Zeitreihe das

korrekte Ergebnis, nach dem Hinzufügen des Messrauschen jedoch wird γ mit -0.34 deutlich unterschätzt. Nutzt man jedoch das soeben beschriebene Verfahren und das aus den Daten gewonnene Offset von $\gamma_1(x_0, \sigma) = 0.65$, so erhält man nach (3.36) für den Drift: $D^{(1)}(x_0) = -.97x_0$. Diese Funktion ist in guter Übereinstimmung mit der zur Erzeugung der Daten verwendeten.

Dieses Verfahren kommt, im Gegensatz zu den „Maximum-Likelihood-Verfahren“, mit einem sehr geringen Rechenaufwand aus, weil es keine hochdimensionale Integration erfordert. Entsprechende Verfahren lassen sich auch für andere stochastische Prozesse sowie für eine Schätzung der Diffusion entwickeln. Eine Anwendung auf kompliziertere Prozesse kann allerdings die Berücksichtigung höherer Momente notwendig machen.

Kapitel 4

Iteratives Verfahren zur Bestimmung von Drift und Diffusion

In Kapitel 3 bin ich ausführlich auf Probleme eingegangen, die sich aus der direkten Schätzung von Drift- und Diffusions-Funktionen aus den Momenten der bedingten Verteilungen ergeben. In den letzten Jahren sind insbesondere die durch Auswertung von (2.5) für endliche Zeitinkremente entstehenden Abweichungen untersucht worden. So haben z.B. Ragwitz & Kantz [11, 12] und Friedrich et al. [10] diskutiert, auf welche Weise Korrekturen für endliche Inkremente berücksichtigt werden können. Solche Korrekturen lassen sich aus (3.12b) und (3.13b) entwickeln und können die Güte der Schätzungen in gewissen Grenzen erhöhen. Es ist zu vermuten, dass auf diesem Wege für einige spezielle Prozesse Verfahren gewonnen werden können, die bei ähnlich geringem Rechenaufwand ein genauereres Ergebnis liefern als die alleinige Auswertung der bedingten Momente.

In diesem Kapitel wird ein allgemeines Verfahren [13] vorgestellt, das – basierend auf einer aus einer ersten Abschätzung von Drift- und Diffusionsfunktionen gewonnenen Parametrisierung – eine Optimierung dieser Funktionen auf Grundlage der Zweipunkt-Verbundwahrscheinlichkeiten für *endliches* Zeitinkrement verwendet. Dieses Verfahren lässt sich für endliche Zeitinkremente exakt durchführen. Ferner ist es weitgehend unabhängig von der Diskretisierung des Ortsraumes und der geringen Anzahl an Datenpunkten in der Nähe instabiler Fixpunkte.

4.1 Allgemeines Verfahren

Gegeben sei eine gemessene Datenreihe $\mathbf{x}(t) \in \mathbb{R}^n$ eines stationären stochastischen Markow-Prozesses. Sei ferner angenommen, dass die Dynamik des Prozesses sich durch eine Langevin-Gleichung wie (1.22) beschreiben lasse. Folgendes Verfahren wird zur Bestimmung von Drift und Diffusion vorgeschlagen:

- Berechnungen der Abschätzungen (2.5) für $\mathbf{D}^{(1)}(\mathbf{x})$ und $\mathbf{D}^{(2)}(\mathbf{x})$ für das kleinstmögliche verfügbare τ .
- Einbettung von $\mathbf{D}^{(1)}(\mathbf{x}, \boldsymbol{\alpha})$ und $\mathbf{D}^{(2)}(\mathbf{x}, \boldsymbol{\alpha})$ in eine Familie von Funktionen abhängig von einem Satz von Parametern $\boldsymbol{\alpha}$. Diese Parameter sollen nun so angepasst werden, dass der Prozess bestmöglich charakterisiert ist. Dafür ist nun iterativ vorzugehen:

- Berechnung der CPDF

$$p(\mathbf{x}, t | \mathbf{x}_0, t_0, \boldsymbol{\alpha})$$

für einen Satz $\boldsymbol{\alpha}$. Dieses kann z.B. durch Simulation der zugehörigen Langevin-Gleichung (1.29) oder durch (numerische) Integration der FP-Gleichung geschehen und kann für *endliche Zeitinkremente* τ erfolgen.

- Berechnung der stationären Verteilung f , mit Hilfe derer die Verbund-Übergangswahrscheinlichkeit

$$f(\mathbf{x}, t; \mathbf{x}_0, t_0 | \boldsymbol{\alpha}) = p(\mathbf{x}, t | \mathbf{x}_0, t_0, \boldsymbol{\alpha}) f(\mathbf{x}_0, t_0 | \boldsymbol{\alpha})$$

angegeben werden kann.

- Vergleich der so erhaltenen JPDF mit der aus den Daten extrahierten f_E . Ein geeignetes Maß für die Abschätzung des Abstandes der Verteilungen ist z.B. das *Kullback-Maß* [44]

$$K(\boldsymbol{\alpha}, t, t_0) = \int d^n x \int d^n x_0 f_E(\mathbf{x}, t; \mathbf{x}_0, t_0) \ln \left[\frac{f_E(\mathbf{x}, t; \mathbf{x}_0, t_0)}{f(\mathbf{x}, t; \mathbf{x}_0, t_0 | \boldsymbol{\alpha})} \right] . \quad (4.1)$$

- Minimierung von $K(\boldsymbol{\alpha}, t, t_0)$ unter Variation von $\boldsymbol{\alpha}$. Die zum optimalen $\boldsymbol{\alpha}_0$ korrespondierenden Funktionen $D^{(1)}(\mathbf{x}, \boldsymbol{\alpha}_0)$ und $D^{(2)}(\mathbf{x}, \boldsymbol{\alpha}_0)$ stellen die beste Charakterisierung des Prozesses bezüglich der gewählten Parametrisierung dar.

Die auf diese Weise erhaltenen Ergebnisse können z.B. durch Übergang zu $\tilde{\tau} =: 2\tau$ und wiederholte Anwendung des oben beschriebenen Verfahrens für das Inkrement $\tilde{\tau}$ überprüft werden.

4.2 Eigenschaften des Kullback-Maßes

Das Kullback-Maß eignet sich als Abstands-Maß für PDF's. Seine Informations-theoretische Bedeutung wird unter anderem von Haken [45] beschrieben. Es hat folgende Eigenschaften:

$$K(\boldsymbol{\alpha}, t, t_0) \geq 0 \quad \forall \boldsymbol{\alpha} \quad (4.2a)$$

$$K(\boldsymbol{\alpha}, t, t_0) = 0 \Leftrightarrow f_E(\mathbf{x}, t; \mathbf{x}_0, t_0) = f(\mathbf{x}, t; \mathbf{x}_0, t_0 | \boldsymbol{\alpha}) \quad \forall x, x_0 \quad (4.2b)$$

Die Eigenschaft des ausgezeichneten Minimums lässt sich mit Hilfe der Variationsrechnung verifizieren, wenn wir $\boldsymbol{\alpha} \in \mathbb{R}^N$ mit $N \rightarrow \infty$ voraussetzen, sodass alle linearen Variationen von f dargestellt werden können. Wir suchen ein Extremum des Kullback-Maßes unter der Nebenbedingung, dass beide Verteilungen normiert sind. Es soll also geprüft werden, wann dieses Funktional extremal wird:

$$S[f] = \int d^n x \int d^n x_0 f_E(\mathbf{x}, t; \mathbf{x}_0, t_0) \ln \left[\frac{f_E(\mathbf{x}, t; \mathbf{x}_0, t_0)}{f(\mathbf{x}, t; \mathbf{x}_0, t_0 | \boldsymbol{\alpha})} \right] \quad (4.3)$$

$$+ \lambda_1 \left[\int d^n x \int d^n x_0 f_E(\mathbf{x}, t; \mathbf{x}_0, t_0) - 1 \right]$$

$$+ \lambda_2 \left[\int d^n x \int d^n x_0 f(\mathbf{x}, t; \mathbf{x}_0, t_0 | \boldsymbol{\alpha}) - 1 \right] \quad (4.4)$$

Dabei berücksichtigen die *Lagrange'schen Multiplikatoren* $\lambda_i \neq 0$ die Normierung der PDF's. Wird dieses Funktional bezüglich f extremal, so muss die lineare Variation von S nach f verschwinden. Damit muss gelten:

$$\int d^n x \int d^n x_0 \left\{ \lambda_2 - \frac{f_E(\mathbf{x}, t; \mathbf{x}_0, t_0)}{f(\mathbf{x}, t; \mathbf{x}_0, t_0 | \boldsymbol{\alpha})} \right\} \delta f(\mathbf{x}, t; \mathbf{x}_0, t_0 | \boldsymbol{\alpha}) \stackrel{!}{=} 0 \quad (4.5)$$

Diese Bedingung kann für beliebige Variationen nur erfüllt sein, wenn der Ausdruck in der Klammer für alle \mathbf{x}, \mathbf{x}_0 verschwindet. Es folgt:

$$f(\mathbf{x}, t; \mathbf{x}_0, t_0 | \boldsymbol{\alpha}) = \frac{1}{\lambda_2} f_E(\mathbf{x}, t; \mathbf{x}_0, t_0) \quad \forall \mathbf{x}, \mathbf{x}_0 \quad . \quad (4.6)$$

Diese Bedingung ist aufgrund der notwendigen Normierung der beiden Verteilungen nur mit $\lambda_2 \equiv 1$ zu erfüllen. Für diesen Fall liefert das Kullback-Maß den Wert 0. Es lässt sich leicht überlegen, dass dieser Wert ein Minimum des Maßes darstellen muss. Wir haben also gezeigt, dass das Maß nur ein Minimum besitzt. Dieses wird angenommen, falls die beiden Verteilungen exakt gleich sind. In diesem Fall ist der Satz von Parametern $\boldsymbol{\alpha}$ so gewählt, dass der zu Grunde liegende Prozess durch die parametrisierten Funktionen exakt beschrieben wird.

Bei der (numerischen) Auswertung des Kullback-Maßes sind folgende Fälle für den Integranden zu beachten, falls eine der Verteilungen verschwindet:

$$0 \cdot \ln \left(\frac{0}{z} \right) := 0 \quad \forall z \quad (4.7a)$$

$$z \cdot \ln \left(\frac{z}{0} \right) := \infty \quad \forall z \neq 0 \quad . \quad (4.7b)$$

Das Kullback-Maß ist, so wie es verwendet wird, nicht symmetrisch in den beiden JPDF's f_E und f . Es spielt eine wichtige Rolle, an welcher Stelle die aus den Daten gewonnene Verteilung f_E im Integral steht: Bei empirischen Daten kann es immer geschehen, dass für einen gewissen Bereich keine Daten vorliegen. In diesem Fall kann über die Qualität der parametrischen $D^{(i)}$ keine Aussage gemacht werden. Das Kullback-Maß nach obiger Definition liefert an diesen Stellen nach Konvention keinen Beitrag. Ist ein Wertepaar jedoch gemessen worden, wird bei gewähltem Parameter $\boldsymbol{\alpha}$ aber mit der Wahrscheinlichkeit 0 reproduziert, so liegt eine andere Situation vor. In diesem Fall können wir keine Aussage darüber tätigen, *wie* gut oder schlecht der Prozess durch den Satz $\boldsymbol{\alpha}$ beschrieben wird, er wird überhaupt nicht beschrieben. Das Kullback-Maß divergiert für dieses $\boldsymbol{\alpha}$.

4.3 Vereinfachungen für einige Spezialfälle

Das soeben beschriebene Verfahren lässt sich in einigen Fällen erheblich vereinfachen. Untersuchen wir nun im Speziellen Systeme, die schon bei eindimensionaler Einbettung Markow-Eigenschaften zeigen, sowie sogenannte *Potential-Systeme* in mehreren Dimensionen.

4.3.1 Eindimensionale Systeme

In einer Dimension lässt sich die stationäre Verteilung analytisch berechnen. Verschwindet die stationäre Verteilung für große $|x|$, so geht die stationäre Verteilung aus den Funktionen $D^{(1)}$ und $D^{(2)}$ hervor als

$$f(x) = \frac{\mathcal{N}}{D^{(2)}(x)} \exp \left(\int dx \frac{D^{(1)}(x)}{D^{(2)}(x)} \right) \quad . \quad (4.8)$$

Die Konstante \mathcal{N} gewährleistet die Normierung der PDF. Für eine gute Beschreibung des Prozesses durch die Parameter $\boldsymbol{\alpha}$ müssen die stationären Verteilungen übereinstimmen. Es muss also gelten:

$$f_E(x) = f(x | \boldsymbol{\alpha}) = \frac{\mathcal{N}}{D^{(2)}(x, \boldsymbol{\alpha})} \exp \left(\int dx \frac{D^{(1)}(x, \boldsymbol{\alpha})}{D^{(2)}(x, \boldsymbol{\alpha})} \right) \quad . \quad (4.9)$$

Dieser Zusammenhang lässt sich ausnutzen und man erkennt, dass $D^{(1)}$ abhängig von $D^{(2)}$ und der stationären Verteilung der Daten ist:

$$D^{(1)}(x, \boldsymbol{\alpha}) = D^{(2)}(x, \boldsymbol{\alpha}) \frac{d}{dx} \ln [f_E(x)] + \frac{d}{dx} D^{(2)}(x, \boldsymbol{\alpha}) \quad (4.10)$$

Eine Parametrisierung von $D^{(2)}$ reicht deshalb völlig aus. Dazu eignet sich in den meisten Fällen ein Polynom niedriger Ordnung wie z.B.

$$D^{(2)}(x, \boldsymbol{\alpha}) = Q + ax^2 + bx^4 + \dots \quad \text{mit} \quad \boldsymbol{\alpha} = \begin{pmatrix} Q \\ a \\ b \\ \vdots \end{pmatrix} \quad . \quad (4.11)$$

Die Anzahl der Parameter, bezüglich derer K minimalisiert werden muss, ist in diesem Falle erheblich reduziert, weil sich selbst komplizierte Drift-Funktionen aus der parametrischen Darstellung der Diffusion sowie der Ableitung des Logarithmus der stationären Verteilung, die aus den Daten gewonnen werden kann, berechnen lassen. Gleichung (4.9) führt außerdem zu einer vereinfachten Berechnung des Kullback-Abstandes der Verteilungen. Für den hier betrachteten Prozess gilt:

$$K(\boldsymbol{\alpha}, t, t_0) = \int dx \int dx_0 f_E(x, t; x_0, t_0) \ln \left[\frac{p_E(x, t|x_0, t_0)}{p(x, t|x_0, t_0, \boldsymbol{\alpha})} \right] \quad . \quad (4.12)$$

Notwendig für die Anwendbarkeit dieser Vereinfachung ist, dass die Anzahl der Datenpunkte genügend groß für die Berechnung der Ableitung des Logarithmus der stationären Verteilung ist. Andernfalls muss eine vollständige Parametrisierung beider Funktionen gewählt werden.

4.3.2 Potential-Systeme

Diese Methode kann umgehend auf sogenannte *Potential-Systeme* ausgeweitet werden. Das sind Systeme, für die gilt

$$\mathbf{D}^{(1)}(\mathbf{x}) = -\nabla V(\mathbf{x}) \quad \text{und} \quad (4.13a)$$

$$D_{ij}^{(2)}(\mathbf{x}) = Q\delta_{ij} \quad . \quad (4.13b)$$

Auch für diese Systeme existiert eine stationäre Lösung der Fokker-Planck-Gleichung, die sich zur Berechnung des Drifts nutzen lässt. Die stationäre Lösung mit den oben eingeführten Größen lautet

$$f(\mathbf{x}) = \mathcal{N} \exp \left[-\frac{V(\mathbf{x})}{Q} \right] \quad . \quad (4.14)$$

Wählen wir $\alpha = Q$ als freien Parameter und setzen, wie schon im vorigen Abschnitt geschehen, die stationären Verteilungen der parametrisierten Funktionen und des zu analysierenden Prozesses gleich, so ergibt sich das Drift-Potential als

$$V(\mathbf{x}|Q) = Q \ln(\mathcal{N}) - Q \ln [f_E(\mathbf{x})] \quad . \quad (4.15)$$

Der erste Summand ist konstant und liefert somit keinen Beitrag zum Drift. Q ist nun der einzige freie Parameter im System, der optimiert werden muss. Wir können bei der Berechnung wieder ausnutzen, dass sich die stationären Verteilungen nun für jedes Q nach Voraussetzung gleichen. Damit vereinfacht sich das Kullback-Maß, ähnlich wie im letzten Abschnitt:

$$K(Q, t, t_0) = \int d^n x \int d^n x_0 f_E(\mathbf{x}, t; \mathbf{x}_0, t_0) \ln \left[\frac{p_E(\mathbf{x}, t|\mathbf{x}_0, t_0)}{p(\mathbf{x}, t|\mathbf{x}_0, t_0, Q)} \right] \quad . \quad (4.16)$$

Das allgemeine Verfahren, das auf den ersten Blick aufwändig scheint, lässt sich für zwei Fälle also erheblich vereinfachen.

Kapitel 5

Beispiele

Im vorangegangenen Kapitel ist ein Verfahren zur Analyse stochastischer Prozesse vorgestellt worden. Ich untersuche nun drei Beispiele, für die die in Abschnitt 4.3.1 und 4.3.2 diskutierten Vereinfachungen des allgemeinen Verfahrens verwendet werden können. Dieses führt zu einer Reduktion der freien Parameter und macht jeweils eine grafische Lösung des Minimierungsproblems möglich.

In den Fällen, in denen diese Vereinfachungen nicht genutzt werden können, ist die Minimierung des Kullback-Maßes in einem höherdimensionalen Parameterraum notwendig, die in der Regel grafisch nicht mehr zu bewerkstelligen ist. In diesem Fall kann das Minimum aber durch Gradienten-Verfahren bestimmt werden. Da aus der Parametrisierung der Schätzungen bereits ein geeigneter Startpunkt bekannt ist, konvergieren diese Verfahren im Allgemeinen gegen den optimalen Satz von Parametern. Ferner beschreibt Weinstein [46] eine iterative und an das Kullback-Maß angepasst Methode zur Minimierung der Parameter.

Die numerische Umsetzung des Minimalisierungsproblems ist entscheidend für den Rechenaufwand der Analyse und die Genauigkeit der erhaltenen Ergebnisse. Zunächst werden deshalb einige numerische Grundlagen, die bei Umsetzung der drei Beispiele verwendet werden, eingeführt und an unsere Problemstellung angepasst. Im Anhang B ist der Quelltext eines Programms, das die Minimierung des Kullback-Abstandes bezüglich zweier Parameter durchführt, abgedruckt. Mit diesem Programm wurde im zweiten Beispiel gearbeitet, die Konzepte, die in die numerische Umsetzung eingeflossen sind, werden im Rahmen dieses Beispiels erläutert. Ich hoffe, mit der ausführlichen Einführung und dem kommentierten Programmcode Interessierten, die sich in Zukunft mit der Kullback-Minimierung beschäftigen, den Einstieg zu erleichtern.

5.1 Numerische Grundlagen

5.1.1 Integration von Langevin-Gleichungen

Zur Berechnung der CPDF's für endliche Zeitintervalle wird in den hier präsentierten Beispielen die numerische Simulation der Langevin-Gleichungen (1.29) verwendet. Diese werden für ein Ensemble von Zeitreihen, die mit der Anfangsbedingung verträglich sind, simuliert und die CPDF als Ensemble-Mitte über die Menge der nach Ende der Simulation erhaltenen Werte gebildet. Die genaue Implementierung dieses Verfahrens ist dem Programmbeispiel sowie den Erläuterungen zu entnehmen.

Bei der numerischen Integration der Gleichung muss der Zeitschritt hinreichend klein gewählt werden. Das Inkrement, für das die Daten benötigt werden, muss deshalb im Allgemeinen zunächst in wesentlich feinere Zeitschritte unterteilt werden.

5.1.2 Numerische Integration der Fokker-Planck-Gleichung

Statt dessen können die CPDF's für endliche Zeiten auch durch numerische Integration der FP-Gleichung berechnet werden. Auch diese Vorgehensweise hat ihre Stärken und kann, vor allem bei Problemen, die eine höherdimensionale Einbettung erfordern, effektiver sein als die Simulation von Ensembles. Der Vollständigkeit halber soll die numerische Integration der Gleichung mit Hilfe eines effektiven, impliziten Verfahrens für eine eindimensionale Einbettung an dieser Stelle erläutert werden. Die Verallgemeinerung auf höherdimensionale Probleme ist entsprechend.

Wir müssen uns zunächst mit der numerischen Differentiation beschäftigen. Man kann Koonin [47] Formeln für Vorwärts- und Rückwärts-Rekursion, sowie die numerisch stabilere symmetrische Formel zur Differentiation entnehmen:

$$\frac{d}{dx} f(x) \approx \frac{f(x + \Delta x) - f(x)}{\Delta x} + \mathcal{O}(\Delta x) \quad (\text{Vorwärts-Rekursion}) \quad (5.1a)$$

$$\frac{d}{dx} f(x) \approx \frac{f(x) - f(x - \Delta x)}{\Delta x} + \mathcal{O}(\Delta x) \quad (\text{Rückwärts-Rekursion}) \quad (5.1b)$$

$$\frac{d}{dx} f(x) \approx \frac{f(x + \Delta x) - f(x - \Delta x)}{2\Delta x} + \mathcal{O}(\Delta x^2) \quad (\text{Symmetrische R.}) \quad (5.1c)$$

Die symmetrische Formel ergibt sich als Mittelwert von Vorwärts- und Rückwärts-Rekursionsformeln. Im Grenzwert $\Delta x \rightarrow 0$ liefern diese Ausdrücke die üblichen Ableitungen der Funktionen. Numerisch müssen sie für endliche Δx ausgewertet werden. Der Vorteil des symmetrischen Verfahrens liegt im Verschwinden des Fehlers in erster Ordnung von τ . Für die zweite Ableitung ergibt sich folgender Zusammenhang:

$$\frac{d^2}{dx^2} f(x) = \frac{f(x + \Delta x) - 2f(x) + f(x - \Delta x)}{\Delta x^2} + \mathcal{O}(\Delta x^2) \quad . \quad (5.2)$$

Mit Hilfe dieser Ausdrücke lässt sich ein Verfahren für die numerische Integration der eindimensionalen FP-Gleichung angeben. Seien Ort und Zeit diskretisiert, sodass $x(i) = x_0 + i\Delta x$ und $t(j) = t_0 + j\tau$ gilt. Statt $f(x(i), t(j))$ wird nun abkürzend $f(i, j)$ verwendet. Setzen wir nun die obenstehenden Formeln für die Ableitungen ein und wählen für die partielle Ableitung nach der Zeit die Vorwärtsrekursion, so erhalten wir:

$$\begin{aligned} f(i, j + 1) - f(i, j) &= -\frac{D^{(1)}(i + 1)f(i + 1, j) - D^{(1)}(i - 1)f(i - 1, j)}{2\Delta x} \\ &+ \frac{D^{(2)}(i + 1)f(i + 1, j) - 2D^{(2)}(i)f(i, j) + D^{(2)}(i - 1)f(i - 1, j)}{\Delta x^2} \end{aligned} \quad (5.3)$$

Es soll die Entwicklung von PDF's für endliche Zeiten untersucht werden. Dazu ergibt sich folgende Rekursion für die Integration der FP-Gleichung:

$$\begin{aligned} f(i, j + 1) &= f(i, j) + \tau \left\{ -\frac{D^{(1)}(i + 1)f(i + 1, j) - D^{(1)}(i - 1)f(i - 1, j)}{2\Delta x} \right. \\ &\left. + \frac{D^{(2)}(i + 1)f(i + 1, j) - 2D^{(2)}(i)f(i, j) + D^{(2)}(i - 1)f(i - 1, j)}{\Delta x^2} \right\} \end{aligned} \quad (5.4)$$

Diese Formel stellt eine explizite Methode, auch Integration nach *Euler* genannt, dar und lässt sich zur Lösung der FP-Gleichung verwenden. Das Verfahren lässt sich aber noch leicht modifizieren und damit der Konvergenz der Lösung bezüglich der Größe des Zeitinkrements τ erheblich verbessern. Dazu geht man in der zweiten Zeile von (5.3) von $f(i, j)$ über zu $f(i, j + 1)$. Wir verwenden also auf der rechten

Seite der Gleichung bereits den neuen Wert der Funktion zur Zeit $j + 1$. Das liefert folgende Rekursion für f :

$$f(i, j + 1) = \left\{ \frac{1}{\tau} + 2 \frac{D^{(2)}(i)}{\Delta x^2} \right\}^{-1} \left\{ -\frac{D^{(1)}(i + 1)f(i + 1, j) - D^{(1)}(i - 1)f(i - 1, j)}{2\Delta x} + \frac{f(i, j)}{\tau} + \frac{D^{(2)}(i + 1)f(i + 1, j) + D^{(2)}(i - 1)f(i - 1, j)}{\Delta x^2} \right\} \quad (5.5)$$

Man beachte, dass in die Berechnung von f an den Stellen gerader i nur Verteilungen an Stellen ungerader i und vice versa eingehen. Die Integration ist daher sinnvoller Weise im „Schachbrett-Muster“¹ durchzuführen. Mit Hilfe dieser Iteration lässt sich die zeitliche Entwicklung von PDF's simulieren. Eine tiefergehende Diskussion dieser Methoden sowie weitere Verfahren zur Integration (parabolischer) partieller Differentialgleichungen finden sich z.B. in [48].

5.1.3 Numerische Ableitung von Daten

Die Zusammenhänge (5.1) liefern für die Ableitung von analytischen Funktionen gute Ergebnisse. Müssen jedoch verrauchte Daten differenziert werden, eignen sich die Ausdrücke nur bedingt. Bei den in 4.3 diskutierten Fällen, die Gegenstand der folgenden Untersuchungen sein werden, ist jeweils die Ableitung des Logarithmus der stationären Verteilung zu berechnen. Durch die Logarithmierung ist die Qualität der Daten vor allem an den Minima der PDF sowie zu den Rändern hin sehr schlecht. Wir benötigen deswegen andere Verfahren zur Berechnung der Ableitung.

Ein gängiges Verfahren ist die Ausführung der Differentiation im Fourierraum. Gehen wir davon aus, dass wir die Ableitung der analytischen, quadratintegrablen Funktion $f(x)$ berechnen wollen. Weil sie an den Grenzen verschwinden muss, existiert die Fourier-Transformierte

$$\tilde{f}(k) = \frac{1}{\sqrt{2\pi}} \int dx f(x) e^{-ikx} \quad . \quad (5.6)$$

Zur Berechnung der Ableitung können wir nun die Rücktransformation einsetzen und die Ableitung mit der Integration vertauschen:

$$\frac{d}{dx} f(x) = \frac{d}{dx} \int dk \tilde{f}(k) e^{ikx} = \int dk \frac{\partial}{\partial x} \tilde{f}(k) e^{ikx} = i \int dk k \tilde{f}(k) e^{ikx} \quad (5.7)$$

Die Ableitung nach dem Ort entspricht also einer Multiplikation mit ik im Fourierraum. Die Berechnung der Fourier-Transformierten und die Rücktransformation ist numerisch nicht schwierig und insbesondere bei höherdimensionalen Problemen oder in den Fällen, in denen Ableitungen verschiedener Ordnung einer Funktion gebildet werden müssen, hilfreich.

Im Folgenden wird jedoch auf ein anderes Verfahren zur Differentiation zurückgegriffen, das in [49] beschrieben ist. Es beruht auf einer linearen Approximation der Funktion an der Stelle x , an der die Ableitung zu berechnen ist. Für kleine Abweichungen h soll gelten:

$$f(x + h) \approx a + bh \quad (5.8)$$

¹Zunächst werden die Verteilungen an Stellen gerader, dann an Stellen ungerader i durchgeführt. Bei Erweiterung der Rekursion auf zwei Dimensionen errinnert diese Vorgehensweise an ein Schachbrett.

Die Ableitung an der Stelle x berechnet sich dann als:

$$\frac{d}{dy} f(y) \Big|_{y=x} = b \quad (5.9)$$

Wir müssen b nun aus diskreten Daten bestimmen. Das soll mit Hilfe der Minimierung des quadratischen Abstandes der linearen Approximation geschehen und entspricht der Berechnung einer *Ausgleichsgeraden*. Es ist zu beachten, dass unsere Parameter a und b damit von den Werten abweichen können, die sich im Sinne der Koeffizienten einer Taylor-Reihe ergeben würden. Seien die x -Werte wieder diskret mit $x(i) = x_0 + i\Delta x$. Berücksichtigt man zur Berechnung der Approximation an f an der Stelle i die $2n + 1$ Funktionswerte an den Stellen $i - n, \dots, i + n$, so ist folgendes Optimierungsproblem zu lösen:

$$\text{Minimalisiere } \sum_{k=-n}^n [a + bk\Delta x - f(i+k)]^2 \text{ unter Variation von } a \text{ und } b. \quad (5.10)$$

Im gesuchten Minimum muss dann gelten:

$$\frac{\partial}{\partial a} \sum_{k=-n}^n [a + bk\Delta x - f(i+k)]^2 = \frac{\partial}{\partial b} \sum_{k=-n}^n [a + bk\Delta x - f(i+k)]^2 \stackrel{!}{=} 0 \quad (5.11)$$

Es ist lediglich der Parameter b zu bestimmen. Dieser lässt sich, abhängig von i und n , berechnen als²

$$b(i, n) = \frac{[2n+1] \sum_{k=-n}^n x(i+k)f(i+k) - \left[\sum_{k=-n}^n x(i+k) \right] \left[\sum_{k=-n}^n f(i+k) \right]}{[2n+1] \sum_{k=-n}^n [x(i+k)]^2 - \left[\sum_{k=-n}^n x(i+k) \right]^2} \quad (5.12)$$

und ist eine gute Näherung für die Steigung an der Stelle i . Diese Formel stellt eine numerisch leicht zu implementierende Berechnung der Ableitung an der Stelle i dar, deren Empfindlichkeit gegen Rauschen sich auf Kosten der Lokalität der Ableitung mit Hilfe der Parameters n einstellen lässt. Dieses Verfahren findet bei der Berechnung der Driftfunktionen Verwendung, weil es eng mit einer geometrischen Anschauung verknüpft ist und sich die Präzision leicht aus der Qualität der Daten abschätzen lässt. Eine Erweiterung auf die partielle Differentiation im Mehrdimensionalen verläuft entsprechend.

5.1.4 Monte-Carlo-Techniken zur Integration

Bei Durchführung der zur Bestimmung der Kullback-Abstandes notwendigen Integration kann man sich einige Elemente aus der Strategie der Monte-Carlo-Integration (siehe z.B. [47]) zu Nutze machen.

Monte-Carlo-Methoden nutzen aus, dass sich das Integral

$$I = \int_0^1 dx f(x) \quad (5.13)$$

näherungsweise als

$$I \approx \frac{1}{N} \sum_{i=1}^N f(x_i) \quad (5.14)$$

²Gleichung (5.12) wurde in dieser überarbeiteten Version berichtigt.

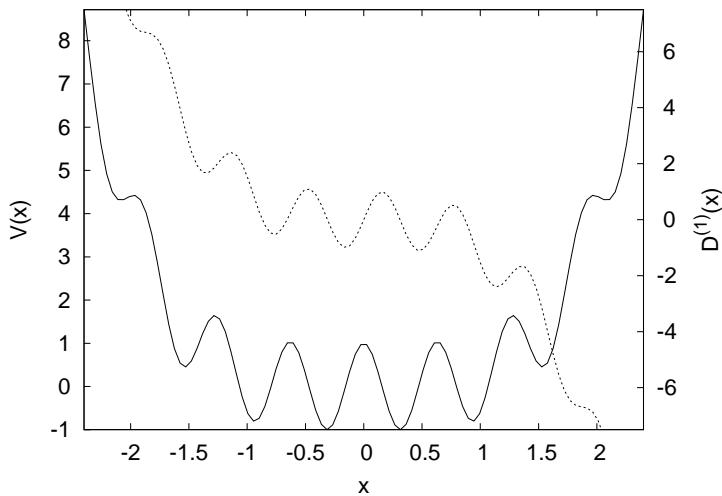


Abbildung 5.1: Zeitreihe I: Zur Simulation verwendeteter Drift (gestrichelt) und zu gehöriges Drift-Potential (durchgezogene Linie).

bestimmen lässt, wobei $x_i \in [0, 1]$ gleichmäßig über das Intervall verteilt, von einander unabhängige Zufallszahlen sein sollen. Dabei gilt für die Varianz σ_I^2 des Integrals:

$$\sigma_I^2 = \frac{1}{N} \left\{ \frac{1}{N} \sum_{i=1}^N [f(x_i)]^2 - \left[\frac{1}{N} \sum_{i=1}^N f(x_i) \right]^2 \right\} = \frac{1}{N} \sigma_f^2 \quad . \quad (5.15)$$

Wir können also die Berechnung des Integrals als Mittelung von f über das Intervall betrachten. Wenn wir die Stützstellen zufällig über unser Integrationsintervall verteilen, so konvergiert der Fehler bei der Integration mit $N^{-1/2}$ gegen 0. Maßgeblich für die Konvergenz ist der Fehler σ_f , der angibt, wie stark f innerhalb des Integrationsintervalls um seinen Mittelwert schwankt. Durch die Einführung einer geeigneten Gewichtungsfunktion lässt sich die Schwankung um den Mittelwert in der Regel verringern und damit die Geschwindigkeit der Konvergenz erhöhen. Verfahren, die auf diesem Prinzip beruhen, bieten insbesondere bei der Integration über viele Integrationsvariablen Vorteile und sind gut untersucht und dokumentiert.

Bei der im Rahmen dieser Arbeit gewählten Implementierung der Kullback-Integration ist der Ortsraum von Beginn an diskretisiert, eine Auswertung des Integrals an durch Zufallsvariable festgelegten Stellen ist deshalb nicht sinnvoll. Auf der anderen Seite ist es bei diskretisiertem Ortsraum möglich, selbstständig für eine gleiche Gewichtung der verschiedenen Regionen zu sorgen. Deshalb wurden von der Monte-Carlo-Strategie folgende Prinzipien übernommen: Der Kullback-Abstand wird nur an einzelnen, gleichmäßig verteilten Stellen ausgewertet. Das „Zwischenergebnis“ kann mit der Anzahl der bis dahin durchgeföhrten Messungen gewichtet werden. Auf diese Art und Weise ist in den meisten Fällen eine schnelle Konvergenz des Verfahrens zu erkennen.

5.2 Eine Dimension, additives Rauschen

In diesem ersten Beispiel wird die Anwendung des in Kapitel 4 beschriebenen Verfahrens auf eine eindimensionale Zeitreihe gezeigt. Diese Daten werden synthetisch

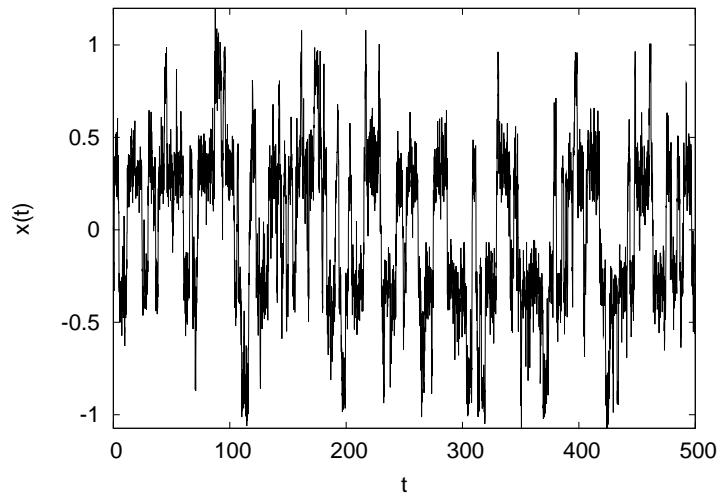


Abbildung 5.2: Ausschnitt aus der Zeitreihe I.

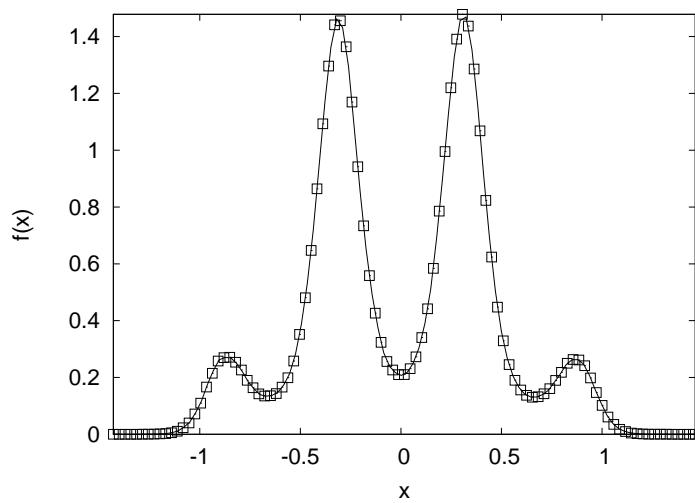


Abbildung 5.3: Zeitreihe I: Aus den Daten gewonnene stationäre Verteilung.

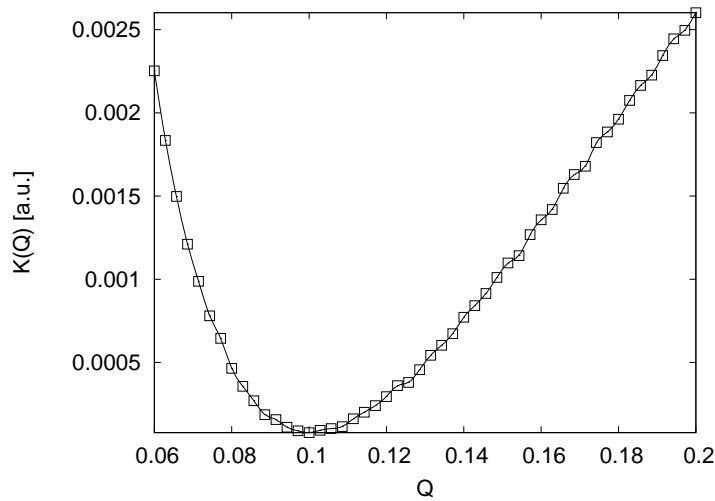


Abbildung 5.4: Zeitreihe I: Optimierung des Kullback-Maßes bezüglich Q für additives Rauschen. Das Minimum an der Stelle $Q = 0.1$ ist stark lokalisiert.

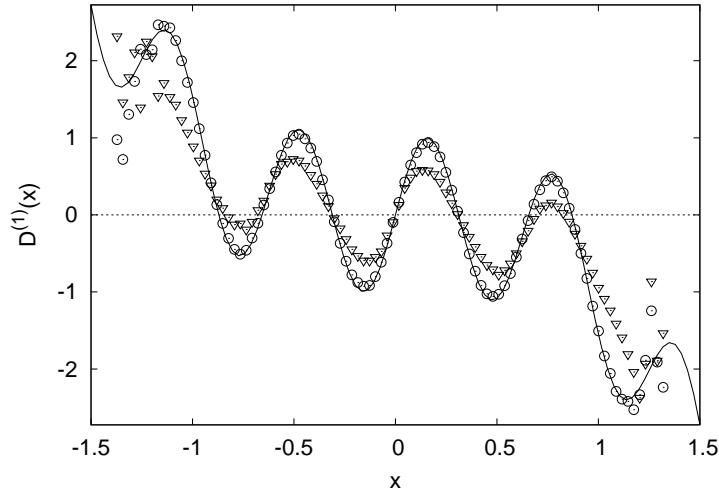


Abbildung 5.5: Zeitreihe I: Drift-Funktion, die zur Simulation der Daten verwendet wurde (durchgezogene Linie), und die Werte, die mittels einer direkten Schätzung für das kleinstmögliche τ (Dreiecke) ermittelt sowie unter Verwendung der mit Hilfe des Kullback-Maßes optimierten konstanten Diffusion Q aus der statio-nären Verteilung berechnet wurden. Eine qualitative Verbesserung der Resultate ist offensichtlich.

durch numerische Integration der Langevin-Gleichung gewonnen. Ziel ist es, die zur Erzeugung der Daten verwendeten Funktionen bestmöglich zu reproduzieren. Folgendes System wurde simuliert:

$$D^{(1)}(x) = \sin(10x) - x^3 \quad (5.16a)$$

$$D^{(2)}(x) = 0.1 \quad (5.16b)$$

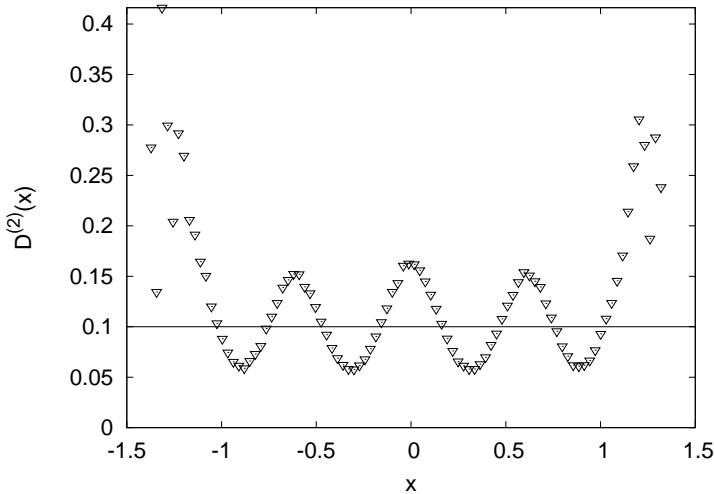


Abbildung 5.6: Zeitreihe I: Konstante Diffusion (Linie), die für die Simulation verwendet und durch Kullback-Mimimierung reproduziert wurde. Die Diffusion, die sich für das minimale τ aus den bedingten Momenten gewinnen lässt (Dreiecke), zeigt aufgrund des hohen Zeitinkrements deutliche Artifakte der Drift-Funktion.

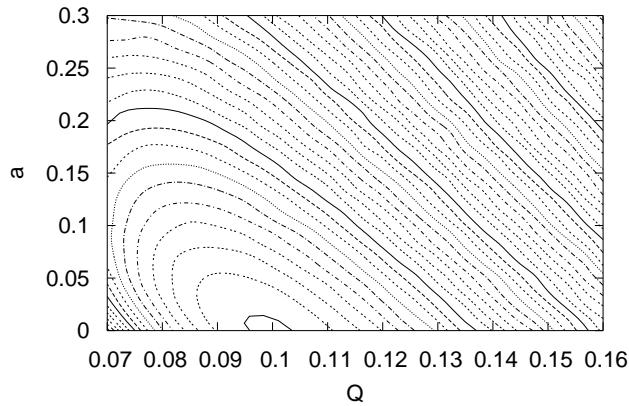


Abbildung 5.7: Zeitreihe I: Minimierung der Diffusionsfunktion $D^{(2)}(x|Q, a) = Q + ax^2$ bezüglich der Parameter Q und a . Die Linien sind Höhenlinien, die, startend von 0.0004 in der Bildmitte, äquidistant im Abstand von 0.00034 liegen. Es ist zu vermuten, dass Terme $\sim x^2$ keinen wichtigen Beitrag zur Diffusion liefern.

Die Drift-Funktion und das zugehörige Drift-Potential $V(x) = x^4 + \frac{1}{10} \cos(10x)$ sind in Abb. 5.1 skizziert. Die Minima entsprechen stabilen, die Maxima instabilen Fixpunkten der Bewegung. Man sieht, dass es sich hier um ein multistabiles Potential handelt. Die Stärke der Diffusion ist so gewählt, dass die stationäre Verteilung schnell erreicht wird.

Die Daten sind mit dem Zeitinkrement $\tau = 0.1$ abgespeichert. Dieses ist so groß, dass die direkte Schätzung der Diffusion aus den Daten keine eindeutigen Aussagen

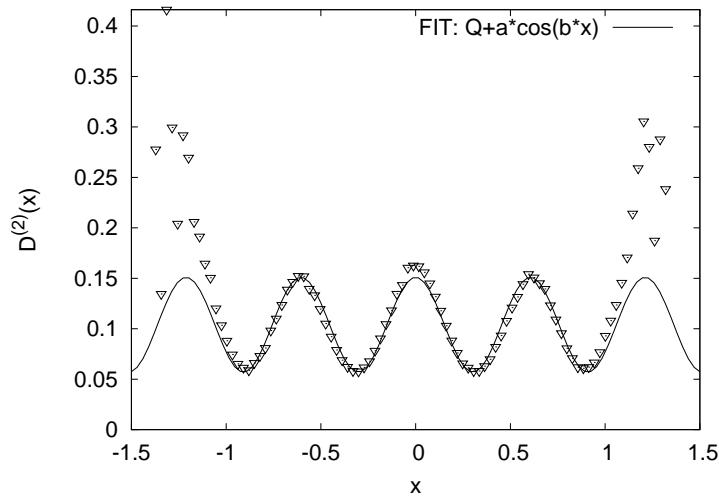


Abbildung 5.8: Zeitreihe I: Diffusion $D^{(2)}$, für das kleinstmögliche τ aus der Zeitreihe extrahiert. Die Datenpunkte lassen sich, im relevanten Bereich, durch den von GNUMplot berechneten Fit beschreiben. Gewählt wird deswegen für den Drift die Parametrisierung $D^{(2)}(x|Q, a) = Q + a \cos(10.3772x)$.

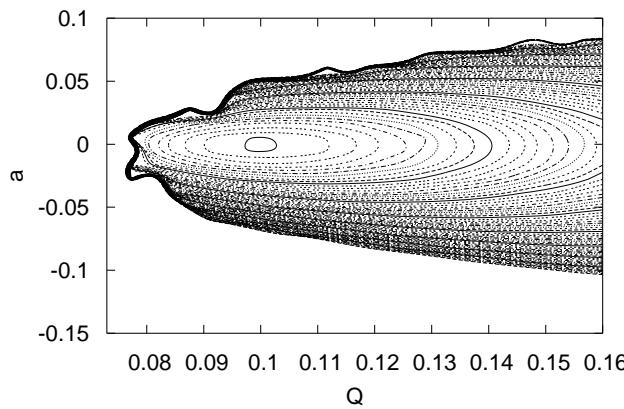


Abbildung 5.9: Zeitreihe I: Minimierung der Parameter Q und a der in Abb. 5.8 skizzierten Parametrisierung. Die Höhenlinien starten bei 0.0004 und liegen äquidistant im Abstand 0.0005. Das Minimum bei $a \equiv 0$ ist deutlich zu erkennen.

mehr zulässt. Das lässt sich wie folgt verstehen: In (3.13b) treten in der Korrektur proportional zu τ^n immer Ableitungen des Drifts bis zur Ordnung n auf. In unserem Fall reproduziert sich der Drift, bis auf eine Verschiebung der Phase um $-\frac{\pi}{20}$ und den Faktor 10, bei jeder Ableitung selbst. Auf der anderen Seite kommt für jede Ordnung von n ein Faktor τ hinzu. Es ist deutlich, dass die in diesem Beispiel gewählte Abspeicherung der Daten mit $\tau = \frac{1}{10}$, die eine untere Grenze bei der Auswertung der Daten darstellt, die Vernachlässigung aller Terme $\sim \tau$ in (3.13b) nicht erlaubt und damit die direkte Schätzung nicht zum richtigen Ergebnis führen kann. Damit ist

die Analyse der Daten mit der im Rahmen dieser Arbeit vorgeschlagenen Methode interessant.

5.2.1 Analyse der Zeitreihe

Die simulierte Zeitreihe umfasst 10^6 Datenpunkte. Einen Ausschnitt aus der Zeitreihe zeigt Abb. 5.2. Zunächst wurde aus den Daten die stationäre Verteilung gewonnen, die in Abb. 5.3 zu sehen ist. Es zu erkennen, dass die Struktur und die Symmetrien der Verteilung denen des Drift-Potentials in Abb. 5.1 entsprechen.

Die Analyse der Daten erfolgt nun für drei verschiedene Parametrisierungen, für

1. additives Rauschen,
2. multiplikatives Rauschen der Form $D^{(2)}(x) = Q + ax^2$,
3. multiplikatives Rauschen, dessen Form aus der Schätzung der Diffusion aus den bedingten Momenten gewonnen werden soll.

In allen drei Fällen lässt sich der durch die gewählte Parametrisierung und die stationäre PDF festgelegte Drift verwenden.

Annahme: Diffusion der Form $D^{(2)}(x|Q) = Q$

Zunächst gehen wir davon aus, dass es sich um additives Rauschen handelt. Es verbleibt dann Q als einziger freier Parameter, bezüglich dessen der Kullback-Abstand minimiert werden muss. Abb 5.4 zeigt den Abstand als Funktion von Q in der Gegend des Minimums, das an der Stelle $Q = 0.1$ liegt. Der Wert, der zur Simulation verwendet wurde, wird also reproduziert. Außerdem fällt die charakteristische Struktur des steilen Anstiegs für kleinere Q auf. Der Grund für diese Form der Kurve liegt in der lokalisierteren Verteilung für kleinere Q . Beim Kullback-Abstand wird der Mittelwert über den Logarithmus des Verhältnisses von zwei Verteilungen berechnet. Wenn die Verteilung, die den Nenner bildet, sehr klein wird – was bei Lokalisierung um die Potentialminima an den Maxima des Potentials geschehen muss – liefern diese Stellen sehr hohe Beiträge. Für große Q ist die Steigung moderater.

Wir können für diesen Parameter aus der stationären Verteilung nun den Drift gewinnen. In Abb. 5.5 ist der auf diese Weise extrahierte Drift mit dem zur Simulation verwendeten und dem aus der ersten Schätzung aus den bedingten Momenten gewonnenen Drift verglichen. Auffällig ist zunächst, dass sich der mit Hilfe der Momente berechnete Drift qualitativ von der analytischen Driftfunktion unterscheidet. Mit Hilfe der im Rahmen dieser Arbeit vorgeschlagenen Methode gelingt jedoch die Reproduktion der analytischen Funktion. Lediglich an den Stellen am Rand des Intervalls, zu denen nur wenige Messwerte vorlagen, treten signifikante Unterschiede auf.

Des Weiteren können wir natürlich die von uns ermittelte, konstante Diffusion mit der ersten Schätzung vergleichen, die uns aus den Momenten zur Verfügung gestanden hätte. Die beiden Funktionen sind in Abb. 5.6 skizziert. In der ersten Schätzung sind deutliche Artefakte der Driftfunktion zu sehen, die die Annahme einer konstanten Diffusion eigentlich nicht rechtfertigen können. Wir wissen aus der Entwicklung des zweiten Momentes nach dem Zeitinkrement τ (3.10), dass solche Terme auftreten, und haben bereits diskutiert, warum sie beim hier verwendeten Zeitinkrement nicht mehr zu vernachlässigen sind.

Wir haben in diesem Falle also die dem Prozess zugrundeliegenden Drift- und Diffusionsfunktionen fehlerfrei reproduziert. Die Minimalisierung der Diffusion zeigt ein deutliches Ergebnis. Mit Blick auf die aus den Daten geschätzte Diffusionsfunktion ist aber leider die Annahme konstanter Diffusion, die Ausgangspunkt dieser Berechnungen war, nicht zu rechtfertigen.

Annahme: Diffusion der Form $D^{(2)}(x|Q, a) = Q + ax^2$

Eine genaue Betrachtung von Abb. 5.6 lässt vermuten, dass es sich bei diesem Beispiel um eine Zeitreihe mit multiplikativem Rauschen handelt. Die einfachstmögliche sinnvolle Parametrisierung für multiplikatives Rauschen ist, wie in 4.3 vorgeschlagen, eine Diffusion der Form

$$D^{(2)}(x) = Q + ax^2 \quad . \quad (5.17)$$

Wiederum lässt sich die stationäre Verteilung zur Berechnung des Drifts verwenden. Es verbleiben, im Gegensatz zur vorherigen Parametrisierung, nun zwei freie Parameter. In Abb. 5.7 ist der Kullback-Abstand abhängig von den Parametern Q und a grafisch dargestellt.

Das Kullback-Maß wächst mit dem Parameter a an, das Minimum befindet sich auf der Q -Achse des Graphen. Damit scheint der Term $\sim x^2$ nicht zur Diffusion beizutragen und es sich damit wiederum um additives Rauschen zu handeln. Wir reproduzieren den im vorigen Abschnitt erhaltenen Wert von $Q = 0.1$.

Annahme: Diffusion der Form $D^{(2)}(x|Q, a, b) = Q + a \cos(bx)$

In Abb. 5.6 ist eine klare Struktur der ersten Schätzung der Diffusion zu erkennen. Laut der Vorgehensweise, die ich in Kapitel 4 beschrieben habe, ist nun eine Entwicklung dieses Ausdrucks in eine Familie von Funktionen abhängig von einem Satz α von Parametern durchzuführen. Die aus den bedingten Momenten gewonnene Schätzung wurde deshalb durch die Funktion

$$D^{(2)}(x|Q, a, b) = Q + a \cos(bx) \quad (5.18)$$

genähert. GNUplot hat für diese Parametrisierung die Werte

$$\begin{pmatrix} Q \\ a \\ b \end{pmatrix} = \begin{pmatrix} 0.104 \\ 0.47 \\ 10.3772 \end{pmatrix} \quad (5.19)$$

als beste Approximation berechnet. Der Fit ist in Abb. 5.8 aufgetragen. Der Parameter b wurde nun festgehalten und der Kullback-Abstand der PDF's bezüglich der Parameter Q und a minimiert. Das Ergebnis ist – wiederum grafisch – in Abb. 5.9 zu sehen.

Das Kullback-Maß besitzt wieder ein deutliches Minimum bei $Q = 0.1$ und $a = 0$. Dieses entspricht dem schon in den beiden vorhergehenden Analysen erzielten Ergebnis. Bei der hier gewählten Parametrisierung ist deutlich die komplizierte Struktur der „Kullback-Landschaft“ um das Minimum herum zu erkennen. Für $|a| > Q$ kann der Abstand nicht berechnet werden, da in diesem Falle $D^{(2)}$ mitunter negativ wird. Daraus ergibt sich die Form des in der Abbildung zu erkennenden Kegels.

5.2.2 Numerische Umsetzung

Die numerische Umsetzung gleicht der, die in Abschnitt 5.3.2 für das nächste Beispiel in Verbindung mit dem im Anhang B abgedruckten Programmbeispiel erläutert wird. Für die Integration der Langevin-Gleichungen wurden jedoch 500 interne Schritte angesetzt, weil das intrinsische Zeitinkrement der Daten im Vergleich zur Geschwindigkeit der Dynamik hoch war.

Für die Optimierung wurden im ersten Fall einer, bei den beiden letzten zwei Parameter variiert. In allen Fällen wurde der Ortsraum zunächst in 100 Bins aufgelöst und dann an 20 Stellen untersucht, die äquidistant verteilt lagen. An diesen

Stellen wurde das Kullback-Maß nun für 50 bzw. 30×30 verschiedene Parameter bestimmt. Im nächsten Abschnitt wird gezeigt, wie die CPDF's numerisch berechnet und sinnvoll gespeichert werden können. Für das sogennannte „feine Gitter“ habe ich bei diesen Beispielen 20 Bins verwendet. Auf den Rechnern, die hier im Hause zur Verfügung stehen, betrug die Laufzeit der Programme³:

Parametrisierung	Laufzeit	Berechnungen	Zeit/Ber.
$D^{(2)}(x Q) = Q$	22h	20×50	1.32min
$D^{(2)}(x Q, a) = Q + ax^2$	92h	20×900	0.31min
$D^{(2)}(x Q, a) = Q + a \cos(bx)$	383h	20×900	1.27min

Während der Laufzeit hat sich die Belegung der Rechner und damit die den Programmen zur Verfügung stehende Rechenleistung laufend verändert, weil die Rechner *Ampere* und *Tesla*⁴ auch von anderen Institutsmitgliedern genutzt wurden. Maßgeblich für die stark variierende Dauer der einzelnen Analysen ist aber die numerische Umsetzung der Minimierung, die die Präzision, mit der die einzelnen CPDF's berechnet werden, anpasst. Ist die Diffusion einer Parametrisierung kleiner als die des Prozesses, wie es z.B. in Abb. 5.4 links vom Minimum des Kullback-Abstandes der Fall ist, so steigt die benötigte Rechenzeit auf ein Vielfaches. Die Form der Kurve in diesem Bereich ist jedoch zur Lokalisierung des Minimums nicht erforderlich und hier nur berechnet worden, um die Eindeutigkeit der grafischen Lösung zu demonstrieren. Zur effektiven Anwendung sollte man sich einem Minimum von hohen Diffusionen aus nähern, dann ließe sich – bei der hier gewählten numerischen Umsetzung – Rechenzeit sparen.

5.3 Eine Dimension, multiplikatives Rauschen

Als zweites Beispiel wird eine Zeitreihe mit multiplikativem Rauschterm untersucht. Auch diese Reihe wurde durch numerische Integration einer Langevin-Gleichung gewonnen, dabei wurde

$$D^{(1)}(x) = x - x^3 \quad \text{und} \quad (5.20a)$$

$$D^{(2)}(x) = 1 + x^2 \quad (5.20b)$$

verwendet. Dieser Drift gehört zu einem symmetrischen Doppelmuldenpotential, wie es nach *Pitchfork-Bifurkationen* auftritt. Zur Analyse wurden 10^6 Punkte mit der Samplingrate $\tau = 10^{-2}$ erzeugt. Ein Ausschnitt aus den Daten ist in Abb. 5.10 zu sehen.

Die Analyse von Daten, die multiplikatives, also in der Stärke ortsabhängiges Rauschen enthalten, ist mit vielen einfachen Verfahren, wie z.B. den verbreiteten *Maximum-Likelihood-Verfahren*, nicht ohne weiteres möglich. Mittels der direkten Schätzung aus den bedingten Momenten gelingt sie hingegen umgehend. Das Zeitintervall ist wesentlich kleiner als im vorangegangenen Beispiel, damit ist die Präzision dieser Schätzung besser. Problematisch ist allerdings die quantitative Auswertung dieser Schätzung, weil die geschätzten Werte aufgrund verschiedener in Kapitel 2 diskutierter Fehler für Werte fernab des Ursprungs von schlechter Qualität sind und divergieren. Die Minimierung des Kullback-Abstandes von Verteilungen hingegen ist für die quantitative Analyse der Daten geeignet.

³ „Berechnungen“ bezeichnet die Anzahl der Summanden, die in das Kullback-Maß eingehen und die einzeln berechnet werden müssen. Sie setzt sich zusammen aus der Anzahl der Stellen, an denen der Abstand ausgewertet wird, und der Menge der Parameter, die jeweils zu prüfen sind.

⁴Bei den Rechnern *Ampere* und *Tesla*, auf denen die Rechnungen durchgeführt wurden, handelt es sich um Rechenserver des Typs *COMPAQ AlphaServer DS20E*, die jeweils mit zwei Prozessoren mit einer Taktfrequenz von 833 MHz bestückt sind und allen Institutsmitgliedern zur Verfügung stehen.

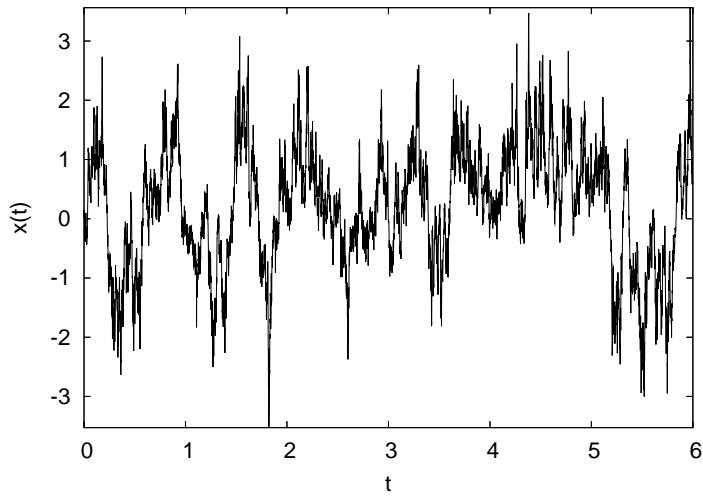
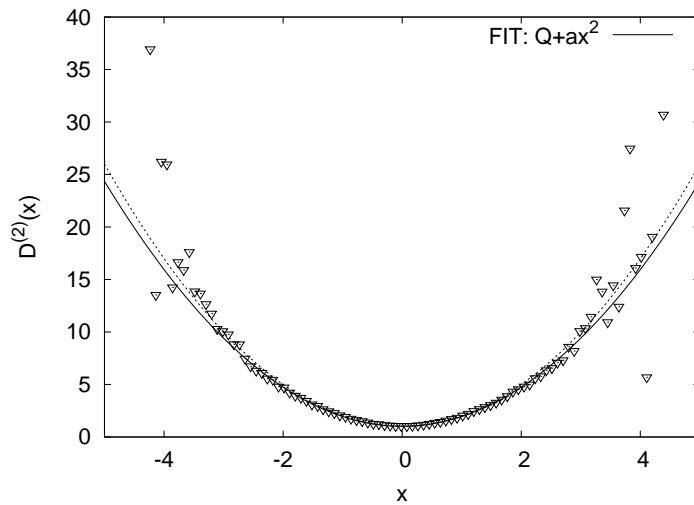


Abbildung 5.10: Ausschnitt aus Zeitreihe II, erzeugt mit multiplikativem Rauschen

Abbildung 5.11: Zeitreihe II: Aus der Auswertung der bedingten Momente gewonnene Abschätzung (Dreiecke) für die Diffusionfunktion. Der Fit $D^{(2)}(x) = 1.01 + 0.93x^2$ (durchgezogene Linie), der von *GNUpot* berechnet wird, weicht von der zur Synthesierung verwendeten Funktion (gestrichelte Linie) ab.

5.3.1 Analyse der Zeitreihe

Aus den Daten können für das kleinstmögliche Inkrement τ zunächst erste Abschätzungen für Drift und Diffusion gewonnen werden. Auch in diesem Falle genügt die Parametrisierung der Diffusionsfunktion. In Abb. 5.11 sind die aus dem zweiten bedingten Moment gewonnene Abschätzung und die ursprünglich zur Simulation der Daten verwendete Diffusion skizziert. Es ist zu erkennen, dass die Abschätzung qualitativ das richtige Verhalten zeigt und sehr nah an der analytischen Funktion liegt.

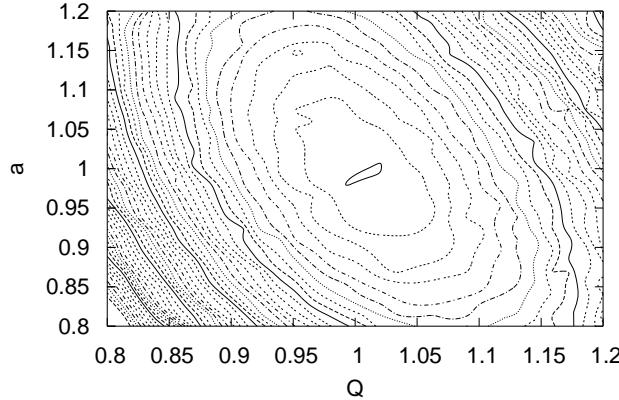


Abbildung 5.12: Zeitreihe II: Kullback-Minimierung der Diffusion $D^{(2)}(x|q, a) = Q + ax^2$ bezüglich der Parameter Q und a . Der Abstand wird minimal für $Q = a = 1$.

Zur quantitativen Bestimmung der Diffusion wird diese parametrisiert als:

$$D^{(2)}(x|Q, a) = Q + ax^2 \quad . \quad (5.21)$$

Die beiden Parameter Q und a müssen mit Hilfe des Kullback-Abstandes optimiert werden. Das Ergebnis dieser Optimierung ist in Abb. 5.12 zu sehen. Wir finden ein deutliches Minimum bei

$$\begin{pmatrix} Q \\ a \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \end{pmatrix} \quad (5.22)$$

und reproduzieren damit die zur Simulation verwendeten Parameter. Das hier vorgestellte Verfahren eignet sich also zur quantitativen Analyse von Daten, die multiplikatives Rauschen enthalten.

5.3.2 Numerische Umsetzung

Das Programm, das zur Analyse dieser Zeitreihe verwendet wurde, ist in Anhang B abgedruckt. In diesem Abschnitt werden die Struktur des Programmes und im Rahmen dessen die verwendeten Datenstrukturen sowie die numerische Umsetzung des Minimalisierungsproblems erläutert. Dazu werde ich mich auf die im Anhang durchgeführte Numerierung der Zeilen im Quelltext beziehen.

Zunächst werden die maximalen Dimensionen der Felder festgelegt. Die Größe dieser Felder, die tatsächlich verwendet wird, wird später als Parameter von der Kommandozeile abgefragt. Die PDF's werden jeweils als Integer- und als Real-Felder angelegt, die Integer-Felder mit der Endung *Int* gekennzeichnet. Nun werden ein sogenanntes „grobes“ und ein „feines“ Raster verwendet: Mit Hilfe des groben Rasters wird zunächst das gesamte Intervall, für das Daten vorliegen, in einzelne, gleichgroße Bins unterteilt. Dafür werden Minimum *aGrob* und Maximum *bGrob* gespeichert und dann das Inkrement *DeltaGrob* als Quotient aus der Breite des Intervalls und der Anzahl der Bins ermittelt. Dieses Gitter spielt beim Extrahieren der stationären Verteilung eine wichtige Rolle.

Die Einführung des feinen Gitters erhöht die Effizienz und Präzision des Programms enorm: es nimmt später die bedingten Verteilungen $p(x|x_0)$ auf und wird für jedes x_0 neu aus den Daten extrahiert. Dabei werden jeweils die Variablen *aFein*,

bFein und *DeltaFein* angepasst. Aus diesem Grunde müssen keine Felder höherer Dimension zur Speicherung der CPDF's angelegt werden. Dieses Vorgehen minimiert den Bedarf des Programms an Speicher und ermöglicht die Erweiterung auf höherdimensionale Einbettungen.

Der aktuelle Wert für Q wird in der Variablen *D1*, der für a in *D2* abgespeichert. Da das Problem grafisch gelöst werden soll, werden die Variablen *D*min* und *D*max* angelegt. Diese beiden Intervalle werden später in *DDim* Stellen unterteilt. Da später zunächst für jede Stelle das Kullback-Maß für alle Werte von Q und a berechnet werden soll, müssen die Ergebnisse für die einzelnen Kombinationen im Speicher hinterlegt werden. Dieses geschieht in *KullmackMat*, einer zweidimensionalen Matrix. In der ebenfalls zweidimensionalen Matrix *KullValueOK*, die bool'schen Typs ist, wird vermerkt, ob an einzelnen Stellen bei der Berechnung des Kullback-Abstandes der Verteilungen Fehler aufgetreten sind. Für diese Parameter wird die Berechnung nicht fortgeführt.

Nun werden die zur Ausführung benötigten Parameter von der Kommandozeile oder einer Eingabedatei abgefragt (Zeilen 34-54). Für die in Abb. 5.12 präsentierte Analyse wurden folgende Parameter verwendet: *DimGrob*=100, *DimFein*=20, *GridSkip*=5, *MinIterate*=50000, *D1Min*=0.8, *D1Max*=1.2, *D2Min*=0.8, *D2Max*=1.2, *DDim*=30, *CutOff*=0, *MaxIterate*=10000000, *DerivOrder*=2. Eine Abfrage der Parameter von der Kommandozeile macht die automatisierte Ausführung des Programms mit verschiedenen Sätzen von Parametern möglich.

Im nächsten Abschnitt werden die Intervallgrenzen für die Diskretisierung des Datenbereichs aus den Daten ermittelt (58-105). Des Weiteren wird die zur Berechnung des Drifts aus der parametrisierten Diffusion benötigte stationäre PDF berechnet und in eine Datei geschrieben.

Nun sollen der Reihe nach die einzelnen Stellen *xpos* untersucht werden. Dazu werden zunächst die bedingten Verteilungen für $x_0 \in [xpos - DeltaGrob/2; xpos + DeltaGrob/2]$ aus den Daten bestimmt (130-172). Dann wird für diese Stelle das Kullback-Maß für alle (Q, a) , die untersucht werden sollen, berechnet:

Zunächst wird, mit Hilfe des in 5.1.3 eingeführten Verfahrens, der Drift aus der Ableitung des Logarithmus der stationären Verteilung berechnet (186-209). Dann werden die Übergangswahrscheinlichkeiten simuliert. Dazu wird mit Hilfe des Zufallszahlengenerators ein Startwert *xPosFein* aus dem Intervall $[xpos - DeltaGrob/2; xpos + DeltaGrob/2]$ gewählt. Für diesen wird dann die Langevin-Gleichung über die Zeit τ integriert und das Ergebnis in die Verteilung *IntPdfInt* eingetragen, falls es zwischen den für diese Verteilung aus den Daten extrahierten Grenzen liegt. Dieses Procedere (219-256) wird nun wiederholt, bis

- es keine Stellen mehr gibt, an denen die im Gegensatz zur externen die simulierte Verteilung verschwindet und mindestens *MinIterate* Iterationen durchgeführt worden sind oder
- die Anzahl der Iterationen *MaxIterate* überschreitet.

Im letzten Fall wird das Ergebnis in *KullbValueOK* als fehlerhaft vermerkt. Nun kann aus den beiden „feinen“ Verteilungen und der stationären Wahrscheinlichkeitsdichte das Kullback-Maß für diese Stelle und den jeweiligen Satz von Parametern berechnet werden (260-271).

Auf diese Weise werden eine Anzahl Stellen, die gleichmäßig über den Datenbereich verteilt sind, ausgewertet und die einzelnen Beiträge zum Kullback-Maß addiert. An den Stellen, für die bei vorgegebener Anzahl maximaler Iterationen der Kullback-Abstand der Verteilungen nicht ermittelt werden konnte, wird später bei der Ausgabe der Daten in eine Datei das Kullback-Maß auf den maximalen korrekt berechneten Abstand gesetzt, um die spätere Auswertung der Daten durch Programme wie *GNUploat* zu erleichtern.

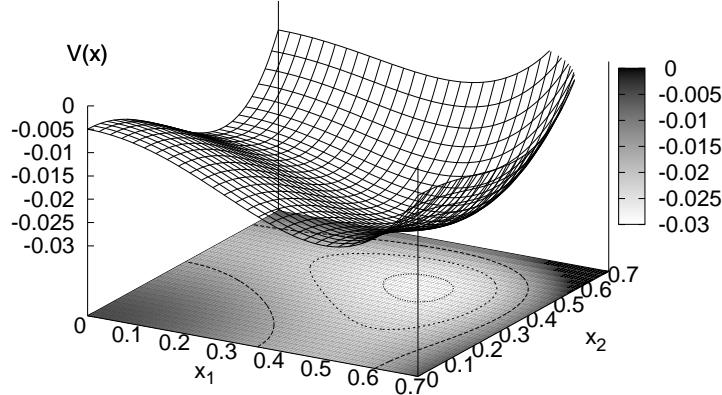


Abbildung 5.13: Potential (5.23) für $\varepsilon = 0.25$, $B = 0.25$.

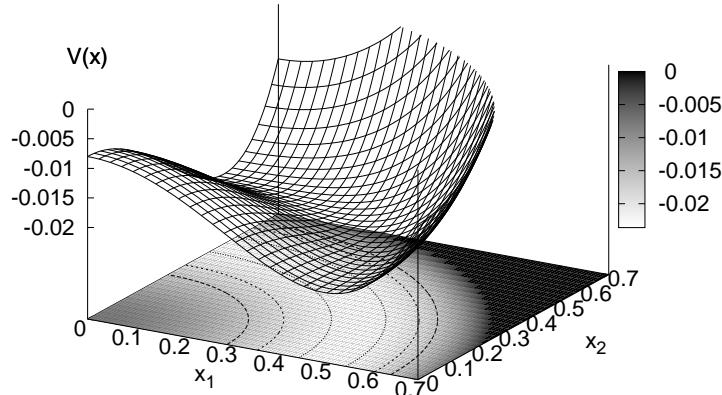
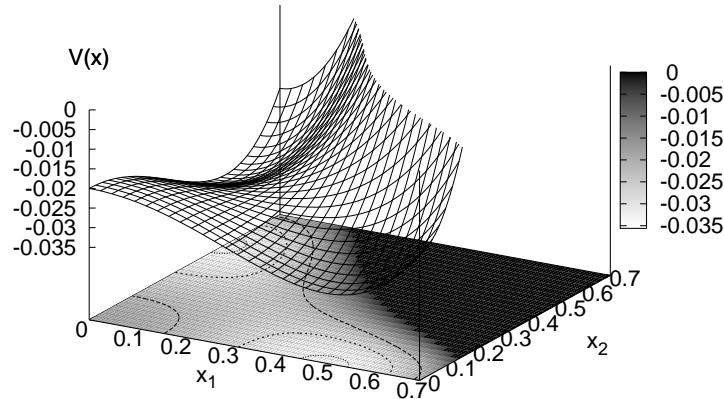
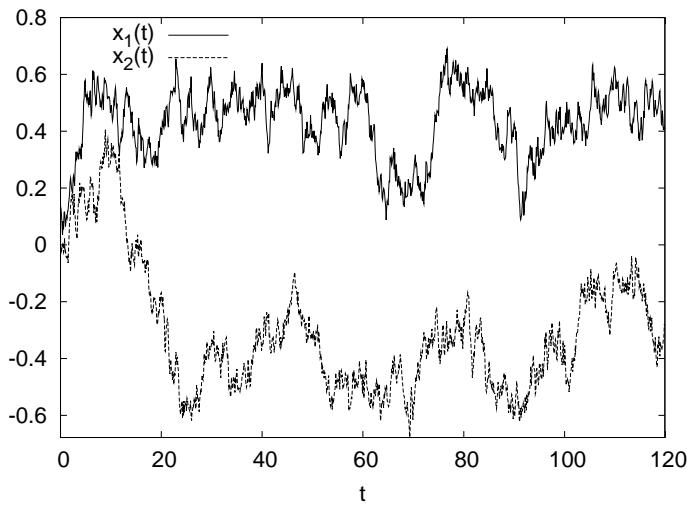


Abbildung 5.14: Potential (5.23) für $\varepsilon = 0.25$, $B = 1$.

Vorteil des hier verwendeten Verfahrens ist, dass sich bei akzeptabler Rechenzeit der Kullback-Abstand auch für Diffusionen, die unterhalb der Optimalen liegen, berechnen lassen. Die geringere Größe der Stichprobe für größere Diffusionen hat, wenn *MinIterate* nicht zu gering gewählt wird, keinen Einfluss auf das erzielte Resultat. Das Verfahren ist damit für die Erzeugung von Daten für die grafische Auswertung geeignet. Die Rechenzeit für dieses Beispiel betrug etwa 16h.

5.4 Zweidimensionales System mit additivem Rauschen

Mit Hilfe des hier beschriebenen Verfahrens und einer leicht modifizierten numerischen Implementierung wurde ferner ein zweidimensionales Problem untersucht,

Abbildung 5.15: Potential (5.23) für $\varepsilon = 0.25$, $B = 4$.Abbildung 5.16: Ausschnitt Zeitreihe IIIa, erzeugt mit $B = 0.25$.

das in die Klasse der sogenannten Potential-Systeme fällt. Wir haben in Kapitel 4 gesehen, dass sich das Vorgehen auch für diese Systeme erheblich vereinfachen lässt und für die Rekonstruktion die Stärke der Diffusion als einziger freier Parameter verbleibt. Beschäftigen wir uns mit dem stochastischen System, das von dem Drift-Potential

$$V(\mathbf{x}) = -\frac{\varepsilon}{2} \mathbf{x}^2 + \frac{1}{4} \mathbf{x}^4 + \frac{B-1}{2} x_1^2 x_2^2 \quad (5.23)$$

und der Diffusion

$$D_{ij}^{(2)}(\mathbf{x}) = Q \delta_{ij} \quad (5.24)$$

für $\mathbf{x} \in \mathbb{R}^2$ gebildet wird. Es handelt sich um ein System mit additivem Rauschen, das seine Wurzeln in der Theorie über der Selbstorganisation in komplexen Systemen [8] hat. Die Eigenschaften des Potentials und damit der Zustand des Systems hängen von den beiden Parametern ε und B ab.

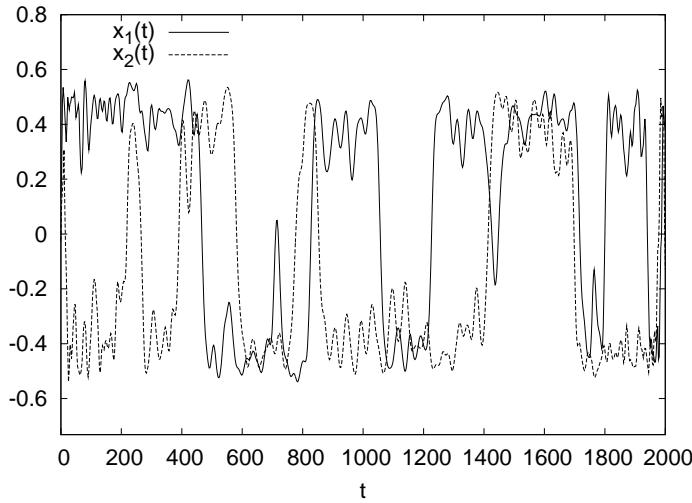


Abbildung 5.17: Geglätteter Langzeit-Ausschnitt aus Zeitreihe IIIa. Es ist zu erkennen, dass es sich hier idealisiert um zwei gekoppelte Zweiniveau-Systeme handelt.

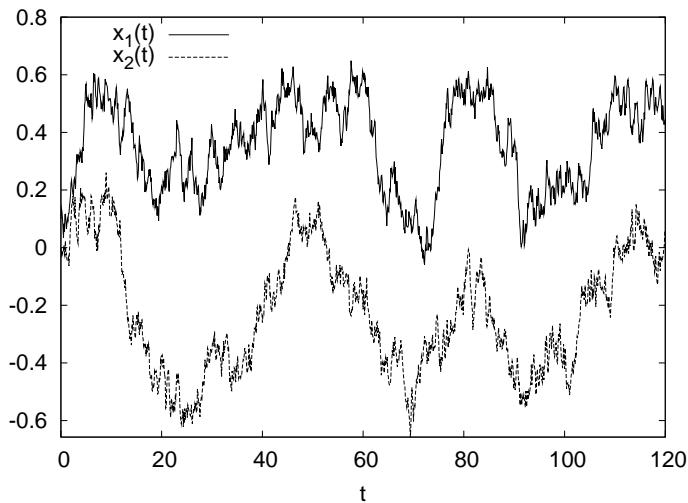
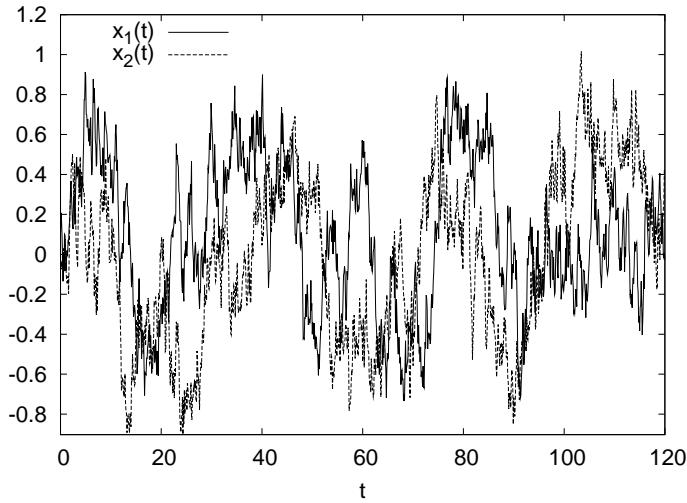
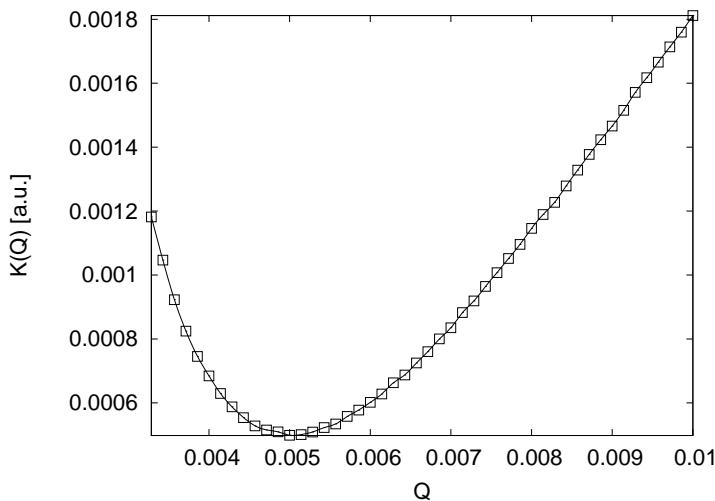


Abbildung 5.18: Ausschnitt Zeitreihe IIIb, erzeugt mit $B = 0.96$.

5.4.1 Diskussion des Potentials

Mit Hilfe des Potentials, das ich in diesem Beispiel zur Erzeugung meiner Daten verwende, lässt sich *Selektion* in komplexen Systemen verstehen: Sei ein System fernab vom thermodynamischen Gleichgewicht, das Strukturbildung zeigen kann, in mehreren Moden angeregt. Nach der Theorie von Haken koexistieren diese Moden zunächst [8, 7]. Es gibt aber einen kritischen Energiefloss, oberhalb dessen sich selbstorganisierte Strukturen bilden können. Der Übergang von einem ungeordneten zu einem selbstorganisiert geordneten System lässt sich in der Gegend des Phasenübergangs durch das Verhalten von Ordnungsparametern beschreiben.

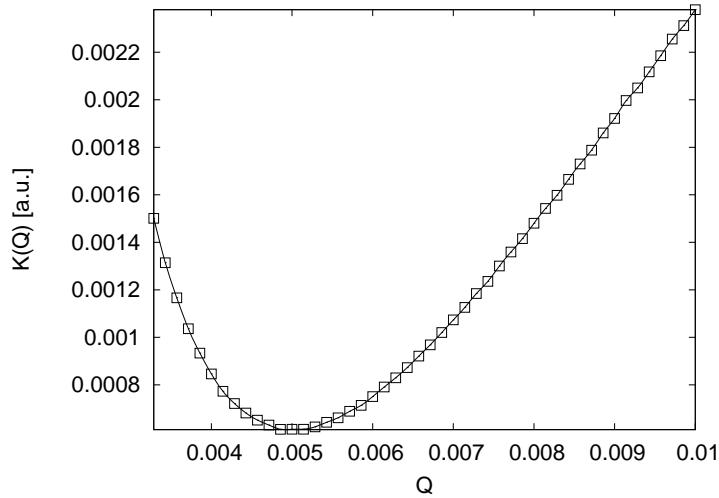
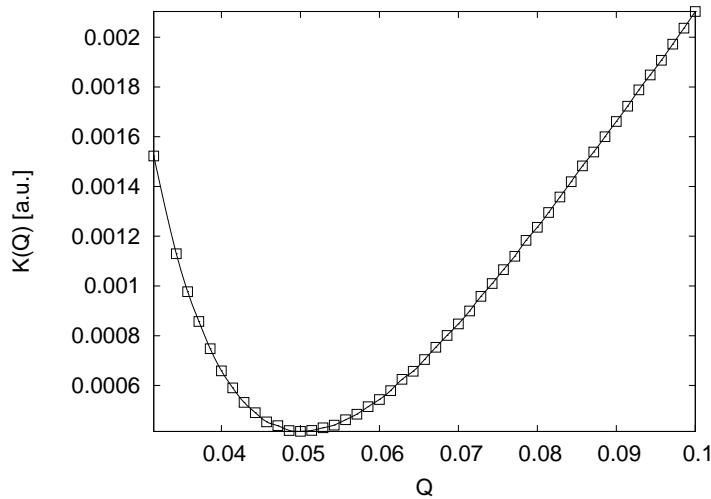
Betrachten wir nun die Dynamik zweier Ordnungparameter, die verschiedenen Moden des Systems entsprechen können, zunächst als überdämpfte Bewegung ohne

Abbildung 5.19: Ausschnitt Zeitreihe IIIc, erzeugt mit $B = 2$.Abbildung 5.20: Zeitreihe IIIa: Optimierung der Diffusion Q .

Diffusion im Potential (5.23). Mit Hilfe des Parameters B , der maßgeblich ist für die Struktur des Potentials, kann der Phasenübergang bei der Strukturbildung nun beschrieben werden. Untersuchen wird das Potential für drei unterschiedliche Werte von B :

$B = 0.25$: Dieser Fall ist in Abb. 5.13 dargestellt. Die Minima des Potentials in den vier Quadranten liegen jeweils auf der Winkelhalbierenden. Das System wird, abhängig von der Anfangsbedingung, in eines dieser Minima getrieben. Die Ordnungsparameter koexistieren in diesem stabilen Fixpunkt des Systems, es herrscht dynamisches Gleichgewicht. Die Ursache für die Koexistenz liegt in der Wechselwirkung der Moden, wie sie aus der klassischen Thermodynamik bekannt ist.

$B = 1.00$: Mit wachsendem B werden die Minima flacher, bis sich bei $B = 1$ ei-

Abbildung 5.21: Zeitreihe IIIb: Optimierung der Diffusion Q .Abbildung 5.22: Zeitreihe IIIc: Optimierung der Diffusion Q .

ne stabile, eindimensionale Mannigfaltigkeit bildet (Abb. 5.14). Dieser Punkt markiert den kritischen Punkt des Systems. Würde eine schwache Diffusion das System treiben, so könnte das System auf dem Boden des sogenannten „Weinflaschen-Potentials“ wandern und sich jeder Zustand einstellen.

$B = 4.00$: Für weiter steigende B setzt nun Selektion ein, es entstehen vier neue Minima auf den beiden Achsen (Abb. 5.15). Abhängig von der Anfangsbedingung wird das System nun in ein Minimum des Potentials getrieben, in dem ein Ordnungsparameter verschwindet. Der verbleibende Ordnungsparameter bestimmt die Struktur des Systems.

Mit Hilfe dieses Formalismus lassen sich viele Beispiele aus dem Bereich der Synergetik, in denen selbstorganisierte Strukturbildung auftritt, verstehen.

Zum Test der Kullback-Minimierung in zwei Dimensionen wir das Potential (5.23)

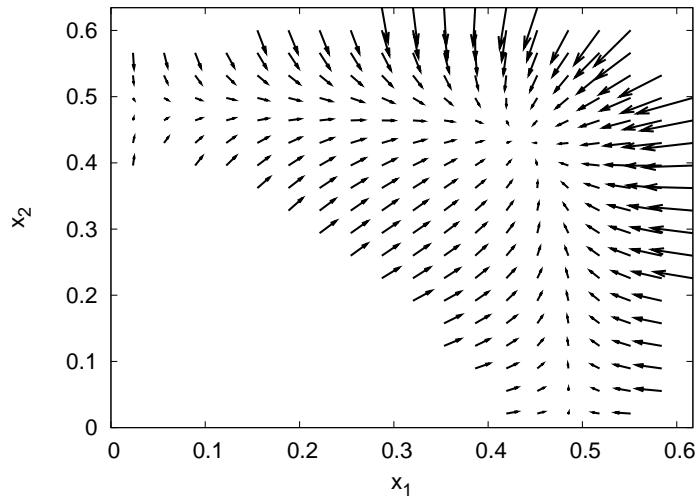


Abbildung 5.23: Zeitreihe IIIa: Ausschnitt aus dem Drift, extrahiert aus der stationären Verteilung für $Q = 0.005$. Die Fixpunkte der Dynamik sind deutlich zu erkennen.

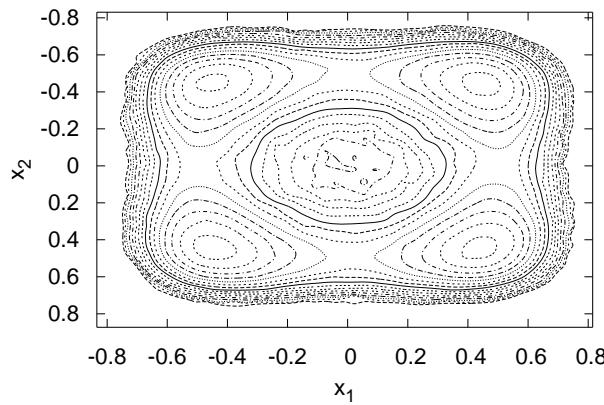


Abbildung 5.24: Zeitreihe IIIa: Zum in Abb. 5.23 skizzierten Drift gehörendes Drift-Potential, ebenfalls aus den Daten gewonnen. Die Höhenlinien liegen ausgehend von -0.04 äquidistant im Abstand von 0.002 .

mit $\varepsilon = 0.25$ für verschiedene Werte des Parameters B untersucht. Die Diffusion wird jeweils so eingestellt, dass die stationäre Verteilung hinreichend schnell unabhängig von der Anfangsbedingung wird. Sie ist also in jedem Fall stark genug, um das System aus den Minima des Drift-Potentials zu treiben.

Für folgende Parameter wurden zur späteren Analyse Daten simuliert:

Reihe	B	Q	Abbildung
IIIa	0.25	0.005	5.16, 5.17
IIIb	0.96	0.005	5.18
IIIc	2.00	0.050	5.19

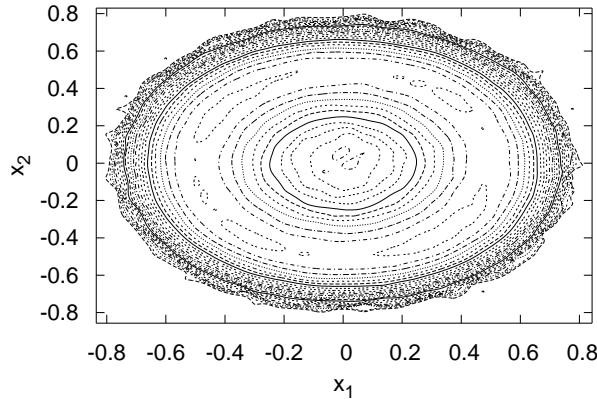


Abbildung 5.25: Zeitreihe IIIb: Drift-Potential, gewonnen aus der stationären Verteilung für $Q = 0.005$. Die Höhenlinien starten bei -0.0374 und haben den Abstand 0.0015 . Es sind kleine Abweichungen vom rotationssymmetrischen Zustand mit $B = 1$ zu erkennen.

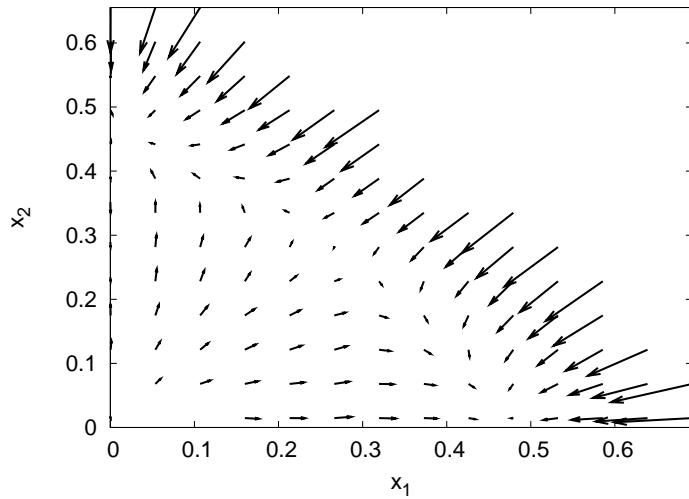


Abbildung 5.26: Zeitreihe IIIc: Ausschnitt aus dem Drift, extrahiert aus der stationären Verteilung für $Q = 0.05$. Die stabilen Fixpunkte der Dynamik liegen nun auf den beiden Achsen.

Dabei wurden jeweils Zeitreihen mit $5 \cdot 10^6$ Datenpunkten und einem Zeitinkrement von $\tau = 0.1$ erzeugt, die interne Auflösung bei der Simulation der Daten lag um einen Faktor 100 höher.

5.4.2 Numerische Umsetzung

Aufgrund der resourcenschonenden Struktur und Speichernutzung des im letzten Beispiel vorgestellten Programms ist eine Erweiterung auf zwei Dimensionen ohne

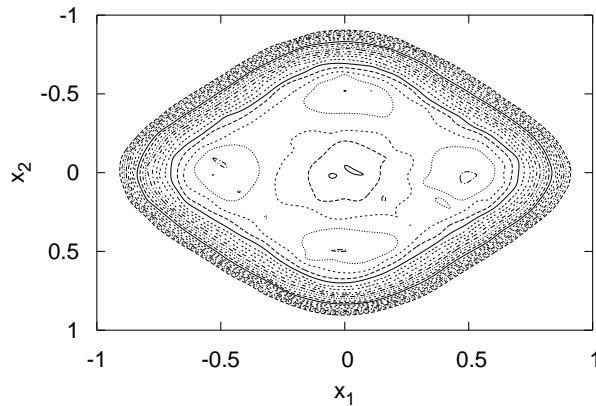


Abbildung 5.27: Zeitreihe IIIc: Zum in Abb. 5.26 skizzierten Drift gehörendes Drift-Potential, ebenfalls aus den Daten gewonnen. Die Höhenlinien liegen ausgehend von -0.385 äquidistant im Abstand von 0.004 .

Weiteres möglich. Die Diffusion wurde dann als

$$D_{ij}^{(2)}(x|Q) = Q\delta_{ij} \quad (5.25)$$

parametrisiert und das Kullback-Maß als Funktion von Q jeweils berechnet. Mit Hilfe von (4.15) können für das optimale Q Driftpotential und Drift rekonstruiert werden.

5.4.3 Analyse der Zeitreihe

Die berechneten Kullback-Abstände als Funktion der Diffusionsstärke Q für die drei verschiedenen Datensätze sind in den Abbildungen 5.20 bis 5.22 aufgetragen. Allen drei Graphen ist das Minimum eindeutig zu entnehmen, es werden jeweils die zur Synthesierung der Daten verwendeten Diffusionen reproduziert. Die Rechenzeit pro Analyse betrug für diese Qualität der Kurven ca. 20h. Das Verfahren konvergiert allerdings schon nach sehr kurzer Zeit und liefert schon zu Beginn ein quantitativ verwertbares Ergebnis.

Um den Nutzen der hier angewendeten Methode zu verdeutlichen, sind aus den Daten mit Hilfe der stationären Verteilung für das optimale Q der Drift bzw. das Potential des Drifts berechnet worden, diese lassen sich grafisch auswerten. Diese Grafiken lassen eine Analyse des deterministischen Anteils der Dynamik des Systems zu. Auf die drei Fälle gehe ich deswegen hier kurz ein:

- **Zeitreihe IIIa:** In Abb. 5.23 ist der Drift, dem die Dynamik unterliegt, für einen Quadranten skizziert. Die beiden instabilen Fixpunkte auf den Achsen sowie der stabile auf der Winkelhalbierenden sind deutlich zu erkennen. Im Potential des Driftes (Abb. 5.24) lässt sich ferner der instabile Fixpunkt im Ursprung identifizieren, für große $|x|$ wird ein stark repulsiver Drift deutlich. Die vier stabilen Punkte sind auch in der geglätteten Langzeitdarstellung der Zeitreihe in Abb. 5.17 zu sehen. Mit Hilfe des Drift-Potentials lässt sich nun auch die Kopplung der beiden Variablen verstehen.

- **Zeitreihe IIIb:** Das System befindet sich hier nahe dem sogenannten „Weinflaschen-Zustand“, der rotationssymmetrisch ist. Der einzige deutliche Fixpunkt ist instabil und liegt im Ursprung. Bei genauer Betrachtung des in Abb. 5.25 dargestellten Potentials lässt sich aber die leichte Asymmetrie im System erkennen, die sich in vier schwachen, attraktiven Fixpunkten auf den Winkelhalbierenden äußert. Es ist deshalb eindeutig, dass das hier untersuchte System zu einem $B < 1$ (verwendet wurde: $B = 0.96$) gehört und sich damit noch unterhalb des kritischen Punktes befindet.
- **Zeitreihe IIIc:** Bei dieser Konfiguration des Systems sind die Umstände denen der Daten IIIa entgegengesetzt: Deterministisch wird das System zu den stabilen Punkten auf den Achsen hin getrieben. Drift (Abb. 5.26) und Potential (Abb. 5.27) bestätigen die in Abschnitt 5.4.1 motivierten Erwartungen.

Wir sehen also, dass in den drei hier untersuchten Fällen die Minimierung des Kullback-Abstandes eindeutig durchzuführen ist und eine detaillierte Analyse der Dynamik des Systems ermöglicht.

Kapitel 6

Zusammenfassung

Es ist deutlich, dass die Analyse von Daten stochastischer Prozesse in vielen Bereichen der Natur-, aber auch anderer Wissenschaften eine wichtige Rolle spielt, weil stochastische Differentialgleichungen die zeitliche Entwicklung von Ordnungsparametern komplexer Systeme beschreiben. Die Grundlage dieser Arbeit bildet ein Verfahren zur Schätzung von Drift- und Diffusionsfunktionen aus bedingten Momenten, die im Grenzübergang kleiner Zeitinkremente auszuwerten sind. Stellvertretend für die Durchführung dieses Verfahrens sind drei Beispiele aus verschiedenen Bereichen herausgegriffen.

Im Rahmen dieser Arbeit sind zunächst die Fehler in Drift und Diffusion, die sich bei dieser Methode ergeben, zusammengestellt und im einzelnen untersucht. Neben einer Diskretisierung der Zeit muss bei einer endlichen Menge an Daten auch der Ortsraum in sogenannte Bins unterteilt werden. Es stellt sich heraus, dass beide Diskretisierungen kritisch sind: Sowohl bei zu klein als auch bei zu groß gewählten Inkrementen treten in den Schätzungen für Drift- und Diffusionsfunktionen Abweichungen auf. Es muss nach einem Satz von Inkrementen gesucht werden, für den diese so klein werden wie möglich.

Es ist bekannt, dass die Anwendung dieser Methode ferner auf Daten mit nicht verschwindendem Messrauschen nicht möglich ist. In Zusammenarbeit mit Frank Böttcher ist der Einfluss dieses Rauschens auf die bedingten Momente untersucht worden. Es ergibt sich ein modifiziertes Verfahren, mit Hilfe dessen sich einfache stochastische Prozesse selbst bei starkem Messrauschen charakterisieren lassen.

Zentral soll aber Kapitel 4 der vorliegenden Arbeit sein: Hier wird ein allgemeines Verfahren zur quantitativen Gewinnung von Drift- und Diffusionsfunktionen vorgestellt, das auf der Bewertung von Zweipunkt-Verbundwahrscheinlichkeiten beruht und auf einer Parametrisierung von Drift und Diffusion auf Basis der Schätzung für endliche Samplingraten aufbaut. Die Optimierung der Parameter lässt sich durch Minimierung des Kullback-Abstandes zwischen der aus den Daten gewonnenen und der durch die Parameter festgelegten Verteilungen iterativ durchführen. Dieses Verfahren leidet nicht unter den Problemen, die für die Schätzungen aus den Momenten offensichtlich geworden sind. Insbesondere kann es für endliche Samplingfrequenzen durchgeführt werden.

Die numerische Umsetzung der Kullback-Minimierung ist anhand dreier Beispiele erläutert. Dabei werden zwei eindimensionale Systeme mit additivem und multiplikativem Rauschen sowie ein zweidimensionales System, das in die Klasse der sogenannten Potential-Systeme fällt, diskutiert. In allen Fällen gelingt die Rekonstruktion der Dynamik der Systeme.

Anhang A

Ausgegliederte Berechnungen

A.1 Berechnung von Integralen wie (3.10c) durch partielle Integration

An dieser Stelle wird erläutert, wie die partielle Integration, die für die Entwicklung der Momente für endliches Zeitinkrement mehrfach benötigt wird, durchzuführen ist. Um das Prinzip der Integration von Produkten, in denen Ableitungen von δ -Funktionen auftauchen, prinzipiell zu verdeutlichen, wird nun die Berechnung des Summanden (3.10c) der Entwicklung des 2. Momentes,

$$\frac{\tau^2}{2} \int dx (x - x_0)^2 \hat{L}^2 \delta(x - x_0) ,$$

durchgeführt.

Bei der Integration müssen die distributionellen Eigenschaften von $\delta(x)$ beachtet werden. Eine Funktion, die für alle stetigen Funktionen f

$$\int dx f(x) \delta(x) = f(0) \quad (A.1)$$

erfüllt, heißt δ -Funktion. Es folgt sofort:

$$\int dx f(x) \delta(x - x_0) = f(x_0) . \quad (A.2)$$

$\delta(x)$ ist keine Funktion im eigentlichen Sinne, sondern eine *Distribution*. Eine mögliche Darstellung ist der Grenzwert einer beliebig schmalen, auf 1 normierten Gaußverteilung:

$$\delta(x) = \lim_{\alpha \rightarrow \infty} \sqrt{\frac{\alpha}{\pi}} \exp[-\alpha x^2] \quad (A.3)$$

$\delta(x)$ divergiert an der Stelle 0 und ist nur im Sinne von (A.1) definiert. Für die Berechnung von Integralen, die Ableitungen von $\delta(x)$ enthalten, lässt sich jedoch ausnutzen, dass $\frac{d^n}{dx^n} \delta(x)$ für $x \neq 0$ für beliebige n verschwindet. Deswegen lassen sich Ableitungen mit Hilfe partieller Integration auf die anderen Funktionen im Integranden übertragen:

$$\int_{-\infty}^{\infty} dx f(x) \frac{d}{dx} \delta(x) = f(x) \delta(x) \Big|_{x=-\infty}^{x=\infty} - \int_{-\infty}^{\infty} dx \left[\frac{d}{dx} f(x) \right] \delta(x) \quad (A.4a)$$

$$= - \int_{-\infty}^{\infty} dx \left[\frac{d}{dx} f(x) \right] \delta(x) \quad (A.4b)$$

Der erste Summand in (A.4a) fällt heraus, weil δ schneller verschwindet als f . Dieses Verhalten wird nun mehrfach ausgenutzt.

Gehen wir nun in der Reihenfolge vor, in der die Summanden von \hat{L}^2 in (3.5a) auftreten, und berechnen die vier Integrale der Reihe nach. An der mit \star gekennzeichneten Stelle wird zum Ende der Rechnungen hin jeweis die in Kapitel 3 abkürzend eingeführte Notation eingesetzt. Beginnen wir mit dem Integral über den ersten Summanden des Differentialoperators:

$$\begin{aligned}
& \frac{\tau^2}{2} \int dx (x - x_0)^2 \frac{\partial}{\partial x} D^{(1)}(x, t) \frac{\partial}{\partial x} D^{(1)}(x, t) \delta(x - x_0) \\
&= \frac{\tau^2}{2} (x - x_0)^2 D^{(1)}(x, t) \frac{\partial}{\partial x} D^{(1)}(x, t) \delta(x - x_0) \Big|_{x=-\infty}^{x=\infty} \\
&\quad - \frac{\tau^2}{2} \int dx D^{(1)}(x, t) \left[\frac{\partial}{\partial x} (x - x_0)^2 \right] \frac{\partial}{\partial x} D^{(1)}(x, t) \delta(x - x_0) \\
&= 0 - \frac{\tau^2}{2} \int dx 2D^{(1)}(x, t) (x - x_0) \frac{\partial}{\partial x} D^{(1)}(x, t) \delta(x - x_0) \\
&= -\tau^2 D^{(1)}(x, t) (x - x_0) D^{(1)}(x, t) \delta(x - x_0) \Big|_{x=-\infty}^{x=\infty} \\
&\quad + \tau^2 \int dx D^{(1)}(x, t) \left[\frac{\partial}{\partial x} D^{(1)}(x, t) (x - x_0) \right] \delta(x - x_0) \\
&= -0 + \tau^2 \int dx D^{(1)}(x, t) \left[\frac{\partial}{\partial x} D^{(1)}(x, t) \right] (x - x_0) \delta(x - x_0) \\
&\quad + \tau^2 \int dx D^{(1)}(x, t) D^{(1)}(x, t) \left[\frac{\partial}{\partial x} (x - x_0) \right] \delta(x - x_0)
\end{aligned}$$

In den Integralen, die nun verblieben sind, tauchen keine Ableitungen der δ -Funktion mehr auf, sie können deshalb ausgewertet werden. Der erste Summand verschwindet, weil der Integrand nach Auswertung der δ -Funktion für die Stelle $x = x_0$ betrachtet werden muss. Es ergibt sich also schließlich:

$$\dots = \tau^2 D^{(1)}(x_0, t) D^{(1)}(x_0, t) \stackrel{\star}{=} \tau^2 D^{(1)} D^{(1)} \quad (\text{A.5})$$

Ebenso gehe ich bei der Berechnung des zweiten Integrals vor:

$$\begin{aligned}
& -\frac{\tau^2}{2} \int dx (x - x_0)^2 \frac{\partial}{\partial x} D^{(1)}(x, t) \frac{\partial^2}{\partial x^2} D^{(2)}(x, t) \delta(x - x_0) \\
&= -\frac{\tau^2}{2} (x - x_0)^2 D^{(1)}(x, t) \frac{\partial^2}{\partial x^2} D^{(2)}(x, t) \delta(x - x_0) \Big|_{x=-\infty}^{x=\infty} \\
&\quad + \frac{\tau^2}{2} \int dx D^{(1)}(x, t) \left[\frac{\partial}{\partial x} (x - x_0)^2 \right] \frac{\partial^2}{\partial x^2} D^{(2)}(x, t) \delta(x - x_0) \\
&= -0 + \tau^2 \int dx D^{(1)}(x, t) (x - x_0) \frac{\partial^2}{\partial x^2} D^{(2)}(x, t) \delta(x - x_0) \\
&= \tau^2 D^{(1)}(x, t) (x - x_0) \frac{\partial}{\partial x} D^{(2)}(x, t) \delta(x - x_0) \Big|_{x=-\infty}^{x=\infty} \\
&\quad - \tau^2 \int dx \left[\frac{\partial}{\partial x} D^{(1)}(x, t) (x - x_0) \right] \frac{\partial}{\partial x} D^{(2)}(x, t) \delta(x - x_0) \\
&= 0 - \tau^2 \int dx \left[\frac{\partial}{\partial x} D^{(1)}(x, t) \right] (x - x_0) \frac{\partial}{\partial x} D^{(2)}(x, t) \delta(x - x_0) \\
&\quad - \tau^2 \int dx D^{(1)}(x, t) \left[\frac{\partial}{\partial x} (x - x_0) \right] \frac{\partial}{\partial x} D^{(2)}(x, t) \delta(x - x_0)
\end{aligned}$$

Es muss nun ein drittes Mal partiell integriert werden. Randterme, die bis hierher wie auch in den nächsten Rechnungen jeweils verschwinden, werden von nun an nicht mehr notiert.

$$\begin{aligned}
&= \tau^2 \int dx D^{(2)}(x, t) \left\{ \frac{\partial}{\partial x} \left[\frac{\partial}{\partial x} D^{(1)}(x, t) \right] (x - x_0) \right\} \delta(x - x_0) \\
&\quad + \tau^2 \int dx D^{(2)}(x, t) \left[\frac{\partial}{\partial x} D^{(1)}(x, t) \right] \delta(x - x_0) \\
&= \tau^2 \int dx D^{(2)}(x, t) \left[\frac{\partial^2}{\partial x^2} D^{(1)}(x, t) \right] (x - x_0) \delta(x - x_0) \\
&\quad + \tau^2 \int dx D^{(2)}(x, t) \left[\frac{\partial}{\partial x} D^{(1)}(x, t) \right] \left[\frac{\partial}{\partial x} (x - x_0) \right] \delta(x - x_0) \\
&\quad + \tau^2 \int dx D^{(2)}(x, t) \left[\frac{\partial}{\partial x} D^{(1)}(x, t) \right] \delta(x - x_0) \\
&= 0 + \tau^2 2D^{(2)}(x_0, t) \frac{\partial}{\partial x_0} D^{(1)}(x_0, t) \stackrel{*}{=} 2\tau^2 D^{(2)} D_x^{(1)} \tag{A.6}
\end{aligned}$$

Für den dritten Summanden von (3.5a) ist folgenden Integral auszuwerten:

$$\begin{aligned}
&- \frac{\tau^2}{2} \int dx (x - x_0)^2 \frac{\partial^2}{\partial x^2} D^{(2)}(x, t) \frac{\partial}{\partial x} D^{(1)}(x, t) \delta(x - x_0) \\
&= + \frac{\tau^2}{2} \int dx \left[\frac{\partial}{\partial x} (x - x_0)^2 \right] \frac{\partial}{\partial x} D^{(2)}(x, t) \frac{\partial}{\partial x} D^{(1)}(x, t) \delta(x - x_0) \\
&= - \tau^2 \int dx D^{(2)}(x, t) \left[\frac{\partial}{\partial x} (x - x_0) \right] \frac{\partial}{\partial x} D^{(1)}(x, t) \delta(x - x_0) \\
&= + \tau^2 \int dx D^{(1)}(x, t) \left[\frac{\partial}{\partial x} D^{(2)}(x, t) \right] \delta(x - x_0) \\
&= \tau^2 D^{(1)}(x_0, t) \frac{\partial}{\partial x_0} D^{(2)}(x_0, t) \stackrel{*}{=} \tau^2 D^{(1)} D_x^{(2)} \tag{A.7}
\end{aligned}$$

Die Berechnung des vierten Integrals schließlich unterscheidet sich nur unwesentlich von der des dritten:

$$\begin{aligned}
&\frac{\tau^2}{2} \int dx (x - x_0)^2 \frac{\partial^2}{\partial x^2} D^{(2)}(x, t) \frac{\partial^2}{\partial x^2} D^{(2)}(x, t) \delta(x - x_0) \\
&= - \frac{\tau^2}{2} \int dx \left[\frac{\partial}{\partial x} (x - x_0)^2 \right] \frac{\partial}{\partial x} D^{(2)}(x, t) \frac{\partial^2}{\partial x^2} D^{(2)}(x, t) \delta(x - x_0) \\
&= + \tau^2 \int dx D^{(2)}(x, t) \left[\frac{\partial}{\partial x} (x - x_0) \right] \frac{\partial^2}{\partial x^2} D^{(1)}(x, t) \delta(x - x_0) \\
&= + \tau^2 \int dx D^{(2)}(x, t) \left[\frac{\partial^2}{\partial x^2} D^{(2)}(x, t) \right] \delta(x - x_0) \\
&= \tau^2 D^{(2)}(x_0, t) \frac{\partial^2}{\partial x_0^2} D^{(2)}(x_0, t) \stackrel{*}{=} \tau^2 D^{(2)} D_{x^2}^{(2)} \tag{A.8}
\end{aligned}$$

Die einzelnen Summanden liefern die Terme proportional zu τ^2 des zweiten bedingten Momentes, wie sie in (3.10c) notiert sind. Die Berechnung der übrigen Summanden, die in (3.7a) bis (3.7d) sowie (3.10a) bis (3.10d) auftauchen, verläuft ähnlich.

A.2 Auswirkung der Diskretisierung: Auswertung von (3.22)

Zu berechnen ist das Integral

$$D_E^{(1)}(x_0, \Delta x) \approx \frac{\int_{x_0-\Delta x/2}^{x_0+\Delta x/2} dx D^{(1)}(x) f(x)}{\int_{x_0-\Delta x/2}^{x_0+\Delta x/2} dx f(x)} \quad (\text{A.9})$$

mit den in Abschnitt 3.3 eingeführten Drift- und Diffusionsfunktionen. Es ist nun:

$$(A.9) \approx [\Delta x f(x_0)]^{-1} \sqrt{\frac{\gamma}{2\pi Q}} \int_{x_0-\Delta x/2}^{x_0+\Delta x/2} dx (-\gamma) x e^{-\frac{\gamma}{2Q}(x)^2} \quad (\text{A.10a})$$

$$= - \left[\Delta x e^{-\frac{\gamma}{2Q}(x_0)^2} \right]^{-1} \gamma \int_{x_0-\Delta x/2}^{x_0+\Delta x/2} dx x e^{-\frac{\gamma}{2Q}(x)^2} \quad (\text{A.10b})$$

$$= - \left[\Delta x e^{-\frac{\gamma}{2Q}(x_0)^2} \right]^{-1} \frac{\gamma}{2} \int_{(x_0-\Delta x/2)^2}^{(x_0+\Delta x/2)^2} dy e^{-\frac{\gamma}{2Q}y} \quad (\text{A.10c})$$

$$= - \left[\Delta x e^{-\frac{\gamma}{2Q}(x_0)^2} \right]^{-1} \frac{\gamma}{2} \left[-\frac{2Q}{\gamma} e^{-\frac{\gamma}{2Q}y} \right]_{y=(x_0-\Delta x/2)^2}^{y=(x_0+\Delta x/2)^2} \quad (\text{A.10d})$$

$$= \left[\Delta x e^{-\frac{\gamma}{2Q}(x_0)^2} \right]^{-1} Q \left[e^{-\frac{\gamma}{2Q}(x_0+\frac{\Delta x}{2})^2} - e^{-\frac{\gamma}{2Q}(x_0-\frac{\Delta x}{2})^2} \right] \quad (\text{A.10e})$$

$$= \left[\Delta x e^{-\frac{\gamma}{2Q}(x_0)^2} \right]^{-1} 2Q e^{-\frac{\gamma}{2Q}\{x_0^2 + (\frac{\Delta x}{2})^2\}} \sinh\left(-\frac{\gamma}{Q}x_0 \frac{\Delta x}{2}\right) \quad (\text{A.10f})$$

$$= \frac{2}{\Delta x} Q e^{-\frac{\gamma}{2Q}(\frac{\Delta x}{2})^2} \sinh\left(-\frac{\gamma}{Q}x_0 \frac{\Delta x}{2}\right) \quad (\text{A.10g})$$

Für eine Untersuchung der Konvergenz für kleine Δx ist eine Entwicklung der Funktionen im Zähler nach deren Argument hilfreich. Seien nun $\Delta x \ll 1$ und $x_0 \Delta x \ll 1$. Setzen wir dann

$$e^z \approx 1 + \sum_{k=1}^{k_0} \frac{(z)^k}{k!} + \mathcal{O}(z^{k_0+1}) \quad (\text{A.11a})$$

$$\sinh(z) \approx \sum_{k=0}^{k_0} \frac{(z)^{2k+1}}{(2k+1)!} + \mathcal{O}(z^{2k_0+3}) , \quad (\text{A.11b})$$

so liefert eine Entwicklung unseres Mittelwertes für $D^{(1)}$ nach Potenzen von Δx :

$$D_E^{(1)}(x_0, \Delta x) \approx -\gamma x_0 + \left[\frac{\gamma^2 x_0}{2Q} - \frac{(\gamma x_0)^3}{6Q^2} \right] \left(\frac{\Delta x}{2} \right)^2 + \mathcal{O}(\Delta x^4) \quad (\text{A.12})$$

Anhang B

Kommentierter Programmcode

Folgendes Programm wurde zur Analyse von der in Abb. 5.10 dargestellten Zeitreihe verwendet. Erläuterungen zur Numerik sowie zu den verwendeten Parametern finden sich in Kapitel 5, insbesondere in Abschnitt 5.3.2.

Das Programm wurde mit dem Compaq Fortran Compiler *fort* auf den *Alpha*-Rechnern hier im Institut kompiliert und dort ausgeführt.

Programmcode: Kullback.F90 (FORTAN 90)

```
PROGRAM Kullback
IMPLICIT NONE

5      INTEGER,PARAMETER:: InFile=10, OutFile=52, GridFile=53,MaxDimGrob=150
      INTEGER,PARAMETER:: MaxDimFein=50,DDimMax=50

10     INTEGER:: i ,k ,m ,o ,p ,FileError
      INTEGER:: DerivOrder ,LocalDerivOrder ,DerivN ,i ,k ,m ,o ,p ,FileError
      INTEGER:: DimFein ,DimGrob ,GridSkip ,MinIterate ,DDim ,CutOff
      INTEGER,DIMENSION( 0:MaxDimGrob ):: SPdfINT      ! Integer-Felder
      INTEGER,DIMENSION( 0:MaxDimFein ):: ExtPdfINT ,IntPdfINT
      INTEGER(KIND=8):: ExtNormInt ,IntNormInt ,MaxIterate

15     REAL:: randzahl
      DOUBLE PRECISION:: x ,xAlt ,xPos ,xPosFein
      DOUBLE PRECISION:: tau ,tauSim
      DOUBLE PRECISION:: aGrob ,bGrob ,DeltaGrob      ! Grobes Raster
      DOUBLE PRECISION:: aFein ,bFein ,DeltaFein      ! Feines Raster
      DOUBLE PRECISION:: D1 ,D2 ,D1min ,D1max ,D2min ,D2max
      DOUBLE PRECISION:: sumx ,sumy ,sumxx ,sumxy      ! Ableitung
20     DOUBLE PRECISION,DIMENSION( 0:MaxDimGrob ):: SPDF
      DOUBLE PRECISION,DIMENSION( 0:MaxDimGrob ):: N      ! Drift
      DOUBLE PRECISION,DIMENSION( 0:MaxDimFein ):: ExtPDF ,IntPDF
      DOUBLE PRECISION,DIMENSION( 1:DDimMax ,1:DDimMax ):: KullbackMat
25     REAL(KIND=8):: Pi ,Kullback ,MaxKullb

      LOGICAL:: RunIterate
      LOGICAL,DIMENSION( 1:DDimMax ,1:DdimMax ):: KullValueOk ! Speichert Fehler

30     CHARACTER(6) :: filename
      ! Initialisierung , Einlesen der Parameter

      pi=QACOS( 0 .q0 ) *2 .q0

35     READ(*,*),filename      ! Zu untersuchender Datensatz
      READ(*,*),DimGrob      ! Dimension des groben Gitters
      READ(*,*),DimFein      ! Dimension des feinen Gitters
      READ(*,*),GridSkip      ! Schrittweite , Abstand zwischen den Bins ,
40     ! fuer die die Berechnung durchgefuehrt wird
```

```

1 READ(*,*) ,MinIterate      !Minimale Anzahl an Iterationen der Langevin-Gl
2 READ(*,*) ,D1Min           !Untere und ...
3 READ(*,*) ,D1Max           !... obere Grenze fuer Q
4 READ(*,*) ,D2Min           !Untere und ...
5 READ(*,*) ,D2Max           !... obere Grenze fuer a
6 READ(*,*) ,DDim             !Dimension des Kullback-Gitters
7 READ(*,*) ,CutOff           !Minimale Anzahl an Messungen pro Bin,
8                               !unterhalb wird der Bin uebersprungen
9 READ(*,*) ,MaxIterate      !Maximale Iterationen pro Bin, darueber wird
10                             !Bin als 'nicht zu berechnen' markiert
11 READ(*,*) ,derivorder      !Nachbarn bei Ableitung (entspricht 'n')

12 OPEN( InFile ,file=filename//".dat")
13 OPEN( GridFile ,file=filename//".grid")
14
15 !Berechnung der stationaeren Verteilung

16 FileError=0
17 READ( InFile ,"(E15.7,E15.7)" ,IOSTAT=FileError) ,x ,tau
18
19 TauSim=Tau/100

20 IF ( FileError.NE.0) THEN
21   WRITE(*,*) , "CalcKullb: Dateifehler"
22   STOP
23 ENDIF

24 aGrob=x
25 bGrob=x
26
27 ExtNormInt=1

28 DO WHILE ( FileError==0)
29   READ( InFile ,"(E15.7)" ,IOSTAT=FileError) ,x
30   IF ( FileError==0) THEN
31     aGrob=MIN( x , aGrob)
32     bGrob=MAX( x , bGrob)
33     ExtNormInt=ExtNormInt+1
34   ENDIF
35 ENDDO

36 DeltaGrob=(bGrob-aGrob)/DimGrob

37 REWIND( InFile )
38 FileError=0

39 DO k=0,DimGrob
40   SPDFInt(k)=0
41 ENDDO

42 DO WHILE ( FileError==0)
43   READ( InFile ,"(E15.7)" ,IOSTAT=FileError) ,x
44   IF ( FileError==0) THEN
45     k=NINT((x-aGrob)/DeltaGrob)
46     SPDFInt(k)=SPDFInt(k)+1
47   ENDIF
48 ENDDO

49 OPEN ( OutFile ,file=filename//".spdf")

50 DO k=0,DimGrob
51   SPDF(k)=1.*SPDFInt(k)/(ExtNormInt*DeltaGrob)
52   WRITE( OutFile ,*) ,aGrob+1.*k*DeltaGrob ,SPDF(k)
53 ENDDO

54 CLOSE( OutFile )

55 !Untersuchung der einzelnen Stellen

```

```

110  DO o=1,DDim !Initialisierung: Alle Werte in Ordnung
111  DO p=0,DDim
112    kullbackMat(o,p)=0.
113    KullValueOk=.true.
114  ENDDO
115  ENDDO

116  DO k=DerivOrder ,DimGrob-DerivOrder ,GridSkip
117    !Randwerte werden ausgelassen, hier Ableitung nicht sicher genug...
118    ExtNormInt=0

119  xPos=aGrob+k*DeltaGrob !Zu untersuchende Stelle

120  IF ( 1.*SPDFInt(k) .GT. CutOff) THEN
121
122    WRITE( GridFile ,*) ,xPos ,SPDFInt(k)

123    !Bestimmung von der CPDF aus Datei:
124    !Nomenklatur: EXT* heisst extern, aus den Daten gewonnen,
125    !INT* hingegen im Programm (intern) simuliert

126  REWIND( InFile )
127  FileError=0

128  aFein=0. !noch nicht sauber...
129  bFein=0.

130  x=0. !Nicht exakt, i.A. aber in Ordnung...

131  DO WHILE ( FileError==0)
132    xAlt=x
133    READ( InFile ,"(E15.7)" ,IOSTAT=FileError) ,x
134    IF ( (2.*ABS(xAlt-xPos)/DeltaGrob.LE.1.) .AND. ( FileError==0) ) THEN
135      aFein=MIN(x-xAlt ,aFein)
136      bFein=MAX(x-xAlt ,bFein)
137      ExtNormInt=ExtNormInt+1
138    ENDIF !in der Box, Fileerror
139  ENDDO

140  IF ( ExtNormInt.GE.2) THEN
141
142    DeltaFein=(bFein-aFein)/DimFein

143    REWIND( InFile )
144    FileError=0

145  DO i=0,DimFein
146    ExtPdflnt(i)=0
147  ENDDO

148  x=0. !s.o.

149  DO WHILE ( FileError==0)
150    xAlt=x
151    READ( InFile ,"(E15.7)" ,IOSTAT=FileError) ,x
152    IF ( (2.*ABS(xAlt-xPos)/DeltaGrob.LE.1.) .AND. ( FileError==0) ) THEN
153      i=NINT((x-xAlt-aFein)/DeltaFein)
154      ExtPdflnt(i)=ExtPdflnt(i)+1
155    ENDIF
156  ENDDO

157  DO i=0,DimFein
158    ExtPDF(i)=1.*ExtPdflnt(i)/(ExtNormInt*DeltaFein)
159  ENDDO

160  !Berechnung von K fuer verschiedene Parametrisierungen

```

```

175      DO o=1,DDim
176      DO p=1,DDim
177
178      IF ( KullValueOk ( o , p )==.TRUE.) THEN
179
180          D1=D1Min+1.*(o-1)*(D1Max-D1Min) /(DDim-1)
181          D2=D2Min+1.*(p-1)*(D2Max-D2Min) /(DDim-1)
182
183          Kullback=.0q0
184
185      DO i=1,DimGrob-1
186
187          ! Berechnung: Drift aus Ableitung von log (SPDF) und D^2
188
189          sumx=0.
190          sumxx=0.
191          sumxy=0.
192          sumy=0.
193          DerivN=0
194          localDerivOrder=MIN( DerivOrder , i , DimGrob-i )
195          DO m=LocalDerivOrder , LocalDerivOrder
196              IF ( SPDF( i+m ) .NE. 0 ) THEN
197                  DerivN=DerivN+1
198                  sumy=sumy+LOG( SPDF( i+m ) )
199                  sumxy=sumxy+(agrob+(i+m)*DeltaGrob ) *LOG( SPDF( i+m ) )
200                  sumx=sumx+(agrob+(i+m)*DeltaGrob )
201                  sumxx=sumxx+(agrob+(i+m)*DeltaGrob ) **2
202              ENDIF
203          ENDDO !m
204          IF ( DerivN .GE. MAX(2*LocalDerivOrder ,2 ) ) THEN
205              N( i )=1.* ( D1+D2*( aGrob+i*DeltaGrob ) **2 ) *( DerivN *SumXY-Sumx*&
206                  &Sumy ) / ( DerivN *SumXX-Sumx**2 ) +2.*D2*( aGrob+i*DeltaGrob )
207          ENDIF
208
209      ENDDO
210
211      ! Simulation der Werte:
212
213      RunIterate=.TRUE. ! Initialisierung der Variablen ...
214      IntNormInt=0
215      DO i=0,DimFein
216          IntPdfInt ( i )=0
217      ENDDO
218
219      DO WHILE ( RunIterate .OR. ( IntNormInt .LE. MinIterate ) )
220
221          CALL RANDOM_NUMBER( RandZahl )
222          xPosFein=xpos+DeltaGrob*(.5-RandZahl)
223          ! Startposition im Bin wird zufaellig gewaehlt
224
225          x=xPosFein
226
227          DO m=1,tau/tausim
228              Xalt=x
229              x=x+TauSim*n1 ( x )+SQRT( TauSim*( D1+D2*x **2 ) ) *random ( )
230          ENDDO !SimNum
231
232          IntNormInt=1+IntNormInt
233          i=NINT(( x-xPosFein-aFein )/DeltaFein )
234
235          IF (( i .GE. 0 ) .AND. ( i .LE. DimFein ) ) THEN
236              IntPdfInt ( i )=IntPdfInt ( i )+1
237
238          IF (( IntPdfInt ( i )==1 ) .AND. ( ExtPdf( i ) .NE. 0 ) .AND. runIterate )THEN
239              runIterate=.FALSE.
240          DO i=0,DimFein

```

```

145      IF ((ExtPdf(i).NE.0).AND.(IntPdfInt(i)==0)) THEN
146        IF ((RunIterate==.FALSE.).AND.(IntNormInt.GE.MinIterate))&
147          &THEN
148            WRITE(*,*),o,p,IntNormInt
149          ENDIF
150          RunIterate=.TRUE.
151        ENDIF
152      ENDDO
153    ENDIF
154    ENDIF !Im Bereich von ExtPdf
155
156    IF (IntNormInt.GE.MaxIterate) THEN
157      KullValueOk(o,p)=.FALSE.
158      RunIterate=.FALSE.
159    ENDIF
160
161    ENDDO !(RunIterate.OR.(IntN...
162
163    !Berechnung des Kullback-Abstandes
164
165    DO i=0,DimFein
166      IntPdf(i)=1.*IntPdfInt(i)/(IntNormInt*DeltaFein)
167    ENDDO
168
169    IF (KullValueOk(o,p)==.TRUE.) THEN
170      DO i=0,DimFein
171        IF (ExtPdf(i).NE.0) THEN
172          kullback=kullback+1.*Spdf(k)*ExtPdf(i)*LOG(ExtPdf(i)/(IntPdf&
173            &(i)))*DeltaFein*DeltaGrob
174        ENDIF
175      ENDDO
176      WRITE(*,*),o,p,IntNormInt,Kullback
177    ENDIF
178
179    KullbackMat(o,p)=KullbackMat(o,p)+Kullback
180
181    ENDIF !KullValueOk?
182
183    ENDDO !p
184    ENDDO !o
185
186    ENDIF !(ExtNormInt.GE.2)...
187
188    !In Datei schreiben
189
190    OPEN(OutFile,file=filename//".out")
191
192    WRITE(OutFile,*),"#Stand:",k
193
194    maxKullb=0.
195
196    DO o=1,DDim          !Maximum bestimmen
197      DO p=1,DDim
198        IF (KullValueOk(o,p)==.TRUE.) THEN
199          MaxKullb=MAX(KullbackMat(o,p),MaxKullb)
200        ENDIF
201      ENDDO
202    ENDDO
203
204    DO o=1,DDim
205      DO p=1,DDim
206        IF (KullValueOk(o,p)==.TRUE.) THEN
207          WRITE(OutFile,*),D1Min+1.*(o-1)*(D1Max-D1Min)/(DDim-1),D2Min&
208            &+1.*(p-1)*(D2Max-D2Min)/(DDim-1),KullbackMat(o,p)
209        ELSE !Fehlende Werte werden durch den maximalen Wert ersetzt
210          WRITE(OutFile,*),D1Min+1.*(o-1)*(D1Max-D1Min)/(DDim-1),D2Min&
211            &+1.*(p-1)*(D2Max-D2Min)/(DDim-1),MaxKullb,"#CutOff!!! "
212        ENDIF
213      ENDDO
214    ENDDO

```

```

305      ENDIF
ENDDO
      WRITE(OutFile,*) ,
ENDDO
      CLOSE(OutFile)

310  ENDIF !(1.*SPDFInt(k).GT. CutOff . .
      WRITE(*,*) ,k ,ExtNormInt

ENDDO !k=DerivOrder ,DimGrob-DerivOrder ,Gr . .

315  CLOSE(GridFile)

CONTAINS

DOUBLE PRECISION FUNCTION random
320  !Erzeugt gaussverteilte Zufallszahlen mit Varianz 2 nach Box-Mueller
  REAL :: y1 ,y2
  CALL RANDOM_NUMBER(y1)
  CALL RANDOM_NUMBER(y2)
  random=2.*SQRT(-LOG(y1))*COS(2.*Pi*y2)
325  END FUNCTION random

DOUBLE PRECISION FUNCTION n1(xfunc)
  !Sucht den Drift, der kontinuierlichen xfunc am naechsten liegt
DOUBLE PRECISION:: xfunc ,yfunc
330  IF ((xfunc.GE.agrob).AND.(xfunc.LE.bgrob)) THEN
    n1=n(NINT((xfunc-aGrob)/DeltaGrob))
  ELSE
    n1=0.
  ENDIF
END FUNCTION n1

335  END PROGRAM Kullback

```

Literaturverzeichnis

- [1] M. Planck, Annalen der Physik **4**, 553 (1901).
- [2] A. Einstein, Annalen der Physik **17**, 132 (1905).
- [3] F. Böttcher, C. Renner, H.-P. Waldl, and J. Peinke, Bound.-Layer Meteorology **108**, 163 (2003).
- [4] S. J. Linz, Phys. Rev. E **54**, 2925 (1996).
- [5] R. Friedrich, J. Peinke, and C. Renner, Phys. Rev. Lett **84**, 5224 (2000).
- [6] H. Haken, *Brain Dynamics, Springer Series in Synergetics* (Springer-Verlag, ADDRESS, 2002).
- [7] H. Haken, *Synergetics, Springer Series in Synergetics* (Springer-Verlag, Berlin, 2004), pp. xvi+763, introduction and advanced topics, Reprint of the third (1983) edition [*Synergetics*] and the first (1983) edition [*Advanced synergetics*].
- [8] H. Haken, *Synergetic Computers and Cognition, Springer Series in Synergetics*, 1st ed. (Springer-Verlag, Berlin, 1991), a Top-Down Approach to Neural Nets.
- [9] R. Friedrich *et al.*, Physics Letters A **271**, 217 (2000).
- [10] R. Friedrich, C. Renner, M. Siefert, and J. Peinke, Physical Review Letters **89**, 149401 (2002).
- [11] M. Ragwitz and H. Kantz, Phys. Rev. Lett. **87**, 254501 (2001).
- [12] M. Ragwitz and H. Kantz, Physical Review Letters **89**, 149402 (2002).
- [13] D. Kleinhans, R. Friedrich, A. Nawroth, and J. Peinke, available on arXiv.org (unpublished).
- [14] H. Risken, *The Fokker-Planck equation*, Vol. 18 of *Springer Series in Synergetics*, 2nd ed. (Springer-Verlag, Berlin, 1989), pp. xiv+472, methods of solution and applications.
- [15] C. W. Gardiner, *Handbook of stochastic methods for physics, chemistry and the natural sciences*, Vol. 13 of *Springer Series in Synergetics*, 3rd ed. (Springer-Verlag, Berlin, 2004), pp. xviii+415.
- [16] R. Friedrich and T. D. Frank, vorlesung an der WWU Münster im Wintersemester 2004/05 (unpublished).
- [17] W. Paul and J. Baschnagel, *Stochastic processes* (Springer-Verlag, Berlin, 1999), pp. xiv+231, from physics to finance.
- [18] H. Dinges and H. Rost, *Prinzipien der Stochastik* (B. G. Teubner, Stuttgart, 1982).

- [19] W. T. Coffey, Y. P. Kalmykov, and J. T. Waldron, *The Langevin equation*, Vol. 14 of *World Scientific Series in Contemporary Chemical Physics*, 2nd ed. (World Scientific Publishing Co. Inc., River Edge, NJ, 2004), pp. xxiv+678, with applications to stochastic problems in physics, chemistry and electrical engineering.
- [20] P. E. Kloeden and E. Platen, *Numerical solution of stochastic differential equations*, Vol. 23 of *Applications of Mathematics (New York)* (Springer-Verlag, Berlin, 1992), pp. xxxvi+632.
- [21] J. Hartung, B. Elpelt, and K.-H. Klösener, *Statistik* (R. Oldenbourg Verlag München Wien, ADDRESS, 1995), lehr- und Handbuch der angewandten Statistik.
- [22] D. S. Lemons, *An introduction to stochastic processes in physics* (Johns Hopkins University Press, Baltimore, MD, 2002), pp. xiv+110, containing “On the theory of Brownian motion” by Paul Langevin, translated by Anthony Gythiel.
- [23] A. Einstein, Annalen der Physik **17**, 549 (1905), reprinted in: Einstein’s Annalen Papers, WILEY-VCH Verlag, 2005.
- [24] H. Kantz and T. Schreiber, *Nonlinear time series analysis*, Vol. 7 of *Cambridge Nonlinear Science Series* (Cambridge University Press, Cambridge, 1997), pp. xvi+304.
- [25] S. Siegert, Ph.D. thesis, Universität Stuttgart, 2001.
- [26] M. Wächter, F. Riess, H. Kantz, and J. Peinke, Europhys. Lett. **64**, 579 (2003).
- [27] T. Kuusela, Physical Review E **69**, 031916 (2004).
- [28] J. Gradišek, S. Siegert, R. Friedrich, and I. Grabec, Phys Rev E **62**, 3146 (2000).
- [29] S. Kriso, R. Friedrich, J. Peinke, and P. Wagner, Physics Letters A **299**, 287 (2002).
- [30] G. R. Jafari *et al.*, Physical Review Letters **91**, 226101 (2003).
- [31] P. Sura and S. Gille, Journal of Marine Research **61**, 313 (2003).
- [32] S. Siegert, R. Friedrich, and J. Peinke, Physics Letters A **243**, 275 (1998).
- [33] S. Lück and J. Peinke, Phys. Rev. Lett. **83**, 5495 (1999).
- [34] T. D. Frank, M. Sondermann, T. Ackemann, and R. Friedrich, submitted to: Nonlinear Phenomena in Complex Systems (unpublished).
- [35] H. U. Bödeker *et al.*, Physical Review E (Statistical, Nonlinear, and Soft Matter Physics) **67**, 056220 (2003).
- [36] J. Gradišek, R. Friedrich, E. Govekar, and I. Grabec, Meccanica **38**, 33 (2003).
- [37] M. Siefert, A. Kittel, R. Friedrich, and J. Peinke, Europhys. Lett. **61**, 466 (2003).
- [38] F. Böttcher, J. Peinke, D. Kleinhans, and R. Friedrich, in preparation (unpublished).
- [39] R. Schlittgen and B. H. J. Streitberg, *Zeitreihenanalyse*, 6 (R. Oldenbourg Verlag, München, 1995).

- [40] M. E. Davies, *Chaos* **8**, 775 (1998).
- [41] R. Meyer and N. Christensen, *Phys. Rev. E* **62**, 3535 (2000).
- [42] R. Meyer and N. Christensen, *Phys Rev E* **65**, 16206 (2001).
- [43] F. Böttcher, Ph.D. thesis, Carl von Ossietzky Universität, Oldenburg, 2005.
- [44] S. Kullback, in *Information Theory and Statistics*, edited by W. A. Shewhart and S. S. Wilks (Wiley Publications in Statistics, ADDRESS, 1959).
- [45] H. Haken, *Information and self-organization, Springer Series in Synergetics*, 2nd ed. (Springer-Verlag, Berlin, 2000), pp. xiv+222, a macroscopic approach to complex systems.
- [46] E. Weinstein, M. Feder, and A. V. Oppenheim, *IEEE Transactions on Acoustics, Speech, and Signal Processing* **38**, 1652 (1990).
- [47] S. Koonin and D. Meredith, *Physik auf dem Computer* (Oldenbourg Verlag München Wien, ADDRESS, 1990), FORTRAN-Version.
- [48] G. D. Smith, *Numerical solution of partial differential equations* (Oxford University Press, New York, 1965), pp. viii+179.
- [49] *A closer look at the WinDaq derivative algorithm*, DATAQ Instruments, Inc.

Danksagung

An dieser Stelle möchte ich mich zunächst bei Herrn Professor Dr. Rudolf Friedrich bedanken. Er hat mich in der Endphase meines Studiums freundlich betreut und stets gefördert. Auf diese Weise hat er dazu beigetragen, dass meine Motivation für das Anfertigen dieser Arbeit stets aufrecht erhalten blieb.

Außerdem danke ich Herrn Dr. Till Frank für die guten Ratschläge und die sorfältige Durchsicht dieser Arbeit.

Ich danke Guido Krüger, mit dem ich das Büro im Institut geteilt habe, für die angenehme Atmosphäre in unserem Zimmer. Seine konstruktive Kritik war mir beim Niederschreiben dieser Arbeit eine große Hilfe.

Ferner möchte ich mich bei all denjenigen bedanken, die an der Entwicklung professioneller Software mitwirken und diese frei zur Verfügung stellen. Es ist bemerkenswert, dass diese Produkte ihre kommerziell entwickelte Konkurrenz häufig in den Schatten stellen.

Mein besonderer Dank gilt meinen Eltern, die mich immer von Herzen unterstützt und mir das Studium ermöglicht haben.

Erklärung zur Diplomarbeit:

Hiermit versichere ich, diese Arbeit selbständig angefertigt und keine anderen als die angegebenen Hilfsmittel verwendet zu haben.

Münster, im Juni 2005