The optimal PF

We want PFs with proposals of the form:

$$q(x_{0:k} \mid y_{1:k}, x_{0:k-1}) \propto q(x_0) \prod_{s=1}^{k} q(x_k \mid x_{0:k-1}, y_{1:k}) \quad (*).$$

We know that we need

$N_{eff} = N/\rho$ to be "large", i.e. we want

$$\rho = \frac{E[\omega^2]}{E[\omega]^2} = \frac{Var(\omega)}{E[\omega]^2} + 1 \quad \text{to be close to } 1.$$

since $E_q[\omega] = \int \frac{p}{q} q \, dx = \int p \, dx = 1$

↳ $\rho = 1 + Var(\omega) \Rightarrow Var(\omega)$ should be as small as possible

↳ The optimal PF minimizes $Var(\omega)$ over proposal distributions of type (*).

Consider the first step of a PF:

$$p(x_{0:1} \mid y_1) \propto p(x_0)\, p(x_1 \mid x_0)\, p(y_1 \mid x_1)$$

$$q(x_{0:1} \mid y_1) \propto q(x_0)\, q(x_1 \mid x_0, y_1)$$

For simplicity, assume $q(x_0) = p(x_0)$.

Then: $$\omega_1 \propto \frac{p(x_{0:1} \mid y_1)}{q(x_{0:1} \mid y_1)} \propto \frac{p(x_0)}{p(x_0)} \frac{p(x_1 \mid x_0)\, p(y_1 \mid x_1)}{q(x_1 \mid x_0, y_1)}$$

$$Var(\omega_1) = E_q[\omega_1^2] - \underbrace{E_q[\omega_1]^2}$$

$$= \left[\iint \frac{p(x_{0:1} \mid y_1)}{q(x_{0:1} \mid y_1)} q(x_{0:1} \mid y_1)\, dx_0\, dx_1\right]^2 = 1^2 = 1$$

$$E[\omega^2] = \iint \left(\frac{p(x_1|x_0)\, p(y_1|x_1)\, p(x_0)}{q(x_1|x_0,y_1)\, p(x_0)} \right)^2 q(x_1|x_0,y_1)\, p(x_0)\, dx_0\, dx_1$$

$$= \iint \frac{\left(p(x_1|x_0)\, p(y_1|x_1)\, p(x_0)\right)^2}{q(x_1|x_0,y_1)\, p(x_0)}\, dx_0\, dx_1$$

<u>Note:</u> $p(a,b|c) = p(a|b,c)\, p(b|c)$

use this:

$$p(x_0,x_1|y_1) = p(x_1|y_1,x_0)\, p(x_0|y_1)$$

from before:

$$p(x_0,x_1|y_1) = p(x_0)\, p(x_1|x_0)\, p(y_1|x_1)$$

$$\Rightarrow E[\omega^2] = \iint \frac{\left(p(x_1|y_1\, x_0)\, p(x_0|y_1)\right)^2}{q(x_1|y_1,x_0)\, p(x_0)}\, dx_0\, dx_1$$

$$= \int \frac{p(x_0|y_1)}{p(x_0)} \underbrace{\left(\int \frac{p(x_1|y_1,x_0)^2}{q(x_1|y_1,x_0)}\, dx_1 \right)}_{\circledast}\, dx_0$$

Define: $q^* = p(x_1|y_1,x_0)$

Then, for $q = q^*$, the integral $\circledast$ becomes

$$\int \frac{p(x_1|y_1,x_0)^2}{p(x_1|y_1,x_0)}\, dx_1 = \int p(x_1|y_1\, x_0)\, dx_1 = 1$$

In general: $\displaystyle f(x_0, x_1, y_1) = \int \frac{p(x_1|x_0, y_1)^2}{q(x_1|x_0, y_1)} dx_1$

Define: $\Delta = p - q$.

Then:

$$f(x_0, x_1, y_1) = \int \frac{(\Delta + q)^2}{q} dx_1$$

$$= \int \frac{\Delta^2 + 2\Delta q + q^2}{q} dx_1$$

$$= \underbrace{\int \frac{\Delta^2}{q} dx_1}_{\geq 0} + 2 \underbrace{\int \Delta \, dx_1} + \underbrace{\int q \, dx_1}_{=1}$$

Δ^2 and q are positive fcns

$$\int \Delta \, dx_1 = \int p(x_1|y_1 x_0) - q(x_1|y_1 x_0) \, dx_1$$

$$= \int p(x_1|y_1 x_0) \, dx_1 - \int q(x_1|y_1 x_0) \, dx_1$$

$$= 1 - 1 = 0$$

$$= 1 + \int \frac{\Delta^2}{q} dx_1 \geq 1.$$

Lower bound achieved when $q = q^* = p$

(then $\Delta = 0$).

Thus: $$E[\omega^2] = \int \frac{p(x_0|y_1)}{p(x_0)} \underbrace{\left(\int \frac{p(x_1|x_0,y_1)^2}{q(x_1|x_0,y_1)} dx_1 \right)}_{\geq 1} dx_0$$

$$\geq \int \frac{p(x_0|y_1)}{p(x_0)} dx_0$$

Lower bound achieved for $q(x_1|x_0,y_1) = q^* = p(x_1|y_1,x_0)$

Back to $var(\omega)$:

$$var(\omega) = E[\omega^2] - E[\omega]^2$$

$$\geq \int \frac{p(x_0|y_1)}{p(x_0)} dx_0 - 1$$

Lower bound achieved for $q = q^*$.

$\Rightarrow$ For 1st step, the optimal proposal is

$$q^* = p(x_1|y_1, x_0).$$

Use induction to show the general case!

$$w_k = \frac{p(x_{0:k} \mid y_{1:k})}{p(x_0) \prod_{j=1}^{k} q(x_j \mid x_{j-1}, y_j)}$$

$$\mathrm{Var}(w_k) = E[w_k^2] - E[w_k]^2 = E[w_k^2] - 1$$

$$E[w_k^2] = \int \cdots \int \frac{p(x_{0:k-1} \mid y_{1:k-1})^2 \, p(x_{k-1} \mid y_k)^2 \, p(x_k \mid y_k, x_{k-1})^2}{q(x_{0:k-1} \mid y_{1:k-1}) \cdot q(x_k \mid y_k, x_{k-1})} dx_{0:k}$$

$$= \int \cdots \int \frac{p(x_{0:k-1} \mid y_{1:k-1})^2 \, p(x_{k-1} \mid y_k)^2}{q(x_{0:k-1} \mid y_{1:k-1})} \underbrace{\left(\int \frac{p(x_k \mid y_k, x_{k-1})^2}{q(x_k \mid y_k, x_{k-1})} dx_k \right)}_{\geq 1 \text{ for any } q \neq q^* = p(x_k \mid y_k, x_{k-1})} dx_0 \cdots dx_{k-1}$$

$$\geq \int \cdots \int \frac{p(x_{0:k-1} \mid y_{1:k-1})^2 \, p(x_{k-1} \mid y_k)^2}{q(x_{0:k-1} \mid y_{1:k-1})} dx_0 \cdots dx_{k-1}$$

$\Rightarrow$ $\mathrm{Var}(w_k)$ is as small as possible if

$q = q^* = p(x_k \mid y_k, x_{k-1})$ at each step!

$\rightarrow$ the optimal proposal is:

$$\boxed{q^*(x_{0:k} \mid y_{1:k}) \propto p(x_0) \prod_{j=1}^{k} p(x_k \mid y_k, x_{k-1})}$$

For a special case, optimal proposal is easy to compute.

Set up: model : $x_k = f(x_{k-1}) + v_k$, $v_k \sim N(0, Q)$ iid

obs : $y_k = H(x_k) + \eta_k$, $\eta_k \sim N(0, R)$ iid.

Note: ~~obs~~ Linear obs available at each model step!

$$p(x_k | x_{k-1})\, p(y_k | x_k) \propto \exp\left(-\tfrac{1}{2}\left(x_k - f(x_{k-1})\right)^T Q^{-1}\left(x_k - f(x_{k-1})\right)\right)$$
$$\times \exp\left(-\tfrac{1}{2}\left(y_k - H(x_k)\right)^T R^{-1}\left(y_k - H(x_k)\right)\right)$$
$$\propto \exp(-F(x_k)).$$

For fixed x_{k-1}, $F(x_k)$ is quadratic!

$$\nabla F = Q^{-1}(x_k - f(x_{k-1})) + H^T R^{-1}(H x_k - y_k)$$

$$= \left(Q^{-1} + H^T R^{-1} H\right) x_k - Q^{-1} f(x_{k-1}) - H^T R^{-1} y_k = 0$$

$$\hookrightarrow x_k^* = \underbrace{\left(Q^{-1} + H^T R^{-1} H\right)^{-1}}_{=(I-KH)Q} \left(Q^{-1} f(x_{k-1}) + H^T R^{-1} y_k\right)$$

$= (I - KH)Q$, $K = QH^T(HQH^T + R)^{-1}$
see KF derivation.

$$x_k^* = (I - KH) f(x_{k-1}) + (I - KH) Q H^T R^{-1} y_k$$

$$= (I - KH) f(x_{k-1}) + \underbrace{\left(QH^TR^{-1} - QH^T(R + HQH^T)^{-1} HQH^TR^{-1} \right)}_{} y_k$$

$$QH^T\left(I - (R + HQH^T)^{-1} HQH^T\right) R^{-1} y_k$$

$$= \underbrace{QH^T(R + HQH^T)^{-1}}_{K} \underbrace{\left((R + \cancel{HQH^T}) - \cancel{HQH^T} \right) R^{-1}}_{= I} y_k$$

$$= K y_k$$

$$= (I - KH) f(x_{k-1}) + K y_k$$

$$\boxed{x_k^* = f(x_{k-1}) + K\left(y_k - H f(x_{k-1})\right)}$$

familiar from KF!

$$\nabla^2 F = \left(Q^{-1} + H^TR^{-1}H\right) := P^{-1}$$

$$\hookrightarrow F(x_k) = \underset{\substack{\uparrow \\ F(x_k^*).}}{\text{const}} + \frac{1}{2}(x_k - x_k^*)(Q^{-1} + H^TR^{-1}H)^{-1}(x_k - x_k^*)$$

$$F(x_k^*) = \frac{1}{2}\left(\cancel{f(x_{k-1})} + K(y_k - Hf(x_{k-1})) \cancel{- f(x_{k-1})}\right)^T Q^{-1}(\cdots)$$

$$+ \frac{1}{2}\left(y_k - H(f(x_{k-1}) + K(y_k - Hf(x_{k-1})))\right)^T R^{-1}(\cdots)$$

$$= \frac{1}{2}(y_k - Hf(x_{k-1}))^T K^T Q^{-1} K (y_k - Hf(x_{k-1}))$$

$$+ \frac{1}{2}(y_k - H((I-KH)f(x_{k-1}) + Ky_k)^T R^{-1}(\cdots\cdot)$$

$$\Downarrow$$

$$= \frac{1}{2}(y_k - Hf(x_{k-1}))^T (R + K^T Q K)^{-1}(y_k - Hf(x_{k-1}))$$

$$\Rightarrow p(x_k|x_{k-1})\,p(y_k|x_k) \propto \exp\left(-\tfrac{1}{2}(x_k - x_k^*)P^{-1}(x_k - x_k^*)\right)$$

$$\times \exp\left(-\tfrac{1}{2}(y_k - Hf(x_{k-1}))^T(R + K^T Q K)^{-1}(y_k - Hf(x_{k-1}))\right)$$

optimal proposal:

$$q^* = p(x_k|x_{k-1}, y_k) \propto \exp\left(-\tfrac{1}{2}(x_k - x_k^*)P^{-1}(x_k - x_k^*)\right)$$

weight:

$$w_k \propto w_{k-1}\, p(x_{k-1}|y_k)$$

$$\propto w_{k-1} \exp\left(-\tfrac{1}{2}(y_k - Hf(x_{k-1}))^T(R + K^T Q K)^{-1}(y_k - Hf(x_{k-1}))\right)$$

<u>Doable!</u>

How to implement it? $K = QH^T(HQH^T + R)^{-1}$

1) Compute $x^* = f(x_{k-1}) + K(y_k - Hf(x_{k-1}))$ for each sample.

2) Compute weight: $w \propto \exp\left(-\tfrac{1}{2}(y_k - Hf(x_k))^T R^{-1}(\ldots)\right)$

How "bad" is this for our simple example?

$$x_k = x_{k-1} + v_k . \quad v_k \sim N(0, I), \text{ n-dimensional}$$

$$y_k = Hx_k + \eta_k , \quad \eta_k \sim N(0, I), \text{ l-dimensional.}$$

<u>weights:</u> $K = QH^T(HQH^T + I)^{-1}$

$$= I_n H^T (I_l + I_l)^{-1} = \tfrac{1}{2} \begin{pmatrix} I_l \\ 0 \end{pmatrix} \}\ \text{first } l \text{ components.}$$

$$K^T Q K + R = \tfrac{1}{4}[I \;\; 0] \begin{pmatrix} I_l & 0 \\ 0 & I_n \end{pmatrix} \begin{pmatrix} I_l \\ 0 \end{pmatrix} + I_l =$$

$$= \tfrac{1}{4} [I_l \;\; 0] \begin{pmatrix} I_l \\ 0 \end{pmatrix} + I_l = \tfrac{5}{4} I_l .$$

Define : $s = y - Hf(x_{k-1}) \approx N(0, I_l)$

log weights: $-\log w = \frac{1}{2}(y - Hf(x_{k-1}))^T (K^T Q K + R)^{-1} (y - Hf(x_{k-1}))$

$$= \frac{1}{2} \cdot \frac{4}{5} s^T s$$

$$= \tfrac{2}{5} s^T s$$

$$\text{Var}(-\log w) = \text{var}\left(\tfrac{2}{5} s^T s\right) = \frac{4}{25} \sum_{i=1}^{l} \text{Var}(s_i^2)$$

$$= \frac{4}{25} \cdot 2l = \frac{8}{25} l .$$

better than PF ($\frac{1}{2} l$)

but still, this collapses in high dimensions & with a large number of observations.

Summary:

(1) There is a well defined optimal proposal.

(2) optimal proposal "easy" to implement for a class of problems (Gaussian noise, linear obs at every time step)

(3) optimal proposal out of reach when problem is not of this class.

(4) Even for simple problems, optimal proposal may not be good enough.