FILTERING USING EXPONENTIAL PROBABILITY DENSITIES

Jochen Bröcker† and Ulrich Parlitz‡

Drittes Physikalisches Institut der Universität Göttingen,
Bürgerstrasse 42–44, 37073 Göttingen, Germany
†Broecker@dpi.physik.uni-goettingen.de, ‡Parlitz@dpi.physik.uni-goettingen.de

Abstract— In this paper we consider an approximative method to solve nonlinear filter and observer problems. This method is closely related to the *projection filter* proposed by BRIGO, HANZON and LEGLAND in [3]. However, we will present a more direct way to obtain the filter which can be applied also in discrete time. In the first section we give a short introduction into filtering. In the following section some basic results of the theory of *exponential families* are summarized. After presenting and discussing the main steps of our filter approximation we study some numerical examples.

I. Filtering

The filter or observer problem [5] consist of estimating the state x_t of a dynamical system at any time t (t is real or integer valued). However, x_t is not available but only a noisy measurement

$$y_t = G(x_t) + \text{noise},$$

where G is some measurement function. To estimate the state x_t we use not only y_t but all data collected up to time t, i.e. $\{y_\tau\}_{\tau\leq t}$. Since the added noise is a random variable we cannot solve this problem exactly. Rather we shall give the solution as a probability density function (pdf) for x_t. In fact, all information about x_t that is available given $\{y_\tau\}_{\tau\leq t}$ is contained in the conditional pdf $p(x_t|\{y_\tau\}_{\tau\leq t})$. In the following we try to calculate this pdf. However, we will consider only the case of discrete time observations, where we only have measurements $\mathcal{Y}_{t_1}^{t_n} := \{y_{t_1}, \ldots, y_{t_n}\}_{t_n \leq t}$ available to estimate x_t.

We consider two cases of dynamical systems: (autonomous) stochastic differential equations (SDEs) and iterated maps with dynamical noise.

SDEs

By a dynamical system in continuous time we think at first of an ODE on an open subset U of $\mathbb{R}^d$ given by a vector field $v : U \to \mathbb{R}^d$. Furthermore, we can think of stochastic influences not only in the initial values of the trajectories but also during their temporal evolution (*dynamic noise*). Such systems are subject to the theory of *stochastic differential equations* (SDEs). For an introduction see for example [2]. Suppose the SDE has the form

$$\frac{\mathrm{d}x}{\mathrm{d}t} = v(x) + a(x)\cdot r_t, \tag{1}$$

where r_t is a vector valued gaussian white noise process with covariance equal to the identity matrix and $a : U \to Gl(d)$ is a matrix valued function. The pdf $p_t(x)$ of x_t at time t satisfies the following FOKKER–PLANCK equation (FPE):

$$\frac{\partial p_t}{\partial t} + \sum_{i=1}^{n} \frac{\partial(v_i \cdot p_t)}{\partial x_i} - \frac{1}{2}\sum_{i,j=1}^{n} \frac{\partial^2 (s_{ij}\cdot p_t)}{\partial x_i \partial x_j} = 0, \tag{2}$$

where $s = a \cdot a^t$. In the following we will use the forward operator defined by

$$\mathcal{L}^* p = -\sum_{i=1}^{n} \frac{\partial(v_i\cdot p)}{\partial x_i} + \frac{1}{2}\sum_{i,j=1}^{n} \frac{\partial^2(s_{ij}\cdot p)}{\partial x_i\partial x_j},$$

and the backward operator defined by

$$\mathcal{L}p = \sum_{i=1}^{n} v_i \frac{\partial p}{\partial x_i} + \frac{1}{2}\sum_{i,j=1}^{n} s_{ij}\frac{\partial^2 p}{\partial x_i \partial x_j}.$$

Note that $\mathcal{L}^*$ is in fact the adjoint of $\mathcal{L}$. Now the FPE reads as $\frac{\partial p_t}{\partial t} = \mathcal{L}^* p_t$.

Iterated maps

The situation is far more easier in discrete time. We consider dynamical systems of the following form:

$$x_{n+1} = F(x_n) + a(x_n)\cdot r_n, \tag{3}$$

where F is a diffeomorphism of $\mathbb{R}^d$, $a : \mathbb{R} \to Gl(d)$ and r_n is a series of independent gaussian random vectors with mean zero and covariance $\mathbf{1}$. Additionally r_n is independent of x_n. Suppose x_n is distributed according to $p_n(x)$. Then using elementary stochastics we obtain

$$p_{n+1}(x) = \int_{\mathbb{R}^d} \varphi(x,y)\, p_n(y)\,\mathrm{d}y. \tag{4}$$

where

$$\begin{aligned}\varphi(x,y) &= p(x_{n+1}=x|x_n=y)\\ &= \frac{1}{\sqrt{(2\pi)^d \det\Gamma_n}}\cdot\\ &\quad \exp(-\tfrac{1}{2}(x-F(y))\Gamma_n^{-1}(x-F(y))),\\ \Gamma_n(y) &= a(y)\cdot a(y)^t.\end{aligned}$$

Equation (4) is often referred to as FROBENIUS–PERRON–equation. Since it is the discrete–time–version of the FPE we will use the same abbreviation. If there is no noise in the dynamics this FPE reduces to the following transformation rule

$$p_{n+1}(x) = p_n(F^{-1}(x))\cdot\det \mathrm{D}F^{-1}(x).$$

Filter update

We now know how to deal with probability densities of variables that are subject to dynamical laws. By solving f.e. the FPE between t_n and t_{n+1} taking $p(x_{t_n}|\mathcal{Y}_{t_1}^{t_n})$ as an initial value we obtain $p(x_{t_{n+1}}|\mathcal{Y}_{t_1}^{t_n})$. This can be considered as a *prediction density* of $x_{t_{n+1}}$ based on the measurements $\mathcal{Y}_{t_1}^{t_n}$. Suppose we have a further measurement $y_{t_{n+1}}$. We now want to *update* the prediction density to get the desired density $p(x_{t_{n+1}}|\mathcal{Y}_{t_1}^{t_{n+1}})$. By using elementary Bayes rules one obtains the following identity:

$$\begin{aligned}&p(x_{t_{n+1}}|\mathcal{Y}_{t_1}^{t_{n+1}})\\ &\quad= p(x_{t_{n+1}}|\mathcal{Y}_{t_1}^{t_n}, y_{t_{n+1}})\\ &\quad= c\cdot p(y_{t_{n+1}}|x_{t_{n+1}})\cdot p(x_{t_{n+1}}|\mathcal{Y}_{t_1}^{t_n}),\end{aligned} \tag{5}$$

where c is a normalisation constant. The required density $p(y_{t_{n+1}}|x_{t_{n+1}})$ depends on the measurement process. Suppose the measurements are taken according to the following equation:

$$y_{t_n} = G(x_{t_n}) + b(x_{t_n})\cdot s_n$$

where $G:\mathbb{R}^d\to\mathbb{R}^m$ is a map, $b:\mathbb{R}\to Gl(m)$ a matrix valued function and s_n is a series of independent gaussian random vectors with zero mean and covariance $\mathbf{1}$ independent of x_{t_n}. Then we have

$$\begin{aligned}&p(y_{t_n}|x_{t_n}) = \frac{1}{\sqrt{(2\pi)^m\det\Lambda_n}}\\ &\quad\cdot\exp(-\tfrac{1}{2}(y_{t_n}-G(x_{t_n}))\Lambda_n^{-1}(y_{t_n}-G(x_{t_n})))\\ &\Lambda_n = b(x_{t_n})\cdot b(x_{t_n})^t.\end{aligned}$$

The filter problem with continuous time observations is far more complicated. The update and prediction step must be performed somewhat simultaneously. Therefore the measurements occur already in the FOKKER–PLANCK equation. They act as a kind of stochastic driving. This equation is known as KUSHNER–STRATONOVIC equation [3].

At the end of this section we list the required steps to obtain the conditional density.

prediction step : Solve the FPE between t_n and t_{n+1} taking $p(x_{t_n}|\mathcal{Y}_{t_1}^{t_n})$ as an initial value. t_{n+1} is the time instant at which the next measurement is available. In discrete time use the equation (4) instead of the FPE.

update step : Update the density according to equation (5) to obtain $p(x_{t_{n+1}}|\mathcal{Y}_{t_1}^{t_{n+1}})$. Here the new measurement $y_{t_{n+1}}$ is needed. Then return to the prediction step.

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 difficult step is the prediction step. Yet the FPEs are 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 FPEs may suffer from the insufficient storing of the complicated pdf on a grid of finite points.

II. Exponential families of pdfs

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 S of pdfs on $\mathbb{R}^d$ and choose a kind of metric. Then search the element $s\in S$ 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 S is a set of pdfs, $\Theta\subset\mathbb{R}^k$ is an open set and $c_i:\mathbb{R}^d\to\mathbb{R}$, $i=1\ldots k$ 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

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

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

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

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

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

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 η_i are diffeomorphic functions of the θ_i. Therefore the expectation parameters form another coordinate system for S, which is of great use in the following. As a metric $d(p_1, p_2)$ between two pdfs p_1, p_2 we introduce the KULLBACK–LEIBLER distance

$$d(p_1, p_2) := \int \log \frac{p_1}{p_2}\, p_1 \,\mathrm{d}x.$$

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 pdf and $p_2 = p_2(x, \theta)$ belong to an exponential family S. 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

$$\eta_i(\theta) = \int c_i(x)\, p_2(x, \theta) \,\mathrm{d}x = \int c_i(x)\, p_1(x) \,\mathrm{d}x \tag{6}$$

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.

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. We only discuss the case of discrete time in detail. Consider the prediction step first. Suppose an exponential family S is chosen and $p_n(x) = p(x, \theta^n) \in S$ (θ^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 $p_{n+1}(x) \notin S$. We approximate $p_{n+1}(x)$ by $p(x, \theta^{n+1})$. According to eq. (6) this yields

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

The r.h.s. is a function of θ^n and the l.h.s. are the c_i–moments of the new pdf. According to the previous section the parameters θ^n are invertible functions of $\eta_i(\theta^n)$. The same holds for θ^{n+1} and $\eta_i(\theta^{n+1})$. So in principle the above equation gives the prediction step as an iterative map $P : \Theta \to \Theta; \theta^{n+1} = P(\theta^n)$.

The case of continuous time FPE is worked out in detail in [3]. We give just an intuitive argument to get the required ODEs. By multiplying eq. (2) with c_i and integrating one obtains:

$$\begin{aligned} \dot\eta_i &= \frac{\partial}{\partial t} \int c_i\, p \,\mathrm{d}x = \int c_i\, \frac{\partial p}{\partial t} \,\mathrm{d}x \\ &= \int c_i\, \mathcal{L}^* p \,\mathrm{d}x = \int p\, \mathcal{L} c_i \,\mathrm{d}x. \end{aligned} \tag{7}$$

Now assume that $p = p(x, \theta) \in S$. Then the r.h.s. is a function of θ^n, which itselv is a function of η. So the above equation is an ODE for η. Since $\dot\eta_i = \sum_j \frac{\partial\eta_i}{\partial\theta_j} \dot\theta_j$ and $\frac{\partial\eta_i}{\partial\theta_j} = g_{ij}$ the above equation can be written as

$$\begin{aligned} \sum_j g_{ij} \dot\theta_j &= \frac{\partial}{\partial t} \int c_i\, p \,\mathrm{d}x = \int c_i\, \frac{\partial p}{\partial t} \,\mathrm{d}x \\ &= \int c_i\, \mathcal{L}^* p \,\mathrm{d}x = \int p\, \mathcal{L} c_i \,\mathrm{d}x \end{aligned}$$

Since the product of two exponential pdfs is again an exponential pdf the update step is simplified considerably if $p(y_n|x_n)$ viewed as a function of x_n is in S for every y_n. Then according to equation (5) the update step can be performed without leaving S.

Examples

We tested the filter using the LORENZ system (for an example using the HÉNON map see [4]). The Lorenz system is given by

$$\begin{aligned} \dot x_1 &= 10(x_2 - x_1) \\ \dot x_2 &= 28x_1 - x_2 - x_1 x_3 \\ \dot x_3 &= x_1 x_2 - \tfrac{8}{3} x_3 \end{aligned}$$

Our observable is given by a linear map:

$$\begin{array}{l} y_n = G x(t_n) + r_n, \\ t_n - t_{n-1} = 0.01, \end{array} \qquad G = \begin{pmatrix} 2 & 1 & 0 \\ 1 & 3 & 0 \end{pmatrix}.$$

For S we used three dimensional normal pdfs, i.e. for the c_i we used $x_1, x_2, x_3, x_1^2, x_2^2, x_3^2, x_1x_2, x_1x_3, x_2x_3$. Then we applied eq. (7). The r.h.s. of eq. (7) can be calculated explicitely giving nine equations for the expectation parameters. Since $p(y_n|x(t_n))$ is also a normal pdf, the update step can be performed exactly. As an estimator for the state x we used the expectation parameters $(\eta_1, \eta_2, \eta_3) =: \hat x$. $G\hat x = \int Gx\, p(x, \theta) \,\mathrm{d}x$ is the expectation of Gx which we used as an estimator for y. Note that if the measurement G is nonlinear then $G(\hat x)$ is in general *not* the expectation of $G(x)$. Figure 3 and 1 show the first component of y (referred to as the signal). The second component has the same behaviour. Starting with an SNR of 5 dB the filter achieves a noise reduction of 7 dB. Note that the estimation of x works quite well (see figure 2) despite the fact that x_3 does not enter the function G explicitely.

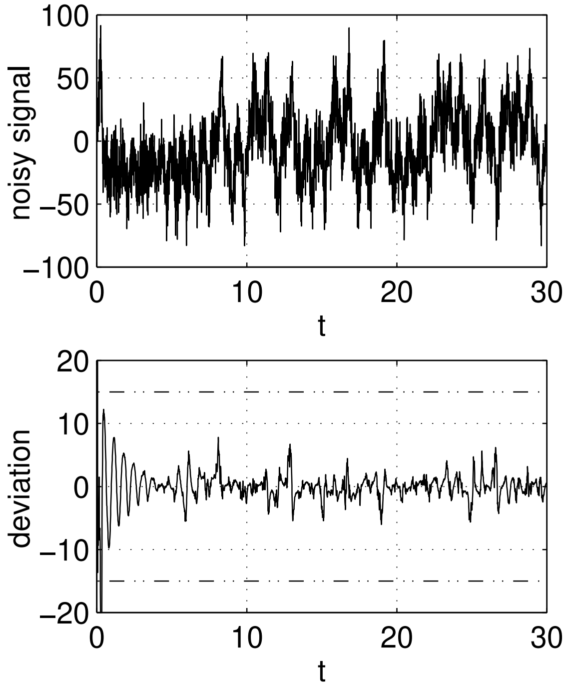

Figure 1: The upper panel shows the noisy signal (first component of y, compare to Fig. 3). The lower panel shows the difference between the true and the estimated signal. A transient is visible before the filter works properly. The dash–dotted line gives the value of the standard deviation of the measurement noise.

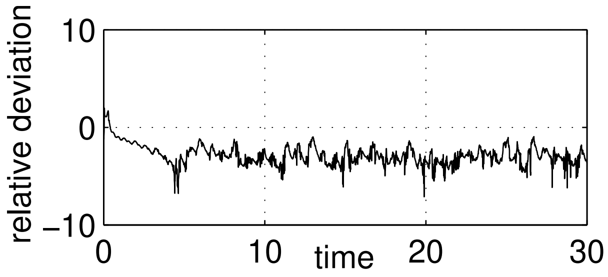

Figure 2: Relative deviation $\log(|x - \hat{x}|/s)$ between the estimated and the true state of the Lorenz system, where s denotes the standard deviation of x.

IV. Conclusion

We presented a new method to approximate nonlinear filters that may be used, for example, as observers in noisy environments and which seem to work quite well for the systems investigated. 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.

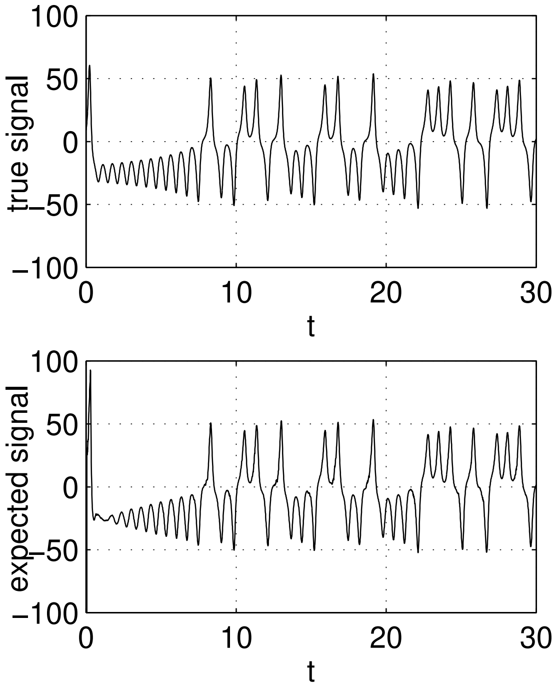

Figure 3: The upper panel shows the true signal (first component of Gx). The lower panel shows the estimated signal (first component of $G\hat{x}$).

References

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

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

[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] Jochen Bröcker. Rauschreduktion in nichtlinearen Systemen. Diploma thesis, Drittes Physikalisches Institut Göttingen, 1999.

[5] H. Nijmeijer and T.I. Fossen, editors. *New Directions in Nonlinear Observer Design*, volume 244 of *Lecture Notes in Control and Information Sciences*. Springer, 1999.

Acknowledgements

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