What Makes Filtering Hard?

Andrew M Stuart

Mathematics Institute, University of Warwick

cw/ S. Agapiou, O. Papaspiliopoulos, D. Sanz-Alonso

`arXiv.1511.06196`

—

cw/ D. Kelly

`In Preparation`

—

Funded by: DARPA, EPSRC, ERC, ONR

Table of Contents

Table of Contents

Problem Statement

Stochastic Dynamics Model

$$v_{j+1} = \psi(v_j) + \xi_j, \quad \xi_j \sim N(0, Q)$$
$$y_{j+1} = h(v_{j+1}) + \eta_{j+1}, \quad \eta_j \sim N(0, R).$$

Objective of Filtering

Define $\mu_0 = \mathbb{P}(v_0), \quad \mu_j = \mathbb{P}(v_j | Y_j), \quad Y_j := (y_1, \ldots, y_j).$

Sequentially Update $\mu_j \mapsto \mu_{j+1}$.

Propagation of Probabilities I

Using

$$\mathbb{P}(v_{j+1}|Y_j) = \mathbb{P}(v_{j+1}|v_j)\mathbb{P}(v_j|Y_j)$$

we obtain

Densities

$$\mathbb{P}(v_{j+1}|Y_{j+1}) = \frac{1}{\mathbb{P}(y_{j+1}|Y_j)}\mathbb{P}(y_{j+1}|v_{j+1}) \int \mathbb{P}(v_{j+1}|v_j)\mathbb{P}(v_j|Y_j)dv_j.$$

Measures

$$\mu_{j+1} = L_j P \mu_j.$$

P. Del Moral, A. Doucet and A. Jasra *Sequential Monte Carlo samplers*, J. Royal Stat. Soc. B **68**(2006).

Propagation of Probabilities II

Using

$$\mathbb{P}(v_{j+1}|v_j, y_{j+1}) = \frac{1}{\mathbb{P}(y_{j+1}|v_j)}\mathbb{P}(y_{j+1}|v_{j+1})\mathbb{P}(v_{j+1}|v_j)$$

we also find that

Densities

$$\mathbb{P}(v_{j+1}|Y_{j+1}) = \frac{1}{\mathbb{P}(y_{j+1}|Y_j)}\int \mathbb{P}(v_{j+1}|v_j, y_{j+1})\mathbb{P}(y_{j+1}|v_j)\mathbb{P}(v_j|Y_j)dv_j.$$

Measures

$$\mu_{j+1} = Q_j\mu_j.$$

A. Doucet, S. Godsill and C. Andrieu *On sequential Monte Carlo sampling methods for Bayesian filtering*, Statistics and Computing **10**(2000).

Sampling and Particle Filters

Monte Carlo Sampling

$$S^N \mu(du) = \sum_{i=1}^{N} \frac{1}{N} \delta_{u^i}(du),$$
$$u^i \sim \mu.$$

Standard Proposal

$$\mu_{j+1}^N = L_j S^N P \mu_j^N, \quad \mu_0^N = \mu_0.$$

Optimal Proposal

$$\mu_{j+1}^N = S^N Q_j \mu_j^N, \quad \mu_0^N = S^N \mu_0.$$

The Standard Proposal

$$
\begin{aligned}
\mu_0^N &= \mu_0;\ \ \mu_j^N(du) = \sum_{i=1}^{N} w_j^i \delta_{u_j^i}(du),\ \ j \geq 1,\\
u_{j+1}^i &= \psi(u^i) + \xi_j^i, \quad u_j^i \sim \mu_j^N,\ \ \xi_j^i \sim N(0, Q),\\
\widehat{w}_{j+1}^i &= \exp\Big(-\frac{1}{2}|y_{j+1} - h(u_{j+1}^i)|_R^2\Big),\\
w_{j+1}^i &= \widehat{w}_{j+1}^i \Big/ \sum_{m=1}^{N} \widehat{w}_{j+1}^m .
\end{aligned}
$$

The Optimal Proposal (case $h(\cdot) = H\cdot$)

$$\mu_j^N(du) = \sum_{i=1}^{N} \frac{1}{N}\delta_{u_j^i}(du),\ j \geq 0,$$

$$\widehat{u}_{j+1}^i = (I - KH)\psi(u_j^i) + Ky_{j+1} + \zeta_j^i,\ \zeta_j^i \sim N(0, C)$$

$$\widehat{w}_{j+1}^i = \exp\Big(-\frac{1}{2}|y_{j+1} - H\psi(u_j^i)|_S^2\Big)$$

$$w_{j+1}^i = \widehat{w}_{j+1}^i \Big/ \sum_{m=1}^{N} \widehat{w}_{j+1}^m$$

$$\mu_{j+1}^N(du) = S^N\Big(\sum_{i=1}^{N} w_{j+1}^i \delta_{\widehat{u}_{j+1}^i}(du)\Big).$$

Here

$$S = HQH^* + R,\ K = QH^*S^{-1},$$

$$C = (I - KH)Q.$$

EnKF (case $h(\cdot) = H\cdot$)

$$\mu_j^N(du) = \sum_{i=1}^{N} \frac{1}{N}\delta_{u_j^i}(du),\ j \geq 0,$$
$$u_{j+1}^i = (I - K_{j+1}H)\psi(u_j^i) + K_{j+1}y_{j+1} + \zeta_j^i,\ \zeta_j^i \sim N(0, C_{j+1})$$

Here $\widehat{C}_{j+1}$ is the empirical covariance of the predicted particle positions

$$\widehat{u}_{j+1}^i = \psi(u_j^i) + \zeta_j^i$$

and

$$S_{j+1} = H\widehat{C}_{j+1}H^* + R,\ K_{j+1} = \widehat{C}_{j+1}H^*S^{-1},$$
$$C_{j+1} = (I - K_{j+1}H)Q.$$

References

It’s all about the weights

T. Bengtsson, P. Bickel, B. Li, *Curse-of-dimensionality revisited: Collapse of the particle filter in very large scale systems*, Probability and Statistics: Essays in Honour of D.A. Freeman, 2008.

C. Snyder, T. Bengtsson, P. Bickel, *Obstacles to high-dimensional particle filtering*, Monthly Weather Review, **136**(2008).

C. Snyder, *Particle filters, the ‘optimal’ proposal and high-dimensional systems*, In Proceedings of the ECMWF Seminar on Data Assimilation for Atmosphere and Ocean, 2011.

M. Ades, P.J. van Leeuwen, *An exploration of the equivalent weights particle filter*, Quarterly Journal of the Royal Meteorological Society, **139**(2013).

A. Chorin, M. Morzfeld and X.Tu *Implicit particle filters for data assimilation*, Comm. Appl. Math. and Comp. Sci., **5**(2014).

A. Chorin, M. Morzfeld, *Conditions for successful data assimilation*, J. Geophysical Research, **118**(2014).

S. Reich, *A dynamical systems framework for intermittent data assimilation*, BIt Numerical Mathematics, **51**(2011).

Table of Contents

Metric on Random Measures

Let $(\mathcal{X}, \mathcal{F})$ be an arbitrary measurable space and ν a priobability measure. Then

$$\nu(\phi) := \int_{\mathcal{X}} \phi(u)\nu(du).$$

If ν_1, ν_2 are random probability measures then we introduce a metric:

$$d(\nu_1, \nu_2) := \Big(\sup_{|\phi| \le 1} \mathbb{E}\Big[\big(\nu_1(\phi) - \nu_2(\phi)\big)^2 \Big] \Big)^{\frac{1}{2}}.$$

P. Rebeschini and R. Van Handel. Can local particle filters beat the curse of dimensionality? Ann. Appl. Prob. **25**(2015).

P. Del Moral. Feynman-Kac formulae: genealogical and interacting particle systems with applications 2004.

The Standard Proposal

Recall

$$\mu_{j+1} = L_j P \mu_j,$$
$$\mu_{j+1}^N = L_j S^N P \mu_j^N, \mu_0^N = \mu_0.$$

Theorem 1

Assume that h is a bounded nonlinear function. Then there exists $\kappa = \kappa(Y_J)$ *such that*

$$d(\mu_j, \mu_j^N)^2 \leq \frac{\kappa}{N}.$$

P. Rebeschini and R. Van Handel. Can local particle filters beat the curse of dimensionality? Ann. Appl. Prob. **25**(2015).

$P\mu_j$ is proposal for importance sampling of μ_{j+1}.

The Optimal Proposal

Recall

$$\mu_{j+1} = Q_j \mu_j,$$
$$\mu_{j+1}^N = S^N Q_j \mu_j^N, \mu_0^N = S^N \mu_0.$$

Theorem 2

Assume that ψ is a bounded nonlinear function and $h(\cdot) = H\cdot$ is linear. Then there exists $\kappa = \kappa(Y_J)$ such that

$$d(\mu_j, \mu_j^N)^2 \leq \frac{\kappa}{N}.$$

D.T.B. Kelly and A.M. Stuart. In preparation (2016).

Q_j factors into importance sampling followed by conditioned dynamics when applied to a sum of Dirac measures.

The EnKF

Theorem 3

Assume that $\psi(\cdot) = M\cdot$ and $h(\cdot) = H\cdot$ is linear. Then there exists $\kappa = \kappa(Y_J)$ such that

$$d(\mu_j, \mu_j^N)^2 \leq \frac{\kappa}{N}.$$

F. Le Gland, V. Monbet and V.-D. Tran. INRIA Report 7014(2009).

The paper also identifies the limit of μ_j^N obtained in the nonlinear, non-Gaussian setting and shows that it is not μ_j. Suggests using EnKF as a proposal for importance sampling.

Autonormalized Importance Sampling (IS)

π and μ probability measures in the arbitrary measurable space $(\mathcal{X}, \mathcal{F})$. Assume $\mu \ll \pi$ so that

$$\mu(du) = g(u)\pi(du) \Big/ \int_{\mathcal{X}} g(u)\pi(du).$$

Applying the Monte Carlo sampling operator twice gives

$$\mu(\phi) = \frac{\pi(\phi g)}{\pi(g)} \approx \frac{S^N \pi(\phi g)}{S^N \pi(g)} := \mu^N(\phi)$$

IS Convergence

Thus we have the approximation of μ defined by, for $u^i \sim \pi$,

$$\mu^N = \sum_{i=1}^{N} w^i \delta_{u^i}, \qquad w^i = \frac{g(u^i)}{\sum_{k=1}^{N} g(u^k)}.$$

Define $\rho := \frac{\pi(g^2)}{\pi(g)^2} \in [1, \infty]$.

Theorem 4

$$d(\mu^N, \mu)^2 \leq \frac{4}{N}\rho.$$

Thus ρ, and its dependence on the problem at hand, is key.

S. Agapiou, O. Papaspiliopoulos, D. Sanz-Alonso and A.M. Stuart. arXiv.1511.06196

Table of Contents

IS for Filtering

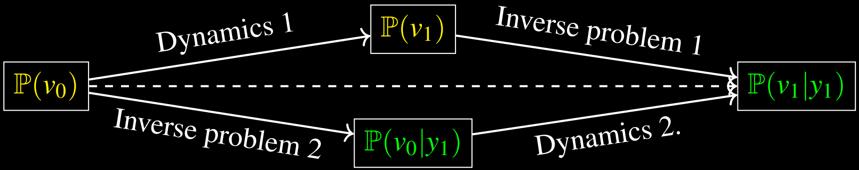
Filtering

Signal: $v_1 = Mv_0 + N(0,Q), \qquad v_0 \sim N(0,P) = \mathbb{P}_0.$

Data: $y_1 = Hv_1 \; + N(0,R)\;. \qquad$ Find: $\mathbb{P}(v_1|y_1).$

Standard proposal: Top Route: dynamics, then inverse problem.

Optimal proposal: Bottom Route: inverse problem, then dynamics.



C. Snyder. Particle filters, the “optimal” proposal and high-dimensional systems. In Proceedings of the ECMWF Seminar on Data Assimilation for Atmosphere and Ocean, 2011.

Inverse Problem: Standard Proposal

Filtering

Signal: $v_1 = Mv_0 + N(0, Q), \qquad v_0 \sim N(0, P) = \mathbb{P}_0.$

Data: $y_1 = Hv_1 \; + N(0, R)\;. \qquad$ Target: $\mu := \mathbb{P}(v_1 | y_1).$

Behaviour of filtering model determined by inverse problem

$$y_1 = Ku + \eta, \quad u \sim N(0, \Sigma), \quad \eta \sim N(0, \Gamma),$$

with

- Target $\mu = \mathbb{P}(v_1 | y_1) : K = H, \Gamma = R, \Sigma = MPM^* + Q$
- Proposal $\pi_{st} = \mathbb{P}(v_1).$

Inverse Problem: Optimal Proposal

Filtering

Signal: $v_1 = Mv_0 + N(0, Q), \qquad v_0 \sim N(0, P) = \mathbb{P}_0.$

Data: $y_1 = Hv_1 \; + N(0, R) \; . \qquad$ Target: $\mu := \mathbb{P}(v_0|y_1).$

Behaviour of filtering model determined by inverse problem

$$y_1 = Ku + \eta, \quad u \sim N(0, \Sigma), \quad \eta \sim N(0, \Gamma),$$

with

- Target $\mu = \mathbb{P}(v_0|y_1)$: $K = HM$, $\Gamma = HQH^* + R$, $\Sigma = Q$
- Proposal $\pi_{st} = \mathbb{P}(v_0)$.

Linear Bayesian Inverse Problems in Hilbert Space

Interested in recovering $u \in \mathcal{H}$ from data $y = Ku + \eta \in \mathcal{H}$.

Bayesian Inverse Problem

$$\left.\begin{aligned} &\text{Prior (proposal)}: u \sim \pi = N(0, \Sigma) \\ &\text{Data}: y = Ku + \eta \in \mathcal{Y}, \quad \eta \sim N(0, \Gamma) \end{aligned}\right\} u|y \sim \mu = N(m, C).$$

Linear Bayesian Inverse Problems in Hilbert Space

Interested in recovering $u \in \mathcal{H}$ from data $y = Ku + \eta \in \mathcal{H}$.

Bayesian Inverse Problem

$$\left.\begin{array}{l} \text{Prior (proposal)}: u \sim \pi = N(0, \Sigma) \\ \text{Data}: y = Ku + \eta \in \mathcal{Y}, \quad \eta \sim N(0, \Gamma) \end{array}\right\} u|y \sim \mu = N(m, C).$$

Let $y_0 = \Gamma^{-\frac{1}{2}} y$ and $S = \Gamma^{-\frac{1}{2}} K \Sigma^{\frac{1}{2}}$. Then

Equivalent Bayesian Inverse Problem

$$\left.\begin{array}{l} \text{Prior (proposal)}: u_0 \sim \pi_0 = N(0, I) \\ \text{Data}: y_0 = Su_0 + \eta_0 \in \mathcal{Y}_0, \quad \eta_0 \sim N(0, I) \end{array}\right\} u_0|y_0 \sim \mu_0 = N(m_0, C_0).$$

Intrinsic Dimension

Key: Eigenvalues of $A := S^*S = \Sigma^{1/2}K^*\Gamma^{-1}K\Sigma^{1/2}$.

$$\boxed{\text{efd} = \text{Tr}\left((I+A)^{-1}A\right)} \qquad \boxed{\tau = \text{Tr}\,(A)}$$

(efd) Spiegelhalter, Best, Carlin, Van der Linde, Bayesian measures of model complexity and fit, 2002.

(τ) Bengtsson, Bickel, and Li, Curse-of-dimensionality revisited: Collapse of the particle filter in very large systems, 2008.

Lemma 5

$$\frac{1}{\|I+A\|}\tau \le \text{efd} \le \tau.$$

$$\text{efd} = \text{Tr}\left((\Sigma - C)\Sigma^{-1}\right) \le \min\{d_u, d_y\},$$
$$\tau = \text{Tr}\left((\Sigma - C)C^{-1}\right).$$

S.Agapiou, O.Papaspiliopoulos, D. Sanz-Alonso and A.M. Stuart. arXiv.1511.06196

Link Between τ, efd, ρ and $\mu \ll \pi$

Theorem 6

Let $\nu(du, dy) = \mathbb{P}(dy|u)\pi(du)$ *and assume* A *bdd. Equivalent:*

i) efd $< \infty$.

ii) $\tau < \infty$.

iii) *For* ν*-a.a.* y, $\mu(du) \propto g(u;y)\pi(du)$ *and*

$$0 < \pi\big(g(\cdot;y)\big) < \infty.$$

iv) *It holds* $0 < g(u;y) < \infty$ ν*-a.s. and for* ν*-a.a.* y

$$\rho := \frac{\pi\left(g(\cdot;y)^2\right)}{\pi\big(g(\cdot;y)\big)^2} < \infty.$$

S.Agapiou, O.Papaspiliopoulos, D. Sanz-Alonso and A.M. Stuart. arXiv.1511.06196

Link Between τ, efd, ρ and $\mu \ll \pi$

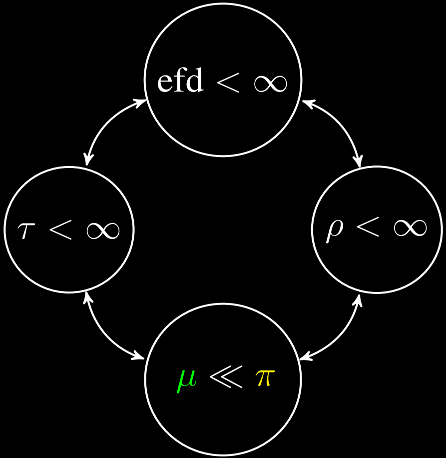


Table of Contents

Ordering by τ

Filtering

Signal: $v_1 = Mv_0 + N(0, Q)$, $\quad v_0 \sim N(0, P) = \mathbb{P}_0$.

Data: $y_1 = Hv_1 + N(0, R)$. $\quad$ Target: $\mathbb{P}(u|y_1)$, $u = (v_0, v_1)$.

	Standard Proposal	Optimal Proposal
S	$R^{-\frac{1}{2}}H(MPM^* + Q)^{\frac{1}{2}}$	$(HQH^* + R)^{-\frac{1}{2}}HMP^{\frac{1}{2}}$

Recall that $A = S^*S$, efd $= \mathrm{Tr}\left((I + A)^{-1}A\right)$, $\tau = \mathrm{Tr}\,(A)$.

Theorem 7

$$\tau_{st} \geq \tau_{op}$$

Diagonal Example: Small Noise and High Dimensions

$M = H = I \in \mathbb{R}^{d\times d}$, and $R = rI$, $Q = qI$, with $r, q > 0$.

Stationary regime $P_\infty = \frac{\sqrt{q^2+4qr}-q}{2} I$,

Chorin and Morzfeld. Conditions for successful data assimilation, 2013.

$$A_{st} = \frac{\sqrt{q^2+4qr}+q}{2r} I, \qquad A_{op} = \frac{\sqrt{q^2+4qr}-q}{2(q+r)} I.$$

Outside stationarity: $P = pI$, with $p > 0$ then

$$A_{st} = \frac{p+q}{r} I, \qquad A_{op} = \frac{p}{q+r} I.$$

Diagonal Example: Small Noise and High Dimensions

Regime	Param.	$\mathrm{eig}(A_{st})$	$\mathrm{eig}(A_{op})$	ρ_{st}	ρ_{op}
Noise	$r \to 0$	$\mathbf{r^{-1}}$	r	$\mathcal{O}\left(\mathbf{r^{-d/2}}\right)$	$\mathcal{O}(1)$
	$r = q \to 0$	1	1	$\mathcal{O}(1)$	$\mathcal{O}(1)$
Large d	$d \to \infty$	1	1	$\mathcal{O}(\exp(\mathbf{d}))$	$\mathcal{O}(\exp(\mathbf{d}))$

Table: $P = P_\infty$.

Regime	Param.	$\mathrm{eig}(A_{st})$	$\mathrm{eig}(A_{op})$	ρ_{st}	ρ_{op}
Noise	$r \to 0$	$\mathbf{r^{-1}}$	1	$\mathcal{O}\left(\mathbf{r^{-d/2}}\right)$	$\mathcal{O}(1)$
	$r = q \to 0$	$\mathbf{r^{-1}}$	$\mathbf{r^{-1}}$	$\mathcal{O}\left(\mathbf{r^{-d/2}}\right)$	$\mathcal{O}\left(\mathbf{r^{-d/2}}\right)$
Large d	$d \to \infty$	1	1	$\mathcal{O}(\exp(\mathbf{d}))$	$\mathcal{O}(\exp(\mathbf{d}))$

Table: $P = pI$.

Table of Contents

Highlights

1. GENERAL FRAMEWORK
   - Standard and optimal filters converge in the large N limit.
   - Cost determined by embedded importance sampling problem.
2. INVERSE PROBLEMS
   - Established connection between notions of dimension: efd and τ.
   - Linked intrinsic dimensions to importance sampling through ρ.
   - Link with absolute continuity in Hilbert space setting.
3. FILTERING
   - Theory from inverse problems determines proposal performance.
   - Optimal proposal has lower dimension than standard proposal.
   - Effect of dimension and small noise can be studied.