cours SOD333

Filtrage Bayésien et Approximation Particulaire

Introduction — Filtre de Kalman

François Le Gland
INRIA Rennes et IRMAR

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

17 septembre 2021

François Le Gland

- téléphone : 02 99 84 73 62 / 06 95 02 13 16
- e–mail : `francois.le_gland@inria.fr`

formation

- ingénieur Ecole Centrale Paris (1978)
- DEA de Probabilités à Paris 6 (1979)
- thèse en Mathématiques Appliquées à Paris Dauphine (1981)

carrière professionnelle : chercheur à l'INRIA (directeur de recherche depuis 1991)

- à Rocquencourt jusqu'en 1983
- à Sophia Antipolis de 1983 à 1993
- à Rennes depuis 1993

membre de l'IRMAR depuis 2012

membre de l'équipe

SIMSMART : Simulating Stochastic Models with Particles

team.inria.fr/simsmart/

thèmes

- filtrage particulaire, et applications en
  - localisation, navigation et poursuite
  - assimilation de données séquentielle
- inférence statistique des modèles de Markov cachés
- simulation d'évènements rares, et extensions en
  - simulation moléculaire
  - optimisation globale
- analyse mathématique des méthodes particulaires

contrats industriels

- avec France Télécom R&D, sur la localisation de terminaux mobiles
- avec Thalès, sur la navigation par corrélation de terrain
- avec DGA / Techniques Navales, sur l'optimisation du positionnement et de l'activation de capteurs
- avec Naval Group (ex-DCNS), sur le pistage TBD multi–cible

projets ANR

- FIL, sur la fusion de données pour la localisation
- PREVASSEMBLE, sur les méthodes d'ensemble pour l'assimilation de données et la prévision

projets européens

- HYBRIDGE et iFLY, sur les méthodes de Monte Carlo conditionnelles pour l'évaluation de risque en trafic aérien

collaboration avec ONERA

organisation pratique du cours : deux types de séances

- cours
- cours + séance de TP
  - par binôme
  - rapport écrit + code source MATLAB ou Python
  - en cas de difficulté, e–mail à `francois.le_gland@inria.fr`

support de cours

- polycopié (aspects théoriques)
- planches présentées en cours
- énoncés des TP

ressources : références bibliographiques, archives (sujets d'examen), etc.

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

contrôle des connaissances

- TP (1er et 15 octobre)
- examen final (22 octobre)

TP proposés cette année

- recalage altimétrique de navigation inertielle
- suivi visuel par histogramme de couleur

autres TP proposés les années précédentes

- assimilation de données séquentielle pour le modèle de Lorenz
- poursuite d'une cible furtive (track–before–detect)
- navigation en environnement intérieur
- valorisation d'options européennes dans le modèle de Black–Scholes
- évaluation du risque de collision

Introduction au filtrage
nécessité / utilité d'un modèle a priori
exploitation du modèle a priori

Estimation bayésienne

Systèmes linéaires gaussiens

Extensions

Cadre général

il s'agit d'estimer, si possible de manière *récursive*, l'état X_n d'un système (par exemple, position et vitesse d'un mobile) au vu d'observations bruitées $Y_{0:n} = (Y_0 \cdots Y_n)$ reliées à l'état par une relation du genre

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

nombreuses applications en *localisation, navigation et poursuite* de mobiles

- poursuite de cible (avion, missile, drône, bâtiment de surface, sous–marin)
- suivi d'objets dans des séquences vidéo
- navigation en environnement intérieur (robot mobile, piéton)
- navigation inertielle, recalage avec un modèle numérique de terrain
- navigation de terminaux mobiles, recalage avec une carte de couverture

dans toutes ces applications, il s'agit de déterminer position et vitesse d'un mobile, en utilisant comme observations des mesures *partielles* issues de capteurs

(i) mesures *directes*

- distance par rapport à une ou plusieurs stations
- angle par rapport à une direction de référence
- proximité par rapport à une ou plusieurs stations
- altitude par rapport à un niveau de référence ou par rapport au relief

(ii) mesures *indirectes*, combinées avec une base de mesures *géo–référencées*, disponibles par exemple sous la forme d'une carte numérique

- altitude du relief + modèle numérique de terrain
- atténuation de la puissance du signal reçu + carte de couverture

contrats industriels avec

France Télécom R&D, Thalès Communications (PEA NCT), Thalès Alénia Space (projet ANR FIL), CEA LETI, Naval Group Research

exemples (en fin de séance / à la prochaine séance)

Nécessité d'un modèle a priori

sans information supplémentaire, observer

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

pose un nouveau problème d'estimation à chaque nouvel instant (pas d'accumulation des observations) + dans chacun de ces problèmes d'estimation séparément, la dimension m de l'état caché est (souvent) plus grande que la dimension d de l'observation

estimer l'état caché X_k au vu de la seule observation Y_k et même sans bruit d'observation V_k, est un problème mal–posé

- la relation possède plus d'inconnues que d'équations
- même dans le cas favorable où $m = d$, la relation non–linéaire peut posséder plusieurs solutions distinctes
- la suite reconstituée peut être globalement peu pertinente (en tant que suite), même si séparément chacune des estimations marginales est pertinente (incohérence temporelle)

approche variationnelle pour lever l'*indétermination* ou l'*incohérence temporelle* : introduire des informations supplémentaires sur la suite cachée, sous la forme de fonctions de coût portant sur l'état initial et sur les transitions entre deux états successifs

par exemple, le critère à minimiser par rapport à la suite $x_{0:n} = (x_0, x_1, \cdots, x_n)$

$$J(x_{0:n}) = c_0(x_0) + \sum_{k=1}^{n} c_k(x_{k-1}, x_k) + \sum_{k=0}^{n} d_k(x_k)$$

combine des fonctions de coût qui représentent

- une information *a priori* sur la suite recherchée
- un terme d'*attache aux données*, par exemple de la forme

$$d_k(x) = \tfrac{1}{2} \, |Y_k - h_k(x)|^2$$

  avec l'interprétation que la suite recherchée doit vérifier à chaque instant l'équation d'observation en un sens approché

ou qui peuvent juste représenter une contrainte (ou une propriété) que la suite recherchée devrait vérifier (ou posséder)

typiquement, l'information *a priori* est représentée par

$$c_0(x) = \tfrac{1}{2}\,|x-\mu|^2 \qquad \text{et} \qquad c_k(x,x') = \tfrac{1}{2}\,|x'-f_k(x)|^2$$

avec l'interprétation que

- l'état initial x_0 recherché doit être proche de μ
- et que la transition (x_{k-1}, x_k) recherchée doit vérifier l'équation $x_k = f_k(x_{k-1})$ en un sens approché

cette régularisation (du point de vue de l'*optimisation*) peut s'interpréter comme l'ajout d'une information a priori (du point de vue de l'*estimation statistique*)

plus généralement, si l'information *a priori* est représentée par

$$c_0(x) = -\log p_0(x) \qquad \text{et} \qquad c_k(x, x') = -\log p_k(x' \mid x)$$

alors minimiser le critère

$$J(x_{0:n}) = -\log p_0(x_0) - \sum_{k=1}^{n} \log p_k(x_k \mid x_{k-1}) + \sum_{k=0}^{n} d_k(x_k)$$

revient à maximiser (estimateur MAP (*maximum a posteriori*))

$$\exp\{-J(x_{0:n})\} = \underbrace{p_0(x_0) \prod_{k=1}^{n} p_k(x_k \mid x_{k-1})}_{p_{0:n}(x_{0:n})} \exp\{-\sum_{k=0}^{n} d_k(x_k)\}$$

où $p_{0:n}(x_{0:n})$ représente la densité de probabilité conjointe des états successifs $(X_0, X_1, \cdots, X_n)$ de la chaîne de Markov caractérisée par

- la densité de probabilité initiale $p_0(x_0)$
- et les densités de probabilité de transition $p_k(x_k \mid x_{k-1})$

approche statistique avec un point de vue pragmatique : au lieu de rechercher un maximum de la densité a posteriori, on choisit d'échantillonner cette densité, ce qui permet par exemple de calculer des espérances (ou des intégrales) du type

$$\int_E \cdots \int_E f(x_{0:n}) \ \exp\{-J(x_{0:n})\} \ dx_{0:n}$$

$$= \int_E \cdots \int_E f(x_{0:n}) \ \exp\{-\sum_{k=0}^{n} d_k(x_k)\} \ p_{0:n}(x_{0:n}) \ dx_{0:n}$$

$$= \mathbb{E}[f(X_{0:n}) \ \exp\{-\sum_{k=0}^{n} d_k(X_k)\}]$$

on verra comment résoudre ce problème de manière approchée, en simulant des échantillons de variables aléatoires distribuées (approximativement) selon la distribution de Gibbs–Boltzmann trajectorielle définie ci–dessus

lien entre les deux approches : méthode asymptotique de Laplace

résumé : nécessité / utilité d'un *modèle a priori*

- pour compléter l'information apportée par les observations
- pour faire le lien entre des observations reçues à des instants différents (compte tenu que le mobile s'est en principe déplacé entre ces instants)

ce modèle peut être *rustique* ou *grossier*, et il est souvent bruité (on accepte l'idée que le modèle est nécessairement faux, et on essaye de quantifier, de manière statistique, l'erreur de modélisation)

il suffit en pratique de savoir *simuler* à chaque pas de temps des v.a. indépendantes selon le modèle proposé : comme ce modèle est bruité (aléatoire) des simulations différentes proposeront des états possibles différents

effet désiré : permettre l'exploration rationnelle de l'espace d'état pour proposer des solutions réalistes, qui seront ensuite confirmées, renforcées ou au contraire abandonnées : il suffira pour cela de savoir *évaluer* la fonction de vraisemblance en chacune des solutions proposées

Modèles

modèle (du plus simple au plus général)

modèle d'état (espace d'état continu $\mathbb{R}^m$)

- linéaire, bruits gaussiens
- non–linéaire, bruits gaussiens
- modèle d'état général : non–linéaire, bruits non–gaussiens

modèle de Markov caché (HMM), et extensions

- modèle de Markov caché (état caché markovien, observations indépendantes / markoviennes conditionnellement aux états cachés)
- chaîne de Markov partiellement observée (états cachés et observations conjointement markoviens)

espace d'état général

fini, dénombrable, continu (espace euclidien), variété, hybride continu / discret, contraint, graphique (collection de nœuds et d'arêtes)

Exploitation du modèle a priori

approche bayésienne, fusion d'information : loi *a posteriori* obtenue en combinant

- loi *a priori* (modèle d'état)
- *vraisemblance* (adéquation entre mesures et états)

mise en œuvre récursive grâce à la propriété de Markov

principe général : à l'aide de la formule de Bayes, la loi conditionnelle de $X_{0:n}$ sachant $Y_{0:n}$ s'exprime à partir de

- la loi de $X_{0:n}$
- la loi conditionnelle de $Y_{0:n}$ sachant $X_{0:n}$, souvent facile à évaluer par exemple dans le modèle additif

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

en principe, la loi conditionnelle de X_n sachant $Y_{0:n}$ s'obtient facilement par marginalisation

en pratique, cette loi conditionnelle n'a d'expression *explicite* que dans certains cas particuliers

- chaîne de Markov à espace d'état fini
- modèle linéaire gaussien

d'où l'intérêt pour les méthodes de simulation de type Monte Carlo dans le cas général : le filtrage particulaire offre une manière numériquement efficace de mettre en œuvre les méthodes bayésiennes

Introduction au filtrage

Estimation bayésienne
- estimation bayésienne
- borne de Cramér–Rao a posteriori
- cadre gaussien

Systèmes linéaires gaussiens

Extensions

Estimation bayésienne

soit $Z = (X, Y)$ à valeurs dans $E \times F$ (par exemple $E = \mathbb{R}^m$, $F = \mathbb{R}^d$) de loi jointe connue (pas nécessairement donnée par une densité $p(x, y)$)

objectif : exploiter au mieux l'observation de Y pour améliorer la connaissance de la composante cachée X

un estimateur ψ de la statistique $\phi = \phi(X)$ à valeurs dans $\mathbb{R}^p$ est une application définie sur F, à valeurs dans $\mathbb{R}^p$

mesure de l'écart entre $\psi(Y)$ et $\phi(X)$: erreur quadratique moyenne

$$\mathbb{E}|\psi(Y) - \phi(X)|^2$$

trace de la matrice de corrélation de l'erreur d'estimation

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

symétrique semi–définie positive, de dimension $p \times p$

estimateur MMSE (*minimum mean square error*) : estimateur $\widehat{\phi}$ tel que

$$\mathbb{E}[\,(\widehat{\phi}(Y)-\phi(X))\,(\widehat{\phi}(Y)-\phi(X))^*\,] \leq \mathbb{E}[\,(\psi(Y)-\phi(X))\,(\psi(Y)-\phi(X))^*\,]$$

pour tout autre estimateur ψ, au sens des matrices symétriques
a fortiori

$$\mathbb{E}|\widehat{\phi}(Y)-\phi(X)|^2 \leq \mathbb{E}|\psi(Y)-\phi(X)|^2$$

pour tout autre estimateur ψ

Proposition l'estimateur MMSE de la statistique $\phi=\phi(X)$ au vu de l'observation Y est la moyenne conditionnelle de $\phi(X)$ sachant Y

$$\widehat{\phi}(Y)=\mathbb{E}[\phi(X)\mid Y]=\int_E \phi(x)\,\mathbb{P}[X\in dx\mid Y]$$

Preuve pour un estimateur ψ quelconque, la décomposition

$$\psi(Y) - \phi(X) = \psi(Y) - \widehat{\phi}(Y) + \widehat{\phi}(Y) - \phi(X)$$

entraîne

$$\begin{aligned}
&\mathbb{E}[\,(\psi(Y) - \phi(X))\,(\psi(Y) - \phi(X))^*\,] \\
&\quad = \mathbb{E}[\,(\psi(Y) - \widehat{\phi}(Y))\,(\psi(Y) - \widehat{\phi}(Y))^*\,] \\
&\qquad + \mathbb{E}[\,(\widehat{\phi}(Y) - \phi(X))\,(\widehat{\phi}(Y) - \phi(X))^*\,] \\
&\qquad + \mathbb{E}[\,(\psi(Y) - \widehat{\phi}(Y))\,(\widehat{\phi}(Y) - \phi(X))^*\,] + (\cdots)^*
\end{aligned}$$

si les deux produits croisés sont nuls, alors

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

au sens des matrices symétriques, avec égalité pour $\psi(Y) = \widehat{\phi}(Y)$

il reste à vérifier que

$$\mathbb{E}[\,(\widehat{\phi}(Y)-\phi(X))\,(\psi(Y)-\widehat{\phi}(Y))^*\,]$$

$$=\mathbb{E}[\,\mathbb{E}[\widehat{\phi}(Y)-\phi(X)\mid Y]\,(\psi(Y)-\widehat{\phi}(Y))^*\,]$$

$$=\mathbb{E}[\,(\widehat{\phi}(Y)-\mathbb{E}[\phi(X)\mid Y])\,(\psi(Y)-\widehat{\phi}(Y))^*\,]=0$$

par définition de

$$\widehat{\phi}(Y)=\mathbb{E}[\phi(X)\mid Y] \qquad \square$$

pros : solution universelle (la loi conditionnelle de X sachant Y permet de construire l'estimateur MMSE de n'importe quelle statistique $\phi=\phi(X)$)
cons : solution infini–dimensionnelle (distribution ou densité de probabilité)

Borne de Cramér–Rao a posteriori

l'erreur quadratique moyenne associée à l'estimateur bayésien fournit une borne inférieure, atteinte (par l'estimateur bayésien) mais souvent difficile à évaluer

la borne de Cramér–Rao a posteriori (différente de la borne paramétrique) est plus facilement calculable mais pas nécessairement atteinte

hypothèse : la loi jointe de $Z = (X, Y)$ possède une densité (assez régulière) sur $\mathbb{R}^m \times F$

$$\mathbb{P}[X \in dx, Y \in dy] = p(x, y)\, dx\, \lambda(dy)$$

obtenue souvent grâce à la factorisation (formule de Bayes)

$$p(x, y) = p(y \mid x)\, p(x) = p(x \mid y)\, p(y)$$

et cette densité vérifie

$$\int_{\mathbb{R}^m} \int_F \frac{\partial^2}{\partial x^2} p(x, y)\ \lambda(dy)\, dx \ = \ \int_{\mathbb{R}^m} p''(x)\, dx = 0$$

Remarque en utilisant l'identité

$$\frac{\partial^2}{\partial x^2}\log p(x,y) = \frac{1}{p(x,y)}\frac{\partial^2}{\partial x^2}p(x,y) - (\frac{\partial}{\partial x}\log p(x,y))^* \ \frac{\partial}{\partial x}\log p(x,y)$$

entre matrices de dimension $m \times m$, et en intégrant membre à membre par rapport à la loi jointe $p(x,y)\,dx\,\lambda(dy)$ du couple (X,Y), il vient

$$\mathbb{E}[\,(\frac{\partial}{\partial x}\log p(X,Y))^* \ \frac{\partial}{\partial x}\log p(X,Y)\,]$$

$$= \int_{\mathbb{R}^m}\int_F \frac{\partial^2}{\partial x^2}p(x,y)\ \lambda(dy)\,dx - \mathbb{E}[\,\frac{\partial^2}{\partial x^2}\log p(X,Y)\,]$$

par hypothèse on a donc

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

et ces matrices sont symétriques et définies positives

on définit le biais de l'estimateur ψ comme

$$b(\psi, x) = \mathbb{E}[\psi(Y) \mid X = x] - \phi(x) = \int_F (\psi(y) - \phi(x))\, p(y \mid x)\, \lambda(dy)$$

Théorème si la matrice d'information de Fisher

$$J = -\mathbb{E}[\frac{\partial^2}{\partial x^2} \log p(X, Y)] \qquad \text{matrice de dimension } m \times m$$

est inversible, alors

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

avec

$$M = \mathbb{E}[\phi'(X)] \qquad \text{matrice de dimension } p \times m$$

pour tout estimateur ψ tel que

$$\int_{\mathbb{R}^m} (b(\psi, x)\, p(x))'\, dx = 0$$

Preuve par définition

$$b(\psi, x)\, p(x) = \int_F (\psi(y) - \phi(x))\, p(x, y)\ \lambda(dy)$$

et la matrice jacobienne (de dimension $p \times m$) associée vérifie

$$(b(\psi, x)\, p(x))'$$

$$= -\phi'(x) \int_F p(x, y)\ \lambda(dy) + \int_F (\psi(y) - \phi(x))\, \frac{\partial}{\partial x} p(x, y)\ \lambda(dy)$$

$$= -\phi'(x)\, p(x) + \int_F (\psi(y) - \phi(x))\, \frac{\partial}{\partial x} \log p(x, y)\ p(x, y)\ \lambda(dy)$$

en intégrant par rapport à la variable $x \in \mathbb{R}^m$ et avec l'hypothèse, il vient

$$-\int_{\mathbb{R}^m} \phi'(x)\, p(x)\, dx + \int_{\mathbb{R}^m} \int_F (\psi(y) - \phi(x))\, \frac{\partial}{\partial x} \log p(x, y)\ p(x, y)\ \lambda(dy)\, dx$$

$$= -\mathbb{E}[\phi'(X)] + \mathbb{E}[\,(\psi(Y) - \phi(X))\, \frac{\partial}{\partial x} \log p(X, Y)\,] = 0$$

c'est–à–dire que

$$\mathbb{E}[\,(\psi(Y)-\phi(X))\,\frac{\partial}{\partial x}\log p(X,Y)\,]=M$$

où la matrice de sensibilité M ne dépend pas de l'estimateur ψ
d'autre part, il résulte de la remarque faite plus haut que

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

on introduit ainsi le vecteur aléatoire (de dimension $p+m$)

$$\begin{pmatrix}\psi(Y)-\phi(X)\\(\frac{\partial}{\partial x}\log p(X,Y))^*\end{pmatrix}$$

et sa matrice de corrélation $\begin{pmatrix}C & M\\ M^* & J\end{pmatrix}$

avec

$$C=\mathbb{E}[\,(\psi(Y)-\phi(X))\,(\psi(Y)-\phi(X))^*\,]$$

compte tenu que cette matrice est symétrique semi–définie positive, on vérifie que pour tout vecteur $u \in \mathbb{R}^p$

$$
\begin{aligned}
0 \leq & \begin{pmatrix} u^* & -(J^{-1} M^* u)^* \end{pmatrix} \begin{pmatrix} C & M \\ M^* & J \end{pmatrix} \begin{pmatrix} u \\ -J^{-1} M^* u \end{pmatrix} \\
= & \begin{pmatrix} u^* & -(J^{-1} M^* u)^* \end{pmatrix} \begin{pmatrix} (C - M J^{-1} M^*) u \\ 0 \end{pmatrix} \\
= & \; u^* (C - M J^{-1} M^*) u
\end{aligned}
$$

c'est–à–dire que le complément de Schur $\Delta = C - M J^{-1} M^*$ est également une matrice symétrique semi–définie positive, d'où

$$C \geq M J^{-1} M^* \qquad \square$$

Utilisation pratique

pour évaluer *à l'avance* la performance d'un estimateur donné $\psi(Y)$ de la variable cachée X au vu de l'observation Y, on réalise N simulations indépendantes

$$(X^j, Y^j) \qquad \text{pour tout } j = 1 \cdots N$$

pour avoir une idée de la performance, on évalue empiriquement la matrice de corrélation des erreurs d'estimation

$$\mathbb{E}[\,(\psi(Y) - X)\,(\psi(Y) - X)^*\,] \approx \frac{1}{N} \sum_{j=1}^{N} (\psi(Y^j) - X^j)\,(\psi(Y^j) - X^j)^*$$

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

pour calculer la matrice d'information de Fisher J intervenant dans la borne, on évalue empiriquement la matrice

$$-J = \mathbb{E}[\frac{\partial^2}{\partial x^2} \log p(X, Y)] \approx \frac{1}{N} \sum_{j=1}^{N} \frac{\partial^2}{\partial x^2} \log p(X^j, Y^j)$$

borne calculable à l'avance, pas de connaissance nécessaire de la "vraie" valeur (qui n'existe d'ailleurs pas dans le cadre bayésien)
moyenne sur toutes les réalisations possibles

de l'observation et de la composante cachée conjointement

par contraste, dans la borne de Cramér–Rao paramétrique (vs. borne de Cramér–Rao a posteriori) il n'y a pas de loi a priori sur la composante cachée θ

Théorème si la matrice d'information de Fisher

$$J(\theta) = -\mathbb{E}[\frac{\partial^2}{\partial\theta^2} \log p(Y \mid \theta) \mid \theta] \qquad \text{matrice de dimension } m \times m$$

est inversible, alors

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

pour tout estimateur ψ sans biais de θ, c–à–d tel que $\mathbb{E}[\psi(Y) \mid \theta] = \phi(\theta)$

décomposition "loi jointe" = "loi conditionnelle" $\times$ "loi marginale"

$$p(x, y) = p(x \mid y)\, p(y) = p(y \mid x)\, p(x)$$

d'où les expressions équivalentes suivantes

$$\begin{aligned} J &= -\mathbb{E}[\frac{\partial^2}{\partial x^2} \log p(X, Y)] \\ &= -\mathbb{E}[\frac{\partial^2}{\partial x^2} \log p(X \mid Y)] \\ &= -\mathbb{E}[\frac{\partial^2}{\partial x^2} \log p(Y \mid X)] - \mathbb{E}[\frac{\partial^2}{\partial x^2} \log p(X)] \\ &= \mathbb{E}[\underbrace{-\mathbb{E}[\frac{\partial^2}{\partial x^2} \log p(Y \mid X) \mid X]}_{J(X)}] - \mathbb{E}[\frac{\partial^2}{\partial x^2} \log p(X)] \end{aligned}$$

pour la matrice d'information de Fisher

Exemple soit X et V deux vecteurs aléatoires gaussiens indépendants de moyenne $\bar{X}$ et 0, de matrice de covariance Q_X et Q_V, et on pose $Y = h(X) + V$
si Q_X et Q_V sont inversibles, alors

$$p(y \mid x) \propto \exp\{-\tfrac{1}{2}\,(y - h(x))^*\, Q_V^{-1}\,(y - h(x))\}$$

$$p(x) \propto \exp\{-\tfrac{1}{2}\,(x - \bar{X})^*\, Q_X^{-1}\,(x - \bar{X})\}$$

d'où

$$-\log p(x,y) = \tfrac{1}{2}\,(y-h(x))^*\, Q_V^{-1}\,(y-h(x)) + \tfrac{1}{2}\,(x-\bar{X})^*\, Q_X^{-1}\,(x-\bar{X}) + \text{cste}$$

$$-\frac{\partial^2}{\partial x^2} \log p(x,y) = (h'(x))^*\, Q_V^{-1}\, h'(x) - (y - h(x))^*\, Q_V^{-1}\, h''(x) + Q_X^{-1}$$

et la matrice d'information de Fisher

$$J = -\mathbb{E}[\frac{\partial^2}{\partial x^2} \log p(X,Y)] = \mathbb{E}[(h'(X))^*\, Q_V^{-1}\, h'(X)] - \underbrace{\mathbb{E}[V^*\, Q_V^{-1}\, h''(X)]}_{0} + Q_X^{-1}$$

dans le cas particulier où $h(x) = H\,x$

$$J = H^*\,Q_V^{-1}\,H + Q_X^{-1} \qquad \text{et} \qquad J^{-1} = Q_X - Q_X\,H^*\,(H\,Q_X\,H^* + Q_V)^{-1}\,H\,Q_X$$

repose sur le lemme d'inversion matricielle suivant

Lemme soit Q et R deux matrices symétriques définies positives, de dimension m et d respectivement, et soit H une matrice $d \times m$, alors

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

Remarque cette formule d'inversion permet de remplacer l'inversion de la matrice $(H^*\,R^{-1}\,H + Q^{-1})$ de dimension m, par l'inversion de la matrice $(H\,Q\,H^* + R)$ de dimension d, avec en général $d \leq m$

Preuve on remarque d'abord que

$$H\,Q\,H^* + R \geq R \qquad \text{et} \qquad H^*\,R^{-1}\,H + Q^{-1} \geq Q^{-1}$$

au sens des matrices symétriques, ce qui prouve que les matrices

$$(H\,Q\,H^* + R) \qquad \text{et} \qquad (H^*\,R^{-1}\,H + Q^{-1})$$

sont inversibles

on vérifie alors que

$$
\begin{aligned}
&[\,Q - Q\,H^*\,(H\,Q\,H^* + R)^{-1}\,H\,Q\,]\,[H^*\,R^{-1}\,H + Q^{-1}] \\
&\quad = Q\,H^*\,R^{-1}\,H + I \\
&\qquad - Q\,H^*\,(H\,Q\,H^* + R)^{-1}\,(H\,Q\,H^* + R - R)\,R^{-1}\,H \\
&\qquad - Q\,H^*\,(H\,Q\,H^* + R)^{-1}\,H \\
&\quad = Q\,H^*\,R^{-1}\,H + I \\
&\qquad - Q\,H^*\,R^{-1}\,H + Q\,H^*\,(H\,Q\,H^* + R)^{-1}\,H \\
&\qquad - Q\,H^*\,(H\,Q\,H^* + R)^{-1}\,H \\
&\quad = I \qquad \square
\end{aligned}
$$

Cadre gaussien

estimateur MMSE dans le cas gaussien
soit $Z = (X, Y)$ vecteur aléatoire gaussien à valeurs dans $\mathbb{R}^m \times \mathbb{R}^d$ de moyenne $\bar{Z} = (\bar{X}, \bar{Y})$, de matrice de covariance $Q_Z = \begin{pmatrix} Q_X & Q_{XY} \\ Q_{YX} & Q_Y \end{pmatrix}$

Proposition la loi conditionnelle de X sachant Y est gaussienne, de moyenne $\widehat{X}(Y)$ et de matrice de covariance R (ne dépendant pas de Y), et si la matrice de covariance Q_Y est *inversible*, alors

$$\widehat{X}(Y) = \bar{X} + Q_{XY}\, Q_Y^{-1}\, (Y - \bar{Y})$$

et

$$0 \leq R = Q_X - Q_{XY}\, Q_Y^{-1}\, Q_{YX} \leq Q_X$$

complément de Schur de la matrice Q_Y dans la matrice–bloc Q_Z

Remarque on vérifie que

$$0 \leq R \leq Q_X$$

au sens des matrices symétriques, c'est–à–dire que l'utilisation de l'information supplémentaire $(Y = y)$, permet de réduire l'incertitude sur le vecteur aléatoire X
la majoration $R \leq Q_X$ est évidente, et la minoration $R \geq 0$ résulte de l'identité

$$\begin{aligned} 0 \leq & \begin{pmatrix} u^* & -(Q_Y^{-1} Q_{YX} u)^* \end{pmatrix} \begin{pmatrix} Q_X & Q_{XY} \\ Q_{YX} & Q_Y \end{pmatrix} \begin{pmatrix} u \\ -Q_Y^{-1} Q_{YX} u \end{pmatrix} \\ = & \begin{pmatrix} u^* & -(Q_Y^{-1} Q_{YX} u)^* \end{pmatrix} \begin{pmatrix} (Q_X - Q_{XY} Q_Y^{-1} Q_{YX}) u \\ 0 \end{pmatrix} \\ = & \; u^* R u \end{aligned}$$

pour tout vecteur $u \in \mathbb{R}^m$, ce qui permet de conclure que $R \geq 0$

Remarque la matrice R ne dépend pas de y, et peut donc être calculée avant même de disposer de la valeur y prise par l'observation Y

Remarque soit $\widehat{X} = \widehat{X}(Y)$ l'estimateur MMSE de X sachant Y compte tenu que

$$\widehat{X} = \bar{X} + Q_{XY}\, Q_Y^{-1}\,(Y - \bar{Y})$$

dépend de façon affine du vecteur aléatoire Y, on en déduit que $(X, \widehat{X}, Y)$ est un vecteur aléatoire gaussien, comme transformation affine du vecteur aléatoire gaussien $Z = (X, Y)$

Preuve de la Proposition on pose $\Xi = X - Q_{XY}\, Q_Y^{-1}\, Y$, et on vérifie que le vecteur aléatoire (Ξ, Y) est gaussien, comme transformation affine du vecteur aléatoire gaussien $Z = (X, Y)$
on vérifie que

$$\bar{\Xi} = \mathbb{E}[\Xi] = \bar{X} - Q_{XY}\, Q_Y^{-1}\, \bar{Y}$$

et par différence

$$\Xi - \bar{\Xi} = (X - \bar{X}) - Q_{XY}\, Q_Y^{-1}\, (Y - \bar{Y})$$

on en déduit facilement la matrice de covariance

$$
\begin{aligned}
Q_{\Xi} &= \mathbb{E}[(\Xi - \bar{\Xi})\,(\Xi - \bar{\Xi})^*] \\
&= \mathbb{E}[((X - \bar{X}) - Q_{XY}\,Q_Y^{-1}\,(Y - \bar{Y}))\,((X - \bar{X}) - Q_{XY}\,Q_Y^{-1}\,(Y - \bar{Y}))^*] \\
&= \mathbb{E}[(X - \bar{X})\,(X - \bar{X})^*] - \mathbb{E}[(X - \bar{X})\,(Y - \bar{Y})^*]\,Q_Y^{-1}\,Q_{YX} \\
&\qquad - Q_{XY}\,Q_Y^{-1}\,\mathbb{E}[(Y - \bar{Y})\,(X - \bar{X})^*] \\
&\qquad + Q_{XY}\,Q_Y^{-1}\,\mathbb{E}[(Y - \bar{Y})\,(Y - \bar{Y})^*]\,Q_Y^{-1}\,Q_{YX} \\
&= Q_X - Q_{XY}\,Q_Y^{-1}\,Q_{YX} = R
\end{aligned}
$$

et la matrice de corrélation

$$
\begin{aligned}
Q_{\Xi\,Y} &= \mathbb{E}[(\Xi - \bar{\Xi})\,(Y - \bar{Y})^*] \\
&= \mathbb{E}[((X - \bar{X}) - Q_{XY}\,Q_Y^{-1}\,(Y - \bar{Y}))\,(Y - \bar{Y})^*] \\
&= \mathbb{E}[(X - \bar{X})\,(Y - \bar{Y})^*] - Q_{XY}\,Q_Y^{-1}\,\mathbb{E}[(Y - \bar{Y})\,(Y - \bar{Y})^*] = 0
\end{aligned}
$$

en particulier, les vecteurs aléatoires gaussiens Ξ et Y sont décorrélés, donc indépendants

il suffit alors d'exprimer la fonction caractéristique de la distribution de probabilité conditionnelle du vecteur aléatoire $X = \Xi + Q_{XY}\, Q_Y^{-1}\, Y$ sachant Y, soit

$$\begin{aligned}
\mathbb{E}[\exp\{i\, u^* X\} \mid Y] &= \mathbb{E}[\exp\{i\, u^* (\Xi + Q_{XY}\, Q_Y^{-1}\, Y)\} \mid Y] \\
&= \exp\{i\, u^*\, Q_{XY}\, Q_Y^{-1}\, Y\}\, \mathbb{E}[\exp\{i\, u^* \Xi\}] \\
&= \exp\{i\, u^*\, Q_{XY}\, Q_Y^{-1}\, Y\}\, \exp\{i\, u^* (\bar{X} - Q_{XY}\, Q_Y^{-1}\, \bar{Y}) - \tfrac{1}{2}\, u^*\, R\, u\} \\
&= \exp\{i\, u^* (\bar{X} + Q_{XY}\, Q_Y^{-1}\, (Y - \bar{Y})) - \tfrac{1}{2}\, u^*\, R\, u\} \\
&= \exp\{i\, u^* \widehat{X}(Y) - \tfrac{1}{2}\, u^*\, R\, u\}
\end{aligned}$$

on reconnait la fonction caractéristique d'un vecteur aléatoire gaussien de moyenne $\widehat{X}(Y)$ et de matrice de covariance R □

Exemple soit X et V deux vecteurs aléatoires gaussiens indépendants de moyenne $\bar{X}$ et 0, de matrice de covariance Q_X et Q_V, et on pose $Y = H\,X + V$
alors $Z = (X, Y)$ vecteur aléatoire gaussien à valeurs dans $\mathbb{R}^m \times \mathbb{R}^d$ de moyenne $\bar{Z} = (\bar{X}, H\,\bar{X})$ et de matrice de covariance

$$Q_Z = \begin{pmatrix} Q_X & Q_X\,H^* \\ H\,Q_X & H\,Q_X\,H^* + Q_V \end{pmatrix}$$

en effet, par différence

$$Y - \bar{Y} = H\,(X - \bar{X}) + V$$

on en déduit facilement la matrice de covariance

$$\begin{aligned} Q_Y &= \mathbb{E}[\,(Y - \bar{Y})\,(Y - \bar{Y})^*\,] \\ &= \mathbb{E}[\,(H\,(X - \bar{X}) + V)\,(H\,(X - \bar{X}) + V)^*\,] \\ &= H\,\mathbb{E}[\,(X - \bar{X})\,(X - \bar{X})^*\,]\,H^* + \mathbb{E}[V\,V^*] \\ &\quad + H\,\mathbb{E}[\,(X - \bar{X})\,V^*\,] + \mathbb{E}[\,V\,(X - \bar{X})^*\,]\,H^* = H\,Q_X\,H^* + Q_V \end{aligned}$$

et la matrice de corrélation

$$\begin{aligned} Q_{XY} &= \mathbb{E}[\,(X-\bar{X})\,(Y-\bar{Y})^*\,] \\ &= \mathbb{E}[\,(X-\bar{X})\,(H\,(X-\bar{X})+V)^*\,] \\ &= \mathbb{E}[\,(X-\bar{X})\,(X-\bar{X})^*\,]\,H^* + \mathbb{E}[\,(X-\bar{X})\,V^*\,] = Q_X\,H^* \end{aligned}$$

si Q_V est inversible, alors a fortiori $H\,Q_X\,H^* + Q_V$ est inversible, et

$$\widehat{X}(Y) = \bar{X} + Q_X\,H^*\,(H\,Q_X\,H^* + Q_V)^{-1}\,(Y - H\,\bar{X})$$

et

$$0 \leq R = Q_X - Q_X\,H^*\,(H\,Q_X\,H^* + Q_V)^{-1}\,H\,Q_X \leq Q_X$$

complément de Schur de la matrice $H\,Q_X\,H^* + Q_V$ dans la matrice–bloc Q_Z

la moyenne conditionnelle $\widehat{X}(Y)$ est l'estimateur MMSE de X sachant Y, donc pour tout estimateur ψ on a

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

dans le cas gaussien : on remarque que $R = J^{-1}$, i.e. la borne de Cramér–Rao a posteriori J^{-1} est atteinte

Remarque autre démonstration de $J^{-1} = R$

$$p(x \mid y) \propto \exp\{-\tfrac{1}{2}\,(x-\widehat{X}(y))^*\,R^{-1}\,(x-\widehat{X}(y))\}$$

d'où

$$-\log p(x \mid y) = \tfrac{1}{2}\,(x-\widehat{X}(y))^*\,R^{-1}\,(x-\widehat{X}(y)) + \text{cste}$$

$$-\frac{\partial^2}{\partial x^2}\log p(x \mid y) = R^{-1}$$

et la matrice d'information de Fisher

$$J = -\mathbb{E}[\frac{\partial^2}{\partial x^2}\log p(X \mid Y)] = R^{-1}$$

Introduction au filtrage

Estimation bayésienne

Systèmes linéaires gaussiens
- systèmes linéaires gaussiens
- filtre de Kalman
- illustration
- lisseur de Kalman

Extensions

Systèmes linéaires gaussiens

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

$$X_k = F_k X_{k-1} + f_k + W_k$$

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

$$Y_k = H_k X_k + h_k + V_k$$

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 Q_k^W
- 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

interprétation du modèle a priori

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

en terme de propagation des incertitudes

- même si l'état $X_{k-1} = x$ est connu exactement à l'instant $(k-1)$, on peut seulement dire que l'état X_k à l'instant k est incertain, et distribué comme un vecteur aléatoire gaussien, de moyenne $F_k\, x + f_k$ et de matrice de covariance Q_k^W
- si l'état X_{k-1} est incertain à l'instant $(k-1)$, et distribué comme un vecteur aléatoire gaussien, de moyenne $\bar{X}_{k-1}$ et de matrice de covariance Q_{k-1}^X, alors cette incertitude se propage à l'instant k : même en absence de bruit, l'état X_k à l'instant k est incertain, et distribué comme un vecteur aléatoire gaussien, de moyenne $F_k\, \bar{X}_{k-1} + f_k$ et de matrice de covariance $F_k\, Q_{k-1}^X\, F_k^*$

Proposition la suite $\{Z_k = (X_k, Y_k)\}$ est un processus aléatoire gaussien

Preuve pour tout instant n, le vecteur aléatoire $(Z_0, Z_1, \cdots, Z_n)$ peut s'exprimer comme transformation affine du vecteur aléatoire $(X_0, W_1, \cdots, W_n, V_0, V_1, \cdots, V_n)$ qui par hypothèse est un vecteur aléatoire gaussien, donc le vecteur aléatoire $(Z_0, Z_1, \cdots, Z_n)$ est gaussien, comme transformation affine d'un vecteur aléatoire gaussien

Résumé on rappelle qu'il s'agit d'estimer le vecteur aléatoire X_k à partir de $Y_{0:k} = (Y_0, \cdots, Y_k)$, de façon *optimale* et *récursive*, dans le modèle

$$X_k = F_k X_{k-1} + f_k + W_k$$

$$Y_k = H_k X_k + h_k + V_k$$

si on adopte le critère MMSE, alors il s'agit de calculer la distribution de probabilité conditionnelle du vecteur aléatoire X_k sachant $Y_{0:k}$, et comme le cadre est gaussien, il suffit de calculer la moyenne et la matrice de covariance déterministe

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

on définit également les quantités suivantes

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

en revanche, le calcul direct à partir des formules de conditionnement dans les vecteurs aléatoires gaussiens est

- de taille croissante
- non récursif

Filtre de Kalman

supposons connue la distribution de probabilité conditionnelle du vecteur aléatoire X_{k-1} sachant $Y_{0:k-1}$: pour calculer la distribution de probabilité conditionnelle du vecteur aléatoire X_k sachant $Y_{0:k}$, on procède en deux étapes

- dans l'étape de *prédiction*, on calcule la distribution de probabilité conditionnelle du vecteur aléatoire X_k sachant les observations passées $Y_{0:k-1}$, ce qui est facile à partir de l'équation d'état
- dans l'étape de *correction*, on utilise la nouvelle observation Y_k, et particulièrement la composante de l'observation Y_k qui apporte une information nouvelle par rapport aux observations passées $Y_{0:k-1}$, c–à–d

$$I_k = Y_k - \mathbb{E}[Y_k \mid Y_{0:k-1}]$$

d'après l'équation d'observation, on a

$$\mathbb{E}[Y_k \mid Y_{0:k-1}] = H_k\,\mathbb{E}[X_k \mid Y_{0:k-1}] + h_k + \mathbb{E}[V_k \mid Y_{0:k-1}] = H_k\,\widehat{X}_k^- + h_k$$

où on a utilisé l'indépendance de V_k et de $Y_{0:k-1}$

Remarque par définition, toute fonction des variables $(Y_0, \cdots, Y_{k-1}, Y_k)$ peut s'exprimer en fonction des variables $(Y_0, \cdots, Y_{k-1}, I_k)$, et réciproquement
on en déduit que $(Y_{0:k-1}, I_k)$ contient exactement la même information que $Y_{0:k}$

Lemme la suite $\{I_k\}$ est un processus gaussien, appelé *processus d'innovation*
en particulier à l'instant k, le vecteur aléatoire I_k est gaussien, de moyenne nulle et de matrice de covariance

$$Q_k^I = H_k \, P_k^- \, H_k^* + Q_k^V$$

et indépendant de $Y_{0:k-1}$
plus généralement, le v.a. $(X_k - \widehat{X}_k^-, I_k)$ est gaussien, de moyenne nulle et de matrice de covariance

$$\begin{pmatrix} P_k^- & P_k^- \, H_k^* \\ H_k \, P_k^- & H_k \, P_k^- \, H_k^* + Q_k^V \end{pmatrix}$$

et indépendant de $Y_{0:k-1}$

Preuve d'après la théorie du conditionnement dans les v.a. gaussiens, l'observation prédite $\mathbb{E}[Y_k \mid Y_{0:k-1}]$ dépend de façon affine des observations passées $(Y_0, Y_1, \cdots, Y_{k-1})$, et donc l'innovation I_k dépend de façon affine des observations $(Y_0, Y_1, \cdots, Y_k)$

on en déduit que le v.a. $(I_0, I_1, \cdots, I_k)$ est gaussien, comme transformation affine du v.a. gaussien $(Y_0, Y_1, \cdots, Y_k)$

toujours d'après la théorie du conditionnement dans les v.a. gaussiens, l'état prédit $\widehat{X}_k^- = \mathbb{E}[X_k \mid Y_{0:k-1}]$ dépend de façon affine des observations passées $(Y_0, Y_1, \cdots, Y_{k-1})$

on en déduit que le v.a. $(Y_0, Y_1, \cdots, Y_{k-1}, X_k - \widehat{X}_k^-, I_k)$ est gaussien — et en particulier le v.a. $(X_k - \widehat{X}_k^-, I_k)$ est gaussien — comme transformation affine d'un v.a. gaussien, et compte tenu que

$$\mathbb{E}[X_k - \widehat{X}_k^- \mid Y_{0:k-1}] = 0 \qquad \text{et} \qquad \mathbb{E}[I_k \mid Y_{0:k-1}] = 0$$

par définition, le v.a. $(X_k - \widehat{X}_k^-, I_k)$ est indépendant de $Y_{0:k-1}$

d'après l'équation d'observation, on a

$$I_k = Y_k - (H_k\,\widehat{X}_k^- + h_k) = H_k\,(X_k - \widehat{X}_k^-) + V_k$$

et on en déduit que

$$\begin{aligned} Q_k^I &= \mathbb{E}[I_k\,I_k^*] \\ &= \mathbb{E}[(H_k\,(X_k - \widehat{X}_k^-) + V_k)\,(H_k\,(X_k - \widehat{X}_k^-) + V_k)^*] \\ &= H_k\,\mathbb{E}[(X_k - \widehat{X}_k^-)\,(X_k - \widehat{X}_k^-)^*]\,H_k^* + \mathbb{E}[V_k\,V_k^*] \\ &\qquad + \mathbb{E}[V_k\,(X_k - \widehat{X}_k^-)^*]\,H_k^* + H_k\,\mathbb{E}[(X_k - \widehat{X}_k^-)\,V_k^*] \\ &= H_k\,P_k^-\,H_k^* + Q_k^V \end{aligned}$$

où on a utilisé dans la dernière égalité, l'indépendance de $(X_k - \widehat{X}_k^-)$ et de V_k, donc $\mathbb{E}[(X_k - \widehat{X}_k^-)\,V_k^*] = 0$

on a également

$$\begin{aligned}\mathbb{E}[(X_k - \widehat{X}_k^-)\, I_k^*] &= \mathbb{E}[(X_k - \widehat{X}_k^-)\, (H_k\,(X_k - \widehat{X}_k^-) + V_k)^*] \\ &= \mathbb{E}[(X_k - \widehat{X}_k^-)\, (X_k - \widehat{X}_k^-)^*]\, H_k^* + \mathbb{E}[(X_k - \widehat{X}_k^-)\, V_k^*] \\ &= P_k^-\, H_k^*\end{aligned}$$

où on a utilisé dans la dernière égalité, l'indépendance de $(X_k - \widehat{X}_k^-)$ et de V_k, donc $\mathbb{E}[(X_k - \widehat{X}_k^-)\, V_k^*] = 0$ □

Remarque si la matrice de covariance Q_k^V est inversible, alors a fortiori la matrice de covariance $Q_k^I = H_k\, P_k^-\, H_k^* + Q_k^V$ est inversible

Théorème [Filtre de Kalman] on suppose que la matrice de covariance Q_k^V est *inversible*, pour tout instant k
alors les suites $\{\widehat{X}_k\}$ et $\{P_k\}$ vérifient les équations récurrentes suivantes

$$\widehat{X}_k^- = F_k \widehat{X}_{k-1} + f_k$$

$$P_k^- = F_k P_{k-1} F_k^* + Q_k^W$$

et

$$\widehat{X}_k = \widehat{X}_k^- + K_k (Y_k - (H_k \widehat{X}_k^- + h_k))$$

$$P_k = (I - K_k H_k) P_k^-$$

où la matrice

$$K_k = P_k^- H_k^* [H_k P_k^- H_k^* + Q_k^V]^{-1}$$

est appelée *gain de Kalman*, avec les initialisations

$$\widehat{X}_0^- = \bar{X}_0 = \mathbb{E}[X_0] \qquad \text{et} \qquad P_0^- = Q_0^X = \mathrm{cov}(\mathrm{X}_0)$$

Remarque la suite $\{P_k\}$ ne dépend pas des observations : elle peut donc être pré–calculée

Remarque au vu de l'expression développée

$$P_k = P_k^- - P_k^- H_k^* [H_k P_k^- H_k^* + Q_k^V]^{-1} H_k P_k^-$$

on vérifie aisément que $P_k \leq P_k^-$, c'est–à–dire que la matrice de covariance de l'erreur de filtrage est plus petite (au sens des matrices symétriques) que la matrice de covariance de l'erreur de prédiction

Remarque si les coefficients F_k, f_k et Q_k^W dans l'équation d'état et les coefficients H_k, h_k et Q_k^V dans l'équation d'observation dépendent des observations passées $Y_{0:k-1}$, alors la suite $\{Z_k = (X_k, Y_k)\}$ n'est plus gaussienne, mais conditionnellement à $Y_{0:k-1}$ le couple (X_k, Y_k) est gaussien : on dit que la suite $\{X_k\}$ est *conditionnellement gaussienne* on peut vérifier que la distribution de probabilité conditionnelle du v.a. X_k sachant $Y_{0:k}$ est gaussienne, de moyenne $\widehat{X}_k$ et de matrice de covariance P_k données encore par les équations du Théorème

Preuve on procède en plusieurs étapes, en s'appuyant sur la théorie du conditionnement dans les v.a. gaussiens

- initialisation expression de $\widehat{X}_0$ et P_0 en fonction de $\widehat{X}_0^-$ et P_0^-

le v.a. (X_0, Y_0) est gaussien, de moyenne et de matrice de covariance données par

$$\begin{pmatrix} \widehat{X}_0^- \\ H_0\,\widehat{X}_0^- + h_0 \end{pmatrix} \quad \text{et} \quad \begin{pmatrix} P_0^- & P_0^-\,H_0^* \\ H_0\,P_0^- & H_0\,P_0^-\,H_0^* + Q_0^V \end{pmatrix}$$

respectivement
on en déduit que la distribution de probabilité conditionnelle du v.a. X_0 sachant Y_0 est gaussienne, de moyenne

$$\widehat{X}_0 = \widehat{X}_0^- + P_0^-\,H_0^*\,[H_0\,P_0^-\,H_0^* + Q_0^V]^{-1}\,(Y_0 - (H_0\,\widehat{X}_0^- + h_0))$$

et de matrice de covariance déterministe

$$P_0 = P_0^- - P_0^-\,H_0^*\,[H_0\,P_0^-\,H_0^* + Q_0^V]^{-1}\,H_0\,P_0^-$$

▶ étape de prédiction expression de $\widehat{X}_k^-$ et P_k^- en fonction de $\widehat{X}_{k-1}$ et P_{k-1}

le v.a. $(X_k, Y_0, \cdots, Y_{k-1})$ est gaussien, de sorte que la distribution de probabilité conditionnelle du v.a. X_k sachant $Y_{0:k-1}$ est gaussienne, de moyenne $\widehat{X}_k^-$ et de matrice de covariance déterministe P_k^-
d'après l'équation d'état

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

on a

$$\widehat{X}_k^- = \mathbb{E}[X_k \mid Y_{0:k-1}] = F_k\, \mathbb{E}[X_{k-1} \mid Y_{0:k-1}] + f_k + \mathbb{E}[W_k \mid Y_{0:k-1}]$$

$$= F_k\, \widehat{X}_{k-1} + f_k$$

compte tenu que W_k et $Y_{0:k-1}$ sont indépendants
par différence

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

on en déduit que

$$
\begin{aligned}
P_k^- &= \mathbb{E}[(X_k - \widehat{X}_k^-)\,(X_k - \widehat{X}_k^-)^*] \\
&= \mathbb{E}[(F_k\,(X_{k-1} - \widehat{X}_{k-1}) + W_k)\,(F_k\,(X_{k-1} - \widehat{X}_{k-1}) + W_k)^*] \\
&= F_k\,\mathbb{E}[(X_{k-1} - \widehat{X}_{k-1})\,(X_{k-1} - \widehat{X}_{k-1})^*]\,F_k^* + \mathbb{E}[W_k\,W_k^*] \\
&\qquad + \mathbb{E}[W_k\,(X_{k-1} - \widehat{X}_{k-1})^*]\,F_k^* + F_k\,\mathbb{E}[(X_{k-1} - \widehat{X}_{k-1})\,W_k^*] \\
&= F_k\,P_{k-1}\,F_k^* + Q_k^W
\end{aligned}
$$

où on a utilisé dans la dernière égalité, l'indépendance de $(X_{k-1} - \widehat{X}_{k-1})$ et de W_k, donc $\mathbb{E}[(X_{k-1} - \widehat{X}_{k-1})\,W_k^*] = 0$

▶ étape de correction expression de $\widehat{X}_k$ et P_k en fonction de $\widehat{X}_k^-$ et P_k^-

le v.a. $(X_k, Y_0, \cdots, Y_k)$ est gaussien, de sorte que la distribution de probabilité conditionnelle du v.a. X_k sachant $Y_{0:k}$ est gaussienne, de moyenne $\widehat{X}_k$ et de matrice de covariance déterministe P_k
compte tenu de l'indépendance entre I_k et $Y_{0:k-1}$, on a

$$\begin{aligned}
\widehat{X}_k &= \mathbb{E}[X_k \mid Y_{0:k}] \\
&= \widehat{X}_k^- + \mathbb{E}[X_k - \widehat{X}_k^- \mid Y_{0:k}] \\
&= \widehat{X}_k^- + \mathbb{E}[X_k - \widehat{X}_k^- \mid Y_{0:k-1}, I_k] \\
&= \widehat{X}_k^- + \mathbb{E}[X_k - \widehat{X}_k^- \mid I_k]
\end{aligned}$$

par différence

$$X_k - \widehat{X}_k = (X_k - \widehat{X}_k^-) - (\widehat{X}_k - \widehat{X}_k^-) = (X_k - \widehat{X}_k^-) - \mathbb{E}[X_k - \widehat{X}_k^- \mid I_k]$$

de sorte que

$$P_k = \mathbb{E}[\,(X_k - \widehat{X}_k)\,(X_k - \widehat{X}_k)^*\,]$$

$$= \mathbb{E}[\,((X_k - \widehat{X}_k^-) - \mathbb{E}[X_k - \widehat{X}_k^- \mid I_k])\,((X_k - \widehat{X}_k^-) - \mathbb{E}[X_k - \widehat{X}_k^- \mid I_k])^*\,]$$

il suffit donc de calculer la moyenne conditionnelle et la matrice de covariance conditionnelle du v.a. $(X_k - \widehat{X}_k^-)$ sachant I_k, or le v.a. $(X_k - \widehat{X}_k^-, I_k)$ est gaussien, de moyenne nulle et de matrice de covariance

$$\begin{pmatrix} P_k^- & P_k^-\,H_k^* \\ H_k\,P_k^- & H_k\,P_k^-\,H_k^* + Q_k^V \end{pmatrix}$$

si la matrice Q_k^V est inversible, alors a fortiori la matrice $Q_k^I = H_k\,P_k^-\,H_k^* + Q_k^V$ est inversible, et on en déduit que

$$\widehat{X}_k = \widehat{X}_k^- + P_k^-\,H_k^*\,[H_k\,P_k^-\,H_k^* + Q_k^V]^{-1}\,I_k$$

et

$$P_k = P_k^- - P_k^-\,H_k^*\,[H_k\,P_k^-\,H_k^* + Q_k^V]^{-1}\,H_k\,P_k^- \qquad \square$$

Lisseur de Kalman

il s'agit estimer l'état caché X_k pour un instant k, intermédiaire entre l'instant initial 0 et l'instant final n, au vu de *toutes* les observations $Y_{0:n} = (Y_0, \cdots, Y_n)$
si on adopte le critère MMSE, alors il s'agit de calculer la distribution de probabilité conditionnelle du vecteur aléatoire X_k sachant $Y_{0:n}$, et comme le cadre est gaussien, il suffit de calculer la moyenne et la matrice de covariance déterministe

$$\widehat{X}_k^n = \mathbb{E}[X_k \mid Y_{0:n}] \qquad \text{et} \qquad P_k^n = \mathbb{E}[(X_k - \widehat{X}_k^n)\,(X_k - \widehat{X}_k^n)^*]$$

clairement, $\widehat{X}_n^n = \widehat{X}_n$ et $P_n^n = P_n$ pour $k = n$

Remarque si la matrice de covariance Q_k^W est inversible, alors a fortiori la matrice de covariance $P_k^- = F_k\, P_{k-1}\, F_k^* + Q_k^W$ est inversible

Théorème [Lisseur de Kalman (Rauch–Tung–Striebel)] on suppose que les matrices de covariance Q_k^W et Q_k^V sont *inversibles*, pour tout instant k alors les suites $\{\widehat{X}_k^n\}$ et $\{P_k^n\}$ vérifient les équations récurrentes rétrogrades suivantes

$$\widehat{X}_{k-1}^n = \widehat{X}_{k-1} + L_k\,(\widehat{X}_k^n - \widehat{X}_k^-)$$

$$P_{k-1}^n = P_{k-1} + L_k\,(P_k^n - P_k^-)\,L_k^*$$

avec la matrice de gain

$$L_k = P_{k-1}\,F_k^*\,(P_k^-)^{-1}$$

et avec les initialisations

$$\widehat{X}_n^n = \widehat{X}_n \qquad \text{et} \qquad P_n^n = P_n$$

Remarque on vérifie par récurrence arrière que $P_k^n \leq P_k$, c'est–à–dire que la matrice de covariance de l'erreur de lissage est plus petite (au sens des matrices symétriques) que la matrice de covariance de l'erreur de filtrage

par définition $P_n^n = P_n$, c'est–à–dire que la relation est vraie au rang n

si la relation est vraie au rang k, c'est–à–dire si $P_k^n \leq P_k$, alors nécessairement $P_k^n \leq P_k^-$ compte tenu que $P_k \leq P_k^-$: en d'autres termes, la différence $(P_k^n - P_k^-)$ est semi–définie négative, de sorte que la différence

$$P_{k-1}^n - P_{k-1} = L_k \left(P_k^n - P_k^-\right) L_k^*$$

aussi est semi–définie négative : en d'autres termes $P_{k-1}^n \leq P_{k-1}$, c'est–à–dire que la relation est vraie au rang $(k-1)$

cette formulation fait l'hypothèse que la matrice de covariance Q_k^W est inversible, et utilise explicitement l'inverse de la matrice de covariance P_k^-

formulation alternative : pour tout $k = 1, \cdots, n$, on introduit les variables

$$r_{k-1} = F_k^* \left(P_k^-\right)^{-1} \left(\widehat{X}_k^n - \widehat{X}_k^-\right)$$

et

$$\Pi_{k-1} = -F_k^* \left(P_k^-\right)^{-1} \left(P_k^n - P_k^-\right) \left(P_k^-\right)^{-1} F_k$$

et on pose $r_n = 0$ et $\Pi_n = 0$ par convention

on rappelle que la différence $\left(P_k^n - P_k^-\right)$ est semi–définie négative, de sorte que la matrice Π_{k-1} est semi–définie positive

clairement

$$\begin{aligned}\widehat{X}_k^n &= \widehat{X}_k + L_{k+1}\,(\widehat{X}_{k+1}^n - \widehat{X}_{k+1}^-)\\ &= \widehat{X}_k + P_k\,F_{k+1}^*\,(P_{k+1}^-)^{-1}\,(\widehat{X}_{k+1}^n - \widehat{X}_{k+1}^-)\\ &= \widehat{X}_k + P_k\,r_k\end{aligned}$$

et de même

$$\begin{aligned}P_k^n &= P_k + L_{k+1}\,(P_{k+1}^n - P_{k+1}^-)\,L_{k+1}^*\\ &= P_k + P_k\,F_{k+1}^*\,(P_{k+1}^-)^{-1}\,(P_{k+1}^n - P_{k+1}^-)\,(P_{k+1}^-)^{-1}\,F_{k+1}\,P_k\\ &= P_k - P_k\,\Pi_k\,P_k\end{aligned}$$

on pose

$$\Xi_k = [H_k\,P_k^-\,H_k^* + Q_k^V]^{-1}$$

Théorème [Lisseur de Kalman (Fraser–Potter)] on suppose que la matrice de covariance Q_k^V est *inversible*, pour tout instant k
alors les suites $\{\widehat{X}_k^n\}$ et $\{P_k^n\}$ sont données par les expressions suivantes

$$\widehat{X}_k^n = \widehat{X}_k + P_k\, r_k \qquad \text{et} \qquad P_k^n = P_k - P_k\, \Pi_k\, P_k$$

où les suites $\{r_k\}$ et $\{\Pi_k\}$ vérifient les équations récurrentes rétrogrades suivantes

$$r_k^- = (I - K_k\, H_k)^*\, r_k + H_k^*\, \Xi_k\, (Y_k - (H_k\, \widehat{X}_k^- + h_k))$$

$$\Pi_k^- = (I - K_k\, H_k)^*\, \Pi_k\, (I - K_k\, H_k) + H_k^*\, \Xi_k\, H_k$$

et

$$r_{k-1} = F_k^*\, r_k^- \qquad \text{et} \qquad \Pi_{k-1} = F_k^*\, \Pi_k^-\, F_k$$

avec les initialisations

$$r_n = 0 \qquad \text{et} \qquad \Pi_n = 0$$

Remarque il est également possible d'obtenir une équation récurrente pour le lisseur, dans le sens direct (et pas dans le sens rétrograde) et autonome (ne faisant pas intervenir ni le filtre ni la matrice de covariance de l'erreur de filtrage)

$$\widehat{X}_k^n = F_k \, \widehat{X}_{k-1}^n + Q_k^W \, r_k^-$$

mais il n'existe pas de résultat similaire pour la matrice de covariance de l'erreur de lissage

comparaison des formulations Rauch–Tung–Striebel vs. Fraser–Potter

- les deux formulations requièrent dans la phase aller une même condition d'inversibilité et l'inversion de systèmes linéaires de dimension d
- la formulation de Rauch–Tung–Striebel requiert dans la phase retour une condition d'inversibilité supplémentaire et l'inversion de systèmes linéaires de dimension m
- les deux formulations utilisent dans la phase retour les valeurs numériques du filtre et de la matrice de covariance d'erreur de filtrage — ces valeurs numériques sont calculées dans la phase aller, et doivent donc être conservées en mémoire pour être utilisées dans la phase retour
- la formulation de Fraser–Potter utilise dans la phase retour la valeur numérique de l'observation ou de manière équivalente celle de l'innovation — ces valeurs numériques sont calculées dans la phase aller, et doivent donc être conservées en mémoire pour être utilisées dans la phase retour

Introduction au filtrage

Estimation bayésienne

Systèmes linéaires gaussiens

Extensions

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

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 les matrices de covariance $Q_k^W(x) = \sigma_k(x)\,(\sigma_k(x))^*$ sont *inversibles* pour tout $x \in \mathbb{R}^m$, alors il existe des *densités de transition* définies par

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

et si les matrices de covariance Q_k^V sont *inversibles*, alors il existe des *densités d'émission* définies par

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

clairement

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

et

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

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

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

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 $x \in \mathbb{R}^m$, et si les matrices de covariance Q_k^W et Q_k^V sont *inversibles*, alors 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 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

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

on évalue empiriquement l'erreur quadratique moyenne

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

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

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

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

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

$$\approx \frac{1}{M}\;\sum_{j=1}^{M}(h_k'(X_k^j))^*\,(Q_k^V)^{-1}\,h_k'(X_k^j)$$

et on calcule récursivement

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

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

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
- 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

$$\widehat{X}_k^- = b_k(\widehat{X}_{k-1})$$

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

et

$$\widehat{X}_k = \widehat{X}_k^- + K_k\,(Y_k - h_k(\widehat{X}_k^-))$$

$$P_k = (I - K_k\, h_k'(\widehat{X}_k^-))\; P_k^-$$

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 vecteur aléatoire X_0