Mathematics and numerics for data assimilation and state estimation – Lecture 13

Summer semester 2020

Overview

Summary of lecture 12

- Monte Carlo methods for sampling π:

$$\pi_{MC}^{M}[f] = \sum_{k=1}^{M} \frac{f(U_k)}{M}, \qquad \text{where} \quad U_k \overset{iid}{\sim} \pi$$

- Sampling the target (exactly or approximately) π indirectly through change of measure or an auxiliary/proposal distribution $\hat{\pi}$.

- Discrete-time continuous-space Markov chains

- Metropolis Hastings MCMC method.

Overview

Markov Chain Monte Carlo method (MCMC)

Input: target pdf π, a conditional proposal $q(y|x)$ (i.e., $q(\cdot|x) \in \mathcal{M}$ for every $x \in \mathbb{R}^d$).

Output: Markov chain $X_0, X_1, \ldots$ with objective that $\pi_{MCMC}^M = \frac{1}{M}\sum_{k=1}^M \delta_{X_k}$ approximates measure associated to π.

Metropolis-Hastings algorithm

Given X_n,

1. generate proposal $Y_n \sim q(\cdot|X_n)$
2. set
$$X_{n+1} = \begin{cases} Y_n & \text{with probability} \quad \rho(X_n, Y_n) \\ X_n & \text{with probability} \quad 1-\rho(X_n, Y_n) \end{cases}$$
where the M-H acceptance probability is defined by
$$\rho(x, y) = \min\left(\frac{\pi(y)}{\pi(x)}\frac{q(x|y)}{q(y|x)}, 1\right)$$

Assumptions and properties of Metropolis Hastings

Assumptions

- must be able to sample from $q(\cdot|x)$ for relevant x
- π must be known up to a constant (i.e., relevant for posterior densities with Z unknown),
- $q(\cdot|x)$ must be known up to a constant that is independent of x.

Properties:

- When $q(x|y) = q(y|x)$ the test ratio becomes
$$\frac{\pi(y)}{\pi(x)}\frac{q(x|y)}{q(y|x)} = \frac{\pi(y)}{\pi(x)}.$$
- If $q(x|y) > q(y|x)$, then (compared to not having a q ratio in the acceptance probability), the probability accepting transitions $x \mapsto y$ increases. So transitions for which the reverse transition $q(x|y)$ is more often proposed than the transition itself, increases likelihood.
- If $q(x|y) < q(y|x)$, then (compared to not having a q ratio in the acceptance probability), the probability accepting transitions $x \mapsto y$ decreases.

Effect of M-H acceptance

Top row: Markov chain with kernel density $k(u, v) = q(v|y)$.

Bottom row: M-H transforms kernel density to new kernel "density"

$$p(u, v) = \rho(u, v)q(v|u), \quad \text{with} \quad \rho(u, v) = \min\left(\frac{\pi(v)}{\pi(u)}\frac{q(u|v)}{q(v|u)}, 1\right)$$

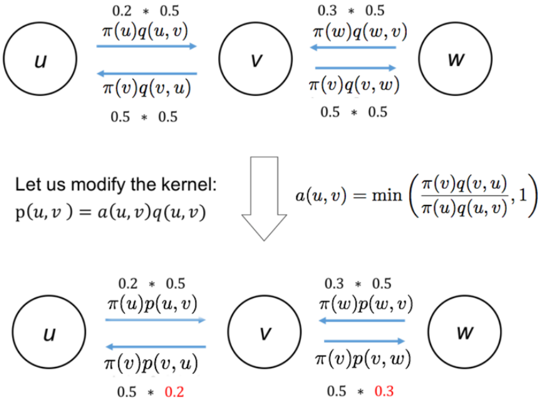


Figure: From Data Assimilation and Inverse Problems, Sanz-Alonso et al.

M-H dynamics is associated to the transition kernel (ubung 5)

$$K(x, A) = \underbrace{\int_A \rho(u, v) q(y|x) dy}_{r(x,A)} + \Big(1 - r(x, \mathbb{R}^d)\Big)\delta_x(A)$$

Idea:

$$\begin{aligned} K(x, A) &= \mathbb{P}(X_1 \in A \mid X_0 = x) \\ &= \mathbb{P}(Y_0 \in A, X_1 = Y_0 \mid X_0 = x) + \mathbb{P}(x \in A, X_1 = x \mid X_0 = x) \end{aligned}$$

M-H properties

If $q(\cdot|x)$ dominates π for all x, then the M-H kernel satisfies detailed balance wrt π:

$$\int_A K(x, B)\pi(x)dx = \int_B K(x, A)\pi(x)dx \qquad \forall A, B \in \mathcal{B}^d,$$

and π is an invariant pdf of the M-H Markov chain.

Sketch of proof: Assume that $X_0 \sim \pi$. Then

$$\begin{aligned}
\mathbb{P}_1(A) &= \int_{\mathbb{R}^d} K(x, A)\mathbb{P}_{X_0}(dx) \\
&= \int_{\mathbb{R}^d} K(x, A)\pi(x)dx \\
&= \int_A K(x, \mathbb{R}^d)\pi(x)\, dx \\
&= \int_A \pi(x)dx = \mathbb{P}_0(A)
\end{aligned}$$

Remarks

Challenges in real applications: Choosing a proposal such that (1) one achieves convergence $\pi^n \to \pi$, (2) the convergence is fast in n, and (3) that acceptance of the proposal is frequent (for efficiency of MCMC).

See SST 6.4.2 for assumptions on prior and likelihood for $\pi(\cdot|y)$ in combination with Gaussian proposal $q(\cdot|x)$ which ensures convergence of the chain distribution.

If interested, "Monte Carlo Statistical Methods" by Robert and Casella is a good book on Monte Carlo and MCMC methods.

Overview

Dynamics and observation setting

Continuous state-space dynamics: A mapping $\Psi \in C(\mathbb{R}^d, \mathbb{R}^d)$ is associated to the dynamics

$$\begin{aligned} V_{j+1} &= \Psi(V_j) + \xi_j, \quad j = 0, 1, \ldots \\ V_0 &\sim N(m_0, C_0) \end{aligned} \tag{1}$$

where $\{\xi_j\}$ is iid $\xi \sim N(0, \Sigma)$-distributed and $V_0 \perp \{\xi_j\}$.

Observations:

$$Y_j = h(V_j) + \eta_j, \quad j = 1, 2, \ldots, \tag{2}$$

where $h \in C(\mathbb{R}^d, \mathbb{R}^k)$ and $\{\eta_j\}$ is iid with $\eta_1 \sim N(0, \Gamma)$.

Independence assumptions:

$$\{\eta_j\} \perp \{\xi_j\} \quad \text{and} \quad \{\eta_j\} \perp V_0.$$

Objectives: Study the smoothing pdf of $V_{0:J} | Y_{1:J} = y_{1:J}$.

Examples of Ψ

In many applications, Ψ can be associated to a solution of a time-invariant ODE:

$$\begin{aligned} \dot{v} &= f(v), \quad t \geq 0 \\ v(0) &= v_0 \end{aligned} \tag{3}$$

Viewing v_0 as a **variable**, let us denote the solution of (3) at time s by $\Psi(v_0; s)$.

For a fixed interval $\tau > 0$ and any $V \in \mathbb{R}^d$, we define

$$\Psi(V) := \Psi(V; \tau).$$

For later reference, let us also introduce

$$\Psi^{(j)}(V) := \underbrace{\Psi \circ \Psi \circ \ldots \circ \Psi}_{j \text{ times}}(V) = \Psi(V; j\tau).$$

Guiding examples

The scalar-valued ODE

$$\begin{aligned} \dot{v} &= \log(\lambda) v, \quad t \geq 0 \\ v(0) &= v_0 \end{aligned} \tag{4}$$

and $\tau = 1$ yields

$$\Psi(V) = e^{\log(\lambda)\tau} V = \lambda V.$$

The dynamics

$$V_{j+1} = \lambda V_j + \xi_j, \quad j = 0, 1, \ldots$$

with $\xi \sim N(0, \sigma^2)$ is fundamentally different when $|\lambda| < 1$ and $|\lambda| > 1$.

Using that

$$\mathbb{E}\left[V_{j+1}\right] = \lambda \mathbb{E}\left[V_j\right], \quad \mathbb{E}\left[V_{j+1}^2\right] = \lambda^2 \mathbb{E}\left[V_j^2\right] + \sigma^2,$$

one can show that when $|\lambda| < 1$,

$$\mathbb{P}_{V_n} \Rightarrow N(0, \frac{\sigma^2}{1 - \lambda^2}).$$

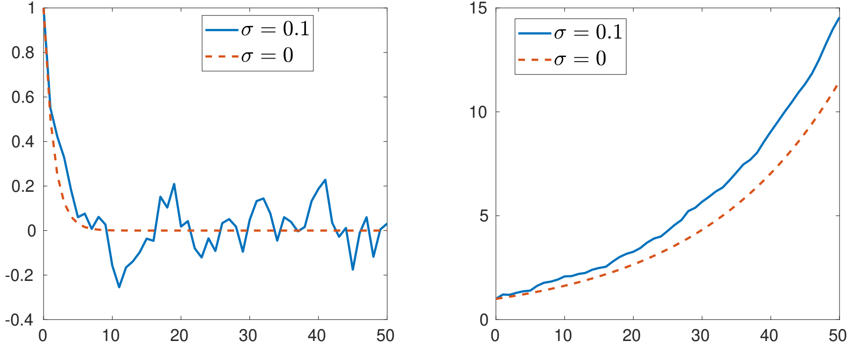


Figure: Dynamics of $V_{j+1} = \lambda V_j + \xi_j$ with $\lambda = 0.5$ (left) and $\lambda = 1.05$ (right).

Nonlinear dynamics

For

$$\Psi(v) = \alpha \sin(v)$$

the deterministic dynamics

$$V_{j+1} = \alpha \sin(V_j)$$

is sensitive to the initial condition.

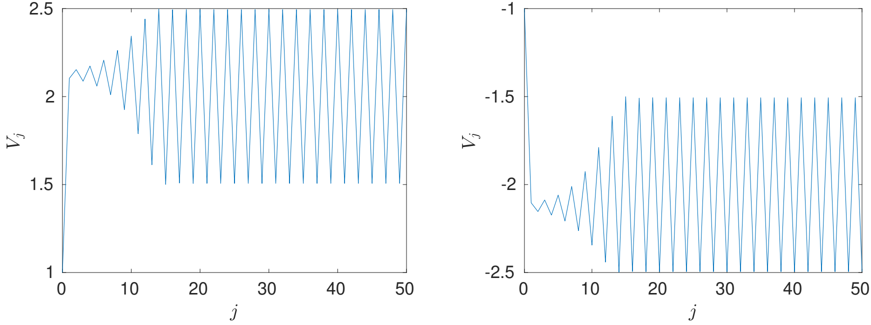


Figure: Dynamics of with $\alpha = 2.5$ and $V_0 = 1$ (left) and $V_0 = -1$ (right)

The stochastic dynamics

$$V_{j+1} = \alpha \sin(V_j) + \xi_j, \quad \xi \sim N(0, \sigma^2).$$

is not sensitive to the initial condition, if one views

$$\mathbb{P}_V(\cdot) := \lim_{J \to \infty} \frac{1}{J} \sum_{j=1}^{J} \delta_{V_j}(\cdot)$$

as the relevant feature (take as soft motivation, have not even shown that this measure exists).

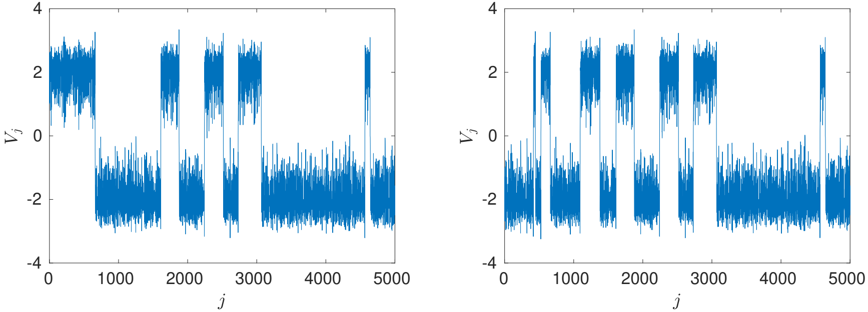


Figure: $\alpha = 2.5$, $\sigma = 1/4$ and $V_0 = 1$ (left) and $V_0 = -1$ (right)

The stochastic dynamics

$$V_{j+1} = \alpha \sin(V_j) + \xi_j, \quad \xi \sim N(0, \sigma^2).$$

is not sensitive to the initial condition, if one views

$$\mathbb{P}_V(\cdot) := \lim_{J \to \infty} \frac{1}{J} \sum_{j=1}^{J} \delta_{V_j}(\cdot)$$

as the relevant feature (take as soft motivation; we have not shown that this measure exists).

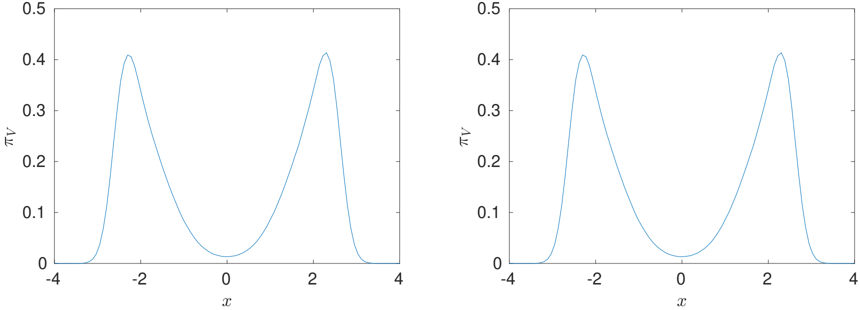


Figure: $\pi_V = PDF(\mathbb{P}_V)$ for $\alpha = 2.5$, $\sigma = 1/4$, $J = 10^7$, and $V_0 = 1$ (left) and $V_0 = -1$ (right)

Lorenz '63

Is the system of ODE

$$\left.\begin{aligned}\dot{v}_1 &= a(v_2 - v_1)\\ \dot{v}_2 &= -av_1 - v_2 - v_1 v_3\\ \dot{v}_3 &= v_1 v_2 - bv_3 - b(r+a)\end{aligned}\right\} =: f(v), \qquad t \geq 0,$$

where $a, b, r > 0$ and $v(0) \in \mathbb{R}^3$.

For some $\alpha, \beta > 0$, depending on vector field, it can be shown that

$$f(v)^T v \leq \alpha - \beta |v|^2.$$

This ensures that [LSZ Example 1.22]

$$\limsup_{t \to \infty} |v(t)|^2 \leq \frac{\alpha}{\beta}$$

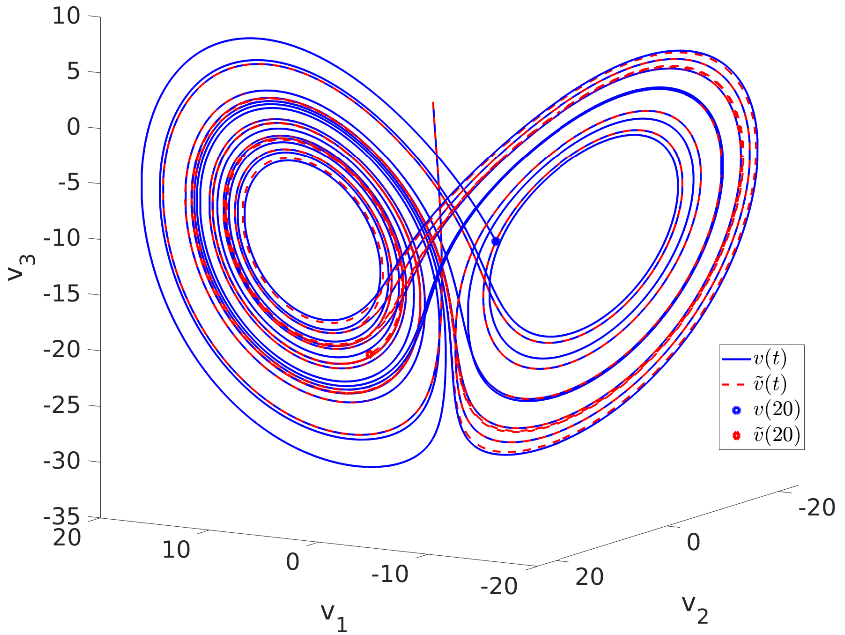
For any $|v(0)| \leq \alpha/\beta$ there exists a unique solution, **see ubung 6**, but $v(t)$ is very sensitive to the initial condition!

Integration in Matlab with parameter values $(a, b, r) = (10, 8/3, 28)$
$v(0) = (1, 1, 1)$ and $\tilde{v}(0) = (1, 1, 1 + 10^{-5})$:

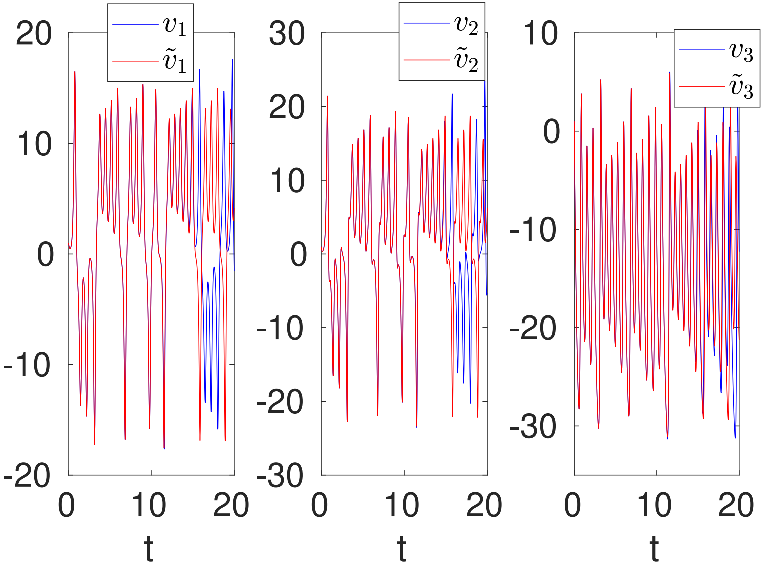
```
options = odeset('RelTol',1e-12,'AbsTol',1e-10);
a = 10;
b = 8/3;
r = 28;
f = @(t,v) [a*(v(2)-v(1));
            -a*v(1)-v(2)-v(1)*v(3);
            v(1)*v(2)-b*v(3)-b*(r+a)];

[t,v]=ode45(f,[0 20],[1 1 1], options);%RK 4/5 order ODE solve
[t2,vTilde] = ode45(f,[0 20],[1 1 1+1e-5], options);
```

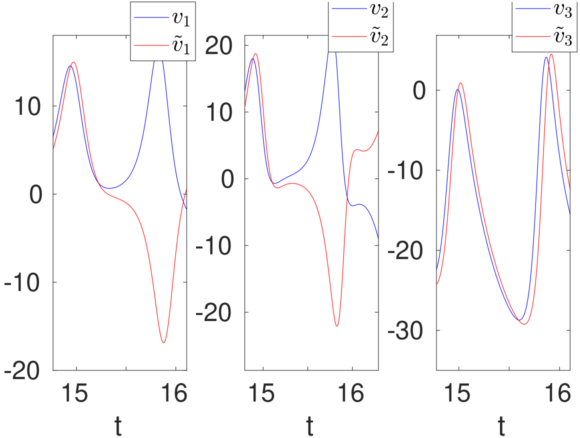
Result: $|v(0) - \tilde{v}(0)| = 10^{-5}$ and $|v(20) - \tilde{v}(20)| \approx 15.7$



v3
v1
v2
10
5
0
-5
-10
-15
-20
-25
-30
-35
20
10
0
-10
-20
-20
0
20
$v(t)$
$\tilde{v}(t)$
$v(20)$
$\tilde{v}(20)$



v_1
$\tilde{v}_1$
v_2
$\tilde{v}_2$
v_3
$\tilde{v}_3$
20
10
0
-10
-20
30
-30
0
10
20
t
t
t



If $v_1(s) = v_2(s) = 0$, then $(v_1, v_2) = (0,0)$ for all later times: when v_3 is sufficiently negative, it is an unstable stationary point on the (v_1, v_2)-subspace.

$$\dot{v}_1 = a(v_2 - v_1)$$
$$\dot{v}_2 = -av_1 - v_2 - v_1 v_3$$
$$\dot{v}_3 = v_1 v_2 - bv_3 - b(r + a)$$

Overview

Smoothing

Given dynamics

$$V_{j+1} = \Psi(V_j) + \xi_j, \quad \xi \sim N(0, \Sigma)$$
$$V_0 \sim N(m_0, C_0)$$

and observations

$$Y_j = h(V_j) + \eta_j, \quad \eta \sim N(0, \Gamma)$$

with $h \in C(\mathbb{R}^d, \mathbb{R}^k)$ and $V_0 \perp \{\eta_j\} \perp \{\xi_j\}$.

Objectives: given $y_{1:J} \in \mathbb{R}^{k \times J}$,

- derive the pdf for smoothing problem:

$$\pi_{V_{0:J}|Y_{1:J}}(v_{0:J}|y_{1:J}) =: \pi(v_{0:J}|y_{1:J})$$

- verify that the smoothing problem is stable wrt perturbations in $y_{1:J} \in \mathbb{R}^{k \times J}$. That is, show that

$$|y_{1:J} - \tilde{y}_{1:J}| = \mathcal{O}(\delta) \implies d_H(\pi(\cdot|y_{1:J}), \pi(\cdot|\tilde{y}_{1:J})) = \mathcal{O}(\delta)$$

The smoothing pdf

By Bayes' rule and the Bayesian viewpoint

$$\pi(v_{0:J}|y_{1:J}) \propto \underbrace{\pi(y_{1:J}|v_{0:J})}_{\text{Likelihood}} \underbrace{\pi(v_{0:J})}_{\text{Prior}}$$

Prior: Note that $\{V_j\}$ is a Markov chain, hence

$$\begin{aligned}\pi(v_{0:J}) &= \pi(v_J|v_{0:J-1})\pi(v_{0:J-1}) = \pi(v_J|v_{J-1})\pi(v_{0:J-1}) \\ &= \ldots = \prod_{j=0}^{J-1} \pi(v_{j+1}|v_j)\pi_{V_0}(v_0).\end{aligned}$$

And

$$V_0 \sim N(m_0, C_0) \implies \pi_{V_0}(v_0) \propto \exp(-\frac{1}{2}|v_0 - m_0|^2_{C_0}),$$

and

$$V_{j+1}|(V_j = v_j) \;=\; (\Psi(V_j) + \underbrace{\eta_j}_{\sim N(0,\Sigma)})|(V_j = v_j) \sim N(\Psi(v_j), \Sigma)$$

$$\implies \pi(v_{j+1}|v_j) \propto \exp(-\frac{1}{2}|v_{j+1} - \Psi(v_j)|^2_{\Sigma})$$

Prior:

$$\pi(v_{0:J}) = \frac{1}{Z_P}\exp(-\mathrm{R}(v_{0:J}))$$

where

$$\mathrm{R}(v_{0:J}) := \frac{1}{2}|v_0 - m_0|^2_{C_0} + \frac{1}{2}\sum_{j=0}^{J-1}|v_{j+1} - \Psi(v_j)|^2_{\Sigma}.$$

Next,

$$\pi(v_{0:J}|y_{1:J}) \propto \underbrace{\pi(y_{1:J}|v_{0:J})}_{\text{Likelihood}}\underbrace{\pi(v_{0:J})}_{\text{Prior}}$$

Likelihood: Since $Y_j = h(V_j) + \eta_j$ and $V_0 \perp \{\eta_j\} \perp \{\xi_j\}$,

$$
\begin{aligned}
Y_{1:J}|(V_{0:J} = v_{0:J}) \;\; &= \;\; (\, Y_1|(V_1 = v_1)\,,\; \ldots \;,\; Y_J|(V_J = v_J)\,) \\
&= (h(v_1) + \eta_1, \ldots, h(v_J) + \eta_J)
\end{aligned}
$$

with independent components and $h(v_j) + \eta_j \sim N(h(v_j), \Gamma)$.
Hence,

$$
\pi(y_{1:J}|v_{0:J}) = \prod_{j=1}^{J} \pi(y_j|v_j) \propto \exp(-\mathsf{L}(v_{1:J}; y_{1:J}))
$$

with

$$
\mathsf{L}(v_{1:J}; y_{1:J}) := \frac{1}{2} \sum_{j=1}^{J} |h(v_j) - y_j|_\Gamma^2.
$$

Smoothing pdf

Theorem 1

For the dynamics-observation sequence (1) *and* (2) *with* $Y_{1:J} = y_{1:J}$, *we obtain*

$$\begin{aligned}\pi(v_{0:J}|y_{1:J}) &= \frac{1}{Z}\exp(-L(v_{1:J}; y_{1:J}) - R(v_{0:J}))\\ &= \frac{1}{Z}\exp\Bigg(-\frac{1}{2}\sum_{j=1}^{J}|h(v_j) - y_j|_{\Gamma}^2 \\ &\qquad -\frac{1}{2}|v_0 - m_0|_{C_0}^2 - \frac{1}{2}\sum_{j=0}^{J-1}|v_{j+1} - \Psi(v_j)|_{\Sigma}^2\Bigg)\end{aligned}$$

where $v_{0:J} \in \mathbb{R}^{d\times(J+1)}$ *and the normalizing constant* Z *depends on* $y_{1:J} \in \mathbb{R}^{k\times J}$

Next question: How stable is the pdf wrt perturbations in $y_{1:J}$?

Well-posedness of the smoothing pdf

Theorem 2 (LSZ 2.15)

Fix $J \in \mathbb{N}$, a pair of observation sequences $y_{1:J}, \tilde{y}_{1:j} \in \mathbb{R}^{k\times J}$, and assume that the dynamics V_j satisfies

$$\mathbb{E}\left[\sum_{j=0}^{J}(1 + |h(V_j)|^2)\right] < \infty.$$

Then there exists a constant $c > 0$ that depends on $y_{1:J}$ and $\tilde{y}_{1:J}$) such that

$$d_H(\pi(\cdot|y_{1:J}), \pi(\tilde{|}\tilde{y}_{1:J}) \leq c\sqrt{\sum_{j=1}^{J}|y_j - \tilde{y}_j|^2}$$

Proof ideas:

$$\pi(v_{0:J}|y_{1:J}) = \frac{1}{Z}\exp(-\mathsf{L}(v_{1:J}; y_{1:J}) - \mathsf{R}(v_{0:J}))$$

and

$$\pi(v_{0:J}|\tilde{y}_{1:J}) = \frac{1}{\tilde{Z}}\exp(-\mathsf{L}(v_{1:J}; \tilde{y}_{1:J}) - \mathsf{R}(v_{0:J}))$$

Results follows from showing that

$$Z, \tilde{Z} > K > 0 \quad \text{and} \quad |Z - \tilde{Z}| = \mathcal{O}(|y_{1:J} - \tilde{y}_{1:J}|)$$

and that

$$\begin{aligned}|\mathsf{L}(v_{1:J}; y_{1:J}) - \mathsf{L}(v_{1:J}; \tilde{y}_{1:J})| &= \frac{1}{2}\sum_{j=1}^{J}\Big||h(v_j) - y_j|_\Gamma^2 - |h(v_j) - \tilde{y}_j|_\Gamma^2\Big| \\ &= \mathcal{O}(|y_{1:J} - \tilde{y}_{1:J}|).\end{aligned}$$

Hint for bounding the loss-term difference: for $u, v \in \mathbb{R}^k$,

$$|u|_\Gamma^2 - |v|_\Gamma^2 = \langle u + v, u - v\rangle_\Gamma$$

where $\langle u, v\rangle_\Gamma := u^T\Gamma^{-1}v$.

Overview

Smoothing problem – deterministic dynamics

Consider the simplified version of (1) where the dynamics is deterministic (but with random initial data):

$$
\begin{aligned}
V_{j+1} &= \Psi(V_j), \qquad j = 0, 1, \dots, \\
V_0 &\sim N(m_0, C_0)
\end{aligned}
$$

with observations $j = 1, 2, \dots$

$$Y_j = h(V_j) + \eta_j, \quad \eta \sim N(0, \Gamma)$$

with $h \in C(\mathbb{R}^d, \mathbb{R}^k)$ and $V_0 \perp \{\eta_j\}$.

Then, given $Y_{1:J} = y_{1:J}$, we now have that $V_{0:J}$ **only is random in** V_0, since using that $V_j = \Psi^{(j)}(V_0)$,

$$V_{0:J} = (V_0, \Psi(V_0), \Psi^{(2)}(V_0), \dots, \Psi^{(J)}(V_0)).$$

Consequently, we now seek to determine the pdf of $V_0 | Y_{1:J} = y_{1:J}$:

$$\pi(v_0 | y_{1:J}) \propto \underbrace{\pi(y_{1:J} | v_0)}_{\text{Likelihood}} \underbrace{\pi_{V_0}(v_0)}_{\text{Prior}}.$$

Likelihood : Since

$$Y_j = h(V_j) + \eta_j = h(\Psi^{(j)}(V_0)) + \eta_j$$

we obtain that

$$Y_{1:J}|(V_0 = v_0) \; = \; (h(\Psi^{(1)}(v_0)) + \eta_1, h(\Psi^{(2)}(V_0)) + \eta_2, \ldots, h(\Psi^{(J)}(V_0)) + \eta_J).$$

This yields

$$\pi(y_{1:J}|v_0) = \prod_{j=1}^{J} \pi(y_j|v_0) \propto \exp\left(\underbrace{- \frac{1}{2} \sum_{j=1}^{J} \left| y_{j+1} - h(\Psi^{(j)}(v_0)) \right|_{\Gamma}^{2}}_{= \mathsf{L}(v_0; y_{1:J})} \right)$$

and the posterior

$$\pi(v_0|y_{1:J}) \propto \exp\Big(- \mathsf{L}(v_0; y_{1:J}) - \underbrace{\frac{1}{2} |v_0 - m_0|_{C_0}^{2}}_{= \mathsf{R}(v_0)} \Big)$$

Numerical study

For the dynamics

$$V_{j+1} = \lambda V_j$$

with $V_0 \sim N(m_0, \sigma_0^2)$ and

$$Y_j = V_j + \eta_j, \quad \eta \sim N(0, \gamma^2)$$

it can be shown that

$$\pi(v_0|y_{1:J}) \propto \exp\Big(-\frac{1}{2\gamma^2}\sum_{j=1}^{J} |y_{j+1} - \lambda^j v_0|^2 - \underbrace{\frac{1}{2\sigma_0^2}|v_0 - m_0|^2_{C_0}}_{=\mathrm{R}(v_0)} \Big)$$

and completing squares in the exponent yields that

$$V_0|(Y_{1:J} = y_{1:J}) \sim N(m, \sigma_{post}^2)$$

If $|\lambda| < 1$, then

$$\lim_{J\to\infty} \sigma_{post}^2 \overset{a.s.}{=} \frac{\gamma^2}{\lambda^2/(1-\lambda^2) + \gamma^2/\sigma_0^2} \quad \text{(so uncertainty remains for large } J\text{).}$$

But cases when either $\lambda^2 \approx 1$ and/or $\gamma \approx 0$ reduce uncertainty (ubung 6).

Numerical test with $\lambda = 1/2$,

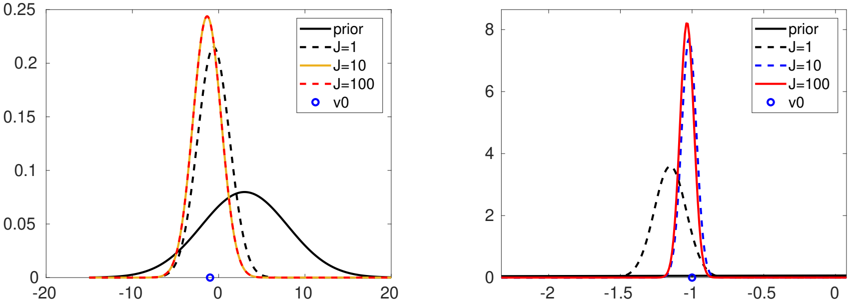


Figure: Numerical tests with $m_0 = 3$, $\sigma_0 = 5$ from $v_0 = -1$ and [**left** $\lambda = 1/2$ and $\gamma = 1$], [**right** $\lambda = 0.9$ and $\gamma = 0.1$].

See LSZ 2.8 for more illustrations of smoothing pdfs for $V_0|Y_{1:J}$.

Next time

We will talk about the filtering pdf $\pi(v_j|y_{1:j})$ and Kalman filtering – i.e., filtering in the Gaussian-linear setting.