Approximations and Applications of Nonlinear Filters

Approximations and Applications of Nonlinear Filters

Dissertation
zur Erlangung des Doktorgrades
der Mathematisch-Naturwissenschaftlichen Fakultäten
der Georg-August-Universität zu Göttingen
vorgelegt von

Jochen Bröcker
aus Kiel

Göttingen 2003

D7
Referent: Prof. Dr. Ulrich Parlitz
Korreferent: Prof. Dr. Theo Geisel
Tag der mündlichen Prüfung: 30. Januar 2003

> "It is always the unexpected that happens," I said in a propitiatory tone. My obtuseness provoked him into a contemptuous "Pshaw!" I suppose he meant that the unexpected couldn't touch him; nothing less than the unconceivable itself could get over his perfect state of preparation.
>
> *Lord Jim*

Summary

In this thesis a certain estimation problem for nonlinear stochastic dynamical systems in discrete time, known as *filtering* in the literature, is considered. The objective is to reconstruct the current state of the system by means of observations. The observations are noise corrupted measurements of a function of the state.

In the first chapter we will provide concepts from probability theory necessary to define and investigate the nonlinear filtering problem. The definition of the nonlinear filter is presented in the second section of this chapter. The second chapter explains in detail why filtering is an extremely difficult problem in a nonlinear context. It turns out that a finite dimensional representation of the filter is possible in very special circumstances, only. This chapter summarizes known results and is intended mainly as a motivation for the necessity of investigating approximation schemes for the nonlinear filter, considered in the following two chapters.

Numerical approximation methods are then investigated from a general point of view in the third chapter. A general framework to obtain error bounds for approximation schemes is presented. Necessary for this investigation are metrics for probability distributions we will make extensive use of. Furthermore, an essential property of the filter required to give a bounded approximation error turns out to be a negative Lyapunov exponent of the filter dynamics.

The fourth chapter provides some classes of approximation methods.

Common to all these methods are projection techniques on parametrized families of probability distributions. This approach was carried out for continuous time systems already by several authors. However, the error analysis carried out in this thesis is, to the best of our knowledge, new.

The last two chapters present (aside from two small sections devoted to a Monte Carlo approach), two interesting and important applications of nonlinear filtering. The first is estimation of an unknown parameter in the dynamics. This problem is very important in all branches of science, and we will present two numerical examples. The second application is reconstruction of a sent message in telecommunications. To this problem a chapter is devoted, including results on the bit error probability obtained using methods from nonlinear filtering theory for a simple transmitter model.

To summarize, the aim of this thesis is threefold. First, to show that filtering of nonlinear dynamical system is a nontrivial and interesting problem from a mathematical point of view, second to show methods to overcome the difficulties arising in applications, and third to show that filtering is not a purely artificial mathematical problem but has a great significance in science and engineering.

Contents

Chapter 1

Introduction

1.1 The problem of estimation

A quotation from Mood, Graybill and Boes' *Introduction to the Theory of Statistics* [51] may serve as a basic statement on the necessity of statistics in science.

> Progress in science is often ascribed to experimentation. The research worker performs an experiment and obtains some data. On the basis of data, certain conclusions are drawn. The conclusions usually go beyond the materials and operations of the particular experiment. In other words, the scientist may generalize from a particular experiment to a class of similar experiments. This sort of extensions from the particular to the general is called inductive inference. It is one way in which new knowledge is found.
>
> Inductive inference is well known to be a hazardous process. (...) One function of statistics is the provision of techniques for making inductive inference and for measuring the degree of uncertainty of such inferences. Uncertainty is measured in terms of probability, and that is the reason we have devoted so much time to the theory of probability.

This may be well enough reasoning that coping with uncertainty is something a scientist should have a basic understanding of. Furthermore, in physics, statistics and probability theory has become a central tool as important as, say, theory of differential equations, not only to verify (or falsify) theories by inductive inference, but also since fundamental laws of quantum mechanics and statistical mechanics (as the name suggests) are formulated using probability theoretical concepts.

Finally, probability theory and statistics have experienced fruitful application in many fields of engineering. Since the fundamental works of Shannon and Wiener, statistics is an indispensable concept in fields like control theory, communication and computer science where we are concerned with such things as systems with uncertain state, noise corrupted messages or algorithms with unknown input.

The inference or *estimation* problems we are going to investigate in this thesis belong to a special class. In general, any estimation problem can roughly be formulated as follows. We want to know a quantity X (future stock prices, membrane potential, fetal heartrate, internal state of a combustion engine,...). What we have available, however, is just a quantity Y (present stock prices, patch clamp information, noisy sound signal, temperature,...) which features only incomplete information about X. Analysis now states that this problem is uniquely solvable if the functional relationship between Y and X is known and invertible. However, in the beforementioned problems it is quite unlikely that the quantity Y is a function of just X alone. We merely have to assume that not only X but a large amount of different further influences determine the actual value of Y. It is obvious that under such circumstances the problem is not solvable in an analytical sense.[1]

In many cases the influences obscuring the dependence of Y on X strongly fluctuate. Although at first sight this seems to make the problem even worse, it is often a reasonable assumption that these quantities obey certain average laws. The internal fluctuations yet amount to fluctuations of the measured quantity Y, but the average behaviour of Y should be determined by the average behaviour of the fluctuations *and the value of* X. Thus it is intuitively clear that a sequence of consecutive measurements of Y may allow to determine the unknown X.

[1] We would like to remark that what we have in mind is to be distinguished from what is known as *ill posed* problems. In an ill posed problem the relationship between X and Y is invertible but the inverse is not continuous. Then a small variation of Y leads to large deviations in X, which obviously exacerbates the problem.

This discussion already suggests the concept of *noise* as random unpredictable fluctuations. Assuming the reader to have at least an intuitive understanding of noise we will explain now more precisely the subject of this thesis. In our case the quantity X that is assumed to be unknown is the state of a dynamical system, i.e. X depends on time. As dynamical systems we may first (for sake of simplicity) consider a finite dimensional iterated map. This is, $X = \{X_1, X_2, \ldots\}$ is recursively defined by an equation of the form

$$X_{n+1} = f(X_n),$$

where X_0 is an *unspecified* quantity. In order to estimate X_n we need data. In this thesis we assume the data Y_n to be dependent on X_n in the form

$$Y_n = h(X_n) + \text{noise}.$$

The central question considered in this thesis is

> Assume a sample of values $Y_1, \ldots, Y_n$ has been recorded. Furthermore, suppose f and h are known functions. What is the value of the system state X_n ?

This special estimation problem is called *filtering*. Of course, the answer to the basic question in filtering cannot be given with infinite accuracy due to the unknown noise. What is desired are estimators (i.e. functions $\hat{X}_n = \hat{X}_n(Y_1, ..., Y_n)$) having a good *average* performance. It should be mentioned that the indices n of X_n and $Y_1 \ldots Y_n$ are not accidentally the same. Estimating X_n from $Y_1 \ldots Y_k$ where $k < n$ or $k > n$ are different (but quite related) problems. They are called *prediction* and *smoothing*, respectively.

We would like to remark that even in the deterministic case (where no noise enters the dynamics or the observations) the filtering problem is not trivial. Note that we do *not* assume that h is invertible, so in general more than one measurement is necessary to recover the underlying system state. In the theory of deterministic control this is known as the *observer problem* (see e.g. [52, 53]). The filtering problem therefore can be seen as a generalisation of the observer poblem.

1.2 Two motivating examples

We will now give two simple but intuitive examples. Consider the filtering problem for

$$\begin{aligned} X_{n+1} &= 0.5X_n + V_n, \\ Y_n &= X_n + 0.1W_n, \end{aligned}$$

where W_n and V_n are standard normal (Gaussian) random variables with probability density

$$p_V(x) = p_W(x) = \frac{1}{\sqrt{2\pi}} \exp(-0.5x^2).$$

Since the system is stable, it is already a reasonable estimate to assume that $X_n = 0$ which is true in average for n large. This estimate has an asymptotic variance of 1.33. However, we can do better. It turns out that $\hat{X}_n$ given by the equations

$$\begin{aligned} \hat{X}_{n+1} &= 0.5\hat{X}_n + \Gamma_{n+1}(Y_{n+1} - \hat{X}_n), \\ \frac{1}{\Gamma_{n+1}} &= \frac{1}{0.25\Gamma_n + 1} + 0.01, \end{aligned}$$

is a superior estimator. Its variance is Γ_n which has asymptotic value 0.57. This value becomes smaller if the observation noise variance decreases, while the naive estimate $X_n = 0$ does not take into account the observations and has a constant variance. The estimator has the structure of an error feedback system with a suitably chosen gain Γ_n. It is an example of the *Kalman* filter and is optimal in the sense that it has the *least average square error* among all possible estimators (see [17]). A filter with this property will be referred to as *optimal* in the following. The relatively simple structure of the optimal filter is unfortunately destroyed when nonlinear effects enter the stage.

In a second example we investigate a nonlinear system. This example is given in order to demonstrate that nonlinear systems amount to far more complicated filtering problems. Consider the *Hénon* system

$$X_{n+1}^{(1)} = 1 - a\left[X_n^{(1)}\right]^2 + bX_n^{(2)}, \tag{1.1}$$

$$X_{n+1}^{(2)} = X_n^{(1)}, \tag{1.2}$$

with parameters $a = 1.4$ and $b = 0.3$. This example features a chaotic signal. The results of different filter approaches will be illustrated using time series $\{X_n\}$ generated by this map. The chaotic attractor reconstructed from the clean time series $\{X_n^{(1)}\}$ is shown in Figure 1.1a. Figure 1.1b shows a reconstruction from a time series

$$Y_n = X_n^{(1)} + W_n,$$

where Gaussian noise $\{W_n\}$ with SNR 13 dB[2] has been added to the data. The chaotic dynamics generates a broadband signal and the added noise occupies the same frequency band (inband noise) and can thus not be removed using linear (spectral) methods. Thus, any linear method (like the Kalman filter) is expected to be suboptimal.

That this filtering problem is actually much harder that the beforementioned linear problem can as well be understood by looking at the attractor Figure 1.1a. This plot can be seen as a sample from the invariant density of the Hénon system. In a filtering problem we know beforehand that the system states lay on the attractor, and the filter should reproduce this fact. For a linear system, the "attractor" is a single point, and this is a set much more easily to cope with than the fractal attractor.

Furthermore, it can be shown that the optimal filter providing the least average quadratic error cannot be represented in a finite dimensional form. Thus, a "nonlinear" optimal Kalman filter does not exist. In Sections 3.1 and 3.2 rigorous results will be presented showing that this unfortunately is the case for the majority of nonlinear systems. This negative result was a central motivation for the investigation of *approximative optimal filters* to be discussed in this thesis.

That the situation is not hopeless is illustrated in Figure 1.1c where one of the approximation methods to be presented in the following has been applied.[3]

[2] To quantify the amount of noise in a signal it is convenient to compare the energy content of the noise to that of the signal. The SNR is this ratio measured on a $10 \cdot \log_{10}$ scale, i.e. if X_n is the signal and W_n is the noise,

$$\text{SNR} = 10 \cdot \left(\log_{10}\left(\frac{1}{N}\sum_n (X_n - \hat{X})^2\right) - \log_{10}\left(\frac{1}{N}\sum_n W_n^2\right) \right),$$

where $\hat{X}$ is the mean of X_n.

[3] Readers already aquainted with noise reduction may think that the resulting attractor still looks very noisy compared to other noise reduction methods common in nonlinear

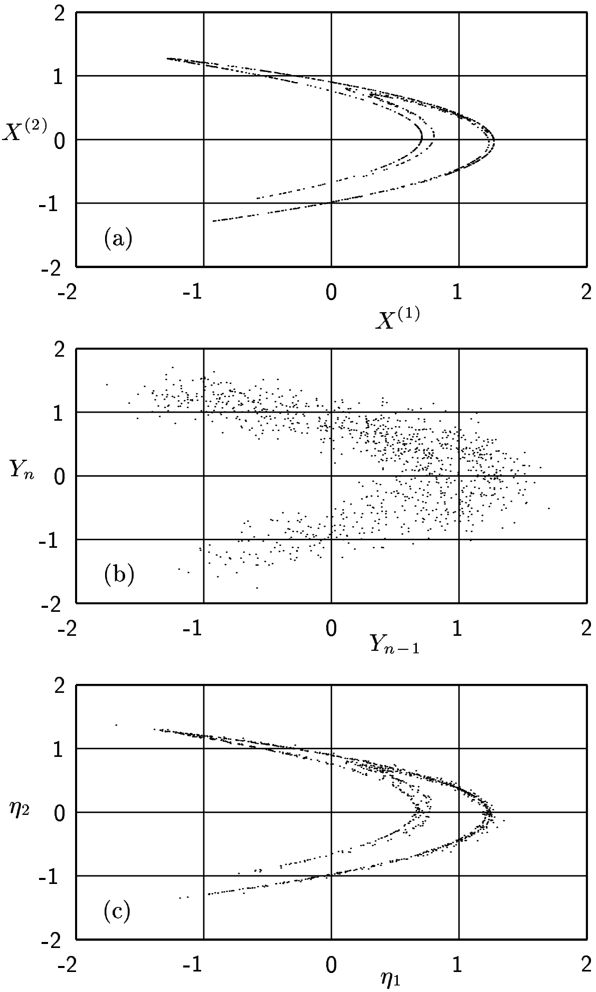


Figure 1.1: Chaotic attractor of the Hénon map (1.1) for $a = 1.4$ and $b = 0.3$. (a) Clean data $\{(X_n^{(1)}, X_n^{(2)})\}$, (b) noisy data (13 dB SNR) $\{Y_n\}$, (c) result of noise reduction using exponential families (see Section 5.2).

We have compared the improvement for a noisy time series from the Hénon system of different algorithms that will be presented in detail in Chapters 5 and 6. Figure 1.2 shows the improvement of the SNR vs the SNR of the given data set consisting of 1024 samples. As a benchmark the dotted curve shows the performance of a linear optimal Wiener filter [55] that turns out to be competitive for negative SNR's, only. Note that for vanishing relative signal power (SNR $\to -\infty$) the dotted curve approaches the straight line given by: SNR improvement = – SNR. This line gives the improvement when using as filtered signal a constant time series given by the empirical mean of the data. Clearly any algorithm should outperform this simple approach.

The unmarked curve in Figure 1.2 shows the SNR improvement of another standard method, the extended Kalman filter (EKF) [17] which gives satisfying results for low noise amplitudes (SNR $>$ 30dB). The dashed-doted curve is obtained with a noise reduction scheme where some underlying probability density functions are approximated by functions from an exponential family. This approach was used in Figure 1.1b and yields good SNR improvement for SNR $>$ 10dB, but fails for large noise amplitudes (SNR < 0). Methods of this kind are subject to Section 5.

The dashed line in Figure 1.2 denotes results obtained with Monte Carlo sampling [19, 20]. A short account on this technique is given in Section 6.1. For large noise amplitudes the performance of this method compares to that of the Wiener filter. For medium noise (0-30 dB) it turned out to be better than the other methods mentioned so far. The sudden decrease of the SNR improvement at about 40dB is due to the finite ensemble used in the Monte Carlo simulation. With a larger ensemble also for higher values of the SNR good improvement is achieved, for smaller ensembles the curve bends already at smaller SNR values. A significant drawback of the Monte Carlo method is the necessity of considerable computer ressources. This of course becomes the more a problem the larger the ensemble.

We emphasize that all the methods mentioned so far provide (less noisy) estimates of the state based on information from the past, only, and thus are filters. These estimates can be improved considerably when more and more future values of the time series are taken into account. For comparison

dynamics. The displayed result, however, was obtained using a causal method, in contrast to most other methods, which are acausal and use the full time series including future values.

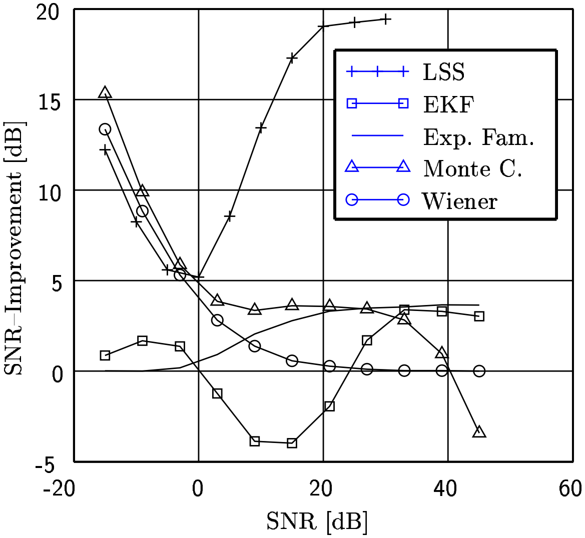


Figure 1.2: SNR–Improvement vs SNR for different noise reduction methods applied to noisy data from a chaotic Hénon map.

the SNR-improvement of an orbit estimation algorithm (called LSS [10]) is shown as a solid curve marked with +–symbols in Figure 1.2. Using the full information from past and future this method outperforms all state estimation schemes.

1.3 Outline of the thesis

With the reader having now an idea of what the thesis is about, let us give a brief overview over the contents. Chapter 2 explains the theoretical background of stochastics and filtering. Everything here is standard textbook material. Chapter 3 discusses the problem of how to *represent* the nonlinear filter as a dynamical system. For applications, it is essential that this dynamical system is finite dimensional, like the Kalman filter. It turns out, however, that this is a very unusual situation. Especially for chaotic systems, a finite dimensional characterisation of the optimal filter is impossible. The material of this chapter is cited from a series of partly recent papers. The main aim of this chapter is merely to motivate the necessity of approximations in nonlinear filtering.

The next two chapters form the heart of the thesis. In Section 4.3 a general bound on the error between approximative and the optimal filter is established. However, for this to be a useful bound, certain restrictions on the filter have to be imposed. A basic property the optimal filter has to satisfy is *insensitivity to its initial condition.* This insensitivity is characterized by means of a (negative) Lyapunov exponent.[4] We will be able to prove rigorous results only for special cases. However, our approach yields a framework for more general results.

Chapter 4 provides a catalogue of approximation schemes. Different circumstances need different tools, and we explain approaches that apply to many interesting situations. Concerning approximation methods for nonlinear filtering in continuous time systems significant work can be found in the literature already, lacking however a complete error analysis. The situation immediately carries over to continuous time signal processes with discrete time observations. This case is obviously of high practical significance. Again an error analysis was, to the best of our knowledge, not available so far. This thesis fills the gap.

Further approaches for approximating the optimal filter already present in the literature are presented in Chapter 6, mainly for the sake of completeness. The last two chapters provide interesting applications of the nonlinear filter. The first is estimation of an unknown parameter in the dynamics. The relation of this problem to filtering is trivial. Our contribution to the subject is mainly to present numerical results employing approximation schemes presented in this thesis. The second application is communication. In (tele)communication the objective is to estimate the original message from the received information. For technical or security reasons, the eventually transmitted data may be a quite complicated function of the message. Furthermore, noise is omnipresent in telecommunication, due to athmospheric disturbances, imperfect semiconductor elements etc. Filtering is therefore an essential tool here. As the preceeding discussion conveys, approximations of the optimal filter are thus of practical relevance in telecommunication.

[4] We are not going to define a Lyapunov *spectrum* for the nonlinear filter. If this was possible, however, the largest Lyapunov exponent is required to be negative.

To come back to the two challenges of statistical estimation theory in science we mentioned in the beginning, namely the inductive inference of natural laws and the control of uncertain systems, the thesis is intended to be contribution to both of them. Although the theoretical considerations will encompass quite a large part of the thesis we will finally end up with some applicable algorithms. We will present numerical studies for parameter estimation in neuron dynamics as well as bit error performance in telecommunication. This may hopefully give a strong indication that filtering and parameter estimation is useful in engineerings practice as well as the understanding of scientific experiments, and thus, of nature.

Chapter 2

Nonlinear Filtering

2.1 A few remarks on probability theory

Let us shortly discuss the main concepts of probability theory and stochastic processes. This is mainly to fix notations. The general reference for this chapter is the book of Breiman [7]. A *measure space* is a tuple $(\Omega, \mathcal{A})$, where Ω is a set of points ω and $\mathcal{A}$ is a system of subsets of Ω called σ-algebra with the following properties

$$\begin{gathered}\Omega \in \mathcal{A}, \\ A \in \mathcal{A} \Rightarrow A^c \in \mathcal{A}, \\ A_i \in \mathcal{A}, i \in \mathbb{N} \Rightarrow \cup_{i \in \mathbb{N}} A_i \in \mathcal{A}.\end{gathered}$$

The elements of $\mathcal{A}$ are the *measurable sets*. The intersection of two σ-algebras is again a σ-algebra. Thus for any arbitrary system $\mathcal{C}$ of subsets of Ω we can consider $\sigma(\mathcal{C})$, the intersection of all σ-algebras of which $\mathcal{C}$ is a subset. On any topological space E the so-called *Borel algebra* can be introduced. It is defined as the smallest σ-algebra containing all open sets (or alternatively, all closed sets) and will be denoted by $\mathcal{B}_E$.

A mapping f between two measure spaces $(\Omega, \mathcal{A})$, $(\Omega', \mathcal{A}')$ is called *measurable*, if for any measurable $A' \in \Omega'$, it holds that $f^{-1}(A')$ is measurable. Note that measurability of f depends on $\mathcal{A}$ and $\mathcal{A}'$.

A *random variable* is a measurable mapping from $(\Omega, \mathcal{A})$ to $(\mathbb{R}, \mathcal{B}_{\mathbb{R}})$. For any arbitrary $f : \Omega \to \mathbb{R}$, the smallest σ-algebra on Ω such that f is a random variable is denoted by $\sigma(f)$. A mapping $g : \Omega \to \mathbb{R}$ is said to be f-measurable if it is $\sigma(f)$-measurable. Random variables are often denoted by capital letters. Suppose X is a random variable and Y is an X-measurable random variable. Then it can be shown that there is a Borel-measurable mapping $f : \mathbb{R} \to \mathbb{R}$ so that $Y = f(X)$.

A *signed measure* μ is a mapping from $\mathcal{A}$ to $\mathbb{R}$ that is σ-additive, this is for any sequence $A_i \in \mathcal{A}$, $i \in \mathbb{N}$ for which $A_i \cap A_j = \emptyset$ it holds that

$$\mu(\cup_i A_i) = \sum_i \mu(A_i).$$

A signed measure is *bounded* if $\sup_{A \in \mathcal{A}} |\mu(A)| < \infty$. For non-negative measures this means $\mu(\Omega) < \infty$. The space of bounded signed measures on $(\Omega, \mathcal{A})$ will be denoted by $\mathcal{M}_\Omega$ and the subset of non-negative measures by $\mathcal{M}_\Omega^+$, where the index may be omitted if clear from the context.

Suppose $\mathcal{C}$ is an arbitrary family of subsets of Ω and we are given a non-negative function μ on $\mathcal{C}$ that is additive. The question arises whether μ can be extended to a measure on $\sigma(\mathcal{C})$. This is indeed possible (theorem of Hahn-Carathéodory). E.g. we can assign to any finite interval $[a, b] \subset \mathbb{R}$ the value $b - a$. This assignment obviously is additive. But the system of intervals is not a σ-algebra. It is easy to extend this definition to all *finite* collections of intervals, but we need the nontrivial result of Hahn-Carathéodory to extend this assignment to a measure on the σ-algebra induced by all finite intervalls (the Borel algebra). The result is called the *Lebesgue* measure. It is not possible to extend the Lebesgue measure to *all* possible subsets of $\mathbb{R}$ in a consistent manner. For a proof of this nontrivial fact see [13].

The measure μ is called a probability measure if it is nonnegative and

$$\mu(\Omega) = 1.$$

Probability measures are denoted by $P, Q, ...$ usually and the subset of $\mathcal{M}_\Omega^+$ of probability measures is denoted by $\mathcal{P}_\Omega$. The triple $(\Omega, \mathcal{A}, P)$ is called a probability space. Any random variable X on a probability space induces a probability measure P_X on $\mathbb{R}$ called the *distribution* of X by the definition

$$P_X(A) := P(\{\omega; X(\omega) \in A\}).$$

Sets like $\{\omega; X(\omega) \in A\}$ or $\{\omega; X(\omega) \leq c\}$ are often denoted by $\{X(\omega) \in A\}$ or $\{X(\omega) \leq c\}$ respectively. The function

$$F_X(x) := P(X < x) := P_X([-\infty, x))$$

is called the distribution function of X. A distribution function has the properties

$$F_X(x+h) - F_X(x) \geq 0 \qquad \text{for } h \geq 0, \tag{2.1}$$

$$F_X(x-) = F_X(x) \qquad \text{“left continuity”}, \tag{2.2}$$

$$\lim_{x\to\infty} F_X(x) = 1. \tag{2.3}$$

A *stochastic process* is a family $\{X_t\}_{t\in I}$ of random variables on a common probability space indexed by a parameter $t \in I$, where I is either an interval of $\mathbb{R}$ or an interval of $\mathbb{Z}$. In the first case the process is said to be a *continuous* time process, while in the latter case we will speak of a *discrete* time process. A stochastic process gives rise to a family of multidimensional distribution functions. Similar conditions like (2.1), (2.2) and (2.3) hold. Furthermore, a certain consistency condition holds, due to the fact that

$$P(X_1 \in A_1, \ldots, X_n \in A_n, X_{n+1} \in \Omega) = P(X_1 \in A_1, \ldots, X_n \in A_n).$$

It can be shown that given a family of distribution functions fulfilling this consistency condition there is always a probability space and a stochastic process with these given distribution functions (Kolmogorov's theorem).

For random variables on a probability space it is possible to define an integral. The construction is explained in [13, 7] and will not be repeated here. It starts with simple functions and is then extended to all possible limits of Cauchy sequences. The integral of a random variable is denoted by

$$\int X \mathrm{d}P.$$

We will also call it the expectation denoted by $E(X)$. For any measure μ and integrable function f also the notation $\mu(f)$ is used. By χ_A we mean the function that is 1 on A and zero elsewhere. We will write

$$\int_A f \mathrm{d}\mu := \int f \cdot \chi_A \, \mathrm{d}\mu.$$

If f is a non-negative function integrable with respect to μ we can introduce the measure $f * \mu$ by the convention

$$f * \mu(A) := \int_A f \mathrm{d}\mu / \mu(f).$$

Any distribution function gives rise to an integral on $\mathbb{R}$. It holds that

$$\int x \,\mathrm{d}F_X = \int X \mathrm{d}P.$$

The space of random variables for which $\int |X|^p \mathrm{d}P$ is finite is denoted by L_p. The L_p-spaces are complete normed vector spaces.

An important concept in estimation theory is the *conditional expectation*. Suppose X is a random variable and $E|X| < \infty$. Suppose $\mathcal{D}$ is a sub-σ-algebra of $\mathcal{A}$. A conditional expectation of X given $\mathcal{D}$ is any $\mathcal{D}$-measurable random variable Z satisfying

$$\int_D Z \mathrm{d}P = \int_D X \mathrm{d}P \qquad \forall D \in \mathcal{D}.$$

Any two such Z differ on a set of measure zero. Thus we can speak of *the* conditional expectation and write $E(X|\mathcal{D})$ for Z. It can be shown that if $E|X|^2 < \infty$, then

$$E(X - E(X|\mathcal{D}))^2 \leq E(X - Z)^2$$

for any $\mathcal{D}$-measurable random variable Z. Furthermore, equality occurs only if $Z = E(X|\mathcal{D})$ almost sure. Thus, $E(X|\mathcal{D})$ is the best approximation of X among all $\mathcal{D}$-measurable random variables. If $\mathcal{D} = \sigma(Y)$ for a random variable Y we will write $E(X|Y)$ instead of $E(X|\sigma(Y))$. Recall that $E(X|Y)$ is an Y-measurable random variable and thus can be represented as $f(Y)$. The conditional expectation should thus be considered as a measurable function $f : \mathbb{R} \to \mathbb{R}$ that minimizes $E(X - f(Y))^2$.

For any two random variables X, Y we can consider the special conditional expectation

$$P(Y \in A|X) = E(\chi_A(Y)|X).$$

This is called the *conditional probability* of Y given X. If $X = x$, then $P(Y \in A|X)$ depends on the value x only, so we write $P(Y \in A|X = x)$. It can be shown that there is a version of $P(Y \in A|X = x)$ which is a L_1 random variable as a function of x and is a probability measure for any x fixed. This version is called *regular*. Regular versions continue to exist as long as Y takes values in a complete separable metric (so called *polish*) space.

The proof of the existence of the conditional expectation relies on the Radon-Nikodym theorem, which we want to finish this section with. A measure Q is *absolutely continuous* with respect to P (write $Q \ll P$) if $P(A) = 0$ always implies $Q(A) = 0$. In this case there is a function denoted by $\frac{\mathrm{d}Q}{\mathrm{d}P}$ (Radon–Nikodym derivative, see [7]) with the property

$$Q(A) = \int_A \frac{\mathrm{d}Q}{\mathrm{d}P}(x)\mathrm{d}P.$$

Two versions of $\frac{\mathrm{d}Q}{\mathrm{d}P}$ coincide P almost sure. If both $Q \ll P$ and $P \ll Q$, they are called *equivalent* and we write $P \lessgtr Q$.

2.2 Nonlinear filtering in discrete time

In this section we present the theoretical background of the thesis, namely the theory of nonlinear filtering in discrete time.

In the introduction we considered already two filtering problems. The basic issue here was to infere from the process Y_n to the process X_n, where Y_n, the measurement process, was a function of X_n corrupted with noise. In both examples of the introduction the underlying systems were of the form

$$X_{n+1} = f(X_n) + \text{noise}.$$

The main point here is that, if X_n is known, the probability distribution of X_{n+1} depends on the noise only and *not* on further past values $X_{n-1}, X_{n-2}, \ldots$. Such processes are called *Markov processes.* More formally, the central property of Markov processes is the following:

$$P(X_n|X_{n-1}\ldots X_1) = P(X_n|X_{n-1}).$$

The conditional probability $P(X_n \in A|X_{n-1} = x) =: \varphi_n(A,x)$ is called the transition kernel. In this thesis,we consider exhaustively polish spaces, so we can assume all conditional probabilities to be regular. Furthermore, if not otherwise stated, we assume the Markov processes to be *homogenous*, this is, φ does not depend on n.

Knowing the transition kernel is equivalent to knowing the state space representation. To define a Markov process we need not only the dynamical law but furthermore the initial value X_0. In general, X_0 is as well assumed to be random with distribution P_{X_0}. The question we consider now is this: given a transition kernel $\varphi(A,x)$ and a probability distribution μ (on the state space), is there always a Markov process with transition kernel $\varphi(A,x)$

and initial distribution $P_{X_0}(A) = \mu(A)$? The answer is "yes". The reason is that if $\{X_n\}_{n\geq 0}$ is a Markov process on a probability space $(\Omega_X, P_X, \mathcal{B}_X)$ in discrete time (i.e. $n \in \mathbb{N}_0$) assumed to have values in a polish space E equipped with a Borel σ–algebra $\mathcal{B}_E$, we can always assume the probability space to be *canonical*, i.e. $\Omega_X = E^\infty$, $\mathcal{B}_X = \mathcal{B}_E^\infty$.[1] According to Kolmogorov's theorem, P_X is well defined by specifying the finite dimensional distributions of $\{X_n\}$. Since $\{X_n\}$ is Markov, the finite dimensional distributions are determined by the distribution ν of X_0 and the transition kernel by the equation

$$\begin{aligned} & P_X^\nu(X_0 \in A_0, \dots, X_k \in A_k) \\ & = \int_{A_k \times \cdots \times A_0} \varphi(\mathrm{d}x_k, x_{k-1}) \cdots \varphi(\mathrm{d}x_1, x_0)\nu(\mathrm{d}x_0), \end{aligned}$$

where $A_0, \dots, A_k \in \mathcal{B}_E$. The dependence on ν is denoted by the superscript and in fact we consider not only one measure P_X on Ω_X but a whole family P_X^ν. If ν assigns probability one to a single point $z \in E$ we write P_X^z. Further properties of Markov processes (mainly concerning their ergodic behaviour) are summarized in the Appendix A.1.

Now we turn to the *measurement* or *observation* process. Let $\{W_n\}_{n\geq 1}$ be a process of i.i.d. random variables having values in $\mathbb{R}$. We assume that the W_n have a probability density function g with respect to Lebesgue measure λ. Let the $\{W_n\}$ be of zero mean and unit standard deviation. By using Kolmogorov's theorem again we can assume the corresponding probability space to be canonical, i.e. $\Omega_W = \mathbb{R}^\infty$, $\mathcal{B}_W = \mathcal{B}^\infty$, where $\mathcal{B}$ is the Borel algebra on $\mathbb{R}$. The probability measure is defined by the finite dimensional distributions:

$$P_W(W_1 \in A_1, \dots, W_k \in A_k) = \prod_{i=1}^{k} \int_{A_i} g(x)\mathrm{d}x.$$

Furthermore, let $\{W_n\}$ be independent of $\{X_n\}$. It is well known that the corresponding probability space covering both $\{X\}$ and $\{W\}$ can be chosen as $\Omega := \Omega_X \times \Omega_W$, $P^\nu := P_X^\nu \times P_W$, $\mathcal{B} := \mathcal{B}_X \otimes \mathcal{B}_W$. Expectation with respect to P^ν or P^x will be denoted by E_ν or E_x, respectively.

[1] By E^∞ we mean the ∞–fold cartesian product of E with itself. E^∞ is the space of all sequences in E. Basically the idea is to consider a stochastic process as a randomly chosen sequence in E.

Finally we introduce the measurement process. Let $h : E \to \mathbb{R}$ be a measurable function. Now define the process $\{Y_n\}_{n \geq 1}$ by

$$Y_n = h(X_n) + \sigma W_n.$$

The σ–algebra $\sigma(Y_1, \ldots, Y_k)$ is denoted by $\mathcal{G}_n$. Note that W_n and also Y_n are defined for $n \geq 1$, while X_n is defined for $n \geq 0$.

To fix notations and conventions let us give the following general

1 Definition (Observed dynamical system (Σ)) *Let $\{X_n\}_{n \in \mathbb{N}_0}$ be a homogenous Markov process on a polish space E. The transition kernel will be denoted by $\varphi(A, x)$. Furthermore, let $\{W_n\}_{n \in \mathbb{N}}$ be a process on $\mathbb{R}$ of independent random variables having identical distribution with zero mean and unit variance. We assume the distribution to be absolutely continuous with respect to Lebesgue measure with density g. Moreover $\{W_n\}$ is independent of $\{X_n\}$. Then we define the process*

$$Y_n = h(X_n) + \sigma W_n,$$

where h is a measurable function on E and σ a positive constant. Such a setup will be called from now on an observed dynamical system. *The process $\{X_n\}$ is called the* signal process, *the process Y_n is called the* observation process.

The basic question of nonlinear filtering now is the following

> Assume a sample of values $Y_1, ..., Y_n$ has been recorded. What is the value of the system state X_n ?

Thus the aim of filtering is to estimate the "hidden" process $\{X_n\}$ from the measurements $\{Y_n\}$ in a causal manner, i.e. the estimator $\hat{X}_n$ of X_n shall depend only on $Y_1, \ldots, Y_n$, that is, it shall be $\mathcal{G}_n$–measurable. Of course, the answer to the basic question in filtering cannot be given with infinite accuracy due to the unknown noise W_n (except if $\sigma = 0$). What is desired are estimators (i.e. functions $\hat{X}_n = \hat{X}_n(Y_1, ..., Y_n)$) having a good *average* performance. We have seen that for any such estimator

$$E[(X_n - \hat{X}_n)^2] \geq E[(X_n - E(X_n|\mathcal{G}_n))^2],$$

while if equality holds, $\hat{X}_n = E(X_n|\mathcal{G}_n)$ almost sure.

To calculate conditional expectations we consider the *filtering process* π_n^ν defined as

$$\pi_n^\nu(A) := P^\nu(X_n \in A|\mathcal{G}_n),$$

where the conditioning on $\mathcal{G}_n$ can be viewed just as a shorthand notation for $Y_1 \dots Y_n$. Define also

$$\pi_n^\nu(f) := E_\nu(f(X_n)|\mathcal{G}_n) = \int f(x)\pi_n^\nu(\mathrm{d}x)$$

for a given bounded continuous function $f : E \to \mathbb{R}$. The filtering process is to be considered as a process on $\mathcal{P}_E$, the space of probability measures on E. The problem now is to give convenient formulas for π_n^ν as an explicit function of $Y_1, \dots, Y_n$.

It follows from the Kallianpur–Striebel formula (see [39]) that

$$\pi_n^\nu(f) = c\cdot \int_E f(z)\cdot g(\frac{Y_n - h(z)}{\sigma}) \int_E \varphi(\mathrm{d}z, x)\pi_{n-1}^\nu(\mathrm{d}x), \tag{2.4}$$

where c is the normalisation constant

$$c = \int_E g(\frac{Y_n - h(z)}{\sigma}) \int_E \varphi(\mathrm{d}z, x)\pi_{n-1}^\nu(\mathrm{d}x).$$

An informal derivation of Equation (2.4) can be found in [37]. If all π_n's and also $\varphi(A, x)$ have densities with respect to a measure λ we have for the densities the formula

$$\pi_n^\nu(x) = c\cdot g(\frac{Y_n - h(x)}{\sigma}) \int_E \varphi(x, z)\pi_{n-1}^\nu(z)\mathrm{d}z.$$

Equation (2.4) should be considered as a dynamical law for the stochastic process π_n. For an abbreviated notation, define the operator

$$\begin{aligned} S &: \mathbb{R} \times \mathcal{M}_E \to \mathcal{P}_E, \\ S(y, \nu)(A) &:= c\cdot \int_A g(\frac{y - h(z)}{\sigma}) \int_E \varphi(\mathrm{d}z, x)\nu(\mathrm{d}x), \end{aligned} \tag{2.5}$$

where c is again the normalizing constant. So $S(y, \cdot)$ maps finite positive measures to probability measures. With this definitions we have the iterative formula

$$\pi_{n+1}^\nu = S(Y_{n+1}, \pi_n^\nu).$$

Furthermore, $\pi_0^\nu = \nu$. A system will be referred to as *regular*, if for any ν having a density with respect to a measure λ, then, for any $y \in \mathbb{R}$, also $S(y, \nu)$ has a density with respect to λ.

As already mentioned, the filtering process π_n^ν is a random process on $\mathcal{P}_E$ and turns out to be a Markov process. Introducing the weak topology on $\mathcal{P}_E$, the transition kernel

$$\Pi(\Lambda, \nu) := P^\nu(\pi_1^\nu \in \Lambda)$$

turns out to be Feller, i.e. for any function $F : \mathcal{P}_E \to \mathbb{R}$ bounded and continuous in the weak topology, also $\Pi F(\nu) := \int F(\mu)\Pi(\mathrm{d}\mu, \nu)$ is bounded and continuous in the weak topology. To compute average quantities in the filtering problem like average filtering errors or approximation errors, ergodic properties of the filtering process are required. The main results needed in this thesis are due to Stettner [69] and Kunita [47] to which we refer the interested reader. A few results are summarized in Appendix A.2. This section is finished with the presentation of three examples.

2 Example (CSK–scheme) As a very simple model of a message transmitting device, let us consider the following setup. Suppose $\{M_n\}$ is a sequence of independent identically distributed random variables assuming only the values 0 or 1 with probability p_0 and p_1, respectively ($\{M_n\}$ should be seen as a *binary message*). Moreover, let f_0, f_1 be two continuous mappings of a closed interval I (which might be the whole real line) to itself. Let X_0 be a real valued random variable and define the process

$$X_{n+1} = f_{M_{n+1}}(X_n).$$

It is obviously a Markov Process. The measurement process is taken as

$$Y_n = X_n + \sigma W_n,$$

where W_n has standard normal distribution. The transition kernel of X_n is

$$\varphi(A, x) = p_1 \delta_{f_1(x)}(A) + p_0 \delta_{f_0(x)}(A),$$

where the *delta measure* $\delta_z(A)$ is 1 if $z \in A$ and 0 else. Often I is chosen as the unit interval and f_0, f_1 are piecewise expanding Markov maps (see [56]). In this case this setup is called *Chaotic Shift Keying* (CSK) scheme. CSK–schemes play an important role in telecommunication engineering [65]. In most of the work on this topic, however, it is assumed that M_n remains constant for more than one n. The number of n's for which M_n remains constant can be seen as a reciprocal of the bandwidth. For small bandwidth, the time series emerging for different M's may be considered as independent. This is usually assumed in the analysis of CSK–schemes. In our setup however we can not assume the Y_n's to be independent. How to recover the *message* from the Y_n's is subject to Chapter 7.

If the distribution of X_n has a density h with respect to Lebesgue measure, then so has the distribution of X_{n+1}, and the Markov transition kernel translates into an operator on L_1, called the *Frobenius–Perron–Operator* (FPO). The FPO of a CSK–scheme is given by

$$\mathcal{L}h(x) = p_1 \sum_{y \in f_1^{-1}(x)} \frac{h(y)}{|f_1'(y)|} + p_0 \sum_{y \in f_0^{-1}(x)} \frac{h(y)}{|f_0'(y)|}.$$

If the distribution ν of X_0 has an L^1 density $\pi_0(x)$ with respect to Lebesgue measure, then also the filtering process π_n^ν has a representation in terms of densities (denoted by $\pi_n^\nu(x)$) given by

$$\pi_n^\nu(x) = c \cdot g(\frac{Y_n - x}{\sigma}) \mathcal{L}\pi_{n-1}^\nu(x),$$

where again c is normalisation and g the density of W_n.

Piecewise expanding Markov maps are thoroughly investigated in [56]. It is shown that there exists an invariant measure ν on the unit interval which has a density h with respect to Lebesgue measure that is of bounded variation. Furthermore, if f is aperiodic, this measure is exact (in particular, ergodic and the only one having a density with respect to Lebesgue measure). The density h is everywhere positive and for any continuous function f

$$\mathcal{L}^n f(z) \to \int f \mathrm{d}x \cdot h(z)$$

uniformly in z. This analysis depends entirely on the FPO, and it turns out that much of it carries over to our setup. Especially there is an invariant measure ν on the unit interval which has a density h with respect to Lebesgue measure that is of bounded variation. The corresponding measure P^ν is therefore stationary and the finite dimensional distributions have all densities. Furthermore, it can be shown that under a modified aperiodicity assumption, any function $g \in L_1(\nu)$ on the interval which is invariant under φ is ν–almost sure equal to a constant. It follows then from Lemma 48 of the Appendix that P^ν is even ergodic.

The relevance of CSK–schemes as models for a real time electronic transmitting device may of course be doubted. They are hovewer subject to vivid research on a more abstract level. They are used to generate signals having desired statistical properties (see e.g. [43])

3 Example (Uniform ergodic process) Let W'_n be a process of iid random variables on $\mathbb{R}^d$ having a continuous and strictly positive pdf $d(x)$ with respect to Lebesgue measure. Let $f : \mathbb{R}^d \to \mathbb{R}^d$ be a continuous and bounded function. Then the process

$$X_{n+1} = f(X_n) + W'_n$$

is a Markov process for which the Theorem 45, namely the property (A.7) holds. The transition kernel is given by

$$\varphi(A, x) = \int_A d(z - f(x))\mathrm{d}z$$

and the FPO by

$$\mathcal{L}h(x) = \int_{\mathbb{R}^d} d(x - f(z))h(z)\mathrm{d}z.$$

This setup can also be extended to a message transmission scheme by letting $\{M_n\}$ be the usual message process and taking *two* functions $f_0, f_1 : \mathbb{R}^d \to \mathbb{R}^d$, both bounded and continuous. $\{X_n\}$ is now defined by

$$X_{n+1} = f_{M_{n+1}}(X_n) + W'_n.$$

Again $\{X_n\}$ is a Markov process satisfying the conditions of theorem (45). The transition kernel is given by

$$\begin{aligned}\varphi(A, x) &= p_0\varphi_0(A, x) + p_1\varphi_1(A, x) \\ &= \int_A p_0{\cdot}d(z - f_0(x)) + p_1{\cdot}d(z - f_1(x))\mathrm{d}z\end{aligned}$$

and the FPO by

$$\mathcal{L}h(x) = \int_{\mathbb{R}^d} \left(p_0{\cdot}d(x - f_0(z)) + p_1{\cdot}d(x - f_1(z))\right) h(z)\mathrm{d}z.$$

Again, if the distribution ν of X_0 has an L^1 density $\pi_0(x)$ with respect to Lebesgue measure, then also the filtering process π_n^ν has a representation in terms of densities (denoted by $\pi_n^\nu(x)$) given by

$$\pi_n^\nu(x) = c \cdot g(\frac{Y_n - x}{\sigma})\mathcal{L}\pi_{n-1}^\nu(x),$$

where again c is normalisation.

4 Example (Linear Gaussian process) The first system class for which the filtering process was calculated explicitely was of course the linear Gaussian case, i.e.

$$X_{n+1} = F_n X_n + a_n + W_n',$$

where W_n' has a Gaussian distribution with covariance matrices $\{R_n\}$, $\{F_n\}$ is a sequence of $d \times d$–matrices and $\{a_n\}$ a sequence of d–dimensional vectors. Furthermore assume X_0 has a Gaussian distribution with covariance matrix Γ_0. Let the measurement process be given by the equation

$$Y_n = G_n X_n + b_n + W_n,$$

where W_n has a Gaussian distribution with covariance matries S_n, $\{G_n\}$ is a sequence of $d \times l$–matrices and $\{b_n\}$ a sequence of l–dimensional vectors. Then

$$\pi_n(x) = \frac{1}{\sqrt{(2\pi)^d \det \Gamma_n}} \exp\left[-0.5\,(x-\mu_n)\,\Gamma_n^{-1}\,(x-\mu_n)\right],$$

where Γ_n and μ_n are given by

$$\begin{aligned}
\Gamma_{n+1}^{-1} &= \left(F_n \Gamma_n F^t + R_n\right)^{-1} + G_n^t S_n^{-1} G_n, \\
\mu_{n+1} &= F_n \mu_n + a_n \\
&\qquad + \Gamma_{n+1} G_n^t S_n^{-1}\left(Y_{n+1} - G_n(F_n \mu_n + a_n) - b_n\right).
\end{aligned}$$

These equations are due to Kalman [40] and a direct consequence of Equation (2.4).

The Kalman filter is an example where the filtering process admits a parametrisation. This is, $\pi_n(x) = \pi(x, \theta_n)$ and θ_n is given iteratively by a finite dimensional dynamical system of the form $\theta_n = F(Y_n, \theta_{n-1})$. We will discuss in Chapter 4 that this is in some sense a very unusual situation.

From an application point of view, the following example is maybe the most interesting.

5 Example (continuous time system) Consider a continuous time process on $\mathbb{R}^d$ defined by the stochastic differential equation

$$\mathrm{d}Z_t = f(Z_t)\mathrm{d}t + \rho(Z_t)\mathrm{d}V_t,$$

where $f : \mathbb{R}^d \to \mathbb{R}^d$ and $\rho : \mathbb{R}^d \to \mathbb{R}^{d\times d}$ are mappings of sufficient regularity in order to uniquely define the stochastic process Z_t, and V_t is a Wiener process. For a comprehensive treatment on stochastic differential equations see [2]. Suppose we are given observations of the usual form

$$Y_n := h(Z_{t_n}) + \sigma W_n,$$

where t_n are equidistant time points. We can assume $t_n = n{\cdot}\delta$, where $\delta > 0$ is the *sampling interval.* Define the process

$$X_n := Z_{t_n}.$$

It can be shown that Z_t and especially X_n are Markov processes. The filtering problem for X_n is given again by (2.4). It remains to specify φ. It turns out that

$$\varphi(A, z) = P(Z_{t_{n+1}} \in A | Z_{t_n} = z) = \Phi(A, z, \delta),$$

where $\Phi(A, x, t)$ is the transition probability of Z_t. Imposing regularity conditions it can be shown that this quantity has a density $\varphi(x, z, t)$ with respect to Lebesgue measure that can be calculated from the Fokker-Planck equation

$$\frac{\partial\varphi}{\partial t}(x, z, t) = -\sum_i \frac{\partial}{\partial x_i}(f_i\varphi) + \frac{1}{2}\sum_i \frac{\partial^2}{\partial x_i \partial x_i}(r_{ij}\varphi),$$

where $r := \rho\rho^T$. In this case, $\varphi(A, z)$, the kernel of X_n, has a density given by

$$\varphi(x, z) = \varphi(x, z, \delta).$$

Chapter 3

Finite Dimensional Filters

3.1 Finite dimensional filter systems

In the last chapter the problem of nonlinear filtering was discussed and basic formulae where given. It was however already stated informally that the filtering process in general has a very high complexity rendering it unfeasible for direct applications. We will make this a little more precise in this section by discussing some well known results about (non)existence of finite dimensional filters [28, 27, 58, 62, 49].

For this suitable approximations of the filtering process turn out to be essential. This is the main subject of this chapter. The basic concept in finite dimensional filtering is the idea of parametric probability distributions which will be presented in this section. A large variety of approximation schemes however can be considered as an approximation of the true optimal filtering process by a finite dimensional filter, whence the concepts introduced in this section also form the natural basis for the approximation schemes we will be concerned with in Chapter 4.

The concept of finite dimensional filter systems emerges more or less naturally when practical nonlinear filtering problems are considered. The Kalman filter appears not as a functional equation for π_n but, since π_n is known to be Gaussian, as a system of equations for the parameters of this Gaussian, namely the mean and the covariance. Thus the parameters obey a finite dimensional dynamics with the observations $Y_1, Y_2, \ldots$ acting as inputs. This representation of the filter is obviously very convenient for practical application. Encouraged by this example one may ask whether such representations are possible also for nonlinear systems.

We will first precisely define the kind of representations we are striving for.

6 Definition (Finite dimensional filter systems) *Let Θ be a subset of a vector space. Let*

$$Q(\cdot,\cdot\cdot) : \mathcal{B} \times \Theta \to \mathbb{R}_{\geq 0}$$

be a mapping such that for any $\theta \in \Theta$, $Q(\cdot,\theta)$ is a probability measure on E. Recall that $\mathcal{B}$ is the set of all measurable sets on E. The pair (Q,Θ) is called a parametrized family of probability distributions. *Parametric sets of probability distributions are well known in statistics, especially in parametric estimation theory, see [1, 51]. A mapping*

$$F : \mathbb{N} \times \mathbb{R} \times \Theta \to \Theta$$

is called a finite dimensional filter system *for the observed dynamical system (Σ) if $\pi_n = Q(\cdot,\theta_n)$, where θ_n is given iteratively by*

$$\theta_{n+1} = F_n(Y_{n+1},\theta_n),$$

and Θ (or more precisely its affine hull) is finite dimensional. The filter is called autonomous *if F does not depend on n.*

In many interesting cases, Q has the form

$$\mathrm{d}Q(\cdot,\theta) = q(\cdot,\theta)\cdot\mathrm{d}\lambda \tag{3.1}$$

where λ is a carrier measure and $q \in C^0(\Theta; C^0(E,\mathbb{R}_+))$. We will say that such a Q is defined by a *family of densities.* In [49] Q depends as well on n, and

$$\mathrm{d}Q_n(\cdot,\theta) = q_n(\cdot,\theta)\mathrm{d}\lambda_n$$

where $q \in C^1(\Theta; C^0(E,\mathbb{R}_+))$. Here the filter is autonomous if q does not depend on n explicitely. The reader may convince himself that the Kalman filter is a finite dimensional filter system. In fact, Q is given by the Gaussian densities and the Lebesgue measure as carrier measure. The filter is autonomous if (Σ) is. The parameter θ is given by μ and Γ.

3.2 On (non)existence of finite dimensional filter systems

The question considered in this section is on necessary and sufficient conditions for existence of a finite dimensional filter system for a given observed nonlinear system (Σ). We know that linearity of (Σ) is a sufficient criterion. Consequently, all systems that can be transformed to a linear system by a transformation of the state and the output admit a finite dimensional filter system. A simple analysis yields that the emerging finite dimensional filter systems are of exponential form, i.e.

$$\mathrm{d}Q = \exp(\theta c(x) - \psi(\theta)) \cdot \mathrm{d}\lambda,$$

where $c(x) \in C^0(E, \mathbb{R}^d)$, $\Theta \subset \mathbb{R}^d$ and ψ is a function to yield $\int \mathrm{d}Q = 1$. Thus,

$$\frac{\mathrm{d}\pi_n}{\mathrm{d}\lambda}(x) = \exp(\theta_n c(x) - \psi(\theta_n)).$$

Basically, all results concerning existence of finite dimensional filter systems state that a filter is finite dimensional if and only if it is of exponential form. If the signal process is a deterministic system, then Levine and Pignie [49] proved that (Σ) is in a certain sense equivalent to a linear system if and only if it admits a finite dimensional filter system.

Unfortunately, exhaustive criteria involving the state space representation of (Σ) have not been obtained so far. The exponentiality criterion does not lend itself to a useful description of all state space models admitting a finite dimensional filter. Construction of a finite dimensional filter system may be commenced with an exponential family. Equations for the filter dynamics are then easily obtained, in contrast to the state space dynamics the filter is connected to. Certain approaches to obtain state space models have, however, been conceived and will be presented later in this section.

Before stating the mentioned results in detail, let us first make a few necessary conventions.

7 Definition (Exponential families) *A family (Q, Θ) of probability measures on E is of exponential form with respect to a σ–finite measure λ if*

$$\frac{\mathrm{d}Q(\cdot, \theta)}{\mathrm{d}\lambda}(x) = q(x, \theta) = \exp(\theta c(x) - \psi(\theta)),$$

where $c : E \to \mathbb{R}^d$ is a measurable function and $\Theta \subset \mathbb{R}^d$. The function ψ is defined by the relation

$$1 = \int \mathrm{d}Q(\cdot, \theta),$$

which yields

$$\psi(\theta) = \log \int \exp(\theta c(x))\mathrm{d}\lambda.$$

We always assume

$$\Theta := \{\theta \in \mathbb{R}^d; \psi(\theta) < \infty\}.$$

We will refere to d as the order *of the exponential family.*

An application of Hölders inequality yields that Θ is convex and ψ as well as $\exp \circ \psi$ are convex functions. By the exponential form of the relative density, all $Q(\cdot, \theta)$ are mutually absolutely continuous. Further properties will be discussed in Section 5.2.

The most general setup was investigated by Ferrante and Runggaldier in [27]. A general observed nonlinear system (Σ) with $E \subset R^n$ is considered. It is generally assumed that π_n has a density with respect to a dominating measure λ for all n. Since

$$\pi_{n+1}(x) = c{\cdot}g(Y_n; x){\cdot}\varphi\pi_n(x),$$

we can assume that

$$\pi_n^+ := \varphi\pi_n(x)$$

has also densities with respect to λ. Suppose now that π_n admits a k-dimensional representation with a family of densities $q(x, \theta)$ (see Equation (3.1)) and filter system $\theta_{n+1} = F_n(Y_{n+1}, \theta_n)$. Then the following theorem holds

8 Theorem *Suppose the following conditions are in force:*

1. *Suppose there is a point $\theta_0 \in \Theta$ so that $\frac{\partial F_n}{\partial \theta}(y, \theta_0)$ exists and is invertible for all y and n*

2. *$q_n(x; \theta)$ is differentiable with respect to θ for all x and all θ of the form $\theta = F_n(y, \theta_0)$*

3. *The observation density $g(y; x)$ is C^1 in y for all x*

Then $g(y;x)$ is of exponential class of order $k' \leq k$.

The second theorem of [27], which we present without giving the least restrictive regularity requirements, concerns the distribution π_n^+.

9 Theorem *Suppose the following conditions are in force:*

1. *For any y and n, $F_n(y,\cdot)$ are diffeomorphisms*
2. *If*
$$F_n(y_1,\theta_1) = \ldots = F_n(y_k,\theta_k) = c,$$
then the matrix
$$\left[\frac{\partial F_n}{\partial y}(y_1,\theta_1),\ldots,\frac{\partial F_n}{\partial y}(y_k,\theta_k)\right]$$
is nonsingular
3. *$\pi(x;\theta)$ and $\pi^+(x;\theta) := \varphi\pi(x;\theta)$ are of class C^1 in θ for all x*
4. *The observation density $g(y;x)$ is C^1 in y for all x*

Then π_n^+ is of exponential class of order k.

These theorems especially imply that π_n is of exponential form. They generalize former results of Sawitzki [62] treating the case of one dimensional filter systems, only.

An explicit formalism to construct systems admitting a finite dimensional filter system was introduced in [58]. The problem considered in this work is the following: Given $g(y;x)$ of exponential class, construct a Markov semigroup φ so that the corresponding filter is finite dimensional (which then is exponential as well). Without going into the mathematical details the idea will be explained now. The authors restrict themselves to observation densities of the form

$$g(x;y) = a(x)b(y)\exp(xy).$$

For the filter the ansatz

$$\pi_n(x) = a(x)^n\cdot\exp(x\theta_n)\cdot c_n(\theta)$$

is made, where $c_n(\theta)$ is for normalisation and

$$\theta_{n+1} = F_n(\theta_n) + Y_{n+1}$$

with an F supposed to be given. Application of the filtering equations yields

$$\pi_{n+1} = \frac{a(x)b(Y_{n+1})\exp(xY_{n+1})\int \varphi(x;z)a(z)^n \exp(z\theta_n)c_n(\theta_n)\mathrm{d}z}{\int[\text{numerator}]}.$$

With the convention

$$\varphi(x;z) = \frac{a(x)^n}{a(z)^n}K_n(x,z)$$

we get, after a little algebra

$$\pi_{n+1} = \frac{a(x)^{n+1}\exp(xY_{n+1})\int K_n(x,z)\exp(z\theta_n)\mathrm{d}z}{\int[\text{numerator}]}.$$

Now the crucial equation is

$$\int K_n(x,z)\exp(z\theta)\mathrm{d}z = \exp(xF_n(\theta))\xi_n(\theta), \tag{3.2}$$

where the factor $\xi_n(\theta)$ ensures that φ is a probability kernel. It turns out that

$$\xi_n(\theta) = \frac{c_n(F_n(\theta))}{c_n(\theta)}.$$

The technique now exploited in [58] is to apply the inverse Laplace transform to $\exp(xF_n(\theta))$ and $\xi_n(\theta)$ and then by convolution of the results obtain $K(x,z)$.

This idea may be generalized to output densities $g(x;z)$ of more general form. Finally we remark that for linear output and Gaussian observation noise, the state space model obtained with this approach is linear as well.

The paper of Levine and Pignie [49] now to be discussed gives a quite exhaustive treatment of finite dimensional filters for deterministic state space models, i.e. without dynamic noise. The results are particularly interesting in view of chaotic dynamics. It turns out that for a finite dimensional filter system to exist the dynamics must be, in a certain sense (see Theorem 10), equivalent to a linear signal process with (in general nonlinear) observations. This, however, is inconsistent with any common understanding of chaos which requires a nonlinear signal process. In [49] a system of the form

$$\begin{aligned} X_{n+1} &= f(X_n) \\ Y_n &= h(X_n) + \sigma(X_n) W_n \end{aligned} \tag{3.3}$$

is considered, where f is a diffeomorphism on a simply connected smooth manifold E, h is a continuous and σ a strictly positive function on E. Consider the space H of functions generated by the functions

$$\begin{aligned} &\frac{1}{\sigma^2 \circ f^k}(x), \qquad k = 0, 1, \dots \\ &\frac{h \circ f^k}{\sigma^2 \circ f^k}(x), \qquad k = 0, 1, \dots \end{aligned}$$

10 Theorem (Levine and Pignie [49]) *The following statements are equivalent*

1. *The system (3.3) admits a finite dimensional filter system*
2. *H is finite dimensional*
3. *There is $r \in \mathbb{N}$, $\eta_1 \dots \eta_r \in H$ and $R \in GL(r)$ with the property*

$$\begin{aligned} \eta \circ f(x) = R\eta(x), \qquad \textit{where } \eta = (\eta_1 \dots \eta_r), \\ \frac{1}{\sigma^2}(x) = \sum_{i=1}^{r} \theta_1^{(i)} \eta_i(x), \\ \frac{h}{\sigma^2}(x) = \sum_{i=1}^{r} \theta_2^{(i)} \eta_i(x), \end{aligned}$$

where $\theta_1^{(i)}, \theta_2^{(i)}$ are real numbers.

The proof is analytical in nature and relies on a factorisation of the unnormalized π_n in two terms that depend on the observations and on the initial density respectively. Basically this is

$$\pi_n(x) \propto p_{Y_1 \dots Y_n | X_n}(y_1 \dots y_n; x) \cdot p_{X_n}(x).$$

It is then shown that if one (and hence all) assertions of the theorem holds, the first factor is proportional to

$$\exp\left(\sum_{i=1}^{r}\theta_n^{(i)}\eta_i(x)\right).$$

Thus, π_n is exponential.

Of course, signal processes assuming only a finite number of states also obey a finite dimensional filter system. Here $\pi_n(k), k = 1 \dots N$, where N is the number of states, is, for any n, a normalized vector of N non-negative components, in other words, it is an element of the $N-1$–standard symplex, which itself can be viewed as the parameter space. The equations for the parameters are given by the filtering equations.

Chapter 4

Approximations and Error Bounds

4.1 General remarks on approximations

We have seen in the last section that an optimal filter being finite dimensional is a very unusual event. At least, in a given application an infinite dimensional filter is likely to appear, especially if the signal model is known to be chaotic. Since an infinite dimensional filter is impossible to realize on a computer, approximations are essential. The approximative filter has to be of course finite dimensional and as optimal as possible. To the best of our knowledge, all algorithms so far developed employ either Monte Carlo like ideas or approximations of π_n by distributions of a finite dimensional family. Monte Carlo methods are *not* subject of this thesis. We will present however the basic ideas in Section 6.1. A detailed error analysis was carried out by several authors, and the interested reader will find the references in Section 6.1.

The purpose of the remainder of this chapter is to analyze schemes featuring approximations of π_n by members of a finite dimensional family of probability distributions. The approximative filtering process will be denoted by $\tilde{\pi}_n$. Although a large variety of methods have been conceived they share a basic and quite natural idea. Let $Q(\cdot, \theta)$ be the parametrized family and $\tilde{\pi}_n = Q(\cdot, \theta_n)$. The idea is to replace the exact prediction and update step by simplified prediction and update steps in order to keep $\tilde{\pi}_n$ a member of the parametrized family. Thus the process $\tilde{\pi}_n$ is obtained inductively as follows

1. Approximate π_0 by $\tilde{\pi}_0$.
2. Suppose $\tilde{\pi}_n$ is already given. Then approximate $S(Y_{n+1}, \tilde{\pi}_n)$ (the correct prediction and update applied to $\tilde{\pi}_n$) by $\tilde{\pi}_{n+1}$.

Obviously, at every step an error is made. A priori it is not clear whether this will amount to an infinite increase of the total error. To investigate this, the distance between two probabilities has to be quantified. This will be the subject of the next section. But suppose for a moment that a distance $d(\mu, \nu)$ between two probability measures is given. Our goal is to get a bound on the *total error* $d(\tilde{\pi}_n, \pi_n)$. However, a direct calculation is of course impossible since π_n is unavailable. Our approximation algorithm approximates $S(Y_{n+1}, \tilde{\pi}_n)$ by $\tilde{\pi}_{n+1}$, so what we may have at hand is $d(S(Y_{n+1}, \tilde{\pi}_n), \tilde{\pi}_{n+1})$, the *approximation residual*, for any n. In Section 4.3 a connection between the total error and the approximation residuals will be established. This leads to bounds on the total error, if certain stability conditions on the nonlinear filter are imposed.

4.2 Metrics for probability distributions

We consider σ-additive set functions on a polish space $(E, \mathcal{B})$ and introduce some metrics for σ-additive set functions that will be convenient later for investigation of the nonlinear filter.

4.2.1 The uniform metric and the total variation distance

Let ν be a σ-additive set function (signed measure) on $\mathcal{B}$. By

$$|\nu| := \sup_{A \in \mathcal{B}} |\nu(A)|$$

one defines a norm for σ-additive set functions. Let $\mathcal{M}_E$ be the space of all σ-additive set functions on E having a finite norm. By $\mathcal{M}_E^+$ we will denote the set of all positive measures.[1] Obviously, $\mathcal{M}$ is a vector space,

[1] The index E will be dropped if no ambiguity is to be expected.

and $|\cdot|$ is a norm that turns $\mathcal{M}$ into a complete metric space. Let X be a bounded measurable function. Let ν_1, ν_2 be finite positive measures on E. If $|X| \leq C$ is a bounded measurable function we get

$$\int X \mathrm{d}\nu_1 - \int X \mathrm{d}\nu_2 \leq 2C|\nu_1 - \nu_2|.$$

Thus $|\cdot|$ controls the accuracy up to which expectations like $\int X \mathrm{d}\nu_1$ are reproduced using ν_2 instead of ν_1. If however $\nu_1 \ll \nu_2$ and X bounded, then

$$\begin{aligned}
&|\int X \mathrm{d}\nu_1 - X \mathrm{d}\nu_2| \\
&\leq \int |X||\frac{\mathrm{d}\nu_1}{\mathrm{d}\nu_2} - 1|\mathrm{d}\nu_2 \\
&\leq \max|X| \int |\frac{\mathrm{d}\nu_1}{\mathrm{d}\nu_2} - 1|\mathrm{d}\nu_2.
\end{aligned}$$

The quantity $\mathsf{TV}(\nu_1, \nu_2) := \int |\frac{\mathrm{d}\nu_1}{\mathrm{d}\nu_2} - 1|\mathrm{d}\nu_2$ is called the *Total Variation Distance.* If ν_1, ν_2 have densities $\nu_1(x)$ and $\nu_2(x)$ with respect to Lebesgue measure it can be written in the form

$$\mathsf{TV}(\nu_1, \nu_2) := \int |\nu_1(x) - \nu_2(x)|\mathrm{d}x.$$

It turns out that TV is symmetric, convex in both arguments, vanishes iff $\nu_1 = \nu_2$ and satisfies the triangle inequality. The discussion suggests that there may be a connection between $|\cdot|$ and $\mathsf{TV}(\cdot, \cdot\cdot)$. Although this fact is easy to prove and apparently assumed to be true in the literature, a proof could not be found.

11 Lemma

$$\mathsf{TV}(\mu, \nu) = 2|\mu - \nu|$$

Proof Set $\frac{\mathrm{d}\mu}{\mathrm{d}\nu} - 1 = [\frac{\mathrm{d}\mu}{\mathrm{d}\nu} - 1]_+ - [\frac{\mathrm{d}\mu}{\mathrm{d}\nu} - 1]_-$. Since

$$\int \frac{\mathrm{d}\mu}{\mathrm{d}\nu} - 1\mathrm{d}\nu = 0,$$

we have

$$\int [\frac{\mathrm{d}\mu}{\mathrm{d}\nu} - 1]_+ \mathrm{d}\nu = \int [\frac{\mathrm{d}\mu}{\mathrm{d}\nu} - 1]_- \mathrm{d}\nu.$$

Therefore

$$\begin{aligned} \int |\frac{\mathrm{d}\mu}{\mathrm{d}\nu} - 1| \nu &= \int [\frac{\mathrm{d}\mu}{\mathrm{d}\nu} - 1]_+ \mathrm{d}\nu + \int [\frac{\mathrm{d}\mu}{\mathrm{d}\nu} - 1]_- \mathrm{d}\nu \\ &= 2 \int [\frac{\mathrm{d}\mu}{\mathrm{d}\nu} - 1]_+ \mathrm{d}\nu. \end{aligned} \tag{4.1}$$

Now there is a set $B \in \mathcal{B}$ so that $(\mu - \nu)(\cdot \cap B)$ and $(\nu - \mu)(\cdot \cap B^c)$ are positive measures (see [22]). Thus

$$\begin{aligned} \sup_A |\mu(A) - \nu(A)| &= \sup_{A \subset B, A' \subset B^c} |(\mu - \nu)(A) + (\mu - \nu)(A')| \\ &= \max\{(\mu - \nu)(B), (\nu - \mu)(B^c)|\}. \end{aligned}$$

But

$$(\mu - \nu)(B) + (\mu - \nu)(B^c) = (\mu - \nu)(E) = 1 - 1 = 0,$$

whence

$$\sup_A |\mu(A) - \nu(A)| = (\mu - \nu)(B). \tag{4.2}$$

However, since $\frac{\mathrm{d}\mu}{\mathrm{d}\nu} \geq 1$ on B and $\frac{\mathrm{d}\mu}{\mathrm{d}\nu} \leq 1$ on B^c we have

$$(\mu - \nu)(B) = \int [\frac{\mathrm{d}\mu}{\mathrm{d}\nu} - 1]_+ \mathrm{d}\nu. \tag{4.3}$$

Now (4.1), (4.2) and (4.3) yield the result □

Two properties of the TV distance will prove to be useful in connection with filtering. First it is easy to see that for any measurable set A we have

$$\int \varphi(A, x) \mathrm{d}\nu(x) - \int \varphi(A, x) \mathrm{d}\mu(x) \leq \mathsf{TV}(\nu, \mu),$$

which implies the following (well known) fact:

12 Lemma

$$\mathsf{TV}(\varphi\nu, \varphi\mu) \leq \mathsf{TV}(\nu, \mu)$$

Furthermore, let $\mu \ll \nu$ and let g be a non-negative function so that $\mu(g)$ and $\nu(g)$ are finite. Define the measures $g * \mu$ and $g * \nu$ by the conventions

$$g * \mu(A) := \frac{1}{\mu(g)} \int_A g \mathrm{d}\mu,$$
$$g * \nu(A) := \frac{1}{\nu(g)} \int_A g \mathrm{d}\nu.$$

Then the following holds:

13 Lemma

$$\mathsf{TV}(g * \mu, g * \nu) \leq \frac{\max g}{\max\{\mu(g), \nu(g)\}} \mathsf{TV}(\mu, \nu) \leq \frac{\max g}{\min g} \mathsf{TV}(\mu, \nu)$$

PROOF See [48]. □

The total variation distance measures the mean deviation of $\frac{\mathrm{d}\mu}{\mathrm{d}\nu}$ from its mean value 1. Deviations however can be measured by means of other functions than $|\cdot|$ as well. This is the concept of the f-distances introduced in the next subsection.

4.2.2 The f–distances

In this section, let f be a convex function on $\mathbb{R}_{\geq 0}$ that vanishes at $x = 1$. Let μ, ν two measures (i.e. positive members of $\mathcal{M}$), and assume $\mu \ll \nu$. We define the *f–distance* between μ and ν by

$$f(\mu, \nu) := \int f(\frac{\mathrm{d}\mu}{\mathrm{d}\nu}) \mathrm{d}\nu.$$

Let $\nu = \nu(E) \cdot \bar{\nu}$, where $\bar{\nu}$ is a probability measure ($\nu(E)$ is by assumption finite). Since

$$f(\mu, \nu) = \nu(S) \int f(\frac{\mathrm{d}\bar{\mu}}{\mathrm{d}\bar{\nu}}) \mathrm{d}\bar{\nu}.$$

it suffices to consider f–distances for probability measures. We will do so in the following.

For, if $\mu = \nu$ we have $\frac{\mathrm{d}\mu}{\mathrm{d}\nu} = 1$, we see that $f(\mu, \nu)$ vanishes in this case. Furthermore, $f(\mu, \nu)$ is non-negative. Indeed, by Jensen's inequality we have

$$0 = f(1) = f(\int \frac{\mathrm{d}\mu}{\mathrm{d}\nu}\mathrm{d}\nu) \le \int f(\frac{\mathrm{d}\mu}{\mathrm{d}\nu})\mathrm{d}\nu = f(\mu, \nu).$$

We remark that $f(\mu, \nu)$ may be infinite. Furthermore $f(\mu, \nu)$ may vanish even if $\mu \neq \nu$. To exclude the latter, we have to impose further conditions on f.

14 Lemma *Suppose there is an $a \in \mathbb{R}$ so that the function*

$$g(x) := f(x) - a(x - 1)$$

is non-negative and vanishes only if $x = 1$, then $f(\mu, \nu)$ vanishes only if $\mu = \nu$.

Proof The function $g(x)$ is convex as well. Furthermore $f(\mu, \nu) = g(\mu, \nu)$. But since g is non-negative,

$$g(\mu, \nu) = \int g(\frac{\mathrm{d}\mu}{\mathrm{d}\nu})\mathrm{d}\nu$$

can only vanish if $g(\frac{\mathrm{d}\mu}{\mathrm{d}\nu})$ is identical to zero, which implies $\frac{\mathrm{d}\mu}{\mathrm{d}\nu} = 1$ ν-a.s. But this means $\mu = \nu$. □

The conditions can be relaxed, but we will not need it.

Common choices for f are

$(\sqrt{x} - 1)^2$	Hellinger–distance HE
$\lvert x - 1\rvert$	total–variation–distance TV
$x \cdot \log(x)$	Kullback–Leibler–distance KL

The total variation distance plays a central role, since all f–distances allow for an estimate against TV. Although the following fact is quite useful and easy to prove, it seems to be unknown in the literature.

15 Theorem *For two probability measures μ, ν, it holds in general that*

$$f(1 + \frac{1}{2}\mathsf{TV}(\mu, \nu)) + f(1 - \frac{1}{2}\mathsf{TV}(\mu, \nu)) \le f(\mu, \nu).$$

Proof The proof of this fact is a generalisation of the method used in [72] to prove the special case of the KL distance. Since $f(1) = 0$, we have the general property that

$$f(x) = f(\max\{x, 1\}) + f(\min\{x, 1\}).$$

To check the validity of this assertion consider the cases $x < 1$, $x = 1$ and $x > 1$ separately. Using this fact and the convexity of f we get the general estimate

$$\begin{aligned} f(\mu, \nu) &= \int f(\frac{\mathrm{d}\mu}{\mathrm{d}\nu})\mathrm{d}\nu \\ &= \int f(\max\{\frac{\mathrm{d}\mu}{\mathrm{d}\nu}, 1\})\mathrm{d}\nu + \int f(\min\{\frac{\mathrm{d}\mu}{\mathrm{d}\nu}, 1\})\mathrm{d}\nu \\ &\geq f(\int \max\{\frac{\mathrm{d}\mu}{\mathrm{d}\nu}, 1\}\mathrm{d}\nu) + f(\int \max\{\frac{\mathrm{d}\mu}{\mathrm{d}\nu}, 1\}\mathrm{d}\nu). \end{aligned}$$

Now use the facts

$$\begin{aligned} \max\{x, 1\} &= \frac{1 + x + |1 - x|}{2} \\ \min\{x, 1\} &= \frac{1 + x - |1 - x|}{2} \end{aligned}$$

to complete the theorem. □

We fix an estimate between TV and KL as a lemma

16 Lemma (Bretagnole–Huber and Furstemberg inequality)

$$\mathsf{TV}(\mu, \nu) \leq 2\sqrt{1 - \exp\left(-\mathsf{KL}(\mu, \nu)\right)} \leq 2\sqrt{\mathsf{KL}(\mu, \nu)}$$

Proof Follows from Theorem 15 as an easy consequence. □

A further useful estimate concerns the Hellinger distance

17 Lemma *For the Hellinger distance* HE *the estimate*

$$2\left[(\sqrt{\mathsf{HE}} + 1)^2 - 1\right] \geq \mathsf{TV}$$

holds.

PROOF Theorem 15 gives the inequality

$$\mathsf{HE} \geq (\sqrt{1+0.5\mathsf{TV}}-1)^2 + (\sqrt{1-0.5\mathsf{TV}}-1)^2.$$

The function $(\sqrt{1+x}-1)^2 + (\sqrt{1-x}-1)^2$ is invertible on $[0,1]$ and the inverse is monotonically increasing (recall that $\mathsf{TV} \leq 2$ always). However, the inverse function is transcendent and the resulting estimate not so convenient. But the right hand side is larger than $(\sqrt{1+0.5\mathsf{TV}}-1)^2$, thus we have

$$\mathsf{HE} \geq (\sqrt{1+0.5\mathsf{TV}}-1)^2,$$

which eventually yields the result. □

4.2.3 The Hilbert distance

In contrast to the former metrics, the *Hilbert metric* compares two probability measures on a point by point basis. Call two finite measures μ, ν *comparable*, if there are positive constants c_1, c_2 so that $c_1\mu \leq \nu \leq c_2\mu$. The *Hilbert distance* is defined as

$$\mathsf{H}(\mu,\nu) := \inf(\log(c_2) - \log(c_1)),$$

where the infimum is taken over all such choices of c_1 and c_2.

As an obvious consequence of comparability, $\mu \lessgtr \nu$. We will see that

$$c_1 \leq \frac{\mathrm{d}\nu}{\mathrm{d}\mu} \leq c_2 \qquad \text{a.s.}$$

and furthermore

$$\mathsf{H}(\mu,\nu) := \operatorname{ess\,sup} \log \frac{\mathrm{d}\mu}{\mathrm{d}\nu} + \operatorname{ess\,sup} \log \frac{\mathrm{d}\nu}{\mathrm{d}\mu} = \operatorname{ess\,sup} \log \frac{\mathrm{d}\mu}{\mathrm{d}\nu} - \operatorname{ess\,inf} \log \frac{\mathrm{d}\mu}{\mathrm{d}\nu}.$$

To see this, we need some remarks concerning essentially bounded functions. Let f be integrable and μ a non-negative measure. Suppose an estimate of the form

$$\int_A f \mathrm{d}\mu \leq C{\cdot}\mu(A)$$

holds for all $A \in \mathcal{B}$. The infimum over all such C is called the essential supremum of f. It holds that

$$\operatorname{ess\,sup} f = \inf\{C; \mu(f > C) = 0\}.$$

Furthermore, there is a set $N \in \mathcal{B}$ which is a subset of a μ-nullset and

$$\operatorname{ess\,sup} f = \sup_{x \in S-N} f.$$

The essential supremum of a given function depends also on the measure μ. It is easy to see that if $\mu \ll \nu$, then $\operatorname{ess\,sup}_\mu f \leq \operatorname{ess\,sup}_\nu f$. Back to the Hilbert distance, if $\nu \leq c_2\mu$, then

$$\int_A \frac{\mathrm{d}\nu}{\mathrm{d}\mu}\mathrm{d}\mu = \nu(A) \leq c_2\mu(A),$$

so

$$\operatorname{ess\,sup} \frac{\mathrm{d}\nu}{\mathrm{d}\mu} \leq c_2,$$

and the infimum of all such c_2 is $\operatorname{ess\,sup} \frac{\mathrm{d}\nu}{\mathrm{d}\mu}$. Furthermore

$$\int_A \frac{\mathrm{d}\mu}{\mathrm{d}\nu}\mathrm{d}\nu = \mu(A) \leq \nu(A)/c_1,$$

so

$$\operatorname{ess\,sup} \frac{\mathrm{d}\mu}{\mathrm{d}\nu} \leq 1/c_1.$$

and the supremum of all such c_1 is $1/\operatorname{ess\,sup} \frac{\mathrm{d}\mu}{\mathrm{d}\nu}$. But

$$\operatorname{ess\,inf} \frac{\mathrm{d}\nu}{\mathrm{d}\mu} = 1/\operatorname{ess\,sup} \frac{\mathrm{d}\mu}{\mathrm{d}\nu} \geq c_1.$$

This shows the stated assertions. Along the same lines it can be shown that μ and ν are comparable if $\frac{\mathrm{d}\mu}{\mathrm{d}\nu}$ exists and $\operatorname{ess\,sup} \frac{\mathrm{d}\mu}{\mathrm{d}\nu}$ and $\operatorname{ess\,inf} \frac{\mathrm{d}\mu}{\mathrm{d}\nu}$ are both finite.

The following easily seen facts make the Hilbert distance very useful for filtering. It follows readily from the definitions that

$$\mathsf{H}(c\mu, \nu) = \mathsf{H}(\mu, \nu)$$

for any positive constant c. Thus, the Hilbert distance is *projective*. Especially, when dealing with the Hilbert distance, the unnormalized filtering process can be used. Moreover, it is easily seen that, if μ and ν are comparable, then so are $\varphi\mu$ and $\varphi\nu$ and it holds that

$$\mathsf{H}(\varphi\mu, \varphi\nu) \leq \mathsf{H}(\mu, \nu).$$

To see this, note that if $c_1 \leq \frac{\mathrm{d}\mu}{\mathrm{d}\nu} \leq c_2$, then

$$c_1 \int \varphi(\cdot, x)\mathrm{d}\nu \leq \int \varphi(\cdot, x)\frac{\mathrm{d}\mu}{\mathrm{d}\nu}(x)\mathrm{d}\nu(x) \leq c_2 \int \varphi(\cdot, x)\mathrm{d}\nu.$$

The term in the middle however is $\int \varphi(\cdot, x)\mathrm{d}\mu(x)$, thus dividing by $\int \varphi(\cdot, x)\mathrm{d}\nu$ gives the assertion. Furthermore,

$$H(\gamma * \mu, \gamma * \nu) = H(\mu, \nu).$$

This will also prove to be a useful property for analyzing filtering processes. The technique of using H in connection with filtering was (to our knowledge) first used in [3].

Finally, the following is easily proved

18 Lemma *Let μ, ν be mutually absolutely continuous measures so that*

$$\operatorname{ess\,inf} \frac{\mathrm{d}\mu}{\mathrm{d}\nu} \leq 1 \leq \operatorname{ess\,sup} \frac{\mathrm{d}\mu}{\mathrm{d}\nu}.$$

This e.g. holds if μ, ν are probability measures. Here $\operatorname{ess\,sup} \frac{\mathrm{d}\mu}{\mathrm{d}\nu}$ *may be infinite. Then*

$$\mathsf{KL}(\mu, \nu) \leq \nu(S)\mathsf{H}(\mu, \nu).$$

PROOF

$$\begin{aligned}
\mathsf{KL}(\mu, \nu) &= \int -\log(\frac{\mathrm{d}\mu}{\mathrm{d}\nu})\,\mathrm{d}\nu \\
&= \int [\log(\frac{\mathrm{d}\mu}{\mathrm{d}\nu})]_-\,\mathrm{d}\nu - \int [\log(\frac{\mathrm{d}\mu}{\mathrm{d}\nu})]_+\,\mathrm{d}\nu \\
&\leq \int [\log(\frac{\mathrm{d}\mu}{\mathrm{d}\nu})]_+\,\mathrm{d}\nu + \int [\log(\frac{\mathrm{d}\mu}{\mathrm{d}\nu})]_-\,\mathrm{d}\nu \\
&\leq \nu(S)\left(\operatorname{ess\,sup}[\log(\frac{\mathrm{d}\mu}{\mathrm{d}\nu})]_+ + \operatorname{ess\,sup}[\log(\frac{\mathrm{d}\mu}{\mathrm{d}\nu})]_-\right).
\end{aligned}$$

Now, if $\operatorname{ess\,sup} \frac{\mathrm{d}\mu}{\mathrm{d}\nu} \geq 1$, $\operatorname{ess\,sup} \log(\frac{\mathrm{d}\mu}{\mathrm{d}\nu}) \geq 0$. Hence, $\operatorname{ess\,sup} \log(\frac{\mathrm{d}\mu}{\mathrm{d}\nu}) = \operatorname{ess\,sup} [\log(\frac{\mathrm{d}\mu}{\mathrm{d}\nu})]_+$. As well, if $\operatorname{ess\,inf} \frac{\mathrm{d}\mu}{\mathrm{d}\nu} \leq 1$, $\operatorname{ess\,inf} \log(\frac{\mathrm{d}\mu}{\mathrm{d}\nu}) \leq 0$. Hence, $\operatorname{ess\,inf} \log(\frac{\mathrm{d}\mu}{\mathrm{d}\nu}) = -\operatorname{ess\,sup} [\log(\frac{\mathrm{d}\mu}{\mathrm{d}\nu})]_-$ which yields

$$\mathsf{KL}(\mu, \nu) \leq \nu(S) \left(\operatorname{ess\,sup} \log(\frac{\mathrm{d}\mu}{\mathrm{d}\nu}) - \operatorname{ess\,inf} \log(\frac{\mathrm{d}\mu}{\mathrm{d}\nu}) \right),$$

which is the desired result. □

4.3 A general error bound on the approximative filtering process

Now we are ready to embark for the first estimate on the error of our approximative filtering process. We apply a technique that was used already in [14] in very special circumstances. It essentially uses only the triangle inequality. Let

1. π_n be the true filtering process,

2. $\tilde{\pi}_n$ be a process obtained by approximation with a parametrized family Q,

3. $S_k^n(\mu) := S(Y_n, S(Y_{n-1}, S(\dots S(Y_k, \mu) \dots)$ be the $n-k+1$ fold iterate of S with arguments μ and $Y_k \dots Y_n$, where, if $k > n$ we set $S_k^n(\mu) = \mu$. We also write $S_n(\mu) := S(Y_n, \mu)$. Recall that $\pi_{n+k} = S_{n+1}^{n+k}(\pi_n)$,

4. $\mathcal{S}$ be the family of measures

$$\mathcal{S} := \{\pi_0\} \cup Q \cup \{S(y, q); q \in Q, y \in \mathbb{R}\},$$

5. D be a metric for probability measures satisfying the triangle inequality,

6. $\mathsf{D}(S(y, \mu), S(y, \nu)) \leq \mathsf{D}(\mu, \nu)$, for all $\mu, \nu \in \mathcal{S}$ and $y \in \mathbb{R}$. In other words, S is a weak contraction w.r.t. D.

Then a direct application of the triangle inequality yields

$$\begin{aligned} \mathsf{D}(\pi_n, \tilde{\pi}_n) \leq & \mathsf{D}(S_n(\pi_{n-1}), \tilde{\pi}_n) \\ & + \mathsf{D}(S_n(\pi_{n-1}), S_n(\tilde{\pi}_{n-1})). \end{aligned}$$

The second term can be bounded as follows

$$\begin{aligned} \mathsf{D}(S_n(\pi_{n-1}), S_n(\tilde{\pi}_{n-1})) \leq & \mathsf{D}(S^n_{n-1}(\tilde{\pi}_{n-2}), S_n(\tilde{\pi}_{n-1})) \\ & + \mathsf{D}(S^n_{n-1}(\pi_{n-2}), S^n_{n-1}(\tilde{\pi}_{n-2})). \end{aligned}$$

Again the second term can be bounded in the same manner. Continuing in this way we get

$$\begin{aligned} \mathsf{D}(\pi_n, \tilde{\pi}_n) \leq & \sum_{k=1}^{n} \mathsf{D}(S^n_k(\tilde{\pi}_{k-1}), S^n_{k+1}(\tilde{\pi}_k)) \\ & + \mathsf{D}(S^n_1(\pi_0), S^n_1(\tilde{\pi}_0)). \end{aligned} \tag{4.4}$$

The term in the sum can be written as

$$\begin{aligned} \mathsf{D}(S^n_k(\tilde{\pi}_{k-1}), S^n_{k+1}(\tilde{\pi}_k)) &= \mathsf{D}(S^n_{k+1}(S_k\tilde{\pi}_{k-1}), S^n_{k+1}(\tilde{\pi}_k)) \\ &= \frac{\mathsf{D}(S^n_{k+1}(S_k\tilde{\pi}_{k-1}), S^n_{k+1}(\tilde{\pi}_k))}{\mathsf{D}(S_k\tilde{\pi}_{k-1}, \tilde{\pi}_k)} \cdot \mathsf{D}(S_k\tilde{\pi}_{k-1}, \tilde{\pi}_k) \\ &\leq \sup_{\mu,\nu\in\mathcal{S}} \left(\frac{\mathsf{D}(S^n_{k+1}(\mu), S^n_{k+1}(\nu))}{\mathsf{D}(\mu,\nu)} \right) \cdot \mathsf{D}(S_k\tilde{\pi}_{k-1}, \tilde{\pi}_k) \\ &=: \tau^n_k \cdot \mathsf{D}(S_k\tilde{\pi}_{k-1}, \tilde{\pi}_k). \end{aligned} \tag{4.5}$$

The second term of Equation (4.4) can be treated in a similar manner. The term $\mathsf{D}(S_k\tilde{\pi}_{k-1}, \tilde{\pi}_k)$ will be called the *approximation residual* in the metric D. Thus, equations (4.4) and (4.5) establish a general connection between the total approximation error and the approximation residuals in the sense of Section 4.1, which reads as

$$\begin{aligned} \mathsf{D}(\pi_n, \tilde{\pi}_n) \leq & \sum_{k=1}^{n} \tau^n_k \cdot \mathsf{D}(S_k\tilde{\pi}_{k-1}, \tilde{\pi}_k) \\ & + \tau^n_0 \mathsf{D}(\pi_0, \tilde{\pi}_0). \end{aligned} \tag{4.6}$$

The crucial point is now a sufficient smallness of the random quantity τ^n_k. Whether the error is bounded or not will depend on whether, roughly speaking, there is $\tau < 1$ so that for $n - k$ large, $\frac{1}{n-k} \log(\tau^n_k) \cong \log \tau$ or not.

The basic property of τ_k^n is that it is *submultiplicative*, i.e.

$$\tau_n^k \leq \tau_n^l \cdot \tau_l^k \qquad n < l < k \in \mathbb{N}$$

Since all τ's are nonnegative we can write this as

$$\log(\tau_n^k) \leq \log(\tau_n^l) + \log(\tau_l^k), \tag{4.7}$$

a property known as *subadditivity*. The central fact about subadditive processes is the following theorem

19 Theorem (Kingman's subadditive ergodic theorem)
Let $Y_1, Y_2, \ldots$ be a stationary, $\mathbb{R}$-valued process and

$$g_k : \mathbb{R}^k \to \mathbb{R}, \qquad k \in \mathbb{N}$$

measurable functions. Set

$$g_n^{n+k} := g_k(Y_{n+1} \ldots Y_{n+k}) \qquad n \in \mathbb{N}_0, k \in \mathbb{N}.$$

Suppose the subadditivity condition

$$g_n^k \leq g_n^l + g_l^k \qquad n < l < k \in \mathbb{N}_0 .$$

holds. We ask for convergence of

$$\frac{1}{n} g_n(Y_1, \ldots, Y_n) = \frac{1}{n} g_0^n$$

to a limit random variable g_*.

1. *If $E[g_1(Y_1)]_+ < \infty$, then convergence takes place a.s.*
2. *If $\inf \frac{1}{n} E g_n(Y_1 \ldots Y_n) > -\infty$, then convergence takes place a.s. and in L_1.*

In both cases, $Eg_* = \inf \frac{1}{n} E g_0^n$.

Proof See [45] □

For convenience let us introduce the abreviations

$$\begin{aligned} D_k &:= \mathsf{D}(S_k \tilde{\pi}_{k-1}, \tilde{\pi}_k), \\ D_0 &:= \mathsf{D}(S_k \tilde{\pi}_0, \tilde{\pi}_0), \end{aligned}$$

that allow us to write the error bound (4.6) as

$$\mathsf{D}(\pi_n, \tilde{\pi}_n) \leq R_n := \sum_{k=0}^{n} \tau_k^n D_k. \tag{4.8}$$

This representation is also convenient since it allows for a separate investigation of the terms τ_k^n and D_k. The error-damping τ_k^n depends on the dynamical system only and may be investigated without specifying an actual approximation process.

Since $\log \tau_k^n$ is subadditive and nonpositive, we can always apply the first version of Kingman's theorem to $\log \tau_k^n$. The limit $\log \tau_*$ will be called the *Lyapunov exponent* of the nonlinear filter. It may be that $E \log \tau_* = -\infty$. The Lyapunov exponent depends of course on D and on $\mathcal{S}$. Lyapunov exponents are intensively studied in the theory of nonlinear deterministic dynamical systems as well. For their significance in practical nonlinear time series analysis see [41]. In a finite dimensional system a whole *spectrum* of Lyapunov exponents can be considered (see [46]). In the remainder of this section we want to discuss why $\log \tau_*$ may have a significance on the total filtering error.

First let us mention what can be expected from equation (4.8). A filtering error converging to zero is surely a bit too much to hope. At every step k, the approximation residual D_k is added to the error, and if D_k does not converge to zero for some intrinsic reasons of the approximation algorithm, then also R_k will not. We see already from the trivial case where D_k is constant that a bounded error will emerge only if τ_k^n is something like t^{n-k} with a $t < 1$. Thus let us assume that $\tau_k^n \leq Ct^{n-k}$ with $t < 1$ and nonrandom $C > 0$. Furthermore, assume that D_k is stationary, then it is plausible from Equation (4.8) that we should have asymptotically

$$\begin{aligned} R_k &= C \sum_{k=0}^{n} t^{n-k} D_k \\ &= C \sum_{k=0}^{n} t^k D_{n-k} \\ &\sim C \sum_{k=0}^{\infty} t^k D_{n-k}. \end{aligned}$$

We see that in the last equation it is necessary to define D_k also for negative values of k. This is indeed possible for stationary processes. More especially, there is always a process on $\mathbb{Z}$ which has the same *distribution* as D_k on $\mathbb{N}$. The process $\sum_{k=0}^{\infty} t^k D_{n-k}$ (if well defined) obviously is stationary, so we see from the above considerations that R_k should have asymptotically a stationary distribution. Of course, the case of a random τ_k^n requires a much more elaborated analysis. Furthermore, the assumption that D_k is stationary is usually not fulfilled. The approximation process (that D_k essentially depends on) is initialized with some quantity that is probably nonrandom. However, it may well be the case that this initialisation dies away with time and thus, D_k is asymptotically stationary. Then the above considerations again apply. That the process R_k is asymptotically stationary with a finite expectation is the best we can hope to hold under not too restrictive assumptions.

We will not carry out these problems exhaustively in this thesis. The sense of the preceeding discussion was merely to motivate the statement: To obtain a stationary filtering error it is necessary that $\log \tau_*$, the Lyapunov exponent of the filter, is negative. In the next section we will present a class of systems where this actually can be proved. We will finish this section by showing numerically that for a piecewise constant Markov map, the well known tent map, the Lyapunov exponent of the filter is negative.

20 Numerical example (Tent map) Consider as a dynamical system the iterations

$$X_{n+1} = f(X_n)$$

of the tent map

$$f : [0,1] \to [0,1], \qquad x \to 1 - |2x - 1|.$$

For the observation noise we take random variable $\{W_n\}$ which are independent, have a centered normal distribution with unit variance and are independent of $\{X_n\}$. The observations are assumed to be

$$Y_n = X_n + \sigma W_n.$$

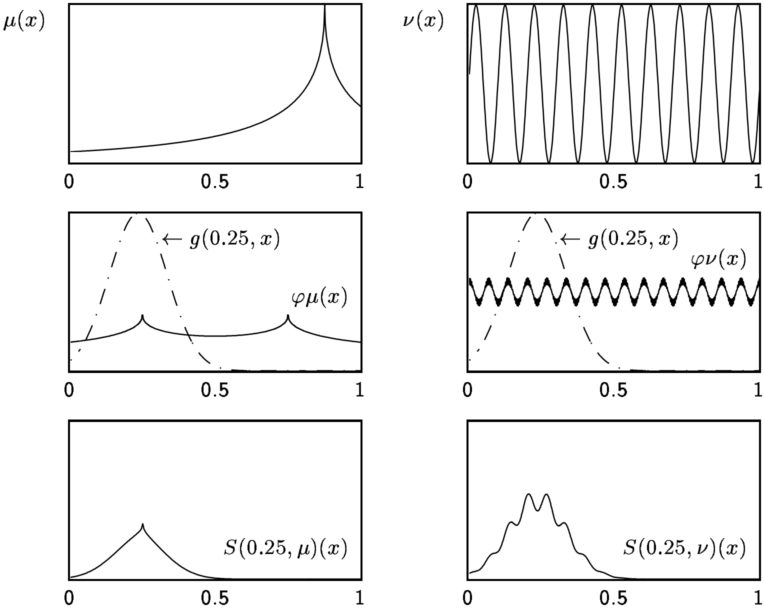


Figure 4.1: Two probability density functions (first row), after application of the Markov kernel (second row) and after application of the update step (third row). The resulting two probability densities look quite similar.

We now took simply two probability density functions $\mu(x)$ and $\nu(x)$ and applied the filtering algorithm $S(0.25, \cdot)$. In Figure 4.1 the result is shown. The probability density functions $\mu(x)$ and $\nu(x)$ are shown in the first row. Application of the Markov kernel gives the solid lines in the second row. The update density $g(y, x)$ is shown by the dotted line, for $y = 0.25$. Application of the update step yields two probability densities which already look quite similar. This hints on the stability of the filter associated with this system.

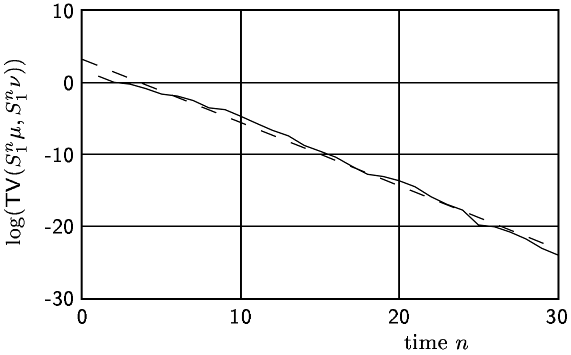


Figure 4.2: The total variation distance TV of two filtering processes associated with the tent map, but started at different initial conditions. The ordinate is logarithmic, so the total variation distance decays exponentially and the Lyapunov exponent of the filter is thus negative.

Figure 4.2 now gives a numerical estimate of the Lyapunov exponent for the TV metric. We plotted $\log \mathsf{TV}(S_1^n\mu, S_1^n\nu)$ for two (actually quite different) initial probability density functions μ and ν. To get this result, the filtering process has to be calculated exactly for the required number of steps. This was accomplished by using an extremly fine mesh. Although the numerical accuracy of the result is not sufficient for a quantitative statement the experiment shows that stability of the filter in this case is likely to be present. We remark that the tent map is ergodic with respect to the Lebesgue measure on the unit interval. The observations were taken from a run of the tent map initialized with a randomly chosen init value, and can thus be seen as ergodic. Therefore the generalisation from a sample mean to an ensemble mean is possible. The definition of the Lyapunov exponent however requires a supremum over $\mathcal{S}$ which we have carried out. The example of the tent map also shows that the Lyapunov exponent depends on the choice of $\mathcal{S}$. If we take as initial probability distributions two measures that distribute equal mass on the discrete points of two disjoint periodic orbits of the tentmap, the filter will consistently reproduce the fact that these orbits are invariant sets and the filtering processes will not become equal asymptotically. The filter is thus insensitive to misspecified initial distributions of a sufficient regularity only. We conjecture that the initial distributions need to be of bounded variation. We have however not lined out the analysis.

In the next sections we will present an important case where the Lyapunov exponent of the filter can actually be proved to be negative.

4.4 On the stability of mixing processes

A large class of Markov processes enjoy a certain mixing property which leads to strong results concerning the ergodic properties of these processes. For a thorough discussion of the ergodic properties of Markov kernels see [21], where especially Theorem 45 is proved. For our analysis we employ the Hilbert metric. The results of this section are essentially due to [3].

The properties of H immediately yield for the filtering process

$$\mathsf{H}(S(y,\mu),S(y,\nu)) = \mathsf{H}(\varphi\mu,\varphi\nu),$$

since the normalisation and the multiplication with $g(\cdots)$ cancels out (see discussion page 47). Furthermore, the Hilbert distance has outstanding properties in connection with positive operators. We have seen that if μ, ν are comparable, then so are $\varphi\mu, \varphi\nu$ and

$$\mathsf{H}(\varphi\mu,\varphi\nu) \leq \mathsf{H}(\mu,\nu).$$

Thus, S is a weak contraction for H. Furthermore,

$$\sup_{0<\mathsf{H}(\mu,\nu)<\infty} \frac{\mathsf{H}(\varphi\mu,\varphi\nu)}{\mathsf{H}(\mu,\nu)} \leq \tanh(\frac{\Delta}{4}), \tag{4.9}$$

where

$$\Delta := \sup_{\mu,\nu\in\mathcal{P}_E} \mathsf{H}(\varphi\mu,\varphi\nu) \tag{4.10}$$

is the projective diameter of φ. This fact is due to G. Birkhoff (see [3, 6, 26]). It can be shown that if φ posesses a density $\varphi(x,z)$ with respect to a measure λ, then

$$\Delta = \log\left[\sup_{z,z'} \operatorname{ess\,sup}_{x,x'} \frac{\varphi(x,z)\varphi(x',z)}{\varphi(x,z')\varphi(x',z')}\right]. \tag{4.11}$$

A Markov kernel with the property that there is a finite measure η and positive constants c_1, c_2 with the property

$$c_1\eta \leq \varphi(\cdot, x) \leq c_2\eta \qquad \forall x \in E$$

will be called *mixing Markov kernel.* For such a kernel we have $\Delta \leq 2\log(c_1/c_2)$, whence

$$\mathsf{H}(\varphi\mu, \varphi\nu) \leq t\mathsf{H}(\mu, \nu).$$

where $t = \tanh(\frac{\log(c_1/c_2)}{2}) < 1$.

Our analysis shows that such Markov kernels have a negative Lyapunov exponent with respect to the Hilbert metric. According to the properties of the Hilbert metric, this behaviour immediately carries over to the filter and yields the following statement

21 Theorem *For mixing Markov kernels we have*

$$\mathsf{H}(\pi_n, \tilde{\pi}_n) \leq \sum_{k=0}^{n} t^{n-k}\mathsf{H}(S_k\tilde{\pi}_{k-1}, \tilde{\pi}_k).$$

For this to be a useful bound we have to assume that $S_k\tilde{\pi}_{k-1}$ and $\tilde{\pi}_k$ are comparable. This is a drawback of the above result that can however be overcomed using the total variation distance.

A connection between the Hilbert metric and the total variation distance is given in the following lemma

22 Lemma *In general*

$$\mathsf{TV}(\mu, \nu) \leq \frac{2}{\log 3}\mathsf{H}(\mu, \nu),$$

where the right hand side is maybe infinite. Furthermore, if φ is a mixing Markov process, then

$$\mathsf{H}(\varphi\mu, \varphi\nu) \leq 2\log(1 + \frac{c_2}{c_1}\mathsf{TV}(\mu, \nu)).$$

Proof The first inequality is due to Atar and Zeitouni [3]. The second inequality is due to Kushner and Budhiraja [14]. □

These results together yield

23 Theorem *The process τ_k^n for the* TV *distance and mixing Markov kernels satisfies*

$$\tau_k^n \leq Ct^{n-k}$$

with

$$t = \tanh(\frac{\log(c_1/c_2)}{2}) < 1$$

and

$$C = 4\frac{c_2}{tc_1 \log 3}.$$

Proof Let μ, ν be two probability measures. Then $S_k^n(\mu)$ and $S_k^n(\nu)$ are always comparable as long as $n \leq k$. Thus for any $1 \leq k \leq n$

$$\begin{aligned}
&\mathsf{TV}(S_k^n(\mu), S_k^n(\nu)) \\
&\leq \frac{2}{\log 3}\mathsf{H}(S_k^n(\mu), S_k^n(\nu)) \\
&\leq \frac{2}{\log 3}t^{n-k}\mathsf{H}(S_k(\mu), S_k(\nu)) \\
&\leq \frac{2}{\log 3}t^{n-k}\mathsf{H}(\varphi\mu, \varphi\nu) \\
&\leq \frac{4}{\log 3}t^{n-k}\log\left(1 + \frac{c_2}{c_1}\mathsf{TV}(\mu, \nu)\right).
\end{aligned}$$

Where we first applied the inequality of Atar and Zeitouni, then the cited contraction properties of the Hilbert metric and eventually the inequality of Kushner and Budhiraja. Here

$$t = \tanh(\frac{\log(c_1/c_2)}{2}).$$

Since $\log(1 + \frac{c_2}{c_1}x) \leq \frac{c_2}{c_1}x$ we get finally

$$\mathsf{TV}(S_k^n(\mu), S_k^n(\nu)) \leq 4\frac{c_2}{c_1 \log 3}t^{n-k}\mathsf{TV}(\mu, \nu).$$

Thus

$$\tau_{k-1}^n \leq \sup_{\mu,\nu \in \mathcal{P}} \left(\frac{\mathsf{TV}(S_k^n(\mu), S_k^n(\nu))}{\mathsf{TV}(\mu, \nu)}\right) \leq 4\frac{c_2}{c_1 \log 3}t^{n-k},$$

which yields

$$\tau_k^n \leq 4\frac{c_2}{tc_1 \log 3}t^{n-k}.$$

□

Finally we will give two important examples of mixing processes.

24 Example Consider again the model of Example 3. Then the transition kernel has a finite projective diameter if

$$\log \frac{d(x-a)}{d(x-b)} \leq D \qquad \forall a,b \in \{x \in \mathbb{R}, |x| \leq C\}.$$

This condition was overseen in [11]. However, we already mentioned that Theorem 45 holds (see [21], chapter V,§5 Case (b)) without this condition. To see that this condition amounts to a finite Δ, employ equation (4.11). This yields

$$\begin{aligned}\Delta &= \log \left[\sup_{z,z'} \operatorname*{ess\,sup}_{x,x'} \frac{d(x-f(z))d(x'-f(z))}{d(x-f(z'))d(x'-f(z'))} \right] \\ &= \log \left[\sup_{|z|\leq C,|z'|\leq C} \operatorname*{ess\,sup}_{x,x'} \frac{d(x-z)d(x'-z)}{d(x-z')d(x'-z')} \right] \\ &\leq 2D.\end{aligned}$$

The second example is basically a discretely sampled continuous system.

25 Example We have seen in Chapter 2 that discretely sampled stochastic differential equations amount to interesting nonlinear filtering problems. Let δ be the sampling time. The question is whether $\Phi(A,z,\delta)$ is a mixing Markov kernel. A sufficient condition is that E, the state space of the stochastic differential equation is compact and the generator of φ is strictly *elliptic*, i.e. the diffusion coefficient is strictly positive. Then (see [3] and references therein) we have

$$c_1 \leq \varphi(x,z,\delta) \leq c_2$$

for positive c_1 and c_2 depending on δ, which especially imply the mixing property.

4.5 Does stability depend only on φ?

It was shown in the last section that Markov kernels satisfying a certain mixing condition lead to filters having a negative Lyapunov exponent. This negative Lyapunov exponent tells us that the filter is insensitive with respect to its initial condition. The approach used solely depends on the Markov kernel. Why mixing Markov kernels imply a negative Lyapunov exponent of the filter can be also understood as follows. The mixing condition is well known to imply an exponential decay of correlations in the process X_n (see [21]). This essentially means that the future evolution of the process is only weakly dependent on its values in the remote past, or determinism is weak in this process. It is therefore quite logical that also the filter does not need to take into account initial values, which means that the distribution of initial values should not determine the long time behaviour of the filter. In this analysis it does not play any role how the observations are taken. As a very extreme example consider the following (completely useless) system:

$$\begin{aligned} X_{n+1} &= V_n, \\ Y_n &= h(X_n) + \sigma W_n, \end{aligned}$$

where V_n are iid normal random variables. In this filter π_{n+1} does not depend on π_n, thus it has Lyapunov exponent $-\infty$. However, to obtain good estimates, the optimal filter usually exploits the underlying determinism in the dynamics. But since there is no determinism in this system, filtering is quite frustrating since even the optimal filter (which is easy to built in this case) yields poor results. So we may be led to the conclusion that the filter is the more stable the weaker the determinism in the dynamics and thus the larger the filtering error.

This however is not generally true. Consider again the Hénon system

$$\begin{aligned} X^{(1)}_{n+1} &= 1 - a\left[X^{(1)}_n\right]^2 + bX^{(2)}_n, \\ X^{(2)}_{n+1} &= X^{(1)}_n, \end{aligned}$$

with observations

$$Y_n = X^{(1)}_n + \sigma W_n.$$

Now if the noise amplitude σ was zero, then we would have the equalities

$$\begin{aligned} X^{(1)}_n &= Y_n \\ X^{(2)}_n &= \frac{1}{b}\left(Y_{n+1} - 1 + a\left[Y_n\right]^2\right), \end{aligned}$$

so the filtering problem is actually solved in two steps, no matter what the initial distribution of $(X^{(1)}_0, X^{(2)}_0)$ was. Thus again here the filter is insensitive with respect to its initial condition. But now this is due to good observations. We see that stability of the filter may be due to either a weak determinism in the Markov process or good observations.

It should be noted that low noise does not necesarily imply good observations. What we mean by good observations actually is not so easy to define. In the above example, the underlying system is deterministic, so if the observation noise goes to zero, filtering theory actually becomes obsolete. For general filtering problems it is necessary to investigate how the observations enter the process τ_k^n. For one–dimensional systems a nice result was obtained in [3]. They considered a mixing Markov process on the unit interval with transition density $\varphi(x,z)$ and observations of the form

$$Y_n = X_n + \sigma W_n.$$

We know alrady that $\tau_k^k \leq t < 1$. To study the influence of the observations on τ_k^n, the idea in [3] is to consider

$$\tau_k^{k+1} = \sup_{0<\mathsf{H}(\mu,\nu)<\infty} \frac{\mathsf{H}(S_k^{k+1}\mu, S_k^{k+1}nu)}{\mathsf{H}(\mu,\nu)}.$$

Since Birkhoff's estimate (4.9) is true for any positive operator we have that

$$\tau_k^{k+1} \leq \tanh(\frac{\Delta_k}{4}),$$

with

$$\Delta_k = \log\left[\sup_{z,z'} \operatorname{ess\,sup}_{x,x'} \frac{\Phi(x,z,Y_k)\Phi(x',z,Y_k)}{\Phi(x,z',Y_k)\Phi(x',z',Y_k)}\right],$$

where

$$\Phi(x,z,Y_k) := \int_0^1 \varphi(x,\xi)\frac{1}{\sigma}g(\frac{\xi - Y_n}{\sigma})\varphi(\xi,z)\mathrm{d}\xi.$$

This can also be written as

$$\tau_k^{k+1} \leq \frac{\sqrt{\Psi_k}-1}{\sqrt{\Psi_k}+1},$$

where

$$\Psi_k := \left[\sup_{z,z'} \operatorname{ess\,sup}_{x,x'} \frac{\Phi(x,z,Y_k)\Phi(x',z,Y_k)}{\Phi(x,z',Y_k)\Phi(x',z',Y_k)}\right].$$

Basically what is proved now in [3] is that if σ goes to zero, then

$$\frac{1}{\sigma}g(\frac{\xi - Y_n}{\sigma}) \to \delta(\xi - Y_n),$$

where δ is the Dirac function. Thus

$$\Phi(x, z, Y_k) \to \varphi(x, Y_n)\varphi(Y_n, z).$$

Replacing this in the expression for Ψ_k we see that this goes to one, thus τ_k^{k+1} goes to zero. In [3] moreover the rate of convergence is computed.

This ideas can be generalized to systems of a form that strongly resemble systems in *observer canonical form* (see [36]). We will present the central idea, omitting technical details. Suppose a Markov process on the d–cube $[0,1]^d$ has a transition kernel that can be represented as a product of d terms as follows:

$$\varphi(x, z) = \varphi_1(x^{(1)}; z^{(d)}) \prod_{k=2}^{d} \varphi_k(x^{(1)}; z^{(1)} \dots z^{(k-1)}, z^{(d)}), \tag{4.12}$$

where $z^{(k)}$ means the k'th component of z. Moreover, let the observations again be of the form

$$Y_n = X_n^{(d)} + \sigma W_n.$$

Now the idea is to consider τ_k^{k+d}. We get exactly the same formulae but now with

$$\begin{aligned} &\Psi_k(x, z; Y_k, Y_{k+1}, \dots Y_{k+d-1}) \\ &= \int \cdots \int \varphi(x, x_1)\frac{1}{\sigma}g(\frac{x_1^{(d)} - Y_k}{\sigma}) \\ &\quad \times \varphi(x_1, x_2)\frac{1}{\sigma}g(\frac{x_2^{(d)} - Y_{k+1}}{\sigma}) \\ &\quad \times \cdots \\ &\quad \times \varphi(x_d, z)\, \mathrm{d}x_1 \cdots \mathrm{d}x_d. \end{aligned}$$

Now, again we use

$$\frac{1}{\sigma}g(\frac{x - y}{\sigma}) \to \delta(x - y).$$

Due to the structure of φ, it turns out that Ψ_k factorizes in two terms depending only on x and z respectively. Thus, τ_k^{k+d} goes to zero again.

The connection with deterministic control theory is as follows. A system of the form

$$\begin{aligned} x_{n+1} &= Ax_n + f(y_n), \\ y_n &= Cx_n, \end{aligned}$$

with $x_n \in \mathbb{R}^d$, $y_n \in \mathbb{R}$ is said to be in observer canonical form if the *observability matrix*

$$O := [C, CA, CA^2 \ldots CA^{d-1}]$$

is invertible. A system in this form allows for a “deterministic filter”, a so called *observer*, of the form

$$\begin{aligned} \hat{x}_{n+1} &= A\hat{x}_n + f(y_n) + K(\hat{y}_n - y_n), \\ \hat{y}_n &= C\hat{x}_n. \end{aligned}$$

It is readily seen that for the error signal $e_n = x_n - \hat{x}_n$ the dynamics

$$e_{n+1} = (A - KC)e_n$$

holds. The nonsingularity of the observability matrix O now ensures that there is a vector K so that the above equations are stable and thus $e_k \to 0$.

Any system in observer canonical form can be transformed to the representation

$$\begin{aligned} x^{(1)}_{n+1} &= \tilde{f}_1(x^{(d)}_n), \\ x^{(2)}_{n+1} &= x^{(1)}_n + \tilde{f}_2(x^{(d)}_n), \\ &\vdots \\ x^{(d)}_{n+1} &= x^{(d-1)}_n + \tilde{f}_d(x^{(d)}_n), \\ y_n &= x^{(d)}_n. \end{aligned} \tag{4.13}$$

If we add independent dynamical noise on the right hand side of this equation, we get a Markov process with transition operator of the form (4.12). Thus we see that there is a connection between deterministic observer theory and stability of the filter.

In control theory now a great deal of work is done to identify conditions that guarantee the existence of a state space transformation which brings the system into canonical observer form. Identifying conditions guaranteeing that a Markov kernel can be transformed to the form (4.12) with a transformation of the state space remains an interesting open problem.

Chapter 5

Parametric Approximations

5.1 Basic definitions

Consider a parametrized set of probability distributions

$$Q : \mathcal{B} \times \Theta \to \mathbb{R}_{\geq 0},$$

where Θ is a subset of a finite dimensional vector space. The parametrisation is called *faithful* if $\theta_1 \neq \theta_2$ necessarily yields $Q(\cdot, \theta_1) \neq Q(\cdot, \theta_2)$. Let us denote the set of measures $\{Q(\cdot, \theta); \theta \in \Theta\}$ by $\mathcal{Q}$.

The basic idea of parametric approximation is to chose a parametrized family and replace π_n by a finite dimensional filter system. As already discussed in Section 3.1, in general this can be carried out only approximately. We will consider two different classes of $\mathcal{Q}$'s, namely *exponential* and *linear* families. They will be introduced after the general schemes have been explained.

26 Definition (Scheme I) *The idea in this scheme is to project π_n on a parametrized family by means of a minimisation technique. Recall the notations of Section 4.3, namely, set*

$$\mathcal{S} := \{\pi_0\} \cup \mathcal{Q} \cup \{S(y, q); q \in \mathcal{Q}, y \in \mathbb{R}\}$$

and let D *be a metric on* $\mathcal{S}$ *that however is* not *assumed to satisfy the triangle inequality. Suppose that for any* $\mu \in \mathcal{S}$ *the minimisation problem*

$$\min_{\theta \in \Theta} \mathsf{D}(\mu, Q(\cdot, \theta))$$

has a unique solution denoted by $m(\mu)$. *The value* $\min_{\theta\in\Theta} \mathsf{D}(\mu, Q(\cdot,\theta))$ *will be called the* projection residual. *Then we set recursively*

$$\tilde{\pi}_0 := Q(\cdot, m(\pi_0)), \tag{5.1}$$

$$\tilde{\pi}_{n+1} := Q(\cdot, m(S(Y_{n+1}, \tilde{\pi}_n))). \tag{5.2}$$

The finite dimensional filter has the dynamical variable θ *and reads as*

$$\theta_{n+1} = m(S(Y_{n+1}, Q(\cdot, \theta_n))).$$

The second scheme differs from the first only due to the fact that the approximation is applied between the prediction and update step.

27 Definition (Scheme II) *Now let, different from Scheme I,*

$$\mathcal{S} := \{\pi_0\} \cup \mathcal{Q} \cup \{\varphi(g(\cdot; y) * q); q \in \mathcal{Q}, y \in \mathbb{R}\}^{1}.$$

Thus, $\mathcal{S}$ *contains all of* Q *plus all measures that emerge from* Q *by first updating and then applying the Markov kernel.* D *is a metric on* S *that again is not assumed to satisfy the triangle inequality. Suppose again that for any* $\mu \in \mathcal{S}$ *the minimisation problem*

$$\min_{\theta \in \Theta} \mathsf{D}(\mu, Q(\cdot, \theta))$$

has a unique solution denoted by $m(\mu)$, *where again* $\min_{\theta\in\Theta} \mathsf{D}(\mu, Q(\cdot,\theta))$ *is referred to as the projection residual. Then we set recursively*

$$\tilde{\pi}_0 := Q(\cdot, m(\pi_0)), \tag{5.3}$$

$$\mathrm{d}\tilde{\pi}_{n+1} := c \cdot g(x, Y_{n+1}) \mathrm{d}Q(\cdot, m(\varphi \tilde{\pi}_n)). \tag{5.4}$$

Thus the approximation takes place between prediction and update step. To define finite dimensional filter system we can consider the affine family $\bar{Q}$ *defined by*

$$\mathrm{d}\bar{Q}(\cdot, \theta, y) := c \cdot g(x, y) \mathrm{d}Q(\cdot, \theta),$$

[1] Recall that by definition $g * q(A) = \int_A g\mathrm{d}q / q(g)$ (see Section 2.1).

with normalisation c. The finite dimensional filter now has dynamical variables (θ_n, y_n) and reads as

$$\begin{aligned} \theta_{n+1} &= m(\varphi\bar{Q}(\cdot, \theta_n, y_n)), \\ y_{n+1} &= Y_{n+1}. \end{aligned}$$

A few remarks seem to be in order.

1. The reason why we dropped the triangle inequality is that we will use, among others, the Kullback-Leibler distance for D. However, due to theorem 15 we will establish a relationship to TV and thus make applicable the analysis of Section 4.3.

2. The difference between the two schemes has important practical and theoretical consequences. The first scheme is appealing from a theoretical point of view since the quantity that is minimized is in fact what we have called the approximation residual $\mathsf{D}(S_{k+1}(\tilde{\pi}_k), \tilde{\pi}_k)$ in Section 4.3. However, the second scheme has the nice property of performing the update step *exactly* which may be a conceptual advantage.

3. Filtering problems arising from discretely sampled continuous time problems allow for another scheme called *projection filtering*. This scheme was proposed by Brigo, Hanzon and LeGland in [9, 8]. It however applies to exponential families only and will be discussed in the corresponding Subsection 5.2.2.

In this thesis, for $\mathcal{Q}$ only linear and exponential families will be considered. It is furthermore assumed from now on that the system is regular. As already mentioned this means that $S(y, \mu)$ has a density with respect to a dominating σ-finite measure λ if μ has. For both families we will discuss some choices of the metric D, possible restrictions on $\mathcal{Q}$ and how to solve the minimisation problem to find θ. The Table 5.1 gives an overview over the discussed setups. In the remainder of this section we will explain this table columnwise, referring to the case of, for example, exponential families and the KL-distance as Case 1A etc.

To each case, a small subsection is devoted. The subsection considers restrictions on μ and $\mathcal{Q}$ ensuring that the minimisation problem

$$\min_\theta \mathsf{D}(\mu, Q(\cdot, \theta))$$

can be solved. The considerations will imply then restrictions on $\mathcal{S}$. Since $\mathcal{S}$ is determined by the family $\mathcal{Q}$ and the dynamical system, we discuss which choices of systems and $\mathcal{Q}$ yield suitable $\mathcal{S}$. An exhaustive treatment of all possible cases however cannot be given here. This would amount to a treatise on convex analysis. We will merely demonstrate the usefulness of the concepts by some examples.

Furthermore, for each case a connection to the results of Section 4.3 is established. The analysis carried out there requires a metric that fulfills the triangle inequality. Of the metrics discussed in this thesis however KL and HE do not share this feature, but play an essential role in parametric approximation methods. So the problem is basically to establish a connection between the projection residuals computed in this metrics to the approximation residuals with respect to one of the metrics considered that fulfills the triangle inequality.

	A: Exp. Families	B: Linear Families
1.: KL	Regular systems, especially of exponential form. Establish bound on approximation residual in TV norm.	Regular systems. Establish bound on approximation residual in TV norm.
2.: HE	Discretely observed continuous time systems. Establish bound on approximation residual in TV norm.	—
3.: H	Strictly regular systems, especially on compact sets. An explicite solution of the minimisation problem is not given.	Strictly regular systems, especially on compact sets.

Table 5.1: Possible metrics, families and the systems they are applied to

5.2 Exponential Families

Exponential families are very important in statistical estimation theory, since many important random quantities obey an exponential distribution law. A thorough discussion of this subject is given in [4]. Their application to filtering of discrete time series was discussed in [12]. We already mentioned the succesful application to the continuous time case in [9, 8]. Furthermore, existence results on finite dimensional filters established in Chapter 4 yield that exponential families are in a certain sense a natural choice.

Let λ be a σ-finite measure on E. Recall from Section 3.2 that a family $Q(\cdot,\cdot\cdot)$ of probability measures on E is of exponential form with respect to λ if

$$\frac{\mathrm{d}Q(\cdot,\theta)}{\mathrm{d}\lambda}(x) = q(x,\theta) := \exp(\theta c(x) - \psi(\theta)),$$

where $c : E \to \mathbb{R}^d$ is a measurable function and $\Theta \subset \mathbb{R}^d$. The function ψ is defined by the relation

$$1 = \int \mathrm{d}Q(\cdot,\theta). \tag{5.5}$$

It turns out that ψ as well as $\exp(\psi)$ are convex functions on the convex set

$$\Theta := \{\theta \in \mathbb{R}^d ; \psi(\theta) < \infty\}.$$

By the exponential form of the relative density, all $Q(\cdot,\theta)$ are mutually absolutely continuous.

If we require the components $c_i(x)$ of $c(x)$ to be affinely independent, i.e. the function $\theta_0 + \sum \theta_i c_i(x)$ vanishes identically if and only if all θ are equal to zero, then the parametrisation turns out to be faithful. Taking the derivative with respect to θ_i on both sides of (5.5) one obtains

$$\eta_i := \int c_i(x)\, q(x,\theta)\, \mathrm{d}\lambda(x) = \frac{\partial \psi}{\partial \theta_i}(\theta). \tag{5.6}$$

The η_i are called the *c_i–moments* or *expectation parameters*, in contrast to the θ_i, which are called *canonical parameters*. The c_i's are called *canonical statistics*. One easily obtains the following identity:

$$g_{ij} := \frac{\partial \eta_i}{\partial \theta_j} = \int \frac{\partial \log q}{\partial \theta_i} \frac{\partial \log q}{\partial \theta_j}\, q\, \mathrm{d}\lambda = \frac{\partial^2 \psi}{\partial \theta_i \partial \theta_j}.$$

Since the c_i's are affinely independent, the functions $\partial \log q / \partial \theta_i$ are linear independent and therefore g_{ij} is a nonsingular positive definite matrix, called the Fisher *metric*. Hence, the function ψ is *strictly* convex. Furthermore, it is easy to see that the expectation parameters η and the canonical parameters θ are connected by a Legendre transform of the function ψ which is defined as

$$\psi^*(\eta) = \sup_\theta \left[\theta\eta - \psi(\theta)\right].$$

In fact, if (5.6) holds, then θ is a critical point for $\theta\eta - \psi(\theta)$. But this function (for η fixed) is strictly concave. A strictly concave function, however, can have at most one critical point. The Legendre transform of a strictly convex function thus uniquely connects θ and η, hence the expectation parameters η_i are *globally* diffeomorphic functions of the θ_i. Therefore the expectation parameters form another coordinate system for $\mathcal{Q}$, which is of great use in the following. It can be shown that

$$\eta = \frac{\partial \psi}{\partial \theta} \qquad \text{if and only if} \qquad \theta = \frac{\partial \psi^*}{\partial \eta}.$$

Thus we see that a given η can be transformed to θ–coordinates if it is in the domain of $\frac{\partial \psi^*}{\partial \eta}$. This domain is not convex and usually smaller than the *effective domain* of ψ^*, the region where it is finite. It can be shown that it is larger than the *relative interior* of the effective domain of ψ^*, which is convex. Similary, the domain of $\frac{\partial \psi}{\partial \theta}$ is smaller than Θ, the effective domain of ψ. Nevertheless, it is larger than the relative interior of Θ. Thus, pathological points appear only at the boundary. For this assertions see the book of Rockafellar [57].

Unfortunately, for many interesting exponential families there is neither a closed form expression for ψ nor for the diffeomorphism $\theta(\eta)$. Therefore numerical schemes have to be employed to compute them. We will briefly discuss some possible approaches in Appendix A.3.

We have seen that in the approximation scheme II $\tilde{\pi}_n$ is not a member of the parametrized family. One may ask if for some special cases one can achieve that even in scheme II, $\tilde{\pi}_n$ stays in the parametrized family. This is possible if $g(y, x)$ as a function of x is of exponential type. One may then chose an exponential family containing also g, and since the multiplication of two exponential densities again yields an exponential density the update step in scheme II will keep $\tilde{\pi}_{n+1}$ a member of the exponential family. Following the conventions in [8], such families will be called *convenient*.

5.2.1 The Kullback-Leibler distance (Case 1A)

We consider now Table 5.1, row 1, column 1. Let $Q(\cdot,\theta)$ belong to an exponential family and μ be an arbitrary measure having a density with respect to λ. Since $Q(\cdot,\theta) \lessgtr \lambda$, μ has a density with respect to Q as well, thus

$$\begin{aligned}\mathsf{KL}(\mu,Q) &= \int \frac{\mathrm{d}\mu}{\mathrm{d}Q}\log\frac{\mathrm{d}\mu}{\mathrm{d}Q}\,\mathrm{d}Q \\ &= \int \frac{\mathrm{d}\mu}{\mathrm{d}\lambda}\log\frac{\mathrm{d}\mu}{\mathrm{d}\lambda}\mathrm{d}\lambda - (\theta\int c(x)\mathrm{d}\mu - \psi(\theta)).\end{aligned}$$

Minimizing this expression w.r.t. θ is again related to a Legendre transform of ψ and therefore a convex optimisation problem. Setting equal to zero the derivative of this expression with respect to θ we get the condition

$$\eta(\theta) \overset{!}{=} \int c(x)\mathrm{d}\mu. \tag{5.7}$$

We are thus faced with two problems:

1. $\mathsf{KL}(\mu,\theta)$ has to be finite for at least one θ.
2. The quantity $\int c(x)\mathrm{d}\mu$ has to be in the domain of $\theta(\eta)$, which, as we have seen, coincides with the domain of $\frac{\partial\psi^*}{\partial\eta}$.

We will not carry out a detailed analysis on choosing the family and the system so that 1 and 2 hold for all elements of $\mathcal{S}$ but from now on merely assume that this is the case. At the end of this section we will give an important example where the method can be seen to work and furthermore we will give a hint how more general systems can be found to which the method is applicable.

To solve the minimisation problem, apply the diffeomorphism θ to Equation 5.7 and get the expression for m (see schemes I and II)

$$m(\mu) = \theta(\int c(x)\mathrm{d}\mu).$$

We now discuss scheme I. The finite dimensional filter system in case of scheme I now reads as

$$\theta_{n+1} = \theta\left(\int c(x)\cdot c\cdot g(Y_{n+1},x)\int \varphi(\mathrm{d}x,z)q(z,\theta_n)\mathrm{d}\lambda(z)\right).$$

At first sight this looks messy, but it is possible for some cases to carry out the integral over $\mathrm{d}\lambda(x)$ explicitely. Define the functions

$$\gamma_1(y,z) := \int g(y,x)\varphi(\mathrm{d}x,z),$$
$$\gamma_c(y,z) := \int c(x)\cdot g(y,x)\varphi(\mathrm{d}x,z).$$

These functions can be computed offline, which means that their calculation does not require the explicit value of Y_n. This is important in applications. If $\varphi(x,z)$ and $g(y,x)$ are normal densities in x, then this calculation can be carried out even explicitely. The finite dimensional filter then reads as

$$\theta_{n+1} = \theta\left(\frac{\int \gamma_c(Y_{n+1},z)q(z,\theta_n)\mathrm{d}\lambda(z)}{\int \gamma_1(Y_{n+1},z)q(z,\theta_n)\mathrm{d}\lambda(z)}\right). \tag{5.8}$$

The finite dimensional filter system in case of scheme II reads as

$$\begin{aligned}\theta_{n+1} &= \theta\left(\int c(x)\int \varphi(\mathrm{d}x,z)c\cdot g(y_n,z)q(z,\theta_n)\mathrm{d}\lambda(z)\right),\\ y_{n+1} &= Y_{n+1}.\end{aligned}$$

Here we can define the functions

$$\gamma_c(z) := \int c(x)\varphi(\mathrm{d}x,z)$$

in order to write the finite dimensional filter as

$$\begin{aligned}\theta_{n+1} &= \theta\left(\frac{\int \gamma_c(z)g(y_n,z)q(z,\theta_n)\mathrm{d}\lambda(z)}{\int g(y_n,z)q(z,\theta_n)\mathrm{d}\lambda(z)}\right),\\ y_{n+1} &= Y_{n+1}.\end{aligned} \tag{5.9}$$

To show that despite the mentioned open problems the method is useful consider the following

28 Example (Gaussian exponential family) Let an observed dynamical system in $\mathbb{R}_d$ be given by the relations

$$\begin{aligned}X_{n+1} &= f(X_n) + \rho V_n,\\ Y_n &= X_n^{(1)} + \sigma W_n,\end{aligned}$$

with standard normal random variables $V_n \in \mathbb{R}^d$, $W_n \in \mathbb{R}$. Here $X_n^{(1)}$ means the first component of X_n. Assume p_{X_0}, the probability density of X_0 has finite first and second moments. For $\mathcal{Q}$ we consider d-dimensional Gaussian families. Thus λ is Lebesgue measure and $c(x) = (x, x_i x_j), i, j = 1 \dots d$. The domain of $\theta(\eta)$ consists of all pairs (μ, Γ), where $\mu \in \mathbb{R}^d$ and $\Gamma_{ij} - \mu_i \mu_j$ is positive definite, i.e. a covariance matrix. Furthermore, $\theta(\eta)$ can be given in closed form. It turns out that

$$g(y, x) = \exp(\frac{yx}{\sigma^2} - \frac{1}{2}\frac{x^2}{\sigma^2})$$

and

$$\varphi(x, z) = \frac{1}{\sqrt{2\pi}^d |\det \rho|} \exp\left[-\frac{1}{2}(x - f(z))(\rho\rho^T)^{-1}(x - f(z))\right].$$

Now a straightforward but lengthy calculation shows that for scheme I γ_0 and γ_c are well defined and moreover the argument of θ in the right hand side of (5.8) actually is in the domain of θ.

For scheme II this is even more obvious. It turns out that

$$\gamma_c(z) = (f(z), \rho\rho^T + f(z)f(z)^T).$$

Now integrating this over a normal density (as required in (5.9)) gives always an η of the required kind. In fact, the finite dimensional filter can be represented in η-coordinates, which amounts to equations for Γ and μ. For scheme II, for example, they read as

$$\begin{aligned} \mu_{n+1} &= \int f(z)q(z, \mu_n, \Gamma_n)\mathrm{d}z + \frac{1}{\sigma^2}\Gamma_{n+1}e_1(Y_{n+1} - \int f_1(z)q(z, \mu_n, \Gamma_n)\mathrm{d}z), \\ \Gamma_n^+ &= \rho\rho^T + \int f(z)f(z)^T q(z, \mu_n, \Gamma_n)\mathrm{d}z - \mu_{n+1}^T\mu_{n+1}, \\ \Gamma_{n+1}^{-1} &= \left(\Gamma_n^+\right)^{-1} + \frac{1}{\sigma^2}e_1^T e_1. \end{aligned} \tag{5.10}$$

if the integrals over f and ff^T can be carried out either numerically or even explicitely (as it is the case for polynomial f) this is a very fast algorithm. If the underlying system has N dimensions, the filter has $N + \frac{N(N+1)}{2}$ dimensions and is implemented straightforwardly. This filter does not need more CPU-time or storage than the Extended Kalman Filter but shows superior results, as we have seen already in the introduction.

29 Example (Systems of exponential form) The systems of the last example where of *exponential form*, this is

$$\varphi(x,z) = \exp(T(z)c(x) - \psi(T(z)))$$

for certain functions $c(x)$ and $T(z)$. Assuming that

$$\exp(\theta c(x) - \psi(\theta))\mathrm{d}\lambda$$

forms an exponential family with a nontrivial parameter set Θ, we can consider the Case 1A for this setup. We restrict our attention to scheme II. For γ_c we get

$$\gamma_c = \int c(x)\exp(T(z)c(x) - \psi(T(z)))\mathrm{d}\lambda = \eta(T(z)).$$

Now, assuming that $T(z) \in \Theta$ for all $z \in E$ we compute the argument of θ in Equation (5.9) and obtain

$$\frac{\int \eta(T(z))g(y_n,z)q(z,\theta_n)\mathrm{d}\lambda}{\int g(y_n,z)q(z,\theta_n)\mathrm{d}\lambda}. \tag{5.11}$$

Note that this is a convex combinaton over several η. In order to define the finite dimensional filter Equation (5.9) we have to require that the expression (5.11) is in the domain of $\theta(\eta)$ which is equal to the domain of $\frac{\partial\psi^*}{\partial\eta}$. Now $\eta(T(z))$ is always in this domain for any z, by construction. As we mentioned already, this domain is not convex. However, we do not loose much if we require $\eta(T(z))$ to be in the relative interior of the domain of ψ^* for all z, which amounts to the requirement that $T(z)$ is in the relative interior of Θ for all z. In our case this is just the topological interior of Θ. With this requirement, the algorithm of scheme II is well defined.

30 Numerical example (Hénon system) The mentioned Hénon system can be analyzed by Gaussian density functions. The noise reduction results where already shown in Figure 1.1b and also the dash–dotted SNR improvement curve shown in Figure 1.2 was computed in this way. The finite dimensional filter allows for an explicite representation since the Hénon system is polynomial, thus the right hand side of equations (5.10) can be calculated explicitely.

Error analysis of Case 1A

The purpose of this section is to make the error bound of Section 4.3 applicable to Case 1A. The projection residuals are in the KL-metric, which does not satisfy the triangle inequality. Thus, the analysis of Section 4.3 is not applicable directly. Therefore, a connection to the TV-metric will be established now. To the approximation residual $\mathsf{TV}(S_k\tilde{\pi}_{k-1}, \tilde{\pi}_k)$ we apply the Bretagnole–Huber inequality (Lemma 16) to get

$$\mathsf{TV}(S_k\tilde{\pi}_{k-1}, \tilde{\pi}_k) \leq 2\sqrt{1 - \exp\left(-\mathsf{KL}(S_k\tilde{\pi}_{k-1}, \tilde{\pi}_k\right))}.$$

But in scheme I $\mathsf{KL}(S_k\tilde{\pi}_{k-1}, \tilde{\pi}_k)$ is exactly the quantity we minimize. The minimum value is

$$\mathsf{KL}(S_kQ(\cdot, \theta_{k-1}), \lambda) - \psi^*(\theta_k),$$

where $\psi^*(\theta_k)$ is the Legendre transform of ψ. This Legendre transform can be computed in course of the calculations necessary for carrying out the minimisation of KL. However, the first term of the above equation requires application of S_k to $Q(\cdot, \theta_{k-1})$, i.e. *one filtering step* has to be carried out.

Scheme II does *not* explicitely minimize the approximation residual. Hence there is no advantage of computing the approximation residual in the KL-metric rather than the TV-metric.

The interesting question whether the approximation residual is stationary or ergodic may now in principal be investigated. It is quite possible that conditions on the system can be identified for which the *compound process* (X_n, θ_n) satisfies the conditions of Theorem 45. Then ergodicity is established. However, we have not carried out the analysis in more detail.

5.2.2 The Hellinger distance (Case 2A)

We consider now Table 5.1, row 2, column 1. The Hellinger distance was used in the investigation of the *projection filter* by Brigo, Hanzon and LeGland in [9, 8]. The projection filter is a device applicable to continuous time systems. We will give only a very brief explanation of the concept. We suppose the setup of Example 5. Furthermore, we assume λ to be the Lebesgue measure. The exponential family is as well fixed and assumed to be convenient. This is we assume $g(y, x)$ to be of the form

$$\log g(y, x) = \gamma_1(y) \cdot c(x) + \gamma_2(y).$$

A finite dimensional filter for this setup can be established straightforwardly if the solution $\varphi(x,z,t)$ of the Fokker–Planck equation is available at $t=\delta$, the sampling time. Usually however this is not the case. The technique proposed in [9, 8] to overcome this problem is to *project* the Fokker–Planck equation onto the *tangent space* of the exponential familiy. An exponential family in fact admits the structure of a differential Riemannian manifold. Eventually this amounts to a differential equation for the parameter θ. A "shortcut" to obtain these equations goes as follows. Suppose $p_t(x)$ is a solution of the Fokker–Planck equation with initial condition $q(x,\theta)$ of exponential form. Write this as

$$\frac{\partial p_t}{\partial t}(x) = \mathcal{L}p_t(x)$$

with the differential operator $\mathcal{L}$ given by the right hand side of the Fokker–Planck equation. Usually $p_t(x)$ will not be of exponential form for $t>0$. Nevertheless we may calculate

$$\eta_t = \int c(x)p_t(x)\mathrm{d}x.$$

Differentiating with respect to t gives

$$\dot{\eta}_t = \int c(x)\mathcal{L}p_t(x)\mathrm{d}x = \int \mathcal{L}^*c(x)p_t(x)\mathrm{d}x,$$

where $\mathcal{L}^*$ is the adjoint of $\mathcal{L}$. Now we *assume* $p_t(x)$ to be of exponential form with time varying parameter. Thus

$$\dot{\eta}_t = \int c(x)\mathcal{L}p_t(x)\mathrm{d}x = \int \mathcal{L}^*c(x)q(x,\theta_t)\mathrm{d}x.$$

The right hand side is now a function of θ_t. On the left hand side we apply the identity

$$\dot{\theta}_t = \frac{\mathrm{d}\theta}{\mathrm{d}\eta}\dot{\eta}_t = g^{-1}\dot{\eta}_t,$$

where g is the Fisher metric. Thus we get the following differential equation for θ_t (written out in coordinates)

$$\sum_j g_{i,j}^{-1}\dot{\theta}_t^j = \int \mathcal{L}^*c_j(x)q(x,\theta_t)\mathrm{d}x. \tag{5.12}$$

A finite dimensional filter is now established by the following scheme

1. Set $\theta_0 = m(\pi_0)$ where m is defined using the metrics KL or H.

2. Solve 5.12 on the interval $t = [0, \delta]$ with initial condition θ_0. Call the result θ_t^-. At $t = \delta$ the first observation Y_1 is available.

3. Set
$$\theta_1 := \theta_\delta^- + \gamma_1(Y_1).$$

4. Repeat steps 2 and 3 using now θ_1 as initial condition etc.

31 Numerical example (Lorenz system) As in the discrete time case (Case 1A) the normal densities are a special case of exponential densities and lead to useful filters. Especially for polynomial systems, the finite dimensional filter allows for an explicit representation. We developed a software tool for the automatic generation of such finite dimensional filters. This package provides a `matlab` m-file that generates a C-code for the Equation (5.12) from symbolic information about the system under consideration. This C-code can then be compiled and linked to `matlab` to be solved with an ode solver.

We applied this tool to the Lorenz system

$$\begin{aligned}\dot{X}_1 &= s(X_2 - X_1),\\ \dot{X}_2 &= rX_1 - X_2 - X_1X_3,\\ \dot{X}_3 &= X_1X_2 - bX_3,\end{aligned}$$

with $s = 10$, $r = 28$ and $b = 8/3$, the standard parameters. Observations where taken according to

$$Y_n = X_1(t_n) + \sigma W_n,$$

where we used in fact not equidistant t. This makes no difference except for Equation (5.12), which simply has to be integrated from t_n to t_{n+1}. The results are plotted in Figure 5.1. This result was obtained for SNR = 20, which amounts to a σ of about 0.8. In fact, we furthermore assumed the parameters s, r and b to be unknown as well. How the estimation of the parameter was accomplished is discussed in Section 6.2.

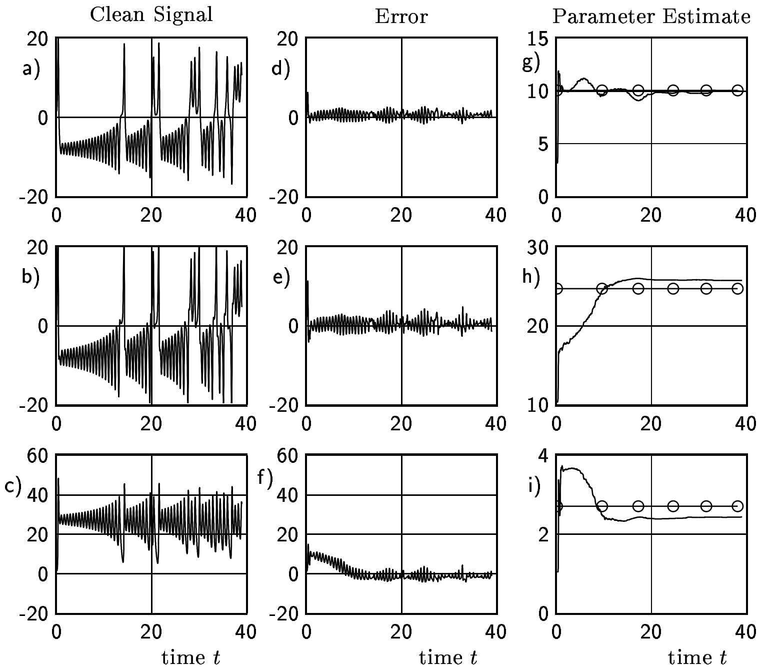


Figure 5.1: Results of an approximative filter using exponential families. The first column shows the three components of the Lorenz system. The second column shows the error between the filter output and the true signal. Small fluctuations remain, but the bias is negligible. The third column shows an estimate of the parameters s, r and b. The correct values are marked with circles on a line. We see that a small bias remains.

Error analysis of Case 2A

The error analysis for the approximative solution of the Fokker–Planck equation given by Equation (5.12) was carried out in [48] where the following result is obtained

32 Lemma *Let $t \in [0, \delta]$ and p_t be a solution of the Fokker–Planck equation. Let $q_t = q(x, \theta_t)$ be of exponential form, where θ_t is a solution of the corresponding Equation (5.12). Then*

$$\frac{\mathrm{d}}{\mathrm{d}t}\sqrt{\mathsf{HE}(q_t, p_t)} \leq \|\mathcal{R}_t(\sqrt{q_t})\|,$$

where $\|\cdot\|$ denotes the L_2–norm with respect to x and $\mathcal{R}$ is the projection residual operator. *This operator is a differential operator and turns out to be*

$$\frac{1}{\sqrt{q_t}}\mathcal{R}_t(\sqrt{q_t}) = \frac{1}{2}\frac{\mathcal{L}q_t}{q_t} - \frac{1}{2}(c(x) - \eta(\theta_t))\dot{\theta}_t.$$

In the filtering algorithm we have the same initial conditions, i.e. $q(\cdot, \theta_0) = p_0$, which means $\mathsf{HE}(q_0, p_0) = 0$. The lemma then yields

$$\sqrt{\mathsf{HE}(q_\delta, p_\delta)} \leq \int_0^\delta \|\mathcal{R}(\sqrt{q_t})\|\mathrm{d}t.$$

We have then applying Lemma 17

$$\mathsf{TV}(q_\delta, p_\delta) \leq 2\left[\left(\int_0^\delta \|\mathcal{R}(\sqrt{q_t})\|\mathrm{d}t + 1\right)^2 - 1\right].$$

To obtain eventually an estimate for the approximation residual TV we apply the result of Lemma 13 and obtain

33 Theorem *For the projection filter the following approximation residual is obtained*

$$\mathsf{TV}(S_k\tilde{\pi}_{k-1}, \tilde{\pi}_k) \leq 2\frac{\max g(Y_k, x)}{\int q_\delta(x)g(Y_k, x)\mathrm{d}x}\left[\left(\int_0^\delta \|\mathcal{R}(\sqrt{q_t})\|\mathrm{d}t + 1\right)^2 - 1\right],$$

where $q_0(x) := \tilde{\pi}_{k-1}(x)$.

This error bound looks messy. However, a direct calculation of the left hand side would amount to calculate $S_k(\tilde{\pi}_k)$ which requires to solve the Fokker–Planck equation in the interval $[0, \delta]$ *exactly*.

5.2.3 The Hilbert metric (Case 3A)

We will not say much about this case, and a few numerical problems will be left open. Recall from the definition of H that $\mathsf{H}(\mu, \nu)$ may be set to ∞ if μ and ν are not comparable. However, useful bounds will emerge of course only if this quantity is finite. If μ is comparable to $Q(\cdot, \theta_c)$ then by defining the "recentered" exponential family

$$Q'(\cdot, \theta) = Q(\cdot, \theta - \theta_c)$$

we can always assume $\theta_c = 0$. The reader may convince himself that Q' is in fact an exponential family with parameter space $\Theta' = \Theta - \theta_c$ and $\lambda' = \exp(\theta_c c(x) - \psi(\theta_c))\lambda$.

For the filtering problem we will have a nontrivial $\mathcal{S}$ if the system is supposed to be *strictly regular*, defined as follows

34 Definition *A system is* strictly regular *with respect to a σ-finite measure λ if, for any μ comparable to λ, then, also $S(y, \mu)$ is comparable to λ for any $y \in \mathbb{R}$.*

Note that this definition implies that λ is finite. On a compact state space a system is regular if $S(y, \mu)$ has a positive density with respect to λ whenever μ has. This is a practically quite relevant case.

The following lemma gives conditions under which the problem of finding the minimum of $\mathsf{H}(\mu, Q(\cdot, \theta))$ is a convex optimisation problem.

35 Lemma *Suppose μ is a finite measure that is comparable to λ. Then*

$$\mathsf{H}(\mu, Q(\cdot, \theta))$$

is convex.

Proof Obviously, if μ is comparable to λ, then $\mu(x) := \frac{\mathrm{d}\mu}{\mathrm{d}\lambda}$ exists and

$$\mathsf{H}(\mu, Q(\cdot, \theta)) = \max_x(\log(\mu(x)) - \theta c(x)) - \min_x(\log(\mu(x)) - \theta c(x)).$$

This expression may be infinite but is definitely finite for $\theta = 0$. Thus $\mathsf{H}(\mu, Q(\cdot, \theta))$ is a function of θ having values in $\mathbb{R} \cup \{\infty\}$. We will show that it is convex. For any $t \in [0,1]$ it holds that

$$\begin{aligned}
&\max_x (\log \mu(x) - (t\theta_1 + (1-t)\theta_2)c(x)) \\
&= \max_x ((t + (1-t)) \log \mu(x) - (t\theta_1 + (1-t)\theta_2)c(x)) \\
&= \max_x (t(\log \mu(x) - \theta_1 c(x)) + (1-t)(\log \mu(x) - \theta_2 c(x))) \\
&\leq t \max_x (\log \mu(x) - \theta_1 c(x)) + (1-t) \max_x (\log \mu(x) - \theta_2 c(x)) \, .
\end{aligned}$$

The same appears to be true for min with the $\geq$-sign. Thus

$$\mathsf{H}(\mu, Q(\cdot, t\theta_1 + (1-t)\theta_2)) \leq t\mathsf{H}(\mu, Q(\cdot, \theta_1)) + (1-t)\mathsf{H}(\mu, Q(\cdot, \theta_2))$$

□

Error analysis of Case 3A

There is not much to say since the Hilbert metric satisfies the triangle inequality and thus the error bound (4.6) is directly applicable. Furthermore it should be remarked that the projection residual is in *both cases* (scheme I and scheme II) the same as the approximation residual in (4.6). For scheme I this is clear, but for scheme II this is true as well since the update step applied after the minimisation does not change the value of H.

5.3 Linear families

The use of these families is motivated by the quite natural idea of approximating the pdf π_n by a piecewise constant function. For systems with finite state space, the exact filter has this form. The piecewise constant function can also be considered as a convex combination of fixed characteristic functions. To generalize this idea, let λ be a σ-finite measure, $p_i(x), i = 1 \ldots k$ be a set of normalized positive functions, i.e. $\int p_i(x) \, \mathrm{d}\lambda = 1$ for all $i = 1 \ldots k$. Then the resulting *linear family* $\mathcal{Q}$ consists of all convex combinations of the form

$$q(x, \theta)\mathrm{d}\lambda = \sum_{i=1}^{k} \theta_i p_i(x) \mathrm{d}\lambda,$$

where the θ's form the convex simplex

$$\Theta = \{\theta \in \mathbb{R}^k ; 0 \leq \theta_i \leq 1, \sum_{i=1}^{k} \theta_i = 1\}.$$

As a special class of linear models we consider *only* the case of nonoverlapping support of the p_i's. This is, for example the case when a pdf is represented by a piecewise constant function (histogram) on a grid.

5.3.1 The Kullback-Leibler distance (Case 1B)

For any probability measure μ having density $\mu(x)$ w.r.t. λ the Kullback–Leibler distance $\mathsf{KL}(Q(\cdot,\theta),\mu)$ reads as

$$\mathsf{KL}(Q(\cdot,\theta),\mu) = \sum_i \theta_i \int \frac{p_i(x)}{\mu(x)} \log \sum_i \theta_i \frac{p_i(x)}{\mu(x)} \mathrm{d}\lambda.$$

Note that $\mathsf{KL}(Q,\mu)$ is used instead of $\mathsf{KL}(\mu,Q)$. The reason is that $\mathsf{KL}(Q,\mu)$ works as well if Q has only compact support, i.e. is zero outside a compact set. Thus using this approach is useful if truncating the probability densities is desired. Furthermore, the mentioned piecewise constant functions have compact support. Problems will appear if μ has noncompact support, but $\frac{\mathrm{d}\mu}{\mathrm{d}Q}$ has to be considered. This quantity is then undefined.

Using the method of Lagrange multipliers it is easy to see that to minimize the Kullback–Leibler distance we have to set

$$\theta_i = c \cdot \exp(-\mathsf{KL}(Q(\cdot,e_i),\mu)),$$

where c is a normalisation to fulfill the requirement $\sum_j \theta_j = 1$. This expression is always defined. It yields the quite reasonable criterion that $p_i(x)$ should be weighted the less, the larger its KL-distance to μ. The finite dimensional filters for scheme I and scheme II can now be written down in a straightforward manner.

Error analysis of Case 1B

The error analysis of Case 1A in principle carries over to this case. The main point here was the Bretagnole–Huber inequality. Let us calculate the minimum value of KL. It turns out that for nonoverlapping supports,

$$\mathsf{KL}(Q(\cdot,\theta),\mu)=\sum_i \theta_i \log\theta_i+\sum_i \theta_i \mathsf{KL}(Q(\cdot,e_i),\mu).$$

If KL is minimal we have to replace θ_i by $c\cdot\exp(-\mathsf{KL}(Q(\cdot,e_i),\mu))$ which yields

$$\mathsf{KL}(Q(\cdot,\theta),\mu)=\log c=-\log\sum_i \exp(-\mathsf{KL}(Q(\cdot,e_i),\mu)).$$

This quantity can be calculated during the approximation process without any further effort.

5.3.2 The Hilbert metric (Case 3B)

The Hilbert metric may well be used in approximation using the linear families. Let $\mathcal{Q}$ be a linear family with carrier measure λ and suppose all $Q(\cdot,e_i)$ have mutually nonoverlapping support. Suppose now that μ is a measure having a density $\mu(x)$ with respect to λ and that μ is comparable to $Q(\cdot,\theta)$ for at least one θ. Denote by A_i the support of $Q(\cdot,e_i)$. Then we can write (due to the nonloverlapping supports)

$$\begin{aligned}\mathsf{H}(Q(\cdot,\theta),\mu)&=\log\frac{\sup_x\left[\sum_i \theta_i p_i(x)/\mu(x)\right]}{\inf_x\left[\sum_i \theta_i p_i(x)/\mu(x)\right]}\\&=\log\frac{\max_i \theta_i \sup_{x\in A_i}\left[p_i(x)/\mu(x)\right]}{\min_i \theta_i \inf_{x\in A_i}\left[p_i(x)/\mu(x)\right]}.\end{aligned}$$

We will now show that this expression is minimized if

$$\theta_i=c\cdot\sqrt{\sup_{x\in A_i}\left[\frac{\mu(x)}{p_i(x)}\right]\cdot\inf_{x\in A_i}\left[\frac{\mu(x)}{p_i(x)}\right]}. \tag{5.13}$$

To see this, note that for any we have $\frac{x_i}{y_i}\le\frac{x_i}{\min y_i}$, whence $\max\frac{x_i}{y_i}\le\frac{\max x_i}{\min y_i}$. This yields

$$\max\frac{x_i}{y_i}\le\frac{\max\theta_i x_i}{\min\theta_i y_i}.$$

However, equality appears here if

$$\theta_i=c\cdot\frac{1}{\sqrt{x_i y_i}}.$$

Thus, this θ must be the minimizer. Replacing x_i by $\sup_{x\in A_i}[p_i(x)/\mu(x)]$ and y_i by $\inf_{x\in A_i}[p_i(x)/\mu(x)]$ in this calculation we get the assertion.

Equation (5.13) can now directly be applied to get a finite dimensional filter. Of course we have again to require that the system is regular with respect to λ. Furthermore, in order to ensure that all θ are always well defined we need to require that

$$\sup_{x\in A_i}[p_i(x)/\mu(x)] < \infty \qquad \forall \mu \in \mathcal{S},$$

which is essentially a restriction on φ.

Error analysis of Case 3B

The same remarks as for the Case 3A apply here.

5.4 Concluding remarks

It turns out that the presented framework unifies a lot of known different approaches. E.g. approximation by linear families (in our language) was already proposed in the 1970's (see [67]). Furthermore, using exponential families amounts to compute a few moments of the actual distribution and discard the higher order ones (see the Appendix). This is in fact the main idea behind the *assumed density principle* (see [37, 74]). This approach has however been carried out only for Gaussian densities. Based on the assumed Gaussian density filter, a further simplification has been proposed by Julier and Ullmann (the *unscented filter*, see [38]). The main idea here is to replace the exact calculation of the moments by an approximation that is applicable also in cases where the discrete time dynamical system equations are not given in mathematically closed form (e.g. if a continuous time system is investigated and a numerical integration scheme is employed).

Chapter 6

Further Approximation Schemes and Applications

6.1 Monte Carlo methods

Classically Monte Carlo methods where conceived to evaluate certain integrals that can be understood as the expectation value of a random quantity. Suppose for example f is a function and p a probability density and we want to calculate

$$\int f(x) \cdot p(x) \, \mathrm{d}x.$$

The idea of Monte Carlo simulation simply is to generate a large number of other independent random variables $X_1, X_2, \ldots, X_M$ (called the *ensemble*) featuring the same statistical properties, that is in this case having the distribution $p(x)$. Then (according to the law of large numbers) the expectation is approximately given by the empirical mean over the ensemble, which in our case means

$$\int f(x) \cdot p(x) \, \mathrm{d}x \cong \frac{1}{M} \sum_{k=1}^{M} f(X_k).$$

The problem is of course how to generate the ensemble, which may be quite difficult for complicated p.

These ideas can be modified for the purpose of state estimation in two ways. The idea of weighted particles is, roughly speaking, to work with a large number of independent Markov Processes featuring the same statistical properties as the original signal process $\{X_n\}$. The ensemble does not provide an approximation of π_n, but allows for approximative calculation of any integral of the form

$$\int f(x) \cdot \pi_n(x) \mathrm{d}x$$

by a *weighted* average over the ensemble of Markov processes. The weights depend on the observations Y_n.

The SNR improvement result in Figure 1.2 denoted “Monte C.” was computed with the weighted particle approach. The decrease of the SNR improvement around 40 dB is due to the (finite) ensemble size of $M = 1000$. Using more “particles” one may achieve better SNR improvement also for higher SNR of the time series (with higher computational costs, of course).

An alternative is the method of evolutionary particles which in contrast to the weighted particles directly provides an ensemble approximating π_n. This method consists of two steps resembling the prediction and the update step. The ensemble points are not independent like in the weighted particle method.

The general drawback of Monte Carlo Methods is the required computer power. It is necessary to store the ensemble points and, as we will see, some weight vectors associated to each ensemble point. Furthermore, the dynamical equations (e.g. iterated maps or stochastic differential equations) have to be solved for all ensemble points in parallel. This obviously requires more power than the previously discussed low dimensional filters.

6.1.1 Weighted Particle Method

The method of weighted particles was proposed and investigated theoretically in [19]. The idea is to generate M independent Markov processes $\{X_n^{(k)}\}_{n\leq 0, k=1\dots M}$, where n is, as before, the time and k denotes the k'th member of the ensemble. All copies have the same statistics as the original signal process $\{X_n\}$, i.e. the same initial distribution and the same transition pdf. Let $f : \mathbb{R}^d \to \mathbb{R}$ be an arbitrary function. As already mentioned before, the method provides an approximation to the quantity

$$E(f(X_n)|Y_1,\dots,Y_n) = \int f(x)\cdot\pi_n(x)\,\mathrm{d}x$$

by a weighted average over the ensemble points $\{X_n^{(k)}\}_{k=1\dots M}$ for the fixed time n, i.e.

$$\int f(x)\cdot\pi_n(x)\,\mathrm{d}x \cong \sum_{k=1}^{M} w_n^{(k)}\cdot f(X_n^{(k)}).$$

It only remains to give an expression for $w_n^{(k)}$. Define the quantities

$$g_j^{(k)} := g(Y_j - h(X_j^{(k)})) \qquad \text{for all } j \leq 0, k = 1,\dots,M.$$

Theoretically one can prove that

$$w_n^{(k)} = c\cdot \prod_{j=1}^{n} g_j^{(k)},$$

where c is a constant chosen to yield

$$\sum_k w_n^{(k)} = 1.$$

A profound analysis of the problem however shows that this method tends to diverge, and one should rather implement a *limited memory version* of the filter, where the memory depends on the ensemble size M. This is done as follows: Let q_M be a certain positive integer depending on the ensemble size M. Then define the weights to be

$$w_n^{(k)} = c\cdot \prod_{j=n-q_M}^{n} g_j^{(k)},$$

where we define $g_j^{(k)} := 1$ if j is negative or zero. Practically this method can be implemented as follows: Choose M and q_M = integer closest to $2\sqrt{\log(M)}$.

Initial condition: Let $X_0^{(1)},\dots,X_0^{(M)}$ be independent samples of the pdf p_{X_0}. For each $k = 1,\dots,M$ allocate a *weight vector*

$$g^{(k)} := [g_1^{(k)},\dots,g_{q_M}^{(k)}]$$

and set all entries equal to one.

From n to $n+1$: Assume the ensemble $X_n^{(1)}, \ldots, X_n^{(M)}$ and the weight vectors $g^{(1)}, \ldots, g^{(M)}$ for time instant n are given. For each $k = 1, \ldots, M$ let

1. $X_{n+1}^{(k)}$ be a sample point of the pdf $\varphi(\cdot, X_n^{(k)})$,
2. $\bar{g}_j^{(k)} = g_{j+1}^{(k)}$ for $j = 1, \ldots, q_M - 1$,
3. $\bar{g}_{q_M}^{(k)} = g(Y_{n+1} - h(X_{n+1}^{(k)}))$.

Then $X_{n+1}^{(1)}, \ldots, X_{n+1}^{(M)}$ is the new ensemble and $\bar{g}^{(1)}, \ldots, \bar{g}^{(M)}$ the new weight vectors at time $n+1$, which are renamed $g^{(1)}, \ldots, g^{(M)}$ for the next time step.

For any function f, the conditional expectation then is approximately given by

$$E(f(X_n)|Y_1 \ldots Y_n) \cong \frac{\sum_{k=1}^M \prod_{j=1}^{q_M} g_j^{(k)} \cdot f(X_n^{(k)})}{\sum_{k=1}^M \prod_{j=1}^{q_M} g_j^{(k)}}.$$

6.1.2 Evolutionary Particle Method

The evolutionary particle method uses an ensemble of particles generated by a sampling procedure resembling the prediction/update mechanism of equation (2.4) which renders the ensemble a particle approximation of π_n itself. Let again $X_n^{(1)}, \ldots, X_n^{(M)}$ denote the ensemble at time n. Fix another positive integer M_C called the *number of children*. The dynamics of the particle ensemble is given by the following evolutionary process:

Initialisation: Let $X_0^{(1)}, \ldots, X_0^{(M)}$ be independent samples of the pdf p_{X_0}.

Prediction step: For each $k = 1, \ldots, M$ fixed produce an ensemble $X_{n+1}^{(k,1)}, \ldots, X_{n+1}^{(k,M_C)}$ of M_C *children* of $X_n^{(k)}$ by sampling M_C times independently from the pdf $\varphi(\cdot, X_n^{(k)})$.

Update step: To each $X_{n+1}^{(k,l)}$ assign the probability

$$p_{k,l} := \frac{g(Y_{n+1} - h(X_{n+1}^{(k,l)}))}{\sum_{k,l}[\text{numerator}]}$$

and select M of the children $X_{n+1}^{(k,l)}$ each with probability $p_{k,l}$. This gives the new ensemble $X_{n+1}^{(1)}, \dots, X_{n+1}^{(M)}$.

For any function f, the conditional expectation then is approximately given by

$$E(f(X_n)|Y_1 \dots Y_n) \cong \frac{1}{M} \sum_{k=1}^{M} f(X_n^{(k)}).$$

As far as we know, the Evolutionary Particle Method remains to be investigated numerically as well as theoretically. Hopefully, also large deviation results like in [19] can be obtained.

6.2 Estimation of unknown parameters

So far we assumed the measurement function h and the dynamical system (or equivalently the transition Markov kernel) to be known. Of course, the presented approaches cannot be applied directly if the underlying equations are not known. Therefore, in this and the next section we shall briefly discuss different approaches for filtering with limited pre–knowledge.

All methods assuming no or little knowledge of the signal process need a certain idea of what kind the signal process should be. Basically it is implicitely or explicitely assumed that the underlying signal belongs to a certain class $\mathcal{D}$. Then the method finds the member of $\mathcal{D}$ that is "at most in accordance" with the given data with respect to a certain measure of accordance.

If a model for the data is available (derived from first principles, for example) and only some parameters are unknown, then the filtering task can be incorporated into the framework of this thesis. The idea is to specify the model up to an unknown parameter vector α and to treat this α as a further state vector to be reconstructed. So suppose the dynamical system has the following form:

$$X_{n+1} = F(X_n, \alpha) + R_{n+1}.$$

Now introducing a further state equation of the form

$$\alpha_{n+1} = \alpha_n,$$

the unknown parameter can be considered as a part of the unknown state and is recovered by the usual filtering process. More general, if X_n is a Markov process with transition kernel $\varphi(A; x, \alpha)$ for a fixed but unknown α, then also the compound process (X_n, α_n) is a Markov process with transition kernel

$$P(X_{n+1} \in A, \alpha_{n+1} \in B | X_n = x, \alpha_n = \alpha) = \varphi(A; x, \alpha_n) \cdot \delta_B(\alpha).$$

Recall from the discussion about general probability theory that the best estimator of the parameter in a mean square sense is the conditional expectation $E(\alpha | Y_1 \ldots Y_n)$. Using the filtering approach mentioned above, we can calculate $E(X_n, \alpha_n | Y_1 \ldots Y_n)$, if the optimal filter was possible to use. But due to the underlying dynamics of α, namely $\alpha_{n+1} = \alpha_n$, this expectation is equal to $E(X_n, \alpha_0 | Y_1 \ldots Y_n)$, and the last component of this vector is the desired quantity. Therefore the filtering approach to parameter estimation yields in principle the optimal estimator.

Usually the optimal filter is impossible to realize, and this of course remains true if a further parameter in the dynamics is unknown. However, using the approximation schemes presented in this thesis, reliable parameter estimation is still possible. We have already presented the Lorenz system (numerical Example 31), where the three parameters were assumed to be unknown. Although the estimate was quite good, it turned out that still after a long time a small bias remained. However, the exact conditional expectation should be bias free in *average over many realisations.* So the small deviation from the correct values could be either due to the approximation we used or the fact that the system parameters are not uniquely defined by the outputs, i.e. even an infinitely long series $Y_1, Y_2, \ldots$ of measurements does not determine the parameters uniquely. We will show a further example where this seems to be the case as well.

36 Numerical example (Hindmarsh–Rose system) For the activity of biological neurons many models have been proposed. Recently nonlinear deterministic dynamical systems received an increasing attention, although dynamical equations describing the action potential activity where proposed already by Hodgkin and Huxley in 1952 [34]. As simplified three dimensional neuron model that nevertheless is capable of reproducing many essential features of the neuron dynamics was proposed by Hindmarsh and Rose [33]. Although the model is phenomenologically in the sense that it is not derived from first biological principles but merely by trying to reproduce

	a_0	a_1	a_2	a_3	b_0	b_1	b_2	c_0	c_1
model	-2.19	0	3	-1	-1	0	5	-0.001	0.004
estimate	9.2795	-0.2793	3.1196	-1.1580	0.8189	3.0950	-0.3338	0.0125	-0.0001

Table 6.1: Estimated and true parameters for the Hindmarsh–Rose model

the dynamic behaviour by means of simple components, it has already been shown to produce results consistend with action potentials of real neurons. This, however, requires a careful tuning of the free parameters the model has. Thus, for systematic construction of a Hindmarsh–Rose model that simulates a given real neuron, systematic parameter estimation is required, and unfortunately this turns out to be a difficult problem.

A general three dimensional Hindmarsh–Rose model looks as follows:

$$\begin{aligned}
\dot{v} &= a_0 + a_1 v + a_2 v^2 + a_3 v^3 - x + y, \\
\dot{x} &= b_0 x + b_1 v + b_2 v^2, \\
\dot{y} &= c_0 y + c_1 v.
\end{aligned}$$

The problem now is to estimate the parameters $a_0 \ldots a_3$, $b_0 \ldots b_2$ and c_0, c_1 from a sample of the first component $v_{t_1}, v_{t_2}, \ldots$ probably corrupted by noise. First results to achieve this goal where obtained in [71]. The results of our approach using Gaussian exponential families is shown in Figure 6.1. The left column panels show the output of a Hindmarsh–Rose model with parameters as in Table 6.1, first row. The error between filter output and the Hindmarsh–Rose model is shown in the right colum. As can be seen, the error in the second component is already significant and in the third component it is quite large. Furthermore, some of the estimated parameters (Table 6.1, second row) are very different from the true parameters. A free run of the Hindmarsh-Rose model using the parameters estimatied by the filter is shown in Figure 6.2. It turns out that the model running with the estimated parameters gives (at least on medium time scales) qualitatively the same output behaviour as the original model. E.g. the frequence of spikes

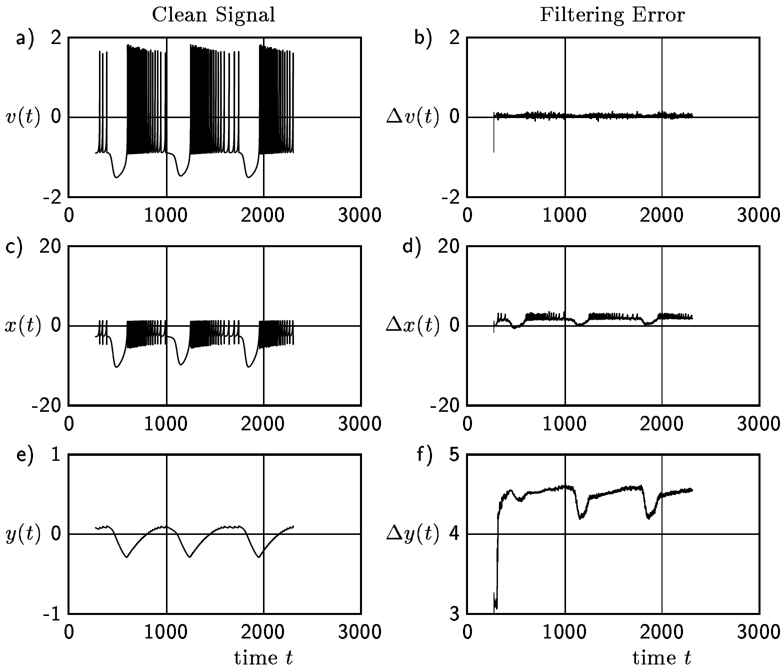


Figure 6.1: The results of our approach using Gaussian exponential families. The left column panels show the output of a Hindmarsh–Rose model with parameters as in Table 6.1, first row. The error between filter output and the Hindmarsh–Rose model is shown in the right colum. The error in the second component is already significant and in the third component it is quite large.

in each spike train is too small compared with the original data. Thus, it seems that some of the parameters have in fact a very weak influence on the output, at least on short time scales. It may be doubted whether *any* parameter estimation method is able to overcome this problem and to recover the true parameters. In fact, other methods seem to suffer from exactly the

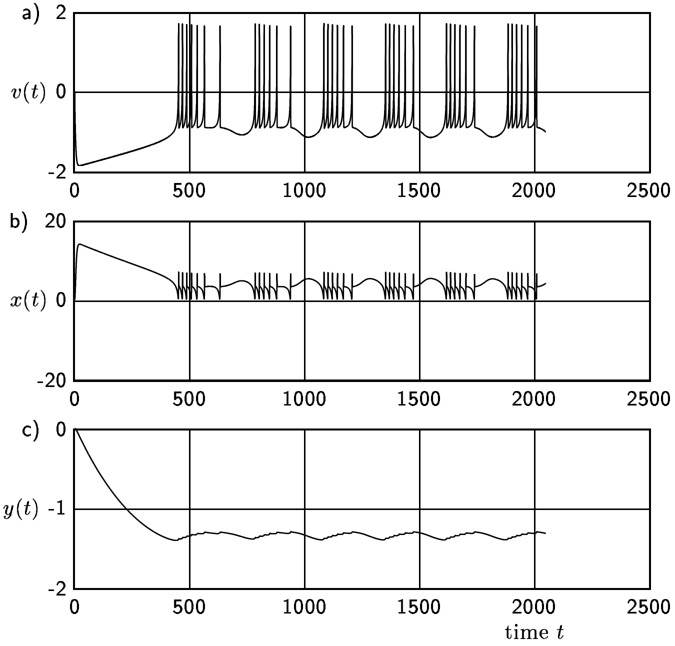


Figure 6.2: A free run of the Hindmarsh-Rose model using the parameters estimatied by the filter. The model running with the estimated parameters gives (at least on medium time scales) qualitatively the same output behaviour as the original model. The frequence of spikes in each spike train is too small compared with the original data.

same problems (see [71]), so our observation seems to be a general problem. In a certain sense, the filter does its best to give us what we deserve: A model that is able to reproduce the data. So probably it is necessary to think about the model again. A model should either be motivated by physical (or biological) considerations, whence the parameters can be identified with meaningful physical quantities (but may be hard to estimate) or the model should be chosen in order to allow for an easy parameter estimation. A model that is just phenomenological and has *furthermore* the problem of being hard to cope with is probably not worth to struggle with. This ends the discussion of the example.

The method of parameter estimation considered here is by far not the only one possible for parameter estimation for Markov processes. Let us mention three further approaches. First, one may ask whether a direct calculation of the probability $P(\alpha|Y_1, \ldots, Y_n)$ is possible. An expression for it can in fact sometimes be given (if the dynamical system is deterministic and the noise appears only in the observation), but it turns out to feature an enormous complexity, rendering a calculation of the expectation by direct quadrature an impossible task in practically all cases. In [18], however, an interesting Monte Carlo approach to this problem was suggested. The difficulty is of course to sample from the very complex conditional probability. This is a general problem often encountered also in other fields where Monte Carlo methods are applied, for example thermodynamics. Therefore sophisticated methods have been conceived to overcome this problem, namely the Markov–Chain–Monte–Carlo or Metropolis–Hastings method. For a general introduction into these methods see e.g. [31]. To our knowledge, parameter estimation using this techniques has been investigated only by [18], and we think that this approach merits further research.

A further quite appealing method was presented in a paper by Bibby and Soerensen in [5]. They proposed estimators which conserve the *martingale* property of the conditional expectation. A stochastic process $\hat{\alpha}_n$ is a martingale with respect to the observations Y_n if, for $k \leq n$,

$$E(\hat{\alpha}_n|Y_1, \ldots, Y_k) = \hat{\alpha}_k.$$

In connection with estimators this means that, if only measurements up to time k are available, the expected future value of the estimator should be the present value. The optimal estimator $E(\alpha|Y_1, \ldots, Y_k)$ is obviously a martingale. The authors of [5] now consider estimators which are suboptimal, but conserve that property. It should be mentioned that in this work continuous time systems are considered only. Stochastic analysis is extensively used here. Furthermore, the approach works only if the observations are of the form $Y_t = X_t + \text{noise}$, so a full state information is necessary.

In the case of an observed dynamical system in a state space with only a finite number of states, the whole model can be described by a finite set of parameters. Such models are known as *Hidden Markov Models*. Identification of Hidden Markov Models has been subject to vivid research. A general reference for this topic is [23]. For a problem with a continuous state space however, identification of the dynamics can be considered as estimating an infinite dimensional parameter. A small account on approaches for this problem is given in the next section.

6.3 Approaches for unknown dynamics

If no dynamical equations are known for the data, some modeling has to be done in combination with or previous to filtering. In the common literature on nonlinear time series analysis[1] this modeling is usually done assuming the dynamics to be esssentially deterministic. For deterministic dynamical systems state space reconstruction from scalar time series is possible using *delay coordinates* [54, 70, 61, 60]. The resulting states $Z_n = (Y_n, Y_{n-1}, \ldots, Y_{n-d})$ consist of sequences of measurement values and provide a faithful representation of the underlying dynamics if the dimension d is large enough.

Having successfully embedded the underlying dynamical system in a reconstructed state space one may approximate the induced flow using *globally* or *locally* defined models. Global modelling using a superposition of radial basis functions has, for example, be used in Ref. [35] for subsequent smoothing.

Most smoothing methods based on delay reconstruction, however, use *local* approximations of the flow in reconstruction space [63, 59, 24], or of some (sub–)manifold containing the reconstructed states [16, 15], because local approximations are very efficient and flexible (and suffer not from a possible poor choice of basis functions). Well written reviews comparing different implementations of the basic idea may be found in Refs. [44, 42, 30]. Such methods have been applied succesfully for denoising speech signals [32] or EEG data [64] and in studies on chaos based communication schemes [25]. However, they have to applied very carefully to avoid artefacts and misinterpretations [50]. When applied iteratively (what is typically done)

[1] Usually in nonlinear time series analysis, smoothing (often referred to as "Noise Reduction" in these contexts) is considered rather than filtering.

most of the algorithms first improve the SNR but start to destroy/scramble the data completely when iterations are continued.

Finally we want to mention some recent approaches for noise reduction exploiting the existence of (unstable) periodic orbits in chaotic systems [73] and the availability of some simultaneously measured reference data set [68].

Chapter 7

Communication

7.1 Introduction

Since the invention of telecommunication its technical aspects have been subject to vivid research. Usually the telecommunication engineers goal is to quantify and to optimally payoff between the demands of low cost, low error and high rate of information transfer. Of course, to obtain nontrivial results certain restrictions on the given setup have to be imposed.

A new area of information theory was heralded by the pioneering works of C. E. Shannon and W. Weaver [66] and N. Wiener [75]. The book of Shannon and Weaver contains the basic ideas and results on channel (and source) coding. Wiener's work adresses the problem of reconstructing a stationary time series that was received in error due to corruption by noise. Although the aim of both works is to combat a nonreliable transmission channel, the respective setups and assumptions are quite different in detail. While Wiener solves his problem by salient handling of elaborated stochastic tools, Shannon applied elementary methods and a couple of completely new and ingenious ideas.

We will briefly review both concepts now. The main reason is that the reader may have the (completely justified) question, how the presented results are related to Shannon's theory. Probably to his or her disappointment, however, it will turn out that this chapter, although concerned with the transmission of messages, is more in spirit of Wiener's work. The remainder of the introduction is intended to explain why and how the connection to nonlinear filtering emerges.

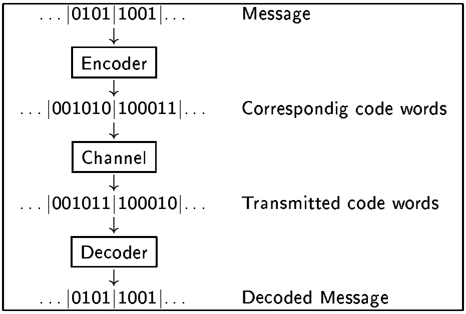
In *channel coding theory* the problem of information transmission over a not fully reliable channel is considered, i.e. it is assumed that with a certain probability the transmitted message is decoded in error. It is pretty obvious that some errors can be corrected at the receiver's side if a certain amount of the transmitted message is redundant. E.g. every bit (assuming the message is represented as a stream of zeros and ones) can be send twice. The surprising result of Shannon and Weaver was that a fixed amount of redundancy is sufficient to achieve an *arbitrarily small* amount of errors. The basic idea is to use a *code* as follows (see also Figure 7.1). Consider all possible words of, say, N bits. There are 2^N such words. Let $R \leq 1$ and specify a subset containing only $2^{\lfloor RN \rfloor}$ words. Here $\lfloor \cdot \rfloor$ means the integer part. This subset is called a *code of rate* R. The elements of this set are called code words, hence there are $2^{\lfloor RN \rfloor}$ code words. We can transmit a message using this code by simply dividing the message into blocks of length $\lfloor RN \rfloor$ (at most $2^{\lfloor RN \rfloor}$ different blocks can appear) and assigning a code word to each such block. Now the code word can be sent through the channel. Recall that the code word has length N, but the message block that is assigned to the code word has length $\lfloor RN \rfloor$, only. So using the code effectively reduces the transmission rate by a factor of R. In Figure 7.1 we used $N = 6$ and $R = 2/3$, i.e. $1/3$ of the bits are redundant.

If a code word is transmitted, at the receiver's side a word of N bits obtains. However, some of the N bits are received in error (in Figure 7.1 the last bits of both code words are incorrect). So a received block of N bits forms a word that is usually *not* a code word (although this may accidentally be the case). Here in general a *decoder* is needed that maps any word of length N back onto a code word. For example, we may take the code word that has the smallest amount of bits different from the received word (minimum Hamming distance). Finally, inverting the message–code assignment, we get back what is supposed to be the transmitted message.

For a given channel, the performance of this scheme obviously depends on the rate R, the length N, the chosen set of codewords and the decoder. The outstanding theorem of Shannon states that associated to the channel there is a number C called the *capacity* with the following property: By taking N sufficiently large we can find a code of rate R arbitrarily close to C and a decoder yielding arbitrarily small transmission error. This is called the direct part of the coding theorem. If R is larger than C, the error is bounded away from zero. This statement is called the converse part.

All Words of Length 4	code words of Length 6
0000	010001
0001	000011
0010	010100
0011	000111
0100	011001
0101	001010
0110	011101
0111	001111
1000	110000
1001	100011
1010	110101
1011	100110
1100	111001
1101	101011
1110	111100
1111	101111

(a) Codetable



(b) Communication Scheme

Figure 7.1: The panel 7.1(a) shows all words of length 4 (first column). The second column contains possible code words of length 6. Panel 7.1(b) shows a communication setup using the code.

Actually, Shannon and Weaver proved this result (together with an explicit expression for C) in the case of memoryless channels, i.e. the probability that a transmission error occurs at time n does not depend on what has happened in the past.

Note that in a practical situation to establish a reliable communication with rates close to the channel capacity it is necessary to manipulate the information carrying quantity *before and after* it is sent. The situation considered by N. Wiener that will now be described briefly does not permit a manipulation of the signal before transmission, which from a communication theoretic viewpoint constitutes the main difference between Wiener's and Shannon's setup.

As in Shannon and Weavers work, Wiener considers a stationary stochastic process $X_n, n \in \mathbb{Z}$ as the quantity carrying the desired information. In contrast to Shannon however, it is not explicitly attributed to as a message. He assumes that at the receiver the process $Y_n = X_n + S_n$ obtains, where S_n is the unwanted part or the noise. Hence effectively he assumes a very specific form of a channel.

Wiener now considers the problem of reconstructing X_n from Y_n in a linear manner. More specifically, taking the ansatz

$$\hat{X}_n := \sum_{k=-\infty}^{\infty} a_k^{(n)} Y_k \tag{7.1}$$

and the least mean square optimality criterion

$$E(X_n - \hat{X}_n)^2 = \min!$$

he obtains an equation (Wiener–Hopf equation) for the coefficients, $a_k^{(n)}$, where $E(\cdot)$ denotes the mathematical expectation. It turns out that only auto and cross correlations of X_n and Y_n enter the Wiener–Hopf equation. Wiener studies a variety of related problems, mainly differing in how much Y's enter the right hand side of the ansatz (7.1). The resulting Wiener–Hopf equations are tackled by spectral methods.

Wiener finished his work already in 1942, but due to its significance for war time issues (radar tracking, automatic fire control) it was classified and published not until 1950. Since then, a huge amount of improvements and generalisations to Wiener's theory have been conceived. The most important where probably the Kalman filter [40], where Gaussian processes admitting a state space description where considered and the extension to nonlinear stochastic differential equations given by Stratonovich and independently by Kushner (see [37] for an overwiev).

Basically, the theory of nonlinear filtering can be seen as a generalisation of Wiener's work. Nonlinear filtering is an attempt to the solution of the general problem: Which was the signal that led to the aquired data? Looking back to Shannon's setup, we see that this is basically the decoding problem: which was the codeword that led to the received data? Actually, Shannon's result can be established using a quite suboptimal receiver. Nevertheless, the trade–off between N and the error depends heavily on the decoder, which is important in practical applications. Furthermore, Shannon's result is valid in full generality only for memoryless channels. For channels having memory, the problem turns out to be quite difficult. In general, a different C appears in the direct and the converse part of the coding theorem, i.e. it is stated that at rates $< C_1$, reliable communication is definitely possible and at rates $> C_2$ definitely not, but in general $C_1 < C_2$

[29]. Furthermore, the results usually depend heavily on the employed decoders. In general, to obtain a larger C in the direct coding theorem, more sophisticated decoders are necessary, probably having a complexity prohibiting their practical implementation.

Thus for practical application and extension of Shannon's theorem, good decoders are mandatory. By good we mean as reliable as necessary to obtain the direct coding theorem at high rates, but as simple as possible to be implementable in applications. Of course, in this thesis we will not solve the problem completely. The basic aim of this chapter is to convince the reader that a possible route to good decoders goes via the theory of filtering.

7.2 Message transmission

In general message transmission is done employing a (usually electronic) device called the transmitter. The internal state of the transmitter at time instant $n \in \mathbb{N}$ is assumed to be determined by a variable X_n in an appropriate space. The state X_n depends on its predecessors $X_1 \dots X_{n-1}$, the message to be transmitted and some additional random influences. In this chapter we will only allow for the simplest possible messages, namely a sequence $\{M_n\}$ of independent, identically distributed random variables assuming the values 0 or 1, only. Furthermore, we assume M_n, the message element at time n, to be independent of $X_1 \dots X_{n-1}$, whence the message element has influence only on the present and future evolution of the transmitter state.

Based on this general considerations a lot of transmitter models can be considered differing basically in how much past information enters the future evolution of X_n. The simplest model of interest for a channel with memory obtains if we assume that X_n is, up to random disturbances, determined by M_n and X_{n-1}.

Usually a transmitter is necessary to generate a signal that is capable of passing through a channel. For example, consider a radio transmitter. The channel here is the atmosphere and the signal transmitted by the channel is the voltage at the antenna, which is a function of the transmitter state. Of course, atmospheric disturbances will take place and lead to a corruption of the transmitted signal. In our model channel noise is taken into account by additive iid random variables. Thus our model of a transmission channel is again a very simple one, namely we assume that the channel output Y_n is a function of the transmitter state corrupted by additive noise. As a simple example let us consider the following stochastic process on the unit interval

$$X_{n+1} = f_{M_{n+1}}(X_n), \tag{7.2}$$

where

$$\begin{aligned} f_0 &: [0,1] \to [0,1], \quad x \to |2x-1|, \\ f_1 &: [0,1] \to [0,1], \quad x \to 1-|2x-1|, \end{aligned} \tag{7.3}$$

are the usual and the inverted tent map. As received signal we take simply X_n itself. As random noise due to channel disturbances we take random variables $\{W_n\}$ which are independent, have a centered normal distribution with unit variance and are independent of $\{X_n\}$. The signal arriving at the receiver is assumed to be

$$Y_n = X_n + \sigma W_n,$$

where σ is a given positive constant. The basic question, the *receiver problem*, now is:

> Assume a sample of values $Y_1, ..., Y_n$ has been recorded. What is the value of the message M_n ?

It will turn out that this problem can be encompassed by calculating the conditional probability of X_n given $Y_1, ..., Y_n$. This in turn is the main aim of filtering theory. How it can be employed to solve the receiver problem will be explained now.

Let us first formalize our basic model of a transmitter. Let $(\Omega, P, \mathcal{A})$ be a probability space. Let E be a complete separable metric space and $\{X_n\}_{n\in\mathbb{N}_0} : \Omega \to E$ (the transmitter state) as well as $\{M_n\}_{n\in\mathbb{N}} : \Omega \to \{0,1\}$ (the message) be random processes. Furthermore we assume that the joint process $\{M_{n+1}, X_n\}_{n\in\mathbb{N}_0}$ is Markov, the variables $\{M_n\}$ are all identically distributed and M_{n+1} is independent of $\{M_{k+1}, X_k\}_{k=0...n-1}$. In practice this is accomplished by means of compression. Let $\mu(A) := P(X_0 \in A)$ and $p_i := P(M_n = i)$, where $i = 0$ or 1. Then the initial distribution of the process $\{M_{n+1}, X_n\}$ is given by $P(X_0 \in A, M_1 = i) = \mu(A)p_i$ and the transition probability is

$$P(X_n \in A, M_{n+1} = i | X_{n-1} = x, M_n = j) = p_i \cdot \varphi_j(A, x),$$

where we define

$$\varphi_j(A, x) := P(X_n \in A | X_{n-1} = x, M_n = j). \tag{7.4}$$

It is easy to see that $\{X_n\}$ alone is a Markov process with transition probability $\varphi(A, x) := P(X_n \in A | X_{n-1} = x) = \sum_j \varphi_j(A, x) \cdot p_j$.

The channel in this context is nothing more than the usual observations, i.e. the information obtained at the receiver is the process Y_n given by

$$Y_n = h(X_n) + \sigma \cdot W_n.$$

7.3 The optimal receiver

The *receiver* is any device that produces a reasonable estimate $\hat{M}_n$ for the actual message M_n based on the time series $Y_1, \ldots, Y_n$. We will show that this problem can be solved if the conditional probability $\rho_n(m) := P(M_n = m | \mathcal{G}_n)$ is known.

We now give an expression for $\rho_n(m)$ in terms of the filtering process. This establishes the beforementioned condition between the receiver problem and the theory of nonlinear filtering.

37 Lemma *Let*

$$\varphi = p_1 \varphi_1 + p_0 \varphi_0$$

be the transition kernel defined as before. Then the conditional probability of receiving zero ($m = 0$) or one ($m = 1$) at time instant n given the observations $Y_1 \ldots Y_n$ is given by

$$\rho_n^\nu(m) = p_m \int \frac{\mathrm{d}\varphi_m \pi_{n-1}^\nu}{\mathrm{d}\varphi \pi_{n-1}^\nu} \pi_n^\nu(\mathrm{d}x),$$

where the superscript (as in Section 2.2) indicates the initial probability distribution.

Proof This follows easily using change of measure like in the Kallianpur–Striebel formula. An informal derivation is given in [12]. □

38 Remark (Concerning notation) For simplicity of notation we write from now on φ_j instead of $\varphi_j \cdot p_j$ for both $j = 0$ and $j = 1$. So formally we (re)define

$$\varphi_j(A, x) = P(X_n \in A, M_{n+1} = i | X_{n-1} = x, M_n = j).$$

The performance of a binary communication channel is usually measured by the Bit Error Rate (BER), which is defined as

$$\mathrm{BER} = \frac{1}{N} \sum_{k=1}^{N} |M_k - \hat{M}_k|,$$

where M_k is the transmitted message and $\hat{M}_k$ is the received message. It should be kept in mind that in general the bit error rate is a random quantity and depends on N. It is an interesting question whether the bit error rate converges to a (possibly random) limit or not. We will briefly consider this question now for $\hat{M}$ beeing the optimal receiver.

In any case (ergodic or stationary or nothing) we will call

$$P^{\nu}(M_k \neq \hat{M}_k)$$

the bit error probability (denoted by BEP_k^{ν}) where $\hat{M}_k$ is used as an estimator for M_k and ν is the distribution of X_0. We now define the receiver $\hat{M}_k$ we will use throughout the rest of this chapter.

39 Definition *We set $\hat{M}_k = 1$ if $\rho_k^{\nu}(1) > \rho_k^{\nu}(0)$ and $\hat{M}_k = 0$ else. Since in fact $\hat{M}_k$ depends on ν we will write $\hat{M}_k^{\nu}$ in the following.*

Obviously, $\hat{M}_k$ is a function of $Y_1 \dots Y_k$. Furthermore, this estimator turns out to have a certain minimum property. If $\bar{M}_k$ is an estimator depending on $Y_1 \dots Y_k$ and taking the values 0 or 1 only, it can be shown that

$$P^{\nu}(M_k = \bar{M}_k) = E_{\nu}(\rho_k^{\nu}(\bar{M}_k)),$$

whence we have that for any such estimator

$$P(M_k = \bar{M}_k) \leq P(M_k = \hat{M}_k^{\nu}).$$

Hence the estimator $\hat{M}_k^{\nu}$ yields the least bit error probability and, in this sense, provides an optimal estimator. Our first theorem concerning ergodicity of the bit error rate can be obtained using ergodic theory of nonlinear filtering.

40 Theorem *Suppose μ is a φ–invariant measure. Furthermore suppose that the compound process (M_k, π_k) is stationary. Then the bit error rate*

$$\mathrm{BER}_N = \frac{1}{N}\sum_{k=1}^{N} |M_k - \hat{M}_k^{\mu}|$$

converges almost surely to a (possibly random) limit. If μ is even a unique φ–invariant measure satisfying condition (A.8), then the limit is almost sure equal to a constant.

Proof If μ is φ–invariant it follows from Theorem 1 in [69] that the distribution of π_n^{μ} converges to an invariant measure of Π, the transition semigroup of the filter. Calling this invariant measure Φ it turns out that the joint random variable (M_{n+1}, π_n^{μ}) has asymptotic distribution $p_i{\cdot}\Phi$. Since $|M_k - \hat{M}_k|$ can be expressed as a function of M_k, π_k^{μ} and π_{k-1}^{μ} it turns out to be stationary as well. The first assertion now follows from Birkhoff's theorem. The second assertion follows if the filtering process turns out to be ergodic. Under condition (A.8), the invariant measure Φ of Π having barycenter μ is unique (see [69], Theorem 2). However, any other Π–invariant measure *must* have a barycenter which is φ–invariant. Since there are no such measures except for μ it turns out that Φ is the unique invariant measure of the filtering process. By Lemma 47, (3) the filtering process is ergodic. □

The stationarity of the compound process (M_{k+1}, π_k) will not be investigated here. However, we conjecture that the stationarity of the compound process $(M_k, \rho_k(m))$, which is needed in Theorem 40, can as well be deduced without further assumptions. A proof of this may proceed along the following lines: First, as discussed in Section 4.3, the stationary processes Y_n and X_n can be extended to $-\infty$ in time. It is then quite logical that the process $(M_n, P(M_n = m|\mathcal{G}_n))$ has asymptotically the same distribution as $(M_n, P(M_n = m|Y_n, n \in \mathbb{Z}_{\leq n}))$ which is stationary. Concerning the asymptotic properties of the bit error probability we have the following theorem

41 Theorem *If μ is an φ–invariant measure satisfying condition (A.8), then the BEP_k^{μ} is convergent and decreasing in k. Call the limit BEP^{μ}. If furthermore ν satisfies the assumption $\nu\varphi^k \to \mu$, then $\mathrm{BEP}_k^{\nu} \to \mathrm{BEP}^{\mu}$.*

PROOF This follows from the fact that the bit error probability BEP_k^ν can be written as

$$\begin{aligned}&\mathrm{BEP}_k^\nu\\ &= \frac{1}{2}E_\nu\left[1 - \int \Big| \int \frac{1}{\sigma} g(\frac{y-h(x)}{\sigma})(\varphi_1 - \varphi_0)\pi_{n-1}^\nu(\mathrm{d}x)\Big|\mathrm{d}y\right],\end{aligned}$$

which is an expectation over a concave function of π_{n-1}^ν. The theorem now follows from the results in [69]. □

We remark that the transmitter model introduced in Example 3 actually satisfies the conditions of Theorem 45 (see Appendix), hence there is a unique invariant measure satisfying the condition (A.8). Thus, both theorems apply.

Theorems 40 and 41 may be of restricted practical use since a quite restricted receiver model is assumed. However, the main purpose was to show that theoretical methods of nonlinear filtering translate into the framework of message transmission.

To tackle a message transmission problem numerically one has to calculate the conditional probability $\rho^\nu(m)$ replacing π_n by any of the approximations $\tilde{\pi}_n$ introduced in the preceeding sections. We will now give a numerical example.

42 Numerical example (Bell shaped map) Figure 7.2 shows in the first panel the Bit Error Rate that is achievable with approximation of the optimal filter for a chaotic signal contaminated by additive Gaussian noise. The model was as follows: The dynamical system constituting the transmitter is given by a bell shaped map

$$f_m(x) = \exp\left[-\left(\frac{x - B - 0.1{\cdot}m}{0.3}\right)^2\right],$$

For h we used the identity map, i.e.

$$Y_n = X_n + S_n.$$

We first set $B = 0.3$. In this case the map is chaotic for $m = 0$ as well as for $m = 1$. To approximate the optimal filter and the optimal receiver we applied a simple approximation on a grid, i.e. a linear family with piecewise

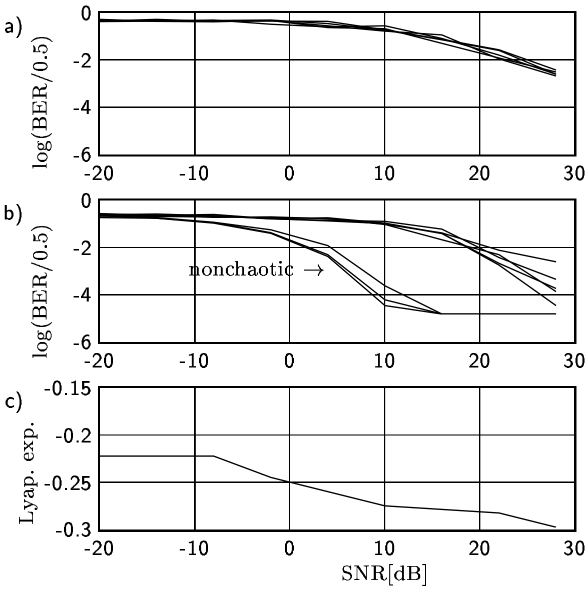


Figure 7.2: The panels show results for the receiver problem for the bell shaped map. The optimal receiver was approximated on a grid. The first panel a) shows BER/0.5 on a log-scale over SNR. The performance is more or less the same for different grid sizes $(100, 200, 400 \dots 6400$ points). The second panel b) shows BER/0.5 for different offsets B. In a nonchaotic regime the preformance is much better than in the chaotic. The third panel shows the estimated Lyapunov exponent of the filter for the TV–norm.

constant functions. It turns out that increasing the number of grid points does not significantly improve the resulting BER obtained for only 100 gridpoints (see first panel of Figure 7.2). We conjecture that this is due to a negative Lyapunov exponent of the nonlinear optimal filter associated with this system. An approximation of the Lyapunov exponent of the filter was calculated as well using a very fine mesh and is plotted on panel c). It turns out to be negative for all considered SNR values. Thus the “good news” is that a 100-point approximation already yields a reliable approximation of the optimal receiver. On the other hand we can conclude that *no* receiver will be able to outperform the (in fact a little disapointing) results plotted in Figure 7.2, first panel.

An interesting result is shown in the second panel b). Here we varied the value of B from 0.3 to 0.7 in steps of 0.05. For $B = 0.3 \ldots 0.5$ the transmitter is chaotic for both values of m, for $B = 0.55$ the map is chaotic only if bit 0 is sent and for larger values of B the map is periodic. The BER performance does not depend so much on the value of B but on the regime the map operates in. The curves of good BER performance correspond to nonchaotic, the other ones to chaotic behaviour. Thus we see that chaos strongly degrades the performance of this transmitter.

This numerical example is another example of a chaotic-shift-keying (CSK) scheme already mentioned in Section 2.2. As was already discussed there, our receiver model corresponds to a very wide transmission bandwidth setup. Several contributions in Refs. [43] and [65] consider setups with smaller bandwidth (i.e. M_n is kept constant for longer epochs), where it is assumed that the signal blocks with different values of M_n are independent. This however is only justified if the transmitter is a mixing system. Thus, in a certain sense, nonchaotic-shift-keying schemes cannot be investigated with this approach. We think that this is a reason why a comparison between chaotic and nonchaotic transmitters has not been carried out yet.

7.4 A Bound on the Bit Error Rate

We have seen how to build the optimal causal receiver using the nonlinear filtering process. Since the nonlinear filtering process cannot be calculated in general, we suggested several approximation schemes. In Section 4.3 we gave a bound on the error between the true and the approximative filtering process. In this section we show the implications of this result on the bit error rate obtained by receivers based on approximative filtering processes rather than the true one. Consider the function

$$f_n(y) := |\int \frac{1}{\sigma} g(\frac{y - h(x)}{\sigma})(\varphi_1 - \varphi_0)\pi^{\nu}_{n-1}(\mathrm{d}x)|.$$

From Section 7.2 we know that

$$\mathrm{BEP}_n^\nu = \frac{1}{2} E_\nu \left[1 - \int f_n(y)\mathrm{d}y \right].$$

we now have using the triangle inequality

$$\begin{aligned}
&|\int \frac{1}{\sigma} g(\frac{y-h(x)}{\sigma})(\varphi_1 - \varphi_0)\pi^\nu_{n-1}(\mathrm{d}x)| \\
&= |\int \frac{1}{\sigma} g(\frac{y-h(x)}{\sigma}) \\
&\qquad \times (\varphi_1 - \varphi_0)(\pi^\nu_{n-1} - \tilde{\pi}^\nu_{n-1} + \tilde{\pi}^\nu_{n-1})(\mathrm{d}x)| \qquad (7.5) \\
&\leq |\int \frac{1}{\sigma} g(\frac{y-h(x)}{\sigma})(\varphi_1 - \varphi_0)\tilde{\pi}^\nu_{n-1}(\mathrm{d}x)| \\
&\qquad + |\int \frac{1}{\sigma} g(\frac{y-h(x)}{\sigma})(\varphi_1 - \varphi_0)(\pi^\nu_{n-1} - \tilde{\pi}^\nu_{n-1})(\mathrm{d}x)|.
\end{aligned}$$

The second term can be bounded using the triangle inequality

$$\begin{aligned}
&|\int \frac{1}{\sigma} g(\frac{y-h(x)}{\sigma})(\varphi_1 - \varphi_0)(\pi^\nu_{n-1} - \tilde{\pi}^\nu_{n-1})(\mathrm{d}x)| \\
&\leq \int \frac{1}{\sigma} g(\frac{y-h(x)}{\sigma})|\varphi(\pi^\nu_{n-1} - \tilde{\pi}^\nu_{n-1})|(\mathrm{d}x).
\end{aligned}$$

The integrant is an integarble function of x and y so we can replace the second term in (7.5), integrate over y and reverse the order of integration in the second term to get

$$\begin{aligned}
&\int |\int \frac{1}{\sigma} g(\frac{y-h(x)}{\sigma})(\varphi_1 - \varphi_0)\pi^\nu_{n-1}(\mathrm{d}x)|\mathrm{d}y \\
&\leq \int |\int \frac{1}{\sigma} g(\frac{y-h(x)}{\sigma})(\varphi_1 - \varphi_0)\tilde{\pi}^\nu_{n-1}(\mathrm{d}x)|\mathrm{d}y \\
&\qquad + \mathsf{TV}(\varphi\pi^\nu_{n-1}, \varphi\tilde{\pi}^\nu_{n-1}) \\
&\leq \int |\int \frac{1}{\sigma} g(\frac{y-h(x)}{\sigma})(\varphi_1 - \varphi_0)\tilde{\pi}^\nu_{n-1}(\mathrm{d}x)|\mathrm{d}y \\
&\qquad + \mathsf{TV}(\pi^\nu_{n-1}, \tilde{\pi}^\nu_{n-1}),
\end{aligned}$$

since $\mathsf{TV}(\varphi\cdot, \varphi\cdot\cdot) \leq \mathsf{TV}(\cdot, \cdot\cdot)$ (Lemma 16). In exactly the same manner (exchanging the role of π and $\tilde{\pi}$) one obtains

$$\begin{aligned}
&\int |\int \frac{1}{\sigma} g(\frac{y-h(x)}{\sigma})(\varphi_1 - \varphi_0)\pi^\nu_{n-1}(\mathrm{d}x)|\mathrm{d}y \\
&\geq \int |\int \frac{1}{\sigma} g(\frac{y-h(x)}{\sigma})(\varphi_1 - \varphi_0)\tilde{\pi}^\nu_{n-1}(\mathrm{d}x)|\mathrm{d}y \\
&\quad - \mathsf{TV}(\pi^\nu_{n-1}, \tilde{\pi}^\nu_{n-1}).
\end{aligned}$$

If we define the quantity

$$\begin{aligned}
&\widetilde{\mathrm{BEP}}^\nu_k \\
&:= \frac{1}{2} E_\nu \left[1 - \int |\int \frac{1}{\sigma} g(\frac{y-h(x)}{\sigma})(\varphi_1 - \varphi_0)\tilde{\pi}^\nu_{n-1}(\mathrm{d}x)|\mathrm{d}y \right],
\end{aligned}$$

which is the same as BEP^ν_k but with π replaced by $\tilde{\pi}$ we can write our estimate as

$$|\mathrm{BEP}^\nu_k - \widetilde{\mathrm{BEP}}^\nu_k| \leq \frac{1}{2} E_\nu \mathsf{TV}(\pi^\nu_{n-1}, \tilde{\pi}^\nu_{n-1}).$$

We assume now the setup of Theorem (23) with a deterministic τ. Then we get

$$|\mathrm{BEP}^\nu_k - \widetilde{\mathrm{BEP}}^\nu_k| \leq C \sum_{k=0}^{n} \tau^{n-k} E_\nu(R_k).$$

So far this estimate is of restricted practical use since both the approximated bit error probability $\widetilde{\mathrm{BEP}}$ as well as the right hand side of the above estimate involve the *true* expectation E_ν.

Under the additional assumption that the compound process $\{Y_n, \theta_n\}$ is ergodic, then we can compute $E_\nu R_n$ and $\widetilde{\mathrm{BEP}}_n$ in an offline experiment, since R_n is a function of θ_{n-1} and Y_n, and furthermore $\widetilde{\mathrm{BEP}}_n$ is the expectation over a function depending on θ_{n-1} only. Then both $\widetilde{\mathrm{BEP}}_n$ and $E_\nu R_n$ are asymptotically equal to a constant depending on the system and the approximation algorithm and can be computed numerically by an empirical mean over a long realisation.

Note that we obtained the bound without actually using any direct bound on $|\rho_n(m) - \tilde{\rho}_n(m)|$, the error between the true and the approximated conditional probability. In order to do this, the TV norm turns out to be not so convenient. Using the Hilbert metric however one can, starting from the representation of $\rho_n(m)$ in Lemma 7.5, establish a bound on $|\rho_n(m) - \tilde{\rho}_n(m)|$. The main ingredients is the observation

$$\exp(-\mathsf{H}(\mu,\nu)) \leq \frac{\mathrm{d}\mu(x)}{\mathrm{d}\nu(x)} \leq \exp(\mathsf{H}(\mu,\nu))$$

for two comparable probability measures μ and ν. Thus, for any positive function f

$$\exp(-\mathsf{H}(\mu,\nu)) \int f \mathrm{d}\nu \leq \int f \mathrm{d}\mu \leq \exp(\mathsf{H}(\mu,\nu)) \int f \mathrm{d}\nu.$$

From this comes

$$\exp(-3\mathsf{H}(\pi_n,\tilde{\pi}_n))\tilde{\rho}_n(m) \leq \rho_n(m) \leq \tilde{\rho}_n(m)\exp(3\mathsf{H}(\pi_n,\tilde{\pi}_n)),$$

which gives the desired relation.

Conclusion

Concluding remarks

This thesis is devoted to a couple of aspects concerning the optimal nonlinear filter. We presented the optimal filter as a statistical estimation problem for dynamical systems. In particular, the objective is to reconstruct the current state of the system by means of noise corrupted observations in a causal manner.

The first chapter was concerned with standard textbook material on probability theory. We presented important statistical concepts such as measure spaces, measures, integration and the conditional expectation. These concepts are necessary for a rigorous formulation and treatment of the nonlinear filtering problem, which was done in Section 2.2. We emphasized the viewpoint of considering the optimal filter as a dynamical system on the space of measures on E, the state space of the underlying system.

The second chapter again presented known material, which is quite relevant in applications and was also intended to be a motivation of the subsequent chapters. The main point in the discussion was that filtering of nonlinear systems is, in general, an infinite dimensional problem. In other words, the filter, seen as a dynamical system, cannot be represented in a finite dimensional state space. Especially for chaotic deterministic systems the filter dynamics is always infinite dimensional. In applications, however, a finite dimensional representation is mandatory, thus approximations are essential for nonlinear filtering to be an applicable tool.

Investigating new and existing approximation schemes in a general framework was the issue of the next two chapters. Any approximation of the optimal filter can be (as the filter itself) considered as a dynamical system on the space of measures on E (denoted by $\mathcal{P}_E$). In order to quantify the quality of the approximation, a metric has to be introduced on $\mathcal{P}_E$. This was done in Section 4.2. Actually, for several reasons a couple of metrics where considered, all having advantages and drawbacks in connection with nonlinear filtering. The metrics are all used in the literature. Connections, bounds and properties of these metrics where partly taken from several publications, but partly are (to the best of our knowledge) new. In the next section an error bound for a wide class of approximation schemes was established. The "generic approximation scheme" was assumed to be a projection of the dynamical equation governing the nonlinear filter on a space of parametrized probability distributions. Roughly speaking, we assumed a finite dimensional set of probability distributions $\mathcal{Q}$ to be specified. By projection we mean that if the optimal filter dynamics is applied to a member of $\mathcal{Q}$, then the result is projected back onto $\mathcal{Q}$ by a minimum error criterion. Thus we get an approximative filter dynamics on $\mathcal{Q}$. The objective of Section 4.3 then was to establish a connection between the total error between the correct and the approximative optimal filter and the approximation residual measuring the error that is made by applying once the approximative filter dynamics rather than the true.

The error bound is established basically using only the triangle inequality. Similar techniques where used already by other authors in more special circumstances. The error bound is a sum over two multiplied terms depending only on the optimal filter and the approximation algorithm, respectively. It turns out that a basic property of the filter to yield small total approximation errors is a stability of the optimal filter dynamics, i.e. an insensitivity with respect to misspecified initial conditions. This stability can be quantified by a Lyapunov exponent. The existence of the Lyapunov exponent is guaranteed by Kingman's ergodic theorem. If the Lyapunov exponent is negative, then, roughly speaking, one step errors made by the approximation algorithm are damped out by the filter dynamics, leading to a bounded (or more general asymptotically stationary) filtering error. This statement was made rigorous for special so called mixing Markov processes. The Lyapunov exponent for optimal filters of mixing Markov processes was already investigated in the literature. We think, however, that the general approach and the representation of the total error in this way are new. We furthermore presented a generalisation of an interesting result of Atar and Zeitouni [3] where it is shown that low noise observations lead also to a negative Lyapunov exponent of the optimal filtering dynamics. Our analysis shows that this continues to be true for systems that can in turn be considered as a stochastic analogon to systems in observer canonical form known in deterministic control theory.

In Chapter 4 we presented a large variety of approximation methods, which are partly new. As parametrized sets of probability distributions we used only exponential and linear families. We gave to some extend exhaustive formulas of the emerging finite dimensional filters. Much more needs to be done here, which will probably require sophisticated methods from convex analysis. We would like to point out that for the case of continuous time dynamical systems with discrete time observations (which is a practically very relevant case) a satisfying error analysis could be established. The problem of how to incorporate the update step into the error analysis of the prediction step (which was carried out already in [48]) could be solved.

In Chapter 6 we considered further approximation schemes based on the Monte Carlo idea. These known results where included mainly for the sake of completeness and for comparison. Furthermore, in this chapter we presented the interesting application of the nonlinear filter to parameter estimation problems. The connection between filtering and parameter estimation is not difficult to see, but using approximations of the optimal filter it becomes an applicable tool for this important task. We investigated numerically two nonlinear differential equations with unknown parameters. It turned out that for the first, the Lorenz system, reliable parameter estimation was possible, although the results where not unbiased. For the second example, the Hindmarsh–Rose neuron model, parameter estimation turned out to be difficult. We however have strong indication that for this example any parameter estimation method will suffer from the same problems as ours.

The last chapter was devoted to problems in telecommunication. It was shown that for certain receiver models the optimal receiver can conveniently be represented by the optimal nonlinear filter. Thus this problem emerges as an interesting sub–problem of the theory of nonlinear filtering. We demonstrated that this leads to fruitful interesting theoretical results concerning the least possible bit error probability. Furthermore, by applying some algorithms proposed in the preceeding chapters we could establish approximatively optimal receivers. A bound on the bit error rate using the bounds on the filtering error was obtained as well.

The appendices concern known material on ergodic theory of Markov processes and the filtering process. The appendix on exponential families contains suggestions how to cope numerically with exponential families.

Outlook

Let us shortly mention the problems and questions that in our opinion merit further investigation, that we could not carry out due to lack of time or that we simply failed to solve.

First we think that the approaches mentioned in this thesis merit more numerical simulations to check how tight the error bounds are in practice. The bounds, nonetheless, have theoretical significance as well since they show that the better we carry out the one step approximation, the better the total error will be.

Concerning the general error bound, the central question is still open: Which requirements on the Lyapunov exponent of the filter and the approximation residuals have to be imposed in order to yield an asymptotically stationary error? It is clear that the largest Lyapunov exponent needs to be negative and the residuals need to be at least asymptotically stationary.

Lyapunov exponents for nonlinear filters have been investigated already, but a general approach seems not to be available. Interesting problems for future research would be to investigate, e.g., piecewise expanding Markov maps or other non–mixing models.

Furthermore many problems involving the approximation schemes are still to be investigated. We did not fully address the question for which setups the approximation algorithms are actually well defined. Especially for the exponential families in connection with the Hilbert and the Kullback–Leibler distance these are challenging problems. As already mentioned, more convex analysis seems to be necessary here.

How to cope with parameter uncertainties remains still a challenging task. We have seen that problems appear also for our parameter estimation approach when the parameters have only weak influence on the output. We think that before applying any parameter estimation algorithm it is necessary to investigate a model more theoretically and establish a kind of "canonical parameter estimation form" in which parameters with weak or no influence on the output can easily be recognized. Furthermore, we think that for parameter estimation problems it should be possible to establish a reduced nonlinear filter that circumvents the unneccessary estimation of all dynamic variables. It is a paradigm of estimation theory to estimate as much as necessary, but not more.

In communication, the connection to coding theory is interesting and, to our knowledge, completely open. Another interesting question is how does chaos actually affect the performance of the optimal receiver? Our results provide a framework to investigate this question. It was shown how to build approximately optimal receivers and how to calculate the error between these and the optimal receivers. Thus, by numerical simulations an estimate of the maximal achievable bit error rate for a given transmitter model can be established. This may be subject to future research. As a preliminary result it was shown that the bell shaped map shows much better performance in the nonchaotic than in the chaotic regime.

Acknowledgements

I would like to thank all people that helped to finish this work, either by financial or mental support, either advice, or simply good friendship. Special thanks are due to Prof. Dr. Ulrich Parlitz for his support, his ideas, his doubts and his omnipresence, even at uncommon daytimes. His widespread interest in time series analysis and nonlinear dynamics was most pertinent to my work.

Also the members of his group at the III. Physikalisches Institut, namely Karsten Peters, Alexander Hornstein, Immo Wedekind, Jörg Dittmar, David Engster and Gerrit Langer, merit many thanks for fruitful discussions as well as thought provoking questions and doubts. Having been a member of this community was always a pleasure to me.

Furthermore I gratefully acknowledge financial support from the Graduiertenkolleg "Strömungsinstabilitäten und Turbulenz" during three years.

The III. Physikalisches Institut was always a place I liked to work at, due to the community of nice people there that know how to work and know how to enjoy life, football and parties. Especially I would like to mention Reinhardt Geisler, Dagmar Krefting, Jakob Großer and Hironori Tokuno.

My dear friends Gisa Kirschman-Schröder, Dr. Jörg Wichard, Dr. Robert Mettin, Dr. Thomas Kurz, Dr. Fabian Evert, Kevin Bube, Dr. Martin Voss and Mario Kuduz from the Band of the Institute merit a great thank you for uncounted evenings of joy and beer and, sometimes, music as well.

During travels I had the pleasure to make the acquaintance of a couple of researchers, and I would like to acknowledge the fruitful discussions I had with Lucas Illing, Ljupco Kocarev, Isao Tokuda, Henry Abarbanel, Henning Voss and eespecially Henk Nijmeijer, who invited me to the Netherlands for a three month introductory course in mathematical control theory.

Further people that have been with the Institute and are now spread over the whole world that I want to mention for help and friendship are Dr. Christian Merkwirth, Dr. Lutz Junge, Dr. Stephan Luther and Dr. Claus-Dieter Ohl.

I am indebted to my parents for both mental and financial support, for a home that is a place where I always like to be and to relax, to my two sisters who are always willing to share sorrows and problems, and to my grandparents for their love. My father provided a crash course in convex analysis, and I want to thank him for this as well.

Finally I would like to thank Stephanie Baum, a person I definitely could not do without, for her continuous and unbroken affection and support concerning all aspects of life and work, despite her busy day.

Appendix

A.1 Ergodic Theory of Markov Processes

We recall some results about ergodic properties of Markov processes. We will keep the same notation as in the thesis, namely let

- E a polish (i.e. complete separable metric) space
- $\mathcal{B}_E$ the Borel field
- $\mathcal{P}_E$ the space of probability measures on E
- $C_b(E)$ the spaces of continuous bounded functions on E
- $\mathcal{B}(\mathcal{P}_E)$ the Borel field of $\mathcal{P}_E$ enowned with the weak topology

We write as usual

$$\int f(x)P(\mathrm{d}x)$$

for the integral of f over P.

43 Definition *A random process $\{X_n\}$ is* stationary *if, for any k and sets $A_j \in \mathcal{B}_E$ the probability $P(X_{n+1} \in A_1, \ldots, X_{n+k} \in A_k)$ does not depend on n, i.e. is invariant with respect to time shifts.*

44 Lemma *A Markov process is stationary iff the probability measure $\nu(A) := P(X_0 \in A)$ has the property*

$$\nu(A) = \int \varphi(A, x)\nu(\mathrm{d}x).$$

Such a measure is called invariant.

PROOF See [7] □

The question arises whether for a given transition kernel $\varphi(A,x)$ there is an invariant measure ν so that the canonical process on $(E^\infty, \mathcal{B}_E^\infty, P^\nu)$ is stationary. A fruitful idea is to consider iterates of the kernel: Define $\varphi^{(1)}(A,x) := \varphi(A,x)$ and iteratively

$$\varphi^{(n)}(A,x) := \int \varphi(A,z)\cdot\varphi^{(n-1)}(\mathrm{d}z,x).$$

The following theorem gives conditions under which the sequence $\varphi^{(n)}(A,x)$ generated by a Markov transition kernel converges to an invariant measure:

45 Theorem *Suppose there is a finite nonzero measure μ and a set $C \in E$ with $\mu(C) > 0$. Let $\varphi(x,z)$ be the density of $\varphi(A,z)$ with respect to μ. Suppose now that*

$$\varphi(x,z) \geq \delta \qquad \textit{for all } z \in E \textit{ and } x \in C. \tag{A.6}$$

then there is an invariant probability measure s absolutely continuous with respect to μ. Furthermore, there are constants $K \geq 0$ and $0 < \delta < 1$ independent of x with

$$\sup_{A\in\mathcal{B}_E} |\varphi^{(n)}(A,x) - s(A)| \leq K\delta^n. \tag{A.7}$$

PROOF The theorem is a slight modification of results presented in [21], chapter V,§5. □

A trivial verification shows that if $\varphi^{(n)}(\cdot,x)$ satisfies the conditions of Theorem 45, then we also have the property

$$\lim_{n\to\infty} \int |f\varphi^{(n)}(x) - s(f)| s(\mathrm{d}x) = 0 \qquad \forall f C_b(E). \tag{A.8}$$

Condition (A.8) (which is weaker than the result of Theorem 45) will prove to be essential for ergodic properties of the filtering process.

Starting with s as the initial distribution, the resulting probability on the probability space $(E^\infty, \mathcal{B}_E^\infty)$ is denoted by P^s, as for every probability measure ν on E the resulting probability on $(E^\infty, \mathcal{B}_E^\infty)$ is denoted by P^ν.

Stationary processes may or may not be *ergodic*. We recall the basic concepts of ergodic theory. Let $\{X_n\}_{n\in\mathbb{N}}$ be a stationary process. An event A is *invariant* if there is a fixed $B \in \mathcal{B}_\infty$ so that for any k, A can be represented as

$$A := \{\omega \in \Omega; (X_k, X_{k+1}, \ldots) \in B\}.$$

The invariant events form a σ–algebra denoted by $\mathcal{I}$. This is the basis for the following famous result:

46 Theorem (Birkhoff's ergodic theorem) *Let X_n be a stationary process, $E|X_1| < \infty$. Then the following limit holds a.s. and in L_1:*

$$\frac{1}{n}\sum_{k=1}^{n} X_k \to E(X_1|\mathcal{I}).$$

Proof See [7] □

If X_n are iid random variables, all invariant events have probability zero or one (Kolmogorov's zero–one law). Obviously, then $E(X_1|\mathcal{I}) = E(X_1)$, and Birkhoff's ergodic theorem translates into the strong law of large numbers. To generalize this, call a process *ergodic*, if all invariant events have probability zero or one. Obviously, a process is ergodic iff all random variables measurable with respect to $\mathcal{I}$ are a.s. constant. Hence, if $E|X_1| < \infty$ and the process is ergodic, $E(X_1|\mathcal{I}) = E(X_1)$ and Birkhoff's theorem gives

$$\frac{1}{n}\sum_{k=1}^{n} X_k \to E(X_1)$$

both a.s. and in L_1.

Obviously, conditions for ergodicity are quite essential:

47 Lemma 1. *Let $f : E^\infty \to \mathbb{R}$ be measurable. Then the process*

$$Y_n := f(X_n, X_{n+1}, \ldots)$$

is stationary (ergodic) if X_n is stationary (ergodic).

2. *A stationary process X_n is ergodic iff all random variables measurable with respect to $\mathcal{I}$ are a.s. constant.*

3. *If a process* X_n *admits a* unique *stationary measure it must be ergodic*

Back to Markov processes we have the following more special criteria

48 Lemma
1. *If* E *is compact, there are always invariant measures for* φ.
2. *If an invariant measure* ν *for* φ *is unique, then* P^ν *must be ergodic*
3. *Let* ν *be an invariant measure for* φ*. Then if any* $f \in L_1(E, \nu)$ *with the property*
$$f\varphi^{(n)}(x) = f(x)$$
is ν*–almost sure constant, then* P^ν *must be ergodic.*

A.2 Ergodic Theory of Filtering Processes

We will briefly review the basic results about invariant measures for filtering processes. The basic references are [47] for compact state spaces and [69] for noncompact but locally compact state spaces. We will keep the same notation as in the previous section. Let $\{X_n\}_{n\in\mathbb{N}_0}$ be a Markov process with transition probability $\varphi(A, x)$ and Feller transition semigroup, i.e. for any $f \in C_b(E)$ we have that $f\varphi^{(n)}(z) := \int f(x)\varphi^{(n)}(\mathrm{d}x, z) \in C_b(E)$.

For a measurable function $h : E \to \mathbb{R}$ and an iid process $\{W_n\}_{n\in\mathbb{N}}$ independend of $\{X\}$ define the measurement process

$$Y_n = h(X_n) + W_n.$$

Let $\mathcal{G}_n := \sigma(Y_1 \dots Y_n)$.

49 Definition *For any* $\nu \in \mathcal{P}_E$ *we define the* minimal *and* maximal filtering processes *given by*
$$\pi_n^\nu(f) = E_\nu(f(X_n)|\mathcal{G}_n)$$
and
$$\tilde{\pi}_n^\nu(f) = E_\nu(f(X_n)|\mathcal{G}_n, X_0)$$
respectively, for $f \in C_b(E)$*. By*
$$m_n^\nu(\Lambda) := P_\nu(\pi_n^\nu \in \Lambda)$$

and

$$M_n^\nu(\Lambda) := P_\nu(\tilde{\pi}_n^\nu \in \Lambda)$$

we denote the probabilities of the minimal and maximal filtering process, respectively. Finally, the transition semigroup of minimal and maximal filtering process is the same and is denoted by $\Pi(\Lambda, \mu)$.

It turns out that both the minimal and maximal filtering process are Markov processes on $\mathcal{P}_E$ with Feller semigroup, i.e. if $\mu_t \to \mu$ in the weak topology of $\mathcal{P}_E$, then $\Pi(\Lambda, \mu_t) \to \Pi(\Lambda, \mu)$ in the weak topology of $\mathcal{P}(\mathcal{P}_E)$.

50 Definition *Let* $\Phi \in \mathcal{P}(\mathcal{P}_E)$. *A measure* $\nu \in \mathcal{P}_E$ *is a* barycenter *of* Φ *if for every* $f \in C_b(E)$ *we have*

$$\nu(f) = \int \nu'(f)\Phi(\mathrm{d}\nu').$$

The central theorem on invariant measures of Π is the following:

51 Theorem *Assume* μ *is* φ*–invariant. Then* $m_n^\mu \to m^\mu$ *and* $M_n^\mu \to M^\mu$ *as* $n \to \infty$ *in the weak topology, where* m^μ *and* M^μ *are* Π*–invariant and have barycenter* μ. *Furthermore, if* Φ *is any other* Π*–invariant measure with barycenter* μ *we have*

$$m^\mu(F) \leq \Phi(F) \leq M^\mu(F)$$

for any $F \in C_c(\mathcal{P}_E)$.

A.3 Design of exponential families

Let $p(x,\theta)$ be the parametrisation of an exponential family with canonical statistics $c_i(x)$, $i = 1 \ldots k$ and carrier measure λ. Let q be an arbitrary measure with density $q(x)$ and define the quantities

$$\eta_i := \int c_i(x) \cdot q(x) \cdot \mathrm{d}\lambda, \qquad i = 1 \ldots k.$$

Then for the Kullback–Leibler distance $\mathsf{KL}(p,q)$ we have

$$\mathsf{KL}(p,q) = \int \log(\frac{\mathrm{d}q}{\mathrm{d}\lambda})\frac{\mathrm{d}q}{\mathrm{d}\lambda}\mathrm{d}\lambda - \left(\sum_{i=1}^{k} \theta_i \eta_i - \psi(\theta)\right).$$

So the minimisation of the Kullback–Leibler distance is equivalent to

$$\psi^*(\mu) := \sup_\theta [\sum_{i=1}^k \theta_i \eta_i - \psi(\theta)], \tag{A.9}$$

which is a Legendre transform of ψ. This appendix will be concerned with numerical methods to compute ψ as well as the Legendre transform, or more exactly the minimizing argument θ^*.

Suppose first that convenient expressions for ψ as well as its first and second derivatives, i.e. the expectation parameters η and their jacobian $\frac{\partial \eta}{\partial \theta}$ respectively, are available. To solve the maximisation problem (A.9), consider a scheme of the form

$$\theta^{(n+1)} = \theta^{(n)} + \delta_n \cdot \Delta_n,$$

where Δ_n is the Newton–direction

$$\Delta_n := \frac{\partial \eta}{\partial \theta}^{-1} [\eta - \eta(\theta^{(n)})],$$

and δ_n is a damping factor taken from the one dimensional maximisation problem

$$\delta_n := \arg\max_\delta \left[(\theta^{(n)} + \delta \Delta_n) \cdot \eta - \psi(\theta^{(n)} + \delta \Delta_n) \right].$$

So δ is chosen to maximize the problem not globally but along the Newton–direction. This scheme can be proved to converge globally. Locally the convergence is even quadratic, i.e. the number of valid digits doubles at every iteration step. There are further modifications simplifying both the computation of δ as well as the Newton–direction, which may be both very costly.

Note that the Newton algorithm in principle *cannot* fail. However, the problem is that all quantities needed for the Newton scheme are given only approximately, and we observed that computing ψ and its derivatives far from the usual range with necessary accuracy is not easy. Therefore, if the Newton scheme does not converge, the reason is often failure of the routines computing the quantities entering Newtons scheme. We will now turn to the problem of computing them for a certain class of families called *adapted.*

Design of exponential families

The exponential family we want to use for a state estimation problem is not given in general but has to be designed. Usually it is required that certain functions $c_i(x), i = 1 \dots k$ are among the canonical statistics. Is it possible to extend them to canonical statistics of an exponential family? Suppose the following growth condition

$$||c(x)|| \leq K(1 + |x|^s)$$

is fulfilled for a certain $s \geq 0$ and $K > 0$. Then setting

$$\xi(x) := \sum_{i=1}^{d} |x_i|^r$$

with $r > s$, we have that

$$p(x, \theta) = \exp(\sum_{i=1}^{k} \theta_i c_i(x) - \theta_{k+1}\xi(x) - \psi(\theta))$$

is an integrable exponential family with parameter space

$$\Theta := \{(\theta_1 \dots \theta_{k+1}) \in \mathbb{R}^{k+1}, \theta_{k+1} > 0\}$$

and $\lambda(x) = 1$. We can also set θ_{k+1} to a constant positive value a and define an exponential family

$$p(x, \theta) = \lambda(x) \exp(\sum_{i=1}^{k} \theta_i c_i(x) - \psi(\theta)),$$

with $\lambda(x) = \exp(-a\xi(x))$. We will call both exponential families *adapted* to the $c_i(x)$. Of course, our growth condition and function ξ is only one possible choice to get an exponential family with a prescribed set of canonical statistics. However, our ξ has the advantage to factorize.

The potential function for adapted exponential families

We now proceed to derive a power series expansion for the function $\varphi := \exp(\psi)$ for adapted exponential families. The notation φ will be adopted only in this appendix and is not to be confused with the transition pdf. The coefficients in the power series expansion of φ are closely related to higher moments of $p(x,\theta)$. For calculations involving higher order moments, the concept of multiindices is useful. A k–dimensional multiindex $\alpha = (\alpha_1,\dots,\alpha_k)$ is an element of $\mathbb{N}^k$. For a multiindex α use the following notations

$$\begin{aligned}
\alpha! &:= \alpha_1!\cdots\alpha_k!,\\
|\alpha| &:= \alpha_1+\ldots+\alpha_k,\\
x^\alpha &:= x_1^{\alpha_1}\cdots x_k^{\alpha_k},\\
\frac{\partial^{|\alpha|}}{\partial x^\alpha} &:= \frac{\partial^{|\alpha|}}{\partial x_1^{\alpha_1}\cdots x_k^{\alpha_k}}.
\end{aligned}$$

Now we let

$$\eta_\alpha(\theta) = \int c_\alpha(x)\,p(x,\theta)\,\mathrm{d}x = \int c_{\alpha_1}(x)\cdots c_{\alpha_k}(x)\,p(x,\theta)\,\mathrm{d}x.$$

Defining the point $\bar\theta := [0\ldots 0,\theta_{k+1}]$ we have the following straight forward identity

$$\begin{aligned}
\varphi(\theta) &= \int \exp\left(\sum_{i=1}^{k}\theta_i c_i(x) - \theta_{k+1}\xi(x)\right)\mathrm{d}x\\
&= \int\left(\sum_{\alpha\in\mathbb{N}^k}\frac{1}{\alpha!}(\theta_1\ldots\theta_k)^\alpha c_\alpha(x)\right)\exp(-\theta_{k+1}\xi(x))\,\mathrm{d}x\\
&= \sum_{\alpha\in\mathbb{N}^k}\frac{1}{\alpha!}\eta_\alpha(\bar\theta)(\theta_1\ldots\theta_k)^\alpha, \qquad\qquad (A.10)
\end{aligned}$$

where $\eta_0 := \varphi$. This reduces the problem of computing ψ at an arbitrary point θ to the computation of ψ as well as higher order moments at a certain fixed point $\bar\theta = [0\ldots 0,\theta_{k+1}]$. For these quantities we have

$$\begin{aligned}
\eta_\alpha(\bar\theta) &= \int c_\alpha(x)\cdot\exp(-\theta_{k+1}\xi(x) - \psi(\bar\theta))\cdot\mathrm{d}x,\\
\varphi(\bar\theta) &= \int \exp(-\theta_{k+1}\xi(x))\cdot\mathrm{d}x.
\end{aligned}$$

Now ξ is a homogenous function of degree r. Substituting

$$x = \theta_{k+1}^{-1/r} z, \qquad \mathrm{d}x = \theta_{k+1}^{-d/r} \mathrm{d}z,$$

we get

$$\varphi(\bar{\theta}) = \theta_{k+1}^{-d/r} \int \exp(-\xi(x))\, \mathrm{d}x.$$

The integral is a constant which may be computed offline. So far we used only the homogenity of ξ. Now for our adapted exponential families we have

$$\int \exp(-\xi(x))\, \mathrm{d}x = \left(\int \exp(-|t|^r)\, \mathrm{d}t\right)^d = \left(\frac{2\Gamma(\frac{1}{r})}{r}\right)^d,$$

which finally yields

$$\psi(\bar{\theta}) = d\cdot \left(\log(\frac{2\Gamma(\frac{1}{r})}{r}) - \frac{\log(\theta_{k+1})}{r}\right).$$

To calculate the $\eta_\alpha(\bar{\theta})$ it can be very helpful to exploit symmetries and invariance properties of the c_i's. For example, the c_i's may be functions that factorize into functions depending on one coordinate only. Then all c_α factorize as well, and finally by choice of ξ, all η_α factorize into integrals over only one coordinate. If for example the c_i's are monomials, these integrals can be expressed in closed form using again the Γ–function.

The problem of representing φ is now solved. In principle we could get the moments $\eta_\alpha(\theta)$ for general θ from computing higher formal derivatives of the power series for φ. i.e. use the general equation

$$\eta_\alpha(\theta) = \frac{1}{\varphi(\theta)} \frac{\partial^{|\alpha|}}{\partial \theta_\alpha} \varphi(\theta)$$

and apply it to Eqation (A.10). The approximation by taking a truncated power series expansion for φ and compute formal derivatives is however of less order than the approximation for φ itself. There are, however, other methods exploiting the fact that the moments cannot be completely independent in general.

We will investigate $\varphi = \exp(\psi)$ in the following for the special case that the canonical statistics are monomials and λ is the Lebesgue measure. To denote these monomials, let $A = \{\alpha^{(1)} \ldots \alpha^{(k)}\}$ be a set of d–dimensional multiindices, i.e. each $\alpha^{(i)}$ denotes a whole d–dimensional multiindex and $\alpha_j^{(i)}$ denotes the j'th entry so that the canonical statistics can be written as $c_i(x) = x^{\alpha^{(i)}}$.

$$\varphi(\theta) = \int \exp(\sum_{i=1}^{k} \theta_i x^{\alpha^{(i)}}) \mathrm{d}x,$$

It is readily seen that φ satisfies a set of partial differential equations in this case. The partial differential equations will be valid in the interior of the parameter space, so we can assume it to be open. Now substitute $z_j := \tau_j x_j$ for every $j = 1, \dots, d$ and τ_j close to one. Then we have

$$\varphi(\theta) = \int \tau \exp(\sum_j \theta_j \tau^{\alpha^{(i)}} z^{\alpha^{(i)}}) \mathrm{d}z = \tau\varphi(\tau^{\alpha^{(1)}}\theta_1, \dots, \tau^{\alpha^{(k)}}\theta_k),$$

where $\tau := \tau_1 \cdots \tau_d$. This property will be referred to as the *invariance property* of φ. Now take the gradient of both sides with respect to $(\tau_1, \dots, \tau_d)$ and set them all equal to 1, which yields

$$0 = \varphi(\theta) + \sum_{i=1}^{k} \alpha_j^{(i)} \theta_i \frac{\partial\varphi}{\partial\theta_i}, \qquad j = 1, \dots, d. \tag{A.11}$$

This is the desired set of equations. These equations can be extremely helpful for computing higher moments. Since $\eta_i = \varphi(\theta)^{-1} \frac{\partial\varphi}{\partial\theta_i}$ we get from the preceding equation

$$0 = 1 + \sum_{i=1}^{k} \alpha_j^{(i)} \theta_i \eta_i, \qquad j = 1, \dots, d.$$

by dividing through $\varphi(\theta)$. This equation may be analyzed by linear tools for the solution space. To obtain similar relations for higher order moments, take the derivative $\frac{\partial^{|\beta|}}{\partial\theta_\beta}$ of eq. (A.11) and aqain divide by $\varphi(\theta)$. There are again further symmetries in this problem: Let β, γ be two k–dimensional multiindices. Now if

$$\beta_1\alpha^{(1)} + \ldots + \beta_k\alpha^{(k)} = \gamma_1\alpha^{(1)} + \ldots + \gamma_k\alpha^{(k)},$$

then $\eta_\beta = \eta_\gamma$. The problem of course becomes more and more involved and may require computational algebraic methods. It seems that symmetry plays an essential role here, and maybe employing group theory one can get more general results.

Bibliography

[1] Amari, S.-I. *Differential Geometric Methods in Statistics*, vol. 28 of *Lecture Notes in Statistics.* Springer–Verlag, Berlin, 1985.

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

[3] Atar, R., and Zeitouni, O. Exponential stability for nonlinear filters. *Ann. Inst. H. Poincaré Prob. Statist. 36* (1997), 691–725.

[4] Barndorff-Nielsen, O. E. *Information and exponential families.* John Wiley & Sons, New York, 1978.

[5] Bibby, B. M., and Sørensen, M. Martingale estimation functions for discretely observed diffusion processes. *Bernoulli* (1996).

[6] Birkhoff, G. *Lattice Theory*, 3 ed., vol. 25 of *AMS Colloquium Publications.* American Mathematical Society, Providence, Rhode Island, 1967.

[7] Breiman, L. *Probability.* Addison-Wesley-Publishing, 1973.

[8] Brigo, D. *Filtering by projecting on the manifold of exponential densities.* PhD thesis, Vrije Universiteit Amsterdam, 1996.

[9] Brigo, D., Hanzon, B., and LeGland, F. A differential geometric approach to nonlinear filtering: the projection filter. Tech. rep., Institut de Recherche en Informatique et Systèmes Aléatoires, 1995.

[10] Bröcker, J., and Parlitz, U. Efficient noncausal noise reduction for deterministic time series. *Chaos 11*, 2 (2001), 319–326.

[11] Bröcker, J., and Parlitz, U. Analyzing communication schemes using methods from nonlinear filtering. *Chaos (to appear) 13*, 3 (2003).

[12] Bröcker, J., Parlitz, U., and Ogorzałek, M. Nonlinear noise reduction. *Proceedings of the IEEE 90*, 5 (May 2002), 898–918.

[13] Bröcker, T. *Analysis*, vol. I, II, III. BI-Verlag, 1992.

[14] Budhiraja, A., and Kushner, H. Robustness of nonlinear filters over the infinite time interval. *SIAM journal on Control and Optimisation 36*, 5 (1998), 1618–1637.

[15] Cawley, R., and Hsu, G.-H. Local–geometric–projection method for noise reduction in chaotic maps and flows. *Phys. Rev. A 46*, 6 (1992), 3057–3082.

[16] Cawley, R., and Hsu, G.-H. Snr performance of a noise reduction algorithm applied to coarsely sampled chaotic data. *Phys. Lett. A 166* (1992), 188–196.

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

[18] Davies, M. Nonlinear noise reduction through monte carlo sampling. *Chaos 8*, 4 (1998), 775–781.

[19] Del Moral, P. Nonlinear filtering: monte-carlo particle resolution. Tech. Rep. 02, Laboratoire de Statistique et Probabilités, Université Paul Sabatier, 31062 Toulouse, 1996.

[20] Del Moral, P. A uniform convergence theorem for the numerical solving of the nonlinear filtering problem. Tech. Rep. 14, Laboratoire de Statistique et Probabilités, Université Paul Sabatier, 31062 Toulouse, 1996.

[21] Doob, J. *Stochastic Processes.* John Wiley & Sons, Inc., New York, 1953.

[22] Dunford, N., and Schwartz, J. T. *Linear Operators.* Interscience Publishers, Inc., New York, 1958.

[23] Elliott, R. J., Aggoun, L., and Moore, J. B. *Hidden Markov Models*, vol. 29 of *Applied Mathematical Sciences.* Springer Verlag, 1995.

[24] Enge, N., Buzug, T., and Pfister, G. Noise reduction on chaotic attractors. *Phys. Lett. A 175* (1993), 178–186.

[25] E.Rosa, Hayes, S., and Grebogi, C. Noise filtering in communication with chaos. *Phys. Review Letters 78*, 7 (1997), 1247–1250.

[26] Eveson, S. P. Ergodic theory of chaos and strange attractors. *Proceedings of the London Mathematical Society 3*, 70 (1995), 411–440.

[27] Ferrante, M., and Runggaldier, W. J. On necessary conditions for the existence of finite dimensional filters in discrete time. *Systems and Control Letters*, 14 (1990), 63–69.

[28] Ferrante, M., and Vidoni, P. Finite dimensional filters for nonlinear stochastic difference equations with multiplicative noises. *Stochastic Processes and their Applications*, 77 (1998), 69–81.

[29] Gallager. *information theory and reliable communication.* Wiley, New York, 1968.

[30] Grassberger, P., Hegger, R., Kantz, H., Schaffrath, C., and Schreiber, T. On noise reduction methods for chaotic data. *Chaos 3*, 2 (1993), 127–141.

[31] Hammersley, J. M. *Monte Carlo Methods.* John Wiley & Sons, New York, 1964.

[32] Hegger, R., Kantz, H., and Matassini, L. Denoising human speech signals using chaoslike features. *Phys. Rev. Lett. 84* (2000), 3197–3200.

[33] Hindmarsh, J., and Rose, R. A model of neuronal bursting using three coupled first order differential equations. *Proceedings of the Royal Society of London 221* (1984), 87–102.

[34] Hodgkin, A. L., and Huxley, A. F. *J. Physiol. London 117* (1952), 500.

[35] Holzfuss, J., and Kadtke, J. Global nonlinear noise reduction using radial basis functions. *Int. J. of Bif. and Chaos 3*, 3 (1993), 589–596.

[36] Huijberts, H. On existence of extended observers for nonlinear discrete time systems. In *New Directions in Nonlinear Observer Design*, H. Nijmeijer and T. e. Fossen, Eds., vol. 244 of *Lecture Notes in Control and Information Sciences*. Springer, 1999, pp. 73–92.

[37] Jazwinsky. *Stochastic Processes and Filtering Theory*, vol. 64 of *Mathematics in Science and Engineering*. Academic Press, 1970.

[38] Julier, S., and Uhlmann, J. A general method for approximating nonlinear transformations of probability distributions. Tech. rep., Department of engineering Science, University of Oxford, 1996.

[39] Kallianpur, G. *Stochastic Filtering Theory*. No. 13 in Applications of Mathematics. Springer Verlag, 1980.

[40] Kalman, R. E. A new approach to linear filtering and prediction problems. *Trans. ASME, Ser. D: J. Basic Eng. 82* (1960), 35–45.

[41] Kantz, H., and Schreiber, T. *Nonlinear Time Series Analysis*. Cambridge UP, Cambridge, 1997.

[42] Kantz, H., Schreiber, T., Hoffmann, I., Buzug, T., Pfister, G., Flepp, L., Simonet, J., Badii, R., and Brun, E. Nonlinear noise reduction: A case study on experimental data. *Phys. Rev. E 48* (1993), 1529–1538.

[43] Kennedy, M. P., Rovatti, R., and Setti, G. *Chaotic Electronics in Telecommunications*. CRC Press, 2000.

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

[45] Krengel, U. *Ergodic Theorems*. de Gruyter, 1985.

[46] Krengel, U. *Wahrscheinlichkeitstheorie und Statistik*. Vieweg, 1991.

[47] Kunita, H. Asymptotic behavior of the nonlinear filtering errors of markov processes. *Journal of Multivariate Analysis 1* (1971), 365–393.

[48] LEGLAND, F. Stability and approximation of nonlinear filters: an information theoretic approach. Tech. rep., IRISA/INRIA, 1999.

[49] LEVINE, J., AND PIGNIE, G. Exact finite–dimensional filters for a class of nonlinear discrete–time systems. *Stochastics 18* (1986), 97–132.

[50] MEES, A., AND JUDD, K. Dangers of geometric filtering. *Physica D 68* (1993), 427–436.

[51] MOOD, A. M., GRAYBILL, F. A., AND BOES, D. C. *Introduction to the Theory of Statistics.* McGraw-Hill Series in Probability and Statistics. McGraw-Hill, 1974.

[52] NIJMEIJER, H., AND FOSSEN, T. E., Eds. *New Directions in Nonlinear Observer Design*, vol. 244 of *Lecture Notes in Control and Information Sciences.* Springer, 1999.

[53] NIJMEIJER, H., AND VAN DER SCHAFT, A., Eds. *Nonlinear Control Systems.* Springer, 1990.

[54] PACKARD, N. H., CRUTCHFIELD, J. P., FARMER, J. D., AND SHAW, R. S. Geometry from a time series. *Phys. Rev. Lett. 45* (1980), 712.

[55] PAPOULIS, A. *Signal Analysis.* McGraw–Hill Book Company, 1977.

[56] POLLICOTT, M., AND YURI, M. *Dynamical systems and ergodic theory*, vol. 40 of *Mathematical Society Student Texts.* Cambridge University Press, 1998.

[57] ROCKAFELLAR, T. *Convex Analysis.* Princeton University Press. Princeton, 1970.

[58] RUNGGALDIER, W. J., AND SPIZZICHINO, F. Sufficient conditions for finite dimensionality of filters in discrete time: A Laplace transform-based approach. *Bernoulli 7* (2001), 211–221.

[59] SAUER, T. A noise reduction method for signals from nonlinear systems. *Physica D 58* (1992), 193–201.

[60] SAUER, T. How many delay coordinates do you need? *Int. J. Bif. Chaos 3*, 3 (1993), 737–744.

[61] SAUER, T., YORKE, J., AND CASDAGLI, M. Embedology. *J. Stat. Phys. 65*, 3/4 (1991), 579–616.

[62] SAWITZKI, G. Finite-dimensinal fiters in discrete time. *Stochastics 5* (1981), 107–114.

[63] SCHREIBER, T. An extremely simple nonlinear noise reduction method. *Phys. Rev. E 47* (1993), 2401.

[64] SCHREIBER, T., AND KAPLAN, D. T. Nonlinear noise reduction for electrocardiograms. *Chaos 6*, 1 (1996), 87–92.

[65] SETTI, G., Ed. *Applications Of Nonlinear Dynamics to Electronic and Information Engineering. Proceedings of the IEEE 90,* No. 5 (May 2002).

[66] SHANNON, C. E., AND WEAVER, W. *The Mathematical Theory of Communication.* Univ. of Illinois Press, 1949.

[67] SORENSON, H. W. On the development of practical nonlinear filters. *Information Sciences 7* (1974), 253–270.

[68] STERNICKEL, K., EFFERN, A., LEHNERTZ, K., SCHREIBER, T., AND DAVID, P. Nonlinear noise reduction using reference data. *Phys. Review E 63* (2001).

[69] STETTNER, L. On invariant measures of filtering processes. In *Stochastic Differential Systems, Proc. 4th Bad Honnef Conf.* (1989), K. Helmes, N. Christopeit, and M. Kohlmann, Eds., Lectures in Control and Information Sciences, pp. 279–292.

[70] TAKENS, F. Detecting strange attractors in turbulence. *Lecture Notes in Mathematics 898* (1981).

[71] TOKUA, I., PARLITZ, U., ILLING, L., KENNEL, M., AND ABARBANEL, H. D. I. Parameter estimation for neuron models. In *Proc. Europ. Control Conf002* (2002).

[72] VAPNIK, V. N. *Statistical Learning Theory.* John Wiley & Sons, Inc., New York, 1998.

[73] Walker, D., Allie, S., and Mees, A. Exploiting the periodic structure of chaotic systems for noise reduction of nonlinear signals. *Phys. Lett. A 242* (1998), 63–73.

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

[75] Wiener, N. *Extrapolation, Interpolation and Smoothing of stationary time series.* MIT Press and John Wiley & Sons Inc., New York, 1950.

Lebenslauf

Jochen Bröcker

30.05.1973	geboren in Kiel
1979–1983	Besuch der Grundschule in Neumünster
1983–1992	Besuch der Integrierten Gesamtschule in Neumünster–Brachenfeld, Abschluß mit Abitur
1992–1995	Studium der Physik an der Christian–Albrechts–Universität Kiel
seit 1995	Studium der Physik an der Georg–August–Universität Göttingen
seit 1997	am III. Physikalischen Institut der Universität Göttingen
4.2.1999	Diplom in Physik
1.9.1999–31.8.2002	Promotionsstipendiat des Graduiertenkollegs “Strömungsinstabilitäten und Turbulenz” an der Universität Göttingen