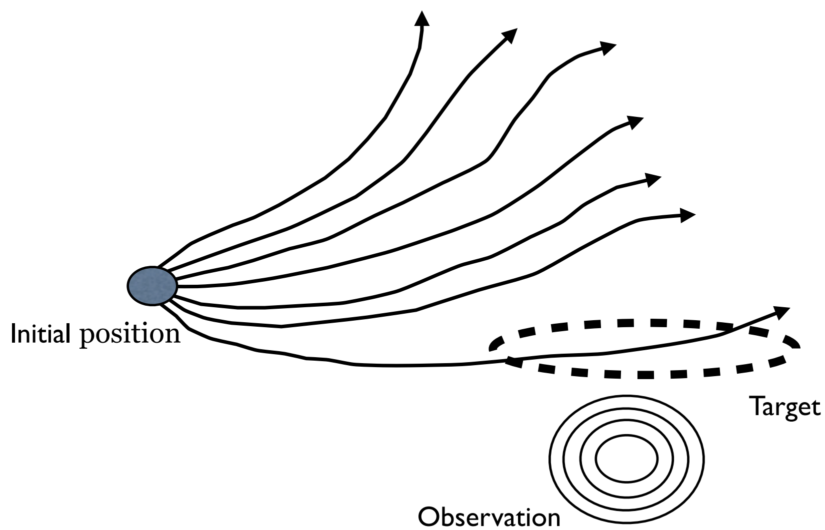
Implicit sampling for particle filters

Alexandre Chorin, Mathias Morzfeld, Xuemin Tu, Ethan Atkins
University of California at Berkeley

Example: Try to find people in a boat in the middle of the ocean knowing approximately their initial position, knowing approximately the directions of the currents and the winds, and having an uncertain observations of a later position.



Equations of motion :

$$dx = f(x,t)dt + gdW,$$

or, in a discrete approximation,

$$x^{n+1} = x^n + \delta f(x^n) + GW,$$

(from now focus on the discrete approximation).

Observations: $b^{n+1} = h(x^{n+1}) + V$, V random.

Special case: equations linear, pdf Gaussian, $\longrightarrow$ Kalman filter.

Extensions: extended Kalman filter, ensemble Kalman filter, try to fit a non-Gaussian situation into a Gaussian framework.

(more on ensemble Kalman filter later)

Simple particle filter:

Follow a bunch of ”particles” (samples, replicas) whose empirical density at time $t = n\delta$ approximates a pdf P_n determined by the equations of motion and conditioned by the observations.

Given P_n:

First evolve the particles by the equations of motion alone: (generates a “prior” density).

Take the observations into account by weighting the particles. (generates a “posterior” density).

Fails, in particular when there are many variables.

To avoid following irrelevant particles, resample, so that you have again a bunch of particles with equal weights. Given weights W_k, pick $\theta_k \sim [0,1]$, and set $\widehat{x}_k^{n+1} = x_j^{n+1}$ for j such that $A^{-1}\sum_1^{i-1} W_j < \theta_k < A^{-1}\sum_1^{i} W_j$, where $A = \sum A_k$.

Bayes theorem:

$$P(n+1|n+1) = P(n|n)P(x^{n+1}|x^n)P(b^{n+1}|x^{n+1})/Z.$$

where

$$P(n|n) = P(x^1, x^2, \cdots, x^n|b^1, b^2, \cdots, b^n).$$

Usual remedy: better choice of prior.

Q=importance density; sample Q, weight $P(X^{n+1}|X^n, b^{n+1})/Q$.
Problem: Q may be hard to find.

Our remedy: implicit sampling.

General idea:

Rather than find samples and then estimate their probability, first pick a probability and then find a sample that carries it.

Given P_n, sample P_{n+1} as follows:

1. Pick a sample ξ from a known, fixed, pdf, e.g. a Gaussian $\exp(-\xi^T\xi/2)/(2\pi)^{m/2}$.

2. Write the (unnormalized) pdf seen by the i-th particle at step n as $P(b^{n+1}|X_i^{n+1})P(X_i^{n+1}|X^n)$ in the form $\exp(-F(X))$, where $X = X_i^{n+1}$ and $F = F_{i,n}$.

3. Solve $F(X) - \min F = \xi^T\xi/2$ (F varies from particle to particle and at each time). This yields high probability samples.

The right pdf is sampled if map $\xi \to X$ is one-to-one and onto.

Each value of X appears with a probability $\exp(-\xi^T\xi/2)/(2\pi)^{m/2}$ divided by J, the Jacobian of the map $\xi \rightarrow X$. The probability we want to sample is $\exp(-\xi^T\xi/2)/(2\pi)^{m/2}$ multiplied by $\exp(-\min F)$ The sampling weight is $\exp(-\min F)J$.

What has been gained: Each sample requires the solution of an algebraic equation for the given particle, not a global estimate of the whole pdf. The samples are stitched together into a global pdf by the common reference pdf. The "prior" is represented as an infinite collections of functions of a fixed Gaussian, a separate function for each particle and step. The samples have high probability.

Solution of $F(X) - \min F = \xi^T \xi/2$ (and evaluation of J):

(Note equation is underdetermined).

Special case: h is linear and there are data at each step. The basic equation reduces to $(X-a)^T A (X-a) = \xi^T \xi$, where A is symmetric postive definite. This is a single linear equation connecting $2M$ variables. Do Choleski: $A = LL^T$, and solve $L^T(X-a) = \xi^T$. J can be read from L.

F convex:

Find $\min F$. Solve the equation $F(X) - \min F = \xi^T \xi/2$ for each ξ. Often quite easy, specially when h is sparse. Obtain J either numerically or by implicit differentiation.

Note again the equation $F(X) - \min F = \xi^T \xi/2$ is underdetermined; any map $\xi \to X$ that solves it will do, provided (i) it is one-to-one (with probability 1), (ii) it is smooth at the origin, (iii) the Jacobian J is easy to compute.

Examples of the use of this freedom:

Suppose you can minimize F by Newton's method (F is generally sparse, and one always has a good starting guess). This yields a Hessian matrix H at the minimum r. Construct

$$F_0(X) = \min F + (1/2)(X - r)^T H(X - r).$$

Solve $F_0(X) - \min F = \xi^T \xi/2$, (this is linear !). The weighting is $\exp(-\phi_0)J$, where

$$\phi_0 = \min F - F_0(X) + F(X).$$

The neighborhood of $\xi = 0$ is still mapped on the high probability region of X, and

$$F(X) - \min F = F_0(X) - \phi_0(X),$$

so there is no bias.

More implementations:

Make the ansatz: $X_j = \mu_j + \lambda_j L_j^T \eta_j$, where μ_j is the minimum of F: $F(\mu_j) = \phi_j$, λ_j is a random unit vector: $\lambda_j = \xi_j/|\xi|$, and L is a suitable matrix. J is found from a 1D computation. Best choice of L: $LL^t = H^{-1}$, where H is the Hessian available from the minimization of F.

More on ensemble Kalman filter:

In the ensemble Kalman filter one solves the Fokker-Planck equation for the evolution of a pdf under the SDE, one extracts an approximate mean and variance, and one does a Kalman step as if the system were linear.

Implicit sampling is equivalent to solving the Zakkai equation (evolution under both the SDE and the observations). Not harder thna solving the FP equation, linearization is no longer needed.

Sparse data in time: Same devices, more dimensions:

If data available every p steps, solve

$$F\left(X^{n+1},\ldots,X^{n+p}) - \min F(X^{n+1},\ldots,X^{n+p})\right) = \xi^T\xi/2.$$

Higher-order accurate approximations to the SDE:

Suppose you are solving $dx = f(x)dt + dW$ by Klauder-Petersen:

$$\begin{aligned} x^{n+1,*} &= x^n + \delta f(x^n) + \eta_1, & (1)\\ x^{n+1} &= x^n + (\delta/2)\left(f(x^n) + f(x^{n+1,*})\right) + \eta_2, & (2)\\ b^{n+1} &= h(x^{n+1}) + \eta_3. & (3) \end{aligned}$$

The probability of $(X^{n+1,*}, X^{n+1})$ is $\exp(-F)$ where

$$F = \left(X^{n+1,*} - X^n - \delta f(X^n)\right)^2 / (2\delta) +$$

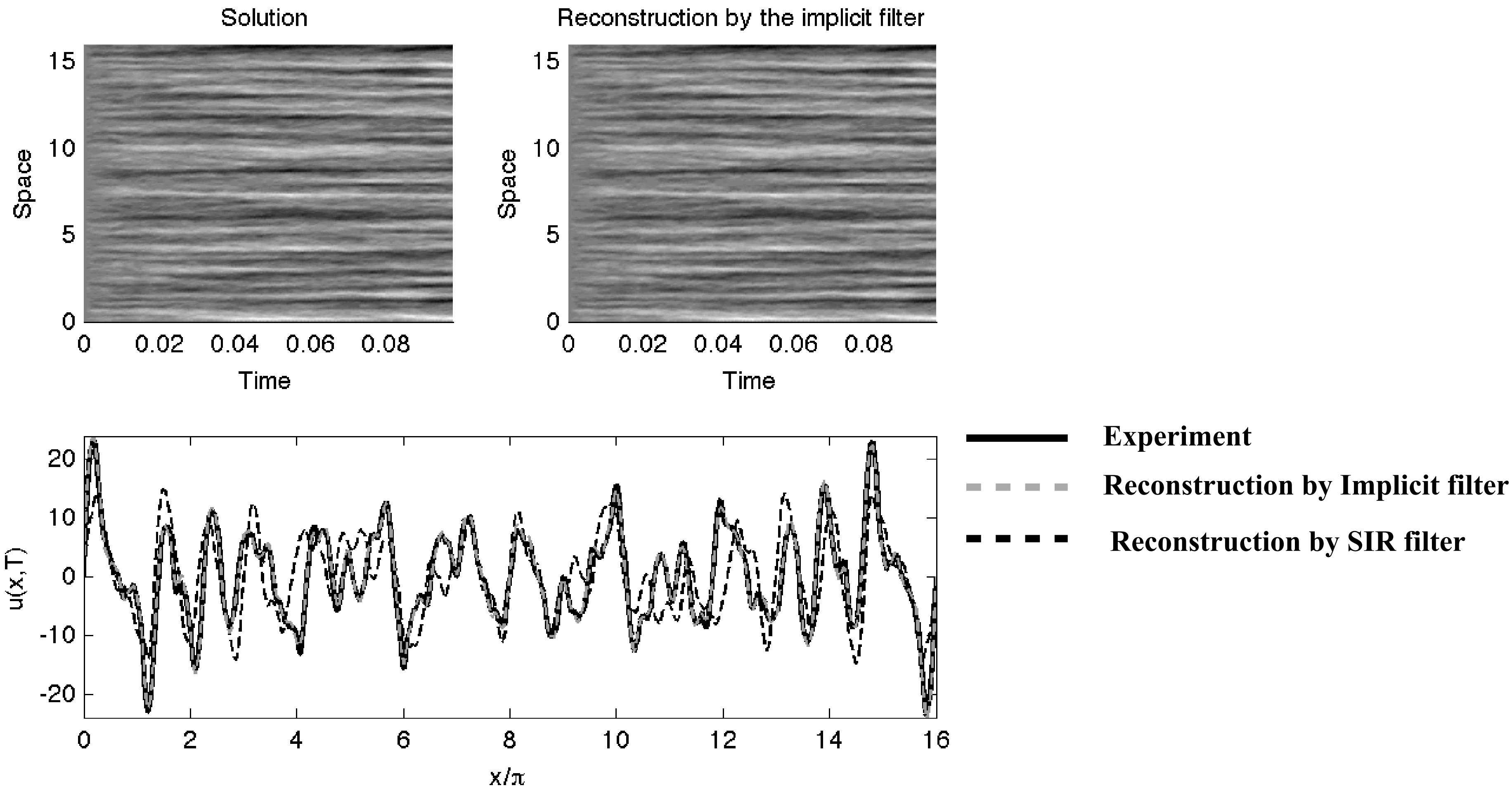
$$\left(X^{n+1} - X^n - (\delta/2)(f(X^n) + f(X^{n+1,*})\right)^2 / (2\delta)$$

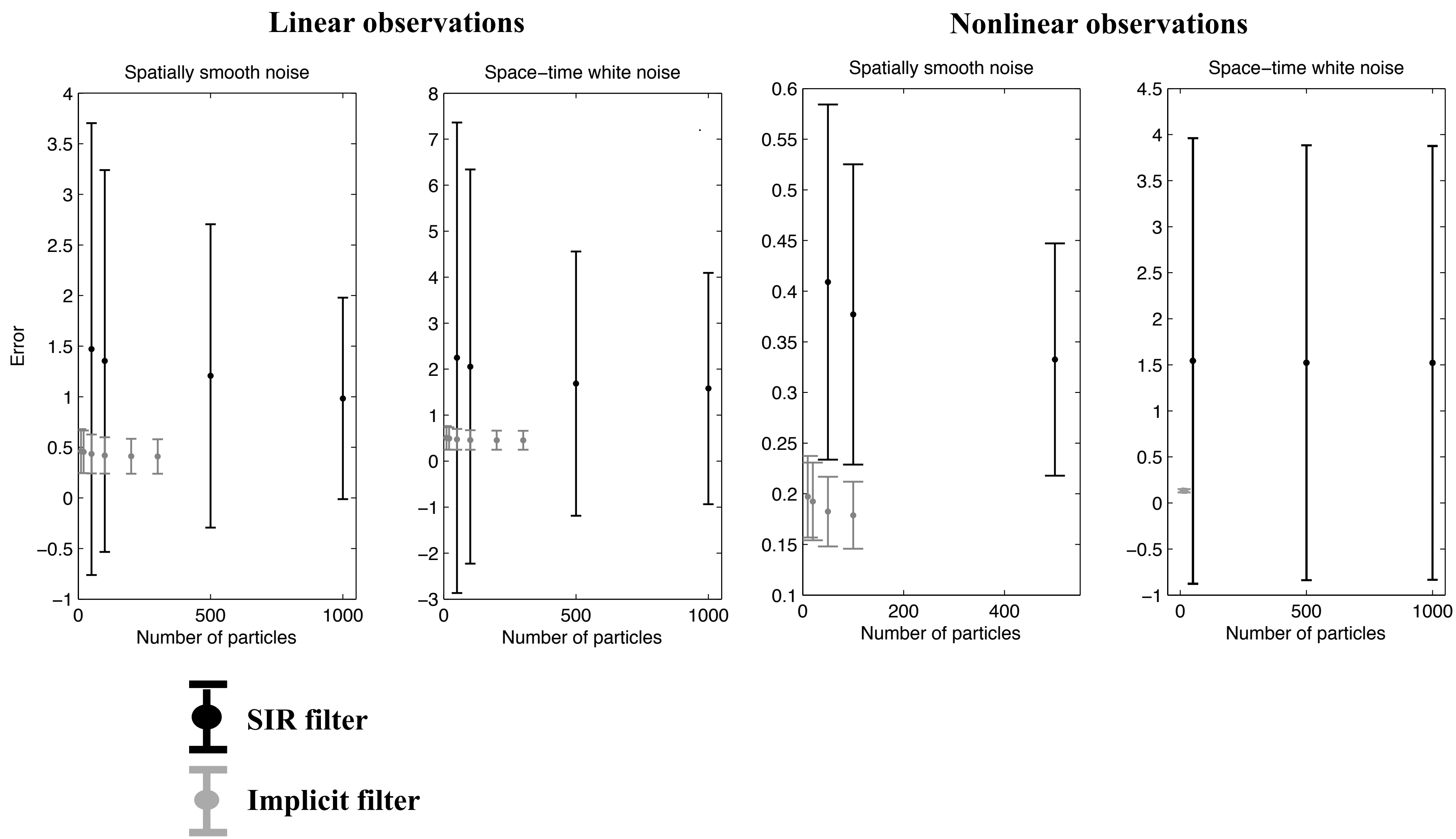
$$\left(h(X^{n+1}) - b)\right)^2 / (2s).$$

Again solve $F - \min F = \xi^T \xi / 2$.

Example: Stochastic Kuramoto-Sivashinski equation $u_t + uu_x + u_{xx} + \nu uu_{xxxx} = gW(x,t)$, W = space-time white noise. Data such that there are 31 linearly unstable modes. Solved by Fourier expansion with 1024 modes, sparse observations of u (!) in space. Comparison of mean norm of error and variance of error with those of SIR.

Sparse linear observations




Linear observations
Nonlinear observations
Spatially smooth noise
Space−time white noise
Spatially smooth noise
Space−time white noise
Error
Number of particles
Number of particles
Number of particles
Number of particles
SIR filter
Implicit filter