ENSAI, filière Génie Statistique, 2021–2022

Filtrage Linéaire et Non–Linéaire
2ème partie : Extensions du Filtrage de Kalman

François Le Gland
INRIA Rennes et IRMAR

people.rennes.inria.fr/Francois.Le_Gland/ensai/

Extensions aux systèmes non–linéaires

- borne de Cramér–Rao a posteriori
- filtre de Kalman étendu (linéarisation)
- filtre de Kalman *unscented* (quadrature)

Borne de Cramér–Rao a posteriori

suite d'états cachés $\{X_k\}$ à valeurs dans $\mathbb{R}^m$, vérifiant

$$X_k = b_k(X_{k-1}) + \sigma_k(X_{k-1})\, W_k$$

et suite d'observations $\{Y_k\}$ à valeurs dans $\mathbb{R}^d$, vérifiant

$$Y_k = h_k(X_k) + V_k$$

hypothèses :

- état initial X_0 pas nécessairement gaussien
- bruit d'état $\{W_k\}$ blanc gaussien, de matrice de covariance identité
- bruit d'observation $\{V_k\}$ blanc gaussien, de matrice de covariance Q_k^V
- suites $\{W_k\}$ et $\{V_k\}$ et état initial X_0 mutuellement indépendants
- fonctions b_k, σ_k et h_k dérivables

si la matrice de covariance $Q^W_k(x) = \sigma_k(x)\, \sigma^*_k(x)$ est *inversible* pour tout $x \in \mathbb{R}^m$, alors il existe une *densité de transition* définie par

$$\mathbb{P}[X_k \in dx' \mid X_{k-1} = x] = p_k(x' \mid x)\, dx'$$

et si la matrice de covariance Q^V_k est *inversible*, alors il existe une *densité d'émission* définie par

$$\mathbb{P}[Y_k \in dy \mid X_k = x] = q_k(y \mid x)\, dy$$

clairement

$$p_k(x' \mid x) = \frac{1}{\sqrt{\det(2\pi\, Q^W_k(x))}}\ \exp\{-\tfrac{1}{2}\,(x' - b_k(x))^*\, (Q^W_k(x))^{-1}\, (x' - b_k(x))\}$$

et

$$q_k(y \mid x) = \frac{1}{\sqrt{\det(2\pi\, Q^V_k)}}\ \exp\{-\tfrac{1}{2}\,(y - h_k(x))^*\, (Q^V_k)^{-1}\, (y - h_k(x))\}$$

Théorème pour tout estimateur $\psi(Y_{0:n})$ de la statistique $\phi(X_n)$ à partir des observations $Y_{0:n}$, l'erreur quadratique moyenne est minorée par

$$\mathbb{E}[\,(\psi(Y_{0:n}) - \phi(X_n))\,(\psi(Y_{0:n}) - \phi(X_n))^*\,] \geq M_n\; J_n^{-1}\; M_n^*$$

avec la matrice de sensibilité $M_n = \mathbb{E}[\phi'(X_n)]$ et la matrice d'information de Fisher J_n peut se calculer de la façon *récursive* suivante

$$J_k^- = D_k^{22} - D_k^{21}\,(J_{k-1} + D_k^{11})^{-1}\,D_k^{12} \qquad \text{et} \qquad J_k = J_k^- + E_k$$

avec

$$D_k = \begin{pmatrix} D_k^{11} & D_k^{12} \\ D_k^{21} & D_k^{22} \end{pmatrix} = -\mathbb{E}[\,\frac{\partial^2}{\partial x_{k-1:k}^2} \log p_k(X_k \mid X_{k-1})\,]$$

$$E_k = -\mathbb{E}[\,\frac{\partial^2}{\partial x_k^2} \log q_k(Y_k \mid X_k)\,]$$

et avec l'initialisation

$$J_0^- = -\mathbb{E}[\,\frac{\partial^2}{\partial x_0^2} \log p_0(X_0)\,] \qquad \text{et} \qquad J_0 = J_0^- + E_0$$

Remarque dans le cas particulier où la matrice de covariance $Q_k^W(x) \equiv Q_k^W$ ne dépend pas de $x \in \mathbb{R}^m$, on a les expressions explicites

$$D_k^{11} = \mathbb{E}[\,(b_k'(X_{k-1}))^*\,(Q_k^W)^{-1}\,b_k'(X_{k-1})\,]$$

$$D_k^{12} = -\mathbb{E}[\,(b_k'(X_{k-1}))^*\,]\,(Q_k^W)^{-1}$$

$$D_k^{22} = (Q_k^W)^{-1}$$

$$E_k = \mathbb{E}[\,(h_k'(X_k))^*\,(Q_k^V)^{-1}\,h_k'(X_k)\,]$$

la matrice D_k^{11} est semi–définie positive, la matrice D_k^{22} est définie positive (et donc la matrice–bloc D_k est elle–même semi–définie positive) et la matrice E_k est semi–définie positive

Utilisation pratique

pour évaluer *à l'avance* la performance d'un estimateur donné $\psi(Y_{0:n})$ de l'état caché X_n au vu des observations $Y_{0:n} = (Y_0, \cdots, Y_n)$, on réalise M simulations indépendantes

$$X_n^j \qquad \text{et} \qquad Y_{0:n}^j = (Y_0^j, \cdots, Y_n^j) \qquad \text{pour tout } j = 1 \cdots M$$

pour avoir une idée de la performance de cet estimateur, on évalue empiriquement la matrice de covariance de l'erreur d'estimation

$$\mathbb{E}[\,(\psi(Y_{0:n}) - X_n)\,(\psi(Y_{0:n}) - X_n)^*\,]$$

$$\approx \frac{1}{M} \sum_{j=1}^{M} (\psi(Y_{0:n}^j) - X_n^j)\,(\psi(Y_{0:n}^j) - X_n^j)^*$$

et pour avoir une idée de la marge d'amélioration possible, on compare cette estimation empirique avec une estimation empirique de la borne J_n^{-1}

Calcul numérique approché

pour calculer la matrice d'information de Fisher J_n intervenant dans la borne, on réalise M simulations indépendantes

$$X_{0:n}^j = (X_0^j, \cdots, X_n^j) \qquad \text{pour tout } j = 1 \cdots M$$

on évalue empiriquement, pour tout instant $k = 1 \cdots n$, les matrices

$$D_k^{11} = \mathbb{E}[\,(b_k'(X_{k-1}))^*\,(Q_k^W)^{-1}\,b_k'(X_{k-1})\,]$$

$$\approx \frac{1}{M}\,\sum_{j=1}^{M}(b_k'(X_{k-1}^j))^*\,(Q_k^W)^{-1}\,b_k'(X_{k-1}^j)$$

$$D_k^{12} = -\mathbb{E}[\,(b_k'(X_{k-1}))^*\,]\,(Q_k^W)^{-1}$$

$$\approx -\frac{1}{M}\,\sum_{j=1}^{M}(b_k'(X_{k-1}^j))^*\,(Q_k^W)^{-1}$$

et

$$E_k = \mathbb{E}[\,(h'_k(X_k))^* \,(Q^V_k)^{-1}\, h'_k(X_k)\,]$$

$$\approx \frac{1}{M} \sum_{j=1}^{M} (h'_k(X^j_k))^* \,(Q^V_k)^{-1}\, h'_k(X^j_k)$$

et on calcule récursivement

$$J^-_k = D^{22}_k - D^{21}_k \,(J_{k-1} + D^{11}_k)^{-1}\, D^{12}_k \qquad \text{et} \qquad J_k = J^-_k + E_k$$

en utilisant les approximations empiriques des matrices D^{11}_k, D^{12}_k et E_k

la preuve du théorème (borne de Cramér–Rao a posteriori) repose sur le lemme d'inversion suivant pour une matrice–bloc

Lemme si la matrice A est inversible, alors

$$M = \begin{pmatrix} A & B \\ C & D \end{pmatrix} = \begin{pmatrix} I & 0 \\ C\,A^{-1} & I \end{pmatrix} \begin{pmatrix} A & 0 \\ 0 & \Delta \end{pmatrix} \begin{pmatrix} I & A^{-1}\,B \\ 0 & I \end{pmatrix}$$

où la matrice $\Delta = D - C\,A^{-1}\,B$ est appelée complément de Schur de la matrice A dans la matrice–bloc M

en particulier, $\det M = \det \Delta \cdot \det A$ de sorte que la matrice M est inversible si et seulement si la matrice Δ (et la matrice A par hypothèse) est inversible, et

$$M^{-1} = \begin{pmatrix} \star & \star \\ \star & \Delta^{-1} \end{pmatrix}$$

Remarque en inversant les rôles, si la matrice D est inversible, alors

$$M = \begin{pmatrix} A & B \\ C & D \end{pmatrix} = \begin{pmatrix} I & B\,D^{-1} \\ 0 & I \end{pmatrix} \begin{pmatrix} \Delta & 0 \\ 0 & D \end{pmatrix} \begin{pmatrix} I & 0 \\ D^{-1}\,C & I \end{pmatrix}$$

où la matrice $\Delta = A - B\,D^{-1}\,C$ est appelée complément de Schur de la matrice D dans la matrice–bloc M

en particulier, la matrice M est inversible si et seulement si la matrice Δ (et la matrice D par hypothèse) est inversible

Preuve on vérifie que

$$\begin{pmatrix} I & 0 \\ C\,A^{-1} & I \end{pmatrix} \begin{pmatrix} A & 0 \\ 0 & \Delta \end{pmatrix} \begin{pmatrix} I & A^{-1}\,B \\ 0 & I \end{pmatrix}$$

$$= \begin{pmatrix} I & 0 \\ C\,A^{-1} & I \end{pmatrix} \begin{pmatrix} A & B \\ 0 & \Delta \end{pmatrix}$$

$$= \begin{pmatrix} A & B \\ C & C\,A^{-1}\,B + \Delta \end{pmatrix} = \begin{pmatrix} A & B \\ C & D \end{pmatrix}$$

on en déduit que

$$M^{-1} = \begin{pmatrix} I & -A^{-1}B \\ 0 & I \end{pmatrix} \begin{pmatrix} A^{-1} & 0 \\ 0 & \Delta^{-1} \end{pmatrix} \begin{pmatrix} I & 0 \\ -CA^{-1} & I \end{pmatrix}$$

et il suffit de remarquer que

$$\begin{pmatrix} I & \star \\ 0 & I \end{pmatrix} \begin{pmatrix} \star & 0 \\ 0 & \Delta^{-1} \end{pmatrix} \begin{pmatrix} I & 0 \\ \star & I \end{pmatrix}$$

$$= \begin{pmatrix} I & \star \\ 0 & I \end{pmatrix} \begin{pmatrix} \star & 0 \\ \star & \Delta^{-1} \end{pmatrix} = \begin{pmatrix} \star & \star \\ \star & \Delta^{-1} \end{pmatrix} \qquad \square$$

Preuve du Théorème on se propose d'utiliser le résultat vu dans le cas statique, soit

$$\mathbb{E}[\,(\psi(Y) - \phi(X))\,(\psi(Y) - \phi(X))^*\,] \geq M\,J^{-1}\,M^*$$

avec la matrice d'information de Fisher

$$J = -\mathbb{E}[\frac{\partial^2}{\partial x^2} \log p(X, Y)]$$

supposée inversible, et avec la matrice de sensibilité

$$M = \mathbb{E}[\phi'(X)]$$

et de transposer ce résultat au cas dynamique, avec

- les variables trajectorielles $X = X_{0:n}$ et $Y = Y_{0:n}$
- et une statistique $\phi(X) = \phi(X_n)$ ne dépendant que de l'état final X_n

le résultat ciblé peut se re–formuler comme

$$\mathbb{E}[\,(\psi(Y_{0:n}) - \phi(X_n))\,(\psi(Y_{0:n}) - \phi(X_n))^*\,] \geq M_{0:n}\; J_{0:n}^{-1}\; M_{0:n}^*$$

avec la matrice d'information de Fisher

$$J_{0:n} = -\mathbb{E}[\frac{\partial^2}{\partial x_{0:n}^2} \log p_{0:n}(X_{0:n}, Y_{0:n})]$$

clairement, la matrice jacobienne

$$\phi'_{0:n}(x_{0:n}) = \begin{pmatrix} 0 & \phi'(x_n) \end{pmatrix}$$

ne dépend aussi que de l'état final x_n, et la matrice de sensibilité s'exprime comme

$$M_{0:n} = \mathbb{E}[\phi'_{0:n}(X_{0:n})] = \begin{pmatrix} 0 & \mathbb{E}[\phi'(X_n)] \end{pmatrix} = \begin{pmatrix} 0 & M_n \end{pmatrix}$$

la loi jointe des états cachés $X_{0:n}$ et des observations $Y_{0:n}$ a pour densité

$$p_{0:n}(x_{0:n}, y_{0:n}) = p_0(x_0)\, q_0(y_0 \mid x_0) \prod_{k=1}^{n} p_k(x_k \mid x_{k-1})\, q_k(y_k \mid x_k)$$

d'où l'expression récurrente pour la log–densité

$$\log\, p_{0:n}(x_{0:n}, y_{0:n}) \;=\; \log\, p_{0:n-1}(x_{0:n-1}, y_{0:n-1})$$

$$+ \log\, p_n(x_n \mid x_{n-1}) + \log\, q_n(y_n \mid x_n)$$

la prochaine étape consiste à calculer chacune des trois matrices hessiennes correspondant aux trois termes figurant dans le membre de droite premièrement

$$\frac{\partial^2}{\partial x_{0:n}^2} \log p_{0:n-1}(x_{0:n-1}, y_{0:n-1})$$

$$= \begin{pmatrix} \dfrac{\partial^2}{\partial x_{0:n-1}^2} & \dfrac{\partial^2}{\partial x_{0:n-1}\,\partial x_n} \\ \star & \dfrac{\partial^2}{\partial x_n^2} \end{pmatrix} \log p_{0:n-1}(x_{0:n-1}, y_{0:n-1})$$

$$= \begin{pmatrix} \dfrac{\partial^2}{\partial x_{0:n-1}^2} \log p_{0:n-1}(x_{0:n-1}, y_{0:n-1}) & 0 \\ 0 & 0 \end{pmatrix}$$

donc

$$-\mathbb{E}[\, \frac{\partial^2}{\partial x_{0:n}^2} \log p_{0:n-1}(X_{0:n-1}, Y_{0:n-1}) \,] = \begin{pmatrix} J_{0:n-1} & 0 \\ 0 & 0 \end{pmatrix}$$

deuxièmement

$$\frac{\partial^2}{\partial x_{0:n}^2} \log p_n(x_n \mid x_{n-1})$$

$$= \begin{pmatrix} \dfrac{\partial^2}{\partial x_{0:n-2}^2} & \dfrac{\partial^2}{\partial x_{0:n-2}\,\partial x_{n-1:n}} \\ \star & \dfrac{\partial^2}{\partial x_{n-1:n}^2} \end{pmatrix} \log p_n(x_n \mid x_{n-1})$$

$$= \begin{pmatrix} 0 & 0 \\ 0 & \dfrac{\partial^2}{\partial x_{n-1:n}^2} \log p_n(x_n \mid x_{n-1}) \end{pmatrix}$$

donc

$$-\mathbb{E}[\, \frac{\partial^2}{\partial x_{0:n}^2} \log p_n(X_n \mid X_{n-1}) \,] = \begin{pmatrix} 0 & 0 \\ 0 & D_n \end{pmatrix}$$

troisièmement et dernièrement

$$
\begin{aligned}
&\frac{\partial^2}{\partial x_{0:n}^2} \log q_n(y_n \mid x_n) \\
&= \begin{pmatrix} \dfrac{\partial^2}{\partial x_{0:n-1}^2} & \dfrac{\partial^2}{\partial x_{0:n-1}\, \partial x_n} \\ \star & \dfrac{\partial^2}{\partial x_n^2} \end{pmatrix} \log q_n(y_n \mid x_n) \\
&= \begin{pmatrix} 0 & 0 \\ 0 & \dfrac{\partial^2}{\partial x_n^2} \log q_n(y_n \mid x_n) \end{pmatrix}
\end{aligned}
$$

donc

$$
-\mathbb{E}[\, \frac{\partial^2}{\partial x_{0:n}^2} \log q_n(Y_n \mid X_n) \,] = \begin{pmatrix} 0 & 0 \\ 0 & E_n \end{pmatrix}
$$

on a donc

$$
\begin{aligned}
J_{0:n} &= -\mathbb{E}[\, \frac{\partial^2}{\partial x_{0:n}^2} \log p_{0:n}(X_{0:n}, Y_{0:n}) \,] \\
&= -\mathbb{E}[\, \frac{\partial^2}{\partial x_{0:n}^2} \log\, p_{0:n-1}(X_{0:n-1}, Y_{0:n-1}) \,] \\
&\quad -\mathbb{E}[\, \frac{\partial^2}{\partial x_{0:n}^2} \log\, p_n(X_n \mid X_{n-1}) \,] - \mathbb{E}[\, \frac{\partial^2}{\partial x_{0:n}^2} \log\, q_n(Y_n \mid X_n) \,] \\
&= \begin{pmatrix} J_{0:n-1} & 0 \\ 0 & E_n \end{pmatrix} + \begin{pmatrix} 0 & 0 \\ 0 & D_n \end{pmatrix}
\end{aligned}
$$

attention : les dimensions des différents blocs figurant dans le membre de droite ne sont pas consistantes !

soit u, v, et w trois vecteurs de dimension appropriées, et les notations

$$[uvw] = \begin{pmatrix} u \\ v \\ w \end{pmatrix} \qquad [uv] = \begin{pmatrix} u \\ v \end{pmatrix} \qquad \text{et} \qquad [vw] = \begin{pmatrix} v \\ w \end{pmatrix}$$

premièrement

$$\begin{pmatrix} [uv]^* & w^* \end{pmatrix} \begin{pmatrix} J_{0:n-1} & 0 \\ 0 & E_n \end{pmatrix} \begin{pmatrix} [uv] \\ w \end{pmatrix} = [uv]^* \, J_{0:n-1} \, [uv] + w^* \, E_n \, w$$

deuxièmement et dernièrement

$$\begin{pmatrix} u^* & [vw]^* \end{pmatrix} \begin{pmatrix} 0 & 0 \\ 0 & D_n \end{pmatrix} \begin{pmatrix} u \\ [vw] \end{pmatrix} = [vw]^* \, D_n \, [vw]$$

de sorte que l'expression

$$[uvw]^* \, J_{0:n} \, [uvw] = [uv]^* \, J_{0:n-1} \, [uv] + w^* \, E_n \, w + [vw]^* \, D_n \, [vw]$$

se décompose comme la somme de trois expressions positives ou nulles clairement, si l'expression $[uvw]^* \, J_{0:n} \, [uvw]$ s'annule, alors nécessairement chacune des trois expressions

$$[uv]^* \, J_{0:n-1} \, [uv] \qquad w^* \, E_n \, w \qquad \text{et} \qquad [vw]^* \, D_n \, [vw]$$

s'annule

si la première expression $[uv]^* \, J_{0:n-1} \, [uv]$ s'annule, alors nécessairement $[uv] = 0$ car la matrice $J_{0:n-1}$ est définie positive d'après l'hypothèse de récurrence

si $[uv] = 0$, alors la troisième expression $[vw]^* \, D_n \, [vw]$ se réduit à $w^* \, D_n^{22} \, w$ et si cette expression s'annule, alors nécessairement $w = 0$ car la matrice D_n^{22} est inversible

on en déduit que la matrice $J_{0:n}$ est définie positive, donc inversible en introduisant la décomposition sous forme de matrice–bloc

$$J_{0:n} = -\mathbb{E}[\, \frac{\partial^2}{\partial x_{0:n}^2} \log p_{0:n}(X_{0:n}, Y_{0:n}) \,] = \begin{pmatrix} A_n & B_n \\ B_n^* & C_n \end{pmatrix}$$

on obtient

$$\begin{aligned} J_{0:n} &= \begin{pmatrix} J_{0:n-1} & 0 \\ 0 & E_n \end{pmatrix} + \begin{pmatrix} 0 & 0 \\ 0 & D_n \end{pmatrix} \\ &= \begin{pmatrix} A_{n-1} & B_{n-1} & 0 \\ B_{n-1}^* & C_{n-1} & 0 \\ 0 & 0 & E_n \end{pmatrix} + \begin{pmatrix} 0 & 0 & 0 \\ 0 & D_n^{11} & D_n^{12} \\ 0 & D_n^{21} & D_n^{22} \end{pmatrix} \\ &= \begin{pmatrix} A_{n-1} & B_{n-1} & 0 \\ B_{n-1}^* & C_{n-1} + D_n^{11} & D_n^{12} \\ 0 & D_n^{21} & E_n + D_n^{22} \end{pmatrix} \end{aligned}$$

par comparaison des deux décompositions de $J_{0:n}$ sous forme de matrice–bloc, on obtient les identifications

$$A_n = \begin{pmatrix} A_{n-1} & B_{n-1} \\ B^*_{n-1} & C_{n-1} + D^{11}_n \end{pmatrix} \qquad B_n = \begin{pmatrix} 0 \\ D^{12}_n \end{pmatrix} \qquad C_n = E_n + D^{22}_n$$

d'après l'hypothèse de récurrence, (i) la matrice A_{n-1} est inversible et on peut définir son complément de Schur

$$J_{n-1} = C_{n-1} - B^*_{n-1}\, A^{-1}_{n-1}\, B_{n-1}$$

dans la matrice–bloc $J_{0:n-1}$ et (ii) la matrice–bloc $J_{0:n-1}$ et la matrice A_{n-1} sont inversibles de sorte que le complément de Schur J_{n-1} est inversible, en utilisant le lemme d'inversion pour une matrice–bloc

d'après l'hypothèse de récurrence, (i) la matrice A_{n-1} est inversible et on peut définir son complément de Schur

$$\Delta_{n-1} = C_{n-1} + D^{11}_n - B^*_{n-1}\, A^{-1}_{n-1}\, B_{n-1} = J_{n-1} + D^{11}_n$$

dans la matrice–bloc A_n et (ii) la matrice symétrique J_{n-1} est inversible, donc définie positive, et a fortiori la matrice symétrique Δ_{n-1} est définie positive, donc inversible la matrice A_{n-1} et le complément de Schur Δ_{n-1} sont inversibles de sorte que la matrice–bloc A_n est inversible, en utilisant le lemme d'inversion pour une matrice–bloc, et

$$A_n^{-1} = \begin{pmatrix} A_{n-1} & B_{n-1} \\ B_{n-1}^* & C_{n-1} + D_n^{11} \end{pmatrix}^{-1} = \begin{pmatrix} \star & \star \\ \star & \Delta_{n-1}^{-1} \end{pmatrix}$$

par ailleurs

$$J_{0:n}^{-1} = \begin{pmatrix} A_n & B_n \\ B_n^* & C_n \end{pmatrix}^{-1} = \begin{pmatrix} \star & \star \\ \star & J_n^{-1} \end{pmatrix}$$

avec

$$J_n = C_n - B_n^* \, A_n^{-1} \, B_n$$

complément de Schur de la matrice A_n dans la matrice–bloc $J_{0:n}$

on a donc

$$M_{0:n}\, J_{0:n}^{-1}\, M_{0:n}^* \;=\; \begin{pmatrix} 0 & M_n \end{pmatrix} \begin{pmatrix} \star & \star \\ \star & J_n^{-1} \end{pmatrix} \begin{pmatrix} 0 \\ M_n^* \end{pmatrix}$$

$$= M_n\, J_n^{-1}\, M_n^*$$

et

$$J_n \;=\; C_n - B_n^*\, A_n^{-1}\, B_n$$

$$= E_n + D_n^{22} - \begin{pmatrix} 0 & D_n^{21} \end{pmatrix} \begin{pmatrix} \star & \star \\ \star & \Delta_{n-1}^{-1} \end{pmatrix} \begin{pmatrix} 0 \\ D_n^{12} \end{pmatrix}$$

$$= E_n + D_n^{22} - D_n^{21}\, (J_{n-1} + D_n^{11})^{-1}\, D_n^{12} \qquad \square$$

Filtre de Kalman étendu (linéarisation)

suite d'états cachés $\{X_k\}$ à valeurs dans $\mathbb{R}^m$, vérifiant

$$X_k = b_k(X_{k-1}) + \sigma_k(X_{k-1})\, W_k$$

et suite d'observations $\{Y_k\}$ à valeurs dans $\mathbb{R}^d$, vérifiant

$$Y_k = h_k(X_k) + V_k$$

hypothèses :

- état initial X_0 pas nécessairement gaussien
- bruit d'état $\{W_k\}$ blanc gaussien, de matrice de covariance identité
- bruit d'observation $\{V_k\}$ blanc gaussien, de matrice de covariance Q_k^V *inversible*
- suites $\{W_k\}$ et $\{V_k\}$ et état initial X_0 mutuellement indépendants
- fonctions b_k et h_k dérivables

idée : linéariser les fonctions b_k et σ_k autour de $\widehat{X}_{k-1}$, c'est–à–dire

$$b_k(x) \simeq b_k(\widehat{X}_{k-1}) + b'_k(\widehat{X}_{k-1})\,(x - \widehat{X}_{k-1}) \quad \text{et} \quad \sigma_k(x) \simeq \sigma_k(\widehat{X}_{k-1})$$

et linéariser la fonction h_k autour de $\widehat{X}_k^-$, c'est–à–dire

$$h_k(x) \simeq h_k(\widehat{X}_k^-) + h'_k(\widehat{X}_k^-)\,(x - \widehat{X}_k^-)$$

on introduit le système conditionnellement linéaire gaussien

$$X_k = F_k^{\mathrm{L}}\,X_{k-1} + f_k^{\mathrm{L}} + W_k^{\mathrm{L}}$$

$$Y_k = H_k^{\mathrm{L}}\,X_k + h_k^{\mathrm{L}} + V_k$$

avec

$$F_k^{\mathrm{L}} = b'_k(\widehat{X}_{k-1}) \quad \text{et} \quad f_k^{\mathrm{L}} = -b'_k(\widehat{X}_{k-1})\,\widehat{X}_{k-1} + b_k(\widehat{X}_{k-1})$$

et avec

$$H_k^{\mathrm{L}} = h'_k(\widehat{X}_k^-) \quad \text{et} \quad h_k^{\mathrm{L}} = -h'_k(\widehat{X}_k^-)\,\widehat{X}_k^- + h_k(\widehat{X}_k^-)$$

on remarque que

$$F_k^{\mathrm{L}}\,\widehat{X}_{k-1} + f_k^{\mathrm{L}} = b_k(\widehat{X}_{k-1}) \quad \text{et} \quad H_k^{\mathrm{L}}\,\widehat{X}_k^- + h_k^{\mathrm{L}} = h_k(\widehat{X}_k^-)$$

le vecteur aléatoire $W_k^{\mathrm{L}} = \sigma_k(\widehat{X}_{k-1})\, W_k$ est gaussien, centré, de matrice de covariance $Q_k^L = \sigma_k(\widehat{X}_{k-1})\, \sigma_k^*(\widehat{X}_{k-1})$
on applique alors exactement le filtre de Kalman à ce nouveau système, d'où l'algorithme sous–optimal suivant

$$\begin{aligned}
\widehat{X}_k^- &= b_k(\widehat{X}_{k-1}) \\
P_k^- &= b_k'(\widehat{X}_{k-1})\, P_{k-1}\, (b_k'(\widehat{X}_{k-1}))^* + \sigma_k(\widehat{X}_{k-1})\, \sigma_k^*(\widehat{X}_{k-1})
\end{aligned}$$

et

$$\begin{aligned}
\widehat{X}_k &= \widehat{X}_k^- + K_k\, (Y_k - h_k(\widehat{X}_k^-)) \\
P_k &= (I - K_k\, h_k'(\widehat{X}_k^-))\; P_k^-
\end{aligned}$$

avec la matrice de gain

$$K_k = P_k^-\, (h_k'(\widehat{X}_k^-))^*\, [h_k'(\widehat{X}_k^-)\, P_k^-\, (h_k'(\widehat{X}_k^-))^* + Q_k^V]^{-1}$$

on choisit l'initialisation $\widehat{X}_0^-$ et P_0^- de telle sorte que $\mathcal{N}(\widehat{X}_0^-, P_0^-)$ soit une bonne approximation de la distribution de probabilité du v.a. X_0

Approximation gaussienne

suite d'états cachés $\{X_k\}$ à valeurs dans $\mathbb{R}^m$, vérifiant

$$X_k = b_k(X_{k-1}) + \sigma_k(X_{k-1})\, W_k$$

et suite d'observations $\{Y_k\}$ à valeurs dans $\mathbb{R}^d$, vérifiant

$$Y_k = h_k(X_k) + V_k$$

hypothèses :

- état initial X_0 gaussien, de moyenne $\bar{X}_0$ et de matrice de covariance Q_0^X
- bruit d'état $\{W_k\}$ blanc gaussien, de matrice de covariance identité
- bruit d'observation $\{V_k\}$ blanc gaussien, de matrice de covariance Q_k^V *inversible*
- suites $\{W_k\}$ et $\{V_k\}$ et état initial X_0 mutuellement indépendants
- fonctions b_k et h_k pas nécessairement dérivables

idée : au lieu de s'appuyer sur une linéarisation des fonctions autour de l'estimateur courant, on se propose ici

- de remplacer les différentes distributions de probabilité conditionnelles par des distributions de probabilité gaussiennes ayant même moyenne et même matrice de covariance
- d'utiliser des formules de quadrature, développées initialement pour le calcul numérique d'intégrales, pour approcher ces moyennes et ces matrices de covariance conditionnelles

le premier point peut s'interpréter comme une projection, au sens de la distance de Kullback–Leibler, sur la famille des distributions de probabilité gaussiennes

Moments

le calcul des deux premiers moments de la distribution de probabilité conditionnelle $\mu_k^-(dx) = \mathbb{P}[X_k \in dx \mid Y_{0:k-1}]$ est facile

Lemme

$$\widehat{X}_k^- = \mathbb{E}[X_k \mid Y_{0:k-1}] = \int b_k(x)\, \mu_{k-1}(dx)$$

et

$$\begin{aligned} P_k^- &= \mathbb{E}[\,(X_k - \widehat{X}_k^-)\,(X_k - \widehat{X}_k^-)^* \mid Y_{0:k-1}] \\ &= \int (b_k(x) - \widehat{X}_k^-)\,(b_k(x) - \widehat{X}_k^-)^*\, \mu_{k-1}(dx) \\ &\qquad + \int \sigma_k(x)\, \sigma_k^*(x)\, \mu_{k-1}(dx) \end{aligned}$$

Preuve clairement

$$\mathbb{E}[X_k \mid Y_{0:k-1}] = \mathbb{E}[b_k(X_{k-1}) \mid Y_{0:k-1}] + \mathbb{E}[\sigma_k(X_{k-1})\, W_k \mid Y_{0:k-1}]$$

et on remarque que

$$\mathbb{E}[\sigma_k(X_{k-1})\, W_k \mid Y_{0:k-1}]$$

$$= \mathbb{E}[\,\mathbb{E}[\sigma_k(X_{k-1})\, W_k \mid X_{k-1}, Y_{0:k-1}] \mid Y_{0:k-1}]$$

$$= \mathbb{E}[\sigma_k(X_{k-1})\, \mathbb{E}[W_k \mid X_{k-1}, Y_{0:k-1}] \mid Y_{0:k-1}] = 0$$

où on a utilisé dans la dernière égalité l'indépendance de $(X_{k-1}, Y_0, \cdots, Y_{k-1})$ et de W_k, de sorte que $\mathbb{E}[W_k \mid X_{k-1}, Y_{0:k-1}] = 0$

il reste

$$\mathbb{E}[X_k \mid Y_{0:k-1}] = \mathbb{E}[b_k(X_{k-1}) \mid Y_{0:k-1}] = \int b_k(x)\, \mu_{k-1}(dx)$$

par différence

$$X_k - \widehat{X}_k^- = (b_k(X_{k-1}) - \widehat{X}_k^-) + \sigma_k(X_{k-1})\, W_k$$

et clairement

$$\begin{aligned}
&\mathbb{E}[\,(X_k - \widehat{X}_k^-)\,(X_k - \widehat{X}_k^-)^* \mid Y_{0:k-1}] \\
&= \mathbb{E}[\,((b_k(X_{k-1}) - \widehat{X}_k^-) + \sigma_k(X_{k-1})\, W_k) \\
&\qquad ((b_k(X_{k-1}) - \widehat{X}_k^-) + \sigma_k(X_{k-1})\, W_k)^* \mid Y_{0:k-1}] \\
&= \mathbb{E}[\,(b_k(X_{k-1}) - \widehat{X}_k^-)\,(b_k(X_{k-1}) - \widehat{X}_k^-)^* \mid Y_{0:k-1}] \\
&\qquad + \mathbb{E}[\sigma_k(X_{k-1})\, W_k\ W_k^*\, \sigma_k^*(X_{k-1}) \mid Y_{0:k-1}] \\
&\qquad + \mathbb{E}[\sigma_k(X_{k-1})\, W_k\ (b_k(X_{k-1}) - \widehat{X}_k^-)^* \mid Y_{0:k-1}] \\
&\qquad + \mathbb{E}[\,(b_k(X_{k-1}) - \widehat{X}_k^-)\ W_k^*\, \sigma_k^*(X_{k-1}) \mid Y_{0:k-1}]
\end{aligned}$$

on remarque que

$$\mathbb{E}[\sigma_k(X_{k-1})\, W_k\ W_k^*\, \sigma_k^*(X_{k-1}) \mid Y_{0:k-1}]$$

$$= \mathbb{E}[\,\mathbb{E}[\sigma_k(X_{k-1})\, W_k\ W_k^*\, \sigma_k^*(X_{k-1}) \mid X_{k-1}, Y_{0:k-1}] \mid Y_{0:k-1}]$$

$$= \mathbb{E}[\,\sigma_k(X_{k-1})\, \mathbb{E}[W_k\ W_k^* \mid X_{k-1}, Y_{0:k-1}]\, \sigma_k^*(X_{k-1}) \mid Y_{0:k-1}]$$

$$= \mathbb{E}[\,\sigma_k(X_{k-1})\, \sigma_k^*(X_{k-1}) \mid Y_{0:k-1}]$$

et

$$\mathbb{E}[\sigma_k(X_{k-1})\, W_k\ (b_k(X_{k-1}) - \widehat{X}_k^-)^* \mid Y_{0:k-1}]$$

$$= \mathbb{E}[\,\mathbb{E}[\sigma_k(X_{k-1})\, W_k\ (b_k(X_{k-1}) - \widehat{X}_k^-)^* \mid X_{k-1}, Y_{0:k-1}] \mid Y_{0:k-1}]$$

$$= \mathbb{E}[\sigma_k(X_{k-1})\, \mathbb{E}[W_k \mid X_{k-1}, Y_{0:k-1}]\, (b_k(X_{k-1}) - \widehat{X}_k^-)^* \mid Y_{0:k-1}] = 0$$

où on a utilisé dans chacune des dernières égalités l'indépendance de $(X_{k-1}, Y_0, \cdots, Y_{k-1})$ et de W_k, de sorte que $\mathbb{E}[W_k\ W_k^* \mid X_{k-1}, Y_{0:k-1}] = I$ et $\mathbb{E}[W_k \mid X_{k-1}, Y_{0:k-1}] = 0$

il reste

$$\mathbb{E}[\,(X_k - \widehat{X}_k^-)\,(X_k - \widehat{X}_k^-)^* \mid Y_{0:k-1}]$$

$$= \mathbb{E}[\,(b_k(X_{k-1}) - \widehat{X}_k^-)\,(b_k(X_{k-1}) - \widehat{X}_k^-)^* \mid Y_{0:k-1}]$$

$$+ \,\mathbb{E}[\sigma_k(X_{k-1})\,\sigma_k^*(X_{k-1}) \mid Y_{0:k-1}]$$

$$= \int (b_k(x) - \widehat{X}_k^-)\,(b_k(x) - \widehat{X}_k^-)^*\,\mu_{k-1}(dx)$$

$$+ \int \sigma_k(x)\,\sigma_k^*(x)\,\mu_{k-1}(dx) \qquad \square$$

le calcul des deux premiers moments de la distribution de probabilité conditionnelle $\mu_k(dx) = \mathbb{P}[X_k \in dx \mid Y_{0:k}]$ n'est pas facile, et on commence par calculer les deux premiers moments de la distribution de probabilité conditionnelle jointe du vecteur aléatoire (X_k, Y_k) sachant $Y_{0:k-1}$ (on a déjà calculé $\widehat{X}_k^-$ et P_k^- plus haut)

Lemme

$$\widehat{Y}_k^- = \mathbb{E}[Y_k \mid Y_{0:k-1}] = \int h_k(x)\, \mu_k^-(dx)$$

et

$$\begin{aligned} \Xi_k &= \mathbb{E}[\,(Y_k - \widehat{Y}_k^-)\,(Y_k - \widehat{Y}_k^-)^* \mid Y_{0:k-1}] \\ &= \int (h_k(x) - \widehat{Y}_k^-)\,(h_k(x) - \widehat{Y}_k^-)^*\, \mu_k^-(dx) + Q_k^V \end{aligned}$$

$$\begin{aligned} C_k &= \mathbb{E}[\,(X_k - \widehat{X}_k^-)\,(Y_k - \widehat{Y}_k^-)^* \mid Y_{0:k-1}] \\ &= \int (x - \widehat{X}_k^-)\,(h_k(x) - \widehat{Y}_k^-)^*\, \mu_k^-(dx) \end{aligned}$$

Preuve clairement

$$\mathbb{E}[Y_k \mid Y_{0:k-1}] = \mathbb{E}[h_k(X_k) \mid Y_{0:k-1}] + \mathbb{E}[V_k \mid Y_{0:k-1}]$$

et on remarque que $\mathbb{E}[V_k \mid Y_{0:k-1}] = 0$

il reste

$$\mathbb{E}[Y_k \mid Y_{0:k-1}] = \mathbb{E}[h_k(X_k) \mid Y_{0:k-1}] = \int h_k(x)\, \mu_k^-(dx)$$

par différence

$$Y_k - \widehat{Y}_k^- = (h_k(X_k) - \widehat{Y}_k^-) + V_k$$

clairement

$$\mathbb{E}[\,(Y_k - \widehat{Y}_k^-)\,(Y_k - \widehat{Y}_k^-)^* \mid Y_{0:k-1}]$$

$$= \mathbb{E}[\,((h_k(X_k) - \widehat{Y}_k^-) + V_k)\,((h_k(X_k) - \widehat{Y}_k^-) + V_k)^* \mid Y_{0:k-1}]$$

$$= \mathbb{E}[\,(h_k(X_k) - \widehat{Y}_k^-)\,(h_k(X_k) - \widehat{Y}_k^-)^* \mid Y_{0:k-1}]$$

$$+ \mathbb{E}[V_k\ V_k^* \mid Y_{0:k-1}]$$

$$+ \mathbb{E}[V_k\ (h_k(X_k) - \widehat{Y}_k^-)^* \mid Y_{0:k-1}]$$

$$+ \mathbb{E}[\,(h_k(X_k) - \widehat{Y}_k^-)\ V_k^* \mid Y_{0:k-1}]$$

et

$$\mathbb{E}[\,(X_k - \widehat{X}_k^-)\,(Y_k - \widehat{Y}_k^-)^* \mid Y_{0:k-1}]$$

$$= \mathbb{E}[\,(X_k - \widehat{X}_k^-)\,(h_k(X_k) - \widehat{Y}_k^-)^* \mid Y_{0:k-1}]$$

$$+ \mathbb{E}[\,(X_k - \widehat{X}_k^-)\ V_k^* \mid Y_{0:k-1}]$$

on remarque que

$$\mathbb{E}[V_k\,(h_k(X_k) - \widehat{Y}_k^-)^* \mid Y_{0:k-1}]$$

$$= \mathbb{E}[\,\mathbb{E}[V_k\,(h_k(X_k) - \widehat{Y}_k^-)^* \mid X_k, Y_{0:k-1}] \mid Y_{0:k-1}]$$

$$= \mathbb{E}[\,\mathbb{E}[V_k \mid X_k, Y_{0:k-1}]\,(h_k(X_k) - \widehat{Y}_k^-)^* \mid Y_{0:k-1}] = 0$$

et

$$\mathbb{E}[\,(X_k - \widehat{X}_k^-)\,V_k^* \mid Y_{0:k-1}]$$

$$\mathbb{E}[\,\mathbb{E}[(X_k - \widehat{X}_k^-)\,V_k^* \mid X_k, Y_{0:k-1}] \mid Y_{0:k-1}]$$

$$\mathbb{E}[\,(X_k - \widehat{X}_k^-)\,\mathbb{E}[V_k^* \mid X_k, Y_{0:k-1}] \mid Y_{0:k-1}] = 0$$

où on a utilisé dans chacunes des dernières égalités l'indépendance de $(X_k, Y_0, \cdots, Y_{k-1})$ et V_k, de sorte que $\mathbb{E}[V_k \mid X_k, Y_{0:k-1}] = 0$

il reste

$$
\begin{aligned}
\mathbb{E}[\,(Y_k - \widehat{Y}_k^-)\,(Y_k - \widehat{Y}_k^-)^* \mid Y_{0:k-1}] \\
&= \mathbb{E}[\,(h_k(X_k) - \widehat{Y}_k^-)\;(h_k(X_k) - \widehat{Y}_k^-)^* \mid Y_{0:k-1}] \\
&\qquad + \mathbb{E}[V_k\;V_k^* \mid Y_{0:k-1}] \\
&= \int (h_k(x) - \widehat{Y}_k^-)\;(h_k(x) - \widehat{Y}_k^-)^*\,\mu_k^-(dx) + Q_k^V
\end{aligned}
$$

où on a utilisé dans la dernière égalité l'indépendance de $(Y_0, \cdots, Y_{k-1})$ et V_k, de sorte que $\mathbb{E}[V_k\;V_k^* \mid Y_{0:k-1}] = Q_k^V$, et

$$
\begin{aligned}
\mathbb{E}[\,(X_k - \widehat{X}_k^-)\,(Y_k - \widehat{Y}_k^-)^* \mid Y_{0:k-1}] \\
&= \mathbb{E}[\,(X_k - \widehat{X}_k^-)\,(h_k(X_k) - \widehat{Y}_k^-)^* \mid Y_{0:k-1}] \\
&= \int (x - \widehat{X}_k^-)\,(h_k(x) - \widehat{Y}_k^-)^*\,\mu_k^-(dx) \qquad \square
\end{aligned}
$$

si on remplace la distribution de probabilité conditionnelle jointe du vecteur aléatoire (X_k, Y_k) sachant $Y_{0:k-1}$ par la distribution de probabilité gaussienne de moyenne et de matrice de covariance

$$\begin{pmatrix} \widehat{X}_k^- \\ \widehat{Y}_k^- \end{pmatrix} \qquad \text{et} \qquad \begin{pmatrix} P_k^- & C_k \\ C_k^* & \Xi_k \end{pmatrix}$$

alors on obtient par conditionnement les approximations suivantes

$$\widehat{X}_k = \widehat{X}_k^- + C_k \, \Xi_k^{-1} \, (Y_k - \widehat{Y}_k^-) \qquad \text{et} \qquad P_k = P_k^- - C_k \, \Xi_k^{-1} \, C_k^*$$

pour les deux premiers moments de la distribution de probabilité conditionnelle $\mu_k(dx) = \mathbb{P}[X_k \in dx \mid Y_{0:k}]$

Remarque ces équations ne sont pas fermées, c'est–à–dire que les moments $\widehat{X}_k^-$ et P_k^- ne s'expriment pas en fonction des moments $\widehat{X}_{k-1}$ et P_{k-1} seulement, mais en fonction de toute la distribution de probabilité conditionnelle μ_{k-1}

de même, les moments $\widehat{X}_k$ et P_k ne s'expriment pas en fonction des moments $\widehat{X}_k^-$ et P_k^- seulement, mais en fonction de toute la distribution de probabilité conditionnelle μ_k^-

Principe de fermeture

on adopte le principe de projection énoncé plus haut

▶ on remplace la distribution de probabilité conditionnelle μ_{k-1} par la distribution de probabilité gaussienne de moyenne $\widehat{X}_{k-1}$ et de matrice de covariance $P_{k-1} = S_{k-1}\, S_{k-1}^*$

en effectuant le changement de variable $x = \widehat{X}_{k-1} + S_{k-1}\, u$, on obtient les approximations

$$\widehat{X}_k^- \approx \int \widehat{b}_k(u)\ \exp\{-\tfrac{1}{2}\,|u|^2\}\ \frac{du}{(2\pi)^{m/2}}$$

et

$$P_k^- \approx \int (\widehat{b}_k(u) - \widehat{X}_k^-)\,(\widehat{b}_k(u) - \widehat{X}_k^-)^*\ \exp\{-\tfrac{1}{2}\,|u|^2\}\ \frac{du}{(2\pi)^{m/2}}$$

$$+ \int \widehat{\sigma}_k(u)\,\widehat{\sigma}_k^*(u)\ \exp\{-\tfrac{1}{2}\,|u|^2\}\ \frac{du}{(2\pi)^{m/2}}$$

où par définition

$$\widehat{b}_k(u) = b_k(\widehat{X}_{k-1} + S_{k-1}\, u) \qquad \text{et} \qquad \widehat{\sigma}_k(u) = \sigma_k(\widehat{X}_{k-1} + S_{k-1}\, u)$$

- de même, on remplace la distribution de probabilité conditionnelle μ_k^- par la distribution de probabilité gaussienne de moyenne $\widehat{X}_k^-$ et de matrice de covariance $P_k^- = S_k^- \, (S_k^-)^*$
en effectuant le changement de variable $x = \widehat{X}_k^- + S_k^- \, u$, on obtient les approximations

$$\widehat{Y}_k^- \approx \int_{\mathbb{R}^m} \widehat{h}_k(u) \; \exp\{-\tfrac{1}{2} \, |u|^2\} \; \frac{du}{(2\pi)^{m/2}}$$

et

$$\Xi_k \approx \int_{\mathbb{R}^m} (\widehat{h}_k(u) - \widehat{Y}_k^-) \, (\widehat{h}_k(u) - \widehat{Y}_k^-)^* \; \exp\{-\tfrac{1}{2} \, |u|^2\} \; \frac{du}{(2\pi)^{m/2}} + Q_k^V$$

$$C_k \approx S_k^- \; \int_{\mathbb{R}^m} u \, (\widehat{h}_k(u) - \widehat{Y}_k^-)^* \; \exp\{-\tfrac{1}{2} \, |u|^2\} \; \frac{du}{(2\pi)^{m/2}}$$

où par définition

$$\widehat{h}_k(u) = h_k(\widehat{X}_k^- + S_k^- \, u)$$

Résumé il reste donc à calculer les intégrales des fonctions non–linéaires

$$\widehat{b}_k(u),\ \widehat{b}_k(u)\,\widehat{b}_k^*(u),\ \widehat{\sigma}_k(u)\,\widehat{\sigma}_k^*(u),\ \widehat{h}_k(u),\ u\,\widehat{h}_k^*(u) \text{ et } \widehat{h}_k(u)\,\widehat{h}_k^*(u)$$

par rapport à la densité gaussienne réduite centrée

Remarque si on suppose que les fonctions b_k et h_k sont dérivables, et qu'on utilise un développement limité au premier ordre au voisinage de $u = 0$ dans les intégrales ci–dessus, on retrouve les équations du filtre de Kalman étendu

l'idée ici est de *ne pas linéariser*, et de calculer les intégrales en utilisant des formules de quadrature

Formules de quadrature

en dimension m, la densité de probabilité gaussienne centrée réduite (de matrice de covariance identité) est représentée par $2m + 1$ points de quadrature $(u_{-m}, \cdots, u_0, \cdots, u_m)$ appelés σ–points, et définis par

$$u_0 = 0 \qquad \text{et} \qquad u_{\pm i} = \pm e_i \sqrt{m + \kappa}$$

où e_i désigne le i–ème vecteur de base, affectés des poids

$$w_0 = \frac{\kappa}{m + \kappa} \qquad \text{et} \qquad w_{\pm i} = \frac{1}{2\,(m + \kappa)} \qquad (\star)$$

pour tout $i = 1 \cdots m$ (d'autres choix de σ–points sont possibles)
on vérifie que les deux premiers moments sont pris en compte exactement

$$\sum_{i=-m}^{+m} w_i = 1\,, \qquad \sum_{i=-m}^{+m} w_i\ u_i = 0 \qquad \text{et} \qquad \sum_{i=-m}^{+m} w_i\ u_i\ u_i^* = \sum_{i=1}^{m} e_i\ e_i^* = I$$

concrètement

$$\int \phi(u)\ \exp\{-\tfrac{1}{2}\,|u|^2\}\ \frac{du}{(2\pi)^{m/2}} \approx \sum_{i=-m}^{+m} w_i\ \phi(u_i)$$

pour toute fonction ϕ, et un changement de variable évident donne

$$\int \phi(\mu + \Sigma^{1/2}\,u)\ \exp\{-\tfrac{1}{2}\,|u|^2\}\ \frac{du}{(2\pi)^{m/2}} \approx \sum_{i=-m}^{+m} w_i\ \phi(\mu + \Sigma^{1/2}\,u_i)$$

pour toute fonction ϕ : les σ–points $(x_{-m}, \cdots, x_0, \cdots, x_m)$ associés à la distribution de probabilité gaussienne de vecteur moyenne μ et de matrice de covariance Σ, sont définis par la relation $x_i = \mu + \Sigma^{1/2}\,u_i$, soit

$$x_0 = \mu \qquad \text{et} \qquad x_{\pm i} = \mu \pm \Sigma^{1/2}\,e_i\ \sqrt{m+\kappa}$$

on vérifie que les deux premiers moments sont pris en compte exactement

$$\sum_{i=-m}^{+m} w_i\ x_i = \mu \quad \text{et} \quad \sum_{i=-m}^{+m} w_i\ (x_i - \mu)\ (x_i - \mu)^* = \sum_{i=1}^{m} \Sigma^{1/2}\,e_i\ (\Sigma^{1/2}\,e_i)^* = \Sigma$$

Filtre de Kalman *unscented*

avec ces formules de quadrature, on obtient l'algorithme de filtrage sous–optimal suivant

▶ expression de $\widehat{X}_k^-$ et P_k^- en fonction de $\widehat{X}_{k-1}$ et $P_{k-1} = S_{k-1}\, S_{k-1}^*$

on introduit les σ–points

$$x_0 = \widehat{X}_{k-1} \qquad \text{et} \qquad x_{\pm i} = \widehat{X}_{k-1} \pm S_{k-1}\, e_i\, \sqrt{m+\kappa}$$

affectés des poids $(\star)$ pour tout $i = 1 \cdots m$, on définit le vecteur moyenne

$$\widehat{X}_k^- = \sum_{i=-m}^{+m} w_i\; b_k(x_i)$$

et la matrice de covariance

$$\begin{aligned} P_k^- &= \sum_{i=-m}^{+m} w_i\; (b_k(x_i) - \widehat{X}_k^-)\, (b_k(x_i) - \widehat{X}_k^-)^* + \sum_{i=-m}^{+m} w_i\; \sigma_k(x_i)\, \sigma_k^*(x_i) \\ &= S_k^-\, (S_k^-)^* \end{aligned}$$

▶ expression de $\widehat{X}_k$ et P_k en fonction de $\widehat{X}_k^-$ et $P_k^- = S_k^- \, (S_k^-)^*$
on introduit les σ–points

$$x_0 = \widehat{X}_k^- \qquad \text{et} \qquad x_{\pm i} = \widehat{X}_k^- \pm S_k^- \, e_i \, \sqrt{m + \kappa}$$

affectés des poids $(\star)$ pour tout $i = 1 \cdots m$, on définit le vecteur moyenne

$$\widehat{Y}_k^- = \sum_{i=-m}^{+m} w_i \; h_k(x_i)$$

et les matrices de covariance et de corrélation

$$\Xi_k = \sum_{i=-m}^{+m} w_i \, (h_k(x_i) - \widehat{Y}_k^-) \, (h_k(x_i) - \widehat{Y}_k^-)^* + Q_k^V$$

$$C_k = \sum_{i=-m}^{+m} w_i \, (x_i - \widehat{X}_k^-) \, (h_k(x_i) - \widehat{Y}_k^-)^*$$

et on pose

$$\widehat{X}_k = \widehat{X}_k^- + C_k \, \Xi_k^{-1} \, (Y_k - \widehat{Y}_k^-) \qquad \text{et} \qquad P_k = P_k^- - C_k \, \Xi_k^{-1} \, C_k^* = S_k \, S_k^*$$