Mathematics and numerics for data assimilation and state estimation – Lecture 17

Summer semester 2020

Overview

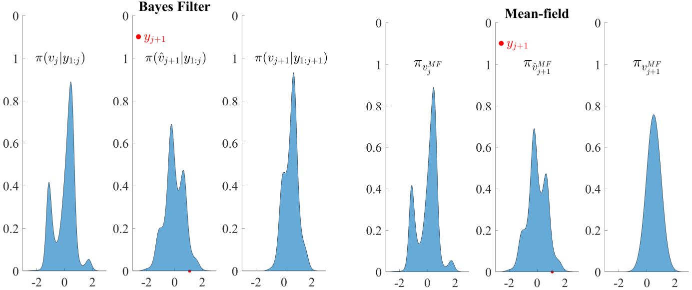
1. Bootstrap particle filter
2. Convergence of the BPF
3. Other proposals for particle filters

Summary lecture 16

- Described the extended KF and introduced ensemble KF.
- Studied convergence properties of the methods, particularly showing that the Gaussian approximation in the analysis of EnKF leads to errors for nonlinear problems:

$$\text{BF:} \quad \pi(v_{j+1}|y_{1:j+1}) \propto \pi_{N(y_{j+1},\Gamma)}(v_{j+1})\pi(v_{j+1}|y_{1:j})$$

$$\text{MFEnKF:} \quad \pi_{v_{j+1}^{\mathrm{MF}}}(v_{j+1}) \;\propto \pi_{K_{j+1}^{\mathrm{MF}} N(y_{j+1}-H\hat{v}_{j+1}^{\mathrm{MF}},\Gamma)} * \pi_{v_j^{\mathrm{MF}}}(v_{j+1}).$$



Plan for today

Particle filtering: a nonlinear filtering method which, in essence, treats the prediction step as EnKF, and reweights particles in the analysis step.

Recall that for the Bayes Filter,

$$\pi(v_j|y_{1:j}) \propto \pi(y_j|v_j)\pi(v_j|y_{1:j-1}).$$

(Bootstrap) particle filters consists of collection of weights and particles: $\{(w_j^{(i)}, \hat{v}_j^{(i)})\}_{i=1}^M$ with empirical measure

$$\pi_j^M(dv) = \sum_{i=1}^M w_j^{(i)} \delta_{\hat{v}_j^{(i)}}(dv)$$

where the weights sum to 1, and

$$w_j^{(i)} \propto \pi_{Y_j|V_j}(y_j|\hat{v}_j^{(i)}),$$

and $\pi_{\hat{v}_j^{(i)}} \approx \pi_{V_j|Y_{1:j-1}}(\cdot|y_{1:j-1})$.

Overview

1. Bootstrap particle filter
2. Convergence of the BPF
3. Other proposals for particle filters

Filtering setting

$V_0 \sim \pi_0$ and mappings $F : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}^d$ and $G : \mathbb{R}^k \times \mathbb{R}^k \to \mathbb{R}^k$ such that for for $j = 0, 1, \ldots$ and the hidden Markov model

$$\begin{aligned} V_{j+1} &= F(V_j, \xi_j) \\ Y_{j+1} &= G(V_{j+1}, \eta_{j+1}) \end{aligned} \tag{1}$$

with iid $\{\xi_j\}$ and iid $\{\eta_j\}$ where $V_0 \perp \{\xi_j\} \perp \{\eta_j\}$.

"Classic" setting obtained with $F(v, \xi) = \Psi(v) + \xi$ and $G(v, \eta) = h(v) + \eta$ with Gaussian ξ and η.

Note that the Markov chain $\{V_j\}$ may be associated to a time-independent kernel density function

$$\pi_{V_{j+1}|V_j}(v_{j+1}|v_j) = p(v_j, v_{j+1}).$$

and that, as in the classic setting,

$$\pi(y_{j+1}|v_{j+1}, y_{1:j}) = \pi(y_{j+1}|v_{j+1}).$$

Bayes filter – in operator notation

Notation for analysis and prediction Bayes filter pdfs:

$$\pi_j(v) := \pi_{V_j|Y_{1:j}}(v|y_{1:j}) \quad \text{and} \quad \hat{\pi}_{j+1}(v) := \pi_{V_j|Y_{1:j}}(v|y_{1:j})$$

The transition $\pi_j \mapsto \pi_{j+1}$ consists of two steps:

1. Prediction:

$$\hat{\pi}_{j+1}(v_{j+1}) = (\mathscr{P}\pi_j)(v_{j+1}) := \int_{\mathbb{R}^d} p(v_j, v_{j+1})\pi_j(v_j)dv_j$$

2. Analysis

$$\pi_{j+1}(v_{j+1}) = (\mathscr{A}_{j+1}\hat{\pi}_j)(v_{j+1}) := \frac{\pi(y_{j+1}|v_{j+1})\hat{\pi}_j(v_{j+1})}{\int_{\mathbb{R}^d} \pi_{Y_{j+1}|V_{j+1}}(y_{j+1}|v)\hat{\pi}_{j+1}(v)dv}$$

   where the subscript in $\mathscr{A}_{j+1}$ relates to the value of y_{j+1}.

Summary: $\pi_{j+1} = \mathscr{A}_{j+1}\mathscr{P}\pi_j$, which may also connect to

$$\pi(v_{j+1}|y_{1:j+1}) \propto \pi(y_{j+1}|v_{j+1})\pi(v_{j+1}|y_{1:j}).$$

Bootstrap particle filter

Given a probability measure or density π, we define for any $M \in \mathbb{N}$, the empirical probability measure

$$\mathcal{S}^M \pi(dv) := \frac{1}{M} \sum_{i=1}^{M} \delta_{v^i}(dv) \quad \text{where} \quad v^{(i)} \overset{iid}{\sim} \pi.$$

Approximation ideas for particle filtering: Given π_j,

1. Approximate $\pi_j^M = \mathcal{S}^M \pi \approx \pi_j$
2. Prediction $\hat{\pi}_{j+1}^M = \mathcal{S}^M(\mathscr{P}\pi_j^M) \approx \mathscr{P}\pi_j$
3. Analysis $\pi_{j+1}^M = \mathscr{A}_{j+1}\hat{\pi}_{j+1}^M \approx \mathscr{A}\hat{\pi}_{j+1}$.

Problem: Have only defined $\mathscr{P}$ and $\mathscr{A}_{j+1}$ as mappings from pdfs to pdfs, but π_j^M and $\hat{\pi}_{j+1}^M$ a measures.

Extension of mappings

$\mathscr{P}$ as mapping from empirical probability measures (epms) to pdfs: For any

$$\pi(dv) = \sum_{i=1}^{M} w^{(i)} \delta_{v^{(i)}}(dv)$$

with $\sum_{i=1}^{M} w^{(i)} = 1$,

$$(\mathscr{P}\pi)(u) := \int_{\mathbb{R}^d} p(v,u)\pi(dv) = \sum_{i=1}^{M} w^{(i)} p(v^{(i)}, u)$$

Example: In the classic setting $p(v,u) \propto \exp(-|\Psi(v) - u|_{\Sigma}^2/2)$ and thus

$$(\mathscr{P}\pi)(u) \propto \sum_{i=1}^{M} w^{(i)} \exp\Big(- |\Psi(v^{(i)}) - u|_{\Sigma}^2/2 \Big).$$

$\mathscr{A}_j$ as mapping from epms to epms

For any epm

$$\pi(dv) = \sum_{i=1}^{M} w^{(i)} \delta_{v^{(i)}}(dv)$$

we define

$$(\mathscr{A}_j \pi)(du) := \frac{\pi_{Y_j|V_j}(y_j|u)\pi(du)}{\int_{\mathbb{R}^d} \pi_{Y_j|V_j}(y_j|v)\pi(dv)}$$

$$=$$

$$= \sum_{i=1}^{M} \frac{w^{(i)} \pi_{Y_j|V_j}(y_j|v^{(i)})}{Z} \delta_{v^{(i)}}(du)$$

with $Z = \sum_{i=1}^{M} w^{(i)} \pi_{Y_j|V_j}(y_j|v^{(i)})$.

Approximation ideas for particle filtering revisited

Given the BF $\pi_j^M = \sum_{i=1}^M w_j^{(i)} \delta_{v_j^i}$, we compute π_{j+1}^M by the following steps

1. Resampling $\pi_j^M = \mathcal{S}^M \pi_j^M$ $\qquad \left(= \frac{1}{M} \sum_{i=1}^M \delta_{v_j^i} \right)$

2. Prediction $\hat{\pi}_{j+1}^M = \mathcal{S}^M (\mathscr{P} \pi_j^M)$ $\qquad \left(= \frac{1}{M} \sum_{i=1}^M \delta_{\hat{v}_{j+1}^i} \right)$

3. Analysis

$$\pi_{j+1}^M = \mathscr{A}_{j+1} \hat{\pi}_{j+1}^M \qquad \left(= \sum_{i=1}^M \underbrace{\frac{\pi_{Y_{j+1}|V_{j+1}}(y_{j+1}|\hat{v}_{j+1}^{(i)})}{Z}}_{w_{j+1}^{(i)}} \delta_{\hat{v}_{j+1}^{(i)}} \right).$$

Note that $\pi_{j+1}^M = \mathscr{A}_{j+1} \mathcal{S}^M \mathscr{P} \pi_j^M$ is described by $\{(w_{j+1}^{(i)}, \hat{v}_{j+1}^{(i)})\}$.

Importance sampling viewpoint:

$$\begin{aligned}\pi_{j+1}(v_{j+1}) &\propto \pi(y_{j+1}|v_{j+1})\pi(v_{j+1}|y_{1:j})\\ &= \underbrace{\pi(y_{j+1}|v_{j+1})}_{\text{“weight”}} \int_{\mathbb{R}^d} \underbrace{\pi(v_{j+1}|v_j)\pi_j(v_j)}_{\text{“sampling density”}} dv_j,\end{aligned}$$

and for the particle filters this is approximated by

$$\pi_{j+1}^M = \sum_{i=1}^{M} w_{j+1}^{(i)} \delta_{\hat{v}_{j+1}^{(i)}}$$

with $\hat{v}_{j+1}^{(i)} \sim \int \pi_{V_{j+1}|V_j}(\cdot|v_j)\pi_j^M(v_j)dv_j$ and $w_{j+1}^{(i)} \propto \pi_{Y_{j+1}|V_{j+1}}(y_{j+1}|\hat{v}_{j+1}^{(i)})$

Bootstrap particle filter (BPF) algorithm [SST 11.1]

- **Input:** Initial distribution π_0 (which we also write π_0^M), obs sequence $y_1, y_2, \ldots$, and M.
- **Particle generation:** For $j = 0, 1, \ldots$

  1. **Resampling** Draw $v_j^{(i)} \overset{iid}{\sim} \pi_j^M$ for $i = 1, \ldots, M$.
  2. Simulate $\hat{v}_{j+1}^{(i)} = F(v_j^{(i)}, \xi_j^{(i)})$ with iid $\xi_j^{(i)}$.
  3. Set $\bar{w}_{j+1}^{(i)} = \pi_{Y_{j+1}|V_{j+1}}(y_{j+1}|\hat{v}_{j+1}^{(i)})$
  4. and $w_{j+1}^{(i)} = \bar{w}_{j+1}^{(i)} / \sum_{k=1}^{M} \bar{w}_{j+1}^{(k)}$.
  5. Set $\pi_{j+1}^M = \sum_{i=1}^{M} w_{j+1}^{(i)} \delta_{\hat{v}_{j+1}^{(i)}}$.

- **Output:** π_j^M approximating the distribution of $V_j | Y_{1:j} = y_{1:j}$.

BPF algorithm classic setting

- **Input:** Initial distribution π_0 (which we also write π_0^M), obs sequence $y_1, y_2, \ldots$, and M.
- **Particle generation:** For $j = 0, 1, \ldots$

  1. **Resampling** Draw $v_j^{(i)} \overset{iid}{\sim} \pi_j^M$ for $i = 1, \ldots, M$.
  2. Simulate $\hat{v}_{j+1}^{(i)} = \Psi(v_j^{(i)}) + \xi_j$ with $\xi_j^{(i)} \overset{iid}{\sim} N(0, \Sigma)$.
  3. Set $\bar{w}_{j+1}^{(i)} = \exp(-\frac{1}{2}|y_{j+1} - h(\hat{v}_{j+1}^{(i)})|_\Gamma^2)$
  4. and $w_{j+1}^{(i)} = \bar{w}_{j+1}^{(i)} / \sum_{k=1}^M \bar{w}_{j+1}^{(k)}$.
  5. Set $\pi_{j+1}^M = \sum_{i=1}^M w_{j+1}^{(i)} \delta_{\hat{v}_{j+1}^{(i)}}$.
- **Output:** π_j^M.

Sequential importance sampling (SIS) vs sequential importance resampling (SIR)

- Bootstrap particle filter is a special case of SIR (can have more general "proposals" in step 2.).
- Without the resampling step 1., the particle weights multiply every step, and one may risk very uneven particle weights: this is called **the degeneracy problem**.
- With resampling, uneven weights are avoided, but (1) one may lose information and (2) the variance of the resulting particle distribution π_j^M can be shown to increase.

- Adaptive resampling can for instance be based on estimating the effective number of particles

$$n_{eff,j} \approx \frac{1}{\sum_{i=1}^{M} (w_j^{(i)})^2}$$

and employing the SIR resampling step to SIS only when $n_{eff,j} < M/10$. (Motivation: if $w_j^{(i)} = 1/M$ for all i, then $n_{eff,j} = M$.)

Sequential importance sampling algorithm 1

- **Input:** Initial distribution π_0, obs sequence $y_1, y_2, \ldots$, and M.
- **Initialization:** Draw $\hat{v}_j^{(i)} \overset{iid}{\sim} \pi_0$ and set $w_0^{(i)} = 1/M$ for $i = 1, \ldots, M$. (Hat notation here is formally "wrong" but practical.)
- **Particle and weight dynamics:** For $j = 0, 1, \ldots,$,

  1. Simulate $\hat{v}_{j+1}^{(i)} = F(\hat{v}_j^{(i)}, \xi_j^{(i)})$ with iid $\xi_j^{(i)}$.
  2. Set $\bar{w}_{j+1}^{(i)} = w_j^{(i)} \pi_{Y_{j+1}|V_{j+1}}(y_{j+1}|\hat{v}_{j+1}^{(i)})$
  3. and $w_{j+1}^{(i)} = \bar{w}_{j+1}^{(i)} / \sum_{k=1}^{M} \bar{w}_{j+1}^{(k)}$.
  4. Set $\pi_{j+1}^M = \sum_{i=1}^{M} w_{j+1}^{(i)} \delta_{\hat{v}_{j+1}^{(i)}}$.
- **Output:** π_j^M.

Adaptive resampling algorithm

- **Input:** Initial distribution π_0, obs sequence $y_1, y_2, \ldots$, and M.
- **Initialization:** Draw $\hat{v}_j^{(i)} \overset{iid}{\sim} \pi_0$ and set $w_0^{(i)} = 1/M$ for $i = 1, \ldots, M$. (Hat notation here is formally "wrong" but practical.)
- **Particle and weight dynamics:** For $j = 0, 1, \ldots,$,

  1. Compute $n_{eff,j}$. If $n_{eff,j} < M/10$, then **resample:** draw $\hat{v}_j^{(i)} \overset{iid}{\sim} \pi_j^M$ for $i = 1, \ldots, M$ and set $w_j^{(i)} = 1/M$ for $i = 1, \ldots, M$.
  2. Simulate $\hat{v}_{j+1}^{(i)} = F(\hat{v}_j^{(i)}, \xi_j^{(i)})$ with iid $\xi_j^{(i)}$.
  3. Set $\bar{w}_{j+1}^{(i)} = w_j^{(i)} \pi_{Y_{j+1}|V_{j+1}}(y_{j+1}|\hat{v}_{j+1}^{(i)})$
  4. and $w_{j+1}^{(i)} = \bar{w}_{j+1}^{(i)} / \sum_{k=1}^M \bar{w}_{j+1}^{(k)}$.
  5. Set $\pi_{j+1}^M = \sum_{i=1}^M w_{j+1}^{(i)} \delta_{\hat{v}_{j+1}^{(i)}}$.
- **Output:** π_j^M.

Example implementation of BPF

Consider **Dynamics:**

$$\begin{aligned} V_{j+1} &= 2.5\sin(V_j) + \xi_j \\ V_0 &\sim N(0,1) \end{aligned} \tag{2}$$

where $\xi_j \sim N(0, 0.09)$ **Observations:**

$$Y_j = h(V_j) + \eta_j, \quad j = 1, 2, \dots,$$

with $\eta_j \sim N(0,1)$.

Boostrap PF:

1. Sample iid $v_0^{(i)} \sim N(0,1)$ for $i = 1, 2, \dots, M$
2. Simulate $\hat{v}_1^{(i)} = 2.5\sin(v_0^{(i)}) + \xi_0^{(i)}$ for $i = 1, 2, \dots, M$.

Bootstrap PF continued

3. Set $w_1^{(i)} \propto \exp(-\frac{1}{2}|y_{j+1} - h(\hat{v}_{j+1}^{(i)})|_\Gamma^2)$ and normalize weights to sum to unity.

4. Set $\pi_1^M(du) = \sum_{i=1}^M w_1^{(i)} \delta_{\hat{v}_1^{(i)}}(du)$.

5. **Resampling:** Sample iid $v_1^{(i)} \sim \pi_1^M$ for $i = 1, 2, \ldots, M$

6. Simulate $\hat{v}_2^{(i)} = 2.5 \sin(v_1^{(i)}) + \xi_1^{(i)}$ for $i = 1, 2, \ldots, M$, and so forth.

How to sample from an empirical probability measure $\pi_j^M(du)$? Similar as sampling a transition in a finite state space Markov chain, cf. Lecture 5 Annotated, p. 35-36, and [SST 11.4].

Overview

Notation:

- Recall that $\mathcal{P}$ denotes the space of probability measures on $\mathbb{R}^d$.
- Let now π_j denote the distribution of $V_j | Y_{1:j} = y_{1:j}$ (rather than, as before, the pdf),
- and let π_j^M denote the corresponding BPF approximation.
- Recall that for any $f : \mathbb{R}^d \to \mathbb{R}$,

$$\pi_j[f] = \mathbb{E}^{\pi_j}[f] \quad \text{and} \quad \pi_j^M[f] = \mathbb{E}^{\pi_j^M}[f].$$

In order to study the large-particle-limit convergence of $\pi_j^M \to \pi_j$, we introduce the following metric on $\mathcal{P}$ (or, equivalently, on the space of pdfs $\mathcal{M}$)

$$d(\pi, \tilde{\pi}) := \sup_{\|f\|_\infty \leq 1} \sqrt{\mathbb{E}\left[\left(\pi_j[f] - \pi_j^M[f] \right)^2 \right]},$$

for $\pi, \tilde{\pi} \in \mathcal{P}$ (or $\in \mathcal{M}$).
Exercise: Verify that the triangle inequality holds.

Theorem 1 (SST 11.6)

Consider the dynamics-observation setting (1)*, and for a given sequence* $y_{1:J}$*, assume there exists a* $\kappa \in (0,1)$ *such that*

$$\kappa \leq \sup_{u \in \mathbb{R}^d} \pi_{Y_j|V_j}(y_j|u) \leq \kappa^{-1} \quad \text{for all } j \in \{0, 1, \ldots, J\}. \tag{3}$$

Then, for all $j \in \{0, 1, \ldots, J\}$*, it holds for the SIS algorithm 1 that*

$$d(\pi_j, \pi_j^M) \leq \frac{c(J, \kappa)}{\sqrt{M}}.$$

Remark: The assumption (3) never holds in the classic setting! See ubung 8 for settings where an adapted assumption holds.

Sketch of proof: Recall that

$$\pi_{j+1} = \mathscr{A}_{j+1} \mathscr{P} \pi_j \quad \text{and} \quad \pi_{j+1}^M = \mathscr{A}_{j+1} \mathcal{S}^M \mathscr{P} \pi_j^M.$$

Proof of Thm 1

Hence,

$$\begin{aligned} d(\pi_{j+1}, \pi_{j+1}^M) &= d\Big(\mathscr{A}_{j+1}\mathscr{P}\pi_j,\, \mathscr{A}_{j+1}\mathcal{S}^M\mathscr{P}\pi_j^M\Big) \\ &\leq d\Big(\mathscr{A}_{j+1}\mathscr{P}\pi_j,\, \mathscr{A}_{j+1}\mathscr{P}\pi_j^M\Big) + d\Big(\mathscr{A}_{j+1}\mathscr{P}\pi_j^M,\, \mathscr{A}_{j+1}\mathcal{S}^M\mathscr{P}\pi_j^M\Big) \\ &\leq \frac{2}{\kappa^2}\Big[d\Big(\mathscr{P}\pi_j,\, \mathscr{P}\pi_j^M\Big) + d\Big(\mathscr{P}\pi_j^M,\, \mathcal{S}^M\mathscr{P}\pi_j^M\Big)\Big], \end{aligned}$$

where in the last line we used that for any $\pi, \tilde{\pi} \in \mathcal{P}$, and $0 \leq j \leq J$,

$$d\Big(\mathscr{A}_j\pi,\, \mathscr{A}_j\tilde{\pi}\Big) \leq \frac{2}{\kappa^2} d\Big(\pi,\, \tilde{\pi}\Big). \tag{4}$$

Verification of (4): Let us write $g_j(u) := \pi_{Y_j|V_j}(y_j|u)$, and note that $\kappa \le g_j \le \kappa^{-1}$, and recall that for any $\tilde{\pi} \in \mathcal{P}$, the analysis operator is defined by

$$(\mathscr{A}_j\tilde{\pi})(du) = \frac{\pi_{Y_j|V_j}(y_j|u)\tilde{\pi}(du)}{\int \pi_{Y_j|V_j}(y_j|u)\tilde{\pi}(du)} = \frac{g_j(u)\tilde{\pi}(du)}{\tilde{\pi}[g_j]}.$$

Hence,

$$(\mathscr{A}_j\tilde{\pi})[f] = \int_{\mathbb{R}^d} f(u)(\mathscr{A}\tilde{\pi})(du) = \int_{\mathbb{R}^d} f(u)\frac{g_j(u)\tilde{\pi}(du)}{\tilde{\pi}[g]} = \frac{\tilde{\pi}[g_j f]}{\tilde{\pi}[g_j]}.$$

and

$$\begin{aligned}
|(\mathscr{A}_j\pi)[f] - (\mathscr{A}_j\tilde{\pi})[f]| &= \left|\frac{\pi[g_j f]}{\pi[g_j]} - \frac{\tilde{\pi}[g_j f]}{\tilde{\pi}[g_j]}\right| \\
&= \left|\frac{\pi[g_j f]}{\pi[g_j]} - \frac{\tilde{\pi}[g_j f]}{\pi[g_j]} + \frac{\tilde{\pi}[g_j f]}{\pi[g_j]} \quad - \quad \frac{\tilde{\pi}[g_j f]}{\tilde{\pi}[g_j]}\right| \\
&= \left|\frac{\pi[\kappa g_j f] - \tilde{\pi}[\kappa g_j f]}{\kappa\pi[g_j]} + \frac{\tilde{\pi}[g_j f]}{\tilde{\pi}[g_j]}\frac{(\tilde{\pi}[\kappa g_j] - \pi[\kappa g_j])}{\kappa\pi[g_j]}\right| \\
&\overset{\tilde{\pi}[g],\pi[g_j]>\kappa}{\le} \frac{\left|\pi[\kappa g_j f] - \tilde{\pi}[\kappa g_j f]\right|}{\kappa^2} + \left|\frac{\tilde{\pi}[g_j f]}{\tilde{\pi}[g_j]}\right|\frac{\left|\tilde{\pi}[\kappa g_j] - \pi[\kappa g_j]\right|}{\kappa^2}
\end{aligned}$$

Since

$$\left|\frac{\tilde{\pi}[g_j f]}{\tilde{\pi}[g_j]}\right| = |(\mathscr{A}_j\tilde{\pi})[f]| \leq 1,$$

we obtain that

$$\Big((\mathscr{A}_j\pi)[f] - (\mathscr{A}_j\tilde{\pi})[f]\Big)^2 \leq \frac{2}{\kappa^4}\left(\big(\pi[\kappa g_j f] - \tilde{\pi}[\kappa g_j f]\big)^2 + \big(\tilde{\pi}[\kappa g_j] - \pi[\kappa g_j]\big)^2\right)$$

Since $g_j \leq \kappa^{-1}$, it holds that $\|\kappa g_j\|_\infty \leq 1$ and $\|\kappa g_j f\|_\infty \leq \|f\|_\infty$, it follows that

$$\begin{aligned}
d(\mathscr{A}_j\pi, \mathscr{A}_j\tilde{\pi})^2 &= \sup_{\|f\|_\infty \leq 1} \mathbb{E}\left[\Big((\mathscr{A}_j\pi)[f] - (\mathscr{A}_j\tilde{\pi})[f]\Big)^2\right] \\
&\leq \sup_{\|f\|_\infty \leq 1} \frac{2}{\kappa^4}\left(\mathbb{E}\left[\big(\pi[\kappa g_j f] - \tilde{\pi}[\kappa g_j f]\big)^2 + \big(\tilde{\pi}[\kappa g_j] - \pi[\kappa g_j]\big)^2\right]\right) \\
&\leq \frac{4}{\kappa^4} \sup_{\|f\|_\infty \leq 1} \mathbb{E}\left[\Big(\pi[f] - \tilde{\pi}[f]\Big)^2\right].
\end{aligned}$$

Conclusion: $d(\mathscr{A}_j\pi, \mathscr{A}_j\tilde{\pi}) \leq \frac{2}{\kappa^2} d(\pi, \tilde{\pi})$.

We have reached

$$d(\pi_{j+1}, \pi^M_{j+1}) = \frac{2}{\kappa^2}\Big[d\Big(\mathscr{P}\pi_j,\ \mathscr{P}\pi^M_j\Big) + d\Big(\mathscr{P}\pi^M_j,\ \mathcal{S}^M\mathscr{P}\pi^M_j\Big)\Big].$$

For the last term, it follows by $\mathcal{S}^M\mathscr{P}\pi^M_j$ being an epm with iid dirac points, that

$$\begin{aligned} d\Big(\mathscr{P}\pi^M_j, \mathcal{S}^M\mathscr{P}\pi^M_j\Big) &= \sup_{\|f\|\leq 1} \mathbb{E}\left[\left((\mathscr{P}\pi^M_j)[f] - \sum_{i=1}^{M}\frac{f(\hat{v}^{(i)}_{j+1})}{M}\right)^2\right] \\ &\leq \frac{\mathrm{Var}^{\mathscr{P}\pi^M_j}[f]}{\sqrt{M}} \leq \sup_{\|f\|_\infty\leq 1}\frac{\mathrm{Var}^{\mathscr{P}\pi^M_j}[f]}{\sqrt{M}} \leq \frac{1}{\sqrt{M}}. \end{aligned}$$

And for the first term, we will show that

$$d\Big(\mathscr{P}\pi_j,\ \mathscr{P}\pi^M_j\Big) \leq d\Big(\pi_j,\ \pi^M_j\Big), \tag{5}$$

Verfication of (5), for any $\pi, \tilde{\pi} \in \mathcal{P}$,

$$
\begin{aligned}
(\mathscr{P}\pi)[f] - (\mathscr{P}\tilde{\pi})[f] &= \int_{\mathbb{R}^d} f(v)\Big(\mathscr{P}\pi)(v) - (\mathscr{P}\tilde{\pi})(v)\Big)dv \\
&= \int_{\mathbb{R}^d} f(v) \int_{\mathbb{R}^d} p(u,v)(\pi(du) - \tilde{\pi}(du))dv \\
&= \int_{\mathbb{R}^d} \left(\int_{\mathbb{R}^d} f(v)p(u,v)dv\right) (\pi(du) - \tilde{\pi}(du)) \\
&= \int_{\mathbb{R}^d} q_f(u)(\pi(du) - \tilde{\pi}(du)) = \pi[q] - \tilde{\pi}[q].
\end{aligned}
$$

and $\|q_f\|_\infty$ whenever $\|f\|_\infty \leq 1$.
Consequently,

$$
\begin{aligned}
d\Big(\pi, \tilde{\pi}\Big)^2 &= \sup_{\|f\| \leq 1} \mathbb{E}\left[\Big((\mathscr{P}\pi)[f] - (\mathscr{P}\tilde{\pi})[f]\Big)^2\right] \\
&= \sup_{\|f\| \leq 1} \mathbb{E}\left[\Big(\pi[q_f] - \tilde{\pi}[q_f]\Big)^2\right] \\
&\leq \sup_{\|q\| \leq 1} \mathbb{E}\left[\Big(\pi[q] - \tilde{\pi}[q]\Big)^2\right] = d\Big(\pi, \tilde{\pi}\Big)^2.
\end{aligned}
$$

Conclusion

$$
\begin{aligned}
d(\pi_{j+1}, \pi_{j+1}^M) &= d\Big(\mathscr{A}_{j+1}\mathscr{P}\pi_j,\, \mathscr{A}_{j+1}\mathcal{S}^M\mathscr{P}\pi_j^M\Big) \\
&\leq \frac{2}{\kappa^2}\Big[d\Big(\mathscr{P}\pi_j,\, \mathscr{P}\pi_j^M\Big) + d\Big(\mathscr{P}\pi_j^M,\, \mathcal{S}^M\mathscr{P}\pi_j^M\Big)\Big] \\
&\leq \frac{2}{\kappa^2}\Big(d\Big(\pi_j,\, \pi_j^M\Big) + \frac{1}{\sqrt{M}}\Big) \\
&\leq \ldots \leq \left(\frac{2}{\kappa^2}\right)^{j+1} \underbrace{d\Big(\pi_0,\, \pi_0^M\Big)}_{=0} + \frac{\sum_{k=0}^{j}\left(\frac{2}{\kappa^2}\right)^k}{\sqrt{M}}.
\end{aligned}
$$

End of proof.

Overview

Other proposals

In the SIS and SIR algorithms we have considered, given $\{(w_j^{(i)}, v_j^{(i)})\}$, the dynamics simulation for the next step reads

- "Simulate $\hat{v}_{j+1}^{(i)} = F(v_j^{(i)}, \xi_j^{(i)})$ with iid $\xi_j^{(i)}$"

  This could also have been written
- "Draw independent $\hat{v}_{j+1}^{(i)} \sim \pi_{V_{j+1}|V_j}(\cdot|v_j^{(i)})$ for $i = 1, \dots, M$".
- This idea has a weakness: for SIS, the particles $\hat{v}_j^{(i)}$ have precisely the same distribution as the true dynamics V_j for every $j \geq 0$, ignoring completely the information from observations.
- This often leads to degeneracy: $n_{eff,j} \ll M$.
- To avoid degeneracy, one can sample from other "dynamics"/kernel density than $\pi_{V_{j+1}|V_j}(\cdot|v_j^{(i)})$ that takes $y_{1:j+1}$ into account.
- Generic notation for kernel density: $\rho(v_{j+1}|v_j, y_{1:j+1})$, it can for instance be

$$\rho(v_{j+1}|v_j, y_{1:j+1}) = \pi_{V_{j+1}|V_j, Y_{1:j+1}}(\cdot|v_j^{(i)}, y_{1:j+1})$$

Change of dynamics/kernel density

Recall that for the Bayes filter

$$\begin{aligned}\pi_{j+1}(v_{j+1}) &\propto \pi(y_{j+1}|v_{j+1})\pi(v_{j+1}|y_{1:j})\\ &= \int_{\mathbb{R}^d} \underbrace{\pi(y_{j+1}|v_{j+1})}_{\text{“weight”}} \underbrace{\pi(v_{j+1}|v_j)}_{\text{“kernel density”}} \pi_j(v_j)dv_j,\end{aligned}$$

and for the particle filters this is approximated by

$$\pi_{j+1}^M = \mathscr{A}_{j+1}\mathcal{S}^M\mathscr{P}\pi_j^M = \sum_{i=1}^M w_{j+1}^{(i)}\delta_{\hat{v}_{j+1}^{(i)}}$$

with $\hat{v}_{j+1}^{(i)} \sim \int \pi_{V_{j+1}|V_j}(\cdot|v_j)\pi_j^M(v_j)dv_j$ and $w_{j+1}^{(i)} \propto \pi_{Y_{j+1}|V_{j+1}}(y_{j+1}|\hat{v}_{j+1}^{(i)})$

We replace the kernel density by $\rho(v_{j+1}|v_j, y_{1:j+1})$ as follows

$$\pi_{j+1}(v_{j+1}) \propto \int \pi(y_{j+1}|v_{j+1})\pi(v_{j+1}|v_j)\pi_j(v_j)dv_j$$

$$= \int \underbrace{\frac{\pi(y_{j+1}|v_{j+1})\pi(v_{j+1}|v_j)}{\rho(v_{j+1}|v_j, y_{1:j+1})}}_{\text{“weight”}} \underbrace{\rho(v_{j+1}|v_j, y_{1:j+1})}_{\text{“dynamics”}} \pi_j(v_j)dv_j$$

Constraint for the kernel density: Given $y_{1:j+1}$, it must hold for any $v_j, v_{j+1} \in \mathbb{R}^d$ such that

$$\pi(y_{j+1}|v_{j+1})\pi(v_{j+1}|v_j) > 0, \quad \text{that also} \quad \rho(v_{j+1}|v_j, y_{1:j+1}) > 0.$$

Essential idea for the modified particle filter:

$$\pi^M_{j+1} = \sum_{i=1}^{M} w^{(i)}_{j+1}\delta_{\hat{v}^{(i)}_{j+1}}, \qquad \text{with } \hat{v}^{(i)}_{j+1} \sim \int \rho(\cdot|v_j, y_{1:j+1})\pi^M_j(v_j)dv_j$$

$$\text{and} \qquad w^{(i)}_{j+1} \propto \frac{\pi_{Y_{j+1}|V_{j+1}}(y_{j+1}|\hat{v}^{(i)}_{j+1})\pi_{V_{j+1}|V_j}(\hat{v}^{(i)}_{j+1}|v^{(i)}_j)}{\rho(\hat{v}^{(i)}_{j+1}|v^{(i)}_j, y_{1:j+1})}$$

More general sequential importance resampling algorithm

- **Input:** Initial distribution π_0 (which we also write π_0^M), obs sequence $y_1, y_2, \ldots$, and M.
- **Particle generation:** For $j = 0, 1, \ldots,$,

1. **Resampling** Draw $v_j^{(i)} \overset{iid}{\sim} \pi_j^M$ for $i = 1, \ldots, M$.
2. Draw independent $\hat{v}_{j+1}^{(i)} \sim \rho(\cdot|v_j^{(i)}, y_{1:j+1})$ for $i = 1, \ldots, M$.
3. Set
$$\bar{w}_{j+1}^{(i)} = \frac{\pi_{Y_{j+1}|V_{j+1}}(y_{j+1}|\hat{v}_{j+1}^{(i)})\pi_{V_{j+1}|V_j}(\hat{v}_{j+1}^{(i)}|v_j^{(i)})}{\rho(\hat{v}_{j+1}^{(i)}|v_j^{(i)}, y_{1:j+1})}$$
4. and $w_{j+1}^{(i)} = \bar{w}_{j+1}^{(i)} / \sum_{k=1}^M \bar{w}_{j+1}^{(k)}$.
5. Set $\pi_{j+1}^M = \sum_{i=1}^M w_{j+1}^{(i)} \delta_{\hat{v}_{j+1}^{(i)}}$.

- **Output:** π_j^M approximating the distribution of $V_j | Y_{1:j} = y_{1:j}$.

Modified Sequential importance sampling algorithm

- **Input:** Initial distribution π_0, obs sequence $y_1, y_2, \ldots$, and M.
- **Initialization:** Draw $\hat{v}_j^{(i)} \overset{iid}{\sim} \pi_0$ and set $w_0^{(i)} = 1/M$ for $i = 1, \ldots, M$. (Hat notation here is formally "wrong" but practical.)
- **Particle and weight dynamics:** For $j = 0, 1, \ldots,$,

  1. Draw independent $\hat{v}_{j+1}^{(i)} \sim \rho(\cdot|\hat{v}_j^{(i)}, y_{1:j+1})$ for $i = 1, \ldots, M$.
  2. Set
  $$\bar{w}_{j+1}^{(i)} = \frac{\pi_{Y_{j+1}|V_{j+1}}(y_{j+1}|\hat{v}_{j+1}^{(i)})\pi_{V_{j+1}|V_j}(\hat{v}_{j+1}^{(i)}|\hat{v}_j^{(i)})}{\rho(\hat{v}_{j+1}^{(i)}|\hat{v}_j^{(i)}, y_{1:j+1})}$$
  3. and $w_{j+1}^{(i)} = \bar{w}_{j+1}^{(i)} / \sum_{k=1}^M \bar{w}_{j+1}^{(k)}$.
  4. Set $\pi_{j+1}^M = \sum_{i=1}^M w_{j+1}^{(i)} \delta_{\hat{v}_{j+1}^{(i)}}$.

- **Output:** π_j^M approximating the distribution of $V_j | Y_{1:j} = y_{1:j}$.

Sampling from a different kernel density

- Sampling from the kernel density

$$\pi_{V_{j+1}|V_j,Y_{j+1}}(\cdot|v_j, y_{j+1}) \tag{6}$$

in SIS gives you the so called optimal particle filter. Meaning

$$\mathrm{Var}^{\pi_{V_{j+1}|V_j,Y_{1:j}}(\cdot|\hat{v}_j^{(i)},y_{1:j+1})}[\bar{w}_{j+1}^{(i)}] = \inf_{\rho(\cdot|\hat{v}_j^{(i)},y_{1:j+1})} \mathrm{Var}^{\rho(\cdot|\hat{v}_j^{(i)},y_{1:j+1})}[\bar{w}_{j+1}^{(i)}]$$

- In other words, of all possible kernel densities $\rho(\cdot|v_j, y_{1:j+1})$, sampling from (6) leads to the minimum variance in $\bar{w}_j^{(i)}$.

- See [SST 12.3] for a setting where it actually is possible to sample from $\pi_{V_{j+1}|V_j,Y_{j+1}}(\cdot|v_j, y_{j+1})$.

Summary and next lecture

- Particle filter is an unbiased filtering method which converges weakly to the Bayes filter in the large-ensemble limit.

- It is applicable also in settings both with nonlinear Ψ and h, and also for more general hidden Markov models.

- Degeneracy is an important issue for particle filters, particularly for high-dimensional problems. It is an ongoing research topic to understand this phenomenon and develop more robust particle filters.

- Next time: Continuous time stochastic processes in the form of Wiener processes, Ito integration and Ito stochastic differential equations.