

**Nichtlineare Methoden  
zur Quantifizierung von  
Abhängigkeiten und Kopplungen  
zwischen stochastischen Prozessen  
basierend auf Informationstheorie**

**Dissertation**

ZUR ERLANGUNG DES DOKTORGRADES  
DER FAKULTÄT DER NATURWISSENSCHAFTEN (DR. RER. NAT.)  
DER BERGISCHEN UNIVERSITÄT GESAMTHOCHSCHULE  
WUPPERTAL

WUB-DIS 2002-10

vorgelegt von

**Andreas Kaiser**

geboren am 19. März 1972 in Melle

Dezember 2002

erstellt am

MAX-PLANCK-INSTITUT FÜR PHYSIK KOMPLEXER SYSTEME  
DRESDEN



## Abstract

In order to determine the relation between two stochastic processes information theory offers an appropriate framework in which of the relationships can be interpreted in terms of information. The dependence can be measured with the mutual information, giving the amount of information which both processes share, i.e. the degree of similarities. Mutual information can also give hints for the coupling direction, however, due to serial correlation in time the results might be misleading. Additionally, only non-coupled systems can be distinguished from coupled systems. To determine the coupling directions, the dynamics of the processes have to be taken into account which leads to the transfer entropy. By considering the past, transfer entropy measures the direct impact which the driving process has on the future state of the driven process by excluding any influence due to the serial correlations. Based on information theory, coupling strength is quantified as the amount of effective information transmission from one process to the other. Thus, transfer entropy allows to distinguish between unidirectional coupling and bidirectional coupling.

While values for mutual information and transfer entropy can be easily archived for processes with discrete state space, their estimations from finite data sets are difficult. Partitioning the state space, mutual information and transfer entropy of the discretised processes converge to the corresponding values of the continuous processes if the partitions are refined. Furthermore, mutual information shows monotonically increasing convergence and thus can be used to reject the assumption of both processes being independent. For transfer entropy no similar monotonic convergence seems to hold. Kernel estimators represent an alternative approach in order to estimate information theoretical quantities. They are easy to implement and the bias of the estimators due to serial correlations in the data sets can be suppressed easily too.

A special class of stochastic processes are point processes. Here, the discrete times at which an event occurs are of interest. Again mutual information can be used to quantify dependence between two point processes but this time the increments, i.e. the number of events within a certain time interval, have to be considered. A weaker measure for dependence is the covariance of the increments leading in a special case to the number of coincidences. Considering increments, coupling directions can be determined with the transfer entropy as well. Unfortunately, due to the large bias in the estimators the exact values of the information transmissions cannot reliably be given. When using increments, the time scale on which dependence is detected is given by the length of the time intervals. As an alternative method inter-event intervals and cross-event intervals are introduced which are ordered to one discrete time index congruently. By calculating the mutual information between these event intervals dependence between point processes is detectable without choosing a certain time scale.



# Inhaltsverzeichnis

<b>1</b>	<b>Einleitung</b>	<b>1</b>
<b>2</b>	<b>Grundlagen der Informationstheorie</b>	<b>5</b>
2.1	Diskrete stochastische Prozesse . . . . .	5
2.1.1	Verteilung diskreter Prozesse . . . . .	5
2.1.2	Dynamische Eigenschaften diskreter Prozesse . . . . .	11
2.1.3	Diskretes Beispiel . . . . .	17
2.2	Kontinuierliche stochastische Prozesse . . . . .	21
2.2.1	Verteilung kontinuierlicher Prozesse . . . . .	21
2.2.2	Dynamische Eigenschaften kontinuierlicher Prozesse . . . . .	24
2.2.3	Transformationsinvarianz . . . . .	27
<b>3</b>	<b>Schätzen von Entropien und Informationen</b>	<b>31</b>
3.1	Schätzen bei einer und mehreren Beobachtungen . . . . .	31
3.2	Schätzer für diskrete Prozesse . . . . .	34
3.3	Partitionierung des Zustandsraums . . . . .	36
3.4	Parametrische Verteilungen . . . . .	43
3.5	Kontinuierliches Beispiel: AR(1)-Prozesse . . . . .	45
3.6	Kernschätzer für Dichten . . . . .	46
3.7	Räumliche Abhängigkeit des Windes . . . . .	52
<b>4</b>	<b>Exkurs: Dynamische Systeme</b>	<b>61</b>
4.1	Entropie eines dynamischen Systems . . . . .	61
4.2	Interdependenz, verallgemeinerte Synchronisation . . . . .	66
<b>5</b>	<b>Punktprozesse</b>	<b>73</b>
5.1	Definition eines Punktprozesses . . . . .	73
5.2	Momente und Ereignisraten . . . . .	77
5.3	Gekoppelte Punktprozesse . . . . .	79
5.4	Nachweis von Abhängigkeiten mittels Zuwächsen . . . . .	82
5.4.1	Abhängigkeitsnachweis über Zuwächse . . . . .	82
5.4.2	Anwendung von Momente und Koinzidenzen . . . . .	89
5.5	Nachweis von Kopplung mittels Zuwächsen . . . . .	95

5.6	Abhängigkeitsnachweis über Ereignisintervalle . . . . .	102
<b>6</b>	<b>Zusammenfassung</b>	<b>109</b>
<b>A</b>	<b>Mathematische Werkzeuge</b>	<b>113</b>
	<b>Literatur</b>	<b>117</b>

# Kapitel 1

## Einleitung

Zum Verständnis komplexer Systeme, wie des Erdklimas, neuronaler Netzwerke, Kreislaufsysteme von Lebewesen oder Finanzmärkten, ist es notwendig, die Wechselwirkungen der einzelnen Subsysteme, also die Abhängigkeiten und Kopplungen der Subsysteme untereinander zu bestimmen. Eine Vielzahl von Arbeiten beschäftigt sich mit der Bestimmung der Kopplungsstärke und Kopplungsrichtung von dynamischen Systemen aus Zeitreihen, insbesondere im Kontext von Synchronisation tritt diese Fragestellung häufig auf [Rosenblum et al. (1996), Pikovsky et al. (2001), Quian Quiroga et al. (2000)]. Für Systeme wie Phasenoszillatoren sind verlässliche Antworten erhältlich [Rosenblum & Pikovsky (2001)]. Ist das angetriebene System nichtsingulär, so kann die Kreuzvorhersage [Rulkov et al. (1995)] oder die Interdependenz [Arnhold et al. (1999), Le Van Quyen et al. (1999)] herangezogen werden, um Aussagen über die Kopplungsrichtung zu bekommen. Oft wird zur Identifikation des treibenden und des angetriebenen Systems versucht, die Kopplungsstärken zu quantifizieren. Hierzu nehmen die meisten Autoren an, dass ein bestimmter Parameter in der zugrunde liegenden Bewegungsgleichung die Auswirkung der Kopplung bestimmt. Aber gerade im Zusammenhang mit nichtlinearen Prozessen ist die Quantifizierung des Kopplungsterms nicht eindeutig, vielmehr hängt die Kopplungsstärke von der verwendeten Methode und den gemessenen Systemvariablen ab.

Zur Bestimmung der Eigenschaften und zur Analyse von stochastischen Systemen haben sich informationstheoretische Techniken [Shannon & Weaver (1949), Cover & Thomas (1991), Jumarie (1990)] wie Shannon- und Kolmogorov-Sinai-Entropien bewährt und sind daher weit verbreitet [Katok & Hasselblatt (1995), Ott (1993), Palus (1996b), Brandt & Pompe (1993), La Rosa et al. (2000)]. Da diese Methoden die Systemeigenschaften in Form von Informationen oder Informationsraten quantifizieren, ist hierdurch eine anschauliche Interpretation der Werte möglich. Aussagen über die Kopplung zwischen stochastischen Prozessen basieren daher häufig auf der Betrachtung der stochastischen Abhängigkeit zweier Prozesse voneinander, welche mit der gegenseitigen Infor-

mation quantifiziert wird [Prichard & Theiler (1995), Fraser & Swinney (1986), Darbellay (1999), Palus (1996a), Pompe (1993)]. Allerdings ist es nicht ohne weiteres möglich, allein von stochastischer Abhängigkeit auf eine Kopplung zu schließen. Beispielsweise sind periodische Prozesse, deren Periodendauern sich lediglich um ein ganzzahliges Vielfaches unterscheiden, stets voneinander abhängig, auch wenn sie ungekoppelt sind. Des Weiteren misst die gegenseitige Information den Informationsaustausch zwischen zwei Prozessen, welcher wiederum symmetrisch ist. Folglich kann mit ihr nicht die dominierende Richtung des Informationsflusses bestimmt werden, was notwendig wäre, um Rückschlüsse auf das getriebene und angetriebene System ziehen zu können.

Andere Konzepte verwenden die bedingte Shannon-Entropie, um Aussagen über die Kopplungsstärke zu machen. Hierbei wird die mittlere Unsicherheit ermittelt, mit der ein zukünftiger Zustand von Prozess  $X$  bei bekannter Vergangenheit von  $Y$  beobachtet wird [Porta et al. (1999), Eguia et al. (2000)]. Bei dieser Methode, welche als Analogon zur Kreuzvorhersage betrachtet werden kann, wird der Informationsfluss innerhalb von  $X$  nicht mit berücksichtigt. Zur Behebung dieses Mangels schlug Schreiber (2000) vor, die Prozesse daraufhin zu überprüfen, ob sie die verallgemeinerte Markov-Eigenschaft erfüllen, das heißt ob der zukünftige Zustand von  $X$  nicht nur von seiner eigenen Vergangenheit, sondern auch von der von  $Y$  abhängt. Die Abweichung von der verallgemeinerten Markov-Eigenschaft wird mit der Transferentropie gemessen. Die Transferentropie gibt den Informationsfluss von einem Prozess zum anderen an und ist somit im Gegensatz zur gegenseitigen Information asymmetrisch beim Vertauschen der beiden Prozesse.

Da die Transferentropie nur für stochastische Prozesse mit diskretem Zustandsraum eingeführt wurde, besteht das Hauptanliegen dieser Arbeit darin, das Konzept der Transferentropie auf stochastische Prozesse mit kontinuierlichem Zustandsraum zu übertragen. Dieser Schritt ist notwendig, weil eine Vielzahl von physikalischen Prozessen nicht mit diskreten, sondern mit kontinuierlichen Variablen beschrieben wird. Zu diesem Zweck werde ich zunächst die wichtigsten Größen und Konzepte der Informationstheorie für diskrete Prozesse vorstellen, wobei der gegenseitigen Information und der Transferentropie besondere Aufmerksamkeit geschenkt werden. Insbesondere gehe ich der Frage nach, welche Relation zwischen diesen beiden Größen besteht und wann aus der stochastischen Unabhängigkeit geschlossen werden kann, dass die Prozesse stochastisch ungekoppelt sind, also die verallgemeinerte Markov-Eigenschaft erfüllt ist. Außerdem werden weitere wichtige mathematische Eigenschaften der Transferentropie aufgezeigt. Im Anschluss hieran stelle ich die wichtigsten informationstheoretischen Größen vor, mit denen kontinuierliche Prozesse untersucht werden. In diesem Kontext definiere ich die kontinuierliche Transferentropie. Hierfür werden im Wesentlichen die Übergangsverteilungen der Prozesse durch die Übergangsverteilungsdichten ersetzt. Außerdem überprüfe ich, welche Eigenschaften der (diskreten) Transferentropie auch von der kontinuierlichen erfüllt werden.

Sobald kontinuierliche Prozesse betrachtet werden, tritt eine Schwierigkeit auf, welche bei diskreten Prozessen nicht existiert: die Frage nach der Invarianz von Shannon-Entropie, gegenseitiger Information und Transferentropie gegenüber Koordinatentransformationen. Bekannt ist, dass die Shannon-Entropie im Allgemeinen nicht invariant ist und dass die gegenseitige Information gegenüber Reskalierungen und Verschiebungen der Prozesse invariant bleibt. Handelt es sich beispielsweise um mehrdimensionale Prozesse, so stellt bereits die einfache Rotation der Messanordnung eine Koordinatentransformation dar, die über eine lineare Transformation hinausgeht. Damit die Quantifizierung von Abhängigkeiten und Kopplungen physikalisch interpretierbar ist, müssen die gegenseitige Information und die Transferentropie invariant gegenüber solchen Transformationen bleiben. Aus diesem Grund untersuche ich, welche Eigenschaften eine Transformation erfüllen muss, damit die Invarianz dieser Größen gewährleistet ist.

Während das Schätzen der gegenseitigen Information und der Transferentropie aus Daten für diskrete Prozesse relativ einfach ist, treten bei kontinuierlichen Prozessen eine Vielzahl von Schwierigkeiten auf. Aus diesem Grund werden oft die kontinuierlichen Prozesse in diskrete umgewandelt, indem der Zustandsraum partitioniert wird [Cover & Thomas (1991), Darbellay (1999), Fraser & Swinney (1986)]. Ist die Verteilungsdichte eines Prozesses stetig und existiert die Shannon-Entropie als Riemann-Integral, so ist bekannt, dass die diskretisierte Shannon-Entropie gegen die kontinuierliche konvergiert, wenn die Partitionselemente immer kleiner gewählt werden [Cover & Thomas (1991)]. Aufbauend auf dieser Tatsache wird untersucht, unter welchen Bedingungen eine Konvergenz von gegenseitiger Information und Transferentropie zu erwarten ist.

Besitzt man *a priori* Kenntnis über die Verteilung der Prozesse, so können die Dichten durch Schätzen der Verteilungsparameter bestimmt werden. Anschließend kann beispielsweise die Transferentropie durch numerische Integration berechnet werden. Handelt es sich um Gauß-Prozesse, so existieren für die Shannon-Entropie und die gegenseitige Information analytische Ausdrücke, in die nur die Kovarianzmatrizen eingehen [Cover & Thomas (1991), Prichard & Theiler (1995)]. Dies gilt ebenfalls für die Transferentropie, was ich in einer kurzen Rechnung zeigen werde.

Eine weitere Technik, mit der die Transferentropie aus Daten berechnet werden kann, basiert auf der Verwendung von Kernschätzern. Dabei handelt es sich um nichtparametrische Dichteschätzer [Silverman (1986)]. Im Gegensatz zur Partitionierung des Zustandsraums kann hier der Bias des Schätzers aufgrund serieller Korrelationen innerhalb der Zeitreihe durch eine einfache Modifikation unterdrückt werden. An einem Beispielsystem, bestehend aus zwei autoregressiven Prozessen, die unidirektional gekoppelt sind, wird demonstriert, wie gegenseitige Information und Transferentropie mit Hilfe von Kernschätzern aus Daten berechnet werden können. Insbesondere ist die Modellstruktur bezüglich der Kopplung identifizierbar.

Mit einem Anwendungsbeispiel wird außerdem gezeigt, wie die gegenseitige

Information auf reale Daten angewendet werden kann, um Abhängigkeiten zwischen den Windgeschwindigkeiten an verschiedenen Orten zu untersuchen.

Zu Beginn der Einleitung habe ich erwähnt, dass die Interdependenz benutzt werden kann, um die Kopplungsrichtung in einem dynamischen System zu bestimmen. In einem kurzen Exkurs möchte ich anhand von logistischen Abbildungen zeigen, dass dieses Verfahren auch dann noch funktioniert, wenn die Parameter des dynamischen Systems sowie die Kopplungsparameter einen langsamen Drift aufweisen und nur eine Zeitreihe für die Untersuchung zur Verfügung steht.

Zum Abschluss dieser Arbeit widme ich mich dem Nachweis von Abhängigkeiten und Kopplungen sowie der Identifizierung der Kopplungsrichtung zwischen Punktprozessen. Hierbei handelt es sich um eine besondere Klasse von stochastischen Prozessen, bei denen die Zeitpunkte, in denen jeweils ein Ereignis eintritt, die zentrale Rolle spielen [Lewis (1972), Bauer (1991), Billingsley (1995), Grün et al. (2002)]. Typische Beispiele für Punktprozesse sind radioaktiver Zerfall, Elektrokardiogramme oder Membranpotentiale von spikenden Neuronen.

Dabei beschäftige ich mich mit der Frage, wie eine Abhängigkeit zwischen Punktprozessen nachgewiesen und quantifiziert werden kann, wenn deren Zuwächse, also die Anzahl der Ereignisse in einem Zeitfenster, betrachtet werden. In diesem Zusammenhang gehe ich auch auf das Auszählen von Koinkidenzen ein, welches ein weit verbreitetes Verfahren ist, um Korrelationen zwischen Punktprozessen aufzuspüren [Grün et al. (1999), Riehle et al. (1997), Schreiber & Schmitz (2000a)].

Unter bestimmten Voraussetzungen kann mit der Transferentropie die Kopplungsrichtung zwischen Punktprozessen allein durch die Betrachtung der Zuwächse bestimmt werden. Dies wird anhand von zwei gekoppelten Hindmarch-Rose-Oszillatoren gezeigt.

Die Zuwächse von Punktprozessen stellen Ereignisgruppen dar, welche von der Fensterbreite bestimmt werden. Demzufolge werden die Prozesse in den bisher erwähnten Verfahren bezüglich dieser Gruppen analysiert, es sei denn, die Fensterbreiten werden so klein gewählt, dass alle Zuwächse höchstens ein Ereignis enthalten. Oft leidet darunter aber die Signifikanz der Verfahren, insbesondere bei relativ kurzen Zeitreihen. Als Alternative hierzu stelle ich daher eine neue Methode vor, bei der der Nachweis von Abhängigkeit stets auf der Ebene von einzelnen Ereignissen vonstatten geht. Zu diesem Zweck werden spezielle Zeitintervalle zwischen den Ereignissen beider Punktprozesse definiert, welche mit der gegenseitigen Information auf Abhängigkeit hin untersucht werden.

# Kapitel 2

## Grundlagen der Informationstheorie

### 2.1 Diskrete stochastische Prozesse

#### 2.1.1 Verteilung diskreter Prozesse

Eine Vielzahl von physikalischen, biologischen, technologischen oder soziologischen Systemen lassen sich durch stochastische Prozesse mit diskreten Zuständen beschreiben. Dabei kann die diskrete Struktur natürlichen Ursprungs sein, zum Beispiel Quantenzustände, digitale Datenverarbeitung, Populationszahlen oder aber aufgrund endlicher Meßauflösungen durch digitale Datenaufzeichnung, bedingt sein.

Gerade wegen des diskreten Zustandsraums sind Methoden aus der Informationstheorie geeignet, um solche stochastischen Prozesse zu untersuchen. Daher werden in diesem Abschnitt die grundlegenden Begriffe der Informationstheorie vorgestellt. Hierauf aufbauend werden Techniken entwickelt, mit denen es möglich ist, die Wechselwirkung von stochastischen Subsystemen untereinander zu analysieren.

Im Folgenden sei  $(\Omega, \mathcal{A}, P)$  ein *Wahrscheinlichkeitsraum*, wobei  $\Omega$  eine beliebige Menge ist,  $\mathcal{A}$  eine  $\sigma$ -Algebra in  $\Omega$  und  $P$  ein auf Eins normiertes Maß auf  $\mathcal{A}$ , welches als *Wahrscheinlichkeitsmaß* bezeichnet wird, siehe beispielsweise [Bauer (1991), Behnen & Neuhaus (1995), Billingsley (1995)]. Die einzelnen Elemente  $\omega \in \Omega$  stellen dabei die verschiedenen zufälligen Realisierungen des beobachteten Experimentes dar. Die Wahrscheinlichkeit, mit der alle Realisierungen des Ereignisses  $A \in \mathcal{A}$ ,  $A \subset \Omega$  beobachtet werden, ist durch  $P(A)$  gegeben.

Des Weiteren sei  $X = (X_i)_{i \in \mathbb{Z}}$  eine Familie von Zufallsvariablen (messbaren Abbildungen) auf dem Wahrscheinlichkeitsraum  $(\Omega, \mathcal{A}, P)$  mit Werten in einem

gemeinsamen Messraum  $(\mathcal{X}, \mathcal{B})$ :

$$\begin{aligned} X_i : (\Omega, \mathcal{A}, P) &\longrightarrow (\mathcal{X}, \mathcal{B}), \\ X_i : \omega &\longrightarrow X_i(\omega) \in \mathcal{X}. \end{aligned} \quad (2.1)$$

Die Menge  $\mathcal{X}$  wird *Zustandsraum* genannt und  $\mathcal{B}$  ist eine  $\sigma$ -Algebra in  $\mathcal{X}$ . Im Folgenden sei der Zustandsraum abzählbar, also diskret und die  $\sigma$ -Algebra  $\mathcal{B}$  sei gleich der Potenzmenge  $\mathbb{P}(\mathcal{X}) = \{B \subset \mathcal{X}\}$ , also gleich der Menge aller Teilmengen von  $\mathcal{X}$ .

Dann bezeichnet man das Quadrupel  $(\Omega, \mathcal{A}, P, (X_i)_{i \in \mathbb{Z}})$  als einen *stochastischen Prozess mit diskreter Zeitmenge und diskretem Zustandsraum* oder kurz als *diskreten stochastischen Prozess mit diskreter Zeitmenge*, siehe zum Beispiel [Bauer (1991), Behnen & Neuhaus (1995), Billingsley (1995)]. Die Zeitmenge ist hier durch  $\mathbb{Z}$  gegeben. Oft wird solch ein stochastischer Prozess einfach mit  $X = (X_i)_{i \in \mathbb{Z}}$  bezeichnet, wobei man sich den Wahrscheinlichkeitsraum  $(\Omega, \mathcal{A}, P)$  dazu denken muss. Die Wahrscheinlichkeit, mit der der Prozess  $X$  zur Zeit  $i$  in den Zuständen aus  $B \in \mathcal{B}$  beobachtet wird, ist durch dessen *Verteilung*  $P^{X_i}$  gegeben:  $P^{X_i}(B) = P(\{\omega \in \Omega : X_i(\omega) \in B\}) = P(X_i^{-1}(B))$ , siehe auch [Bauer (1991), Behnen & Neuhaus (1995)]. Die Messbarkeit von  $X_i$  gewährleistet dabei, dass  $X_i^{-1}(B) \in \mathcal{A}$  ist, siehe Gl. (2.1). Für die Wahrscheinlichkeit, dass  $X_i$  einen bestimmten Zustand  $x \in \mathcal{X}$  einnimmt, wird im Folgenden  $P\{X_i = x\}$  statt  $P^{X_i}(\{x\})$  geschrieben. Insbesondere gilt für die Verteilungen von allen  $X_i$ :

$$\begin{aligned} P^{X_i}(B) &\geq 0 && \forall B \in \mathcal{B}, \\ P^{X_i}(B) &= \sum_{x \in B} P\{X_i = x\}, \\ P^{X_i}(\mathcal{X}) &= 1. \end{aligned}$$

Für jedes  $\omega \in \Omega$  heißt die durch  $i \longrightarrow X_i(\omega)$  definierte Abbildung *Realisierung* oder auch *Pfad* des Prozesses  $X$ .

Da jeder ein-dimensionale, bzw.  $d$ -dimensionale diskrete Zustandsraum eindeutig auf eine Teilmenge der ganzen Zahlen  $\mathbb{Z}$  bzw.  $\mathbb{Z}^d$  abgebildet werden kann, wird ohne Beschränkung der Allgemeinheit angenommen, dass  $\mathcal{X} = \mathbb{Z}$  bzw.  $\mathcal{X} = \mathbb{Z}^d$  ist.

Im Gegensatz zu deterministischen Prozessen sind die Zustände  $X_i(\omega)$  eines stochastischen Prozesses selbst bei gleichem Anfangswert  $X_0(\omega) \equiv x_0$  für verschiedene Realisierungen  $\omega_1, \omega_2 \in \Omega$  unterschiedlich:  $X_i(\omega_1) \neq X_i(\omega_2)$ . Stattdessen weist  $X_i$  zu jedem Zeitpunkt  $i$  eine bestimmte Verteilung auf. Im Allgemeinen werden auch die Zustände von  $X_0$  über den Zustandsraum verteilt sein.

Somit kann der diskrete stochastische Prozess  $X = (X_i)_{i \in \mathbb{Z}}$  zu jedem Zeitpunkt  $i \in \mathbb{R}$  durch die Verteilung  $P\{X_i = x\}$  seiner Zustände  $x \in \mathcal{X}$ , also durch die Funktion  $x \longrightarrow P\{X_i = x\}$  beschrieben werden. Eine einfachere Charakterisierung der Verteilung von  $X_i$  kann mit einem von Shannon & Weaver (1949)

eingeführten Funktional erfolgen. Dazu wird die *Unsicherheit*  $-\log P\{X_i = x\}$  des Zustandes  $x$  eingeführt. Sie ist ein Maß, das angibt, wie unwahrscheinlich es ist, bei einer Beobachtung  $X_i$  im Zustand  $x$  zu finden. Die Mittelung dieser Unsicherheiten über alle  $x \in \mathcal{X}$  liefert das Funktional

$$H(X_i) = - \sum_{x \in \mathcal{X}} P\{X_i = x\} \log P\{X_i = x\}, \quad (2.2)$$

das die Verteilung  $P^{X_i}$  auf eine nichtnegative Zahl abbildet. Dieses Funktional wird *Shannon-Entropie* genannt. Shannon & Weaver (1949) zeigten, dass diese Entropie bis auf einen Vorfaktor das einzige Funktional ist, welches aus folgenden Axiomen hervorgeht:

- A1:  $H(X_i)$  ist stetig in den Wahrscheinlichkeiten  $(P\{X_i = x\})_{x \in \mathcal{X}}$ .  
(Kleine Veränderungen in der Verteilung sollen nur kleine Änderungen in der mittleren Unsicherheit verursachen.)
- A2: Sind alle Zustände gleich verteilt,  $P\{X_i = x_1\} = \dots = P\{X_i = x_m\} = 1/m$ , dann soll  $H(X_i)$  eine monoton steigende Funktion in  $m$  sein.  
(Je größer die Anzahl der möglichen Zustände ist, desto größer ist die Unsicherheit.)
- A3:  $H(X_i)$  lässt sich als eine gewichtete Linearkombination von Teilentropien beim Zusammenfassen von Zuständen darstellen.  
Hierzu sei  $X_i$  eine Zufallsvariable mit Zuständen in  $\{x_1, \dots, x_m\}$ .  $Y_i, Z_i$  seien zwei weitere Zufallsvariablen mit Werten in  $\{y_1, \dots, y_{m-1}\}$  bzw.  $\{z_1, z_2\}$ , deren Verteilungen durch

$$P\{Y_i = y_j\} = \begin{cases} P\{X_i = x_j\} & \text{falls } j \leq m-2 \\ P\{X_i = x_{m-1}\} + P\{X_i = x_m\} & \text{falls } j = m-1 \end{cases},$$

$$P\{Z_i = z_1\} = \frac{P\{X_i = x_{m-1}\}}{P\{X_i = x_{m-1}\} + P\{X_i = x_m\}},$$

$$P\{Z_i = z_2\} = \frac{P\{X_i = x_m\}}{P\{X_i = x_{m-1}\} + P\{X_i = x_m\}}$$

gegeben seien. Dann gilt

$$H(X_i) = H(Y_i) + (P\{X_i = x_{m-1}\} + P\{X_i = x_m\}) H(Z_i).$$

(Somit ist die Unsicherheit der Zustände von  $X_i$  gleich der Unsicherheit, mit der bei einer ersten Messung bestimmt wird, ob einer der Zustände  $x_1, \dots, x_{m-2}$  oder ein Zustand aus  $\{x_{m-1}, x_m\}$  vorliegt plus der Unsicherheit, mit der bei einer zweiten Beobachtung, welche mit Wahrscheinlichkeit  $P\{X_i = x_{m-1}\} + P\{X_i = x_m\}$  notwendig ist, bestimmt wird, ob  $x_{m-1}$  oder  $x_m$  auftrat.)

Siehe hierzu auch [Jumarie (1990)].

Für diskrete deterministische Prozesse,

$$P\{X_i = x\} = \begin{cases} 1 & \text{falls } x = x' \\ 0 & \text{sonst} \end{cases},$$

gilt stets  $H(X_i) = 0$ , da das Besetzen eines Zustandes eindeutig bestimmt ist, also keine Unsicherheit aufweist. Ist der Ereignisraum endlich,  $\mathcal{X} = \{x_1, \dots, x_m\}$ , und sind alle Zustände gleichverteilt,  $P\{X_i = x_1\} = \dots = P\{X_i = x_m\} = 1/m$ , das heißt die Beobachtung aller Zustände von  $X_i$  ist gleich unsicher, so ist die Shannon-Entropie mit  $H(X_i) = \log(m)$  maximal. Für alle anderen diskreten stochastischen Prozesse mit endlichem Zustandsraum gilt  $0 \leq H(X_i) \leq \log(m)$ , wobei  $H(X_i)$  umso größer ist, je unsicherer die Vorhersage der Zustände von  $X_i$  im Mittel ist.

Dieses Konzept kann auf endlich-dimensionale Randverteilungen  $P^{X_{i_1}, \dots, X_{i_j}}$  von  $X$  erweitert werden. Hier ist  $j \in \mathbb{N}$  und  $i_1 < i_2 < \dots < i_j \in \mathbb{Z}$  beliebig. Denn mit  $X_i$ ,  $i \in \mathbb{Z}$  und abzählbarem Zustandsraum  $\mathcal{X}$  besitzt  $(X_{i_1}, \dots, X_{i_j})$  den abzählbaren Zustandsraum  $\mathcal{X}^j$ . Demnach muss in Gl. (2.2) lediglich  $P\{X_i = x\}$  durch  $P\{X_{i_1} = x_1, \dots, X_{i_j} = x_j\} = P^{X_{i_1}, \dots, X_{i_j}}(\{(x_1, \dots, x_j)\})$  ersetzt und über alle Zustände aus  $\mathcal{X}^j$  gemittelt werden, um die Shannon-Entropie von  $(X_{i_1}, \dots, X_{i_j})$  zu erhalten.

Häufig liegt einem beobachteten stochastischen Prozess eine Modellvorstellung zugrunde und damit auch eine *a priori* Verteilung  $Q^{X_i}$  des Prozesses. Auch hier wird  $Q\{X_i = x\}$  statt  $Q^{X_i}(\{x\})$  für die *a priori* Wahrscheinlichkeit, mit der  $X_i$  in  $x$  zu finden ist, geschrieben. Der Approximationsfehler, wenn die *a priori* Verteilung  $Q^{X_i}$  statt der *wahren* Verteilung  $P^{X_i}$  angenommen wird, kann durch die mittlere Differenz der Unsicherheiten  $-\log P\{X_i = x\} + \log Q\{X_i = x\}$  quantifiziert werden:

$$K_{P||Q}(X_i) = \sum_{x \in \mathcal{X}} P\{X_i = x\} \log \frac{P\{X_i = x\}}{Q\{X_i = x\}}. \quad (2.3)$$

Dies Funktional wird *Kullback-Entropie* [Kullback (1959), Jumarie (1990)] bzw. *Kullback-Leibler-Abstand* [Cover & Thomas (1991)]. Damit es existiert, muss  $Q\{X_i = x\} > 0$  für alle  $x \in \{x' \in \mathcal{X} : P\{X_i = x'\} > 0\}$  gefordert werden. Mit häufig getroffenen der Konvention  $0 \log 0 = 0$ ,  $a \log(a/0) = \infty$  falls  $a > 0$  und  $0 \log(0/0) = 0$  ist dies Funktional auch definiert, falls  $Q\{X_i = x\} \equiv 0$  gilt. In diesem Fall ist  $K_{P||Q}(X_i) = \infty$ , was aber eine relativ geringe Aussage hat. Die Kullback-Entropie nimmt nur nichtnegativen Werte an. Dies folgt aus der Log-Summen-Ungleichung [Cover & Thomas (1991)], die wiederum aus der Jensenschen Ungleichung [Bauer (1991), Cover & Thomas (1991)] hervorgeht. Die Beweise für beide Ungleichungen sind in Anhang A zu finden.

Die *Log-Summen-Ungleichung* besagt, dass für beliebige nichtnegative Zahlen  $a_1, \dots, a_n$  und  $b_1, \dots, b_n$  die Ungleichung

$$\sum_{i=1}^n a_i \log \frac{a_i}{b_i} \geq \left( \sum_{i=1}^n a_i \right) \log \frac{\sum_{i=1}^n a_i}{\sum_{i=1}^n b_i} \quad (2.4)$$

gilt. Wobei die Gleichheit dann und nur dann erfüllt ist, wenn  $a_i/b_i = \text{konstant}$  sind. Entsprechend der zuvor eingeführten Konversion ist die Ungleichung auch dann noch erfüllt, wenn  $b_i = 0$  für ein  $i$  gilt.

Hieraus folgt unmittelbar, dass

$$K_{P||Q}(X_i) \geq \left( \sum_{x \in \mathcal{X}} P\{X_i = x\} \right) \log \frac{\sum_{x \in \mathcal{X}} P\{X_i = x\}}{\sum_{x \in \mathcal{X}} Q\{X_i = x\}} = 0$$

ist und  $K_{P||Q}(X_i)$  genau dann verschwindet, wenn *wahre* und *a priori* Verteilung identisch sind,  $P^{X_i} = Q^{X_i}$ . An dieser Stelle sei darauf hingewiesen, dass es sich bei der Kullback-Entropie nicht um einen Abstand im Sinne einer Metrik oder Norm handelt.

Mit Hilfe der Kullback-Entropie ist es möglich, die Abhängigkeit zweier stochastischer Prozesse mit diskreten Zuständen zu quantifizieren. Sind  $X = (X_i)_{t \in \mathbb{Z}}$  und  $Y = (Y_i)_{i \in \mathbb{Z}}$  stochastische Prozesse mit abzählbarem Zustandsraum  $\mathcal{X}$  bzw.  $\mathcal{Y}$ , dann ist auch der Produktprozess  $(X, Y) = ((X_i, Y_i))_{i \in \mathbb{Z}}$  ein stochastischer Prozess mit dem diskreten Zustandsraum  $\mathcal{X} \times \mathcal{Y}$ . Für die gemeinsame Randverteilung von  $X$  und  $Y$  wird analog zu oben  $P\{X_{i_1} = x_1, \dots, X_{i_k} = x_k, Y_{j_1} = y_1, \dots, Y_{j_l} = y_l\} = P^{X_{i_1}, \dots, X_{i_k}, Y_{j_1}, \dots, Y_{j_l}}(\{(x_1, \dots, x_k, y_1, \dots, y_l)\})$  geschrieben, wobei auch hier  $k, l \in \mathbb{N}$ ,  $i_1 < i_2 < \dots < i_k$  und  $j_1 < j_2 < \dots < j_l$  aus  $\mathbb{Z}$  beliebig sind.

Die Prozesse  $X$  und  $Y$  sind genau dann *stochastisch unabhängig*, wenn jede Randverteilung von  $(X, Y)$  als Produktmaß, bestehend aus den Randverteilungen von  $X$  und  $Y$ , geschrieben werden kann, siehe zum Beispiel [Bauer (1991), Behnen & Neuhaus (1995), Billingsley (1995)]:

$$P^{X_{i_1}, \dots, X_{i_k}, Y_{j_1}, \dots, Y_{j_l}} = P^{X_{i_1}, \dots, X_{i_k}} \otimes P^{Y_{j_1}, \dots, Y_{j_l}}. \quad (2.5)$$

Aufgrund der diskreten Zustandsräume ist dies äquivalent zu

$$\begin{aligned} & P\{X_{i_1} = x_1, \dots, X_{i_k} = x_k, Y_{j_1} = y_1, \dots, Y_{j_l} = y_l\} \\ &= P\{X_{i_1} = x_1, \dots, X_{i_k} = x_k\} \cdot P\{Y_{j_1} = y_1, \dots, Y_{j_l} = y_l\}, \end{aligned} \quad (2.6)$$

für alle Zustände aus  $\mathcal{X}$  und  $\mathcal{Y}$ , siehe insbesondere [Behnen & Neuhaus (1995)].

Ein Spezialfall ist die stochastische Unabhängigkeit des Prozesses  $X$  zum Zeitpunkt  $i$  vom Prozess  $Y$  zum Zeitpunkt  $j$ . Diese ist gegeben durch

$$P^{X_i, Y_j} = P^{X_i} \otimes P^{Y_j} \quad (2.7)$$

$$\iff P\{X_i = x, Y_j = y\} = P\{X_i = x\} \cdot P\{Y_j = y\} \quad (2.8)$$

und soll im Folgenden Gegenstand der Untersuchung sein.

Für die Produktabbildung  $(X_i, Y_j)$  stellt  $P\{X_i = x, Y_j = y\}$  die wahre Verteilung dar. Zur Quantifizierung der gegenseitigen Abhängigkeit wird die Modellvorstellung zugrunde gelegt, dass  $X_i$  und  $Y_j$  stochastisch unabhängig sind. Somit lautet die *a priori* Verteilung  $Q\{X_i = x, Y_j = y\} = P\{X_i = x\} \cdot P\{Y_j = y\}$ . Werden diese Verteilungen in Gl. (2.3) eingesetzt, wobei über alle  $x$  und  $y$  zu mitteln ist, so folgt für diese spezielle Kullback-Entropie

$$M(X_i, Y_j) = \sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{Y}} P\{X_i = x, Y_j = y\} \log \frac{P\{X_i = x, Y_j = y\}}{P\{X_i = x\} \cdot P\{Y_j = y\}}. \quad (2.9)$$

Diese spezielle Kullback-Entropie ist nichts anderes als die *gegenseitige Information* zwischen  $X_i$  und  $Y_j$  [Cover & Thomas (1991), Jumarie (1990), Shannon & Weaver (1949)]. In der Informationstheorie werden die Werte dieses Maßes als Information, die  $X_i$  und  $Y_j$  miteinander gemeinsam haben, interpretiert. Um zeitverzögerte Abhängigkeiten zu untersuchen, wird häufig die *Zeitverschiebung*  $i - j = \tau$  eingeführt und  $M(X_i, Y_{i+\tau})$  als Funktion von  $\tau$  studiert [Fraser & Swinney (1986), Prichard & Theiler (1995)].

Unter Ausnutzung der Identität  $P\{X_i = x\} = \sum_{y \in \mathcal{Y}} P\{X_i = x, Y_j = y\}$  lässt sich die gegenseitige Information als Summe von Shannon-Entropien schreiben:

$$M(X_i, Y_j) = H(X_i) + H(Y_j) - H(X_i, Y_j). \quad (2.10)$$

Offensichtlich ist die gegenseitige Information symmetrisch in ihren Argumenten,  $M(X_i, Y_j) = M(Y_j, X_i)$ .

Für unabhängige Prozesse ist die gegenseitige Information Null und damit minimal. Der entgegengesetzte Fall liegt vor, wenn  $X_i$  und  $Y_j$  durch eine Abbildung miteinander verknüpft sind, zum Beispiel  $Y_j = f \circ X_i$ . Dann gilt

$$\begin{aligned} P\{X_i = x, Y_j = y\} &= P\{X_i = x, X_i = f^{-1}(y)\} = P\{X_i \in (\{x\} \cap \{f^{-1}(y)\})\} \\ &= P\{X_i = x\} \delta(y, f(x)), \\ P\{Y_j = y\} &= \sum_{\tilde{x} \in \mathcal{X}} P\{X_j = \tilde{x}\} \delta(y, f(\tilde{x})), \end{aligned}$$

wobei  $\delta(\cdot, \cdot)$  das Kronecker- $\delta$  ist. Dies eingesetzt in Gl. (2.9) liefert für die gegenseitige Information:

$$\begin{aligned} M(X_i, f \circ X_i) &= - \sum_{x \in \mathcal{X}} P\{X_i = x\} \log \left( \sum_{\tilde{x} \in \mathcal{X}} P\{X_i = \tilde{x}\} \delta(f(x), f(\tilde{x})) \right) \\ &\leq H(X_i). \end{aligned} \quad (2.11)$$

Die zweite Summe in Gl. (2.11) reduziert sich genau dann zu  $P\{X_i = x\}$ , wenn  $f$  injektiv ist. In diesem Fall wird die Ungleichung zur Gleichung und die gegenseitige Information ist maximal, also gleich der Shannon-Entropie von  $X_i$ .

### 2.1.2 Dynamische Eigenschaften diskreter Prozesse

Die dynamische Struktur eines stochastischen Prozesses  $X$  ist durch seine Übergangsverteilung  $P^{X_{i+1}|X_i=x_i, X_{i-1}=x_{i-1}, \dots}$  gegeben. Sie gibt die Verteilung von  $X_{i+1}$  an, wenn  $X_i, X_{i-1}$  usw. die Zustände  $x_i, x_{i-1}$  usw. besetzt haben.

Im Allgemeinen müssen unendlich viele Zeitpunkte in der Vergangenheit berücksichtigt werden, um  $X$  vollständig zu beschreiben. Bei zeitlich abnehmenden Korrelationen innerhalb des Prozesses reicht häufig die Berücksichtigung von nur endlich vielen Zuständen aus der Vergangenheit aus, um die Dynamik des Prozesses ausreichend gut approximieren zu können,

$$P^{X_{i+1}|X_i=x_i, X_{i-1}=x_{i-1}, \dots} \approx P^{X_{i+1}|X_i=x_i, \dots, X_{i-k+1}=x_{i-k+1}}. \quad (2.12)$$

Prozesse, bei denen die linke und rechte Seite exakt gleich sind, heißen Markov-Prozesse  $k$ -ter Ordnung. Man spricht häufig auch von einer Markovkette  $k$ -ter Ordnung.

Im Folgenden wird  $P\{X_{i+1} = x_{i+1} | \mathbf{X}_i^{(k)} = \mathbf{x}_i^{(k)}\}$  für die Wahrscheinlichkeit  $P^{X_{i+1} | \mathbf{X}_i^{(k)} = \mathbf{x}_i^{(k)}}(\{x_{i+1}\})$ , mit der  $X_{i+1}$  unter den Nebenbedingungen  $X_i = x_i, \dots, X_{i-k+1} = x_{i-k+1}$  in  $x_{i+1}$  anzutreffen ist, geschrieben. Zur kürzeren Schreibweise wurde  $(X_i, \dots, X_{i-k+1})$  mit  $\mathbf{X}_i^{(k)}$  und  $(x_i, \dots, x_{i-k+1})$  mit  $\mathbf{x}_i^{(k)}$  abgekürzt. Die Dimension der Vektoren  $\mathbf{X}_i^{(k)}$  und  $\mathbf{x}_i^{(k)}$  ist jeweils  $k$ . Des Weiteren werden häufig die Beziehung

$$P\{\mathbf{X}_{i+1}^{(k+1)} = \mathbf{x}_{i+1}^{(k+1)}\} = P\{X_{i+1} = x_{i+1} | \mathbf{X}_i^{(k)} = \mathbf{x}_i^{(k)}\} \cdot P\{\mathbf{X}_i^{(k)} = \mathbf{x}_i^{(k)}\} \quad (2.13)$$

zwischen den endlich-dimensionalen Randverteilungen und den Übergangsverteilungen sowie die Identität

$$\begin{aligned} \sum_{x_i \in \mathcal{X}} P\{\dots, X_{i+1} = x_{i+1}, X_i = x_i, X_{i-1} = x_{i-1}, \dots\} \\ = P\{\dots, X_{i+1} = x_{i+1}, X_{i-1} = x_{i-1}, \dots\} \end{aligned} \quad (2.14)$$

verwendet [Behnen & Neuhaus (1995), Bauer (1991)].

Da es sich bei der Übergangsverteilung ebenfalls um Wahrscheinlichkeitsmaße handelt, kann wie in Abschn. 2.1.1 die Unsicherheit eines Zustandswechsels bei gegebenen vorherigen Zuständen durch  $-\log P\{X_{i+1} = x_{i+1} | \mathbf{X}_i^{(k)} = \mathbf{x}_i^{(k)}\}$  definiert werden. Dies eingesetzt in die Shannon-Entropie, Gl. (2.2), und zusätzliches Mitteln über alle vergangenen Zustände  $\mathbf{x}_i^{(k)}$  ergibt die *bedingte Shannon-Entropie*,

siehe beispielsweise [Cover & Thomas (1991), Jumarie (1990)]

$$\begin{aligned}
& H(X_{i+1}|\mathbf{X}_i^{(k)}) \\
&= - \sum_{\mathbf{x}_i^{(k)} \in \mathcal{X}^k} P\{\mathbf{X}_i^{(k)} = \mathbf{x}_i^{(k)}\} \sum_{x_{i+1} \in \mathcal{X}} P\{X_{i+1} = x_{i+1} | \mathbf{X}_i^{(k)} = \mathbf{x}_i^{(k)}\} \\
&\quad \times \log P\{X_{i+1} = x_{i+1} | \mathbf{X}_i^{(k)} = \mathbf{x}_i^{(k)}\} \\
&= - \sum_{\mathbf{x}_{i+1}^{(k+1)} \in \mathcal{X}^{k+1}} P\{\mathbf{X}_{i+1}^{(k+1)} = \mathbf{x}_{i+1}^{(k+1)}\} \\
&\quad \times \log P\{X_{i+1} = x_{i+1} | \mathbf{X}_i^{(k)} = \mathbf{x}_i^{(k)}\} \tag{2.15}
\end{aligned}$$

$$= H(\mathbf{X}_{i+1}^{(k+1)}) - H(\mathbf{X}_i^{(k)}). \tag{2.16}$$

Für die beiden letzten Gleichungen in Gl. (2.15) und Gl. (2.16) wurde die Relation Gl. (2.13) ausgenutzt.

Analog zur Shannon-Entropie ist die bedingte Shannon-Entropie ein Maß für die Unvorhersagbarkeit des Prozesses bezüglich seiner Dynamik. Man kann zeigen, dass für alle Prozesse, auch solche die nicht markovsche sind, die *Entropierate*

$$\lim_{k \rightarrow \infty} H(X_{i+1}|\mathbf{X}_i^{(k)}) = H(X_{i+1}|\mathbf{X}_i^{(\infty)})$$

existiert, siehe auch [Jumarie (1990), Kolmogorov (1993)]. Für Markov-Prozesse  $k$ -ter Ordnung gilt stets  $H(X_{i+1}|\mathbf{X}_i^{(l)}) = H(X_{i+1}|\mathbf{X}_i^{(k)})$  für alle  $l \geq k$ .

Wird  $X_{i+1} = X$  und  $X_i = Y$  gesetzt, so lässt sich die gegenseitige Information, Gl. (2.9), mit Hilfe der bedingten Shannon-Entropie umschreiben in

$$M(X, Y) = H(X) - H(X|Y), \tag{2.17}$$

wofür Gl. (2.16) in Gl. (2.10) eingesetzt wurde. Hierbei gibt  $H(X)$  die Unsicherheit an, mit der  $X$  beobachtet wird und  $H(X|Y)$  die Unsicherheit von  $X$  wenn der Zustand von  $Y$  bekannt ist. Somit ist  $M(X, Y)$  die Abnahme der Unsicherheit von  $X$ , also der Informationsgewinn aufgrund der zusätzlichen Beobachtung von  $Y$ .

Wird ein unbekannter Prozess beobachtet, dann muss zur Modellierung seiner Dynamik oft eine *a priori* Übergangswahrscheinlichkeit  $Q^{X_{i+1}|\mathbf{X}_i^{(k)}=\mathbf{x}_i^{(k)}}$  statt der wahren Übergangswahrscheinlichkeit  $P^{X_{i+1}|\mathbf{X}_i^{(k)}=\mathbf{x}_i^{(k)}}$  angenommen werden. Dies führt zu einer Erhöhung der mittleren Unsicherheit (Entropie). Als Maß für den Informationsverlust aufgrund der fehlerhaften Annahme kann analog zur Kullback-Entropie (Gl. (2.3)) und mit der gleichen Argumentation wie bei der bedingten Shannon-Entropie die *bedingte Kullback-Entropie* konstruiert werden

[Cover & Thomas (1991), Jumarie (1990)]:

$$K_{P||Q}(X_{i+1}|\mathbf{X}_i^{(k)}) = \sum_{\mathbf{x}_{i+1}^{(k+1)} \in \mathcal{X}^{k+1}} P\{\mathbf{X}_{i+1}^{(k+1)} = \mathbf{x}_{i+1}^{(k+1)}\} \log \frac{P\{X_{i+1} = x_{i+1} | \mathbf{X}_i^{(k)} = \mathbf{x}_i^{(k)}\}}{Q\{X_{i+1} = x_{i+1} | \mathbf{X}_i^{(k)} = \mathbf{x}_i^{(k)}\}}, \quad (2.18)$$

wobei auch hier gefordert werden muss,  $Q\{X_{i+1} = x_{i+1} | \mathbf{X}_i^{(k)} = \mathbf{x}_i^{(k)}\} > 0$  gilt, falls  $P\{X_{i+1} = x_{i+1} | \mathbf{X}_i^{(k)} = \mathbf{x}_i^{(k)}\} > 0$  ist, damit dieser Abstand existiert. Die bedingte Kullback-Entropie lässt sich in eine Differenz von Kullback-Entropien zerlegen, wozu lediglich Gl. (2.13) und Gl. (2.14) anzuwenden sind,

$$K_{P||Q}(X_{i+1}|\mathbf{X}_i^{(k)}) = K_{P||Q}(\mathbf{X}_{i+1}^{(k+1)}) - K_{P||Q}(\mathbf{X}_i^{(k)}).$$

Wie die Kullback-Entropie ist auch die bedingte nichtnegativ und genau dann Null, wenn *a priori* und *wahre* Übergangsverteilung identisch sind. Dies folgt ebenfalls aus der Log-Summen-Ungleichung (Gl. (2.4)), wobei jetzt lediglich über die Zustände  $x_{i+1}$  zu summieren ist.

Als Nächstes soll ein stochastisches System betrachtet werden, das in zwei Subsysteme zerlegt werden kann. Diese Subsysteme werden im Folgenden durch die diskreten stochastischen Prozesse  $X$  und  $Y$  beschrieben. Somit wird das gesamte System mit dem Produktprozess  $(X, Y)$  modelliert. Die Zustandsräume  $\mathcal{X}$  und  $\mathcal{Y}$  beider Subsysteme können dabei unterschiedlich sein. Basierend auf der bedingten Kullback-Entropie soll nun eine Methode vorgestellt werden, die es ermöglicht, Abhängigkeiten in der Dynamik zwischen beiden Subsystemen zu quantifizieren.

Zunächst wird angenommen, dass die Dynamik des gesamten Systems durch die Übergangsverteilung  $P^{X_{i+1}, Y_{j+1} | \mathbf{X}_i^{(k)} = \mathbf{x}_i^{(k)}, \mathbf{Y}_j^{(l)} = \mathbf{y}_j^{(l)}}$  vollständig beschrieben werden kann. Durch einfaches Ausintegrieren folgt, dass die Dynamik von  $X$  durch die Übergangsverteilung  $P^{X_{i+1} | \mathbf{X}_i^{(k)} = \mathbf{x}_i^{(k)}, \mathbf{Y}_j^{(l)} = \mathbf{y}_j^{(l)}}$  und die von  $Y$  durch  $P^{Y_{j+1} | \mathbf{X}_i^{(k)} = \mathbf{x}_i^{(k)}, \mathbf{Y}_j^{(l)} = \mathbf{y}_j^{(l)}}$  gegeben ist.

Angenommen, der Prozess  $Y$  koppelt nicht in  $X$  (bzw.  $Y$  treibt nicht  $X$  an), dann ist zu erwarten, dass der zukünftige Zustand  $X_{i+1}$  nur von den  $k$  letzten Zuständen von  $X_i, \dots, X_{i-k+1}$  abhängt, nicht aber von den  $l$  Zuständen  $Y_j, \dots, Y_{j-l+1}$  von  $Y$ . Da hier vorausgesetzt wird, dass die Dynamik und insbesondere die Kopplung kausal ist, gilt im Folgenden immer  $j \leq i$ . Das bedeutet, dass  $X$  in Bezug auf  $Y$  die verallgemeinerte Markov-Eigenschaft

$$P^{X_{i+1} | \mathbf{X}_i^{(k)} = \mathbf{x}_i^{(k)}, \mathbf{Y}_j^{(l)} = \mathbf{y}_j^{(l)}} = P^{X_{i+1} | \mathbf{X}_i^{(k)} = \mathbf{x}_i^{(k)}} \quad (2.19)$$

erfüllt, also dass für alle Zustände  $x_{i+1}, \dots, x_{i-k+1} \in \mathcal{X}$  und  $y_j, \dots, y_{j-l+1} \in \mathcal{Y}$

$$P\{X_{i+1} = x_{i+1} | \mathbf{X}_i^{(k)} = \mathbf{x}_i^{(k)}, \mathbf{Y}_j^{(l)} = \mathbf{y}_j^{(l)}\} = P\{X_{i+1} = x_{i+1} | \mathbf{X}_i^{(k)} = \mathbf{x}_i^{(k)}\} \quad (2.20)$$

ist. Gilt Gl. (2.19) bzw. Gl. (2.20) für zwei stochastische Prozesse  $X$  und  $Y$ , so wird gesagt,  $X$  ist *stochastisch ungekoppelt von  $Y$* , siehe auch [Behnen & Neuhaus (1995)]. Ist Gl. (2.19) nicht erfüllt, so *koppelt  $Y$  in  $X$* . Ist andererseits Gl. (2.19) erfüllt, so bedeutet dies, dass zur Beschreibung von  $X$  der Prozess  $Y$  keine weiteren Informationen liefert. In Abb. 2.1 ist die Kopplung zweier stochastischer Prozesse schematisch dargestellt.

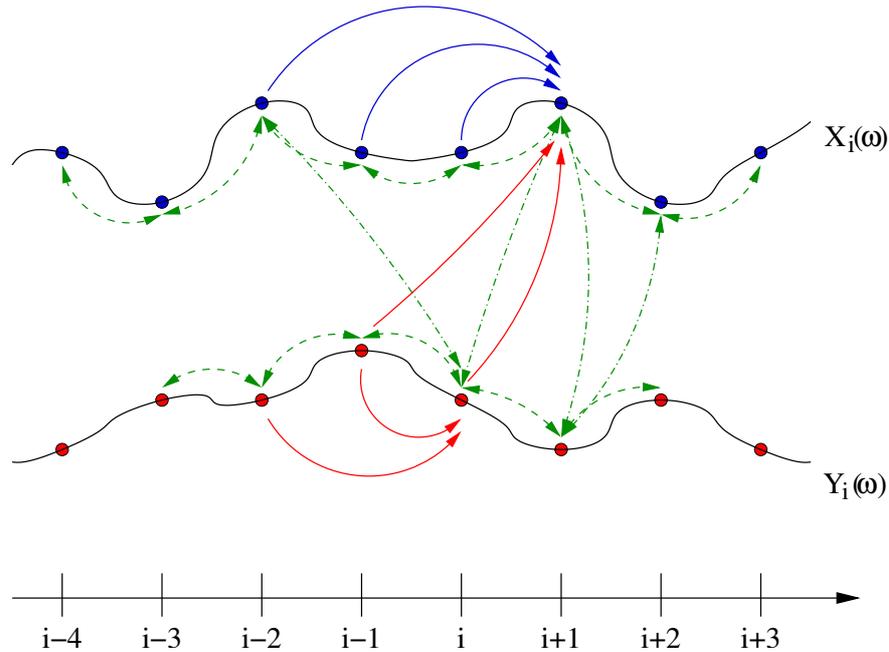


Abbildung 2.1: Schematische Darstellung der Abhängigkeit und Kopplung von stochastischen Prozessen. Die gestrichelten Pfeile symbolisieren die gegenseitige Abhängigkeit der Zustände zum Beispiel aufgrund von Korrelation. (Nicht alle Abhängigkeiten sind eingetragen.). Die durchgezogenen Pfeile symbolisieren den Einfluss, den die vergangenen Zustände durch Kopplung auf den zukünftigen haben.

Liegt eine stochastische Kopplung vor, zum Beispiel durch eine physikalische Wechselwirkung, so kann diese quantifiziert werden, indem in Gl. (2.18) als *a priori* Übergangsverteilung  $Q\{X_{i+1} = x_{i+1} | \mathbf{X}_i^{(k)} = \mathbf{x}_i^{(k)}, \mathbf{Y}_j^{(l)} = \mathbf{y}_j^{(l)}\} = P\{X_{i+1} = x_{i+1} | \mathbf{X}_i^{(k)} = \mathbf{x}_i^{(k)}\}$  eingesetzt wird. Die so erhaltene bedingte Kullback-Entropie

lautet

$$\begin{aligned}
T(X_{i+1}|\mathbf{X}_i^{(k)}, \mathbf{Y}_j^{(l)}) &= \sum_{\mathbf{x}_{i+1}^{(k+1)} \in \mathcal{X}^{k+1}} \sum_{\mathbf{y}_j^{(l)} \in \mathcal{Y}^l} P\{X_{i+1} = x_{i+1}, \mathbf{X}_i^{(k)} = \mathbf{x}_i^{(k)}, \mathbf{Y}_j^{(l)} = \mathbf{y}_j^{(l)}\} \\
&\quad \times \log \frac{P\{X_{i+1} = x_{i+1} | \mathbf{X}_i^{(k)} = \mathbf{x}_i^{(k)}, \mathbf{Y}_j^{(l)} = \mathbf{y}_j^{(l)}\}}{P\{X_{i+1} = x_{i+1} | \mathbf{X}_i^{(k)} = \mathbf{x}_i^{(k)}\}} \quad (2.21)
\end{aligned}$$

mit  $j \leq i$  und wird *Transferentropie* genannt [Schreiber (2000b)]. Als Spezialfall der Kullback-Entropie liefert sie nichtnegative Werte und ist genau dann Null, wenn Gl. (2.20) erfüllt ist. Dabei können ihre Werte als Grad für die Abweichung der Verteilungen in Gl. (2.20) voneinander interpretiert werden, weshalb sie ein Maß für die Abhängigkeit der Dynamiken von  $X$  und  $Y$  darstellt.

Die Transferentropie lässt sich als Differenz von bedingten Shannon-Entropien

$$T(X_{i+1}|\mathbf{X}_i^{(k)}, \mathbf{Y}_j^{(l)}) = H(X_{i+1}|\mathbf{X}_i^{(k)}) - H(X_{i+1}|\mathbf{X}_i^{(k)}, \mathbf{Y}_j^{(l)}). \quad (2.22)$$

Somit misst sie den Grad der Abweichung zwischen der Unsicherheit in der Dynamik von  $X$ , wenn das Wissen von  $Y$  aus der Vergangenheit zusätzlich mit berücksichtigt wird. Folglich gibt die Transferentropie die Menge an Information an, die man über die Dynamik von  $X$  erhält, wenn man zusätzlich Kenntnis über  $Y$  hat. Da andererseits dieser Informationsgewinn ein Maß für den Einfluss von  $Y$  auf  $X$  aufgrund der Kopplung ist, kann die Transferentropie als Informationsfluss von  $Y$  nach  $X$  interpretiert werden. Im Gegensatz zur gegenseitigen Information ist die Transferentropie explizit nichtsymmetrisch bezüglich der Vertauschung von  $X$  und  $Y$ . Aus diesem Grund kann mit ihr die Richtung der Abhängigkeit und somit des Informationsflusses festgestellt werden.

Sind  $\mathbf{X}_{i+1}^{(k+1)}$  und  $\mathbf{Y}_j^{(l)}$  stochastisch unabhängig,  $P^{\mathbf{X}_{i+1}^{(k+1)}, \mathbf{Y}_j^{(l)}} = P^{\mathbf{X}_{i+1}^{(k+1)}} \otimes P^{\mathbf{Y}_j^{(l)}}$ , dann folgt mit Gl. (2.13)

$$\begin{aligned}
&P\{X_{i+1} = x_{i+1} | \mathbf{X}_i^{(k)} = \mathbf{x}_i^{(k)}\} \cdot P\{\mathbf{X}_i^{(k)} = \mathbf{x}_i^{(k)}\} \cdot P\{\mathbf{Y}_j^{(l)} = \mathbf{y}_j^{(l)}\} \\
&= P\{\mathbf{X}_{i+1}^{(k+1)} = \mathbf{x}_{i+1}^{(k+1)}\} \cdot P\{\mathbf{Y}_j^{(l)} = \mathbf{y}_j^{(l)}\} = P\{\mathbf{X}_{i+1}^{(k+1)} = \mathbf{x}_{i+1}^{(k+1)}, \mathbf{Y}_j^{(l)} = \mathbf{y}_j^{(l)}\} \\
&= P\{X_{i+1} = x_{i+1} | \mathbf{X}_i^{(k)} = \mathbf{x}_i^{(k)}, \mathbf{Y}_j^{(l)} = \mathbf{y}_j^{(l)}\} \\
&\quad \times P\{\mathbf{X}_i^{(k)} = \mathbf{x}_i^{(k)} | \mathbf{Y}_j^{(l)} = \mathbf{y}_j^{(l)}\} \cdot P\{\mathbf{Y}_j^{(l)} = \mathbf{y}_j^{(l)}\}.
\end{aligned}$$

Da diese Gleichung für alle Zustände gelten muss, erhält man durch Ausintegrieren von  $X_{i+1}$  zunächst  $P\{\mathbf{X}_i^{(k)} = \mathbf{x}_i^{(k)} | \mathbf{Y}_j^{(l)} = \mathbf{y}_j^{(l)}\} = P\{\mathbf{X}_i^{(k)} = \mathbf{x}_i^{(k)}\}$  und anschließend, dass  $X$  von  $Y$  stochastisch ungekoppelt ist, also Gl. (2.20). Somit

kann, wenn  $\mathbf{X}_{i+1}^{(k+1)}$  und  $\mathbf{Y}_j^{(l)}$  voneinander unabhängig sind, darauf geschlossen werden, dass  $X$  von  $Y$  ungekoppelt ist. Dies kann in der Implikation

$$M(\mathbf{X}_{i+1}^{(k+1)}, \mathbf{Y}_j^{(l)}) = 0 \implies T(X_{i+1} | \mathbf{X}_i^{(k)}, \mathbf{Y}_j^{(l)}) = 0, \quad (2.23)$$

zwischen der gemeinsamen Information und der Transferentropie zusammen gefasst werden. Der Umkehrschluss ist falsch, so reicht  $M(X_{i+1}, \mathbf{Y}_j^{(l)}) \neq 0$  oder  $M(X_{i+1}, Y_{j-\tau}) \neq 0$ ,  $\tau = 0, 1, \dots, l-1$  nicht aus, um auf Kopplung schließen zu können. Siehe hierzu auch das Beispiel in Absch. 2.1.3.

Demzufolge werden mit *stochastisch unabhängig* und *stochastisch ungekoppelt* zwei nicht äquivalente Eigenschaften zwischen stochastischen Prozessen beschrieben. Der Hauptunterschied basiert vor allem darauf, dass bei der stochastischen Kopplung die Abhängigkeit zwischen  $X_{i+1}$  und  $\mathbf{Y}_j^{(l)}$  ohne den Einfluss (Informationsfluss) von  $\mathbf{Y}_j^{(l)}$  auf  $X_{i+1}$  über  $\mathbf{X}_i^{(k)}$  betrachtet wird, siehe Abb. 2.1 für eine schematische Darstellung von Abhängigkeit und Kopplung. Dies wird vor allem durch die Äquivalenz zwischen Transferentropie und bedingter Transferinformation deutlich:

$$T(X_{i+1} | \mathbf{X}_i^{(k)}, \mathbf{Y}_j^{(l)}) = 0 \iff M(X_{i+1}, \mathbf{Y}_j^{(l)} | \mathbf{X}_i^{(k)}) = 0, \quad (2.24)$$

wobei die *bedingte Transferinformation*  $M(X_{i+1}, \mathbf{Y}_j^{(l)} | \mathbf{X}_i^{(k)})$ , oft auch *bedingte Transinformation* genannt, durch

$$\begin{aligned} M(X_{i+1}, \mathbf{Y}_j^{(l)} | \mathbf{X}_i^{(k)}) &= \sum_{\mathbf{x}_{i+1}^{(k+1)} \in \mathcal{X}^{k+1}} \sum_{\mathbf{y}_j^{(l)} \in \mathcal{Y}^l} P\{\mathbf{X}_{i+1}^{(k+1)} = \mathbf{x}_{i+1}^{(k+1)}, \mathbf{Y}_j^{(l)} = \mathbf{y}_j^{(l)}\} \\ &\times \log \frac{P\{X_{i+1} = x_{i+1}, \mathbf{Y}_j^{(l)} = \mathbf{y}_j^{(l)} | \mathbf{X}_i^{(k)} = \mathbf{x}_i^{(k)}\}}{P\{X_{i+1} = x_{i+1} | \mathbf{X}_i^{(k)} = \mathbf{x}_i^{(k)}\} \cdot P\{\mathbf{Y}_j^{(l)} = \mathbf{y}_j^{(l)} | \mathbf{X}_i^{(k)} = \mathbf{x}_i^{(k)}\}} \end{aligned} \quad (2.25)$$

definiert ist und die Abhängigkeit von  $X_{i+1}$  und  $\mathbf{Y}_j^{(l)}$  ohne Korrelationen über  $\mathbf{X}_i^{(k)}$  quantifiziert [Jumarie (1990), Palus (1996a)].

Die Äquivalenz in Gl. (2.24) ergibt sich unmittelbar aus Gl. (2.14), da für alle Zustände  $x_{i+1}$ ,  $\mathbf{x}_i^{(k)}$  und  $\mathbf{y}_j^{(l)}$

$$\begin{aligned} &P\{X_{i+1} = x_{i+1}, \mathbf{Y}_j^{(l)} = \mathbf{y}_j^{(l)} | \mathbf{X}_i^{(k)} = \mathbf{x}_i^{(k)}\} \cdot P\{\mathbf{X}_i^{(k)} = \mathbf{x}_i^{(k)}\} \\ &= P\{X_{i+1} = x_{i+1}, \mathbf{X}_i^{(k)} = \mathbf{x}_i^{(k)}, \mathbf{Y}_j^{(l)} = \mathbf{y}_j^{(l)}\} \\ &= P\{X_{i+1} = x_{i+1} | \mathbf{X}_i^{(k)} = \mathbf{x}_i^{(k)}, \mathbf{Y}_j^{(l)} = \mathbf{y}_j^{(l)}\} \\ &\quad \times P\{\mathbf{Y}_j^{(l)} = \mathbf{y}_j^{(l)} | \mathbf{X}_i^{(k)} = \mathbf{x}_i^{(k)}\} \cdot P\{\mathbf{X}_i^{(k)} = \mathbf{x}_i^{(k)}\} \end{aligned}$$

gelten muss. Hieraus folgt für die Übergangsverteilungen

$$\begin{aligned} &P\{X_{i+1} = x_{i+1}, \mathbf{Y}_j^{(l)} = \mathbf{y}_j^{(l)} | \mathbf{X}_i^{(k)} = \mathbf{x}_i^{(k)}\} \\ &= P\{X_{i+1} = x_{i+1} | \mathbf{X}_i^{(k)} = \mathbf{x}_i^{(k)}, \mathbf{Y}_j^{(l)} = \mathbf{y}_j^{(l)}\} \cdot P\{\mathbf{Y}_j^{(l)} = \mathbf{y}_j^{(l)} | \mathbf{X}_i^{(k)} = \mathbf{x}_i^{(k)}\} \end{aligned}$$

und somit Gl. (2.24). Wird  $P\{X_{i+1} = x_{i+1}, \mathbf{Y}_j^{(l)} = \mathbf{y}_j^{(l)} | \mathbf{X}_i^{(k)} = \mathbf{x}_i^{(k)}\}$  in Gl. (2.25) eingesetzt, so erhält man sogar die Gleichheit von Transferentropie und bedingter Transferinformation

$$T(X_{i+1} | \mathbf{X}_i^{(k)}, \mathbf{Y}_j^{(l)}) = M(X_{i+1}, \mathbf{Y}_j^{(l)} | \mathbf{X}_i^{(k)}). \quad (2.26)$$

An dieser Stelle sei noch einmal explizit darauf hingewiesen, dass zur Quantifizierung der Kopplung mittels Transferentropie  $k$  mindestens so groß gewählt werden muss, dass jeglicher indirekte Einfluss von  $\mathbf{Y}_j^{(l)}$  auf  $X_{i+1}$  über  $\mathbf{X}_i^{(k)}$  unterbunden ist. Handelt es sich bei  $X$  um einen Markov-Prozess  $m$ -ter Ordnung, so ist  $k \geq m$  zu wählen. Andernfalls kann die Transferentropie  $T(X_{i+1} | \mathbf{X}_i^{(k)}, \mathbf{Y}_j^{(l)})$  auch positive Werte liefern, obwohl  $Y$  nicht in  $X$  koppelt. Auf diese Problematik wird bei der Bestimmung der Kopplungsrichtung zwischen zwei Punktprozessen in dem Beispiel von Absch. 5.5 noch einmal eingegangen.

Mit Gl. (2.16) kann die Transferentropie als Summe von Shannon-Entropien umgeschrieben werden:

$$\begin{aligned} T(X_{i+1} | \mathbf{X}_i^{(k)}, \mathbf{Y}_j^{(l)}) \\ = H(\mathbf{X}_i^{(k)}, \mathbf{Y}_j^{(l)}) - H(\mathbf{X}_{i+1}^{(k+1)}, \mathbf{Y}_j^{(l)}) + H(\mathbf{X}_{i+1}^{(k+1)}) - H(\mathbf{X}_i^{(k)}). \end{aligned} \quad (2.27)$$

Diese Beziehung wird später noch benötigt. Des Weiteren kann sie in eine Summe von gegenseitigen Informationen zerlegt werden:

$$\begin{aligned} T(X_{i+1} | \mathbf{X}_i^{(k)}, \mathbf{Y}_j^{(l)}) &= M(\mathbf{X}_{i+1}^{(k+1)}, \mathbf{Y}_j^{(l)}) - M(\mathbf{X}_i^{(k)}, \mathbf{Y}_j^{(l)}) \\ &= M(X_{i+1}, (\mathbf{X}_i^{(k)}, \mathbf{Y}_j^{(l)})) - M(X_{i+1}, \mathbf{X}_i^{(k)}). \end{aligned} \quad (2.28)$$

### 2.1.3 Diskretes Beispiel

An einem einfachen Beispiel, das einem erlaubt, alle Verteilungen zu berechnen, soll das unterschiedliche Verhalten von gegenseitiger Information und Transferentropie studiert werden. Hierzu werden zwei stationäre Markovprozesse  $X$  und  $Y$  betrachtet, die nur die Werte 0 oder 1 annehmen,  $\mathcal{X} = \mathcal{Y} = \{0, 1\}$ . Der Prozess  $Y$  ist autonom und wechselt bei jedem Zeitschritt  $i$  seinen Zustand:

$$P\{Y_{i+1} = 1 | Y_i = 0\} = P\{Y_{i+1} = 0 | Y_i = 1\} = 1.$$

Im ungekoppelten Fall ( $c = 0$ ) macht auch der Prozess  $X$  einen Zustandswechsel, allerdings nur mit Wahrscheinlichkeit  $1/2$ ,

$$P\{X_{i+1} = 1 | X_i = 0\} = P\{X_{i+1} = 0 | X_i = 1\} = 1/2.$$

Wird die Kopplung  $c > 0$  eingeschaltet, so soll die Dynamik von  $X$  zusätzlich von den Zuständen von  $Y$  beeinflusst werden, wobei bei voller Kopplungsstärke ( $c = 1$ )  $X$  den vorherigen Zustand von  $Y$  einnehmen soll, das heißt  $X$  synchronisiert

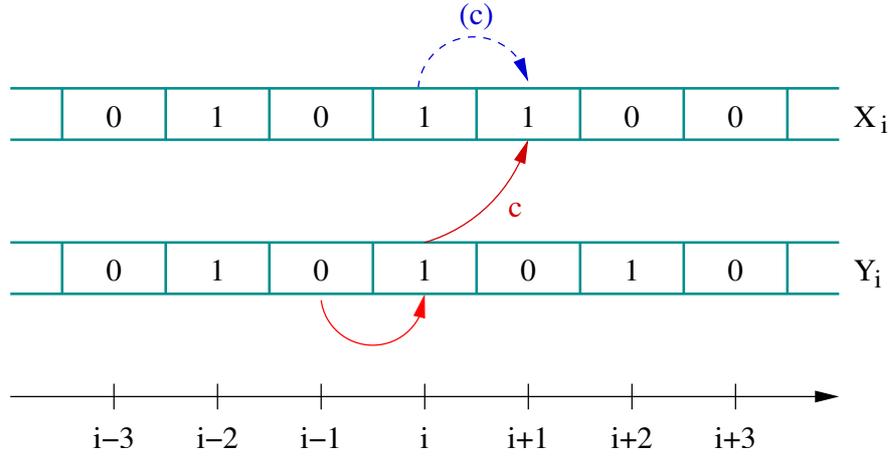


Abbildung 2.2: Zeitreihen der binären stochastischen Prozesse  $X$  und  $Y$ .  $Y$  ist autonom und koppelt in  $X$ .

mit  $Y$  ( $X_{i+1} = Y_i$ ). Für  $0 < c < 1$  soll  $X_{i+1} = Y_i$  mit Wahrscheinlichkeit  $c$  sein und mit Wahrscheinlichkeit  $1-c$  soll  $X_{i+1}$  zufällig und unabhängig von  $Y$  gewählt werden. Zur Illustration der Prozesse siehe Abb. 2.2. Dementsprechend lautet die Übergangsverteilung des Systems

$$\begin{aligned} P\{X_{i+1} = x_{i+1}, Y_{i+1} = y_{i+1} | X_i = x_i, Y_i = y_i\} \\ = (1-c) \cdot P\{X_{i+1} = x_{i+1} | X_i = x_i\} \cdot P\{Y_{i+1} = y_{i+1} | Y_i = y_i\} \\ + c \cdot P\{Y_{i+1} = y_{i+1} | Y_i = y_i\} \cdot \delta(x_{i+1}, y_i). \end{aligned}$$

Hieraus folgt für die jeweilige Dynamik von  $X$  und  $Y$  im System

$$P\{X_{i+1} = x_{i+1} | X_i = x_i, Y_i = y_i\} = (1-c) \cdot P\{X_{i+1} = x_{i+1} | X_i = x_i\} + c \cdot \delta(x_{i+1}, y_i),$$

$$P\{Y_{i+1} = y_{i+1} | X_i = x_i, Y_i = y_i\} = P\{Y_{i+1} = y_{i+1} | Y_i = y_i\}.$$

Als Nächstes werden die stationären Lösungen dieses Systems gesucht. Mit Gl. (2.13) folgt, dass hierzu die Gleichungen

$$\begin{aligned} P\{X_i = x, Y_i = y\} &\equiv P\{X_{i+1} = x, Y_{i+1} = y\} \\ &= \sum_{x_i=0}^1 \sum_{y_i=0}^1 P\{X_{i+1} = x_{i+1}, Y_{i+1} = y_{i+1} | X_i = x_i, Y_i = y_i\} \\ &\quad \times P\{X_i = x_i, Y_i = y_i\}, \end{aligned}$$

für  $x, y = 0, 1$  unter den Nebenbedingungen

$$\begin{aligned} \sum_{x=0}^1 \sum_{y=0}^1 P\{X_i = x, Y_i = y\} &= 1, \\ \sum_{x=0}^1 \sum_{y=0}^1 P\{X_{i+1} = x, Y_{i+1} = y | X_i = x_i, Y_i = y_i\} &= 1. \end{aligned}$$

nach  $P\{X_{i+1} = x, Y_i = y\}$  aufzulösen sind. Die Lösungen lauten

$$\begin{aligned} P\{X_i = 0, Y_i = 0\} &= P\{X_i = 1, Y_i = 1\} = (1 - c)/4, \\ P\{X_i = 0, Y_i = 1\} &= P\{X_i = 1, Y_i = 0\} = c/4. \end{aligned}$$

Zusammen mit Gl. (2.13) und Gl. (2.14) lassen sich hieraus alle anderen benötigten Verteilungen berechnen. Insbesondere erhält man aufgrund der Kopplung die folgenden Übergangverteilungen  $P\{X_{i+1} = x_{i+1} | X_i = x_i\}$ :

$$\begin{aligned} P\{X_{i+1} = 0 | X_i = 0\} &= P\{X_{i+1} = 1 | X_i = 1\} = \frac{(1 - c)(1 + c)}{4}, \\ P\{X_{i+1} = 1 | X_i = 0\} &= P\{X_{i+1} = 0 | X_i = 1\} = \frac{(1 - c)^2 + 2c}{2}. \end{aligned}$$

Des Weiteren folgt  $P\{X_i = 1\} = P\{Y_i = 1\} = 1/2$ . Da  $Y$  zu jedem Zeitschritt zwischen den Zuständen 0 und 1 mit Wahrscheinlichkeit 1 wechselt, folgt mit der letzten Gleichung, dass  $Y$  ein stochastischer Prozess ist, dessen Dynamik deterministisch ist.

Werden diese Verteilungen in Gl. (2.9) bzw. in Gl. (2.21) eingesetzt, so erhält man für die gemeinsamen Informationen:

$$\begin{aligned} M(X_{i+1}, Y_i) &= M(Y_{i+1}, X_i) = M(X_i, Y_i) \\ &= \frac{1}{2} \{(1 + c) \log(1 + c) + (1 - c) \log(1 - c)\}. \end{aligned}$$

und für die Transferentropien:

$$\begin{aligned} T(X_{i+1} | X_i, Y_i) &= \frac{c}{2} \{(1 + c) \log(1 + c) - (1 - c) \log(1 - c)\} \\ &\quad - \frac{1 + c^2}{2} \log(1 + c^2), \\ T(Y_{i+1} | Y_i, X_i) &= 0. \end{aligned}$$

Die gemeinsamen Informationen und Transferentropien sind in Abb. 2.3 als Funktion der Kopplungskonstante  $c$  dargestellt. Da die Pfade von  $Y$  periodisch sind,

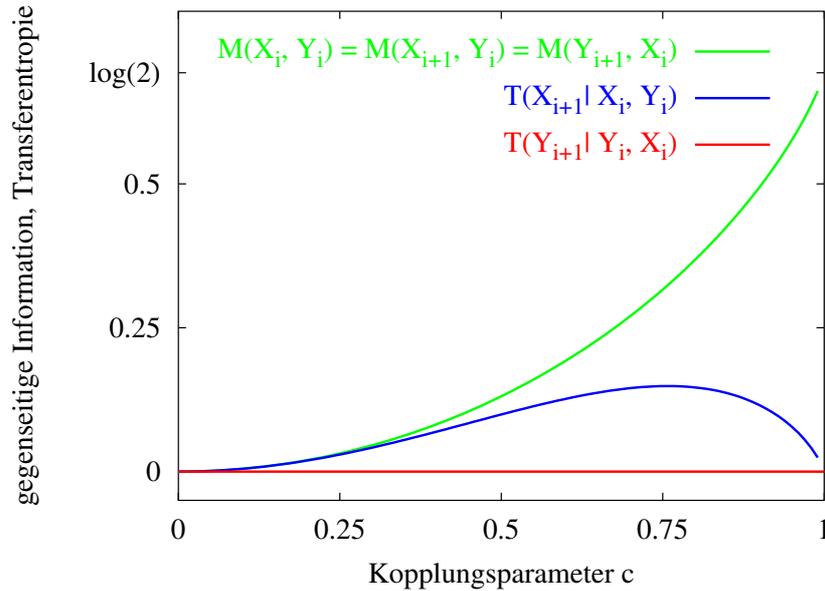


Abbildung 2.3: Gegenseitige Information und Transferentropie als Funktion des Kopplungsparameters  $c$  für zwei binäre stochastische Prozesse  $X$  und  $Y$ , wobei  $Y$  autonom ist und stochastisch in  $X$  koppelt.

sind die gleichzeitige gegenseitige Information  $M(X_i, Y_i)$  sowie die zeitverzögerte  $M(X_{i+1}, Y_i)$  und  $M(Y_{i+1}, X_i)$  für alle Kopplungsstärken  $c$  identisch. Somit kann die zeitverzögerte gegenseitige Information allein nicht als Indikator für die Richtung des Informationstransports bzw. der Kopplung verwendet werden, was aber häufig in der Literatur [Kaneko (1986), Vastano & Swinney (1988)] zu finden ist. So deutet in dem Beispiel die gemeinsame Information darauf hin, dass Information mit derselben Rate in beide Richtungen ausgetauscht wird. Dies ist aber falsch, denn aufgrund der kausalen Natur kann  $Y$  nicht durch  $X$  beeinflusst werden. Korrekt wird dies hingegen von der Transferentropie  $T(Y_{i+1}|Y_i, X_i)$  wiedergegeben, welche identisch Null ist. Außer für  $c = 0$ , ungekoppelter Fall, und  $c = 1$ , maximale Kopplung, ist die zur Richtung  $Y \rightarrow X$  zugehörige Transferentropie  $T(X_{i+1}|X_i, Y_i)$  positiv. Bei maximaler Kopplung tritt vollständige Synchronisation ein. In solch einem Fall ist es nicht möglich, nachträglich festzustellen, ob eine Kopplung vorliegt oder ob  $X$  ebenfalls autonom und periodisch ist. Insbesondere ist dieser Fall ein Beispiel für stochastische Abhängigkeit ( $M(X_i, Y_i) > 0$ ) ohne Kopplung ( $T(X_{i+1}|X_i, Y_i) = 0$ ), siehe auch Gl. (2.23).

## 2.2 Kontinuierliche stochastische Prozesse

### 2.2.1 Verteilung kontinuierlicher Prozesse

In diesem Abschnitt sollen stochastische Prozesse mit kontinuierlichen Zuständen untersucht werden, das heißt der Zustandsraum ist reellwertig und nicht abzählbar (überabzählbar). Wie zuvor sei der stochastische Prozess  $X = (X_i)_{i \in \mathbb{Z}}$  auf dem Wahrscheinlichkeitsraum  $(\Omega, \mathcal{A}, P)$  definiert, siehe auch [Bauer (1991), Behnen & Neuhaus (1995), Billingsley (1995)]. Für jedes  $i \in \mathbb{Z}$  sind dessen Zufallsvariablen  $X_i$  reellwertig und messbar:

$$X_i : (\Omega, \mathcal{A}, P) \longrightarrow (\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d)).$$

$\mathcal{B}(\mathbb{R}^d)$  ist die Borelsche  $\sigma$ -Algebra auf  $\mathbb{R}^d$ , mit  $d \in \mathbb{N}$ .

Da der Zustandsraum der hier betrachteten Prozesse nicht länger abzählbar ist, können Größen wie Shannon-Entropie, gegenseitige Information und Transferentropie nicht mehr direkt durch die Wahrscheinlichkeiten der einzelnen Zustände ausgedrückt werden. Um dennoch ähnliche Maße zum Beispiel für Kopplung zu erhalten, muss die Untersuchung von kontinuierlichen Prozessen auf solche beschränkt werden, deren Verteilungen eine Lebesgue-Dichte besitzen [Cover & Thomas (1991), Jumarie (1990)]. Andernfalls besteht die Möglichkeit, durch nachträgliche Diskretisierung die Prozesse auf diskrete Prozesse abzubilden. Dies liefert Mehrdeutigkeiten, die aber unter bestimmten Umständen zusätzliche Aussagekraft besitzen [Gaspard & Wang (1993)]. Ein Beispiel hierfür ist die Kolmogorov-Sinai-Entropie, die eine spezielle Form der bedingten Shannon-Entropie darstellt. Mit ihr ist es möglich, Aussagen über die Vorhersagbarkeit von deterministischen Systemen, welche keine Lebesgue-Dichte besitzen, zu machen. Die Kolmogorov-Sinai-Entropie wird in Absch. 4 vorgestellt.

Ein stochastischer Prozess  $X$  besitzt genau dann eine Wahrscheinlichkeitsdichte bezüglich des Lebesgue-Maßes [Bauer (1992)], wenn jede endlich-dimensionale Randverteilung  $P^{X_{i_1}, \dots, X_{i_k}}$  von  $X$  mit  $i_1 < \dots < i_k \in \mathbb{Z}$  als

$$P^{X_{i_1}, \dots, X_{i_k}}(B) = \int_B g_{X_{i_1}, \dots, X_{i_k}}(x_1, \dots, x_k) dx_1 \dots dx_k \quad \forall B \in \mathcal{B}(\mathbb{R}^{d \cdot k}) \quad (2.29)$$

geschrieben werden kann. Dabei ist  $g_{X_{i_1}, \dots, X_{i_k}}$  eine nichtnegative, reellwertige, messbare Funktion. Sie stellt die Lebesgue-Dichte der Randverteilung dar und wird *Wahrscheinlichkeitsdichte* oder kurz *Dichte* genannt [Bauer (1992), Bauer (1991), Behnen & Neuhaus (1995), Billingsley (1995)]. Die hier verwendeten Dichten beziehen sich stets auf das Lebesgue-Maß. Aus Gl. (2.29) folgt die Identität

$$\begin{aligned} \int g_{\dots, X_{i_{k-1}}, X_{i_k}, X_{i_{k+1}}, \dots}(\dots, x_{k-1}, x_k, x_{k+1}, \dots) dx_k \\ = g_{\dots, X_{i_{k-1}}, X_{i_{k+1}}, \dots}(\dots, x_{k-1}, x_{k+1}, \dots), \end{aligned} \quad (2.30)$$

welche hier die Rolle der Identität (2.14) übernimmt.

Entsprechend Gl. (2.29) lässt sich die Dichte  $g_{X_i}(x)$  als Wahrscheinlichkeit pro Volumeneinheit interpretieren, mit der der Prozess  $X$  zur Zeit  $i$  im Volumenelement  $dx$  um  $x$  beobachtet wird. Demnach ist es naheliegend, die Unsicherheit, mit der  $X_i$  in einer infinitesimalen Umgebung um  $x$  anzutreffen ist, als  $-\log g_{X_i}(x)$  zu definieren [Shannon & Weaver (1949), Jumarie (1990), Cover & Thomas (1991)]. Die Mittelung dieser Unsicherheit über alle Zustände liefert die *kontinuierliche Shannon-Entropie*

$$H(X_i) = - \int \log g_{X_i}(x) P^{X_i}(dx) = - \int g_{X_i}(x) \log g_{X_i}(x) dx, \quad (2.31)$$

mit der die Verteilung  $P^{X_i}$  charakterisiert werden kann. In der Literatur ist auch die Bezeichnung *differentielle Entropie* gebräuchlich, zum Beispiel in [Cover & Thomas (1991)].

Im Gegensatz zur Shannon-Entropie kann die kontinuierliche Shannon-Entropie auch negative Werte annehmen. Ist zum Beispiel  $X_i$  auf dem Intervall  $[0, a)$  gleichverteilt, das heißt  $g_{X_i}(x) = 1/a$  falls  $x \in [0, a)$  und sonst 0, so ist  $H(X_i) = \log a$ , also negativ, wenn  $0 < a < 1$  ist. Die physikalische Interpretation der kontinuierlichen Shannon-Entropie ist im Allgemeinen unklar. Vor allem, da jeder kontinuierliche Prozess so umskaliert werden kann, dass seine Entropie stets Null ist. Zu diesem Zweck muss, falls  $d = 1$  ist,  $X_i$  lediglich mit  $\exp(-H(X_i))$  multipliziert werden. Die Addition einer Konstanten zu  $X_i$  ändert hingegen die Entropie nicht. Auf die Transformationseigenschaften wird in Absch. 2.2.2 näher eingegangen.

Die Quantifizierung der Abweichung einer *a priori* Verteilung  $Q^{X_i}$  von der *wahren* Verteilung  $P^{X_i}$  kann bei kontinuierlichen Prozessen ebenfalls über die Dichten erfolgen. Denn die beiden Verteilungen sind genau dann gleich, wenn deren Dichten fast sicher gleich sind [Bauer (1992), Billingsley (1995)]. Der Grad der Abweichung zwischen *wahrer* Dichte  $g_{X_i}$  und *a priori* Dichte  $\tilde{g}_{X_i}$  ist durch die *kontinuierliche Kullback-Entropie*, oft auch *relative Entropie* genannt, gegeben:

$$K_{P||Q}(X_i) = \int g_{X_i}(x) \log \frac{g_{X_i}(x)}{\tilde{g}_{X_i}(x)} dx. \quad (2.32)$$

Damit die kontinuierliche Kullback-Entropie existiert, muss zusätzlich gefordert werden, dass  $\tilde{g}_{X_i}(x) > 0$  für alle  $x \in S = \{x' \in \mathbb{R}^d : g_{X_i}(x') > 0\}$ .

Die kontinuierliche Kullback-Entropie ist nichtnegativ, denn

$$\begin{aligned} K_{P||Q}(X_i) &= - \int g_{X_i}(x) \log \frac{\tilde{g}_{X_i}(x)}{g_{X_i}(x)} = - \int_S g_{X_i}(x) \log \frac{\tilde{g}_{X_i}(x)}{g_{X_i}(x)} \\ &\geq - \log \left( \int_S g_{X_i}(x) \frac{\tilde{g}_{X_i}(x)}{g_{X_i}(x)} dx \right) = - \log \left( \int_S \tilde{g}_{X_i}(x) dx \right) \\ &\geq \log 1 = 0, \end{aligned}$$

siehe auch [Cover & Thomas (1991)]. Die erste Ungleichung folgt dabei aus der Jensenschen Ungleichung. Sie besagt, dass für jede konvexe Funktion  $f$  die Ungleichung  $E[f(X_i)] \geq f(E[X_i])$  erfüllt ist, siehe zum Beispiel [Bauer (1991)]. Der Erwartungswert von  $X_i$  ist mit  $E[X_i] = \int x P^{X_i}$  bezeichnet worden. Somit ist für die erste Ungleichung lediglich zu berücksichtigen, dass  $-\log(x)$  eine konvexe Funktion ist.

Bereits in Absch. 2.1 wurde erwähnt, dass zwei beliebige stochastische Prozesse  $X$  und  $Y$  genau dann stochastisch unabhängig sind, wenn sich die gemeinsame Verteilung als Produktmaß der endlich-dimensionalen Randverteilungen von  $X$  und  $Y$  schreiben lässt, Gl. (2.6). Insbesondere sind die Systemzustände  $X_i$  und  $Y_j$  stochastisch unabhängig, wenn Gl. (2.7) gilt. Besitzen die Verteilungen von  $X$  und  $Y$  Dichten, so erhält man man die Äquivalenz

$$P^{X_i, Y_j} = P^{X_i} \otimes P^{Y_j} \iff g_{X_i, Y_j}(x, y) = g_{X_i}(x) \cdot g_{Y_j}(y), \quad (2.33)$$

das heißt die gemeinsame Dichte von  $(X_i, Y_j)$  kann als Produkt der Dichten von  $X_i$  und  $Y_j$  geschrieben werden [Bauer (1991), Behnen & Neuhaus (1995), Billingsley (1995)]. In Bezug auf Unabhängigkeit ist Gl. (2.33) somit bei kontinuierlichen Prozessen das Analogon zu Gl. (2.6).

Wird  $g_{X_i} \cdot g_{Y_j}$  als *a priori* Dichte in Gl. (2.32) eingesetzt, so erhält man als Maß für die Unabhängigkeit die *kontinuierliche gegenseitige Information* [Shannon & Weaver (1949), Cover & Thomas (1991), Jumarie (1990)]:

$$M(X_i, Y_j) = \int \int g_{X_i, Y_j}(x, y) \log \frac{g_{X_i, Y_j}(x, y)}{g_{X_i}(x) \cdot g_{Y_j}(y)} dx dy. \quad (2.34)$$

Sie ist nichtnegativ und es gilt  $M(X_i, Y_j) = 0$ , falls  $X_i$  und  $Y_j$  unabhängig sind. Des Weiteren kann die kontinuierliche gegenseitige Information als Summe von kontinuierlichen Shannon-Entropien geschrieben werden, wozu lediglich Gl. (2.30) anzuwenden ist. Dies führt ebenfalls auf Gl. (2.10).

Um die kontinuierliche gegenseitige Information berechnen zu können, ist es nicht ausreichend, wenn die Verteilungen von  $X$  und  $Y$  eine Dichte haben, sondern auch die Verteilungen des Produktprozesses  $(X, Y)$  muss eine besitzen. Somit dürfen  $X$  und  $Y$  nicht durch eine Abbildung gekoppelt sein, zum Beispiel  $Y_j = f \circ X_i$ . Auch wenn die Dichte  $g_{X_i}$  glatt ist, so gilt dies nicht für die gemeinsame Dichte  $g_{X_i, Y_j}$ ; sie ist eine  $\delta$ -Distribution

$$g_{X_i, Y_j}(x, y) = g_{X_i}(x) \delta(y - f(x)).$$

Hieraus folgt zwar, dass  $g_{Y_j}(y) = \int g_{X_i}(x) \delta(y - f(x)) dx$  eine reellwertige Funktion ist, werden aber alle Dichten in Gl. (2.34) eingesetzt, so bleibt stets ein nicht definierter  $\log \delta(\cdot)$ -Term im Integranden der kontinuierlichen gegenseitigen Information stehen. Insbesondere wird der Grenzfall  $M(X_i, Y_j) = H(X_i)$  nicht erreicht.

Sowohl kontinuierliche Shannon-Entropie, Kullback-Entropie als auch gegenseitige Information lassen sich auf endlich-dimensionale Randverteilungen verallgemeinern. Hierzu ist zum Beispiel für die Shannon-Entropie  $H(X_{i_1}, \dots, X_{i_k})$  die Dichte  $g_{X_i}$  in Gl. (2.31) durch  $g_{X_{i_1}, \dots, X_{i_k}}$  zu ersetzen und anschließend über alle Zustandsvektoren  $x = (x_1, \dots, x_k)$  zu mitteln. Entsprechendes gilt für die Kullback-Entropie und die gegenseitige Information.

## 2.2.2 Dynamische Eigenschaften kontinuierlicher Prozesse

Zur Untersuchung der dynamischen Eigenschaften eines stochastischen Prozesses müssen seine Übergangsverteilungen studiert werden. Wie in Abschn. 2.1.2, Gl. (2.12) soll auch hier angenommen werden, dass endlich viele Zustände aus der Vergangenheit genügen, um die Dynamik der kontinuierlichen Prozesse ausreichend gut zu approximieren. Des Weiteren wird vorausgesetzt, dass alle betrachteten Übergangsverteilungen eine Lebesgue-Dichte besitzen,

$$P^{X_{i+1}|\mathbf{X}_i^{(k)}=\mathbf{x}_i^{(k)}}(B) = \int_B g_{X_{i+1}|\mathbf{X}_i^{(k)}}(x_{i+1}|\mathbf{x}_i^{(k)}) dx_{i+1} \quad \forall B \in \mathcal{B}(\mathbb{R}^d). \quad (2.35)$$

Bei  $\mathbf{X}_i^{(k)}$  und  $\mathbf{x}_i^{(k)}$  handelt es sich um die in Abschn. 2.1.2 eingeführten Abkürzungen für die Produktabbildungen und Zustandsvektoren.

Da für die folgenden Betrachtungen die Existenz der Übergangsdichten vorausgesetzt werden muss, sind hier deterministische Prozesse auszuschließen. Denn ihre zukünftigen Zustände sind durch ihre Vergangenheit festgelegt, weshalb ihre Übergangsdichten Diracsche Delta-Distributionen darstellen,  $g_{X_{i+1}|\mathbf{X}_i^{(k)}}(x_{i+1}|\mathbf{x}_i^{(k)}) = \delta(x_{i+1} - \varphi(\mathbf{x}_i^{(k)}))$ , wobei  $\varphi$  die deterministische Abbildung ist.

Zwischen den Randverteilungen und den Übergangsverteilungen eines jeden stochastischen Prozesses gilt

$$\begin{aligned} P^{\mathbf{X}_{i+1}^{(k+1)}}(B_{i+1} \times B_i \times \dots \times B_{i-k+1}) \\ = \int_{B_i \times \dots \times B_{i-k+1}} \int_{B_{i+1}} P^{X_{i+1}|\mathbf{X}_i^{(k)}=\mathbf{x}_i^{(k)}}(dx_{i+1}) P^{\mathbf{X}_i^{(k)}}(d\mathbf{x}_i^{(k)}) \end{aligned} \quad (2.36)$$

für beliebige  $B_{i+1}, \dots, B_{i-k+1} \in \mathcal{B}(\mathbb{R}^d)$ , siehe [Behnen & Neuhaus (1995)]. Nach dem Eindeutigkeitssatz für Maße [Bauer (1992), Behnen & Neuhaus (1995), Billingsley (1995)] sowie dem Eindeutigkeitssatz für Dichten [Bauer (1992), Bauer (1991)] folgt nach Einsetzen der jeweiligen Dichten, dass Gl. (2.36) äquivalent zu

$$g_{\mathbf{X}_{i+1}^{(k+1)}}(\mathbf{x}_{i+1}^{(k+1)}) = g_{X_{i+1}|\mathbf{X}_i^{(k)}}(x_{i+1}|\mathbf{x}_i^{(k)}) \cdot g_{\mathbf{X}_i^{(k)}}(\mathbf{x}_i^{(k)}) \quad \text{fast sicher} \quad (2.37)$$

ist. Diese und ähnliche Gleichungen werden im Folgenden wie Gl. (2.13) im Diskreten verwendet.

Die Charakterisierung der Übergangswahrscheinlichkeit erfolgt wie bei den diskreten Prozessen mit der bedingten Shannon-Entropie, wobei jetzt die Unsicherheiten durch die Dichten der Übergangsverteilungen gegeben sind. Unter Ausnutzen der in Gl. (2.37) gegebenen Relation folgt somit die *bedingte kontinuierliche Shannon-Entropie*

$$H(X_{i+1}|\mathbf{X}_i^{(k)}) = \int g_{\mathbf{X}_{i+1}^{(k+1)}}(\mathbf{x}_{i+1}^{(k+1)}) \log g_{X_{i+1}|\mathbf{X}_i^{(k)}}(x_{i+1}|\mathbf{x}_i^{(k)}) d\mathbf{x}_{i+1}^{(k+1)}. \quad (2.38)$$

Sie lässt sich als Differenz von kontinuierlichen Shannon-Entropien schreiben. Dies führt wiederum auf Gl. (2.16). Ebenso lässt sich die kontinuierliche gegenseitige Information in der Form von Gl. (2.17) darstellen. Auch hier kann gezeigt werden, dass  $\lim_{k \rightarrow \infty} H(X_{i+1}|\mathbf{X}_i^{(k)}) = H(X_{i+1}|\mathbf{X}_i^{(\infty)})$  existiert, und dass  $H(X_{i+1}|\mathbf{X}_i^{(l)}) = H(X_{i+1}|\mathbf{X}_i^{(k)})$  für alle  $l \geq k$  ist, wenn  $X$  ein Markov-Prozess  $k$ -ter Ordnung ist.

Die Güte, mit welcher die *a priori* Dichte  $\tilde{g}_{X_{i+1}|\mathbf{X}_i^{(k)}}(x_{i+1}|\mathbf{x}_i^{(k)})$  die *wahre* Dichte  $g_{X_{i+1}|\mathbf{X}_i^{(k)}}(x_{i+1}|\mathbf{x}_i^{(k)})$  approximiert, kann mit der *bedingten kontinuierlichen Kullback-Entropie*

$$K_{P||Q}(X_{i+1}|\mathbf{X}_i^{(k)}) = \int g_{\mathbf{X}_{i+1}^{(k+1)}}(\mathbf{x}_{i+1}^{(k+1)}) \log \frac{g_{X_{i+1}|\mathbf{X}_i^{(k)}}(x_{i+1}|\mathbf{x}_i^{(k)})}{\tilde{g}_{X_{i+1}|\mathbf{X}_i^{(k)}}(x_{i+1}|\mathbf{x}_i^{(k)})} d\mathbf{x}_{i+1}^{(k+1)} \quad (2.39)$$

quantifiziert werden. Auch hier muss  $\tilde{g}_{X_{i+1}|\mathbf{X}_i^{(k)}}(x|\mathbf{x}_i^{(k)}) > 0$  für alle  $x \in \{x' \in \mathbb{R}^d : g_{X_{i+1}|\mathbf{X}_i^{(k)}}(x'|\mathbf{x}_i^{(k)}) > 0\}$  gefordert werden, damit diese bedingte Kullback-Entropie existiert.

Analog zu der kontinuierlichen Kullback-Entropie kann gezeigt werden, dass die bedingte nichtnegativ ist. Des Weiteren folgt  $K_{P||Q}(X_{i+1}) = 0$ , falls  $g_{X_{i+1}|\mathbf{X}_i^{(k)}} = \tilde{g}_{X_{i+1}|\mathbf{X}_i^{(k)}}$  fast sicher ist.

Aufbauend auf der bedingten kontinuierlichen Kullback-Entropie, soll nun eine Version der Transferentropie abgeleitet werden, um die stochastische Kopplung zwischen den kontinuierlichen Prozessen  $X$  und  $Y$  zu studieren. Der Prozess  $X$  ist dabei genau dann von  $Y$  stochastisch ungekoppelt, wenn die Übergangsverteilungen von  $X$  nicht von den Zuständen von  $Y$  abhängen, also wenn Gl. (2.19)

erfüllt ist. Somit gilt im ungekoppelten Fall für jedes  $B \in \mathcal{B}(\mathbb{R}^d)$  die Gleichung

$$\begin{aligned} \int_B g_{X_{i+1}|\mathbf{X}_i^{(k)}, \mathbf{Y}_j^{(l)}}(x_{i+1}|\mathbf{x}_i^{(k)}, \mathbf{y}_j^{(l)}) dx_{i+1} \\ = P^{X_{i+1}|\mathbf{X}_i^{(k)}=\mathbf{x}_i^{(k)}, \mathbf{Y}_j^{(l)}=\mathbf{y}_j^{(l)}}(B) = P^{X_{i+1}|\mathbf{X}_i^{(k)}=\mathbf{x}_i^{(k)}}(B) \\ = \int_B g_{X_{i+1}|\mathbf{X}_i^{(k)}}(x_{i+1}|\mathbf{x}_i^{(k)}) \cdot dx_{i+1} \end{aligned}$$

Der Eindeutigkeitsatz für Dichten und der für Maße [Bauer (1992)] liefern dann, dass Gl. (2.19) äquivalent zu

$$g_{X_{i+1}|\mathbf{X}_i^{(k)}, \mathbf{Y}_j^{(l)}}(x_{i+1}|\mathbf{x}_i^{(k)}, \mathbf{y}_j^{(l)}) = g_{X_{i+1}|\mathbf{X}_i^{(k)}}(x_{i+1}|\mathbf{x}_i^{(k)}) \quad (2.40)$$

ist. Wird somit  $g_{X_{i+1}|\mathbf{X}_i^{(k)}}$  als *a priori* Dichte in Gl. (2.39) eingesetzt, so erhält man die *kontinuierliche Transferentropie*

$$\begin{aligned} T(X_{i+1}|\mathbf{X}_i^{(k)}, \mathbf{Y}_j^{(l)}) = \int \int g_{\mathbf{X}_{i+1}^{(k+1)}, \mathbf{Y}_j^{(l)}}(\mathbf{x}_{i+1}^{(k+1)}, \mathbf{y}_j^{(l)}) \\ \times \log \frac{g_{X_{i+1}|\mathbf{X}_i^{(k)}, \mathbf{Y}_j^{(l)}}(x_{i+1}|\mathbf{x}_i^{(k)}, \mathbf{y}_j^{(l)})}{g_{X_{i+1}|\mathbf{X}_i^{(k)}}(x_{i+1}|\mathbf{x}_i^{(k)})} d\mathbf{x}_{i+1}^{(k+1)} d\mathbf{y}_j^{(l)}. \quad (2.41) \end{aligned}$$

Mit der gleichen Technik, mit der Gl. (2.37) abgeleitet wurde, kann gezeigt werden, dass

$$\begin{aligned} g_{X_{i+1}, \mathbf{Y}_j^{(l)}|\mathbf{X}_i^{(k)}}(x_{i+1}, \mathbf{y}_j^{(l)}|\mathbf{x}_i^{(k)}) \\ = g_{X_{i+1}|\mathbf{X}_i^{(k)}, \mathbf{Y}_j^{(l)}}(x_{i+1}|\mathbf{x}_i^{(k)}, \mathbf{y}_j^{(l)}) \cdot g_{\mathbf{Y}_j^{(l)}|\mathbf{X}_i^{(k)}}(\mathbf{y}_j^{(l)}|\mathbf{x}_i^{(k)}) \quad (2.42) \end{aligned}$$

fast sicher gilt. Wird dies in Gl. (2.41) eingesetzt, so folgt unmittelbar, dass die kontinuierliche Transferentropie identisch mit der *bedingten kontinuierlichen Transferinformation*

$$\begin{aligned} M(X_{i+1}, \mathbf{Y}_j^{(l)}|\mathbf{X}_i^{(k)}) = \int \int g_{\mathbf{X}_{i+1}^{(k+1)}, \mathbf{Y}_j^{(l)}}(\mathbf{x}_{i+1}^{(k+1)}, \mathbf{y}_j^{(l)}) \\ \times \frac{g_{X_{i+1}, \mathbf{Y}_j^{(l)}|\mathbf{X}_i^{(k)}}(x_{i+1}, \mathbf{y}_j^{(l)}|\mathbf{x}_i^{(k)})}{g_{X_{i+1}|\mathbf{X}_i^{(k)}}(x_{i+1}|\mathbf{x}_i^{(k)}) \cdot g_{\mathbf{Y}_j^{(l)}|\mathbf{X}_i^{(k)}}(\mathbf{y}_j^{(l)}|\mathbf{x}_i^{(k)})} d\mathbf{x}_{i+1}^{(k+1)} d\mathbf{y}_j^{(l)} \quad (2.43) \end{aligned}$$

ist. Somit quantifiziert die kontinuierliche Transferentropie wie stark der zukünftige Zustand von  $X$  von den vergangenen von  $Y$  abhängt, wobei jegliche Korrelationen ausgeschlossen werden, die über die vergangenen Zustände von  $X$  weitergeleitet wurden. Insbesondere zeigt sich auch hier, dass stochastische Abhängigkeit

und stochastische Kopplung zwei verschiedene Eigenschaften eines gekoppelten Systems sind, siehe Gl. (2.23).

Die kontinuierliche Transferentropie kann ebenfalls als eine Summe von bedingten kontinuierlichen Shannon-Entropien, kontinuierlichen Shannon-Entropien oder auch gegenseitigen Informationen zerlegt werden. Hierfür sind Gl. (2.30) und Gl. (2.37) anzuwenden. Dies führt ebenfalls auf die Gleichungen Gl. (2.22), Gl. (2.27) bzw. Gl. (2.28).

### 2.2.3 Transformationsinvarianz

Die gegenseitige Information und die Transferentropie besitzen nur dann eine physikalisch relevante Aussagekraft, wenn sie mindestens gegenüber Koordinatentransformationen invariant sind. Für diskrete Prozesse ist dies offensichtlich für alle der hier vorgestellten informationstheoretischen Größen der Fall. Bei kontinuierlichen Prozessen ist dies zumindest bei der gegenseitigen Information zu erwarten, wenn die Prozesse  $X$  und  $Y$  unabhängig sind, denn stochastische Unabhängigkeit ist eine Eigenschaft, die bei Transformationen mit messbaren Abbildungen erhalten bleibt, siehe beispielsweise [Bauer (1991), Behnen & Neuhaus (1995), Billingsley (1995)]. Für alle anderen Fälle sind detaillierte Untersuchungen nötig.

Hierzu soll als erstes studiert werden, wie eine Wahrscheinlichkeitsdichte bezüglich eines  $C^1$ -Diffeomorphismus transformiert wird. Sei  $\psi$  ein  $C^1$ -Diffeomorphismus auf den  $\mathbb{R}^d$ , der  $X_i$  auf  $\psi \circ X_i$  abbildet. Für jede beliebige Borelmenge  $B \in \mathcal{B}(\mathbb{R}^d)$  gilt dann

$$P^{X_i}(B) = P^{X_i}(\psi^{-1}(\psi(B))) = P^{\psi \circ X_i}(\psi(B)).$$

Dabei ist mit  $B$  auch  $\psi(B)$  eine Borelmenge [Bauer (1992)]. Sei  $g_{X_i}$  die Wahrscheinlichkeitsdichte der Verteilung  $P^{X_i}$  und  $g_{\psi \circ X_i}$  diejenige von  $P^{\psi \circ X_i}$ , so folgt mit dem Transformationssatz für Integrale bezüglich des Lebesgue-Maßes [Bauer (1992)]

$$\int_B g_{X_i}(x) dx = \int_{\psi(B)} g_{\psi \circ X_i}(x') dx' = \int_B g_{\psi \circ X_i}(\psi(x)) \cdot |\det(D\psi)(x)| dx.$$

Dabei ist  $\det(D\psi)(x)$  die Jacobi-Determinante von  $\psi$  an der Stelle  $x$ . Da diese Gleichung für jede Borelmenge  $B \in \mathcal{B}(\mathbb{R}^d)$  gilt und die Integranden messbare Funktionen sind, erhält man die Identität

$$g_{X_i}(x) = g_{\psi \circ X_i}(\psi(x)) \cdot |\det(D\psi)(x)| \quad (2.44)$$

fast sicher [Bauer (1992)].

Einsetzen von Gl. (2.44) in Gl. (2.31) und erneute Anwendung des Transformationsatzes liefert

$$H(X_i) = H(\psi \circ X_i) - \int g_{\psi \circ X_i}(x') \cdot \log |\det(D\psi)(\psi^{-1}(x'))| dx'. \quad (2.45)$$

Folglich ändert sich bei der Koordinatentransformation  $\psi$  der Wert der Shannon-Entropie um den Erwartungswert der logarithmierten Jacobi-Determinanten von  $\psi$ ,  $E[-\log |\det(D\psi)(\psi^{-1}(\cdot))|]$ . Somit sind die Shannon-Entropiewerte immer in Bezug auf ein Koordinatensystem anzugeben [Shannon & Weaver (1949), Cover & Thomas (1991), Jumarie (1990)].

Die eigentliche Ursache, weshalb die Shannon-Entropie nicht invariant ist, wenn der Prozess mit einem  $C^1$ -Diffeomorphismus transformiert wird, liegt darin, dass die Jacobi-Determinante der Transformation im Logarithmus auftaucht. In der gegenseitigen Information besteht somit die Möglichkeit, dass sich dieser zusätzliche Term durch eine geschickte Wahl der Transformation heraushebt. Offensichtlich reicht es hierzu aus, wenn  $X_i$  mit einem  $C^1$ -Diffeomorphismus  $\psi_1$  auf  $\psi_1 \circ X_i$  und separat  $Y_j$  mit einem zweiten  $C^1$ -Diffeomorphismus  $\psi_2$  auf  $\psi_2 \circ Y_j$  abgebildet wird. Dann ist  $\psi \circ (X_i, Y_j) = (\psi_1 \circ X_i, \psi_2 \circ Y_j)$  und die Jacobi-Determinante von  $\psi$  lautet

$$\det(D\psi)(x, y) = \det(D(\psi_1, \psi_2))(x, y) = \det(D\psi_1)(x) \cdot \det(D\psi_2)(y). \quad (2.46)$$

Hieraus folgt unmittelbar die Invarianz der gegenseitigen Information bei Koordinatentransformationen

$$M(X_i, Y_j) = M(\psi_1 \circ X_i, \psi_2 \circ Y_j). \quad (2.47)$$

Als Nächstes wird untersucht, wie sich die Größen zur Charakterisierung dynamischer Eigenschaften beim Übergang von einem Koordinatensystem in ein anderes verändern. Durch eine ähnliche Argumentation wie in Gl. (2.45) folgt, dass auch die bedingte, kontinuierliche Shannon-Entropie nicht invariant bezüglich Transformationen mit  $C^1$ -Diffeomorphismen ist. Werden allerdings  $X_{i+1}$  und  $\mathbf{X}_i^{(k)}$  separat mit zwei  $C^1$ -Diffeomorphismen transformiert, das heißt  $\psi \circ (X_{i+1}, \mathbf{X}_i^{(k)}) = (\psi_1 \circ X_{i+1}, \psi_2 \circ \mathbf{X}_i^{(k)})$ , so geht in die Transformation der bedingte kontinuierliche Shannon-Entropie nur noch  $\psi_1$  ein. Denn in diesem Fall lautet die Jacobi-Determinante von  $\psi$

$$\begin{aligned} \det(D\psi)(x_{i+1}, \mathbf{x}_i^{(k)}) &= \det(D(\psi_1, \psi_2))(x_{i+1}, \mathbf{x}_i^{(k)}) \\ &= \det(D\psi_1)(x_{i+1}) \cdot \det(D\psi_2)(\mathbf{x}_i^{(k)}). \end{aligned} \quad (2.48)$$

Ausnutzen der Relationen Gl. (2.37) und Gl. (2.30) liefert unmittelbar

$$\begin{aligned} H(X_{i+1}|\mathbf{X}_i^{(k)}) &= H(\psi_1 \circ X_{i+1}|\psi_2 \circ \mathbf{X}_i^{(k)}) \\ &= \int g_{\psi_1 \circ X_{i+1}}(x') \cdot \log |\det(D\psi_1)(\psi_1^{-1}(x'))| dx'. \end{aligned} \quad (2.49)$$

Sei schließlich  $\psi_3$  ein weiterer  $C^1$ -Diffeomorphismus, der  $\mathbf{Y}_j^{(l)}$  auf  $\psi_3 \circ \mathbf{Y}_j^{(l)}$  abbildet. Dann ist die Jacobi-Determinante des  $C^1$ -Diffeomorphismus  $\psi = (\psi_1, \psi_2, \psi_3)$  gegeben durch

$$\begin{aligned} \det(D\psi)(x_{i+1}, \mathbf{x}_i^{(k)}, \mathbf{y}_j^{(l)}) &= \det(D(\psi_1, \psi_2, \psi_3))(x_{i+1}, \mathbf{x}_i^{(k)}, \mathbf{y}_j^{(l)}) \\ &= \det(D\psi_1)(x_{i+1}) \cdot \det(D\psi_2)(\mathbf{x}_i^{(k)}) \cdot \det(D\psi_3)(\mathbf{y}_j^{(l)}). \end{aligned} \quad (2.50)$$

Werden nun die Dichten der Übergangverteilung in Gl. (2.41) gemäß Gl. (2.37) durch die Dichten der Verteilungen ersetzt, so folgt mit Gl. (2.48) und Gl. (2.50) sowie mit dem Transformationssatz für Integrale die Transformationsinvarianz der kontinuierlichen Transferentropie:

$$T(X_{i+1}|\mathbf{X}_i^{(k)}, \mathbf{Y}_j^{(l)}) = T(\psi_1 \circ X_{i+1}|\psi_2 \circ \mathbf{X}_i^{(k)}, \psi_3 \circ \mathbf{Y}_j^{(l)}). \quad (2.51)$$

Somit sind gegenseitige Information und Transferentropie geeignete Größen, mit denen die stochastische Abhängigkeit bzw. die stochastische Kopplung zwischen zwei kontinuierlichen Prozessen gemessen werden kann. Beide hängen nicht von dem gewählten Koordinatensystem ab, insbesondere sind sie invariant gegenüber der Reskalierung von Daten.



# Kapitel 3

## Schätzen von Entropien und Informationen

### 3.1 Schätzen bei einer und mehreren Beobachtungen

In der Realität können Experimente nur endlich oft wiederholt werden. Des Weiteren ist die Zeitauflösung nach unten hin beschränkt. Möchte man die Transferentropie eines stochastischen Prozesses berechnen, so muss zunächst dessen Verteilung bzw. dessen Verteilungsdichte aus den  $N$  Zuständen  $x_i(n) = X_i(\omega_n)$ ,  $n = 1, \dots, N$ , die in jedem der diskreten Zeitpunkte  $i = 1, \dots, T$  beobachtet wurden, geschätzt werden. Durch die Elemente  $\omega_n \in \Omega$ ,  $n = 1, \dots, N$  des Ereignisraums  $\Omega$  sind die Realisierungen  $i \rightarrow X_i(\omega_n)$ , also die jeweiligen Wiederholungen des Experiments gegeben.

Die grundlegende Idee, mit der die Verteilung  $P^{\mathbf{X}_i^{(k)}}$  aus den beobachteten Zuständen  $\mathbf{x}_i^{(k)}(1), \dots, \mathbf{x}_i^{(k)}(N)$  geschätzt wird, ist folgende: Zunächst wird jeder beobachtete Zustand  $\mathbf{x}_i^{(k)}(n) = \mathbf{X}_i^{(k)}(\omega_n)$  mit einer messbaren Funktion  $\psi_B$  auf eine nichtnegative Zahl abgebildet,  $\mathbf{x}_i^{(k)}(n) \rightarrow \psi_B(\mathbf{x}_i^{(k)}(n))$ . Dabei soll der Erwartungswert von  $\psi_B$  für jede Ereignismenge  $B$  gleich der Wahrscheinlichkeit sein, mit der  $\mathbf{X}_i^{(k)}$  in  $B$  zu finden ist:

$$\mathbb{E}[\psi_B] = \int \psi_B(\mathbf{x}_i^{(k)}) P^{\mathbf{X}_i^{(k)}}(\mathbf{x}_i^{(k)}) = P^{\mathbf{X}_i^{(k)}}(B).$$

Sind die Beobachtungen  $\omega_1, \dots, \omega_N$  voneinander unabhängig und gleich verteilt, dann konvergiert  $\psi_B$  nach dem Gesetz der großen Zahlen, siehe [Bauer (1991), Billingsley (1995)], gemittelt über alle Beobachtungen, gegen  $P^{\mathbf{X}_i^{(k)}}(B)$ ,

$$\frac{1}{N} \sum_{n=1}^N \psi_B(\mathbf{x}_i^{(k)}(n)) \xrightarrow{N \rightarrow \infty} P^{\mathbf{X}_i^{(k)}}(B).$$

Mit dieser Methode kann auch die Dichte  $g_{\mathbf{X}_i^{(k)}}$  aus den Beobachtungen geschätzt werden. Hierfür muss  $\psi_x$  für fast jedes  $x \in \mathbb{R}^{dk}$  die Gleichung  $E[\psi_x] = g_{\mathbf{X}_i^{(k)}}(x)$  erfüllen. Folglich kann die zeitliche Entwicklung von Shannon-Entropie, gegenseitiger Information oder Transferentropie studiert werden, sofern eine ausreichende Anzahl von Beobachtungen vorliegt.

Ist andererseits nur eine Realisierung  $X_i(\omega_0)$  gegeben, so ist es unmöglich ohne weitere Annahmen über den Prozess die Verteilungen von  $X$  zu schätzen. Falls der Prozess stationär ist, das heißt es gilt

$$P^{\mathbf{X}_i^{(k)}} = P^{\mathbf{X}_{i+\tau}^{(k)}} \quad \forall \tau \in \mathbb{Z}, \quad (3.1)$$

kann die in der Zeit verschobene Realisierung  $i \rightarrow X_{i+\tau}(\omega_0)$  als eine weitere Realisierung von  $X$  betrachtet werden,  $X_i(\omega_\tau) = X_{i+\tau}(\omega_0)$ . Weil diese Realisierungen aber nicht unabhängig voneinander sind, kann das oben beschriebene Verfahren zum Schätzen der Verteilungen oder Dichten nicht angewendet werden. Stattdessen ist eine sehr viel stärkere Eigenschaft als die Stationarität zu fordern, nämlich die Ergodizität [Billingsley (1965), Walters (1981), Pollicott & Yuri (1998), Eckmann & Ruelle (1985), Katok & Hasselblatt (1995)]. Die wesentlichste Eigenschaft ergodischer Prozesse ist, dass der Erwartungswert einer beliebigen integrierbaren Funktion  $f$  bezüglich der Verteilung  $P^{\mathbf{X}_i^{(k)}}$  des Prozesses berechnet werden kann, indem über die Zeit gemittelt wird:

$$\lim_{T \rightarrow \infty} \frac{1}{T} \sum_{i=1}^T f(\mathbf{X}_i^{(k)}(\omega_0)) = \int f(\mathbf{x}_1^{(k)}) P^{\mathbf{X}_1^{(k)}}(d\mathbf{x}_1^{(k)}) = \int f(\mathbf{X}_1^{(k)}(\omega)) P(d\omega). \quad (3.2)$$

Beispielsweise lässt sich die Wahrscheinlichkeit  $P\{\mathbf{X}_1^{(k)} \in B\}$  aus Gl. (3.2) berechnen, indem  $f$  gleich der Indikatorfunktion  $\mathbf{1}_B$  gesetzt wird. Diese ist definiert durch  $\mathbf{1}_B(x) = 1$  falls  $x \in B$  und 0 sonst. Da eine Voraussetzung für Ergodizität die Stationarität ist, folgt, dass die Entropien und Informationen zeitlich konstant sind und lediglich von Zeitdifferenzen wie  $i - j$  abhängen.

Liegt nur eine Realisierung vor und ist der beobachtete Prozess nicht ergodisch, so können die Verteilungen nicht geschätzt werden. Resultiert aber dieses nichtergodische Verhalten beispielsweise aus einem sich langsam veränderlichen Parameter im System, so kann häufig der Prozess innerhalb eines kleinen Zeitfensters als "lokal" ergodisch angenommen werden. Eine notwendige Voraussetzung hierfür ist, dass  $P^{\mathbf{X}_i^{(k)}}$  sich nur langsam ändert:  $P^{\mathbf{X}_i^{(k)}} \approx P^{\mathbf{X}_{i+\tau}^{(k)}}$  für kleine  $\tau$ . In diesem Fall können Mittelungen innerhalb eines Zeitfensters um  $i$  trotz statistischer Fehler noch aussagekräftige Werte für beispielsweise gegenseitige Information oder Transferentropie liefern.

Dies soll an dem in Absch. 2.1.3 vorgestellten Beispiel eines gekoppelten diskreten Prozesses demonstriert werden. Zu diesem Zweck wird ein zeitabhängiger

Kopplungsparameter  $c$  verwendet, welcher durch

$$c(i) = \frac{1}{2} \left\{ 1 + \sin \left( \frac{2\pi}{\sqrt{1 - i/10\,000}} \right) \right\}$$

gegeben ist. In Abb. 3.1 sind der exakte und der geschätzte Verlauf von  $M(X_{i+1}, Y_i)$  bzw.  $T(X_{i+1}|X_i, Y_i)$  aufgetragen. Die zeitabhängige gegenseitige

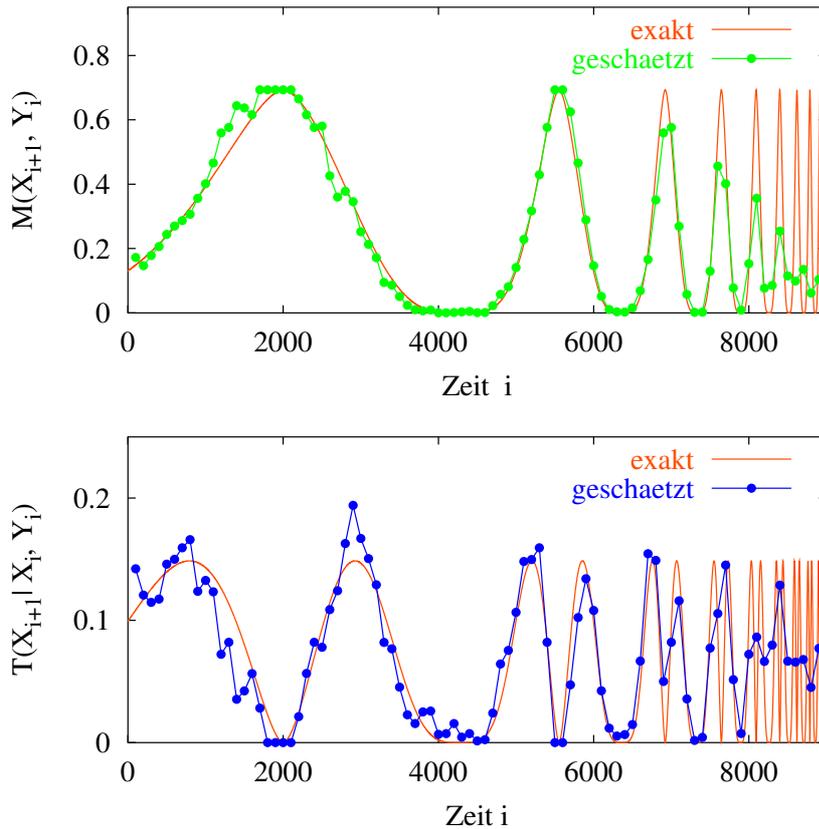


Abbildung 3.1: Gegenseitige Information  $M(X_{i+1}, Y_i)$  und Transferentropie  $T(X_{i+1}|X_i, Y_i)$  von den zwei gekoppelten diskreten stochastischen Prozessen aus Abschn. 2.1.3. Der Kopplungsparameter ist zeitabhängig.

Information und Transferentropie wurden geschätzt, indem Verteilungen wie  $P\{X_{i+1} = x, Y_i = y\}$  mit den beobachteten Werten, die innerhalb eines Zeitfensters lagen, approximiert wurden. Dabei wurde die Fensterbreite mit 200 Zeitpunkten fest gewählt. Die verwendeten Schätzer werden in Abschn. 3.2 vorgestellt.

Bis zu einem Zeitpunkt von  $i \approx 7\,000$  stimmen die genäherten Werte von  $M(X_{i+1}, Y_i)$  und  $T(X_{i+1}|X_i, Y_i)$  innerhalb statistischer Fluktuationen mit den

exakten Werte überein. Insbesondere wird das oszillatorische Verhalten beider Größen richtig wiedergegeben. Für spätere Zeiten werden die Schätzer schlechter und liefern falsche Werte. In diesem Bereich ist aufgrund der schnelleren Oszillationen von  $c(i)$  das Mittelungsfenster zu groß, so dass hier Ergodizität auch nicht näherungsweise vorliegt. Um das richtige Verhalten zu rekonstruieren, muss daher die Fensterbreite reduziert werden, was aber wegen der hieraus hervorgehenden statistischen Fluktuationen nur bis zu einem bestimmten Grad möglich ist. Die geschätzten Werte von  $M(X_i, Y_i)$  und  $M(Y_{i+1}, X_i)$  lieferten ähnliche Werte wie für  $M(X_{i+1}, Y_i)$ . Des Weiteren wurde  $T(Y_{i+1}|Y_i, X_i)$  für alle Zeit mit Null geschätzt, was mit dem Modell übereinstimmt. Zum Abschluss sei noch angemerkt, dass solch kurze Zeitreihen wie die hier verwendeten üblicherweise nur für diskrete Prozesse stabile Ergebnisse liefern.

### 3.2 Schätzer für diskrete Prozesse

Werden die diskreten stochastischen Prozesse  $X$  und  $Y$  beobachtet, so müssen zunächst deren Verteilungen rekonstruiert werden, um Entropien und Informationen berechnen zu können. Angenommen, es liegen  $N$  Beobachtungen  $\omega_1, \dots, \omega_N \in \Omega$  vor, dann können zum Beispiel die Wahrscheinlichkeiten  $P\{X_i = x_i\}$  aus den beobachteten Zuständen  $x_i(n) \equiv X_i(\omega_n)$  geschätzt werden. Ein einfaches Verfahren besteht darin, die relative Häufigkeit, mit der der Zustand  $x_i$  auftritt, zu ermitteln:

$$\hat{P}\{X_i = x_i\} = \frac{1}{N} \sum_{n=1}^N \delta(x_i, x_i(n)). \quad (3.3)$$

Nach dem Gesetz der großen Zahlen (Satz von Etemadi) [Bauer (1991)] konvergiert (3.3) gegen  $P\{X_i = x_i\}$  für  $N \rightarrow \infty$ , sofern die Beobachtungen identisch verteilt und paarweise unabhängig sind. Der entsprechende Schätzer für die endlich-dimensionalen Randverteilungen lautet

$$\hat{P}\{\mathbf{X}_i^{(k)} = \mathbf{x}_i^{(k)}\} = \frac{1}{N} \sum_{n=1}^N \prod_{j=0}^{k-1} \delta(x_{i-j}, x_{i-j}(n)). \quad (3.4)$$

Mit den in Gl. (3.3) bzw. (3.4) gegebenen Schätzern kann beispielsweise die bedingte Shannon-Entropie (2.15) berechnet werden, indem die Wahrscheinlichkeiten durch die entsprechenden Schätzer ersetzt werden. Hierzu ist zunächst die bedingte Wahrscheinlichkeit mit Gl. (2.13) auf Randverteilungen umzuschreiben. Um die Werte der Verteilungsschätzer aus der Zeitreihe zu erhalten sind lediglich Histogramme für jede benötigte Verteilung zu erstellen.

Das Erstellen von Histogrammen für Markov-Prozesse mit großer Ordnung erweist sich in der Praxis oft aufgrund des beschränkten Computerspeichers als

schwierig. Aus diesem Grund verwende ich eine alternative Herangehensweise, bei der die Verteilungen der beobachteten Zustände für jede Beobachtung erneut geschätzt werden. So kann der Erwartungswert einer beliebigen Funktion  $f$  genähert werden, indem über alle Beobachtungen gemittelt wird:

$$\hat{E}[f] = \sum_{\mathbf{x}_i^{(k)} \in \mathcal{X}^k} \hat{P}\{\mathbf{X}_i^{(k)} = \mathbf{x}_i^{(k)}\} \cdot f(\mathbf{x}_i^{(k)}) = \frac{1}{N} \sum_{n=1}^N f(\mathbf{x}_i^{(k)}(n)).$$

Für unabhängige Beobachtungen konvergiert, nach dem Gesetz der großen Zahlen, dieser Schätzer für große  $N$  gegen den Erwartungswert von  $f$ .

Da es sich bei den Entropien und Informationen ebenfalls um die Berechnung von Erwartungswerten handelt, führt dies auf die folgenden Schätzer für Shannon-Entropie, gegenseitige Information, bedingte Shannon-Entropie und Transferentropie:

$$\hat{H}(X_i) = -\frac{1}{N} \sum_{n=1}^N \log \hat{P}\{X_i = x_i(n)\}, \quad (3.5)$$

$$\hat{M}(X_i, Y_j) = \frac{1}{N} \sum_{n=1}^N \log \frac{\hat{P}\{X_i = x_i(n), Y_j = y_j(n)\}}{\hat{P}\{X_i = x_i(n)\} \cdot \hat{P}\{Y_j = y_j(n)\}}, \quad (3.6)$$

$$\hat{H}(X_{i+1} | \mathbf{X}_i^{(k)}) = -\frac{1}{N} \sum_{n=1}^N \log \frac{\hat{P}\{\mathbf{X}_{i+1}^{(k+1)} = \mathbf{x}_{i+1}^{(k+1)}(n)\}}{\hat{P}\{\mathbf{X}_i^{(k)} = \mathbf{x}_i^{(k)}(n)\}}, \quad (3.7)$$

und

$$\begin{aligned} & \hat{T}(X_{i+1} | \mathbf{X}_i^{(k)}, \mathbf{Y}_j^{(l)}) \\ &= \frac{1}{N} \sum_{n=1}^N \log \frac{\hat{P}\{\mathbf{X}_{i+1}^{(k+1)} = \mathbf{x}_{i+1}^{(k+1)}(n), \mathbf{Y}_j^{(l)} = \mathbf{y}_j^{(l)}(n)\} \cdot \hat{P}\{\mathbf{X}_i^{(k)} = \mathbf{x}_i^{(k)}(n)\}}{\hat{P}\{\mathbf{X}_i^{(k)} = \mathbf{x}_i^{(k)}(n), \mathbf{Y}_j^{(l)} = \mathbf{y}_j^{(l)}(n)\} \cdot \hat{P}\{\mathbf{X}_{i+1}^{(k+1)} = \mathbf{x}_{i+1}^{(k+1)}(n)\}}. \end{aligned} \quad (3.8)$$

Für unabhängige Beobachtungen konvergieren diese Schätzer für große  $N$  gegen die entsprechenden Entropien und Informationen.

Üblicherweise stehen nur endlich viele Beobachtungen zur Verfügung, weshalb die geschätzten Werte fluktuieren. So hat die gegenseitige Information bei endlichen Datenmengen einen positiven Bias, da sie nur nicht-negative Werte besitzt. Für die Shannon-Entropie und für die gegenseitige Information sind diesbezüglich Korrekturen entwickelt worden, beispielsweise in [Grassberger (1985), Grassberger (1988), Moddemeijer (1999), Panzeri & Treves (1996), Roulston (1997)]. Des Weiteren kann unter gewissen Annahmen der statistische Fehler berechnet werden [Panzeri & Treves (1996), Roulston (1997), Roulston (1999)]. Diese Annahmen beinhalten, dass der Bias und die Varianz des Fehlers der einzelnen Entropien in Gl. (2.10) und Gl. (2.27)

unabhängig voneinander verteilt sind. Dies ist im Allgemeinen aber nicht der Fall. Außerdem gelten diese Korrekturen nur im Mittel, was dazu führt, dass die gegenseitige Information und die Transferentropie aufgrund der Korrektur negative Werte annehmen können. Daher kann es sich im Einzelfall als zweckdienlicher erweisen, den positiven Bias bewusst beizubehalten.

### 3.3 Partitionierung des Zustandsraums

Die Berechnung von Entropien bzw. Informationen diskreter oder kontinuierlicher Prozesse aus Daten ist mit verschiedenen Schwierigkeiten verbunden. So lassen sich die Wahrscheinlichkeiten von diskreten Zuständen relativ einfach schätzen, wohingegen das Schätzen von Wahrscheinlichkeitsdichten und insbesondere Dichten von Übergangsverteilungen nur sehr schwer möglich ist und im Allgemeinen umfangreiche Datensätze erfordert. Aus diesem Grund sind viele Verfahren entwickelt worden, bei denen die kontinuierlichen Zufallsvariablen mit einer entsprechenden Vergrößerungsprozedur in diskrete umgewandelt werden. Anschließend können die Formalismen für diskrete Prozesse angewendet werden [Grassberger & Procaccia (1983b), Fraser & Swinney (1986), Cover & Thomas (1991), Gaspard & Wang (1993), Darbellay (1999)]. In diesem Kapitel sollen die Voraussetzungen, unter denen dieses Vorgehen gerechtfertigt ist, untersucht werden.

Betrachte die Zufallsvariable  $X_i$  eines kontinuierlichen stochastischen Prozesses  $X$ , wobei  $X$   $d$ -dimensional sei. Aufgrund der Kartesischen Struktur, die dem euklidischen Raum  $\mathbb{R}^d$  zugrunde liegt, wird der Zustandsraum von  $X_i$  in eine abzählbare Anzahl nicht-überlappender Quader  $I_m = B_{m,1} \times \dots \times B_{m,d}$ ,  $m = 1, \dots, M \in \mathbb{N}$  zerlegt,  $\cup_{m=1}^M I_m = \mathbb{R}^d$  bzw. “= Zustandsraum”, siehe Abb. 3.2. Dabei stellen die  $B_{m,j} = [a_{m,j}, b_{m,j}) = \{x \in \mathbb{R} : a_{m,j} \leq x < b_{m,j}\}$ ,  $j = 1, \dots, d$  rechts halboffene Intervalle dar. Diese dürfen sich für verschiedene  $m$  überlappen,  $B_{m_1,j} \cap B_{m_2,j} \neq \emptyset$ , es muss aber stets  $I_{m_1} \cap I_{m_2} = \emptyset$  für  $m_1 \neq m_2$  gelten. Die Menge aller Quader  $\mathcal{I} = \{I_m : m \in \mathbb{N}\}$  heißt *Partition des Zustandsraums*. Sie übernimmt hier die Rolle des diskreten Zustandsraums  $\mathcal{X}$ .

Des Weiteren wird die *Norm der Partition*  $\|\mathcal{I}\|$  benötigt. Sie wird als die größte Kantenlänge aller Quader der Partition definiert,

$$\|\mathcal{I}\| = \max_{m \in \mathbb{N}} \max_{1 \leq j \leq d} \Delta_{m,j}, \quad (3.9)$$

wobei  $\Delta_{m,j} = |B_{m,j}| = b_{m,j} - a_{m,j}$  die  $j$ -te Kantenlänge von  $I_m$ , also die Länge des Intervalls  $B_{m,j}$  ist.

Für Konvergenzaussagen bezüglich der gegenseitigen Information müssen noch Verfeinerungen von Partitionen eingeführt werden. Sind  $\mathcal{J}$  und  $\mathcal{I}$  zwei Partitionen, wobei  $\mathcal{J}$  aus  $\mathcal{I}$  gewonnen werden kann, indem genau ein Quader  $I_{\tilde{m}} \in \mathcal{I}$  in zwei Quader  $J_{\tilde{m}_1}, J_{\tilde{m}_2} \in \mathcal{J}$  mit  $J_{\tilde{m}_1} \cup J_{\tilde{m}_2} = I_{\tilde{m}}$  und  $J_{\tilde{m}_1} \cap J_{\tilde{m}_2} = \emptyset$  aufgespalten wird, dann wird  $\mathcal{J}$  eine *elementare Verfeinerung von  $\mathcal{I}$*  genannt, siehe

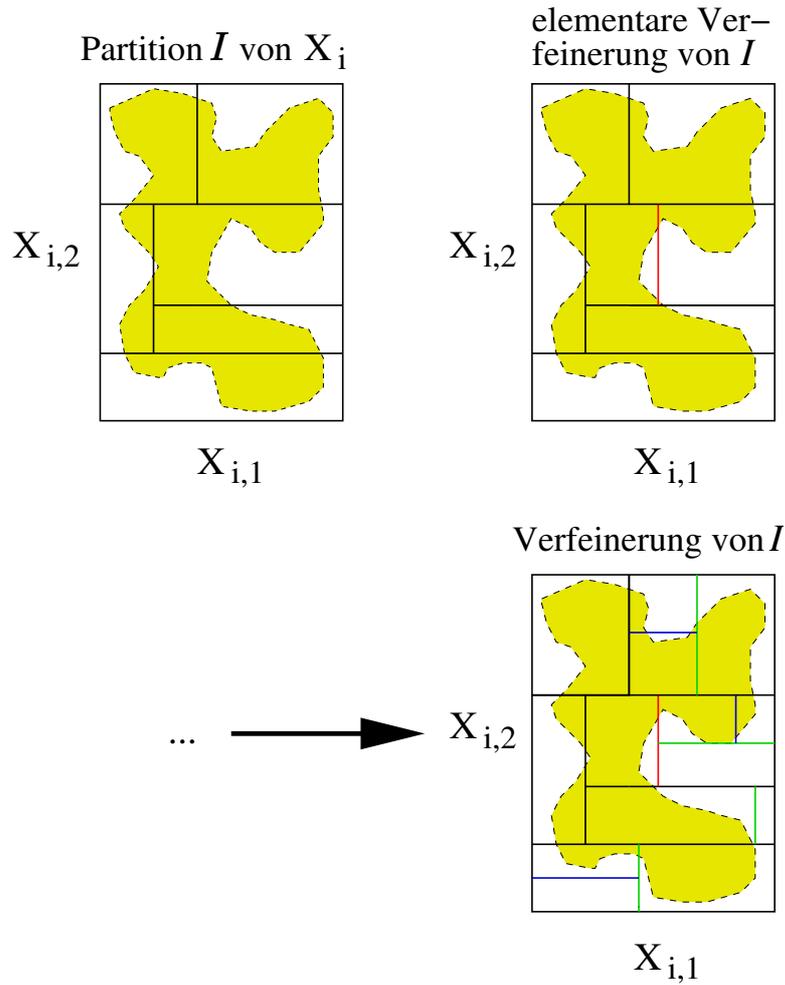


Abbildung 3.2: Beispiel einer Partition des Zustandsraums von  $X_i = (X_{i,1}, X_{i,2})$ , wobei  $X$  zwei-dimensional ist ( $d = 2$ ).

Abb. 3.2.  $\mathcal{J}$  heißt eine *Verfeinerung* von  $\mathcal{I}$ , wenn nach mehreren hintereinander ausgeführten elementaren Verfeinerungen  $\mathcal{J}$  aus  $\mathcal{I}$  hervorgehen kann, in Zeichen:  $\mathcal{J} \prec \mathcal{I}$ .

Die Partitionierung von zum Beispiel  $\mathbf{X}_i^{(k)}$  erfolgt analog wie die von  $X_i$ , denn  $\tilde{X}_i = \mathbf{X}_i^{(k)}$  kann als eine  $(dk)$ -dimensionale Zufallsvariable aufgefasst werden.

Als erstes wird das Konvergenzverhalten der diskretisierten Shannon-Entropie in einer allgemeineren Form als in [Cover & Thomas (1991)] untersucht. Hierzu wird vorausgesetzt, dass die kontinuierliche Shannon-Entropie, Gl. (2.31), als Riemann-Integral existiert. Außerdem soll die Dichte  $g_{X_i}$  eine stetige Funktion sein. Des Weiteren sei  $(\mathcal{I}_n)_{n \in \mathbb{N}}$  eine Folge von Partitionen des Zustandsraums von  $X_i$  mit  $\|\mathcal{I}_n\| \rightarrow 0$  für  $n \rightarrow \infty$ .

Für jedes  $n \in \mathbb{N}$  ist die Wahrscheinlichkeit, mit der  $X_i$  im Quader  $I_{m,n}$  anzu-

treffen ist, gegeben durch

$$P^{X_i}(I_{m,n}) \equiv P\{X_i \in I_{m,n}\} = \int_{I_{m,n}} g_{X_i}(x) dx.$$

Da  $g_{X_i}$  stetig ist, existiert nach dem Mittelwertsatz ein  $\tilde{x}(m, n) \in I_{m,n}$ , so dass

$$P\{X_i \in I_{m,n}\} = g_{X_i}(\tilde{x}(m, n)) \cdot \int_{I_{m,n}} dx = g_{X_i}(\tilde{x}(m, n)) \cdot |I_{m,n}| \quad (3.10)$$

gilt. Hier ist  $|I_{m,n}| = \int_{I_{m,n}} dx$  das Volumen von  $I_{m,n}$ . Einsetzen dieser Wahrscheinlichkeiten in Gl. (2.2) und Mitteln über alle Partitionselemente (Quader) liefert die diskretisierte Shannon-Entropie bezüglich der Partition  $\mathcal{I}_n$

$$H_{\mathcal{I}_n}(X_i) = - \sum_{m=1}^{M_n} g_{X_i}(\tilde{x}(m, n)) \log(g_{X_i}(\tilde{x}(m, n))) \cdot |I_{m,n}| \\ - \sum_{m=1}^{M_n} g_{X_i}(\tilde{x}(m, n)) \log |I_{m,n}| \cdot |I_{m,n}|. \quad (3.11)$$

Da die Riemann-Integrierbarkeit von  $g_{X_i} \cdot \log g_{X_i}$  vorausgesetzt wurde, folgt, dass der erste Summand für  $n \rightarrow \infty$  gegen die kontinuierliche Shannon-Entropie konvergiert. Somit erhält man die Konvergenzaussage

$$H_{\mathcal{I}_n}(X_i) + \sum_{m=1}^{M_n} g_{X_i}(\tilde{x}(m, n)) \log |I_{m,n}| \cdot |I_{m,n}| \xrightarrow{n \rightarrow \infty} H(X_i), \quad (3.12)$$

das heißt die kontinuierliche Shannon-Entropie und die diskretisierten Shannon-Entropien der Partitionenfolge unterscheiden sich im Limes um den Erwartungswert der logarithmierten Quadervolumina dieser Partitionen. Für den Spezialfall, dass als Partitionselemente Würfel mit gleicher Kantenlänge  $\varepsilon_n$  gewählt werden, also  $|I_{m,n}| = \varepsilon_n^d$ , liefert dies die Konvergenzaussage

$$H_{\mathcal{I}_n}(X_i) + d \cdot \log \varepsilon_n \xrightarrow{n \rightarrow \infty} H(X_i).$$

Dementsprechend divergiert die diskretisierte Shannon-Entropie  $H_{\mathcal{I}_n}(X_i)$  wie  $-d \log \varepsilon_n$  für  $n \rightarrow \infty$ , das heißt die Divergenz wird von der Skala  $\varepsilon_n$  der Messauflösung bestimmt. Ist beispielsweise  $X_i$  auf dem Intervall  $[0, 1)$  gleichverteilt ( $g_{X_i}(x) = 1$  falls  $x \in [0, 1)$  und 0 sonst), so ist die Shannon-Entropie  $H(X_i) = 0$ . Wird der Zustandsraum bei der  $n$ -ten Partition in  $n$  gleich große Intervalle zerlegt ( $\varepsilon_n = 1/n$ ), so erhält man für die diskretisierte Shannon-Entropie  $H_{\mathcal{I}_n}(X_i) = \log n$ . Sie divergiert also logarithmisch mit der Anzahl der Partitionselemente.

Um Konvergenzaussagen über die vergrößerte gegenseitige Information machen zu können, muss auch hier vorausgesetzt werden, dass  $g_{X_i, Y_j} \cdot \log(g_{X_i, Y_j}/g_{X_i}g_{Y_j})$  Riemann-integrierbar ist und dass die Dichten  $g_{X_i, Y_j}$ ,  $g_{X_i}$  sowie  $g_{Y_j}$  stetig sind. Diesmal wird der Zustandsraum von  $(X_i, Y_j)$  partitioniert, wobei auch hier wieder  $(\mathcal{I}_n)_{n \in \mathbb{N}}$  eine Folge von Partitionen mit  $\|\mathcal{I}_n\| \rightarrow 0$  für  $n \rightarrow \infty$  ist. Mit dem Mittelwertsatz lassen sich die Wahrscheinlichkeiten, mit denen der Prozess in dem Quader  $\mathcal{I}_n$  anzutreffen ist, durch die Dichten ausdrücken:

$$\begin{aligned} P\{(X_i, Y_j) \in I_{m,n}\} &= g_{X_i, Y_j}(\tilde{x}, \tilde{y})(m, n) \cdot |I_{m,n}|, \\ P\{X_i \in \Pi_{X_i}(I_{m,n})\} &= g_{X_i}(\tilde{x}(m, n)) \cdot |\Pi_{X_i}(I_{m,n})|, \\ P\{Y_j \in \Pi_{Y_j}(I_{m,n})\} &= g_{Y_j}(\tilde{y}(m, n)) \cdot |\Pi_{Y_j}(I_{m,n})|, \end{aligned} \quad (3.13)$$

wobei  $\Pi_{X_i}$  und  $\Pi_{Y_j}$  Projektoren sind, die  $(X_i, Y_j)$  auf  $X_i$  bzw.  $Y_j$  abbilden. Auch hier sind  $(\tilde{x}, \tilde{y})(m, n) \in I_{m,n}$ ,  $\tilde{x}(m, n) \in \Pi_{X_i}(I_{m,n})$  und  $\tilde{y}(m, n) \in \Pi_{Y_j}(I_{m,n})$ . Werden diese Wahrscheinlichkeiten in Gl. (2.9) eingesetzt und wird berücksichtigt, dass für die Quadervolumina nach dem Satz von Fubini

$$|I_{m,n}| = |\Pi_{X_i}(I_{m,n}) \times \Pi_{Y_j}(I_{m,n})| = |\Pi_{X_i}(I_{m,n})| \cdot |\Pi_{Y_j}(I_{m,n})|$$

gilt, so erhält man für die gegenseitige Information bezüglich  $\mathcal{I}_n$

$$\begin{aligned} M_{\mathcal{I}_n}(X_i, Y_j) &= \sum_{m=1}^{M_n} g_{X_i, Y_j}(\tilde{x}(m, n), \tilde{y}(m, n)) \\ &\quad \times \log \frac{g_{X_i, Y_j}(\tilde{x}(m, n), \tilde{y}(m, n))}{g_{X_i}(\tilde{x}(m, n)) \cdot g_{Y_j}(\tilde{y}(m, n))} \cdot |I_{m,n}|. \end{aligned} \quad (3.14)$$

Da die kontinuierliche gegenseitige Information als Riemann-Integral existieren soll, folgt unmittelbar

$$M_{\mathcal{I}_n}(X_i, Y_j) \xrightarrow{n \rightarrow \infty} M(X_i, Y_j). \quad (3.15)$$

Handelt es sich bei  $(\mathcal{I}_n)_{n \in \mathbb{N}}$  um eine Folge verfeinerter Partitionen, so ist diese Konvergenz sogar monoton steigend, denn es gilt

$$\mathcal{I}_{n+1} \prec \mathcal{I}_n \implies M_{\mathcal{I}_{n+1}}(X_i, Y_j) \geq M_{\mathcal{I}_n}(X_i, Y_j), \quad (3.16)$$

wie Darbellay (1999) zeigte.

Ein alternativer Beweis für diese Aussage soll im Folgenden vorgestellt werden. Da jede Verfeinerung einer Partition gewonnen werden kann, indem eine endliche Anzahl von elementaren Verfeinerungen hintereinander ausgeführt wird, reicht es aus, wenn  $M_{\mathcal{J}}(X_i, Y_j) \geq M_{\mathcal{I}}(X_i, Y_j)$  für jede elementare Verfeinerung  $\mathcal{J}$  von  $\mathcal{I}$  gezeigt wird.  $\mathcal{J}$  soll hierbei aus  $\mathcal{I}$  hervorgehen, indem der Quader  $I_m \in \mathcal{I}$  in die Quader  $J_{m_1}, J_{m_2} \in \mathcal{J}$ , mit  $J_{m_1} \cap J_{m_2} = \emptyset$  und  $J_{m_1} \cup J_{m_2} = I_m$  zerlegt wird. Dafür muss genau ein Intervall  $B_{m,k}$ , das die  $k$ -te Kante von  $I_m$  darstellt, in zwei

disjunkte Intervalle aufgeteilt werden. Handelt es sich bei  $B_{m,k}$  um ein Intervall im Bildraum von einer der Komponenten von  $X_i$ , so gilt

$$\begin{aligned} P\{X_i \in \Pi_{X_i}(I_m)\} &= P\{X_i \in \Pi_{X_i}(J_{m_1})\} + P\{X_i \in \Pi_{X_i}(J_{m_2})\}, \\ P\{Y_j \in \Pi_{Y_j}(I_m)\} &= P\{Y_j \in \Pi_{Y_j}(J_{m_1})\} = P\{Y_j \in \Pi_{Y_j}(J_{m_2})\}. \end{aligned}$$

Eine analoge Relation erhält man, wenn sich  $B_{m,k}$  auf eine Komponente von  $Y_j$  bezieht. Für beide Fälle liefern diese Relationen die Gleichung

$$\begin{aligned} P\{X_i \in \Pi_{X_i}(I_m)\} \cdot P\{Y_j \in \Pi_{Y_j}(I_m)\} \\ = P\{X_i \in \Pi_{X_i}(J_{m_1})\} \cdot P\{Y_j \in \Pi_{Y_j}(J_{m_1})\} \\ + P\{X_i \in \Pi_{X_i}(J_{m_2})\} \cdot P\{Y_j \in \Pi_{Y_j}(J_{m_2})\}. \end{aligned}$$

Hieraus folgt mit der Log-Summen-Ungleichung (2.4), siehe auch Anhang A,

$$\begin{aligned} M_{\mathcal{J}}(X_i, Y_j) - M_{\mathcal{I}}(X_i, Y_j) \\ = \frac{P\{(X_i, Y_j) \in J_{m_1}\}}{P\{X_i \in \Pi_{X_i}(J_{m_1})\} \cdot P\{Y_j \in \Pi_{Y_j}(J_{m_1})\}} \\ + \frac{P\{(X_i, Y_j) \in J_{m_2}\}}{P\{X_i \in \Pi_{X_i}(J_{m_2})\} \cdot P\{Y_j \in \Pi_{Y_j}(J_{m_2})\}} \\ - \frac{P\{(X_i, Y_j) \in I_m\}}{P\{X_i \in \Pi_{X_i}(I_m)\} \cdot P\{Y_j \in \Pi_{Y_j}(I_m)\}} \geq 0, \quad (3.17) \end{aligned}$$

also die monoton steigende Konvergenz der diskretisierten gegenseitigen Information gegen die kontinuierliche gegenseitige Information für eine verfeinerte Partitionenfolge.

Sind  $X$  und  $Y$  deterministisch gekoppelt ( $Y_j = f \circ X_i$ ), so folgt aus Gl. (3.15) und mit Gl. (2.11), dass  $M_{\mathcal{I}_n}(X_i, f \circ X_i)$  wie  $-\log \|\mathcal{I}_n\|$  divergiert.

Als Vorbereitung zur Untersuchung des Konvergenzverhaltens der vergrößerten Transferentropie soll zunächst das Verhalten der bedingten Shannon-Entropie bezüglich Partitionierung betrachtet werden. Hierfür wird der Zustandsraum von  $\mathbf{X}_{i+1}^{(k+1)}$  partitioniert, wobei  $(\mathcal{I}_n)_{n \in \mathbb{N}}$  wieder eine Partitionenfolge sein soll, deren Norm für große  $n$  verschwindet. Sind die Dichten  $g_{\mathbf{X}_{i+1}^{(k+1)}}$  und  $g_{\mathbf{X}_i^{(k)}}$  wieder stetig, so ist nach dem Mittelwertsatz

$$\begin{aligned} P\{\mathbf{X}_{i+1}^{(k+1)} \in I_{m,n}\} &= g_{\mathbf{X}_{i+1}^{(k+1)}}(\tilde{\mathbf{x}}_{i+1}^{(k+1)}(m, n)) \cdot |I_{m,n}|, \\ P\{\mathbf{X}_i^{(k)} \in \Pi_{\mathbf{X}_i^{(k)}}(I_{m,n})\} &= g_{\mathbf{X}_i^{(k)}}(\tilde{\mathbf{x}}_i^{(k)}(m, n)) \cdot |\Pi_{\mathbf{X}_i^{(k)}}(I_{m,n})|, \quad (3.18) \end{aligned}$$

mit  $\tilde{\mathbf{x}}_{i+1}^{(k+1)}(m, n) \in I_{m,n}$  und  $\tilde{\mathbf{x}}_i^{(k)}(m, n) \in \Pi_{\mathbf{X}_i^{(k)}}(I_{m,n})$ . Aus Gl. (2.13) sowie  $|I_{m,n}| = |\Pi_{\mathbf{X}_{i+1}}(I_{m,n})| \cdot |\Pi_{\mathbf{X}_i^{(k)}}(I_{m,n})|$  folgt für die diskretisierte Übergangsverteilung

lung

$$\begin{aligned} P\{X_{i+1} \in \Pi_{X_{i+1}}(I_{m,n}) | \mathbf{X}_i^{(k)} \in \Pi_{\mathbf{X}_i^{(k)}}(I_{m,n})\} \\ = \frac{g_{\mathbf{X}_{i+1}^{(k+1)}}(\tilde{\mathbf{x}}_{i+1}^{(k+1)}(m,n))}{g_{\mathbf{X}_i^{(k)}}(\tilde{\mathbf{x}}_i^{(k)}(m,n))} \cdot |\Pi_{X_{i+1}}(I_{m,n})|. \end{aligned} \quad (3.19)$$

Werden Gl. (3.18) und Gl. (3.19) in Gl. (2.15) eingesetzt, so erhält man für die bedingte Shannon-Entropie auf einer Partition

$$\begin{aligned} H_{\mathcal{I}_n}(X_{i+1} | \mathbf{X}_i^{(k)}) \\ = \sum_{m=1}^{M_n} g_{\mathbf{X}_{i+1}^{(k+1)}}(\tilde{\mathbf{x}}_{i+1}^{(k+1)}(m,n)) \times \log \frac{g_{\mathbf{X}_{i+1}^{(k+1)}}(\tilde{\mathbf{x}}_{i+1}^{(k+1)}(m,n))}{g_{\mathbf{X}_i^{(k)}}(\tilde{\mathbf{x}}_i^{(k)}(m,n))} \cdot |I_{m,n}| \\ - \sum_{m=1}^{M_n} g_{\mathbf{X}_{i+1}^{(k+1)}}(\tilde{\mathbf{x}}_{i+1}^{(k+1)}(m,n)) \cdot \log |\Pi_{X_{i+1}}(I_{m,n})| \cdot |I_{m,n}|. \end{aligned} \quad (3.20)$$

Existiert die kontinuierliche bedingte Shannon-Entropie als Riemann-Integral, dann folgt mit Gl. (2.37)

$$\begin{aligned} H_{\mathcal{I}_n}(X_{i+1} | \mathbf{X}_i^{(k)}) + \sum_{m=1}^{M_n} g_{\mathbf{X}_{i+1}^{(k+1)}}(\tilde{\mathbf{x}}_{i+1}^{(k+1)}(m,n)) \cdot \log |\Pi_{X_{i+1}}(I_{m,n})| \cdot |I_{m,n}| \\ \xrightarrow{n \rightarrow \infty} H(X_{i+1} | \mathbf{X}_i^{(k)}). \end{aligned} \quad (3.21)$$

Auch hier unterscheidet sich im Limes die vergrößerte bedingte Shannon-Entropie von der kontinuierlichen durch den Erwartungswert der logarithmierten Intervalllänge  $|\Pi_{X_{i+1}}(I_{m,n})|$  der Partitionen. Insbesondere divergieren beide Terme auf der linken Seite mit  $-\log |\Pi_{X_{i+1}}(I_{m,n})|$  bzw.  $\log |\Pi_{X_{i+1}}(I_{m,n})|$  für  $n \rightarrow \infty$ .

Zum Abschluss wird nun gezeigt, dass die Transferentropie bezüglich einer Partitionsfolge  $(\mathcal{I}_n)_{n \in \mathbb{N}}$  mit  $\|\mathcal{I}_n\| \rightarrow_{n \rightarrow \infty} 0$  gegen die kontinuierliche Transferentropie konvergiert. Hierzu muss vorausgesetzt werden, dass die kontinuierliche Transferentropie als Riemann-Integral existiert und dass die Dichten  $g_{\mathbf{X}_i^{(k)}}$ ,  $g_{\mathbf{X}_{i+1}^{(k+1)}}$ ,  $g_{\mathbf{X}_i^{(k)}, \mathbf{Y}_j^{(l)}}$  und  $g_{\mathbf{X}_{i+1}^{(k+1)}, \mathbf{Y}_j^{(l)}}$  stetig sind.

Wird der Zustandsraum von  $(\mathbf{X}_{i+1}^{(k+1)}, \mathbf{Y}_j^{(l)})$  in beliebige Quader aufgeteilt, so ist

$$\begin{aligned} P\{(\mathbf{X}_{i+1}^{(k+1)}, \mathbf{Y}_j^{(l)}) \in I_{m,n}\} &= g_{\mathbf{X}_{i+1}^{(k+1)}, \mathbf{Y}_j^{(l)}}((\tilde{\mathbf{x}}_{i+1}^{(k+1)}, \tilde{\mathbf{y}}_j^{(l)})(m,n)) \cdot |I_{m,n}|, \\ P\{(\mathbf{X}_i^{(k)}, \mathbf{Y}_j^{(l)}) \in \Pi_{\mathbf{X}_i^{(k)}, \mathbf{Y}_j^{(l)}}(I_{m,n})\} & \\ &= g_{\mathbf{X}_i^{(k)}, \mathbf{Y}_j^{(l)}}((\tilde{\mathbf{x}}_i^{(k)}, \tilde{\mathbf{y}}_j^{(l)})(m,n)) \cdot |\Pi_{\mathbf{X}_i^{(k)}, \mathbf{Y}_j^{(l)}}(I_{m,n})|, \end{aligned} \quad (3.22)$$

mit  $(\tilde{\mathbf{x}}_{i+1}^{(k+1)}, \tilde{\mathbf{y}}_j^{(l)})(m, n) \in I_{m,n}$  und  $(\tilde{\mathbf{x}}_i^{(k)}, \tilde{\mathbf{y}}_j^{(l)})(m, n) \in \Pi_{\mathbf{X}_i^{(k)}, \mathbf{Y}_j^{(l)}}(I_{m,n})$ .

Als Nächstes werden die Übergangswahrscheinlichkeiten in Gl. (2.21) gemäß der Relation (2.13) durch die endlich-dimensionalen Verteilungen ersetzt. Einsetzen von Gl. (3.18) und Gl. (3.22) sowie Mitteln über alle Partitionselemente liefert dann für die Transferentropie einer Partition

$$T_{\mathcal{I}_n}(X_{i+1}|\mathbf{X}_i^{(k)}, \mathbf{Y}_j^{(l)}) = \sum_{m=1}^{M_n} g_{\mathbf{X}_{i+1}^{(k+1)}, \mathbf{Y}_j^{(l)}}((\tilde{\mathbf{x}}_{i+1}^{(k+1)}, \tilde{\mathbf{y}}_j^{(l)})(m, n)) \times \log \left( \frac{g_{\mathbf{X}_{i+1}^{(k+1)}, \mathbf{Y}_j^{(l)}}((\tilde{\mathbf{x}}_{i+1}^{(k+1)}, \tilde{\mathbf{y}}_j^{(l)})(m, n)) \cdot g_{\mathbf{X}_i^{(k)}}(\tilde{\mathbf{x}}_i^{(k)}(m, n))}{g_{\mathbf{X}_i^{(k)}, \mathbf{Y}_j^{(l)}}((\tilde{\mathbf{x}}_i^{(k)}, \tilde{\mathbf{y}}_j^{(l)})(m, n)) \cdot g_{\mathbf{X}_{i+1}^{(k+1)}}(\tilde{\mathbf{x}}_{i+1}^{(k+1)}(m, n))} \right) \cdot |I_{m,n}|. \quad (3.23)$$

Unter Berücksichtigung der Relation Gl. (2.37) konvergiert  $T_{\mathcal{I}_n}(X_{i+1}|\mathbf{X}_i^{(k)}, \mathbf{Y}_j^{(l)})$  als Riemann-Integral gegen die kontinuierliche Transferentropie,

$$T_{\mathcal{I}_n}(X_{i+1}|\mathbf{X}_i^{(k)}, \mathbf{Y}_j^{(l)}) \xrightarrow{n \rightarrow \infty} T(X_{i+1}|\mathbf{X}_i^{(k)}, \mathbf{Y}_j^{(l)}). \quad (3.24)$$

Im Gegensatz zur gegenseitigen Information muss die Konvergenz der Transferentropie bei Verfeinerungen nicht monoton sein. Die Ursache liegt darin, dass die bedingten Wahrscheinlichkeiten (Übergangswahrscheinlichkeiten) nicht additiv in ihrer Bedingung sind, das heißt  $P(A|B) + P(A|C) \neq P(A|B \cup C)$ . Dennoch wurde bei den bisher studierten Beispielsystemen beobachtet, dass die Transferentropie in Kopplungsrichtung bei größeren Partitionierungen größer war, als die Transferentropie für die Gegenrichtung. Dieses Verhalten hat sich bei weiteren Verfeinerungen der Partition erst dann geändert, als aufgrund von endlichen Datenpunkten die Partitionselemente so klein gewählt wurden, dass auf ihnen die Verteilung nicht mehr geschätzt werden konnte.

Aufgrund der Kartesischen Struktur des zugrundeliegenden Zustandsraums sind Quader die natürliche Wahl für eine Partitionierung. Werden Elemente mit einer anderen Form für die Zerlegung verwendet, so erhält man bei den Konvergenzaussagen einen zusätzlichen geometrischen Faktor.

Die Konvergenzaussagen können weiter abgeschwächt werden. Sie gelten offenbar auch dann noch, wenn die Dichten bis auf endlich viele Punkte stetig sind.

Diese Konvergenzaussagen können direkt angewendet werden, um die gegenseitige Information oder Transferentropie aus Daten zu berechnen. In der Praxis ist dies aber mit numerischen Problemen verbunden. Werden die Quader zu klein gewählt, so fallen zu wenige Datenpunkte in diese Partitionselemente und die Verteilungen können nicht mehr zuverlässig geschätzt werden. Folglich wird eine Konvergenz oft nicht beobachtet.

Bei der einfachsten Implementierung wird der Zustandsraum in gleich große Quader zerlegt, wobei von einer Partition zur nächsten die Quadergröße gleichmäßig verkleinert wird. Für inhomogene Verteilungen ist dieses Verfahren

sehr rechenintensiv, denn das Gitter der Partitionierung muss fein genug sein, um die Strukturen der Verteilung noch auflösen zu können. Dies hat zur Folge, dass viele Partitionselemente nur noch sehr wenige Punkte enthalten.

Eine alternative Methode stellen adaptive Partitionen dar. Hierbei werden nur jene Quader weiter zerlegt, die in einem lokalen Gebiet des Zustandsraumes liegen, wo eine statistisch signifikante Substruktur vorhanden ist. Solch ein Algorithmus, bei dem die Quader rekursiv in jeder Dimension des Zustandsraums aufgespalten werden, wurde bereits von Fraser & Swinney (1986) vorgestellt. Eine Variante ist in [Schreiber (1990)] beschrieben. Kürzlich wurde von Darbellay (1999) ein ähnlicher Algorithmus publiziert, welcher die Quader gleichzeitig entlang aller Koordinatenachsen zerlegt. Hierdurch wird ein Symmetriebruch, den die gegenseitige Information in den zuvor erwähnten Methoden beim Vertauschen der Argumente erfährt, vermieden. Als Abbruchkriterium wird ein  $\chi^2$ -Test auf Unabhängigkeit angewandt. Auf Konvergenz wird hingegen nicht getestet, so dass sich hier eine Fehlerquelle auftut, wenn die exakten Werte benötigt werden. Da die gegenseitige Information bei den adaptiven Verfahren monoton steigt, mag dies bei ihr nicht problematisch sein, bei der Transferentropie hingegen wohl. Daher wurde von einer Modifikation der adaptiven Algorithmen zur Berechnung der Transferentropie Abstand genommen.

Ein weiteres Problem bei der Schätzung der Dichten auf Partitionen ist ein zusätzlicher Bias, der durch serielle Korrelationen entstehen kann, wenn die Punkte in den Quadern gezählt werden. Eine detaillierte Diskussion dieser Problematik ist in [Kantz & Schreiber (1997)] zu finden. Das einzig bekannte Korrekturverfahren besteht darin, nur solche Punkte in die Partitionselemente einzusortieren, die in der Zeit dekorreliert sind, wofür allerdings eine enorme Menge an Daten benötigt wird. Kernschätzer, die später noch vorgestellt werden, bieten eine viel ökonomischere Lösung.

Falls nur ein Test auf Abhängigkeit von  $X_t$  und  $Y_s$  gewünscht ist, so ist es wegen der monotonen Konvergenz der gegenseitigen Information ausreichend, wenn für eine beliebige Partition gezeigt werden kann, dass  $M_{\mathcal{I}}(X_i, Y_j) > 0$  innerhalb statistischer Fluktuationen ist. Eine einfache Partition hierfür kann beispielsweise erzeugt werden, indem der Zustandsraum in jeder Dimension in genau zwei Intervalle aufgeteilt wird. Für einen statistischen Test kann die in [Darbellay (1999)] hergeleitete Fehlerformel oder ein Test mit Surrogaten [Schreiber & Schmitz (2000a)] verwendet werden.

### 3.4 Parametrische Verteilungen

Gehört die Verteilung des Prozesses  $(X, Y)$  zu einer Familie von Verteilungen, die durch spezifische Parameter festgelegt sind, so reicht es aus, diese Parameter für das konkrete System zu schätzen, um die gegenseitige Information oder Transferentropie zu berechnen. Für beliebige parametrische Verteilungen sind

für gewöhnlich analytische Ausdrücke für  $M(X_i, Y_j)$  und  $T(X_{i+1}|\mathbf{X}_i^{(k)}, \mathbf{Y}_j^{(l)})$  nicht ableitbar. Aber mit den geschätzten Dichten können sie durch numerische Integration bestimmt werden.

Eine Ausnahme stellen multivariante Gauß-Prozesse dar, denn für sie lassen sich Shannon-Entropie und gegenseitige Information analytisch berechnen [Prichard & Theiler (1995), Cover & Thomas (1991)]. Sei  $X$  ein zentrierter,  $d$ -dimensionaler Prozess, das heißt für alle  $i \in \mathbb{Z}$  sind die Erwartungswerte der einzelnen Komponenten von  $X_i$  Null,  $E[X_{i,\mu}] = 0$ ,  $\mu = 1, \dots, d$ . Des Weiteren sei  $\Sigma(X_i)$  die Kovarianzmatrix von  $X_i$ , deren Elemente durch  $\Sigma_{\mu,\nu}(X_i) = \text{Cov}[X_{i,\mu}, X_{i,\nu}] = E[X_{i,\mu} \cdot X_{i,\nu}] = \Sigma_{\nu,\mu}(X_i)$  gegeben sind. Trivialerweise ist  $\Sigma(X_i)$  symmetrisch. Ist nun  $X_i$  Gauß-verteilt, so lautet deren Dichte [Bauer (1991), Prichard & Theiler (1995)]

$$g_{X_i}(x) = \frac{1}{\sqrt{(2\pi)^d \det \Sigma(X_i)}} \cdot \exp\left(-\frac{x^T \cdot \Xi(X_i) \cdot x}{2}\right),$$

wobei  $\det \Sigma(X_i)$  die Determinante und  $\Xi(X_i)$  die inverse Matrix von  $\Sigma(X_i)$  ist.  $x^T$  bezeichnet den transponierten Vektor von  $x$ .

Wird dies in Gl. (2.31) eingesetzt, so folgt nach einer elementaren Rechnung [Prichard & Theiler (1995), Cover & Thomas (1991)]

$$H(X_i) = \frac{d}{2} \log(2\pi e) + \frac{1}{2} \log(\det \Sigma(X_i)). \quad (3.25)$$

$e$  ist hier die Eulersche Zahl. Da nach Gl. (2.45) die kontinuierliche Shannon-Entropie unter der Transformation  $X_i \rightarrow X_i + a$ , mit  $a \in \mathbb{R}^d$  unverändert bleibt, ist Gl. (3.25) auch für nicht-zentrierte Gauß-Prozesse gültig. Offensichtlich ist Gl. (3.25) auch anwendbar, um zum Beispiel  $H(\mathbf{X}_i^{(k)})$  zu berechnen, sofern  $\mathbf{X}_i^{(k)}$  Gauß-verteilt ist, denn  $\mathbf{X}_i^{(k)}$  stellt nur eine  $(kd)$ -dimensionale Zufallsvariable dar.

Mit Gl. (2.16) folgt für die bedingte Shannon-Entropie, sofern  $\mathbf{X}_{i+1}^{(k+1)}$  Gauß-verteilt ist:

$$H(X_{i+1}|\mathbf{X}_i^{(k)}) = \frac{d}{2} \log(2\pi e) + \frac{1}{2} \log\left(\frac{\det \Sigma(\mathbf{X}_{i+1}^{(k+1)})}{\det \Sigma(\mathbf{X}_i^{(k)})}\right). \quad (3.26)$$

Ist  $(X_i, Y_j)$  eine multivariante Gauß-Variable, so folgt mit Gl. (2.10)

$$M(X_i, Y_j) = \frac{1}{2} \log \frac{\det \Sigma(X_i) \cdot \det \Sigma(Y_j)}{\det \Sigma(X_i, Y_j)}, \quad (3.27)$$

siehe [Prichard & Theiler (1995)]. Entsprechend ist die Transfarentropie gemäß Gl. (2.22) durch

$$T(X_{i+1}|\mathbf{X}_i^{(k)}, \mathbf{Y}_j^{(l)}) = \frac{1}{2} \log \frac{\det \Sigma(\mathbf{X}_i^{(k)}, \mathbf{Y}_j^{(l)}) \cdot \det \Sigma(\mathbf{X}_{i+1}^{(k+1)})}{\det \Sigma(\mathbf{X}_{i+1}^{(k+1)}, \mathbf{Y}_j^{(l)}) \cdot \det \Sigma(\mathbf{X}_i^{(k)})}, \quad (3.28)$$

gegeben, falls  $(\mathbf{X}_{i+1}^{(k+1)}, \mathbf{Y}_j^{(l)})$  Gauß-verteilt ist.

### 3.5 Kontinuierliches Beispiel: AR(1)-Prozesse

Analog zum diskreten Beispiel in Absch. 2.1.3 soll hier das Verhalten der kontinuierlichen gegenseitigen Information und der kontinuierlichen Transferentropie für gekoppelte autoregressive Prozesse untersucht werden, wobei  $Y$  autonom ist und in  $X$  koppelt:

$$\begin{aligned} X_{i+1} &= \alpha X_i + c Y_i + \eta_{i+1}^X, \\ Y_{i+1} &= \beta Y_i + \eta_{i+1}^Y. \end{aligned} \quad (3.29)$$

$c$  ist der Kopplungsparameter. Die Zufallsvariablen  $\eta_i^X$  und  $\eta_i^Y$ , stellen Gaußsches weißes Rauschen dar und sind normalverteilt,  $E[\eta_i^X] = E[\eta_i^Y] = 0$ ,  $E[\eta_i^X \cdot \eta_i^X] = E[\eta_i^Y \cdot \eta_i^Y] = 1$ . Somit sind alle Verteilungen gaußisch, so dass Gl. (3.27) und Gl. (3.28) angewendet werden können. Des Weiteren ist der Prozess  $(X, Y)$  für große Zeiten stationär, woraus unmittelbar folgt, dass alle Momente zeitunabhängig sind. Dies wiederum erlaubt einem, auf einfachste Weise alle benötigten Kovarianzen zu berechnen. Hierfür ist zunächst das Gleichungssystem

$$\begin{aligned} \text{Cov}[Y_{i+1}, Y_{i+1}] &= \beta^2 \text{Cov}[Y_i, Y_i] + 1 && \equiv \text{Cov}[Y_i, Y_i] \\ \text{Cov}[X_{i+1}, X_{i+1}] &= \alpha^2 \text{Cov}[X_i, X_i] + c^2 \text{Cov}[Y_i, Y_i] \\ &\quad + 2c\alpha \text{Cov}[X_i, Y_i] + 1 && \equiv \text{Cov}[X_i, X_i] \\ \text{Cov}[X_{i+1}, Y_{i+1}] &= \alpha\beta \text{Cov}[X_i, Y_i] + c\beta \text{Cov}[Y_i, Y_i] && \equiv \text{Cov}[X_i, Y_i] \end{aligned}$$

zu lösen, was mit  $u = (1 - \beta^2)$ ,  $v = (1 - \alpha^2)$  und  $w = (1 - \alpha\beta)$  auf

$$\begin{aligned} \text{Cov}[Y_i, Y_i] &= \frac{1}{u}, \\ \text{Cov}[X_i, Y_i] &= c \frac{\beta}{uw}, \\ \text{Cov}[X_i, X_i] &= c^2 \frac{1 + \alpha\beta}{uvw} + \frac{1}{v} \end{aligned}$$

führt. Hieraus folgen die restlichen Kovarianzen

$$\begin{aligned} \text{Cov}[Y_{i+1}, Y_i] &= \beta \text{Cov}[Y_i, Y_i] && = \frac{\beta}{u}, \\ \text{Cov}[Y_{i+1}, X_i] &= \beta \text{Cov}[X_i, Y_i] && = c \frac{\beta^2}{uw}, \\ \text{Cov}[X_{i+1}, Y_i] &= \alpha \text{Cov}[X_i, Y_i] + c \text{Cov}[Y_i, Y_i] && = c \frac{1}{uw}, \\ \text{Cov}[X_{i+1}, X_i] &= \alpha \text{Cov}[X_i, X_i] + c \text{Cov}[X_i, Y_i] && = c^2 \frac{\alpha + \beta}{uvw} + \frac{\alpha}{v}. \end{aligned}$$

Hiermit können die gegenseitige Information und die Transferentropie in beide Richtungen gemäß Gl. (3.27) bzw. Gl. (3.28) berechnet werden. Die Ergebnisse für  $\alpha = 0.6$  und  $\beta = 0.5$  sind als Funktion des Kopplungsparameters  $c$  in

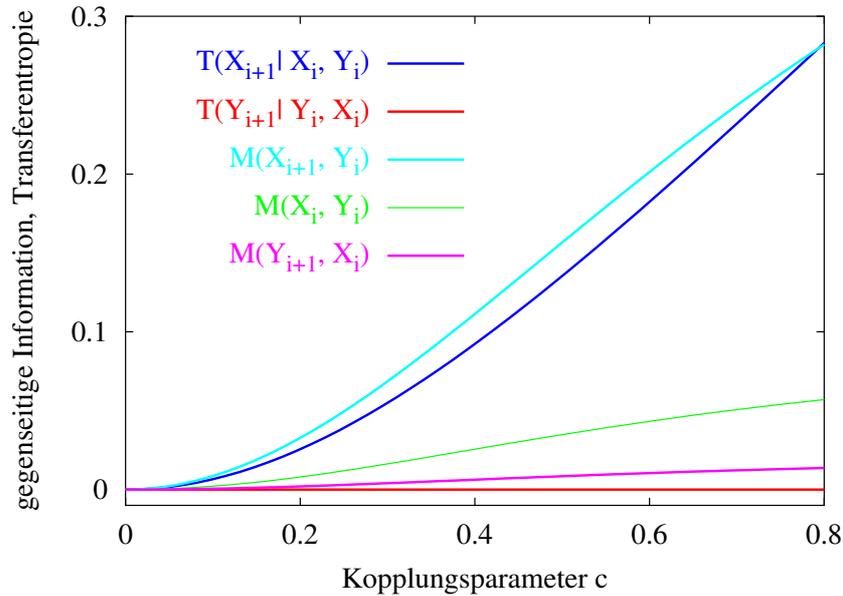


Abbildung 3.3: Gegenseitige Information und Transferentropie als Funktion des Kopplungsparameters  $c$  für zwei autoregressive Prozesse  $X$  und  $Y$ , wobei  $Y$  autonom ist und in  $X$  koppelt.

Abb. 3.3 aufgetragen. Auch hier zeigt sich wieder, dass  $M(Y_{i+1}, X_i) > 0$  ist, was fälschlicherweise dahingehend interpretiert werden könnte, dass  $X$  in  $Y$  koppelt. Die Transferentropie hingegen spiegelt die richtige Struktur des Modells wieder:  $T(Y_{i+1}|Y_i, X_i) = 0$  für alle  $c$ .

### 3.6 Kernschätzer für Dichten

Eine attraktive Alternative zum Diskretisieren der Verteilung ist das nicht-parametrische Schätzen der Verteilungsdichten mit Kernen. Diese Technik ist ausführlich in der Literatur diskutiert worden, zum Beispiel in [Silverman (1986)]. Die einfachste Möglichkeit, die Dichte in dem Punkt  $x$  zu schätzen, besteht darin, die Anzahl der Punkte in einer  $\varepsilon$ -Umgebung um  $x$  zu zählen und anschließend durch das Volumen dieser Umgebung zu dividieren. Mit einem Kern  $K$  können dabei die Punkte beim Zählen entsprechend ihres Abstands von  $x$  gewichtet werden.

Wurde zur Zeit  $i$  der Zustand eines ein-dimensionalen Prozesses  $N$ -mal beobachtet, so lässt sich die Dichte  $g_{X_i}$  in einem willkürlichen Punkt  $x_i$  durch

$$\hat{g}_{X_i;\varepsilon}(x_i) = \frac{1}{N\varepsilon} \sum_{n=1}^N K\left(\frac{x_i - x_i(n)}{\varepsilon}\right) \quad (3.30)$$

parabolischer Kern	$K(u) = \frac{3}{4}(1 - u^2)$	falls $ u  < 1$	und 0 sonst
biparabolischer	$K(u) = \frac{15}{16}(1 - u^2)^2$	falls $ u  < 1$	und 0 sonst
dreieckiger	$K(u) = 1 -  u $	falls $ u  < 1$	und 0 sonst
Gaußscher	$K(u) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}u^2}$		
rechteckiger	$K(u) = \frac{1}{2}$	falls $ u  < 1$	und 0 sonst

Tabelle 3.1: Kerne und ihre Definitionen.

approximieren. Die Zustände der  $N$  Beobachtungen  $\omega_n \in \Omega$ ,  $n = 1, \dots, N$  sind mit  $x_i(n) \equiv X_i(\omega_n)$  bezeichnet worden. Die nichtnegative Funktion  $K$  ist der sogenannte *Kern*, mit dem diese Punkte gewichtet werden. Der Parameter  $\varepsilon > 0$  ist die *Bandbreite*, mit ihr wird die Größe der Umgebung festgelegt, deren Punkte beim Zählen berücksichtigt werden. Erfüllt  $K$

$$\int K(u) du = 1, \quad \int u K(u) du = 0, \quad \int u^2 K(u) du < \infty \quad (3.31)$$

und ist  $g_{X_i}$  zweimal differenzierbar, dann konvergiert  $\hat{g}_{X_i;\varepsilon}$  punktweise gegen  $g_{X_i}$ :

$$\hat{g}_{X_i;\varepsilon}(x_i) \xrightarrow{\varepsilon \rightarrow 0} g_{X_i}(x_i), \quad \forall x_i \in \mathbb{R}. \quad (3.32)$$

Siehe Anhang A für einen Beweis dieser Aussage. Ist die Dichte  $g_{X_i}$  ausreichend glatt, so dass alle Strukturen bereits mit einer festen Bandbreite  $\varepsilon_0$  aufgelöst werden, so wird im Allgemeinen  $\hat{g}_{X_i;\varepsilon} \approx g_{X_i}$  für  $0 < \varepsilon < \varepsilon_0$  gelten. Dementsprechend nähert bei fest vorgegebener Bandbreite  $\varepsilon$  der Kernschätzer die Dichte in dem Sinne, dass Strukturen auf einer kleineren Skala als  $\varepsilon$  ignoriert werden.

Die gebräuchlichsten Kerne sind in Tab. 3.1 aufgelistet. Insbesondere bei der Analyse von dynamischen Systemen wird der rechteckige Kern verwendet [Grassberger & Procaccia (1983a), Theiler (1986), Theiler (1990), Kantz & Schreiber (1997)].

Das Konzept der Kernschätzer kann auf mehrdimensionale Prozesse erweitert werden. Leider hängt hier die Wahl eines effektiven Kerns (gemeint ist ein Kern mit schneller Konvergenz in der Bandbreite und kleiner Varianz) sehr stark von der zugrunde liegenden Struktur und der Dynamik des Prozesses ab [Silverman (1986)]. Ohne Kenntnis über die Dynamik bleibt nur, die Kartesische Struktur des Zustandsraums auszunutzen. Deshalb wird hier der mehr-dimensionale Kern als Produkt von ein-dimensionalen Kernen geschrieben,  $K(u_1, \dots, u_d) = K_1(u_1) \cdots K_d(u_d)$ . Hieraus folgt der  $d$ -dimensionalen Kernschätzer

$$\hat{g}_{X_i,\varepsilon}(x_i) = \frac{1}{N} \sum_{n=1}^N \prod_{j=1}^d \frac{1}{\varepsilon_j} K_j \left( \frac{x_{i,j} - x_{i,j}(n)}{\varepsilon_j} \right), \quad (3.33)$$

wobei jetzt  $X_i = (X_{i,1}, \dots, X_{i,d})$ ,  $\varepsilon = (\varepsilon_1, \dots, \varepsilon_d) \in (\mathbb{R}^+)^d$  und  $x_i = (x_{i,1}, \dots, x_{i,d})$ ,  $x_i(n) = (x_{i,1}(n), \dots, x_{i,d}(n)) \in \mathbb{R}^d$  ist. Dabei ist  $K_j$  der Kern und  $\varepsilon_j$  die Bandbreite der  $j$ -ten Komponente von  $X_i$ , die alle unterschiedlich gewählt werden können. Sind die  $K_j$  rechteckige Kerne, dann ist die Summe proportional zur Anzahl der Datenpunkte, die innerhalb einer Entfernung von  $\varepsilon_j$  in jeder Richtung gefunden werden. Erfüllen alle  $K_j$  die oben gegebenen Bedingungen und ist  $g_{X_i}$  zweimal differenzierbar, so ist auch dieser Schätzer punktweise erwartungstreu,

$$\hat{g}_{X_i;\varepsilon}(x_{i,1}, \dots, x_{i,d}) \xrightarrow{\varepsilon_1, \dots, \varepsilon_d \rightarrow 0} g_{X_i}(x_{i,1}, \dots, x_{i,d}), \quad \forall x_i \in \mathbb{R}^d,$$

siehe hierzu Anhang A.

Zur Berechnung von Shannon-Entropie, gegenseitiger Information und Transferentropie werden zunächst alle benötigten Dichten in den beobachteten Zuständen geschätzt. Anschließend werden die Integrationen in Gl. (2.31), Gl. (2.34) und Gl. (2.41) ausgeführt. Da es sich hierbei um die Berechnung von Erwartungswerten handelt, wird, entsprechend dem Gesetz der großen Zahlen [Bauer (1991), Billingsley (1995)], über alle Beobachtungen gemittelt, siehe auch [Ott (1993)]. Somit sollten die Schätzer

$$\hat{H}_\varepsilon(X_i) = -\frac{1}{N} \sum_{n=1}^N \log \hat{g}_{X_i;\varepsilon}(x_i(n)), \quad (3.34)$$

$$\hat{M}_\varepsilon(X_i, Y_j) = \frac{1}{N} \sum_{n=1}^N \log \frac{\hat{g}_{X_i, Y_j;\varepsilon}(x_i(n), y_j(n))}{\hat{g}_{X_i;\varepsilon_X}(x_i(n)) \cdot \hat{g}_{Y_j;\varepsilon_Y}(y_j(n))}, \quad (3.35)$$

$$\hat{H}_\varepsilon(X_{i+1} | \mathbf{X}_i^{(k)}) = -\frac{1}{N} \sum_{n=1}^N \log \frac{\hat{g}_{\mathbf{X}_{i+1}^{(k+1)};\varepsilon}(\mathbf{x}_{i+1}^{(k+1)}(n))}{\hat{g}_{\mathbf{X}_i^{(k)};\varepsilon_X}(\mathbf{x}_i^{(k)}(n))} \quad (3.36)$$

sowie

$$\begin{aligned} & \hat{T}_\varepsilon(X_{i+1} | \mathbf{X}_i^{(k)}, \mathbf{Y}_j^{(l)}) \\ &= \frac{1}{N} \sum_{n=1}^N \log \frac{\hat{g}_{\mathbf{X}_{i+1}^{(k+1)}, \mathbf{Y}_j^{(l)};\varepsilon}(\mathbf{x}_{i+1}^{(k+1)}(n), \mathbf{y}_j^{(l)}(n)) \cdot \hat{g}_{\mathbf{X}_i^{(k)};\varepsilon_X}(\mathbf{x}_i^{(k)}(n))}{\hat{g}_{\mathbf{X}_i^{(k)}, \mathbf{Y}_j^{(l)};\varepsilon_{X,Y}}(\mathbf{x}_i^{(k)}(n), \mathbf{y}_j^{(l)}(n)) \cdot \hat{g}_{\mathbf{X}_{i+1}^{(k+1)};\varepsilon_{X,X}}(\mathbf{x}_{i+1}^{(k+1)}(n))} \end{aligned} \quad (3.37)$$

für große  $N$  und kleine  $\varepsilon$  gegen die entsprechenden Entropien bzw. Informationen konvergieren. Dabei wurden in den beiden letzten Gleichungen die bedingten Dichten unter Ausnutzung von Gl. (2.37) durch einfache ersetzt. Die Bandbreitenvektoren  $\varepsilon_X$ ,  $\varepsilon_{X,X}$ ,  $\varepsilon_Y$ ,  $\varepsilon_{X,Y}$  werden aus  $\varepsilon$  gebildet, indem die entsprechenden Komponenten zu einem neuen Vektor zusammengefasst werden. Für feste Bandbreite  $\varepsilon$  ist  $\hat{H}_\varepsilon(\mathbf{X}_i^{(k)})$  auch unter dem Namen *Cohen-Procaccia-Entropie* bekannt [Cohen & Procaccia (1985), Gaspard & Wang (1993)].

Häufig kann ein Experiment nur einmal beobachtet werden. Wie bereits in Absch. 3.1 erläutert wurde, können aus diesem einen Experiment  $\omega_1$  weitere Realisierungen durch Zeitverschiebung geschaffen werden, sofern der ( $d$ -dimensionale) Prozess  $X$  stationär ist, das heißt es wird  $x_i(n) \equiv X_i(\omega_n) = X_n(\omega_1) \equiv x_n$  gesetzt. Dementsprechend erfolgt die Mittelung in Gl. (3.30) sowie in Gl. (3.34) – (3.37) nicht mehr über Beobachtungen, sondern über die Zeit. Aufgrund serieller Korrelationen innerhalb der Zeitreihe entsteht allerdings ein Bias, so dass die Kernschätzer im Limes nicht mehr erwartungstreu sind. Diese Problematik wurde bereits im Zusammenhang mit dem Schätzen von fraktalen Dimensionen von Theiler (1986) diskutiert. Deshalb müssen die Kernschätzer diesbezüglich korrigiert werden. Das von Theiler (1986) vorgeschlagene Verfahren besteht darin, in der Summation in Gl. (3.33) all jene Terme auszulassen, die in der Zeit zu nahe am Referenzzeitpunkt liegen. Der hieraus folgende modifizierte Kernschätzer lautet

$$\hat{g}_{X_i;\varepsilon}(x_i) = \frac{1}{\alpha(i, N, w)} \sum_{\substack{n=1 \\ |n-i|>w}}^N \prod_{j=1}^d \frac{1}{\varepsilon_j} K_j \left( \frac{x_{i,j} - x_{n,j}}{\varepsilon_j} \right), \quad (3.38)$$

wobei der Normierungsfaktor durch

$$\alpha(i, N, w) = N - \min(i + w, N) + \max(i - w, 1) - 1$$

gegeben ist. Das *Theiler-Fenster*  $w$  ist größer oder gleich der Dekorrelationszeit zu setzen, siehe auch [Theiler (1990)].

Für den Kernschätzer  $\hat{g}_{\mathbf{X}_i^{(k)}, \mathbf{Y}_j^{(l)}; \varepsilon}$  ist es naheliegend, die Bandbreiten bezüglich jeder Zeitkomponente von  $\mathbf{X}_i^{(k)}$  gleich  $\varepsilon_X$  und für jede Zeitkomponente von  $\mathbf{Y}_j^{(l)}$  gleich  $\varepsilon_Y$  zu wählen, so dass  $\varepsilon = (\varepsilon_X, \dots, \varepsilon_Y, \dots)$  ist. Wenn die physikalische Natur oder das dynamische Verhalten von  $X$  und  $Y$  unterschiedlich ist, ist es im Allgemeinen schwierig, a priori eine geeignete Wahl für das Verhältnis von  $\varepsilon_X$  und  $\varepsilon_Y$  zu treffen damit der Kernschätzer schnell konvergiert.

Die Kernschätzer haben den Vorteil, dass mit ihnen der Zustandsraum gleichmäßig aufgelöst wird. Des Weiteren sind sie einfach in ein Computerprogramm umzusetzen. Nachteilig ist, dass sie lokal hohe statistische Fehler aufweisen können. Stehen nur endlich viele Datenpunkte zur Verfügung, so werden viele  $\varepsilon$ -Umgebungen nur wenige oder gar keine Nachbarn beinhalten. Dies führt zu statistischen und systematischen Fehlern. Für Rechteckkerne wurde von Grassberger (1988) eine Korrekturformel für die Shannon-Entropie abgeleitet. In erster Linie wird hierzu der Logarithmus durch die Digamma-Funktion  $\psi(x) = \Gamma(x)'/\Gamma(x)$  mit  $\psi(0) = 0$  ersetzt. Ähnliche Korrekturterme sind auch in [Panzeri & Treves (1996), Moddemeijer (1999), Roulston (1999)] zu finden. Für die gemeinsame Information und die Transferentropie wurden diese Korrekturen zum Beispiel in [Schreiber (2000b)] und [Roulston (1997), Roulston (1999)] auf

Gl. (2.10) bzw. Gl. (2.27) angewendet, wobei in den beiden letztgenannten Arbeiten das gleiche Prinzip aber eine anderen Korrektur verwendet wurde. Diese Vorgehensweise birgt aber einen zusätzlichen nachteiligen Effekt in sich, denn die Fluktuationen zum Beispiel in  $g_{\mathbf{X}_i^{(k)}, \mathbf{Y}_j^{(l)}; \varepsilon}$  und  $g_{\mathbf{X}_i^{(k)}; \varepsilon}$  sind aufgrund endlicher Daten nicht unabhängig voneinander, weshalb deren Bias nicht separat voneinander korrigiert werden dürfen. Wie sich herausstellte, ist die Auswirkung dieses Effektes bei der Biaskorrektur sogar noch sehr viel problematischer als der Bias aufgrund der Fluktuationen selbst.

Da keine geeignete Korrektur zur Verfügung steht, werden die unkorrigierten Schätzer verwendet. Zusätzlich werden in der Summe von Gl. (3.35) und Gl. (3.37) all jene Terme Null gesetzt, bei denen weniger als  $n_{\min}$  Punkte zum Schätzen der Dichte zur Verfügung stehen, das heißt es wird angenommen, dass die Dichten in diesem Fall lokal faktorisieren. Hierdurch werden die lokalen Fluktuationen der Kernschätzer reduziert, dafür tritt bei kleinen Bandbreiten ein Bias in Richtung Null auf.

Die Anwendung der Kernschätzer soll an einem Beispiel demonstriert werden. Hierzu wird das in Absch. 3.5, Gl. (3.29) vorgestellte System von gekoppelten autoregressiven Prozessen verwendet, wobei diesmal der Kopplungsparameter mit  $c = 0.4$  fest gewählt wird. Zum Schätzen der Dichten werden Rechteckkerne mit gleicher Bandbreite für beide Signale und alle Komponenten verwendet. Die Zeitreihe hat eine Länge von 10 000 Punkten. Das Theilerfenster  $w$  wurde gleich 20 gewählt. Wie sich herausstellte, erhält man die robustesten Resultate mit  $n_{\min} = 4$ . In Abb. 3.4 sind die gegenseitigen Information  $\hat{M}_\varepsilon(X_{i+1}, Y_i)$  und  $\hat{M}_\varepsilon(Y_{i+1}, X_i)$  als auch die Transferentropien  $\hat{T}_\varepsilon(X_{i+1}|X_i, Y_i)$  und  $\hat{T}_\varepsilon(Y_{i+1}|Y_i, X_i)$  über die Bandbreite  $\varepsilon$  aufgetragen.

Mit abnehmender Bandbreite erreichen alle Kurven ein Plateau nahe am theoretischen Wert. Am unteren Ende setzen Fluktuationen aufgrund endlich vieler Datenpunkte ein. Dies ist für die Transferentropie  $\hat{T}_\varepsilon(X_{i+1}|X_i, Y_i)$  stärker ausgeprägt, was damit zusammenhängt, dass hier bereits Dichten auf dem  $\mathbb{R}^3$  zu schätzen sind. Der theoretische Wert von 0.0923 wird von  $\hat{T}_\varepsilon(X_{i+1}|X_i, Y_i)$  nicht erreicht (durchgezogene Linie). Über eine große Skalenbreite hingegen liefert  $\hat{T}_\varepsilon(Y_{i+1}|Y_i, X_i)$  innerhalb statistischer Fluktuationen einen Wert von 0.

Die Verwendung von parabolischen Kernen lieferte, obwohl sie eine kleinere Varianz als rechteckige Kerne haben [Silverman (1986)], lediglich für  $\hat{M}_\varepsilon(X_{i+1}, Y_i)$  ein geringfügig stärker ausgeprägtes Plateau. Für die anderen Größen hingegen waren die Plateaus deutlich schlechter ausgeprägt als in Abb. 3.4, bzw. existierten noch nicht einmal. Insbesondere ist hier die Unterdrückung der Fluktuationen mit  $n_{\min}$  nicht so effektiv wie bei den rechteckigen Kernen.

Bei den hier verwendeten Schätzern wurde die Bandbreite  $\varepsilon$  fest gewählt und anschließend die Anzahl der Nachbarn gezählt, die in der durch  $\varepsilon$  festgelegten Umgebung um  $x_i(n)$  liegen. Alternativ hierzu kann auch die Anzahl der Nachbarn

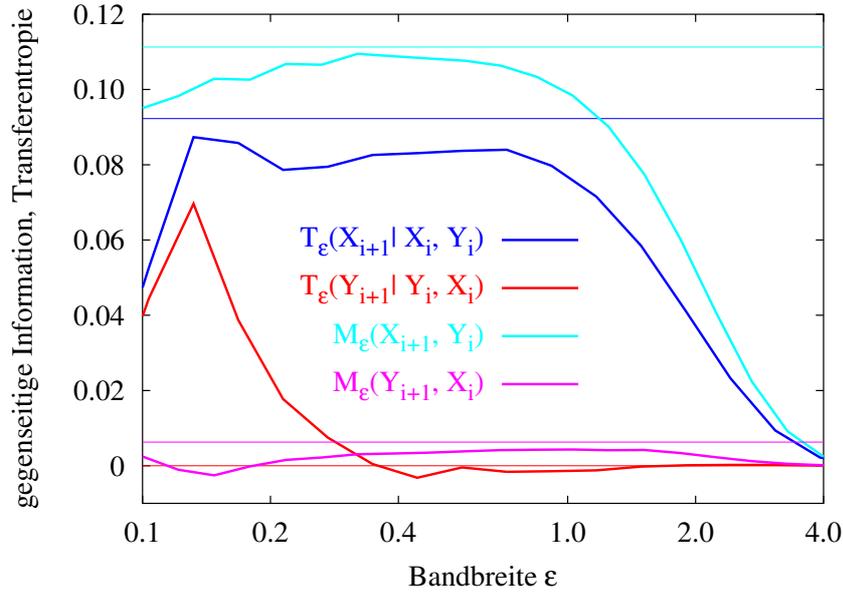


Abbildung 3.4: Gegenseitige Information und Transferentropie in Abhängigkeit der Bandbreite eines Kernschätzers für ein System von autoregressiven Prozessen mit gerichteter Kopplung, Gl. (3.29). Die zwei oberen Kurven repräsentieren  $\hat{M}_\varepsilon(X_{i+1}, Y_i)$  und  $\hat{T}_\varepsilon(X_{i+1}|X_i, Y_i)$ , die zwei unteren  $\hat{M}_\varepsilon(Y_{i+1}, X_i)$  und  $\hat{T}_\varepsilon(Y_{i+1}|Y_i, X_i)$ . Die exakten Werte der gegenseitigen Information und Transferentropie sind als dünnere horizontale Linien eingezeichnet.

$M(x_i(n))$  fest vorgegeben werden und die Umgebung wird so lange vergrößert, bis sie die geforderte Anzahl an Nachbarn enthält. Dann ist

$$\hat{g}_{X_i}(x_i(n)) = \alpha^{-1} \frac{M(x_i(n))}{V(x_i(n))}$$

ebenfalls ein Schätzer für  $g_{X_i}$ . Hier ist  $V(x_i(n))$  das Volumen der Umgebung um  $x_i(n)$  mit  $M(x_i(n))$  Nachbarn und  $\alpha$  ist ein geeigneter Normierungsfaktor. Wird als Umgebung ein Quader gewählt, so ist  $V(x_i(n)) = \prod_{j=1}^d 2\varepsilon_j$ . Dieses Verfahren wird *Fixed-Mass-Methode* genannt [Kantz & Schreiber (1997)]. Wie sich herausstellte, liefert dieser Schätzer keine besseren Resultate als der hier vorgestellte Kernschätzer mit festem  $\varepsilon$ . Stattdessen war seine Anwendung noch sehr viel rechenintensiver. Eine ausführliche Diskussion der Anwendung der Fixed-Mass-Methode auf das hier verwendete Beispielsystem ist in [Kaiser & Schreiber (2002)] zu finden.

### 3.7 Räumliche Abhängigkeit des Windes

Im folgenden Beispiel wird gezeigt, wie die gegenseitige Information angewendet werden kann, um Abhängigkeiten von Windgeschwindigkeiten an verschiedenen Orten in Bodennähe auf kurzen Zeitskalen zu untersuchen. Der hierfür verwendete Datensatz wurde in Lammefjord, Dänemark am Tag 188 aufgezeichnet, siehe [URL (2002)]. Die Messapparatur besteht aus vier Masten, welche auf den relativen Koordinaten (in Metern)

Mast	x [m]	y [m]	Mast	x [m]	y [m]
Mast 1	0	0	Mast 2	-13	15
Mast 3	-17	19	Mast 4	-20	8

positioniert sind, siehe Abb. 3.5. Auf jedem Mast befinden sich in 10 m Höhe über

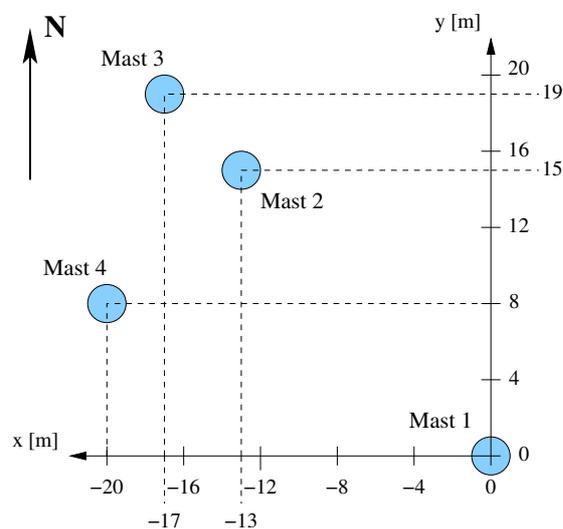


Abbildung 3.5: Position der Messmasten in Lammefjord, Dänemark zur Aufzeichnung der Windgeschwindigkeit und Windrichtung.

NN eine Windfahne und ein Schalenkreuzanemometer, mit denen Windrichtung und Windgeschwindigkeit mit einer Frequenz von 8 Hz abgetastet werden. Die Windgeschwindigkeit wird dabei in Schritten von 0.04 m/s aufgezeichnet.

Die Windrichtungen an allen vier Masten sind für den Tag 188 in Abb. 3.6 dargestellt. Offensichtlich ist an diesem Tag der Wind rechteckig, das heißt er dreht von Nord-Ost über Süd nach West: Zu Beginn der Messung kommt der Wind aus Nord-Ost und dreht in dem Zeitraum von  $t = 250$  min bis 550 min nach Süd, wobei er um  $t = 400$  min aus Osten und um  $t = 500$  min aus Süd-Osten weht. Bis  $t = 800$  min ändert sich die Windrichtung zunächst nicht. Anschließend dreht der Wind auf West-Süd-West bis West. Bei den Sprüngen in den Daten von Mast 1, 2 und 3 von  $\approx 45^\circ$  auf  $\approx 355^\circ$  handelt es sich um rückdrehende Windböen.

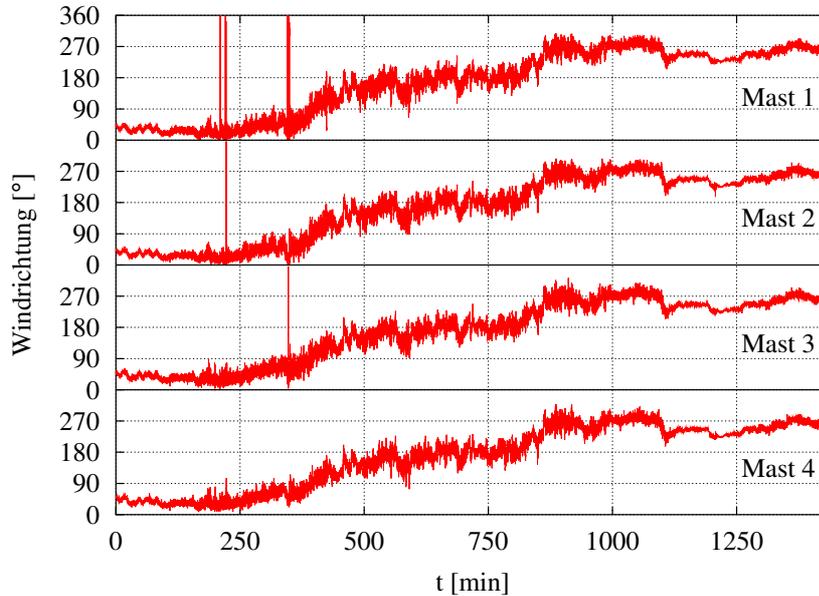


Abbildung 3.6: Windrichtungen an den vier Messmasten in Lammefjord, Dänemark am Tag 188.

Die Windgeschwindigkeiten an den vier Masten sind in Abb. 3.7 aufgetragen. Wie der Grafik zu entnehmen ist, hat die Windgeschwindigkeit über längere Zeiträume jeweils unterschiedliche Drifts und unterschiedlich stark ausgeprägte Fluktuationen. Folglich handelt es sich bei den Windgeschwindigkeiten  $(V_i^{(k)})_{i \in \mathbb{N}_0}$ , wobei  $k = 1, \dots, 4$  die Mastnummer angibt, um einen nichtstationären Prozess.

Vergleicht man  $V_i^{(1)}$  bis  $V_i^{(4)}$  in Abb. 3.7, so verhalten sich alle vier Windgeschwindigkeiten im Mittel gleich, was aufgrund der Nähe der Masten auch zu erwarten ist. Anders sieht es hingegen aus, wenn die Geschwindigkeitswerte auf kleineren Zeitskalen, das heißt Zeitspannen von 50 Minuten und weniger, betrachtet werden, siehe Abb. 3.8. Dann ist der Drift in der mittleren Geschwindigkeit im Vergleich zu den Fluktuationen meist vernachlässigbar. Bei genauerer Betrachtung fällt auf, dass die Form und der Zeitpunkt der Minima und Maxima fluktuieren. Auf noch kleineren Zeitskalen ist eine Korrelation zwischen den Windgeschwindigkeiten an den Masten visuell nicht mehr nachweisbar.

Um Abhängigkeiten zwischen den Windgeschwindigkeiten an den vier Masten zu untersuchen, müssen aufgrund der Nichtstationarität und insbesondere des Drifts in den Geschwindigkeiten kleine Zeitfenster betrachtet werden, innerhalb derer der Prozess als ergodisch angenommen werden kann. Andererseits dürfen die Fenster auch nicht zu klein gewählt werden, da sonst die statistischen Fluktuationen der Schätzer zu groß werden. Zur Analyse der Abhängigkeiten wird hier die gegenseitige Information verwendet, da sie im Gegensatz zur Transferentropie

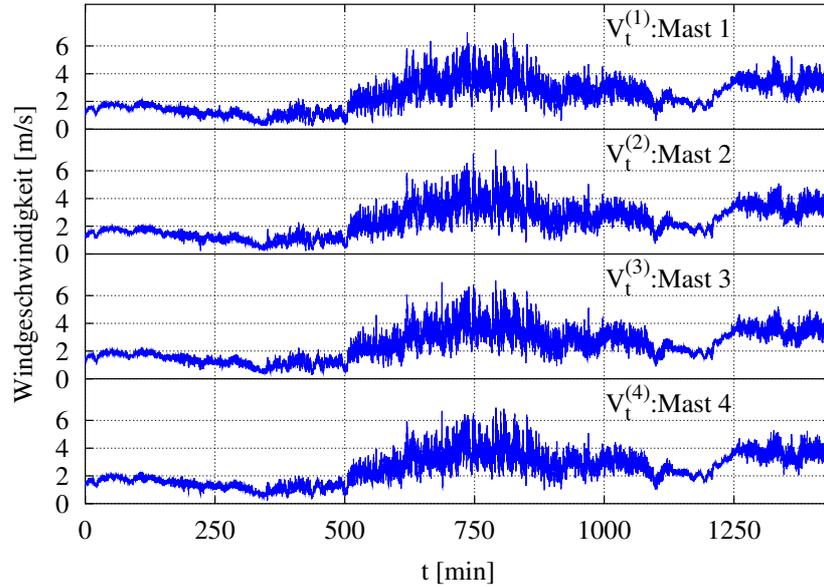


Abbildung 3.7: Windgeschwindigkeiten an den vier Messmasten in Lammefjord, Dänemark am Tag 188.

bei kurzen Zeitreihen deutlich robustere Ergebnisse liefert, vor allem da die Geschwindigkeiten sich nur mit Markov-Prozessen höherer Ordnung approximieren lassen.

Dementsprechend wird folgende Prozedur zur Untersuchung der Windgeschwindigkeiten verwendet: Zunächst wird ein Zeitfenster  $W_t = [t - \Delta, t + \Delta]$  um den Zeitpunkt/Datenpunkt  $t$  mit einer Breite von  $2\Delta$  gewählt, wobei hier  $\Delta$  gleich 10 000 Datenpunkte gesetzt wurde. Da die Abtastrate 8 Hz beträgt, entspricht dies einer Fensterbreite von 41 min und 40 s. Auf diesem Fenster wird die gegenseitige Information  $M(V_i^{(4)}, V_{i+\tau}^{(k)})|_{W_t}$  als Funktion der Zeitverschiebung  $\tau$  zwischen den Windgeschwindigkeiten des 4. Masts  $V_i^{(4)}$  und des  $k$ -ten Masts  $V_{i+\tau}^{(k)}$  gemessen, wobei  $k = 1, 2, 3$  ist. Anschließend wird das Zeitfenster um  $\Delta$  Zeitschritte verschoben und erneut die gegenseitigen Information  $M(V_i^{(4)}, V_{i+\tau}^{(k)})|_{W_{t+\Delta}}$  als Funktion der Zeitverschiebung  $\tau$  berechnet. Da die Windgeschwindigkeiten innerhalb der verschiedenen Fenster  $W_t$  eine unterschiedliche Varianz aufweisen, ist es angebracht, die gegenseitigen Informationen zu normieren, bevor sie miteinander verglichen werden.  $M(V_i^{(4)}, V_{i+\tau}^{(k)})|_{W_t}$  ist aufgrund der diskreten Geschwindigkeitswerte nach oben durch die Shannon-Entropie  $H(V_i^{(4)})|_{W_t}$  von  $V_i^{(4)}$  beschränkt, weshalb sie ein geeigneter Normierungsfaktor ist. Die normierte ge-

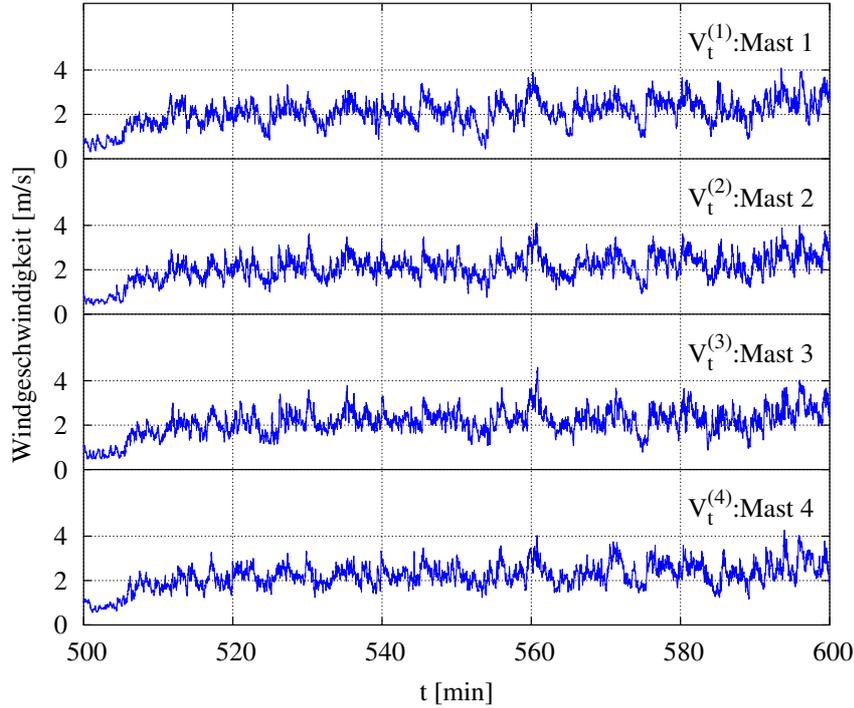


Abbildung 3.8: Windgeschwindigkeiten an den vier Messmasten in Lammefjord, Dänemark am Tag 188 innerhalb eines Zeitraums von 100 Minuten.

gegenseitige Information lautet somit

$$m(V_i^{(4)}, V_{i+\tau}^{(k)})|_{W_t} = \frac{M(V_i^{(4)}, V_{i+\tau}^{(k)})|_{W_t}}{H(V_i^{(4)})|_{W_t}} = 1 - \frac{H(V_i^{(4)} | V_{i+\tau}^{(k)})|_{W_t}}{H(V_i^{(4)})|_{W_t}}, \quad (3.39)$$

wobei die Beziehung (2.17) ausgenutzt wurde. Sind die Geschwindigkeitswerte von zwei Masten bezüglich der Zeitverschiebung  $\tau_0$  am stärksten voneinander abhängig, so ist  $m(V_i^{(4)}, V_{i+\tau}^{(k)})|_{W_t}$  in  $\tau = \tau_0$  maximal und wird im Allgemeinen mit steigendem  $|\tau - \tau_0|$  auf Null abfallen, wobei die Abnahme umso schneller ist, je weniger stark die jeweiligen Windgeschwindigkeitswerte in der Zeit korreliert sind. Gilt sogar  $V_i^{(4)} = \varphi(V_{i+\tau_0}^{(k)})$ , wobei  $\varphi$  eine stetige Abbildung ist, dann ist  $m(V_i^{(4)}, V_{i+\tau_0}^{(k)})|_{W_t} = 1$ . Für unabhängige  $V_i^{(4)}$  und  $V_{i+\tau}^{(k)}$  ist hingegen  $m(V_i^{(4)}, V_{i+\tau}^{(k)})|_{W_t} = 0$  für alle  $\tau$ .

In Abb. 3.9 sind die normierten gegenseitigen Informationen  $m(V_i^{(4)}, V_{i+\tau}^{(k)})|_{W_t}$  zwischen dem 4. Mast und dem 1., 2. bzw. 3. Mast als Funktion der Zeitverschiebung  $\tau$  für verschiedene Zeitfenster  $W_t$  gegeben. Zunächst fällt auf, dass die gegenseitigen Informationen trotz Normierung teilweise erhebliche Sprünge von einem Zeitfenster zum nächsten machen. Dies wird durch die starke Nichtstationarität des Windes verursacht. Desweiteren liegen die Werte der gegenseitigen

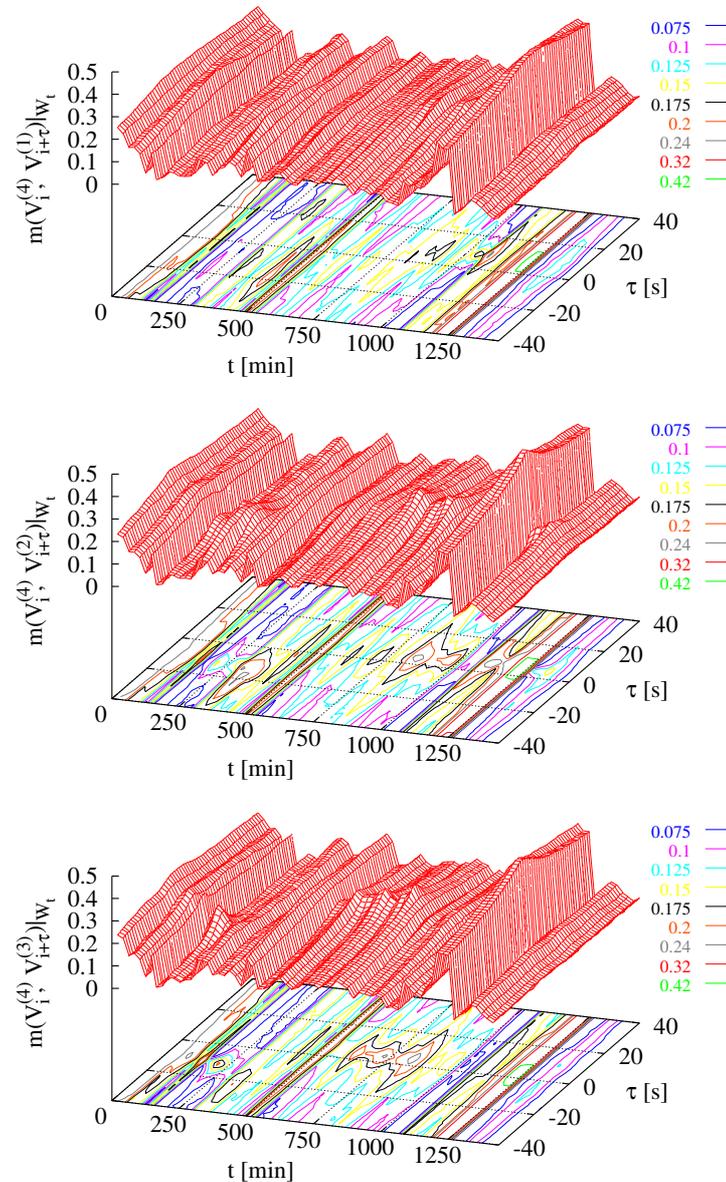


Abbildung 3.9: Normierte gegenseitige Information  $m(V_i^{(4)}, V_{i+\tau}^{(k)})|_{W_t}$  als Funktion der Zeitverschiebung  $\tau$  für verschiedene Zeitfenster  $W_t$  zwischen dem 4. und dem 1. (oben), 2. (mitte) bzw. 3. Mast (unten). Die Zeitfensterbreite beträgt 20 000 Datenpunkte, dies entspricht 40 min und 40 s.

Information zwischen 0 und 0.5. Ein Grund hierfür dürfte sein, dass das Zeitfenster noch zu groß gewählt wurde, um die Windgeschwindigkeiten als ergodisch annehmen zu können. So macht die gegenseitige Information selbst bei kleineren Zeitfenstern mit halber Breite ( $2 \Delta = 10\,000$ ) von einem Zeitfenster zum nächsten erhebliche Sprünge. Da andererseits die Fluktuationen des Schätzers der gegenseitigen Information und der Entropie für kleine Fenster größer werden, wurden hier die breiteren Zeitfenster verwendet. Somit sind die geschätzten Abhängigkeiten als zeitlich gemittelte Abhängigkeiten aufzufassen.

Desweiteren sind die Maxima der gegenseitigen Informationen eines Mastpaares innerhalb der verschiedenen Zeitfenster unterschiedlich stark ausgeprägt. So sind bei  $t \approx 1220$  min, wenn der Wind aus Süd-Westen auffrischt und die Windgeschwindigkeiten kleine Varianzen aufweisen, die gegenseitigen Informationen für alle Mastpaare selbst bei größeren Zeitverschiebungen  $\tau$  relativ hoch. Hingegen zeigt sich bei  $t \approx 750$  min, wenn der Wind aus Süden kommt und die Varianz der Windgeschwindigkeit groß ist, ein ausgeprägtes Maximum in der gegenseitigen Information zwischen  $V_i^{(4)}$  und  $V_i^{(2)}$  bzw.  $V_i^{(3)}$ . Offensichtlich spielen hier die Böigkeit, die Windrichtung und Homogenität des Winds auf dem Areal der Messanordnung eine wesentliche Rolle.

Im Folgenden soll die Zeitverschiebung  $\tau_0(V_i^{(4)}, V_{i+\tau}^{(k)})|_{W_t}$ , bei der  $m(V_i^{(4)}, V_{i+\tau}^{(k)})|_{W_t}$  maximal wird, als Funktion der Zeit  $t$  und damit als Funktion der Windrichtung genauer untersucht werden. Weht der Wind relativ homogen, so ist zu erwarten, dass  $\tau_0(V_i^{(4)}, V_{i+\tau}^{(k)})|_{W_t}$  positiv ist, falls der Wind von  $V_i^{(4)}$  nach  $V_i^{(k)}$  weht, und negativ bei entgegengesetzter Windrichtung. Dabei sollte der Absolutwert von  $\tau_0$  in etwa umgekehrt proportional zur Windgeschwindigkeit sein. Bei schräg einfallendem Wind sollte das Vorzeichen von  $\tau$  unverändert bleiben, der Wert aber abnehmen und schließlich Null werden, wenn der Wind senkrecht zur Verbindungslinie von  $V_i^{(4)}$  nach  $V_i^{(k)}$  einfällt. Ist der Wind hingegen räumlich inhomogen, zum Beispiel aufgrund von starken Turbulenzen, dann wird das Maximum in der gegenseitigen Information sehr schwach ausgeprägt sein und in Folge der statistischen Fluktuationen des Schätzers wird  $\tau_0(V_i^{(4)}, V_{i+\tau}^{(k)})|_{W_t}$  willkürlich hin und her springen.

In Abb. 3.10 ist  $\tau_0(V_i^{(4)}, V_{i+\tau}^{(k)})|_{W_t}$  als Funktion des Zeitindex  $t$  für  $k = 1, 2, 3$  gegeben. Als Erstes wird das Verhalten von  $\tau_0(V_i^{(4)}, V_{i+\tau}^{(3)})|_{W_t}$  als Funktion der Zeit  $t$  untersucht. Bis zur Zeit  $t = 300$  min, also bei Wind aus Nord-Ost, siehe Abb. 3.6, ist  $\tau_0(V_i^{(4)}, V_{i+\tau}^{(3)})|_{W_t}$  negativ, entsprechend der relativen Lage von Mast 3 und 4 zum Wind, siehe Abb. 3.5. Dabei wächst  $|\tau_0|$  im Laufe der Zeit an, was mit der abnehmenden Windgeschwindigkeit zu erklären ist, siehe Abb. 3.7. Anschließend beginnt  $\tau_0(V_i^{(4)}, V_{i+\tau}^{(3)})|_{W_t}$  um Null zu fluktuieren. In diesem Zeitraum weht der Wind in etwa senkrecht zur Verbindungslinie der Masten 3 und 4; die gegenseitige Information zeigt relativ niedrige Werte bei allen Zeitverzögerungen, siehe Abb. 3.9. Bei  $t = 500$  min schläft der Wind kurzfristig ein, die gegenseitige Information macht einen Sprung, dessen Höhe nahezu unabhängig von den Zeit-

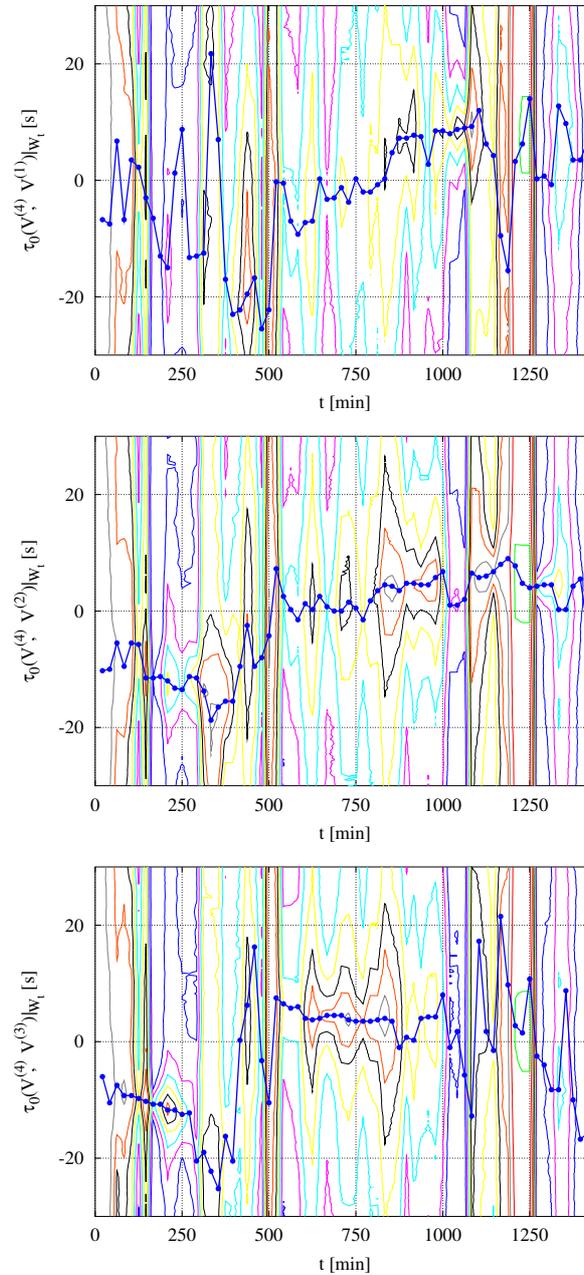


Abbildung 3.10: Zeitverzögerung  $\tau_0(V_i^{(4)}, V_{i+\tau}^{(k)})|_{W_t}$ , bei der  $m(V_i^{(4)}, V_{i+\tau}^{(k)})|_{W_t}$  maximal ist, als Funktion der Zeit  $t$  (breite Linie mit Kreisen) für  $k = 1$  (oben),  $k = 2$  (mitte),  $k = 3$  (unten). Die Breite von  $W_t$  beträgt 20 000 Datenpunkte, dies entspricht 40 min und 40 s. Feine Linien: Höhenlinien von  $m(V_i^{(4)}, V_{i+\tau}^{(k)})|_{W_t}$  als Funktion von  $\tau$  und  $t$ . Die Farbkodierung ist in Abb. 3.9 gegeben.

verschiebungen ist. Dann frischt er aus Süd-Ost wieder auf und dreht nach Süden. Dementsprechend ist  $\tau_0(V_i^{(4)}, V_{i+\tau}^{(3)})|_{W_t}$  in dem Zeitraum von 510 min bis 800 min positiv, wobei die Absolutwerte aufgrund der höheren mittleren Windgeschwindigkeit kleiner sind als zu Beginn der Messung. Danach dreht der Wind nach West und flaut ab ( $t = 1000$  min). Jetzt sind die gegenseitigen Informationen relativ klein und die Zeitverschiebungen  $\tau_0$  fluktuieren um Null. In dem kurzen Zeitraum von  $t = 1210$  min bis 1250 min kommt es zu einem erneuten Auffrischen des Windes, die gegenseitige Information nimmt enorm zu und  $\tau_0(V_i^{(4)}, V_{i+\tau}^{(3)})|_{W_t}$  hat positive Werte, die im Laufe der Zeit aufgrund der größer werdenden Windgeschwindigkeit abnehmen. Im Anschluss hieran beginnt die Zeitverzögerung  $\tau_0$  wieder um Null zu fluktuieren. Ebenso ist die gegenseitige Information minimal.

$\tau_0(V_i^{(4)}, V_{i+\tau}^{(2)})|_{W_t}$  als Funktion der Zeit  $t$  zeigt qualitativ ein ähnliches Verhalten wie  $\tau_0(V_i^{(4)}, V_{i+\tau}^{(3)})|_{W_t}$ . Hier soll nur auf die wesentlichen Unterschiede eingegangen werden. In dem Zeitintervall von 520 min bis 760 min ist  $\tau_0(V_i^{(4)}, V_{i+\tau}^{(2)})|_{W_t}$  nahezu Null und zeigt nur geringe Fluktuationen. In diesem Zeitraum weht der Wind aus Süd-Ost bis Süd und die Masten 2 und 4 sind in Bezug auf den Wind nahezu auf gleicher Höhe, siehe Abb. 3.5. Desweiteren zeigt die gegenseitige Information ein ausgeprägtes Maximum in der Zeit bevor der Wind um  $t = 1210$  min auffrischt. Dies hängt ebenfalls mit der relativen Lage der beiden Masten zum Wind zusammen.

Betrachtet man die Masten 1 und 4, so fluktuiert  $\tau_0(V_i^{(4)}, V_{i+\tau}^{(1)})|_{W_t}$  zunächst, da der Wind in etwa senkrecht zur Verbindungsline der Masten 1 und 4 einfällt. Erst bei östlichem bis süd-östlichem Wind, also zwischen  $t = 400$  min bis 500 min, hat die gegenseitige Information ein ausgeprägteres Maximum und die Zeitverschiebung  $\tau_0(V_i^{(4)}, V_{i+\tau}^{(1)})|_{W_t}$  ist negativ. Im darauf folgenden Zeitraum bleibt  $\tau_0$  negativ und nahe bei Null. Erst wenn der Wind über Süd-West nach West gedreht hat ( $t = 800$  min), wird diese Zeitverschiebung positiv. Eine Ausnahme stellt hier der Zeitraum kurz vor dem Auffrischen des Windes bei  $t = 1220$  min da. Hier ist  $\tau_0(V_i^{(4)}, V_{i+\tau}^{(1)})|_{W_t} < 0$ . Insgesamt sind die Maxima in der gegenseitigen Information  $m(V_i^{(4)}, V_{i+\tau}^{(1)})|_{W_t}$  weniger stark ausgeprägt als in  $m(V_i^{(4)}, V_{i+\tau}^{(2)})|_{W_t}$  oder  $m(V_i^{(4)}, V_{i+\tau}^{(3)})|_{W_t}$ . Dies ist unter anderem auf den größeren Abstand zwischen den Masten 1 und 4 zurückzuführen.

Die Analyse von Winddaten stellt eine große Herausforderung dar, denn es handelt sich um nichttriviale Zeitreihen, bei denen man mit allen Schwierigkeiten der Realität konfrontiert wird. Aufgrund von Turbulenzen sind hier der zeitabhängige Drift, die unterschiedlichen Fluktuationen in den Daten und insbesondere die starke Nichtstationarität der zugrunde liegenden Prozesse zu nennen. Trotz dieser Umstände lässt sich mit der gegenseitigen Information die zu erwartenden Abhängigkeiten bzw. Korrelationen in den Signalen sowie deren Zeitabhängigkeit zeigen. Schwierigkeiten bereitet die Normierung. Zwar ist die gegenseitige Information durch die Shannon-Entropie nach oben beschränkt, aber selbst auf ihr normiert macht die gegenseitige Information von einem Zeitfenster

zum nächsten teilweise erhebliche Sprünge. Mit der Transferentropie können keine Aussagen gemacht werden, da sie eine sehr viel aufwendigere Statistik benötigt. Zu erwarten wäre, dass die Transferentropie in Kopplungsrichtung entgegen dem Wind den Wert Null liefert und in Kopplungsrichtung mit dem Wind einen positiven Wert. Desweiteren sollte die Zeit, welche die Windfront von einem Mast zum nächsten benötigt, durch ein Maximum in der Transferentropie als Funktion der Zeitverzögerung gegeben sein, wenn beide Signale in der Zeit gegeneinander verschoben werden.

# Kapitel 4

## Exkurs: Dynamische Systeme

### 4.1 Entropie eines dynamischen Systems

Eine Vielzahl von physikalischen, mathematischen oder biologischen Phänomenen wird durch gewöhnliche Differentialgleichungssysteme

$$\dot{x}(t) = F(x(t)) \quad (4.1)$$

mit Anfangswerten  $x(0) = x_0$  beschrieben, ( $x(t) \in \mathbb{R}^d$ ). Ist  $F$  Lipschitz-stetig, so existiert eine Lösung  $x(t, x_0)$  für das System, wobei  $x(t, x_0)$  für jeden Zeitpunkt  $t$  eindeutig durch die Anfangsbedingung  $x(0) = x_0$  festgelegt ist. Diese deterministischen Prozessen waren in den bisherigen Betrachtungen weitgehend ausgeschlossen.

In diesem Kapitel werden die wesentlichen Ideen vorgestellt, mit denen es möglich ist, deterministische Prozesse als eine spezielle Klasse von stochastischen Prozessen zu betrachten. Hierauf aufbauend wird anschließend gezeigt, wie sie mit Techniken der Informationstheorie studiert werden können.

Ein allgemeines mathematisches Konzept, mit dem deterministische Prozesse beschrieben werden, sind die *dynamischen Systeme*, [Katok & Hasselblatt (1995), Ott (1993), Schuster (1989)]. Hierunter versteht man eine Familie von Abbildungen  $(X_t)_{t \in \Gamma}$  mit den Eigenschaften

1.  $X_t : M \longrightarrow M \subset \mathbb{R}^d$ ,  $\forall t \in \Gamma$ ,
2.  $X_0(x) = \text{id}(x) = x$ ,  $\forall x \in M$ ,
3.  $X_t(X_s(x)) = X_{t+s}(x)$ ,  $\forall s, t \in \Gamma, x \in M$ ,
4.  $X_t : (M, \mathcal{B}(M)) \rightarrow (M, \mathcal{B}(M))$   $\forall t \in \Gamma$ ,
5.  $t \rightarrow X_t(x_0)$  stetig  $\forall x_0 \in M$ ,

wobei  $\mathcal{B}(M)$  die Borelsche  $\sigma$ -Algebra auf  $M$  ist, siehe hierfür auch [Reitmann (1996), Walters (1981)]. Die kompakte Teilmenge  $M$  von  $\mathbb{R}^d$  wird *Zustands-* oder *Phasenraum* genannt. Ist die Zeitmenge  $\Gamma = \mathbb{R}$  oder  $\mathbb{R}_+$ , so spricht

man von einem *zeitkontinuierlichen dynamischen System*, ist  $\Gamma = \mathbb{Z}$  oder  $\mathbb{Z}_+$ , so wird das System als *zeitdiskret* bezeichnet. Ist die Zeitmenge  $\Gamma = \mathbb{R}$ , so wird das dynamische System auch *Fluß* genannt, ist  $\Gamma = \mathbb{R}_+$ , so spricht man von einem *Halbfluß* [Reitmann (1996)].

Offensichtlich ist ein dynamisches System *invertierbar*, wenn es sich um ein Fluß handelt oder wenn  $\Gamma = \mathbb{Z}$  ist. Denn dann existiert zu jedem  $X_t$  auch die inverse Abbildung  $X_t^{-1} = X_{-t}$ . Dies folgt aus der Gruppeneigenschaft in Gl. (4.2):  $X_{-t}(X_t(x)) = X_0(x) = x$ .

Wenn dynamische Systeme beobachtet werden, so erfolgt dies im Allgemeinen nur zu diskreten Zeitpunkten, wobei die Zeitpunkte einen gleichen Abstand haben. Dieser Abstand wird von der zeitlichen Auflösung der Messapparatur bestimmt. Bei dem so beobachteten System handelt es sich ebenfalls um ein dynamisches System aber diesmal mit diskreter Zeitmenge. So kann jeder Fluß  $(X_t)_{t \in \mathbb{R}}$  in ein zeitdiskretes dynamisches System  $(\tilde{X}_i)_{i \in \mathbb{Z}} \equiv (X_{t_i})_{i \in \mathbb{Z}}$  transformiert werden, wenn äquidistante, diskrete Zeitpunkte  $t_i$  mit  $t_{i+1} - t_i = \Delta$  für alle  $i \in \mathbb{Z}$  betrachtet werden. Offensichtlich erfüllt  $(\tilde{X}_i)_{i \in \mathbb{Z}}$  die Eigenschaften 1 – 4. Insbesondere ist  $\tilde{X}_{i+1}(x) = X_{t_{i+1}}(x) = X_\Delta(\tilde{X}_i(x))$  für alle  $i \in \mathbb{Z}$ . Im Folgendem werden nur dynamische Systeme mit diskreter Zeitmenge betrachtet.

Wird ein dynamisches System beobachtet, so sind zwar seine Zustände  $X_i(x_0) = x_i$  zu jedem Zeitpunkt  $i$  durch den Anfangswert  $X_0(x_0) = x_0$  eindeutig festgelegt, aber die Anfangswerte selbst werden im Allgemeinen zufällig ausgewählt sein. Aus diesem Grund ist es naheliegend,  $M$  als Ereignisraum zu betrachten, um hierauf aufbauend einen stochastischen Prozess zu konstruieren, der das dynamische System repräsentiert.

Hierfür wird  $M$  mit der Borelschen  $\sigma$ -Algebra  $\mathcal{B}(M)$  versehen. Desweiteren sei  $P$  ein Wahrscheinlichkeitsmaß auf  $\mathcal{B}(M)$ , welches die Verteilung der Anfangswerte  $x_0 \in M$  bei Beobachtungen festlegt. Dann ist  $X = (X_i)_{i \in \Gamma}$  mit  $\Gamma = \mathbb{Z}, \mathbb{Z}_+$  ein stochastischer Prozess mit den in Gl. (4.2) gegebenen Eigenschaften.

Die Verteilungen von  $X$  sind durch  $P^{X_i}$  gegeben, wobei die Wahrscheinlichkeit, mit der das dynamische System zur Zeit  $i$  in einem der Zustände aus  $B \in \mathcal{B}(M)$  zu finden ist, durch  $P^{X_i}(B) = P(X_i^{-1}(B))$  gegeben ist. Insbesondere ist  $P^{X_0} = P$ . Besitzen die Verteilungen  $P^{X_i}$ ,  $i \in \Gamma$  eine Dichte, so kann das dynamische System wie andere stochastische Prozesse mit der kontinuierlichen Shannon-Entropie Gl. (2.31) charakterisiert werden.

Genau genommen hängt diese Charakterisierung von der Anfangsverteilung  $P$  und von der Dynamik des Systems ab, an Letzterem ist man aber nur interessiert. Deshalb werden nur solche Systeme betrachtet, bei denen die Anfangsverteilung  $P$  so gewählt werden kann, dass sich die Wahrscheinlichkeit  $P^{X_i}(B)$ , mit der das System in einem Gebiet  $B \in \mathcal{B}(M)$  des Phasenraums beobachtet wird, im Laufe der Zeit nicht verändert [Katok & Hasselblatt (1995), Ott (1993), Reitmann (1996), Walters (1981)]:

$$P^{X_i}(B) = P(X_i^{-1}(B)) \equiv P(B) = P^{X_0}(B), \quad \forall i \in \Gamma. \quad (4.3)$$

Solch ein Maß  $P$  wird *invariant unter  $X$*  genannt. Aus Gl. (4.3) folgt, dass das invariante Wahrscheinlichkeitsmaß  $P$  den stochastischen Prozess, welcher dem dynamischen System zugeordnet ist, zu einem stationär stochastischen Prozess macht.

Viele dynamische Systeme, vor allem solche mit chaotischer Dynamik, haben die Eigenschaft, dass sie ihr invariantes Maß  $P$  selbst erzeugen. Wird für solche Systeme eine beliebige Anfangsverteilung  $P_0$  vorgegeben, so konvergiert  $P_i \equiv P_0^{X_i}$  schwach gegen  $P$ :

$$P_i \equiv P_0^{X_i} \implies P \equiv P^{X_0} \equiv P^{X_i} \quad \text{für } i \rightarrow \infty,$$

siehe [Reitmann (1996), Walters (1981)] und [Bauer (1992), Bauer (1991)] für schwache Konvergenz. Dieses invariante Maß  $P$  wird auch *natürliches invariantes (Wahrscheinlichkeits-)Maß* genannt. Nach dem Portmanteau-Theorem [Bauer (1991), Billingsley (1995)] bedeutet dies, dass  $\lim_{i \rightarrow \infty} P_i(A) = P(A)$  für alle  $A \in \mathcal{B}(M)$  gilt, deren Ränder  $\partial A$  keine Masse tragen,  $P(\partial A) = 0$ .

Besitzt das invariante Maß bzw. das natürliche invariante Maß  $P$  eine Lebesgue-Dichte  $g_{X_i}$ , so kann das dynamische System mit der kontinuierlichen Shannon-Entropie Gl. (2.31) charakterisiert werden.

Wird ein dynamisches System nur einmal beobachtet, so dass das invariante Maß aus einer einzigen Realisierung, die zum Beispiel in Form einer Zeitreihe vorliegt, geschätzt werden soll, so ist dies nur möglich, wenn der Prozess ergodisch ist. Dies bedeutet, dass der Erwartungswert einer jeden integrierbaren Funktion  $f$  bezüglich  $P$  durch Zeitmittelung berechnet werden kann, siehe Absch. 3.1, Gl. (3.2). Insbesondere erhält man  $P(A) = P^{X_0}(A)$ , indem in Gl. (3.2)  $f$  gleich der Indikatorfunktion  $\mathbf{1}_A$  gesetzt wird.

Mit dem invarianten natürlichen Maß  $P$  ist das dynamische System  $X$  charakterisiert, seine Dynamik, die regulär oder chaotisch sein kann, hingegen nicht. Bei einer regulären Bewegung befinden sich alle *Trajektorien (Pfade)*  $i \rightarrow X_i(x'_0)$  mit Anfangswerten  $x'_0$  aus einer  $\varepsilon_0$ -Umgebung um  $x_0$  nach  $i$  Zeitschritten in einer  $\varepsilon_i$ -Umgebung um  $X_i(x_0)$ , wobei die Umgebung langsamer wächst als exponentiell,  $1/i \log \varepsilon_i \rightarrow 0$  für  $i \rightarrow \infty$ . Bei chaotischen Bewegungen wächst der Abstand von zwei Trajektorien, deren Anfangswerte nur infinitesimal auseinander liegen, im Mittel exponentiell mit der Zeit, siehe Abb. 4.1. Die Geschwindigkeit, mit der der Abstand pro Zeiteinheit zunimmt, wird durch den größten Lyapunov-Exponenten  $\lambda$  bestimmt, siehe beispielsweise [Katok & Hasselblatt (1995), Ott (1993), Reitmann (1996), Schuster (1989)]. Bei einer zufälligen Bewegung läuft, im Gegensatz zu deterministischen Bewegungen, eine  $\delta$ -Verteilung auseinander. Somit sind die Zustände von Trajektorien mit gleichem Anfangswert bereits nach kurzen Zeiten über den gesamten Zustandsraum verstreut.

Die Dynamik selbst kann nicht mit der kontinuierlichen bedingten Shannon-Entropie charakterisiert werden, da aufgrund des Determinismus die Übergangs-

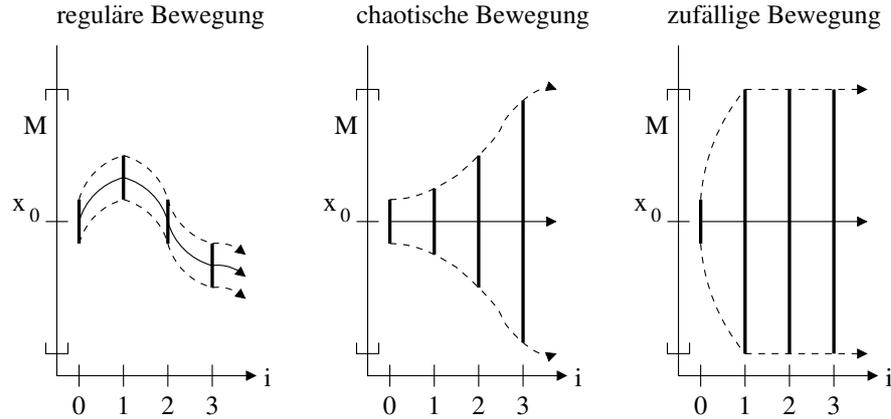


Abbildung 4.1: Schematische Darstellung von regulärer, chaotischer und zufälliger Bewegung.

verteilungen als Lebesgue-Dichte lediglich Diracsche Delta-Distributionen besitzen. Solch eine Charakterisierung kann aber mit der von Kolmogorov eingeführten *metrischen Entropie*, die auch *Kolmogorov-Sinai-Entropie* genannt wird, erfolgen [Katok & Hasselblatt (1995), Ott (1993), Reitmann (1996), Walters (1981)].

Der grundlegende Gedanke bei der Konstruktion der Kolmogorov-Sinai-Entropie ist folgender: Angenommen, der beschränkte Zustandsraum  $M$  ist in  $r$  beliebige, nichtleere Untermengen  $W_i \in \mathcal{B}(M)$ ,  $i = 1, \dots, r$  mit  $\cup_{i=1}^r W_i = M$  und  $W_i \cap W_j = \emptyset$  für  $i \neq j$  partitioniert. Die Menge  $\mathcal{W} = \{W_1, \dots, W_r\}$  wird *Partition* des Zustandsraums genannt. Dann kann jeder Produktabbildung  $\mathbf{X}_{n-1}(x_0) = (X_0(x_0), \dots, X_{n-1}(x_0))$  die Symbolsequenz  $\mathbf{S}_{n-1}^{(n)}(x_0) = (S_0(x_0), \dots, S_{n-1}(x_0)) \in \{1, \dots, r\}^n$  zugeordnet werden, wobei  $S_i(x_0) = j$  falls  $X_i(x_0) \in W_j$  ist.

Als Nächstes werden zwei Trajektorien eines Systems  $X_i(x_0)$  und  $X_i(x'_0)$  mit  $|x_0 - x'_0| \ll 1$  betrachtet. Handelt es sich um Trajektorien eines chaotischen Systems, so werden diese bis zu einem  $\tilde{n} < \infty$  die gleiche Symbolsequenz haben, bevor sie aufgrund des chaotischen Verhaltens so weit auseinander gelaufen sind, dass  $\mathbf{S}_{n-1}^{(n)}(x_0) = \mathbf{S}_{n-1}^{(n)}(x'_0)$  für  $n \leq \tilde{n}$  und  $\mathbf{S}_{n-1}^{(n)}(x_0) \neq \mathbf{S}_{n-1}^{(n)}(x'_0)$  für  $n > \tilde{n}$  gilt. Somit können chaotische Trajektorien für  $n \rightarrow \infty$  mittels Symbolsequenzen unterschieden werden.

Ist  $P$  das invariante Wahrscheinlichkeitsmaß von  $X$ , dann ist die Verteilung der Symbolsequenzen aus  $n$  Zeitschritten durch

$$\begin{aligned} P\{\mathbf{S}_{n-1}^{(n)} = \mathbf{s}_{n-1}^{(n)}\} &= P\{\mathbf{X}_{n-1}^{(n)} \in \prod_{i=0}^{n-1} W_{k_i}\} \\ &= P\{W_{k_0} \cap X_1^{-1}(W_{k_1}) \cap \dots \cap X_{n-1}^{-1}(W_{k_{n-1}})\} \quad (4.4) \end{aligned}$$

gegeben,  $W_{k_0}, \dots, W_{k_{n-1}} \in \mathcal{W}$ . Aus Gl. (2.13) folgt für die Übergangswahrschein-

lichkeit, mit der  $X_n$  in einem Partitionselement  $W_{k_n}$  zu finden ist,

$$\begin{aligned} & P\{X_n \in W_{k_n} | \mathbf{X}_{n-1}^{(n)} \in \prod_{i=0}^{n-1} W_{k_i}\} \\ &= P\{X_n \in W_{k_n} | W_{k_0} \cap X_1^{-1}(W_{k_1}) \cap \dots \cap X_{n-1}^{-1}(W_{k_{n-1}})\} \\ &= \frac{P\{W_{k_0} \cap X_1^{-1}(W_{k_1}) \cap \dots \cap X_{n-1}^{-1}(W_{k_{n-1}}) \cap X_n^{-1}(W_{k_n})\}}{P\{W_0 \cap X_1^{-1}(W_{k_1}) \cap \dots \cap X_{n-1}^{-1}(W_{k_{n-1}})\}}. \end{aligned} \quad (4.5)$$

Werden diese Wahrscheinlichkeiten in die bedingte Shannon-Entropie, Gl. (2.15), eingesetzt, wobei jetzt über die Partitionselemente zu summieren ist, so erhält man die mittlere Unsicherheit, mit der die zukünftigen Zustände bezüglich der Partitionierung des Zustandsraums bei bekannter Vergangenheit beobachtet werden. Diese diskretisierte bedingte Shannon-Entropie wird mit  $H_{\mathcal{W};P}(X_n | \mathbf{X}_{n-1}^{(n)})$  bezeichnet.

Nach dem oben Gesagten lassen sich in einem chaotischen System die Trajektorien und damit deren Anfangswerte  $X_0(x_0) = x_0$  umso besser unterscheiden, je mehr Zeitbeobachtungen vorliegen. Da andererseits die Systemzustände  $X_n$  durch  $x_0$  eindeutig bestimmt sind, folgt hieraus, dass für große  $n$  die Entropie  $H_{\mathcal{W};P}(X_n | \mathbf{X}_{n-1}^{(n)})$  konvergiert:

$$h(\mathcal{W}; P) = \lim_{n \rightarrow \infty} H_{\mathcal{W};P}(X_n | \mathbf{X}_{n-1}^{(n)}). \quad (4.6)$$

Dieser Grenzwert hängt von der Partition  $\mathcal{W}$  und dem invarianten Maß  $P$  ab. Letzteres ist für die hier betrachteten dynamischen Systeme eindeutig bestimmt, so dass nur noch die Willkürlichkeit der Partitionswahl beseitigt werden muss. Hierzu wird das Supremum von  $h(\mathcal{W}; P)$  über alle Partitionen  $\mathcal{W}$  genommen,

$$h_{\text{KS}}(P) = \sup_{\mathcal{W}} h(\mathcal{W}; P). \quad (4.7)$$

Diese Größe wird *Kolmogorov-Sinai-Entropie* genannt. Sie charakterisiert die Unsicherheit, mit der die Trajektorie eines dynamischen Systems identifiziert werden kann, wenn das System zu beliebig vielen Zeitpunkten beobachtet wird und wenn eine beliebig genaue Messauflösung zur Verfügung steht.

Entsprechend Gl. (2.16) kann die diskretisierte bedingte Shannon-Entropie als Differenz von Shannon-Entropien geschrieben werden,

$$H_{\mathcal{W};P}(X_n | \mathbf{X}_{n-1}^{(n)}) = H_{\mathcal{W};P}(\mathbf{X}_n^{(n+1)}) - H_{\mathcal{W};P}(\mathbf{X}_{n-1}^{(n)}).$$

Die Shannon-Entropien lassen sich als Unsicherheiten bei der Identifikation der Trajektorien interpretieren. Da andererseits die Trajektorien durch ihre Anfangswerte  $x_0$  festgelegt sind, ist  $H_{\mathcal{W};P}(X_n | \mathbf{X}_{n-1}^{(n)})$  der Verlust von Unsicherheit über den Anfangswert, wenn das dynamische System zu einem späteren Zeitpunkt ein

zusätzliches Mal beobachtet wird. Dies ist gleichbedeutend mit der Aussage, dass die zusätzliche Beobachtung Information über den Anfangswert liefert. Anders ausgedrückt: chaotische Abbildungen produzieren im Laufe der Zeit Information über ihren Anfangswert. Für große  $n$  divergieren die einzelnen Entropien wie  $n$ :  $H_{\mathcal{W};P}(\mathbf{X}_{n-1}^{(n)}) \approx \text{Konstante} \cdot n$ . Hieraus folgt, dass

$$h_{\text{KS}}(P) = \sup_{\mathcal{W}} \lim_{n \rightarrow \infty} \frac{1}{n} H_{\mathcal{W};P}(\mathbf{X}_{n-1}^{(n)})$$

gilt.

Da Trajektorien regulärer Bewegungen langsamer als exponentiell auseinander laufen, wenn deren Anfangswerte einen beliebig kleinen Abstand voneinander haben, und somit die Anzahl der Symbolsequenzen  $\mathbf{S}_{n-1}^{(n)}$  langsamer als exponentiell zunimmt, gilt  $h_{\text{KS}}(P) = 0$ , siehe Abb. 4.1. Für stochastische Prozesse, bei denen die Zustände zeitlich unabhängig sind, faktorisieren die Verbundwahrscheinlichkeiten, so dass  $H_{\mathcal{W};P}(\mathbf{X}_{n-1}^{(n)}) \approx n \cdot H_{\mathcal{W};P}(X_0)$  ist, also  $H_{\mathcal{W};P}(X_n | \mathbf{X}_{n-1}^{(n)}) \approx H_{\mathcal{W};P}(X_0)$ . Da  $H_{\mathcal{W};P}(X_0)$  wie  $-\log \|\mathcal{W}\|$  divergiert, wobei  $\|\mathcal{W}\|$  die Norm der Partition angibt (das Volumen des größten Partitionselementes), erhält man  $h_{\text{KS}}(P) = \infty$ , siehe Gl. (3.11). Chaotische Systeme sind dadurch charakterisiert, dass  $0 < h_{\text{KS}}(P) < \infty$  gilt, [Ott (1993), Schuster (1989)].

Ruelle (1989) konnte zeigen, dass zwischen der Kolmogorov-Sinai-Entropie eines chaotischen Systems und dessen Lyapunov-Exponenten  $\lambda_i$ ,  $i = 1, \dots, d$  stets

$$h_{\text{KS}}(P) \leq \sum_{\lambda_i > 0} \lambda_i$$

erfüllt ist. Es gilt sogar die Gleichheit, also das Persin-Theorem, wenn das dynamische System dissipativ ist, einen hyperbolischen Attraktor besitzt und die periodischen Orbits dicht in ihm liegen.

Die Kolmogorov-Sinai-Entropie für diverse dynamische Systeme ist in der bereits zitierten Literatur gegeben. Verfahren zur Berechnung der Kolmogorov-Sinai-Entropie aus Zeitreihen sind zum Beispiel in [Grassberger & Procaccia (1983c), Cohen & Procaccia (1985), Kantz & Schreiber (1997)] und für räumlich ausgedehnte Systeme in [Olbrich et al. (2000)] zu finden.

## 4.2 Interdependenz, verallgemeinerte Synchronisation

Im Folgenden wird ein ergodisches dynamisches System  $(X, Y)$  betrachtet, das aus zwei Subsystemen  $X = (X_i)_{i \in \mathbb{Z}}$  und  $Y = (Y_i)_{i \in \mathbb{Z}}$  bestehen soll. Des Weiteren

wird angenommen, dass die Subsysteme nur in einer Richtung gekoppelt seien (gerichtet gekoppelt), das heißt  $Y$  sei ein autonomes System welches  $X$  antreibt:

$$\begin{aligned} x_{i+1} &= X_{i+1}(x_0, y_0) = X_1(x_i, y_i) \\ y_{i+1} &= Y_{i+1}(y_0) = Y_1(y_i). \end{aligned} \quad (4.8)$$

Wird solch ein dynamisches System beobachtet, so stellt sich die Frage, wie eine Abhängigkeit, also eine Kopplung nachgewiesen werden kann. Die Abhängigkeit zweier dynamischer Systeme wird auch *Interdependenz* genannt [Arnhold et al. (1999), Le Van Quyen et al. (1999), Schiff et al. (1996)]. Ist das angetriebene Subsystem nicht singulär, das heißt  $\det(\partial X_1/\partial y_{i,j}) \neq 0$ , wobei  $y_{i,j}$  die  $j$ -te Komponente von  $y_i$  ist, so kann  $X_1$  lokal invertiert werden und man erhält

$$y_i = \phi_K(x_i, x_{i+1}), \quad (4.9)$$

siehe [Arnhold et al. (1999)]. Dabei ist  $\phi_K$  lokal stetig.

Eine andere interessante Eigenschaft ist die Synchronisation von  $X$  und  $Y$  zum Beispiel aufgrund von Kopplung [Quian Quiroga et al. (2000), Rulkov et al. (1995)]. So wird von *verallgemeinerter Synchronisation* gesprochen, wenn es eine stetige Abbildung  $\phi_S$  gibt, so dass

$$x_i = \phi_S(y_i) \quad (4.10)$$

ist.

Üblicherweise liegen die Beobachtungen der Subsysteme  $X$  und  $Y$  als Zeitreihen  $(u_1, \dots, u_N)$  und  $(v_1, \dots, v_N)$  vor. Diese Zeitreihen werden erzeugt, indem zu den Zeitpunkten  $i$  die Systemzustände  $X_i(x_0)$  und  $Y_i(y_0)$  mit den Messfunktionen  $U$  und  $V$  auf dem Skalar  $u_i = U[X_i(x_0)]$  bzw.  $v_i = V[Y_i(y_0)]$  abgebildet werden [Kantz & Schreiber (1997)]. Nach dem Takens-Theorem [Takens (1981), Sauer et al. (1991)] kann der Zustandsraum eines ergodischen dynamischen Systems aus den Zeitreihen rekonstruiert werden, indem man zu den sogenannten Einbettungsvektoren  $\mathbf{u}_i = (u_i, \dots, u_{i-m-1})$  und  $\mathbf{v}_i = (v_i, \dots, v_{i-m-1})$  übergeht. Gilt für die Einbettungsdimension  $m$  die Relation  $m > 2D_f$ , wobei  $D_f$  die fraktale Dimension des dynamischen Systems ist, so gibt es Diffeomorphismen  $\zeta_X$  und  $\zeta_Y$  mit  $(x_i, y_i) = \zeta_X(\mathbf{u}_i) = \zeta_Y(\mathbf{v}_i)$ . Sind  $\phi_S$  und  $\phi_K$  stetig bzw. stetig und injektiv, dann folgen aus Gl. (4.9) die Beziehungen

$$\mathbf{v}_i = \psi_K(\mathbf{u}_i), \quad (4.11)$$

bzw. aus Gl. (4.10)

$$\mathbf{u}_i = \psi_S(\mathbf{v}_i), \quad (4.12)$$

mit stetigem bzw. stetigem, injektiven  $\psi_K$  und  $\psi_S$ . Für die erste Gleichung muss eventuell die Einbettungsdimension von  $\mathbf{u}_i$  erhöht werden, da hier  $x_i$  und  $x_{i+1}$  zu einem höherdimensionalen Zustandsvektor zusammengefasst werden müssen.

Demzufolge kann auf Interdependenz oder verallgemeinerte Synchronisation geschlossen werden, wenn die Existenz einer stetigen bzw. stetigen, injektiven Funktion  $\psi_K$  bzw.  $\psi_S$  nachgewiesen werden kann. Eine modifizierte Version der hierfür von Arnhold et al. (1999) vorgeschlagenen Methode soll am Beispiel der Interdependenz, Gl. (4.11), skizziert werden:

Als Erstes werden alle Nachbarn von  $\mathbf{u}_n$  gesucht, die sich in einer  $\delta$ -Umgebung  $B_\delta(\mathbf{u}_n)$  um  $\mathbf{u}_n$  befinden,  $B_\delta(\mathbf{u}_n) = \{\mathbf{u}_i : \|\mathbf{u}_i - \mathbf{u}_n\| < \delta, i = 1, \dots, N\}$ .  $\|\cdot\|$  ist eine geeignete Norm auf dem Raum der Einbettungsvektoren  $\mathbf{u}_i$ , beispielsweise die Maximumsnorm  $\|\mathbf{u}_i\|_\infty = \max_{j=0, \dots, m-1} |u_{i-j}|$ . Diese Nachbarn seien durch die Indizes  $s_{n,1}, \dots, s_{n,k_n(\delta)}$  gegeben. Dabei ist  $k_n(\delta) = |B_\delta(\mathbf{u}_n)|$  die Anzahl der Elemente in  $B_\delta(\mathbf{u}_n)$ . Ist  $\psi_K$  stetig, so sollten alle Bilder  $\mathbf{v}_{s_{n,i}} = \psi_K(\mathbf{u}_{s_{n,i}})$  von  $\mathbf{u}_{s_{n,i}}$  in einer Umgebung um  $\mathbf{v}_n$  liegen, deren mittlerer quadratischer Abstand durch

$$d_{Y|X}^{(\delta)}(\mathbf{v}_n) = \frac{1}{k_n(\delta)} \sum_{i=1}^{k_n(\delta)} |\mathbf{v}_{s_{n,i}} - \mathbf{v}_n|^2$$

gegeben ist, siehe Abb. 4.2.

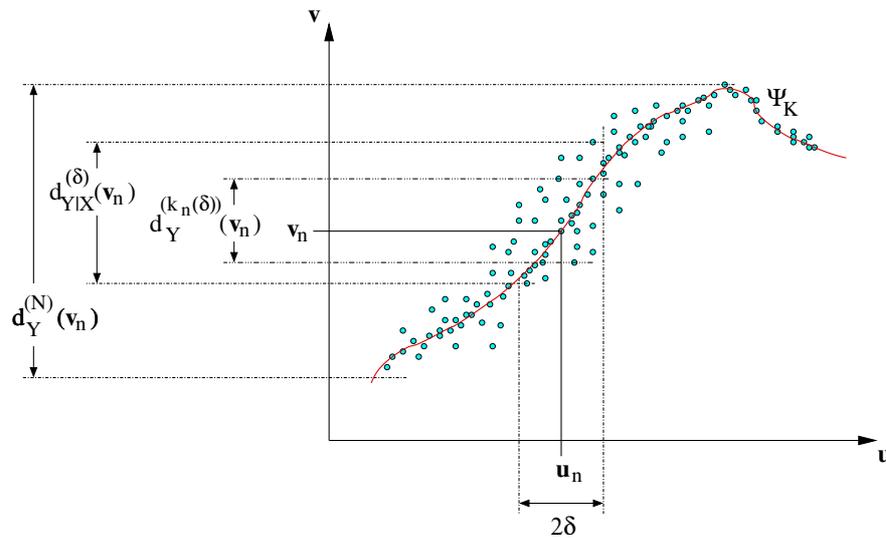


Abbildung 4.2: Schematische Darstellung zur Definition der Interdependenz sowie der hierfür benötigten Abstände.

Anschließend werden die  $k_n(\delta)$  nächsten Nachbarn von  $\mathbf{v}_n$  bestimmt. Diese seien durch die Indizes  $r_{n,1}, \dots, r_{n,k_n(\delta)}$  gegeben. Der mittlere quadratische eukli-

dische Abstand dieser Vektoren

$$d_Y^{(k_n(\delta))}(\mathbf{v}_n) = \frac{1}{k_n(\delta)} \sum_{i=1}^{k_n(\delta)} |\mathbf{v}_{r_{n,i}} - \mathbf{v}_n|^2$$

ist ebenfalls ein Maß für die Größe der Umgebung, aus der diese Punkte stammen.

Ist  $\psi_K$  stetig und zusätzlich injektiv, dann wird  $d_Y^{(k_n(\delta))}(\mathbf{v}_n)$  in etwa  $d_{Y|X}^{(\delta)}(\mathbf{v}_n)$  entsprechen,  $d_Y^{(k_n(\delta))}(\mathbf{v}_n) \approx d_{Y|X}^{(\delta)}(\mathbf{v}_n)$ . Existiert keine stetige, injektive Abbildung  $\psi_K$ , so wird  $d_{Y|X}^{(\delta)}(\mathbf{v}_n) \gg d_Y^{(k_n(\delta))}(\mathbf{v}_n)$  sein. Indem das Verhältnis dieser Abstände über alle Punkte gemittelt wird, erhält man ein Maß für die Abhängigkeit zwischen  $X$  und  $Y$ :

$$S^{(\delta)}(Y|X) = \frac{1}{N} \sum_{n=1}^N \frac{d_Y^{(k_n(\delta))}(\mathbf{v}_n)}{d_{Y|X}^{(\delta)}(\mathbf{v}_n)}. \quad (4.13)$$

Nach Konstruktion der Abstände gilt  $d_{Y|X}^{(\delta)}(\mathbf{v}_n) \geq d_Y^{(k_n(\delta))}(\mathbf{v}_n)$  und somit  $0 \leq S^{(\delta)}(Y|X) \leq 1$ . Dabei ist der Wert von  $S^{(\delta)}(Y|X)$  umso größer, je stärker die Kopplung ist, siehe hierzu [Arnhold et al. (1999), Quian Quiroga et al. (2000)].

Ist  $\psi_K$  stetig aber nicht injektiv, dann kann die Interdependenz gemessen werden, indem  $d_{Y|X}^{(\delta)}(\mathbf{v}_n)$  mit dem mittleren quadratischen Abstand  $d_Y^{(N)}(\mathbf{v}_n)$ , den  $\mathbf{v}_n$  zu allen Einbettungsvektoren hat, verglichen wird:

$$H^{(\delta)}(Y|X) = \frac{1}{N} \sum_{n=1}^N \log \frac{d_Y^{(N)}(\mathbf{v}_n)}{d_{Y|X}^{(\delta)}(\mathbf{v}_n)}. \quad (4.14)$$

Sind  $X$  und  $Y$  ungekoppelt, gemeint ist  $\psi_K$  ist nicht stetig, so wird im Allgemeinen  $d_{Y|X}^{(\delta)}(\mathbf{v}_n) \approx d_Y^{(N)}(\mathbf{v}_n)$  sein und somit  $H^{(k)}(Y|X) \approx 0$ . Andererseits ist  $H^{(\delta)}(Y|X) \gg 0$  wenn  $X$  und  $Y$  gekoppelt sind [Arnhold et al. (1999), Quian Quiroga et al. (2000)].

Bei dem ursprünglich von Arnhold et. al. (1999) vorgeschlagenen Verfahren wurde für jeden Abstand  $d_{Y|X}^{(\cdot)}(\mathbf{v}_n)$  eine fest vorgegebene Anzahl von  $k_n(\delta) \equiv k$  nächsten Nachbarn gesucht und keine  $\delta$ -Umgebung verwendet. Da für längere Zeitreihen die  $k$  nächsten Nachbarn von  $\mathbf{u}_n$  bzw.  $\mathbf{v}_n$  aus einer kleineren Umgebung stammen, sind die Abstände  $d_{Y|X}^{(\cdot)}(\mathbf{v}_n)$  bzw.  $d_Y^{(\cdot)}(\mathbf{v}_n)$  und somit die Interdependenzen  $S^{(\cdot)}(Y|X)$ ,  $H^{(\cdot)}(Y|X)$  von der Länge der Zeitreihe abhängig. Indem  $\delta$ -Umgebungen zur Definition des Abstandes  $d_{Y|X}^{(\cdot)}(\mathbf{v}_n)$  verwendet werden, wird dieser Effekt eliminiert.

Um einen Bias aufgrund serieller Korrelationen zu unterdrücken, sind bei der Berechnung der Abstände  $d_Y^{(k_n(\delta))}(\mathbf{v}_n)$ ,  $d_{Y|X}^{(\delta)}(\mathbf{v}_n)$  und  $d_Y^{(N)}(\mathbf{v}_n)$  nur jene Nachbarn zu berücksichtigen, die zeitlich weit genug auseinander liegen,  $|n - r_{n,i}| > w$  und  $|n - s_{n,i}| > w$ , wobei  $w$  das Theiler-Fenster ist [Theiler (1986)].

Die von Arnhold et al. (1999) vorgeschlagene Methode zum Nachweis von Kopplung bzw. von verallgemeinerter Synchronisation kann auch auf dynamische Systeme

$$x_{n+1} = X_{n+1}(x_0; p_0) = X_1(x_n; p_n), \quad \{p_0, p_1, p_2, \dots\} \text{ gegeben} \quad (4.15)$$

angewendet werden, deren Parameter  $p_n$  sich nur langsam in der Zeit verändern,

$$p_{n+1} \approx p_n. \quad (4.16)$$

Hier ist zu beachten, dass die Parameter  $p_0, p_1, \dots$  gegeben sind;  $(p_n)_{n \in \mathbb{N}_0}$  ist kein dynamisches System. Kann  $X_1$  lokal nach  $p_n$  aufgelöst werden,  $p_n = \zeta(x_n, x_{n+1})$ , so folgt mit Gl. (4.16)

$$x_{n+2} = X_1(x_{n+1}; p_{n+1}) \approx X_1(x_{n+1}; \zeta(x_n, x_{n+1})).$$

Werden  $x_n$  und  $x_{n+1}$  zu  $\tilde{x}_{n+1}$  zusammengefasst und wird  $\tilde{X}_1(\tilde{x}_n) = (X_1(x_n; \zeta(x_{n-1}, x_n)), \text{id}(x_n))$  gesetzt, so ergibt sich hieraus ein neues parameterfreies dynamisches System [Hegger et al. (2000)],

$$\tilde{x}_{n+1} = \tilde{X}_1(\tilde{x}_n). \quad (4.17)$$

Liegen die Beobachtungen von  $X$  als Zeitreihe vor, so kann das System (4.17) mittels Übereinbettung rekonstruiert werden. Hierzu ist jetzt eine Einbettungsdimension  $m > 2(D_f + P)$  notwendig. Dabei ist  $D_f$  die fraktale Dimension von  $X$  und  $P$  die Anzahl der Parameter [Hegger et al. (2000)].

Dies Verfahren soll an einem zeitdiskreten System von gekoppelten chaotischen Oszillatoren mit gerichteter Kopplung demonstriert werden,

$$\begin{aligned} x_{n+1} &= X_1(x_n) + c \cdot (Y_1(y_n) - X_1(x_n)) \\ y_{n+1} &= Y_1(y_n). \end{aligned} \quad (4.18)$$

Der Kopplungsparameter  $c$  wird mit  $c = 0.1$  zunächst fest gewählt. Die einzelnen Subsysteme werden mit der logistische Abbildung modelliert, siehe beispielsweise [Ott (1993)],

$$\begin{aligned} X_1(x) &= a \cdot x(1-x) \\ Y_1(y) &= b \cdot y(1-y), \end{aligned}$$

wobei  $a = 3.81$  und  $b = 3.80$  gesetzt werden, so dass beide Subsysteme sich im chaotischen Regime befinden.

Als Erstes wird das Verhalten der Interdependenz  $H^{(\delta)}(Y|X)$  untersucht, wenn der Kopplungsparameter einen zeitlichen Drift aufweist:

$$c \longrightarrow c_n = c + \gamma \sin(2\pi\nu n). \quad (4.19)$$

Hierbei wurde die Driftfrequenz mit  $\nu = 10^{-4}$  fest gewählt. Das System in (4.18) wurde numerisch iteriert, um eine Zeitreihe mit 10 000 Punkten zu erhalten. Somit oszilliert  $c_n$  genau einmal in dem verwendeten Datensatz.

Für verschiedene  $\gamma$ ,  $0 \leq \gamma \leq 0.1$ , wurden die Interdependenzen  $H^{(\delta)}(Y|X)$  und  $H^{(\delta)}(X|Y)$  berechnet. Die Ergebnisse für die Interdependenzen sind in Abb. 4.3 aufgetragen. Als Parameter wurde  $m = 4$  (Einbettungsdimension von  $X$  und  $Y$ ),  $w = 10$  (Theiler-Fenster) und  $\delta = 1/20$  des Wertebereich von  $X$  bzw.  $Y$  (Umgebungsgröße) verwendet. Die Interdependenz nimmt mit zunehmender Driftam-

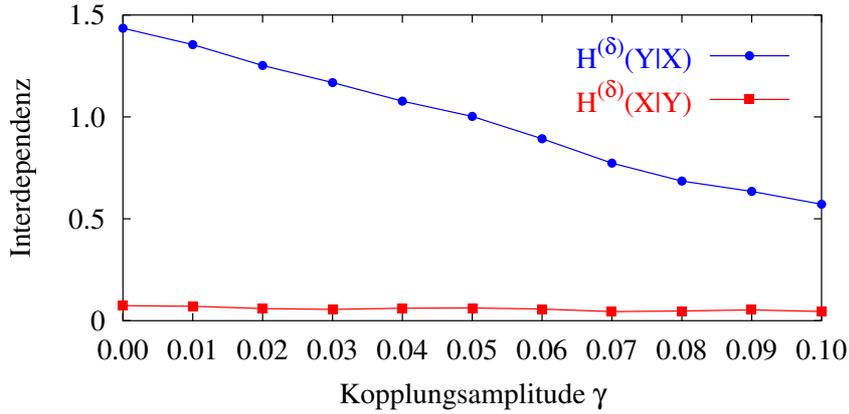


Abbildung 4.3: Interdependenz  $H^{(\delta)}(Y|X)$  (Linie mit Kreisen) und  $H^{(\delta)}(X|Y)$  (Linie mit Quadraten) für das in Gl. (4.18) gegebene System in Abhängigkeit der Kopplungsamplitude  $\gamma$  (Gl. (4.19)).  $\delta$  wurde gleich  $1/20$  des Wertebereichs von  $X$  bzw.  $Y$  gesetzt.

plitude  $\gamma$  ab, liefert aber dennoch in Kopplungsrichtung größere Werte als entgegen dieser Richtung, wo sie nahezu konstant ist. Somit gibt die Interdependenz  $H^{(\delta)}(\cdot|\cdot)$  die Struktur des dynamischen Systems bei driftender Kopplungsstärke richtig wieder. Die Abnahme der Interdependenz  $H^{(\delta)}(Y|X)$  mit der Amplitude  $\gamma$  tritt auf, da mit wachsendem  $\gamma$  der zeitliche Drift von  $c_n$  zunimmt, insbesondere wird die Näherung  $c_{n+1} \approx c_n$  immer schlechter. Für längere Zeitreihen ändern sich die Interdependenzen lediglich innerhalb geringer statistischer Fluktuationen.

Desweiteren wurde die Interdependenz  $H^{(\delta)}(\cdot|\cdot)$  bei festgehaltenem  $\gamma$  und variierender Driftfrequenz  $\nu$  untersucht. Auch hier spiegelte die Interdependenz die zugrunde liegende Modellstruktur richtig wieder.

Als Nächstes wurde das Verhalten der Interdependenz bei fester Kopplung und driftenden Parametern des dynamischen Systems untersucht. Die Ergebnisse, bei denen der Parameter  $b$  des antreibenden Systems durch

$$b \longrightarrow b_n = b + \beta \sin(2\pi\nu n) \quad (4.20)$$

mit  $\nu = 10^{-4}$  ersetzt wurde, sind in Abb. 4.4 zu sehen. Die Amplituden  $\beta$  des

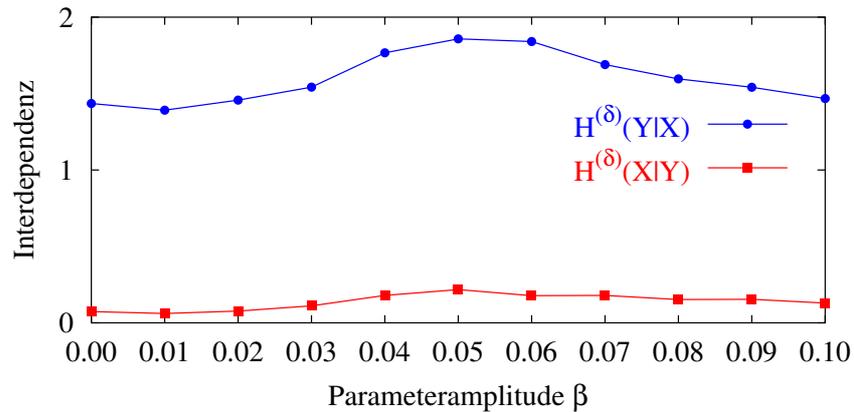


Abbildung 4.4: Interdependenz  $H^{(\delta)}(Y|X)$  (Linie mit Kreisen) und  $H^{(\delta)}(X|Y)$  (Linie mit Quadraten) für das in Gl. (4.18) gegebene System in Abhängigkeit der Parameteramplitude  $\beta$  (Gl. (4.20)).  $\delta$  wurde gleich  $1/20$  des Wertebereichs von  $X$  bzw.  $Y$  gesetzt.

Driftparameters wurden zwischen 0 und 0.1 variiert. Somit bleibt das treibende System bei kleinen  $\beta$  im chaotischen Regime, bei mittleren  $\beta$  erreicht es das periodische Regime, welches es bei großen  $\beta$  durchfährt.

Im gesamten Parameterbereich spiegelte die Interdependenz zusammen mit der Übereinbettung die Struktur des Systems richtig wieder. Insbesondere ist in diesem Bereich die kontinuierliche Abnahme der  $H^{(\delta)}(Y|X)$ -Werte bei wachsendem Drift, wie es bei der Kopplung der Fall war, nicht zu beobachten.

# Kapitel 5

## Punktprozesse

### 5.1 Definition eines Punktprozesses

Bei den bisher untersuchten Prozessen spielte der Zustand, den ein System zu einer bestimmten Zeit annimmt, die entscheidende Rolle. Denkt man aber an den  $\alpha$ -Zerfall, Elektrokardiogramme oder das Eintreffen eines Telefonats in der Vermittlungszentrale, so ist nicht der Zustand des Prozesses, sondern vielmehr der Zeitpunkt, an dem solch ein Ereignis eintritt, von Interesse. Die Beschreibung von stochastischen Systemen durch ihre Ereigniszeiten erfolgt mit einer speziellen Prozessklasse, den sogenannten stochastischen Punktprozessen, siehe [Bauer (1991), Billingsley (1995)] und [Lewis (1972)]. Diese Prozesse sollen im Folgenden genauer betrachtet werden, insbesondere soll die gegenseitige Abhängigkeit zweier Punktprozesse anhand ihrer Ereigniszeiten genauer untersucht werden.

Wird zum Beispiel 1mol Thorium betrachtet, so findet zu aufeinander folgenden Zeitpunkten  $0 = t_0 < t_1 < t_2 < \dots$  ein  $\alpha$ -Zerfall statt, wobei angenommen wird, dass die Wahrscheinlichkeit, mit der solche Ereignisse gleichzeitig vorkommen, Null ist. Außerdem sollen in einem endlichen Zeitintervall auch nur endlich viele Zerfallsprozesse auftreten. Würde dieses Experiment wiederholt, so erhielte man erneut eine Sequenz von Ereignissen, im Allgemeinen aber mit anderen Ereigniszeiten  $t_k, k \in \mathbb{N}_0$  als zuvor. Dies ist in der stochastischen Natur dieses Prozesses begründet.

Dementsprechend sind die Zeitpunkte  $(t_k)_{k \in \mathbb{N}_0}$  von Ereignissen eines Prozesses, wie den oben beschriebenen, durch eine Familie von nichtnegativ reellwertigen Zufallsvariablen  $(T_k)_{k \in \mathbb{N}_0}$  gegeben. Für diese gilt

$$0 = T_0(\omega) < T_1(\omega) < T_2(\omega) < \dots, \quad \sup_k T_k(\omega) = \infty \quad (5.1)$$

für jedes  $\omega \in \Omega$  (siehe Abb. 5.1) [Bauer (1991), Billingsley (1995)]. Dabei ist  $(\Omega, \mathcal{A}, P)$  der zugrunde liegende Wahrscheinlichkeitsraum, der nicht weiter spezifiziert werden muss. Die Zeitpunkte  $T_k(\omega), k \in \mathbb{N}_0$  heißen *Ereigniszeiten* oder

auch *Ankunftszeiten*.

Die Zeitintervalle zwischen benachbarten Ereignissen, die durch den Prozess  $D_k = T_k - T_{k-1}$  gegeben sind, werden *Wartezeiten* genannt, siehe Abb. 5.1. Mit  $D_0 = 0$  folgt unmittelbar aus Gl. (5.1)

$$D_1(\omega) > 0, D_2(\omega) > 0, \dots, \quad \sum_{k \geq 0} D_k(\omega) = \infty \quad (5.2)$$

für alle  $\omega \in \Omega$ .

Prinzipiell lassen sich solche stochastischen Systeme mit den Zufallsvariablen für Ereigniszeiten  $T_k$  oder den Wartezeiten  $D_k$ ,  $k \in \mathbb{N}_0$  beschreiben. Im Allgemeinen können die Ereigniszeiten aber nicht nach dem Index  $k$  geordnet werden. Eher ist zu erwarten, dass zum Beispiel beim ersten Experiment das sechste Ereignis früher eintritt als das fünfte bei der Wiederholung, das heißt  $T_6(\omega_1) < T_5(\omega_2)$ , während gleichzeitig  $T_4(\omega_1) > T_3(\omega_2)$  gilt. Diese Situation ist in Abb. 5.1 illustriert.

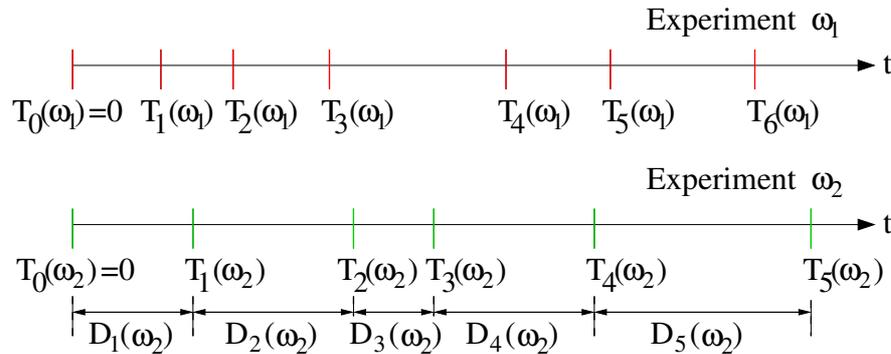


Abbildung 5.1: Ereigniszeiten von zwei Experimenten.

Um diese Schwierigkeit zu umgehen, wird eine Größe benötigt, die auf der reellen Zeitachse definiert ist und gleichzeitig zu jedem Zeitpunkt  $t \geq 0$  den Zustand des Prozesses beschreibt. Diese Größe ist die *Ereigniszahl*  $N_t(\omega)$ ; sie gibt die Anzahl der Ereignisse an, die im Zeitintervall  $[0, t] = \{t' \in \mathbb{R}_+ : 0 \leq t' \leq t\}$  auftreten. Die Ereigniszahl ist durch die größte ganze Zahl  $k$ , für die  $T_k(\omega) \leq t$  gilt, gegeben,

$$N_t(\omega) = \max\{k \in \mathbb{N}_0 : T_k(\omega) \leq t\}, \quad (5.3)$$

siehe Abb. 5.2. Entsprechend der Konstruktion von  $N_t$  ist insbesondere  $N_0 = 0$  fast sicher. Da es sich bei  $\{N_t : t \geq 0\}$  um eine Familie von Zufallsvariablen mit Werten in  $\mathbb{N}_0$  handelt, ist durch  $N = (N_t)_{t \in \mathbb{R}_+}$  ein stochastischer Prozess gegeben, der *Punktprozess* oder auch *Zählprozess* genannt wird [Bauer (1991), Billingsley (1995)]. Insbesondere ist der *Pfad* bzw. die *Realisierung eines Punktprozesses*, das heißt die Abbildung  $t \rightarrow N_t(\omega)$ , eine rechtsseitig

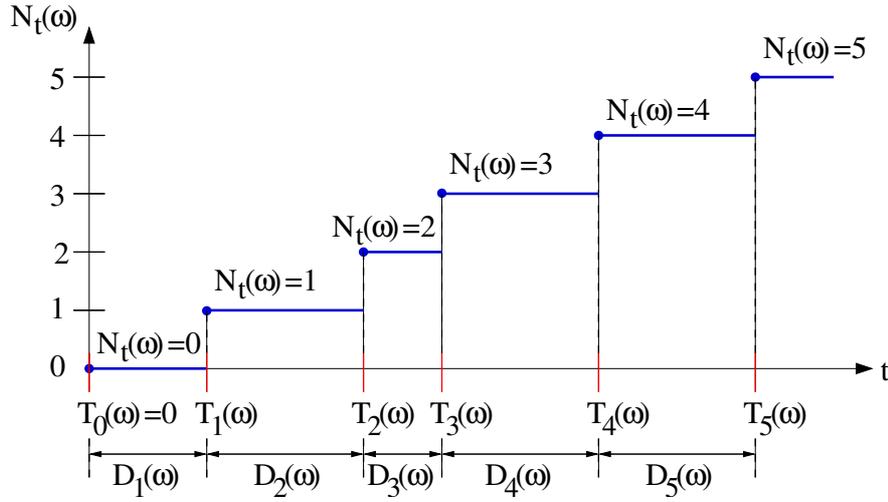


Abbildung 5.2: Ereigniszahl  $N_t(\omega)$  in Abhängigkeit der Zeit  $t$ . Die Ereigniszeiten sind mit  $T_k(\omega)$  gekennzeichnet und die Wartezeiten mit  $D_k(\omega)$ ,  $k = 0, 1, 2, \dots$

stetige Abbildung mit ganzzahligen, positiven Werten. Desweiteren macht  $N_t$  in ihren Unstetigkeitsstellen einen Sprung um genau 1 und bleibt sonst konstant:  $N_t(\omega) - \sup_{s < t} N_s(\omega) \in \{0, 1\}$ .

Ist andererseits  $N = (N_t)_{t \in \mathbb{R}_+}$  ein Punktprozess, so sind seine Ereigniszeiten durch

$$T_k(\omega) = \inf\{t \in \mathbb{R}_+ : N_t(\omega) \geq k\} \quad (5.4)$$

bestimmt. Die Ereigniszeiten  $T_k(\omega)$  sind somit die Unstetigkeitsstellen der Pfade von  $N$ , also von  $N_t(\omega)$  als Funktion von  $t$ .

Offensichtlich ist jeder Punktprozess entweder durch seine Ereigniszahlen  $N_t, t \geq 0$ , seine Ereigniszeiten  $T_k, k \in \mathbb{N}_0$  oder seine Wartezeiten  $D_k, k \in \mathbb{N}_0$  eindeutig festgelegt.

Aufgrund endlicher Beobachtungszeiträume, endlicher Zeitauflösung von Messapparaturen usw. ist man häufig nicht so sehr an der Anzahl der Ereignisse bis zu einem bestimmten Zeitpunkt  $t$  interessiert, als vielmehr an der Anzahl von Ereignissen, die in einem bestimmten Zeitfenster stattfanden. Für das Zeitintervall  $(s, t] = \{t' \in \mathbb{R}_+ : s < t' \leq t\}$  ist diese Anzahl einer Beobachtung durch den sogenannten *Zuwachs eines Punktprozesses*  $N_t(\omega) - N_s(\omega)$  gegeben.

Es wird gesagt, ein Punktprozess habe *stationär verteilte Zuwächse*, wenn die Verteilung  $P^{N_t - N_s}$  nur von der Zeitdifferenz  $t - s$  abhängt:

$$\begin{aligned} P\{N_t - N_s = n\} &= P\{N_{t-s} - N_0 = n\} \\ &= P\{N_{t-s} = n\} \equiv \pi_{t-s} \quad \forall n \in \mathbb{N}_0. \end{aligned} \quad (5.5)$$

Die dritte Gleichung folgt dabei aus  $N_0 = 0$  fast sicher.

Die Menge aller Intervalle  $(s, t], s, t \in \mathbb{R}_+$  stellt einen Erzeuger der Borelschen  $\sigma$ -Algebra  $\mathcal{B}(\mathbb{R}_+)$  auf  $\mathbb{R}_+$  dar [Bauer (1992)]. Weiterhin ist  $N_t(\omega)$  eine monotone, rechtsseitig stetige Abbildung von  $\mathbb{R}_+ \rightarrow \mathbb{N}_0$ . Somit kann jedem  $\omega \in \Omega$  eindeutig ein ganzzahliges,  $\sigma$ -endliches Maß  $\eta(\omega, \cdot)$  auf  $\mathcal{B}(\mathbb{R}_+)$  zugeordnet werden, so dass  $N_t(\omega) - N_s(\omega) = \eta(\omega, (s, t])$  gilt. Dabei gibt  $\eta(\omega, B)$  die Anzahl der Ereignisse an, die in dem Zeitfenster  $B \in \mathcal{B}(\mathbb{R}_+)$  auftreten. Daley & Vere-Jones (1972) bezeichnen die messbare Abbildung von einem Wahrscheinlichkeitsraum in den Raum aller ganzzahligen,  $\sigma$ -endlichen Maße, der mit einer geeigneten  $\sigma$ -Algebra versehen wurde, als *stochastischen Punktprozess*.

Oft werden Punktprozesse in Form von Zeitreihen dargestellt (zum Beispiel Schreiber & Schmitz (2000)), die nur die Werte 0 und 1 annehmen (siehe Abb. 5.3):

$$X(t, \omega) = \sum_{k=0}^{\infty} \delta(t, T_k(\omega)). \quad (5.6)$$

Offensichtlich kann aus jeder Beobachtung eines Punktprozesses, das heißt für

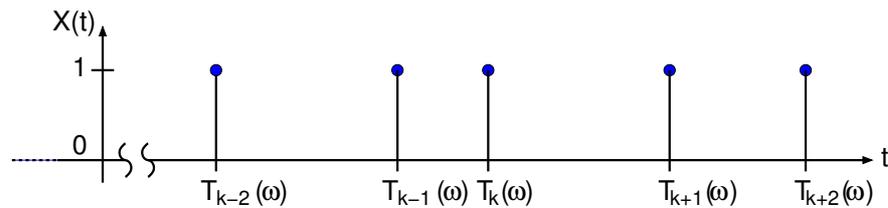


Abbildung 5.3: Realisierung eines Punktprozesses. Die Ereigniszeiten sind mit  $T_k$  bezeichnet.

jedes  $\omega$ , solch eine Darstellung gewonnen werden. Daher wird die Funktion  $X(t, \omega)$  im Folgenden *Zeitreihe eines Punktprozesses* genannt. Andererseits kann aus einer Zeitreihe, die durch Gl. (5.6) gegeben ist, die Ereigniszahl  $N_t(\omega)$  rekonstruiert werden:

$$N_t(\omega) = |\{t' \geq 0 : X(t') = 1\}|. \quad (5.7)$$

Dabei ist  $|\cdot|$  die Kardinalzahl einer Menge, das heißt die Anzahl ihrer Elemente.

**Beispiel: Poisson-Prozess** Ein in der Literatur ([Bauer (1991), Billingsley (1995)]) ausführlich diskutierter Punktprozess ist der sogenannte *Poisson-Prozess*. Seine Entwicklung geht auf den Dänen Erlang (1878-1929) zurück, der den Telefonverkehr untersuchte.

Ein Punktprozess  $N = (N_t)_{t \geq 0}$  heißt Poisson-Prozess, wenn

1. seine Zuwächse  $N_{t_1} - N_0 = N_{t_1}, N_{t_2} - N_{t_1}, \dots, N_{t_k} - N_{t_{k-1}}$  für alle  $0 < t_1 < \dots < t_k$  unabhängig verteilt sind,

2. die Zuwächse  $N_t - N_s$  mit  $0 \leq s < t$  Poisson-verteilt sind. Das heißt die Verteilung von  $N_t - N_s$  ist gegeben durch

$$P\{N_t - N_s = n\} = e^{-\alpha(t-s)} \frac{(\alpha(t-s))^n}{n!}, \quad n \in \mathbb{N}_0. \quad (5.8)$$

Da die hier eingeführten Punktprozesse die Normalität  $N_0 = 0$  fast sicher erfüllen, wird dieser Poisson-Prozess auch *normal* genannt [Bauer (1991)].

Insbesondere ist jeder Punktprozess mit unabhängig und exponentiell verteilten Wartezeiten  $D_k$  ein Poisson-Prozess. Ebenso handelt es sich bei einem Punktprozess  $N$  um einen Poisson-Prozess, wenn seine Zuwächse  $N_{t_1}, N_{t_2} - N_{t_1}, \dots, N_{t_k} - N_{t_{k-1}}$  für beliebige  $0 < t_1 < t_2 < \dots < t_k$  unabhängig und stationär verteilt sind. Letzteres besagt, dass die Verteilung von  $N_t - N_s$  für  $0 \leq s < t$  nur von der Zeitdifferenz  $t - s$  abhängt, das heißt  $P^{N_t - N_s} = \pi_{t-s}$ , siehe Gl. (5.5) sowie [Bauer (1991), Billingsley (1995)]. Dementsprechend stellt die Verteilung der Zuwächse eines Poisson-Prozesses eine Faltungshalbgruppe dar:  $\pi_{t-s} = P^{N_t - N_s} = P^{N_t - N_u + N_u - N_s} = P^{N_t - N_u} * P^{N_u - N_s} = \pi_{t-u} * \pi_{u-s}$  für  $0 \leq s < u < t$ , wobei “\*” die Faltung zweier Maße bezeichnet, siehe hierzu [Bauer (1992), Bauer (1991)].

## 5.2 Momente und Ereignisraten

Im vorherigen Abschnitt wurde dargelegt, dass die Anzahl der Ereignisse einer Beobachtung im Zeitintervall  $(s, t]$ ,  $0 \leq s < t$  durch den Zuwachs  $N_t(\omega) - N_s(\omega)$  gegeben ist. Um eine generelle Aussage über ein stochastisches System zu machen, ist nicht so sehr der Wert einer Beobachtung von Interesse, als viel mehr der Erwartungswert gemittelt über alle möglichen Realisierungen. Daher ist der Zuwachs  $N_t(\omega) - N_s(\omega)$  über alle Beobachtungen  $\omega \in \Omega$  zu mitteln. Dieser Erwartungswert von Ereignissen im Zeitintervall  $(s, t]$  lautet

$$M((s, t]) = E[N_t - N_s] = \int_{\Omega} N_t(\omega) - N_s(\omega) P(d\omega). \quad (5.9)$$

Im Folgenden wird die mittlere Anzahl von Ereignissen in einem Zeitintervall als *1. Moment des Punktprozesses* bezeichnet.

Zwar ist bei dem 1. Moment die Zufallskomponente herausgemittelt, dennoch stellt sie nicht die gewünschte physikalische Größe dar, denn das Moment bezieht sich immer auf ein Intervall  $(s, t]$ . Dementsprechend ist das Moment nicht invariant gegenüber Zeitskalierungen. Für physikalische Interpretationen ist die mittlere Anzahl von Ereignissen pro Zeiteinheit erwünscht, also die *Ereignisrate*. Diese ist gegeben durch

$$m(t) = \lim_{\Delta \rightarrow 0+} \frac{M((t - \Delta, t])}{\Delta}, \quad (5.10)$$

wobei vorausgesetzt wird, dass dieser Grenzwert existiert. Dies ist für die meisten physikalischen Systeme der Fall. Im Zusammenhang mit der Untersuchung von Neuronen wird die Ereignisrate auch Spikedichtefunktion genannt.

Offensichtlich kann bei gegebener Ereignisrate  $m(t)$  die mittlere Anzahl der Ereignisse im Intervall  $(s, t]$  mittels Integration berechnet werden,

$$M((s, t]) = \int_s^t m(t') dt'.$$

Ist  $N$  ein Punktprozess mit stationär verteilten Zuwächsen, so folgt unmittelbar

$$M((s, t]) = M((0, t - s]), \quad \forall 0 \leq s < t, \quad (5.11)$$

das 1. Moment hängt demnach nur von der Zeitdifferenz  $t - s$ . Bildet man für die 1. Momente  $M((t - \tau, t]) = M((s - \tau, s])$  den Grenzübergang  $\tau \rightarrow 0_+$ , so folgt aus Gl. (5.10) unmittelbar  $m(t) = m(s)$ . Somit ist die Ereignisrate eines Punktprozesses mit stationär verteilten Zuwächsen eine Konstante,  $m(t) = m_0$  für alle  $t \geq 0$ .

Zeitliche Korrelationen innerhalb eines Punktprozesses, also die Häufigkeiten, mit denen ein Ereignis im Zeitintervall  $(s_1, t_1]$ , ein weiteres im Intervall  $(s_2, t_2]$  usw. vorkommt, können mit *Momenten höherer Ordnung* beschrieben werden. So gibt

$$M((s_1, t_1] \times \dots \times (s_l, t_l]) = E[(N_{t_1} - N_{s_1}) \cdot \dots \cdot (N_{t_l} - N_{s_l})] \quad (5.12)$$

die mittlere Anzahl der Ereignistupel  $(T_{k_1}(\omega), \dots, T_{k_l}(\omega))$  an, bei denen  $T_{k_1}(\omega)$  in  $(s_1, t_1]$  fällt,  $T_{k_2}(\omega)$  in  $(s_2, t_2]$ , ... und  $T_{k_l}(\omega)$  in  $(s_l, t_l]$  [Daley & Vere-Jones (1972)].

Die Rate  $m(t_1, \dots, t_l)$ , mit der diese Tupel pro Zeiteinheiten auftreten, ist durch die Dichte dieser Momente gegeben, das heißt

$$M((s_1, t_1] \times \dots \times (s_l, t_l]) = \int_{s_1}^{t_1} \dots \int_{s_l}^{t_l} m(t'_1, \dots, t'_l) dt'_1 \dots dt'_l.$$

Auch hier gilt, dass für Punktprozesse mit stationär verteilten Zuwächsen die Momente höherer Ordnung nur von den Intervallbreiten abhängen und dass deren Dichten Konstante sind.

**Beispiel: radioaktiver Zerfall.** Der radioaktive Zerfall einen nicht-stationären Punktprozess dar. Die mittlere Anzahl der noch nicht zerfallenen, instabilen Isotope ist durch  $N_{\text{isotop}}(t) = N_{\text{isotop}}(0) e^{-\lambda t}$  gegeben, wobei  $\lambda$  die Zerfallskonstante ist. Die Ereignisrate (Zerfallsrate) lautet  $m(t) = \lambda N_{\text{isotop}}(t) = N_{\text{isotop}}(0) \lambda e^{-\lambda t}$ . Hieraus folgt für die mittlere Anzahl von Zerfallsprozessen  $E[N_t] = M((0, t]) = N_{\text{isotop}}(0) (1 - e^{-\lambda t})$ .

**Beispiel: Poisson-Prozess.** Durch einfaches Integrieren mit der in Gl. (5.8) gegebenen Verteilung folgt für das 1. Moment eines Poissonprozesses

$$M((s, t]) = \mathbb{E}[N_t - N_s] = \alpha(t - s)$$

und somit, dass die Ereignisrate durch den Parameter  $\alpha$  gegeben ist,  $m(0) = \alpha$ . Desweiteren ist  $M((s, t] \times (s, t]) = \mathbb{E}[(N_t - N_s)^2] = \alpha^2(t - s)^2 + \alpha(t - s)$ . Somit folgt für die Varianz der Zuwächse eines Poisson-Prozesses:

$$\text{Var}[N_t - N_s] = \mathbb{E}[(N_t - N_s - \mathbb{E}[N_t - N_s])^2] = \alpha(t - s).$$

Leider erweist sich das Schätzen der Ereignisrate aus Beobachtungen häufig als sehr schwierig, insbesondere wenn nur eine Realisierung (Messreihe) vorliegt. Denn der Grenzübergang in Gl. (5.10) würde zu einer Funktion führen, die nur die Werte 0 und  $\infty$  annimmt. Folglich ist man auf ein endliches Zeitfenster  $\Delta > 0$  angewiesen, um  $m(t)$  schätzen zu können, zum Beispiel mit  $m(t) \approx (N_t - N_{t-\Delta})/\Delta$ . Andererseits kann  $m(t)$  auch zeitabhängig sein, wie beim radioaktiven Zerfall. Dementsprechend darf  $\Delta$  auch nicht zu groß angenommen werden, um noch eine gute Approximation für  $m(t)$  zu erhalten. Aber ohne weiteres Wissen über den vorliegenden Prozess ist es häufig nicht möglich, das Zeitfenster entsprechend zu wählen. Somit bleibt in den meisten Fällen die Untersuchung von Punktprozessen auf eine festgelegte Zeitskala und damit auf die Momente beschränkt, es sei denn, die Zuwächse des Punktprozesses sind stationär verteilt.

## 5.3 Gekoppelte Punktprozesse

Werden zwei Punktprozesse gleichzeitig beobachtet, so stellt sich häufig die Frage, ob die beiden Prozesse voneinander abhängen. In diesem und in den folgenden Abschnitten soll auf diese Fragestellung näher eingegangen werden. Insbesondere werden verschiedene Methoden vorgestellt, mit denen eine Abhängigkeit und eine Kopplung zwischen Punktprozessen nachgewiesen werden kann.

Hierzu seien  $X = (X_t)_{t \geq 0}$  und  $Y = (Y_t)_{t \geq 0}$  zwei Punktprozesse, deren Ereigniszeiten durch die Zufallsvariablen  $T_k$  bzw.  $S_k$ ,  $k \in \mathbb{N}_0$  gegeben sind. Koppelt  $Y$  in  $X$ , so wird im Allgemeinen die Wahrscheinlichkeit, dass das nächste Ereignis von  $X$  zum Zeitpunkt  $T_{k+1} = t_{k+1}$  eintritt, eine Funktion aller Ereignisse von  $X$  und  $Y$  sein, die zuvor aufgetreten sind:

$$P\{T_{k+1} = t_{k+1}\} = \mathcal{F}[t_{k+1}, T_k = t_k, \dots, S_{j_{k+1}-1} = s_k, \dots], \quad (5.13)$$

mit  $j_{k+1}(\omega) = \min\{l \in \mathbb{N}_0 : S_l(\omega) \geq T_{k+1}(\omega)\}$ , siehe Abb. 5.4. Der Index  $j_{k+1}(\omega)$  gibt hierbei das nächste Ereignis von  $Y$  an, das dem  $(k+1)$ -ten Ereignis von  $X$  zur Zeit  $T_{k+1}(\omega)$  folgt. Dementsprechend ist  $S_{j_{k+1}(\omega)-1}(\omega)$  der Zeitpunkt des

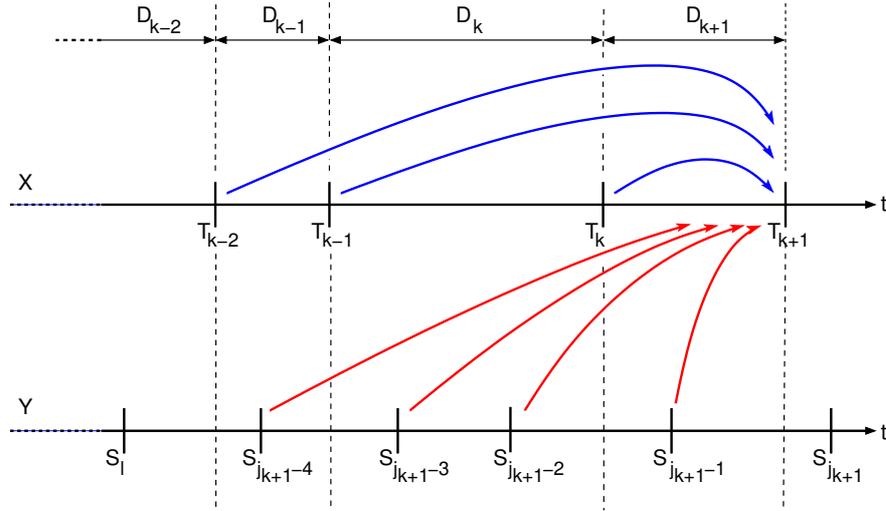


Abbildung 5.4: Kopplung zwischen zwei Punktprozessen. Die Pfeile symbolisieren den Einfluss der vorherigen Ereignisse auf das nächste Ereignis, das zur Zeit  $T_{k+1}(\omega)$  auftritt.

letzten Ereignisses von  $Y$ , das die Ankunftszeit  $T_{k+1}(\omega)$  aufgrund von Kausalität noch beeinflussen kann.

Entsprechend dieser Modellvorstellung kann das System bezüglich seiner Dynamik dahingehend untersucht werden, ob die Punktprozesse  $X$  und  $Y$  stochastisch ungekoppelt sind. Dies ist genau dann der Fall, wenn

$$P^{T_{k+1}|T_k=t_k, T_{k-1}=t_{k-1}, \dots, S_{j_{k+1}-1}=s_k, S_{j_{k+1}-2}=s_{k-1}, \dots} = P^{T_{k+1}|T_k=t_k, T_{k-1}=t_{k-1}, \dots} \quad (5.14)$$

gilt.

Eine weitere Möglichkeit besteht darin, die Verteilung von  $(X, Y)$  auf Unabhängigkeit zu untersuchen. So ist entsprechend Gl. (5.13) zu erwarten, dass  $X$  und  $Y$  nur dann stochastisch unabhängig sind, wenn wenigstens

$$P^{T_{k+1}, T_k, T_{k-1}, \dots, S_{j_{k+1}-1}, S_{j_{k+1}-2}, \dots} = P^{T_{k+1}, T_k, T_{k-1}, \dots} \otimes P^{S_{j_{k+1}-1}, S_{j_{k+1}-2}, \dots} \quad (5.15)$$

erfüllt ist.

Prinzipiell können Gl. (5.14) und Gl. (5.15) benutzt werden, um Methoden zum Nachweis von Kopplungen und Abhängigkeiten zu entwickeln. Die Tatsache, dass die Ereigniszeit  $S_{j_{k+1}-1}$  von  $T_{k+1}$  abhängt, erschwert dies aber erheblich. Des Weiteren handelt es sich bei den Ereigniszeiten um Zufallsvariablen mit kontinuierlichem Wertebereich.

Diese Schwierigkeit wird umgangen, indem nicht die Ereigniszeiten, sondern die Ereigniszahlen  $X_t$  und  $Y_t$  der beiden Punktprozesse betrachtet werden. So ist auch hier ein entsprechender funktioneller Zusammenhang wie in Gl. (5.13)

für die Verteilung von  $X_{t_{i+1}}$  zu erwarten, wobei für  $i, j \in \mathbb{N}$  die Zeitpunkte  $0 \leq \dots, t_{i-1} < t_i < t_{i+1}$  und  $0 \leq \dots, s_{j-1} < s_j < t_{i+1}$  beliebig sind:

$$\begin{aligned} P\{X_{t_{i+1}} = x_{i+1}\} \\ = \mathcal{F}[x_{i+1}, X_{t_i} = x_i, \dots, X_{t_1} = x_1, Y_{s_j} = y_j, \dots, Y_{s_1} = y_1]. \end{aligned} \quad (5.16)$$

Dementsprechend wird  $X$  in seiner Dynamik nicht von  $Y$  beeinflusst, wenn

$$\begin{aligned} P\{X_{t_{i+1}} = x_{i+1} | X_{t_i} = x_i, \dots, X_{t_1} = x_1, Y_{s_j} = y_j, \dots, Y_{s_1} = y_1\} \\ = P\{X_{t_{i+1}} = x_{i+1} | X_{t_i} = x_i, \dots, X_{t_1} = x_1\} \end{aligned} \quad (5.17)$$

gilt, das heißt wenn  $X$  stochastisch ungekoppelt von  $Y$  ist, siehe Absch. 2.1.2, Gl. (2.20). Des Weiteren muss nach der Modellvorstellung in Gl. (5.16) wenigstens

$$\begin{aligned} P\{X_{t_{i+1}} = x_{i+1}, \dots, X_{t_1} = x_1, Y_{s_j} = y_j, \dots, Y_{s_1} = y_1\} \\ = P\{X_{t_{i+1}}, \dots, X_{t_1} = x_1\} \otimes P\{Y_{s_j} = y_j, \dots, Y_{s_1} = y_1\} \end{aligned} \quad (5.18)$$

erfüllt sein, wenn  $X$  und  $Y$  stochastisch unabhängig sein sollen, wobei hier  $s_j \geq t_{i+1}$  zugelassen ist, siehe Absch. 2.1.1, Gl. (2.6).

Die Kopplung zwischen zwei Punktprozesse bzw. die Abhängigkeit beider Prozesse voneinander kann prinzipiell mit Gl. (5.17) und Gl. (5.18) untersucht werden. In der Praxis erweist sich diese Herangehensweise allerdings als sehr schwierig. Dies hängt vor allem damit zusammen, dass die Pfade von Punktprozessen monoton steigende Funktionen darstellen und folglich eine ausreichende Zahl von Realisierungen vorhanden sein muss, um die Übergangsverteilungen in Gl. (5.17) und die Verteilungen in Gl. (5.18) schätzen zu können. Häufig wird aber ein Experiment nur einmal beobachtet. Aus diesem Grund werden die Punktprozesse nur auf Kopplung und Abhängigkeit in Bezug auf ihre Zuwächse untersucht. Wie in den folgenden Abschnitten gezeigt wird, impliziert die Gültigkeit von Gl. (5.18), dass auch die Zuwächse der jeweiligen Prozesse unabhängig sind. Dagegen muss für die Kopplung eine andere Modellvorstellung zugrunde gelegt werden, um Kopplung auf Basis der Zuwächse nachweisen zu können.

Bevor auf die verschiedenen Methoden zum Nachweis von Abhängigkeiten eingegangen wird, sei darauf hingewiesen, dass es sich um Abhängigkeiten im stochastischen Sinn handelt. So ist der mathematisch definierte Begriff *stochastisch unabhängig* in seiner Bedeutung nicht mit dem Begriff *physikalisch unabhängig* äquivalent, der für *ungekoppelt/stochastisch ungekoppelt* steht. Denn *stochastisch unabhängig* besagt nur, dass die Verteilungen der Systemvariablen nicht voneinander abhängen, während *ungekoppelte* Systeme physikalisch voneinander getrennt sind bzw. sich nicht mittels Kraftfeldern gegenseitig beeinflussen.

Sind  $X$  und  $Y$  ungekoppelte Systeme, die miteinander synchronisieren, das heißt  $X_t = \psi(Y_t)$ , wobei  $\psi$  eine stetige Funktion ist, so ist  $X$  nicht stochastisch unabhängig von  $Y$ . Ein einfaches Beispiel stellen zwei getrennte Frequenzgeneratoren dar, die ein periodisches Signal geben, deren eine Frequenz sich nur um

ein ganzzahliges Vielfaches von der Frequenz des zweiten Generators unterscheidet. Ebenso sind  $X$  und  $Y$  (stochastisch) abhängig, wenn diese ungekoppelt sind, aber von einem dritten Prozess  $Z$  getriggert werden. Damit von Abhängigkeit auf Kopplung geschlossen werden darf, müssen Fälle wie diese ausgeschlossen werden. In den meisten Fällen ist solch eine Schlussfolgerung möglich, wenn eine entsprechende Modellvorstellung der zu untersuchenden Systeme zugrunde liegt.

## 5.4 Nachweis von Abhängigkeiten mittels Zuwächsen

### 5.4.1 Abhängigkeitsnachweis über Zuwächse

Da es sich bei den Pfaden der Punktprozesse um monoton wachsende Funktionen handelt (siehe Abb. 5.2), können die Wahrscheinlichkeiten in Gl. (5.18) nur dann geschätzt werden, wenn eine ausreichende Anzahl von Beobachtungen vorliegt. Anders sieht es mit den Zuwächsen aus. Sind die zugrundeliegende Dynamik der Punktprozesse  $X$  und  $Y$  sowie die Kopplung zwischen ihnen zeitunabhängig, sind also die Zuwächse  $X_{t_i} - X_{t_i - \Delta}$  und  $X_{t_{i+1}} - X_{t_{i+1} - \Delta}$  stationär verteilt, so lautet ein Schätzer für die Verteilung der Zuwächse  $X_t - X_{t - \Delta}$  bezüglich des Zeitfensters  $(t - \Delta, t]$

$$P\{X_t - X_{t - \Delta} = x\} \approx \frac{1}{n} \sum_{i=1}^n \delta(X_{t_i}(\omega) - X_{t_i - \Delta}(\omega), x).$$

Dementsprechend kann die Verteilung eines stationär verteilten Zuwachses aus einer Beobachtung geschätzt werden.

Um ein Kriterium für Abhängigkeit zu erhalten, wird zunächst Gl. (5.18) von Ereigniszahlen auf Zuwächse umgeschrieben. Hierzu ist lediglich zu berücksichtigen, dass die Subtraktion eine stetige Abbildung ist. Folglich ist im Falle der Unabhängigkeit auch die gemeinsame Verteilung der Zuwächse von  $X$  und  $Y$  stochastisch unabhängig [Bauer (1991), Behnen & Neuhaus (1995), Billingsley (1995)]:

$$\begin{aligned} & P^{X_{t_{i+1}} - X_{t_i}, \dots, X_{t_2} - X_{t_1}, Y_{s_j} - Y_{s_{j-1}}, \dots, Y_{s_2} - Y_{s_1}} \\ &= P^{X_{t_{i+1}} - X_{t_i}, \dots, X_{t_2} - X_{t_1}} \otimes P^{Y_{s_j} - Y_{s_{j-1}}, \dots, Y_{s_2} - Y_{s_1}}, \end{aligned} \quad (5.19)$$

mit beliebigen  $0 \leq t_1 < \dots < t_i < t_{i+1}$ ,  $0 \leq s_1 < \dots < s_j$  und  $i, j \in \mathbb{N}$ . Die Abhängigkeit  $X_{t_{i+1}} - X_{t_i}$  von  $X_{t_i} - X_{t_{i-1}}$ ,  $\dots$  und  $Y_{t_{i+1}} - Y_{t_i}$ ,  $\dots$  ist in Abb. 5.5 illustriert, wobei  $s_j = t_{i+1}, \dots$  ist.

Durch Ausintegrieren von Gl. (5.19) folgt unmittelbar, dass die Punktprozesse  $X$  und  $Y$  nicht unabhängig sein können, wenn für beliebige  $0 \leq t_{i-1} < t_i$  und  $0 \leq s_{j-1} < s_j$  die Zuwächse  $X_{t_i} - X_{t_{i-1}}$  und  $Y_{s_j} - Y_{s_{j-1}}$  voneinander abhängen,

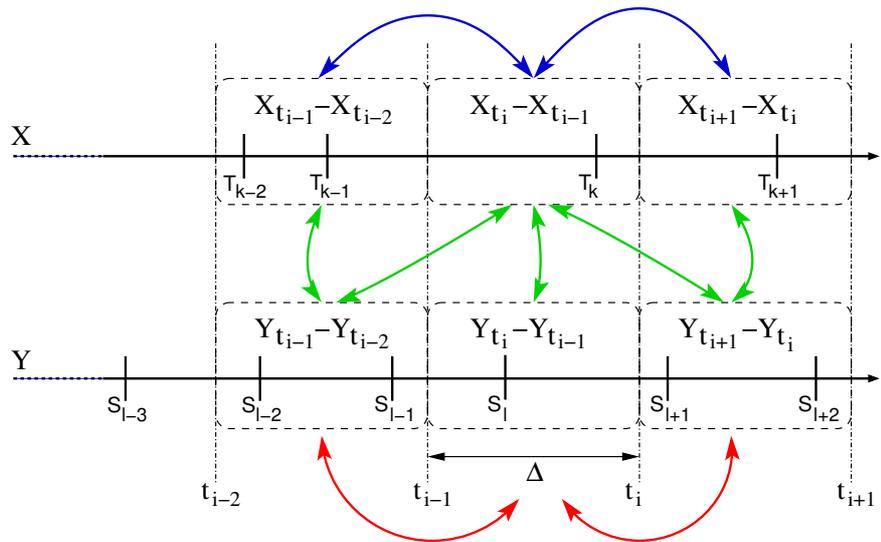


Abbildung 5.5: Abhängigkeit zwischen den Zuwächsen zweier Punktprozesse. Die Pfeile symbolisieren Abhängigkeiten bzw. Informationsaustausch zwischen den Zuwächsen von  $X$  und  $Y$ . Nicht alle Abhängigkeiten sind eingezeichnet. Die Zeitpunkte  $t_i$  wurden äquidistant gewählt ( $t_{i+1} - t_i = \Delta$ ).

das heißt wenn die Gleichung

$$\begin{aligned}
 P\{X_{t_i} - X_{t_{i-1}} = x, Y_{s_j} - Y_{s_{j-1}} = y\} \\
 = P\{X_{t_i} - X_{t_{i-1}} = x\} \cdot P\{Y_{s_j} - Y_{s_{j-1}} = y\} \quad (5.20)
 \end{aligned}$$

für irgendein  $x, y \in \mathbb{N}_0$  nicht erfüllt ist.

Zur Überprüfung der Gültigkeit von Gl. (5.20) ist es ausreichend, die gegenseitige Information  $M(X_{t_i} - X_{t_{i-1}}, Y_{s_j} - Y_{s_{j-1}})$  zwischen den Zuwächsen zu berechnen (siehe Gl. (2.9) in Absch. 2.1.1). Ist diese größer als Null, so sind beide Prozesse voneinander abhängig. In dem Fall, dass die gegenseitige Information der Zuwächse Null ist, kann keine Aussage getroffen werden. Eventuell wurden die Zeitpunkte  $t_{i-1}, t_i, s_{j-1}, s_j$  nur ungünstig gewählt, so dass lediglich  $X_{t_i} - X_{t_{i-1}}$  und  $Y_{s_j} - Y_{s_{j-1}}$  unabhängig sind.

**Beispiel: Gekoppelte Hindmarsh-Rose-Oszillatoren.** Im Folgenden soll am Beispiel zweier gekoppelter Hindmarsh-Rose-Oszillatoren [Hindmarsh & Rose (1984)] demonstriert werden, wie mittels Gl. (5.20) eine Abhängigkeit zwischen den beiden Oszillatoren nachgewiesen werden kann.

Ein (ungekoppelter) Hindmarsh-Rose-Oszillator  $Y$  ist durch ein dreidimensionales System von nichtlinearen gewöhnlichen Differentialgleichungen ge-

geben ([Hindmarsh & Rose (1984), Huerta et al. (1998)]),

$$\begin{aligned}\dot{y}_1 &= y_2 + 3y_1^2 - y_1^3 - y_3 + \varepsilon_2 \\ \dot{y}_2 &= 1 - 5y_1^2 - y_2 \\ \frac{1}{\mu} \dot{y}_3 &= -y_3 + \sigma(y_1 + \delta) - 1.\end{aligned}\tag{5.21}$$

Ursprünglich wurde dieser Oszillator als Modell für ein Neuron entwickelt. Dabei wird die Variable  $y_1$  als Membranpotential interpretiert und  $y_2$  repräsentiert die schnellen,  $y_3$  die langsamen Ionenströme. Hierbei ist das Subsystem  $(y_1, y_2)$  für die schnellen Oszillationen (Spikes) verantwortlich, während die langsamen (Bursts) von  $(y_1, y_3)$  verursacht werden. Die Zeitskalen sowie das Verhalten des Oszillators werden durch die Parameter  $\mu, \sigma, \delta, \varepsilon_2$  bestimmt. Sie werden auf  $\mu = 0.0021$ ,  $\sigma = 4$ ,  $\delta = 1.6$  gesetzt, so dass mit  $\varepsilon_2$  der Oszillator in die verschiedenen Regimes gesetzt werden kann: Ruhezustand  $\varepsilon_2 \leq 1.268$ , periodisches Bursten  $1.269 \leq \varepsilon_2 \leq 3.220$ , chaotisches Bursten  $3.221 \leq \varepsilon_2 \leq 3.325$ , kontinuierliches Spiken  $\varepsilon_2 > 3.325$ .

Als zweiter Oszillator  $X$  wird ebenfalls ein Hindmarsh-Rose-Oszillator mit den selben Parametern wie in Gl. (5.21) verwendet,

$$\begin{aligned}\dot{x}_1 &= x_2 + 3x_1^2 - x_1^3 - x_3 + \varepsilon_1 + gS(y_1)(r - x_1) \\ \dot{x}_2 &= 1 - 5x_1^2 - x_2 \\ \frac{1}{\mu} \dot{x}_3 &= -x_3 + \sigma(x_1 + \delta) - 1,\end{aligned}\tag{5.22}$$

wobei der Oszillator  $Y$  in den Oszillator  $X$  über eine Synapse hineinkoppelt. Hierzu wird die Kopplung durch die nichtlineare, gewöhnliche Differentialgleichung

$$(1 - S_\infty(z)) \dot{S}(z) = \frac{1}{100} (S_\infty(z) - S(z)),\tag{5.23}$$

mit

$$S_\infty(z) = \begin{cases} \tanh(z + 0.5) & \text{falls } z + 0.5 > 0 \\ 0 & \text{sonst} \end{cases}$$

modelliert [Pinto et al. (2000)]. Mit dem Schwellparameter  $r$  in Gl. (5.22) kann festgelegt werden, ob es sich um eine hemmende ( $r < 0$ ) oder um eine erregende ( $r > 0$ ) Synapse handelt. Hier wird  $r = 0.3$  fest gewählt. Der zweite Kopplungsparameter  $g$  ist die synaptische Leitfähigkeit, mit der die Kopplungsstärke eingestellt wird.

Für die folgenden Untersuchungen werden beide Oszillatoren in ein chaotisches Regime gesetzt,  $\varepsilon_1 = \varepsilon_2 = 3.280$ , und die Kopplungsstärke  $g$  wird zwischen 0 und 0.5 variiert. Die Lösung des Differentialgleichungssystems, bestehend aus Gl. (5.21), Gl. (5.22) und Gl. (5.23), erfolgt numerisch für 50 000 Zeiteinheiten.

Die hieraus resultierenden Zeitreihen werden mit einer Samplerate von einem Datenpunkt pro Zeiteinheit aufgenommen, so dass die Zeitreihen ebenfalls aus 50 000 Datenpunkten bestehen. Alle nachfolgenden Zeitangaben werden in der Zeiteinheit gegeben, in der das Differentialgleichungssystem gelöst wird.

Als Ereignis wird das Auftreten eines Spikes in dem Membranpotential  $x_1$  bzw.  $y_1$  angenommen, wobei der Ereigniszeitpunkt als jener Zeitpunkt definiert ist, in dem das Membranpotential einen Wert von 0.6 aufsteigend überschreitet. Ein Ausschnitt aus den Zeitreihen der Membranpotentiale  $x_1$  von Oszillator  $X$  und  $y_1$  von Oszillator  $Y$  sowie die entsprechenden Zeitreihen der Punktprozesse (siehe Gl. (5.6)) sind in Abb. 5.6 für den gekoppelten Fall  $g = 0.3$  dargestellt.

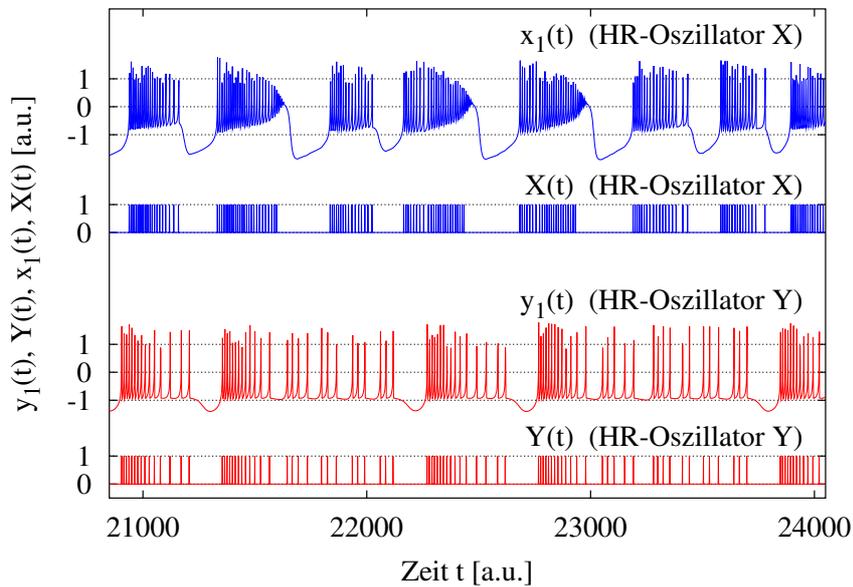


Abbildung 5.6: Membranpotential  $x_1$ ,  $y_1$  sowie die entsprechenden Zeitreihen der Punktprozesse zweier synaptisch gekoppelter Hindmarsh-Rose-Oszillatoren  $X$  und  $Y$ . Kopplungsstärke  $g = 0.3$ .

Im Folgenden wird der zu Oszillator  $X$  gehörige Punktprozess ebenfalls mit  $X$  und der zu Oszillator  $Y$  gehörige Punktprozess mit  $Y$  bezeichnet. Zur Untersuchung der Kopplung zwischen Oszillator  $X$  und  $Y$  werden ausschließlich die aus ihnen hervorgehenden Punktprozesse betrachtet. Hierfür wird die gegenseitige Information  $M(X_{t_i} - X_{t_{i-1}}, Y_{t_i+\tau} - Y_{t_{i-1}+\tau})$  zwischen den Zuwächsen  $X_{t_i} - X_{t_{i-1}}$  von Punktprozess  $X$  und den Zuwächsen  $Y_{t_i+\tau} - Y_{t_{i-1}+\tau}$  von  $Y$  als Funktion der Zeitverschiebung  $\tau$  berechnet. Die Zeitpunkte  $t_i$  sind dabei äquidistant mit  $t_i - t_{i-1} = \Delta$ . In Abb. 5.7 ist das Verhalten von  $M(X_{t_i} - X_{t_{i-1}}, Y_{t_i+\tau} - Y_{t_{i-1}+\tau})$  in Abhängigkeit von  $\tau$  für verschiedene Kopplungsstärken  $g = 0.0$  (ungekoppelt),  $g = 0.1, 0.3, 0.5$  (gekoppelt) dargestellt. Die Zuwächse von  $X$  und  $Y$  werden mit

einer fest gewählten Fensterbreite  $\Delta$  mit  $\Delta = 10$  berechnet.

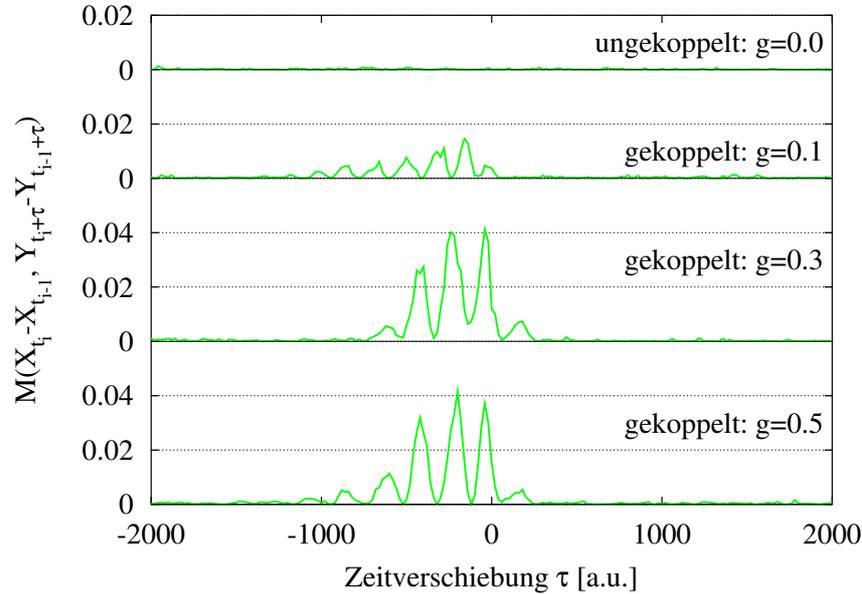


Abbildung 5.7: Gegenseitige Information  $M(X_{t_i} - X_{t_{i-1}}, Y_{t_i+\tau} - Y_{t_{i-1}+\tau})$  zwischen den Zuwächsen  $X_{t_i} - X_{t_{i-1}}$  und  $Y_{t_i+\tau} - Y_{t_{i-1}+\tau}$  als Funktion der Zeitverschiebung  $\tau$  für verschiedene Kopplungsstärken  $g$ . Die Fensterbreite wurde mit  $t_i - t_{i-1} = \Delta = 10$  fest gewählt.

Während für den ungekoppelten Fall ( $g = 0.0$ ) die gegenseitige Information Werte nahezu um Null liefert, ist eine deutliche temporale Struktur im gekoppelten Fall  $g = 0.1$  für Zeitverschiebungen  $\tau \approx -1000 \dots 0$  zu sehen. Insbesondere steigen die Werte von  $M(X_{t_i} - X_{t_{i-1}}, Y_{t_i+\tau} - Y_{t_{i-1}+\tau})$  mit der Kopplungsstärke  $g$ , wobei bei  $g > 0.1$  auch eine nichtkausale Abhängigkeit für  $\tau = 0 \dots 200$  zu erkennen ist. Diese ist auf die deterministische Dynamik der Hindmarsh-Rose-Oszillatoren zurückzuführen. Wegen der zugrunde liegenden Kausalität und der zeitlich abnehmenden Korrelation in den Dynamiken von Oszillator  $X$  und  $Y$  fallen für große  $|\tau|$  die Werte der gegenseitigen Information auf nahezu Null ab.

Der Wert Null selbst wird nur selten erreicht, da eine endliche Zeitreihe analysiert wird. Denn aufgrund von statistischen Fluktuationen in den Schätzungen der Verteilungen bleibt stattdessen, da die gegenseitige Information nichtnegativ ist, ein positiver Bias. Dieser Bias kann prinzipiell mit der “Finite sample”-Korrektur [Grassberger (1988), Panzeri & Treves (1996), Moddemeijer (1999), Roulston (1999)] herausgerechnet werden. Da diese Korrekturen aber nur im Mittel gelten, ist der exakte Bias nach wie vor unbekannt, was dazu führt, dass die gegenseitige Information sogar negative Werte annehmen kann. Daher wird hier keine Biaskorrektur vorgenommen, sondern die Werte der gegenseitigen Informa-

tion für kleine Zeitverschiebungen  $|\tau|$  werden mit jenen für große  $|\tau|$  verglichen. Letztere liegen, wie oben dargelegt wurde, in der Größenordnung des Biases.

Die Oszillationen in der gegenseitigen Information werden von der Dynamik des angetriebenen Prozesses  $X$  bestimmt. Dabei zeigt  $X$  aufgrund der Kopplung im zeitlichen Mittel ein quasi-periodisches Verhalten, bei dem sich stets Zeitintervalle ohne Ereignisse mit Intervallen, in denen sehr viele Ereignisse auftreten, abwechseln, siehe Abb. 5.6. Die Längen beider Intervalle variieren in der Zeit, im Mittel beträgt sie ungefähr 200 Zeiteinheiten. Dieser Wert entspricht der Periodendauer in den Oszillationen der gegenseitigen Information. Da sowohl die Länge der Zeitintervalle mit und ohne Ereignissen, als auch der Abstand zweier Ereignisse von  $X$  innerhalb der einzelnen Intervalle sich ändern, verlaufen die Abhängigkeiten mit der Zeit, so dass die Amplitude der Oszillationen mit  $\tau$  abnehmen. Insbesondere besitzt die gemeinsame Information der Zuwächse des angetriebenen Systems,  $M(X_{t_i} - X_{t_{i-1}}, X_{t_i+\tau} - X_{t_{i-1}+\tau})$ , die gleiche Periodizität.

Als Nächstes wird untersucht, welchen Einfluss die Fensterbreite der Zuwächse  $\Delta$  auf den Nachweis der Abhängigkeit hat. Hierzu wird abermals die gegenseitige Information  $M(X_{t_i} - X_{t_{i-1}}, Y_{t_i+\tau} - Y_{t_{i-1}+\tau})$  als Funktion der Zeitverschiebung  $\tau$  berechnet. Jetzt hingegen bleibt  $g = 0.3$  konstant und  $t_i - t_{i-1} = \Delta$  wird als Parameter von 1 bis 100 variiert. Die Ergebnisse sind in Abb. 5.8 zu sehen.

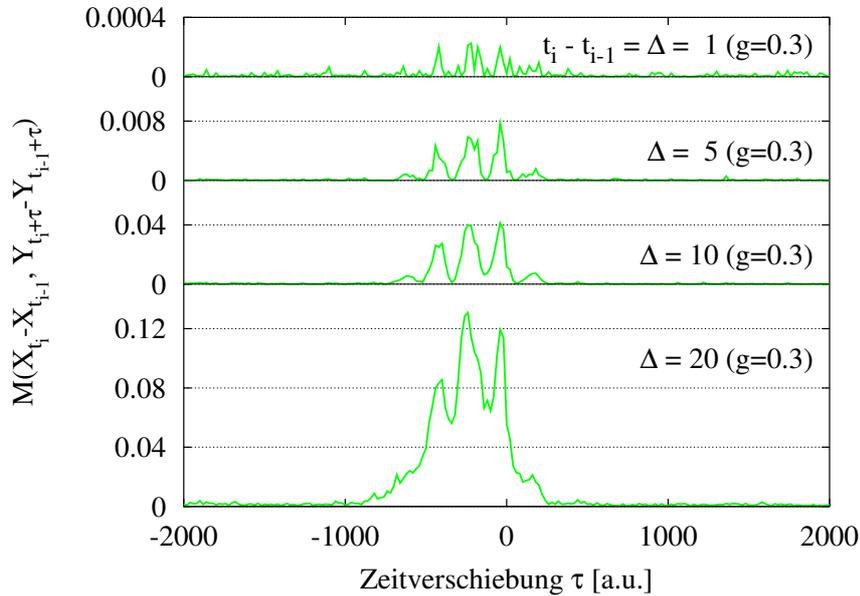


Abbildung 5.8: Gegenseitige Information  $M(X_{t_i} - X_{t_{i-1}}, Y_{t_i+\tau} - Y_{t_{i-1}+\tau})$  zwischen den Zuwächsen  $X_{t_i} - X_{t_{i-1}}$  und  $Y_{t_i+\tau} - Y_{t_{i-1}+\tau}$  als Funktion der Zeitverschiebung  $\tau$  für verschiedene Fensterbreiten  $t_i - t_{i-1} = \Delta$  und mit einer festgelegten Kopplung von  $g = 0.3$ .

Hier zeigt sich eine temporale Struktur im Bereich von  $\tau = -1000 \dots 200$ . Mit zunehmender Fensterbreite  $t_i - t_{i-1} = \Delta$  ist sie ausgeprägter. Ab  $\Delta = 20$  bleibt der Kurvenverlauf von  $M(X_{t_i} - X_{t_{i-1}}, Y_{t_i+\tau} - Y_{t_{i-1}+\tau})$  qualitativ unverändert. Lediglich ihre Werte und die des Biases nehmen zu. Zwar nehmen die Werte der gegenseitigen Information mit der Fensterbreite ab, dennoch kann bei  $\Delta = 1$  eine Abhängigkeit nachgewiesen werden. So ergab die gegenseitige Information bei  $\tau \approx -200$  Werte um  $2 \cdot 10^{-4}$ , während der Bias für  $|\tau| > 1000$  unterhalb von  $5 \cdot 10^{-5}$  lag, siehe Abb. 5.8. Dabei waren die temporalen Strukturen umso stärker ausgeprägt, je größer der Kopplungsparameter war. Insbesondere konnte eine Kopplung bereits bei  $g = 0.1$  nachgewiesen werden. Zur Verifizierung der Robustheit dieser Methode ist des Weiteren die gegenseitige Information der Zuwächse im ungekoppelten Fall für verschiedene Fensterbreiten berechnet worden. Hier zeigten sich keine temporalen Strukturen.

Die optimale Wahl der Fensterbreite  $\Delta$  wird im Allgemeinen von der Zeitskala der Strukturen abhängen, die den stärksten Einfluss aufeinander haben. Findet eine Wechselwirkung zwischen den Punktprozessen über einzelne Ereignisse (Spikes) statt, so wird  $\Delta$  meist in der Größenordnung der Zeitauflösung liegen. Sind hingegen ganze Ereignisgruppen für die Kopplung verantwortlich, dann werden Werte innerhalb der Größenordnung dieser Gruppen eine geeignete Wahl für  $\Delta$  sein. Andererseits kann durch die Wahl der Fensterbreite die Untersuchung auf Strukturen mit einer festgelegten Zeitskala eingeschränkt werden.

Zum Abschluss wurde die Signifikanz der Methode betrachtet. Hier zeigte sich, dass bei kürzeren Zeitreihen mit einer Länge bis zu 10 000, dies entspricht in etwa 500 Ereignisse, die Signifikanz im Wesentlichen unverändert blieb, für  $g = 0.3$  siehe Abb. 5.9.

Zur Berechnung des Informationstransfers zwischen gekoppelten Neuronen benutzten Eguia et al. (2000) ebenfalls Methoden der Informationstheorie. Hierzu teilten sie die Zeitreihen in Bins der Breite  $\Delta$  auf. Diesen Bins wurde der Wert 0 bzw. 1 zugewiesen, je nachdem, ob in ihnen ein Ereignis (Spike) stattgefunden hatte oder nicht. Anschließend wurden  $l$  Bins zu einem Wort zusammengefasst und die gegenseitige Information bezüglich dieser Wörter berechnet.

Diese Methode lässt sich auf die Untersuchung von Gl. (5.19) zurückführen. Hierzu ist  $i + 1 = j = l + 1$  zu setzen und die Zeitpunkte  $t_k$  und  $s_k$  sind so zu wählen, dass  $t_k - t_{k-1} = s_k - s_{k-1} = \Delta$  für alle  $k = 2, \dots, l + 1$  gilt.

Es zeigte sich aber, dass bei diesem Verfahren deutlich längere Zeitreihen für eine robuste Berechnung der gemeinsamen Informationen notwendig sind, als wenn lediglich die Zuwächse  $X_{t_{l+1}} - X_{t_1}$  und  $Y_{s_{l+1}} - Y_{s_1}$  verwendet werden. Der Grund hierfür ist, dass bei Eguia et al. (2000) die gemeinsame Information zwischen  $l$ -dimensionalen Zufallsvariablen berechnet werden muss, aber aufgrund eines endlichen Abstands zwischen den Spikes im Allgemeinen weniger als  $l$  Spikes in die Zeitintervalle  $(t_1, t_{l+1}]$  bzw.  $(s_1, s_{l+1}]$  fallen. Folglich lassen sich die Verteilungen von  $X_{t_{l+1}} - X_{t_1}$  und  $Y_{s_{l+1}} - Y_{s_1}$  besser schätzen als die Verteilungen der

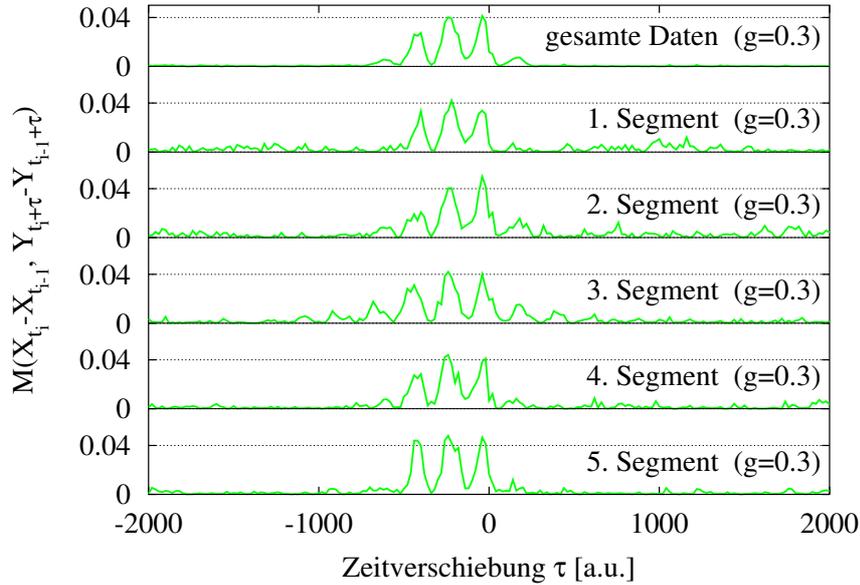


Abbildung 5.9: Gegenseitige Information  $M(X_{t_i} - X_{t_{i-1}}, Y_{t_i+\tau} - Y_{t_{i-1}+\tau})$  zwischen den Zuwächsen  $X_{t_i} - X_{t_{i-1}}$  und  $Y_{t_i+\tau} - Y_{t_{i-1}+\tau}$  als Funktion der Zeitverschiebung  $\tau$ . Fensterbreiten  $t_i - t_{i-1} = \Delta = 10$ , Kopplungsstärke  $g = 0.3$ . Oberste Kurve: gesamte Zeitreihe (50000 Punkte  $\approx$  2500 Ereignisse). Untere Kurven: verschiedene Segmente der gesamten Zeitreihe mit 10000 Punkten  $\approx$  500 Ereignisse.

von Eguia et al. verwendeten Wörter.

## 5.4.2 Anwendung von Momente und Koinzidenzen

### 5.4.2.1 Kovarianz zwischen Zuwächsen

In Abschn. 5.4.1 wurde eine Methode vorgestellt, mit der es möglich ist, Abhängigkeiten zwischen zwei Punktprozessen nachzuweisen. Hierzu wurde die Unabhängigkeit der Zuwächse von Punktprozessen (Gl. (5.20)) überprüft. Dies erfolgte, indem die gegenseitige Information zwischen den Zuwächsen  $X_{t_i} - X_{t_{i-1}}$  und  $Y_{s_j} - Y_{s_{j-1}}$  mit  $0 \leq t_{i-1} < t_i$  und  $0 \leq s_{j-1} < s_j$  berechnet wurde.

Insbesondere bei langen Zeitreihen kann die Berechnung der gegenseitigen Information sehr rechenintensiv werden, weshalb ein einfacheres Verfahren wünschenswert ist. Dieses kann ebenfalls aus Gl. (5.20) abgeleitet werden. So folgt aus der Gültigkeit von Gl. (5.20) unmittelbar, dass die Kovarianzen zwischen den Zuwächsen

$$\begin{aligned} \text{Cov} [X_{t_i} - X_{t_{i-1}}, Y_{s_j} - Y_{s_{j-1}}] \\ = \text{E} [(X_{t_i} - X_{t_{i-1}} - \text{E} [X_{t_i} - X_{t_{i-1}}]) \cdot (Y_{s_j} - Y_{s_{j-1}} - \text{E} [Y_{s_j} - Y_{s_{j-1}}])] \end{aligned} \quad (5.24)$$

verschwinden müssen [Bauer (1991)], das heißt

$$\text{Cov} [X_{t_i} - X_{t_{i-1}}, Y_{s_j} - Y_{s_{j-1}}] = 0. \quad (5.25)$$

Dabei ist  $E[Z] = \int_{\Omega} Z(\omega) P(d\omega)$  der Erwartungswert der Zufallsvariablen  $Z$ . Hierbei ist zu beachten, dass stets aus der Unabhängigkeit (Gl. (5.20)) die Unkorreliertheit (Gl. (5.25)) folgt. Die Umkehrung gilt im Allgemeinen nicht.

Aus einer elementaren Rechnung folgt, dass die Kovarianz der Zuwächse durch Momente höherer Ordnung (siehe Abschn. 5.2) gegeben ist,

$$\begin{aligned} & \text{Cov} [X_{t_i} - X_{t_{i-1}}, Y_{s_j} - Y_{s_{j-1}}] \\ &= E [(X_{t_i} - X_{t_{i-1}}) \cdot (Y_{s_j} - Y_{s_{j-1}})] - E [X_{t_i} - X_{t_{i-1}}] E [Y_{s_j} - Y_{s_{j-1}}]. \end{aligned} \quad (5.26)$$

Folglich müssen für einen Test auf Abhängigkeit nur noch die Momente der Zuwächse berechnet und nicht deren Verteilungen geschätzt werden, wie es bei der gegenseitigen Information notwendig war (siehe Abschn. 5.4.1). Aufgrund der Linearität der Momente ist allerdings zu erwarten, dass bei nichtlinearer Kopplung die Signifikanz eines Tests auf Abhängigkeit, basierend auf der Kovarianz, geringer ausfallen wird als bei einem, dem die gegenseitige Information zugrunde liegt.

Wie aus Gl. (5.26) ersichtlich ist, hängt der Wert der Kovarianz von den Fensterbreiten  $t_i - t_{i-1}$  und  $s_j - s_{j-1}$  ab. Hier bietet sich die Möglichkeit, diese Abhängigkeit zu eliminieren und somit einen absoluten Wert für die Korrelation zwischen den Punktprozessen  $X$  und  $Y$  zu erhalten. Denn wenn die einzelnen Momente in Gl. (5.26) eine Dichte  $m_{X,Y}(t, s)$ ,  $m_X(t)$  bzw.  $m_Y(s)$  besitzen (siehe Abschn. 5.2), dann besitzt auch die Kovarianz der Zuwächse eine Dichte  $c(t, s) = m_{X,Y}(t, s) - m_X(t) m_Y(s)$  [Daley & Vere-Jones (1972)] mit

$$\text{Cov} [X_{t_i} - X_{t_{i-1}}, Y_{s_j} - Y_{s_{j-1}}] = \int_{t_{i-1}}^{t_i} \int_{s_{j-1}}^{s_j} c(t, s) ds dt.$$

Sie gibt die Korrelation zwischen den Punktprozessen pro Zeiteinheiten an und kann mittels Grenzwertbildung ermittelt werden:

$$c(t, s) = \lim_{\substack{\Delta_1 \rightarrow 0_+ \\ \Delta_2 \rightarrow 0_+}} \frac{\text{Cov} [X_t - X_{t-\Delta_1}, Y_s - Y_{s-\Delta_2}]}{\Delta_1 \Delta_2}.$$

Insbesondere folgt aus  $c(t, s) > 0$  bzw.  $c(t, s) < 0$  für  $t_{i-1} < t \leq t_i$ ,  $s_{j-1} < s \leq s_j$ , dass  $\text{Cov} [X_{t_i} - X_{t_{i-1}}, Y_{s_j} - Y_{s_{j-1}}] \neq 0$  ist, also die Abhängigkeit beider Punktprozesse voneinander.

Analog zu den Dichten der Momente ist ein Schätzer für die Kovarianzdichte  $c(t, s)$  durch

$$\hat{c}_{\Delta_1, \Delta_2}(t, s) = \frac{\text{Cov} [X_t - X_{t-\Delta_1}, Y_s - Y_{s-\Delta_2}]}{\Delta_1 \Delta_2}$$

gegeben, wobei die Fensterbreiten  $\Delta_1$  und  $\Delta_2$  klein zu wählen sind. Allerdings fehlt auch hier im Falle endlicher Zeitreihen ein Kriterium für die optimale Wahl von  $\Delta_1$  und  $\Delta_2$ .

Werden Experimente mehrfach wiederholt, so begegnet einem oft folgendes Problem: Die Ereignisse in einem Zeitfenster variieren teilweise erheblich von Experiment zu Experiment. Dies hat zur Folge, dass die Kovarianzwerte der Zuwächse eine sehr hohe Streuung aufweisen. Um dem zu begegnen, ist es häufig angebracht, die Kovarianzen auf die Standardabweichungen der Zuwächse zu normieren [Kantz & Schreiber (1997)]:

$$\begin{aligned} \text{Corr} [X_{t_i} - X_{t_{i-1}}, Y_{s_j} - Y_{s_{j-1}}] \\ = \frac{\text{Cov} [X_{t_i} - X_{t_{i-1}}, Y_{s_j} - Y_{s_{j-1}}]}{\sqrt{\text{Var} [X_{t_i} - X_{t_{i-1}}] \cdot \text{Var} [Y_{s_j} - Y_{s_{j-1}}]}}. \end{aligned} \quad (5.27)$$

Hierbei ist  $\text{Var} [Z] = \text{E} [(Z - \text{E} [Z])^2]$  die Varianz von  $Z$  [Bauer (1991), Behnen & Neuhaus (1995)]. Die Größe  $\text{Corr} [X_{t_i} - X_{t_{i-1}}, Y_{s_j} - Y_{s_{j-1}}]$  wird Korrelation zwischen den Zuwächsen  $X_{t_i} - X_{t_{i-1}}$  und  $Y_{s_j} - Y_{s_{j-1}}$  genannt. In der Literatur wird sie oft auch als Kreuzkorrelation bezeichnet. Mit einer elementaren Rechnung folgt aus  $\text{Var} [X \pm Y] \geq 0$ , dass stets  $|\text{Corr} [X, Y]| \leq 1$  gilt.

**Beispiel: Gekoppelte Hindmarsh-Rose-Oszillatoren.** Die Anwendung der Korrelation zum Nachweis von Kopplung wird am Beispiel der Hindmarsh-Rose-Oszillatoren, die in Abschn. 5.4.1 vorgestellt wurden, gezeigt. Auch hier werden die Zeitpunkte äquidistant gewählt, mit  $t_i - t_{i-1} = \Delta$ ,  $t_i = s_i$ .

Zunächst wird die Korrelation  $\text{Corr} [X_{t_i} - X_{t_{i-1}}, Y_{t_i+\tau} - Y_{t_{i-1}+\tau}]$  zwischen den Zuwächsen als Funktion der Zeitverschiebung  $\tau$  bei fest gewählter Fensterbreite  $\Delta = 10$  für die Kopplungsstärken  $g = 0.0, 0.1, 0.3$  und  $0.5$  berechnet.

Wie aus Abb. 5.10 zu ersehen ist, tritt im gekoppelten Fall  $g > 0.0$  eine temporale Struktur im Bereich  $\tau = -1000 \dots 200$  auf. Wie bei der gegenseitigen Information (siehe Abschn. 5.4.1) ist die nichtkausale Abhängigkeit bei  $g > 0.1$  auf den zugrunde liegenden Determinismus in der Dynamik der Oszillatoren zurückzuführen. Des Weiteren ist die Periode, mit der die temporalen Strukturen oszillieren, bei der Korrelation doppelt so lang wie bei der gegenseitigen Information, da die gegenseitige Information kein negatives Vorzeichen hat und somit die Minima nach oben klappt. Im Gegensatz zur gegenseitigen Information weist die Korrelation relativ hohe Fluktuationen für große  $|\tau|$  auf.

Auch die Korrelationswerte  $\text{Corr} [X_{t_i} - X_{t_{i-1}}, Y_{t_i+\tau} - Y_{t_{i-1}+\tau}]$  wachsen bei konstanter Kopplung mit der Fensterbreite  $\Delta$  an. Dies ist für  $g = 0.3$  in Abb. 5.11 gezeigt. Bei einer genaueren Untersuchung des ungekoppelten Falls ( $g = 0.0$ ) zeigt sich, dass mit steigendem  $\Delta$  die Korrelation in Abhängigkeit der Zeitverschiebung  $\tau$  eine oszillatorische Struktur herausbildet, deren Amplitude  $\approx 0.06$  bei  $\Delta = 20$

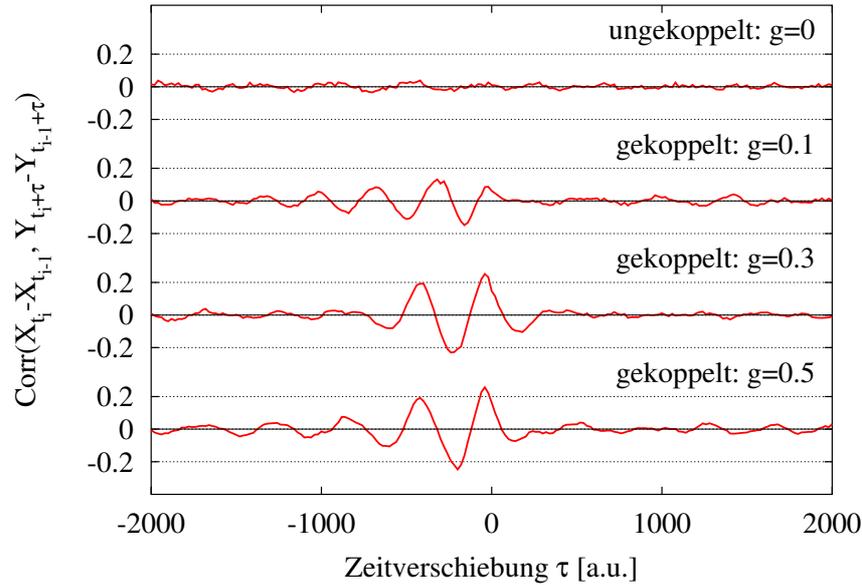


Abbildung 5.10: Korrelation  $\text{Corr}[X_{t_i} - X_{t_{i-1}}, Y_{t_i+\tau} - Y_{t_{i-1}+\tau}]$  zwischen den Zuwächsen  $X_{t_i} - X_{t_{i-1}}$  und  $Y_{t_i+\tau} - Y_{t_{i-1}+\tau}$  als Funktion der Zeitverschiebung  $\tau$  für verschiedene Kopplungsstärken  $g$ . Die Fensterbreite wurde mit  $t_i - t_{i-1} = \Delta = 10$  fest gewählt.

beträgt. Diese Oszillationen treten aufgrund der Endlichkeit der Zeitreihen und der oszillatorischen Natur der Prozesse auf. Erst für längere Zeitreihen mitteln sich diese Oszillationen heraus. Des Weiteren stellte sich heraus, dass bei einer Fensterbreite von  $\Delta = 1$  eine Kopplung von  $g > 0.1$  notwendig ist, um zwischen temporaler Struktur in der Korrelation und den Fluktuationen des Biases unterscheiden zu können.

#### 5.4.2.2 Koinzidenzen

Eine häufig verwendete Methode, um Kopplung zwischen Punktprozessen festzustellen, ist das Zählen von Koinzidenzen. Dieser Methode liegt die Modellvorstellung zugrunde, dass, wenn der Prozess  $Y$  in  $X$  koppelt, ein Ereignis von  $Y$  zum Zeitpunkt  $S_l(\omega)$  innerhalb eines Zeitfensters  $\Delta$  ein Ereignis in  $X$  zur Zeit  $T_k(\omega)$  auslöst. Dies ist in Abb. 5.12 schematisch dargestellt. Dabei kann das Triggern des Ereignisses in  $X$  gegebenenfalls mit einer Zeitverzögerung  $\tau$  erfolgen. Dementsprechend wird ein Paar von Ereigniszeiten  $(S_l(\omega), T_k(\omega))$  eine Koinzidenz genannt, wenn

$$S_l(\omega) \leq T_k(\omega) - \tau < S_l(\omega) + \Delta \quad (5.28)$$

gilt [Grün et al. (1999), Riehle et al. (1997), Schreiber & Schmitz (2000a)].

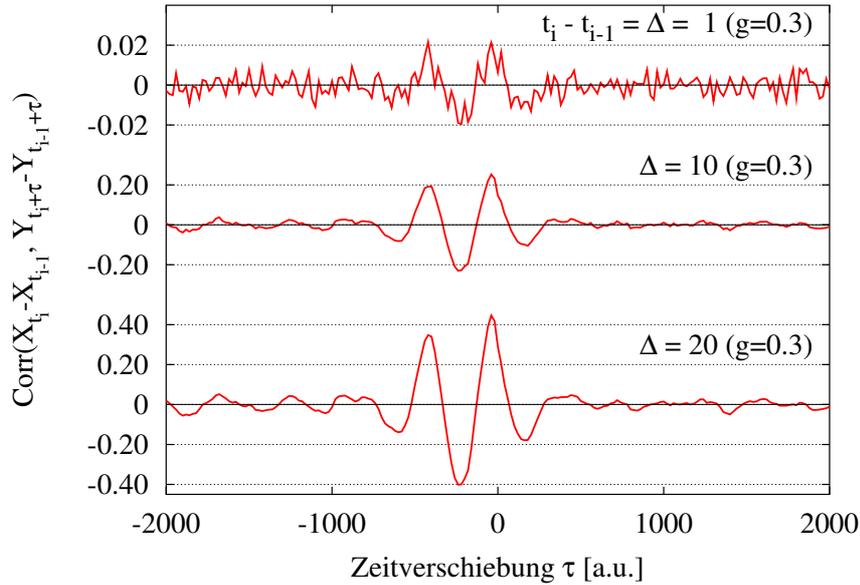


Abbildung 5.11: Korrelation  $\text{Corr}[X_{t_i} - X_{t_{i-1}}, Y_{t_i+\tau} - Y_{t_{i-1}+\tau}]$  zwischen den Zuwächsen  $X_{t_i} - X_{t_{i-1}}$  und  $Y_{t_i+\tau} - Y_{t_{i-1}+\tau}$  als Funktion der Zeitverschiebung  $\tau$  für verschiedene Fenstergrößen  $t_i - t_{i-1} = \Delta$  bei fest gewählter Kopplung  $g = 0.3$ .

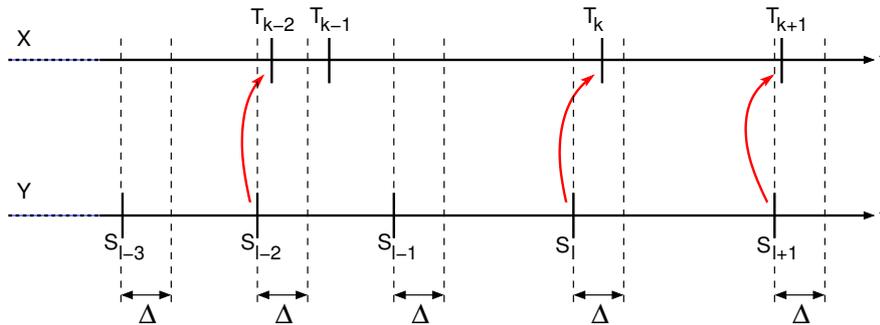


Abbildung 5.12: Schematische Darstellung der Koinzidenzen zwischen zwei Punktprozessen  $X$  und  $Y$ , wobei  $Y$  in  $X$  koppelt (ohne Zeitverzögerung). Die Ereigniszeiten von  $Y$  sind mit  $S_l$  und jene von  $X$  mit  $T_k$  gekennzeichnet.

Werden Zeitreihen von Punktprozessen beobachtet, so haben benachbarte Ereignisse eines Prozesses einen zeitlichen Mindestabstand  $\delta$ , der von der Zeitauflösung herrührt. Daher lassen sich die mittleren Koinzidenzanzahlen durch die Momente der Punktprozesse ausdrücken. Denn mit der Zeitauflösung  $\delta$  gibt  $Y_{t_i+\delta} - Y_{t_i}$  an, ob der entsprechende Datenpunkt ein Ereignis repräsentiert oder nicht. Folglich liefert  $X_{t_i+\tau+\Delta} - X_{t_i+\tau}$  die Anzahl der Ereignisse von  $X$  im Intervall  $(t_i + \tau, t_i + \tau + \Delta]$ . Hier wird vorausgesetzt, dass  $\tau$  und  $\Delta$  ganzzahlige

Vielfache von  $\delta$  sind. Somit ist  $(Y_{t_i+\delta} - Y_{t_i}) \cdot (X_{t_i+\tau+\Delta} - X_{t_i+\tau})$  die Anzahl der Koinzidenzpaare eines Experimentes. Die mittlere Koinzidenzanzahl ist dann durch den Erwartungswert

dieser Koinzidenzpaare gegeben:

$$\text{Koinz-}\# = \text{E}[(Y_{t_i+\delta} - Y_{t_i}) \cdot (X_{t_i+\tau+\Delta} - X_{t_i+\tau})]. \quad (5.29)$$

Im unabhängigen Fall ist die mittlere Koinzidenzanzahl gegeben durch

$$\text{Koinz-}\#_{\text{unabh}} = \text{E}[Y_{t_i+\delta} - Y_{t_i}] \cdot \text{E}[X_{t_i+\tau+\Delta} - X_{t_i+\tau}]. \quad (5.30)$$

Die Differenz beider Werte ist somit ein Maß für die Abhängigkeit. Insbesondere folgt aus einem Vergleich mit Gl. (5.26), dass dieses Maß gleich der Kovarianz ist,

$$\begin{aligned} \text{Koinz-}\# - \text{Koinz-}\#_{\text{unabh}} &= \text{Cov}[Y_{t_i+\delta} - Y_{t_i}, X_{t_i+\tau+\Delta} - X_{t_i+\tau}] \\ &= \sum_{j=1}^{\Delta/\delta} \text{Cov}[Y_{t_i+\delta} - Y_{t_i}, X_{t_i+\tau+j\delta} - X_{t_i+\tau+(j-1)\delta}]. \end{aligned} \quad (5.31)$$

Für die letzte Gleichheit wurde die Additivität der Zuwächse ausgenutzt,  $X_{t_{i+2}} - X_{t_i} = X_{t_{i+2}} - X_{t_{i+1}} + X_{t_{i+1}} - X_{t_i}$  mit  $t_i < t_{i+1} < t_{i+2}$ .

Da auch  $X_{t_i+\tau+j\delta} - X_{t_i+\tau+(j-1)\delta}$  nur die Werte Null oder Eins annimmt, ist der Wert  $(\text{Koinz-}\# - \text{Koinz-}\#_{\text{unabh}})/(\Delta/\delta)$  nichts anderes als die über ein Zeitfenster  $\Delta$  gemittelte Kovarianz einer Zeitreihen von Punktprozessen (Gl. (5.6)), die um 0 bis  $\Delta/\delta$  Zeitschritte gegeneinander verschoben sind.

**Ereignismuster.** Unter der Annahme, dass die Punktprozesse  $X$  und  $Y$  stationär verteilte Zuwächse haben, kann die Wahrscheinlichkeit angegeben werden, wie häufig es vorkommt, dass zum Beispiel Prozess  $X$  ein Ereignis hat und gleichzeitig  $Y$  keins, siehe [Grün et al. (2002)] für Details. Im Folgenden seien die Zeitpunkte  $t_i$  durch die Datenaufzeichnung so vorgegeben, dass  $t_i - t_{i-1} = \delta$  für  $i = 1, \dots, T$  gilt, wobei  $\delta$  die Zeitaufösung der Messung ist und  $T$  die Anzahl der Datenpunkte. Somit repräsentiert das Zeitintervall  $(t_{i-1}, t_i]$  genau einen Datenpunkt.

Die Wahrscheinlichkeit, mit der  $X$  oder  $Y$  in einem Datenpunkt ein Ereignis hat, ist durch die Ereignisrate  $\lambda_X$  bzw.  $\lambda_Y$  gegeben,  $P\{X_{t_i} - X_{t_{i-1}} = 1\} = \lambda_X \delta$ ,  $P\{Y_{t_i} - Y_{t_{i-1}} = 1\} = \lambda_Y \delta$ . Das Ereignismuster  $(x, y) \in \{0, 1\}^2$  tritt, wenn  $X$  und  $Y$  unabhängig sind, mit der Wahrscheinlichkeit  $P\{X_{t_i} - X_{t_{i-1}} = x, Y_{t_i} - Y_{t_{i-1}} = y\} = P\{X_{t_i} - X_{t_{i-1}} = x\} \cdot P\{Y_{t_i} - Y_{t_{i-1}} = y\}$  auf.

Bezeichnet  $N(x, y, T)$  die Zufallsvariable, die angibt, wie häufig innerhalb des Datensatzes das Muster  $(x, y)$  vorkommt, so lautet deren Verteilung für un-

abhängiges  $X$  und  $Y$

$$P\{N(x, y, T) = n\} = \frac{T!}{n! \cdot (T - n)!} \cdot (P\{X_{t_i} - X_{t_{i-1}} = x\} \cdot P\{Y_{t_i} - Y_{t_{i-1}} = y\})^n \\ \times (1 - P\{X_{t_i} - X_{t_{i-1}} = x\} \cdot P\{Y_{t_i} - Y_{t_{i-1}} = y\})^{T-n}.$$

Folglich ist die Wahrscheinlichkeit, mit der im unabhängigen Fall mindestens  $n_{\min}$  bzw. maximal  $n_{\max}$  Ereignismuster  $(x, y)$  gefunden werden, durch

$$P\{\min N(x, y, T) = n_{\min}\} = \sum_{n=n_{\min}}^{\infty} P\{N(x, y, T) = n\},$$

bzw.

$$P\{\max N(x, y, T) = n_{\max}\} = \sum_{n=0}^{n_{\max}} P\{N(x, y, T) = n\}$$

gegeben. Dies ermöglicht es, die Werte  $n_{\min}$  und  $n_{\max}$  so zu bestimmen, dass ein statistischer Test zum Signifikanzlevel  $\alpha$  ausgeführt werden kann, um auf nicht zufällige Koinzidenzen zu testen. Dieses Verfahren lässt sich direkt auf mehr als zwei Punktprozesse erweitern, siehe [Grün et al. (2002)].

## 5.5 Nachweis von Kopplung mittels Zuwächsen

Um Kopplung zwischen Punktprozessen untersuchen zu können, wäre es von Vorteil, wenn Gl. (5.17) ebenfalls auf Zuwächse umgeschrieben werden könnte. Dies ist aber nicht ohne weiteres möglich. Denkt man aber an Neuronen, die über eine Synapse gekoppelt sind, so ist es plausibel anzunehmen, dass das Neuron  $X$  zusätzlich zum Spiken angeregt wird, wenn die Anzahl der Ereignisse von Neuron  $Y$  innerhalb eines Zeitfensters einen bestimmten Schwellenwert übersteigt.

Häufig stellen Punktprozesse Abbilder von dynamischen Systemen oder kontinuierlichen Prozessen in dem Sinne dar, dass die Ereigniszeiten die Zeitpunkte markieren, in denen die dynamischen Variablen einen Schwellenwert überschreiten. Dementsprechend kann jedem dynamischen System ein Punktprozess zugeordnet werden, indem die Durchstoßzeiten eines Poincaré-Schnitts aufgezeichnet werden [Kantz & Schreiber (1997), Hegger & Kantz (1997)]. Dieses Verfahren wurde bereits verwendet, um aus den Hindmarsch-Rose-Oszillatoren des vorherigen Beispiels Punktprozesse zu erzeugen. Unter diesen Bedingungen ist ebenfalls die Annahme naheliegend, dass die Zuwächse eines späteren Zeitfensters von den vorherigen abhängen. Damit in jedem Zeitpunkt nur endlich viele Zuwächse aus der Vergangenheit berücksichtigt werden müssen, das heißt die Zuwächse

sollen Markov-Prozesse darstellen, muss zusätzlich gefordert werden, dass die betrachteten Zeitpunkte  $0 = t_0 < t_1 < t_2 < \dots < t_{i+1}$  äquidistante Abstände haben:

$$t_{j+1} - t_j = \Delta \quad \forall j = 0, \dots, i.$$

Entsprechend dieser Modellvorstellung wird in dem Fall, dass der Prozess  $Y$  in  $X$  koppelt, der Zuwachs  $X_{t_{i+1}} - X_{t_i}$  von den vorherigen Zuwächsen  $X_{t_i} - X_{t_{i-1}}, \dots, X_{t_{i-k+1}} - X_{t_{i-k}}$  sowie von  $Y_{t_{i+1}} - Y_{t_i}, \dots, Y_{t_{i-l+1}} - Y_{t_{i-l}}$  abhängen ( $k, l \in \mathbb{N}$ ), während im ungekoppelten Fall stets

$$\begin{aligned} P^{X_{t_{i+1}} - X_{t_i} | (\mathbf{X}_{t_i} - \mathbf{X}_{t_{i-1}})^{(k) = \mathbf{x}_i^{(k)}, (\mathbf{Y}_{t_{i+1}} - \mathbf{Y}_{t_i})^{(l) = \mathbf{y}_{i+1}^{(l)}}} \\ = P^{X_{t_{i+1}} - X_{t_i} | (\mathbf{X}_{t_i} - \mathbf{X}_{t_{i-1}})^{(k) = \mathbf{x}_i^{(k)}} \end{aligned} \quad (5.32)$$

gilt. Hierbei wurde die Abkürzung  $(\mathbf{X}_{t_i} - \mathbf{X}_{t_{i-1}})^{(k)} = (X_{t_i} - X_{t_{i-1}}, \dots, X_{t_{i-k+1}} - X_{t_{i-k}})$  für den Produktzuwachs verwendet. Zur Illustration siehe Abb. 5.13. Da

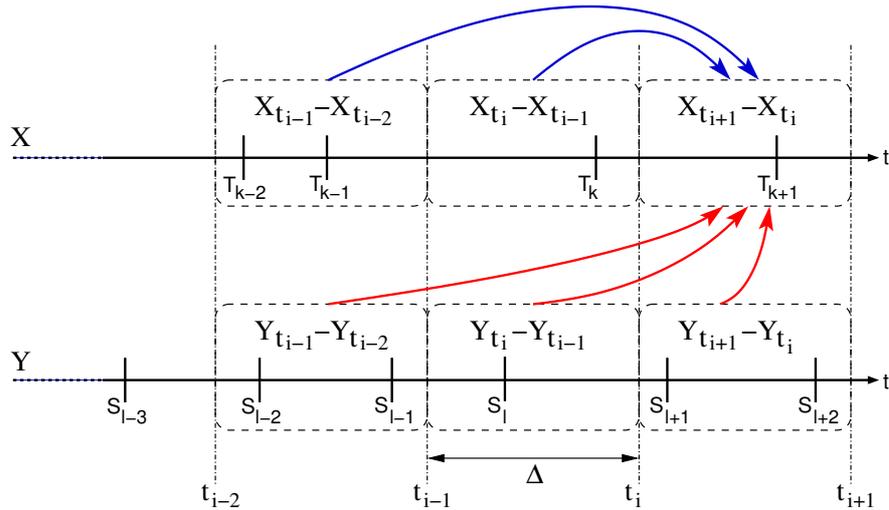


Abbildung 5.13: Abhängigkeit zwischen den Zuwächsen zweier Punktprozesse. Die Pfeile symbolisieren den Einfluss der vorherigen Zuwächse auf den Zuwachs  $X_{t_{i+1}} - X_{t_i}$ . Alle Zeitpunkte  $t_i$  haben einen äquidistanten Abstand ( $t_{i+1} - t_i = \Delta$ ).

Ereignisse von  $Y$  im Intervall  $(t_i, t_{i+1}]$  auch noch einen Einfluss auf den Zuwachs  $X_{t_{i+1}} - X_{t_i}$  haben können, muss in der Bedingung der Übergangsverteilung in Gl. (5.32) der gegenwärtige Zuwachs  $Y_{t_{i+1}} - Y_{t_i}$  von  $Y$  zugelassen werden.

Die Quantifizierung der Kopplung kann wieder mit Hilfe der Transferentropie  $T(X_{t_{i+1}} - X_{t_i} | (\mathbf{X}_{t_i} - \mathbf{X}_{t_{i-1}})^{(k)}, (\mathbf{Y}_{t_{i+1}} - \mathbf{Y}_{t_i})^{(l)})$  erfolgen, indem jetzt die Zuwächse der Prozesse betrachtet werden, siehe auch Abschn. 2.1.2, Gl. (2.21). Ist ein Nachweis der Kopplung ausreichend, so kann hierfür die Transferentropie

$T(X_{t_{i+1}} - X_{t_i} | (\mathbf{X}_{t_i} - \mathbf{X}_{t_{i-1}})^{(k)}, Y_{t_{i+1}-\tau} - Y_{t_i-\tau})$  als Funktion der *Zeitverzögerung*  $\tau \geq 0$  betrachtet werden, wobei wegen der angenommenen kausalen Kopplung nur nichtnegative  $\tau$  berücksichtigt werden müssen. Ist diese Transferentropie für ein  $\tau \geq 0$  signifikant von Null verschieden, so hat mindestens einer der Zuwachs  $Y_{t_{i+1}-\tau} - Y_{t_i-\tau}$  einen direkten Einfluß auf  $X_{t_{i+1}} - X_{t_i}$ , das heißt ein Kopplung von  $Y$  nach  $X$  liegt vor. Ist hingegen  $T(X_{t_{i+1}} - X_{t_i} | (\mathbf{X}_{t_i} - \mathbf{X}_{t_{i-1}})^{(k)}, Y_{t_{i+1}-\tau} - Y_{t_i-\tau}) = 0$  für alle  $\tau \geq 0$ , so kann keine Kopplung nachgewiesen werden.

Während für den Nachweis der Abhängigkeit beider Prozesse voneinander die gegenseitige Information  $M(X_{t_i} - X_{t_{i-1}}, Y_{s_j} - Y_{s_{j-1}})$  auf einen 2-dimensionalen Zustandsraum geschätzt werden musste, muss für die Transferentropie mindestens ein 3-dimensionaler Zustandsraum berücksichtigt werden. Dies führt dazu, dass statistische Fluktuationen bei der Berechnung der Transferentropie sehr viel größer sein werden als bei der gegenseitigen Information. Somit werden für den Nachweis der Kopplung im Allgemeinen deutlich längere Zeitreihen benötigt als zum Nachweis der gegenseitigen Abhängigkeit.

**Beispiel: Gekoppelte Hindmarsh-Rose-Oszillatoren.** Im Folgenden wird an dem in Absch. 5.4.1 eingeführten Beispielsystem von gekoppelten Hindmarsh-Rose-Oszillatoren gezeigt, wie mit Hilfe der Transferentropie das System auf Kopplung untersucht werden kann, indem die Zuwächse betrachtet werden. Hierzu wurde dieses System unter Verwendung der selben Parameter wie in Absch. 5.4.1 für 1 000 000 Zeiteinheiten gelöst. Solch lange Zeitreihen waren diesmal notwendig, um die statistischen Fluktuationen beim Schätzen der Übergangsverteilungen auf einem ausreichend kleinen Niveau zu halten, da hier Zustandsräume mit bis zu 12 Dimensionen betrachtet werden. Weiterhin wurden nichtüberlappende, äquidistante Zeitfenster verwendet, das heißt  $t_i = s_i = \Delta i$ ,  $t_i - t_{i-1} = \Delta$  für alle  $i$ , wobei  $\Delta = 20$  gewählt wurde. Somit enthält jeder Zuwachs bis zu 2, vereinzelt auch 3 Ereignisse.

Zunächst wird für den Kopplungsparameter  $g = 0.3$  der Einfluss, den der Zuwachs  $X_{t_{i+1}-\tau} - X_{t_i-\tau}$ ,  $\tau \geq 0$ , auf  $Y_{t_{i+1}} - Y_{t_i}$  hat, untersucht. Hierzu wird für  $l = 1, \dots, 10$  die Transferentropie  $T(Y_{t_{i+1}} - Y_{t_i} | (\mathbf{Y}_{t_i} - \mathbf{Y}_{t_{i-1}})^{(l)}, X_{t_{i+1}-\tau} - X_{t_i-\tau})$  als Funktion der Zeitverzögerung  $\tau \geq 0$  berechnet. Aufgrund von Kausalität werden hier nur nichtnegative Zeitverzögerungen betrachtet. Wie in Abb. 5.14 zu sehen ist, zeigt die Transferentropie für  $l = 1$  eine temporale Struktur im Bereich  $\tau = 0, \dots, 250$ . Für große  $\tau$  fallen die Werte der Transferentropie innerhalb statistischer Fluktuationen auf einen kleinen positiven Wert ab. Da die Korrelationen innerhalb der Dynamik des Systems mit der Zeit abnehmen, stellt dieser Wert den positiven Bias des Schätzers dar.

Werden weitere Zuwächse aus der Vergangenheit von  $Y$  berücksichtigt, so nimmt die Ausprägung der temporalen Strukturen ab, da jetzt die indirekte Einflussnahme der Zuwächse  $X_{t_{i+1}-\tau} - X_{t_i-\tau}$  auf den Zuwachs  $Y_{t_{i+1}} - Y_{t_i}$  stärker unterbunden wird. Aufgrund des ebenfalls steigenden Biases ist eine direkte Ein-

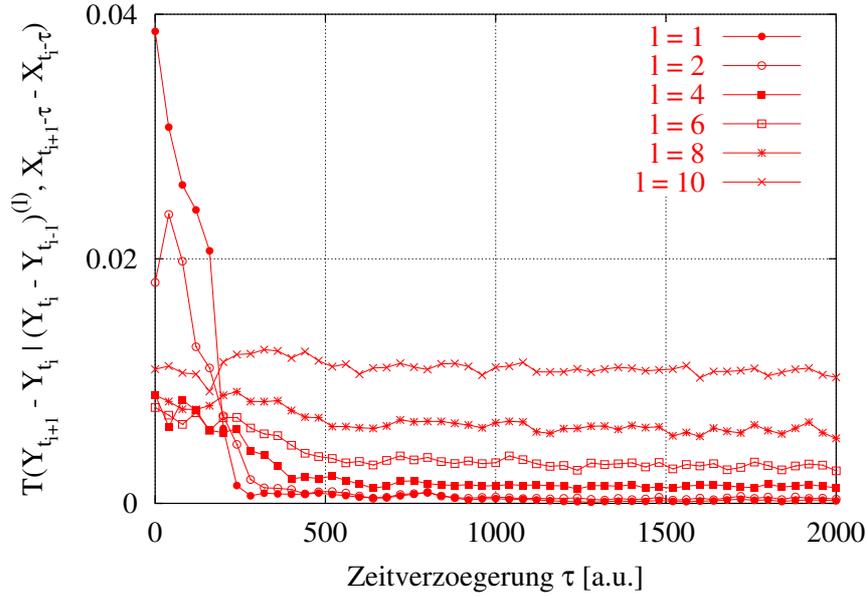


Abbildung 5.14: Transferentropie  $T(Y_{t_{i+1}} - Y_{t_i} | (\mathbf{Y}_{t_i} - \mathbf{Y}_{t_{i-1}})^{(l)}, X_{t_{i+1}-\tau} - X_{t_i-\tau})$  als Funktion der Zeitverzögerung  $\tau$  für  $l = 1, \dots, 10$  von unidirektional gekoppelte Hindmarsh-Rose-Oszillatoren mit Kopplungsparameter  $g = 0.3$ . Die Zeitpunkte  $t_i$  wurden äquidistant gewählt,  $t_{i+1} - t_i = \Delta$ , mit  $\Delta = 20$ .

flussnahme von  $X_{t_{i+1}-\tau} - X_{t_i-\tau}$  auf  $Y_{t_{i+1}} - Y_{t_i}$  für  $l = 10$  nicht mehr nachweisbar. Somit kann innerhalb dieser begrenzten Genauigkeit darauf geschlossen werden, dass  $X$  nicht in  $Y$  koppelt.

Als Nächstes wird die Kopplungsrichtung von  $Y$  nach  $X$  für  $g = 0.3$  untersucht. Hierfür wird der direkte Einfluss von  $Y_{t_{i+1}-\tau} - Y_{t_i-\tau}$  auf  $X_{t_{i+1}} - X_{t_i}$  für die Zeitverzögerungen  $\tau \geq 0$  studiert. Dies erfolgt, indem die Transferentropie  $T(X_{t_{i+1}} - X_{t_i} | (\mathbf{X}_{t_i} - \mathbf{X}_{t_{i-1}})^{(k)}, Y_{t_{i+1}-\tau} - Y_{t_i-\tau})$  als Funktion von  $\tau$  betrachtet wird. Mit dem Parameter  $k = 1, \dots, 10$  wird zusätzlich überprüft, wieviele Zuwächse von  $X$  in der Vergangenheit berücksichtigt werden müssen, um einen indirekten Einfluss von  $Y_{t_{i+1}-\tau} - Y_{t_i-\tau}$  über  $X_{t_{i-k}} - Y_{t_{i-k-1}}$  auf  $X_{t_{i+1}} - Y_{t_i}$  auszuschließen. Die Ergebnisse dieser Berechnungen sind in Abb. 5.15 dargestellt.

Hier zeigt sich zunächst, dass für alle  $k$  die Transferentropie temporale Strukturen im Bereich  $\tau < 500$  aufweist. Desweiteren liefert die Transferentropie in diesem Bereich Werte, die deutlich höher sind als der Bias, der für  $\tau > 1250$  vorliegt. Für etwa  $k > 6$  ändert sich innerhalb statistischer Fluktuationen die Transferentropie für Zeitverzögerungen von  $0 \leq \tau < 500$  qualitativ und quantitativ nicht. Dementsprechend ist es ausreichend  $k > 6$  zu wählen, um den unmittelbaren Einfluss von  $Y_{t_{i+1}-\tau} - Y_{t_i-\tau}$  auf  $X_{t_{i+1}} - X_{t_i}$  zu quantifizieren. Da für kleine  $\tau$  die Transferentropiewerte oberhalb des Biases liegen, kann hieraus

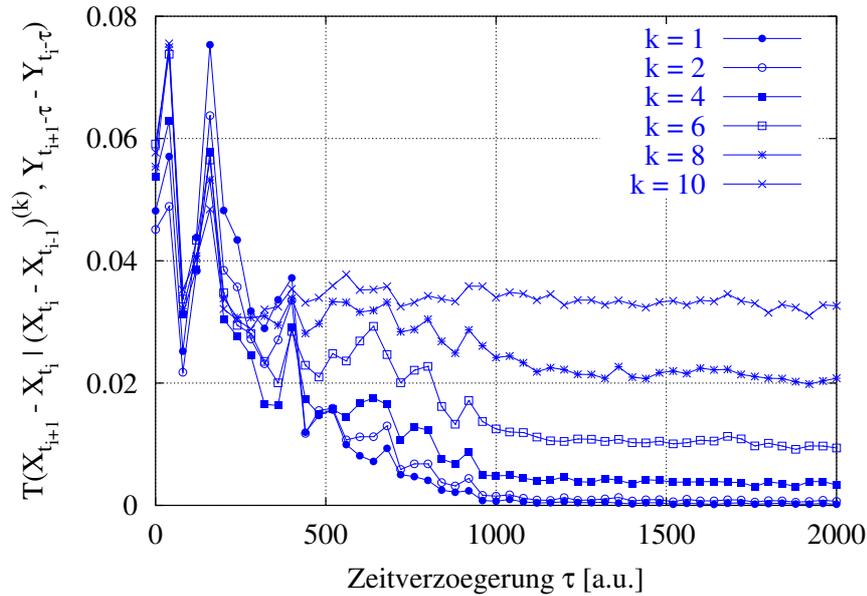


Abbildung 5.15: Transferentropie  $T(X_{t_{i+1}} - X_{t_i} | (\mathbf{X}_{t_i} - \mathbf{X}_{t_{i-1}})^{(k)}, Y_{t_{i+1}-\tau} - Y_{t_i-\tau})$  als Funktion der Zeitverzögerung  $\tau$  für  $k = 1, \dots, 10$  von unidirektional gekoppelte Hindmarsh-Rose-Oszillatoren mit Kopplungsparameter  $g = 0.3$ . Die Zeitpunkte  $t_i$  wurden äquidistant gewählt,  $t_{i+1} - t_i = \Delta$ , mit  $\Delta = 20$ .

die Kopplung von  $Y$  in  $X$  gefolgert werden, wobei der Zuwachs  $Y_{t_i} - Y_{t_{i-1}}$  den größten Einfluss auf  $X_{t_{i+1}} - X_{t_i}$  hat.

Die gleichen Untersuchungen wurden für die Kopplungsparameter  $0 \leq g \leq 0.5$  durchgeführt. Im gekoppelten Fall ( $g > 0.0$ ) spiegelte für  $k > 8$  und  $l = 10$  die Transferentropie stets die richtige Struktur des Modells wider. Im ungekoppelten Fall ( $g = 0.0$ ) lieferten die Transferentropien in Abhängigkeit von der Zeitverzögerung für alle  $k$  und  $l$  konstante Werte, die denen des Biases entsprechen, was auch hier die Modellstruktur richtig wiedergibt. In Abb. 5.16 sind für  $k = 10$  und  $l = 10$  die Transferentropien für beide Kopplungsrichtungen und alle Kopplungsstärken als Funktion der Zeitverzögerung aufgetragen. Die geschätzten Biaswerte der Transferentropien, welche unter anderem auch von der Kopplungsstärke  $g$  abhängen, sind zusätzlich als horizontale Linien eingetragen.

Mit der Transferentropie ist es zwar möglich, die Kopplungsrichtung zwischen zwei Punktprozessen zu identifizieren, störend ist aber der relativ hohe Bias des Schätzers. So beträgt dieser teilweise ein Drittel des Wertes von  $T(X_{t_{i+1}} - X_{t_i} | (\mathbf{X}_{t_i} - \mathbf{X}_{t_{i-1}})^{(10)}, Y_{t_{i+1}-\tau} - Y_{t_i-\tau})$  bei  $\tau = \Delta = 20$ , siehe Abb. 5.16. Der Bias der Transferentropie für die Richtung von  $Y$  nach  $X$ ,  $T(X_{t_{i+1}} - X_{t_i} | (\mathbf{X}_{t_i} - \mathbf{X}_{t_{i-1}})^{(10)}, Y_{t_{i+1}-\tau_\infty} - Y_{t_i-\tau_\infty})$  mit  $\tau_\infty = 2000$ , nimmt zwar mit der Länge der Zeitreihe ab, siehe Abb. 5.17, doch dies geschieht relativ langsam und die Ge-

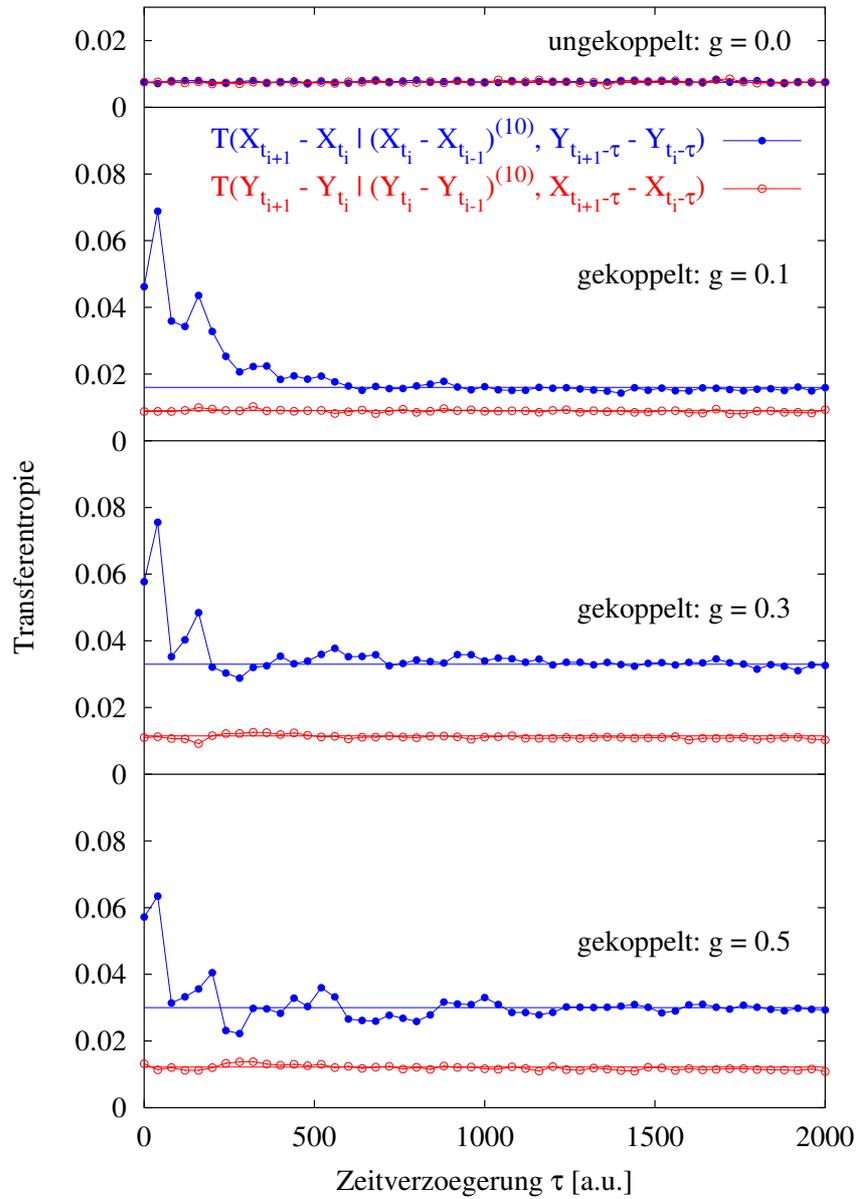


Abbildung 5.16: Transferentropie  $T(X_{t_{i+1}} - X_{t_i} | (\mathbf{X}_{t_i} - \mathbf{X}_{t_{i-1}})^{(10)}, Y_{t_{i+1}-\tau} - Y_{t_i-\tau})$  und  $T(Y_{t_{i+1}} - Y_{t_i} | (\mathbf{Y}_{t_i} - \mathbf{Y}_{t_{i-1}})^{(10)}, X_{t_{i+1}-\tau} - X_{t_i-\tau})$  als Funktion der Zeitverzögerung  $\tau$  für  $g = 0.0$  (ungekoppelt),  $g = 0.1, 0.3, 0.5$  (gekoppelt). Die Zeitpunkte  $t_i$  wurden äquidistant gewählt,  $t_{i+1} - t_i = \Delta$ , mit  $\Delta = 20$ . Horizontale Linien: geschätzter Bias der einzelnen Transferentropien.

schwindigkeit, mit der er abnimmt, hängt von der Kopplungsstärke  $g$ , also von der Dynamik des Prozesses  $X$  ab. Somit müssen Zeitreihen verwendet werden, deren Längen schätzungsweise um mehrere Größenordnungen länger sind als die

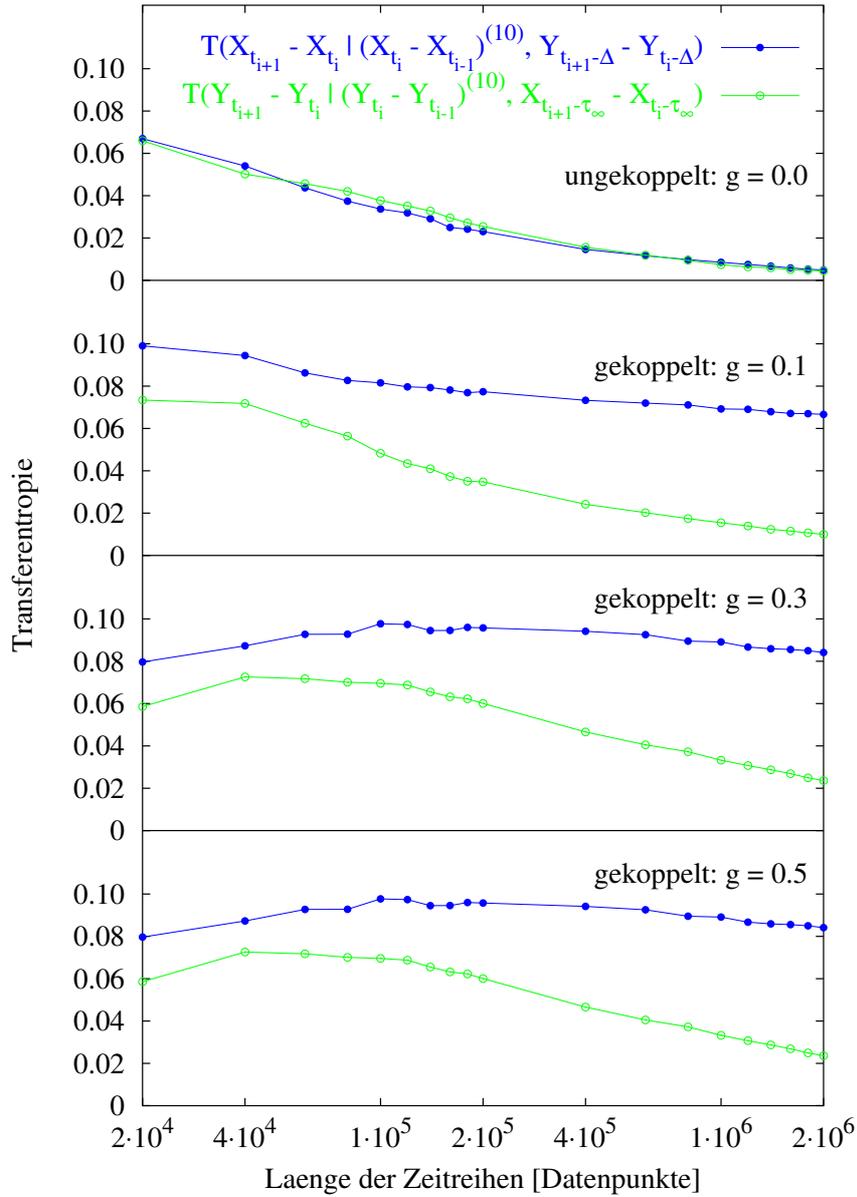


Abbildung 5.17: Transferentropie  $T(X_{t_{i+1}} - X_{t_i} | (\mathbf{X}_{t_i} - \mathbf{X}_{t_{i-1}})^{(10)}, Y_{t_{i+1}-\tau} - Y_{t_i-\tau})$  für  $\tau = \tau_\infty$  und  $\tau = \Delta$  in Abhängigkeit von der Länge der verwendeten Zeitreihen für verschiedene Kopplungsstärken  $g$ ,  $g = 0.0$  (ungekoppelt),  $g = 0.1, 0.3, 0.5$  (gekoppelt). Die Zeitpunkte  $t_i$  wurden äquidistant gewählt,  $t_{i+1} - t_i = \Delta$ , mit  $\Delta = 20$ . Für  $\tau = \tau_\infty = 2000$  gibt die Transferentropie den Bias des Schätzers an. Bei  $\tau = \Delta$  liegt die maximale Transferentropie vor.

hier benutzen, um den Bias auf einen Bruchteil des Wertes der Transferentropie bei  $\tau = \Delta$  zu drücken. Der Grund hierfür ist, dass die Transferentropie auf einem

12-dimensionalen Zustandsraum geschätzt werden muss.

Auch die Werte der Transferentropie für  $\tau = \Delta$ , die den stärksten Einfluss eines Zuwachses von  $Y$  auf einen Zuwachs von  $X$  angibt, sind biasbehaftet, siehe Abb. 5.17. Daher fallen sie ebenfalls mit der Länge der Zeitreihe ab, allerdings deutlich langsamer als die Transferentropie für  $\tau = \tau_\infty$ . Insbesondere klaffen beide Transferentropien für  $g > 0$  mit der Zeitreihenlänge immer weiter auseinander, da die Transferentropie für  $\tau = \Delta$  aufgrund der Kopplung gegen einen Wert größer als Null konvergieren sollte.

Eventuell kann der Bias nach unten korrigiert werden, indem die gleiche Korrektur wie bei den Kernschätzern, siehe Absch. 3.6, gemacht wird. Dies bedeutet bei der Berechnung der Transferentropie, Gl. (3.8), werden in der Summation nur solche Terme berücksichtigt, für die eine ausreichende Anzahl von Punkten zum Schätzen der Verteilungen zur Verfügung steht. Diese Korrektur konnte leider noch nicht implementiert werden.

Für die Kopplungsstärke  $0 \leq g \leq 0.5$  wurde die Analyse mit einer Fensterbreite von  $\Delta = 50$  wiederholt. Dies lieferte quantitativ die gleichen Ergebnisse wie bei  $\Delta = 50$ , wobei jetzt  $k \geq 4$  und  $l \geq 4$  zu wählen sind, damit die Transferentropie, angewendet auf Zuwächse, die Modellstruktur richtig wiedergibt.

## 5.6 Abhängigkeitsnachweis über Ereignisintervalle

Die bisher vorgestellten Methoden zum Nachweis von Abhängigkeiten zwischen Punktprozessen basieren auf der Untersuchung der Ereigniszahlen innerhalb eines Zeitintervalls, also auf den Zuwächsen von Punktprozessen. In diesem Abschnitt soll eine andere Herangehensweise vorgestellt werden, bei der der Nachweis von Abhängigkeiten auf Ereignisintervalle, gemeint sind Zeitintervalle zwischen den Ereignissen, basiert.

Wie bereits in Absch. 5.3 dargelegt wurde, hängt im Allgemeinen die Ankunftszeit  $T_{k+1}(\omega)$  des  $(k+1)$ -ten Ereignisses des Punktprozesses  $X$  von den Zeitpunkten seiner vorherigen Ereignisse  $T_k(\omega), T_{k-1}(\omega), \dots$  ab. Koppelt zusätzlich  $Y$  in  $X$ , so wird  $T_{k+1}(\omega)$  auch von den Ereignissen  $S_{j_{k+1}-1}(\omega), S_{j_{k+1}-2}(\omega), \dots$  von  $Y$  beeinflusst, die vor  $T_{k+1}(\omega)$  auftraten. Dabei kennzeichnet der Index  $j_{k+1}(\omega)$  jenes Ereignis  $S_l(\omega)$  von  $Y$ , das dem Ereignis  $T_{k+1}(\omega)$  folgt,  $j_{k+1}(\omega) = \min\{l \in \mathbb{N}_0 : S_l(\omega) \geq T_{k+1}(\omega)\}$ .

Wird der Zeitpunkt  $T_k(\omega)$  als Referenzzeit verwendet, so bedeutet dies, dass die Länge der Wartezeit  $D_{k+1}(\omega) = T_{k+1}(\omega) - T_k(\omega)$  von den *Intra-Intervallen*

$$D_k^{(i)}(\omega) = T_k(\omega) - T_{k-i}(\omega), \quad i = 1, 2, 3, \dots \quad (5.33)$$

und im gekoppelten Fall zusätzlich von den *Kreuz-Intervallen*

$$E_k^{(i)}(\omega) = S_{j_{k+i}}(\omega) - T_k(\omega), \quad i = 0, \pm 1, \pm 2, \dots \quad (5.34)$$

abhängt, siehe Abb. 5.18. Analog zu Gl. (5.13) folgt somit für die Verteilung der

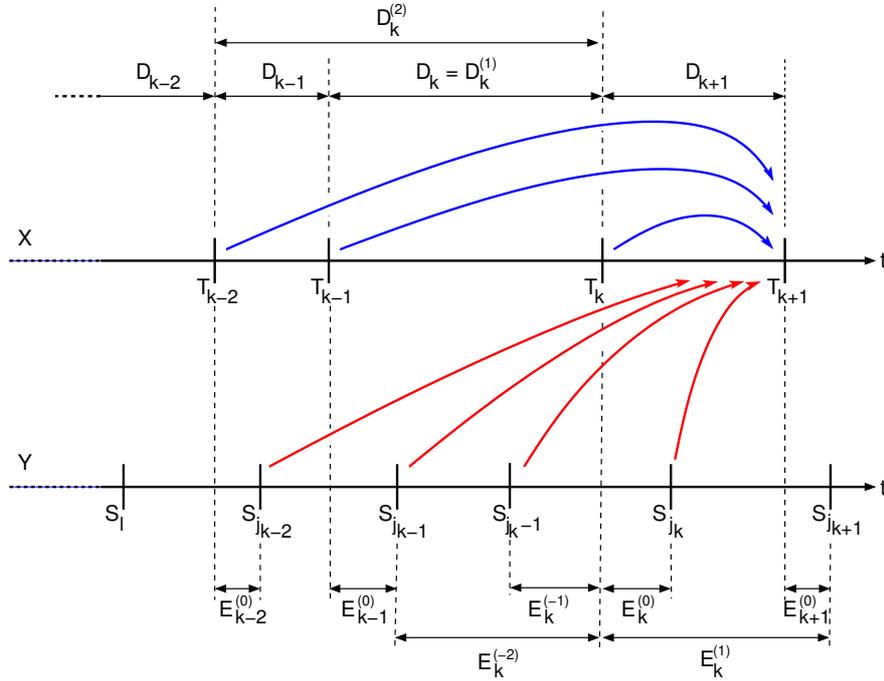


Abbildung 5.18: Definition der Ereignisintervalle  $D_k(\omega)$  und  $E_k^{(i)}$  bezüglich der Punktprozesse  $X$  und  $Y$ . Die Pfeile symbolisieren den Einfluss der vorherigen Ereignisse auf das nächste Ereignis, wobei  $Y$  in  $X$  koppelt.  $T_k(\omega)$  gibt die Ereigniszeiten von  $X$  und  $S_l(\omega)$  jene von  $Y$  an. Der Index  $j_{k+1}$  ist gegeben durch  $j_{k+1}(\omega) = \min\{l \in \mathbb{N}_0 : S_l(\omega) \geq T_{k+1}(\omega)\}$ .

Wartezeiten der funktionelle Zusammenhang

$$P\{D_{k+1} = d_{k+1}\} = \mathcal{F}[d_{k+1}, D_k^{(1)} = d_1, D_k^{(2)} = d_2, \dots, E_k^{(0)} = e_0, E_k^{(1)} = e_1, E_k^{(-1)} = e_{-1} \dots]. \quad (5.35)$$

Wie Gl. (5.33), bzw. Abb. 5.18 zu entnehmen ist, werden die Intra-Intervalle  $D_k^{(i)}$  nur aus den Ereigniszeiten des Prozesses  $X$  gebildet. Hierbei gibt die Intervallordnung  $i$  an, welches Ereignis vor  $T_k$  verwendet wird. Für die Kreuz-Intervalle (Gl. (5.34)) werden hingegen sowohl Ereigniszeiten von  $X$  als auch von  $Y$  benutzt. Insbesondere werden Ereignisse von  $Y$  verwendet, die vor (Intervallordnung  $i < 0$ ) und nach ( $i \geq 0$ ) dem Ereignis  $T_k$  auftreten. Aufgrund von Kausalität ist zu erwarten, dass es eine maximale Intervallordnung  $\hat{i} \geq 0$  gibt, so dass in Gl. (5.35) nicht alle Intervalle berücksichtigt werden müssen.

Entsprechend Absch. 5.3 folgt aus Gl. (5.35), dass die Prozesse  $X$  und  $Y$  nicht unabhängig voneinander sein können, wenn die Wartezeiten  $D_{k+1}$  und die

Kreuz-Intervalle  $E_k^{(i)}$  für irgendeine Intervallordnung  $i \in \mathbb{Z}$  nicht stochastisch unabhängig voneinander sind, das heißt wenn die gemeinsame Verteilung sich nicht als Produktverteilung schreiben lässt. Dementsprechend wird die Abhängigkeit von  $X$  und  $Y$  dadurch nachgewiesen, dass eine signifikante Abweichung von

$$P^{D_{k+1}, E_k^{(i)}} = P^{D_{k+1}} \otimes P^{E_k^{(i)}}, \quad i = 0, \pm 1, \pm 2, \dots \quad (5.36)$$

gesucht wird. Hierzu kann die gegenseitige Information  $M(D_{k+1}, E_k^{(i)})$  zwischen den Intervallen berechnet werden.

Da es sich bei  $D_{k+1}$  und  $E_k^{(i)}$  um reellwertige Zufallsvariablen handelt, erweist sich die Berechnung von  $M(D_{k+1}, E_k^{(i)})$  oft als problematisch, vor allem wenn nur wenige Daten gegeben sind. Diese Schwierigkeit kann mittels Diskretisierung überwunden werden. Sind  $D_{k+1}$  und  $E_k^{(i)}$  voneinander unabhängig, dann sind auch die aus ihnen hervorgehenden diskretisierten Zufallsvariablen  $\tilde{D}_{k+1}$  und  $\tilde{E}_k^{(i)}$  unabhängig. Des Weiteren wurde in Absch. 3.3, Gl. (3.16) gezeigt, dass die gegenseitige Information jeder Vergrößerung von stochastischen Prozessen eine untere Schranke für die gegenseitige Information der kontinuierlichen Prozesse ist, das heißt

$$M(\tilde{D}_{k+1}, \tilde{E}_k^{(i)}) \leq M(D_{k+1}, E_k^{(i)}).$$

Ist somit für eine Vergrößerung der Intervallprozesse die gemeinsame Information positiv, so hängen beide Punktprozesse voneinander ab.

Demzufolge bietet sich folgender Algorithmus zum Nachweis von Abhängigkeiten zwischen Punktprozessen an: Zuerst wird der Zustandsraum von  $D_{k+1}$  in Intervalle mit Länge  $\Delta_D$  unterteilt, wobei als Partitionierungsursprung ein beliebiges  $\tilde{d} \in \mathbb{R}_+$  gewählt wird. Analog wird der Zustandsraum von  $E_k^{(i)}$  partitioniert, wobei eine Intervalllänge  $\Delta_E$  und der Ursprung  $\tilde{e}$  verwendet werden. Die diskreten Zufallsvariablen, die aus dieser Partitionierung resultieren, lauten

$$\begin{aligned} \tilde{D}_{k+1}(\tilde{d}, \Delta_D) &= u, & \text{falls } \tilde{d} + u \Delta_D \leq D_{k+1} < \tilde{d} + (u+1) \Delta_D, \\ \tilde{E}_k^{(i)}(\tilde{e}, \Delta_E) &= v, & \text{falls } \tilde{e} + v \Delta_E \leq E_k^{(i)} < \tilde{e} + (v+1) \Delta_E. \end{aligned}$$

Als Nächstes wird die gegenseitige Information  $M(\tilde{D}_{k+1}(\tilde{d}, \Delta_D), \tilde{E}_k^{(i)}(\tilde{e}, \Delta_E))$  zwischen  $D_{k+1}(\tilde{d}, \Delta_D)$  und  $E_k^{(i)}(\tilde{e}, \Delta_E)$  berechnet. Um mehr als nur eine Vergrößerung zu betrachten, wird der Partitionsprozess mehrfach wiederholt, indem die Ursprünge  $\tilde{d}$  und  $\tilde{e}$  bei unveränderten  $\Delta_D$  und  $\Delta_E$  verschoben werden.

Bezüglich jeder dieser Partitionen wird die gegenseitige Information berechnet und anschließend das Supremum genommen,

$$B_{\Delta_D, \Delta_E}(\tilde{D}_{k+1}, \tilde{E}_k^{(i)}) = \sup_{\substack{0 \leq \tilde{d} < \Delta_D \\ 0 \leq \tilde{e} < \Delta_E}} M(\tilde{D}_{k+1}(\tilde{d}, \Delta_D), \tilde{E}_k^{(i)}(\tilde{e}, \Delta_E)). \quad (5.37)$$

Auch  $B_{\Delta_D, \Delta_E}(\tilde{D}_{k+1}, \tilde{E}_k^{(i)})$  stellt eine untere Grenze für die gegenseitige Information von  $D_{k+1}$  und  $E_k^{(i)}$  dar, das heißt

$$B_{\Delta_D, \Delta_E}(\tilde{D}_{k+1}, \tilde{E}_k^{(i)}) \leq M(D_{k+1}, E_k^{(i)}).$$

Folglich kann aus  $B_{\Delta_D, \Delta_E}(\tilde{D}_{k+1}, \tilde{E}_k^{(i)}) > 0$  geschlossen werden, dass die Punktprozesse  $X$  und  $Y$  voneinander abhängen.

Dieses Verfahren kann auch zur Analyse von Punktprozessen verwendet werden, wenn nur ein Experiment beobachtet wurde, sofern angenommen werden kann, dass die zugrunde liegende Dynamik der Prozesse sowie die Kopplung zeitunabhängig sind. In diesem Fall sind  $D_{k+1}$  bzw.  $E_k^{(i)}$ ,  $k = 0, 1, 2, \dots$ ,  $i = 0, \pm 1, \pm 2, \dots$  stationär verteilt.

Auch wenn bei dieser Methode  $D_{k+1}$  und  $E_k^{(i)}$  mittels  $\Delta_D$  bzw.  $\Delta_E$  diskretisiert werden, ist diese Diskretisierung nicht mit der Zeitdiskretisierung der zuvor besprochenen Methoden zu verwechseln. Insbesondere werden bei der Betrachtung von Zuwächsen häufig mehrere Ereignisse als Ereignisgruppe zusammengefasst. Hier sind  $\Delta_D$  und  $\Delta_E$  Parameter, die die Empfindlichkeit des Tests auf Abhängigkeit bestimmen. Je kleiner diese Intervalllängen gewählt werden, desto besser approximiert  $B_{\Delta_D, \Delta_E}(\tilde{D}_{k+1}, \tilde{E}_k^{(i)})$  die gegenseitige Information  $M(D_{k+1}, E_k^{(i)})$  asymptotisch, die der eigentliche Gegenstand der Untersuchung ist. Aufgrund von endlichen Beobachtungen kann  $\Delta_D$  und  $\Delta_E$  nicht beliebig klein gewählt werden, stattdessen tritt der "Finite sample"-Effekt deutlicher in den Vordergrund. In verschiedenen Untersuchungen hat sich herausgestellt, dass diese Intervalllängen auch nicht beliebig klein gewählt werden müssen, sondern dass eine binäre Partitionierung ausreichend ist, um Abhängigkeiten zu detektieren, das heißt  $\Delta_D = \Delta_E = \Delta$ , wobei  $\Delta$  die Länge des größten betrachteten Intervalls ist (Maximaler Durchmesser des Zustandsraums).

Selbst wenn der Test stets  $B_{\Delta_D, \Delta_E}(\tilde{D}_{k+1}, \tilde{E}_k^{(i)}) \approx 0$  für alle  $i = 0, \pm 1, \pm 2, \dots$  liefert, kann trotzdem nicht auf Unabhängigkeit geschlossen werden. So ist es auch möglich, dass der Test nicht in der Lage ist, eine Abhängigkeit zu erkennen. In diesem Fall sollten  $\Delta_D$  und  $\Delta_E$  verkleinert werden. Des Weiteren ist es empfehlenswert, die Ereigniszeiten von  $X$  und  $Y$  für die Festlegung der Ereignisintervalle in Gl. (5.33) und Gl. (5.34) miteinander zu vertauschen, was zu einem anderen, gleichwertigen Satz von Ereignisintervallen führt.

**Beispiel: Gekoppelte Hindmarsh-Rose-Oszillatoren.** Zum Abschluss soll dieses Verfahren zur Illustration auf die gekoppelten Hindmarsh-Rose-Oszillatoren, die im Absch. 5.4.1 eingeführt wurden, angewendet werden.

Das Supremum  $B_{\Delta, \Delta}(\tilde{D}_{k+1}, \tilde{E}_k^{(i)})$  der gegenseitigen Informationen bezüglich der binär diskretisierten Ereignisintervalle  $D_{k+1}$  und  $E_k^{(i)}$  ist als Funktion der Intervallordnung  $i$  in Abb. 5.19 für verschiedene Kopplungsstärken  $g$  dargestellt.  $\Delta$  ist hier der Durchmesser des Zustandsraums von  $(\tilde{D}_{k+1}, \tilde{E}_k^{(i)})$ . Wie Abb. 5.19 zu

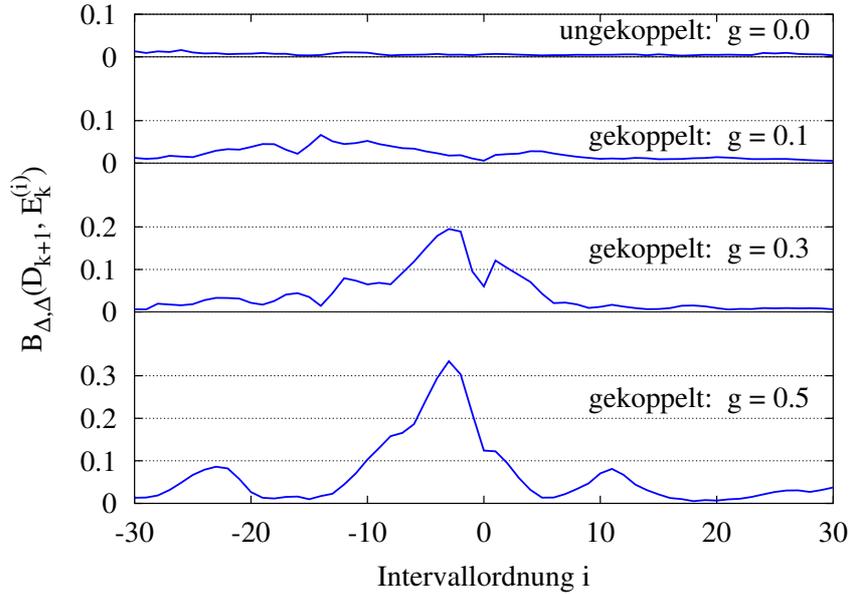


Abbildung 5.19: Supremum  $B_{\Delta, \Delta}(\tilde{D}_{k+1}, \tilde{E}_k^{(i)})$  über die gegenseitigen Informationen aller binär diskretisierten Ereignisintervalle  $D_{k+1}$  und  $E_k^{(i)}$  in Abhängigkeit der Intervallordnung  $i$ . Die Punktprozesse entstammen einem System von unidirektional gekoppelten Hindmarsh-Rose-Oszillatoren. Die Kopplungsstärke wurde mit dem Parameter  $g$  eingestellt.

entnehmen ist, zeigt der Kurvenverlauf von  $B_{\Delta, \Delta}(\tilde{D}_{k+1}, \tilde{E}_k^{(i)})$  über  $i$  eine Struktur im Bereich  $i \approx -28 \dots 15$ , insbesondere bildet sich ein markanter Peak um  $i = -3$  bei Kopplungsstärken  $g > 0.1$  heraus, dessen Höhe mit  $g$  zunimmt. Für große  $|i|$  sinken die Supremawerte auf näherungsweise Null ab.

Da die Korrelation zwischen Ereignissen von  $X$  und  $Y$  umso mehr abnimmt, je weiter sie auseinander liegen, verschwinden die Supremawerte für große  $|i|$  nahezu, wobei sie nicht negativ werden können. Wie in Absch. 5.4.1 bereits erwähnt wurde, kann dieser Bias im Mittel mit der ‘‘Finite sample’’-Korrektur herausgerechnet werden. Aus den gleichen Gründen, die bereits in Absch. 5.4.1 dargelegt wurden, wird auch hier auf eine Korrektur verzichtet, und stattdessen werden die Werte von  $B_{\Delta, \Delta}(\tilde{D}_{k+1}, \tilde{E}_k^{(i)})$  für kleine  $|i|$  mit jenen für große  $|i|$  verglichen, um auf Abhängigkeit schließen zu können. Die nichtkausale Abhängigkeit für  $i = 0 \dots 15$  ist ebenfalls auf den zugrunde liegenden Determinismus in der Dynamik der Oszillatoren zurückzuführen.

Für eine zusätzliche Verifizierung der in Abb. 5.19 beobachteten Strukturen wurden die Ereigniszeiten der beiden Punktprozesse  $X$  und  $Y$  gegeneinander in der Zeit verschoben, das heißt  $S_l(\omega) \rightarrow S_l(\omega) + \tau$ . Letztendlich werden hierdurch

nur die Ereignisintervalle  $E_k^{(i)}$  beeinflusst (siehe Gl. (5.33) und Gl. (5.34)),

$$E_k^{(i)} \rightarrow S_{j_k+i}(\omega) + \tau - T_k(\omega) \equiv E_{k,\tau}^{(i)}.$$

Die Ergebnisse von  $B_{\Delta,\Delta}(\tilde{D}_{k+1}, \tilde{E}_{k,\tau}^{(i)})$  als Funktion der Zeitverschiebung  $\tau$  und der Intervallordnung  $i$  sind in Abb. 5.20 für die Kopplung  $g = 0.3$  dargestellt. Für

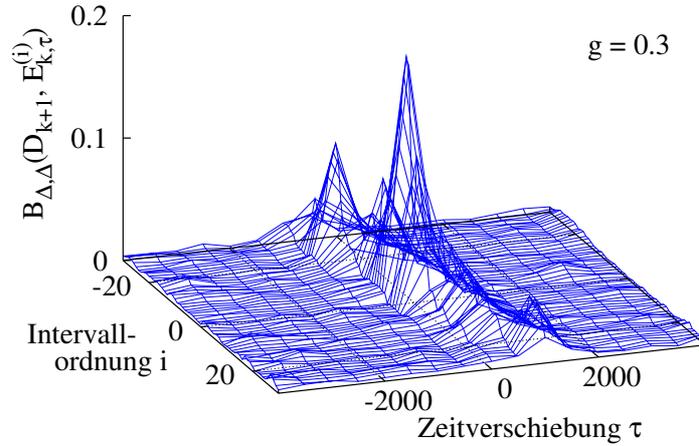


Abbildung 5.20: Supremum  $B_{\Delta,\Delta}(\tilde{D}_{k+1}, \tilde{E}_{k,\tau}^{(i)})$  in Abhängigkeit der Intervallordnung  $i$  und Zeitverschiebung  $\tau$ . Die Kopplungsstärke ist  $g = 0.3$ .

kleine Zeitverschiebungen  $|\tau|$  zeigt sich hier die Abhängigkeit beider Punktprozesse aufgrund der Kopplung deutlich. Aufgrund von Kausalität und Abnahme der Korrelation zwischen beiden Prozessen in der Zeit verschwindet die gegenseitige Information bei großem  $|\tau|$ .

Wie für die zuvor vorgestellten Methoden können auch hier kürzere Zeitreihen verwendet werden. Für eine Kopplung von  $g = 0.3$  zeigte sich, dass bei Zeitreihen mit einer Länge von nur noch 10 000 Punkten, was  $\approx 500$  Ereignissen entspricht, der funktionelle Verlauf des Supremums  $B_{\Delta,\Delta}(\tilde{D}_{k+1}, \tilde{E}_{k,\tau}^{(i)})$  qualitativ unverändert bleibt, siehe Abb. 5.21. Somit liefert diese Methode noch eine ausreichende Signifikanz, um Abhängigkeiten mit wenigen Ereignissen nachzuweisen.

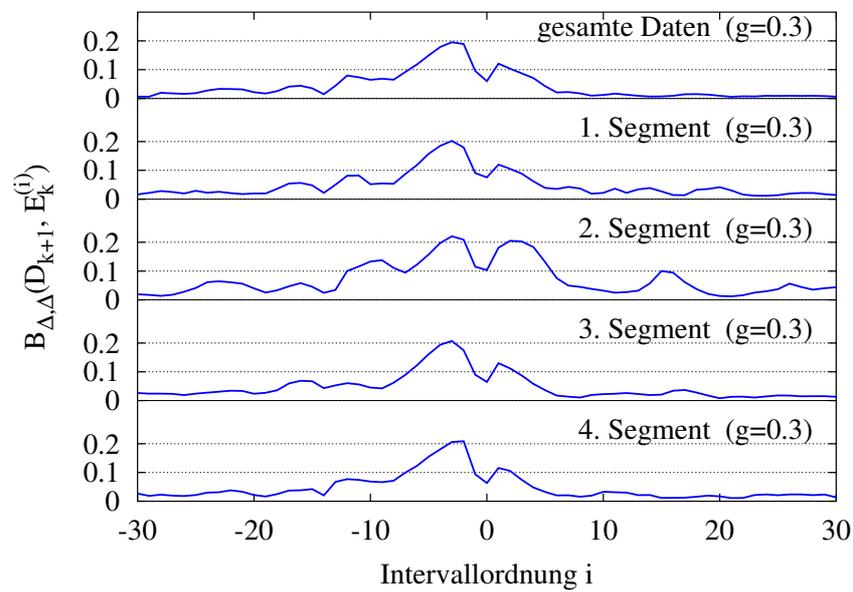


Abbildung 5.21: Supremum  $B_{\Delta, \Delta}(\tilde{D}_{k+1}, \tilde{E}_k^{(i)})$  in Abhängigkeit der Intervallordnung  $i$  für den gesamten Datensatz  $\approx 2500$  Ereignisse und verschiedene Segmente mit  $\approx 500$  Ereignissen. Kopplungsstärke  $g = 0.3$ .

# Kapitel 6

## Zusammenfassung

Die Bestimmung der Kopplungsrichtung und der Kopplungsstärke zwischen zwei beobachteten Prozessen aus Zeitreihen ist mit den unterschiedlichsten Schwierigkeiten verbunden. Handelt es sich um ein dynamisches System, so wird häufig die Ergodizität der Prozesse gefordert. Diese Voraussetzung kann dahingehend abgeschwächt werden, dass sich die System- oder Kopplungsparameter nur langsam in der Zeit ändern dürfen. So habe ich anhand von zwei gekoppelten logistischen Abbildungen demonstriert, dass mit der Interdependenz die Kopplungsrichtung auch in solch einem Fall richtig erkannt wird, wenn der Zustandsraum durch Übereinbettung rekonstruiert wird.

Die Untersuchung der Kopplung zwischen stochastischen Prozessen erfolgt in der Literatur vorwiegend dadurch, dass Abhängigkeiten, also zufällige Übereinstimmungen zwischen beiden Prozessen nachgewiesen bzw. quantifiziert werden. Hierzu wird oft die gegenseitige Information berechnet, welche die Informationsmenge angibt, die beide Prozesse gemeinsam haben. Zur Quantifizierung der Kopplung von Prozess  $Y$  nach  $X$  muss aber die Auswirkung, mit der die vergangenen Zustände von  $Y$  den zukünftigen Zustand von  $X$  direkt beeinflussen, gemessen werden. Eine geeignete Größe hierfür ist die Transferentropie. Ihre Werte können als Menge des Informationsflusses vom ersten Prozess  $Y$  in den zweiten Prozess  $X$  interpretiert werden. Dementsprechend ist die Transferentropie, ebenso wie die gegenseitige Information, stets nichtnegativ. Liegt keine Kopplung von  $Y$  nach  $X$  vor, so liefert die Transferentropie für diese Richtung immer Null, auch wenn eine Kopplung in der entgegengesetzten Richtung existiert. Somit kann sie zur Bestimmung der Kopplungsrichtung sowie zur Quantifizierung der Kopplungsstärke herangezogen werden. Hierin unterscheidet sie sich von der gegenseitigen Information, welche für unabhängige Prozesse stets Null ist, während sie im Falle einer Kopplung im Allgemeinen positiv ist. So kann, wie ich gezeigt habe, nur dann geschlossen werden, dass  $Y$  nicht in  $X$  koppelt, wenn alle vergangenen Zustände von  $Y$  von allen Zuständen von  $X$  einschließlich des zukünftigen unabhängig sind. Außerdem ist die gegenseitige Information symmetrisch bezüglich des Vertauschens der Prozesse, so dass mit ihr keine Aussage über die

Kopplungsrichtung gemacht werden kann.

Da die Transferentropie ursprünglich nur für stochastische Prozesse mit diskretem Zustandsraum definiert wurde, war es notwendig, sie für Prozesse mit kontinuierlichem Zustandsraum zu formulieren. Zu diesem Zweck habe ich die in der Transferentropie auftretenden Übergangswahrscheinlichkeiten durch deren Dichten bezüglich des Lebesgue-Maßes ersetzt. Hierdurch haben sich alle Eigenschaften der diskreten Transferentropie auf die kontinuierliche Transferentropie übertragen. Da hierfür die Existenz der Dichten gefordert wird, können nur solche kontinuierlichen Prozesse betrachtet werden, deren Verteilungen und Übergangsverteilungen eine Lebesgue-Dichte besitzen.

Damit die Quantifizierung der Abhängigkeit und der Kopplung physikalisch interpretierbare Werte liefert, wird gefordert, dass sich die Werte der gegenseitigen Information und der Transferentropie bei Koordinatentransformationen nicht ändern. Für diskrete Prozesse ist dies trivialerweise erfüllt. Wie ich zeigen konnte, bleiben beide informationstheoretischen Größen ebenfalls invariant gegenüber Koordinatentransformationen, wenn die Zustände beider Prozesse separat mit einem  $C^1$ -Diffeomorphismus transformiert werden. Im Besonderen ändern nichtlineare Reskalierungen oder Rotationen der gemessenen Daten bei mehrdimensionalen Prozessen die Werte der gegenseitigen Information und der Transferentropie nicht.

Möchte man die gegenseitige Information oder die Transferentropie aus Zeitreihen schätzen, so ist man mit einer Vielzahl von unterschiedlichen Schwierigkeiten konfrontiert. Die Verteilungen von diskreten Prozessen lassen sich relativ einfach schätzen, indem die Häufigkeiten, mit denen die Zustände beobachtet werden, gezählt werden. Werden bei zeitlich langsam driftenden Kopplungsparametern die Verteilungen auf kleinen Zeitfenstern geschätzt, so liefern gegenseitige Information und Transferentropie noch aussagekräftige Werte, so dass das zeitliche Verhalten von Abhängigkeit und Kopplung studiert werden kann.

Hingegen ist das Schätzen von Transferentropie und gegenseitiger Information für kontinuierliche Prozesse sehr viel schwieriger, da die Dichten in überabzählbar vielen Zuständen aus endlich vielen Beobachtungen geschätzt werden müssen. Bei der Partitionsmethode wird zunächst der Zustandsraum in nichtüberlappende Quader zerlegt. Anschließend werden die Wahrscheinlichkeiten betrachtet, mit denen die Prozesse in diesen Partitionselementen beobachtet werden. Dies entspricht einer Transformation der kontinuierlichen Prozesse in Symbolsequenzen, also in diskrete Prozesse. Sind die Dichten der Übergangsverteilungen stetig und existiert die kontinuierliche Transferentropie als Riemann-Integral, so konvergiert auf Partitionen mit immer kleiner werdenden Elementen die Transferentropie gegen die kontinuierliche Transferentropie. Ebenso konvergiert die gegenseitige Information auf einer Partition gegen die kontinuierliche gegenseitige Information, sofern die Dichten stetig sind. Handelt es sich bei der Partitionenfolge um Partitionen, deren Elemente in kleinere Quader zerlegt werden, so ist die Konvergenz der gegenseitigen Information sogar monoton steigend. Folglich kann eine Abhängigkeit

zwischen zwei Prozessen nachgewiesen werden, indem geprüft wird, ob die gegenseitige Information auf einer beliebigen Partition positiv ist. Eine analoge Aussage für die Transferentropie kann hingegen nicht gemacht werden. Insbesondere kann für bidirektional gekoppelte Systeme die Richtung des Nettoinformationsflusses nur dann bestimmt werden, wenn die Transferentropie konvergiert.

Besitzt man bereits im Voraus Kenntnisse über die Verteilungen der Prozesse, so müssen für parametrische Verteilungen lediglich deren Parameter aus den Daten geschätzt werden. Handelt es sich um Gauß-Prozesse, so lassen sich die gegenseitige Information und die Transferentropie allein aus der Kovarianzmatrix analytisch berechnen.

Als eine weitere Methode habe ich Kernschätzer vorgestellt, mit denen die informationstheoretischen Größen aus Daten berechnet werden können. Hierbei handelt es sich um nichtparametrische Schätzer für die Dichten. Gegenüber der Partitionierung des Zustandsraums hat dieses Verfahren den Vorteil, dass der Bias des Schätzers aufgrund serieller Korrelationen innerhalb der Zeitreihe mit einem Theiler-Fenster auf einfache Weise unterdrückt werden kann. Die größte Schwierigkeit bei der Verwendung von Kernschätzern liegt in der Konstruktion eines effektiven Kerns für mehr-dimensionale Prozesse. Anhand von gekoppelten autoregressiven Prozessen wurde dennoch demonstriert, dass es trotz eines einfachen Produktansatzes für den Kern möglich ist, die Kopplungsrichtung sowie die Kopplungsstärke zu bestimmen. Des Weiteren konnten die Abhängigkeiten näherungsweise richtig quantifiziert werden.

Die Untersuchung der Abhängigkeiten zwischen stochastischen Prozessen stellt insbesondere dann eine große Herausforderung dar, wenn die Prozesse nicht-stationär sind und nur eine Realisierung vorliegt. Anhand von Zeitreihen, in denen die Windgeschwindigkeiten an verschiedenen Orten gegeben sind, wurde demonstriert, wie die gegenseitige Information angewendet werden kann, um räumliche Abhängigkeiten zu untersuchen. Hierzu wurde die gegenseitige Information auf kleinen Zeitfenstern geschätzt. Indem dies Zeitfenster über die Zeitreihen verschoben wurde, war es möglich, die Abhängigkeiten zwischen den Windgeschwindigkeiten an verschiedenen Orten als Funktion der Zeit zu ermitteln. Insbesondere ist es möglich, die Windrichtung aus den Windgeschwindigkeiten zu bestimmen. Hierzu wurden die Zeitreihen gegeneinander verschoben und die gegenseitige Information als Funktion der Zeitverschiebung untersucht.

Zu einer speziellen Klasse von stochastischen Prozessen gehören die Punktprozesse. Bei ihnen ist nicht der Zustand, sondern der Zeitpunkt, in dem ein Ereignis eintritt, von Interesse. Die Realisierungen von Punktprozessen sind Funktionen  $N_t$ , deren Werte die Anzahl der Ereignisse bis zum beobachteten Zeitpunkt  $t$  angeben. Diese Ereigniszahlen sind zwar ganzzahlige Funktionen, aber da sie monoton wachsen, muss eine große Anzahl von Beobachtungen vorhanden sein, damit die Verteilungen der Punktprozesse geschätzt werden können, um Abhängigkeiten oder Kopplungen zwischen zwei Punktprozessen nachzuweisen. Kann aber angenommen werden, dass dem Punktprozess ein ergodischer Prozess zugrunde

liegt, so dass die Ereignisse jene Zeitpunkte markieren, in denen ein Schwellenwert in der Dynamik überschritten wird, so können die Verteilungen der Zuwächse  $N_t - N_s$  aus nur einer Zeitreihe geschätzt werden. Der Zuwachs  $N_t - N_s$  gibt dabei die Anzahl der Ereignisse an, die innerhalb des Zeitfensters  $(s, t]$  stattgefunden haben. Da, wie ich gezeigt habe, zwei Punktprozesse nicht voneinander unabhängig sein können, wenn deren Zuwächse voneinander abhängen, müssen lediglich die Verteilungen der Zuwächse geschätzt werden, um Abhängigkeiten nachzuweisen. An einem Beispielsystem, bestehend aus gekoppelten Hindmarch-Rose-Oszillatoren, wurde demonstriert, dass es durch Betrachten der Zuwächse möglich ist, Abhängigkeiten zwischen zwei Punktprozessen mit Hilfe der gegenseitigen Information nachzuweisen und zu quantifizieren. Hierzu reichten sogar relativ kurze Zeitreihen aus, die ungefähr 500 Ereignisse enthielten.

Häufig werden Koinzidenzen zwischen Punktprozessen gesucht, um auf Abhängigkeiten schließen zu können. Dabei versteht man unter Koinzidenzen das nahezu gleichzeitige Auftreten von Ereignissen in beiden Prozessen. Wie ich gezeigt habe, lässt sich die mittlere Anzahl von Koinzidenzen auf die Momente der Zuwächse zurückführen. Insbesondere kann sie durch die Kovarianz der Zuwächse quantifiziert werden.

Damit die Stärke der Kopplung sowie die Kopplungsrichtung zwischen Punktprozessen aus nur einer Realisierung bestimmt werden können, muss angenommen werden, dass sich alle dynamischen Eigenschaften der Prozesse in den Zuwächsen widerspiegeln. Dies wurde an dem oben genannten Beispielsystem illustriert. Da hierfür aufgrund der Markov-Eigenschaft der Prozesse mehrere Zuwächse in der Vergangenheit bei der Berechnung der Transferentropie berücksichtigt werden müssen, sind relativ lange Zeitreihen nötig, um Aussagen über die Kopplung und somit über die Struktur des Systems treffen zu können.

Der Nachweis von Abhängigkeiten durch Zuwächse erfolgt auf der Basis von Ereignisgruppen, welche durch die Fensterbreite der Zuwächse bestimmt werden. Um den Nachweis auf der Ebene einzelner Ereignisse machen zu können, muss daher die Fensterbreite sehr klein gewählt werden, was wiederum auf Kosten der Signifikanz geht. Aus diesem Grund wurde ein neues Verfahren entwickelt, bei dem zwei Sequenzen von Ereignisintervallen zwischen den Ereignissen beider Prozesse definiert wurden, und zwar so, dass beide Sequenzen mit nur einem Index geordnet werden können. Anschließend wurden diese Ereignisintervalle mit der gegenseitigen Information auf Abhängigkeit hin untersucht. An dem Beispielsystem zeigte sich ebenfalls, dass dieses Verfahren auch bei kürzeren Zeitreihen, die nur noch bis zu 500 Ereignisse enthalten, robuste Aussagen lieferte und kaum an Signifikanz verlor.

# Anhang A

## Mathematische Werkzeuge

**Definition 1** Eine Funktion  $f$  heißt *konvex* auf einem offenem Intervall  $(a, b)$ , wenn für beliebige  $x_1, x_2 \in (a, b)$  und  $0 \leq \lambda \leq 1$

$$f(\lambda x_1 + (1 - \lambda)x_2) \leq \lambda f(x_1) + (1 - \lambda)f(x_2) \quad (\text{A.1})$$

gilt. Sie heißt *streng convex*, wenn die Gleichheit nur für  $\lambda = 0$  oder  $\lambda = 1$  erfüllt ist. Eine Funktion  $f$  heißt *konkav*, wenn  $-f$  convex ist.

**Anmerkung:** Ist  $f$  streng convex und  $f(\lambda x_1 + (1 - \lambda)x_2) = \lambda f(x_1) + (1 - \lambda)f(x_2)$  für  $0 < \lambda < 1$  so ist  $x_1 = x_2$ .

**Satz 1 (Jensensche Ungleichung für diskrete Zufallsvariablen)** Ist  $f$  eine convexe Funktion und  $X$  eine Zufallsvariable mit diskreten Werten, dann gilt

$$\mathbb{E}[f(X)] \geq f(\mathbb{E}[X]). \quad (\text{A.2})$$

Hier ist  $\mathbb{E}[f(X)] = \sum_{i=1}^m p_i f(x_i)$ ,  $p_i = P\{X = x_i\}$  der Erwartungswert von  $f(X)$ . Ist  $f$  eine streng convexe Funktion, dann folgt aus der Gleichheit, dass  $X = \mathbb{E}[X]$  fast sicher ist, das heißt,  $X$  ist eine Konstante.

**Beweis:** Der Beweis erfolgt durch vollständiger Induktion. Für eine Zweimassenverteilung  $\{p_1, p_2\}$  folgt aus Gl. (A.1) unmittelbar

$$\begin{aligned} f(p_1 x_1 + p_2 x_2) &= f(p_1 x_1 + (1 - p_1)x_2) \\ &\leq p_1 f(x_1) + (1 - p_1)f(x_2) = p_1 f(x_1) + p_2 f(x_2), \end{aligned}$$

also die Behauptung.

Angenommen, die Behauptung sei für eine beliebige  $k$ -Massenverteilung  $\{p'_1, \dots, p'_k\}$  erfüllt und  $\{p_1, \dots, p_k, p_{k+1}\}$  sei eine Verteilung mit  $k + 1$  Massen.

Aus ihr kann eine  $k$ -Massenverteilung konstruiert werden, indem  $p'_i = p_i/(1-p_{k+1})$  für  $i = 1, \dots, k$  gesetzt wird. Dann folgt,

$$\begin{aligned}
\sum_{i=1}^{k+1} p_i f(x_i) &= p_{k+1} f(x_{k+1}) + (1 - p_{k+1}) \sum_{i=1}^k p'_i f(x_i) \\
&\geq p_{k+1} f(x_{k+1}) + (1 - p_{k+1}) f\left(\sum_{i=1}^k p'_i x_i\right) \\
&\geq f\left(p_{k+1} x_{k+1} + (1 - p_{k+1}) \sum_{i=1}^k p'_i x_i\right) \\
&= f\left(\sum_{i=1}^{k+1} p_i x_i\right). \tag{A.3}
\end{aligned}$$

Hierbei folgt die erste Ungleichung aus der Induktionshypothese und die zweite aus der Voraussetzung, dass  $f$  convex ist.

Ist insbesondere  $f$  eine streng convexe Funktion, dann ist (A.3) eine echte Ungleichung, und damit  $E[f(X)] > f(E[X])$ . Hieraus folgt, dass die Gleichheit in Gl. (A.2) dann und nur dann gilt, wenn  $X$  fast sicher eine Konstante ist.  $\square$

**Satz 2 (Log-Summen-Ungleichung)** Für beliebige nichtnegative Zahlen  $a_1, \dots, a_n$  und  $b_1, \dots, b_n$  gilt

$$\sum_{i=1}^n a_i \log \frac{a_i}{b_i} \geq \left( \sum_{i=1}^n a_i \right) \log \frac{\sum_{i=1}^n a_i}{\sum_{i=1}^n b_i}. \tag{A.4}$$

Die Gleichheit ist dann und nur dann erfüllt, wenn  $a_1/b_1 = \dots = a_n/b_n$  ist.

**Beweis:** Unter Einführung der Konventionen  $0 \log 0 = 0$ ,  $a \log(a/0) = \infty$  falls  $a > 0$  und  $0 \log(0/0) = 0$  kann ohne Beschränkung der Allgemeinheit angenommen werden, dass  $a_i > 0$  und  $b_i > 0$  für  $i = 1, \dots, n$  ist.

Die Funktion  $f(t) = t \log t$ ,  $t > 0$  ist streng convex, da  $f'(t) = \log(t) + 1$  und  $f''(t) = 1/t > 0$  für alle positiven  $t$  ist. Somit folgt aus der Jensenschen Ungleichung [Bauer (1991), Billingsley (1995), Cover & Thomas (1991)] für die diskreten Wahrscheinlichkeiten  $p_i \geq 0$ ,  $\sum_{i=1}^n p_i = 1$

$$\sum_{i=1}^n p_i f(t_i) \geq f\left(\sum_{i=1}^n p_i t_i\right).$$

Die Gleichheit gilt genau dann, wenn alle  $t_i$  gleich sind. Wird nun  $p_i = b_i / \sum_{i=1}^n b_i$  und  $t_i = a_i/b_i$  gesetzt, so folgt

$$\sum_{i=1}^n \frac{a_i}{\sum_{i=1}^n b_i} \log \frac{a_i}{b_i} \geq \sum_{i=1}^n \frac{a_i}{\sum_{i=1}^n b_i} \log \sum_{i=1}^n \frac{a_i}{\sum_{i=1}^n b_i}.$$

Dies ist die Log-Summen-Ungleichung.  $\square$

**Satz 3 (Ein-dimensionaler Kernschätzer)** *Ist die Wahrscheinlichkeitsdichte  $g$  einer ein-dimensionalen Zufallsvariablen  $X$  zweimal differenzierbar und gilt für den Kern  $K$*

$$\int K(u) du = 1, \quad \int u K(u) du = 0, \quad \int u^2 K(u) du < \infty,$$

so ist der Kernschätzer der Dichte  $g$

$$\hat{g}_\varepsilon(x) = \frac{1}{N\varepsilon} \sum_{n=1}^N K\left(\frac{x - X(\omega_n)}{\varepsilon}\right)$$

erwartungstreu, dass heißt

$$\mathbb{E}[g_\varepsilon(x)] \xrightarrow{\varepsilon \rightarrow 0} g(x).$$

**Beweis:** Für den Erwartungswert von  $g_\varepsilon(x)$  gilt:

$$\begin{aligned} \mathbb{E}[g_\varepsilon(x)] &= \int \frac{1}{N\varepsilon} \sum_{n=1}^N K\left(\frac{x - X(\omega_n)}{\varepsilon}\right) P(d\omega_1, \dots, d\omega_N) \\ &= \frac{1}{N\varepsilon} \sum_{n=1}^N \int K\left(\frac{x - X(\omega_n)}{\varepsilon}\right) P(d\omega_n) \\ &= \frac{1}{\varepsilon} \int K\left(\frac{x - X(\omega_1)}{\varepsilon}\right) P(d\omega_1) = \frac{1}{\varepsilon} \int K\left(\frac{x - y}{\varepsilon}\right) g(y) dy \\ &= \int K(u) g(x - \varepsilon u) du. \end{aligned}$$

Da  $g$  zweimal differenzierbar ist, kann die Dichte in eine Taylorsche Reihe um  $x$  entwickelt werden,

$$g(x - \varepsilon u) = g(x) - \frac{\partial g(x)}{\partial x} \varepsilon u + \frac{1}{2} \frac{\partial^2 g(\tilde{x})}{\partial x^2} \varepsilon^2 u^2,$$

mit  $\tilde{x} \in [x - \varepsilon u, x]$ . Dies eingesetzt in die vorherige Gleichung und unter Berücksichtigung der Eigenschaften von  $K$  liefert die Behauptung

$$\mathbb{E}[g_\varepsilon(x)] = g(x) + \frac{\varepsilon^2}{2} \cdot \text{Rest} \xrightarrow{\varepsilon \rightarrow 0} g(x).$$

$\square$

**Satz 4 (*d*-dimensionaler Kernschätzer)** *Ist die Wahrscheinlichkeitsdichte  $g$  einer  $d$ -dimensionalen Zufallsvariablen  $X = (X_1, \dots, X_d)$  zweimal differenzierbar und gilt für den Kern  $K$*

$$\begin{aligned} \int K(u_1, \dots, u_d) du_1 \dots du_d &= 1, \\ \int u_j K(u_1, \dots, u_d) du_1 \dots du_d &= 0, & j = 1, \dots, d, \\ \int u_i u_j K(u_1, \dots, u_d) du_1 \dots du_d &< \infty, & i, j = 1, \dots, d, \end{aligned}$$

so ist der Kernschätzer der Dichte  $g$

$$\hat{g}_\varepsilon(x_1, \dots, x_d) = \frac{1}{N \varepsilon_1 \dots \varepsilon_d} \sum_{n=1}^N K\left(\frac{x_1 - X_1(\omega_n)}{\varepsilon_1}, \dots, \frac{x_d - X_d(\omega_n)}{\varepsilon_d}\right)$$

erwartungstreu, dass heißt

$$\mathbb{E}[\hat{g}_\varepsilon(x_1, \dots, x_d)] \xrightarrow{\varepsilon_1, \dots, \varepsilon_d \rightarrow 0} g(x_1, \dots, x_d).$$

**Beweis:** Eine analoge Rechnung wie im Satz 3 liefert die Behauptung.  $\square$

# Literaturverzeichnis

- [Arnhold et al. (1999)] J. Arnhold, P. Grassberger, K. Lehnertz, C.E. Elger, “A robust method for detecting interdependences: application to intracranially recorded EEG”, *Physica D* **134**, 419-430 (1999).
- [Bauer (1991)] H. Bauer, *Wahrscheinlichkeitstheorie*, 4.Aufl., Walter de Gruyter, Berlin (1991).
- [Bauer (1992)] H. Bauer, *Maß- und Integrationstheorie*, 2.Aufl., Walter de Gruyter, Berlin (1992).
- [Behnen & Neuhaus (1995)] K. Behnen & G. Neuhaus, *Grundkurs Stochastik*, 3.Aufl., Teubner, Stuttgart (1995).
- [Billingsley (1965)] P. Billingsley, *Ergodic Theory and Information*, Wiley series on the science and technology of materials, Wiley, New York (1965).
- [Billingsley (1995)] P. Billingsley, *Probability and Measure*, Wiley series in probability and mathematical statistics, 3.Aufl., Wiley, New York (1995).
- [Brandt & Pompe (1993)] C. Brandt & B. Pompe, “The Entropy Profile — A Function Describing Statistical Dependences”, *J. Stat. Phys.* **70**, 967-983 (1993).
- [Cohen & Procaccia (1985)] A. Cohen & I. Procaccia, “Computing the Kolmogorov entropy from time signals of dissipative and conservative dynamical systems”, *Phys. Rev. A* **31**, 1872-1882 (1985).
- [Cover & Thomas (1991)] T.M. Cover & J.A. Thomas, *Elements of Information Theory*, Wiley, New York (1991).
- [Daley & Vere-Jones (1972)] D.J. Daley & D. Vere-Jones, “A summary of the theory of point processes” in *Stochastic Point Processes: Statistical Analysis, Theory, and Applications*, ed. v. P.A.W. Lewis, Wiley, New York, 299-383 (1972).

- [Darbellay (1999)] G.A. Darbellay, “An estimator of the mutual information based on a criterion for conditional independence”, *Comp. Stat. & Data Anal.* **32**, 1-17 (1999).
- [Eckmann & Ruelle (1985)] J.-P. Eckmann & D. Ruelle, “Ergodic theory of chaos and strange attractors”, *Rev. Mod. Phys.* **57**, 617-656 (1985).
- [Eguia et al. (2000)] M.C. Eguia, M.I. Rabinovich, H.D.I. Abarbanel, “Information transmission and recovery in neural communications channels”, *Phys. Rev. E* **2000**, 7111-7122 (2000).
- [Fraser & Swinney (1986)] A. M. Fraser & H. L. Swinney, “Independent coordinates for strange attractors from mutual information”, *Phys. Rev. A* **33**, 1134-1140 (1986).
- [Gaspard & Wang (1993)] P. Gaspard & X.-J. Wang, “Noise, chaos, and  $(\varepsilon, \tau)$ -entropy per unit time” *Phys. Reports* **235**, 291-343 (1993).
- [Grassberger & Procaccia (1983a)] P. Grassberger & I. Procaccia, “Characterization of Strange Attractors”, *Phys. Rev. Lett.* **50**, 346-349 (1983).
- [Grassberger & Procaccia (1983b)] P. Grassberger & I. Procaccia, “Measuring the strangeness of strange attractors”, *Physica D* **9**, 189-208 (1983).
- [Grassberger & Procaccia (1983c)] P. Grassberger & I. Procaccia, “Estimating the Kolmogorov entropy from a chaotic signal”, *Phys. Rev. A* **28**, 2591-2593 (1983).
- [Grassberger (1985)] P. Grassberger, “Generalizations of the Hausdorff dimension of fractal measures” *Phys. Lett. A* **107**, 101-104 (1985).
- [Grassberger (1988)] P. Grassberger, “Finite sample corrections to entropy and dimension estimates”, *Phys. Lett. A* **128**, 369-373 (1988).
- [Grün et al. (1999)] S. Grün & M. Diesmann & F. Grammont & A. Riehle & A. Aertsen, “Detecting unitary events without discretization of time”, *J. Neurosci. Meth.* **93**, 67-79 (1999).
- [Grün et al. (2002)] S. Grün & M. Diesmann & A. Aertsen, “Unitary Events in Multiple Single-Neuron Spiking Activity: I. Detection and Significance”, *Neural Computation* **14**, 43-80 (2002).
- [Hegger & Kantz (1997)] R. Hegger & H. Kantz, “Embedding of sequences of time intervals”, *Europhys. Lett.* **38**, 267-272 (1997).
- [Hegger et al. (2000)] R. Hegger & H. Kantz & L. Matassini & T. Schreiber, “Coping with Nonstationarity by Overembedding”, *Phys. Rev. Lett.* **84**, 4092-4095 (2000).

- [Hindmarsh & Rose (1984)] J.L. Hindmarsh & R.M. Rose, "A model of neuronal bursting using three coupled first order differential equations", Proc. R. Soc. London, Ser. B **221**, 87-102 (1984).
- [Huerta et al. (1998)] R. Huerta, M. Bazhenov & M.I. Rabinovich, "Clusters of synchronization and bistability in lattices of chaotic neurons", Europhys. Lett **43**, 719-724 (1998).
- [Jumarie (1990)] G. Jumarie, *Relative information: theories and applications*, Springer series in synergetics, Vol. 47, Springer, Berlin (1990).
- [Kaiser & Schreiber (2002)] A. Kaiser & T. Schreiber, "Information transfer in continuous processes", Physica D **166**, 43-62 (2002).
- [Kaneko (1986)] K. Kaneko, "Lyapunov analysis and information flow in coupled map lattices", Physica D **23**, 436-447 (1986).
- [Kantz & Schreiber (1997)] H. Kantz & T. Schreiber, *Nonlinear time series analysis*, Cambridge University Press, Cambridge MA (1997).
- [Katok & Hasselblatt (1995)] A. Katok & B. Hasselblatt, *Introduction to the modern theory of dynamical systems*, Cambridge University Press, Cambridge (1995).
- [Kolmogorov (1993)] A.N. Kolmogorov, *Information theory and the theory of algorithms, selected works*, Vol. 3, Kluwer, Dordrecht (1993).
- [Kullback (1959)] S. Kullback, *Information theory and statistics*, Wiley, New York (1959).
- [Lewis (1972)] P.A.W. Lewis (Ed.), *Stochastic Point Processes: Statistical Analysis, Theory, and Applications*, Wiley, New York (1972).
- [Moddemeijer (1999)] R. Moddemeijer, "A statistic to estimate the variance of the histogram-based mutual information estimator based on dependent pairs of observations", Sign. Proc. **75**, 51-63 (1999).
- [Olbrich et al. (2000)] E. Olbrich & R. Hegger & H. Kantz, "Local Estimates for Entropy Densities in Coupled Map Lattices", Phys. Rev. Lett. **84**, 2132-2135 (2000).
- [Ott (1993)] E. Ott, *Chaos in Dynamical Systems*, Cambridge University Press, Cambridge (1993).
- [Panzeri & Treves (1996)] S. Panzeri & A. Treves, "Analytical estimates of limited sampling biases in different information measures", Comput. Neural Syst. **7**, 87-107 (1996).

- [Palus (1996a)] M. Palus, “Detecting nonlinearity in multivariate time series”, *Phys. Lett. A* **213**, 138-147 (1996).
- [Palus (1996b)] M. Palus, “Coarse-grained entropy rates for characterization of complex time series”, *Physica D* **93**, 64-77 (1996).
- [Pikovsky et al.(2001)] A.S. Pikovsky & M.G. Rosenblum & J. Kurths, *Synchronization*, Cambridge University Press, Cambridge (2001).
- [Pinto et al. (2000)] R.D. Pinto & P. Varona & A.R. Volkovskii & A. Szücs & H.D.I. Abarbanel & M.I. Rabinovich, “Synchronous behavior of two coupled electronic neurons”, *Phys. Rev. E* **62**, 2644-2656 (2000).
- [Pollicott & Yuri (1998)] M. Pollicott & M. Yuri, *Dynamical Systems and Ergodic Theory*, Cambridge University Press, Cambridge (1998).
- [Pompe (1993)] B. Pompe, “Measuring Statistical Dependences in Time Series”, *J. Stat. Phys.* **73**, 587-610 (1993).
- [Porta et. al. (1999)] A. Porta, G. Baselli, F. Lombardi, N. Montano, A. Malliani, S. Cerutti, “Conditional entropy approach for the evaluation of the coupling strength”, *Biol. Cybern.* **81**, 119-129 (1999).
- [Prichard & Theiler (1995)] D. Prichard & J. Theiler, “Generalized redundancies for time series analysis”, *Physica D* **84**, 476-493 (1995).
- [Quian Quiroga et al. (2000)] R. Quian Quiroga & J. Arnhold & P. Grassberger, “Learning driver-response relationships from synchronization patterns”, *Phys. Rev. E* **61**, 5142-5148 (2000).
- [Le Van Quyen et al. (1999)] M. Le Van Quyen & J. Martinerie & C. Adam & F.J. Varela, “Nonlinear analyses of interictal EEG map the brain interdependences in human focal epilepsy”, *Physica D* **127**, 250-266 (1999).
- [Reitmann (1996)] V. Reitmann, *Reguläre und chaotische Dynamik*, B.G. Teubner, Stuttgart (1996).
- [Riehle et al. (1997)] A. Riehle & S. Grün & M. Diesmann & A. Aertsen, “Spike Synchronization and rate modulation differentially involved in motor cortical function”, *Science* **278**, 1950-1953 (1997).
- [La Rosa et al. (2000)] M. La Rosa, M.I. Rabinovich, R. Huerta, H.D.I. Abarbanel, L. Fortuna, “Slow regularization through chaotic oscillation transfer in an unidirectional chain of Hindmarsh-Rose models”, *Phys. Lett. A* **266**, 88-93 (2000).

- [Rosenblum et al. (1996)] M.G. Rosenblum & A.S. Pikovsky & J. Kurths, “Phase synchronization of chaotic oscillators”, *Phys. Rev. Lett.* **76**, 1804-1807 (1996).
- [Rosenblum & Pikovsky (2001)] M.G. Rosenblum & A.S. Pikovsky, “Detecting direction of coupling in interacting oscillators”, *Phys. Rev. E* **64**, 045202(R) (2001).
- [Roulston (1997)] M.S. Roulston, “Significance testing of information theoretic functionals”, *Physica D* **110**, 62-66 (1997).
- [Roulston (1999)] M.S. Roulston, “Estimating the errors on measured entropy and mutual information”, *Physica D* **125**, 285-294 (1999).
- [Ruelle (1989)] D. Ruelle, *Chaotic Evolution and Strange Attractors*, Cambridge University Press, New York (1989).
- [Rulkov et al. (1995)] N.F. Rulkov & M.M. Sushchik & L.S. Tsimring & H.D.I. Abarbanel, “Generalized synchronization of chaos in directionally coupled chaotic systems”, *Phys. Rev. E* **51**, 980-994 (1995).
- [Sauer et al. (1991)] T. Sauer & J.A. Yorke & M. Casdagli, “Embedology”, *J. Stat. Phys.* **65**, 579-616 (1991).
- [Schiff et al. (1996)] S.J. Schiff & P. So & T. Chang & R.E. Burke & T. Sauer, “Detecting dynamical interdependence and generalized synchrony through mutual prediction in neural ensemble”, *Phys. Rev. E* **54**, 6708-6724 (1996).
- [Schreiber (1990)] T. Schreiber, “Spatio-temporal structure in coupled map lattices: two-point correlation versus mutual information” *J. Phys. A* **23**, L393 (1990).
- [Schreiber & Schmitz (2000a)] T. Schreiber & A. Schmitz, “Surrogate time series”, *Physica D* **142**, 346-382 (2000).
- [Schreiber (2000b)] T. Schreiber, “Measuring information transfer”, *Phys. Rev. Lett.* **85**, 461-464 (2000).
- [Schuster (1989)] H.G. Schuster, *Deterministic Chaos An Introduction*, VCH Verlagsgesellschaft, Weinheim (1989).
- [Shannon & Weaver (1949)] C.E. Shannon, W. Weaver, *The mathematical theory of information*, University of Illinois Press, Urbana IL (1949).
- [Silverman (1986)] B.W. Silverman, *Density Estimation for Statistics and Data Analysis*, Monographs on statistics and applied probability Vol 26, Chapman and Hall, London (1986).

- [Takens (1981)] F. Takens, *Detecting Strange Attractors in Turbulence*, Lecture Notes in Mathematics Vol. 898, pp 366-381 ed. v. A. Rand & L.-S. Young, Springer, New York (1981).
- [Theiler (1986)] J. Theiler, "Spurious dimension from correlation algorithms applied to limited time-series data", *Phys. Rev. A* **34**, 2427-2432 (1992).
- [Theiler (1990)] J. Theiler, "Estimating fractal dimension", *J. Opt. Soc. Am. A* **7**, 1055-1073 (1990)
- [Theiler et al. (1992)] J. Theiler & S. Eubank & A. Longtin & B. Galdrikian & J.D. Farmer, "Testing for nonlinearity in time series: the method of surrogate data", *Physica D* **58**, 77-94 (1992).
- [URL (2002)] <http://www.winddata.com>.
- [Vastano & Swinney (1988)] J.A. Vastano, H.L. Swinney, "Information transport in spatiotemporal systems", *Phys. Rev. Lett.* **60**, 1773-1776 (1988).
- [Walters (1981)] P. Walters, *An Introduction to Ergodic Theory*, Springer, New York (1981).

## Danksagung

Ich möchte mich bei allen Personen bedanken, die mich bei der Erstellung dieser Doktorarbeit unterstützt haben. Mein besonderer Dank gilt Priv. Doz. Holger Kantz und Priv. Doz. Thomas Schreiber für die Betreuung der Arbeit sowie für anregende und kritische Diskussionen. Weiterhin bedanke ich mich bei Prof. Dr. Ramón Huerta für die Einführung auf dem Gebiet der Neurowissenschaften und Modellierung von Neuronen am *Institute for Nonlinear Science, University of California, San Diego*. Bei Hannah Köllmer und Dr. Anja Riegert möchte ich mich für ihre Geduld und Ausdauer beim Korrektur lesen bedanken. Für seine Bereitschaft, das Zweitgutachten zu übernehmen, danke ich Herrn Prof. Dr. Karl-Heinz Mütter. Zum Abschluss möchte ich mich noch bei allen Personen bedanken, die mir beim Lösen von Computerproblemen oder durch die Bereitstellung von freiverfügbaren Softwareprodukten geholfen haben.