Lecture II : Section 3 ..

In the first lecture we studied the Bayesian approach to the filtering problem, which gave expressions for

$$P(u_{j+1} | y_{1:j+1}) \propto P(y_{j+1} | u_{j+1}) P(u_{j+1} | y_{1:j})$$

If the model is linear, the problem has an exact solution, the Kalman filter.

In the nonlinear case, one typically must evaluate the integral

$$P(u_{j+1} | y_{1:j}) = \int P(u_{j+1} | u_j) P(u_j | y_{1:j}) \, du_j$$

When the state space is large, this integral becomes messy to compute.

eg. In NWP models, $\dim(X) \approx O(10^9)$, discretized spatial model.

To compute the integral w/ mesh size $1/N$ would require N^{100} grid points!

Instead of finding the exact density of $u_j | y_{1:j}$, practioners in geosciences seek approximate solutions to the filtering problem.

Approximate nonlinear filters :

The Kalman filter can be obtained by a minimization procedure :

$$m_{j+1} = \underset{v}{\operatorname{argmin}}\, I(v)$$

where $I(v) = \frac{1}{2}\left|y_{j+1} - Hv\right|_{\Gamma}^{2} + \frac{1}{2}\left|v - \hat{m}_{j+1}\right|_{\hat{C}_{j+1}}^{2}$.

This idea can be generalized to nonlinear systems by taking

$$\hat{m}_{j+1} = \Psi(m_j).$$

and choosing $\hat{C}_{j+1}$ in some way.

Note that this minimization procedure is explicitly solvable, same calculation as Kalman

$$m_{j+1} = (I - K_{j+1}H)\hat{m}_{j+1} + K_{j+1}y_{j+1}$$

$$K_{j+1} = \hat{C}_{j+1}H^T(H\hat{C}_{j+1}H^T + \Gamma)^{-1}.$$

Scanned by CamScanner

Two popular methods of choosing $\hat{C}_{j+1}$:

3DVar : Simply fix $\hat{C}_{j+1} = \hat{C}$ for all j, so we have

$$m_{j+1} = (I - KH)\Psi(m_j) + K y_{j+1}$$

$$K = \hat{C} H^T (H \hat{C} H^T + \Gamma)^{-1}$$

{3DVar - 3 dim variational}.

Developed by ~~the~~ UK Met office, ECMWF, US NOAA in 80s, 90s.

$\hat{C}$ is picked using ensemble forecasting over large time windows.

~~We have an approx~~ $\cancel{z_{j+1} = m_j + \xi_j, \quad \xi_j \sim N(0, C)}$

$\cancel{\hat{C}^{-1} = \hat{C}^{-1} + H^T \Gamma^{-1} H}$

Extended KF : Propagate the covariance under linearized dynamics.

$$\hat{C}_{j+1} = \mathbb{E}\left[(u_{j+1} - \hat{m}_{j+1})(u_{j+1} - \hat{m}_{j+1})^T \,\middle|\, y_{1:j} \right]$$

$$= \mathbb{E}\left[(\Psi(u_j) - \Psi(m_j) + \eta_j)(\Psi(u_j) - \Psi(m_j) + \eta_j)^T \,\middle|\, y_{1:j} \right]$$

$$\approx \mathbb{E}\left[(D\Psi(m_j)(u_j - m_j) + \eta_j)(\dots \,''\, \dots) \,\middle|\, y_{1:j} \right].$$

$$= D\Psi(m_j)\,\mathbb{E}\left((u_j - m_j)(u_j - m_j)^T \mid y_{1:j}\right) D\Psi(m_j)^T + \Sigma.$$

$$= D\Psi(m_j)\, C_j\, D\Psi(m_j)^T + \Sigma.$$

~~$C_{j+1}^{-1} = \hat{C}_{j+1}^{-1} + H^T \Gamma^{-1} H.$~~

~~Similarly, $z_j = m_j + C_j^{1/2} \xi_j$, $\xi_j \sim N(0, C_j)$,~~

ExKF works quite well at tracking the true state in low dimensional models

eg. ~~$~~ space trajectories, GPS, robotics

But it is too expensive to compute $D\Psi$ for high dimensional models, so cruder approx. like 3DVar are preferred.

How should we evaluate approximate filters?

~~Sur~~ Accuracy:

Suppose the observations are produced by a given trajectory

$$u_{j+1}^{\dagger} = \Psi(u_j^{\dagger}) + \xi_j$$

$$y_{j+1} = H u_{j+1}^{\dagger} + \xi_{j+1}$$

Let m_j be the approximate ~~nonlinear~~ filter. ~~Can~~ Can we show that

$$|m_j - u_j^{\dagger}| \text{ is small as } j \to \infty\ ??.$$

<u>Stability</u> : Does the filter depend on initialization?

if $m_j, \tilde{m}_j$ are approx. filters that see the same obs but different $m_0, \tilde{m}_0$

Do we have $|m_j - \tilde{m}_j| \to \cancel{0}$ small as $j \to \infty$?

<u>Statistical stability / Ergodicity</u> :

Do the statistics of the filter depend on initialization?

~~eg. If u is an random approx. of the random variable $u_j | y_{1:j}$~~

~~w/ law $\mu_j^{u_0}$. Do we have $d(\mu_j^{u_0}, \mu_j^{\tilde{u}_0}) \to 0$ as $j \to \infty$??~~

Let $z_j = m_j + \xi_j$ and let μ_j^x be the law of z_j w/ $z_0 = x$; w/ fixed set of observations.

Do we expect $d(\mu_j^x, \mu_j^y) \to 0$ as $j \to \infty$??

6

$$\boxed{u_{j+1} = \Psi(u_j)}$$

Accuracy for 3DVAR:

$$\boxed{\text{Suppose } y_{j+1} = Hu^{\dagger}_{j+1} + \varepsilon_{j+1} \quad \text{and} \quad \sup_{j\in\mathbb{N}} |\varepsilon_j| = \varepsilon.}$$

~~determ model~~

Thm: ~~Ass~~ Suppose $\hat{C}$ is chosen in such a way that $(I-KH)\Psi : \mathbb{R}^n \to \mathbb{R}^n$ is globally Lipschitz w/ constant $a<1$ in some norm $\|\cdot\|$.

$\Downarrow$

Then there is a constant $c>0$ s.t.

$$\limsup_{j\to\infty} \|m_j - v^{\dagger}_j\| < \frac{c}{1-a}\varepsilon.$$

Proof: We have

$$\begin{aligned} m_{j+1} &= (I-KH)\Psi(m_j) + Ky_{j+1} \\ &= (I-KH)\Psi(m_j) + KHu^{\dagger}_{j+1} + K\varepsilon_{j+1}. \\ &= (I-KH)\Psi(m_j) + KH\Psi(u^{\dagger}_j) + K\varepsilon_{j+1}. \end{aligned}$$

$$m_{j+1} - u^{\dagger}_{j+1} = (I-KH)(\Psi(m_j) - \Psi(u^{\dagger}_j)) + K\varepsilon_{j+1}.$$

$$\therefore \|m_{j+1} - u^{\dagger}_{j+1}\| \le a\|m_j - u^{\dagger}_j\| + \|K\varepsilon_{j+1}\|.$$

$$\le a\|m_j - u^{\dagger}_j\| + c\varepsilon.$$

Discrete Gronwall: $\{v_j\}_{j\in\mathbb{Z}}$ pos. $\lambda>0$, $k\in\mathbb{R}$

If $v_{j+1} \le \lambda v_j + k$ then $v_j \le \lambda^j v_0 + k\dfrac{1-\lambda^j}{1-\lambda}$ $(\lambda\neq 1)$ and $v_j \le v_0 + jk$ $(\lambda=1)$.

(7)

Why is the assumption $(I-KH)\Psi(\cdot)$ Lipschitz reasonable?

eg. Suppose $H = (I, 0)^T$ (for some partition of the state space).

Let $\Gamma = \gamma^2 I$, $\hat{C} = \sigma^2 I$. then

$$I - KH = I - \hat{C} H^T (H \hat{C} H^T + \Gamma)^{-1} H.$$

$$\cancel{= I - \sigma^2 (I, 0)\left(\sigma^2 \begin{pmatrix} I \\ 0 \end{pmatrix}(I\ 0) + \gamma^2 I\right)^{-1} \begin{pmatrix} I, 0 \\ 0 \end{pmatrix}}$$

$$\cancel{= \begin{pmatrix} I & 0 \\ 0 & I \end{pmatrix} - \sigma^2 (I, 0)\left(\sigma^2\right(}$$

$$\cancel{= \begin{pmatrix} I & 0 \\ 0 & I \end{pmatrix} - \sigma^2 \left(I\ (I, 0)\right)\left(\left(I\right.\right.}$$

$$= I - \sigma^2 \begin{pmatrix} I \\ 0 \end{pmatrix}\left(\sigma^2 (I\ 0)\begin{pmatrix} I & 0 \\ 0 & I \end{pmatrix}\begin{pmatrix} I \\ 0 \end{pmatrix} + \gamma^2 I\right)^{-1}(I, 0)$$

$$= I - \sigma^2 \begin{pmatrix} I \\ 0 \end{pmatrix}\left(\sigma^2 I + \gamma^2 I\right)^{-1}(I\ 0). \qquad \eta^2 = \gamma^2/\sigma^2.$$

$$= I - \frac{\sigma^2}{\sigma^2+\gamma^2}\begin{pmatrix} I & 0 \\ 0 & 0 \end{pmatrix} = \begin{pmatrix} \frac{\gamma^2}{\sigma^2+\gamma^2} I & 0 \\ 0 & I \end{pmatrix} = \begin{pmatrix} \frac{\eta^2}{1+\eta^2} I & 0 \\ 0 & I \end{pmatrix}$$

$$\cancel{\text{Now suppose } \Psi(u) = Mu}, \quad M = \begin{pmatrix} 2I & 0 \\ 0 & aI \end{pmatrix} \quad |a| < 1.$$

$$\cancel{\text{then } (I-KH)M} = \begin{pmatrix} \frac{2\eta^2}{1+\eta^2} I & 0 \\ 0 & aI \end{pmatrix}.$$

✓

8

When the unobserved directions are stable $|a|<1$, we can make ~~the~~ $(I-KH)L$ a contraction by picking γ^2 suff. small.

When the unobserved modes are unstable $|a|\geq 1$, we can't do anything.

This can be extended using the notion of observability from control theory.

Ergodicity via coupling arguments:

How can we prove statistical stability for approx filtering algorithms?

~~There is a nice 'coupli~~

For Markov chains, there is a nice coupling technique ~~the~~ due to Doeblin.

~~Let~~ $\{X_n\}$ be a Markov chain w/ transition prob. ~~invariant mea~~

$P(x, A)$ and invariant measure π.

Let X'_n, X''_n be two copies of the MC w/ $X'_0 = x$ and $X''_0 \sim \pi$; but coupled in

such a way that once they meet, they stay equal

8

Ergodicity via coupling arguments:

How can we prove statistical stability for approx. filtering distributions?

Coupling technique due to Doeblin.

Let $\{X_k\}$ be an inhomogeneous MC w/ transition prob.

$$\mathbb{P}(X_{k+1} \in A \mid X_k = x) = P_k(x, A).$$

and law $P^k(x, A)$ (when started w/ $X_0 = x$).

$$\left(P^k(x, A) = \int P_{k-1}(y, A) P^{k-1}(x, dy). \right)$$

$$\boxed{\begin{aligned} &d_{TV}(\mu_1, \mu_2) \\ &= \sup_{|\varphi| \le 1} \left| \mathbb{E} \int \varphi \, d\mu_1 - \int \varphi \, d\mu_2 \right|. \end{aligned}}$$

Thm: Suppose $\exists$ a measure ν and const $\varepsilon_k \in (0,1)$. ~~$\varepsilon_k > 0$~~ s.t.

$$P_k(x, A) \geq \varepsilon_k \nu(A) \quad \forall\, x, A; k.$$

Then

$$d_{TV}(P^k(x, \cdot), P^k(y, \cdot)) \leq \prod_{i=0}^{k-1} (1 - \varepsilon_i).$$

e.g. If $\varepsilon_k = \varepsilon$, this gives ~~geometric~~ convergence ....

(9)

$\boxed{\text{eg.}}$ Consider the "randomized 3d Var" algorithm:

~~$Z_{k+1} = Z_k +$~~

$$Z_{k+1} = (1-KH)\left(\Psi(Z_k) + \eta_k\right) + KH\left(y_{k+1} + \tilde{\xi}_{k+1}\right), \quad \tilde{\xi}_{k+1} \sim N(0,\Gamma)$$

iid. and ind. of noise in obs.

Assume Ψ bdd.

One can show that in the linear case ($\Psi(u) = Mu$)

Z_{k+1} is Gaussian whose mean and covariance satisfy the Kalman update formulas.

We have $P_k(z,A) = \mathbb{P}\left[Z_{k+1} \in A \mid Z_k = z\right]$.

$$= Z_C^{-1} \int_A \exp\left(-\frac{1}{2}\left|z' - m_{k+1}\right|_C^2\right) dz', \text{ where } C^{-1} = \hat{C}^{-1} + H^T\Gamma^{-1}H > 0$$

(provided $\hat{C} > 0$).

$$= Z_C^{-1} \int_A \exp\left(-\frac{1}{2}\left|z' - (1-KH)\Psi(z) - KH\Psi(u_k^+) - KH\xi_{k+1}\right|_C^2\right)$$

$$\geq Z_C^{-1} \exp\left(-2\left|(1-KH)\Psi(z) + KH\Psi(u_k^+)\right|_C^2\right) \exp\left(-2\left|KH\xi_{k+1}\right|^2\right.$$

$$\int_A \exp\left(-\frac{1}{4}\left|z'\right|_C^2\right) dz'.$$

(10)

Since Ψ is bdd we have

$$R = \sup_{(u,v)} \left| (1-KH)\Psi(u) + KH\Psi(v) \right|_C < \infty.$$

$$\therefore P_k(z, A) \geq Z_C^{-1} e^{-2R^2} e^{-2|KH\xi_{k+1}|_C^2} Z_{C/2}\, \nu(A)$$

$$\boxed{\text{where } \nu(A) = Z_{C/2} \int_A \exp\left(-\tfrac{1}{2}|z'|_{C/2}^2\, dz'\right).}$$

$$= (Z_C^{-1} Z_{C/2})\, e^{-2R^2} e^{-2|KH\xi_{k+1}|_C^2}\, \nu(A).$$

$$|KH\xi_{k+1}|_C^2 \leq c|\xi_{k+1}|^2 \; ; \; (Z_C^{-1} Z_{C/2}) = 2^{-n/2}.$$

$$\Rightarrow P_k(z, A) \geq \varepsilon_k\, \nu(A)$$

where $\varepsilon_k = 2^{-n/2} e^{-c|\xi_{k+1}|^2} \vartheta$

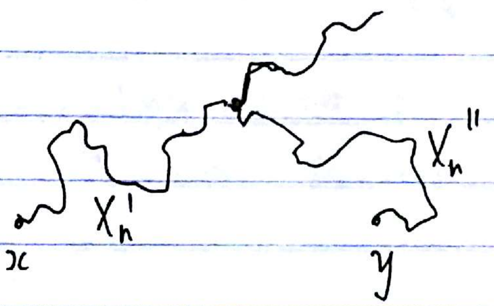
It follows that

$$d_{TV}(P^k(x,\cdot), P^k(y,\cdot)) \underset{k\to\infty}{\lesssim} r^k$$

where $r = \lim_{k\to\infty} \left(\prod_{i=1}^{k} (1-\varepsilon_i) \right)^{1/k}$ (One can show $r \in (0,1)$ by LoLN)

Proof of Doeblin thm:

Let X_k', X_k'' be two copies of the MC w/ $X_0' = x$, $X_0'' = y$, and suppose they are coupled in such a way that once they meet, they stay equal forever.



~~dez~~ Let A_k be the event that $X_n' \neq X_n''$ ~~have not met for an~~ $\forall\, n \leq k$.

Then we have

$$d_{TV}(P^k(x,\cdot), P^k(y,\cdot)) = \frac{1}{2} \sup_{|\varphi| \leq 1} \left| \mathbb{E}\varphi(X_k') - \mathbb{E}\varphi(X_k'') \right|.$$

$$= \frac{1}{2} \sup_{|\varphi| \leq 1} \left| \mathbb{E}\left(\varphi(X_k') - \varphi(X_k'')\right) \mathbb{1}_{A_k} + \underbrace{\mathbb{E}\left(\varphi(X_k') - \varphi(X_k'')\right) \mathbb{1}_{A_k^c}}_{= 0} \right|$$

$$\leq \frac{1}{2} 2\, \mathbb{E}\, \mathbb{1}_{A_k} = \mathbb{P}(A_k)$$

We now use the minorization condition to construct a coupling for which $\mathbb{P}(A_k) = \prod (1-\varepsilon_i)$.

Let $$\tilde{P}_k(x,A) = \frac{P_k(x,A) - \varepsilon\nu(A)}{1-\varepsilon_k}$$

The minorization condition guarantees that this is a Markov kernel. Let $\tilde{X}_k$ be corresponding MC.

We can write $\tilde{X}_k$ as a "random dynamical system"

$$\tilde{X}_{k+1} = \tilde{f}(\tilde{X}_k, \omega).$$

Let β_k be an $(\overset{0}{\varepsilon_k}, \overset{1}{1-\varepsilon_k})$ Bernoulli r.v., let $\overset{V_k}{\cancel{\xi}} \sim \nu(\cdot)$

Then not hard to see that

$$X_{k+1} = \beta_k \tilde{f}(\tilde{X}_k, \omega) + (1-\beta_k)\overset{V_k}{\cancel{\xi_k}}. \qquad (*)$$

is ~~equal~~ ~~to~~ a copy of the MC w/ kernel $P(x,A)$.

~~Let~~ Pick X_k', X_k'' to be given by $(*)$ but coupled by using the same β_k and V_k.

Thus, as soon as $\beta_k = 0$, we have $X_k' = X_k''$.

And $\mathbb{P}(A_k) = (1-\varepsilon_0)(1-\varepsilon_1)\cdots(1-\varepsilon_k)$. $\square$