Université de Rennes 1
Master Recherche SISEA

Examen du cours UE S3-2
“Filtre de Kalman et modèles de Markov cachés”
Jeudi 3 février 2011, 14:00 à 16:00

Exercice

Le but de cet exercice est de montrer qu'il est possible (et avantageux) de traiter séquentiellement plutôt que simultanément les composantes d'une observation multi–dimensionnelle, pourvu que les bruits d'observation associés aux différentes composantes soient indépendants.

On se limite au cas statique suivant, dans lequel l'observation Y est donnée par la relation

$$Y = H\,X + V\ ,$$

où X et V sont deux vecteurs aléatoires gaussiens indépendants, de dimension m et d, de moyenne $\bar{X}$ et 0, et de matrice de covariance P et Q, respectivement. On suppose que la matrice Q est inversible.

(i) **Rappeler l'expression donnée dans le cours pour la moyenne conditionnelle $\widehat{X}$ et la matrice de covariance R, pour l'estimation MMSE de l'état caché X sachant l'observation Y.**

On suppose maintenant que le vecteur aléatoire V se décompose en deux composantes V_1 et V_2 indépendantes, de dimension d_1 et d_2 respectivement (avec $d_1 + d_2 = d$), et on introduit les partitions suivantes des matrices

$$H = \begin{pmatrix} H_1 \\ H_2 \end{pmatrix} \qquad \text{et} \qquad Q = \begin{pmatrix} Q_1 & 0 \\ 0 & Q_2 \end{pmatrix}\ .$$

Il en résulte que l'observation Y se décompose en deux composantes Y_1 et Y_2, de dimension d_1 et d_2 respectivement (avec $d_1 + d_2 = d$), données par les relations

$$Y_1 = H_1\,X + V_1 \qquad \text{et} \qquad Y_2 = H_2\,X + V_2\ .$$

(ii) **Rappeler l'expression donnée dans le cours pour la moyenne conditionnelle $\widehat{X}_2$, la matrice de gain de Kalman K_2 et la matrice de covariance R_2, pour l'estimation MMSE de l'état caché X sachant l'observation Y_2 seulement.**

(iii) **Donner la décomposition par blocs de la matrice $d \times d$**

$$\Xi = H\,P\,H^* + Q = \begin{pmatrix} \star & \star \\ \star & \star \end{pmatrix}$$

en blocs diagonaux de dimensions $d_1 \times d_1$ et $d_2 \times d_2$ et en blocs hors–diagonaux de dimensions appropriées $d_1 \times d_2$ et $d_2 \times d_1$ (en d'autres termes, remplacer les $\star$ par leurs expressions).

(iv) **Donner aussi la décomposition par blocs de la moyenne conditionnelle $\widehat{X}$ et de la matrice de covariance R, rappelées en réponse à la question (i).**

Pour un calcul efficace de la matrice inverse Ξ^{-1}, on rappelle le résultat d'inversion matricielle suivant : Dans la matrice–bloc

$$M = \begin{pmatrix} A & B \\ B^* & D \end{pmatrix}$$

si la matrice D est inversible, et si la matrice $\Delta = A - B\,D^{-1}\,B^*$ (appelée complément de Schur de la matrice D dans la matrice–bloc M) est inversible, alors la matrice–bloc M est inversible, et la matrice inverse est donnée par

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

(v) **Appliquer ce résultat au calcul de la matrice inverse Ξ^{-1}. Mettre en évidence dans l'expression obtenue : la matrice de gain de Kalman K_2 et la matrice de covariance R_2, rappelées en réponse à la question (ii).**

(vi) **Reporter l'expression obtenue en réponse à la question (v) dans la décomposition par blocs de la moyenne conditionnelle $\widehat{X}$ et de la matrice de covariance R, obtenues en réponse à la question (iv). Mettre en évidence dans l'expression obtenue : la moyenne conditionnelle $\widehat{X}_2$ et la matrice de covariance R_2, rappelées en réponse à la question (ii).**

(vii) **Interpréter le résultat obtenu, en terme de la moyenne conditionnelle et de la matrice de covariance pour l'estimation MMSE de l'état caché** X **dans le modèle artificiel suivant, dans lequel l'observation** Y **est donnée par la relation**

$$Y = H_1 \, X + V_1 \ ,$$

où X **et** V_1 **sont deux vecteurs aléatoires gaussiens indépendants, de dimension** m **et** d_1**, de moyenne** $\widehat{X}_2$ **et** 0**, et de matrice de covariance** R_2 **et** Q_1**, respectivement.**

(viii) **Quel est l'intérêt pratique de cette méthode séquentielle ?**

Problème

L'estimateur MMSE dans un modèle gaussien (et plus généralement le filtre de Kalman pour l'estimation récursive de l'état caché dans un système linéaire gaussien) minimise l'erreur quadratique moyenne pourvu que le modèle utilisé pour construire l'estimateur coïncide avec le modèle réel, qui a servi a générer l'observation. Que dire de l'erreur quadratique moyenne, évaluée pour le modèle réel et pour un estimateur construit avec un modèle de substitution, appelé modèle de conception, par exemple parce le modèle réel est inconnu ?

L'objectif de ce problème est de proposer un autre estimateur, minimisant un autre critère que l'erreur quadratique moyenne et n'utilisant pas le modèle réel, mais dont on puisse néanmoins contrôler l'erreur quadratique moyenne, évaluée pour le modèle réel.

Attention ! Les questions (iii), (iv) et (v) demandent des calculs sans réelle difficulté mais assez longs, et il est conseillé de n'aborder ces questions que si le temps le permet.

Sous le modèle de conception, utilisé dans un but pratique et par ignorance du modèle réel, l'observation Y est donnée par la relation

$$Y = H\,X + V\ ,$$

où X et V sont deux vecteurs aléatoires gaussiens indépendants, de dimension m et d, de moyenne $\bar{X}$ et 0, et de matrice de covariance P et Q, respectivement. On suppose pour simplifier que les deux matrices P et Q sont *inversibles*. Par construction, le vecteur $\bar{X}$ et les matrices H, P et Q sont choisies par le concepteur, et sont donc connues.

Le critère proposé pour mesurer la qualité d'un estimateur quelconque $\psi = \psi(Y)$ est de la forme exponentielle suivante

$$\mathbb{E}[\,\exp\{\mu\,|X - \psi(Y)|^2\}\,]\ ,$$

où $\mu > 0$ est un paramètre suffisamment petit pour que le critère soit bien défini. On notera $\widehat{X}^\mu = \widehat{X}^\mu(Y)$ l'estimateur optimal au sens de ce critère, appelé aussi estimateur *risk–sensitive*, c'est–à–dire tel que

$$\mathbb{E}[\,\exp\{\mu\,|X - \widehat{X}^\mu(Y)|^2\}\,] \leq \mathbb{E}[\,\exp\{\mu\,|X - \psi(Y)|^2\}\,]\ ,$$

pour tout estimateur ψ. Cette propriété d'optimalité propose seulement une définition implicite de l'estimateur $\widehat{X}^\mu$, et la première étape consiste donc à en proposer une expression explicite.

(i) **Donner l'expression de la densité jointe $p_{X,Y}(x,y)$ du vecteur aléatoire (X,Y) sous le modèle de conception.**

(ii) **Pour un estimateur ψ quelconque, écrire le critère comme une intégrale double par rapport à la densité jointe $p_{X,Y}(x,y)$.**

(iii) [calculs faciles, mais longs] **En intégrant d'abord par rapport à la variable $x \in \mathbb{R}^m$, écrire le critère comme une intégrale par rapport à la variable $y \in \mathbb{R}^d$ seulement.**

(iv) [calculs faciles, mais longs] **En déduire une expression explicite pour l'estimateur *risk–sensitive* $\widehat{X}^\mu$.**

(v) [calculs faciles, mais longs] **Donner l'expression de la valeur minimale du critère**

$$c^\mu = \mathbb{E}[\exp\{\mu\,|X - \widehat{X}^\mu(Y)|^2\}] = \min_\psi \mathbb{E}[\exp\{\mu\,|X - \psi(Y)|^2\}] \ ,$$

obtenue pour l'estimateur *risk–sensitive* $\widehat{X}^\mu$.

(vi) **À titre de comparaison, rappeler l'expression donnée dans le cours pour l'estimateur MMSE $\widehat{X}$, calculé sous le modèle de conception.**

En réalité, c'est–à–dire sous le modèle réel, l'observation Y est donnée par la relation

$$Y = H_0\, X + V \ ,$$

où X et V sont deux vecteurs aléatoires gaussiens indépendants, de dimension m et d, de moyenne $\bar{X}_0$ et 0, et de matrice de covariance P_0 et Q_0, respectivement. Ici aussi, on suppose que les deux matrices P_0 et Q_0 sont *inversibles*. En revanche, le vecteur $\bar{X}_0$ et les matrices H_0, P_0 et Q_0 sont inconnues (sinon, on utiliserait directement ce modèle, au lieu du modèle de conception).

L'objectif est de contrôler l'erreur quadratique moyenne, évaluée sous le modèle réel, de l'estimateur *risk–sensitive*, c'est–à–dire de contrôler

$$\mathbb{E}_0|X - \widehat{X}^\mu|^2 \ .$$

(vii) **À titre de comparaison, rappeler l'expression donnée dans le cours pour l'estimateur MMSE $\widehat{X}_0$, calculé sous le modèle réel.**

(viii) **En déduire l'expression exacte de l'erreur quadratique moyenne, évaluée sous le modèle réel, des deux estimateurs MMSE $\widehat{X}$ et $\widehat{X}_0$, calculés sous le modèle de conception et sous le modèle réel, respectivement, c'est–à–dire**

$$\mathbb{E}_0|X - \widehat{X}|^2 \qquad \textbf{et} \qquad \mathbb{E}_0|X - \widehat{X}_0|^2 \ .$$

On pourrait faire de même pour l'erreur quadratique moyenne, évaluée sous le modèle réel, de l'estimateur *risk–sensitive*, c'est–à–dire pour

$$\mathbb{E}_0|X - \widehat{X}^\mu|^2 \ ,$$

mais l'intérêt de ces expressions est néanmoins limité, puisque le modèle réel est inconnu. À défaut d'expression exacte, on voudrait obtenir une borne ne nécessitant pas la connaissance du modèle réel, donc *calculable*.

(ix) **Donner l'expression de la densité jointe $p^0_{X,Y}(x,y)$ du vecteur aléatoire (X,Y) sous le modèle réel.**

(x) **Pour un estimateur ψ quelconque, écrire le critère comme une intégrale double par rapport à la densité jointe $p^0_{X,Y}(x,y)$, en faisant apparaître le rapport des deux densités $p_{X,Y}(x,y)$ et $p^0_{X,Y}(x,y)$.**

On rappelle l'inégalité de Jensen (inégalité de convexité)

$$\mathbb{E}_0[\exp\{Z\}] \geq \exp\{\mathbb{E}_0[Z]\} \qquad \text{soit} \qquad \mathbb{E}_0[Z] \leq \log \mathbb{E}_0[\exp\{Z\}] \ .$$

(xi) **En déduire que**

$$\mu\, \mathbb{E}_0|X - \widehat{X}^\mu|^2 \leq \log \mathbb{E}[\exp\{\mu\, |X - \widehat{X}^\mu|^2\}\,] + I(M_0\|M) \ ,$$

où

$$I(M_0\|M) = \int\int p^0_{X,Y}(x,y)\ \log \frac{p^0_{X,Y}(x,y)}{p_{X,Y}(x,y)}\ dx\, dy$$

s'interprète comme la distance de Kullback–Leibler entre le modèle de conception et le modèle réel. Quel est l'intérêt d'une telle majoration ?

(xii) **Donner l'expression de la distance de Kullback–Leibler $I(M_0\|M)$ en fonction des vecteurs $\bar{X}$ et $\bar{X}_0$ et des matrices H, H_0, P, P_0, Q et Q_0.**

Petit formulaire

inversion matricielle et gain de Kalman

$(P^{-1} + H^*\, Q^{-1}\, H)^{-1} = P - P\, H^*\, (H\, P\, H^* + Q)^{-1}\, H\, P$
$(H\, P\, H^* + Q)^{-1} = Q^{-1} - Q^{-1}\, H\, (P^{-1} + H^*\, Q^{-1}\, H)^{-1}\, H^*\, Q^{-1}$
$K = P\, H^*\, (H\, P\, H^* + Q)^{-1} = (P^{-1} + H^*\, Q^{-1}\, H)^{-1}\, H^*\, Q^{-1}$