Implicit Sampling with Application to Filtering and Data Assimilation

Ethan Atkins*°, Alexandre J. Chorin*°, Matthias Morzfeld°, Xuemin Tu‡

° Lawrence Berkeley National Laboratory
‡ Department of Mathematics, University of Kansas
* Department of Mathematics, University of California at Berkeley

Introduction

The task in data assimilation is to use available data to update the forecast of a numerical model. The numerical model is typically given by a discretization of a stochastic differential equation (SDE)

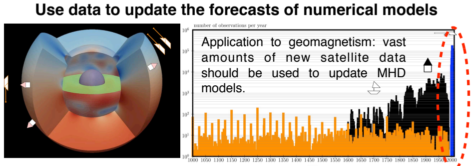
$$x^{n+1} = R(x^n, t^n) + G(x^n, t^n)\Delta W^{n+1}$$

where x is an m-dimensional vector, called the state, t^n, $n = 0,1,2,...$ is a sequence of times, R is an n-dimensional vector function, G is an $m \times m$ matrix and dW is an m-dimensional vector, whose elements are independent standard normal variates. The random vectors $G(x^n,t^n)\Delta W$ represent the uncertainty in the system. The data

$$z^l = h(x^{q(l)}, t^{q(l)}) + Q(x^{q(l)}, t^{q(l)})V^l$$

are collected at times $t^{q(l)}$, $l = 0,1,2,...$; for simplicity, we assume that the data are collected at a subset of the model steps, i.e. $q(l) = rl$, with r being a constant. In the above equation, z is a k-dimensional vector ($k \leq m$), h is a k-dimensional vector function, V is a k-dimensional vector whose components are independent standard normal variates, and Q is a $k \times k$ matrix.

Use data to update the forecasts of numerical models



Particle filters are sequential Monte Carlo methods for data assimilation [1] which approximate the probability density function (pdf) of the state given the observations, $p(x^{0:q(l)} \mid z^{1:l})$. The state estimate is a statistic (e.g. the mean, median, mode etc.) of this pdf. Most particle filters rely on the recursive relation

$$p(x^{0:q(l+1)} \mid z^{1:l+1}) \propto p(x^{0:q(l)} \mid z^{1:l})\, p(z^{l+1} \mid x^{q(l+1)})\, p(x^{q(l)+1:q(l+1)} \mid x^{q(l)})$$

In the above equation $p(x^{0:q(l+1)} \mid z^{1:l+1})$ is the pdf of the state trajectory up to time $t^{q(l+1)}$ given all available observations and is called the target density; $p(z^{l+1} \mid x^{q(l+1)})$ is the probability density of the current observation given the current state and can be obtained from the observation equations. The pdf $p(x^{q(l)+1:q(l+1)} \mid x^{q(l)})$ is the density of the state trajectory from the previous assimilation step to the current observation, conditioned on the state at the previous assimilation step, and is determined by the model equations.

Implicit Sampling

The basic idea of implicit sampling [2,3,4] is to use the available observations to find regions of high probability in the target density and look for samples within this region. This implicit sampling strategy generates a thin particle beam within the high probability domain and, thus, keeps the number of particles required manageable, even if the state dimension is large.

Assume we are given a collection of M particles X_j, whose empirical density approximates the conditional density and suppose the next observation is available the after r time steps. Using Bayes' theorem, one can show that

$$p(X_j) \propto p(z^{l+1} \mid X_j^{q(l+1)})\, p(X_j^{q(l)+1:q(l+1)} \mid X_j^{q(l)})$$

is the pdf of the j^{th} particle given its previous state and the available observations. Implicit sampling is a recipe for obtaining high probability samples of this pdf. To draw a sample we define, for each particle, a function F_j by

$$F_j(X_j) = -\log p(X_j)$$

With this F_j we solve the equation

$$F_j(X_j) - \phi_j = \frac{1}{2}\xi_j^T\xi_j$$

where ξ_j is a realization of the reference variable $\xi \sim N(0, I)$ and where $\phi_j = \min F_j$.

Implicit Sampling (ctd.)

The minimum of F_j can often be obtained by standard methods (e.g. Newton–Raphson, gradient descent, or trust-region methods). The reference variable ξ is known and easy to sample, and by definition most of its samples will be high-probability samples near the origin. The corresponding values of F_j will be near the minimum ϕ_j and therefore will have a high probability, so that with high probability we will have high probability samples.

The empirical density defined by the new particle positions differs from the target density so that each sample must be weighted by the ratio of its probability with respect to the target density to its proposal probability [1]. The weight for each particle is $w_j = \exp(-\phi_j)J$, where J is the Jacobian of the map $x \to \xi$. Here we choose this map to be random:

$$X_j = \mu_j + \lambda_j L_j \eta_j$$

where $F_j(\mu_j) = \phi_j$ and $\eta_j = \xi_j / |\xi_j|$ is uniformly distributed on the unit sphere. The invertible square matrix L_j is deterministic, under our control, and remains to be chosen. We compute λ_j by substitution and obtain a single equation with the single unknown λ_j. A data assimilation problem of arbitrary dimension boils down to the solution of a single algebraic equation per particle. For its solution, we can again use standard methods (e.g. Newton–Raphson).

It remains to compute the Jacobian. After some algebra, it can be shown that

$$J = \det(L_j)\rho_j^{1-rm/2}\left|\lambda_j^{rm-1}\frac{\partial\lambda_j}{\partial\rho_j}\right|$$

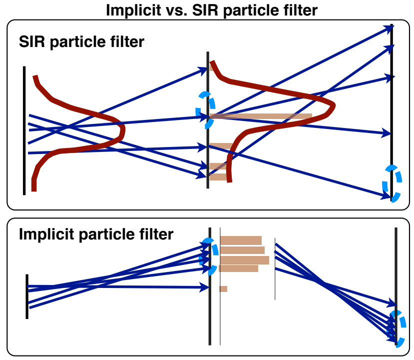
where $\rho_j = \xi_j^T\xi_j$ and $\det(A)$ denotes the determinant of the square matrix A. The scalar derivative can be computed by

$$\frac{\partial\lambda_j}{\partial\rho_j} = \frac{\rho}{2(\nabla F_j)L_j^T\eta_j}$$

Implicit Sampling vs. SIR

Traditional Bayesian particle filters [1] (Sampling-Importance-Resampling (SIR)), follow replicas of the model (called particles) to construct a prior probability density function. The prior is updated by sampling weights determined by the observations to yield a posterior density, to approximate the pdf of the state, conditioned on the observations.

Implicit vs. SIR particle filter



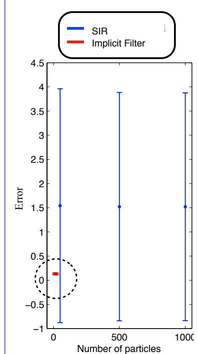
If the dimension of the model is large, particles are likely to stray into regions of low probability and the number of particles required can grow catastrophically.

The standard procedure is reversed in the implicit particle filter. Rather than first generating a sample and then computing its probability, it first picks a probability and then looks for a sample that carries it, taking the observations into account in that search. It generates a thin particle beam, sharply focussed towards the observations. This focussing effect makes the number of required particles manageable.

Application to the SKS Equation

We test our implementation of the implicit filter by applying it to the stochastic Kuramoto–Sivashinsky (SKS) equation of combustion theory [5,6]

$$u_t + uu_x + u_{xx} + \nu u_{xxxx} = W(x,t)$$



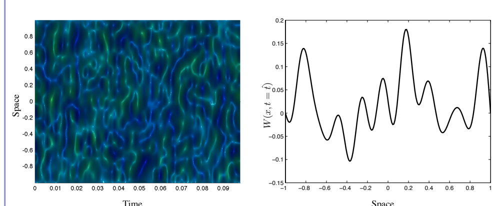
on the strip $x \in [0,L]$, $t \geq 0$ with L-periodic boundary conditions. In the SKS equation, ν denotes the viscosity and $W(x,t)$ is a space-time white noise process. Projection of the stochastic partial differential equation (SPDE) into an N-dimensional subspace spanned by N Fourier modes yields an Itô–Galerkin approximation to the SKS equation in the form of an stochastic differential equation. We require 512 Fourier modes when representing the solution and use the exponential Euler scheme [7] with time step $\delta = 0.001$ for time discretization. We consider nonlinear observations in physical space ($h(x) = x+x^3$) available at every model step. The results of 500 twin experiments are shown on the left. We observe that the implicit filter requires only about 10 particles to yield estimates with an accuracy for which SIR requires thousands of particles. The experiments confirm that the desired focussing effect of the particles can indeed be achieved. Further experiments with the SKS equation can be found in [4].

Geomagnetic Data Assimilation

Data assimilation has been recently applied to geomagnetic applications and there is a need to find out which data assimilation technique is most suitable [8]. We apply the implicit particle filter to a test problem very similar to the one first introduced by Fournier and his colleagues in [9]. The model is given by two SDE's

$$\partial_t u + u\partial_x u = b\partial_x b + \nu\partial_x^2 u + g_u\partial_t W(x,t)$$
$$\partial_t b + u\partial_x b = b\partial_x u + \partial_x^2 b + g_b\partial_t W(x,t)$$

where, g_u, g_b are scalars, and where W is a stochastic process. Physically, u represents the velocity field and b represents the secular variation of the magnetic field. We study the above equations with $\nu = 10^{-3}$ as in [9], and with $g_u = 0.01$, $g_b = 1$. The noise process W is spatially smooth.

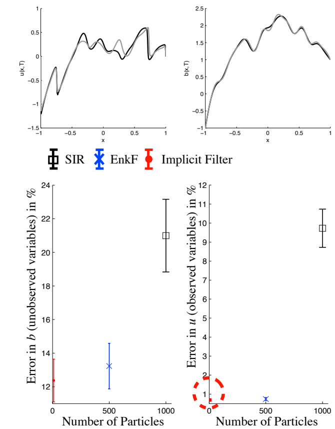


We discretize the equations using Legendre spectral elements of order 300 and an implicit-explicit scheme with time step $\delta = 0.02$ for time discretization. The data are the values of the magnetic field b, measured at k equally spaced locations.

We can exploit the smoothness property of the noise to reduce the dimension of the implicit filter by focussing attention on the variables driven by the largest noise [10]. In this application, the dimension of the filter could be reduced from 600 to 50. The minimization required for implicit sampling was carried out using a gradient descent method with line-search. The minimization was initialized by a free model run and converged typically in 10 steps.

Geomagnetic Data Assimilation (ctd.)

To assess the performance of the filters, we ran 100 twin experiments. A twin experiment amounts to: (*a*) drawing a sample from the initial state and running the model forward in time until $T = 0.2$ (one fifth of a magnetic diffusion time) (*b*) collecting the data from this free model run; and (*c*) using the data as the input to a filter and reconstructing the state trajectory.



We ran experiments varying the availability of data in time and space and observed that the implicit filter performs well with very few (~10) particles while competing methods, such as SIR or the ensemble Kalman filter (EnKF) required significantly more particles for similar accuracy.

Conclusions

We have presented implicit sampling for sequential data assimilation. The implicit filter requires, for each particle, the minimization of a known, real, scalar function and the solution of an underdetermined scalar equation. Using random maps, a solution of this equation can be found by solving one equation with one variable. The numerical experiments with the SKS equation and the geomagnetic model suggest that the implicit filter works well in large dimensional problems. The implicit filter outperforms SIR and EnKF in all cases considered.

References

1. A. Doucet, N. de Freitas, N. Gordon, eds., *Sequential Monte Carlo Methods in Practice*, (2001), Springer, New York.

2. A. J. Chorin, X. Tu, *Implicit sampling for particle filters*, PNAS **106** (2009), no. 41, pp. 17249–17254.

3. A. J. Chorin, M. Morzfeld, X. Tu, *Implicit particle filters for data assimilation*, CAMCoS 5 (2010), no. 2, pp. 221–240.

4. M. Morzfeld, X. Tu, E. Atkins, A. J. Chorin, *A random map implementation of implicit filters*, JCP (2011), accepted for publication.

5. Y. Kuramoto, T. Tsuzuki, *On the formation of dissipative structures in reaction-diffusion systems*, Progr. Theoret. Phys. **54** (1975), pp. 687–699.

6. G. Sivashinsky, *Nonlinear analysis of hydrodynamic instability in laminar flames, Part I. Derivation of basic equations*, Acta Astronaut **4** (1977), pp. 1177–1206.

7. A. Jentzen, P. E. Klöden, *Overcoming the order barrier in the numerical approximation of stochastic partial differential equations with additive space-time noise*, Proc. R. Soc. A **465** (2009), pp. 649–667.

8. A. Fournier, G. Hulot, D. Jault, W. Kuang, A. Tangborn, N. Gillet, E. Canet, J. Aubert, F. Lhuillier, *An introduction to data assimilation and predictability in geomagnetism*, Space Sci. Rev. 155 (1-4) (2010), pp. 247–291.

9. A. Fournier. C. Eymin, T. Alboussiere, *A case for variational geomagnetic data assimilation: insights from a one-dimensional, nonlinear, and sparsely observed MHD system*, Nonlin. Proc. Geophys. 14 (2007), pp. 163–180.

10. M. Morzfeld, A. J. Chorin, Implicit particles filtering for models with partial noise, and an application to geomagnetic data assimilation, Nonlin. Proc. Geophys. (2011), submitted for publication.

This work was supported by the Director, Office of Science, Computational and Technology Research, U.S. Department of Energy under Contract No. DE-AC02-05CH11231. Email enquiries to mmo@math.lbl.gov.