# Noise Reduction and Filtering of Chaotic Time Series

Jochen Bröcker and Ulrich Parlitz

*Drittes Physikalisches Institut, Universität Göttingen*
*Bürgerstr. 42-44, 37073 Göttingen, Germany*
Parlitz@physik3.gwdg.de
Broecker@physik3.gwdg.de

***Abstract***— We present a general method for state and parameter estimation for discrete time dynamical systems perturbed by noise. A rigorous approach to these problems would be to consider the corresponding probabilities. It turns out that for chaotic systems these probabilities become the more complicated functions the longer the system evolves and the smaller the uncertainty becomes. In fact, in general they follow an infinite dimensional dynamics. The algorithm proposed uses parametrized families of probabilities, the well known *exponential families*, to approximate the true probabilities. The dynamics of the probabilities then is carried over to the finite dimensional parameter space of the exponential family. This approach was suggested in [3] for continuous time filtering. It turns out that finding the parameters of the exponential family is a convex optimisation problem. We consider a simple example and compare an implementation of our filter to the extended Kalman filter.

## I. Introduction

The dynamical systems we deal with in this paper are randomly perturbed iterated maps. They are special cases of markov processes, and in fact most of what will be said in this paper is valid for general Markov processes. We, however, consider only dynamical systems of the following form:

$$X_{n+1} = f(X_n, \alpha) + A(X_n) \cdot r_n, \tag{1}$$

where $f(\cdot, \alpha)$ is a diffeomorphism of $\mathbb{R}^d$, $\alpha$ is a parameter, $A : \mathbb{R} \to GL(d)$ a matrix valued function and $r_n$ is a series of independent gaussian random vectors with mean zero and covariance $I$. Additionally $r_n$ is independent of $\{X_k\}_{k \leq n}$. Now to put this equation into a stochastic framework, we define the measures $P_n(M) := \text{Prop}(X_n \in M)$, where $M$ is a subset of $\mathbb{R}^d$. It is easy to see that these probabilities are determined by the initial distribution $P_0(M) := \text{Prop}(X_0 \in M)$ of $X_0$ and the *transition* probability $\Phi(M, \xi) := \text{Prop}(X_{n+1} \in M | X_n = \xi)$.

Very often in applications one deals in fact with probability density functions (pdf's) with respect to a certain *carrier measure* $\lambda$. We fix $\lambda$ from now on throughout the rest of this paper. Then $p_X(x)$ always denotes the pdf of $X$ w.r.t. $\lambda$. It turns out [2] that the pdfs $p_{X_n}(x)$ satisfiy the following dynamical equation:

$$\begin{aligned} &p_{X_{n+1}}(x) \\ &\quad = \int_{\mathbb{R}^d} \varphi(x, \xi, \alpha) p_{X_n}(\xi) \mathrm{d}\lambda(\xi) \\ &\quad =: \mathcal{L}^* p_{X_n}(\xi), \end{aligned} \tag{2}$$

with $\mathcal{L}^*$ denoting the integral operator on the r.h.s. The transition probability $\varphi(x, \xi, \alpha)$ can be calculated from the equation (1). It turns out that

$$\begin{aligned} \varphi(x, \xi, \alpha) &= \mathcal{N}(x, f(\xi, \alpha), R(\xi)) \\ R(\xi) &= A(\xi) \cdot A(\xi)^{tr}, \end{aligned}$$

where $\mathcal{N}(\cdot, \mu, \Gamma)$ is a gaussian pdf with mean $\mu$ and covariance $\Gamma$.

In applications the underlying dynamics of a measured time series is often are modeled by a process like (1). The problem is then to estimate the state $X_n$ and the unknown parameter $\alpha$ from the measurements. Suppose for example we collect measurements of the form

$$Y_n = G(X_n) + B(X_n) \cdot s_n, \tag{3}$$

where $s_n$ is a random variable independent of the whole process $\{X\}_0^\infty$. Let $\mathcal{Y}^n := \{Y_1, \ldots, Y_n\}$. Then a rigorous approach to estimate $X_n$ or $\alpha$ is to consider the *conditional probability* $\text{Prop}(X_n, \alpha | \mathcal{Y}^n)$ or the corresponding pdf denoted by $p_n(x, a)$. It turns out [2] that this pdf satisfies the following iterative equation:

$$p_{n+1}(x, a) = c \cdot q(Y_{n+1}, x) \cdot \mathcal{L}^* p_n(x, a) \tag{4}$$

where $q(Y_{n+1}, x)$ can be calculated from equation (3). In fact

$$\begin{aligned} q(y, x) &= p_{Y_{n+1}|X_{n+1}=x}(y) \\ &= \mathcal{N}(y, G(x), S(x)) \\ S(x) &= B(x) \cdot B(x)^{tr}, \end{aligned}$$

where $\mathcal{N}$ is again a gaussian probability density. Closed form solutions to the problem stated above are rarely found. A well known example is the KALMAN Filter for linear systems. In this case one is concerned with normal pdfs only. The filter is described by dynamical equations for the mean and covariance.

In the nonlinear case the difficulty is the quadrature in equation (4). Yet the equation is linear but in general infinite dimensional. Especially for chaotic systems the pdf is subject to the stretch and fold mechanism and becomes a quite complicated function. Therefore grid methods to solve the equation (4) may suffer from the insufficient storing of the complicated pdf on a grid of finite points.

## II. Exponential Families

The main goal of this paper is to give an approximative solution of the filter problem by approximating the pdfs. This means: Specify a set $\mathcal{E}$ of pdfs on $\mathbb{R}^d$ and choose a kind of metric. Then search the element $p \in \mathcal{E}$ closest to the true density with respect to the metric. Although other choices seem to have good properties for special purposes, throughout this paper we work only with the well known *exponential families*, defined as follows (for a global overview see [1]): Suppose $\mathcal{E}$ is a set of pdfs with respect to a reference measure $\lambda$, $\Theta \subset \mathbb{R}^k$ is an open set and $c_i : \mathbb{R}^d \to \mathbb{R}$, $i = 1 \dots k$ is a set of random variables so that

$$p : \Theta \to S, \theta \to \exp(\sum_i \theta_i c_i - \psi(\theta))$$

is a bijective map, where

$$\psi(\theta) := \log \int \exp(\sum_i \theta_i c_i(x)) \, \lambda(\mathrm{d}x).$$

Then $\mathcal{E}$ is an exponential family. We write $p(x, \theta) := p(\theta)(x)$ in the following. By the definition of $\psi$ the distribution $p$ is normalized: $\int p(x, \theta) \, \mathrm{d}x = 1$. Taking the derivative with respect to $\theta_i$ on both sides one obtains

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

The $\eta_i$ are called the *$c_i$–moments* or *expectation parameters*. One easily obtains the following identity:

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

If we additionally require the random variables $\partial \log p / \partial \theta_i$ to be linear independent then $g_{ij}$ turns out to be a nonsingular positiv definite matrix, called the FISHER *metric*. Furthermore the expectation parameters $\eta_i$ are *globally* diffeomorphic functions of the $\theta_i$. Therefore the expectation parameters form another coordinate system for $\mathcal{E}$, which is of great use in the following.

As a metric $d(P_1, P_2)$ between two mutually absolutely continuous probability measures $P_1, P_2$ we introduce the KULLBACK–LEIBLER distance

$$d(P_1, P_2) := \int \log \frac{\mathrm{d}P_1}{\mathrm{d}P_2} \, P_1(\mathrm{d}x).$$

The Kullback–Leibler distance is neither symmetric nor fulfills the triangle inequality. Nevertheless, $d(P_1, P_2)$ is always positive (Kullback–Leibler inequality) and vanishes if and only if $P_1 = P_2$ a.s. Furthermore, let $P_1$ be an arbitrary measure and $\frac{\mathrm{d}P_2}{\mathrm{d}P_1} = p(x, \theta)$ belong to an exponential family $\mathcal{E}$. Suppose $P_2$ is a critical point of $d(P_1, P_2)$, i.e. $\partial d(P_1, P_2)(x, \theta) / \partial \theta_i = 0$. Necessary conditions are

$$\eta_i(\theta) = \int c_i(x) \, p(x, \theta) \, \lambda(\mathrm{d}x) = \int c_i(x) \, P_1(\mathrm{d}x). \quad (5)$$

Then taking the second derivative we obtain

$$\frac{\partial^2}{\partial \theta_i \partial \theta_j} d(P_1, P_2)(x, \theta) = \frac{\partial^2 \psi}{\partial \theta_i \partial \theta_j},$$

which is the Fisher metric. Since this is a positiv definite matrix, the critical point turns out to be a minimum. It is in fact easy to see that minimizing the Kullback–Leibler distance w.r.t. $\theta$ is a convex optimisation problem and related to a Legendre transform of $\psi$. For such problems, there exist always a unique solution as well as numerical algorithms to find it.

## III. Approximative Filters

In this section we approximate the true pdfs by pdfs of an exponential family by minimising the Kullback–Leibler distance between them. Consider the equation (2) first. The parameter $\alpha$ doesn't play any exeptional role and is ommitted. Suppose an exponential family $\mathcal{E}$ is chosen and $p_n(x) = p(x, \theta^n) \in \mathcal{E}$ ($\theta^n$ is not the $n$–th power but the parameters of $p_n$). Then we have

$$p_{n+1}(x) = \int \varphi(x, y) \, p(y, \theta^n) \, \mathrm{d}y.$$

In general $p_{n+1}(x) \notin \mathcal{E}$. We approximate $p_{n+1}(x)$ by $p(x, \theta^{n+1})$. According to eq. (5) this yields

$$\begin{aligned} \eta_i(\theta^{n+1}) &= \int c_i(x) \, p_{n+1}(x) \, \mathrm{d}x \\ &= \int (\int c_i(x) \, \varphi(x, y) \, \mathrm{d}x) \, p(y, \theta^n) \, \mathrm{d}y \\ &= \int \mathcal{L} c_i(y) \, p(y, \theta^n) \, \mathrm{d}y \end{aligned}$$

The r.h.s. is a function of $\theta^n$ and the l.h.s. are the $c_i$–moments of the new pdf. According to the previous section the parameters $\theta^n$ are invertible functions of $\eta_i(\theta^n)$. The same holds for $\theta^{n+1}$ and $\eta_i(\theta^{n+1})$. So in principle the above equation gives the equation (2) approximately as an iterative map $\Pi : \Theta \to \Theta; \theta^{n+1} = \Pi(\theta^n)$. Since the product of two exponential pdfs is again an exponential pdf the equation (4) is simplified considerably if $q(y, x)$ viewed as a function of $x$ is in $\mathcal{E}$ for every $y$. Then according to equation (4) the multiplication with $q$ can be performed without leaving $\mathcal{E}$.

## IV. Numerical Examples

We studied as a numerical example the Henon map

$$\begin{aligned} x_{n+1} &= 1 - \alpha \cdot x_n^2 + 0.3 \cdot y_n \\ y_{n+1} &= x_n \end{aligned}$$

with oservable

$$\begin{aligned} v_n &= G \cdot (x_n, y_n) = x_n + y_n \\ G &= (1\,1)\,. \end{aligned}$$

The parameter $\alpha$ is assumed to be unknown. Practically this problem can equivalently be treated by introducing a further dynamical equation

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

for $\alpha$. Furthermore, we assume a Gaussian observation noise (of 18.5 dB) as well as a small dynamic noise ($\cong$ 30 dB) to be present. As an exponential family we use 3–dimensional Gaussian distributions

$$p((x_n, y_n, \alpha_n), \Gamma_n, \mu_n) = \mathcal{N}((x_n, y_n, \alpha_n), \Gamma_n, \mu_n),$$

where $\Gamma_n, \mu_n$ play the roles of the canonical parameters $\theta$. It turns out that all moments can be calculated analytically and therefore an expression for the map $\Pi$ can be obtained explicitely giving iterative equations for $\mu_n$ and $\Gamma_n$. Since the observation is linear, the function $q$ is gaussian and the multiplication with $q$ in equation 4 yields again a gaussian.

Figure 1 shows the performance. The upper panel shows the remainig error in the noise–cleaned output. The quadratic deviation is plotted on a logarithmic scale and normalized with the variance of the observation noise. The lower panel shows the performance for the parameter estimation. We plotted $\log(frac\alpha_n - 1.41.4)^2$. The value 1.4 was the true value.

The results where compared to the performance of an extended Kalman filter (EKF). The EKF is obtained by linearisation of the nonlinear system equations and then applying a usual Kalman filter. The detailed procedure of this very common technique is described in [4].

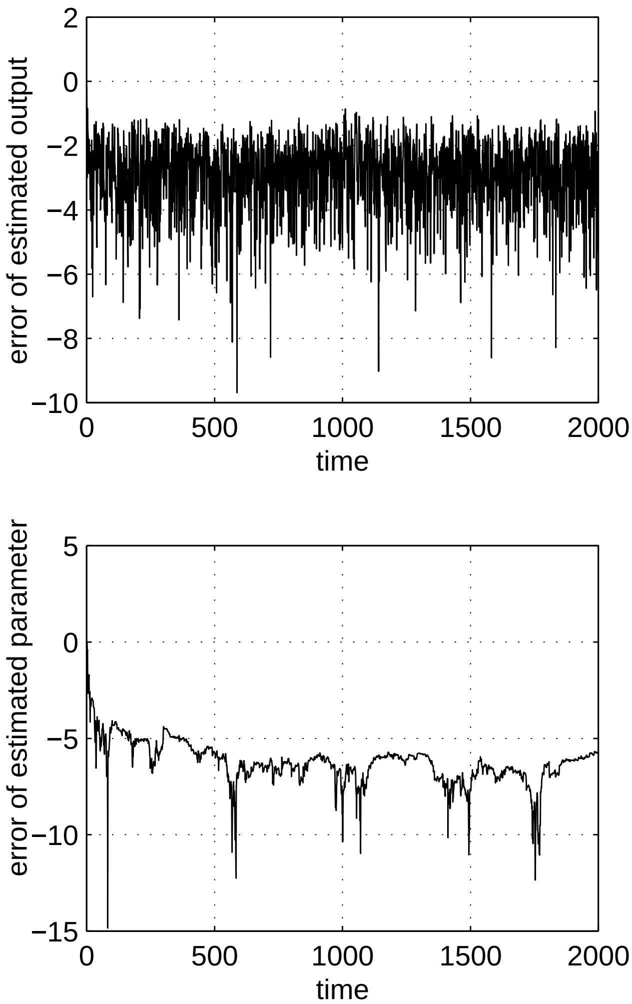


Figure 1: Performance for the new filter

Figure 2 shows the performance for the Kalman filter. The plots and scales are exactly the same. It is easy to see that the EKF works a few magnitudes worse than our algorithm. In the parameter estimate, even a systematic error remains, while the parameter estimate of the new algorithm seems to be more or less convergent.

Finally we remark that the computing time of the new filter was only 25 percent larger than for the EKF. The memory consumption was more or less the same.

## V. Conclusion

We presented a new method to approximate nonlinear filters that may be used, for example, as observers in noisy environments and for system identification. They seem to work quite well for the systems investigated, in fact better than the usual technique of extended Kalman filtering. Open questions are for example the choice of a proper exponential family to simplify the required calculations and to achieve better performance. This will be subject to further studies.

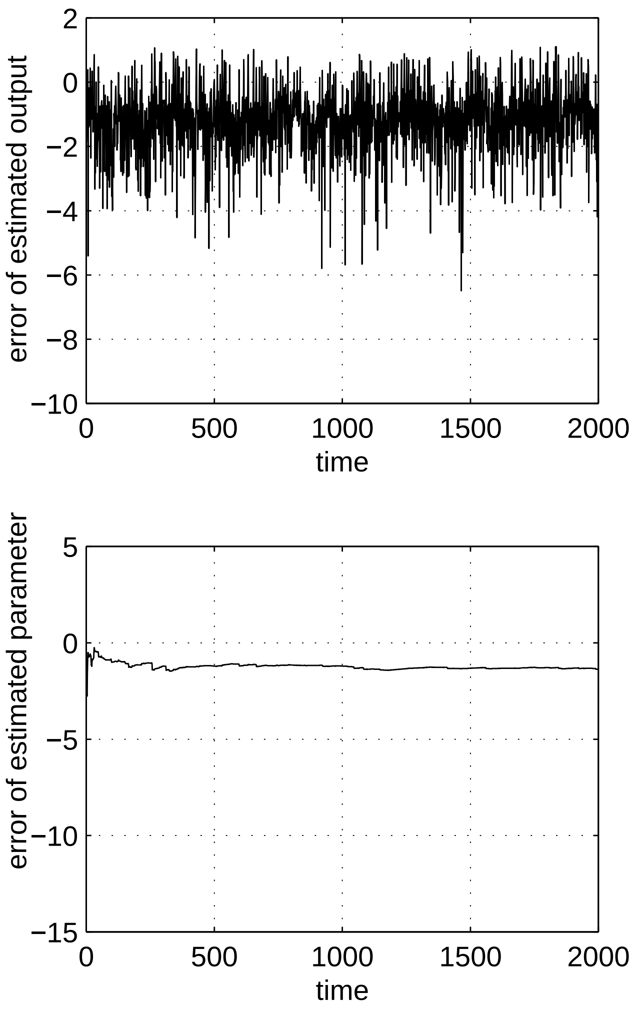


Figure 2: Performance for the EKF

## Acknowledgements

We acknowledge support by the *Deutsche Forschungsgemeinschaft* (Graduiertenkolleg “Strömungsinstabilitäten und Turbulenz”).

## References

[1] Sun-Ichi Amari. *Differential Geometric Methods in Statistics*, volume 28 of *Lecture Notes in Statistics*. Springer–Verlag, Berlin, 1985.

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

[3] Damiano Brigo, Bernard Hanzon, and Francois LeGland. A differential geometric approach to nonlinear filtering: the projection filter. Technical report, Institut de Recherche en Informatique et Systèmes Aléatoires, 1995.

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