Implicit sampling for particle filters

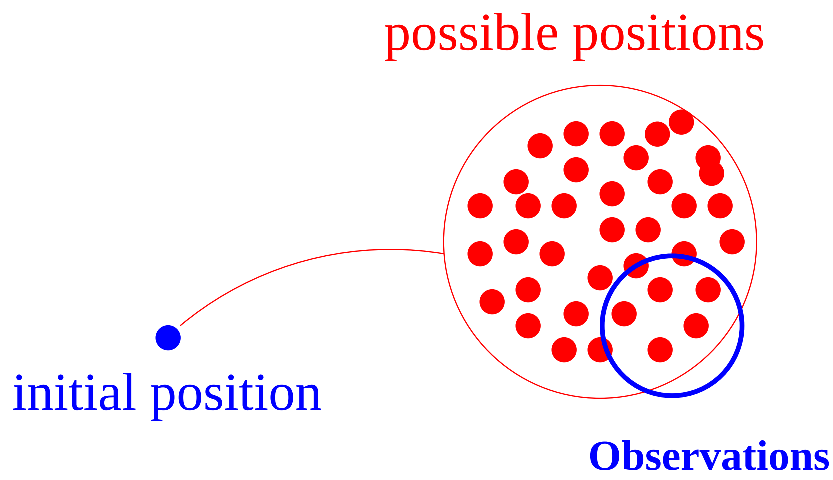
Alexandre Chorin, Xuemin Tu, Mathias Morzfeld
University of California at Berkeley

Example:

Try to rescue people in a boat in the middle of the ocean from approximate knowledge of their initial position and from uncertain observations of a later position.

Equations of motion : $dx = f(x,\omega)dt + gdW$, or, in a discrete approximation, $x_{n+1} = x_n + \delta F(x_n) + GdW$,

Observations: $b^n = b(n\delta) = g(x^n) + V$, 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 moved by the equations of motion constrained by the observations.

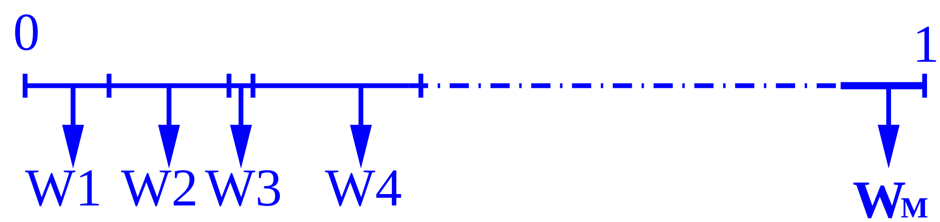
Given P_n:

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).

To avoid following irrelevant particles, resample, so that you have again a bunch of particles with equal weights. (for $\theta_k \sim [0,1]$, pick $\widehat{x}^{n+1} = x^{n+1}$ such that

$A^{-1}\sum_1^{i-1} W_j < \theta_k < A^{-1}\sum_1^{i} W_j$, where $A = \sum A_j$.

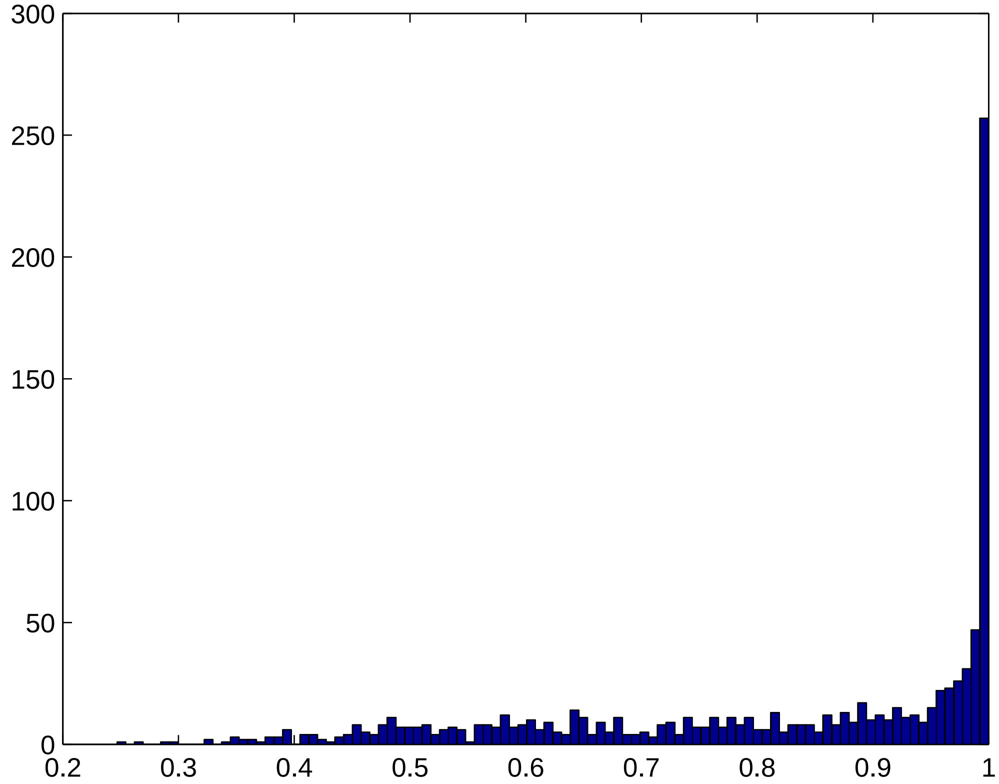


Final (important) step: Go back, smooth the past, and resample.

Bayes theorem:

$$P(x^{n+1}|x^n, b^{n+1}) = \frac{P(x^{n+1}, b^n) P(b^{n+1}|x^{n+1})}{P(b^{n+1}|b^n)}$$

Fails, in particular when there are many variables: (example from Bickel at al.)



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, (special case of Markov field sampling).

General idea: rather than find samples and then estimate their probability, first pick a probability and then find a sample that carries it.

The pdf one is trying to sample is almost known, in the sense that if a sample is given, its probability can be found (up to a constant). We solve (sometimes by iteration) the implicit equation (assumed probability)=probability of the sample X, yielding X.

First task: given P_n, sample P_{n+1}.

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 $F(X) - \phi = \xi^T\xi/2a$ (both F and ϕ are functions of i, n).

(ϕ needed for a solution to exist)

The right pdf is sampled if map $\xi \rightarrow X$ 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 $\xi \to X$. The probability we want to sample is $\exp(-\xi^T\xi/2)/(2\pi)^{m/2}$ multiplied by $\exp(-\phi)$. The sampling weight is $\exp(-\phi)J$.

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.

Solution of $F(X) - \phi = \xi^T \xi/2$:

(Note equation is underdetermined).

(i) If F is convex:

Iteration $X^j \rightarrow X = X^{n+1}$: $X^0 = 0$.

Linearize h around X^j: $h(X^{j+1}) = h(X^j) + h_x \cdot (X^{j+1} - X^j)$ (exact if converges). F is the sum of 2 quadratics in X^{j+1}. Completing squares gives a single quadratic $(x^{j+1} - \bar{a})^2/(2\bar{v}) + \phi(X^j)$ (a "pseudo-Gaussian"). J is easy to compute.

Special case: h linear (it does not matter whether the SDE is linear). Iteration converges in one step. J is independent of the particle. ϕ is explicitly known. Easy implementation of optimal sequential sampling.

(ii) F is U-shaped: Find $\phi = \min F$ (not a global minimum for the pdf !). Solve $F(X) - \phi = \xi^T \xi / 2$ by Newton's method. J pops out.

(iii) F not U-shaped: $F \rightarrow F_0$, F_0 convex, $\phi \rightarrow \phi + F(X) - F_0(X)$. Sampling still exact !

Note the equation $F(X)-\phi = \xi^T\xi/2$ 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.

Example of the use of this freedom:

Suppose the observation function is linear and dense $h(x) = Mx$, M dense. $F(X) - \phi = F(x) - \min F = ((x-a)^T A(x-a)/2$, where A is positive definite and symmetric, A, a known. Write $y = x - a$. You have to solve

$$y^T Ay = |\xi|^2.$$

Ansatz: $y = \lambda\eta$, λ scalar, $\eta = \xi/|\xi|$ unit random vector, so that $\lambda^2(\eta^T A\eta) = |\xi|^2$.

$E[\eta^T A\eta] = \text{trace}(A)/m = \Lambda$.

Solve instead

$\lambda^2\Lambda = |\xi|^2$;

a solution is $\lambda = |\xi|/\sqrt{\Lambda}$, and $y_i = \xi_i/\sqrt{\Lambda}$, $J = \Lambda^{-m/2}$.

To remove bias, $\phi \to \phi + \lambda^2[(\eta^T A\eta) - \Lambda)]$; note: $(\eta^T A\eta) - \Lambda) \to 0$ as $m \to \infty$!!

Backward sampling:

Suppose we go one step, resampling X^n given X^{n-1} and X^{n+1}. Set $X = X^n$, and define $\exp(-F) = P(X|X^{n+1})P(X|X^{n-1})P(b^n|X^n)$. Same steps follow.

Sparse observations:

Suppose no observations are available at time X^n. Sample X^n and X^{n+1} simultanuoulsy given data at times X^{n-1} and X^{n+1}- analogous construction.

Parameter estimation: usual augmentation does not work, solve falsification equation by stochastic approximation.

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.

The ship azimuth problem

A ship sets out from a point (x_0, y_0)

$$\begin{aligned} x^{n+1} &= x^n + u^{n+1}, \quad u^{n+1} \sim N(u^n, \beta) \\ y^{n+1} &= y^n + v^{n+1}, \quad v^{n+1} \sim N(v^n, \beta) \end{aligned}$$

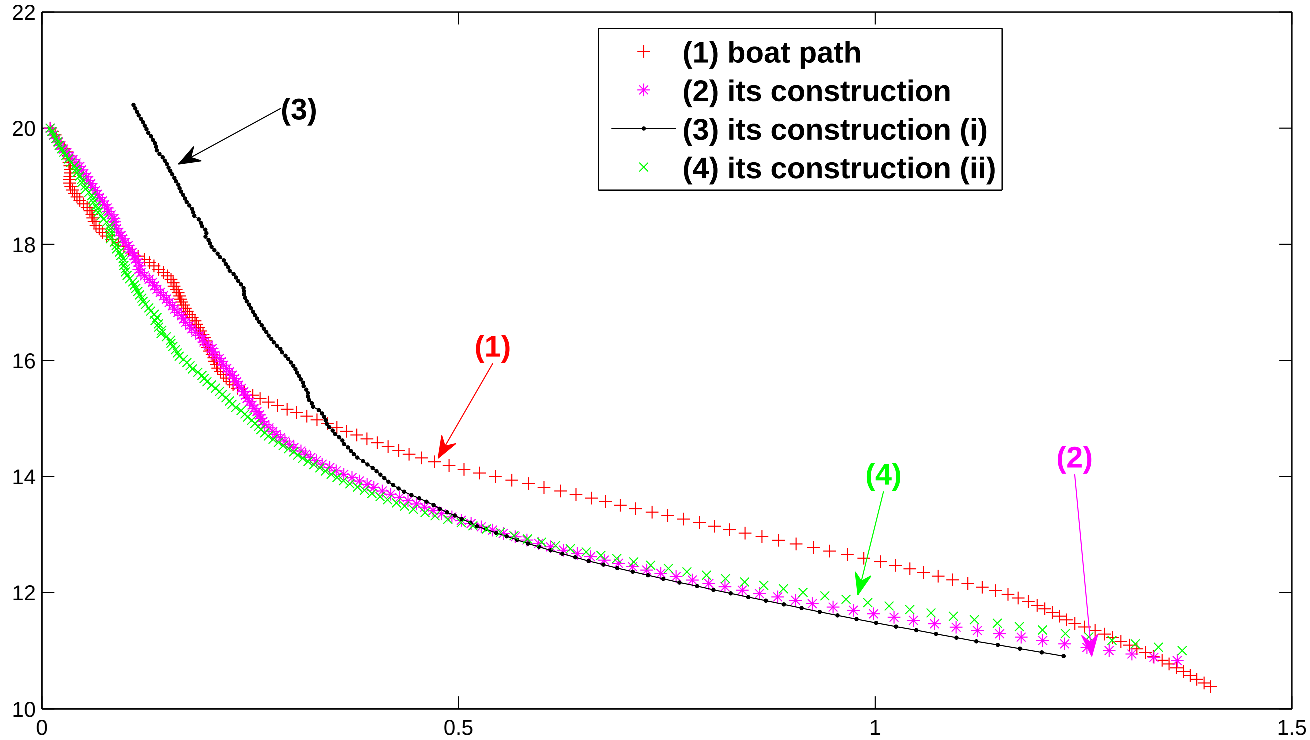
Observations:

$$b^n = \arctan \frac{y^n}{x^n} + N(0, s).$$

Take

- $x_0 = 0.01$, $y_0 = 20$, $u^1 = 0.002$, $v^1 = -0.06$,
- $\beta = 1 \cdot 10^{-6}$, $s = 25 \cdot 10^{-6}$.
- M particles all starting from $X_i^0 = x_0, Y_i^0 = y_0$

Sample boat path and its reconstruction



Construction (i): with perturbation of the initial positions

Construction (ii): with perturbation of the system variances

$$
\begin{aligned}
\frac{dP}{dt} &= \frac{N}{0.2+N}\gamma P - 0.1P - 0.6\frac{P}{0.1+P}Z + N(0,\sigma_P^2) \\
\frac{dZ}{dt} &= 0.18\frac{P}{0.1+P}Z - 0.1Z + N(0,\sigma_Z^2) \\
\frac{dN}{dt} &= 0.1D + 0.24\frac{P}{0.1+P}Z - \gamma P\frac{N}{0.2+N} + 0.05Z + N(0,\sigma_N^2) \\
\frac{dD}{dt} &= -0.1D + 0.1P + 0.18\frac{P}{0.1+P}Z + 0.05Z + N(0,\sigma_D^2)
\end{aligned}
$$

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

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

Observation: the concentration of plant pigments in the eastern tropical Pacific from late 1997 to mid 2002 (NASA's SeaWiFS)

Results (10, 10, 1000 particles)