Rauschreduktion in nichtlinearen Systemen

Diplomarbeit

vorgelegt von

Jochen Bröcker

aus

Kiel

angefertigt am

Dritten Physikalischen Institut

der Georg–August–Universität zu Göttingen

1999

Komm, Valerio, wir müssen was treiben! Wir wollen uns mit tiefen Gedanken abgeben; wir wollen untersuchen, wie es kommt, daß der Stuhl auf drei Beinen steht und nicht auf zweien. Komm, wir wollen Ameisen zergliedern, Staubfäden zählen!

Leonce

1.Auflage
vom Autor korrigierte fünfte Fassung
Göttingen, den 18. März 1999

Inhaltsverzeichnis

Einleitung

Jede Form von Signalübertragung, sei es zur Kommunikation oder bei der Aufnahme von Daten in physikalischen Experimenten, ist potentiell dem Einfluß von Störungen ausgesetzt. Diese Störungen können vielfältigster Natur sein, gemeinsam ist ihnen im allgemeinen nur, daß sie unerwünscht sind, und zwar besonders dann, wenn sie Teile der Information des Signals zerstören. Man denke sich zum Beispiel einen Informationskanal gegeben, über den mit Wahrscheinlichkeit $1/2$ eine Null und mit Wahrscheinlichkeit $1/2$ eine Eins übertragen wird. Auf halber Strecke in diesem Kanal sitzt ein Dämon, der mit Wahrscheinlichkeit $1/2$ die Nachricht verändert (aus der Eins macht er eine Null und umgekehrt), mit Wahrscheinlichkeit $1/2$ aber die Nachricht unverändert läßt. Es ist anschaulich klar und läßt sich mit wahrscheinlichkeitstheoretischen Methoden verifizieren, daß in diesem (zugegebenermaßen extremen) Beispiel die empfangene und die gesendete Nachricht nichts mehr miteinander zu tun haben. Offensichtlich geht keine Information verloren, wenn der Dämon *stets* eine Vertauschung vornimmt (sofern das Verhalten des Dämons bekannt ist). Gefährlich sind also solche Störungen, deren Verhalten nicht genau bekannt ist. Diese Störungen bezeichnet man als zufällig. Andere Störungen, deren Verhalten bekannt ist oder sich identifizieren läßt, bezeichnet man als systematisch. Es ist nicht unbedingt sinnlos, Störungen, deren Verhalten man vollständig klären könnte, trotzdem als zufällig anzusehen, da die Identifikation ihrer Ursachen oft schwieriger ist als ihre Bekämpfung mit wahrscheinlichkeitstheoretischen Methoden.

In dieser Arbeit geht es um eine gewisse Klasse von Signalen, nämlich solche, die durch eine iterierte Abbildung zustande kommen. Besondere Aufmerksamkeit wird dabei nichtlinearen Abbildungen gewidmet, die im allgemeinen größere Probleme aufwerfen als lineare. Ein Grund ist, daß nichtlineare Abbildungen ein ähnlich irreguläres Verhalten zeigen können wie zufällige Signale. Das Verhalten iterierter Abbildungen ist natürlich theoretisch immer vorhersagbar, in der Praxis jedoch zeigt sich, daß viele Systeme extrem sensitiv auf kleine Störungen des Zustandes reagieren. Aufgrund der endlichen Rechengenauigkeit werden diese Systeme, für die man allgemein den Begriff *chaotische Systeme* benutzt, praktisch unvorhersagbares Verhalten zeigen.

Im ersten Kapitel dieser Arbeit werden erst einmal die wahrscheinlichkeitstheoretischen Methoden vorgestellt, die in dieser Arbeit benötigt werden. Die

Wahrscheinlichkeitstheorie gestattet die Behandlung von Größen, deren Verhalten nur in Form von Verteilungen bekannt ist. Dieser Formalismus dient dann zur Beschreibung der Störeinflüsse auf das eigentlich gewünschte Signal.

Im zweiten Kapitel geht es um die Systeme, welche die Signale erzeugen, die eigentlich gemessen werden sollen. Das sind, wie gesagt, nichtlineare iterierte Abbildungen, die auch *dynamische Systeme* genannt werden. Zur Beschreibung ihres Verhaltens werden die notwendigen Begriffe eingeführt. Davon wird später nur ein Teil benötigt. Um jedoch ein allgemeines Bild dieser Abbildungen zu zeichnen werden auch zusätzliche Dinge angeführt, die von allgemeinem Interesse sind und deshalb meines Erachtens hierher gehören.

Im dritten und vierten Kapitel werden die Methoden zur Reduktion der Störeinflüsse bzw. des *Rauschens*, wie im folgenden immer gesagt werden wird, entwickelt und in exemplarischen Beispielen zur Anwendung gebracht. Dabei geht es im dritten Kapitel um *Filterung*. Damit ist hier gemeint, daß man als Information zur Rekonstruktion des momentanen Systemzustandes im wesentlichen die Beobachtungen des gestörten Signals aus der Vergangenheit zur Verfügung hat. Im vierten Kapitel geht es dann um den Fall, daß man das gestörte Signal in seiner vollen Länge zur Verfügung hat. Die Aufgabe ist dann die Rekonstruktion des Ursprungssignals. Es wird dort auch kurz darauf eingegangen, was zu tun ist, wenn das dynamische System nicht oder nicht vollständig bekannt ist.

In dieser Arbeit können weder alle Aspekte dieses Themenkreises noch alle technischen Einzelheiten angesprochen werden. Ziel der Arbeit ist es vielmehr, die bestehenden Verfahren in ein allgemeines Konzept einzuordnen. Dieses Konzept besteht in der Anwendung einfacher statistischer und wahrscheinlichkeitstheoretischer Werkzeuge. Daraus ergeben sich auch einige weitere Verfahren, die getestet werden.

Es sei an dieser Stelle darauf hingewiesen, daß in dieser Arbeit *ausschließlich* dynamische Systeme in diskreter Zeit besprochen werden. Einige wenige Worte zu dynamischen Systemen und Reduktion von Rauschen in kontinuierlicher Zeit sind im Anhang A zu finden.

Einige technische Hinweise

In dieser Arbeit werden Signal–Rauschverhältnisse (SNR *signal–to–noise–ratio*) in dB angegeben, und zwar auf folgende Weise: Das Rauschen sei $r = \{r_n\}_{n=1,\ldots,N}$, das Signal sei $s = \{s_n\}_{n=1,\ldots,N}$. Diese Größen seien mittelwertfrei. Dann ist

$$\mathrm{SNR}(r+s) := 10 \cdot \log_{10}\left(\frac{\sigma(s)}{\sigma(r)}\right)$$

mit

$$\sigma(s) := \sqrt{\frac{1}{N}\sum_{i=1}^{N} \|s_n\|^2} \qquad \sigma(r) := \sqrt{\frac{1}{N}\sum_{i=1}^{N} \|r_n\|^2}$$

Mit der *dB–Verbesserung* eines Rauschreduktionsalgorithmus ist die Differenz der SNR vor und nach der Anwendung des Algorithmus auf eine Zeitreihe gemeint.

Zeichen und Symbole

$p(x\|y)$	Wahrscheinlichkeit von x gegeben y
$p(x,y)$	Wahrscheinlichkeit von x und y
Ω	Wahrscheinlichkeitsraum
	Menge der nichtwandernden Punkte
M^m	differenzierbare Mannigfaltigkeit der Dimension m
$\mathrm{E}(X)$	Erwartungswert von X
$\mathrm{Cov}(X)$	Kovarianz von X
x_k^n	$= \{x_i\}_{i=k,\ldots,n}$
x^n	$= \{x_i\}_{i=0,\ldots,n}$
$x^{(\alpha)}$	α–te Komponente des Vektors x
Γ^t	Transponierte von Γ
$x^t\Gamma x$	$= \sum_{ij} \Gamma^{(ij)} x^{(i)} x^{(j)}$
$\mathcal{D}p(x)$	$= \int p_F(x,y)p(y)\,\mathrm{d}y$ (Vorwärtsoperator)
$\mathcal{R}p(x)$	$= \int p_F(y,x)p(y)\,\mathrm{d}y$ (Rückwärtsoperator)
$\log$	Logarithmus zur Basis e
$\log_{10}$	Logarithmus zur Basis 10

1 Stochastik, Statistik und die Bayes–Strategie

Es soll zuerst eine kurze Übersicht über die in dieser Arbeit verwendeten Methoden der Stochastik und der Statistik gegeben werden. Dies ist erstens notwendig, um die allgemeine Aufgabenstellung dieser Arbeit adäquat formulieren zu können und zweitens praktisch, um wesentliche Formeln zitierfähig zu halten. Allgemein sei hier auf [4] verwiesen. Zuerst werden die wichtigsten Konzepte der Stochastik vorgestellt, dann werden einige wichtige Formeln zur GAUSS–Verteilung gebracht. Zum Schluß wird auf Schätzungen und die BAYES–Formel eingegangen.

1.1 Etwas Theorie

Zu diesem Abschnitt sei allgemein auf das Buch von BREIMAN, [4] verwiesen. In der Stochastik betrachtet man *Wahrscheinlichkeitsräume* oder kurz *W–Räume*. Das sind Punktmengen Ω mit etwas zusätzlicher Struktur, nämlich einer σ–Algebra Σ und einem Maß. Eine σ–Algebra ist ein System von Teilmengen von Ω, die man auch Ereignisse nennt. Für die σ–Algebra fordert man spezielle Eigenschaften:

$$
\begin{aligned}
\Omega &\in \Sigma, \\
A \in \Sigma &\Rightarrow A^c \in \Sigma, \\
A,\, B \in \Sigma &\Rightarrow A \cup B \in \Sigma, \\
\{A_i\}_{i\in\mathbb{N}} \in \Sigma &\Rightarrow \bigcup_{i\in\mathbb{N}} A_i \in \Sigma.
\end{aligned}
$$

Hier bezeichnet A^c das Komplement von A. Die erste Eigenschaft besagt, das Ω selbst ein Ereignis ist. Die zweite Eigenschaft besagt, daß wenn A ein Ereignis ist, das auftreten kann, dann ist es auch ein Ereignis, daß A nicht auftritt. Die dritte Eigenschaft besagt, daß wenn A und B Ereignisse sind, so ist auch „A oder B“ ein Ereignis. Die ersten drei Eigenschaften wird man also natürlicherweise von Ereignissen fordern, die vierte ist eine Verallgemeinerung der dritten, die aus technischen Gründen äußerst wichtig ist. Hat man ein System von Mengen, genannt $\mathcal{M}$, welches noch keine σ–Algebra ist, so kann man die kleinste σ–Algebra betrachten, die $\mathcal{M}$ enthält. Diese bezeichnet man als die von $\mathcal{M}$ *erzeugte* σ–Algebra und schreibt $\sigma(\mathcal{M})$.[1]

Ein Maß μ ist eine reelle Funktion auf Σ mit den folgenden Eigenschaften:

$$\begin{aligned} \mu(\Omega) &= 1, \\ \mu(A) &\geq 0, \\ A \cap B = \emptyset &\Rightarrow \mu(A \cap B) = \mu(A) + \mu(B), \\ A_i \cap A_j = \emptyset \, \forall i,j &\Rightarrow \mu(\bigcup_{i \in \mathbb{N}} A_i) = \sum_{i \in \mathbb{N}} \mu(A_i). \end{aligned}$$

Das Maß nennt man auch die *Wahrscheinlichkeit* und schreibt dann $P(A)$ statt $\mu(A)$. Die erste Eigenschaft sagt nur, daß mit Sicherheit irgendein Ereignis eintritt. In anderen Maßräumen (z.B. im $\mathbb{R}^n$) läßt man diese Eigenschaft weg. Das Maß kann dann auch unendlich sein. Maßräume mit endlichem Maß führt man immer auf den Fall mit $\mu(\Omega) = 1$ zurück. In dieser Arbeit sei immer ein Raum mit endlichem Maß gegeben. Die nächsten beiden Eigenschaften sind anschaulich für eine Wahrscheinlichkeit einleuchtend. Die vierte Eigenschaft ist wiederum technischer Natur; sie ist eine Verallgemeinerung der dritten.

Interessant wird die Maßtheorie, wenn es um die Konstruktion von Maßen geht. Hierbei steht man oft vor folgendem Problem: Man hat ein System von Mengen gegeben, die man schon messen kann, für die man ein kanonisches Volumen hat. Z.B. die Rechtecke auf dem $\mathbb{R}^n$ und ihre endlichen Vereinigungen und Schnitte. Ihnen kann man ein kanonisches Volumen zuordnen. Diese Mengen bilden jedoch keine σ–Algebra. Im Hinblick auf das nächste Kapitel sollen dazu folgende Definitionen gebracht werden:

Definition 1.1 *Ein* äußeres Maß μ^* *ist eine Funktion, die jeder Teilmenge* $A \subset \Omega$ *eine positive Zahl zuordnet mit den Eigenschaften:*

1. $\mu^*(\emptyset) = 0$,
2. $\mu^*(\bigcup_{i=1}^{\infty} A_i) \leq \sum_{i=1}^{\infty} \mu^*(A_i)$.

Ist darüber hinaus Ω *ein metrischer Raum, so heißt* μ^* *ein* metrisches *äußeres Maß, wenn für zwei Mengen* A, B *mit dist*$(A,B) > 0$ *gilt:* $\mu^*(A \cup B) = \mu^*(A) + \mu^*(B)$.

Eine Menge $A \subset \Omega$ *heißt* μ^*–meßbar, *wenn für jede beliebige Menge* $Z \subset \Omega$ *gilt:*

$$\mu^*(Z) = \mu^*(Z \cap A) + \mu^*(Z \setminus A).$$

Der Witz ist nun, daß man aus einem äußeren Maß, welches oft leicht zu konstruieren ist, immer ein richtiges Maß bekommt. Die zugehörige σ–Algebra wird gerade von den μ^*–meßbaren Mengen gebildet. Genauer gilt:

[1] Dieses $\sigma(\mathcal{M})$ ist eindeutig definiert. Enthalten, zwei σ–Algebren $\mathcal{M}$, so auch ihr Schnitt. Der Schnitt ist jedoch wieder eine σ–Algebra.

Satz 1.2 (Lemma von Carathéodory) *Für ein äußeres Maß μ^* bilden die μ^*–meßbaren Mengen eine σ–Algebra und das äußere Maß ist darauf ein Maß. Ist darüber hinaus μ^* ein metrisches äußeres Maß, so sind offene Mengen stets μ^*–meßbar.*

Hierauf werden wir im zweiten Kapitel zurückkommen.

Tatsächlich sind aber nicht nur die Wahrscheinlichkeitsräume wichtig, sondern auch Funktionen zwischen Ω und $\mathbb{R}$. Eine solche Funktion heißt *Zufallsvariable*, wenn sie meßbar ist, d.h. Urbilder meßbarer Mengen sind ebenfalls meßbar.[2] Ist X eine integrierbare Zufallsvariable, dann nennt man

$$\int_\Omega X(\omega)\ \mathrm{d}P(\omega) =: \mathrm{E}(X)$$

den *Erwartungswert* von X.

Die Funktion $F(x) := P(\omega; X(\omega) < x)$ nennt man die *Verteilung* von X. Weiter wird durch $\tilde{P}(B) := P(\omega; X(\omega) \in B)$ mit meßbarem $B \subset \mathbb{R}$ ein Maß auf $\mathbb{R}$ definiert, welches eindeutig durch Kenntnis der Verteilungsfunktion definiert ist. Unter gewissen Voraussetzungen besitzt eine Verteilungsfunktion eine *Dichte*. Das ist eine Funktion $p(x)$ mit

$$F(x) = \int_{-\infty}^{x} p(y)\ \mathrm{d}y.$$

Es gilt ferner die Formel:

$$\mathrm{E}(X) := \int_\mathbb{R} x\ \mathrm{d}\tilde{P}(x) = \int_\mathbb{R} x\ \mathrm{d}F(x) = \int_\mathbb{R} x\, p(x)\, \mathrm{d}x.$$

Der Erwartungswert ist also ebenfalls durch Kenntnis der Verteilungsfunktion festgelegt. Da praktisch alle anderen stochastisch relevanten Größen ebenfalls durch die Verteilung bestimmt sind, erhebt sich die Frage, wozu überhaupt Ω eingeführt werden muß. In der Praxis begnügt man sich ja mit Aussagen wie: „X ist eine Gauß–verteilte Zufallsgröße" und hat obendrein meist sogar noch eine Dichte. Tatsächlich ist der W–Raum nicht eindeutig bestimmt. Und wenn man eine einzige Zufallsvariable betrachtet und deren Verteilung kennt, kann man sich immer $\Omega = \mathbb{R}$, $P(A) = \int_A \mathrm{d}F(x)$ als gegeben denken. Ω wird erst wichtig, wenn man mehrere Zufallsvariable betrachtet, unter Umständen unendlich viele mit verschiedenen Wertebereichen. Aus den Einzelverteilungen geht dann offensichtlich nichts mehr über die wechselseitigen Beziehungen der Variablen hervor. Wichtig ist noch ein Begriff:

Die Zufallsvariablen $X_1, ..., X_n$ heißen *unabhängig*, wenn gilt

$$P(X_1 \in B_1, ..., X_n \in B_n) = \prod_{i=1}^{n} P(X_i \in B_i).$$

[2] Oder anders gesagt, Urbilder von Ereignissen sind wieder Ereignisse. Hier und im folgenden sei als σ–Algebra auf $\mathbb{R}$ immer die Borel–Algebra angenommen.

Natürlich kann $X_1, ..., X_n$ auch als eine Zufallsvariable mit Werten im $\mathbb{R}^n$ aufgefaßt werden. Die Verteilungsfunktion und ggf. die Dichte sind dann Funktionen auf $\mathbb{R}^n$. Der Erwartungswert ist dann einfach komponentenweise zu bilden.

1.2 Erzeugende Funktionen und die Gauß–Verteilung

Als ein *Moment k–ter Ordnung* oder einfach *k–tes Moment* der Zufallsvariablen $X := X_1, \ldots, X_n$ oder auch als ein *k–tes Moment* ihrer gemeinsamen Verteilung bezeichnet man das Integral

$$M_{i_1,\ldots,i_k} := \mathrm{E}(X_{i_1} \cdot \ldots \cdot X_{i_k}) = \int_{\mathbb{R}^n} X_{i_1} \cdot \ldots \cdot X_{i_k} \, \mathrm{d}P(X_{i_1}, \ldots, X_{i_k})$$

falls es konvergiert. Der Erwartungswert besteht also aus den Momenten erster Ordnung. Die Funktion $\eta : \mathbb{R}^n \to \mathbb{R}$ mit

$$\eta(y) := \mathrm{E}(e^{\sum y_i X_i}) = \mathrm{E}(e^{yX})$$

bezeichnet man als *erzeugende Funktion* der $M_{i_1,\ldots,i_k}$. Diese Bezeichnung wird deutlich an folgendem

Satz 1.3 *Es gilt*

$$M_{i_1,\ldots,i_k} = \frac{\partial^k}{\partial x_{i_1}, \ldots, \partial x_{i_k}} \eta(0).$$

Entwickelt man e^{yX} in eine Taylorreihe und wendet die Definition der erzeugenden Funktion an, so sieht man dieses unmittelbar ein.[3]

Die praktisch wichtigste Verteilung ist die Gauß–Verteilung. Eine Zufallsvariable $X : \Omega \to \mathbb{R}^n$ heißt *Gauß–verteilt*, wenn sie eine erzeugende Funktion η besitzt mit

$$\begin{aligned} \eta(y) &= e^{\frac{1}{2} y \Gamma y + \mu y}, \text{ wobei} \\ \Gamma &\geq 0 \text{ ist.} \end{aligned}$$

Hierbei bedeutet $\Gamma \geq 0$ einfach $y\Gamma y \geq 0 \, \forall y$. Ebenso sind Bezeichnungen wie $\Gamma \leq 0$, $\Gamma > 0$ und $\Gamma < 0$ zu lesen. Γ heißt *Kovarianzmatrix*, μ heißt *Mittelwert*. Man beachte, daß die hier gegebene Definition der Gauß–verteilten Zufallsvariable auch dann noch sinnvoll ist, wenn die Kovarianzmatrix singulär ist. Der folgende Satz charakterisiert eine Gauß–verteilte Zufallsvariable auf etwas bekanntere Weise:

[3] Über die Mächtigkeit dieses Werkzeugs und der eng damit verwandten *charakteristischen Funktion* siehe [4].

Satz 1.4 *Besitzt eine Gauß–verteilte Zufallsvariable X mit Mittelwert μ eine invertierbare Kovarianzmatrix Γ ($\Gamma > 0$), so besitzt X die Dichte*

$$p(x) = \frac{1}{\sqrt{(2\pi)^n \det \Gamma}} e^{-\frac{1}{2}(x-\mu)\Gamma^{-1}(x-\mu)}.$$

Diesen Sachverhalt rechnet man einfach nach. Ebenso sind die folgenden Aussagen nicht schwer zu verifizieren:

Satz 1.5 *Seien X, Y Gauß–verteilte Zufallsvariablen mit invertierbaren Kovarianzmatrizen Γ, Υ und Mittelwerten μ, ν, so gelten die folgenden Aussagen:*

1. $\mu = \mathrm{E}(X)$.
2. $\Gamma_{i,j} = \mathrm{Cov}(X) := \mathrm{E}(X_i - \mu_i)(X_j - \mu_j)$.
3. *Sind X, Y unabhängig, so gilt* $\mathrm{E}(X \cdot Y) = \mathrm{E}(X) \cdot \mathrm{E}(Y)$.[4] *Die Umkehrung ist ebenfalls richtig.*
4. *Ist A eine beliebige Matrix und b ein beliebiger Vektor, so ist $Z := AX + b$ ebenfalls Gauß–verteilt mit* $\mathrm{E}(Z) = A\mathrm{E}(X) + b$ *und* $\mathrm{Cov}(Z) = A\Gamma A^t$.
5. $X+Y$ *ist ebenfalls Gauß–verteilt mit* $\mathrm{E}(X+Y) = \mathrm{E}(X)+\mathrm{E}(Y)$ *und* $\mathrm{Cov}(X+Y) = \mathrm{Cov}(X) + \mathrm{Cov}(Y) + \mathrm{E}(X-\mu)(Y-\nu)$.
6. $\mathrm{E}(X_i - \mu_i)(X_j - \mu_j)(X_k - \mu_k)(X_l - \mu_l) = \Gamma_{ij}\Gamma_{kl} + \Gamma_{il}\Gamma_{jk} + \Gamma_{ik}\Gamma_{jl}$.

Somit ist klar, wie sich Gauß–verteilte Zufallsvariablen unter linearen Transformationen verhalten. Für spätere Anwendungen notieren wir noch die folgenden Ergebnisse:

Satz 1.6 *Sind $R, S > 0$ $(n \times n)$–Matrizen und A eine beliebige $(d \times n)$–Matrix, so gilt*

$$(\mu - x)R^{-1}(\mu - x) + (\nu - Ax)S^{-1}(\nu - Ax) = (\delta - x)T^{-1}(\delta - x) + const.$$

mit

$$T^{-1} = R^{-1} + A^t S^{-1} A$$
$$\delta = \mu + T A^t S^{-1}(\nu - A\mu).$$

Die Funktion

$$p(x) = \int e^{-\frac{1}{2}(\hat{x}-x)R^{-1}(\hat{x}-x)} \cdot e^{-\frac{1}{2}(\mu-\hat{x})S^{-1}(\mu-\hat{x})} \, \mathrm{d}\hat{x}$$

ist ebenfalls eine (unnormierte) Gauß–Verteilung mit Mittelwert

$$\delta = \mu$$

und Kovarianzmatrix

$$T^{-1} = S^{-1} + R^{-1},$$

wie man einfach nachrechnet.

[4] Dieses ist bei *beliebigen* Zufallsvariablen richtig.

1.3 Schätzprobleme und die Bayes–Formel

In der Statistik betrachtet man u. a. folgende Situation: Gegeben sei eine Zufallsvariable X mit Wahrscheinlichkeitsdichte $p_\theta(x)$, wobei θ ein Parameter aus einem beliebigen Parameterraum Θ sei. Dieser Parameter besitze nun einen festen, jedoch unbekannten Wert. Ziel ist es nun, diesen Wert zu *schätzen*, und zwar mit Hilfe einer *Stichprobe*. Eine Stichprobe $\xi = (\xi_1, \ldots, \xi_s)$ ist eine endliche Anzahl von Realisierungen der Zufallsvariable X, m. a. W. $\xi_1 = X(\omega_1), \xi_2 = X(\omega_2), \ldots$. Als einen *Schätzer* von θ bezeichnet man dann eine Funktion $\hat{\theta}(\xi)$, die jeder Stichprobe einen Wert in Θ zuordnet. Später wird die gesuchte Zufallsvariable der Zustand des dynamischen Systems sein. Ein Schätzer ist ebenfalls eine Zufallsvariable , da er von der zufälligen Stichprobe abhängt. Geeignete Schätzer zu finden bzw. Kriterien dafür anzugeben, wann ein Schätzer geeignet ist, ist Aufgabe der Schätztheorie (siehe hierzu vor allem [24] und auch [18] für eine Einführung).[5]

Der vielleicht wichtigste Schätzer ist der intuitiv sehr einleuchtende *Maximum–Likelyhood–Estimator*. Hierzu bestimmt man zu einer gegebenen Stichprobe ξ die Wahrscheinlichkeit $P_\theta(\xi)$. Sind z.B. die einzelnen Elemente der Stichprobe unabhängig, so ist $P_\theta(\xi) = p_\theta(\xi_1) \cdot \ldots \cdot p_\theta(\xi_s)$. Als Funktion von θ betrachtet nennt man diese Funktion die *Likelyhood–Funktion*, bezeichnet mit $L(\theta)$.

Definition 1.7 *Der* Maximum–Likelyhood–Estimator $\hat{\theta}(\xi)$ *ist der Wert von* θ*, für den die Likelyhood–Funktion* $L(\theta)$ *ein Maximum annimmt.*

Man schätzt also θ so, daß die gegebene Stichprobe gerade die mit der größten Wahrscheinlichkeit wird. Ist z.B. X eine reelle Gauß–verteilte Zufallsvariable mit unbekannten Parametern μ (Erwartungswert) und σ (standartabweichung) und ξ eine Stichprobe mit s unabhängigen Werten ξ_i, so ist

$$L(\mu, \sigma) = \prod_{i=1}^{s} \frac{1}{\sqrt{(2\pi\sigma^2)^s}} e^{-\frac{1}{2\sigma^2}(\xi_i - \mu)^2}$$

und als Maximum–Likelyhood–Estimator ergibt sich, wie man leicht nachrechnet[6]

$$\begin{aligned} \hat{\mu}(\xi) &= \frac{1}{s} \sum_{i=1}^{s} \xi_i, \\ \hat{\sigma}(\xi) &= \sqrt{\frac{1}{s} \sum_{i=1}^{s} (\xi_i - \hat{\mu}(\xi))^2}. \end{aligned}$$

[5] In der Schätztheorie wird nicht nur θ allein, sondern auch der Wert einer gegebenen Funktion $g(\theta)$ geschätzt.

[6] Hierzu bildet man praktischerweise die *Log–Likelyhood–Funktion*. Das ist der natürliche Logarithmus der Likelyhood–Funktion. Da der Logarithmus monoton steigt, hat die Log–Likelyhood–Funktion genau dieselben Maxima wie die Likelyhood–Funktion.

Der Maximum–Likelyhood–Estimator kann jedoch auch systematisch falsche Schätzungen liefern. Angenommen, man möchte schätzen, wieviele Streifenwagen die Göttinger Polizei besitzt. Diese unbekannte Zahl sei N. Jeder Wagen besitzt eine Funkkennung, bei der es sich um eine fortlaufende Nummer handelt. Durch Abhören des Funkverkehrs verschafft man sich einen Überblick über die derzeit im Einsatz befindlichen Wagen. Da sich jeder Wagen zumindest einmal während des Einsatzes über Funk meldet, hat man bald alle Kennungen zusammen. Es sei ν die größte beobachtete Kennummer. Wie auch immer die Wahrscheinlichkeit der Stichprobe von N abhängt, sie muß verschwinden, wenn $N < \nu$ ist. Nehmen wir weiter an, daß sie für $N \geq \nu$ mit wachsendem N kleiner wird, so ist der Maximum-Likelyhood-Schätzer $\hat{N} = \nu$, was systematisch zu klein ist.

Obwohl der Name dies suggeriert, kann man im allgemeinen nicht davon sprechen, der Maximum-Likelyhood-Schätzer sei der wahrscheinlichste Wert für den Parameter, da auf dem Parameterraum im Allgemeinen kein Maß gegeben ist. Gerade ein solches Maß kann jedoch sehr hilfreich sein, um eventuell vorhandene *a-priori-Information* über mögliche Parameterwerte einzubringen. Im eben angeführten Polizeiauto-Beispiel will man evtl. die Information ausnutzen, daß jüngstens ein Auto einen Totalschaden hatte und die Nummer noch nicht wieder vergeben wurde. Oder man hat gestern schon eine Schätzung gemacht und eine gewisse Verteilung für N ermittelt, die man nun einbringen möchte. Hierzu dient das Konzept der *bedingten Wahrscheinlichkeit*, also der Wahrscheinlichkeit für ein Ereignis unter der Bedingung, daß ein anderes Ereignis bereits stattgefunden hat. Gegeben sei also ein Ereignis $A \in \sigma$ mit $P(A) > 0$. Für ein beliebiges Ereignis $B \in \sigma$ definiert man die bedingte Wahrscheinlichkeit für B unter der Bedingung A

$$P(B|A) := P(A, B)/P(A).$$

Man verifiziert unmittelbar die BAYES–Formel

$$P(B|A) \cdot P(A) = P(A|B) \cdot P(B),$$

denn beide Seiten sind gleich $P(A, B)$. Man macht sich leicht klar, daß für festes A die bedingte Wahrscheinlichkeit $P(\cdot|A)$ wieder eine Wahrscheinlichkeit ist. Da man in der Schätztheorie mit Zufallsvariablen arbeitet, ist man an bedingten Wahrscheinlichkeiten der Form $P(X \in A|Y \in B)$ interessiert. Ohne Beweis (siehe hierzu [4], Kap. 4) zitieren wir hier den

Satz 1.8 *Seien X, Y Zufallsvariable und es existiere eine gemeinsame Dichte p mit $P(X \in A, Y \in B) = \int_A\int_B p(x, y)\,\mathrm{d}x\,\mathrm{d}y$. Dann ist für festes y durch $P(X \in A|Y = y)$ eine Wahrscheinlichkeit gegeben mit Dichte $p(x|y) = \frac{p(x,y)}{p(y)}$, d. h. es gilt*

$$P(X \in A|Y = y) = \int_A \frac{p(x, y)}{p(y)}\,\mathrm{d}x.$$

Außerdem gilt die Formel

$$p(x|z) = \int p(x|y,z)p(y|z)\,\mathrm{d}y.$$

Die Bayes–Formel lautet nunmehr

$$p(x|y) \cdot p(y) = p(y|x) \cdot p(x).$$

Integriert man diese Formel über x und bedenkt $\int p(x|y)\,\mathrm{d}x = 1$, so folgt

$$p(x|y) = \frac{p(y|x) \cdot p(x)}{\int p(y|x) \cdot p(x)\,\mathrm{d}x}. \tag{1.9}$$

Kehren wir zurück zu den Schätzproblemen. Der gesuchte Parameter ist eine Zufallsvariable, und man kann nach der Wahrscheinlichkeit für θ unter der Bedingung ξ fragen. Mit der Bayes–Formel (1.9) ergibt sich

$$p(\theta|\xi) = k \cdot p(\xi|\theta) \cdot p(\theta),$$

wobei k nur ein Normierungsfaktor ist. Diese Formel ist die Grundlage aller *Bayesian Estimations.* Die linke Seite enthält gewissermaßen die volle Information über θ, die wir mit Hilfe der Stichprobe gewinnen können. Rechts steht die *a-priori-Wahrscheinlichkeit* $p(\theta)$, die die vorher schon vorhandene Information über den Parameter enthält und die bedingte Wahrscheinlichkeit $p(\xi|\theta)$, die nichts anderes darstellt als die Funktion $p_\theta(\xi)$, von der oben die Rede war und die unser Modell darstellt. Einen Schätzer gewinnt man jetzt aus $p(\theta|\xi)$, wobei es verschiedene Möglichkeiten gibt. Man kann wieder das Maximum bezüglich θ wählen, womit man eine Art verallgemeinerten Maximum–Likelyhood–Estimator hätte. Ebenso wichtig ist auch der *Posterior-Bayes-Estimator*, der durch den Erwartungswert

$$\hat{\theta}(\xi) := \int \theta\, p(\theta|\xi)\,\mathrm{d}\theta$$

gegeben ist.

Zum Schluß des Abschnitts noch eine Verallgemeinerung der Bayes–Formel, die *Bayes–Kettenregel* :

Satz 1.10 $$p(x_n, ..., x_2|x_1) = \prod_{i=1}^{n} p(x_i|x_{i-1}, ..., x_1),$$

welche man sich durch elementares Nachrechnen ebenso klar macht wie die normale Bayes–Formel.

2 Nichtlineare Dynamik

Dieses Kapitel stellt erst einmal einige wichtige Begriffe vor, die beim Studium von dynamischen Systemen wichtig sind. Obwohl oder gerade weil diese Dinge zum Grundwissen der Theorie der dynamischen Systeme gehören, sollen sie hier der Vollständigkeit halber aufgeführt werden. Hierzu gehören auch einige Tatsachen über lineare Systeme, die trotz des Namens dieses Kapitels erst einmal besprochen werden. In dieser Arbeit werden Techniken für iterierte Abbildungen entwickelt, die deshalb auch in diesem Kapitel die Hauptrolle spielen werden. Bevor es losgeht, gleich eine generelle Definition:

Definition 2.1 *Ein* dynamisches System *ist eine m-dimensionale differenzierbare Mannigfaltigkeit M^m und darauf ein Diffeomorphismus F. M^m wird auch* Zustandsraum *oder* Phasenraum *genannt, F auch* Dynamik.[1]

Allgemein geht es hier, solange nichts gegenteiliges gesagt wird, um autonome dynamische Systeme. Zu jedem Anfangswert $x_0 \in M$ induziert ein dynamisches System eine Folge von Punkten $x_{-\infty}^{\infty} := (\ldots, x_{-1}, x_0, x_1, \ldots)$ durch die Vorschrift $x_{n+1} = F(x_n)$. Aufgrund der Diffeomorphie ist durch ein beliebiges Element des *Orbits* $x_{-\infty}^{\infty}$ dieser eindeutig festgelegt. Hängt allgemeiner F auch noch von n ab, so spricht man von *nichtautonomen* dynamischen Systemen.

2.1 Lineare iterierte Abbildungen

Ist $M^m = \mathbb{R}^m$, so hat man dort die Klasse der *linearen* dynamischen Systeme, die durch $F(x) = Ax + b$ gegeben sind. A ist hierbei eine (notwendigerweise reguläre) Matrix, b ein beliebiger fester Vektor. Ist $b = 0$, so spricht man von *homogenen* linearen dynamischen Systemen.[2] Für alle linearen Abbildungen hat man eine Norm, die folgendermaßen erklärt ist:

Definition 2.2 *Sei $\|\cdot\|$ eine Norm auf dem $\mathbb{R}^m$. Dann induziert diese Norm eine Matrixnorm, die* Operatornorm*, die wieder mit $\|\cdot\|$ bezeichnet sei, und die durch folgende Vorschrift gegeben ist:*

$$\|A\| := \max(\|Ax\|, \|x\| = 1).$$

[1] Zu Mannigfaltigkeiten findet man eine allgemeine Einführung in [5] und in [6].

[2] Die in diesem Abschnitt besprochenen Resultate findet man vor allem in Numerikbüchern, z.B. [29]

Im Folgenden ist mit Ausdrücken wie $\|A\|$ immer die Operatornorm gemeint.

Man prüft leicht nach, daß es sich wirklich um eine Norm handelt. Mit Hilfe dieser Norm formuliert man folgendes zentrales

Theorem 2.3 (über lineare dynamische Systeme) *Ist $Ax + b$ ein lineares dynamisches System auf dem $\mathbb{R}^m$, so gelten folgende Aussagen:*

1. *Ist x_0 der Anfangswert, so gilt*
$$x_n = A^n x_0 + \sum_{k=0}^{n-1} A^k b \tag{2.4}$$

2. *Ist $\|A\| < 1$, so gilt*
$$\lim_{n\to\infty} x_n = (1 - A)^{-1} b.$$

3. *Ist hingegen $\|A^{-1}\| < 1$, so ist identisch $x_n = (1 - A)^{-1} b$ oder*
$$\lim_{n\to\infty} x_n = \infty.$$

Man beachte, daß diese Aussagen für jede *Norm richtig sind.*

Beweis: Die erste Formel prüft man durch elementares Nachrechnen. Die zweite Aussage kann man auf zwei Arten einsehen: Man prüfe, daß die Bedingung garantiert, daß die Abbildung *kontrahierend* ist. Deshalb konvergiert x_n nach dem BANACHschen Fixpunktsatz gegen einen Fixpunkt x_f von $Ax + b$. Es gilt also $Ax_f + b = x_f$, woraus $x_f = (1 - A)^{-1} b$ folgt. Man kann aber auch direkt schließen: Berücksichtigt man die Abschätzung $\|A^n b\| \leq \|A\|^n \|b\|$ sowie die Dreiecksungleichung, so erkennt man die Abschätzung

$$\|\sum_{k=m}^{n-1} A^k\| \leq \sum_{k=m}^{n-1} \|A\|^k.$$

Da $\|A\| < 1$ ist, kann man die rechte Seite nach dem CAUCHY-Kriterium, angewandt auf die geometrische Reihe, durch Wahl von m beliebig klein machen und sieht so ein, daß auch die linke Seite beliebig klein wird und damit $\sum_{k=0}^{n-1} A^k b$ eine Cauchyfolge ist. Ebenso ist offenbar $A^n b$ eine Nullfolge. Außerdem kann man mit einem Argument wie bei der geometrischen Reihe zeigen, daß

$$(1 - A) \cdot \sum_{k=0}^{n-1} A^k b = (1 - A^n) \cdot b \tag{2.5}$$

ist. Auf beiden Seiten kann man nun (für beliebiges b) zum Limes übergehen und erhält

$$\sum_{k=0}^{\infty} A^k b = (1-A)^{-1} b.$$

Also erhalten wir aus Aussage (2.4) durch Übergang zum Limes die Aussage (2).

Zum Beweis der dritten Aussage mache man sich zuerst klar, daß auch in diesem Fall $1-A$ invertierbar ist. Angenommen, es wäre $Ax = x$ für ein $x \neq 0$, so wäre $x = A^{-1}x$ und damit $\|x\| \leq \|A^{-1}\|\|x\|$. Dann ergäbe sich aber nach Definition der Operatornorm der Widerspruch $1 \leq \|A^{-1}\|$. Außerdem gilt für $x \neq 0$ die Formel

$$\lim_{n\to\infty} A^n x = \infty,$$

denn aus $A^n x = y_n$ ergibt sich $x = A^{-n} y_n$ und damit $\|x\| \leq \|A^{-1}\|^n \|y_n\|$, was wiederum $(\frac{1}{\|A^{-1}\|})^n \|x\| \leq \|y_n\|$ nach sich zieht. Die linke Seite jedoch divergiert, also auch die rechte. Nach Gleichung (2.5) gilt nun

$$\begin{aligned} x_n &= A^n x_0 + (1-A)^{-1}(1-A^n)\cdot b \\ &= A^n \cdot (x_0 - (1-A)^{-1} b) + (1-A)^{-1} b, \end{aligned}$$

und der erste Term rechter Hand divergiert, es sei denn, $x_0 = (1-A)^{-1}b$, womit dann auch identisch $x_n = (1-A)^{-1}b$ gilt □

In den oben besprochenen Fällen ist also stets $1-A$ invertierbar bzw. 1 kein Eigenwert von A. Dann ist $(1-A)^{-1}b$ der einzige Fixpunkt des Systems, und der obige Satz gibt an, wann er stabil bzw. instabil ist.[3]

Es deutete sich bereits an, daß die Eigenwerte eine zentrale Rolle spielen bei der Entscheidung, ob der Fixpunkt stabil bzw. instabil ist. Dies wird in dem gleich folgenden Lemma etwas präzisiert. Vorerst eine

Definition 2.6 *Sei A eine lineare Abbildung und $\lambda_{\max}$ der Eigenwert von A mit dem größten Betrag. Dann nennt man $\rho(A) := |\lambda_{\max}|$ den* Spektralradius *von A.*

...und nun das

Lemma 2.7 *Sei A eine lineare Abbildung und $\rho(A)$ der Spektralradius von A. Dann gilt:*

1. $\rho(A) \leq \|A\|.$

   Ist insbesondere $\|A\| < 1$, so ist auch $\rho(A) < 1$.

[3] Von stabilem (bzw. instabilem) Fixpunkt soll die Rede sein, wenn das System für alle Anfangswerte gegen den Fixpunkt konvergiert (bzw. für alle Anfangswerte, die ungleich dem Fixpunkt sind, divergiert).

2. *Zu jedem ϵ existiert eine Norm $\|\cdot\|_\epsilon$ auf $\mathbb{R}^m$, so daß für die zugehörige Operatornorm gilt*

$$\|A\|_\epsilon \leq \rho(A) + \epsilon.$$

Ist insbesondere $\rho(A) < 1$, so existiert ein ϵ mit $\|A\|_\epsilon < 1$.

3. *Ist A symmetrisch oder unitär, so gilt $\rho(A) = \|A\|$ für die euklidische Norm.*

Beweis:

1. Man wähle x mit $\lambda_{\max} x = Ax$ und wende darauf die Norm an. Es ergibt sich

$$|\lambda_{\max}|\|x\| = \|Ax\| \leq \|A\|\|x\|.$$

Division durch $\|x\|$ ergibt die Behauptung.

2. Es sei $\|\cdot\|$ die euklidische Norm. Nach einer Variante des Satzes von JORDAN existiert zu jedem ϵ ein Isomorphismus T_ϵ, so daß gilt

$$T_\epsilon^{-1} A T_\epsilon = D + N,$$

wobei D eine Diagonalmatrix mit den Eigenwerten von A und $\|N\| < \epsilon$ ist. Man stellt nun leicht fest, daß mit $\|\cdot\|$ auch $\|\cdot\|_\epsilon := \|T_\epsilon^{-1} \cdot\|$ eine Norm ist. Es gilt

$$\begin{aligned} \|Ax\|_\epsilon &= \|T_\epsilon^{-1} A T_\epsilon \cdot T_\epsilon^{-1} x\| \\ &\leq \|T_\epsilon^{-1} A T_\epsilon\| \cdot \|T_\epsilon^{-1} x\| \\ &= \|D + N\| \cdot \|x\|_\epsilon \\ &\leq (\rho(A) + \epsilon) \cdot \|x\|_\epsilon. \end{aligned}$$

Nun wählt man ein x mit $\|x\|_\epsilon = 1$ und $\|Ax\|_\epsilon = \|A\|_\epsilon$ und erhält daraus die Behauptung.

3. Es sei nun wieder $\|\cdot\|$ die euklidische Norm. Offenbar ist für ein x_0 mit $\|x_0\| = 1$ die Funktion $\|Ax\|$ genau dann maximal, wenn $\langle Ax, Ax\rangle$ maximal ist für dasselbe x_0 mit $\langle x_0, x_0\rangle = 1$. Nach der Methode der Multiplikatoren von LAGRANGE ist dann notwendigerweise $Ax_0 = \lambda x_0$ für ein $\lambda \in \mathbb{R}$. Aus dieser Gleichung ergibt sich dann unmittelbar

$$\|A\| = \sqrt{\langle Ax_0, Ax_0\rangle} = \sqrt{\langle \lambda x_0, \lambda x_0\rangle} = |\lambda| = \rho(A).$$

Für unitäre Matrizen ist natürlich $\|A\| = \rho(A) = 1$

□

Mit diesem Lemma ausgerüstet zieht man nun folgende in der Praxis recht wichtige

Folgerung 2.8 *1. Es gibt stets eine Norm mit*

$$\rho(A) < 1 \Leftrightarrow \|A\| < 1.$$

2. Es gibt stets eine Norm mit

$$|\lambda_{min}(A)| > 1 \Leftrightarrow \|A^{-1}\| < 1,$$

wobei $\lambda_{min}(A)$ der betragsmäßig kleinste Eigenwert ist.

Beweis: Der Beweis ergibt sich direkt aus dem vorhergehenden Lemma und der Tatsache $\rho(A^{-1}) = \frac{1}{|\lambda_{min}(A)|}$. □

Um die Stabilität des Fixpunktes zur ermitteln, muß man also die Eigenwerte studieren.

Das Konzept der linearen Abbildungen ist nicht nur deswegen wichtig, weil man Lösungen explizit berechnen kann[4], sondern auch, weil man mit Hilfe der Ableitung auch nichtlineare Probleme innerhalb gewisser Grenzen in lineare transformieren kann.

2.2 Iterierte Abbidungen auf Mannigfaltigkeiten

Eine allgemeine Mannigfaltigkeit ist kein Vektorraum, und deshalb ist der Begriff der linearen Abbildung hier nicht sinnvoll. Trotzdem interessiert man sich für Lösungen und insbesondere für deren Langzeitverhalten. Dies ist die Fragestellung der *Ergodentheorie*, aus der einige Resultate im nächsten Kapitel kurz aufgeführt werden.

Ab jetzt sei mit Mannigfaltigkeit stets eine kompakte (unberandete) Mannigfaltigkeit gemeint, es sei denn, es wird explizit etwas anderes gesagt. Auf einer solchen Mannigfaltigkeit kann man immer eine RIEMANN*sche Metrik* einführen. Das bedeutet, daß auf den Tangentialräumen T_pM ein Skalarprodukt $(\ ,\ )_p$ gegeben ist, welches differenzierbar vom Punkte p abhängt. Man kann dann von

[4] Für eine allgemeine Lösung kann man sich natürlich auf Formel (2.4) zurückziehen. Um diesen Ausdruck etwas expliziter zu berechnen, kommt man normalerweise nicht umhin, die Matrix auf eine praktische Normalform zu bringen, was aber auch nur „im Prinzip“ möglich ist.

Längen von Tangentialvektoren, Normen linearer Abbildungen (des Tangentialraumes) und orthogonalen Transformationen sprechen. Außerdem kann man ein invariantes Integral konstruieren, was für eine Funktion f auf M^m lokal durch

$$\int_A f := \int_{\mathbb{R}^m} (f \cdot g) \circ \varphi^{-1} \, \mathrm{d}x_1, \ldots, x_n$$

gegeben ist. Hierbei sind φ die lokalen Koordinaten und g die GRAMsche Determinante. Insbesondere kann man die Funktion $f = 1$ integrieren und erhält so ein Maß μ auf M^m, das gerade dem LEBESGUE–Maß auf dem $\mathbb{R}^m$ entspricht und ab jetzt auch so genannt wird. Da M^m kompakt ist, ist dieses Maß endlich, also o.B.d.A. $\mu(M^m) = 1$. Die Metrik bringt weitere Struktur auf die Mannigfaltigkeit, doch in unserem Falle kommt es eher darauf an, daß die Aussagen, die mit Hilfe der Metrik getroffen werden, von einer speziellen Wahl der Metrik unabhängig sind. Die Metrik gestattet es auch, die Abstände zwischen Punkten auf der Mannigfaltigkeit zu definieren. Für uns ist es jedoch praktischer, die Mannigfaltigkeit in den $\mathbb{R}^m$ einzubetten (dieses Problem wird in Kapitel 2.5 erörtert). Die Mannigfaltigkeit erbt dann vom $\mathbb{R}^m$ eine Metrik, welche die Topologie der Mannigfaltigkeit induziert. Mit der Bezeichnung $d(p, q)$ sei im Folgenden immer diese Metrik gemeint. Es sollen nun einige wichtige Eigenschaften erklärt werden, die man in vielen dynamischen Systemen vorfindet und die sozusagen qualitative Merkmale darstellen. Wir bleiben zunächst beim Volumen und betrachten, wie sich das Volumen einer Menge unter der Abbildung transformiert. Die Volumenform transformiert sich mit der Funktionaldeterminante $\lambda(p)$, und an ihr kann man erkennen, ob das Volumen kontrahiert oder expandiert oder konstant bleibt:

Definition 2.9 *Eine dynamisches System heißt* konservativ*, falls identisch* $\lambda(p) = 1$. *Ist* $\lambda(p) < 1$, *so nennt man es* dissipativ.

Definition 2.10

1. *Ein Punkt* $p \in M$ *heißt* Fixpunkt *des Systems, wenn* $F(p) = p$ *ist. In diesem Falle ist* $\mathrm{D}F(p)$ *eine Abbildung von* T_pM *in sich und man kann die Eigenwerte von* $\mathrm{D}F(p)$ *studieren.*

2. *Ein Punkt* $p \in M$ *heißt* periodischer Punkt der Ordnung n *des Systems, wenn* p *ein Fixpunkt von* F^n *ist. Der induzierte Orbit heißt periodischer Orbit. Hier kann man natürlich wieder die Eigenwerte von* $\mathrm{D}F^n(p)$ *studieren.*

3. *Eine Teilmenge* $A \subset M$ *heißt invariant, wenn* $A = F^{-1}(A)$.

Fixpunkte und periodische Orbits sind also invariante Mengen. Häufig entsteht die Frage, wie das Verhalten von F in der Nähe dieser Mengen aussieht. Konvergiert die Trajektorie gegen den Fixpunkt, läuft sie davon weg oder beides nicht? Hierzu betrachten wir folgendes

Lemma 2.11 *Sei $F : U \subset \mathbb{R}^m \to \mathbb{R}^m$, und es gelte für eine kleine Kugel $B(x^*) \subset U$ und ein $K < 1$ die Abschätzung*

$$\|F(x) - x^*\| \leq K\|x - x^*\| \, \forall x \in B(x^*),$$

mit anderen Worten F sei kontrahierend *mit Kontraktionskonstante K. Dann ist x^* ein Fixpunkt von F und die Folge $F^n(x)$ konvergiert gegen x^*, falls $x \in B(x^*)$.*

Beweis: Es gilt $\|F(x^*) - x^*\| \leq 0$, also ist x^* ein Fixpunkt. Außerdem ist natürlich $\|F^n(x) - x^*\| \leq K^n\|x - x^*\|$, also $\|F^n(x) - x^*\| \to 0$ für $n \to \infty$ □

Um die Kontraktionseigenschaften der Abbildung nachzuweisen, benutzt man die Ableitung, genauer führt man das Problem auf eine kontrahierende lineare Abbildung zurück:

Satz 2.12 *Sei p ein Fixpunkt des dynamischen Systems und es gelte $\rho(\mathrm{D}F(p)) < 1$, dann ist p stabil, d.h. es existiert eine Umgebung U von p mit $F^n(q) \to p$, falls $q \in U$. Umgekehrt impliziert die Ungleichung $\rho(\mathrm{D}F(p)^{-1}) < 1$, daß $F^{-n}(q) \to p$, falls $q \in U$.*

Beweis: Nach Einführen lokaler Koordinaten φ zieht man sich lokal auf eine Umgebung $U \subset \mathbb{R}^m$ zurück und hat dort für $\mathrm{D}F(p)$ auch dieselben Eigenwerte, da sich $\mathrm{D}F(p)$ lokal als $\mathrm{D}\varphi(p)\mathrm{D}F(p)\mathrm{D}\varphi(p)^{-1}$ darstellt, und das ist eine Ähnlichkeitstransformation. Auch kann man die Umgebung U so klein wählen, daß man komplett in der Situation des Lemmas ist, nur die Ungleichung muß noch etabliert werden. Wir wählen eine Norm $\|\cdot\|$ so daß $\|\mathrm{D}F(x^*)\| < K < 1$ ist. Das ist möglich nach (2.7). Da die Funktion $\|\mathrm{D}F\|$ stetig ist, kann man durch weiteres Verkleinern von U erreichen, daß $\|\mathrm{D}F(x)\| < K < 1$ in ganz U gilt. Nun wendet man den Mittelwertsatz an:

$$F(x) - F(x^*) = \int_0^1 \mathrm{D}F(x^* + t(x - x^*)) \, \mathrm{d}t \cdot (x - x^*),$$

und daraus folgt

$$\|F(x) - x^*\| \leq \int_0^1 \|\mathrm{D}F(x^* + t(x - x^*))\| \, \mathrm{d}t \cdot \|(x - x^*)\| \leq K \cdot \|(x - x^*)\|.$$

Die anderen Aussagen beweißt man analog. □

Genauso kann man die Stabilität von m-fach periodischen Orbits untersuchen, indem man die eben bewiesenen Aussagen auf F^m anwendet.

Eine wesentlich allgemeinere Situation liegt im Fall des *hyperbolischen* Fixpunktes vor. Es sei gleich allgemeiner eine hyperbolische Menge definiert:

Definition 2.13 *Eine abgeschlossene Menge $\Lambda \subset M$ heißt* hyperbolisch*, falls $F(\Lambda) = \Lambda$ und für jedes $x \in \Lambda$ gilt: Der Tangentialraum T_xM ist eine direkte Summe $T_xM = E^u_x \oplus E^s_x$ von Unterräumen mit*

1. $\mathrm{D}F(x)E_x^u = E_{F(x)}^u$, $\mathrm{D}F(x)E_x^s = E_{F(x)}^s$.

2. *Es existieren Konstanten c, $\lambda \in (0,1)$ mit*

$$\begin{aligned} \|\mathrm{D}F^n v\| &\le c\lambda^n \|v\| \quad \textit{wenn} \quad v \in E_x^s,\ n \ge 0, \\ \|\mathrm{D}F^{-n} v\| &\le c\lambda^n \|v\| \quad \textit{wenn} \quad v \in E_x^u,\ n \ge 0. \end{aligned}$$

3. E_x^u, E_x^s *hängen stetig von x ab.*

Der Tangentialraum gestattet also eine invariante Zerlegung in stabile bzw. instabile Unterräume. Die Frage ist nun wieder, ob diese Eigenschaften der Linearisierung sich lokal auf die nichtlineare Abbildung F ausdehnen lassen. Hierzu definiert man für $x \in M$ die folgenden Objekte:

$$\begin{aligned} W^s(x) &:= \{y \in M; d(F^n(x), F^n(y)) \to 0 \text{ wenn } n \to \infty\} \\ W_\epsilon^s(x) &:= \{y \in M; d(F^n(x), F^n(y)) < \epsilon \text{ für alle } n \ge 0\} \\ W^u(x) &:= \{y \in M; d(F^n(x), F^n(y)) \to 0 \text{ wenn } n \to -\infty\} \\ W_\epsilon^u(x) &:= \{y \in M; d(F^n(x), F^n(y)) < \epsilon \text{ für alle } n \le 0\} \end{aligned}$$

Man nennt $W^s(x)$ (bzw. $W^u(x)$) die *stabile* (bzw. *instabile*) *Mannigfaltigkeit* von x und $W_\epsilon^s(x)$ (bzw. $W_\epsilon^u(x)$) die *lokal stabile (bzw. lokal instabile) Mannigfaltigkeit* von x. Hierbei handelt es sich wirklich um Mannigfaltigkeiten, genauer um *C^r–Disks*. Eine C^r–Disk ist eine Mannigfaltigkeit, die C^r–diffeomorph zu einer Kreisscheibe im euklidischen Raum ist. Die genaue Situation sieht folgendermaßen aus:

Theorem 2.14 (Hirsch, Pugh, zitiert nach [3]) *Sei Λ eine hyperbolische Menge für ein r-fach stetig differenzierbares dynamisches System F. Dann gilt für genügend kleines ϵ:*

1. $W_\epsilon^s(x)$, $W_\epsilon^u(x)$ *sind C^r-Disks für $x \in \Lambda$ mit* $T_x W_\epsilon^s(x) = E_x^s$, $T_x W_\epsilon^u(x) = E_x^u$.

2. $d(F^n(x), F^n(y)) \le \lambda^n d(x,y)$ *für* $y \in W_\epsilon^s(x), n \ge 0$,
   $d(F^{-n}(x), F^{-n}(y)) \le \lambda^n d(x,y)$ *für* $y \in W_\epsilon^u(x), n \ge 0$,
   und

3. $W_\epsilon^s(x)$, $W_\epsilon^u(x)$ *hängen stetig von x ab.*

Dieses Theorem hat weitreichende Konsequenzen und kann als Krönung der lokalen Theorie dynamischer Systeme angesehen werden. Von hier aus startet man zum Studium der globalen Eigenschaften dynamischer Systeme. Es werden nun einige Folgerungen aus dem Theorem von Hirsch zusammengestellt, die später im Zusammenhang mit dem Problem des *shadowing* eine gewisse Rolle spielen.

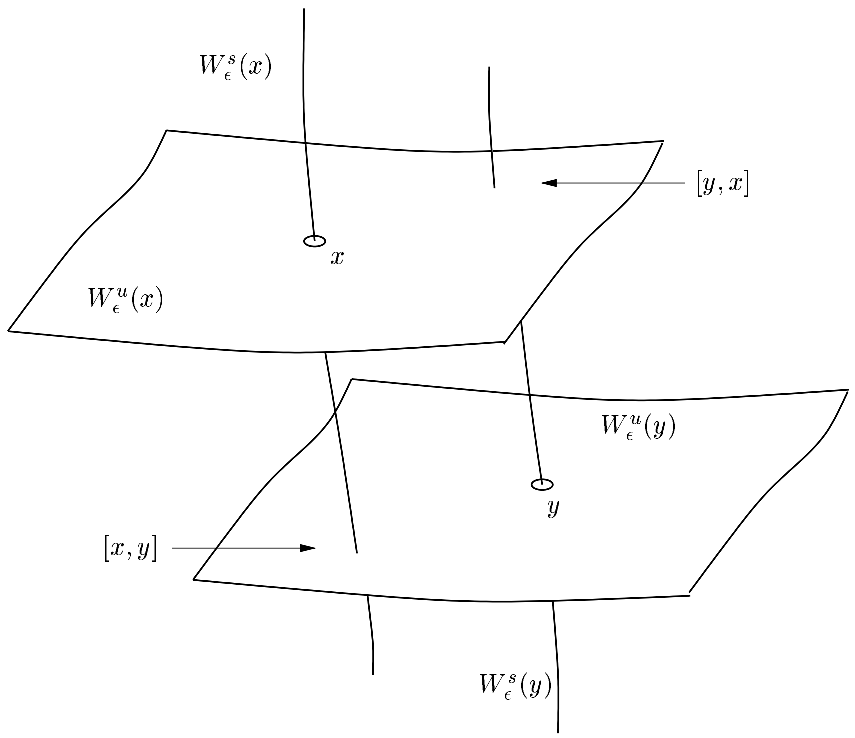


Abbildung 2.1: Lokale stabile und instabile Mannigfaltigkeiten

Folgerung 2.15 *Sei Λ eine invariante hyperbolische Menge und $x, y \in \Lambda$.*

1. *Es ist $W_\epsilon^s(x) \subset W^s(x)$ und $W_\epsilon^u(x) \subset W^u(x)$.*

2. *Für genügend kleines ϵ existiert ein δ, so daß $W_\epsilon^s(x) \cap W_\epsilon^u(y)$ aus genau einem Punkt $[x, y]$ besteht, falls $d(x, y) < \delta$.*

3. *F ist* expansiv*, d.h. es existiert ein festes $\epsilon > 0$, so daß, falls $x \neq y$ ist, stets ein $n \in \mathbb{Z}$ existiert mit $d(F^n(x), F^n(y)) > \epsilon$.*

Beweis: siehe [3]. Die erste Aussage ist klar nach dem Theorem von Hirsch. Die zweite Aussage folgt daraus, daß $W_\epsilon^s(x)$ und $W_\epsilon^u(x)$ transversal sind, und diese Eigenschaft ist unempfindlich gegen kleine Störungen. Die Situation wird in Abb. 2.1 dargestellt. Die dritte Eigenschaft folgt aus der zweiten. In der Tat, denn sonst wäre $y \in W_\epsilon^u(x) \cap W_\epsilon^s(x)$, und darin befindet sich auch x selbst. Also muß $y = x$ sein $\Box$

Für weitere Resultate ist es notwendig, daß Λ möglichst viele („wichtige") Punkte enthält, z.B. ganz M:

Definition 2.16 *Eine Abbildung F heißt* hyperbolisch *bzw. ein* ANOSOV-

Diffeomorphismus, falls ganz M hyperbolisch ist.[5]

Diese Definitionen und Aussagen erläutern die lokale Situation hyperbolischer Systeme. Als Anwendung soll nun das Konzept des *shadowing* vorgestellt werden, welches sowohl in der Theorie als auch in der Numerik eine große Rolle spielt. Von jetzt ab bis zum Ende dieses Abschnitts sei F hyperbolisch.

Definition 2.17 *Eine Folge $x_k^l := \{x_i \in M; i = k, \ldots, l\}$ heißt ein δ–*Pseudo–Orbit*, falls stets $d(F(x_i), x_{i+1}) < \delta$ gilt.*

*Ein Punkt $x \in M$ heißt ein ϵ–*Schatten *von x_k^l, falls $d(F^i(x), x_i) < \epsilon$ gilt für alle $i = k, \ldots, l$.*

Satz 2.18 (Bowens Shadowing–Lemma) *Für genügend kleines $\epsilon > 0$ existiert ein $\delta > 0$, so daß jeder δ–Pseudo–Orbit x_k^l einen ϵ–Schatten $x \in M$ besitzt. Hierbei ist $k = -\infty$, $l = \infty$ erlaubt.*

Beweis: Siehe [3]. Die wesentlich Idee sei hier kurz erläutert. Der Einfachheit halber sei $k = 0$ und l endlich. Man konstruiert nun induktiv einen zweiten Pseudo–Orbit y_k^l, und zwar setzt man zuerst $y_0 := x_0$. Ist nun y_i schon konstruiert, so setzt man $y_{i+1} := [x_{i+1}, F(y_i)]$, also y_{i+1} ist der Schnittpunkt der stabilen Mannigfaltigkeit von x_{i+1} mit der instabilen Mannigfaltigkeit von $F(y_i)$ (siehe hierzu Abb. 2.2). Dann setzt man schließlich $x := F^{-l}(y_l)$. Dieses x hat dann die gewünschten Eigenschaften, denn $d(F^i x, x_i) \leq d(F^i x, y_i) + d(y_i, x_i)$. Der erste Term läßt sich umschreiben zu

$$\begin{aligned} d(F^i x, y_i) &= d(F^{i-l} y_l, y_i) \\ &\leq \sum_{s=i+1}^{l} d(F^{i-s} y_s, F^{i-s+1} y_{s-1}) \\ &= \sum_{s=i+1}^{l} d(F^{i-s} y_s, F^{i-s}(F y_{s-1})) \\ &\leq \sum_{s=i+1}^{l} \lambda^{i-s} d(y_s, F y_{s-1}). \end{aligned}$$

Hierbei wurde ausgenutzt, daß $y_s \in W_\epsilon^u(F y_{s-1})$. Die verbleibende Distanz kann man ebenfalls geeignet abschätzen und dann die geometrische Reihe anwenden. Für den Fall $k = -\infty$, $l = \infty$ bildet man zuerst einen ϵ–Schatten $x(m)$ für x_{-m}^m und bekommt dann eine Folge $x(m)$, die sich irgendwo häufen muß. Ein solcher Punkt hat dann die gewünschten Eigenschaften □

Im folgenden wird auch öfters von *Shadowing–Orbits* zu einem gegebenen Pseudo–

[5] Eine etwas weitere Klasse sind die sogenannten *Axiom–A*–Diffeomorphismen, für die im wesentlichen dasselbe gilt. Anosov-Diffeomorphismen sind immer Axiom–A–Diffeomorphismen (*Anosov's closing lemma*).

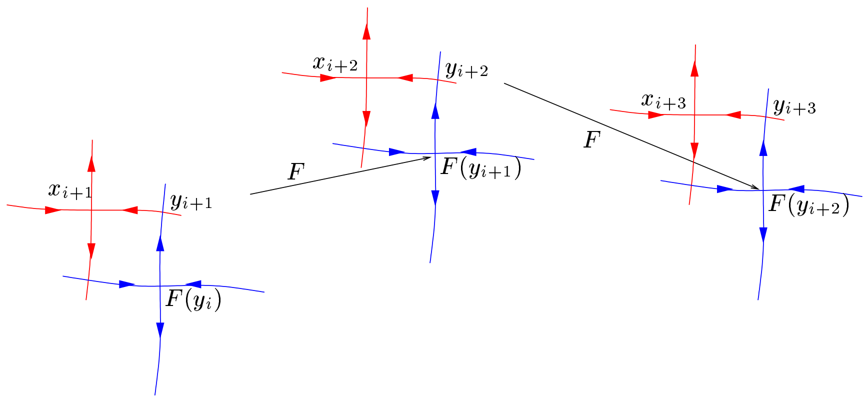


Abbildung 2.2: Zum Shadowing–Lemma

Orbit die Rede sein. Damit sind Orbits gemeint, die von den Schatten des Pseudo–Orbits induziert werden.

Folgerung 2.19 *Hat man in der Situation des Shadowing–Lemmas $k = -\infty$, $l = \infty$, so ist x eindeutig definiert.*

Beweis: Angenommen, z sei ein zweiter Schatten. Dann ist

$$d(F^k x, F^k z) < d(F^k x, x_k) + d(x_k, F^k z) \leq 2\epsilon$$

für alle $k \in \mathbb{Z}$. Da F jedoch expansiv ist, muß für genügend kleines ϵ gelten $z = x$. $\square$

Die Menge aller möglichen Schatten für x_{-m}^m besteht also nur aus einem Punkt, falls $m = \infty$. Wie verändert sie sich mit m? Diese Frage ist wichtig für numerische Berechnungen von Schatten, da man stets nur endliche Pseudo–Orbits betrachten kann und natürlich trotzdem den Schatten x möglichst genau approximieren möchte. Man definiert hierzu

$$W_\epsilon^m(x) := \{y \in M, d(F^k x, F^k y) \leq \epsilon \, \forall k \in [-m, m]\}.$$

Lemma 2.20 *Sei $x \in M$ fest. Ist ϵ genügend klein, so gilt folgendes: Für jedes $\delta > 0$ existiert ein m_0, so daß $y \in W_\epsilon^{m_0}(x)$ impliziert $d(y, x) < \delta$.*

Beweis: Betrachte $B := \{y \in M; d(y, x) \geq \delta\}$. Die Mengen

$$B_m := \{y \in M; d(F^m y, F^m x) > \epsilon\}$$

sind offen und überdecken B (da F expansiv ist). B ist kompakt, deshalb existiert ein endliches m_0, so daß

$$B \subset \bigcup_{-m_0}^{m_0} B_m.$$

Das heißt: Für jedes $y \in B$ existiert ein $m \in [-m_0, m_0]$ mit $d(F^m y, F^m x) > \epsilon$. Deshalb ist $W_\epsilon^{m_0}(x) \subset M \setminus B = \{y \in M; d(y,x) < \delta\}$ □

Mit diesem Lemma können wir folgende Aussage beweisen:

Folgerung 2.21 *Angenommen $y_{-\infty}^{\infty}$ ist ein Pseudo–Orbit, x sein eindeutig bestimmter ϵ–Schatten und ϵ ist genügend klein. Dann gilt:*

Für jedes $\delta > 0$ existiert ein m_0, so daß für jeden Schatten z von $y_{m_0}^{m_0}$ gilt: $d(x,z) < \delta$.

Beweis: $d(F^k x, F^k z) \leq d(F^k x, y_k) + d(y_k, F^k z) < 2\epsilon$ für alle $k \in [-m_0, m_0]$. Deshalb ist $z \in W_{2\epsilon}^{m_0}(x)$. Ist ϵ genügend klein, so existiert nach dem vorangegangenen Lemma ein m_0 mit $d(x,z) < \delta$ □

Die vorangegangenen Aussagen und Beweise sind hier im Hinblick auf spätere Anwendungen vorgestellt worden. Der Zusammenhang der folgenden Aussagen mit dem eigentlichen Thema dieser Arbeit ist eher indirekt. Trotzdem sollen sie hier gebracht werden, um das allgemeine Bild der nichtlinearen Abbildungen abzurunden.

Definition 2.22

1. *Ein Punkt x heißt* nichtwandernd*, falls für jede offene nichtleere Umgebung U von x ein $n > 0$ existiert mit $U \cup F^n U \neq \emptyset$. Die Menge der nichtwandernden Punkte sei mit $\Omega(F)$ bezeichnet.*

2. *Eine Menge $T \subset M$ heißt* topologisch transitiv*, falls für zwei beliebige offene nichtleere Mengen U, V stets ein n existiert mit $U \cup F^n V \neq \emptyset$.*

3. *Eine Menge $T \subset M$ heißt* (topologisch) mischend*, falls für zwei beliebige offene nichtleere Mengen U, V stets ein N existiert mit $U \cup F^n V \neq \emptyset$ für alle $n \geq N$.*

Es gelten offenbar die Implikationen: T ist topologisch mischend $\Rightarrow$ T ist topologisch transitiv $\Rightarrow$ T besteht nur aus nichtwandernden Punkten. Für hyperbolische Abbildungen gilt für die Menge der nichtwandernden Punkte der folgende Zerlegungssatz:

Satz 2.23 (Spektralzerlegung)

1. *Die periodischen Orbits sind dicht in Ω.*

2. *Es existiert eine endliche Zerlegung $\Omega = \Omega_1 \cup \ldots \cup \Omega_s$ von paarweise disjunkten abgeschlossenen invarianten topologisch transitiven Mengen.*

Es läßt sich unter gewissen Voraussetzungen zeigen, daß eine dieser Mengen eine Umgebung besitzt, die vollständig aus stabilen Mannigfaltigkeiten gebildet wird, d.h. die Menge $\cup_{x \in \Omega_s} W_\epsilon^s(x)$ ist eine Umgebung von Ω_s. Eine solche Menge heißt ein *Attraktor*. Die in diesem Kapitel besprochenen Aussagen gelten in vielen Fällen nur für die Menge der nichtwandernden Punkte. Da die Spektralzerlegung der nichtwandernden Punkte immer einen Attraktor enthält, kann man davon ausgehen, daß physikalische Realität sich immer auf einem Attraktor abspielt (sofern überhaupt ein hyperbolisches System vorliegt). Alles andere ist also höchst untypisch. Deshalb stellt die Beschränkung auf die Menge der nichtwandernden Punkte keine wirkliche Einschränkung dar.

Motiviert durch diese Situation bei hyperbolischen Systemen kann man folgende allgemeine Definition für einen Attraktor geben:

Definition 2.24 *Sei F ein dynamisches System auf einer Mannigfaltigkeit M. Eine Teilmenge A von M heißt ein* Attraktor*, falls A abgeschlossen, invariant und topologisch transitiv ist und eine Umgebung U besitzt mit den Eigenschaften*

1. *$F(U) \subset U$.*
2. *Für jedes $x \in U$ gilt $d(F^n x, A) \to 0$ für $n \to \infty$.*

Ein Attraktor kann (ebenso wie eine stabile oder instabile Mannigfaltigkeit) eine geometrisch äußerst komplizierte Struktur haben. Dies wird im nächsten Kapitel etwas näher beleuchtet. Die Abbildung (2.2) zeigt den HENON–Attraktor, der später noch besprochen wird. Obwohl nachwievor unklar ist, ob es sich wirklich um einen Attraktor im Sinne der mathematischen Definition handelt, bildet der Henon–Attraktor das Objekt vieler Untersuchungen, da er durch eine äußerst einfache Abbildung zustande kommt:

$$(x_{n+1}, y_{n+1}) = (1 - 1.4 \cdot x_n^2 + 0.3 \cdot y_n, x_n)$$

Sie wird später wieder auftauchen.

2.3 Ergodentheorie differenzierbarer Abbildungen

Die Verallgemeinerung eines dynamischen Systems auf einer Mannigfaltigkeit ist aus der Sicht der Maßtheorie die *meßbare Transformation*, also eine meßbare Abbildung T des Wahrscheinlichkeitsraumes Ω in sich selbst.

Definition 2.25 *Eine meßbare Transformation T heißt* maßerhaltend *(bzw. das Maß μ heißt invariant unter T), wenn für alle meßbaren A gilt: $\mu(T^{-1}(A)) = \mu(A)$.*

Eines der ersten Resultate über das Langzeitverhalten von Trajektorien, die durch maßerhaltende Transformationen erzeugt wurden, war das sogenannte POINCARÉsche Rekurrenztheorem, welches besagt, daß eine Trajektorie, die in einer Menge A startet, fast sicher dahin zurückkehrt. Man verschärft das Rekurrenztheorem leicht zu der Aussage, daß fast jeder Punkt unendlich oft nach A zurückkehrt (man wendet das Theorem auf die Potenzen von T an). Dies führt zu dem Versuch, eine Art mittlere Aufenthaltsdauer $d_n(x) := \frac{1}{n} \sum_{i=0}^{n-1} \chi(T^n(x))$ einzuführen.

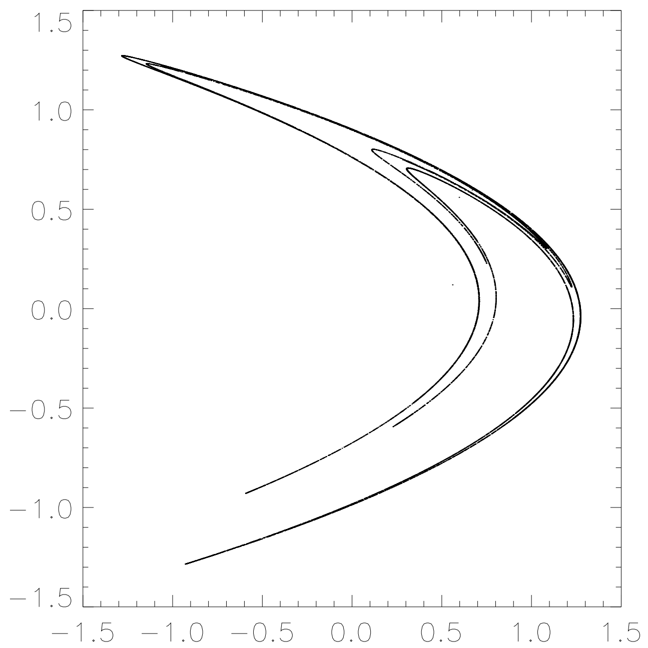


Abbildung 2.3: Henon–Attraktor

χ soll hierbei die charakteristische Funktion von A sein, also $\chi(x) = 1$ wenn $x \in A$ und sonst Null. Was ergibt sich nun im Limes n gegen unendlich? Man kann, falls die Folge überhaupt punktweise konvergiert, nicht erwarten, daß der Limes unabhängig von x ist, sich die Folge also überall mit asymptotisch gleicher Häufigkeit aufhält. In der Tat kann die Folge nämlich in invarianten Mengen hängen bleiben, also Mengen, für die $T^{-1}(A) = A$ gilt. Man kann also höchstens hoffen, daß die Funktion $d(x)$ nicht direkt von x, sondern nur von der invarianten Menge abhängt, in der x sich befindet. Als nächstes schließt sich natürlich die Hoffnung an, Mittelwerte von *beliebigen* Funktionen bilden zu können, also Ausdrücke wie $\frac{1}{N}\sum_{k=0}^{N-1} f(T^n(x))$ für beliebige meßbare f. Die invarianten Mengen bilden eine σ–Algebra $\mathcal{I}$, und tatsächlich konvergieren solche *Zeitmittel* gegen eine $\mathcal{I}$-invariante Funktion, genauer:

Theorem 2.26 (Birkhoff) *Sei T eine maßerhaltende Transformation und $\mathcal{I}$ die σ–Algebra der invarianten Mengen. Dann gilt für beliebiges meßbares f mit* $\mathrm{E}|f| < \infty$*:*

$$\frac{1}{N}\sum_{k=0}^{N-1} f(T^n(x)) \longrightarrow \mathrm{E}(f|\mathcal{I})(x)$$

fast überall.

Einige Bemerkungen zum Satz von Birkhoff:

1. Der ursprünglich wohl sehr lange Beweis wurde im Laufe der Zeit erheblich abgekürzt. Er stützt sich im wesentlichen auf das *Maximal Ergodic Theorem*, für das A. GARSIA 1965 einen elementaren und kurzen Beweis gefunden hat, den man mit dem Beweis des Satzes von Birkhoff in [4] findet.

2. Die Funktion $\mathrm{E}(f|\mathcal{I})$ ist der *bedingte Erwartungswert* von f bezüglich $\mathcal{I}$ (siehe hierzu [4]). $\mathrm{E}(f|\mathcal{I})$ ist $\mathcal{I}$-meßbar und es gilt für alle $D \in \mathcal{I}$: $\int_D f \, \mathrm{d}\mu = \int_D \mathrm{E}(f|\mathcal{I}) \, \mathrm{d}\mu$.

3. Man beachte, daß „fast überall“ vom Maß μ abhängt.

4. Es gibt Varianten, die die Konvergenz im p–ten Mittel statt der fast–überall–punktweise–Konvergenz sicherstellen. Für einen elementaren Beweis einer L_2–Version des Ergodensatzes siehe [13].

Eine besondere Situation liegt vor, wenn die invarianten Mengen stets das Maß Null oder Eins haben. Man nennt dann T *ergodisch (bezüglich μ)* bzw. μ ein *ergodisches Maß* für T. In diesem Fall gilt nämlich $\mathrm{E}(f|\mathcal{I}) = \mathrm{E}f$ und es ergibt sich die

Folgerung 2.27 *Ist μ ergodisch, so gilt*

$$\frac{1}{N} \sum_{k=0}^{N-1} f(T^n(x)) \longrightarrow \mathrm{E}(f) = \int f \, \mathrm{d}\mu$$

fast überall bezüglich μ.

In unserer Situation liegt ein dynamisches System vor, und es sei, wenn nichts gegenteiliges gesagt wird, wieder vorausgesetzt, daß die Mannigfaltigkeit M^n kompakt ist. Im letzten Kapitel wurde erläutert, daß man dann ein endliches Maß μ auf M^n hat, welches man sich als Verallgemeinerung des Lebesgue-Maßes vorstellen kann. Dieses Maß ist jedoch nicht invariant, geschweige denn ergodisch. Ein invariantes Maß kann man sich jedoch basteln: Man betrachtet die Maße $\mu_N(A) := \frac{1}{N} \sum_{k=0}^{N-1} \mu(T^{-k}A)$. Diese Folge konvergiert nicht. Allerdings existieren konvergente Teilfolgen, die dann, wie man leicht zeigt, invariante Maße darstellen.[6]

Das Auffinden von ergodischen Maßen ist weit schwieriger. Im letzten Abschnitt wurden die sogenannten Axiom–A–Diffeomorphismen und als Spezialfall davon die hyperbolischen Diffeomorphismen eingeführt, für die man ergodische

[6] Bevor man von konvergenten Teilfolgen reden kann, muß man erst einmal eine geeignete Topologie auf dem Raum der Maße einführen, die sogenannte *schwache Topologie*. Dann erhält man einen kompakten Raum (siehe [22]), und in solchen Räumen besitzt bekanntlich jede Folge eine konvergente Teilfolge.

Maße mit besonders schönen Eigenschaften (sogenannte GIBBS–Maße) finden kann (siehe hierzu [3]). Was nun nützt einem ein ergodisches Maß? Man kann, wie bereits bemerkt wurde, von Zeitmitteln zu *Scharmitteln* übergehen, also zum Erwartungswert. Andererseits muß man bedenken, daß das nicht überall, sondern nur fast überall gilt, d.h. bis auf Mengen vom Maß null. Diese Mengen können aber bezüglich des Lebesgue-Maßes sehr wohl ein von null verschiedenes Maß haben.

2.4 Entropie, Dimensionsbegriffe und Lyapunov–Exponenten

Im letzten Abschnitt wurde die Bedeutung von invarianten Maßen für meßbare Transformationen bzw. dynamische Systeme dargestellt. Es sollen nun einige Größen vorgestellt werden, die das Langzeitverhalten eines dynamischen Systems charakterisieren und natürlich vom Konzept der invarianten Maße Gebrauch machen.

Als erstes soll erläutert werden, daß ein solches System in gewisser Weise Information produzieren kann, die man ähnlich wie in der Informationstheorie mit Hilfe der *Entropie* messen kann. Hierzu stelle man sich folgende Situation vor: Ω sei eingeteilt in paarweise disjunkte meßbare Mengen C_j, und wir wollen einen Punkt $x \in M$ finden. Die Aussage „x liegt in C_3“ liefert uns eine Information, die monoton mit $\mu(C_3)$ fallen muß. Je kleiner die Wahrscheinlichkeit von $x \in C_3$ ist, desto größer der Informationsgewinn, wenn dies tatsächlich der Fall ist. Angenommen, zwei Mengen C_i, C_j sind unabhängig, dann ist $\mu(C_i \cap C_j) = \mu(C_i)\mu(C_j)$. Die Information sollte in diesem Fall additiv sein. Die Funktion $-\log(x)$ ist die einzige Funktion mit diesen Eigenschaften. In der Informationstheorie ist die Basis 2 populär, ansonsten wird jedoch allgemein e benutzt. Man definiert die Entropie $H(\mathcal{C})$ der Partition $\mathcal{C} = \{C_1, \ldots, C_n\}$ als die mittlere Information, die durch eine Antwort der Art „x liegt in C_3“ gegeben ist:

$$H(\mathcal{C}) = \sum_{i=1}^{n} -\mu(C_i)\log(\mu(C_i)).$$

Nun kommt das dynamische System T ins Spiel. Angenommen, man erhält auf der Suche nach x eine Antwort wie: „x ist zuerst in C_1, wird dann von T nach C_2 abgebildet, dann nach C_3,...und schließlich landet es in C_n“. Die Information dieser Aussage ist genau $-\log(\mu(C_1 \cap T^{-1}C_2 \cap \ldots \cap T^{-n+1}C_n)$. Wir führen deshalb folgende Schreibweise ein: Ist $\mathcal{C}$ eine endliche meßbare Partition paarweise disjunkter Mengen und $\mathcal{D}$ ebenfalls, so ist $\mathcal{C} \vee \mathcal{D} := \{C_i \cap D_j : C_i \in \mathcal{C}, D_j \in \mathcal{D}\}$ ebenfalls eine endliche meßbare Partition paarweise disjunkter Mengen. Nun kann man den folgenden Limes bilden (siehe [3], Lemma 1.19):

$$h(T, \mathcal{C}) := \lim_{m\to\infty} \frac{1}{m} H(\mathcal{C} \vee T^{-1}\mathcal{C} \vee \ldots \vee T^{-m+1}\mathcal{C})$$

Diese Größe hängt noch von der Partition ab. Zur Entropie gelangt man schließlich auf folgende Weise:

Definition 2.28 (Kolmogorov–Sinai–Entropie) *Die Größe*

$$h(T) := \sup_{\mathcal{C}} h(T, \mathcal{C}),$$

wobei das Supremum über alle endlichen Partitionen gebildet wird, heißt Entropie *von T (bezüglich μ).*

Die praktische Berechnung der Entropie nach den dieser Definition dürfte etwas schwer fallen. Die Situation vereinfacht sich jedoch erheblich, wenn man in einem kompakten metrischen Raum arbeitet, also z.B. auf einer Mannigfaltigkeit. Dann kann man den *Durchmesser* $\mathrm{diam}\mathcal{D} := \max\{|D_i|; D_i \in \mathcal{D}\}$[7] einer Partition definieren. Es gilt dann:

Satz 2.29 *Es sei M ein kompakter metrischer Raum, $T : M \to M$ stetig und μ ein T–invariantes Maß. Dann gilt:*

1. *Ist $\mathcal{D}_n$ eine Folge von Partitionen mit $\mathrm{diam}\mathcal{D}_n \to 0$, dann gilt*

$$h(T, \mathcal{C}_n) \to h(T) \text{ für } n \to \infty.$$

2. *Ist T expansiv mit Konstante ϵ, so gilt*

$$h(T, \mathcal{C}) = h(T), \text{ falls } \mathrm{diam}\mathcal{D} \leq \epsilon.$$

Beweis: Siehe [3]. Er beruht auf einigen Ungleichungen für $h(T, \mathcal{C})$ und $H(\mathcal{C})$ und der Tatsache, daß sich eine beliebige Partition $\mathcal{C}$ dadurch approximieren läßt, daß man eine Partition $\mathcal{D}$ mit kleinem Durchmesser wählt und dann zu geeigneten Vereinigungen übergeht □

Als zweites nun soll eine weitere „Kenngröße“ dynamischer Systeme besprochen werden, die mit der Geometrie der invarianten Mengen zusammenhängt, die, wie in Abschnitt 2.2 bereits angesprochen wurde, sehr kompliziert sein kann. Der erste Schritt zur Untersuchung geometrisch merkwürdiger Mengen war die Verallgemeinerung des Dimensionsbegriffs durch Felix Hausdorff[8]. Wir geben die

[7] Mit $|U|$ für eine Teilmenge U eines metrischen Raumes ist der Durchmesser von U gemeint.

[8] Carathéodory (1914) führte die äußeren Maße ein, die wir schon betrachtet haben. Er definierte 1-dimensionale Maße im euklidischen Raum, die sich auch auf n Dimensionen verallgemeinern ließen. Hausdorff (1919) erkannte, daß diese Definition auch für nichtganzzahliges n sinnvoll ist und zeigte, daß die bekannte Cantor-Menge ein endliches n–dimensionales Maß hat, wenn $n = \log 2 / \log 3$ (übernommen aus [10]).

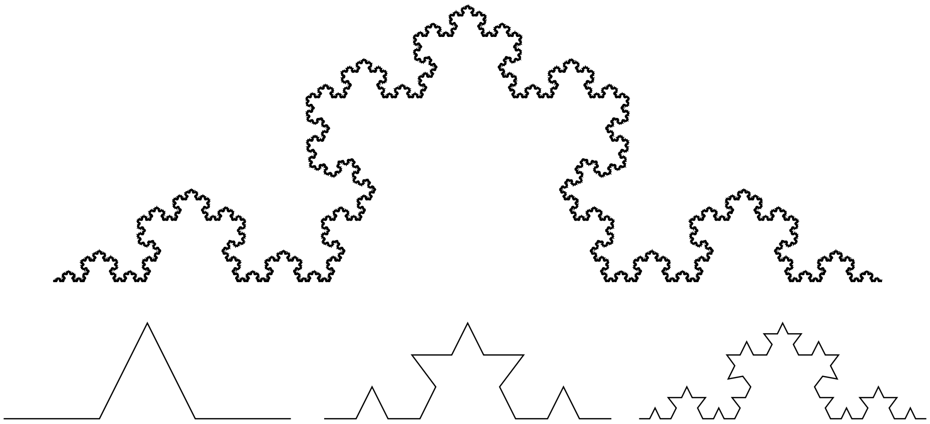


Abbildung 2.4: Die Kochkurve

Definition der Hausdorff–Dimension nach [10]: Sei A eine beliebige Teilmenge eines beliebigen metrischen Raumes M. Ein abzählbares System $\mathcal{D}$ von nichtleeren Teilmengen von M heißt ein δ–*Cover* von A, falls diam$\mathcal{D} \leq \delta$ und $A \subset \cup\mathcal{D}$. Sei nun $\delta > 0$ und $s \geq 0$. Dann definieren wir

$$\mathcal{H}^s_\delta(A) := \inf \sum_{U \in \mathcal{D}} |U|^s,$$

wobei das Infimum über alle δ–Cover von A gebildet wird. Das *äußere s-dimensionale Hausdorff–Maß* von A ist definiert als $\mathcal{H}^s(A) := \lim_{\delta \to 0} \mathcal{H}^s_\delta(A)$. Es ist leicht einzusehen, daß es sich dabei um ein metrisches äußeres Maß im Sinne von Lemma 1.2 handelt. Also ist das äußere Hausdorff–Maß ein richtiges Maß (eben das *s-dimensionale Hausdorff–Maß*) auf der σ–Algebra, die von den offenen Mengen erzeugt wird. $\mathcal{H}^s(A)$ ist eine nichtwachsende Funktion, wenn s von 0 nach ∞ läuft. Außerdem haben wir für $s > t$ folgende Ungleichung:

$$\sum_{i=1}^{\infty} |U_i|^s = \sum_{i=1}^{\infty} |U_i|^{s-t} |U_i|^t \leq \delta^{s-t} \sum_{i=1}^{\infty} |U_i|^t,$$

was die Ungleichung

$$\delta^{t-s} \mathcal{H}^s_\delta(A) \leq \mathcal{H}^t_\delta(A)$$

impliziert. Ist also $\mathcal{H}^s(A)$ positiv, so ist $\mathcal{H}^t(A) = \infty$ für alle $t < s$. Ist außerdem M n–dimensional, so ist $\mathcal{H}^s(A) = 0$ für $s > n$. Deshalb gilt folgende Aussage:

Lemma 2.30 *Es existiert genau eine Zahl* $\dim_H(A) \in \mathbb{R}_+$, *so daß* $\mathcal{H}^s(A) = 0$ *für* $s > \dim_H(A)$ *und* $\mathcal{H}^s(A) = \infty$ *für* $s < \dim_H(A)$. *Diese Zahl* $\dim_H(A)$ *heißt die* Hausdorff–Dimension *und* $\mathcal{H}^{\dim_H(A)}(A)$ *das* Hausdorff–Maß *von* A.

In [10] wird die Geometrie von Mengen mit ganzzahliger und nichtganzzahliger Hausdorff–Dimension untersucht, z.B. der in Abb. 2.4 gezeigten Koch-Kurve. Zitiert sei hier nur ein einfaches Lemma über das Verhalten von $\dim_H(A)$ unter Lipschitz-stetigen Abbildungen:

Lemma 2.31 *Sei F surjektiv und es existiere ein $C > 0$ mit*

$$d(F(x), F(y)) < C \cdot d(x, y)$$

für alle $x, y \in M$. Dann ist $\mathcal{H}^s(F(A)) \leq C^s \mathcal{H}^s(A)$.

Ist F diffeomorph und M eine kompakte Mannigfaltigkeit, so ist $\mathcal{H}^s(F(A)) = \mathcal{H}^s(A)$.

Die Hausdorff–Dimension ist im allgemeinen schwer zu berechnen, weshalb man sich in der Praxis oft mit leicht modifizierten Größen begnügen muß, für die man eine Abschätzung gegen die Hausdorff–Dimension kennt (oder vermutet). Der Zusammenhang mit dynamischen Systemen besteht über die *Informationsdimension*:

Definition 2.32 *Sei F ein dynamisches System und μ ein invariantes Maß. Die Größe*

$$\dim_H \mu := \inf_{A \subset M; \mu(A)=1} \dim_H(A)$$

heißt Informationsdimension *von (F, μ).*

Diese Größe ist natürlich im allgemeinen sehr schwer zu handhaben. Eine Näherung liefert der Algorithmus von Grassberger und Procaccia (siehe [9]).

Als drittes wird hier das Konzept der Lyapunov–Exponenten vorgestellt. Sie hängen mit der Frage zusammen, wie sich eine unendlich kleine Störung des Systemzustandes mit der Zeit entwickelt. M. a. W., wir betrachten für ein $u \in T_x M$ die Größe

$$\delta_n(x, u) := \sqrt[n]{\|\mathrm{D}F^n(x)u\|}$$

in der Annahme, daß die erwartete Expansion bzw. Kontraktion mit einem Potenzgesetz verläuft. Wir definieren, falls der Grenzwert existiert:

Definition 2.33 *Der Limes*

$$\lambda(x, u) := \lim_{n \to \infty} \log \delta_n(x, u) = \lim_{n \to \infty} \frac{1}{n} \log \|\mathrm{D}F^n(x)u\|$$

heißt Lyapunov-Exponent *von F an der Stelle x in Richtung u.*

Man kann nun die Matrix $\Delta^n_{i,j}(x) := \langle \mathrm{D}F^n(x)e_i, \mathrm{D}F^n(x)e_j\rangle_x$ betrachten. Angenommen, es existiert der Limes

$$\Lambda_x := \lim_{n\to\infty} \sqrt[2n]{\Delta^n(x)},$$

dann kann man die Eigenwerte $\mu_i(x)$ und Eigenvektoren $v_i(x)$ von Λ_x berechnen und bekommt

$$\lambda(x, v_i) = \log \mu_i(x)$$

woran man sieht, daß eine orthogonale Zerlegung $T_xM = \oplus E^i_x$ von Unterräumen existiert, auf denen $\lambda(x, v_i) = \log \mu_i(x)$ konstant ist, nämlich gerade die Eigenräume von Λ_x.

Außerdem stellt man fest, daß die Exponenten invariant sind gegenüber der Abbildung $(x, u) \to (F(x), \mathrm{D}F(x)u)$. Das gilt auch für die orthogonale Zerlegung, d.h es gilt $\mathrm{D}F(x)E^i_x = E^i_F x$. Punkte in denen die besprochene Zerlegung sowie die Lyapunov–Exponenten existieren, heißen *regulär*. Die zentrale Aussage über reguläre Punkte ist das folgende

Theorem 2.34 (Oseledec [25], [9]) *Angenommen, es existiert ein invariantes Maß $\mu(x)$ für F. Dann sind fast alle x regulär.*

Da $\lambda(x, v)$ wie gesagt invariant ist, ist für ein ergodisches Maß diese Größe für fast alle x konstant.

2.5 Einbettungen

In der praktischen Untersuchung dynamische Systeme ist man letztendlich an Zahlen interessiert. Zahlen, die den Systemzustand wiedergeben oder Größen, die sich in Zahlen ausdrücken lassen und in irgendeiner Weise das System charakterisieren, z.B. Dimensionen und Lyapunov–Exponenten. Wir haben gesehen, daß diese Größen einerseits unabhängig von einer speziellen Einbettung sind, andererseits für ihre praktische Berechnung eine Einbettung mit gewissen Regularitätseigenschaften benötigt wird. Auch in theoretischen Betrachtungen über dynamische Systeme und Mannigfaltigkeiten können geeignete Einbettungen die Argumentation wesentlich vereinfachen. Zuerst wird geklärt, welchen Regularitätseigenschaften eine Einbettung genügen sollte. Dann werden die wichtigsten Einbettungssätze zitiert.

Es sei M^n eine n–dimensionale differenzierbare unberandete Mannigfaltigkeit und $\Phi : M^n \to \mathbb{R}^d$ eine differenzierbare Abbildung mit Differential $\mathrm{D}\Phi$, welches in jedem Punkt eine lineare Abbildung des Tangentialraumes nach $\mathbb{R}^d$ darstellt. Angenommen, die Sphäre (Abb. 2.5) soll eingebettet werden. Weder die Figur 2 noch die Figur 3 sind als gute Einbettungen anzusehen: Figur 2 weist eine Selbstdurchdringung auf (Φ ist nicht injektiv), und Figur 3 hat eine Spitze, also der

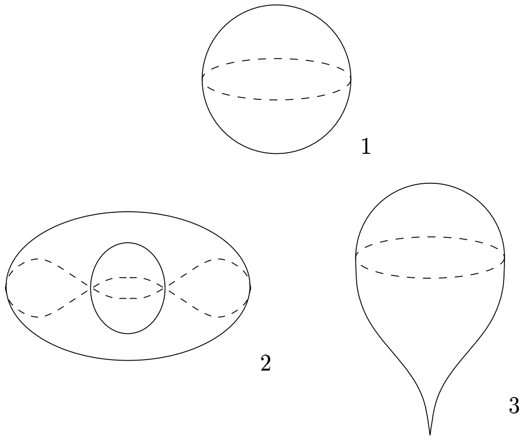


Abbildung 2.5:

Tangentialraum dort hat Dimension Null. DΦ hat dort offenbar zu kleinen Rang. Einen etwas verwickelteren Fall soll Abb. 2.5 darstellen: Die Mannigfaltigkeit M wird durch Φ gefaltet. Denkt man sich den Streifen M nach links und rechts unendlich fortgesetzt, so kann man eine unendlich dicht zusammengefaltete Struktur $\Phi(M)$ erzeugen. Punkte, die ursprünglich auf M getrennt lagerten, werden durch Φ beliebig dicht zusammengebracht, m. a. W. die Umkehrabbildung Φ^{-1} ist unstetig. Die beobachteten Phänomene fassen wir zusammen:

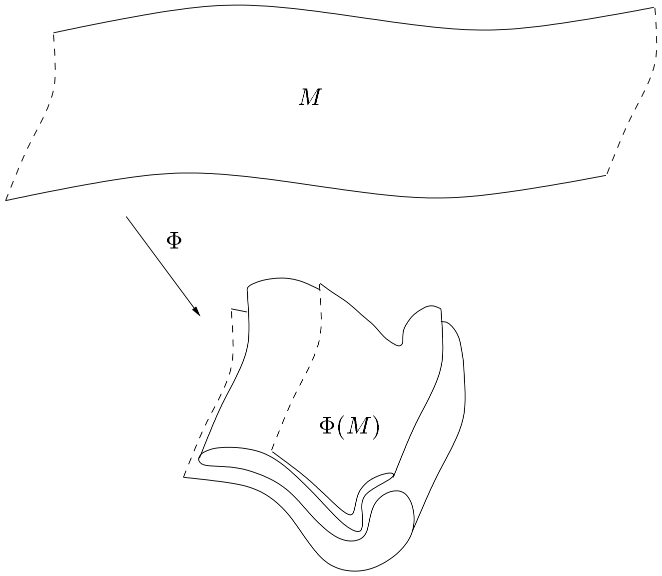


Abbildung 2.6:

Definition 2.35 *Sei M^n eine n–dimensionale differenzierbare Mannigfaltigkeit und $\Phi : M^n \to \mathbb{R}^d$ eine differenzierbare Abbildung.*

1. *Φ heißt* Immersion, *wenn* $\mathrm{D}\Phi$ *überall injektiv ist (den Rang n hat).*

2. *Φ heißt* Einbettung, *wenn Φ eine injektive Immersion mit stetiger Umkehrabbildung ist (Φ ist ein Homöomorphismus auf sein Bild).*

Eine differenzierbare Abbildung muß keine Immersion, eine Immersion nicht injektiv und eine injektive Immersion keine Einbettung sein. Ist jedoch M kompakt, so ist jede bijektive stetige Abbildung automatisch ein Homöomorphismus. Deshalb ist in dieser Situation eine injektive Immersion automatisch eine Einbettung. Die guten Eigenschaften von Einbettungen stellt folgender Satz klar:

Satz 2.36 *Das Bild einer Einbettung ist eine differenzierbare Untermannigfaltigkeit des $\mathbb{R}^d$.*

Die Frage, ob es immer Einbettungen gibt, kann im wesentlichen positiv beantwortet werden. Es gelten nämlich die folgenden Aussagen:

Theorem 2.37 (Immersionssätze von Whitney) *Sei M^n eine n–dimensionale differenzierbare Mannigfaltigkeit. Weiter sei $C^\infty(M^n, \mathbb{R}^d)$ der Raum der differenzierbaren Abbildungen von M^n nach $\mathbb{R}^d$ ausgestattet mit der C^0-Topologie*[9] *und $\Phi \in C^\infty(M^n, \mathbb{R}^d)$.*

1. *Ist $\mathrm{d} \geq 2 \cdot n$, so sind die Immersionen dicht in $C^\infty(M^n, \mathbb{R}^d)$. Ist M kompakt, so ist Φ typischerweise eine Immersion.*

2. *Ist $\mathrm{d} > 2 \cdot n$, so sind die injektiven Immersionen dicht in $C^\infty(M^n, \mathbb{R}^d)$. Ist M kompakt, so ist Φ typischerweise eine injektive Immersion.*

Beweis: Siehe [6] □

Der Begriff *typischerweise* ist hier folgendermaßen gemeint:

Definition 2.38 *Sei T ein topologischer Raum und A eine Teilmenge. Man nennt die Elemente $x \in A$* typische Elemente *von T, wenn A offen und dicht in T ist. Die Eigenschaft „$x \in A$“ nennt man auch eine* typische *oder* generische *Eigenschaft der Elemente von T.*

Typische Eigenschaften sind unempfindlich gegen kleine Störungen, und einem beliebigen $x \in T$ kann durch beliebig kleine Störungen eine typische Eigenschaft gegeben werden. Ab und an wird „generisch“ mit „fast sicher“ gleichgesetzt. Das ist schon deshalb sinnlos, weil auf T kein Maß existiert. Außerdem konstruiert man schon auf $\mathbb{R}$ Mengen mit beliebig kleinem Lebesgue-Maß, deren Punkte generisch sind, zum Beispiel (nach [5]): Betrachtet man das Einheitsintervall $[0, 1]$, so findet man sicher eine Abzählung p_j der rationalen Punkte. Um jeden dieser Punkte legt man ein Intervall $I_j \subset [0, 1]$ mit dem Maß $\mu(I_j) \leq \epsilon \cdot 1/2^j$ und betrachte

[9] Siehe [6]. Diese Topologie beschreibt im wesentlichen die Konvergenz der Funktionswerte.

$I := \bigcup_j I_j$. Diese Menge ist offen und dicht, also ihre Punkte generisch. Aber $\mu(I) = \mu(\bigcup_j I_j) \leq \sum_j \mu(I_j) = \epsilon$, also beliebig klein.

Im Experiment ist man darauf angewiesen, eine solche Einbettung zu konstruieren. Wie gesagt ist jedoch der Systemzustand im allgemeinen kein Vektor. Vielmehr mißt man meistens eine skalare Funktion λ des Systemzustandes. 1980 wurden von Packard et al in [26] (auf Vorschlag von Ruelle) zur Rekonstruktion eines Rössler-Oszillators die Methode der *Delay-Koordinaten* vorgeschlagen: Man bildet für ein dynamisches System F und eine skalare Funktion $\lambda : M \to \mathbb{R}$ die *Delay-Einbettung*

$$\Phi_{F,\lambda,\Delta} : M \to \mathbb{R}^d;\ x \to \big(\lambda(x), \lambda(F^{\Delta}x), \ldots, \lambda(F^{\Delta\cdot(d-1)}x)\big)$$

für ein geeignetes d und Δ. Δ heißt das *Delay* der Einbettung. Daß es sich dabei typischerweise um eine Einbettung handelt, konnte Takens 1981 zeigen:

Theorem 2.39 (Takens Delay–Embedding–Theorem, [31]) *Sei M kompakt und $d > 2 \cdot n$. Dann ist für typisches $(F, \lambda) \in C^\infty(M) \times C^\infty(M, \mathbb{R})$ die Delay-Einbettung $\Phi_{F,\lambda,\Delta}$ eine Einbettung im Sinne von Definition 2.35.*

Beweis: siehe [31]. □

3 Filterung gestörter Signale

In diesem Kapitel werden zuerst die wesentlichen Formeln zur Reduktion von Rauschen, also zum Aufbau von Schätzern für den Systemzustand gegeben. Wenn auch in der Praxis vielerlei Konzessionen an die jeweiligen Umstände gemacht werden müssen, so lassen sich daraus zumindest theoretisch optimale Schätzer finden. Dann wird kurz der lineare Kalman–Filter hergeleitet und besprochen. Im dritten Abschnitt geht es um einen grundsätzlichen Näherungsansatz, bei dem man die entsprechenden Verteilungen durch Gauß–Verteilungen approximiert. Im letzten Abschnitt schließlich werden einige bereits existierende Implementationen besprochen und die Ergebnisse einer eigenen Implementation dargestellt.

3.1 Die Bayes–Filterformel

In dieser Arbeit geht es um eine spezielle Klasse von stochastischen Prozessen, die nun definiert werden soll. Sei $F : M \to M$ ein dynamisches System und $G : \mathbb{M} \to \mathbb{R}^d$ eine differenzierbare Abbildung, die wir als eine Messung der Größe G auffassen. Für einen Orbit x_k^n messen wir die gestörten Größen $y_n = G(x_n) + s_n$. Die Problemstellung besteht nun darin, von y_n auf x_n zu schließen. Das wird später noch präzisiert. Bevor es weitergeht noch eine Bemerkung zu „dynamischem Rauschen“. Darunter versteht man ein Rauschen, welches die Dynamik stört, also ein Prozeß wie

$$x_{n+1} = F(x_n) + r_n.$$

Auf dem $\mathbb{R}^d$ induziert ein solches System eine *Übergangswahrscheinlichkeit*

$$\begin{aligned} p_F(x,y) &:= p(x_{n+1}=x|x_n=y) \\ &= \frac{1}{\sqrt{(2\pi)^n \det R}}\, e^{-\frac{1}{2}(x-F(y))R^{-1}(x-F(y))}, \end{aligned} \tag{3.1}$$

wobei R die Kovarianzmatrix von r_n ist. Aus dieser Übergangswahrscheinlichkeit und der Anfangsverteilung $p_0(x)$ lassen sich sämtliche Verteilungen $p_n(x)$ für die x_n realisieren:

$$p_{n+1}(x) = \int p_F(x,y) \cdot p_n(y) \cdot \,\mathrm{d}y.$$

Der Operator $\mathcal{D}p(x) := \int p_F(x|y) \cdot p(y) \cdot \mathrm{d}y$ soll im folgenden *Vorwärtsoperator* heißen. Einen solchen Zufallsprozess nennt man MARKOV–Prozess. Es ist nicht ganz offensichtlich, wie man einen solches dynamisches Rauschen für ein dynamisches System auf einer Mannigfaltigkeit verallgemeinern soll, da man dort keine Addition hat. Die Betrachtungen dieses Abschnitts hängen jedoch nicht von den dynamischen Gleichungen ab, sondern lediglich davon, daß es sich bei den x_n um einen Markov–Prozess handelt. Um ein allgemeines System auf einer Mannigfaltigkeit mit dynamischem Rauschen einzuführen, muß man also seinen Vorwärtsoperator angeben. Die grundsätzlichen Formeln, die später hergeleitet werden, können mit Hilfe des Vorwärtsoperators ausgedrückt werden und sind deshalb allgemein gültig. Numerische Berechnungen mit dynamischem Rauschen wurden für diese Arbeit jedoch nur an Systemen auf dem $\mathbb{R}^d$ durchgeführt, bei denen die Übergangswahrscheinlichkeit die explizite Form (3.1) hat. Bei anderen Berechnungen wurden die Eigenschaften von Abbildungen auf Mannigfaltigkeiten ausgenutzt und der Einfachheit halber dynamisches Rauschen nicht betrachtet. In diesem Falle degeneriert die Übergangswahrscheinlichkeit einfach zur Delta–Funktion $p_F(x,y) = \delta(x - F(y))$. Wir treffen nun ein für allemal folgende

Generalvoraussetzung 3.2 *Sei x_k^n ein Markov–Prozeß mit Wahrscheinlichkeitsdichten p_n induziert durch eine Anfangswahrscheinlichkeit p_0 und eine Übergangswahrscheinlichkeit $p(x,y)$.*[1] *Weiter sei ein Messvorgang durch*

$$y_n = G(x_n) + s_n \tag{3.3}$$

gegeben, wobei $G : M \to \mathbb{R}^d$ differenzierbar ist. s_n ist ein stochastischer Prozeß bestehend aus unabhängigen, Gauß–verteilten Zufallsvariablen mit $\mathrm{E}(s_n) = 0$, $\mathrm{Cov}(s_n) =: S_n$, wobei außerdem s_n unabhängig von x_n ist. y_n heißt im folgenden Observable.

Es wird in diesem Kapitel folgende Notation benutzt: Mit $y_k^n := \{y_i\}_{i=k,\ldots,n}$ bezeichnen wir eine Folge von Observablen von k bis n und sprechen von d-dimensionalen Zeitreihen. Ist $d = 1$, so heißt y_k^n auch einfach Zeitreihe. Ist $k = 0$, so schreiben wir y^n.

Das Ziel ist nun, aus dem Prozeß y_n, der als bekannt vorausgesetzt wird, Aussagen über den unbekannten Prozeß x_n zu gewinnen.

Angenommen wir wollen auf der Basis einer Folge y_k^n den Zustand x_t schätzen. Die Philosophie der Bayes-Strategie ist, wie bereits erläutert, die bedingte Wahrscheinlichkeitsdichte

$$p(x_t|y_k^n)$$

zu betrachten und daraus geeignete Schätzer zu konstruieren. In dieser Arbeit werden zwei Schätzer betrachtet:

[1] Oder $x_k^n = (x_k, \ldots, x_n)$ sei eine Folge von Punkten auf einer Mannigfaltigkeit M induziert durch ein dynamisches System $F : M \to M$, also $x_{n+1} = F(x_n)$. Das ist wie gesagt ein Spezialfall eines Markov–Prozesses.

Maximum-Likelyhood-Schätzer Dieser Schätzer ist gegeben durch $\hat{x}_t$ mit der Eigenschaft: $p(x_t = x | y_k^n)$ ist maximal für $x = \hat{x}_t$.

Bedingter Erwartungswert Dieser Schätzer ist gegeben durch

$$\hat{x}_t = \mathrm{E}(x_t | y_k^n) = \int x \cdot p(x_t = x | y_k^n) \, \mathrm{d}x.$$

Von Filterung (*tracking*) spricht man eigentlich nur, wenn $t = n$ ist. Der Fall $t < n$ heißt *Interpolation*, der Fall $t > n$ heißt *Vorhersage*. Eine scharfe Trennung werden wir in dieser Arbeit jedoch nicht vornehmen. Sei o.B.d.A. $k = 0$.

Zuerst werden folgende Abkürzungen eingeführt:

$$\begin{aligned} p_{t,n} &:= p(x_t | y^n) \\ p_t &:= p_{t,t} \\ q_{t,n} &:= p(y_t^n | x_t). \end{aligned}$$

Darüberhinaus erkennt man unmittelbar die Formel

$$p(y_t | x_t) = \frac{1}{\sqrt{(2\pi)^n \det S_t}} \, e^{-\frac{1}{2}(y_t - G(x_t)) S_t^{-1} (y_t - G(x_t))}. \tag{3.4}$$

Als Anfangsverteilung für x setzen wir von nun an stets

$$p_0(x) = p(x_0 = x) = p(y_0 | x_0).$$

Nun wird eine Formel etabliert, mit der man $p(x_t | y^t)$ aus $p(x_{t-1} | y^{t-1})$ berechnen kann. Im wesentlichen wird hier Satz 1.8 verwendet. Nach der Bayes–Formel ergibt sich

$$p(x_t | y^t) = \frac{p(x_t, y_t | y^{t-1})}{p(y_t | y^{t-1})}.$$

Nun ist

$$p(y_t | y^{t-1}) = \int p(x_t, y_t | y^{t-1}) \, \mathrm{d}x_t,$$

also der Nenner ist gleich einem gewissen Integral über den Zähler. Der Zähler schreibt sich unter Benutzung der Bayes–Kettenregel als

$$\begin{aligned} \text{Zähler} &= \int p(y_t, x_t, x_{t-1} | y^{t-1}) \, \mathrm{d}x_{t-1} \\ &= \int p(y_t | x_t, x_{t-1}, y^{t-1}) p(x_t | x_{t-1}, y^{t-1}) p(x_{t-1} | y^{t-1}) \, \mathrm{d}x_{t-1}. \end{aligned}$$

Der erste Term im Integranden kann geschrieben werden als $p(y_t | x_t)$, da bei gegebenem x_t die Größe y_t nicht mehr von x_{t-1}, y^{t-1} abhängt. Mit einem ähnlichen

Argument schreibt man den zweiten Term als $p(x_t|x_{t-1})$, d.h. als Übergangswahrscheinlichkeit. Deshalb gilt[2]

$$p(x_t|y^t) = \frac{p(y_t|x_t)\int p(x_t|x_{t-1})p(x_{t-1}|y^{t-1})\,\mathrm{d}x_{t-1}}{\int [\text{Zähler}]\,\mathrm{d}x_t}. \tag{3.5}$$

Unter Verwendung des Vorwärtsoperators und der Abkürzungen ergibt sich

$$p_t(x) = \alpha \cdot p(y_t|x_t{=}x) \cdot \mathcal{D}p_{t-1}(x)$$

wobei der Faktor α hier und im folgenden so zu wählen ist, daß $\int p_t(x)\,\mathrm{d}x = 1$ ist.

Mittels $p_t(x)$ läßt sich $p_{t,n}(x)$ wie folgt berechnen:

$$\begin{aligned} p_{t,n}(x) &= p(x_t|y^n) \\ &= p(x_t|y^t, y_{t+1}^n) \\ &= \alpha \cdot p(y_{t+1}^n|x_t) \cdot p(x_t|y^t) \\ &= \alpha \cdot p(y_{t+1}^n|x_t) \cdot p_t(x) \end{aligned}$$

Der erste Term rechter Hand schreibt sich mit Verwendung der Übergangswahrscheinlichkeit als

$$p(y_{t+1}^n|x_t) = \int p(y_{t+1}^n|x_{t+1}) \cdot p(x_{t+1}|x_t)\,\mathrm{d}x_{t+1}.$$

Der Operator $\mathcal{R}p(x) := \int p(y) \cdot p_F(y,x)\,\mathrm{d}y$ heißt *Rückwärtsoperator*. Wir können nun unter Verwendung des Rückwärtsoperators schreiben

$$p(y_{t+1}^n|x_t) = \mathcal{R}q_{t+1,n}(x).$$

Als letzte sieht man unmittelbar die Beziehung

$$\begin{aligned} p(y_t^n|x_t) &= p(y_{t+1}^n|x_t, y_t) \cdot p(y_t|x_t) \\ &= p(y_{t+1}^n|x_t) \cdot p(y_t|x_t), \end{aligned}$$

die man in die beiden oben hergeleiteten Formel einsetzen kann. Zusammenfassend kann man diese Formeln schreiben als

$$p_t(x) = \alpha_1 \cdot p(y_t|x_t) \cdot \mathcal{D}p_{t-1}(x), \tag{3.6}$$

$$p_{t,n}(x) = \alpha_2 \cdot \frac{q_{t,n} \cdot p_t(x)}{p(y_t|x_t)} \tag{3.7}$$

$$q_{t,n}(x) = p(y_t|x_t) \cdot \mathcal{R}q_{t+1,n}(x). \tag{3.8}$$

Hier sind α_1 und α_2 wieder Normierungskonstanten. Mit diesen Formeln, die im folgenden *Bayes–Filterformeln* heißen sollen, läßt sich die Bestimmung von

[2] Die Formel 3.6 wird in [1] und in [21] angegeben.

$p(x_{t-1}|y_k^{n-1})$ zurückführen auf die Bestimmung des Vorwärts– und Rückwärts-operators, also auf die Bestimmung der Übergangswahrscheinlichkeit (3.1) und der *Meßwahrscheinlichkeit* (3.4). Die Zahl $n-t$ heißt *Akausalität* des Filters.

Für chaotische Systeme jedoch wird $p_{t,n}$ mit steigendem t und n eine äußerst irreguläre Funktion, was direkte numerische Auswertung problematisch macht. Abb. 3.1 stellt $\log p(x_t|y^t)$ für eine typische verrauschte (18 dB) Trajektorie des Hénon-Systems dar. Der unverrauschte Punkt liegt in der Mitte des Bildes, t läuft in 2er–Schritten von 2 bis 8. Für $t=2$ ist noch erkennbar, daß es sich einst um eine Gauß–Verteilung handelte. Für größere t ist gut zu erkennen, daß die Funktion sehr isolierte Peaks entwickelt. Für $t=8$ sieht man eine Reihe von Peaks entlang eines sehr dünnen Streifens. Es ist kaum zu erwarten, daß mit Hilfe dieser Funktion eine Schätzung des Systemzustandes (der ja in der Mitte des Bildes liegt) möglich ist. Eine Maximum–Likelyhood–Schätzung ist problematisch, da es mehrere Maxima gibt. Zur Bestimmung des bedingten Erwartungswertes wäre eine Integration notwendig, was bei dieser Funktion offensichtlich auch nicht einfach ist.

3.2 Der Kalman–Filter

Der Klassiker unter den Filtern ist der lineare Kalman–Filter. Es handelt sich um eine Anwendung der Formel (3.6) auf lineare Systeme. Wir betrachten den Fall $t=n$ (in der Notation des vorigen Kapitels), also tatsächlich ein Filterproblem. Als dynamisches System haben wir

$$x_{n+1} = F_n x_n + r_n,$$

F_n linear, mit der Übergangswahrscheinlichkeit

$$p_F(x_{n+1}, x_n) = \frac{1}{\sqrt{(2\pi)^n \det R}} e^{-\frac{1}{2}(F_n x_n - x_{n+1}) R^{-1} (F_n x_n - x_{n+1})}$$

Als Observable haben wir

$$y_n = G x_n + s_n$$

mit linearem G und der Meßwahrscheinlichkeit

$$p(y_n|x_n) = \frac{1}{\sqrt{(2\pi)^n \det S}} e^{-\frac{1}{2}(y_n - G x_n) S^{-1} (y_n - G x_n)}.$$

Man beachte, daß wir im Gegensatz zu unserer Generalvoraussetzung hier nicht-autonome Systeme zulassen, was aber lediglich zur Folge hat, daß die Meßwahrscheinlichkeit und die Übergangswahrscheinlichkeit zeitabhängig werden. Angenommen, wir haben $p(x_n|y^n)$ schon berechnet mit

$$p(x_n|y^n) = \frac{1}{\sqrt{(2\pi)^n \det \Gamma_n}} e^{-\frac{1}{2}(x_n - \hat{x}_n) \Gamma_n^{-1} (x_n - \hat{x}_n)},$$

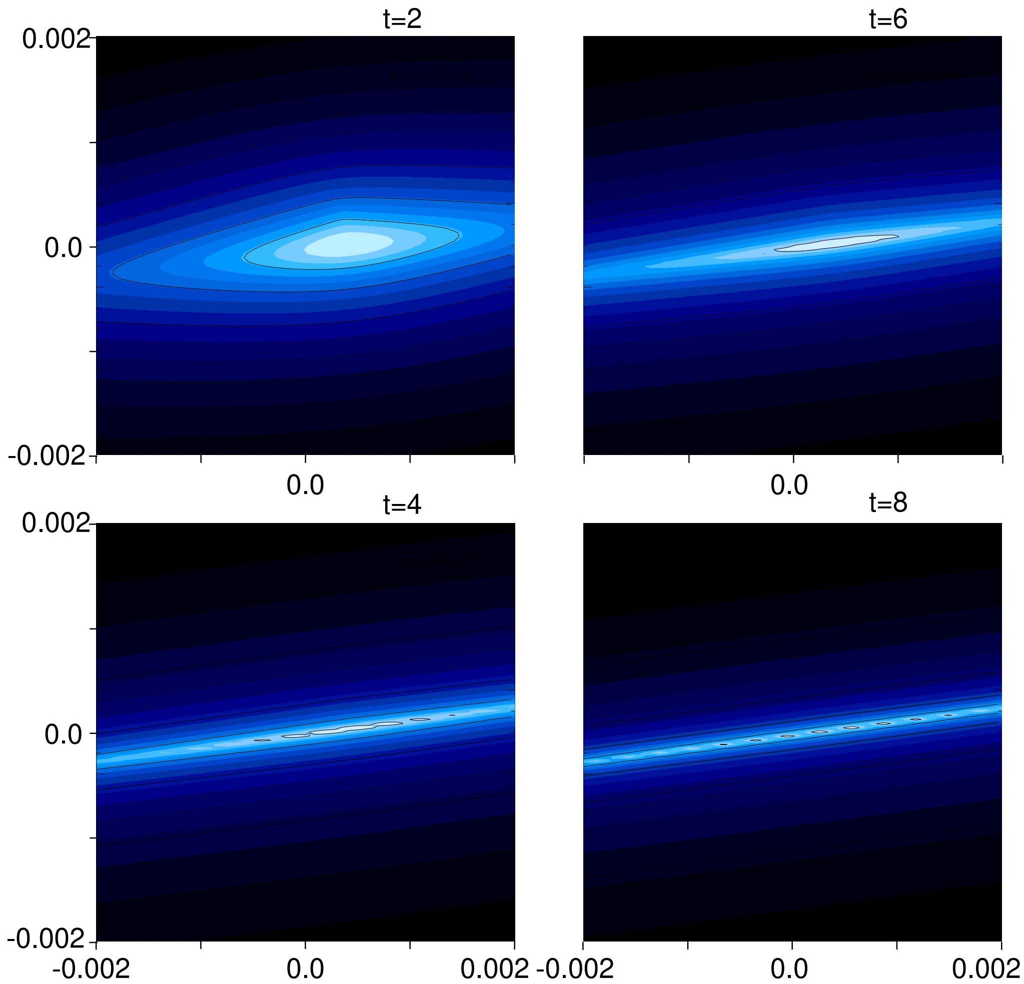


Abbildung 3.1: $\log p(x_t|y^t)$ für $t = 2$ bis 8 in 2er–Schritten

so überzeugt man sich leicht davon, daß die Bayes–Filterformel ausschließlich wieder Gauß–Verteilungen liefert. Es reicht also aus, Iterationsgleichungen für $\Gamma_n, \hat{x}_n$ anzugeben. Dazu setzt man $p(x_n|y^n)$ in die Bayes–Filterformel (3.6) ein. Mit Hilfe der Formeln (1.6) und etwas Algebra überzeugt man sich von den folgenden Identitäten:

$$\hat{x}_{n+1} = F_n\hat{x}_n + \Gamma_{n+1}G^t_{n+1}S^{-1}(y_{n+1} - G_{n+1}F_n\hat{x}_n) \tag{3.9}$$

$$\Gamma^{-1}_{n+1} = (F_n\Gamma_nF^t_n + R)^{-1} + G^t_{n+1}S^{-1}G_{n+1}. \tag{3.10}$$

Einige Bemerkungen zu diesen Formeln:

1. Γ_n hängt nur von den Systemgleichungen, nicht jedoch von den y^n ab. Deshalb kann diese Matrix „offline“ berechnet werden, kann also sozusagen bei Einsatz des Filters schon bereitstehen.

2. $\hat{x}_{n+1}$ hängt explizit nur von y_{n+1} ab und nicht von der gesamten Vorgeschichte. Der Filter ist also ein dynamisches System, welches immer nur mit dem aktuellen Meßwert gefüttert wird.

3. Die Formeln haben die Struktur eines Prädiktor-Korrektor-Verfahrens: Vermittels $\hat{x}_n$ wird y_{n+1} vorhergesagt. Diese Größe wird aber auch direkt gemessen und die Differenz (der Vorhersagefehler) geeignet verstärkt. Dieser Wert liefert dann eine Korrektur zur Vorhersage von $\hat{x}_{n+1}$.

4. Die Gleichungen bleiben richtig, wenn sowohl R als auch S von n abhängen. Darüber hinaus kann man auch inhomogene lineare Systeme betrachten. Hat man die Systemgleichungen

$$\begin{aligned}\tilde{F}_n(x) &= F_n x + f_n,\\ \tilde{G}_n(x) &= G_n x + g_n,\end{aligned}$$

so muß man einfach (3.9) ersetzen durch

$$\hat{x}_{n+1} = \tilde{F}_n(\hat{x}_n) + \Gamma_{n+1} G_{n+1}^t S^{-1} \left(y_{n+1} - \tilde{G}_{n+1}(\tilde{F}_n(\hat{x}_n)) \right).$$

5. Der Maximum-Likelyhood-Schätzer und der bedingte Erwartungswert sind beim Kalman–Filter gerade gleich, da es sich um Gauß–Verteilungen handelt. Beide sind durch (3.9) gegeben.

Das Studium des Kalman–Filters ist im wesentlichen ein Studium der Gleichung (3.10), die man unter Benutzung verschiedener Matrix-Identitäten auf andere Form bringen kann. Ihre Stabilitätseigenschaften, asymptotische Lösungen etc. werden in vielen Büchern untersucht, z.B. [7].

3.3 Entwicklung des Vorwärts– und Rückwärtsoperators

Es existieren viele Verfahren, den Kalman–Filter, der in der Praxis größte Relevanz besitzt, auf nichtlineare Probleme zu erweitern. Dabei soll jedoch zunächst die Beschreibung der Wahrscheinlichkeiten weiterhin ausschließlich durch die Kovarianzmatrix Γ und den Mittelwert $\hat{x}$ erfolgen. Allein dies stellt schon eine Näherung dar, da diese beiden Größen zur Beschreibung einer nicht Gauß–förmigen Verteilung im allgemeinen nicht ausreichen. Man wird also bei jedem Schritt Information verlieren, was sich natürlich auch auf die Genauigkeit von $\hat{x}, \Gamma$ auswirkt.

Ein allgemeineres Prinzip, mit Hilfe von Kovarianz und Mittelwert einen Filter zu konstruieren, kann man etwa folgendermaßen beschreiben: Es sei $p(x_t = x | y^n)$ schon berechnet. Bei Anwendung der Bayes-Filtergleichung (3.6) wendet man den

Vorwärtsoperator auf p_t an. Man geht nun einfach davon aus, das Resultat sei eine Gauß–Verteilung, also

$$\begin{aligned} p(x_t = x|y^{t-1}) &= \mathcal{D}p(x_{t-1} = x|y^{t-1}) \\ &\cong \frac{1}{\sqrt{(2\pi)^n \det \Gamma_{t|t-1}}} e^{-\frac{1}{2}(x-\hat{x}_{t|t-1})\Gamma^{-1}_{t|t-1}(x-\hat{x}_{t|t-1})}. \end{aligned} \tag{3.11}$$

Diesen Ansatz will ich im folgenden den *allgemeinen Kalman–Filter* nennen. Die Notation $\hat{x}_{t+1|t}, \Gamma_{t+1|t}$ soll andeuten, daß es sich um die Vorhersage der Kovarianz und des Mittelwertes handelt. Hierzu betrachten wir folgendes Problem: Sei $\mathcal{D}$ ein Vorwärtsoperator definiert durch eine Übergangswahrscheinlichkeit wie in (3.1) und $p(x)$ eine beliebige Wahrscheinlichkeitsverteilung. Wir suchen eine geeignete Näherung

$$\begin{aligned} \mathcal{D}p(x) &= \alpha \int e^{-\frac{1}{2}(Fy-x)R^{-1}(Fy-x)} p(y)\,\mathrm{d}y \\ &\cong \frac{1}{\sqrt{(2\pi)^n \det \Gamma}} e^{-\frac{1}{2}(x-\mu)\Gamma^{-1}(x-\mu)}. \end{aligned}$$

Nun berechnen wir μ durch einfache Integration:

$$\begin{aligned} \mu &= \alpha \int\int x e^{-\frac{1}{2}(Fy-x)R^{-1}(Fy-x)} p(y)\,\mathrm{d}y\,\mathrm{d}x \\ &= \alpha \int\int x e^{-\frac{1}{2}(Fy-x)R^{-1}(Fy-x)}\,\mathrm{d}x\, p(y)\,\mathrm{d}y \\ &= \int Fy \cdot p(y)\,\mathrm{d}y \end{aligned}$$

und ebenso Γ

$$\begin{aligned} \Gamma^{(ij)} + \mu^{(i)}\mu^{(j)} &= \alpha \int\int x^{(i)}x^{(j)} e^{-\frac{1}{2}(Fy-x)R^{-1}(Fy-x)} p(y)\,\mathrm{d}y\,\mathrm{d}x \\ &= \alpha \int\int x^{(i)}x^{(j)} e^{-\frac{1}{2}(Fy-x)R^{-1}(Fy-x)}\,\mathrm{d}x\, p(y)\,\mathrm{d}y \\ &= \int (R^{(ij)} + F^{(i)}(y)F^{(j)}(y)) \cdot p(y)\,\mathrm{d}y \end{aligned}$$

oder

$$\Gamma^{(ij)} = R^{(ij)} + \int (F^{(i)}(y) - \mu^{(i)})(F^{(j)}(y) - \mu^{(j)}) \cdot p(y)\,\mathrm{d}y.$$

(Hochgestellte Indices bezeichnen die Komponenten).

In unserem Fall ist natürlich $p(x) = p(x_{t-1}{=}x|y^{t-1})$. Bezogen auf unser Ausgangsproblem ergeben sich also mit der oben verwendeten Notation die folgenden

Identitäten:

$$\hat{x}_{t|t-1} = \int F(x) \cdot p(x_{t-1}{=}x|y^{t-1})\, \mathrm{d}x, \tag{3.12}$$

$$\begin{aligned}\Gamma^{(ij)}_{t|t-1} &= R^{(ij)} \\ &\quad + \int (F^{(i)}(x) - \hat{x}^{(i)}_{t|t-1})(F^{(j)}(x) - \hat{x}^{(j)}_{t|t-1}) p(x_{t-1}{=}x|y^{t-1})\, \mathrm{d}x.\end{aligned} \tag{3.13}$$

Dann bildet man nach der Bayes–Filterformel die Wahrscheinlichkeit

$$p(x_t{=}x|y^n) \cong \alpha \cdot p(y^n_t|x_t{=}x) \cdot e^{-\frac{1}{2}(x-\hat{x}_{t|t-1})\Gamma^{-1}_{t|t-1}(x-\hat{x}_{t|t-1})}. \tag{3.14}$$

Der erste Term ist zwar keine Gauß–Verteilung, jedoch in unseren Fällen der eher unkritische Term, da es sich um eine feste Funktion der Argumente y^n_t, x handelt. Wirkliche Schwierigkeiten treten nur dann auf, wenn n sehr viel größer als t ist, was bei Filterung ja nicht der Fall ist. Der Umgang mit diesem Term wird in verschiedenen Spezialfällen besprochen. Grundsätzlich gibt es die Möglichkeit, der Rückwärtsoperator ähnlich zu behandeln, wie es für den Vorwärtsoperator hier vorgeführt wurde. Damit hätte man das Problem dann vollständig auf die Verteilungen $p(y_t|x_t)$ zurückgeführt. Diese Verteilungen geschickt durch Gauß–Verteilungen zu approximieren ist ein Problem, das von Fall zu Fall verschieden gelöst werden muß, da die spezielle Gestalt stark von G abhängt. Konkretere Beispiele werden im nächsten Kapitel besprochen.

3.4 Implementation allgemeiner Kalman–Filter

Ein Ansatz zur Implementation des allgemeinen Kalman–Filters für $t = n$ läßt sich aus den Gleichungen (3.12), (3.13), (3.14) direkt ablesen: Man bildet die Integrale (3.12), (3.13), geht damit in Gleichung (3.14) usw. Die Ergebnisse einer Implementation werden am Ende dieses Abschnitts vorgestellt. Es zeigt sich, daß der Filter für technische Anwendungen in dieser Form natürlich viel zu langsam ist. Es existieren jedoch Verfahren zur näherungsweisen Berechnung von $\Gamma^{ij}_{t|t-1}, \hat{x}_{t|t-1}$, die eine explizite Integration umgehen.

Extended Kalman Filter (EKF) : Wir erläutern nur den Fall $t = n$. Die einfachste und klassische Methode besteht darin, die Dynamik F um $\hat{x}$ zu linearisieren. Dann kann man die Integrale (3.12), (3.13) direkt auswerten. Linearisiert man die Messung G ebenfalls um $\hat{x}$, so hat man für den ersten Term in (3.14) ebenfalls eine Gauß–Verteilung, womit die gewünschte Vereinfachung erreicht ist. Dieser Filter hat sich in der Praxis bewährt, solange es sich um halbwegs „linearisierbare“ Probleme handelt. So groß die persönliche Abneigung gegen einfaches brute-force-Linearisieren auch sein mag, man sollte anerkennen, daß dieser Filter schon das Sonnensystem durchflogen hat, während andere „ausgebufftere“ Filtervarianten den Beweis ihrer praktischen Brauchbarkeit noch schuldig sind.[3]

Higher-Order-Filter (HOF) Das Verfahren des EKF läßt sich fortsetzen: In den Integralen (3.12), (3.13) entwickelt man F in eine Taylorreihe und kann dann $\Gamma^{ij}_{t|t-1}, \hat{x}_{t|t-1}$ durch die Ableitungen von F und die höheren Momente von $p(x_{t-1}|y^{n-1})$ ausdrücken. Das ist aber auch nur dann einfach, wenn man $p(x_{t-1}|y^{n-1})$ durch eine Gauß–Verteilung gegeben hat, d.h. man muß sich auch wieder um die Wahrscheinlichkeiten $p(y^n_t|x_t\!=\!x)$ kümmern. Handelt es sich hierbei um eine Gauß–Verteilung, also um eine lineare Observable, so kann man (3.14) direkt auswerten und bleibt bei Gauß–Verteilungen.

Julier-Filter Eine interessante Methode zur näherungsweisen Berechnung der Integrale (3.12), (3.13) haben J. Ullmann und S. Julier in [15] angegeben. Dort wird das Problem betrachtet, den Mittelwert φ und die Kovarianzmatrix Φ einer Zufallsvariablen $F(x)$ zu schätzen, wobei F nichtlinear ist und x den Mittelwert $\hat{x}$ und die Kovarianzmatrix Γ besitzt. Dazu werden die zwei Matrizen $\pm\sqrt{n\Gamma}$ gebildet und eine Menge von Punkten $X := (x_1, \ldots, x_{2n})$ betrachtet, bestehend aus den Reihen oder Spalten von $\pm\sqrt{n\Gamma}$. Dann wird $y_i := F(x_i + \hat{x})$ gebildet und gezeigt, daß durch

$$
\begin{aligned}
\hat{\varphi} &= \frac{1}{2n}\sum_{i=1}^{2n} y_i \\
\hat{\Phi}^{(\alpha\beta)} &= \frac{1}{2n}\sum_{i=1}^{2n}(y_i^{(\alpha)} - \hat{\varphi}^{(\alpha)})(y_i^{(\beta)} - \hat{\varphi}^{(\beta)})
\end{aligned}
$$

gute Näherungen für φ, Φ gegeben sind. „Gut“ heißt hier Übereinstimmung der Taylorreihen bis zur zweiten Ordnung nach Potenzen von $x - \varphi$. Diese Technik, die weder Ableitungen noch Integrationen erfordert, wird dann benutzt, um einen allgemeinen Kalman–Filter zu implementieren.

Eine Große Anzahl von Implementationen dieser Filtervarianten wurde in [32] vorgenommen. Dort wurde als Observable stets eine lineare Abbildung (meist Projektion auf eine Komponente) benutzt. Als dynamische Systeme dienten u.a. die Henon– und die Ikeda–Abbildung. Es zeigte sich, daß HOF's und Julier–Filter etwa gleich in der Performance waren, in jedem Fall aber dem EKF überlegen. Große Verbesserungen erzielte man auch dort durch „akausale“ Filter, also Filter mit $n > t$. Dies wird durch die entsprechenden Ergebnisse dieser Arbeit bestätigt.

[3] Für Anwendungen des Extended–Kalman–Filter siehe [7]

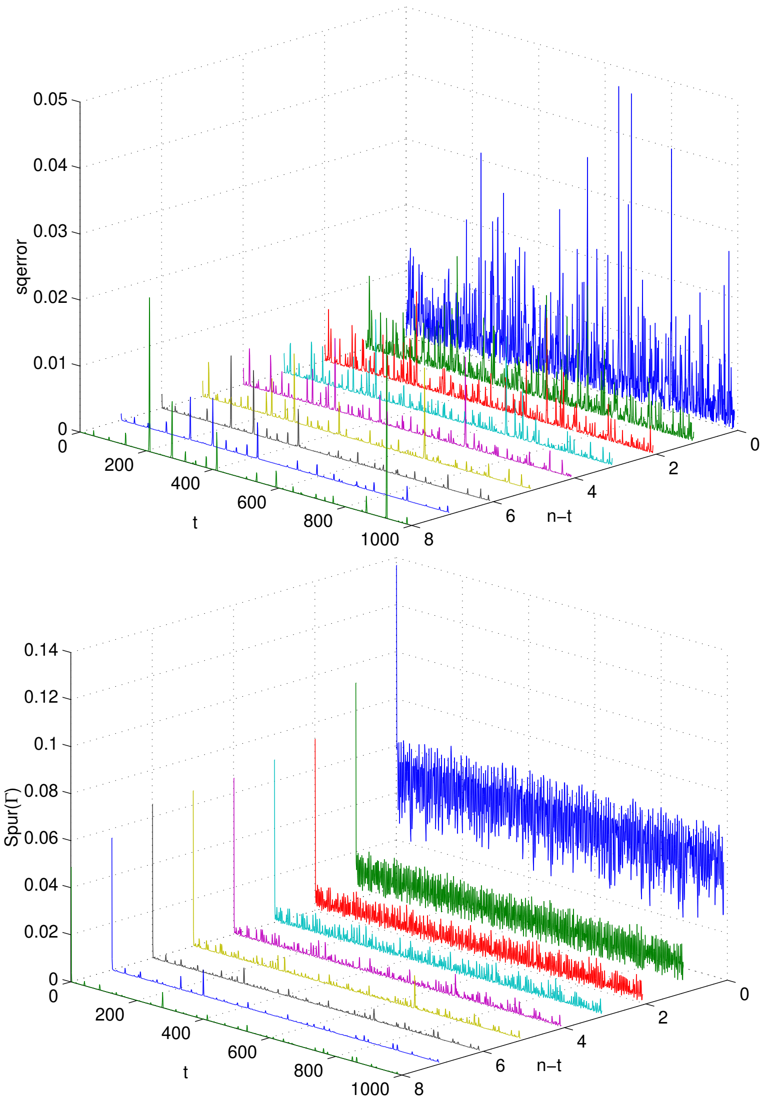


Abbildung 3.2: Quadratischer Fehler und Spur der Fehler-Kovarianzmatrix

Ergebnisse des Kalman–Filters

Als dynamisches System F wurde die Henon-Abbildung gewählt:

$$F:\mathbb{R}^2\to\mathbb{R}^2; F(x_1,x_2):=(1-1.4x_1^2+0.3x_2,x_1)$$

Dynamisches Rauschen war nicht vorhanden. Die Observable war einfach die Identität mit einem additiven Rauschen von 10 dB. Damit ergibt sich

$$p(y_t^n|x_t{=}x)=\alpha\cdot e^{-1/2\cdot\sum_{i=t}^n(y_i-F^{i-t}x)S^{-1}(y_i-F^{i-t}x)}$$

und

$$p(x_{t+1}{=}x|x_t{=}z)=p_F(x,z)=\delta(x-F(z)).$$

Im Prinzip läßt sich die Bayes–Filterformel sogar direkt auswerten:

$$\begin{aligned}\mathcal{D}p(x_t|y^t)(x) &= \int \delta(x-F(z))p(x_t\!=\!z|y^t)\,\mathrm{d}z \\ &= \int \delta(x-z)p(x_t\!=\!F^{-1}z|y^t)\frac{\mathrm{d}z}{\det F\circ F^{-1}z} \\ &\propto p(x_t\!=\!F^{-1}x|y^t),\end{aligned} \tag{3.15}$$

da die Determinante konstant ist. Damit ergibt sich

$$\begin{aligned}p_{t,n} &= \alpha\cdot e^{-\frac{1}{2}\sum_{i=t}^{n}(y_i-F^{i-t}x)S^{-1}(y_i-F^{i-t}x)}\cdot p_{t-1}(F^{-1}x) \\ p_t &= \alpha\cdot e^{-\frac{1}{2}(y_i-x)S^{-1}(y_i-x)}\cdot p_{t-1}(F^{-1}x).\end{aligned} \tag{3.16}$$

Der Filter wurde auf folgende Weise implementiert:

Die Funktion $p(y_t^n|x_t\!=\!x)$ wurde „fest installiert". Weiter stand eine Funktion

$$\Phi(x,\Gamma,\mu) := \frac{1}{\sqrt{(2\pi)^n\det\Gamma}}e^{-\frac{1}{2}(\mu-x)\Gamma^{-1}(\mu-x)}$$

zur Verfügung. Iterativ wurde dann folgendes Schema durchlaufen:

1. Die Funktion

   $$p(x_t\!=\!x|y^n) \cong q_t(x) := p(y_t^n|x_t\!=\!x)\cdot\Phi(x,\Gamma_{t|t-1},\mu_{t|t-1})$$

   wurde betrachtet und eine *Präkonditionierung* vorgenommen, d.h. eine affine Abbildung berechnet, mit der die Funktion transformiert wurde, um die Integration numerisch stabiler durchführen zu können. Dies wird später noch erläutert.

2. Die Funktion $q_t(x)$ wurde bezüglich x normiert.

3. Es wurde

   $$\begin{aligned}\mu_t &:= \int x\cdot p(x_t\!=\!x|y^n)\,\mathrm{d}x \\ &\cong \int x\cdot q_t(x)\,\mathrm{d}x\end{aligned}$$

   berechnet und als Schätzung des Zustandes x_t herausgegeben.

4. Es wurde

   $$\begin{aligned}\Gamma_t^{(ij)} &:= \int x^{(i)}x^{(j)}\cdot p(x_t\!=\!x|y^n)\,\mathrm{d}x - \mu_t^{(i)}\mu_t^{(j)} \\ &\cong \int x^{(i)}x^{(j)}\cdot q_t(x)\,\mathrm{d}x - \mu_t^{(i)}\mu_t^{(j)}\end{aligned}$$

   berechnet und als Schätzung der Fehler-Kovarianzmatrix herausgegeben.

5. Die Integrale (3.12), (3.13) liefern einen neuen Mittelwert $\mu_{t+1|t}$ und eine neue Kovarianzmatrix $\Gamma_{t+1|t}$. Dann wurde zu Schritt 1 zurückgekehrt.

Man bemerke, daß in die Schätzung des t-ten Zustandes nur die Werte $y_t^n = (y_t, \ldots, y_n)$ explizit eingehen.

Die Integrale wurden numerisch auf einem 20×20-Gitter mit Hilfe der Gauß–Quadratur (siehe [28]) ausgewertet. Dazu ist eine adaptive Anpassung des Gitters nötig, was mit Hilfe der schon erwähnten Präkonditionierung geschah. Diese Präkonditionierung liefert eine affine Abbildung $x \to \Delta(x-\delta)$, welche den Mittelwert der Verteilung $q_t(x)$ ungefähr in den Ursprung verschiebt und die Kovarianz etwa auf die Einheitsmatrix normiert. Dazu ist es nötig, den Mittelwert μ und die Kovarianz Γ, die ja in den Schritten 3, 4 erst berechnet werden, in etwa abzuschätzen. Ist das geschehen, so setzt man $\Delta := \Gamma^{-1/2}$, $\delta := \mu$.[4] Die Abschätzung selbst gewinnt man durch Linearisieren der Dynamik, also im Prinzip durch einen EKF.

Da ohne dynamisches Rauschen gearbeitet wurde, sollte die Fehler-Kovarianzmatrix Γ gegen Null gehen, was dann einer Schätzung $\mu_t = F(\mu_{t-1})$ entspräche. Dann jedoch ist Γ nicht mehr invertierbar, was numerische Schwierigkeiten bereitet. Da jedoch die Schätzung niemals genauer sein kann als die Maschinengenauigkeit ϵ, wurde ständig die Matrix $\epsilon \cdot \mathbf{1}$ zur Kovarianz Γ addiert. Trotzdem ist stets Vorsicht angebracht, wenn die Determinante von Γ klein wird, da dies auch auf schlechte Kondition von Γ hinweisen kann. In diesem Fall muß der Filter neu initialisiert werden. Gestartet wurde mit

$$\begin{aligned} \mu_{1|0} &= y_1, \\ \Gamma_{1|0} &= S. \end{aligned}$$

In Abb.3.2 sind die Spur von Γ_t und der quadratische Fehler zwischen den gefilterten und den unverrauschten Ausgangsdaten abgebildet. Die Länge der Zeitreihen betrug 1000 Punkte (es sind nur 990 dargestellt, da der Filter bei Akausalität k keine Schätzungen für die letzten k Werte liefert). Das Signal-Rauschverhältnis der gefilterten Daten ist in folgender Tabelle dargestellt:

Akausalität	SNR/dB
0	11.9336
1	14.5311
2	15.9062
3	16.8956
4	17.6995
5	18.5137
6	19.5465

Akausalität	SNR/dB
7	20.2653
8	18.7672
9	12.9783
10	11.8210
11	10.3803
12	10.2390
13	10.0110

Es zeigt sich, daß der Filter durchaus arbeitet, und zwar mit zunehmender Akausalität zunächst besser. Die Stabilität läßt sich an den Zeitentwicklungen

[4] Ist $U^t D U = \Gamma$ die Eigenwertzerlegung von Γ, so ist $\Gamma^{-1/2} = U^t D^{-1/2}$.

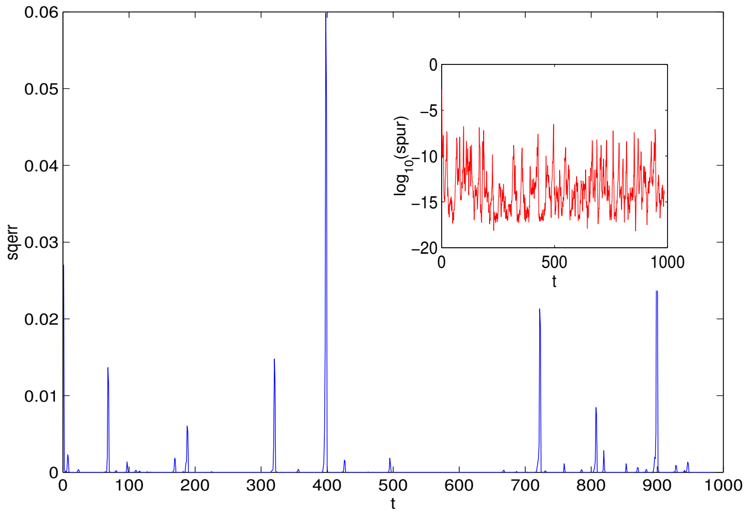


Abbildung 3.3: Quadratischer Fehler und Spur der Fehler-Kovarianzmatrix für Akausalität 12

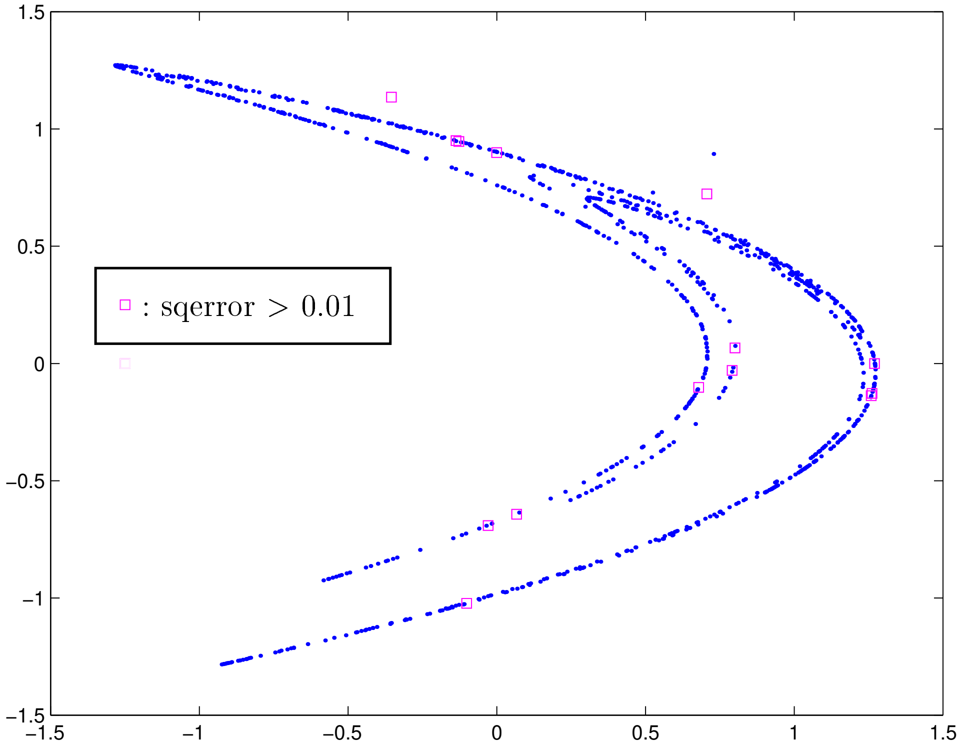


Abbildung 3.4: Henon-Attraktor mit den Stellen, an denen Glitches auftreten

der Spur erkennen (da die Matrix durchweg gut konditioniert war). Bei Akausalität 8 jedoch zeigt sich ein Abfallen des SNR und man sieht im quadratischen Fehler zwei isolierte, aber deutliche *Glitches*, also isolierte Peaks, deren Wert sehr groß ist im Vergleich zu den umliegenden Werten. Diese Glitches treten auch bei höheren Akausalitäten auf. Abb. 3.3 zeigt z.B. für Akausalität 12 noch einmal den quadratischen Fehler und die Spur (im Logarithmus). Es zeigen sich Glitches, die schon wieder in der Größenordnung des Rauschens liegen. Diese Fehler befinden sich auf dem Henon-Attraktor an charakteristischen Stellen, nämlich dort, wo stabile und instabile Mannigfaltigkeiten tangential liegen (siehe Abb. 3.4). Dieses Phänomen tritt bei den im nächsten Kapitel untersuchten Verfahren ebenfalls auf und soll dort näher beleuchtet werden. Bei Akausalitäten > 14 lief der Filter nicht mehr stabil und brach wegen schlechter Kondition der Kovarianzmatrix ab.

3.5 Ausblicke

Zum Ende des Kapitels möchte ich einige Aspekte der Filtertheorie betrachten, die in dieser Arbeit nicht weiterverfolgt werden, aber nicht uninteressant sind. Es handelt sich um zwei weitere Näherungsmethoden und die grundsätzliche Frage nach Fehlerabschätzungen für die Filter.

Weitere Näherungsmethoden zur Lösung des Filterproblems sind zwar denkbar, aber meines Wissens noch nicht implementiert worden. Naheliegend wäre hier zunächst, höhere Momente mit ins Spiel zu bringen. Es ist jedoch nicht so offensichtlich, wie man eine Verteilung in ihre Momente entwickelt, oder mit anderen Worten, welche Art von Verteilung man einer gegebenen Menge von Momenten zuordnen soll. Im Zusammenhang mit dem zentralen Grenzwertsatz wurden mehrere Ansätze entwickelt (Gram-Charlier-Reihe, Edgeworth-Reihe), alle jedoch mit dem Ziel, die Verteilung einer Summe von n unabhängigen Zufallsvariablen nach Potenzen von $1/n$ zu entwickeln. Das hier behandelte Problem kann man natürlich so auffassen und versuchen, diese Entwicklungen gewinnbringend einzusetzen. Eine Darstellung dieser Reihenentwicklungen und der dazugehörigen Theorie findet sich in [8]. Allerdings steht natürlich gegen den Nutzen solcher Methoden der Zuwachs an Rechenaufwand. Um das k-te Moment einer n-dimensionalen Zufallsvariablen zu berechnen, muß man im allgemeinen $\frac{1}{k!}\prod_{i=0}^{k-1}(m+i)$ Integrale berechnen, wenn man nicht irgendwelche Symmetrien ausnutzen kann.

Eine völlig andere Methode gibt die Annäherung durch Gauß–Verteilungen auf und könnte zumindest prinzipiell die Möglichkeit eröffnen, mit Verteilungen zu arbeiten, die dem Problem (und seinen Symmetrien) besser angepaßt sind. Dazu betrachtet man zu einem offenen Parameterbereich $\Theta \subset \mathbb{R}^n$ eine *exponentielle Familie* von Verteilungen. Daß ist eine Abbildung

$$p(\cdot,\theta);\Theta \to L(\Omega);\theta \to \exp(\sum \theta_i \cdot c_i(x) - \psi(\theta)),$$

also eine Schar von Verteilungen auf einem W-Raum Ω, wobei $c_i : \Omega \to \mathbb{R}$ fest gewählte Funktionen sind. Die Funktion $\psi(\theta)$ ist so gewählt, daß $p(\cdot, \theta)$ normiert ist. An die Funktionen c_i müssen natürlich gewisse Regularitätsforderungen gestellt werden, so daß $p(\cdot, \theta)$ wohldefiniert ist. Nun betrachtet man die *c–Momente* der Verteilung, das sind die Größen

$$\eta_i(\theta) := \int c_i(x) p(x, \theta) \,\mathrm{d}x.$$

Aus der Relation

$$\int p(x, \theta) \,\mathrm{d}x = 1$$

folgt durch Differentiation sofort

$$\eta_i(\theta) = \frac{\partial}{\partial \theta_i} \psi(\theta). \tag{3.17}$$

Es läßt sich zeigen, daß die Hesseform von $\psi(\theta)$ nicht ausgeartet ist, was impliziert, daß zumindest lokal eine diffeomorphe Abbildung zwischen den θ und den η existiert. In Filterproblemen könnte man dann so vorgehen: Die Verteilung $p(x_t = x|y^n)$ hat man (näherungsweise) durch einen Parametersatz $\theta^{(t)}$ dargestellt, also $p(x_t = x|y^n) \cong p(x, \theta^{(t)})$. Darauf wendet man die Bayes–Filterformel an und erhält eine Funktion $q(x) \cong p(x_{t+1} = x|y^{n+1})$. Von dieser Funktion bildet man die c-Momente

$$\eta_i(\theta_{t+1}) = \int c_i(x) q(x) \,\mathrm{d}x,$$

setzt diese in (3.17) ein, löst nach θ_{t+1} auf und beginnt von vorne.

Beide hier dargestellten Konzepte stammen aus der Statistik, Implementationen in Filterproblemen jedoch sind mir wie gesagt nicht bekannt. Ideal wäre natürlich, wenn man eine exponentielle Familie findet, bei der man explizit dynamische Gleichungen für die Parameter aufstellen kann.

In keiner mir vorliegenden Arbeit wurden für die approximativen Filter irgendwelche quantitativen Fehlerabschätzungen durchgeführt. Das wird in dieser Arbeit nicht anders sein, doch es wäre sicherlich interessant, Methoden der Statistik, die prinzipielle Fehlerabschätzungen ermöglichen (Rao–Cramér–Ungleichung etc.), auch auf Filterprobleme anzuwenden.

4 Rekonstruktion gestörter Zeitreihen

In der Zeitreihenanalyse hat man oft mit verrauschten Daten zu tun, die vor einer eingehenden Analyse erst einmal von Rauschen befreit werden müssen, da die Methoden der Zeitreihenanalyse oft sehr empfindlich auf Rauschen reagieren. Die im vorigen Kapitel vorgestellten Filter leisten das im Prinzip, haben ihren Platz jedoch eher in online-Anwendungen. In der Zeitreihenanalyse hat man weniger das Problem der Echtzeitfähigkeit, vielmehr möchte man die gesamte Zeitreihe, also die volle Information, die man hat, auch ausnutzen. Dieses Kapitel wird sich mit dieser Problematik beschäftigen. Da es sich wie gesagt nicht um Filterung handelt (weder im strengen Sinne noch in dem Sinne, daß man nur kleine Akausalitäten zuläßt), soll diese Problematik *Rekonstruktion* genannt werden. In den ersten beiden Abschnitten geht es um grundsätzliche Ansätze, wobei man eine Unterscheidung zwischen wahrscheinlichkeitstheoretisch motivierten Methoden (in der Tradition des letzten Kapitels) und der Methode des shadowing unterscheiden kann. Der nächste Abschnitt behandelt ein grundsätzliches Problem von Rauschreduktionsalgorithmen für den Fall, daß tangentiale Berührungen vorliegen. In den drei letzten Kapiteln werden Algorithmen vorgestellt und Ergebnisse besprochen. Dabei geht es im letzten Kapitel um die Analyse von Zeitreihen ohne explizite Kenntnis der dynamischen Gleichungen. Ein Ansatz dieses Kapitel ist auch, bereits bestehende Methoden in das wahrscheinlichkeitstheoretische Konzept einzuordnen (für einen Überblick über gängige Rauschreduktionsmethoden siehe [16]).

4.1 Maximum–Likelyhood–Methoden

Im letzten Kapitel wurde ausgiebig die Funktion $p(x_t = x|y^n)$ betrachtet und dazu benutzt, den Zustand x_t aus der Information y^n zu schätzen. In der Zeitreihenanalyse hat man aber eine ganze Zeitreihe y^N zur Verfügung, und es geht darum, den Orbit $x^N = (x_0, \ldots, x^N)$ herauszufinden. Es soll zuerst geklärt werden, wo die Unterschiede und Gemeinsamkeiten liegen.

Als Konzept bietet sich an, nunmehr die bedingte Wahrscheinlichkeit

$$p(x^N|y^N)$$

zu betrachten. Da der Orbit im allgemeinen selbst eine Zufallsgröße ist, wird auch hier wieder der Bayes-Formalismus benutzt. Es ergibt sich

$$\begin{aligned} p(x^N|y^N) &= \alpha \cdot p(y^N|x^N) \cdot p(x^N) \\ &= \alpha \cdot e^{-\frac{1}{2}\sum_{i=1}^N (y_i - Gx_i)S^{-1}(y_i - Gx_i)} \cdot p(x^N). \end{aligned}$$

Die Funktion $p(x^N)$ läßt sich induktiv berechnen als

$$\begin{aligned} p(x^N) &= \prod_{i=1}^N p(x_i|x_{i-1}) \cdot p_0(x) \\ &= \alpha \cdot e^{-\frac{1}{2}\sum_{i=1}^N (x_i - Fx_{i-1})R^{-1}(x_i - Fx_{i-1})} \cdot p_0(x), \end{aligned}$$

wobei die Bayes-Kettenregel und die Markov-Eigenschaft ausgenutzt wurde. Ein Maximum-Likelyhood-Schätzer ergibt sich damit durch maximieren der Funktion $p(x^N|y^N)$ oder äquivalent durch minimieren der Funktion

$$\begin{aligned} H_{KY}(x^N) &= -\log(p(x^N|y^N)) \\ &= \sum_{i=0}^N (y_i - Gx_i)S^{-1}(y_i - Gx_i) \\ &+ \sum_{i=1}^N (x_i - Fx_{i-1})R^{-1}(x_i - Fx_{i-1}), \end{aligned} \tag{4.1}$$

wobei ausgenutzt wurde, daß die Anfangsverteilung die Gestalt

$$p_0(x) = \alpha \cdot e^{-\frac{1}{2}(y_0 - Gx_0)S^{-1}(y_0 - Gx_0)}$$

besitzt. Diese Funktion wurde von Kostelich und Yorke in [17] zur Reduktion von Rauschen benutzt. Ist kein dynamisches Rauschen vorhanden, so ergibt sich der Schätzer einfach durch minimieren der Funktion

$$H(x^N) = \sum_{i=0}^N (y_i - Gx_i)S^{-1}(y_i - Gx_i) \tag{4.2}$$

unter der Nebenbedingung

$$x_i = F(x_{i-1}).$$

Es sei kurz erwähnt, daß man das Problem der Minimierung von H_{KY} als eine Regularisierung des Minimierungsproblems H mit Nebenbedingungen ansehen kann. In der Funktion H_{KY} taucht die Nebenbedingung in gewissermaßen abgeschwächter Form auf. Läßt man die Eigenwerte von R^{-1} gleichmäßig anwachsen, so erwartet man, daß die Lösung des Minimierungsproblems gegen ein Minimum von H konvergiert und gleichzeitig die Nebenbedingungen immer besser und besser erfüllt. Aus der in dieser Arbeit vertretenen Sicht bedeutet das, daß man das dynamische Rauschen gegen Null gehen läßt, womit sich diese Regularisierungstechnik gut in unsere Sicht einordnet. Im letzten Kapitel wurde die Verteilung $p_{t,n}$ betrachtet. Diese Funktion liefert natürlich auch Maximum-Likelyhood-Schätzer, allerdings für die einzelnen Zustände. Ein durch solche Schätzer induzierter Orbit ist jedoch im allgemeinen *nicht* derselbe wie ein durch minimieren der Funktion $H(x^N)$ ermittelter. Es gilt jedoch die folgende Ausnahme: Ist kein dynamisches Rauschen vorhanden und besitzt F eine konstante Funktionaldeterminante, so induziert der durch $p(x_N = x|y^N)$ gegebene Maximum-Likelyhood-Schätzer den Orbit, der durch minimieren der Funktion $H(x^N)$ gegeben ist, oder äquivalent den Orbit, der die Verteilung $p(x^N|y^N)$ maximiert. Diese Tatsache ist leicht einzusehen, da man genauso vorgehen kann wie in den Gleichungen (3.15), (3.16). Induktiv ergibt sich

$$p_N(x) = \alpha \cdot e^{-\frac{1}{2}\sum_{i=1}^{N}(y_i - GF^{i-N}x)S^{-1}(y_i - GF^{i-N}x)} \cdot p_0(F^{-N}x).$$

Bildet man den negativen natürlichen Logarithmus dieses Ausdrucks und beachtet wieder die Gestalt der Anfangsverteilung, so landet man gerade bei der Funktion

$$H'(x) = \frac{1}{2}\sum_{i=0}^{N}(y_i - GF^{i-N}x)S^{-1}(y_i - GF^{i-N}x).$$

Das Minimieren dieses Ausdrucks ist dann gerade äqivalent dem Minimieren von $H(x^N)$ unter der Nebenbedingung $x_i = F(x_{i-1})$. An dieser Betrachtung wird auch deutlich, daß das Problem nur einen Freiheitsgrad hat, nämlich den Anfangswert (oder einen beliebigen anderen Punkt des Orbits). Es ist klar, daß zur Lösung des Problems eine gewisse Regularisierung notwendig ist.

4.2 Shadowing–Methoden

Einen etwas anderen Ansatz verfolgt man bei Rauschreduktion durch shadowing. Im Gegensatz zu den wahrscheinlichkeitstheoretischen Methoden des vorigen Abschnitts knüpfen wir hier an die Konzepte des Kapitels über nichtlineare Dynamik an. Shadowing hat das allgemeine Ziel, zu einem Pseudo-Orbit einen (wenn möglich in irgendeiner Weise typischen) Schatten zu finden, wobei die Ursache der Störung eher unberücksichtigt bleibt. Der zentrale Unterschied zu den wahrscheinlichkeitstheoretischen Methoden ist, daß *in jedem Falle* das Ergebnis ein deterministischer Orbit ist, auch wenn der Zustand des Systems eine Zufallsvariable ist (also dynamisches Rauschen vorliegt). Diese Fragestellungen sind u.a. in folgenden Zusammenhängen wichtig:

1. In numerischen Simulationen dynamischer Systeme kann man aufgrund der endlichen Genauigkeit niemals einen echten Orbit erzeugen. Man hat vielmehr einen δ-Pseudo-Orbit, wobei δ in etwa der Rechengenauigkeit entspricht. Trotzdem möchte man, daß dieser Pseudo-Orbit in gewisser Weise einem typischen echten Orbit des Systems entspricht. Dies ist gegeben, wenn eindeutige Shadowing–Orbits existieren.

2. In vielen Anwendungen läßt man ein dynamisches System nicht frei laufen, sondern nimmt Steuereingriffe vor, um der Dynamik ein gewisses Verhalten aufzuzwingen. Z.B. kann man sich mit Hilfe einer nichtlinearen Dynamik ein Kommunikationssystem aufgebaut denken. Dazu wird der Phasenraum geeignet partitioniert, im einfachsten Fall in zwei disjunkte Teile. Jeder Teilmenge ordnet man ein bestimmtes Symbol zu. Ein Orbit induziert dann eine Symbolsequenz, die im Falle von zwei Symbolen einfach eine Bitsequenz ist. Die Partitionierung muß natürlich so gewählt sein, daß eine möglichst große Menge von Symbolsequenzen erzeugbar ist. Es läßt sich zeigen (siehe [3]), daß jedes hyperbolische System konjugiert zu einem Raum von Symbolsequenzen *(shift-space)* mit endlich vielen Symbolen ist, die man gerade durch eine geeignete Partition erzeugt wird. Die Dynamik wird auf dem shift-space durch den *shift* gegeben, der die Symbolsequenz einfach um eine Stelle nach links verschiebt. Um nun eine gewünschte Symbolsequenz zu erzeugen, ist es aber im allgemeinen notwendig, durch kleine Steuereingriffe die Dynamik zu beeinflussen, da es praktisch unmöglich ist, den richtigen Anfangswert zu präparieren. Bleiben die Steuereingriffe jedoch klein, so kann man davon ausgehen, daß es einen echten Orbit gibt, der dieselbe Symbolsequenz erzeugt.

Zur Konstruktion eines Shadowing–Orbits kann man im Prinzip die Technik benutzen, die im Beweis des Shadowing–Lemmas verwendet wurde. Dazu muß man die stabilen und instabilen Mannigfaltigkeiten in jedem Punkt kennen, was aber sehr aufwendig ist. Praktisch alle anderen Methoden betrachten einen quadratischen Fehler der Form (4.1) oder (4.2), wobei die Observable G die Identität ist. Dies ist auch in allen Shadowing–Techniken, die in dieser Arbeit besprochen werden, der Fall. Die eigentlich richtige Norm (das richtige Fehlermaß) für die Abweichung eines Pseudo-Orbits vom Determinismus wäre natürlich die Maximumnorm, denn genau genommen ist das Shadowing–Lemma eine Aussage über die Maximumnorm. Diese Norm jedoch ist natürlich analytisch schwer zu handhaben; deshalb ist die L_2–Norm für analytische Zwecke so beliebt. Im übernächsten Abschnitt jedoch wird ein Verfahren vorgestellt, das sich mit *jeder* Norm durchführen läßt, sofern sie die Dreiecksungleichung erfüllt.

4.3 Tangentiale Berührungen

Bevor auf die Einzelheiten der Algorithmen eingegangen wird, soll ein prinzipielles Problem angesprochen werden, das auftreten kann (und typischerweise auch auftritt), wenn man Shadowing–Orbits für nichthyperbolische Systeme konstruieren will, genauer gesagt für Pseudo-Orbits mit Punkten in der Nähe einer Stelle, an der stabile und instabile Mannigfaltigkeiten sich tangential berühren, was ja bei hyperbolischen Systemen nicht passieren kann. Dazu betrachte man die Funktion

$$L(x^N) = \sqrt{\frac{1}{N+1} \cdot \sum_{i=0}^{N} (y_i - x_i) S^{-1} (y_i - x_i)}.$$

Angenommen $x^N = (x_0, \ldots, x_N)$ sei ein Orbit. Sei nun $z^N = (z_0, \ldots, z_N)$ ein zweiter Orbit mit $z_0 \in W^s(x_0)$. Das impliziert $z_i \in W^s(x_i)$, und es gilt

$$L(x^N) \leq \sqrt{\frac{1}{N+1} \cdot \sum_{i=0}^{N} (z_i - x_i) S^{-1} (z_i - x_i)} + L(z^N),$$

was man mit Hilfe der Dreiecksungleichung einsieht. Da nun $z_0 \in W^s(x_0)$ hat man $(z_i - x_i) S^{-1} (z_i - x_i) \to 0$ für $i \to \infty$ und deshalb

$$\sqrt{\frac{1}{N+1} \cdot \sum_{i=0}^{N} (z_i - x_i) S^{-1} (z_i - x_i)} \to 0$$

für $N \to \infty$. Diese Argumente sind völlig symmetrisch in x, z. Außerdem darf man annehmen $M := \max_{x,i} (y_i - x) S^{-1} (y_i - x)$ ist eine endliche Zahl. Deshalb folgt für ein beliebiges $\epsilon > 0$

$$\begin{aligned} L(x^N)^2 &\leq L(z^N)^2 + \epsilon(M + \epsilon) \\ L(z^N)^2 &\leq L(x^N)^2 + \epsilon(M + \epsilon), \end{aligned}$$

wenn man N genügend groß wählt. Daraus sieht man, daß für ein beliebiges $\epsilon > 0$

$$\left| \frac{H(x^N)}{N+1} - \frac{H(z^N)}{N+1} \right| \leq \epsilon(M + \epsilon),$$

wenn man N genügend groß wählt. Die Minima von $H(x^N)/(N+1)$ sind dieselben wie von $H(x^N)$. Diese Überlegung zeigt aber, daß $H(x^N)/(N+1)$ für große N entlang der stabilen Mannigfaltigkeiten von x_0^N nahezu konstant ist. Dieselben Überlegungen kann man nun durchführen für $z_0 \in W^u(x_0)$ und erhält ebenso

$$\left| \frac{H(x^0_{-N})}{N+1} - \frac{H(z^0_{-N})}{N+1} \right| \leq \epsilon(M + \epsilon).$$

Das zeigt, daß $\frac{H(x^0_{-N})}{N+1}$ für große N entlang der instabilen Mannigfaltigkeiten von x^0_{-N} nahezu konstant ist.

Diese Tatsachen haben nun folgende Konsequenz: Sei y^N_{-N} ein Pseudo-Orbit und x^N_{-N} ein Orbit, der die zugehörige Fehlerfunktion $H(x^N_{-N})$ minimiert. Betrachte einen Orbit z^N_{-N} mit z^N_{-N} homoklin zu x^N_{-N}, also

$$z_k \in W^u(x_k) \cap W^s(x_k)\, \forall k \in (-N, N).$$

Dann ist nach den eben angestellten Betrachtungen

$$\begin{aligned} \frac{H(x^N_{-N})}{2N+1} &= \frac{N+1}{2N+1}\left(\frac{H(x^0_{-N})}{N+1} + \frac{H(x^N)}{N+1}\right) \\ &\cong \frac{N+1}{2N+1}\left(\frac{H(z^0_{-N})}{N+1} + \frac{H(z^N)}{N+1}\right) \\ &= \frac{H(z^N_{-N})}{2N+1}. \end{aligned}$$

Deshalb wird H auch in z^N_{-N} (nahezu) minimal sein. Diese Methode unterscheidet also in dieser Form schlecht zwischen dem „richtigen“ Shadowing–Orbit und homoklinen Orbits.

Die hier wesentlichen Eigenschaften homokliner Orbits bei hyperbolischen Systemen werden durch den Satz von Hirsch (2.14) und seine Folgerungen gegeben. Zwei beliebige Orbits, egal ob homoklin oder nicht, müssen einen Maximalabstand $> \bar{\epsilon}$ haben, wobei $\bar{\epsilon}$ eine systemimanente Größe ist. Man wird also zum Auffinden des Shadowing–Orbits für y^N_{-N} stets nur kleine Korrekturen vornehmen und ausnutzen, daß ein korrekter Shadowing–Orbit und ein homokliner Orbit einen Maximalabstand von mehr als $\bar{\epsilon}$ haben, y^N_{-N} und ein korrekter Shadowing–Orbit jedoch einen Maximalabstand von weniger als $\bar{\epsilon}$, jedenfalls wenn das Rauschen in y^N_{-N} nicht zu groß ist. Fehler macht man dann höchstens am Rand, denn x_{-N} kann man immer noch entlang der lokalen stabilen Mannigfaltigkeit $W^s_{\bar{\epsilon}}(x_{-N})$ verschieben. Ebenso kann man x_N entlang der lokalen instabilen Mannigfaltigkeit $W^u_{\bar{\epsilon}}(x_N)$ verschieben. Dieser Randfehler konvergiert jedoch schnell gegen null, je mehr man in die „Mitte“ des Orbits geht.

Bei nicht hyperbolischen Systemen ist die Situation im allgemeinen anders. Hier können u.U. zwei homokline Orbits einen sehr kleinen Abstand haben, oder m. a. W. ein Orbit kann durch kleine Störungen in einen homoklinen Orbit überführt werden.[1] Dann kann man die obige Betrachtung nicht mehr anstellen. Auch die Eindeutigkeit des Shadowing–Orbits für $N \to \infty$ stimmt dann nicht mehr. Hat man einen Pseudo-Orbit $y^\infty_{-\infty}$, dazu einen ϵ–Schatten x und zu diesem x einen homoklinen Punkt z, wobei für die induzierten Orbits $x^\infty_{-\infty}$, $z^\infty_{-\infty}$ gilt $d(x_i, z_i) < \epsilon\, \forall i$, so ist auch z ein ϵ–Schatten von $y^\infty_{-\infty}$.

[1]Man könnte auch behaupten, daß dort die sensitive Abhängigkeit von den Anfangsbedingungen (bzw. die Expansivität) nicht uneingeschränkt gilt.

Das Problem ist nicht so gravierend, wenn sich das System typischerweise auf einer hyperbolischen Menge aufhält. Dies ist jedoch beim Hénon–System nicht der Fall, und es zeigt sich, daß die Rauschreduktionsalgorithmen, die in dieser Arbeit untersucht werden, typischerweise Fehler an Stellen machen, an denen die stabilen und instabilen Mannigfaltigkeiten sich tangential berühren. Nach den angestellten Betrachtungen über homokline Orbits könnte das daran liegen, daß es dort homokline Punkte gibt, die Orbits mit sehr kleinem Maximalabstand induzieren. Es soll nun ein geometrisch–anschauliches Argument gegeben werden, warum das tatsächlich der Fall ist. Es sei ausdrücklich bemerkt, daß die Argumentation nicht mathematisch streng ist und insofern nur als Erklärungsversuch zu werten ist. Die Abbildung (4.1) zeigt die Situation. Im Punkte x liegt eine tan-

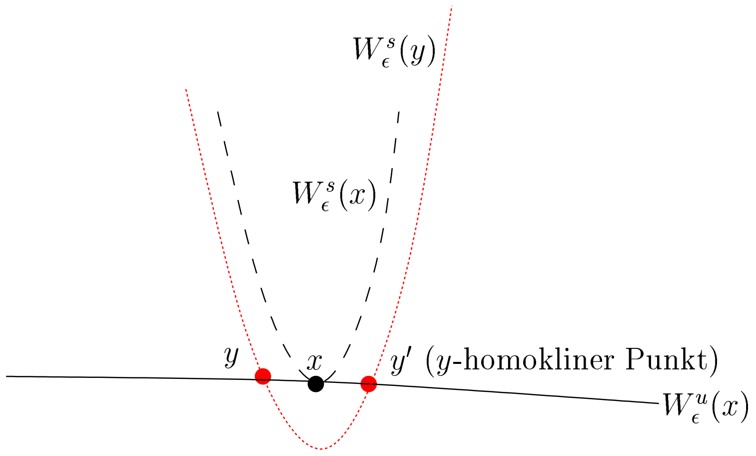


Abbildung 4.1: Stabile und instabile Mannigfaltigkeiten in der Nähe einer tangentialen Berührung

gentiale Berührung von stabiler und instabiler Mannigfaltigkeit vor. Der Punkt y ist eine kleine Störung von x auf der instabilen Mannigfaltigkeit, die ja lokal dem Attraktor entspricht. Angenommen, diese Mannigfaltigkeiten behalten unter kleinen Störungen in etwa ihre Gestalt (in der Zeichnung gepunktet dargestellt).[2] Die Mannigfaltigkeit $W^s_\epsilon(y)$ hat immer noch parabelförmige Gestalt und schneidet deshalb $W^u_\epsilon(y) = W^u_\epsilon(x)$ in einem weiteren Punkt y', der deshalb zu y homoklin ist. Offenbar kann man durch beliebig kleine Störung von x homokline Punktepaare mit beliebig kleinem Abstand erzeugen, die außerdem durch sehr kurze Stücken von stabiler und instabiler Mannigfaltigkeit verbunden sind. Ist der Punkt x eine isolierte tangentiale Berührung, so ist in der Umgebung das dynamische System hyperbolisch. Für festes y kann man dann sogar geometrische Konvergenz erwar-

[2]Genau genommen handelt es sich hier um eine spezielle Sorte von tangentialen Berührungen, um *quadratisch tangentiale* Berührungen. Hier ist die lokale Gestalt der Mannigfaltigkeiten tatsächlich stabil gegen kleine Störungen, aber natürlich ist nicht jede tangentiale Berührung auch quadratisch tangential.

ten, d.h. $d(F^n y, F^n y') \leq \lambda^n d(y,y')$ und $d(F^{-n} y, F^{-n} y') \leq \lambda^n d(y,y')$ für $n \geq 0$ mit einem $\lambda \in (0,1)$. Diese beiden homoklinen Punkte würden dann Orbits mit sehr kleinem Maximalabstand induzieren, was, wie bereits ausgeführt wurde, zu Problemen in den Algorithmen führt. Bei Auswertung der numerischen Ergebnisse in den nächsten Abschnitten wird sich zeigen, daß für das Hénon–System die Algorithmen genau an den Stellen versagten, an denen eine tangentiale Berührung vorlag. Das die fehlerhaften Korrekturen und die tatsächlichen Punkte homoklin sind, ist zwar nicht explizit nachgerechnet worden, aber einige detaillierte Plots der stabilen und instabilen Mannigfaltigkeiten des Hénon–Attraktors legen diese Vermutung sehr nahe (siehe Seite 68 und 70).

4.4 Least–Squares–Algorithmen

In diesem Kapitel geht es um die Lösung des im 1. Kapitel formulierten Problems:

$$\begin{aligned} &\text{minimiere} && H(x^N) = \sum_{n=0}^{N} (y_n - Gx_n) S^{-1} (y_n - Gx_n) \\ &\text{unter der Nebenbedingung} && x_n = F(x_{n-1}). \end{aligned} \tag{4.3}$$

Im folgenden sei wie gesagt G die Identität. Weiter sei $S = \mathbf{1}$. Außerdem verwenden wir die Abkürzung $\theta_n := x_n - y_n$. Das Problem liest sich dann folgendermaßen:

$$\begin{aligned} 0 &= (y_n + \theta_n) - F(y_{n-1} + \theta_{n-1}), \\ \min! &= \sum_{k=1}^{N} \|\theta_k\|^2. \end{aligned}$$

Nun wird die erste Gleichung um y_{n-1} linearisiert und dann die Methode der Lagrangeschen Multiplikatoren angewandt. Das führt auf die Gleichungen

$$\begin{aligned} 0 &= y_n - F(y_{n-1}) + \theta_n - \mathrm{D}F(y_{n-1})\theta_{n-1}, \\ \theta_n &= \lambda_n - \mathrm{D}F^t(y_n)\lambda_{n+1} \; 1 < n < N, \\ \theta_1 &= -\mathrm{D}F^t(y_1)\lambda_2, \\ \theta_N &= \lambda_N. \end{aligned}$$

Den Ausdruck für θ_n setzt man in die erste Gleichung ein und erhält:

$$\begin{aligned} 0 &= y_2 - F(y_1) + \left[\mathbf{1} + \mathrm{D}F\,\mathrm{D}F^t(y_1)\right] \lambda_2 \\ &\quad - \mathrm{D}F^t(y_2)\lambda_3, \\ 0 &= y_n - F(y_{n-1}) + \left[\mathbf{1} + \mathrm{D}F\,\mathrm{D}F^t(y_{n-1})\right] \lambda_n \\ &\quad - \left[\mathrm{D}F^t(y_n)\lambda_{n+1} + \mathrm{D}F(y_{n-1})\lambda_{n-1}\right] \\ &\quad \text{für } 1 < n < N, \\ 0 &= y_N - F(y_{N-1}) + \left[\mathbf{1} + \mathrm{D}F\,\mathrm{D}F^t(y_{N-1})\right] \lambda_N \\ &\quad -\mathrm{D}F(y_{N-1})\lambda_{N-1}. \end{aligned}$$

Dieses ist ein (schwach besetztes) lineares Gleichungssystem für die $\lambda_2 ... \lambda_N$. Da dieses Gleichungssystem obendrein symmetrisch und positiv definit ist, bietet es sich an, ein iteratives Mehrschrittverfahren (GAUSS–SEIDEL–Verfahren, siehe ([30])) zur Lösung zu verwenden. Wenn die Dimension des Phasenraumes nicht zu groß ist, kann man hier auch blockweise vorgehen, indem man als Blöcke die Matrizen $\mathrm{D}F(y_{N-1})$, $\mathrm{D}F^t(y_{N-1})$ (Seitendiagonalblöcke) und $[\mathbf{1} + \mathrm{D}F\,\mathrm{D}F^t(y_{N-1})]$ (Hauptdiagonalblöcke) nimmt. Man bemerke, daß diese Blöcke stets invertierbar sind, falls F diffeomorph ist. Dann ist nämlich $\mathrm{D}F$ isomorph und deshalb auch $\mathrm{D}F\,\mathrm{D}F^t$. Diese Matrix hat jedoch nur positive Eigenwerte, die somit strikt positiv sein müssen. Dieses bleibt natürlich richtig, wenn die Einheitsmatrix addiert wird.

Durch diese Matrixinversion erhält man dann eine Lösung für die λ_n, die man in die Gleichung für die θ einsetzen kann. Die so korrigierte Zeitreihe dient dann als Start für eine neue Iteration. Die Frage ist, wann man das Blockweise–Gauss–Seidel–Verfahren abbricht. In den in dieser Arbeit behandelten Fällen zeigte sich, daß es für die Gesamtperformance ausreichte, nur einen einzigen Schritt durchzuführen, d.h. einfach die Hauptdiagonale zu invertieren und den Rest zu ignorieren. Man kann dan die approximative Lösung für λ_n konkret angeben:

$$\lambda_n \cong \left[\mathbf{1} + \mathrm{D}F\,\mathrm{D}F^t(y_{n-1})\right]^{-1} \left(F(y_{n-1}) - y_n\right).$$

Einsetzen in die Gleichung für θ ergibt

$$\begin{aligned}
\theta_1 &= -\,\mathrm{D}F^t(y_1)\left[\mathbf{1} + \mathrm{D}F\,\mathrm{D}F^t(y_1)\right]^{-1}\left(F(y_1) - y_2\right) \\
\theta_n &= \left[\mathbf{1} + \mathrm{D}F\,\mathrm{D}F^t(y_{n-1})\right]^{-1}\left(F(y_{n-1}) - y_n\right) \\
&\quad - \mathrm{D}F^t(y_n)\left[\mathbf{1} + \mathrm{D}F\,\mathrm{D}F^t(y_n)\right]^{-1}\left(F(y_n) - y_{n+1}\right), \\
\theta_N &= \left[\mathbf{1} + \mathrm{D}F\,\mathrm{D}F^t(y_{N-1})\right]^{-1}\left(F(y_{N-1}) - y_N\right)
\end{aligned} \tag{4.4}$$

Dies ist die Endformel für das *Diagonalverfahren*, wie es im folgenden genannt werden soll. Die wesentlichen Schritte seien noch einmal kurz zusammengefaßt:

1. Das Minimisierungsproblem unter Nebenbedingungen (4.3) wird aufgestellt.
2. Die Nebenbedingungen werden um die approximative Lösung y^N linearisiert und dann die Methode der Lagrangeschen Multiplikatoren angewandt.
3. Es ergeben sich zwei gekoppelte Gleichungssysteme für die Korrekturen und die Multiplikatoren. Ein Gleichungssystem für die Multiplikatoren läßt sich abkoppeln und per Gauss-Seidel-Iteration lösen. Durch Rücksubstitution erhält man eine Formel für die Korrekturen in Abhängigkeit von y^N.
4. Man erhält eine neue approximative Lösung und beginnt mit Punkt 2.
5. Empirische Beobachtung zeigt, daß es (jedenfalls für die hier behandelten Systeme) ausreicht, das Gauss-Seidel-Verfahren schon nach dem ersten Schritt abzubrechen.

Ergebnisse des Diagonalverfahrens

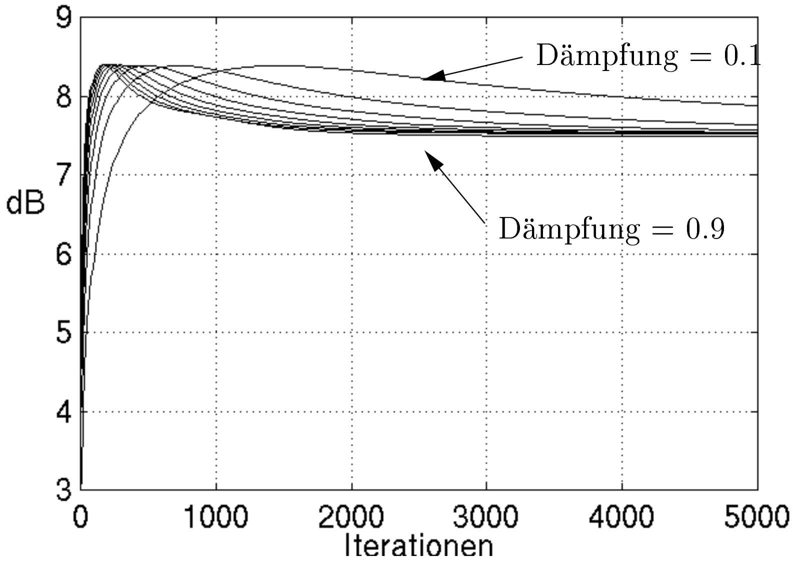


Abbildung 4.2: dB–Verbesserung für Hénon-Zeitreihe mit 8,4 dB SNR

Der Algorithmus wurde in C implementiert und auf das Hénon–System sowie die *Cat–Map* angewendet. Die Gleichungen für das Hénon–System wurden im Abschnitt über den Kalman–Filter bereits besprochen. Die Cat–Map ist durch

$$\begin{pmatrix} x_{n+1} \\ y_{n+1} \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ 1 & 2 \end{pmatrix} \begin{pmatrix} x_n \\ y_n \end{pmatrix} \bmod 1$$

gegeben. Es handelt sich um eine diffeomorphe Abbildung, wenn man das Einheitsquadrat, auf dem die Cat–Map definiert ist, mit dem 2–Torus identifiziert. Die Cat–Map ist hyperbolisch und chaotisch. Für weitere Eigenschaften der Cat–Map siehe [14]. Beide Systeme wurden nur mit additivem Meßrauschen gestört. Die Observable G war einfach die Identität.

Implementiert wurde im wesentlichen die Formel (4.4), allerdings mit Dämpfung, also

$$x_{\mathrm{neu}} = x_{\mathrm{alt}} + \delta \cdot \theta$$

mit Dämpfungsparameter δ. Aus der Herleitung der Formel (4.4) ergibt sich schon, daß häufiges Iterieren notwendig ist. Der Algorithmus läuft jedoch sehr schnell. Für eine 2–dimensionale Zeitreihe des Hénon–Systems der Länge 2000 werden auf einer Ultrasparc 1 nur 2.7 sec benötigt, um 100 Iterationen durchzuführen. Deshalb sind viele Iterationen kein Problem, auch da der Algorithmus völlig stabil lief. Es war jedenfalls in den hier geschilderten Fällen nicht notwendig, zu große Korrekturvorschläge des Algorithmus abzufangen. Abbildung 4.2

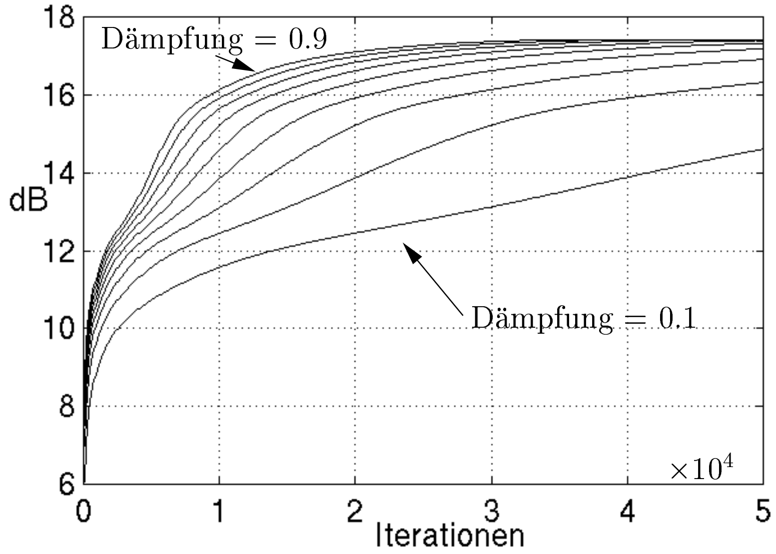


Abbildung 4.3: dB–Verbesserung für Hénon-Zeitreihe mit 11,7 dB SNR

zeigt die dB–Verbesserung des Diagonalverfahrens angewandt auf eine Zeitreihe des Hénon–Systems mit additivem Rauschen von 8,4 dB SNR. Die Länge der Zeitreihe betrug 2000 Punkte. Auf der Abszisse sind die Iterationen aufgetragen. Der Scharparameter ist die Dämpfung δ. Es ist deutlich erkennbar, daß ein kleineres δ keine besseren Ergebnisse liefert und obendrein die Zahl der benötigten Iterationen vergrößert. Es zeigte sich jedoch, daß für $\delta > 0,95$ das Verfahren nicht stabil lief und eine radikale Verschlechterung der Ergebnisse auftrat.

Weiterhin läßt die Graphik vermuten, daß die dB–Verbesserung absättigt. Experimente mit sehr hohen Iterationen bestätigten, daß keine Verbesserung oder Verschlechterung eintritt, sondern das Ergebnis konstant bleibt. Offensichtlich ist jedoch auch, daß dieser asymptotische Wert nicht mit dem Maximum übereinstimmt. Mit anderen Worten, die Kurven sind nicht monoton steigend, was natürlich unbefriedigend ist. Der Grund ist jedoch nicht eine allgemeine Verschlechterung des SNR, sondern das Auftreten von Glitches, welche nach und nach wieder auf die Größenordnung des Rauschens anwachsen und, obwohl sie isoliert liegen, die mittlere dB–Verbesserung beeinträchtigen. Die Glitches, die wir hier vorfinden, scheinen ihre Ursache genau in den Mechanismen zu haben, die im Abschnitt 4.3 erläutert wurden. Bevor das näher betrachtet wird, kurz zu Abbildung 4.3.

Hier wurde das Diagonalverfahren auf eine Zeitreihe des Hénon–Systems mit additivem Rauschen von 11,73 dB SNR angewandt, die Iterationen laufen im Gegensatz zum eben besprochenen Fall bis 50000. Man beobachtet im wesentlichen dasselbe, nur daß die Kurven monoton steigen und wesentlich mehr Iterationen erforderlich sind, bis die dB–Verbesserung stationär wird. Glitches treten hier eben-

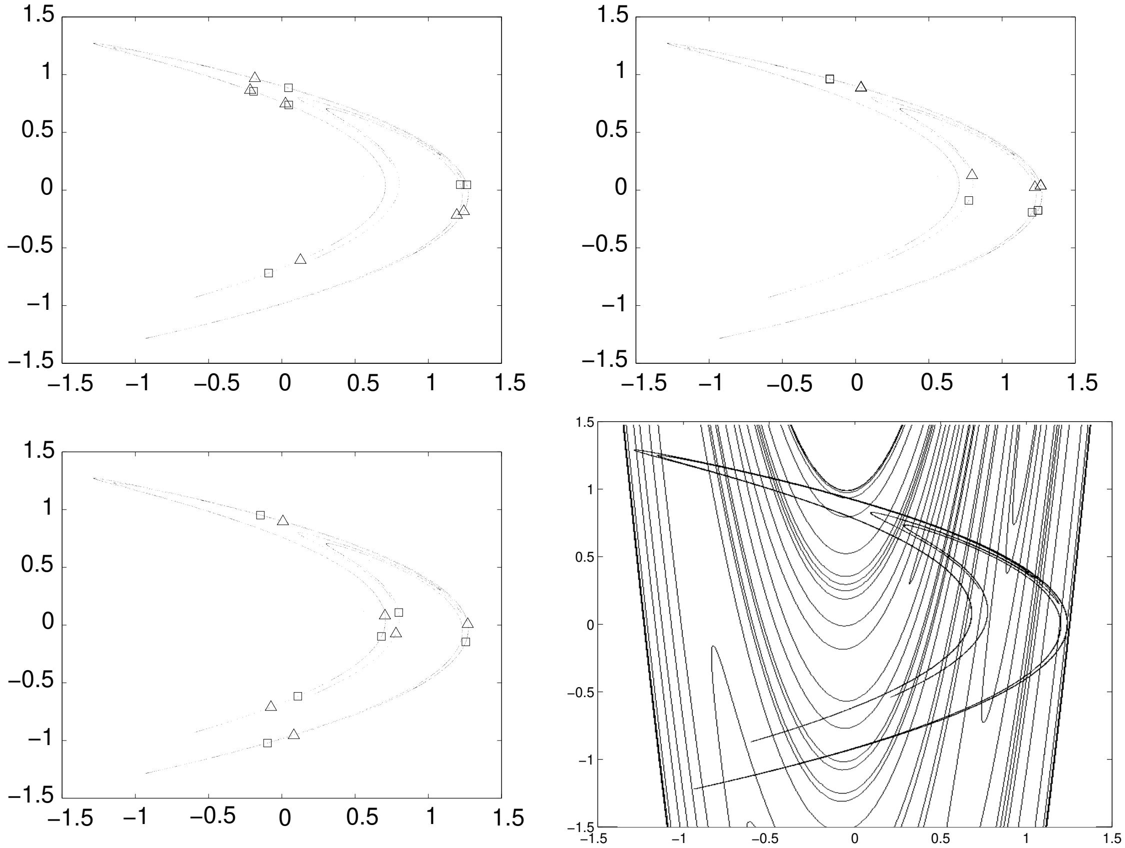


Abbildung 4.4: Glitches und tatsächliche Punkte auf dem Hénon–Attraktor sowie stabile und instabile Mannigfaltigkeiten

falls auf, nur haben sie eine geringere Höhe (ebenfalls in der Größenordnung der Rauschamplitude), was vielleicht ihren geringeren Einfluß zur dB–Verbesserung ausmacht.

Abbildung 4.4 zeigt den Hénon–Attraktor nach Rekonstruktion mit dem Diagonalverfahren. Ausgangspunkt war eine Zeitreihe der Länge 2000 mit additivem Rauschen von 8,4 dB. Iteriert wurde 5000 mal. Abbildung 4.5 zeigt die quadrierte Abweichung zwischen der rekonstruierten und der Originalzeitreihe. Die 18 größten Glitches sind mit einem Quadrat markiert. Die zugehörigen Punkte auf dem Attraktor finden sich in Abbildung 4.4 wieder: Die ersten sechs links oben, die nächsten sechs rechts davon und die wiederum nächsten sechs links unten, jeweils mit Quadraten markiert. die zugehörigen tatsächlichen Werte der Hénon–Zeitreihe sind mit Dreiecken markiert. In der Abbildung 4.4, rechts unten schließlich ist der Attraktor mit einer Schar von stabilen Mannigfaltigkeiten abgebildet (die instabilen Mannigfaltigkeiten bilden den Attraktor selbst). Ein

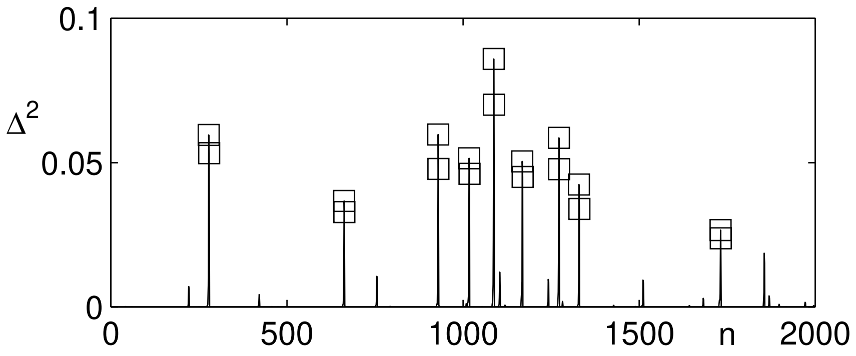


Abbildung 4.5: Glitches im quadratischen Fehler

Vergleich der übrigen Plots mit diesem zeigt, daß die Glitches in Bereichen tangentialer Berührung auftreten, vor allem in solchen, wo die Berührung sehr flach ist. Dort besteht eine kurze homokline Verbindung, welche nach den Ausführungen in Abschnitt 4.3 das Auftreten von Glitches stark begünstigt. Daß die in der Zeichnung abgebildeten Dreiecke und Quadrate jeweils paarweise homoklin sind, läßt sich in der Zeichnung jedoch nur erahnen. Eine Vergrößerung aus dem oberen Ast des Attraktors stellt Abbildung 4.6 dar. Hier ist ebenfalls der rekonstruierte Punkt mit einem Quadrat markiert, der tatsächliche Punkt mit einem Dreieck. Die gestrichelte stabile Mannigfaltigkeit wurde allerdings nicht berechnet, sondern aus den umliegenden interpoliert. Somit ist zwar nicht gesichert, daß die beiden Punkte homoklin sind. Daß sie jedoch sehr nahe an einem homoklinen Punktepaar liegen ist kaum zu bezweifeln.

Der Algorithmus wurde wie gesagt auch auf die Cat–Map angewandt. Da die Cat–Map hyperbolisch ist, ist hier zu erwarten, daß zumindest für kleine Rauschamplituden keine Glitches auftreten, was auch numerisch bestätigt wurde. Allerdings hat die Cat–Map den Nachteil, daß sie aufgrund der modulo-Operation analytisch schwer zu handhaben ist. Bei Bildung von Differenzen zwischen zwei Phasenraumpunkten hat man immer mehrere Möglichkeiten. In der hier vorgestellten Implementation wurde als Differenzvektor zweier Phasenraumpunkte stets der mit dem kleinsten Betrag gewählt. Es zeigt sich, daß der Algorithmus auf jeden Fall für SNR $\geq$ 7,5 dB optimal läuft. Die dB–Verbesserung ist in Abbildung 4.7 dargestellt. Zuerst ist die dB–Verbesserung jedoch negativ. Es zeigen sich in diesem Stadium noch viele Fehler vom Betrag der Größenordnung 1, was ein Problem der modulo-Arithmetrik ist. Mit höherer Iteration sättigt die Kurve ab. Es handelt sich jedoch um ein Artefakt, was durch große Fehler an den Enden der Zeitreihe hervorgerufen wird. Durch Verlängern der Zeitreihe kann die dB–Verbesserung noch erheblich gesteigert werden, da dann die Randfehler weniger stark ins Gewicht fallen. Bis auf diese Randpunkte fällt der Fehler nämlich bis auf die Rechengenauigkeit ab. Isolierte Glitches treten nicht auf.

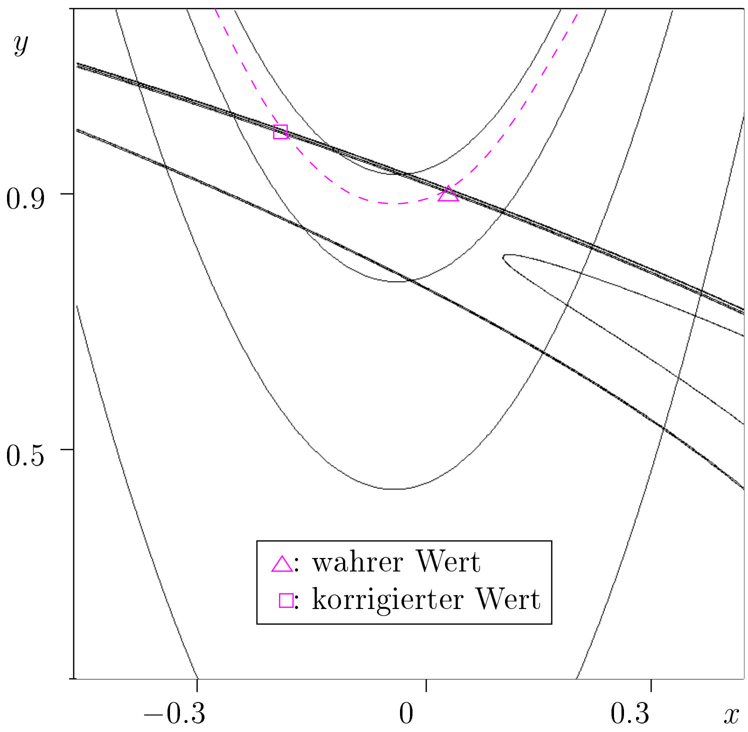


Abbildung 4.6: Vergrößerung des Hénon–Attraktors mit Glitch und tatsächlichem Punkt sowie stabile und instabile Mannigfaltigkeiten

Bei höheren Rauschamplituden (SNR $<$ 7,5 dB) hatte der Algorithmus seine Schwierigkeiten. Eine genauere Untersuchung wäre hier eigentlich noch angebracht. Es ist nämlich nicht auszuschließen, daß durch eine bessere Darstellung des Torus der Algorithmus weniger Schwierigkeiten hat, die korrekten Abstände zwischen zwei Phasenraumpunkten zu berechnen. Es könnten dann möglicherweise noch bessere Ergebnisse erziehlt werden, was natürlich wünschenswert wäre. Es konnte zwar die Vermutung bestätigt werden, daß für kleine Rauschamplituden keine Glitches mehr auftreten. Daß die maximale Rauschamplituden jedoch so klein ausfallen würde, ist eigentlich etwas enttäuschend.

4.5 NNS$_2$ und NNS$_\infty$

Bei den im letzten Kapitel besprochenen Least–Squares–Algorithmen minimiert man numerisch einen Ausdruck unter Berücksichtigung gewisser Nebenbedingungen. Die Nebenbedingungen schränken die Menge der möglichen Lösungen ein auf solche, die der Dynamik genügen. Die Idee der nun vorgestellten Algorithmen ist, diese Nebenbedingungen nicht durch analytische Gleichungen, sondern durch eine (möglichst große) Menge von Referenzpunkten zu manifestieren. Diese Referenzpunkte bilden einfach einen sehr langen typischen Orbit des Systems.

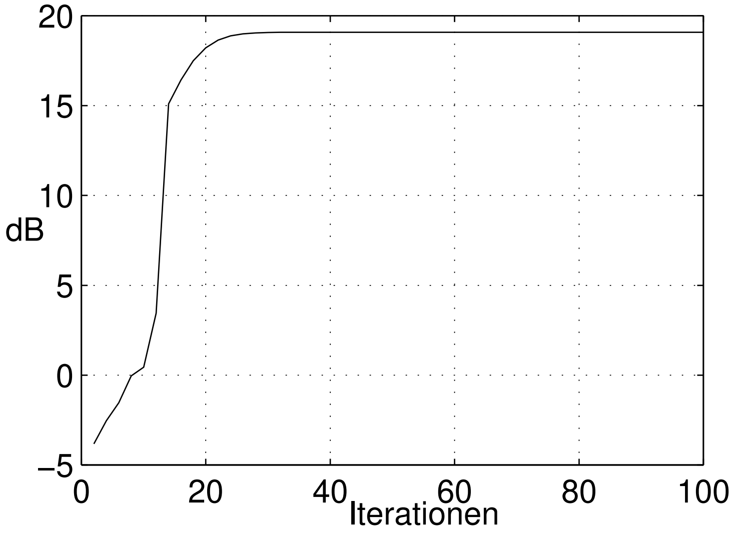


Abbildung 4.7: dB–Verbesserung für die Cat–Map mit 7,5 dB SNR

Der Algorithmus kennt also die dynamischen Gleichungen nicht, kann aber auf einen Referenzorbit zurückgreifen, der erstens möglichst lang, zweitens möglichst rauschfrei und drittens typisch für das System sein sollte.[3] Dieser Referenzorbit sei mit $\hat{x}^K$ bezeichnet. Hat man eine Zeitreihe y^N gegeben, so ist der negative Logarithmus der Likelyhoodfunktion eines Orbits x^N durch

$$H(x^N) = \sum_{i=0}^{N} (y_i - x_i) S^{-1} (y_i - x_i)$$

gegeben. Sind x_n und y_n Vektoren im R^m, so kann dieser Ausdruck auch als Abstand von x^N und y^N in einem $R^{m \cdot N}$–dimensionalen Raum aufgefaßt werden. Der Algorithmus macht nun einfach folgendes: Zu jedem n wird ein Stück Orbit $y_n^{n+\beta}$ betrachtet mit einer noch festzulegenden Zahl β. Zu diesem Orbit wird die Likelyhoodfunktion

$$H(x_n^{n+\beta}) = \sum_{i=n}^{n+\beta} (y_i - x_i) S^{-1} (y_i - x_i)$$

aufgestellt. Diese Likelyhoodfunktion wird nun minimiert, wobei als Orbits $x_n^{n+\beta}$ gerade Stücke des Referenzorbits zugelassen werden.

[3] Die Idee für diesen Algorithmus entstand, als nach einer Möglichkeit gesucht wurde, die Wahrscheinlichkeit $p(x_t|y^t)$ für das Hénon–System zu berechnen. Hierzu muß man die inverse Abbildung iterieren, die jedoch instabil ist. Die Trajektorien, deren Anfangspunkte nicht auf dem Attraktor liegen, divergieren sehr schnell. Andererseits sind nur Punkte auf dem Attraktor wirklich interessant. Die Abbildung wie auch ihre Inverse wurde deshalb mit einem langen Orbit realisiert, dessen Punkte den Attraktor dicht genug bedecken.

Man kann sich leicht überlegen, daß diese Vorgehensweise auch folgendermaßen formuliert werden kann: Der Referenzorbit $\hat{x}^K$ wird β–dimensional eingebettet. Damit ist gemeint, daß man die Vektoren

$$\hat{X}_n := (x_n, x_{n+1}, x_{n+2}, \ldots, x_{n+\beta-1})$$

bildet und als Vektoren im $\mathbb{R}^{m\cdot\beta}$ auffaßt. Das Minimierungsproblem ist nun nichts anderes als eine Nachbarsuche im $\mathbb{R}^{m\cdot\beta}$. Zu jedem $y_n^{n+\beta}$ sucht man nach dem nächsten Nachbar unter den X_n. Als Metrik dient hier gerade die Funktion $H(x_n^{n+\beta})^{1/2}$.

Die Weiterverarbeitung der X_n kann sehr verschieden erfolgen. Ziel ist ja, aus den X_n eine neue Zeitreihe zu erzeugen. In dieser Arbeit wurde folgendermaßen verfahren: Für jedes n wurde die gewichtete Summe

$$\tilde{X}_n = \alpha \cdot y_n^{n+\beta} + (1-\alpha) \cdot X_n$$

gebildet. Diese $\tilde{X}_n$ enthalten dann neue Vorschläge für rekonstruierte Punkte y_n^{neu}, $n = 0 \ldots N$. Um diese zu erhalten, kann man aber die Einbettung nicht einfach rückgängig machen, da die $\tilde{X}_n$ im allgemeinen keine Delaystruktur haben, d.h. es ist im allgemeinen

$$\tilde{X}_n^{(i)} \neq \tilde{X}_{n+k}^{(i-k)}.$$

Um y_n^{neu} zu konstruieren, wurde in dieser Arbeit eine gewichtete Summe aus den entsprechenden Einträgen in den X_n gebildet, also

$$y_n^{\text{neu}} = \sum_{i=0}^{\beta-1} \gamma_i \tilde{X}_{n-i}^{i+1}$$

mit Gewichtsfaktoren γ.

Man beachte, daß auch andere Metriken denkbar sind, sofern der Nachbarsuchalgorithmus dies zuläßt. Es wurde im Abschnitt über shadowing–Methoden bereits erwähnt, daß die richtige Metrik für shadowing eigentlich die Maximumnorm wäre. Tatsächlich wurde in dieser Arbeit auch diese Metrik getestet, wobei die Unterschiede zwischen den Resultaten allerdings kaum nennenswert sind. Die Algorithmen sollen im Folgenden NNS_2 im Falle der durch $H(x_n^{n+\beta})^{1/2}$ gebildeten Norm und NNS_∞ im Falle der Maximumnorm genannt werden. NNS bedeutet einfach *Nearest–Neighbor–Shadowing.*

Die Algorithmen bestehen im wesentlichen aus einem effizienten Nachbarsuchalgorithmus, der mit verschiedenen Metriken arbeiten kann. In dieser Arbeit wurde ein Algorithmus verwendet, der auf den *Anna*–Algorithmus von James McNames basiert (siehe [23]). Die hier verwendete Version wurde von Christian Merkwirth erweitert und implementiert. Das Grundgerüst dieses Algorithmus ist ein binärer Baum, in dessen Blätter die Punkte des Referenzdatensatzes eingeordnet werden. Zum Aufbau des Baumes wird der Datensatz in zwei disjunkte

Teilmengen eingeteilt. Um jede dieser beiden Teilmengen kann man eine Kugel legen, die alle Punkte der jeweiligen Teilmenge enthält. Diese Kugeln denkt man sich so gewählt, daß ihr Radius minimal ist. Die Einteilung erfolgt nun so, daß diese Kugeln möglichst disjunkt liegen. Dann wiederholt man diese Teilung bei den beiden Teilmengen usw., bis man den gesamten Referenzdatensatz in einer baumartigen Struktur von Teilmengen untergebracht hat und bei der letzten Unterteilung in jede Teilmenge etwa noch 50 Punkte fallen. Bei dieser Einteilung muß ein guter Kompromiß zwischen Optimalität der Einteilung (die nicht so einfach zu quantifizieren ist) und dem Rechenaufwand gemacht werden. Zur Suche eines Nachbarpunktes läuft man nun von oben nach unten durch den Baum und entscheidet allein mit Hilfe der Dreiecksungleichung, in welcher der einhüllenden Kugeln (d.h. in welchem Ast des Baumes) noch nach Nachbarn gesucht werden muß. Ist die Einteilung gut gemacht und der Datensatz nicht zu pathologisch, so können mit wenigen Auswertungen der Dreiecksungleichung große Teile des Referenzdatensatzes von der Suche ausgeschlossen werden. Man beachte, daß für einen gegebenen Referenzdatensatz der Baum nur einmal aufgebaut werden muß. Der Algorithmus entfaltet also vor allem dann seine Stärke, wenn viele Nachbarsuchen in ein– und demselben Referenzdatensatz durchgeführt werden müssen. Außerdem ist der Algorithmus für die Anwendung mit jeder beliebigen Metrik geeignet.

Die Algorithmen NNS_2 und NNS_∞ wurden wieder auf die Cat–Map und die Hénon–Abbildung angewandt. Dabei haben sich als Gewichtsfaktoren

$$\gamma_i = 0.1 \text{ für } i = 0 \text{ und } i = \beta - 1$$
$$\gamma_i = 1 \text{ sonst}$$

als brauchbar erwiesen. Als Metriken wurden wie gesagt die Funktion $H(x_n^{n+\beta})^{1/2}$ im Falle von NNS_2 und die Maximumsmetrik im Falle von NNS_∞ verwandt.

Als erstes werden die Ergebnisse von NNS_∞ bei Anwendung auf die Cat–Map besprochen. Hier wurden für den Referenzdatensatz folgende Parameter gewählt:

Länge des Referenzdatensatzes:	50.000
Rauschen im Referenzdatensatz:	14,5 dB

Der Referenzdatensatz war also selbst mit einem Rauschen verunreinigt. Dies wurde im Hinblick auf praktische Anwendung untersucht, wo man damit rechnen muß, daß auch der Referenzdatensatz nicht absolut frei von Rauschen ist. Abbildung 4.8 zeigt die dB–Verbesserungen für zwei Zeitreihen der Cat–Map mit einem SNR von 9,8 dB (obere Graphik) und $-0,2dB$ (untere Graphik). „Einbettung“ bezeichnet β. Mit Iterationen sind die Iterationen des gesamten Algorithmus NNS_∞ gemeint. Es sind nur die Fälle Einbettung $= 5, \ldots, 10$ abgebildet. Für kleinere Einbettungen als 5 lief der Algorithmus nicht gut, bis auf den Fall Einbettung $= 1$ natürlich, in dem der Ausgangsdatensatz einfach durch Punkte des Referenzdatensatzes ersetzt wird, die im Phasenraum am dichtesten liegen. Für Einbettungsdimensionen > 5 fällt die allgemeine dB–Verbesserung in der oberen Graphik wieder stark ab. Der Algorithmus hängt also leider sehr von der Wahl

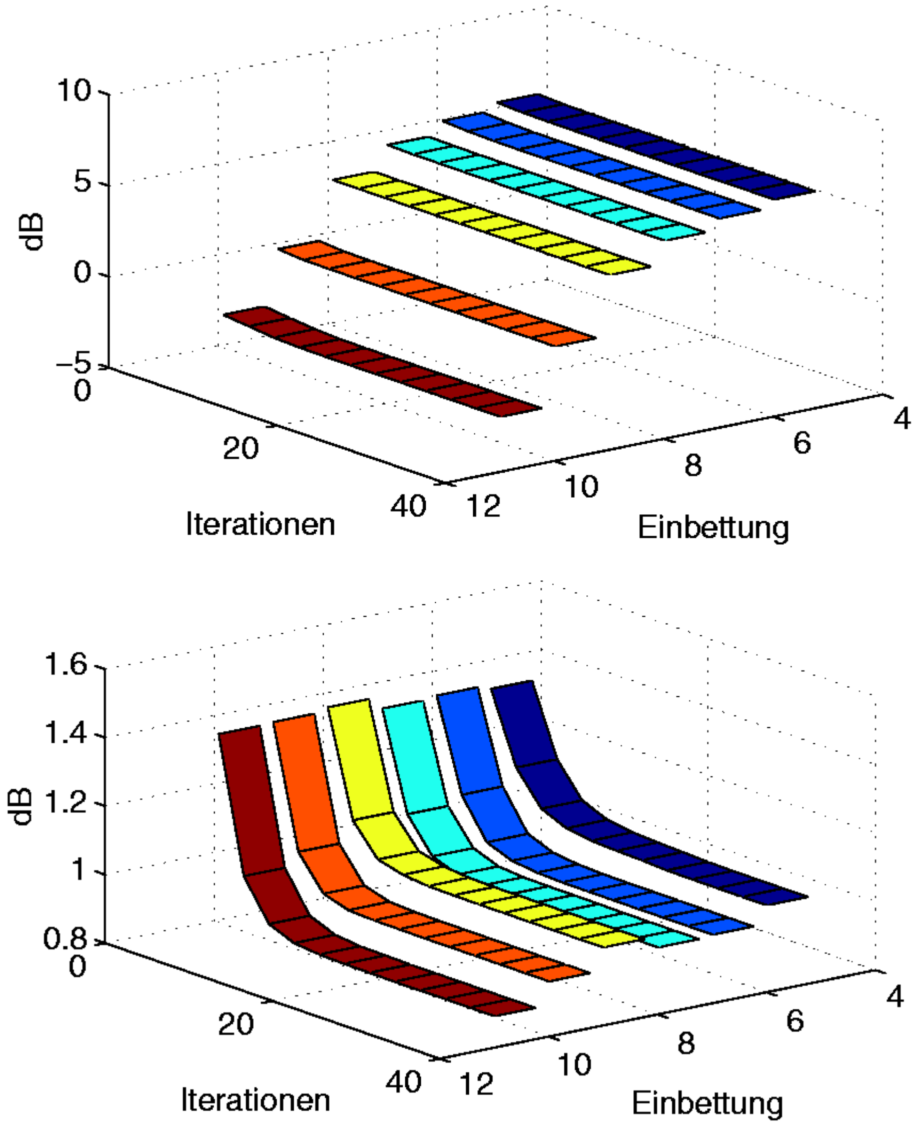


Abbildung 4.8: Performance von NNS_∞ für die Cat–Map

dieses Parameters ab. Im darunter dargestellten Fall sieht man diese Abhängigkeit nicht. Hier ist jedoch die allgemeine Performance schlechter. Man beachte, daß in allen Fällen das viele Iterieren offensichtlich sinnlos ist. Im Falle von $-0,2$ dB Rauschen sinkt die dB–Verbesserung sogar ganz wesentlich wieder ab. Gründe für diese eher mäßigen Ergebnisse für großes Rauschen wurden im letzten Abschnitt bei Erläuterung des Diagonal–Verfahrens schon erklärt. Es sei noch angemerkt, daß man für unverrauschte Referenzsätze qualitativ dasselbe Verhalten beobachtet. Die dB–Verbesserung zumindest für den Fall von 9,8 dB bei optimaler Parameterwahl ist jedoch noch etwas besser, was ja auch zu erwarten ist. Es ergibt sich z.B. für die Parameter Einbettungsdimension = 5, Iterationen = 20 eine dB–Verbesserung von 9,53 dB.

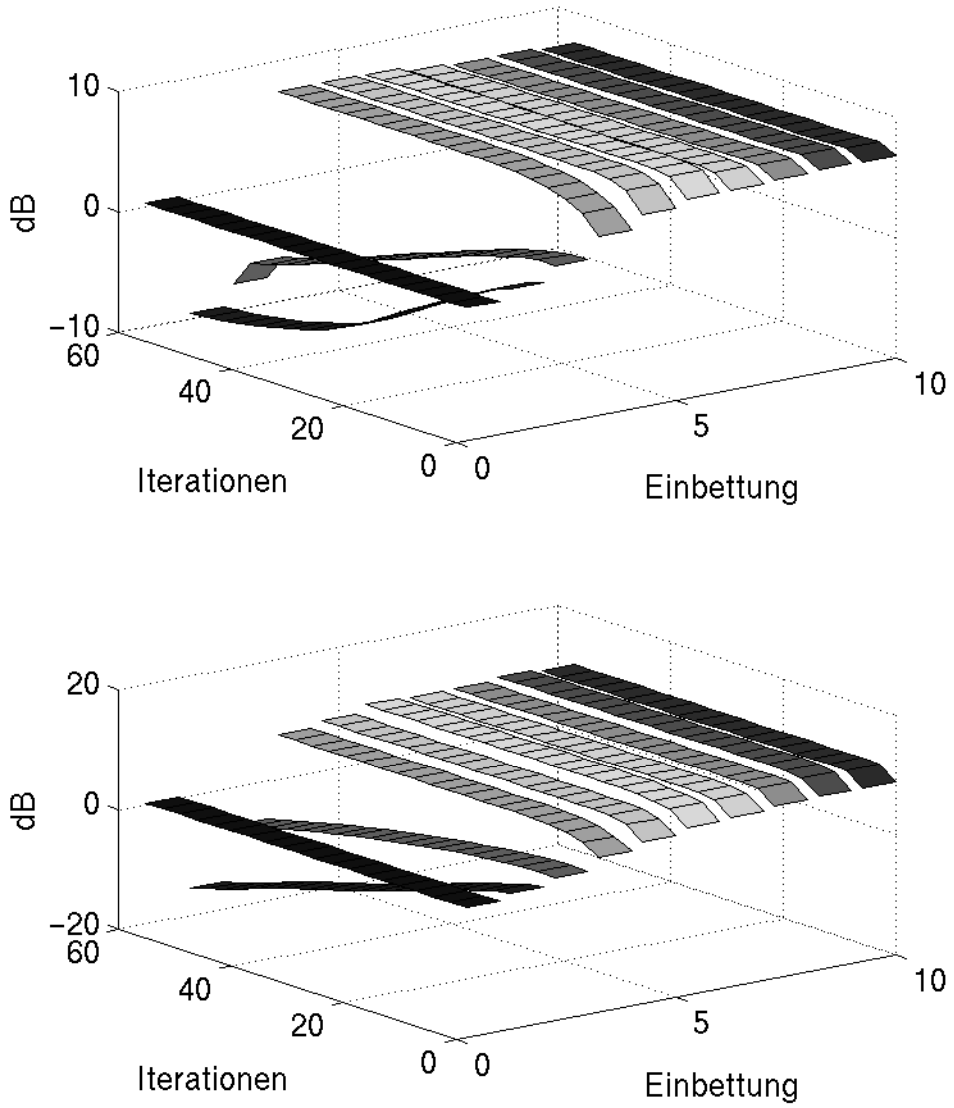


Abbildung 4.9: Performance von NNS_∞ für die Hénon-Abbildung

Für die Hénon-Abbildung brachte NNS_∞ zwar keine quantitativ wirklich besseren Ergebnisse, die Performance schien aber allgemein nicht so stark von der Parameterwahl abzuhängen. Wieder wurde ein leicht verrauschter Referenzdatensatz gewählt:

Länge des Referenzdatensatzes:	50.000
Rauschen im Referenzdatensatz:	18.4 dB

Abbildung 4.9 zeigt die dB–Verbesserungen für zwei Zeitreihen der Hénon-Abbildung mit einem SNR von 8,4 dB (obere Graphik) und 18,4 dB (untere

Graphik). Man beachte, daß im Vergleich mit Abbildung 4.8 hier die Orientierung der x– und y–Achse aus darstellungstechnischen Gründen vertauscht wurde. Die Performance ist ab Einbettungsdimension 4 akzeptabel. Erst dann scheinen die eingebetteten Vektoren des Referenzdatensatzes und des Eingangsdatensatzes genug dynamische Informnation zu enthalten. Danach treten kaum noch Verbesserungen ein. Auch die Iteration des Algorithmus ist hier offenbar überflüssig, sofern sie über etwa 20 hinausgeht. Auch dieser Algorithmus ist gegen die vielbesprochenen Glitches machtlos. Auf dieses Phänomen in den Ergebnissen der NNS–Algorithmen wird noch einmal kurz eingegangen, wenn die Performance von NNS_2 besprochen wird. Da das Verhalten der beiden Algorithmen in diesem Punkt völlig gleich ist, soll die Besprechung erst dort erfolgen.

In Abbildung 4.10 sind die Ergebnisse von NNS_2 für die Cat–Map dargestellt. Die Kurven bezeichnen die dB–Verbesserung aufgetragen über den Iterationen des Algorithmus für verschiedene Einbettungsdimensionen ($7 = \square$, $8 = \star$, $9 = \triangle$, 10 =Strichpunktlinie). Sowohl die Gewichtsfaktoren γ_i als auch die Referenzdaten sind wie für NNS_∞ gewählt. Die zu rekonstruierenden Zeitreihen besaßen wieder ein SNR von 9,8 dB (obere Graphik) und -0,2 dB (untere Graphik). Für andere Einbettungsdimensionen lief der Algorithmus nicht zufriedenstellend. Quantitativ ergeben sich kaum Unterschiede zu NNS_∞, wobei die Performance sowohl im Hinblick auf die optimale dB–Verbesserung als auch auf die brauchbaren Einbettungsdimensionen sogar noch etwas schlechter ist. Dies könnte an der bereits angesprochenen Tatsache liegen, daß für shadowing–Probleme die Maximumsmetrik zumindest theoretisch die bessere ist.

Für die Hénon-Abbildung (Abbildung 4.11) sind die Unterschiede zu NNS_∞ sogar noch weniger nennenswert. Die Daten und Parameter dieser Numerischen Simulation waren auch dieselben wie für den Fall NNS_∞. Wie gesagt erscheinen bei NNS_∞ und auch hier wieder Glitches. Abbildung 4.12 zeigt den quadratischen Fehler einer Zeitreihe des Hénon–Systems nach Anwendung von NNS_2. Es handelte sich um eine Zeitreihe mit 8,4 dB. Die Einbettungsdimension betrug 10. Ansonsten waren die Parameter und Daten dieselben wie für die bereits besprochenen Simulationen. Die Glitches sind nicht zu übersehen. Die eingesetzte Abbildung zeigt einen Ausschnitt. Der Bereich ist ein Abschnitt, in dem keine Glitches auftreten. Man sieht, daß die Rauschreduktion hier um Größenordnungen besser läuft. In diesem Abschnitt betrug die dB–Verbesserung 11,5 dB, während sie insgesamt nur 7,6 dB erreichte.

Als letztes sei noch ganz kurz die Anwendung von NNS_2 auf den Colpitts–Oszillator besprochen. Der Colpitts–Oszillator ist ein chaotischer elektronischer Schwingkreis ohne Anregung. Die Daten wurden mir freundlicherweise von Gerrit Langer zur Verfügung gestellt. Auf den von ihm verfaßten Technical Report [19] sei hier für die Gleichungen und deren Herleitung verwiesen. NNS_2 wurde mit einem Referenzdatensatz der Länge 50.000 ohne Rauschen angewandt. Der erste Test ergab für eine zu rekonstruierende Zeitreihe der Länge 2000 mit SNR=15,5 dB eine dB–Verbesserung von 3,36 dB. Bei einer Zeitreihe der Länge 2000 mit

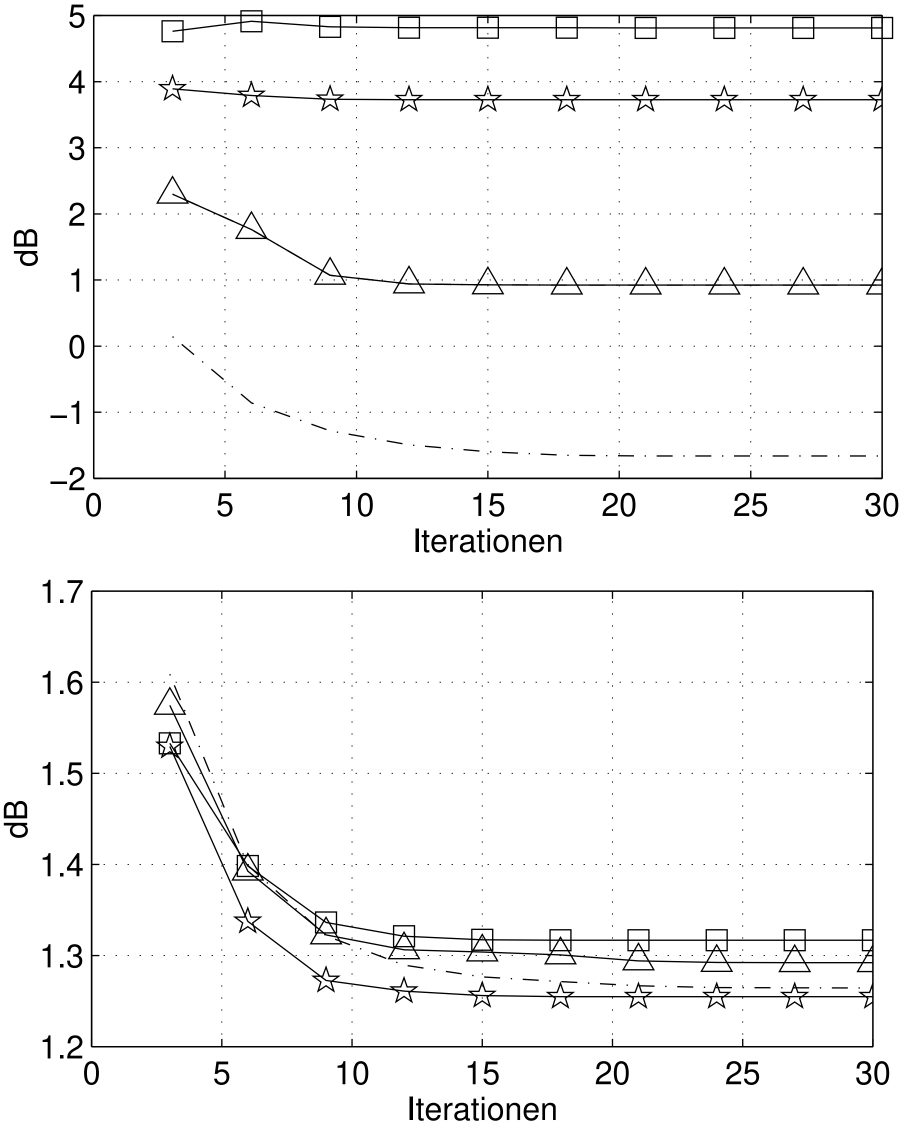


Abbildung 4.10: Performance von NNS_2 für die Cat–Map

SNR=12,5 konnte sogar eine dB–Verbesserung von 5,2 dB erzielt werden.

Zum Schluß soll eine kurze Zusammenfassung des Abschnitts gegeben werden. Insgesamt kann man festhalten, daß der hier vorgestellte Algorithmus mit seinen zwei Varianten durchaus brauchbare Ergebnisse liefert, wobei die Wahl der Parameter nicht immer ganz einfach ist. Der Unterschied zwischen beiden Varianten ist nur graduell. Speziell zu den hier durchgeführten Simulationen läßt sich sagen, daß für die Cat–Map eine besser angepaßte Metrik vielleicht noch etwas bringen könnte. Allerdings kann ein Grund für die Probleme bei der Cat–Map auch die Länge des Referenzdatensatzes sein kann. Während der Hénon–Attraktor mit ei-

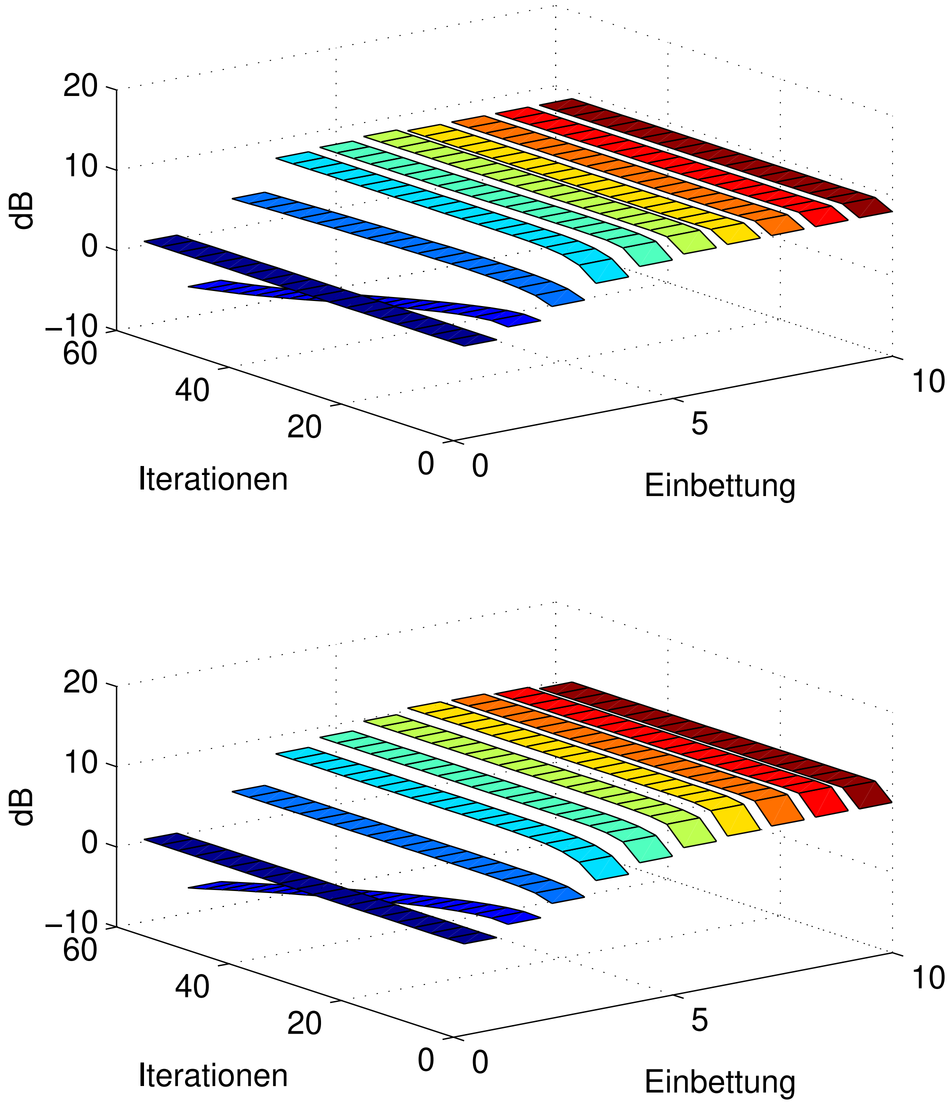


Abbildung 4.11: Performance von NNS_2 für die Hénon-Abbildung

ner Zeitreihe von 50.000 Punkten schon hinreichend dicht besetzt zu sein scheint, ist möglicherweise der Torus, der den Attraktor der Cat–Map bildet (die Cat–Map hat Determinante 1!) mit 50.000 Punkten noch zu dünn besetzt. Dieses Phänomen könnte ein grundsätzliches Problem der NNS–Algorithmen sein. Es ist offensichtlich, daß man für Attraktoren höherer Dimension zu große Referenzdatensätze benötigt, bei denen selbst der beste Nachbarsuchalgorithmus zu lange nachdenken muß.

Die Glitches treten auch bei den NNS–Algorithmen nach den gewohnten Gesetzmäßigkeiten auf. Man könnte sich deshalb einen Algorithmus nach dem Muster der NNS–Algorithmen denken, der gerade dieses Phänomen ausnutzt, um

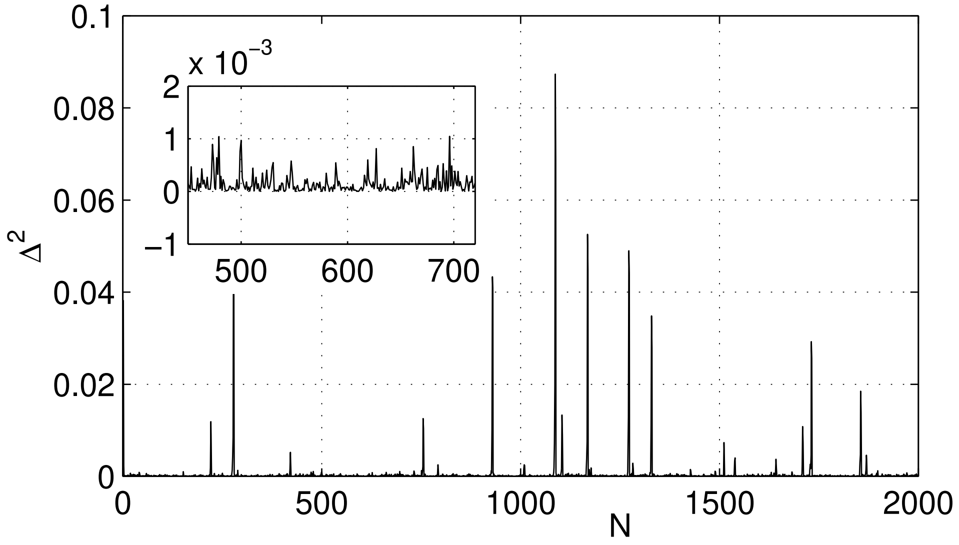


Abbildung 4.12: Der quadratische Fehler in einer Zeitreihe des Hénon–Systems nach Anwendung von NNS_2. Kleines Bild: Auschnitt

allein anhand eines Datensatzes zu entscheiden, ob in einem System tangentiale Berührungen auftreten oder nicht. Etwas euphorischer könnte man auch sagen, daß im Prinzip eine Unterscheidung zwischen hyperbolischen und nichthyperbolischen Systemen anhand eines langen Datensatzes möglich ist. Dazu würde man von dem Referenzdatensatz ein Stück abschneiden und nach der Technik der NNS–Algorithmen mit dem Referenzdatensatz vergleichen. Dieser Algorithmus hätte natürlich mit denselben Schwierigkeiten zu kämpfen wie die NNS–Algorithmen. Zu kurze Referenzdatensätze, die unter Umständen auch nicht ganz frei von Rauschen sind, würden hier Probleme bereiten.

4.6 Analyse von Zeitreihen ohne bekannte Dynamik

In allen bisher behandelten Problemen wurde selbstverständlich vorausgesetzt, daß die Dynamik bekannt ist. Natürlich hat man aber oft den Fall zu behandeln, daß eine explizite Dynamik nicht gegeben ist. Dann hat man vor der Filterung noch das Problem, ein geeignetes Modell zu entwerfen. Dieses Problem soll hier in seiner vollen Allgemeinheit nicht besprochen werden. Ein grundsätzliches, in der Praxis viel benutztes Verfahren besteht darin, sich auf eine Klasse von Dynamiken zu beschränken, die durch eine endliche Anzahl von Parametern beschrieben werden kann. Für diese Parameter versucht man dann anhand der gemessenen Daten möglichst gut konditionierte Gleichungen aufzustellen. Oft wählt man die Klasse so, daß die Parameter linear eingehen und wählt als Optimalitätskriterium den minimalen quadratischen Fehler. So erhält man lineare Gleichungssysteme.

Es zeigt sich, daß zur Lösung dieser Gleichungssysteme die Singulärwertzerlegung das optimale Werkzeug ist (siehe vor allem [28] für einen Überblick). Diese Methoden werden in [27] benutzt, um eine Dynamik zu konstruieren, mit deren Hilfe die Identifikation tatsächlicher und falscher Lyapunov-Exponenten möglich ist.

Darüberhinaus hat hat man in der Zeitreihenanalyse mit eingebetteten Daten zu tun. Das bedeutet, daß die Punkte auf einer Untermannigfaltigkeit liegen, die im allgemeinen weder den ganzen Raum noch einen linearen Unterraum umfaßt. Deshalb ist es oft nicht möglich, die dynamischen Gleichungen in expliziter Form aufzustellen. Es ist deshalb sinnvoll, von vornherein nach lokal gültigen Gleichungen der Form

$$\begin{aligned} G(x_n) &= 0, \\ F(x_n, x_{n+1}) &= 0, \end{aligned}$$

zu suchen. Die erste dieser Gleichungen definiert die Untermannigfaltigkeit, die zweite Gleichung gibt in impliziter Form die Dynamik. Auch für diese Form der Dynamik läßt sich ein Diagonalverfahren wie im letzten Abschnitt angeben. Die Formeln ergeben sich ganz analog.

In dieser Arbeit soll nur die Technik der *lokal linearen* Modelle besprochen werden. Eine ausfürliche Darstellung findet sich in dem Artikel von GRASSBERGER et al ([11]). Obendrein gehen wir davon aus, daß die Einbettung durch eine *Delayeinbettung einer skalaren Zeitreihe* realisiert wird. Die Dynamik ist dann einfach in der Delaystruktur der Zustandsvektoren verpackt. Zusätzlich hat man dann noch eine implizite Gleichung, die lokal die eingebettete Mannigfaltigkeit definiert. Das heißt in Formeln:

$$G(x_n) = 0, \tag{4.5}$$

$$x_n^{(\alpha)} = x_{n+i}^{(\alpha+i)} \text{ für } \alpha = 1 \ldots d \text{ und } i = 1 \ldots d - \alpha, \tag{4.6}$$

wobei $G : \mathbb{R}^m \to \mathbb{R}^d$ eine Submersion ist. Die eigentliche skalare Zeitreihe sei mit ξ_n bezeichnet. Mit anderen Worten $x_n = (\xi_n, \ldots, \xi_{n-d+1})$. Die Gleichung (4.5) wird nun lokal durch eine lineare Approximation ersetzt, also durch eine Gleichung der Form $G(x_n - b) = 0$ mit einer $d \times m$–Matrix G und einem m–dimensionalen Vektor b. Das Minimisierungsproblem lautet jetzt:

$$\text{min!} = \sum_{n=1}^{\nu} (x_n - y_n) P^{-1} (x_n - y_n) \tag{4.7}$$

unter den Nebenbedingungen

$$G(x_n - b) = 0 \tag{4.8}$$

und

$$x_n^{(\alpha)} = x_{n+i}^{(\alpha+i)} \text{ für } \alpha = 1 \ldots d \text{ und } i = 1 \ldots d - \alpha, \tag{4.9}$$

wobei ν die Anzahl der Punkte in einer lokal gewählten Umgebung ist. Man geht hier folgendermaßen vor:

1. Das Problem wird transformiert

$$\begin{aligned}\theta_n &:= \sqrt{P^{-1}}(x_n - y_n)\\ \eta_n &:= \sqrt{P^{-1}}y_n\end{aligned}$$

2. Man bildet den Mittelwert

$$\eta := \frac{1}{\nu}\sum_{n=1}^{\nu}\eta_n$$

   und die Kovarianzmatrix

$$\Gamma^{(\alpha\beta)} := \frac{1}{\nu}\sum_{n=1}^{\nu}(\eta_n - \eta)^{(\alpha)}(\eta_n - \eta)^{(\beta)}$$

3. Man berechnet die Eigenwerte λ_i und die normierten Eigenvektoren v_i von Γ und stellt die Projektion

$$\Psi := \sum_{i=1}^{\sigma}\langle v_i, \cdot\rangle v_i$$

   auf die σ kleinsten Eigenwerte von Γ auf.

4. Es gelten die Formeln

$$\theta_n = \Psi(\eta - \eta_n) \tag{4.10}$$

$$x_n = y_n + \sqrt{P}\theta_n \tag{4.11}$$

$$G = \Psi\sqrt{P^{-1}} \tag{4.12}$$

$$b = \frac{1}{\nu}\sum_{n=1}^{\nu}y_n \tag{4.13}$$

Diese Tatsachen beweist man mit Hilfe elementarer Analysis. Ausfürliche Darstellungen dieser Thematik dürften darüberhinaus in jedem Numerikbuch zu finden sein, z.B. in [28].

Der Rauschreduktionsalgorithmus nutzt die eben beschriebenen Tatsachen folgendermaßen aus: Um jedem (gestörten) Punkt y_n wird eine Kugel $B_n := B(y_n, \epsilon)$ gelegt mit noch zu bestimmendem Radius ϵ. Es wird die Menge $Y_n := \{y_k; y_k \in B_n\}$ aller Punkte bestimmt, die in dieser Kugel enthalten sind. Hierzu bietet sich wieder der im vorigen Abschnitt beschriebene Nachbarsuchalgorithmus an. Dann werden für dieses Y_n die oben aufgelisteten Schritte durchlaufen mit $\nu := \#Y_n$. Das liefert dann (nach Formel 4.11) eine Korrektur x_n für y_n. Zur kompletten

Lösung des Problems fehlt nun noch ein Detail. Die Gleichung (4.9) wurde ignoriert. Die x_n in Gleichung (4.11) haben keine Delaystruktur wie in (4.9) gefordert. Dieses führt zu verschiedenen Korrekturvorschlägen für die einzelnen ξ_n. Um daraus wieser eine Zeitreihe zu gewinnen, wird über alle Korrekturvorschläge gemittelt. Damit ergibt sich

$$\xi_n = \frac{1}{d} \sum_{i=0}^{d-1} x_{n+i}^{(i+1)}$$

Dieser Algorithmus, der im folgenden nach CAWLEY, HSU und SAUER CHS–Algorithmus genannt werden soll, besitzt natürlich einige freie Parameter, die geeignet zu wählen sind. Neben dem Kugeldurchmesser ϵ sind das die Einbettungsdimension m, die Anzahl der Nebenbedingungen d und die Gestalt der Matrix P. Dieses P bezeichnet eigentlich die Kovarianz des Rauschens. Wenn man die Zeitreihe jedoch einbettet, transformiert sich P in recht verwickelter Weise. Man kann dann eigentlich gar nicht mehr von Gauß–verteiltem (weißen) Rauschen ausgehen. Eigentlich bestehen dann Korrelationen zwischen den einzelnen eingebetteten Zustandsvektoren. Es ist deshalb sinnvoll, die Performance des Algorithmus durch Wahl von P zu verbessern.

Der CHS–Algorithmus wurde auf drei Zeitreihen des Henonsystems angewandt. Das SNR betrug 8,48, 10,2 und 13,2 dB (siehe hierzu Abbildungen 4.13, 4.14 und 4.15). Es ist deutlich erkennbar, daß die dB-Verbesserung *nicht* absättigt, sondern vielmehr schon nach wenigen Iterationen wieder schlechter wird. Mit fortschreitender Iteration sinkt das SNR sogar unter das ursprüngliche SNR ab. Der rekonstruierte Attraktor zeigt dann Verzerrungen, die immer stärker werden. Der Algorithmus muß also schon nach 4 bis 5 Iterationen abgebrochen werden.

Für alle hier betrachteten Fälle erreichte der Algorithmus bei einer Einbettungsdimension von 6 mit 3 Nebenbedingungen die beste Performance. Es ist völlig klar, daß extreme Parameterwerte keine guten Ergebnisse produzieren. Sicher ist, daß der Attraktor gut aufgespannt sein muß, Andererseits darf die Dichte der Punkte nicht zu klein sein. Außerdem soll der Algorithmus lokale Streuung der Daten, die durch Rauschen verursacht werden, herausfiltern, andererseits aber nicht diejenige Streuung, die durch die Dynamik entsteht. Warum aber diese Konstellation günstig ist und auch eine recht starke Abhängigkeit von der Parameterwahl zu beobachten ist, ist schwer zu sagen.

Etwas besser funktioniert der Algorithmus bei Daten des DUFFING–Oszillators. Es handelt sich hierbei um ein getriebenes dynamisches System in kontinuierlicher Zeit beschrieben durch die Gleichung

$$\ddot{x}(t) + \delta\dot{x} + x(x^2 - 1) = \gamma \cos \omega t.$$

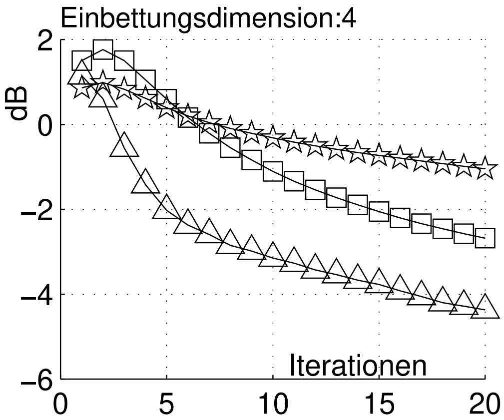
Es beschreibt z.B. ein gedämpftes Masse–Feder–System mit einer nichtlinearen Feder. Eine ausführliche Besprechung dieser Gleichung findet sich in [12]. Das System wurde mit $\omega = 2\pi \cdot 128\text{Hz}$, $\gamma = 760\text{mV}$ betrieben und einr Frequenz von

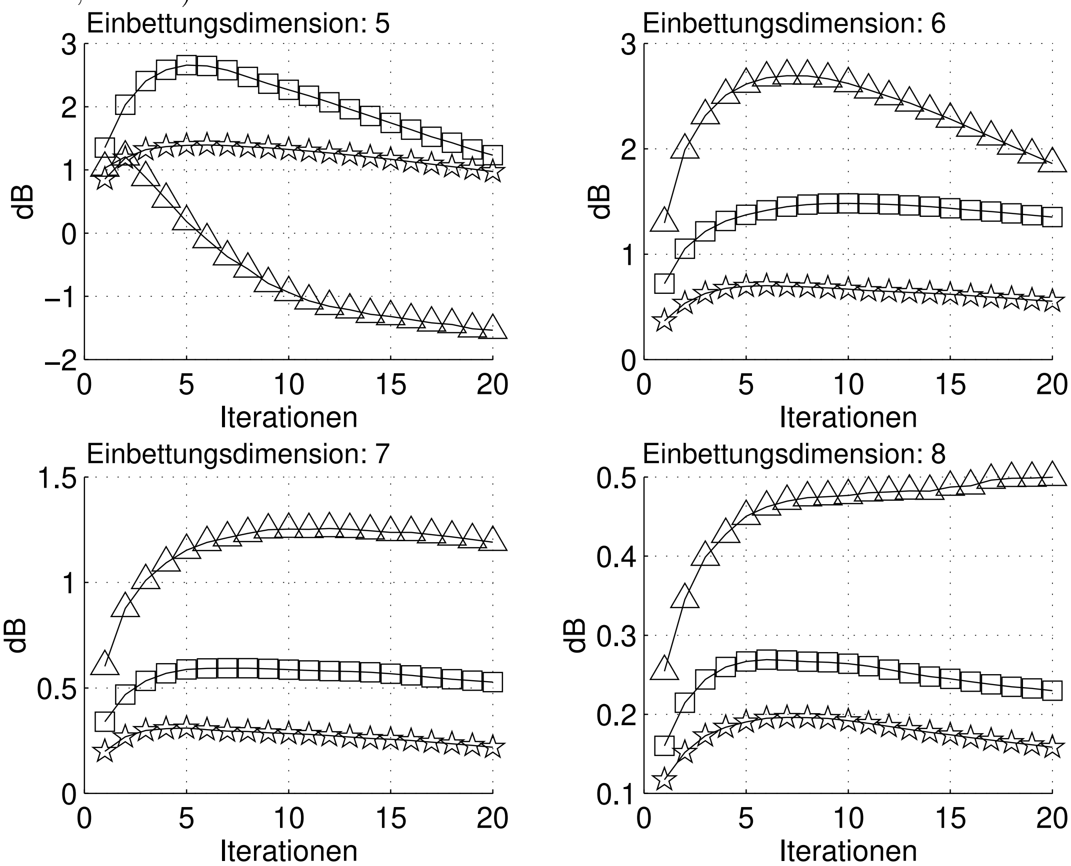
49 kHz abgetastet.[4] Die Ergebnisse des CHS–Algorithmus in diesem Fall finden sich in Abb. 4.16. Die Performance ist hier vor allem etwas stabiler und auch leicht besser. Auch hier sind die genauen Zusammenhänge zwischen Nebenbedingungen, Einbettungsdimension und den Ergebnissen schwer zu erklären. Eine allgemein bessere Performance gegenüber dem Hénon–System ist jedoch zu erwarten, da der Duffing–Oszillator nicht so chaotisch ist wie die Hénon–Abbildung. Ein Lyapunov-Exponent ist null, und das System ist lokal sehr viel besser linearisierbar als das Hénon–System.

Alles in allem stellt der CHS–Algorithmus sicherlich eine gute Methode dar, um Rauschen bei unbekannter Systemdynamik zu reduzieren, wenn man die notwendigen Vorsichtsmaßnahmen trifft. Vor allem bietet sich dieser Algorithmus an, wenn man statt dynamischer Gleichungen irgendein anderes Kriterium hat, ob die Daten gut sind oder nicht, z.B. das geschulte Auge. Es ist auch nicht klar, ob dieser Algorithmus den Varianten unterlegen ist, die lokal nichtlineare oder globale Modelle benutzen. Alle diese Algorithmen scheitern letztendlich daran, daß sie mit sehr sehr wenig Information auskommen müssen (keine Wahrschscheinlichkeitsverteilungen, keine dynamischen Gleichungen, nur die Zeitreihe) und deshalb immer die Tendenz haben, Overfitting zu betreiben, d.h. Information zu produzieren, die eigentlich nicht vorhanden ist.

[4] Die Daten wurden mir freundlicherweise von Gerrit Langer zur Verfügung gestellt. Der experimentelle Aufbau ist in [20] beschrieben.

Abbildung 4.13: Dargestellt sind die dB-Verbesserungen für eine Hénon–Zeitreihe mit 13,2 dB Rauschen. Getestet wurde für Einbettungsdimensionen von 4 bis 8. Scharparameter der Kurven sind die Nebenbedingungen ($1 = \star$, $2 = \square$, $3 = \triangle$).





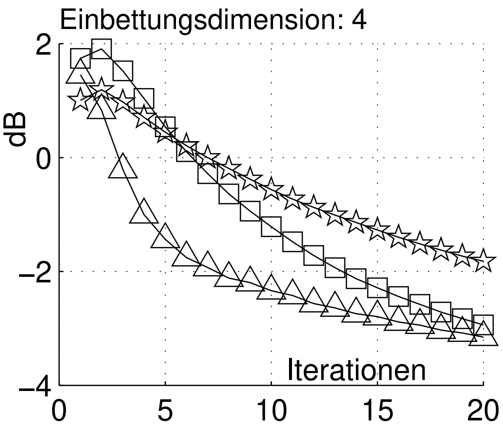
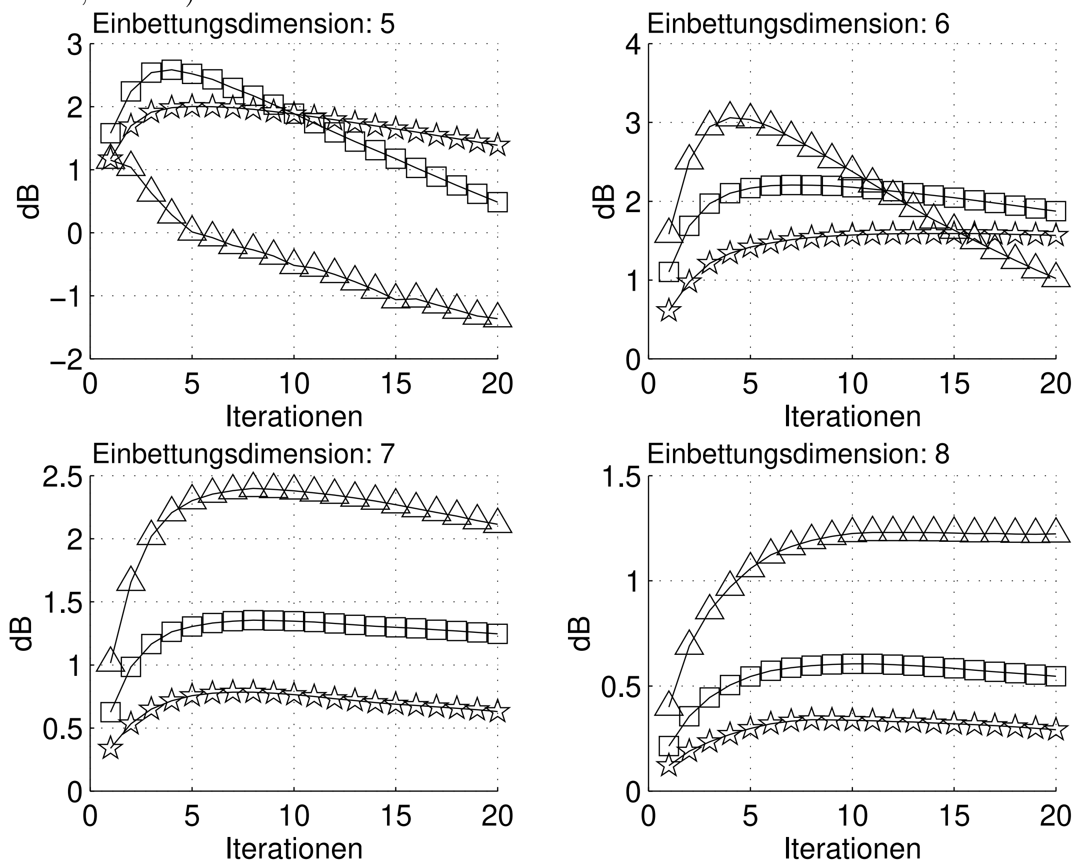


Abbildung 4.14: Dargestellt sind die dB-Verbesserungen für eine Hénon–Zeitreihe mit 10,2 dB Rauschen. Getestet wurde für Einbettungsdimensionen von 4 bis 8. Scharparameter der Kurven sind die Nebenbedingungen ($1 = \star$, $2 = \square$, $3 = \triangle$).



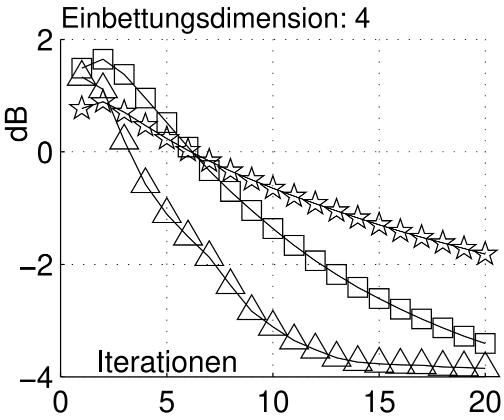


Abbildung 4.15: Dargestellt sind die dB-Verbesserungen für eine Hénon–Zeitreihe mit 8,48 dB Rauschen. Getestet wurde für Einbettungsdimensionen von 4 bis 8. Scharparameter der Kurven sind die Nebenbedingungen ($1 = \star$, $2 = \square$, $3 = \triangle$).

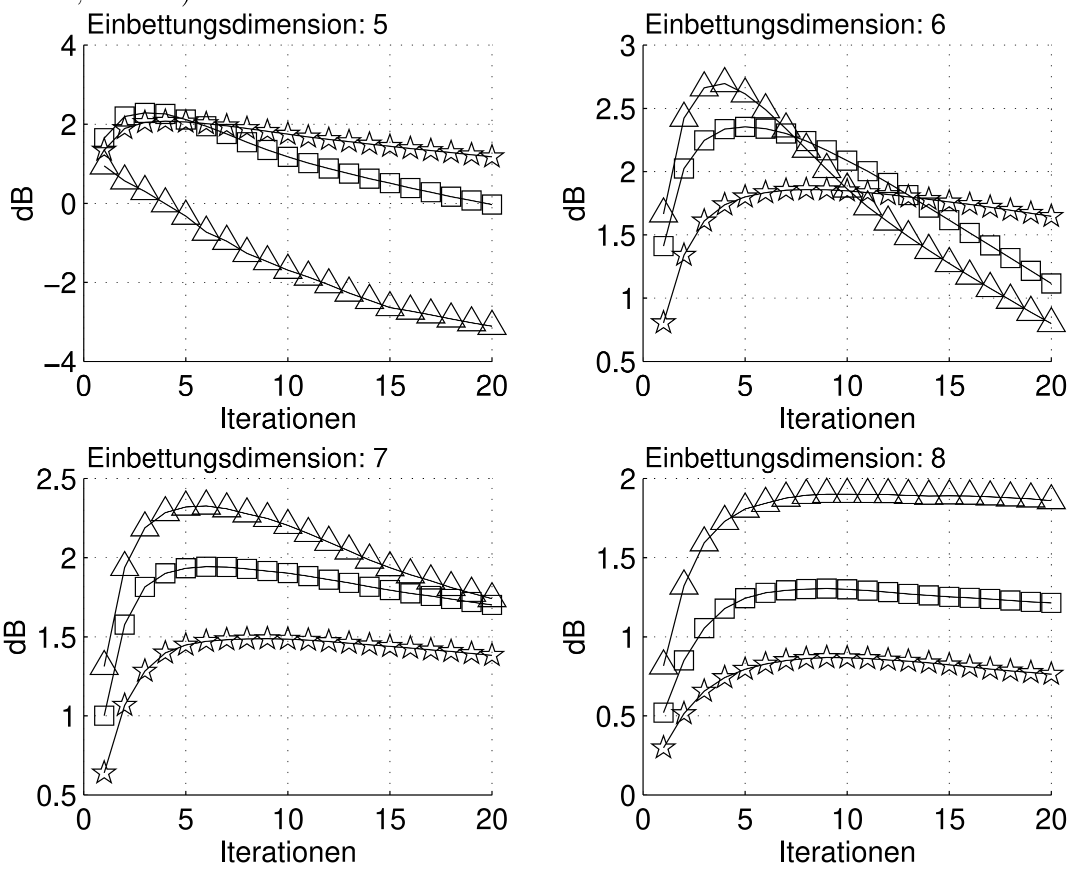
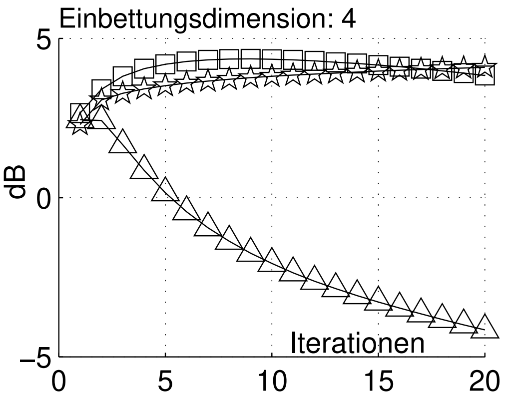
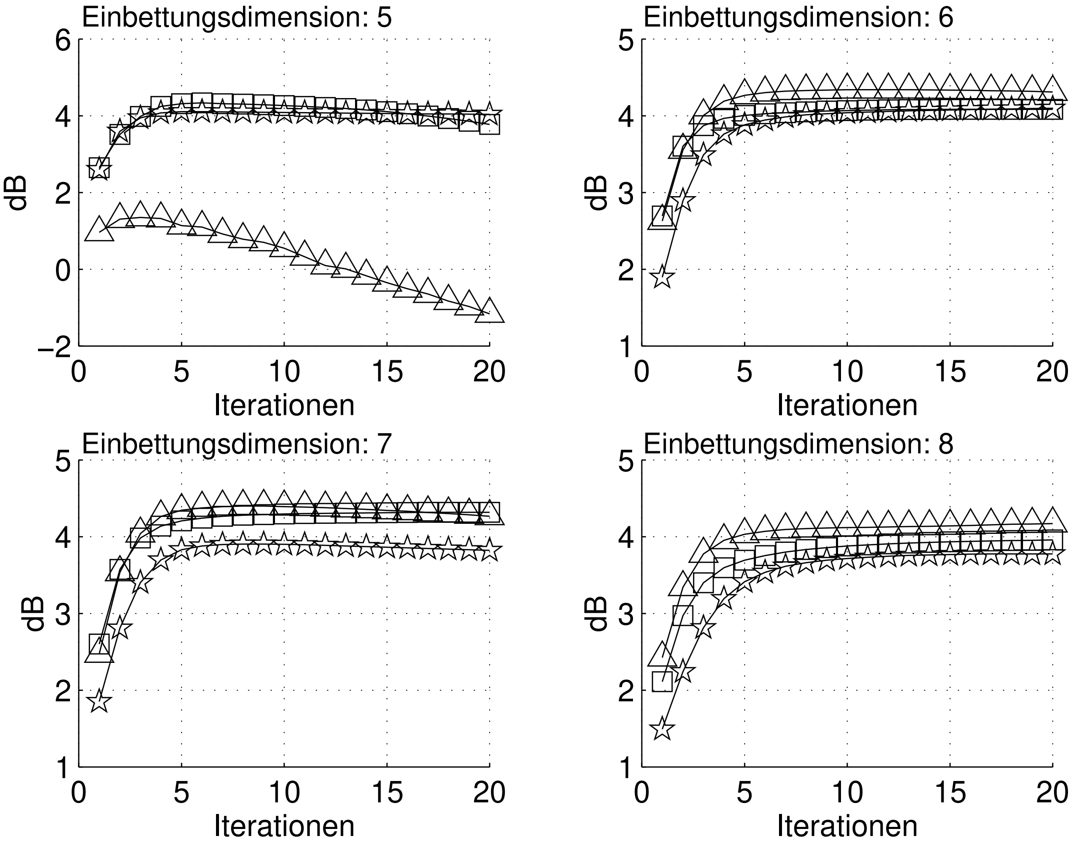


Abbildung 4.16: Dargestellt sind die dB-Verbesserungen für eine Duffing–Zeitreihe mit 10 dB Rauschen. Getestet wurde für Einbettungsdimensionen von 4 bis 8. Scharparameter der Kurven sind die Nebenbedingungen ($1 = \star$, $2 = \square$, $3 = \triangle$).





Zusammenfassung und Ausblick

In dieser Arbeit wurden einige Aspekte der Rauschreduktion in nichtlinearen Systemen behandelt. Die Darstellung war weder umfassend noch wurden die vorgestellten Verfahren bis in die letzten Details beleuchtet oder versucht, die absolut bestmögliche Performance zu erreichen. Die hier vorgestellten Algorithmen, Konzepte und Hilfsmittel, sowohl theoretische wie numerische, wurden mit dem Ziel behandelt, eine Darstellung der Materie zu geben, von der aus der Sprung zu weiteren Methoden, die bereits existieren, nicht besonders weit ist. Dies war vor allem der Hintergrund des theoretischen Teils. Auch die numerischen Simulationen geschahen nicht mit dem Ziel, besonders schnelle oder technisch relevante Algorithmen vorzustellen, sondern um zu zeigen, auf welchem Wege aus den theoretischen Konzepten brauchbare Algorithmen entstehen. Wir betrachten die Kapitel nochmals im einzelnen:

Im ersten Kapitel wurden die notwendigen Konzepte der Stochastik und der Statistik zusammengefaßt. Zentrale Punkte waren zum ersten Maße und ihre Eigenschaften, die im Kapitel über nichtlineare Dynamik zur Anwendung kamen, sowie die Bayes-Formel und die Begriffe der Schätztheorie, welche im Kapitel über Filterung und Rekonstruktion zur Anwendung kamen und hier die zentrale Rolle spielen. Es werden außerdem einige Formeln für das Rechnen mit Gauß–Verteilungen angegeben.

Im zweiten Kapitel wurde die nichtlineare Dynamik behandelt mit einem einleitenden Abschnitt über lineare Systeme. Die linearen Systeme bilden die lokalen Modelle für die nichtlinearen Systeme. Dann wurden die nichtlinearen Systeme besprochen mit besonderer Berücksichtigung der hyperbolischen Systeme. Diese Systeme besitzen einen besonderen Strukturreichtum, wovon vor allem das Konzept des shadowing erläutert wurde, was einen direkten Bezug zur Rauschreduktion besitzt. Dieser Aspekt wurde vor allem im Kapitel 4 wieder aufgegriffen. Desweiteren wurden einige bekannte Tatsachen zur Ergodentheorie nichtlinearer Systeme dargestellt, um das allgemeine Bild, was man sich von nichtlinearen Systemen macht, zu vervollständigen und abzurunden.

Während die ersten beiden Kapitel mehr Reproduktion bekannter Tatsachen waren, stellen die nun folgenden Kapitel das Hauptgewicht der Arbeit dar. Im Kapitel 3 über Filterung wurden zuerst die wesentlichen Formeln hergeleitet, die in dieser Arbeit unter der Bezeichnung *Bayes-Filterformeln* laufen. Der Grundgedanke war, das Problem der Rauschreduktion auf die Behandlung einer bedingten Wahrscheinlichkeitsverteilung zurückzuführen. Ausgangspunkt war, daß der Zustand x_n des Systems nicht direkt beobachtbar war, sondern nur eine Funktion desselben, die Observable, und daß diese Beobachtung y_n durch ein Rauschen gestört war. x_n wurde durch ein Markov-Modell beschrieben. Die betrachtete Verteilung war dann $p(x_t = x|y^n)$, also die Wahrscheinlichkeit, daß x_t gerade den Wert x hat unter der Bedingung, daß man die Werte $y^n = (y_0, \ldots, y_n)$ gemessen (beobachtet) hat. Die Bayes-Filterformeln dienten dazu, diese Verteilung zurückzuführen auf die Verteilungen $p(y_n|x_n)$, d.h. die Wahrscheinlichkeit, daß man y_n mißt unter der Bedingung, daß das System den Zustand x_n angenommen hat, sowie der Verteilung $p(x_n|x_{n-1})$, d.h. der Wahrscheinlichkeit, daß das System von einem Zustand x_{n-1} in den Zustand x_n übergeht. Es wurde dann mit Hilfe dieses Formalismus der Kalman-Filter hergeleitet. Dann wurden Methoden besprochen, um die Bayes-Filterformeln näherungsweise zu berechnen. Hier wurde hauptsächlich auf Verfahren eingegangen, die Approximation durch Gauß–Verteilungen durchführen. Eine Implementation dieses Verfahrens diente dazu, Zeitreihen des Hénon-Systems von Rauschen zu befreien. Hierbei deutete sich an, daß der Algorithmus an bestimmten Stellen des Hénon-Attraktors Schwierigkeiten hatte, daß Rauschen zu reduzieren. Es traten dort isolierte große Fehler auf, die als Glitches bezeichnet wurden. Diese Glitches sollten sich später auch bei anderen Algorithmen zeigen.

In Kapitel 4 wurde auf die Rekonstruktion von bereits gemessenen Zeitreihen eingegangen, d.h. es wurde das Konzept aufgegeben, nur die Messungen bis zu einer bestimmten Zeit n einfließen zu lassen. Vielmehr wurde versucht, aus einer gegebenen Zeitreihe den gesamten Orbit zu rekonstruieren. Zunächst wurden einige der in der Literatur vorkommenden Minimierungsansätze wahrscheinlichkeitstheoretisch gedeutet. Dann wurde für den Fall, daß kein Systemrauschen vorliegt, ein Algorithmus zur Maximierung der Likelyhoodfunktion gegeben, der sich insbesondere durch Einfachheit auszeichnete und in den hier betrachteten Fällen zufriedenstellende Ergebnisse lieferte. Der Algorithmus lief darüberhinaus sehr schnell und war in seiner Performance nicht übermäßig empfindlich gegenüber der Wahl der Parameter. Auch hier traten für das Hénon–System wieder Glitches auf, die eine theoretische Deutung erfuhren. Die Glitches traten stets in der Nähe einer tangentialen Berührung auf. Die Erklärung lautete, daß dort Punkte mit einer sehr kurzen homoklinen Verbindung existieren. Homokline Punkte können jedoch von Algorithmen, die einen mittleren Fehler zu minimieren suchen, nicht geeignet unterschieden werden. Die Vermutung, daß die rekonstruierten Punkte und die tatsächlichen Punkte des Orbits wirklich homoklin sind konnte zumindest graphisch verifiziert werden. Die Begründung für das Auftreten der Glitches wurde weiter durch die Beobachtung gestützt, daß sie bei der Cat-Map nicht entstanden. Weiter wurde kurz erläutert, daß der Ansatz des shadowing verschieden von dem wahrscheinlichkeitstheoretischen ist. Bei Rekonstruktion durch shadowing geht es in jedem Falle darum, einen tatsächlichen Orbit des Systems zu konstruieren (und nicht zu rekonstruieren), auch wenn die Zeitreihe, von der man ausgeht, nicht auf dem Wege einer gestörten Beobachtung eines reinen unverrauschten Orbits zustande gekommen ist. Die letzten beiden Abschnitte dienten der Vorstellung zweier Algorithmen. Der erste, genannt NNS, basierte auf einer Nachbarsuche in einem hochdimensionalen Raum. Im Gegensatz zu den vorher vorgestellten Algorithmen benötigte dieser Algorithmus die dynamischen Gleichungen nicht. Vielmehr verglich der Algorithmus durch die Nachbarsuche die gegebene verrauschte Zeitreihe mit Punkten aus einem sehr großen Referenzdatensatz desselben dynamischen Systems. Es konnte gezeigt werden, daß auch dann, wenn der Referenzdatensatz durch ein kleines Rauschen verunreinigt war, brauchbare Ergebnisse erzielt wurden. Zum Abschluß wurde der Vollständigkeit halber ein aus der Literatur bekannter Algorithmus vorgestellt, der weder dynamische Gleichungen noch einen langen Referenzdatensatz verlangt. Die grundsätzliche Idee ist, die Dynamik mit Hilfe der verrauschten Zeitreihe zu rekonstruieren. In dieser Arbeit wurde nur der Fall von lokal linearen Modellen besprochen. Es zeigten sich vor allem im Fall des Duffing–Oszillators gute Ergebnisse. Diese Tatsache erfuhr eine kurze Deutung.

Zum Schluß dieser Arbeit möchte ich einen kurzen Ausblick geben. Das Kapitel über Filterung wurde schon mit einem kurzen Ausblick abgeschlossen, um den Gedankengang dort nicht zu unterbrechen. Diese Dinge sollen hier nicht wiederholt werden. Eine Anwendung der dort benutzten Methoden jedoch hat engen Zusammenhang mit der Zeitreihenanalyse und soll hier erläutert werden. Der Formalismus der Filtertheorie gestattet es nämlich, neben dem Zustand des Systems auch unbekannte Parameter des Systems zu schätzen. Angenommen man modelliert die Dynamik durch eine Schar von Funktionen $F(\cdot, a), a \in \mathbb{R}^p$, wobei a der unbekannte Parametersatz ist. Die dynamischen Gleichungen lauten dann

$$\begin{aligned} x_{n+1} &= F(x_n, a) + r_n, \\ y_n &= G(x_n) + s_n. \end{aligned}$$

Diesen Gleichungen kann man nun einfach „dynamische“ Gleichungen für die Parameter hinzufügen:

$$a_{n+1} = a_n + \rho_n$$

mit einem Rauschen ρ_n, welches die Unsicherheit in der Parameterschätzung repräsentiert. Zusammen liest sich das dann

$$\begin{aligned} \tilde{x}_{n+1} &= \tilde{F}(\tilde{x}_n) + \tilde{r}_n, \\ y_n &= \tilde{G}(\tilde{x}_n) + s_n, \end{aligned}$$

mit

$$\begin{aligned} \tilde{x} &:= (x, a), \\ \tilde{F}(\tilde{x}) &:= (F(x, a), a), \\ \tilde{G}(\tilde{x}) &:= G(x), \\ \tilde{r}_n &:= (r_n, \rho_n). \end{aligned}$$

Implementationen dieses Verfahrens zur Modellierung dynamischer Systeme sind in [32] vorgenommen worden. Einen allgemeinen Einzug in die Methoden der Zeitreihenanalyse hat diese Technik aber noch nicht so recht gehalten.

Bei der Rekonstruktion von Zeitreihen im Sinne des letzten Kapitels ist nicht so unmittelbar klar, wie es weitergehen könnte. Die Algorithmen, die mit analytischen Methoden eine Likelyhoodfunktion maximieren, sind viel diskutiert und ausprobiert worden und liegen in zahlreichen Varianten vor. Diese Algorithmen werden in nächster Zeit wahrscheinlich vor allem von der Zunahme der Rechenleistung als von zündenden Ideen profitieren. Große Defizite bestehen meiner Meinung nach jedoch vor allem bei der geschickten Ausnutzung der Symboldynamik. Angenommen, man hat die Symboldynamik für ein dynamisches System F gegeben. D.h. man hat einen Raum S von Symbolsequenzen mit endlich vielen Symbolen und eine Abbildung Φ, welche eine Symbolsequenz s eindeutig auf einen Punkt $x \in M$ abbildet. Diese Symboldynamik sei durch eine Partition $\mathcal{P} := (P_1, \ldots, P_\sigma)$ zustande gekommen, d.h. jeder Punkt $x \in M$ induziert einen Orbit $(\ldots, F^{-1}x, x, Fx, \ldots)$ und damit eine Symbolsequenz $s = \Phi^{-1}(x) := (\ldots, i_{-1}, i_0, i_1, \ldots)$ durch die Vorschrift $F^k \in P_{i_k}$. Außerdem besitzt S noch eine *Grammatik*, die definiert, von welcher Menge der Partition das dynamische System in welche andere springen kann. Hat man nun einen Pseudo-Orbit gegeben, so induziert dieser Orbit ebenfalls eine Symbolsequenz s. Handelt es sich dabei um eine erlaubte Symbolsequenz, so hat man mit $\Phi(s)$ sofort den zugehörigen Schatten gegeben. Für hyperbolische Systeme kann man immer eine Partition finden, die die notwendigen Eigenschaften hat: Für δ genügend klein erzeugt ein δ-Pseudo-Orbit immer eine erlaubte Symbolsequenz. Auf den ersten Blick scheint durch diese Konstruktion nichts gewonnen. Statt den Schatten direkt zu berechnen, muß man jetzt die gesamte Symboldynamik aufstellen. Der Vorteil könnte aber sein, daß man damit einen großen Teil der Lösung schon vorfabriziert hat, auf die man im konkreten Fall immer wieder zurückgreifen kann und dann in sehr kurzer Zeit viele Zeitreihen behandeln kann. Dieses ist besonders dann wichtig, wenn man das dynamische System nicht erforschen, sondern für bestimmte Zwecke gebrauchen will, z.B. für Kommunikation.

Anhang A
Systeme mit kontinuierlicher Zeit

Als generelle Referenz sei hier [2] gegeben. Die in dieser Arbeit beschriebenen Ideen über Filterung auf den Fall stochastischer Prozesse mit einem reellen Intervall als Parameterbereich zu übertragen ist möglich, aber technisch bedeutend schwieriger. Man hat nämlich mit Differentialgleichungen statt iterierten Abbildungen zu tun. Der Haken dabei ist, daß eine Differentialgleichung der Form

$$\dot{x}(t) = G(x(t)) + r(t) \tag{4.14}$$

sicherlich keine Lösung besitzt, wenn $r(t)$ eine sehr irreguläre Funktion ist, also besonders dann, wenn es sich um weißes Rauschen handelt. Zur Lösung dieses Problems muß ein neuer Integralbegriff definiert werden, ein sogenanntes stochastisches Integral. Entsprechend nennt man die oben besprochene Gleichung eine stochastische Differentialgleichung. Sie spielen bei der Betrachtung von stochastischen Prozessen eine große Rolle.

Der Unterschied zwischen einem stochastischen Integral und dem Riemann- oder Lebesqueintegral ist nicht nur formal, sondern liefert auch andere Ergebnisse, z.B. gilt der Hauptsatz der Integral- und Differentialrechnung nicht mehr in der alten Form. Ist z.B. $u(t, x)$ eine differenzierbare Funktion, so gilt bekanntlich

$$\frac{du}{dt}(t, x(t)) = \frac{\partial u}{\partial t}(t, x(t)) + \frac{\partial u}{\partial x}(t, x(t))\dot{x}(t),$$

und wenn wir (4.14) als gewöhnliche DGL interprtieren, so folgt

$$\frac{du}{dt}(t, x(t)) = \frac{\partial u}{\partial t}(t, x(t)) + \frac{\partial u}{\partial x}(t, x(t))(G(x(t)) + r(t)).$$

Im Kalkül der stochastischen DGL würde diese Gleichung lauten

$$\frac{du}{dt}(t, x(t)) = \frac{\partial u}{\partial t}(t, x(t)) + \frac{1}{2}\frac{\partial^2 u}{\partial x^2}(t, x(t)) + \frac{\partial u}{\partial x}(t, x(t))(G(x(t)) + r(t)).$$

Es tritt also ein Zusatzterm auf, der zweite Ableitungen von $u(t, x)$ enthält.

Auch für diese Fälle existiert eine Filtertheorie, die in der Idee genauso ist wie im zeitdiskreten Falle. Der Kalman-Bucy-Filter im linearen Fall wird durch eine Matrix bestimmt, die sich im Prinzip aus einer matrixwertigen Differentialgleichung vom RICATTI-Typ bestimmen läßt. Für den Fall, daß zusätzlich die Spektren des Rauschens bekannt sind, kann man das Problem durch Übergang in den Spektralbereich wesentlich vereinfachen. Die Übertragungsfunktion ist dann die Lösung einer Integralgleichung, der WIENER-HOPF-Gleichung. Durch Lösen dieser Gleichung kommt man z.B. auf den bekannten WIENER-Filter.

Danksagung

Ich möchte allen danken, die die Entstehung dieser Arbeit möglich gemacht haben, sei es durch fachliche Diskussionen und Anregungen, materielle oder moralische Unterstützung oder Hilfe bei technischen Problemen. Mein Betreuer Herr PD Dr. Ulrich Parlitz hat durch Vergabe des interessanten Themas wie durch unermüdliche Diskussionen mich ebenso in meinen eigenen Ideen unterstützt wie lenkend eingegriffen. Ihm möchte ich herzlich danken. Herrn Professor Dr. Lauterborn möchte ich für die Aufnahme in seine Arbeitsgruppe und die optimalen Arbeitsbedingungen am Institut danken. Die Mitglieder der Arbeitsgruppe Nichtlineare Dynamik haben durch unzählige Gespräche, Tips und Anregungen das Entstehen dieser Arbeit möglich gemacht, ihnen sei ebenfalls herzlich gedankt. Besonders möchte ich Herrn Dipl.–Phys. Christian Merkwirth und Herrn Dipl.–Phys. Jörg D. Wichard für die massive Unterstützung bei computertechnischen Fragen, für Korrekturlesen und vielerlei größere und kleinere Hilfestellungen danken. Meinen Zimmerkollegen, besonders den Herren Diplomphysikern Olgert Lindau und Reinhard Geisler, möchte ich für die (manchmal zu) angenehme Atmosphäre und für die massive Hilfe im täglichen Kampf mit den Rechenmaschinen danken. Die Institutsband bestehend aus Irena Schönke, Gisa Kirschmann–Schröder, Fabian Evert, Claus–Dieter Ohl, Martin Voss, Jörg D. Wichard und Robert Mettin zeichnet für viele vergnügliche Abende nach getaner Arbeit und unvergessene Institutsfeten verantwortlich. Meinem Vater Johannes Bröcker möchte ich für sein sorgfältiges Korrekturlesen danken, ebenso meiner Freundin Stephanie Baum. Ihr vor allem habe ich es zu verdanken, daß ich die anstrengenden Phasen heil überstanden habe. Dafür möchte ich ihr herzlich danken.

Literaturverzeichnis

[1] Masanao Aoki. *Optimisation of Stochastic Systems.* Mathematics in Science and Engineering, 1972.

[2] Walter Arnold. *Stochastische Differentialgleichungen.* Oldenbourg, 1978.

[3] Rufus Bowen. *Equilibrium States and the Ergodic Theory of Anosov-Diffeomorphisms*, volume 470 of *Lecture Notes in Mathematics.* Springer-Verlag, 1975.

[4] Leo Breiman. *Probability.* Addison-Wesley-Publishing, 1973.

[5] Theodor Bröcker. *Analysis*, volume I, II, III. BI-Verlag, 1992.

[6] Theodor Bröcker and Klaus Jänich. *Einführung in die Differentialtopologie.* Springer-Verlag, 1973.

[7] C. K. Chui and G. Chen. *Kalman Filtering*, volume 17 of *Springer Series in Information Sciences.* Springer-Verlag, 1987.

[8] Harald Cramér. *Mathematical Methods of Statistics.* Princeton University Press, 1946.

[9] J. P. Eckmann and D. Ruelle. Ergodic theory of chaos and strange attractors. *Review of Modern Physics*, 57(3), 1985.

[10] K. J. Falconer. *The Geometry of Fractal Sets.* Cambridge University Press, 1985.

[11] Peter Grassberger, Rainer Hegger, Holger Kantz, Carsten Schaffrath, and Thomas Schreiber. On noise reduction methods for chaotic data. *Chaos*, 3(2):127–141, 1993.

[12] John Guckenheimer and Philip Holmes. *nonlinear Osxillations, Dynamical Systems and Bifurcation of Vector Fields.* Number 42 in Applied Mathematical Sciences. Springer Verlag, 1983.

[13] Paul Halmos. *Ergodic Theory.* Mathematical Society of Japan, 1956.

[14] E. Atlee Jackson. *Perspectives of Nonlinear Dynamics*, volume 2. Cambridge University Press, 1991.

[15] Simon Julier and Jeffrey Uhlmann. A general method for approximating nonlinear transformations of probability distributions. Technical report, Department of engineering Science, University of Oxford, 1996.

[16] Eric J. Kostelich and Thomas Schreiber. Noise reduction in chaotic time series data: A survey of common methods. *Physical Rewiev E*, 48:1752–1763, 1993.

[17] Eric J. Kostelich and James A. Yorke. Noise reduction in dynamical systems. *Physical Review A*, 38:1649, 1988.

[18] Ulrich Krengel. *Wahrscheinlichkeitstheorie und Statistik*. Vieweg, 1991.

[19] Gerrit Langer. *Der Colpitts–Oszillator*. Tech. Report, 1998.

[20] Gerrit Langer. *Experimentelle Bifurkationsanalyse*. Diplomarbeit, 1999.

[21] Robert Liptser. Stochastic processes. Lecture Notes, Tel Aviv University, 1997.

[22] Ricardo Mañé. *Ergodic Theory and Differentiable Dynamics*. Springer-Verlag, 1983.

[23] James McNames. A nearest trajectory strategy for time series prediction. In *International Workshop on Advanced Black-Box Techniques For Nonlinear Modeling*. Katholieke Universiteit Leuven, Belgium, 1998.

[24] Alexander M. Mood, Franklin A. Graybill, and Duane C. Boes. *Introduction to the Theory of Statistics*. McGraw-Hill Series in Probability and Statistics. McGraw-Hill, 1974.

[25] V. I. Oseledec. A multiplicative ergodic theorem. *Trans. Moscow Math. Soc.*, 19, 1968.

[26] N. H. Packard, J. P. Crutchfield, J. D. Farmer, and R. S. Shaw. Geometry from a time series. *Phys. Rev. Lett.*, 45:712, 1980.

[27] Ulrich Parlitz. Identification of true and spurious lyapunov-exponents from time-series. *International Journal of Bifurcation and Chaos*, 2:155–165, 1992.

[28] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery. *Numerical Recipes in C*. Cambridge University Press, 1996.

[29] W.C. Rheinboldt and J.M. Ortega. *Iterative Solution of nonlinear Equations in Several Variables*. Academic Press, 1970.

[30] Josef Stoer and Roland Burlisch. *Numerische Mathematik*, volume I, II. Springer-Verlag, 1990.

[31] Floris Takens. Detecting strange attractors in turbulence. *Lecture Notes in Mathematics*, 898, 1981.

[32] David M. Walker. *Reconstruction and Noise Reduction of Nonlinear Dynamics using Nonlinear Filters*. PhD thesis, University of Western Australia, 1998.