2008 International Symposium on Nonlinear Theory and its Applications
NOLTA'08, Budapest, Hungary, September 7-10, 2008

The Quest for a Shady Place: A guide (using shadowing filters for state estimation)

Thomas Stemler, Kevin Judd

School of Mathematics and Statistics, University of Western Australia,
35 Stirling Highway, Crawley 6009, WA, Australia
Email: thomas@maths.uwa.edu.au kevin@maths.uwa.edu.au

Abstract—Given a time series of a physical system and some model of it, the quality of forecasts rely on a satisfactory model and state estimation. Small alterations in the initial state can result in large forecast errors, even when the model is perfect. We report on a promising new tool for state estimation, the *shadowing filters*. We compare its performance with the classical extended Kalman filter and demonstrate that it is superior. To ensure successful implementation of a shadowing filter in an arbitrary model we advise using several tests. These tests cover many important aspects of a real-world applications, such as: data requirements, identification of errors, and the cost performance of different update routines for sequential state estimation. This talk reports on the superior performance of shadowing filters and provides a "how to" for implementing them.

1. Introduction

Forecasting the future behaviour of a system is of interest in engineering, physics, meteorology, geophysics and other sciences. Formally forecasting can be divided into four tasks: (i) devising an appropriate model of the system; (ii) making appropriate observations; (iii) assimilation of the observations into the model to obtain state estimates; (iv) evolving the model states to obtain forecasts.

Task (i) of devising an appropriate model is non-trivial. Often forecasters approach tasks (iii) and (iv) using statistical techniques that essentially assume one has a perfect model, that is, the model can exactly replicate the system dynamics. In practice, of course, this is a fiction; all models are wrong, but some are useful. Even if the model were perfect, the forecaster must still cope with observational errors. Many non-linear systems display sensitivity to initial conditions, that is, a small error in obtaining an appropriate state estimate can result in exponentially growing forecast errors. In practice, in order to deal with the uncertainties of observations and model error, forecasters often employ ensemble forecasts. That is, one makes multiple forecasts using differnt initial states, and sometimes different models.

In this paper we consider only the problem of task (ii), assimilation of the observations into the model to obtain state estimates. Many modern approachs to task (iii) have been developed from the Kalman filter approach, perhaps looked at from the Bayesian statistical point of view. These methods either use local linearizations or ensembles to deal with uncertainty in various way. All these various methods have the common feature that they are *sequential filters*, that is, there is current state estimate, or ensembles of state estimates, that is sequentially updated as each new observation arrives.

In this paper we present at new and different approach that is not a sequential filter. These shadowing filters are derived from dynamical system view point and are able to avoid many of the intrinsic failings of squential filters. For more details see ...

2. Shadowing filters by gradient descent of indeterminism

Although the shadowing filter can be implemented in more general imperfect model situations, for the purposes of this discussion we will consider the perfect model scenariowith isotropic Gaussian noise.

Consider a discrete time dynamical system on $\mathbb{R}^d$ with a dynamics given by the map $y_{i+1} = g(y_i)$. Assume $s_i = y_i + \xi$ is our observation of y_i, where ξ are independent Gaussian random variates with an isotropic variance σ^2. Also assume to have a model f of the system that is identical to g and that f is differentiable.

The task of a shadowing filter is to find a sequence of states $X = (x_1, \ldots, x_n)$ from a given sequence of observations $S = (s_1, \ldots, s_n)$. X should be a trajectory of the model f and should shadow S. For X to be a trajectory requires $x_i = f(x_i)$ for $i = 1, \ldots, n-1$. The trajectory will shadow S if the distances $\|s_i - x_i\|$ are not large relative to σ.

Their are a number of way to implement a shadowing filter, we consider the implementation of a shadowing filter by gradient descent of indeterminism (GDI).

Define for any sequence of X its indeterminism:

$$I(X) = \frac{1}{n-1}\sum_{i=1}^{n-1}\|x_{i+1} - f(x_i)\|^2. \quad (1)$$

Since X can be considered as a point in $\mathbb{R}^{nd}$ the indeterminism is a scalar function on this $(n \times d)$–dimensional space. Almost surely $I(S)$ will be nonzero. Also $I(X)$ will only be zero if X is a trajectory of f. The gradient descent method takes S as a starting sequence and follows the steepest decent of the gradient $I(X)$ down to the minimum where $I(X) = 0$. One way to achieve this is to solve in the limit as $\tau \to \infty$ the differential equation: $\mathrm{d}X/\mathrm{d}\tau = -\nabla I(X(\tau))$, $X(0) = S$. A more practical method is to solve the differential equation by a Euler iteration until suitable convergence is achieved, which provides an *iterative GDI shadowing filter*. Let $X_0 = S$ and $X_m = (x_{1,m}, \ldots, x_{n,m})$ where

$$\begin{aligned} x_{i,m+1} &= x_{i,m} - \frac{2\Delta}{n-1} \times c_{i,m}, \quad (2) \\ c_{i,m} &= \begin{cases} -A\left(x_{i,m}\right)\left(x_{i+1,m} - f\left(x_{i,m}\right)\right), & i = 1 \\ x_{i,m} - f\left(x_{i-1,m}\right) & 1 < i < n \\ -A\left(x_{i,m}\right)\left(x_{i+1,m} - f\left(x_{i,m}\right)\right), & \\ x_{i,m} - f\left(x_{i-1,m}\right), & i = n. \end{cases} \end{aligned}$$

Here $A(x)$ denotes the adjoint of f (transpose of the Jacobian matrix) evaluated at x, and Δ is an arbitrary step size. Later on we outline a method to find suitable values for the step size Δ, but typically the choice $\frac{2\Delta}{n-1} = 0.1$ will lead to a convergence of the iterative GDI shadowing filter.

Details on the properties of such GDI shadowing filters are given elsewhere [5] Here we just want to mention that the GDI method always converges to a shadowing trajectory of the model and $I(X)$ converges monotonically to zero. Furthermore, given a long observation sequence with sufficiently small bounded measurement noise of a hyperbolic system it can be shown that for perfect models the GDI shadowing filter converges to the true trajectory.

In practice we will iterate eq. (2) until X_m has converged sufficiently. The remaining magnitude I_m is one quantity that measures the quality of the estimated states. In addition we define below three other quantities that measure the quality: The magnitude of mismatch $I_{n,m}$, the root mean square error E_m and the last point error $E_{n,m}$:

$$\begin{aligned} I_{n,m} &= \|x_{n,m} - f(x_{n-1,m})\| \quad (3) \\ E_m &= \sqrt{\frac{1}{n}\sum_{i=1}^{n}\|x_{i,m} - y_i\|^2} \quad (4) \\ E_{n,m} &= \|x_{n,m} - y_n\| \quad (5) \end{aligned}$$

In practice we will not know y_i and therefore are unable to evaluate E_m and $E_{n,m}$. Nevertheless, they can be used to assess successful implementation using artificial data.

We are also interested in comparing the forecast $f^t(x_{n,m})$ with the future states y_{n+t}. To measure this forecast quality one can evaluate the *separation time*. We define the separation time as the largest lead time for which the forecast error remains less than a given threshold:

$$T_m = \max\left\{T\colon \|y_{n+t} - f^t(x_{m,n})\| \leq 2\sigma,\ \forall\ 0 \leq t \leq T\right\} \quad (6)$$

Here we chose the threshold to be 2σ. But the results we quote are not particular sensitive to this choice.

3. Application of the shadowing filter

Shadowing filters have been tested now for some while in simple chaotic dynamical models[8], simple atmospheric models [6] and only recently in an operational weather forecasting model at reduced resolution using real atmospheric observations [4]. In all these cases shadowing filters have proven to be a useful tool for state estimation. Unlike linear systems, each nonlinear system has its own peculiar characteristics. Therefore all these implementations follow different system–specific rules. But we were able to identify certain tests and experiments that can be used to find these specific rules in an arbitrary system. The complete guide with details on time discrete and time continuous models can be found somewhere else [8]. Here we focus on time discrete models and illustrate how to find suitable parameter values of the shadowing filter using the Ikeda map [2]:

$$\begin{aligned} u_{t+1} &= 1 + \mu\left(u_t\cos(\theta_t) - v_t\sin(\theta_t)\right) \quad (7) \\ v_{t+1} &= \mu\left(u_t\sin(\theta_t) + v_t\cos(\theta_t)\right) \\ \theta_t &= a - b/\left(1 + u_t^2 + v_t^2\right). \end{aligned}$$

This two dimensional map shows chaotic dynamics for $a = 0.4, b = 0.6, \mu = 0.83$.

To demonstrate typical behaviour of the four quantities I_m, E_m, $E_{n,m}$ and $I_{n,m}$ we computed their average values from an ensemble of 10000 observation sequences S. Each of the ensemble members consisted of $n = 15$ points and we used $m \geq 500$, $2\Delta/(n-1) = 0.1$ for $\sigma = 0.05, 0.1, 0.2$. The data can be seen in fig. 1.

Note that in fig. 1 all averages values decrease monotonically with m. The noticeably faster rate of decrease in the first 10 iterations is associated with a fast noise–reduction. Thereafter the shadowing filter makes fine adjustments to states toward obtaining a trajectory. After $m = 100$ iterations I_m decreased more than an order of magnitude, $I_{n,m}$ about two orders of magnitude, E_m decreased by a factor of 1/4 and

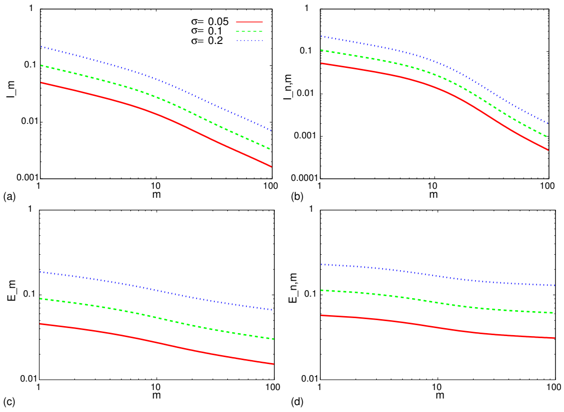

Figure 1: Average values from 10000 time series with $n = 15$ and $\sigma = 0.05, 0.1, 0.2$ for (a) I_m, (b) $I_{n,m}$, (c) E_m, and (d) $E_{n,m}$ as a function of the number of iterative GDI shadowing filter steps m.

Table 1: Various statistics of the separation time T_m, namely the mean, variance, percentage of instances where $T_m \geq T_0$, and the percentage of instances where $T_m > T_0$. The horizontal blocks correspond to data with $\sigma = 0.05, 0.1, 0.2$.

m	$\langle T_m \rangle$	var(T_m)	$\%(T_m \geq T_0)$	$\%(T_m > T_0)$
0	5.4	12.6	0	0
10	6.4	15.7	81	48
100	7.4	19.8	81	59
500	7.9	22.2	82	62
0	4.6	11.1	0	0
10	5.6	14.2	82	49
100	6.6	17.9	82	60
500	7.0	20.7	82	62
0	3.8	9.6	0	0
10	4.7	11.6	81	49
100	5.7	16.3	81	59
500	6.0	18.3	81	62

$E_{n,m}$ decreased by 1/2. From this list and the data in fig. 1 one may note that $I_m > I_{n,m}$ and $E_m < E_{n,m}$. This ordering is significant and a feature arising from the limited information available to the final state of the sequence. The final state $x_{n,m}$ only has to adjust to mismatches on one side, since x_{n+1} is unknown (cf. eq. (2)). Since points in the middle have to adjust to mismatches on both sides it is typically more difficult to achieve the minimum. On the other hand the missing information of the future dynamics leads to a poorer state estimation and consequently $E_{n,m}$ is higher than the average value E_m of the sequence.

With the given estimated states of the ensemble of observations we evaluated the separation time T_m defined above. It is useful to compare T_m for $m > 0$ with the initial value T_0 arising from the forecasts of the unfiltered data. Tab. 1 shows various statistics of T_m.

The statistical indicators shown reveal that the shadowing filter improves the quality of forecasts statistically significant. The major improvement takes place in the first 10 iterations while additional iterations give further improvement. For all noise levels we observe at $m = 500$ an increase of the average separation time of at least 2 times units. In total 60% of the estimated states do have at least 1 time unit longer separation time.

So far we demonstrated that our shadowing filter implementation is able to find state estimates that improve the forecasting. But given a observation sequence with finite length n and some σ the estimated state is not unique. From one observation sequence we find one state consistent with the dynamics of the system and the measurement noise realisation. For other noise realisations we will find another state. In fact there exists a whole set of states which are consistent with the dynamics for a given σ value. This set is called the indistinguishable set [5] and the uncertainty associated with estimating one state of it is has to be compared to the uncertainty of other filters.

The nowadays widely used variants of the extended Kalman filter use a nonlinear forecast model, but linearise the model about the current state to achieve an update of the error covariance. In fig. 2 we show the indistinguishable set and the typical phase space region in which the state estimates of an extended Kalman filter are. Both filter lead to estimates that are close to the true point $(u, v) \approx (0.75, 0.68)$. Note that the indistinguishable states are all on the attractor, while the estimates of the extended Kalman filter can be off the attractor. This leads to a big difference when forecasts are done from these states. In the figure we show three forward iterations of the map (from (u, v) down, to the left and then to the right). Since the area of estimated states of the extended Kalman filter is not on the attractor the forward iterations lead to a much higher spread of the area compared to the spread of the indistinguishable states. In conclusion we have shown, that shadowing filters do not only improve the quality of the forecasts but although are superior compared with extended Kalman filters [3].

4. The Windowing Test

The demonstrated performance of the shadowing filter depend on the particular choices of Δ, n and m. The windowing test is a basic procedure that enables us to find optimal or appropriate values of these parameters. The description below outlines the windowing test for given values of Δ and m and therefore results in an optimal window length n. But the test can be easily modified to estimate the optimal values

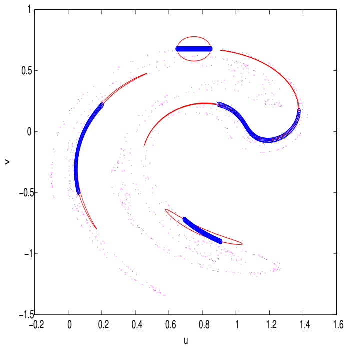

Figure 2: Typical area occupied by the estimated states using a shadowing filter (symbols) and the extended Kalman filter for a true state at $(u, v) \approx (0.75, 0.68)$. The background shows the Ikeda attractor. In addition three forward mappings of the area are shown. Note the bigger growth of the area of estimated states from the extended Kalman filter.

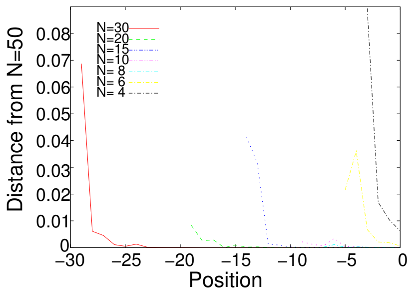

Figure 3: Difference between the estimated trajectory $X_{m,N}$ from a long time series ($N = 50$) and the estimation $X_{m,p}$ for $p \in \{4, 6, 8, 10, 15, 20, 30\}$.

of the other parameters.

The essential idea is to apply a shadowing filter on increasing length windows of a long data sequence and observe convergence [8]. Note that this test can be applied even without knowing the true states, which is typically the case in practice. On the other hand it can be applied from artificial observations from a computed model trajectory using suitable assumptions about the measurement noise.

The basic windowing test is applied as follows. Given an observation sequence $S = (s_1, \dots, s_N)$ apply the shadowing filter to the length n subsequence $S_n = (s_{N-n+1}, \dots, s_N)$, for $2 \leq n \leq N$, to obtain state sequences $X_{m,n} = (x_{N-n+1,m,n}, \dots, x_{N,m,n})$. Now for increasing n compare corresponding states of $X_{m,n}$ to those of $X_{m,p}$ for $2 \leq p < n$, that is, compare the distances $\|x_{N-i,m,n} - x_{N-i,m,p}\|$, for $1 \leq i < p$ and $2 \leq p < n$. possible with numerical model data.

A typical outcome for $m = 100$, $\sigma = 0.1$ and $2\Delta/(n-1) = 0.1$ is shown in fig. 3. We started with a observation sequence $N = 50$ and used $p \in \{4, 6, 8, 10, 15, 20, 30\}$. Observe the convergence of X_m for example and note that $X_{m,15}$ and $X_{m,30}$ are almost identical for the last 10 states of the sequence, hence, to obtain convergence for a 10 point trajectory segment, $n = 15$ is sufficient. A typical application of the windowing test is to use it to optimise n for a state estimation that can be used as an initial state for forecasting. Hence we might be only interested in convergence for the last state of the sequence. From fig. 3 we conclude that $n \geq 8$ for forecasting purposes.

5. Conclusion

We have demonstrated that shadowing filters are useful to enhance the quality of state estimates. Our investigation shows that this leads to longer forecast times. The state estimations from shadowing filters are always on the attractor and therefore are better than the ones of the extended Kalman filter. In addition we introduced the windowing test. This simple test enabled us to optimise the parameters of the shadowing filter. Since the concept of shadowing filters and the windowing test can be easily applied to all kind of systems we hope that this guide will stimulate further applications.

We acknowledge the support from ARC, grant no. DP 0662841.

References

[1] P.L. Houtekamer, L. Lefaivre, J. Derone, H. Ritchie, and H.L. Mitchell. *Monthly Weather Review*, 124:1225–1242, 1996.

[2] K. Ikeda. *Optical Communications*, 30:257–261, 1979.

[3] K. Judd. *Physica D*, 183:273–281, 2003.

[4] K. Judd, C.A. Reynolds, and T. Rosmond. Technical Report NRL/MR/7530-04-18, Naval Research Laboratory, Monterey, CA, 93943-5502 USA., 2004.

[5] K. Judd and L.A. Smith. *Physica D*, 151:125–141, 2001, *Physica D*, 196:224–242, 2004. D. Ridout and K. Judd. *Physica D*, 165:27–48, 2002.

[6] K. Judd, L.A. Smith, and A. Weisheimer. *Physica D*, 190:153–166, 2004.

[7] L. A. Smith. Ensemble predictions and chaotic systems. Technical report, Mathematical Institute, University of Oxford, 1995.

[8] T. Stemler and K. Judd. A guide to using shadowing filters for forecasting and state estimation Preprint.