Estimação e Filtragem Estocástica
Curso de Pós-Graduação em Engenharia Elétrica
Departamento de Engenharia Elétrica
Universidade de Brasília

FILTRAGEM ESTOCÁSTICA NÃO-LINEAR

Geovany A. Borges
gaborges@ene.unb.br

O problema da filtragem não-linear

■ Sistema dinâmico estocástico não-linear em tempo discreto (ruído aditivo):

$$\mathbf{x}_k = \mathbf{f}(\mathbf{x}_{k-1}, \mathbf{u}_{k-1}) + \mathbf{\Gamma}_k \mathbf{w}_k \quad (1)$$
$$\mathbf{y}_k = \mathbf{h}(\mathbf{x}_k) + \mathbf{v}_k \quad (2)$$

com $\mathbf{x}_0$, $\mathbf{w}_k$ e $\mathbf{v}_k$ descorrelacionados.

■ Problema: resolver

$$p(\mathbf{x}_k|\mathbf{y}_1, \cdots, \mathbf{y}_k) = \frac{p(\mathbf{y}_k|\mathbf{x}_k, \mathbf{y}_1, \cdots, \mathbf{y}_{k-1}) p(\mathbf{x}_k|\mathbf{y}_1, \cdots, \mathbf{y}_{k-1})}{p(\mathbf{y}_k|\mathbf{y}_1, \cdots, \mathbf{y}_{k-1})} \quad (3)$$

ou ainda, determinar

$$\hat{\mathbf{x}}_k = E\{\mathbf{x}_k|\mathbf{y}_1, \cdots, \mathbf{y}_k\} \Longrightarrow \text{Estimativa de mínima variância}$$
$$\hat{\mathbf{x}}_k = \arg_{\mathbf{x}_k} \max p(\mathbf{x}_k|\mathbf{y}_1, \cdots, \mathbf{y}_k) \Longrightarrow \text{Estimativa de máximo } \textit{a posteriori}$$

O problema da filtragem não-linear

■ Dificuldades em encontrar uma solução analítica para

$$p(\mathbf{x}_k|\mathbf{y}_1, \cdots, \mathbf{y}_k) = \frac{p(\mathbf{y}_k|\mathbf{x}_k, \mathbf{y}_1, \cdots, \mathbf{y}_{k-1}) p(\mathbf{x}_k|\mathbf{y}_1, \cdots, \mathbf{y}_{k-1})}{p(\mathbf{y}_k|\mathbf{y}_1, \cdots, \mathbf{y}_{k-1})} \quad (4)$$

- Mesmo com $\mathbf{x}_0$ Gaussiano, são fatores que fazem com que $\mathbf{x}_k$ seja não-Gaussiano:
  - Não-linearidade de $\mathbf{f}(\mathbf{x}_{k-1}, \mathbf{u}_{k-1})$, que determina $p(\mathbf{x}_k|\mathbf{y}_1, \cdots, \mathbf{y}_k)$;
  - Não-linearidade de $\mathbf{h}(\mathbf{x}_k)$, que determina $p(\mathbf{y}_k|\mathbf{x}_k, \mathbf{y}_1, \cdots, \mathbf{y}_{k-1})$;
  - Erros de modelamento do processo, que podem implicar que o $\mathbf{w}_k$ real seja não-Gaussiano;
  - Erros de modelamento do medição, que podem implicar que o $\mathbf{v}_k$ real seja não-Gaussiano.

O problema da filtragem não-linear

Exemplo 1: *Modelo estocástico não-linear*

$$\mathbf{x}_{k+1} = \begin{pmatrix} x_{1,k+1} \\ x_{2,k+1} \end{pmatrix} = \begin{pmatrix} x_{1,k} + 0,5 \\ \frac{1}{2}x_{2,k} + \frac{25x_{2,k}}{1+x_{2,k}^2} + 0,8\cos(1,2(t_k - 1)) \end{pmatrix} + \mathbf{w}_{k+1}$$
$$y_k = \frac{x_{2,k}^2}{20} + v_k$$

com $\mathbf{x}_0 \sim N(\mathbf{0}, \mathbf{P}_0)$, $\mathbf{w}_k \sim N(\mathbf{0}, \mathbf{Q})$, $v_k \sim N(\mathbf{0}, r)$ tais que

$$\mathbf{P}_0 = \begin{pmatrix} 0,05 & 0 \\ 0 & 2 \end{pmatrix}, \qquad \mathbf{Q} = \begin{pmatrix} 0,001 & 0 \\ 0 & 2 \end{pmatrix}, \qquad r = 1.$$

Filtros para modelos não-lineares

- Filtro de Kalman Estendido: linearização local de primeira ordem do modelo.
- Filtro de Segunda Ordem: linearização local de segunda ordem do modelo.
- Filtro de Kalman Estendido Iterativo: linearização local com maximização iterativa de $p(\mathbf{x}_k|\mathbf{y}_1, \cdots, \mathbf{y}_k)$.
- Filtro de Kalman por Pontos Sigma: propagação de média e matriz de covariâncias do modelo de processo usando a *Unscented Transform*.
- Filtro Soma de Gaussianas: aproximação de $p(\mathbf{x}_k|\mathbf{y}_1, \cdots, \mathbf{y}_k)$ por uma soma de gaussianas.
- Filtro de Monte Carlo Seqüencial: aproximação de $p(\mathbf{x}_k|\mathbf{y}_1, \cdots, \mathbf{y}_k)$ por amostras.
- e muitos outros ...

Filtro de Kalman Estendido (FKE)

Considerando o modelo de processo

$$\mathbf{x}_k = \mathbf{f}(\mathbf{x}_{k-1}, \mathbf{u}_{k-1}) + \mathbf{\Gamma}_k \mathbf{w}_k$$

em torno de $\hat{\mathbf{x}}_{k-1}$, usando expansão em série de Taylor tem-se

$$\mathbf{f}(\mathbf{x}_{k-1}, \mathbf{u}_{k-1}) \approx \mathbf{f}(\hat{\mathbf{x}}_{k-1}, \mathbf{u}_{k-1}) + \frac{\partial \mathbf{f}(\hat{\mathbf{x}}_{k-1}, \mathbf{u}_{k-1})}{\partial \hat{\mathbf{x}}_{k-1}} (\mathbf{x}_{k-1} - \hat{\mathbf{x}}_{k-1})$$

Assim, o modelo pode ser re-escrito na forma

$$\begin{aligned}
\mathbf{x}_k &= \mathbf{f}(\hat{\mathbf{x}}_{k-1}, \mathbf{u}_{k-1}) + \frac{\partial \mathbf{f}(\hat{\mathbf{x}}_{k-1}, \mathbf{u}_{k-1})}{\partial \hat{\mathbf{x}}_{k-1}} (\mathbf{x}_{k-1} - \hat{\mathbf{x}}_{k-1}) + \mathbf{\Gamma}_k \mathbf{w}_k \\
&= \frac{\partial \mathbf{f}(\hat{\mathbf{x}}_{k-1}, \mathbf{u}_{k-1})}{\partial \hat{\mathbf{x}}_{k-1}} \mathbf{x}_{k-1} + \left[\mathbf{f}(\hat{\mathbf{x}}_{k-1}, \mathbf{u}_{k-1}) - \frac{\partial \mathbf{f}(\hat{\mathbf{x}}_{k-1}, \mathbf{u}_{k-1})}{\partial \hat{\mathbf{x}}_{k-1}} \hat{\mathbf{x}}_{k-1} \right] + \mathbf{\Gamma}_k \mathbf{w}_k \\
&= \mathbf{A}_{k-1} \mathbf{x}_{k-1} + \mathbf{u}'_{k-1} + \mathbf{\Gamma}_k \mathbf{w}_k
\end{aligned}$$

com

$$\mathbf{A}_{k-1} = \frac{\partial \mathbf{f}(\hat{\mathbf{x}}_{k-1}, \mathbf{u}_{k-1})}{\partial \hat{\mathbf{x}}_{k-1}} \qquad \mathbf{u}'_{k-1} = \mathbf{f}(\hat{\mathbf{x}}_{k-1}, \mathbf{u}_{k-1}) - \mathbf{A}_{k-1} \hat{\mathbf{x}}_{k-1}$$

O mesmo pode ser feito com relação ao modelo de medição

$$\mathbf{y}_k = \mathbf{h}(\mathbf{x}_k) + \mathbf{v}_k$$

que em torno da predição $\hat{\mathbf{x}}_{k|k-1}$

$$\mathbf{h}(\mathbf{x}_k) \approx \mathbf{h}(\hat{\mathbf{x}}_{k|k-1}) + \frac{\partial \mathbf{h}(\hat{\mathbf{x}}_{k|k-1})}{\partial \hat{\mathbf{x}}_{k|k-1}} (\mathbf{x}_k - \hat{\mathbf{x}}_{k|k-1})$$

que se torna

$$\begin{aligned}
\mathbf{y}_k &= \mathbf{h}(\hat{\mathbf{x}}_{k|k-1}) + \frac{\partial \mathbf{h}(\hat{\mathbf{x}}_{k|k-1})}{\partial \hat{\mathbf{x}}_{k|k-1}} (\mathbf{x}_k - \hat{\mathbf{x}}_{k|k-1}) + \mathbf{v}_k \\
\mathbf{y}_k - \mathbf{h}(\hat{\mathbf{x}}_{k|k-1}) + \frac{\partial \mathbf{h}(\hat{\mathbf{x}}_{k|k-1})}{\partial \hat{\mathbf{x}}_{k|k-1}} \hat{\mathbf{x}}_{k|k-1} &= \frac{\partial \mathbf{h}(\hat{\mathbf{x}}_{k|k-1})}{\partial \hat{\mathbf{x}}_{k|k-1}} \mathbf{x}_k + \mathbf{v}_k \\
\mathbf{y}'_k &= \mathbf{C}_k \mathbf{x}_k + \mathbf{v}_k
\end{aligned}$$

com

$$\mathbf{C}_k = \frac{\partial \mathbf{h}(\hat{\mathbf{x}}_{k|k-1})}{\partial \hat{\mathbf{x}}_{k|k-1}} \qquad \mathbf{y}'_k = \mathbf{y}_k - \mathbf{h}(\hat{\mathbf{x}}_{k|k-1}) + \mathbf{C}_k \hat{\mathbf{x}}_{k|k-1}$$

Desta forma, temos um sistema linear transformado

$$\begin{aligned}
\mathbf{x}_k &= \mathbf{A}_{k-1} \mathbf{x}_{k-1} + \mathbf{u}'_{k-1} + \mathbf{\Gamma}_k \mathbf{w}_k \quad (5) \\
\mathbf{y}'_k &= \mathbf{C}_k \mathbf{x}_k + \mathbf{v}_k \quad (6)
\end{aligned}$$

sobre o qual pode-se aplicar o Filtro de Kalman, resultando nas seguintes equações:

(i) Fase de predição (entre os instantes t_{k-1} e t_k):

$$\begin{aligned}
\hat{\mathbf{x}}_{k|k-1} &= E\{\mathbf{f}(\mathbf{x}_{k-1}, \mathbf{u}_{k-1}) + \mathbf{\Gamma}_k \mathbf{w}_k | \mathbf{y}_1, \cdots, \mathbf{y}_{k-1}\} && (7)\\
&= \mathbf{f}(\hat{\mathbf{x}}_{k-1}, \mathbf{u}_{k-1}) && (8)\\
\mathbf{P}_{k|k-1} &= \mathbf{A}_{k-1}\mathbf{P}_{k-1}\mathbf{A}_{k-1}^T + \mathbf{\Gamma}_k \mathbf{Q}_k \mathbf{\Gamma}_k^T && (9)
\end{aligned}$$

com

$$\mathbf{A}_{k-1} = \frac{\partial \mathbf{f}(\hat{\mathbf{x}}_{k-1}, \mathbf{u}_{k-1})}{\partial \hat{\mathbf{x}}_{k-1}}$$

(ii) Fase de correção (no instante t_k): dado a medição $\mathbf{y}_k$ com matriz de covariâncias $\mathbf{R}_k$

$$\begin{aligned}
\hat{\mathbf{x}}_k &= \hat{\mathbf{x}}_{k|k-1} + \mathbf{G}_k(\mathbf{y}'_k - \mathbf{C}_k\hat{\mathbf{x}}_{k|k-1}) && (10)\\
&= \hat{\mathbf{x}}_{k|k-1} + \mathbf{G}_k(\mathbf{y}_k - \mathbf{h}(\hat{\mathbf{x}}_{k|k-1}) + \mathbf{C}_k\hat{\mathbf{x}}_{k|k-1} - \mathbf{C}_k\hat{\mathbf{x}}_{k|k-1}) && (11)\\
&= \hat{\mathbf{x}}_{k|k-1} + \mathbf{G}_k(\mathbf{y}_k - \mathbf{h}(\hat{\mathbf{x}}_{k|k-1})) && (12)
\end{aligned}$$

$$\begin{aligned}
\mathbf{P}_k &= (\mathbf{I} - \mathbf{G}_k\mathbf{C}_k)\,\mathbf{P}_{k|k-1} && (13)\\
&= (\mathbf{I} - \mathbf{G}_k\mathbf{C}_k)\,\mathbf{P}_{k|k-1}\,(\mathbf{I} - \mathbf{G}_k\mathbf{C}_k)^T + \mathbf{G}_k\mathbf{R}_k\mathbf{G}_k^T && (14)\\
\mathbf{G}_k &= \mathbf{P}_{k|k-1}\mathbf{C}_k^T\left(\mathbf{C}_k\mathbf{P}_{k|k-1}\mathbf{C}_k^T + \mathbf{R}_k\right)^{-1} && (15)
\end{aligned}$$

com

$$\mathbf{C}_k = \frac{\partial \mathbf{h}(\hat{\mathbf{x}}_{k|k-1})}{\partial \hat{\mathbf{x}}_{k|k-1}}$$

Filtro de Kalman Estendido (FKE)

Algoritmo 1: *Filtro de Kalman Estendido*

Seja o sistema dinâmico não-linear estocástico em tempo discreto

$$\begin{aligned}
\mathbf{x}_k &= \mathbf{f}(\mathbf{x}_{k-1}, \mathbf{u}_{k-1}) + \mathbf{\Gamma}_k \mathbf{w}_k && (16)\\
\mathbf{y}_k &= \mathbf{h}(\mathbf{x}_k) + \mathbf{v}_k && (17)
\end{aligned}$$

com $\mathbf{w}_k \sim N(\mathbf{0}, \mathbf{Q}_k)$, $\mathbf{v}_k \sim N(\mathbf{0}, \mathbf{R}_k)$, $\mathbf{x}_0 \sim N(\hat{\mathbf{x}}_0, \mathbf{P}_0)$ descorrelacionados, ou seja

$$E\{\mathbf{w}_k\mathbf{v}_k^T\} = \mathbf{0}, \qquad E\{\mathbf{w}_k\mathbf{x}_0^T\} = \mathbf{0}, \qquad E\{\mathbf{v}_k\mathbf{x}_0^T\} = \mathbf{0}. \qquad (18)$$

O Filtro de Kalman Estendido é o estimador de $\mathbf{x}_k$, composto de duas fases:

(i) Fase de predição (entre os instantes t_{k-1} e t_k):

$$\begin{aligned}
\hat{\mathbf{x}}_{k|k-1} &= \mathbf{f}(\hat{\mathbf{x}}_{k-1}, \mathbf{u}_{k-1}) && (19)\\
\mathbf{P}_{k|k-1} &= \frac{\partial \mathbf{f}(\hat{\mathbf{x}}_{k-1}, \mathbf{u}_{k-1})}{\partial \hat{\mathbf{x}}_{k-1}}\mathbf{P}_{k-1}\left(\frac{\partial \mathbf{f}(\hat{\mathbf{x}}_{k-1}, \mathbf{u}_{k-1})}{\partial \hat{\mathbf{x}}_{k-1}}\right)^T + \mathbf{\Gamma}_k\mathbf{Q}_k\mathbf{\Gamma}_k^T && (20)
\end{aligned}$$

(ii) Fase de correção (no instante t_k): dado a medição $\mathbf{y}_k$ com matriz de covariâncias $\mathbf{R}_k$

$$\begin{aligned}
\hat{\mathbf{x}}_k &= \hat{\mathbf{x}}_{k|k-1} + \mathbf{G}_k(\mathbf{y}_k - \mathbf{h}(\hat{\mathbf{x}}_{k|k-1})) && (21)\\
\mathbf{P}_k &= \left(\mathbf{I} - \mathbf{G}_k\frac{\partial \mathbf{h}(\hat{\mathbf{x}}_{k|k-1})}{\partial \hat{\mathbf{x}}_{k|k-1}}\right)\mathbf{P}_{k|k-1} && (22)\\
\mathbf{G}_k &= \mathbf{P}_{k|k-1}\left(\frac{\partial \mathbf{h}(\hat{\mathbf{x}}_{k|k-1})}{\partial \hat{\mathbf{x}}_{k|k-1}}\right)^T\left(\frac{\partial \mathbf{h}(\hat{\mathbf{x}}_{k|k-1})}{\partial \hat{\mathbf{x}}_{k|k-1}}\mathbf{P}_{k|k-1}\left(\frac{\partial \mathbf{h}(\hat{\mathbf{x}}_{k|k-1})}{\partial \hat{\mathbf{x}}_{k|k-1}}\right)^T + \mathbf{R}_k\right)^{-1} && (23)
\end{aligned}$$

Filtro de Kalman Estendido (FKE)

Observação 1:

O Filtro de Kalman Estendido obtido por linearização aproxima $p(\mathbf{x}_k|\mathbf{y}_1, \cdots, \mathbf{y}_k)$ por uma Gaussiana, pois usa o modelo (5)-(6).

Observação 2:

O Filtro de Kalman Estendido é um estimador sub-ótimo, devido às aproximações.

Observação 3:

Com $\mathbf{w}_k \sim N(\mathbf{0}, \bar{\mathbf{Q}}_k)$, $\mathbf{v}_k \sim N(\mathbf{0}, \bar{\mathbf{R}}_k)$, $\mathbf{x}_0 \sim N(\hat{\mathbf{x}}_0, \bar{\mathbf{P}}_0)$ descorrelacionados, mesmo se forem usados $\mathbf{P}_0 \geq \bar{\mathbf{P}}_0$, $\mathbf{Q}_k \geq \bar{\mathbf{Q}}_k$, $\mathbf{R}_k \geq \bar{\mathbf{R}}_k$ não se tem como garantir consistência de $\hat{\mathbf{x}}_k$ e $\hat{\mathbf{x}}_{k|k-1}$.

Filtro de Kalman Estendido (FKE)

Exemplo 3: *Estimação sob restrição*

Considere o sistema

$$\begin{aligned} \mathbf{x}_k &= \mathbf{f}(\mathbf{x}_{k-1}, \mathbf{u}_{k-1}) + \mathbf{\Gamma}_k \mathbf{w}_k \quad (26) \\ \mathbf{y}_k &= \mathbf{h}(\mathbf{x}_k) + \mathbf{v}_k \quad (27) \end{aligned}$$

com $\mathbf{w}_k \sim N(\mathbf{0}, \mathbf{Q}_k)$, $\mathbf{v}_k \sim N(\mathbf{0}, \mathbf{R}_k)$, $\mathbf{x}_0 \sim N(\hat{\mathbf{x}}_0, \mathbf{P}_0)$ descorrelacionados, mas com restrição nas variáveis de estado sob a foma de

$$\mathbf{c} = \mathbf{g}(\mathbf{x}_k). \quad (28)$$

Uma forma de incorporar as restrições no problema consiste em usar um artifício conhecido por *pseudo-medição.* Ou seja, o modelo de medição passaria a ter uma componente fictícia:

$$\begin{pmatrix} \mathbf{y}_k \\ \mathbf{c} \end{pmatrix} = \begin{pmatrix} \mathbf{h}(\mathbf{x}_k) \\ \mathbf{g}(\mathbf{x}_k) \end{pmatrix} + \mathbf{v}'_k$$

Filtro de Kalman Estendido (FKE)

Exemplo 2: *Estimação conjunta de estados e parâmetros*

Considere o sistema

$$\begin{aligned} \mathbf{x}_k &= \mathbf{f}(\mathbf{x}_{k-1}, \mathbf{u}_{k-1}, \boldsymbol{\theta}_{k-1}) + \mathbf{\Gamma}_k \mathbf{w}_k \quad (24) \\ \mathbf{y}_k &= \mathbf{h}(\mathbf{x}_k) + \mathbf{v}_k \quad (25) \end{aligned}$$

em que $\boldsymbol{\theta}$ é um vetor de parâmetros que compõem o modelo de evolução do processo. O problema consiste em estimar simultâneamente $\mathbf{x}_k$ e $\boldsymbol{\theta}_k$. Uma solução consiste em aplicar o Filtro de Kalman Estendido (ou qualquer filtro estocástico não-linear) para o sistema estendido

$$\begin{aligned} \mathbf{z}_k &= \begin{pmatrix} \mathbf{x}_k \\ \boldsymbol{\theta}_k \end{pmatrix} = \begin{pmatrix} \mathbf{f}(\mathbf{x}_{k-1}, \mathbf{u}_{k-1}, \boldsymbol{\theta}_{k-1}) \\ \boldsymbol{\theta}_{k-1} \end{pmatrix} + \begin{pmatrix} \mathbf{\Gamma}_k \mathbf{w}_k \\ \mathbf{w}_{\boldsymbol{\theta},k} \end{pmatrix} \\ \mathbf{y}_k &= \mathbf{h}(\mathbf{z}_k) + \mathbf{v}_k \end{aligned}$$

em que $\hat{\mathbf{z}}_k$ contem estimativas de ambos $\mathbf{x}_k$ e $\boldsymbol{\theta}_k$ que se tornam correlacionadas devido a $\mathbf{f}(\mathbf{x}_{k-1}, \mathbf{u}_{k-1}, \boldsymbol{\theta}_{k-1})$.

com $\mathbf{v}'_k \sim N(\mathbf{0}, \mathbf{R}'_k)$ e

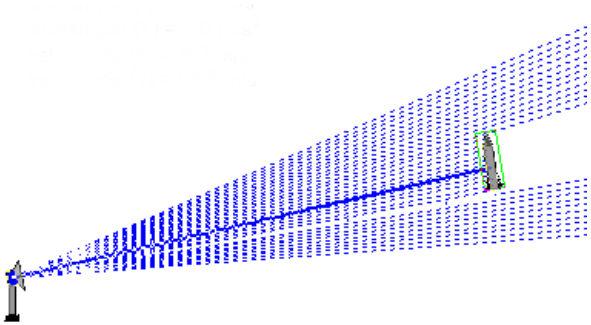
$$\mathbf{R}'_k = \begin{pmatrix} \mathbf{R}_k & \mathbf{0} \\ \mathbf{0} & \mathbf{S}_k \end{pmatrix}$$

Com $\mathbf{S}_k = \mathbf{0}$, a restrição (28) é satisfeita a cada passo de correção do filtro. Se for desejado relaxar a restrição (28), basta escolher $\mathbf{S}_k > \mathbf{0}$. Se $\mathbf{S}_k$ ainda for "muito pequeno" a estimativa não fica longe de satisfazer (28).

Filtro de Kalman Estendido (FKE)

Exemplo 4: *Rastreamento de Foguete*

Discussão em sala de aula. Observar o conceito de medição modificada.



Filtro de Kalman Estendido Iterativo (FKEI)

Leitura em sala de aula.

Filtro de Kalman por Pontos Sigma (FKPS)

Leitura em sala de aula.

Filtros sequenciais de Monte Carlo

Seja o sistema dinâmico não-linear estocástico em tempo discreto

$$\mathbf{x}_k = \mathbf{f}(\mathbf{x}_{k-1}, \mathbf{u}_{k-1}) + \mathbf{\Gamma}_k \mathbf{w}_k \quad (29)$$

$$\mathbf{y}_k = \mathbf{h}(\mathbf{x}_k) + \mathbf{v}_k \quad (30)$$

Os filtros de Monte Carlo pertencem a uma classe de algoritmos que aproximam $p(\mathbf{x}_k|\mathbf{y}_1, \cdots, \mathbf{y}_k)$ por amostras

$$\mathbf{x}_k^{(1)} \quad \mathbf{x}_k^{(2)} \quad \cdots \quad \mathbf{x}_k^{(N_s)}$$

com pesos associados

$$q_k^{(1)} \quad q_k^{(2)} \quad \cdots \quad q_k^{(N_s)}$$

de forma que $q_k^{(i)} \propto p(\mathbf{x}_k^{(i)}|\mathbf{y}_1, \cdots, \mathbf{y}_k)$ e a seguinte aproximação é feita

$$p(\mathbf{x}_k|\mathbf{y}_1, \cdots, \mathbf{y}_k) \approx \sum_{i=1}^{N_s} q_k^{(i)} \delta(\mathbf{x}_k - \mathbf{x}_k^{(i)}) \tag{31}$$

Os pesos devem ser determinados de forma que

$$\begin{aligned} \int_{-\infty}^{\infty} p(\mathbf{x}_k|\mathbf{y}_1, \cdots, \mathbf{y}_k)\, d\mathbf{x}_k &= \int_{-\infty}^{\infty} \left[\sum_{i=1}^{N_s} q_k^{(i)} \delta(\mathbf{x}_k - \mathbf{x}_k^{(i)}) \right] d\mathbf{x}_k \qquad (32) \\ &= \sum_{i=1}^{N_s} q_k^{(i)} = 1 \qquad (33) \end{aligned}$$

Neste contexto, dois filtros MCS serão apresentados:

- Filtro *Sequential Importance Sampling* (SIS)
- Filtro *Sampling Importance Re-sampling* (SIR)

Boas referências sobre este tópico são [1] [2].

A partir das amostras, a estimativa de mínima variância é dada por

$$\begin{aligned} E\{\mathbf{x}_k|\mathbf{y}_1, \cdots, \mathbf{y}_k\} &= \int_{-\infty}^{\infty} \mathbf{x}_k\, p(\mathbf{x}_k|\mathbf{y}_1, \cdots, \mathbf{y}_k)\, d\mathbf{x}_k \\ &\approx \int_{-\infty}^{\infty} \mathbf{x}_k \left[\sum_{i=1}^{N_s} q_k^{(i)} \delta(\mathbf{x}_k - \mathbf{x}_k^{(i)}) \right] d\mathbf{x}_k \\ &\approx \sum_{i=1}^{N_s} q_k^{(i)} \mathbf{x}_k^{(i)} \end{aligned}$$

Por outro lado, a estimativa *máximo a posteriori* seria

$$\begin{aligned} \hat{x}_{MAP} &= \arg_{x_k} \max\, p(\mathbf{x}_k|\mathbf{y}_1, \cdots, \mathbf{y}_k) \\ &\approx \mathbf{x}_k^{(j)} \text{ tal que } q_k^{(j)} > q_k^{(i)},\, j \neq i. \end{aligned}$$

Pela literatura, as amostras $\mathbf{x}_k^{(i)}$ são comumente denominadas de *partículas*. Devido a isto, filtros MCS são também conhecidos por *filtros de partículas*.

Filtro *Sequential Importance Sampling* (SIS)

A densidade $p(\mathbf{x}_k|\mathbf{y}_1, \cdots, \mathbf{y}_k)$ pode ser obtida a partir de

$$p(\mathbf{x}_k|\mathbf{y}_1, \cdots, \mathbf{y}_k) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} p(\mathbf{x}_0, \cdots, \mathbf{x}_{k-1}, \mathbf{x}_k|\mathbf{y}_1, \cdots, \mathbf{y}_k)\, d\mathbf{x}_0 \cdots d\mathbf{x}_{k-1} \tag{34}$$

Desta forma, o problema de filtragem pode ser resolvido a partir da densidade conjunta $p(\mathbf{x}_0, \cdots, \mathbf{x}_{k-1}, \mathbf{x}_k|\mathbf{y}_1, \cdots, \mathbf{y}_k)$, que pode ser aproximada por

$$p(\mathbf{x}_0, \cdots, \mathbf{x}_{k-1}, \mathbf{x}_k|\mathbf{y}_1, \cdots, \mathbf{y}_k) \approx \sum_{i=1}^{N_s} q_k^{(i)} \delta(\{\mathbf{x}_0, \cdots, \mathbf{x}_k\} - \{\mathbf{x}_0^{(i)}, \cdots, \mathbf{x}_k^{(i)}\}) \tag{35}$$

com $\sum_{i=1}^{N_s} q_k^{(i)} = 1$. No entanto, $p(\mathbf{x}_0, \cdots, \mathbf{x}_{k-1}, \mathbf{x}_k|\mathbf{y}_1, \cdots, \mathbf{y}_k)$ é difícil de ser amostrada diretamente. Um dos motivos é que, devido ao grande número de amostras acumuladas desde $k = 0$, o algoritmo de amostragem seria bastante pesado computacionalmente.

O algoritmo de amostragem de importância pode então ser usado nestes casos:

- Gera-se $\{\mathbf{x}_0^{(i)}, \cdots, \mathbf{x}_k^{(i)}\}$ seguindo uma FDP auxiliar $\pi(\mathbf{x}_0, \cdots, \mathbf{x}_k|\mathbf{y}_1, \cdots, \mathbf{y}_k\}$ (função de importância)

- Usa-se $q_k^{(i)}$ dado por

$$q_k^{(i)} = \frac{p(\mathbf{x}_0^{(i)}, \cdots, \mathbf{x}_k^{(i)}|\mathbf{y}_1, \cdots, \mathbf{y}_k)}{\pi(\mathbf{x}_0^{(i)}, \cdots, \mathbf{x}_k^{(i)}|\mathbf{y}_1, \cdots, \mathbf{y}_k)} \tag{36}$$

de forma que fica satisfeita a relação

$$q_k^{(i)} \propto p(\mathbf{x}_0^{(i)}, \cdots, \mathbf{x}_k^{(i)}|\mathbf{y}_1, \cdots, \mathbf{y}_k). \tag{37}$$

O problema com esta solução é que para calcular (36) requer-se que sejam armazenadas todas as $(k+1)N_s$ amostras $\{\mathbf{x}_0^{(i)}, \cdots, \mathbf{x}_k^{(i)}\}$, número este que aumenta com o tempo. Portanto, faz-se necessário uma solução seqüencial.

Considere que exista uma função de importância que possa ser fatorada como

$$\begin{aligned}\pi(\mathbf{x}_0, \cdots, \mathbf{x}_k|\mathbf{y}_1, \cdots, \mathbf{y}_k) &= \pi(\mathbf{x}_k|\mathbf{x}_0, \cdots, \mathbf{x}_{k-1}, \mathbf{y}_1, \cdots, \mathbf{y}_k) \\ &\quad \times\pi(\mathbf{x}_0, \cdots, \mathbf{x}_{k-1}|\mathbf{y}_1, \cdots, \mathbf{y}_{k-1}).\end{aligned} \tag{38}$$

Na verdade, isto seria possível se

$$\pi(\mathbf{x}_0, \cdots, \mathbf{x}_{k-1}|\mathbf{y}_1, \cdots, \mathbf{y}_k) = \pi(\mathbf{x}_0, \cdots, \mathbf{x}_{k-1}|\mathbf{y}_1, \cdots, \mathbf{y}_{k-1}). \tag{39}$$

Verifique isto a partir da relação $\pi(A, B|C) = \pi(A|B, C)\pi(B|C)$.

Mais ainda, $\pi(\cdot)$ pode ser escolhida de forma que

$$\pi(\mathbf{x}_k|\mathbf{x}_0, \cdots, \mathbf{x}_{k-1}, \mathbf{y}_1, \cdots, \mathbf{y}_k) = \pi(\mathbf{x}_k|\mathbf{x}_{k-1}, \mathbf{y}_1, \cdots, \mathbf{y}_k) \tag{40}$$

e assim

$$\pi(\mathbf{x}_0, \cdots, \mathbf{x}_k|\mathbf{y}_1, \cdots, \mathbf{y}_k) = \pi(\mathbf{x}_k|\mathbf{x}_{k-1}, \mathbf{y}_1, \cdots, \mathbf{y}_k)\pi(\mathbf{x}_0, \cdots, \mathbf{x}_{k-1}|\mathbf{y}_1, \cdots, \mathbf{y}_{k-1}) \tag{41}$$

Então, a partir de (41), o conjunto de amostras $\mathbf{x}_0^{(i)}, \cdots, \mathbf{x}_k^{(i)}$ pode ser obtido a partir de $\mathbf{x}_0^{(i)}, \cdots, \mathbf{x}_{k-1}^{(i)}$ acrescentando-se amostras $\mathbf{x}_k^{(i)}$ geradas de

$$\begin{aligned}\mathbf{x}_k^{(i)} &\sim \pi(\mathbf{x}_k|\mathbf{x}_0^{(i)}, \cdots, \mathbf{x}_{k-1}^{(i)}, \mathbf{y}_1, \cdots, \mathbf{y}_k) \qquad i = 1, \cdots, N_s. && (42)\\ &= \pi(\mathbf{x}_k|\mathbf{x}_{k-1}^{(i)}, \mathbf{y}_1, \cdots, \mathbf{y}_k) && (43)\end{aligned}$$

Percebe-se que $\mathbf{x}_k^{(i)}$ é uma predição de $\mathbf{x}_k$ dada toda seqüência de eventos ocorridos desde $k = 0$. Portanto, determinou-se uma forma de gerar amostras seqüencialmente (ou recursivamente). O mesmo precisa ser feito com relação aos pesos $q_k^{(i)}$.

A partir da regra de Bayes

$$p(\mathbf{x}_0, \cdots, \mathbf{x}_k|\mathbf{y}_1, \cdots, \mathbf{y}_k) = \frac{p(\mathbf{y}_k|\mathbf{x}_0, \cdots, \mathbf{x}_k, \mathbf{y}_1, \cdots, \mathbf{y}_{k-1})p(\mathbf{x}_0, \cdots, \mathbf{x}_k|\mathbf{y}_1, \cdots, \mathbf{y}_{k-1})}{p(\mathbf{y}_k|\mathbf{y}_1, \cdots, \mathbf{y}_{k-1})},$$

e considerando

$$\begin{aligned}p(\mathbf{x}_0, \cdots, \mathbf{x}_k|\mathbf{y}_1, \cdots, \mathbf{y}_{k-1}) &= p(\mathbf{x}_k|\mathbf{x}_0, \cdots, \mathbf{x}_{k-1}, \mathbf{y}_1, \cdots, \mathbf{y}_{k-1}) \\ &\quad \times p(\mathbf{x}_0, \cdots, \mathbf{x}_{k-1}|\mathbf{y}_1, \cdots, \mathbf{y}_{k-1}),\end{aligned} \tag{44}$$

então $p(\mathbf{x}_0, \cdots, \mathbf{x}_k|\mathbf{y}_1, \cdots, \mathbf{y}_k)$ pode ser obtido recursivamente:

$$p(\mathbf{x}_0, \cdots, \mathbf{x}_k|\mathbf{y}_1, \cdots, \mathbf{y}_k) = \psi \times p(\mathbf{x}_0, \cdots, \mathbf{x}_{k-1}|\mathbf{y}_1, \cdots, \mathbf{y}_{k-1}) \tag{45}$$

$$\psi = \frac{p(\mathbf{y}_k|\mathbf{x}_0, \cdots, \mathbf{x}_k, \mathbf{y}_1, \cdots, \mathbf{y}_{k-1})p(\mathbf{x}_k|\mathbf{x}_0, \cdots, \mathbf{x}_{k-1}, \mathbf{y}_1, \cdots, \mathbf{y}_{k-1})}{p(\mathbf{y}_k|\mathbf{y}_1, \cdots, \mathbf{y}_{k-1})} \tag{46}$$

Dado o modelo

$$\begin{aligned}\mathbf{x}_k &= \mathbf{f}(\mathbf{x}_{k-1}, \mathbf{u}_{k-1}) + \mathbf{\Gamma}_k\mathbf{w}_k && (47)\\ \mathbf{y}_k &= \mathbf{h}(\mathbf{x}_k) + \mathbf{v}_k && (48)\end{aligned}$$

tem-se que

$$p(\mathbf{y}_k|\mathbf{x}_0,\cdots,\mathbf{x}_k,\mathbf{y}_1,\cdots,\mathbf{y}_{k-1}) = p(\mathbf{y}_k|\mathbf{x}_k) \quad (49)$$

$$p(\mathbf{x}_k|\mathbf{x}_0,\cdots,\mathbf{x}_{k-1},\mathbf{y}_1,\cdots,\mathbf{y}_{k-1}) = p(\mathbf{x}_k|\mathbf{x}_{k-1}) \quad (50)$$

Portanto,

$$p(\mathbf{x}_0,\cdots,\mathbf{x}_k|\mathbf{y}_1,\cdots,\mathbf{y}_k) = \left(\frac{p(\mathbf{y}_k|\mathbf{x}_k)p(\mathbf{x}_k|\mathbf{x}_{k-1})}{p(\mathbf{y}_k|\mathbf{y}_1,\cdots,\mathbf{y}_{k-1})}\right) p(\mathbf{x}_0,\cdots,\mathbf{x}_{k-1}|\mathbf{y}_1,\cdots,\mathbf{y}_{k-1})$$

é uma forma recursiva de obter $p(\mathbf{x}_0,\cdots,\mathbf{x}_k|\mathbf{y}_1,\cdots,\mathbf{y}_k)$ a partir da densidade conjunta acumulada até $k-1$. Juntando este resultado com o da equação (41), obtem-se uma forma seqüêncial para atualização dos pesos:

$$q_k^{(i)} = \frac{p(\mathbf{x}_0^{(i)},\cdots,\mathbf{x}_k^{(i)}|\mathbf{y}_1,\cdots,\mathbf{y}_k)}{\pi(\mathbf{x}_0^{(i)},\cdots,\mathbf{x}_k^{(i)}|\mathbf{y}_1,\cdots,\mathbf{y}_k)}$$

$\sum_{i=1}^{N_s} q_k^{(i)} = 1$, que pode ser obtido fazendo normalização:

$$\tilde{q}_k^{(i)} = \frac{q_k^{(i)}}{\sum_{j=1}^{N_s} q_k^{(j)}} \quad (52)$$

e depois usando $\tilde{q}_k^{(i)}$ como os novos pesos.

$$q_k^{(i)} = \frac{p(\mathbf{y}_k|\mathbf{x}_k^{(i)})p(\mathbf{x}_k^{(i)}|\mathbf{x}_{k-1}^{(i)})}{p(\mathbf{y}_k|\mathbf{y}_1,\cdots,\mathbf{y}_{k-1})\pi(\mathbf{x}_k^{(i)}|\mathbf{x}_{k-1}^{(i)},\mathbf{y}_1,\cdots,\mathbf{y}_k)} \times \frac{p(\mathbf{x}_0^{(i)},\cdots,\mathbf{x}_{k-1}^{(i)}|\mathbf{y}_1,\cdots,\mathbf{y}_{k-1})}{\pi(\mathbf{x}_0^{(i)},\cdots,\mathbf{x}_{k-1}^{(i)}|\mathbf{y}_1,\cdots,\mathbf{y}_{k-1})}$$

$$q_k^{(i)} = \frac{p(\mathbf{y}_k|\mathbf{x}_k^{(i)})p(\mathbf{x}_k^{(i)}|\mathbf{x}_{k-1}^{(i)})}{p(\mathbf{y}_k|\mathbf{y}_1,\cdots,\mathbf{y}_{k-1})\pi(\mathbf{x}_k^{(i)}|\mathbf{x}_{k-1}^{(i)},\mathbf{y}_1,\cdots,\mathbf{y}_k)} \times q_{k-1}^{(i)}$$

Como $p(\mathbf{y}_k|\mathbf{y}_1,\cdots,\mathbf{y}_{k-1})$ não depende dos estados,

$$q_k^{(i)} \propto \frac{p(\mathbf{y}_k|\mathbf{x}_k^{(i)})p(\mathbf{x}_k^{(i)}|\mathbf{x}_{k-1}^{(i)})}{\pi(\mathbf{x}_k^{(i)}|\mathbf{x}_{k-1}^{(i)},\mathbf{y}_1,\cdots,\mathbf{y}_k)} \times q_{k-1}^{(i)} \quad (51)$$

Na verdade, não é necessário conhecer $p(\mathbf{y}_k|\mathbf{y}_1,\cdots,\mathbf{y}_{k-1})$. Mas é preciso garantir

Filtro *Sequential Importance Sampling* (SIS)

Algoritmo 2: *Filtro Sequential Importante Sampling (SIS)*

Seja o sistema dinâmico não-linear estocástico em tempo discreto

$$\mathbf{x}_k = \mathbf{f}(\mathbf{x}_{k-1},\mathbf{u}_{k-1}) + \mathbf{\Gamma}_k\mathbf{w}_k \quad (53)$$

$$\mathbf{y}_k = \mathbf{h}(\mathbf{x}_k) + \mathbf{v}_k \quad (54)$$

com $\mathbf{w}_k \sim N(\mathbf{0},\mathbf{Q}_k)$, $\mathbf{v}_k \sim N(\mathbf{0},\mathbf{R}_k)$, $\mathbf{x}_0 \sim N(\hat{\mathbf{x}}_0,\mathbf{P}_0)$. O algoritmo SIS de filtragem estocástica Monte Carlo se resume nos seguintes passos:

- Inicialização: dado $p(x_0)$, fazer
  - Gerar amostras $x_0^{(i)}$, $i = 1,\cdots,N_s$, de $p(x_0)$.
  - Determinar os pesos $q_0^{(i)}$. A partir de (37), seria natural usar

$$q_0^{(i)} = p(\mathbf{x}_0^{(i)})$$

e depois normalizar seus valores de forma que $\sum_{i=1}^{N_s} q_k^{(i)} = 1$.

- A cada instante $k \geq 1$, atualizar as amostras e pesos por meio de

$$\begin{aligned} \mathbf{x}_k^{(i)} &\sim \pi(\mathbf{x}_k^{(i)}|\mathbf{x}_{k-1}^{(i)}, \mathbf{y}_1, \cdots, \mathbf{y}_k) \\ q_k^{(i)} &= \frac{p(\mathbf{y}_k|\mathbf{x}_k^{(i)})p(\mathbf{x}_k^{(i)}|\mathbf{x}_{k-1}^{(i)})}{\pi(\mathbf{x}_k^{(i)}|\mathbf{x}_{k-1}^{(i)}, \mathbf{y}_1, \cdots, \mathbf{y}_k)} \times q_{k-1}^{(i)} \end{aligned}$$

e depois normalizar $q_k^{(i)}$ de forma que $\sum_{i=1}^{N_s} q_k^{(i)} = 1$. A estimativa de mínima variância do vetor de estados é dada por

$$\hat{\mathbf{x}}_k = \sum_{i=1}^{N_s} q_k^{(i)} \mathbf{x}_k^{(i)}$$

Observação 4:

Sendo $\mathbf{w}_k \sim N(\mathbf{0}, \mathbf{Q}_k)$ e $\mathbf{v}_k \sim N(\mathbf{0}, \mathbf{R}_k)$, então $p(\mathbf{y}_k|\mathbf{x}_k^{(i)})$ e $p(\mathbf{x}_k^{(i)}|\mathbf{x}_{k-1}^{(i)})$ são densidades Gaussianas.

Observação 5:

O desempenho do SIS depende fortemente da escolha da densidade $\pi(\mathbf{x}_k^{(i)}|\mathbf{x}_{k-1}^{(i)}, \mathbf{y}_1, \cdots, \mathbf{y}_k)$. Uma má escolha pode fazer com que o algoritmo se degenere (divergência), implicando que uma dada amostra j pode ter $q_k^{(j)} \approx 1$, e as outras não seriam amostras representativas da distribuição.

Filtro *Sampling Importance Re-sampling* (SIR)

O SIR é um filtro com reduzido risco de divergência quando comparado ao SIS. Isto se deve a duas alterações no SIS:

I. Escolhe-se

$$\pi(\mathbf{x}_k^{(i)}|\mathbf{x}_{k-1}^{(i)}, \mathbf{y}_1, \cdots, \mathbf{y}_k) = p(\mathbf{x}_k^{(i)}|\mathbf{x}_{k-1}^{(i)})$$

de foma que $q_k^{(i)}$ é atualizado como

$$q_k^{(i)} = p(\mathbf{y}_k|\mathbf{x}_k^{(i)}) \times q_{k-1}^{(i)}$$

de forma que as amostras com maior verosimilhança com as medições terão seus pesos aumentados com relação ao outros (após normalização). As amostras podem ser obtidas, por exemplo, usando

$$\mathbf{x}_k^{(i)} = \mathbf{f}(\mathbf{x}_{k-1}^{(i)}, \mathbf{u}_{k-1}) + \mathbf{\Gamma}_k \mathbf{w}_k^{(i)}$$

com $\mathbf{w}_k^{(i)}$ sendo uma amostra de $p(\mathbf{w}_k)$.

II. Reamostragem: esta fase se faz necessária para minimizar o número de amostras com $q_k^{(i)} \approx 0$. A fase de reamostragem consiste em gerar um novo conjunto de amostras $\tilde{\mathbf{x}}_k^{(1)}, \cdots, \tilde{\mathbf{x}}_k^{(N_s)}$ a partir de amostragem com reposição de $\mathbf{x}_k^{(1)}, \cdots, \mathbf{x}_k^{(N_s)}$ em que

$$\Pr\{\tilde{\mathbf{x}}_k^{(i)} = \mathbf{x}_k^{(j)}\} = q_k^{(j)}$$

o que implica que serão replicadas amostras com maior verossimilhança $p(\mathbf{y}_k|\mathbf{x}_k^{(i)})$. Após esta fase, os pesos recebem $q_k^{(i)} = 1/N_s$.

Observação 6:

Existe um risco de degeneração se as medidas possuírem baixíssima incerteza, i.e., $p(\mathbf{y}_k|\mathbf{x}_k^{(i)})$ é muito estreita. Isto reduz a probabilidade de se ter amostras representativas em torno do máximo de $p(\mathbf{y}_k|\mathbf{x}_k^{(i)})$.

Filtro *Sampling Importance Re-sampling* (SIR)

Exemplo 5: *SIR aplicado a um modelo estocástico não-linear*

$$\begin{aligned}\mathbf{x}_{k+1} &= \begin{pmatrix} x_{1,k+1} \\ x_{2,k+1} \end{pmatrix} = \begin{pmatrix} x_{1,k} + 0,5 \\ \frac{1}{2}x_{2,k} + \frac{25x_{2,k}}{1+x_{2,k}^2} + 0,8\cos(1,2(t_k - 1)) \end{pmatrix} + \mathbf{w}_{k+1} \\ y_k &= \frac{x_{2,k}^2}{20} + v_k\end{aligned}$$

com $\mathbf{x}_0 \sim N(\mathbf{0}, \mathbf{P}_0)$, $\mathbf{w}_k \sim N(\mathbf{0}, \mathbf{Q})$, $v_k \sim N(\mathbf{0}, r)$ tais que

$$\mathbf{P}_0 = \begin{pmatrix} 0,05 & 0 \\ 0 & 2 \end{pmatrix}, \qquad \mathbf{Q} = \begin{pmatrix} 0,001 & 0 \\ 0 & 2 \end{pmatrix}, \qquad r = 1.$$

Referências

[1] DOUCET, A.; GODSILL, S.; ANDRIEU, C. On sequential monte carlo sampling methods for bayesian filtering. *Statistics and Computing*, v. 10, p. 197Ű208, 2000.

[2] ARULAMPALAM, M. S. et al. A tutorial on particle filters for online nonlinear/non-gaussian Bayesian tracking. *IEEE Trans. on Signal Processing*, v. 50, n. 2, p. 174–188, February 2002.