Mathematical Modelling and Numerical Analysis Will be set by the publisher
Modélisation Mathématique et Analyse Numérique

e-mail: `chorin@math.berkeley.edu` e-mail: `xuemin@math.berkeley.edu`

AN ITERATIVE IMPLEMENTATION OF THE IMPLICIT NONLINEAR FILTER

Alexandre J. Chorin[1] and Xuemin Tu[1]

Abstract. Implicit sampling is a sampling scheme for particle filters, designed to move particles one-by-one so that they remain in high-probability domains. We present a new derivation of implicit sampling, as well as a new iteration method for solving the resulting algebraic equations.

1991 Mathematics Subject Classification. 60G35, 62M20, 86A05.

The dates will be set by the publisher.

1. Introduction

There are many problems in science in which the state of a system must be identified from an uncertain equation supplemented by a stream of noisy data (see e.g. [7]). A natural model of this situation consists of an Ito stochastic differential equation (SDE):

$$dx = f(x,t)\,dt + g(x,t)\,dw, \tag{1}$$

where $x = (x_1, x_2, \dots, x_m)$ is an m-dimensional vector, w is m-dimensional Brownian motion, f is an m-dimensional vector function, and $g(x,t)$ is an m by m diagonal matrix. The initial state x^0 is assumed given and may be random as well. Furthermore, as the solution of the SDE unfolds, it is observed, and the values b^n of a measurement process are recorded at times $t^n, n = 1, 2, ...$ For simplicity assume $t^n = n\delta$, where δ is a fixed time interval. The measurements are related to the evolving state $x(t)$ by

$$b^n = h(x^n) + QW^n, \tag{2}$$

where h is a k-dimensional, generally nonlinear, vector function with $k \leq m$, Q is a k by k diagonal matrix, $x^n = x(n\delta)$, and W^n is a vector whose components are k independent Gaussian variables of mean zero and variance one, independent also of the Brownian motion in equation (1). The task is to estimate x on the basis of both equation (1) and the observations (2). We shall call the function $h = h(x^n)$ in equation (2) the "observation function". For general information about this problem and its solution, see e.g. [1–4, 7–9, 11–14].

The first thing we do in the present paper is approximate the SDE (1) by a difference scheme of the form

$$x^{n+1} = x^n + R(x^n, t^n)\delta + G(x^n, t^n)V^{n+1}, \tag{3}$$

where we assume temporarily that δ equals the interval between observations, i.e., we assume that there is an observation at every time step. x^n stands for $x(n\delta)$, G is assumed to be diagonal, and x^n, x^{n+1} are m-dimensional vectors. R, G determine the scheme used to solve the SDE, see for example [6]. V^{n+1} is a vector of $N(0,\delta)$ Gaussian variables, independent of each other for each n as well as of the noise W in equation (2).

Keywords and phrases: Implicit sampling, filter, reference density, Jacobian, iteration, particles

[1] Department of Mathematics, University of California at Berkeley and Lawrence Berkeley National Laboratory, Berkeley, CA, 94720

We assume here that the SDE is well-approximated by the difference scheme, in particular δ is small enough, and address exclusively this discrete problem. For simplicity, we assume that δ is fixed.

We now solve the discretized problem (3)-(2). If the system (1) and equation (2) are linear and the data are Gaussian, the solution can be found via the Kalman-Bucy filter (see e.g. [3]). In the general case, it is natural to try to estimate x via its evolving probability density.

This can be done by following "particles" (replicas of the system) whose empirical distribution approximates P_n. (see e.g [1,4,8–13]). Thus, at time $t^n = n\delta$ we have M particles X_i^n, $1 \leq i \leq M$, $n = 0, 1, \ldots$, whose empirical density approximates P_n, the probability density at time $n\delta$ of the particles that obey the discrete evolution equation (3) subject to the observations (2) at times $t = k\delta$ for $k \leq n$. The problem is to find new positions for the particles so that the empirical distribution they define approximates P_{n+1}. A standard construction (see e.g [1,4,8–13]) uses the probability density function (pdf) P_n and equation (1) to approximate a prior density, and then uses the new data b^{n+1} to approximate a posterior density P_{n+1} through weighting and resampling. This can be very expensive, as the number of particles needed can grow catastrophically (see e.g. [2,14]). The challenge is to generate high probability samples so as to minimize the number of particles.

In [6] we briefly introduced implicit sampling, a way to generate high probability samples, working with the particles one-by-one and pulling them to the high probability domain. Roughly speaking, the idea is to define a probability for each particle first, and then force the particle to assume this probability, at least approximately. In [6] we presented the idea briefly in the narrow context of a specific low-dimensional problem, with the resulting algebraic problem solved by a rotation-based algorithm. In the present paper we offer a general derivation, a brief discussion of general methods for doing the algebra, as well as an iteration algorithm that we found to be useful.

2. Implicit sampling

The conditional probability densities $P_n(x)$ at times t^n, determined by the discretized SDE (3) given the observations (2), satisfy the recurrence relation (see e.g. [7], p.6):

$$P(x^{n+1}) = P(x^n)P(x^{n+1}|x^n)P(b^{n+1}|x^{n+1})/Z, \tag{4}$$

where $P(x^n) = P(x^n|b^1, b^2, \ldots, b^n)$ is the probability density at time $n\delta$, $P(x^{n+1}|x^n)$ is the probability density of x^{n+1} given x^n as it is determined by the dynamics, $P(b^{n+1}|x^{n+1})$ is the probability of the next observation given the new position, as per the observation equation, and Z is a normalization constant.

To approximate this, we require that as a particle moves from X_i^n to X_i^{n+1}, the probability of the new position X_i^{n+1} satisfy

$$P(X_i^{n+1}) = P(X_i^n)P(X_i^{n+1}|X_i^n)P(b^{n+1}|X_i^{n+1})/Z_0, \tag{5}$$

with the obvious interpretation of the various terms. We shall soon see that one can set $P(X_i^n|b^n) = 1$ here without loss of generality. The initial state x^0 is known and so is its probability density; all one has to do is evaluate sequentially the density P_{n+1} of x^{n+1} given the probability density P_n of x^n and the new data b^{n+1}. This probability density can be sampled by picking a suitable "prior" density for the particles, for example $P(X_i^{n+1}|X_i^n)$, and then using the observations to weigh the samples and so produce a "posterior" density that satisfies the equations. As is well-known (see the references), the problem is that in most such weighting schemes, most particles end up by having very small weights and therefore very many particles are required. We now propose a sampling scheme that largely avoids this problem.

First, choose once and for all a fixed reference random variable, say ξ, with a given probability density function (pdf), say a Gaussian $\exp(-\xi^*\xi/2)/(2\pi)^{m/2}$), which one knows how to sample so that most samples have high probability. Here the star * denotes a transpose. The choice of a Gaussian here is convenient but not necessary, and in no way prejudges the nature of the density that is being sampled. We then make X_i^{n+1} a function of ξ, a different function of each particle and each step, each function designed so that the map

$\xi \to X_i^{n+1}$ connects highly probable values of ξ to highly probable values of X_i^{n+1}. To that end, write

$$P(X_i^{n+1}|X_i^n)P(b^{n+1}|X_i^{n+1}) = \exp(-F_i(X)),$$

where on the right-hand side X is a shorthand for X_i^{n+1} and all the other arguments are omitted. This defines a function F_i for each particle i and each time t^n. For each i and n, F_i is an explicitly known function of $X = X_i^{n+1}$. Then solve the equation

$$F_i(X) - \phi_i = \xi^*\xi/2, \tag{6}$$

where ξ is a sample of the fixed reference variable (a different sample for each particle) and $\phi_i = \min F_i$ is an additive factor needed to make the equation solvable. The need for ϕ_i becomes obvious if one considers the case of a linear observation function h in equation (2), so that the right side of equation (6) is quadratic but the left is a quadratic plus a constant and without ϕ_i the equation in general has no solution. The additive term ϕ_i has a second role: with our choice of reference density the most likely samples are near zero; the resulting X_i^{n+1} will then be near the minimum of F_i and therefore they will be high probability samples of X_i^{n+1}. We also require that for each particle, the function $X_i^{n+1} = X = X(\xi)$ defined by (6) be one-to-one and onto so that the correct pdf is sampled, in particular, it must have distinct branches for positive values and negative values of each component of ξ. From now on we omit the index i in both F and ϕ, but it should not be forgotten that these functions vary from particle to particle and from one time step to the next.

Once the function $X = X(\xi)$ is determined, each value of $X^{n+1} = X$ (the subscript i is omitted) appears with probability $\exp(-\xi^*\xi/2)J^{-1}/(2\pi)^{m/2}$, where J is the Jacobian of the map $X = X(\xi)$, while the product $P(X^{n+1}|X^n)P(b^{n+1}|X^{n+1})$ evaluated at X^{n+1} equals $\exp(-\xi^*\xi/2)\exp(-\phi)/(2\pi)^{m/2}$. The sampling weight for the particle is therefore $\exp(-\phi)J$. If the map $\xi \to X$ is smooth near $\xi = 0$, so that ϕ and J do not vary rapidly from particle to particle, and if there is an easy way to compute J, then we have an effective way to sample P_{n+1} given P_n. It is important to note that though the functions F and ϕ vary from particle to particle, the probabilities of the various samples are expressed in terms of the fixed reference pdf, so that they can be compared with each other. Picking ξ first is tantamount to (approximately) fixing the probability of the sample and then seeking a corresponding sample.

As usual, the weights can be eliminated by resampling. A standard simple resampling algorithm goes as follows [7]: let the weight of the i-th particle be $W_i, i = 1, \ldots, M$. Define $A = \sum W_i$; for each of M random numbers $\theta_k, k = 1, \ldots, M$ drawn from the uniform distribution on $[0, 1]$, choose a new $\widehat{X}_k^{n+1} = X_i^{n+1}$ such that $A^{-1}\sum_{j=1}^{i-1} W_j < \theta_k \le A^{-1}\sum_{j=1}^{i} W_j$, and then suppress the hat. This justifies the statement following equation (5) that one can set $P(X_n) = 1$.

In conclusion, provided we can efficiently solve equation (6) and evaluate the Jacobian J, we have solved the problem of finding high probability samples. No approximations are made other than approximating the SDE by a difference equation, representing a pdf by a bunch of particle, and performing algebra on a computer.

3. An iteration scheme for solving the algebraic equation

We now discuss how the algebraic equation (6), $F(X)-\phi = \xi^*\xi/2$, is to be solved. First, one should make sure it is solvable. If F is convex and smooth, there is no problem. If it is not (as may happen when the observation function is strongly nonlinear, for example cubic in X, and simultaneously the data have low probability, see the examples in [5]), one has to define a function F_0 to use in (6) that is convex, and then replace ϕ by $\phi - F + F_0$ (so there is no bias) while preserving the property that the map $\xi \to X$ maps the neighborhood of zero onto the neighborhood of the minimum of F. This is generally quite easy, see the examples in [5].

It is important to note here that equation (6) is underdetermined. It is a single equation that links $2M$ variables (the M components of ξ and the M components of X), with some additional not very restrictive conditions thrown in. This freedom can be used to find functions F_0 in the previous paragraph, and to formulate efficient solutions algorithms. Iterations of Newton type can be used both to find $\phi = \min F$ and then solve for X, so that there many possible algorithms. We present one that we found useful and used in the example below.

Its advantages are that it converges very fast when the observation function h is only moderately nonlinear (if h is linear it is not an iteration because it converges in one step), and that it finds ϕ and X simultaneously. Newton's method proceeds by successive linearization, and so does our algorithm, except that what is linearized is h rather than F. In our experience, this iteration converges whenever F is convex, but we have not been able to prove this. Our presentation presents some of the issues in all such iterations, in particular the need to make choices due to the non-uniqueness.

The function G in (3) does not depend on X^{n+1} for an Ito equation, and we assume for simplicity F does not depend on X^{n+1} either. Equation (3) states that $X^{n+1}-X^n$ is an $N(R(X^n,t^n)\delta,\delta G(X^n,t^n)^*G(X^n,t^n))$ vector. Consider the i-th particle and suppress temporarily the index i in the equation. Equation (6) can be written as:

$$\xi^*\xi/2 = F(X^{n+1}) - \min F(X^{n+1}), \tag{7}$$

where

$$F(X^{n+1}) = \left(X^{n+1}-X^n-R_n\right)^*\left(G_n^*G_n\right)^{-1}\left(X^{n+1}-X^n-R_n\right)/2+\left(h(X^{n+1})-b^{n+1}\right)^*\left(Q^*Q\right)^{-1}\left(h(X^{n+1})-b^{n+1}\right)/2,$$

$R_n = R(X^n,t^n)\delta$, and $G_n = \sqrt{\delta}G(X^n,t^n)$.

If h is linear, F is the sum of two quadratic forms; a completion of a square yields a single expression of the form $(X^{n+1}-\bar{m})^*\Sigma^{-1}(X^{n+1}-\bar{m})+\phi$, where Σ^{-1} is now a constant, symmetric, positive definite matrix, $\bar{m}$ is a constant vector, and ϕ is the constant remainder which is also the minimum of F and cancels out in the equation. As pointed out above, we now have a single equation for the M components of X, and a choice has to be made as to how to connect X to ξ. The choice we make is as follows: We perform a Choleski decomposition of Σ^{-1}: $\Sigma = LL^*$, where L is a lower triangular matrix and L^* is its transpose, and then set $X=\bar{m}+L\xi$. It should be noted that h is linear and there are data at every step, our sampling reduces to a particular implementation of sequential importance sampling, see e.g. [8].

In the general case, we proceed by iteration: we find a sequence of approximations X_j^{n+1} ($=X_j$ for brevity) which we hope converges to X^{n+1}; we set $X_0=0$, and now explain how to find X_{j+1} given X_j. First, expand the function h in the observation equation (2) in Taylor series around X_j:

$$h(X_{j+1}) = h(X_j) + H_j\cdot(X_{j+1}-X_j), \tag{8}$$

where H_j is a Jacobian matrix evaluated at X_j. The observation equation (2) can be approximated as:

$$z_j = H_jX_{j+1} + QW^{n+1}, \tag{9}$$

where $z_j = b^{n+1} - h(X_j) + H_jX_j$.

The right side of equation (7) is approximated as:

$$F_j(X_{j+1}) - \min F_j(X_{j+1}), \tag{10}$$

where

$$\begin{aligned} &F_j(X_{j+1}) = \\ &\left(X_{j+1}-X^n-R_n\right)^*\left(G_n^*G_n\right)^{-1}\left(X_{j+1}-X^n-R_n\right)/2+\left(H_jX_{j+1}-z_j\right)^*\left(Q^*Q\right)^{-1}\left(H_jX_{j+1}-z_j\right)/2 \\ =&\left(X_{j+1}-\bar{m}_j\right)^*\Sigma_j^{-1}\left(X_{j+1}-\bar{m}_j\right)/2+\Phi_j, \end{aligned} \tag{11}$$

$$\Sigma_j^{-1} = (G_n^*G_n)^{-1} + H_j^*(Q^*Q)^{-1}H_j, \quad \bar{m}_j = \Sigma_j\left((G_n^*G_n)^{-1}(X^n+R_n) + H_j^*(Q^*Q)^{-1}z_j\right),$$

and

$$K_j = H_jG_n^*G_nH_j^* + Q^*Q, \quad \Phi_j = \left(z_j - H_j(X^n+R_n)\right)^*K_j^{-1}\left(z_j-H_j(X^n+R_n)\right)/2.$$

Since $\min F_j(X_{j+1}) = \Phi_j$, we solve the equation

$$(X_{j+1} - \bar{m}_j)^* \Sigma_j^{-1} (X_{j+1} - \bar{m}_j)/2 = \xi^* \xi/2. \tag{12}$$

This is again a single equation for the M components of X_{j+1}, and we again connect X_{j+1} to ξ via a Choleski decomposition. The iteration is now fully described. Its complexity depends on the observation function h; the more arguments in h, the broader the band structure of Σ_j^{-1}. In the special but frequently encountered case where each component of h has a single argument, Σ_j^{-1} is diagonal.

Suppose the sequence X_j converges to a limit, call the limit X^{n+1}. One can readily check that the approximate equation (9) becomes the exact observation equation in the limit and that the sequence of Φ_j, each one a minimum of an approximate F, converges to the minimum of F.

We next compute the Jacobian determinant $J = \det(\partial X^{n+1}/\partial \xi)$. This can be often done analytically. Equation (12) relates X^{n+1} to ξ implicitly. We have values of ξ and the corresponding values of X^{n+1}; to find J there is no need to solve again for X^{n+1}; an implicit differentiation is all that is needed. Alternately, J can be found numerically, by taking nearby values of ξ, redoing the iteration (which should converge in one step, because one can start from the known value of X^{n+1}), and differencing.

Consider now a situation where we do not have observations at every time step. Assume for example that one has an observation at time $(n+1)\delta$ but not at time $n\delta$. We sample X^n and X^{n+1} together given the observation information at time step $(n+1)\delta$. Consider the i-th particle. Suppose we are given the vector X_i^{n-1} for that particle. Suppress again the particle index i. The joint probability density $P_{n,n+1}$ of X^n and X^{n+1} given X^{n-1} is

$$P_{n,n+1}(X^n, X^{n+1}) = \exp(-F_{n,n+1}(X^n, X^{n+1}))/Z,$$

where

$$\begin{aligned} &F_{n,n+1}(X^n, X^{n+1}) \\ &= \left(X^n - X^{n-1} - R_{n-1}\right)^* (G_{n-1}^* G_{n-1})^{-1} \left(X^n - X^{n-1} - R_{n-1}\right)/2 \\ &\quad + \left(X^{n+1} - X^n - R_n\right)^* (G_n^* G_n)^{-1} \left(X^{n+1} - X^n - R_n\right)/2 + \left(h(X^{n+1}) - b^{n+1}\right)^* (Q^*Q)^{-1} \left(h(X^{n+1}) - b^{n+1}\right)/2, \end{aligned} \tag{13}$$

and Z is the normalization constant. We recall that $R_{n-1} = R(X^{n-1}, t^{n-1})\delta$, $G_{n-1} = \sqrt{\delta} G(X^{n-1}, t^{n-1})$ are known from the approximation of the SDE, R_n and G_n depend on X^n.

In the now familiar sequence of steps, we pick two independent samples ξ_n and ξ_{n+1}, each with probability density $\exp(-\xi^*\xi/2)/(2\pi)^{m/2}$, and solve the equation

$$\xi_n{}^* \xi_n/2 + \xi_{n+1}^* \xi_{n+1}/2 = F_{n,n+1}(X^n, X^{n+1}) - \min F_{n,n+1}(X^n, X^{n+1})$$

to obtain X^n and X^{n+1} as functions of ξ_n and ξ_{n+1}. Note that now the equation is linear only if both h and the SDE are linear, so that an iteration is needed even if h is linear.

4. An example

We apply our filter to a prototypical marine ecosystem model studied in [10]. We set the main parameters equal to the ones in [10]; however, we will also present some results with a range of noise variances to make a particular point. We did the data assimilation with the filter described above, and also by the a standard particle filter SIR (Sampling importance resampling), see [1].

The model involves four state variables: phytoplankton P (microscopic plants), zooplankton Z (microscopic animals), nutrients N (dissolved inorganics), and detritus D (particulate organic non-living matter). At the initial time $t = 0$ we have $P(0) = 0.125$, $Z(0) = 0.00708$, $N(0) = 0.764$, and $D(0) = 0.136$. The system is described by the following nonlinear ordinary differential equations, explained in [10]:

$$\begin{aligned} dP &= \left(\frac{N}{0.2+N}\gamma P - 0.1P - 0.6\frac{P}{0.1+P}Z\right)dt + \sigma_P^2 dW_1, \\ dZ &= \left(0.18\frac{P}{0.1+P}Z - 0.1Z\right)dt + \sigma_Z^2 dW_2, \\ dN &= \left(0.1D + 0.24\frac{P}{0.1+P}Z - \gamma P\frac{N}{0.2+N} + 0.05Z\right)dt + \sigma_N^2 dW_3, \\ dD &= \left(-0.1D + 0.1P + 0.18\frac{P}{0.1+P}Z + 0.05Z\right)dt + \sigma_D^2 dW_4, \end{aligned} \tag{14}$$

where W_1, W_2, W_3, and W_4 are mutually independent Brownian motion with unit variance, and the parameter γ, the "growth rate", is determined by the equations given by

$$\gamma_t = 0.14 + 3\Delta\gamma_t, \quad \Delta\gamma_t = 0.9\Delta\gamma_{t-1} + N(0, \sigma_\gamma^2).$$

The variances of the noise terms are: $\sigma_P^2 = (0.01P(0))^2$, $\sigma_Z^2 = (0.01Z(0))^2$, $\sigma_N^2 = (0.01N(0))^2$, $\sigma_D^2 = (0.01D(0))^2$, and $\sigma_\gamma^2 = (0.01)^2$.

The observations were obtained from NASA's SeaWiFS satellite ocean color images. These observations provide a time series for phytoplankton; the relation between the observations $P(t)_{\text{obs}}$ (corresponding to the vector b^n in the earlier discussion) and the solution $P(t)$ of the equation of the first equation in (14) is assumed to be:

$$\log P(t)_{\text{obs}} = \log P(t) + N(0, \sigma_{\text{obs}}^2),$$

where $\sigma_{\text{obs}}^2 = 0.3^2$. Note that this observation equation is not linear. There are 190 data points distributed from late 1997 to mid 2002. The sample intervals ranged from a week to a month or more, for details see [10]. As in [10], we discretize the system (14) by an Euler method with $\Delta t = 1$ day and prohibit the state variables from dropping below 1 percent of their initial values.

We have compared our filter and SIR in three sets of numerical experiments, all with the same initial values as listed above. In each case we attempted to find a trajectory of the system consistent with the fixed data, and observed how well we succeeded. In the first set of the experiments, we used 100 particles and take $\sigma_P^2 = (0.01P(0))^2$ as in [10]. In this case, the (assumed) variance of the system is much smaller than the (assumed) variance of the observations; the particle paths are bunched close together, and the results from our filter and from SIR are quite close, see Figure 1, where we plotted the P component of the reconstructed solution as well as the corresponding data.

In the second set of the experiments, we still used 100 particle but assumed $\sigma_p^2 = (P(0))^2$. The variance of the system is now comparable to the variance of the observation. For SIR, after resampling, the number of distinct particles is smaller than in the first case, as a result of the loss of diversity after resampling when the weights are very different from each other, see Table 1, where we exhibit the average number of distinct particles left after each resample; there is a resample after each step. Remember that there is some loss of diversity in resampling even if all the weights are equal. With 100 particles, the filtered results with SIR are still comparable to those with our filter. See Figure 2.

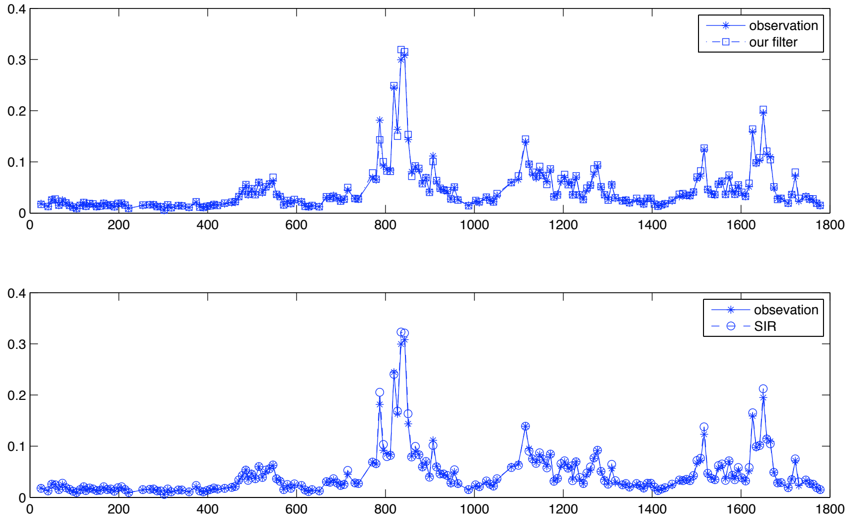
In the third set of the experiments, we used only 10 particles and kept $\sigma_p^2 = (P(0))^2$. As one could have foreseen, our filter does better than SIR, see Figure 3. One should remember however that we are working with a low dimensional problem where the differences between filters are not expected to be very significant; the cost if 100 particles is not prohibitive.

Several other, simpler, examples have been presented in [5].

Table 1. The number of distinct particles after resampling with different system variances and different numbers of particles

σ_p	# particle	average # particles left after resampling	
		SIR	Our filter
$0.01P(0)$	100	61	61
$P(0)$	100	19	63
$P(0)$	10	2.2	6.3

Figure 1. Results with $\sigma_P^2 = (0.01P(0))^2$ and 100 particles



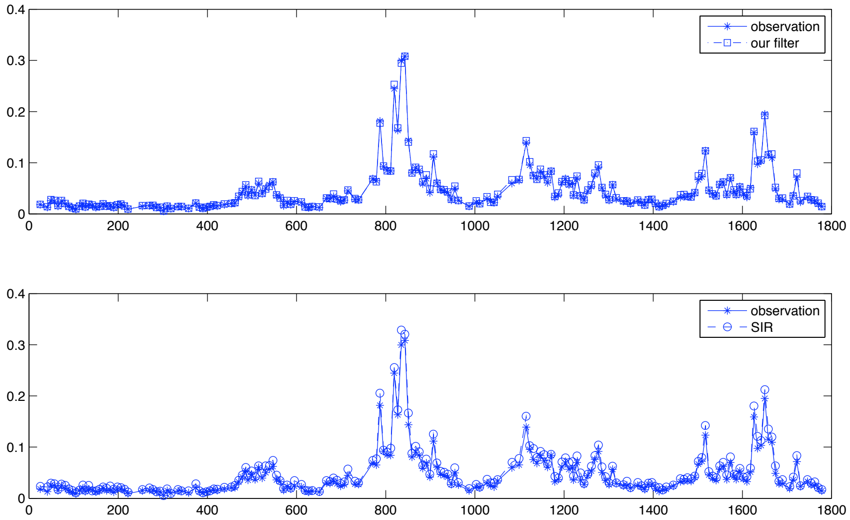
5. Conclusions

We have presented a new derivation of implicit sampling and an iteration process used in our new implicit nonlinear particle filter. The goal is to aim particle paths sharply so that fewer are needed. We conjecture that there is no general way to reduce the variability of the weights in particle sampling further than we have. We also presented a simple example that illustrates the potential of this new sampling. The example is simple in that it is low-dimensional. High-dimensional nonlinear examples are currently in the works.

6. Acknowledgments

We would like to thank Prof. J. Goodman, who urged us to write a more general version of our previous paper and suggested some notations and nomenclature, Profs. R. Miller and Y. Spitz, who suggested that we try Dowd's model plankton problem as a first step toward an ambitious joint effort and helped us set it up, Prof. M. Dowd, who kindly made the data available, and Mr. M. Morzfeld, who made many helpful suggestions. This work was supported in part by the Director, Office of Science, Computational and Technology Research, U.S. Department of Energy under Contract No. DE-AC02-05CH11231, and by the National Science Foundation under grants DMS-0705910 and OCE-0934298.

Figure 2. Results with $\sigma_P^2 = P(0)^2$ and 100 particles



References

[1] M. Arulampalam, S. Maskell, N. Gordon, and T. Clapp. A tutorial on particle filters for online nonlinear/nongaussian Bayesia tracking. *IEEE Trans. Sig. Proc.*, 50:174–188, 2002.

[2] P. Bickel, B. Li, and T. Bengtsson. Sharp failure rates for the bootstrap particle filter in high dimensions. *IMS Collections: Pushing the Limits of Contemporary Statistics: Contributions in Honor of Jayanta K. Ghosh*, 3:318–329, 2008.

[3] S. Bozic. *Digital and Kalman Filtering*. Butterworth-Heinemann, Oxford, 1994.

[4] A.J. Chorin and P. Krause. Dimensional reduction for a Bayesian filter. *Proc. Nat. Acad. Sci. USA*, 101:15013–15017, 2004.

[5] A.J. Chorin, M. Morzfeld, and X. Tu. Implicit sampling for data assimilation. June, 2010. submitted for publication.

[6] A.J. Chorin and X. Tu. Implicit sampling for particle filters. *Proc. Nat. Acad. Sc. USA*, 106:17249–17254, 2009.

[7] A. Doucet, N. de Freitas, and N. Gordon. *Sequential Monte Carlo Methods in Practice*. Springer, New York, 2001.

[8] A. Doucet, S. Godsill, and C. Andrieu. On sequential Monte Carlo sampling methods for Bayesian filtering. *Stat. Comp.*, 10:197–208, 2000.

[9] A. Doucet and A. Johansen. Particle filtering and smoothing: Fifteen years later. *Handbook of Nonlinear Filtering (eds. D. Crisan et B. Rozovsky)*, to appear.

[10] M. Dowd. A sequential Monte Carlo approach for marine ecological prediction. *Environmetrics*, 17:435–455, 2006.

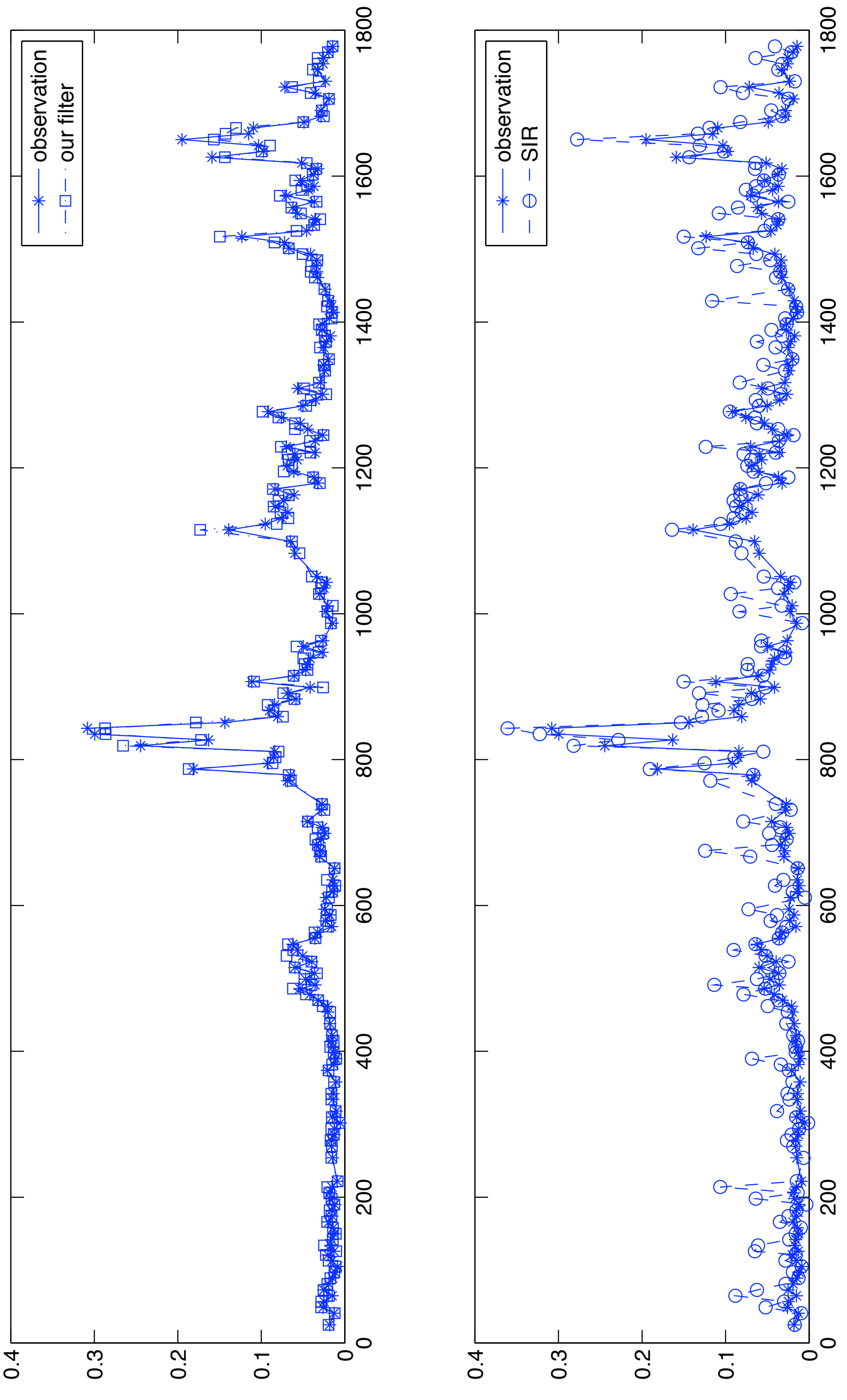
[11] W. Gilks and C. Berzuini. Following a moving target -Monte Carlo inference for dynamic Bayesian models. *J. Roy. Statist. Soc. B*, 63:127–146, 2001.

[12] J. Liu and C. Sabatti. Generalized Gibbs sampler and multigrid Monte Carlo for Bayesian computation. *Biometrika*, 87:353–369, 2000.

[13] S. Maceachern, M. Clyde, and J. Liu. Sequential importance sampling for nonparametric Bayes models: the next generation. *Can. J. Stat.*, 27:251–267, 1999.

[14] C. Snyder, T. Bengtsson, P. Bickel, and J. Anderson. Obstacles to high-dimensional particle filtering. *Mon. Wea. Rev.*, 136:4629–4640, 2008.

Figure 3. Results with $\sigma_P^2 = P(0)^2$ and 10 particles


observation
our filter
observation
SIR