Université de Rennes 1
DEA STIR
Signal — Télécommunications — Images — Radar
Option Signal

Examen du cours
"Filtre de Kalman et modèles de Markov cachés"
Jeudi 9 janvier 1997, 8:00 à 9:30
— Corrigé —

EXERCICE :

Soit $\{X_k\}$ une chaîne de Markov à espace d'état fini $E = \{1, \cdots, N\}$, caractérisée par les données suivantes :

- *loi initiale* $\nu = (\nu_i)$

$$\nu_i = \mathbf{P}[X_0 = i] \;, \qquad \text{pour tout } i \in E,$$

- *matrice de transition* $\pi = (\pi_{i,j})$

$$\pi_{i,j} = \mathbf{P}[X_{k+1} = j \mid X_k = i] \;, \qquad \text{pour tout } i, j \in E.$$

Soit $\{Y_k\}$ une suite d'observations, telle que :

$$Y_k = h(X_k) + V_k \;,$$

où la suite $\{V_k\}$ est un bruit blanc gaussien de dimension d, de moyenne nulle et de matrice de covariance identité, indépendant de la chaîne de Markov $\{X_k\}$. La fonction h définie sur E à valeurs dans $\mathbf{R}^d$ est caractérisée par la donnée d'une famille $h = (h_i)$ de N vecteurs de $\mathbf{R}^d$, et on a

$$\psi_i(y)\, dy = \mathbf{P}[Y_k \in dy \mid X_k = i] = \frac{1}{(\sqrt{2\pi})^d} \; \exp\Big\{ -\tfrac{1}{2}\, |y - h_i|^2 \Big\}\; dy \;,$$

pour tout $i \in E$, et tout $y \in \mathbf{R}^d$.

Le but de cet exercice est de proposer un algorithme d'estimation pour le vecteur des moyennes (h_i).

On considère dans un premier temps le cas où $\{X_0, \cdots, X_n\}$ et $\{Y_0, \cdots, Y_n\}$ sont observés. La fonction de log–vraisemblance pour l'estimation de $h = (h_i)$, à partir de l'observation de $\{X_0, \cdots, X_n\}$ et $\{Y_0, \cdots, Y_n\}$, s'écrit

$$\ell(h) = -\tfrac{1}{2} \sum_{k=0}^{n} |Y_k - h(X_k)|^2 \ .$$

(i) **En utilisant la décomposition**

$$h(X_k) = \sum_{i \in E} h_i \, \mathbf{1}_{[X_k = i]} \ ,$$

montrer que

$$\ell(h) = -\tfrac{1}{2} \sum_{i \in E} \sum_{k=0}^{n} \mathbf{1}_{[X_k = i]} \, |Y_k - h_i|^2 \ .$$

Solution

Par définition

$$\begin{aligned}
\ell(h) &= -\tfrac{1}{2} \sum_{k=0}^{n} |Y_k - h(X_k)|^2 \\
&= -\tfrac{1}{2} \sum_{k=0}^{n} \sum_{i \in E} \mathbf{1}_{[X_k = i]} \, |Y_k - h(X_k)|^2 \\
&= -\tfrac{1}{2} \sum_{i \in E} \sum_{k=0}^{n} \mathbf{1}_{[X_k = i]} \, |Y_k - h_i|^2 \ .
\end{aligned}$$

□

(ii) **En maximisant $\ell(h)$ par rapport au vecteur h_i, montrer que l'estimateur du maximum de vraisemblance pour la moyenne h_i, à partir de l'observation de $\{X_0, \cdots, X_n\}$ et $\{Y_0, \cdots, Y_n\}$, est donné par**

$$\widehat{h}_i = \frac{\displaystyle\sum_{k=0}^{n} Y_k \, \mathbf{1}_{[X_k = i]}}{\displaystyle\sum_{k=0}^{n} \mathbf{1}_{[X_k = i]}} \ .$$

Solution

Le vecteur $\widehat{h}_i$ maximise l'application

$$h_i \longmapsto -\tfrac{1}{2} \sum_{k=0}^{n} \mathbf{1}_{[X_k = i]} \, |Y_k - h_i|^2 \; .$$

Il vérifie l'équation

$$\sum_{k=0}^{n} \mathbf{1}_{[X_k = i]} \, (Y_k - h_i) = 0 \; ,$$

ou de façon équivalente

$$h_i \sum_{k=0}^{n} \mathbf{1}_{[X_k = i]} = \sum_{k=0}^{n} Y_k \, \mathbf{1}_{[X_k = i]} \; ,$$

c'est–à–dire que

$$\widehat{h}_i = \frac{\displaystyle\sum_{k=0}^{n} Y_k \, \mathbf{1}_{[X_k = i]}}{\displaystyle\sum_{k=0}^{n} \mathbf{1}_{[X_k = i]}} \; .$$

□

On considère ensuite le cas où seulement $\{Y_0, \cdots, Y_n\}$ est observé. On pose

$$N_i = \sum_{k=0}^{n} \mathbf{1}_{[X_k = i]} \qquad \text{et} \qquad O_i = \sum_{k=0}^{n} Y_k \, \mathbf{1}_{[X_k = i]} \; .$$

(iii) **En supposant le modèle $\mathbf{M} = (\nu, \pi, h)$ connu, donner l'expression des estimateurs $\widehat{N}_i = \mathbf{E}[N_i \mid \mathcal{Y}_n]$ et $\widehat{O}_i = \mathbf{E}[O_i \mid \mathcal{Y}_n]$ à l'aide des variables forward et backward correspondantes.**

Solution

On a

$$\widehat{N}_i = \mathbf{E}[N_i \mid \mathcal{Y}_n] = \sum_{k=0}^{n} \mathbf{P}[X_k = i \mid \mathcal{Y}_n] = \frac{1}{P} \sum_{k=0}^{n} q_k^i \; ,$$

où $q_k^i = p_k^i \, v_k^i$ pour tout $i \in E$, et pour tout instant $k = 0, \cdots, n$, et où la constante de normalisation

$$P = \sum_{j \in E} q_k^j \; ,$$

ne dépend pas de l'instant $k = 0, \cdots, n$.

De même

$$\widehat{O}_i = \mathbf{E}[O_i \mid \mathcal{Y}_n] = \sum_{k=0}^{n} \mathbf{E}[Y_k \, \mathbf{1}_{[X_k = i]} \mid \mathcal{Y}_n] = \sum_{k=0}^{n} Y_k \, \mathbf{P}[X_k = i \mid \mathcal{Y}_n] = \frac{1}{P} \sum_{k=0}^{n} Y_k \, q_k^i \; .$$

□

(iv) **Proposer un algorithme itératif pour l'estimation de la moyenne h_i à partir de l'observation de $\{Y_0, \cdots, Y_n\}$ seulement.**

Solution

On peut proposer l'algorithme itératif suivant, qui repose sur la notion de re–estimation.

initialisation : choix initial h.

étape E : résolution des équations forward–backward de Baum, pour le modèle $\mathbf{M} = (\nu, \pi, h)$

$$p_{k+1}^j = \psi_j(Y_{k+1}) \sum_{i \in E} \pi_{i,j}\, p_k^i \;, \qquad p_0^i = \nu^i \;,$$

$$v_k^i = \sum_{j \in E} \Big[\, \pi_{i,j}\, \psi_j(Y_{k+1})\, v_{k+1}^j \,\Big] \;, \qquad v_n^i = 1 \;,$$

où on a défini

$$\psi_i(y) = \frac{1}{(\sqrt{2\pi})^d} \, \exp\Big\{ -\tfrac{1}{2}\, |y - h_i|^2 \Big\} \;,$$

pour tout $i \in E$, et pour tout $y \in \mathbf{R}^d$.

étape M : re–estimation du vecteur des moyennes

$$\widehat{h}_i = \frac{\widehat{O}_i}{\widehat{N}_i} = \frac{\displaystyle\sum_{k=0}^{n} Y_k\, q_k^i}{\displaystyle\sum_{k=0}^{n} q_k^i} \;.$$

itération : $h = \widehat{h}$, et retour à l'étape E.

□

Problème :

Le but de ce problème est d'établir les équations de l'estimateur du maximum a posteriori dans le cas linéaire gaussien, et de montrer que cet estimateur coïncide avec le filtre de Kalman.

On considère le système linéaire suivant :

$$X_{k+1} = F\,X_k + W_k\ , \qquad (\text{dans } \mathbf{R}^m)$$

$$Y_k = H\,X_k + V_k\ , \qquad (\text{dans } \mathbf{R}^d)$$

avec les hypothèses habituelles :

- X_0 *variable aléatoire gaussienne, de moyenne* μ_0 *et de matrice de covariance* Σ_0.
- $\{W_k\}$ *bruit blanc gaussien centré de matrice de covariance* Q.
- $\{V_k\}$ *bruit blanc gaussien centré de matrice de covariance* R *inversible.*
- X_0, $\{W_k\}$ *et* $\{V_k\}$ *mutuellement indépendants.*

Pour simplifier la dérivation, on suppose en outre que les matrices de covariance Σ_0 et Q sont *inversibles.*

(i) **Montrer que la loi conditionnelle de X_k sachant $(X_0, \cdots, X_{k-1})$ est donnée par**

$$\mathbf{P}[X_k \in dx \mid X_0 = x_0, \cdots, X_{k-1} = x_{k-1}] =$$

$$= \frac{1}{(\sqrt{2\pi})^m\,\sqrt{\det Q}}\ \exp\Big\{ -\tfrac{1}{2}\,(x - F\,x_{k-1})^*\,Q^{-1}\,(x - F\,x_{k-1}) \Big\}\ dx\ .$$

En déduire que la suite $\{X_k\}$ est une chaîne de Markov à valeur dans $\mathbf{R}^m$.

Solution

Par définition

$$X_k = F\,X_{k-1} + W_{k-1}\ ,$$

et on remarque que $(X_0, \cdots, X_{k-1})$ ne dépend que de $(X_0, W_0, \cdots, W_{k-2})$. Il en résulte que :

- le vecteur aléatoire X_{k-1} dépend de $(X_0, \cdots, X_{k-1})$,
- et le vecteur aléatoire W_{k-1} est indépendant de $(X_0, \cdots, X_{k-1})$.

On peut donc dire que *la connaissance de* $(X_0, \cdots, X_{k-1})$ *n'apprend rien de plus sur* X_k *que la connaissance de* X_{k-1} *seulement.*

Il en résulte que la suite $\{X_k\}$ est une chaîne de Markov à valeurs dans $\mathbf{R}^m$.

Conditionnellement à $(X_{k-1} = x_{k-1})$, le vecteur aléatoire X_k est un vecteur aléatoire gaussien de $\mathbf{R}^m$, de moyenne $F\, x_{k-1}$ et de matrice de covariance Q, d'où la loi conditionnelle

$$\mathbf{P}[X_k \in dx \mid X_0 = x_0, \cdots, X_{k-1} = x_{k-1}] = \mathbf{P}[X_k \in dx \mid X_{k-1} = x_{k-1}] =$$

$$= \frac{1}{(\sqrt{2\pi})^m \sqrt{\det Q}} \exp\Big\{ -\tfrac{1}{2} (x - F\, x_{k-1})^* \, Q^{-1} \, (x - F\, x_{k-1}) \Big\} \; dx \; .$$

□

(ii) **Déduire de (i) que la loi jointe de** $(X_0, \cdots, X_n)$ **est donnée par**

$$\mathbf{P}[X_0 \in dx_0, \cdots, X_n \in dx_n] =$$

$$= \frac{1}{(\sqrt{2\pi})^m \sqrt{\det \Sigma_0}} \Big[\frac{1}{(\sqrt{2\pi})^m \sqrt{\det Q}} \Big]^n$$

$$\exp\Big\{ -\tfrac{1}{2} (x_0 - \mu_0)^* \, \Sigma_0^{-1} \, (x_0 - \mu_0) - \tfrac{1}{2} \sum_{k=1}^{n} (x_k - F\, x_{k-1})^* \, Q^{-1} \, (x_k - F\, x_{k-1}) \Big\}$$

$$dx_0 \; dx_1 \cdots dx_n \; .$$

Solution

La loi jointe de $(X_0, \cdots, X_n)$ est obtenue par application répétée de la formule de Bayes

$$\mathbf{P}[X_0 \in dx_0, \cdots, X_n \in dx_n] =$$

$$= \mathbf{P}[X_n \in dx_n \mid X_0 = x_0, \cdots, X_{n-1} = x_{n-1}] \; \mathbf{P}[X_0 \in dx_0, \cdots, X_{n-1} \in dx_{n-1}]$$

$$= \prod_{k=1}^{n} \mathbf{P}[X_k \in dx_k \mid X_0 = x_0, \cdots, X_{k-1} = x_{k-1}] \; \mathbf{P}[X_0 \in dx_0] \; ,$$

ce qui donne d'après (i)

$$\mathbf{P}[X_0 \in dx_0, \cdots, X_n \in dx_n] =$$

$$= \frac{1}{(\sqrt{2\pi})^m \sqrt{\det \Sigma_0}} \Big[\frac{1}{(\sqrt{2\pi})^m \sqrt{\det Q}} \Big]^n$$

$$\exp\Big\{ -\tfrac{1}{2} (x_0 - \mu_0)^* \, \Sigma_0^{-1} \, (x_0 - \mu_0) - \tfrac{1}{2} \sum_{k=1}^{n} (x_k - F\, x_{k-1})^* \, Q^{-1} \, (x_k - F\, x_{k-1}) \Big\}$$

$$dx_0 \; dx_1 \cdots dx_n \; .$$

$\square$

(iii) **Montrer que la propriété de canal sans mémoire est vérifiée, et que**

$$\mathbf{P}[Y_k \in dy \mid X_k = x] = \frac{1}{(\sqrt{2\pi})^d \sqrt{\det R}} \exp\Big\{ -\tfrac{1}{2} (y - H\,x)^* R^{-1} (y - H\,x) \Big\}\; dy \; .$$

SOLUTION

Par définition

$$\begin{pmatrix} Y_0 \\ \vdots \\ Y_k \end{pmatrix} = \begin{pmatrix} H & 0 & \cdots & \cdots & 0 \\ 0 & H & 0 & & \vdots \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \vdots & & 0 & H & 0 \\ 0 & \cdots & \cdots & 0 & H \end{pmatrix} \begin{pmatrix} X_0 \\ \vdots \\ X_k \end{pmatrix} + \begin{pmatrix} V_0 \\ \vdots \\ V_k \end{pmatrix} ,$$

et par hypothèse $(V_0, \cdots, V_k)$ est indépendant de $(X_0, \cdots, X_k)$. Il en résulte que, conditionnellement à $(X_0, \cdots, X_k)$ les vecteurs aléatoires $Y_0, \cdots, Y_k$ sont des vecteurs aléatoires gaussiens de $\mathbf{R}^d$ mutuellement indépendants. En outre

$$Y_l = H\,X_l + V_l \; ,$$

pour tout $l = 0, \cdots, k$, et le vecteur aléatoire V_l est indépendant de $(X_0, \cdots, X_k)$. On peut donc dire que *la connaissance de* $(X_0, \cdots, X_k)$ *n'apprend rien de plus sur* Y_l *que la connaissance de* X_l *seulement.*

Il résulte de ces deux remarques que la propriété de canal sans mémoire est vérifiée.

En particulier, conditionnellement à $(X_k = x)$, le vecteur aléatoire Y_k est un vecteur aléatoire gaussien de moyenne $H\,x$ et de matrice de covariance R, d'où la loi conditionnelle

$$\mathbf{P}[Y_k \in dy \mid X_k = x] = \frac{1}{(\sqrt{2\pi})^d \sqrt{\det R}} \exp\Big\{ -\tfrac{1}{2} (y - H\,x)^* R^{-1} (y - H\,x) \Big\}\; dy \; .$$

$\square$

(iv) **Déduire de (ii) et (iii) que la loi jointe de** $(X_0, \cdots, X_n, Y_0, \cdots, Y_n)$ **est donnée par**

$$\mathbf{P}[X_0 \in dx_0, \cdots, X_n \in dx_n, Y_0 \in dy_0, \cdots, Y_n \in dy_n] =$$

$$= \frac{1}{(\sqrt{2\pi})^m \sqrt{\det \Sigma_0}} \Big[\frac{1}{(\sqrt{2\pi})^m \sqrt{\det Q}} \Big]^n \Big[\frac{1}{(\sqrt{2\pi})^d \sqrt{\det R}} \Big]^{n+1}$$

$$\exp\Big\{ -\tfrac{1}{2} (x_0 - \mu_0)^* \Sigma_0^{-1} (x_0 - \mu_0) - \tfrac{1}{2} \sum_{k=1}^{n} (x_k - F\,x_{k-1})^* Q^{-1} (x_k - F\,x_{k-1})$$

$$- \tfrac{1}{2} \sum_{k=0}^{n} (y_k - H\,x_k)^* R^{-1} (y_k - H\,x_k) \Big\}$$

$$dx_0\, dx_1 \cdots dx_n\, dy_0\, dy_1 \cdots dy_n \; .$$

Solution

La loi jointe de $(X_0, \cdots, X_n, Y_0, \cdots, Y_n)$ est obtenue par application de la formule de Bayes

$$\mathbf{P}[X_0 \in dx_0, \cdots, X_n \in dx_n, Y_0 \in dy_0, \cdots, Y_n \in dy_n] =$$

$$= \mathbf{P}[Y_0 \in dy_0, \cdots, Y_n \in dy_n \mid X_0 = x_0, \cdots, X_n = x_n] \, \mathbf{P}[X_0 \in dx_0, \cdots, X_n \in dx_n] \ .$$

D'après la propriété de canal sans mémoire, et d'après l'expression obtenue en (iii), la loi de $(Y_0, \cdots, Y_n)$ sachant $(X_0, \cdots, X_n)$ est donnée par

$$\mathbf{P}[Y_0 \in dy_0, \cdots, Y_n \in dy_n \mid X_0 = x_0, \cdots, X_n = x_n] = \prod_{k=0}^{n} \mathbf{P}[Y_k \in dy_k \mid X_k = x_k] =$$

$$= \Big[\frac{1}{(\sqrt{2\pi})^d \sqrt{\det R}} \Big]^{n+1} \exp\Big\{ -\tfrac{1}{2} \sum_{k=0}^{n} (y_k - H\, x_k)^* \, R^{-1} \, (y_k - H\, x_k) \Big\} \; dy_0 \, dy_1 \cdots dy_n \ .$$

On obtient le résultat demandé en utilisant l'expression de la loi jointe de $(X_0, \cdots, X_n)$ obtenue en (ii).

$\square$

Par définition, l'estimateur du maximum a posteriori pour la suite $\{X_0, \cdots, X_n\}$ est la suite d'états $(x_0, \cdots, x_n)$ qui maximise la densité conditionnelle de $(X_0, \cdots, X_n)$ sachant $(Y_0, \cdots, Y_n)$.

(v) **Montrer que maximiser par rapport à $(x_0, \cdots, x_n)$ la densité conditionnelle de $(X_0, \cdots, X_n)$ sachant $(Y_0, \cdots, Y_n)$ donne le même résultat que maximiser par rapport à $(x_0, \cdots, x_n)$ la densité jointe de $(X_0, \cdots, X_n, Y_0, \cdots, Y_n)$.**

Solution

Par définition

$$p_{X_0, \cdots, X_n \mid Y_0 = y_0, \cdots, Y_n = y_n}(x_0, \cdots, x_n) \; p_{Y_0, \cdots, Y_n}(y_0, \cdots, y_n) =$$

$$= p_{X_0, \cdots, X_n, Y_0, \cdots, Y_n}(x_0, \cdots, x_n, y_0, \cdots, y_n) \ .$$

La densité conditionnelle de $(X_0, \cdots, X_n)$ sachant $(Y_0, \cdots, Y_n)$ et la densité jointe de $(X_0, \cdots, X_n, Y_0, \cdots, Y_n)$ ne diffèrent donc que par une constante multiplicative qui ne dépend que de $(y_0, \cdots, y_n)$, d'où le résultat.

$\square$

On se propose donc de minimiser par rapport à $(x_0, \cdots, x_n)$ la fonction suivante

$$\begin{aligned} J_n(x_0, \cdots, x_n) \;=\; & \tfrac{1}{2}\, (x_0 - \mu_0)^* \, \Sigma_0^{-1} \, (x_0 - \mu_0) \\ & + \tfrac{1}{2} \sum_{k=1}^{n} (x_k - F\, x_{k-1})^* \, Q^{-1} \, (x_k - F\, x_{k-1}) \\ & + \tfrac{1}{2} \sum_{k=0}^{n} (Y_k - H\, x_k)^* \, R^{-1} \, (Y_k - H\, x_k) \ . \end{aligned}$$

On définit

$$V_n(x) = \min_{x_0, \cdots, x_{n-1}} J_n(x_0, \cdots, x_{n-1}, x) \ .$$

L'objectif de la fin du problème est de montrer par récurrence que

$$V_n(x) = \tfrac{1}{2}\, (x - \mu_n)^* \, \Sigma_n^{-1} \, (x - \mu_n) + c_n \ , \qquad (\star)$$

et de caractériser la matrice Σ_n, le vecteur μ_n, et le scalaire c_n.

(vi) **Déduire de $(\star)$ que μ_n est l'estimateur du maximum a posteriori pour l'état X_n.**

Solution

Pour tout $x \in \mathbf{R}^m$

$$V_n(x) = \min_{x_0, \cdots, x_{n-1}} J_n(x_0, \cdots, x_{n-1}, x) = J_n(\xi_0(x), \cdots, \xi_{n-1}(x), x) \ ,$$

pour une certaine suite $(\xi_0(x), \cdots, \xi_{n-1}(x))$ qui dépend de x. D'autre part

$$\min_x V_n(x) = \min_{x_0, \cdots, x_n} J_n(x_0, \cdots, x_n) = J_n(X_0^{\mathrm{MAP}}, \cdots, X_n^{\mathrm{MAP}}) \ ,$$

où par définition la suite $(X_0^{\mathrm{MAP}}, \cdots, X_n^{\mathrm{MAP}})$ est appelée *estimateur du maximum a posteriori* de la suite $(X_0, \cdots, X_n)$.

D'après la formule de récurrence $(\star)$, le minimum de l'application V_n est atteint pour $x = \mu_n$, c'est–à–dire que

$$J_n(X_0^{\mathrm{MAP}}, \cdots, X_n^{\mathrm{MAP}}) = V_n(\mu_n) = J_n(\xi_0(\mu_n), \cdots, \xi_{n-1}(\mu_n), \mu_n) \ .$$

On en déduit que

$$(X_0^{\mathrm{MAP}}, \cdots, X_n^{\mathrm{MAP}}) = (\xi_0(\mu_n), \cdots, \xi_{n-1}(\mu_n), \mu_n) \ ,$$

pourvu que le minimum de l'application J_n soit unique. En particulier

$$X_n^{\mathrm{MAP}} = \mu_n \ .$$

□

(vii) **Montrer que**

$$J_{k+1}(x_0,\cdots,x_k,x) \;=\; J_k(x_0,\cdots,x_k) + \tfrac{1}{2}\,(x - F\,x_k)^*\,Q^{-1}\,(x - F\,x_k)$$

$$+\; \tfrac{1}{2}\,(Y_{k+1} - H\,x)^*\,R^{-1}\,(Y_{k+1} - H\,x)\;\;.$$

Solution

Par définition

$$\begin{aligned}
& J_{k+1}(x_0,\cdots,x_k,x) = \\
&= \tfrac{1}{2}\,(x_0-\mu_0)^*\,\Sigma_0^{-1}\,(x_0-\mu_0) + \tfrac{1}{2}\,\sum_{l=1}^{k}(x_l - F\,x_{l-1})^*\,Q^{-1}\,(x_l - F\,x_{l-1}) \\
&\quad + \tfrac{1}{2}\,(x - F\,x_k)^*\,Q^{-1}\,(x - F\,x_k) \\
&\quad + \tfrac{1}{2}\,\sum_{l=0}^{k}(Y_l - H\,x_l)^*\,R^{-1}\,(Y_l - H\,x_l) + \tfrac{1}{2}\,(Y_{k+1} - H\,x)^*\,R^{-1}\,(Y_{k+1} - H\,x) \\
&= J_k(x_0,\cdots,x_k) + \tfrac{1}{2}\,(x - F\,x_k)^*\,Q^{-1}\,(x - F\,x_k) \\
&\quad + \tfrac{1}{2}\,(Y_{k+1} - H\,x)^*\,R^{-1}\,(Y_{k+1} - H\,x)\;\;.
\end{aligned}$$

□

(viii) **En prenant le minimum par rapport à $(x_0,\cdots,x_k)$ et en utilisant l'hypothèse de récurrence $(\star)$ supposée vraie à l'instant k, montrer que**

$$V_{k+1}(x) \;=\; \min_{x_k}\Big[\,\tfrac{1}{2}\,(x_k-\mu_k)^*\,\Sigma_k^{-1}\,(x_k-\mu_k) + \tfrac{1}{2}\,(x - F\,x_k)^*\,Q^{-1}\,(x - F\,x_k)\,\Big]$$

$$+\; \tfrac{1}{2}\,(Y_{k+1} - H\,x)^*\,R^{-1}\,(Y_{k+1} - H\,x) + c_k\;\;.$$

Solution

D'après (vii)

$$\begin{aligned}
V_{k+1}(x) &= \min_{x_0,\cdots,x_k} J_{k+1}(x_0,\cdots,x_k,x) \\
&= \min_{x_0,\cdots,x_k}\Big[\,J_k(x_0,\cdots,x_k) + \tfrac{1}{2}\,(x - F\,x_k)^*\,Q^{-1}\,(x - F\,x_k)\,\Big] \\
&\quad + \tfrac{1}{2}\,(Y_{k+1} - H\,x)^*\,R^{-1}\,(Y_{k+1} - H\,x)
\end{aligned}$$

$$= \min_{x_k} \Big[\min_{x_0,\cdots,x_{k-1}} J_k(x_0,\cdots,x_k) + \tfrac{1}{2}\,(x - F\,x_k)^*\,Q^{-1}\,(x - F\,x_k) \Big]$$

$$+ \tfrac{1}{2}\,(Y_{k+1} - H\,x)^*\,R^{-1}\,(Y_{k+1} - H\,x)$$

$$= \min_{x_k} \Big[V_k(x_k) + \tfrac{1}{2}\,(x - F\,x_k)^*\,Q^{-1}\,(x - F\,x_k) \Big]$$

$$+ \tfrac{1}{2}\,(Y_{k+1} - H\,x)^*\,R^{-1}\,(Y_{k+1} - H\,x)\ ,$$

et d'après l'hypothèse de récurrence $(\star)$

$$V_{k+1}(x) \;=\; \min_{x_k} \Big[\tfrac{1}{2}\,(x_k - \mu_k)^*\,\Sigma_k^{-1}\,(x_k - \mu_k) + c_k + \tfrac{1}{2}\,(x - F\,x_k)^*\,Q^{-1}\,(x - F\,x_k) \Big]$$

$$+ \tfrac{1}{2}\,(Y_{k+1} - H\,x)^*\,R^{-1}\,(Y_{k+1} - H\,x)\ .$$

$\square$

(ix) **Effectuer la minimisation par rapport à x_k, et montrer que**

$$V_{k+1}(x) \;=\; \tfrac{1}{2}\,(x - F\,\mu_k)^*\,(F\,\Sigma_k\,F^* + Q)^{-1}\,(x - F\,\mu_k)$$

$$+ \tfrac{1}{2}\,(Y_{k+1} - H\,x)^*\,R^{-1}\,(Y_{k+1} - H\,x) + c_k\ .$$

Solution

On commence par développer, et par compléter le carré, de la façon suivante

$$\tfrac{1}{2}\,(x_k - \mu_k)^*\,\Sigma_k^{-1}\,(x_k - \mu_k) + \tfrac{1}{2}\,(x - F\,x_k)^*\,Q^{-1}\,(x - F\,x_k) =$$

$$= \tfrac{1}{2}\,x_k^*\,(\Sigma_k^{-1} + F^*\,Q^{-1}\,F)\,x_k - x_k^*\,(\Sigma_k^{-1}\,\mu_k + F^*\,Q^{-1}\,x) + \tfrac{1}{2}\,(\mu_k^*\,\Sigma_k^{-1}\,\mu_k + x^*\,Q^{-1}\,x)$$

$$= \tfrac{1}{2}\,x_k^*\,A\,x_k - x_k^*\,b + \tfrac{1}{2}\,e$$

$$= \tfrac{1}{2}\,(x_k - A^{-1}\,b)^*\,A\,(x_k - A^{-1}\,b) + \tfrac{1}{2}\,e - \tfrac{1}{2}\,b^*\,A^{-1}\,b\ ,$$

avec

$$A = \Sigma_k^{-1} + F^*\,Q^{-1}\,F\ , \qquad b = \Sigma_k^{-1}\,\mu_k + F^*\,Q^{-1}\,x\ , \qquad e = \mu_k^*\,\Sigma_k^{-1}\,\mu_k + x^*\,Q^{-1}\,x\ .$$

On en déduit d'abord que

$$\min_{x_k} \Big[\tfrac{1}{2}\,(x_k - \mu_k)^*\,\Sigma_k^{-1}\,(x_k - \mu_k) + \tfrac{1}{2}\,(x - F\,x_k)^*\,Q^{-1}\,(x - F\,x_k) \Big] = \tfrac{1}{2}\,e - \tfrac{1}{2}\,b^*\,A^{-1}\,b\ .$$

On remarque ensuite, d'après le Lemme 1.1 du cours, que

$$A^{-1} = (\Sigma_k^{-1} + F^*\, Q^{-1}\, F)^{-1} = \Sigma_k - \Sigma_k\, F^*\, (F\, \Sigma_k\, F^* + Q)^{-1}\, F\, \Sigma_k \ ,$$

d'où

$$\begin{aligned}
A^{-1}\, b \ &= \ [\, \Sigma_k - \Sigma_k\, F^*\, (F\, \Sigma_k\, F^* + Q)^{-1}\, F\, \Sigma_k\,]\ (\Sigma_k^{-1}\, \mu_k + F^*\, Q^{-1}\, x) \\
&= \ \mu_k + \Sigma_k\, F^*\, Q^{-1}\, x - \Sigma_k\, F^*\, (F\, \Sigma_k\, F^* + Q)^{-1}\, F\, \mu_k \\
&\quad -\Sigma_k\, F^*\, (F\, \Sigma_k\, F^* + Q)^{-1}\, F\, \Sigma_k\, F^*\, Q^{-1}\, x \\
&= \ \mu_k + \Sigma_k\, F^*\, (F\, \Sigma_k\, F^* + Q)^{-1}\, (x - F\, \mu_k) \ ,
\end{aligned}$$

et finalement

$$\begin{aligned}
b^*\, A^{-1}\, b \ &= \ (\mu_k^*\, \Sigma_k^{-1} + x^*\, Q^{-1}\, F)\ [\, \mu_k + \Sigma_k\, F^*\, (F\, \Sigma_k\, F^* + Q)^{-1}\, (x - F\, \mu_k)\,] \\
&= \ \mu_k^*\, \Sigma_k^{-1}\, \mu_k + \mu_k^*\, F^*\, (F\, \Sigma_k\, F^* + Q)^{-1}\, (x - F\, \mu_k) \\
&\quad + x^*\, Q^{-1}\, F\, \mu_k + x^*\, Q^{-1}\, F\, \Sigma_k\, F^*\, (F\, \Sigma_k\, F^* + Q)^{-1}\, (x - F\, \mu_k) \\
&= \ \mu_k^*\, \Sigma_k^{-1}\, \mu_k + x^*\, Q^{-1}\, x - (x - F\, \mu_k)^*\, (F\, \Sigma_k\, F^* + Q)^{-1}\, (x - F\, \mu_k) \ .
\end{aligned}$$

On en déduit que

$$e - b^*\, A^{-1}\, b = (x - F\, \mu_k)^*\, (F\, \Sigma_k\, F^* + Q)^{-1}\, (x - F\, \mu_k) \ ,$$

et finalement

$$\begin{aligned}
V_{k+1}(x) \ &= \ \tfrac{1}{2}\, (x - F\, \mu_k)^*\, (F\, \Sigma_k\, F^* + Q)^{-1}\, (x - F\, \mu_k) \\
&\quad + \tfrac{1}{2}\, (Y_{k+1} - H\, x)^*\, R^{-1}\, (Y_{k+1} - H\, x) + c_k \ .
\end{aligned}$$

$\square$

On introduit les notations suivantes :

$$\mu_{k+1}^- = F\, \mu_k \qquad \text{et} \qquad \Sigma_{k+1}^- = F\, \Sigma_k\, F^* + Q \ ,$$

ce qui donne

$$\begin{aligned}
V_{k+1}(x) \ &= \ \tfrac{1}{2}\, (x - \mu_{k+1}^-)^*\, (\Sigma_{k+1}^-)^{-1}\, (x - \mu_{k+1}^-) \\
&\quad + \tfrac{1}{2}\, (Y_{k+1} - H\, x)^*\, R^{-1}\, (Y_{k+1} - H\, x) + c_k \ .
\end{aligned}$$

(x) **Montrer que**

$$V_{k+1}(x) = \tfrac{1}{2}\,(x - \mu_{k+1})^*\,\Sigma_{k+1}^{-1}\,(x - \mu_{k+1}) + c_{k+1}\ ,$$

c'est–à–dire que la propriété ($\star$) est vérifiée, avec en particulier

$$\mu_{k+1} = \mu_{k+1}^- + \Sigma_{k+1}^-\,H^*\,(H\,\Sigma_{k+1}^-\,H^* + R)^{-1}\,(Y_{k+1} - H\,\mu_{k+1}^-)\ .$$

Comparer ces équations avec celles du filtre de Kalman, et conclure.

Solution

On développe, et on complète le carré, de la façon suivante

$$\begin{aligned}
V_{k+1}(x) &= \tfrac{1}{2}\,(x - \mu_{k+1}^-)^*\,(\Sigma_{k+1}^-)^{-1}\,(x - \mu_{k+1}^-) \\
&\quad + \tfrac{1}{2}\,(Y_{k+1} - H\,x)^*\,R^{-1}\,(Y_{k+1} - H\,x) + c_k \\
&= \tfrac{1}{2}\,x^*\,[\,(\Sigma_{k+1}^-)^{-1} + H^*\,R^{-1}\,H\,]\,x - x^*\,[\,(\Sigma_{k+1}^-)^{-1}\,\mu_{k+1}^- + H^*\,R^{-1}\,Y_{k+1}\,] \\
&\quad + \tfrac{1}{2}\,[\,(\mu_{k+1}^-)^*\,(\Sigma_{k+1}^-)^{-1}\,\mu_{k+1}^- + Y_{k+1}^*\,R^{-1}\,Y_{k+1}\,] + c_k \\
&= \tfrac{1}{2}\,x^*\,A\,x - x^*\,b + \tfrac{1}{2}\,e + c_k \\
&= \tfrac{1}{2}\,(x - A^{-1}\,b)^*\,A\,(x - A^{-1}\,b) + \tfrac{1}{2}\,e - \tfrac{1}{2}\,b^*\,A^{-1}\,b + c_k\ ,
\end{aligned}$$

avec

$$A = (\Sigma_{k+1}^-)^{-1} + H^*\,R^{-1}\,H\ , \qquad b = (\Sigma_{k+1}^-)^{-1}\,\mu_{k+1}^- + H^*\,R^{-1}\,Y_{k+1}\ ,$$

$$e = (\mu_{k+1}^-)^*\,(\Sigma_{k+1}^-)^{-1}\,\mu_{k+1}^- + Y_{k+1}^*\,R^{-1}\,Y_{k+1}\ .$$

On en déduit que

$$V_{k+1}(x) = \tfrac{1}{2}\,(x - \mu_{k+1})^*\,\Sigma_{k+1}^{-1}\,(x - \mu_{k+1}) + c_{k+1}\ ,$$

avec

$$\Sigma_{k+1} = A^{-1}\ , \qquad \mu_{k+1} = A^{-1}\,b\ , \qquad c_{k+1} = \tfrac{1}{2}\,e - \tfrac{1}{2}\,b^*\,A^{-1}\,b + c_k\ ,$$

c'est–à–dire que l'hypothèse de récurrence ($\star$) est vérifiée.

On remarque ensuite, d'après le Lemme 1.1 du cours, que

$$\begin{aligned}
\Sigma_{k+1} = A^{-1} &= [\,(\Sigma_{k+1}^-)^{-1} + H^*\,R^{-1}\,H\,]^{-1} \\
&= \Sigma_{k+1}^- - \Sigma_{k+1}^-\,H^*\,(H\,\Sigma_{k+1}^-\,H^* + R)^{-1}\,H\,\Sigma_{k+1}^-\ ,
\end{aligned}$$

d'où

$$\mu_{k+1} = A^{-1}\, b = [\, \Sigma^-_{k+1} - \Sigma^-_{k+1}\, H^*\, (H\, \Sigma^-_{k+1}\, H^* + R)^{-1}\, H\, \Sigma^-_{k+1}\,]$$

$$[\, (\Sigma^-_{k+1})^{-1}\, \mu^-_{k+1} + H^*\, R^{-1}\, Y_{k+1}\,]$$

$$= \mu^-_{k+1} + \Sigma^-_{k+1}\, H^*\, R^{-1}\, Y_{k+1} - \Sigma^-_{k+1}\, H^*\, (H\, \Sigma^-_{k+1}\, H^* + R)^{-1}\, H\, \mu^-_{k+1}$$

$$- \Sigma^-_{k+1}\, H^*\, (H\, \Sigma^-_{k+1}\, H^* + R)^{-1}\, H\, \Sigma^-_{k+1}\, H^*\, R^{-1}\, Y_{k+1}$$

$$= \mu^-_{k+1} + \Sigma^-_{k+1}\, H^*\, (H\, \Sigma^-_{k+1}\, H^* + R)^{-1}\, (Y_{k+1} - H\, \mu^-_{k+1})\ .$$

On reconnait les équations du filtre de Kalman.

$\square$