

Optimal Importance Proposals in Particle Filtering

Particle Filtering in Dynamical Systems

Sai Sampath Kedari

sampath@umich.edu

Contents

1	Particle Filtering: Problem Setup and Algorithm	2
2	Limitation of the Bootstrap Particle Filter	3
3	Optimal Proposal Distribution in Particle Filtering	4
3.1	Optimality Criterion	4
3.2	Optimal Proposal	4
3.3	Implications for ESS and Particle Degeneracy	5
3.4	Sampling and Weight Update Under the Optimal Proposal	5
3.5	Required Distributions and Computational Bottlenecks	6

1 Particle Filtering: Problem Setup and Algorithm

We consider a general nonlinear state-space model

$$X_n = \Phi(X_{n-1}) + \xi_n, \quad (1)$$

$$Y_n = h(X_n) + \eta_n, \quad (2)$$

where:

- X_n is the latent state at time n ,
- Y_n is the measurement at time n ,
- $\Phi(\cdot)$ is a (possibly nonlinear) state transition function,
- $h(\cdot)$ is a (possibly nonlinear) measurement function,
- ξ_n and η_n are process and measurement noise terms.

No distributional assumptions are imposed on the noise terms. All that is assumed is the availability of:

- a state transition distribution

$$p(x_n \mid x_{n-1}),$$

- a measurement likelihood

$$p(y_n \mid x_n).$$

The objective is to approximate the filtering distribution

$$p(x_n \mid \mathcal{Y}_n)$$

using a finite set of weighted samples.

Particle Filtering Algorithm (Sequential Importance Sampling with Resampling)

Assume that at time $n - 1$ we have the empirical approximation

$$\hat{P}(x_{n-1} \mid y_{1:n-1}) = \sum_{i=1}^N \bar{w}_{n-1}^{(i)} \delta_{x_{n-1}^{(i)}}.$$

The particle filter proceeds as follows.

Step 1: Initialization ($n = 0$) Draw

$$x_0^{(i)} \sim p(x_0), \quad i = 1, \dots, N,$$

and set

$$\bar{w}_0^{(i)} = \frac{1}{N}.$$

Step 2: Propagation (Proposal Sampling) For each particle, draw

$$x_n^{(i)} \sim \pi\left(x_n \mid \mathcal{X}_{n-1}^{(i)}, \mathcal{Y}_n\right),$$

where

$$\mathcal{X}_{n-1}^{(i)} = (x_0^{(i)}, x_1^{(i)}, \dots, x_{n-1}^{(i)}), \quad \mathcal{Y}_n = (y_0, y_1, \dots, y_n).$$

The new trajectory is formed as

$$\mathcal{X}_n^{(i)} = (\mathcal{X}_{n-1}^{(i)}, x_n^{(i)}).$$

Step 3: Weight Update Update the unnormalized weights as

$$w_n^{(i)} = \bar{w}_{n-1}^{(i)} \frac{p(y_n \mid x_n^{(i)}) p(x_n^{(i)} \mid x_{n-1}^{(i)})}{\pi\left(x_n^{(i)} \mid \mathcal{X}_{n-1}^{(i)}, \mathcal{Y}_n\right)}.$$

Normalize:

$$\bar{w}_n^{(i)} = \frac{w_n^{(i)}}{\sum_{j=1}^N w_n^{(j)}}.$$

Step 4: Resampling (If Needed) Compute

$$N_{\text{eff}} = \left(\sum_{i=1}^N (\bar{w}_n^{(i)})^2 \right)^{-1}.$$

If $N_{\text{eff}} < N_{\text{th}}$, resample particles according to $\{\bar{w}_n^{(i)}\}$ and reset all weights to $1/N$.

Remark

The proposal distribution $x_n^{(i)} \sim \pi\left(x_n \mid \mathcal{X}_{n-1}^{(i)}, \mathcal{Y}_n\right)$ is the only design choice in the algorithm. Its selection determines the variance of the importance weights and the statistical efficiency of the particle filter.

The remainder of this report focuses exclusively on the *optimal proposal* and tractable approximations thereof.

2 Limitation of the Bootstrap Particle Filter

In the bootstrap particle filter, the proposal distribution is chosen as the state transition model:

$$\pi\left(x_n \mid \mathcal{X}_{n-1}^{(i)}, \mathcal{Y}_n\right) = \mathbb{P}\left(x_n \mid x_{n-1}^{(i)}\right).$$

Under this choice, particles are propagated forward solely according to the system dynamics:

$$x_0^{(i)} \rightarrow x_1^{(i)} \rightarrow \dots \rightarrow x_{n-1}^{(i)} \rightarrow x_n^{(i)}.$$

Crucially, the measurement data $\mathcal{Y}_n = (y_0, \dots, y_n)$ plays no role in the generation of new state samples. All information from the observations is injected *after the fact* through the importance weights.

As a result, particles are sampled freely from the transition distribution, regardless of whether the proposed states are compatible with the observed measurements. When the likelihood $p(y_n | x_n)$ is sharply peaked, most proposed particles are assigned negligible weight, leading to rapid weight degeneracy.

This behavior is fundamentally mismatched with the target distribution. The objective is to approximate the smoothing distribution

$$\mathbb{P}(\mathcal{X}_n | \mathcal{Y}_n),$$

which is explicitly conditioned on the measurement data. However, the bootstrap proposal ignores this conditioning during sampling and relies entirely on weight correction.

This mismatch motivates the search for proposal distributions that explicitly incorporate measurement information when generating new samples. The optimal proposal achieves this by conditioning directly on the current observation, thereby minimizing weight variance.

3 Optimal Proposal Distribution in Particle Filtering

In sequential importance sampling, particle degeneracy is driven by the variance of the incremental importance weights. If this variance is large, probability mass concentrates on a small number of particles, causing a rapid collapse of the effective sample size (ESS). Controlling the variance of the weights is therefore the central objective in the design of the proposal distribution.

At time n , particles are proposed according to a distribution of the form

$$\pi(x_n | \mathcal{X}_{n-1}^{(i)}, \mathcal{Y}_n),$$

which may depend on the full previously sampled trajectory and all available measurements. The choice of this proposal directly determines the statistical efficiency of the particle filter.

3.1 Optimality Criterion

Conditional on $\mathcal{X}_{n-1}^{(i)}$ and \mathcal{Y}_n , the incremental importance weight is

$$v_n(x_n) = \frac{\mathbb{P}(y_n | x_n) \mathbb{P}(x_n | x_{n-1}^{(i)})}{\pi(x_n | \mathcal{X}_{n-1}^{(i)}, \mathcal{Y}_n)}.$$

The optimal proposal is defined as the distribution that minimizes the conditional variance of this incremental weight,

$$\pi^* = \arg \min_{\pi} \text{Var}\left(v_n(X_n) \mid \mathcal{X}_{n-1}^{(i)}, \mathcal{Y}_n\right).$$

3.2 Optimal Proposal

The solution of the variance minimization problem is

$$\pi^*(x_n | x_{n-1}^{(i)}, y_n) = \mathbb{P}(x_n | x_{n-1}^{(i)}, y_n),$$

the posterior distribution of the next state conditioned on the previous state and the current measurement.

Under this choice, the randomness in the incremental weight no longer depends on the sampled state x_n , but only on the previous particle through the predictive likelihood.

3.3 Implications for ESS and Particle Degeneracy

The optimal proposal is constructed to make the incremental importance weights as close to equal as possible. In the ideal case, the conditional variance of the weights is zero, implying that all particles carry equal weight.

When all weights are equal, the effective sample size satisfies $\text{ESS} = N$, which is the best possible scenario. As weight variance increases, probability mass concentrates on a small number of particles, leading to a decrease in ESS and particle degeneracy.

By minimizing weight variance at each time step, the optimal proposal slows the decay of ESS and delays particle degeneracy.

3.4 Sampling and Weight Update Under the Optimal Proposal

Under the optimal proposal, particles are sampled according to

$$x_n^{(i)} \sim \mathbb{P}(x_n \mid x_{n-1}^{(i)}, y_n).$$

The generic sequential importance sampling weight update is

$$w_n^{(i)} = \bar{w}_{n-1}^{(i)} \frac{p(y_n \mid x_n^{(i)}) p(x_n^{(i)} \mid x_{n-1}^{(i)})}{\pi(x_n^{(i)} \mid \mathcal{X}_{n-1}^{(i)}, \mathcal{Y}_n)}.$$

Substituting the optimal proposal yields

$$w_n^{(i)} = \bar{w}_{n-1}^{(i)} \frac{p(y_n \mid x_n^{(i)}) p(x_n^{(i)} \mid x_{n-1}^{(i)})}{\mathbb{P}(x_n^{(i)} \mid x_{n-1}^{(i)}, y_n)}.$$

Applying Bayes' rule, the optimal proposal admits the factorization

$$\mathbb{P}(x_n \mid x_{n-1}, y_n) = \frac{\mathbb{P}(y_n \mid x_n, x_{n-1}) \mathbb{P}(x_n \mid x_{n-1})}{\mathbb{P}(y_n \mid x_{n-1})}.$$

From the state-space model, the Bayesian network implies the conditional independence

$$Y_n \perp X_{n-1} \mid X_n,$$

which follows directly from d-separation. Therefore,

$$\mathbb{P}(y_n \mid x_n, x_{n-1}) = \mathbb{P}(y_n \mid x_n).$$

Substituting, we obtain

$$\mathbb{P}(x_n \mid x_{n-1}, y_n) = \frac{\mathbb{P}(y_n \mid x_n) \mathbb{P}(x_n \mid x_{n-1})}{\mathbb{P}(y_n \mid x_{n-1})}.$$

Substituting this expression into the generic importance weight update,

$$w_n^{(i)} = \bar{w}_{n-1}^{(i)} \frac{\mathbb{P}(y_n \mid x_n^{(i)}) \mathbb{P}(x_n^{(i)} \mid x_{n-1}^{(i)})}{\mathbb{P}(x_n^{(i)} \mid x_{n-1}^{(i)}, y_n)},$$

yields

$$w_n^{(i)} = \bar{w}_{n-1}^{(i)} \frac{\mathbb{P}(y_n \mid x_n^{(i)}) \mathbb{P}(x_n^{(i)} \mid x_{n-1}^{(i)})}{\frac{\mathbb{P}(y_n \mid x_n^{(i)}) \mathbb{P}(x_n^{(i)} \mid x_{n-1}^{(i)})}{\mathbb{P}(y_n \mid x_{n-1}^{(i)})}}.$$

Canceling identical terms in numerator and denominator gives

$$w_n^{(i)} = \bar{w}_{n-1}^{(i)} \mathbb{P}(y_n \mid x_{n-1}^{(i)}).$$

Thus, under the optimal proposal, the incremental weight depends only on the predictive likelihood of the current observation given the previous state.

3.5 Required Distributions and Computational Bottlenecks

Under the optimal proposal, the particle filter recursion decomposes into two distinct operations, each requiring a different probabilistic quantity.

Sampling (prediction). For each particle i , the next state is drawn from the optimal proposal,

$$x_n^{(i)} \sim \mathbb{P}(x_n \mid x_{n-1}^{(i)}, y_n).$$

This distribution incorporates the current measurement and is used solely for proposing new particles.

Weight update (correction). The corresponding importance weight update is

$$w_n^{(i)} = \bar{w}_{n-1}^{(i)} \mathbb{P}(y_n \mid x_{n-1}^{(i)}),$$

where the predictive likelihood is

$$\mathbb{P}(y_n \mid x_{n-1}^{(i)}) = \int \mathbb{P}(y_n \mid x_n) \mathbb{P}(x_n \mid x_{n-1}^{(i)}) dx_n.$$

Computational bottlenecks. Implementing the optimal proposal therefore requires computing two quantities:

- the posterior distribution $\mathbb{P}(x_n \mid x_{n-1}, y_n)$ for sampling,
- the predictive likelihood $\mathbb{P}(y_n \mid x_{n-1})$ for weight updating.

Both quantities involve Bayesian inference and, for general nonlinear or non-Gaussian models, admit no closed-form expressions. This intractability is the fundamental obstacle to implementing the optimal proposal exactly.

To proceed, we follow the same strategy used in Gaussian filtering. We impose Gaussian assumptions on the process and measurement noise and study the resulting forms of the two required distributions: the posterior $\mathbb{P}(x_n \mid x_{n-1}, y_n)$ used for sampling, and the predictive likelihood $\mathbb{P}(y_n \mid x_{n-1})$ used for weight updating.

Under these Gaussian noise assumptions, no approximation is required when the measurement model is linear. In this case, both distributions admit exact, closed-form analytic expressions, yielding an exact implementation of the optimal proposal.

For nonlinear measurement models, closed-form solutions no longer exist. In those cases, the same Gaussian framework is retained, but the required distributions are approximated. Local linearization leads to EKF-based constructions, while numerical quadrature leads to UKF-based constructions.

The following chapters develop these cases in sequence, starting with the linear measurement model and progressing to EKF- and UKF-based Gaussian optimal proposals.