

École Nationale de la Statistique et de l'Analyse de l'Information
Filière : Génie Statistique
UE Traitement de l'Image

Filtrage Linéaire et Non-Linéaire

version du 17 janvier 2016

François Le Gland
INRIA Rennes et IRMAR
<http://www.irisa.fr/aspi/legland/ensai/>

Objectif du cours

Le filtrage consiste à estimer de façon récursive un état caché au vu d'observations. Le domaine d'application principal est la localisation, la navigation et la poursuite de mobiles, dans le domaine militaire, mais aussi en robotique mobile, en vision par ordinateur, où il s'agit de combiner : un modèle a priori de déplacement du mobile, des mesures issues de capteurs, et éventuellement une base de mesures de références, disponibles par exemples sous la forme d'une carte numérique (modèle numérique de terrain, carte de couverture, etc.).

Le problème de filtrage possède une solution explicite, appelée filtre de Kalman, dans le cas particulier des systèmes linéaires gaussiens. Dans le cas plus général des modèles de Markov cachés, des méthodes de simulations efficaces sont apparues récemment, sous le nom de filtrage particulaire. L'objectif de ce cours est de présenter différents algorithmes de filtrage linéaire, de filtrage non-linéaire (obtenus par linéarisation ou par quadrature), et de mettre en œuvre ces algorithmes dans le cadre d'un TP.

Table des matières

1	Introduction	1
1.1	Importance de l'information a priori	1
1.2	Estimation bayésienne	4
1.3	Cadre gaussien	9
2	Exemples	15
2.1	Recalage altimétrique de navigation inertielle	15
2.2	Suivi visuel par histogramme de couleur	19
2.3	Poursuite d'une cible furtive (track-before-detect)	22
2.4	Navigation en environnement intérieur	24
3	Filtrage de Kalman	29
3.1	Systèmes linéaires gaussiens	29
3.2	Filtre de Kalman	32
3.3	Lisseur de Kalman	37
4	Extensions aux systèmes non-linéaires	41
4.1	Filtre de Kalman linéarisé, filtre de Kalman étendu	42
4.2	Filtre de Kalman <i>unscented</i>	44
5	Filtrage bayésien	51
5.1	Systèmes non-linéaires à bruits non-gaussiens	51
5.2	Modèles de Markov cachés	54
5.3	Filtre bayésien	55
5.4	Approximation particulière	59
6	Borne de Cramér-Rao a posteriori	65
A	Inversion matricielle	71

Chapitre 1

Introduction

En toute généralité, le filtrage consiste à estimer l'état d'un système dynamique, c'est-à-dire évoluant au cours du temps, à partir d'observations partielles, généralement bruitées.

Typiquement, on dispose d'une suite Y_0, Y_1, \dots, Y_n d'observations, par exemple obtenues après traitement préalable du signal recueilli au niveau des capteurs. Chaque observation Y_n est reliée à l'état inconnu X_n par une relation du type

$$Y_n = h_n(X_n) + V_n, \quad (1.1)$$

où V_n est un *bruit*, qui modélise l'erreur d'observation. On précisera plus loin dans ce cours la notion de *bruit*, en terme de variables aléatoires le plus souvent centrées (de moyenne nulle).

1.1 Importance de l'information a priori

Une hypothèse assez commune est de supposer que les variables aléatoires V_0, V_1, \dots, V_n sont indépendantes entre elles. A cause de cette hypothèse d'indépendance mutuelle des bruits d'observation, et a fortiori en absence de bruit, seule l'observation Y_n participe à l'estimation de l'état caché X_n , c'est-à-dire qu'on se trouve confronté à une succession de problèmes d'estimation découplés : dans la relation (1.1), l'observation Y_n est disponible (par définition) tandis que ni l'état caché X_n ni le bruit V_n ne sont disponibles, et il faut arriver à retrouver (estimer) l'état caché X_n au vu de l'observation Y_n et malgré la présence du bruit V_n .

Tel qu'il est formulé, le problème de l'estimation de l'état caché X_n à partir des observations Y_0, Y_1, \dots, Y_n est en général mal-posé :

- en général, la dimension m de la variable cachée est plus grande que la dimension d de l'observation : même en absence de bruit, on ne peut pas inverser la relation (1.1) qui possède plus d'inconnues que d'équations,
- dans le cas favorable où $m = d$, et même en absence de bruit, il n'est pas toujours possible d'inverser la relation (1.1) qui peut très bien posséder plusieurs solutions distinctes,

- la situation est évidemment encore plus compliquée en présence de bruit : à cause du phénomène de découplage cité plus haut, la suite X_0, X_1, \dots, X_n reconstituée peut ne pas être pertinente en tant que trajectoire, même si chacune des estimations est pertinente séparément.

Pour lever l'indétermination, c'est-à-dire pour garantir l'existence d'une solution unique, et pour résoudre le problème de cohérence temporelle, la solution classique consiste à utiliser des informations supplémentaires sur la suite cachée, par exemple sous la forme de fonctions de coût portant sur l'état initial ou sur les transitions entre deux états successifs. Par exemple, on cherchera à minimiser le critère

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

par rapport à la suite $x_{0:n} = (x_0, x_1, \dots, x_n)$, qui combine des fonctions de coût représentant l'information *a priori* sur la solution avec des fonctions de coût d'une autre nature, qui peuvent représenter par exemple un terme d'*attache aux données*, de la forme

$$h_k(x) = \frac{1}{2} |Y_k - h_k(x)|^2 \quad \text{ou bien} \quad h_k(x) = \frac{1}{2} (Y_k - h_k(x))^* I_k (Y_k - h_k(x)) ,$$

pour tout $k = 0, 1, \dots, n$, avec l'interprétation que la suite recherchée doit également vérifier à chaque instant l'équation d'observation en un sens approché. Plus généralement, ces fonctions de coût peuvent juste représenter une contrainte (ou une propriété) que la suite recherchée devrait vérifier (ou posséder). En absence d'information *a priori*, le critère se réduit simplement à

$$J(x_{0:n}) = \frac{1}{2} \sum_{k=0}^n |Y_k - h_k(x_k)|^2 \quad \text{ou bien} \quad J(x_{0:n}) = \frac{1}{2} \sum_{k=0}^n (Y_k - h_k(x_k))^* I_k (Y_k - h_k(x_k)) ,$$

ce qui revient en absence de couplage à minimiser séparément le critère

$$\frac{1}{2} |Y_k - h_k(x_k)|^2 \quad \text{ou bien} \quad \frac{1}{2} (Y_k - h_k(x_k))^* R_k^{-1} (Y_k - h_k(x_k)) ,$$

par rapport à l'état x_k , pour tout $k = 0, 1, \dots, n$, avec les conséquences déjà évoquées en terme d'indétermination et de possible incohérence temporelle. Un exemple classique de fonctions de coût représentant l'information *a priori* est

$$c_0(x) = \frac{1}{2} |x - \mu|^2 \quad \text{ou bien} \quad c_0(x) = \frac{1}{2} (x - \mu)^* \Sigma_0^{-1} (x - \mu) ,$$

avec l'interprétation que l'état initial x_0 recherché doit être proche de μ , et

$$c_k(x, x') = \frac{1}{2} |x' - f_k(x)|^2 \quad \text{ou bien} \quad c_k(x, x') = \frac{1}{2} (x' - f_k(x))^* Q_k^{-1} (x' - f_k(x)) ,$$

avec l'interprétation que l'état x_k recherché doit être proche de $f_k(x_{k-1})$, ou de manière équivalente que la transition (x_{k-1}, x_k) recherchée doit vérifier l'équation $x_k = f_k(x_{k-1})$ dans un sens approché, pour tout $k = 1, \dots, n$. On remarque que ces fonctions de coût sont (à une constante additive près) de la forme

$$c_0(x) = -\log p_0(x) \quad \text{et} \quad c_k(x, x') = -\log p_k(x' | x) , \quad (1.2)$$

pour tout $k = 1, \dots, n$, où $p_0(x)$ est la densité de probabilité initiale, et où $p_k(x' | x)$ est la densité de probabilité de transition, dans le modèle non-linéaire suivant avec bruits gaussiens additifs

$$X_k = f_k(X_{k-1}) + W_k \quad \text{avec} \quad W_k \sim \mathcal{N}(0, Q_k) ,$$

et avec condition initiale $X_0 \sim \mathcal{N}(\mu, \Sigma)$. En effet (à une constante mutiplicative de normalisation près)

$$\mathbb{P}[X_0 \in dx] \propto \exp\{-\frac{1}{2} (x - \mu)^* \Sigma_0^{-1} (x - \mu)\} dx \propto p_0(x) dx ,$$

et

$$\mathbb{P}[X_k \in dx' | X_{k-1} = x] \propto \exp\{-\frac{1}{2} (x' - f_k(x))^* Q_k^{-1} (x' - f_k(x))\} dx' \propto p_k(x' | x) dx' ,$$

pour tout $k = 1, \dots, n$. En toute généralité, si les relations (1.2) sont vérifiées pour une densité de probabilité $p_0(x)$ et pour des densités de probabilité de transition $p_k(x' | x)$, pour tout $k = 1, \dots, n$, alors le critère à minimiser peut s'écrire

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

ce qui revient à maximiser

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

par rapport à la suite $x_{0:n} = (x_0, x_1, \dots, x_n)$. On remarque que $p_{0:n}(x_{0:n})$ représente la densité de probabilité conjointe des états successifs (X_0, X_1, \dots, 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 $p_k(x' | x)$, pour tout $k = 1, \dots, n$.

Comme alternative au point de vue de l'optimisation déterministe développé jusqu'ici, on adoptera dans ce cours un point de vue d'estimation bayésienne, c'est-à-dire qu'on remplacera le problème de minimisation déterministe, avec prise en compte de l'information *a priori* en terme de fonctions de coût, par le problème du calcul de la distribution de Gibbs-Boltzmann définie (à une constante mutiplicative près) sur l'espace des trajectoires $E_n = E \times \dots \times E$ par

$$\exp\{-J(x_{0:n})\} dx_{0:n} = \underbrace{p_0(x_0) \prod_{k=1}^n p_k(x_k | x_{k-1})}_{p_{0:n}(x_{0:n})} \exp\{-\sum_{k=0}^n d_k(x_k)\} dx_{0:n} . \quad (1.3)$$

En d'autres termes, on remplacera le problème de calculer le mode, c'est-à-dire la trajectoire $x_{0:n} = (x_0, x_1, \dots, x_n)$ de plus forte densité, par le problème de calculer des espérances (ou des

intégrales) du type

$$\begin{aligned} \int_E \cdots \int_E f(x_{0:n}) \exp\{-J(x_{0:n})\} dx_{0:n} &= \int_E \cdots \int_E f(x_{0:n}) \exp\left\{-\sum_{k=0}^n d_k(x_k)\right\} p_{0:n}(x_{0:n}) dx_{0:n} \\ &= \mathbb{E}[f(X_{0:n}) \exp\left\{-\sum_{k=0}^n d_k(X_k)\right\}] , \end{aligned}$$

pour des fonctions-test f définies sur l'espace des trajectoires $E_n = E \times \cdots \times E$. Dans la pratique, 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 (à une constante multiplicative près) par (1.3).

1.2 Estimation bayésienne

Dans de nombreux cas, la prise en compte de l'information a priori peut se ramener au problème statique suivant : étant donnés deux vecteurs aléatoires X et Y , qu'apporte le fait d'observer la réalisation $Y = y$ sur la connaissance que l'on a de X ?

Soit X et Y deux variables aléatoires à valeurs dans E et dans F respectivement, et soit ϕ une application mesurable définie sur E à valeurs dans \mathbb{R}^p . Par définition, un *estimateur* de $\phi(X)$ à partir de l'observation de Y est un vecteur aléatoire $\psi(Y)$, où ψ est une application mesurable définie sur F à valeurs dans \mathbb{R}^p (par abus de notation, la variable aléatoire $\psi(Y)$ sera également notée ψ).

► **Estimateur MMSE** Soit ψ un estimateur de $\phi(X)$ sachant Y . Naturellement $\psi = \psi(Y)$ n'est pas égal à $\phi(X)$: une mesure de l'écart entre l'estimateur et la vraie valeur est fournie par la matrice (de dimension $p \times p$) de corrélation d'erreur

$$\mathbb{E}[(\psi(Y) - \phi(X))(\psi(Y) - \phi(X))^*] , \quad (1.4)$$

dont la trace

$$\text{trace } \mathbb{E}[(\psi(Y) - \phi(X))(\psi(Y) - \phi(X))^*] = \mathbb{E}|\psi(Y) - \phi(X)|^2 ,$$

est l'erreur quadratique moyenne. L'estimateur du minimum d'erreur quadratique moyenne (MMSE, pour *minimum mean-square error*) de $\phi(X)$ sachant Y est un estimateur $\hat{\phi}$ tel que

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

au sens des matrices symétriques, pour tout autre estimateur ψ .

La Proposition 1.1 ci-dessous montre que cet estimateur est obtenu à l'aide de la distribution de probabilité conditionnelle de X sachant $Y = y$, définie à partir de la distribution de probabilité jointe de (X, Y) par la décomposition

$$\mathbb{P}[X \in dx, Y \in dy] = \mathbb{P}[X \in dx | Y = y] \mathbb{P}[Y \in dy] . \quad (1.5)$$

Proposition 1.1 Soit X et Y deux variables aléatoires à valeurs dans E et F respectivement, et soit ϕ une application mesurable définie sur E à valeurs dans \mathbb{R}^p . L'estimateur MMSE de $\phi(X)$ sachant Y est la moyenne conditionnelle de $\phi(X)$ sachant Y , i.e.

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

PREUVE. Pour tout estimateur ψ , la décomposition

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

entraîne

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

et on remarque que

$$\begin{aligned} \mathbb{E}[(\psi(Y) - \widehat{\phi}(Y)) (\widehat{\phi}(Y) - \phi(X))^*] &= \\ &= \int_E \int_F (\psi(y) - \widehat{\phi}(y)) (\widehat{\phi}(y) - \phi(x))^* \mathbb{P}[X \in dx, Y \in dy] \\ &= \int_E \int_F (\psi(y) - \widehat{\phi}(y)) (\widehat{\phi}(y) - \phi(x))^* \mathbb{P}[X \in dx \mid Y = y] \mathbb{P}[Y \in dy] \\ &= \int_F (\psi(y) - \widehat{\phi}(y))^* \left\{ \int_E (\widehat{\phi}(y) - \phi(x)) \mathbb{P}[X \in dx \mid Y = y] \right\} \mathbb{P}[Y \in dy] = 0 , \end{aligned}$$

par définition de $\widehat{\phi}(y)$. On a donc

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

au sens des matrices symétriques, avec égalité pour $\psi = \widehat{\phi}$. □

Remarque 1.2 Compte tenu que le vecteur aléatoire $(\widehat{\phi}(Y) - \phi(X))$ est centré, la matrice de corrélation d'erreur est aussi la matrice de covariance d'erreur, dans le cas particulier de l'estimateur $\widehat{\phi}$.

► **Borne de Cramér–Rao a posteriori** On suppose à présent que $E = \mathbb{R}^m$, c'est-à-dire que X et Y sont des variables aléatoires à valeurs dans \mathbb{R}^m et F respectivement, et soit ϕ une application mesurable définie sur \mathbb{R}^m à valeurs dans \mathbb{R}^p . Le biais de l'estimateur ψ de $\phi(X)$ sachant Y est défini par

$$b(\psi, x) = \mathbb{E}[\psi(Y) \mid X = x] - \phi(x) .$$

On suppose que la distribution de probabilité jointe des vecteurs aléatoires X et Y possède une densité

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

sur $\mathbb{R}^m \times F$, suffisamment régulière par rapport à la variable $x \in \mathbb{R}^m$, avec les deux factorisations alternatives

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

en termes de distributions de probabilités conditionnelles et marginales, et en particulier

$$\mathbb{P}[X \in dx] = p(x) dx \quad \text{avec} \quad p(x) = \int_F p(x, y) \lambda(dy) .$$

On suppose que

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

Proposition 1.3 *Si la matrice d'information de Fisher (de dimension $m \times m$) définie par*

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

est inversible, alors la matrice de corrélation de l'erreur d'estimation est minorée (au sens des matrices symétriques) par la relation suivante

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

avec la matrice de sensibilité (de dimension $p \times m$) définie par

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

pour tout estimateur ψ de $\phi(X)$ sachant Y tel que

$$\int_{\mathbb{R}^m} (b(\psi, x) p(x))' dx = 0 . \tag{1.6}$$

Remarque 1.4 La matrice d'information de Fisher J et la matrice de sensibilité M qui interviennent dans l'expression de la borne ne dépendent pas de l'estimateur ψ .

PREUVE. Par définition

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

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

$$\begin{aligned} (b(\psi, x) p(x))' &= -\phi'(x) \int_F p(x, y) \lambda(dy) + \int_F (\psi(y) - \phi(x)) \frac{\partial}{\partial x} p(x, y) \lambda(dy) \\ &= -\phi'(x) p(x) + \int_F (\psi(y) - \phi(x)) \frac{\partial}{\partial x} \log p(x, y) p(x, y) \lambda(dy) . \end{aligned}$$

En intégrant par rapport à la variable $x \in \mathbb{R}^m$, il vient

$$\begin{aligned} &\int_{\mathbb{R}^m} (b(\psi, x) p(x))' dx \\ &= - \int_{\mathbb{R}^m} \phi'(x) p(x) dx + \int_{\mathbb{R}^m} \int_F (\psi(y) - \phi(x)) \frac{\partial}{\partial x} \log p(x, y) p(x, y) \lambda(dy) dx \\ &= -\mathbb{E}[\phi'(X)] + \mathbb{E}[(\psi(Y) - \phi(X)) \frac{\partial}{\partial x} \log p(X, Y)] , \end{aligned}$$

et si la condition (1.6) est vérifiée, alors

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

où la matrice de sensibilité M ne dépend pas de ψ . D'autre part, il résulte de l'identité

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

entre matrices de dimension $m \times m$, que

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

et par hypothèse on a donc

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

On introduit ensuite le vecteur aléatoire

$$Z = \begin{pmatrix} \psi(Y) - \phi(X) \\ \left(\frac{\partial}{\partial x} \log p(X, Y) \right)^* \end{pmatrix} \quad \text{et on vérifie que} \quad \mathbb{E}[Z Z^*] = \begin{pmatrix} C & M \\ M^* & J \end{pmatrix} .$$

Compte tenu que la matrice symétrique $\mathbb{E}[Z Z^*]$ est semi-définie positive, il résulte du Lemme A.3 d'inversion matricielle que le complément de Schur $\Delta = C - M J^{-1} M^*$ est également une matrice semi-définie positive, c'est-à-dire que

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

Remarque 1.5 Par définition de l'estimateur MMSE, on a nécessairement

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

pour tout estimateur ψ , et la borne la plus à gauche est atteinte pour $\psi = \widehat{\phi}$. La borne donnée par l'estimateur MMSE est donc plus fine que la borne de Cramér–Rao a posteriori, mais aussi plus difficile à calculer : le plus souvent en effet on ne dispose pas de l'expression de l'estimateur MMSE, mais l'expression des matrices J et M est assez facile à obtenir. La borne de Cramér–Rao a posteriori peut même être assez grossière et atteinte par aucun estimateur, et on déduit de l'encadrement ci-dessus que si la borne de Cramér–Rao a posteriori est atteinte, alors elle est nécessairement atteinte pour l'estimateur MMSE $\psi = \widehat{\phi}$.

Exemple 1.6 Soit X et V deux vecteurs aléatoires gaussiens indépendants, de moyenne \bar{X} et 0, et de matrice de covariance Q_X et Q_V , respectivement, et on pose $Y = h(X) + V$. Si les matrices de covariance Q_X et Q_V sont inversibles, alors on a

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

et

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

de sorte que

$$\begin{aligned} -\log p(x, y) &= -\log p(y | x) - \log p(x) \\ &= \frac{1}{2}(y - h(x))^* Q_V^{-1}(y - h(x)) + \frac{1}{2}(x - \bar{X})^* Q_X^{-1}(x - \bar{X}) + \text{cste} , \end{aligned}$$

et

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

d'où l'expression de la matrice d'information de Fisher

$$\begin{aligned} J &= -\mathbb{E}\left[\frac{\partial^2}{\partial x^2} \log p(X, Y)\right] = \mathbb{E}[(h'(X))^* Q_V^{-1} h'(X)] + \mathbb{E}[V^* Q_V^{-1} h''(X)] + Q_X^{-1} \\ &= \mathbb{E}[(h'(X))^* Q_V^{-1} h'(X)] + Q_X^{-1} , \end{aligned}$$

compte tenu que

$$\mathbb{E}[V^* Q_V^{-1} h''(X)] = 0 .$$

Dans le cas particulier où l'application $h(x) = Hx$ est linéaire, on obtient

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

d'après le Lemme A.1 d'inversion matricielle.

1.3 Cadre gaussien

Dans le cas particulier des vecteurs aléatoires gaussiens, le résultat général obtenu ci-dessus peut être précisé de la façon suivante.

Proposition 1.7 *Soit $Z = (X, Y)$ un vecteur aléatoire gaussien de dimension $m + d$, de moyenne et de matrice de covariance*

$$\bar{Z} = \begin{pmatrix} \bar{X} \\ \bar{Y} \end{pmatrix} \quad \text{et} \quad Q_Z = \begin{pmatrix} Q_X & Q_{XY} \\ Q_{YX} & Q_Y \end{pmatrix},$$

respectivement. Si la matrice Q_Y est inversible, alors la distribution de probabilité conditionnelle du vecteur aléatoire X sachant $Y = y$, est une distribution de probabilité gaussienne de moyenne

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

et de matrice de covariance

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

complément de Schur de la matrice Q_Y dans la matrice-bloc Q_Z .

Remarque 1.8 Pour simuler un vecteur aléatoire gaussien de dimension m , de moyenne $\hat{X}(y)$ et de matrice de covariance R , il suffit de simuler un vecteur aléatoire gaussien $Z' = (X', Y')$ de dimension $m + d$, de même moyenne et de même matrice de covariance que $Z = (X, Y)$, et de poser

$$\xi(y) = X' + Q_{XY} Q_Y^{-1} (y - Y').$$

On vérifie en effet que le vecteur aléatoire $\xi(y)$ ainsi défini est gaussien, comme transformation affine du vecteur aléatoire gaussien Z' , de moyenne

$$\mathbb{E}[\xi(y)] = \mathbb{E}[X'] + Q_{XY} Q_Y^{-1} (y - \mathbb{E}[Y']) = \bar{X} + Q_{XY} Q_Y^{-1} (y - \bar{Y}) = \hat{X}(y),$$

et de matrice de covariance

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

compte tenu que

$$\begin{aligned}\xi(y) - \widehat{X}(y) &= X' + Q_{XY} Q_Y^{-1} (y - Y') - (\bar{X} + Q_{XY} Q_Y^{-1} (y - \bar{Y})) \\ &= (X' - \bar{X}) - Q_{XY} Q_Y^{-1} (Y' - \bar{Y}) ,\end{aligned}$$

par différence.

Remarque 1.9 On vérifie aisément que

$$0 \leq R \leq Q_X ,$$

au sens des matrices symétriques (la majoration est immédiate et la minoration résulte du Lemme A.3), c'est-à-dire que l'utilisation de l'information supplémentaire $Y = y$, ne peut que réduire l'incertitude que l'on a sur le vecteur aléatoire X . En outre, la matrice R ne dépend pas de y , et peut donc être calculée avant même de disposer de la valeur prise par l'observation Y .

Remarque 1.10 Soit $\widehat{X} = \widehat{X}(Y)$ l'estimateur du minimum de variance 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 $Z = (X, Y)$.

Remarque 1.11 Si $Y = (Y', Y'')$ où les composantes Y' et Y'' sont indépendantes, alors

$$\bar{Z} = \begin{pmatrix} \bar{X} \\ \bar{Y}' \\ \bar{Y}'' \end{pmatrix} \quad \text{et} \quad Q_Z = \begin{pmatrix} Q_X & Q_{XY'} & Q_{XY''} \\ Q_{Y'X} & Q_{Y'} & 0 \\ Q_{Y''X} & 0 & Q_{Y''} \end{pmatrix} ,$$

et si les matrices $Q_{Y'}$ et $Q_{Y''}$ sont inversibles, alors la distribution de probabilité conditionnelle du vecteur aléatoire X sachant $Y = y$, avec $y = (y', y'')$, est une distribution de probabilité gaussienne de moyenne

$$\begin{aligned}\widehat{X}(y) &= \bar{X} + Q_{XY} Q_Y^{-1} (y - \bar{Y}) \\ &= \bar{X} + \begin{pmatrix} Q_{XY'} & Q_{XY''} \end{pmatrix} \begin{pmatrix} Q_{Y'} & 0 \\ 0 & Q_{Y''} \end{pmatrix}^{-1} \begin{pmatrix} y' - \bar{Y}' \\ y'' - \bar{Y}'' \end{pmatrix} \\ &= \bar{X} + Q_{XY'} Q_{Y'}^{-1} (y' - \bar{Y}') + Q_{XY''} Q_{Y''}^{-1} (y'' - \bar{Y}'') ,\end{aligned}$$

et de matrice de covariance

$$\begin{aligned}
R &= Q_X - Q_{XY} Q_Y^{-1} Q_{YX} \\
&= Q_X - \begin{pmatrix} Q_{XY'} & Q_{XY''} \end{pmatrix} \begin{pmatrix} Q_{Y'} & 0 \\ 0 & Q_{Y''} \end{pmatrix}^{-1} \begin{pmatrix} Q_{Y'X} \\ Q_{Y''X} \end{pmatrix} \\
&= Q_X - Q_{XY'} Q_{Y'}^{-1} Q_{Y'X} - Q_{XY''} Q_{Y''}^{-1} Q_{Y''X} .
\end{aligned}$$

Exemple 1.12 Soit X et V deux vecteurs aléatoires gaussiens indépendants, de moyenne \bar{X} et 0, et de matrice de covariance Q_X et Q_V , respectivement, et on pose $Y = H X + V$. Le vecteur aléatoire $Z = (X, Y)$ est alors gaussien, de moyenne et de matrice de covariance

$$\bar{Z} = \begin{pmatrix} \bar{X} \\ H \bar{X} \end{pmatrix} \quad \text{et} \quad Q_Z = \begin{pmatrix} Q_X & Q_X H^* \\ H Q_X & H Q_X H^* + Q_V \end{pmatrix} ,$$

respectivement. Si la matrice Q_V est inversible, alors a fortiori la matrice $Q_Y = H Q_X H^* + Q_V$ est inversible, et il découle de la Proposition 1.7 que la distribution de probabilité conditionnelle du vecteur aléatoire X sachant Y , est une distribution de probabilité gaussienne de moyenne

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

et de matrice de covariance déterministe

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

complément de Schur de la matrice $Q_Y = H Q_X H^* + Q_V$ dans la matrice-bloc Q_Z . Pour simuler un vecteur aléatoire gaussien de dimension m , de moyenne $\hat{X}(Y)$ et de matrice de covariance R , il découle de la Remarque 1.8 qu'il suffit de simuler deux vecteurs aléatoires gaussiens indépendants X' et V' , de moyenne \bar{X} et 0, et de matrice de covariance Q_X et Q_V , respectivement, c'est-à-dire de même moyenne et de même matrice de covariance que X et V respectivement, et de poser

$$\xi(Y) = X' + Q_X H^* (H Q_X H^* + Q_V)^{-1} (Y - (H X' + V')) .$$

Si en outre la matrice Q_X est inversible, alors il découle du Lemme A.1 d'inversion matricielle que la matrice R est inversible, et

$$R^{-1} = H^* Q_V^{-1} H + Q_X^{-1} = J ,$$

d'après l'expression obtenue dans l'Exemple 1.6 pour la matrice d'information de Fisher. Dans ce cas particulier, la borne de Cramér-Rao a posteriori est donc atteinte, puisque

$$\mathbb{E}[(\hat{X}(Y) - X)(\hat{X}(Y) - X)^*] = R = J^{-1} .$$

Pour finir, on peut montrer directement la relation $J = R^{-1}$ sans utiliser l'expression obtenue dans l'Exemple 1.6. En effet, si la matrice R est inversible, ce qui est garanti dès que les matrices Q_X et Q_Y sont inversibles, alors on a

$$p(x | y) \propto \exp\left\{-\frac{1}{2} (x - \widehat{X}(y))^* R^{-1} (x - \widehat{X}(y))\right\},$$

de sorte que

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

et

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

et on retrouve bien l'expression de la matrice d'information de Fisher

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

PREUVE DE LA PROPOSITION 1.7. 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 calcule facilement la moyenne

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

la matrice de covariance

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

et la matrice de corrélation

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

compte tenu que

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

par différence. 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

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

On reconnaît la fonction caractéristique d'un vecteur aléatoire gaussien de moyenne $\hat{X}(Y)$ et de matrice de covariance R . \square

Conclusion On voit qu'il est important de disposer d'une information *a priori* sur l'état inconnu X_n , par exemple de disposer d'une équation d'état décrivant l'évolution de X_n quand n varie. On peut considérer deux types de modèles :

- les systèmes linéaires gaussiens,
- les chaînes de Markov à espace d'état fini,

et dans chacun de ces deux cas, il est possible de résoudre exactement le problème de filtrage de façon optimale, par la mise en œuvre :

- du filtre de Kalman, dans le cas des systèmes linéaires gaussiens,
- des équations forward–backward de Baum, ou de l'algorithme de Viterbi, dans le cas des chaînes de Markov à état fini.

Ces deux cas peuvent être vus comme des cas particuliers de modèles beaucoup plus généraux :

- les chaînes de Markov à espace d'état quelconque (fini, dénombrable, continu, hybride, etc.),

et dans ce cas il ne sera pas possible de résoudre exactement le problème de filtrage de façon optimale, qui s'exprime pourtant très simplement en termes de distributions de Feynman–Kac, et il faudra avoir recours à la mise en œuvre de méthodes de résolution approchées, en l'occurrence :

- de filtres particulières, c'est-à-dire de méthodes de Monte Carlo avec interaction.

Chapitre 2

Exemples

2.1 Recalage altimétrique de navigation inertielle

Un avion survole une zone dont le relief est connu : la hauteur $h(r)$ du relief en chaque point de coordonnée horizontale r est connue, et enregistrée dans une carte numérique.

Dans la suite, la position horizontale de l'avion est notée r , la position verticale, ou altitude, est notée z , et la vitesse horizontale est notée v . A l'instant 0, la position horizontale initiale de l'avion est r_0 , son altitude initiale est z_0 et sa vitesse horizontale initiale est v_0 . En réalité, l'avion se déplace à l'altitude $z = z_0$ constante et à la vitesse horizontale constante $v = v_0$.

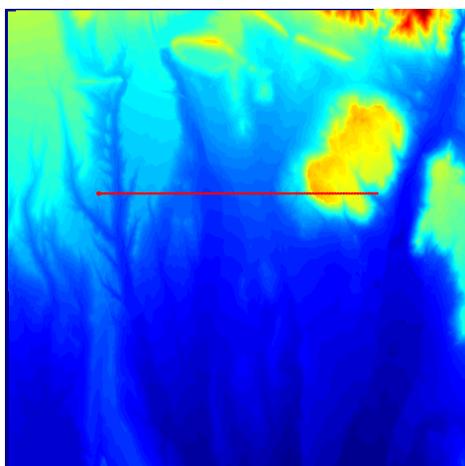


FIGURE 2.1 – Modèle numérique de terrain, et trajectoire réelle

Pour effectuer la navigation, c'est-à-dire pour permettre à l'avion d'estimer sa propre position horizontale r_k et sa propre vitesse horizontale v_k à chaque instant t_k , on recueille (au

moyen d'accéléromètres et de gyroscopes installés à bord) avec un pas de temps $\Delta = t_k - t_{k-1}$ et jusqu'à l'instant final T , l'accélération horizontale de l'avion avec une erreur additive modélisée par un bruit blanc gaussien centré de matrice de covariance $\sigma_{\text{INS}}^2 I_2$ (on dénote par I_2 la matrice identité de dimension 2×2). L'écart-type σ_{INS} est une caractéristique supposée connue de la centrale de navigation inertielle utilisée.

En respectant les caractéristiques statistiques données ci-dessus, la suite a_k^{INS} d'accélération bruitées vérifie

$$a_k^{\text{INS}} = a_k + w_k^{\text{INS}} ,$$

où a_k dénote l'accélération réelle de l'avion, ici $a_k \equiv 0$ compte que l'avion se déplace en réalité à vitesse constante, et où la suite w_k^{INS} est un bruit blanc gaussien centré de matrice de covariance $\sigma_{\text{INS}}^2 I_2$.

L'estimation r_k^{INS} de la position horizontale exacte r_k est obtenue simplement en intégrant les mesures d'accélération horizontale, à l'aide du modèle d'état suivant

$$\begin{pmatrix} r_k^{\text{INS}} \\ v_k^{\text{INS}} \end{pmatrix} = \begin{pmatrix} I_2 & \Delta I_2 \\ 0 & I_2 \end{pmatrix} \begin{pmatrix} r_{k-1}^{\text{INS}} \\ v_{k-1}^{\text{INS}} \end{pmatrix} + \Delta \begin{pmatrix} 0 \\ a_k^{\text{INS}} \end{pmatrix} \quad \text{avec} \quad \begin{pmatrix} r_0^{\text{INS}} \\ v_0^{\text{INS}} \end{pmatrix} = \begin{pmatrix} r_0 \\ v_0 \end{pmatrix} .$$

Si on représente sur le même graphique la position horizontale exacte r_k de l'avion et l'estimation inertielle r_k^{INS} , pour chaque instant entre 0 et T , on remarque que la trajectoire estimée s'écarte de la trajectoire réelle, juste parce que les erreurs sur l'estimation de l'accélération s'accumulent au cours du temps.

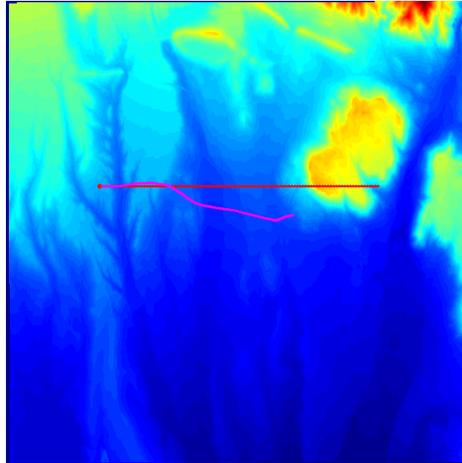


FIGURE 2.2 – Modèle numérique de terrain, trajectoire réelle et trajectoire inertielle

On introduit comme nouvelles variables d'état les erreurs d'estimation inertielle en position horizontale $\delta r_k = r_k^{\text{INS}} - r_k$ et en vitesse horizontale $\delta v_k = v_k^{\text{INS}} - v_k$, et le modèle d'état

correspondant est donc donné par

$$\begin{pmatrix} \delta r_k \\ \delta v_k \end{pmatrix} = \begin{pmatrix} I_2 & \Delta I_2 \\ 0 & I_2 \end{pmatrix} \begin{pmatrix} \delta r_{k-1} \\ \delta v_{k-1} \end{pmatrix} + \Delta \begin{pmatrix} 0 \\ w_k^{\text{INS}} \end{pmatrix} \quad \text{avec} \quad \begin{pmatrix} \delta r_0 \\ \delta v_0 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix},$$

où la suite w_k^{INS} est un bruit blanc gaussien centré de matrice de covariance $\sigma_{\text{INS}}^2 I_2$.

On se propose dans la suite d'estimer ces nouvelles variables d'état, en exploitant d'autres mesures, de manière à corriger les estimations inertielles obtenues lors de cette première phase.

Pour corriger la dérive de l'estimation inertielle en position horizontale r_k^{INS} par rapport à la position horizontale exacte r_k , on recueille séparément (au moyen d'un radar altimétrique, ou radio-altimètre, installé à bord) avec le même pas de temps Δ une mesure d_k^{ALT} de la hauteur de l'avion au-dessus du relief situé à la verticale, avec une erreur additive modélisée par un bruit blanc gaussien centré de variance σ_{ALT}^2 . L'écart-type σ_{ALT} est une caractéristique supposée connue du radio-altimètre utilisé.

On recueille également (au moyen d'un baromètre altimétrique, ou baro-altimètre, installé à bord) avec le même pas de temps Δ une mesure z_k^{BAR} de l'altitude de l'avion, avec une erreur additive modélisée par un bruit blanc gaussien centré de variance σ_{BAR}^2 . L'écart-type σ_{BAR} est une caractéristique supposée connue du baro-altimètre utilisé.

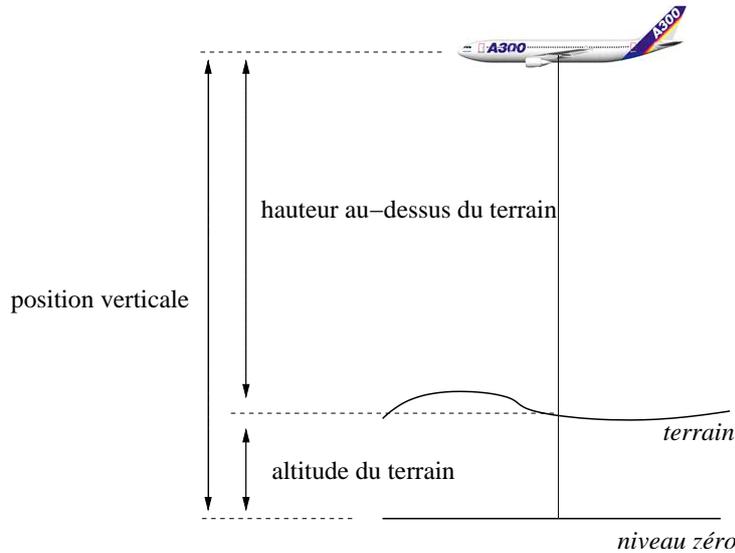


FIGURE 2.3 – Principe du recalage altimétrique

A chaque instant t_k , le radio-altimètre fournit une mesure bruitée d_k^{ALT} de la distance entre l'avion et le relief, c'est-à-dire

$$d_k^{\text{ALT}} = (z_k - h(r_k)) + w_k^{\text{ALT}},$$

où r_k dénote la position horizontale réelle de l'avion, où z_k dénote l'altitude réelle de l'avion, où $h(r_k)$ dénote la hauteur du relief au point de coordonnée horizontale r_k , et où la suite w_k^{ALT} est

un bruit blanc gaussien centré de variance σ_{ALT}^2 . Au même instant t_k , le baro–altimètre fournit une mesure bruitée z_k^{BAR} de l’altitude de l’avion, c’est-à-dire

$$z_k^{\text{BAR}} = z_k + w_k^{\text{BAR}} ,$$

où z_k dénote l’altitude réelle de l’avion, et où la suite w_k^{BAR} est un bruit blanc gaussien centré de variance σ_{BAR}^2 . La hauteur du relief survolé à l’instant t_k déduite à partir des mesures fournies par le radio–altimètre et par le baro–altimètre est donc

$$h_k^{\text{ALT}} = z_k^{\text{BAR}} - d_k^{\text{ALT}} = h(r_k) + w_k^{\text{BAR}} - w_k^{\text{ALT}} ,$$

et peut être reliée à l’erreur de position inertielle horizontale δr_k par

$$h_k^{\text{ALT}} = h(r_k^{\text{INS}} - \delta r_k) + w_k^{\text{BAR}} - w_k^{\text{ALT}} .$$

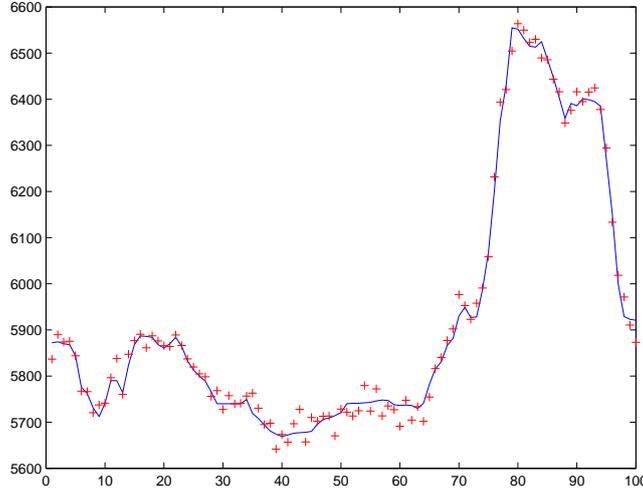


FIGURE 2.4 – Profil réel du terrain survolé et mesures altimétriques

En résumé, le modèle d’état utilisé pour le recalage altimétrique de navigation inertielle comprend :

- l’équation d’état

$$\begin{pmatrix} \delta r_k \\ \delta v_k \end{pmatrix} = \begin{pmatrix} I_2 & \Delta I_2 \\ 0 & I_2 \end{pmatrix} \begin{pmatrix} \delta r_{k-1} \\ \delta v_{k-1} \end{pmatrix} + \Delta \begin{pmatrix} 0 \\ w_k^{\text{INS}} \end{pmatrix} ,$$

où la suite w_k^{INS} est un bruit blanc gaussien centré de variance $\sigma_{\text{INS}}^2 I_2$,

- la condition initiale

$$\begin{pmatrix} \delta r_0 \\ \delta v_0 \end{pmatrix} \text{ gaussienne, centrée, de matrice de covariance } \begin{pmatrix} \sigma_{r_0}^2 I_2 & 0 \\ 0 & \sigma_{v_0}^2 I_2 \end{pmatrix} ,$$

- et l'équation d'observation

$$h_k^{\text{ALT}} = h(r_k^{\text{INS}} - \delta r_k) + w_k^{\text{BAR}} - w_k^{\text{ALT}} .$$

où la suite w_k^{ALT} est un bruit blanc gaussien centré de variance σ_{ALT}^2 , et où la suite w_k^{BAR} est un bruit blanc gaussien centré de variance σ_{BAR}^2 .

L'estimation inertielle horizontale r_k^{INS} fournie par la centrale inertielle, et la mesure h_k^{ALT} de la hauteur du relief fournie par le radio–altimètre et par le baro–altimètre sont disponibles. La fonction $r \mapsto h(r)$ n'est pas connue de façon analytique, mais définie *point–par–point* en allant lire la carte numérique.

2.2 Suivi visuel par histogramme de couleur

On souhaite réaliser un algorithme de suivi dans une séquence d'images numériques couleur. A la lecture de la première image de la séquence, l'utilisateur sélectionne une zone de l'image, et le suivi s'effectue de façon séquentielle sur l'ensemble de la séquence, voir Figure 2.5.

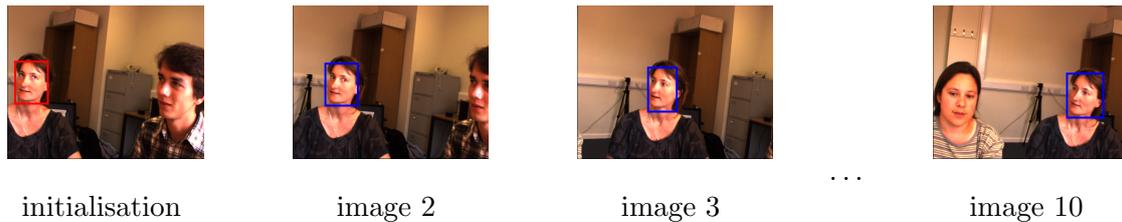


FIGURE 2.5 – Suivi d'un visage dans une séquence de 10 images

La méthode est construite sur l'algorithme SIR (souvent appelé algorithme CONDENSATION, pour *conditional density propagation*, en vision par ordinateur). Elle repose sur l'hypothèse que l'histogramme de couleur de la zone à suivre est constant le long de la séquence. Pour avoir plus d'informations sur cette méthode de suivi visuel, on pourra lire [8].

Introduction aux images numériques

On désigne sous le terme d'image numérique toute image (dessin, icône, photographie, etc.) acquise, créée, traitée ou stockée sous forme binaire. On distingue généralement deux grandes catégories d'images :

- les images vectorielles, dont la description informatique est composée d'objets géométriques individuels (segments de droite, polygones, arcs de cercle, etc.), chacun définis par divers attributs de forme, de position, de couleur, etc.
- les images matricielles, représentées par un tableau à deux dimensions dont chaque case est un pixel (mot dérivé de l'anglais *picture element*, élément d'image). A chaque pixel est associée une ou plusieurs valeurs décrivant son niveau de gris ou sa couleur.

Les images vectorielles sont utilisées essentiellement pour du graphisme ou en CAO. Lorsque l'on s'intéresse au traitement d'images et à la vision par ordinateur, la représentation utilisée est la forme matricielle. Il existe plusieurs standards de codage de la couleur :

bitmap noir et blanc : en stockant un bit dans chaque case, il est possible de définir deux couleurs (noir ou blanc).

bitmap 256 niveaux de gris : en stockant un octet dans chaque case, il est possible de définir 256 dégradés de gris allant du noir au blanc

palette de couleurs (colormap) : grâce à cette méthode, il est possible de définir une palette, ou table des couleurs, contenant l'ensemble des couleurs pouvant être contenues dans l'image, à chacune desquelles est associé un indice. Le nombre de bits réservé au codage de chaque indice de la palette détermine le nombre de couleurs pouvant être utilisées. On appelle ainsi image en couleurs indexées une image dont les couleurs sont codées selon cette technique.

couleurs vraies (true color) : le codage de la couleur est réalisé sur trois octets, chaque octet représentant la valeur d'une composante couleur par un entier de 0 à 255. Ces trois valeurs codent généralement la couleur dans l'espace RVB (rouge, vert, bleu), mais d'autres espaces de couleurs peuvent être utilisés. Le nombre de couleurs différentes pouvant être ainsi représentées est de $256 \times 256 \times 256$ possibilités, soit près de 16 millions de couleurs.

Une image numérique est avant tout un signal 2D. D'un point de vue mathématique, on considère l'image comme une fonction de $\mathbb{R} \times \mathbb{R}$ dans Ω où le couplet d'entrée est une position spatiale sur la grille des pixels, et où Ω est l'espace des valeurs de codage de la couleur (ou du niveau de gris). Par extension, on parlera d'images en dimension $2D+t$ (t pour le temps) pour désigner une séquence d'images numériques (ou vidéo numérique).

Remarque 2.1 Etant donné que l'écran effectue un balayage de gauche à droite et de haut en bas, on désigne généralement par les coordonnées $(0, 0)$ le pixel situé en haut à gauche de l'image, ce qui signifie que les axes de l'image sont orientés de la façon suivante : l'axe X est orienté de gauche à droite, l'axe Y est orienté de haut en bas, contrairement aux notations conventionnelles en mathématiques, où l'axe Y est orienté vers le haut.

Principe de l'algorithme de suivi visuel

Le but de cet algorithme est de suivre une région d'intérêt dans une séquence d'images. Cette région est initialisée par l'utilisateur et sa forme est fixée a priori. On considèrera ici un rectangle, paramétré par la position, en pixel, du centre du rectangle $d = (x, y)$ et un paramètre d'échelle s . Au pas de temps k (i.e. à l'image k), l'état du système à estimer sera donc $X_k = (d_k, s_k)$. Le paramètre d'échelle permet de suivre un objet même si celui-ci avance ou s'éloigne dans l'axe de la caméra (effet de zoom). A l'initialisation, l'utilisateur clique 4 points dans l'image, qui vont définir le rectangle initial. Celui-ci est décrit par les coordonnées du point haut/gauche, une largeur et une hauteur.

Équation d'état On s'intéresse à la situation où aucune information a priori n'est disponible sur la nature de l'objet suivi. Dans ce cas, l'équation dynamique du système doit être peu informative. On supposera donc un modèle à position constante

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

où W_k est un bruit blanc gaussien, centré en 0 et de matrice de covariance C , matrice 3×3 diagonale. Les valeurs sur la diagonale sont c_1 , c_2 et c_3 . Notons que si la nature de l'objet suivi est connu, il est plus intéressant d'utiliser un modèle dynamique approprié. Par exemple, on pourrait imaginer utiliser un modèle à vitesse constante pour le suivi d'une voiture dans une vidéo acquise par une caméra sur autoroute.

Modèle de couleur La zone initiale à suivre est caractérisée par un histogramme de couleur. Cet histogramme de référence est construit sur les Nb couleurs les plus représentatives de cette zone, comme montré sur la Figure 2.6. Cet histogramme de référence est noté $q^* = \{q^*(n), n = 1, \dots, Nb\}$, où $q^*(n)$ représente le nombre normalisé de pixels de la zone initiale dont la couleur la plus proche est la couleur n . On a $\sum_{n=1}^{Nb} q^*(n) = 1$. Pour plus d'informations sur les différents espaces de couleur, on pourra se reporter à la page *color space* sous Wikipedia.

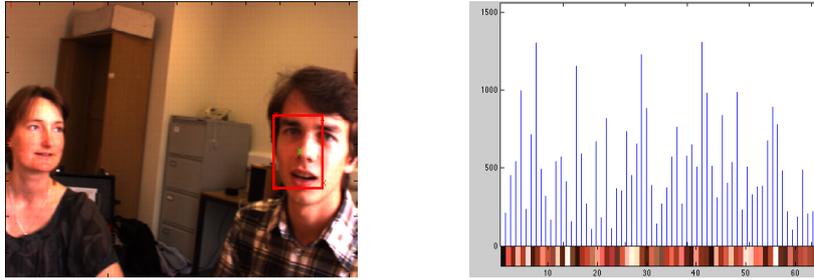


FIGURE 2.6 – Zone de l'image à suivre et histogramme de couleur associé pour $Nb = 64$

Comme décrit précédemment, le but est de suivre une zone de l'image le long de la séquence, sous l'hypothèse que son histogramme de couleur est invariant dans le temps. Au temps k , l'histogramme de couleur $q_k(x)$ d'un état hypothèse x sera comparé au modèle de couleur de référence q^* , et on définit la mesure de distance D entre ces deux histogrammes de couleur normalisés

$$D(q^*, q_k(x)) = \left(1 - \sum_{n=1}^{Nb} \sqrt{q^*(n) q_k(x, n)}\right)^{1/2} ,$$

Pour favoriser les états hypothèses dont l'histogramme de couleur associé est proche de l'histogramme de référence, on introduit la fonction de pondération

$$g_k(x) \propto \exp\{-\lambda D^2(q^*, q_k(x))\} .$$

2.3 Poursuite d'une cible furtive (track-before-detect)

Une image radar est constituée par un tableau rectangulaire de $p \times p$ pixels, où l'intensité de l'écho recueilli en un point est codée par un niveau de gris allant du plus foncé (écho de faible intensité) au plus clair (écho de forte intensité). La même situation se rencontre avec un dispositif opto-électronique, comme une caméra matricielle, où chaque pixel reçoit et affiche une intensité lumineuse différente. En principe, si une cible est présente dans la scène 3D visée, elle apparaîtra dans le plan-image sous la forme d'un pixel plus clair (ou d'un groupe de pixels adjacents plus clairs) que les autres pixels de l'image, lesquels correspondent à l'écho d'objets secondaires de moindre intensité et/ou à un bruit spatial, indépendant ou bien spatialement corrélé d'un pixel à l'autre. Pour détecter (et localiser) la cible, il suffit en principe de rechercher dans l'image le pixel (ou le groupe de pixels adjacents) le plus clair, c'est-à-dire de plus forte intensité. Au lieu d'une recherche exhaustive, on utilise souvent une méthode de seuillage : rechercher les pixels d'intensité supérieure à un seuil bien choisi, permet souvent d'obtenir directement le pixel de plus forte intensité. En répétant cette opération pour chaque image successivement on peut ainsi détecter d'abord, puis suivre, la cible dans une séquence d'images.

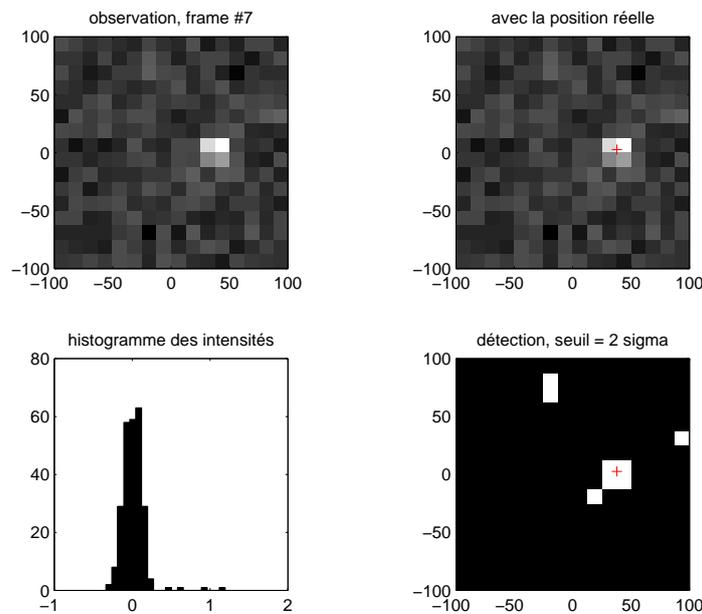


FIGURE 2.7 – Image observée, position réelle, histogramme, détection (cible visible)

On s'intéresse ici au cas d'une cible furtive, caractérisée par un écho de très faible intensité, c'est-à-dire d'une intensité du même ordre de grandeur que l'intensité caractéristique du bruit présent dans l'image, voire même d'un ordre de grandeur inférieur. Dans ce cas, une méthode de seuillage est inefficace : quel que soit le seuil choisi, rechercher les pixels d'intensité supérieure au seuil ne permet plus d'isoler la cible au milieu du bruit. Même un opérateur humain est incapable, sur une image isolée, de détecter la présence et la position de la cible. En revanche, un opérateur

humain est capable dans certains cas de suivre la cible dans une séquence d'images, comme une succession de pixels (un dans chaque image de la séquence) animés d'un mouvement cohérent au milieu de l'agitation incoordonnée des autres pixels. En quelque sorte, l'œil humain suit la cible sans jamais la détecter vraiment : c'est ce genre de performance qu'il s'agit de reproduire ici de manière algorithmique, connue sous le terme de *track-before-detect*, en s'appuyant sur un modèle a priori pour le déplacement de la cible, qui favorise le mouvement cohérent de pixels entre des images successives.

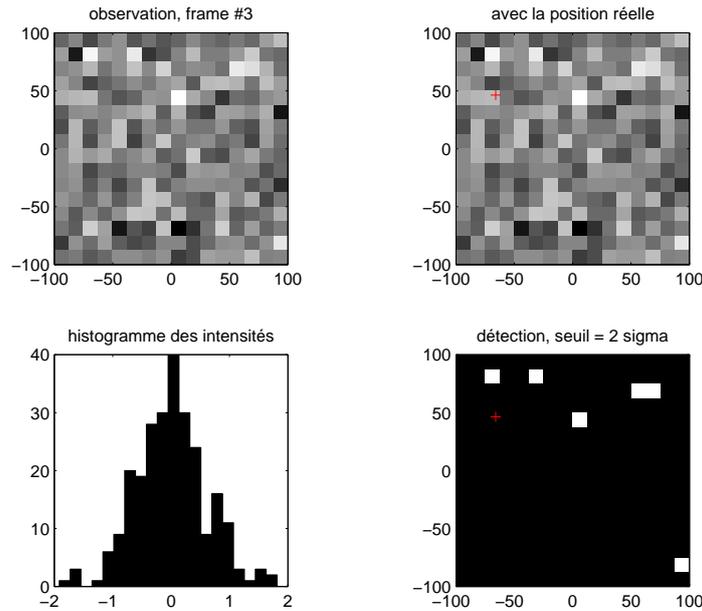


FIGURE 2.8 – Image observée, position réelle, histogramme, détection (cible furtive)

Chaque image peut se représenter comme un champ aléatoire $(Y_k(s), s \in S)$ où l'indice $s \in S$ désigne le pixel ou de manière équivalente le site d'un réseau bi-dimensionnel. Par hypothèse, l'intensité observée au pixel $s \in S$ se décompose comme

$$Y_k(s) = I(r_k, s) + B_k(s) ,$$

c'est-à-dire comme la somme de l'intensité due à la présence de la cible à la position (inconnue) r_k et de l'intensité due au bruit seulement. L'intensité au point $s \in S$ due à la présence de la cible à la position r est modélisée par une fonction d'étalement ponctuelle (ou *point spread function*, PSF)

$$I(r, s) = I_0 \frac{\delta^2}{2\pi\sigma_{\text{PSF}}^2} \exp\left\{-\frac{|r(s) - r|^2}{2\sigma_{\text{PSF}}^2}\right\} 1_{(s \in C(r))} ,$$

où $r(s)$ désigne la position dans l'espace physique du centre du pixel s , où $\delta > 0$ désigne la taille du pixel dans l'espace physique, et où l'ensemble $C(r)$ désigne le voisinage à 9 points dans l'espace-image autour du pixel contenant le point de position r dans l'espace physique.

L'intensité due au bruit seulement est modélisée comme un champ aléatoire gaussien ($B_k(s)$, $s \in S$) centré, de variance σ_B^2 en tout pixel $s \in S$ et décorrélié spatialement, c'est-à-dire

$$\mathbb{E}[B_k(s)] = 0 \quad \text{et} \quad \mathbb{E}[B_k(s) B_k(s')] = \sigma_B^2 \mathbf{1}_{(s = s')} .$$

On définit le rapport signal à bruit (ou *signal to noise ratio*, SNR) en decibel, comme

$$\text{SNR} = 20 \log_{10} \frac{I_0}{\sigma_B} .$$

On pose ici (par convention) $I_0 = 1$ de sorte qu'un rapport signal à bruit de 20 dB correspond à $\sigma_B = 0.1$ tandis qu'un rapport signal à bruit de 0 dB correspond à $\sigma_B = 1$.

La fonction de vraisemblance est donnée à une constante multiplicative près par l'expression, en fonction de la variable r , de la densité du champ aléatoire observé ($Y_k(s)$, $s \in S$) quand la cible occupe la position r dans l'espace physique. On a donc par définition

$$g_k(r) = \exp\left\{-\frac{1}{2\sigma_B^2} \sum_{s \in S} |Y_k(s) - I(r, s)|^2\right\} ,$$

et on remarque que

$$g_k(r) = \exp\left\{\frac{1}{\sigma_B^2} \sum_{s \in C(r)} I(r, s) Y_k(s) - \frac{1}{2\sigma_B^2} \sum_{s \in C(r)} |I(r, s)|^2\right\} ,$$

à une constante multiplicative près, de sorte que le calcul porte seulement sur les 9 pixels du voisinage $C(r)$, et pas sur l'ensemble S de tous les pixels.

Le modèle a priori pour l'évolution de la cible est donné par le modèle d'état suivant

$$\begin{pmatrix} r_k \\ v_k \end{pmatrix} = \begin{pmatrix} I_2 & \Delta I_2 \\ 0 & I_2 \end{pmatrix} \begin{pmatrix} r_{k-1} \\ v_{k-1} \end{pmatrix} + \sigma \sqrt{\Delta} \begin{pmatrix} 0 \\ w_k \end{pmatrix} ,$$

où w_k est un vecteur aléatoire gaussien centré de matrice de covariance I_2 , où la position initiale r_0 est distribuée uniformément dans l'espace physique défini ci-dessus, et où la vitesse initiale v_0 est distribuée uniformément dans le domaine délimité en module par $v_{\min} \leq |v_0| \leq v_{\max}$ et en orientation par $[0, 2\pi)$.

Sur chaque image, on peut rechercher le pixel de plus forte intensité observée, ou bien mettre en œuvre une méthode de seuillage pour détecter les pixels d'intensité supérieure au seuil choisi. On peut aussi extraire l'histogramme des intensités observées aux différents pixels de l'image. Si le rapport signal à bruit est trop faible, alors une simple détection image par image s'avère inefficace. On peut en revanche considérer les images successives comme des observations (matricielles), et mettre en œuvre un algorithme de filtrage pour effectuer directement le suivi.

2.4 Navigation en environnement intérieur

Un utilisateur se déplace à l'intérieur d'un bâtiment dont le plan est disponible sous la forme d'une carte numérique. L'utilisateur est caractérisé à l'instant t_k

- par sa position r_k (un point du plan hors des zones noires),
- par son orientation θ_k (un vecteur unitaire, ou un angle) par rapport à la direction de référence correspondant au vecteur unitaire $u = (1, 0)$, dirigé vers la droite sur la carte.

Un exemple de trajectoire admissible, c'est-à-dire ne rencontrant pas les obstacles (représentés par les zones noires), est représenté sur la Figure 2.9.

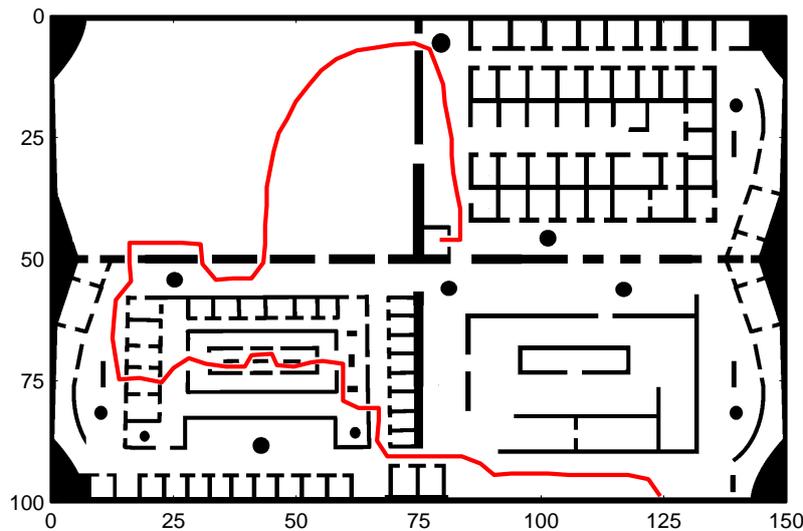


FIGURE 2.9 – Exemple de trajectoire admissible

Le segment numéro k , joignant les positions r_k et r_{k+1} occupées par l'utilisateur aux instants t_k et t_{k+1} respectivement, peut être caractérisé de la manière équivalente

- par son origine r_k , qui s'interprète comme la position de l'utilisateur à l'instant t_k ,
- par sa longueur $d_k = |r_{k+1} - r_k|$, qui peut s'interpréter comme la distance parcourue par l'utilisateur entre les instants t_k et t_{k+1} ,
- et par son orientation θ_k (déjà mentionnée), qui peut être définie de manière équivalente par le vecteur unitaire $u_k = (r_{k+1} - r_k)/d_k$,

et on dénote par $\alpha_k = \theta_k - \theta_{k-1}$ le changement d'orientation entre le segment numéro $(k - 1)$ et le segment numéro k , qui peut s'interpréter comme une rotation effectuée par l'utilisateur à l'instant t_k .

Pour effectuer la navigation, c'est-à-dire pour permettre à l'utilisateur d'estimer sa propre position à chaque instant, celui-ci est équipé d'un module de navigation à l'estime (ou module PNS, pour *pedestrian navigation system*), qui fournit

- une mesure $\hat{\alpha}_k$ de la rotation effectuée par l'utilisateur à l'instant t_k , avec une incertitude caractérisée par un bruit gaussien additif de moyenne nulle et de variance σ_{turn}^2 ,
- et une mesure \hat{d}_k de la distance parcourue par l'utilisateur entre les instants t_k et t_{k+1} , avec une incertitude caractérisée par un bruit gaussien additif de moyenne nulle et de variance σ_{walk}^2 .

En d'autres termes

$$\hat{\alpha}_k = \alpha_k + w_k^{\text{turn}} \quad (\text{modulo } 2\pi) \quad \text{et} \quad \hat{d}_k = d_k + w_k^{\text{walk}}, \quad (2.1)$$

où w_k^{walk} et w_k^{turn} sont deux variables aléatoires gaussiennes indépendantes, de moyenne nulle et de variance σ_{walk}^2 et σ_{turn}^2 , respectivement.

Les mesures bruitées $\hat{d}_1, \dots, \hat{d}_{n_{\text{max}}-1}$ et $\hat{\alpha}_1, \hat{\alpha}_2, \dots, \hat{\alpha}_{n_{\text{max}}-1}$ (avec la convention $\hat{\alpha}_1 = 0$) sont recueillies par l'utilisateur le long de la trajectoire. À partir de ces mesures PNS incrémentales bruitées, et à partir d'estimations de la position initiale r_1 et de l'orientation initiale θ_1 inconnues, on peut essayer de reconstruire la position et l'orientation de l'utilisateur à chaque instant, par intégration

$$\theta_k^{\text{PNS}} = \theta_{k-1}^{\text{PNS}} + \hat{\alpha}_k \quad (\text{modulo } 2\pi) \quad \text{et} \quad r_{k+1}^{\text{PNS}} = r_k^{\text{PNS}} + \hat{d}_k u(\theta_k^{\text{PNS}}),$$

où $u(\theta) = (\cos \theta, \sin \theta)$ désigne le vecteur unitaire associé à l'angle θ . La trajectoire estimée à partir des mesures PNS seulement est représentée sur la Figure 2.10.

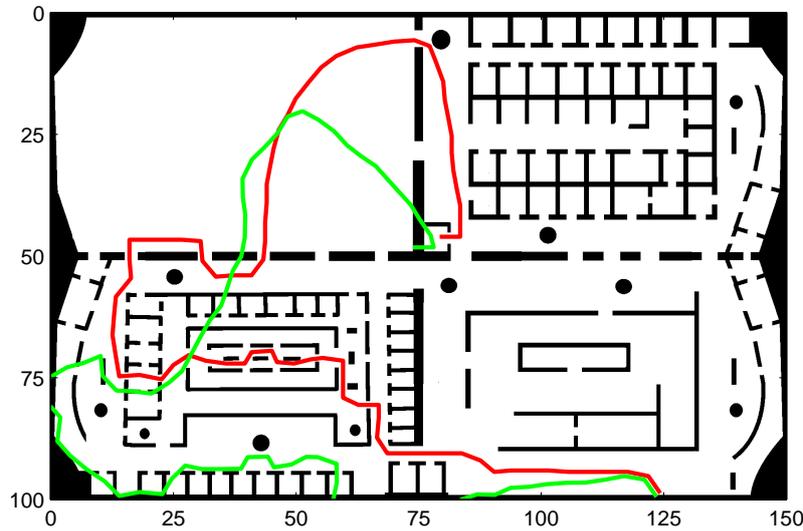


FIGURE 2.10 – Trajectoire estimée à partir des mesures PNS seulement

On remarque que la trajectoire estimée s'écarte de la trajectoire réelle, juste parce que les erreurs sur les mesures PNS incrémentales s'accroissent au cours du temps.

Pour corriger la dérive de la trajectoire estimée à partir des mesures PNS seulement, l'idée consiste à recueillir séparément des mesures fournies par d'autres capteurs. Dans la solution proposée ici, à l'intérieur du bâtiment sont disposées des balises de *ranging* identiques, dont les positions sont connues. Chaque balise est caractérisée par sa portée R , de sorte que

- tout utilisateur se trouvant à une distance inférieure à R par rapport à une balise est détecté par cette balise,
- et symétriquement, tout utilisateur se trouvant à une distance supérieure à R par rapport à une balise n'est pas détecté par cette balise.

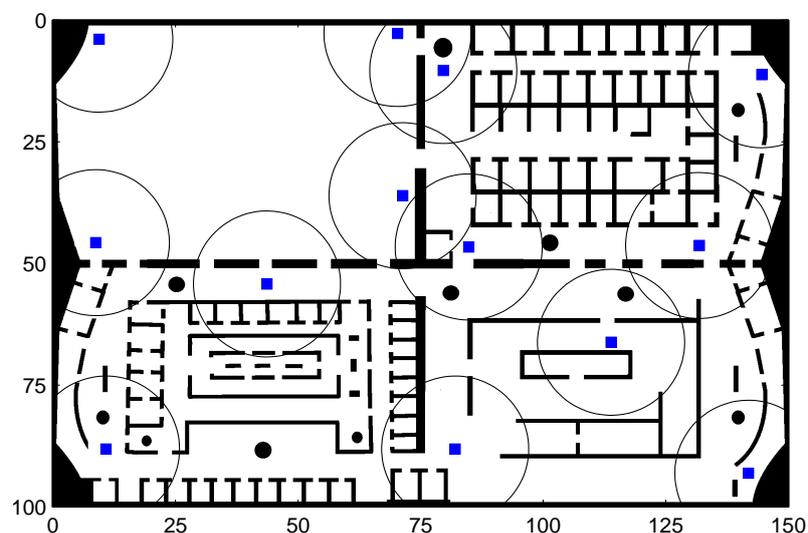


FIGURE 2.11 – Balises de *ranging* à portée limitée

En outre, si une balise détecte un utilisateur alors une mesure de la distance entre l'utilisateur et cette balise est également disponible, avec une incertitude caractérisée par un bruit gaussien additif de moyenne nulle et de variance σ_{range}^2 . Les éventuelles détections et mesures de distance bruitées sont recueillies par l'utilisateur le long de la trajectoire, et sont disponibles pour le reclage de navigation.

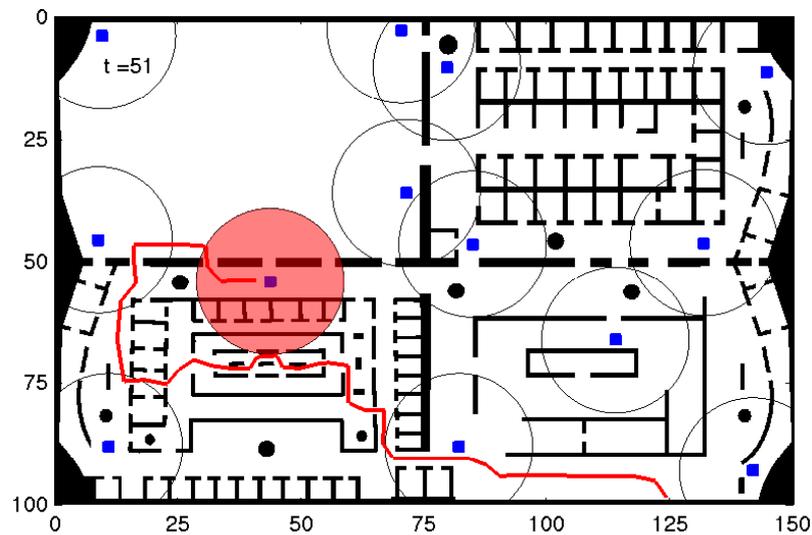


FIGURE 2.12 – Détection par une balise de *ranging*

Pour réaliser le recalage de navigation, on dispose des informations suivantes

- un modèle *a priori* pour l'évolution de la position et de l'orientation de l'utilisateur, utilisant les mesures PNS incrémentales bruitées définies en (2.1),
- une fonction de vraisemblance associée à chaque balise active, c'est-à-dire à chaque balise déclenchée par l'utilisateur,

et on peut également prendre en compte

- la détection (ou la non-détection) de l'utilisateur par une balise,
- et les contraintes sur l'évolution de l'utilisateur dues à la présence d'obstacles, typiquement les murs et cloisons intérieures du bâtiment, information disponible à partir de la carte numérique.

Chapitre 3

Filtrage de Kalman

Le problème de filtrage (en temps discret) se présente en général de la manière suivante : on considère $\{X_k\}$, un processus (dont les caractéristiques statistiques sont connues) représentant l'état d'un système non observé. A l'instant k , on recueille une observation Y_k qui est formée d'un signal (i.e. une fonction $h(X_k)$ de l'état X_k) et d'un bruit additif

$$Y_k = h(X_k) + V_k .$$

Les caractéristiques statistiques du bruit de mesure $\{V_k\}$ sont également supposées connues. A l'instant k , on dispose de l'information $Y_{0:k} = (Y_0, \dots, Y_k)$ et le but est d'obtenir *le plus d'information possible* sur l'état du système X_k (on veut, par exemple, pouvoir calculer un estimateur \hat{X}_k de X_k). On a vu à la Section 1.2 que la solution consiste à calculer la distribution de probabilité conditionnelle de la variable aléatoire X_k sachant $Y_{0:k}$.

Dans le cas des systèmes décrits à la Section 3.1, le cadre est gaussien et l'évolution de cette distribution de probabilité conditionnelle (déterminée par sa moyenne et sa matrice de covariance) est régie par les équations du filtre de Kalman, présentées à la Section 3.2 et très simples à mettre en œuvre. Les techniques développées dans le cas linéaire peuvent parfois s'étendre au cas non linéaire par des méthodes de linéarisation, présentées à la Section 4.1. Les filtres ainsi obtenus sont très souvent utilisés en pratique mais ont parfois des performances peu satisfaisantes.

3.1 Systèmes linéaires gaussiens

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

$$X_k = F_k X_{k-1} + f_k + G_k W_k , \tag{3.1}$$

où $\{X_k\}$ et $\{W_k\}$ prennent respectivement leurs valeurs dans \mathbb{R}^m et \mathbb{R}^p , et une suite d'observations $\{Y_k\}$ à valeurs dans \mathbb{R}^d , vérifiant

$$Y_k = H_k X_k + h_k + V_k , \tag{3.2}$$

et on suppose que

- la condition initiale X_0 est gaussienne, de moyenne \bar{X}_0 et de matrice de covariance Q_0^X ,
- la suite $\{W_k\}$ est un bruit blanc gaussien, de matrice de covariance Q_k^W ,
- la suite $\{V_k\}$ est un bruit blanc gaussien, de matrice de covariance Q_k^V ,
- les suites $\{W_k\}$ et $\{V_k\}$ et la condition initiale X_0 sont mutuellement indépendants.

La signification du modèle (3.1) est la suivante

- 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 $G_k Q_k^W G_k^*$,
- 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, c'est-à-dire même si $G_k = 0$, 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 3.1 *La suite $\{Z_k = (X_k, Y_k)\}$ est un processus aléatoire gaussien à valeurs dans \mathbb{R}^{m+d} . En particulier à l'instant k , le vecteur aléatoire Z_k est gaussien, de moyenne et de matrice de covariance*

$$\begin{pmatrix} \bar{X}_k \\ \bar{Y}_k \end{pmatrix} \quad \text{et} \quad \begin{pmatrix} Q_k^X & Q_k^{XY} \\ Q_k^{YX} & Q_k^Y \end{pmatrix},$$

respectivement, avec

$$\bar{X}_k = F_k \bar{X}_{k-1} + f_k \quad \text{et} \quad \bar{Y}_k = H_k \bar{X}_k + h_k,$$

et

$$Q_k^X = F_k Q_{k-1}^X F_k^* + G_k Q_k^W G_k^*, \quad Q_k^{XY} = Q_k^X H_k^* \quad \text{et} \quad Q_k^Y = H_k Q_k^X H_k^* + Q_k^V.$$

PREUVE. Comme sortie d'un système linéaire à entrées gaussiennes, la suite $\{Z_k\}$ est un processus aléatoire gaussien. En effet, pour tout instant n , le vecteur aléatoire (Z_0, Z_1, \dots, Z_n) peut s'exprimer comme transformation affine du vecteur aléatoire $(X_0, W_1, \dots, W_n, V_0, V_1, \dots, V_n)$ qui par hypothèse est un vecteur aléatoire gaussien, donc le vecteur aléatoire (Z_0, Z_1, \dots, Z_n) est gaussien, comme transformation affine d'un vecteur aléatoire gaussien. Par ailleurs, d'après (3.1)

$$\bar{X}_k = \mathbb{E}[X_k] = F_k \mathbb{E}[X_{k-1}] + f_k + G_k \mathbb{E}[W_k] = F_k \bar{X}_{k-1} + f_k,$$

et d'après (3.2)

$$\bar{Y}_k = \mathbb{E}[Y_k] = H_k \mathbb{E}[X_k] + h_k + \mathbb{E}[V_k] = H_k \bar{X}_k + h_k.$$

Par différence

$$X_k - \bar{X}_k = F_k (X_{k-1} - \bar{X}_{k-1}) + G_k W_k \quad \text{et} \quad Y_k - \bar{Y}_k = H_k (X_k - \bar{X}_k) + V_k,$$

de sorte que

$$\begin{aligned}
Q_k^X &= \mathbb{E}[(X_k - \bar{X}_k)(X_k - \bar{X}_k)^*] \\
&= \mathbb{E}[(F_k(X_{k-1} - \bar{X}_{k-1}) + G_k W_k)(F_k(X_{k-1} - \bar{X}_{k-1}) + G_k W_k)^*] \\
&= F_k \mathbb{E}[(X_{k-1} - \bar{X}_{k-1})(X_{k-1} - \bar{X}_{k-1})^*] F_k^* + G_k \mathbb{E}[W_k W_k^*] G_k^* \\
&\quad + G_k \mathbb{E}[W_k(X_{k-1} - \bar{X}_{k-1})^*] F_k^* + F_k \mathbb{E}[(X_{k-1} - \bar{X}_{k-1}) W_k^*] G_k^* \\
&= F_k Q_{k-1}^X F_k^* + G_k Q_k^W G_k^* ,
\end{aligned}$$

où on a utilisé dans la dernière égalité le fait que $(X_{k-1} - \bar{X}_{k-1})$ est indépendant de W_k , donc $\mathbb{E}[(X_{k-1} - \bar{X}_{k-1}) W_k^*] = 0$. Par ailleurs

$$\begin{aligned}
Q_k^{XY} &= \mathbb{E}[(X_k - \bar{X}_k)(Y_k - \bar{Y}_k)^*] \\
&= \mathbb{E}[(X_k - \bar{X}_k)(H_k(X_k - \bar{X}_k) + V_k)^*] \\
&= \mathbb{E}[(X_k - \bar{X}_k)(X_k - \bar{X}_k)^*] H_k^* + \mathbb{E}[(X_k - \bar{X}_k) V_k^*] \\
&= Q_k^X H_k^* ,
\end{aligned}$$

où on a utilisé dans la dernière égalité le fait que $(X_k - \bar{X}_k)$ est indépendant de V_k , donc $\mathbb{E}[(X_k - \bar{X}_k) V_k^*] = 0$. Finalement

$$\begin{aligned}
Q_k^Y &= \mathbb{E}[(Y_k - \bar{Y}_k)(Y_k - \bar{Y}_k)^*] \\
&= \mathbb{E}[(H_k(X_k - \bar{X}_k) + V_k)(H_k(X_k - \bar{X}_k) + V_k)^*] \\
&= H_k \mathbb{E}[(X_k - \bar{X}_k)(X_k - \bar{X}_k)^*] H_k^* + \mathbb{E}[V_k V_k^*] \\
&\quad + \mathbb{E}[V_k(X_k - \bar{X}_k)^*] H_k^* + H_k \mathbb{E}[(X_k - \bar{X}_k) V_k^*] \\
&= H_k Q_k^X H_k^* + Q_k^V ,
\end{aligned}$$

où on a encore utilisé dans la dernière égalité le fait que $(X_k - \bar{X}_k)$ est indépendant de V_k , donc $\mathbb{E}[(X_k - \bar{X}_k) V_k^*] = 0$. \square

3.2 Filtre de Kalman

On considère un système linéaire du type (3.1) (3.2), c'est-à-dire

$$X_k = F_k X_{k-1} + f_k + G_k W_k, \quad (3.3)$$

$$Y_k = H_k X_k + h_k + V_k, \quad (3.4)$$

avec les hypothèses faites à la Section 3.1. A l'instant k , on dispose de l'information

$$Y_{0:k} = (Y_0, Y_1, \dots, Y_k).$$

L'objectif est d'estimer de façon optimale et récursive le vecteur aléatoire X_k à partir de $Y_{0:k}$. Si on adopte le critère du minimum de variance, il s'agit d'après la Section 1.2 de calculer la distribution de probabilité conditionnelle du vecteur aléatoire X_k sachant $Y_{0:k}$. Comme le cadre est gaussien, il suffit de calculer la moyenne et la matrice de covariance

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

On définit également les quantités suivantes

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

D'après la Remarque 1.9, les matrices de covariances conditionnelles P_k et P_k^- ne dépendent pas des observations, c'est-à-dire que

$$P_k = \mathbb{E}[(X_k - \widehat{X}_k)(X_k - \widehat{X}_k)^*] \quad \text{et} \quad P_k^- = \mathbb{E}[(X_k - \widehat{X}_k^-)(X_k - \widehat{X}_k^-)^*].$$

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 (3.3),
- dans l'étape de *correction*, on utilise la nouvelle observation Y_k , et en particulier, on considère la composante de l'observation Y_k qui apporte une information nouvelle par rapport aux observations passées $Y_{0:k-1}$, c'est-à-dire

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

et d'après (3.4), on a

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

compte tenu que V_k et $Y_{0:k-1}$ sont indépendants.

Remarque 3.2 Par définition, toute fonction des variables $(Y_0, \dots, Y_{k-1}, Y_k)$ peut s'exprimer en fonction des variables $(Y_0, \dots, Y_{k-1}, I_k)$, et réciproquement. On en déduit que $(Y_{0:k-1}, I_k)$ contient exactement la même information que $Y_{0:k}$.

Lemme 3.3 *Le processus $\{I_k\}$ est un processus gaussien à valeurs dans \mathbb{R}^d , appelé processus d'innovation. En particulier, 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 vecteur aléatoire $(X_k - \widehat{X}_k^-, I_k)$ 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 Remarque 1.10, l'observation prédite $\mathbb{E}[Y_k | Y_{0:k-1}]$ dépend de façon affine des observations passées $(Y_0, Y_1, \dots, Y_{k-1})$, de sorte que l'innovation I_k dépend de façon affine des observations (Y_0, Y_1, \dots, Y_k) . On en déduit que le vecteur aléatoire (I_0, I_1, \dots, I_k) est gaussien, comme transformation affine d'un vecteur aléatoire gaussien.

Toujours d'après la Remarque 1.10, l'état prédit $\widehat{X}_k^- = \mathbb{E}[X_k | Y_{0:k-1}]$ dépend de façon affine des observations passées (Y_0, \dots, Y_{k-1}) , de sorte que le vecteur aléatoire $(Y_0, \dots, Y_{k-1}, X_k - \widehat{X}_k^-, I_k)$ dépend de façon affine du vecteur $(Y_0, Y_1, \dots, Y_k, X_k)$ formé de l'état courant X_k et des observations (Y_0, Y_1, \dots, Y_k) . On en déduit que le vecteur aléatoire $(Y_0, \dots, Y_{k-1}, X_k - \widehat{X}_k^-, I_k)$ est gaussien, et donc a fortiori le vecteur aléatoire $(X_k - \widehat{X}_k^-, I_k)$ est gaussien, comme transformation affine d'un vecteur aléatoire gaussien. Compte tenu que

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

par définition, on en déduit que le vecteur aléatoire $(X_k - \widehat{X}_k^-, I_k)$ est indépendant de $Y_{0:k-1}$. D'après l'équation (3.4), on a

$$I_k = Y_k - (H_k \widehat{X}_k^- + h_k) = H_k (X_k - \widehat{X}_k^-) + V_k , \quad (3.5)$$

et on en déduit que

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

Dans cette dernière égalité, on a utilisé le fait que $(X_k - \widehat{X}_k^-)$ est indépendant de V_k , donc $\mathbb{E}[(X_k - \widehat{X}_k^-) V_k^*] = 0$. On déduit également de (3.5) que

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

Dans cette dernière égalité, on a de nouveau utilisé le fait que $(X_k - \widehat{X}_k^-)$ est indépendant de V_k , donc $\mathbb{E}[(X_k - \widehat{X}_k^-) V_k^*] = 0$. \square

Remarque 3.4 Compte tenu que la distribution de probabilité conditionnelle du vecteur aléatoire Y_k sachant $Y_{0:k-1}$ est gaussienne, de moyenne $H_k \widehat{X}_k^- + h_k$ et de matrice Q_k^I inversible, on obtient l'expression suivante

$$\begin{aligned} L_n &= \prod_{k=0}^n \exp\left\{-\frac{1}{2} (Y_k - (H_k \widehat{X}_k^- + h_k))^* (Q_k^I)^{-1} (Y_k - (H_k \widehat{X}_k^- + h_k))\right\} \\ &= \prod_{k=0}^n \exp\left\{-\frac{1}{2} I_k^* (Q_k^I)^{-1} I_k\right\} , \end{aligned}$$

pour la vraisemblance du modèle, a une constante multiplicative près.

Théorème 3.5 (Filtre de Kalman) *On suppose que la matrice de covariance Q_k^V est inversible, pour tout instant k . Alors $\{\widehat{X}_k\}$ et $\{P_k\}$ sont définis par les équations suivantes*

$$\begin{aligned} \widehat{X}_k^- &= F_k \widehat{X}_{k-1} + f_k , \\ P_k^- &= F_k P_{k-1} F_k^* + G_k Q_k^W G_k^* , \end{aligned}$$

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, et avec les initialisations

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

Remarque 3.6 On vérifie que la suite $\{P_k\}$ ne dépend pas des observations : elle peut donc être pré-calculée, en particulier dans le cas simple où les coefficients $F_k = F$, $G_k = G$, $H_k = H$, $Q_k^W = Q^W$ et $Q_k^V = Q^V$ sont constants.

Remarque 3.7 Si les coefficients F_k , f_k et G_k dans l'équation (3.1) et les coefficients H_k et h_k dans l'équation (3.2) dépendent des observations $Y_{0:k-1}$, alors la suite $\{Z_k = (X_k, Y_k)\}$, et a fortiori la suite $\{X_k\}$, 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*, et on vérifie facilement que la distribution de probabilité conditionnelle du vecteur aléatoire X_k sachant $Y_{0:k}$ est gaussienne, de moyenne \hat{X}_k et de matrice de covariance P_k données encore par les équations du Théorème 3.5.

PREUVE. On procède en plusieurs étapes. Le point central est la Proposition 1.7 qui sera constamment utilisée.

Expression de \hat{X}_0 et P_0 en fonction de \hat{X}_0^- et P_0^- :

Le vecteur aléatoire (X_0, Y_0) est gaussien, de moyenne et de matrice de covariance données par

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

respectivement. D'après la Proposition 1.7, la distribution de probabilité conditionnelle du vecteur aléatoire X_0 sachant Y_0 est gaussienne, de moyenne

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

et de matrice de covariance

$$P_0 = Q_0^X - Q_0^X H_0^* [H_0 Q_0^X H_0^* + Q_0^V]^{-1} H_0 Q_0^X.$$

Expression de \hat{X}_k^- et P_k^- en fonction de \hat{X}_{k-1} et P_{k-1} :

Le vecteur aléatoire $(X_k, Y_0, \dots, Y_{k-1})$ est gaussien, et d'après la Proposition 1.7, la distribution de probabilité conditionnelle du vecteur aléatoire X_k sachant $Y_{0:k-1}$ est gaussienne, de moyenne \hat{X}_k^- et de matrice de covariance P_k^- . D'après l'équation (3.3), c'est-à-dire

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

on a

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

compte tenu que W_k et Y_{k-1} sont indépendants. Par différence

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

de sorte que

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

Dans cette dernière égalité, on a utilisé le fait que $(X_{k-1} - \widehat{X}_{k-1}^-)$ est indépendant de W_k , donc $\mathbb{E}[(X_{k-1} - \widehat{X}_{k-1}^-) W_k^*] = 0$.

Expression de \widehat{X}_k et P_k en fonction de \widehat{X}_k^- et P_k^- :

Le vecteur aléatoire (X_k, Y_0, \dots, Y_k) est gaussien, et d'après la Proposition 1.7, la distribution de probabilité conditionnelle du vecteur aléatoire X_k sachant $Y_{0:k}$ est gaussienne, de moyenne \widehat{X}_k et de matrice de covariance déterministe P_k . D'après la Remarque 3.2, on a

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

compte tenu que les vecteurs aléatoires $(X_k - \widehat{X}_k^-)$ et I_k sont indépendants de $Y_{0:k-1}$, d'après le Lemme 3.3. Par différence

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

de sorte que

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

Pour calculer la moyenne conditionnelle et la matrice de covariance conditionnelle du vecteur aléatoire X_k sachant $Y_{0:k}$, il suffit donc de calculer la moyenne conditionnelle et la matrice de covariance conditionnelle du vecteur aléatoire $(X_k - \widehat{X}_k^-)$ sachant I_k . En d'autres termes, pour estimer l'état caché X_k au vu des observations $Y_{0:k}$ il suffit d'estimer de quelle quantité, exprimée

en fonction de l'écart I_k constaté entre la nouvelle observation et l'observation prédite, corriger l'estimation prédite \widehat{X}_k^- . C'est de cette propriété que découle la forme récursive du filtre de Kalman. D'après le Lemme 3.3, le vecteur aléatoire $(X_k - \widehat{X}_k^-, I_k)$ 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 d'après la Proposition 1.7 on a immédiatement

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

et

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

ce qui termine la démonstration. \square

3.3 Lisseur de Kalman

On dispose désormais de l'information

$$Y_{0:n} = (Y_0, Y_1, \dots, Y_n),$$

et l'objectif est d'estimer de façon optimale le vecteur aléatoire X_k à partir de $Y_{0:n}$, pour un instant k intermédiaire entre l'instant initial 0 et l'instant final n . Si on adopte le critère du minimum de variance, il s'agit d'après la Section 1.2 de calculer la distribution de probabilité conditionnelle du vecteur aléatoire X_k sachant $Y_{0:n}$. Comme le cadre est gaussien, il suffit de calculer la moyenne et la matrice de covariance

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

et clairement, $\widehat{X}_n^n = \widehat{X}_n$ et $P_n^n = P_n$ pour $k = n$. D'après la Remarque 1.9, la matrice de covariance conditionnelle P_k^n ne dépend pas des observations, c'est-à-dire que

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

Théorème 3.8 (Lisseur de Kalman) *On suppose que les matrices de covariance Q_k^V et P_k^- sont inversibles, pour tout instant k . Alors $\{\widehat{X}_k^n\}$ et $\{P_k^n\}$ sont définis par les équations rétrogrades suivantes*

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

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

avec la matrice de gain

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

et avec les initialisations

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

PREUVE. On remarque que le vecteur aléatoire Y_k peut s'exprimer comme transformation affine du vecteur aléatoire (X_k, V_k) , et donc a fortiori comme transformation affine du vecteur aléatoire $(Y_{0:k-1}, X_k - \widehat{X}_k^-, V_k)$. De même, le vecteur aléatoire Y_{k+p} peut s'exprimer comme transformation affine du vecteur aléatoire (X_{k+p}, V_{k+p}) , et par transitivité comme transformation affine du vecteur aléatoire $(X_k, W_{k+1}, \dots, W_{k+p}, V_{k+p})$, et donc a fortiori comme transformation affine du vecteur aléatoire $(Y_{0:k-1}, X_k - \widehat{X}_k^-, W_{k+1}, \dots, W_{k+p}, V_{k+p})$. On en déduit que le vecteur aléatoire $Y_{0:n} = (Y_{0:k-1}, Y_k, \dots, Y_n)$ peut s'exprimer comme transformation affine du vecteur aléatoire $(Y_{0:k-1}, X_k - \widehat{X}_k^-, Z_{k+1:n})$ où $Z_{k+1:n} = (W_{k+1}, \dots, W_n, V_k, V_{k+1}, \dots, V_n)$ par définition. Les vecteurs aléatoires $Y_{0:k-1}$, $X_k - \widehat{X}_k^-$ et $Z_{k+1:n}$ sont mutuellement indépendants, et il résulte de la Remarque 1.11 que

$$\begin{aligned}
U_{k-1}^n &= \mathbb{E}[X_{k-1} \mid Y_{0:k-1}, X_k - \widehat{X}_k^-, Z_{k+1:n}] \\
&= \widehat{X}_{k-1} + \mathbb{E}[X_{k-1} - \widehat{X}_{k-1} \mid Y_{0:k-1}, X_k - \widehat{X}_k^-, Z_{k+1:n}] \\
&= \widehat{X}_{k-1} + \mathbb{E}[X_{k-1} - \widehat{X}_{k-1} \mid Y_{0:k-1}] + \mathbb{E}[X_{k-1} - \widehat{X}_{k-1} \mid X_k - \widehat{X}_k^-] \\
&\quad + \mathbb{E}[X_{k-1} - \widehat{X}_{k-1} \mid Z_{k+1:n}] \\
&= \widehat{X}_{k-1} + \mathbb{E}[X_{k-1} - \widehat{X}_{k-1} \mid X_k - \widehat{X}_k^-],
\end{aligned}$$

compte tenu que $\mathbb{E}[X_{k-1} - \widehat{X}_{k-1} \mid Y_{0:k-1}] = 0$ par définition, et où on a utilisé dans la dernière égalité le fait que $(X_{k-1} - \widehat{X}_{k-1})$ est indépendant de $Z_{k+1:n}$, donc $\mathbb{E}[X_{k-1} - \widehat{X}_{k-1} \mid Z_{k+1:n}] = 0$. Par différence

$$\begin{aligned}
X_{k-1} - U_{k-1}^n &= (X_{k-1} - \widehat{X}_{k-1}) - (U_{k-1}^n - \widehat{X}_{k-1}) \\
&= (X_{k-1} - \widehat{X}_{k-1}) - \mathbb{E}[X_{k-1} - \widehat{X}_{k-1} \mid X_k - \widehat{X}_k^-],
\end{aligned}$$

de sorte que

$$\begin{aligned}
&\mathbb{E}[(X_{k-1} - U_{k-1}^n)(X_{k-1} - U_{k-1}^n)^*] \\
&= \mathbb{E}[(X_{k-1} - \widehat{X}_{k-1}) - \mathbb{E}[X_{k-1} - \widehat{X}_{k-1} \mid X_k - \widehat{X}_k^-]] \\
&\quad ((X_{k-1} - \widehat{X}_{k-1}) - \mathbb{E}[X_{k-1} - \widehat{X}_{k-1} \mid X_k - \widehat{X}_k^-])^*].
\end{aligned}$$

Pour calculer la moyenne conditionnelle et la matrice de covariance conditionnelle du vecteur aléatoire X_{k-1} sachant $(Y_{0:k-1}, X_k - \widehat{X}_k^-, Z_{k+1:n})$, il suffit donc de calculer la moyenne conditionnelle et la matrice de covariance conditionnelle du vecteur aléatoire $(X_{k-1} - \widehat{X}_{k-1})$ sachant $(X_k - \widehat{X}_k^-)$. D'après la Remarque 1.10, l'état prédit $\widehat{X}_k^- = \mathbb{E}[X_k \mid Y_{0:k-1}]$ dépend de façon affine des observations passées (Y_0, \dots, Y_{k-1}) , et l'état estimé $\widehat{X}_k = \mathbb{E}[X_k \mid Y_{0:k}]$ dépend de façon affine des observations (Y_0, \dots, Y_k) , de sorte que le vecteur aléatoire $(X_{k-1} - \widehat{X}_{k-1}, X_k - \widehat{X}_k^-)$ dépend de façon affine du vecteur $(Y_0, \dots, Y_k, X_{k-1}, X_k)$. On en déduit que le vecteur aléatoire

$(X_{k-1} - \widehat{X}_{k-1}, X_k - \widehat{X}_k^-)$ est gaussien, comme transformation affine d'un vecteur aléatoire gaussien. Par différence

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

de sorte que

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

Dans cette dernière égalité, on a utilisé le fait que $(X_{k-1} - \widehat{X}_{k-1})$ et W_k sont indépendants, donc $\mathbb{E}[(X_{k-1} - \widehat{X}_{k-1}) W_k^*] = 0$. On en déduit que le vecteur aléatoire $(X_{k-1} - \widehat{X}_{k-1}, X_k - \widehat{X}_k^-)$ est gaussien, de moyenne nulle et de matrice de covariance

$$\begin{pmatrix} P_{k-1} & P_{k-1} F_k^* \\ F_k P_{k-1} & P_k^- \end{pmatrix} .$$

Par hypothèse la matrice P_k^- est inversible, et d'après la Proposition 1.7 on a immédiatement

$$U_{k-1}^n = \widehat{X}_{k-1} + P_{k-1} F_k^* (P_k^-)^{-1} (X_k - \widehat{X}_k^-) = \widehat{X}_{k-1} + L_k (X_k - \widehat{X}_k^-) ,$$

et

$$\mathbb{E}[(X_{k-1} - U_{k-1}^n) (X_{k-1} - U_{k-1}^n)^*] = P_{k-1} - P_{k-1} F_k^* (P_k^-)^{-1} F_k P_{k-1} = P_{k-1} - L_k P_k^- L_k^* .$$

et on en déduit que

$$\widehat{X}_{k-1}^n = \mathbb{E}[X_{k-1} | Y_{0:n}] = \mathbb{E}[U_{k-1}^n | Y_{0:n}] = \widehat{X}_{k-1} + L_k (\widehat{X}_k^n - \widehat{X}_k^-) ,$$

compte tenu que $(Y_{0:k-1}, X_k - \widehat{X}_k^-, Z_{k+1:n})$ contient davantage d'information que $Y_{0:n}$. Par différence

$$X_{k-1} - \widehat{X}_{k-1}^n = (X_{k-1} - U_{k-1}^n) + (U_{k-1}^n - \widehat{X}_{k-1}^n) \quad \text{et} \quad U_{k-1}^n - \widehat{X}_{k-1}^n = L_k (X_k - \widehat{X}_k^n) ,$$

de sorte que

$$\begin{aligned} P_{k-1}^n &= \mathbb{E}[(X_{k-1} - \widehat{X}_{k-1}^n) (X_{k-1} - \widehat{X}_{k-1}^n)^*] \\ &= \mathbb{E}[((X_{k-1} - U_{k-1}^n) + (U_{k-1}^n - \widehat{X}_{k-1}^n)) ((X_{k-1} - U_{k-1}^n) + (U_{k-1}^n - \widehat{X}_{k-1}^n))^*] \\ &= \mathbb{E}[(X_{k-1} - U_{k-1}^n) (X_{k-1} - U_{k-1}^n)^*] + \mathbb{E}[(U_{k-1}^n - \widehat{X}_{k-1}^n) (U_{k-1}^n - \widehat{X}_{k-1}^n)^*] \\ &\quad + \mathbb{E}[(U_{k-1}^n - \widehat{X}_{k-1}^n) (X_{k-1} - U_{k-1}^n)^*] + \mathbb{E}[(X_{k-1} - U_{k-1}^n) (U_{k-1}^n - \widehat{X}_{k-1}^n)^*] \\ &= (P_{k-1} - L_k P_k^- L_k^*) + L_k P_k^n L_k^* . \end{aligned}$$

Dans cette dernière égalité, on a utilisé le fait que

- $(U_{k-1}^n - \widehat{X}_{k-1}^n)$ dépend de $(Y_{0:k-1}, X_k - \widehat{X}_k^-, Z_{k+1:n})$,
- et $\mathbb{E}[X_{k-1} - U_{k-1}^n \mid Y_{0:k-1}, X_k - \widehat{X}_k^-, Z_{k+1:n}] = 0$ par définition,

donc $\mathbb{E}[(X_{k-1} - U_{k-1}^n) (U_{k-1}^n - \widehat{X}_{k-1}^n)^*] = 0$.

□

Chapitre 4

Extensions aux systèmes non-linéaires

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

$$X_k = b_k(X_{k-1}) + \sigma_k(X_{k-1}) W_k , \quad (4.1)$$

où $\{X_k\}$ et $\{W_k\}$ prennent respectivement leurs valeurs dans \mathbb{R}^m et \mathbb{R}^p , et une suite d'observations $\{Y_k\}$ à valeurs dans \mathbb{R}^d , vérifiant

$$Y_k = h_k(X_k) + V_k , \quad (4.2)$$

et on suppose que

- la condition initiale X_0 est gaussienne, de moyenne \bar{X}_0 et de matrice de covariance Q_0^X ,
- la suite $\{W_k\}$ est un bruit blanc gaussien, de matrice de covariance Q_k^W ,
- la suite $\{V_k\}$ est un bruit blanc gaussien, de matrice de covariance Q_k^V inversible,
- les suites $\{W_k\}$ et $\{V_k\}$ et la condition initiale X_0 sont mutuellement indépendants.

La signification du modèle (4.1) est la suivante

- 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 $b_k(x)$ et de matrice de covariance $\sigma_k(x) Q_k^W (\sigma_k(x))^*$.

La plupart des propriétés obtenues à la Section 3.1 ne sont pas vraies pour le système décrit par les équations (4.1) et (4.2). En particulier, le processus $\{Z_k = (X_k, Y_k)\}$ n'est pas gaussien (ni même conditionnellement gaussien), et les moments conditionnels de X_k sachant $Y_{0:k}$ ne peuvent pas être calculés de manière simple. Deux approches pragmatiques sont présentées dans ce chapitre, qui permettent d'obtenir des estimateurs sous-optimaux, c'est-à-dire qui n'atteignent pas nécessairement le minimum de l'erreur quadratique moyenne, mais qui sont néanmoins très

largement utilisés en pratique. La première approche présentée à la Section 4.1 repose sur des techniques de linéarisation, et donne lieu au filtre de Kalman linéarisé et au filtre de Kalman étendu. La deuxième approche présentée à la Section 4.2 repose sur des techniques d'approximation gaussienne et de quadrature numérique, et donne lieu au filtre de Kalman dit *unscented*. Dans les chapitres suivants, on abandonnera ce point de vue de linéarisation ou d'approximation gaussienne, et on s'attachera d'abord à caractériser la distribution de probabilité conditionnelle de l'état caché sachant les observations, soit par une représentation probabiliste, soit par une équation récurrente dans l'espace des distributions de probabilité, et on proposera ensuite des approximations numériques reposant sur méthodes de simulation de type Monte Carlo.

4.1 Filtre de Kalman linéarisé, filtre de Kalman étendu

On considère le système non linéaire

$$\begin{aligned} X_k &= b_k(X_{k-1}) + \sigma_k(X_{k-1}) W_k, \\ Y_k &= h_k(X_k) + V_k, \end{aligned} \tag{4.3}$$

et on suppose que les fonctions b_k et h_k sont dérivables. En linéarisant le système (4.3) autour d'une suite déterministe donnée, ou bien autour de l'estimateur courant, on peut obtenir des algorithmes sous-optimaux, qui sont décrits ci-dessous.

Filtre de Kalman linéarisé

On se donne une suite (déterministe) $\{\bar{x}_k\}$ à valeurs dans \mathbb{R}^m , appelée *trajectoire nominale* (on peut prendre par exemple \bar{x}_k comme une approximation de la moyenne de X_k). La méthode consiste à linéariser les fonctions b_k et σ_k autour de \bar{x}_{k-1} , c'est-à-dire

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

et la fonction h_k autour de \bar{x}_k , c'est-à-dire

$$h_k(x) \simeq h_k(\bar{x}_k) + h'_k(\bar{x}_k)(x - \bar{x}_k).$$

Le système non-linéaire (4.3) est alors remplacé par le système linéaire gaussien

$$\begin{aligned} X_k &= F_k X_{k-1} + f_k + G_k W_k, \\ Y_k &= H_k X_k + h_k + V_k, \end{aligned}$$

avec

$$F_k = b'_k(\bar{x}_{k-1}), \quad f_k = -b'_k(\bar{x}_{k-1}) \bar{x}_{k-1} + b_k(\bar{x}_{k-1}) \quad \text{et} \quad G_k = \sigma_k(\bar{x}_{k-1}),$$

et avec

$$H_k = h'_k(\bar{x}_k) \quad \text{et} \quad h_k = -h'_k(\bar{x}_k) \bar{x}_k + h_k(\bar{x}_k).$$

On applique alors exactement le filtre de Kalman à ce nouveau système, d'où l'algorithme sous-optimal suivant

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

et

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

avec la matrice de gain

$$K_k = P_k^- h'_k(\bar{x}_k)^* [h'_k(\bar{x}_k) P_k^- (h'_k(\bar{x}_k))^* + Q_k^V]^{-1} .$$

A la place de la première et la troisième de ces équations, on peut utiliser

$$\begin{aligned}\widehat{X}_k^- &= b_k(\widehat{X}_{k-1}) , \\ \widehat{X}_k &= \widehat{X}_k^- + K_k [Y_k - h_k(\widehat{X}_k^-)] .\end{aligned}$$

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 .

Filtre de Kalman étendu

Au lieu de linéariser autour d'une trajectoire nominale déterministe $\{\bar{x}_k\}$, on peut utiliser l'estimateur courant. La méthode consiste à 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 à linéariser la fonction h_k autour de \widehat{X}_k^- , c'est-à-dire

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

Le système non-linéaire (4.3) est alors remplacé par le système conditionnellement linéaire gaussien

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

$$Y_k = H_k + h_k + V_k ,$$

avec

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

et avec

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

On applique alors exactement le filtre de Kalman à ce nouveau système, au vu de la Remarque 3.7, d'où l'algorithme sous-optimal suivant

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

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

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 .

Remarque 4.1 Dans cet algorithme, la suite $\{P_k\}$ dépend des observations, et ne peut donc pas être pré-calculée.

4.2 Filtre de Kalman *unscented*

On considère à nouveau le système non linéaire (4.3), c'est-à-dire

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

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

et on ne suppose plus que les fonctions b_k et h_k sont dérivables, mais on suppose que les fonctions b_k , h_k et σ_k et certaines fonctions associées, peuvent être intégrées par rapport à certaines distributions de probabilité gaussiennes.

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.

► Le calcul des deux premiers moments (moyenne et matrice de covariance) de la distribution de probabilité conditionnelle $\mu_k^-(dx) = \mathbb{P}[X_k \in dx \mid Y_{0:k-1}]$, c'est-à-dire le calcul de la moyenne conditionnelle et de la matrice de covariance conditionnelle du vecteur aléatoire X_k sachant $Y_{0:k-1}$, est facile. Par définition

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

compte tenu que

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

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

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

de sorte que

$$\begin{aligned} P_k^- &= \mathbb{E}[(X_k - \widehat{X}_k^-) (X_k - \widehat{X}_k^-)^* \mid Y_{0:k-1}] \\ &= \mathbb{E}[(b_k(X_{k-1}) - \widehat{X}_k^-) + \sigma_k(X_{k-1}) W_k] ((b_k(X_{k-1}) - \widehat{X}_k^-) + \sigma_k(X_{k-1}) W_k)^* \mid Y_{0:k-1}] \\ &= \mathbb{E}[(b_k(X_{k-1}) - \widehat{X}_k^-) (b_k(X_{k-1}) - \widehat{X}_k^-)^* \mid Y_{0:k-1}] \\ &\quad + \mathbb{E}[\sigma_k(X_{k-1}) W_k (b_k(X_{k-1}) - \widehat{X}_k^-)^* \mid Y_{0:k-1}] \\ &\quad + \mathbb{E}[(b_k(X_{k-1}) - \widehat{X}_k^-) W_k^* \sigma_k^*(X_{k-1}) \mid Y_{0:k-1}] \\ &\quad + \mathbb{E}[\sigma_k(X_{k-1}) W_k W_k^* \sigma_k^*(X_{k-1}) \mid Y_{0:k-1}] \\ &= \int_{\mathbb{R}^m} (b_k(x) - \widehat{X}_k^-) (b_k(x) - \widehat{X}_k^-)^* \mu_{k-1}(dx) + \int_{\mathbb{R}^m} \sigma_k(x) Q_k^W \sigma_k^*(x) \mu_{k-1}(dx) , \end{aligned}$$

compte tenu que

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

où on a utilisé dans la dernière égalité l'indépendance de $(Y_0, \dots, Y_{k-1}, X_{k-1})$ et de W_k , donc $\mathbb{E}[W_k W_k^* \mid X_{k-1}, Y_{0:k-1}] = Q_k^W$, et compte tenu que

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

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

► En revanche, le calcul des deux premiers moments (moyenne et matrice de covariance) de la distribution de probabilité conditionnelle $\mu_k(dx) = \mathbb{P}[X_k \in dx \mid Y_{0:k}]$, c'est-à-dire le calcul de la moyenne conditionnelle et de la matrice de covariance conditionnelle du vecteur aléatoire X_k sachant $Y_{0:k}$, n'est pas immédiat, et on commence par le calcul des deux premiers moments (moyenne et matrice de covariance) de la distribution de probabilité conditionnelle jointe du vecteur aléatoire (X_k, Y_k) sachant $Y_{0:k-1}$, qui est plus facile. On rappelle que

$$\widehat{X}_k^- = \int_{\mathbb{R}^m} b_k(x) \mu_{k-1}(dx) ,$$

a déjà été obtenu plus haut, et par définition

$$\begin{aligned}
\widehat{Y}_k^- &= \mathbb{E}[Y_k \mid Y_{0:k-1}] \\
&= \mathbb{E}[h_k(X_k) \mid Y_{0:k-1}] + \mathbb{E}[V_k \mid Y_{0:k-1}] \\
&= \int_{\mathbb{R}^m} h_k(x) \mu_k^-(dx) .
\end{aligned}$$

On rappelle que

$$P_k^- = \int (b_k(x) - \widehat{X}_k^-) (b_k(x) - \widehat{X}_k^-)^* \mu_{k-1}(dx) + \int_{\mathbb{R}^m} \sigma_k(x) Q_k^W \sigma_k^*(x) \mu_{k-1}(dx) ,$$

a déjà été obtenu plus haut, et par différence

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

de sorte que

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

compte tenu que

$$\begin{aligned}
&\mathbb{E}[V_k (h_k(X_k) - \widehat{Y}_k^-)^* | Y_{0:k-1}] \\
&= \mathbb{E}[\mathbb{E}[V_k (h_k(X_k) - \widehat{Y}_k^-)^* | X_k, Y_{0:k-1}] | Y_{0:k-1}] \\
&= \mathbb{E}[\mathbb{E}[V_k | X_k, Y_{0:k-1}] (h_k(X_k) - \widehat{Y}_k^-)^* | Y_{0:k-1}] = 0.
\end{aligned}$$

où on a utilisé dans la dernière égalité l'indépendance de $(Y_0, \dots, Y_{k-1}, X_k)$ et de V_k , donc $\mathbb{E}[V_k | X_k, Y_{0:k-1}] = 0$, et

$$\begin{aligned}
C_k &= \mathbb{E}[(X_k - \widehat{X}_k^-)(Y_k - \widehat{Y}_k^-)^* | Y_{0:k-1}] \\
&= \mathbb{E}[(X_k - \widehat{X}_k^-)(h_k(X_k) - \widehat{Y}_k^-)^* | Y_{0:k-1}] + \mathbb{E}[(X_k - \widehat{X}_k^-) V_k^* | Y_{0:k-1}] \\
&= \int_{\mathbb{R}^m} (x - \widehat{X}_k^-)(h_k(x) - \widehat{Y}_k^-)^* \mu_k^-(dx).
\end{aligned}$$

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

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

respectivement. Si la matrice Q_k^V est inversible, alors a fortiori la matrice Ξ_k est inversible, et d'après la Proposition 1.7 on obtient immédiatement les approximations suivantes

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

pour les deux premiers moments de la distribution de probabilité conditionnelle μ_k , c'est-à-dire pour la moyenne conditionnelle et de la matrice de covariance conditionnelle du vecteur aléatoire X_k sachant $Y_{0:k}$.

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} , et 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^- . Pour fermer ces équations, 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 $P_{k-1} = S_{k-1} S_{k-1}^*$, et en effectuant le changement de variable $x = \widehat{X}_{k-1} + S_{k-1} u$, on obtient les approximations

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

et

$$\begin{aligned} P_k^- &\approx \int (\widehat{b}_k(u) - \widehat{X}_k^-) (\widehat{b}_k(u) - \widehat{X}_k^-)^* \exp\{-\frac{1}{2}|u|^2\} \frac{du}{(2\pi)^{m/2}} \\ &\quad + \int \widehat{\sigma}_k(u) Q_k^W \widehat{\sigma}_k^*(u) \exp\{-\frac{1}{2}|u|^2\} \frac{du}{(2\pi)^{m/2}} \end{aligned}$$

où par définition

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

► 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 $P_k^- = S_k^- (S_k^-)^*$, et en effectuant le changement de variable $x = \widehat{X}_k^- + S_k^- u$, on obtient les approximations

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

et

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

et

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

où par définition

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

Il reste donc à calculer les intégrales des fonctions non-linéaires

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

par rapport à la densité gaussienne réduite centrée.

Remarque 4.2 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 numérique.

On introduit les formules de quadrature suivantes, reposant sur la notion de σ -points. 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}, \dots, u_m) appelés σ -points, et définis par

$$u_0 = 0, \quad u_i = e_i \sqrt{m + \kappa} \quad \text{et} \quad u_{-i} = -u_i,$$

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

$$w_0 = \frac{\kappa}{m + \kappa} \quad \text{et} \quad w_{-i} = w_i = \frac{1}{2(m + \kappa)}, \quad (4.4)$$

pour tout $i = 1, \dots, m$ (d'autres choix de σ -points sont possibles). On vérifie que

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

c'est-à-dire que les deux premiers moments sont pris en compte exactement. Plus généralement

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

et un changement de variable évident donne aussitôt

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

pour toute fonction ϕ définie sur \mathbb{R}^m , c'est-à-dire que les σ -points (x_{-m}, \dots, 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

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

pour tout $i = 1, \dots, m$. On vérifie que

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

c'est-à-dire que les deux premiers moments sont pris en compte exactement.

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

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

affectés des poids (4.4) pour tout $i = 1, \dots, m$, et on définit le vecteur moyenne

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

et la matrice de covariance

$$P_k^- = \sum_{i=-m}^{+m} w_i (b_k(x_i) - \widehat{X}_k^-) (b_k(x_i) - \widehat{X}_k^-)^* + \sum_{i=-m}^{+m} w_i \sigma_k(x_i) Q_k^W \sigma_k^*(x_i) = S_k^- (S_k^-)^*.$$

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

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

affectés des poids (4.4) pour tout $i = 1, \dots, m$, on définit le vecteur moyenne

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

la matrice de covariance

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

et la matrice de corrélation

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

et on pose

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

Chapitre 5

Filtrage bayésien

On considère d'abord une classe très générale de modèles à espace d'état, qui est aussi un cas particulier de la classe encore plus générale des modèles de Markov cachés, pour lesquels l'espace d'état peut être très général, et qui seront considérés ensuite.

5.1 Systèmes non-linéaires à bruits non-gaussiens

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

$$X_k = f_k(X_{k-1}, W_k) , \quad (5.1)$$

où $\{X_k\}$ et $\{W_k\}$ prennent respectivement leurs valeurs dans \mathbb{R}^m et \mathbb{R}^p , et une suite d'observations $\{Y_k\}$ à valeurs dans \mathbb{R}^d , vérifiant

$$Y_k = h_k(X_k) + V_k , \quad (5.2)$$

et on suppose que

- la condition initiale X_0 n'est pas nécessairement gaussienne,
- la suite $\{W_k\}$ est un bruit blanc, pas nécessairement gaussien,
- la suite $\{V_k\}$ est un bruit blanc, pas nécessairement gaussien,
- les suites $\{W_k\}$ et $\{V_k\}$ et la condition initiale X_0 sont mutuellement indépendants.

On ne suppose pas que les fonctions f_k et h_k sont dérivables. On suppose en revanche que

- il est facile de *simuler* un vecteur aléatoire selon la loi $\eta_0(dx)$ de X_0 ,
- il est facile de *simuler* un vecteur aléatoire selon la loi $p_k^W(dw)$ de W_k ,
- la loi du vecteur aléatoire V_k admet une densité $q_k^V(v)$ qu'il est facile d'*évaluer* pour tout $v \in \mathbb{R}^d$.

Proposition 5.1 *La suite $\{X_k\}$ est une chaîne de Markov à valeurs dans \mathbb{R}^m , c'est-à-dire que la loi conditionnelle par rapport au passé*

$$\mathbb{P}[X_k \in dx' \mid X_0, \dots, X_{k-1}] = \mathbb{P}[X_k \in dx' \mid X_{k-1}] ,$$

ne dépend que du passé immédiat, avec le noyau de probabilités de transition

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

défini par

$$Q_k \phi(x) = \mathbb{E}[\phi(X_k) \mid X_{k-1} = x] = \int_{\mathbb{R}^p} \phi(f_k(x, w)) p_k^W(dw) ,$$

pour toute fonction mesurable bornée ϕ définie sur \mathbb{R}^m .

PREUVE. Compte tenu que W_k est indépendant de (X_0, \dots, X_{k-1}) , on a

$$\begin{aligned} \mathbb{E}[\phi(X_k) \mid X_0, \dots, X_{k-1}] &= \mathbb{E}[\phi(f_k(X_{k-1}, W_k)) \mid X_0, \dots, X_{k-1}] \\ &= \int_{\mathbb{R}^p} \phi(f_k(X_{k-1}, w)) p_k^W(dw) , \end{aligned}$$

pour toute fonction mesurable bornée ϕ définie sur \mathbb{R}^m . Clairement, le résultat ne dépend que de X_{k-1} , c'est-à-dire que

$$\mathbb{E}[\phi(X_k) \mid X_0, \dots, X_{k-1}] = \mathbb{E}[\phi(X_k) \mid X_{k-1}] ,$$

et

$$\mathbb{E}[\phi(X_k) \mid X_{k-1} = x] = \int_{\mathbb{R}^p} \phi(f_k(x, w)) p_k^W(dw) . \quad \square$$

Remarque 5.2 Si $f_k(x, w) = b_k(x) + w$, et si la loi $p_k^W(dw)$ de W_k admet une densité encore notée $p_k^W(w)$, c'est-à-dire si $p_k^W(dw) = p_k^W(w) dw$, alors

$$Q_k(x, dx') = p_k^W(x' - b_k(x)) dx'$$

c'est-à-dire que le noyau $Q_k(x, dx')$ admet une densité. En effet, le changement de variable $x' = b_k(x) + w$ donne immédiatement

$$Q_k \phi(x) = \int_{\mathbb{R}^m} \phi(b_k(x) + w) p_k^W(w) dw = \int_{\mathbb{R}^m} \phi(x') p_k^W(x' - b_k(x)) dx' ,$$

pour toute fonction mesurable bornée ϕ définie sur \mathbb{R}^m .

Remarque 5.3 En général, le noyau $Q_k(x, dx')$ n'admet pas de densité. En effet, conditionnellement à $X_{k-1} = x$, le vecteur aléatoire X_k appartient nécessairement au sous-ensemble

$$\mathcal{M}(x) = \{x' \in \mathbb{R}^m : \text{il existe } w \in \mathbb{R}^p \text{ tel que } x' = f_k(x, w)\} ,$$

et dans le cas où $p < m$ ce sous ensemble $\mathcal{M}(x)$ est généralement, sous certaines hypothèses de régularité, une sous-variété différentielle de dimension p dans l'espace \mathbb{R}^m . Il ne peut donc pas y avoir de densité pour la loi $Q_k(x, dx')$ du vecteur aléatoire X_k .

Proposition 5.4 *La suite $\{Y_k\}$ vérifie l'hypothèse de canal sans mémoire, c'est-à-dire que*

- *conditionnellement aux états cachés X_0, \dots, X_n les observations Y_0, \dots, Y_n sont mutuellement indépendantes,*
- *pour tout $k = 0, \dots, n$, la loi conditionnelle de Y_k sachant X_0, \dots, X_n ne dépend que de X_k , avec la probabilité d'émission*

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

définie par

$$g_k(x, y) = q_k^V(y - h_k(x)) ,$$

et on définit la fonction de vraisemblance

$$g_k(x) = g_k(x, Y_k) = q_k^V(Y_k - h_k(x)) ,$$

qui mesure l'adéquation d'un état quelconque $x \in \mathbb{R}^m$ avec l'observation Y_k .

En d'autres termes

$$\begin{aligned} \mathbb{P}[Y_0 \in dy_0, \dots, Y_n \in dy_n \mid X_0 = x_0, \dots, X_n = x_n] &= \prod_{k=0}^n \mathbb{P}[Y_k \in dy_k \mid X_k = x_k] \\ &= \prod_{k=0}^n g_k(x_k, y_k) dy_0 \cdots dy_n . \end{aligned}$$

PREUVE. Pour toute famille de fonctions mesurables bornées ϕ_0, \dots, ϕ_n définies sur \mathbb{R}^d , et compte tenu que les vecteurs aléatoires V_0, \dots, V_n sont mutuellement indépendants et indépendants des vecteurs aléatoires X_0, \dots, X_n , on a

$$\begin{aligned} &\mathbb{E}[\phi_0(Y_0) \cdots \phi_n(Y_n) \mid X_0, \dots, X_n] \\ &= \mathbb{E}[\phi_0(h_0(X_0) + V_0) \cdots \phi_n(h_n(X_n) + V_n) \mid X_0, \dots, X_n] \\ &= \int_{\mathbb{R}^d} \cdots \int_{\mathbb{R}^d} \phi_0(h_0(X_0) + v_0) \cdots \phi_n(h_n(X_n) + v_n) \mathbb{P}[V_0 \in dv_0, \dots, V_n \in dv_n] \\ &= \prod_{k=0}^n \int_{\mathbb{R}^d} \phi_k(h_k(X_k) + v) \mathbb{P}[V_k \in dv] , \end{aligned}$$

et de même

$$\begin{aligned} \mathbb{E}[\phi_k(Y_k) \mid X_k] &= \mathbb{E}[\phi_k(h_k(X_k) + V_k) \mid X_k] \\ &= \int_{\mathbb{R}^d} \phi_k(h_k(X_k) + v) \mathbb{P}[V_k \in dv] \\ &= \int_{\mathbb{R}^d} \phi_k(h_k(X_k) + v) q_k^V(v) dv = \int_{\mathbb{R}^d} \phi_k(y) q_k^V(y - h_k(X_k)) dy , \end{aligned}$$

de sorte que

$$\mathbb{E}[\phi_0(Y_0) \cdots \phi_n(Y_n) \mid X_0, \dots, X_n] = \prod_{k=0}^n \mathbb{E}[\phi_k(Y_k) \mid X_k] ,$$

et

$$\mathbb{P}[Y_k \in dy \mid X_k = x] = q_k^V(y - h_k(x)) dy . \quad \square$$

5.2 Modèles de Markov cachés

Plus généralement, on peut aussi considérer un modèle de Markov caché où les états cachés $\{X_k\}$ forment une chaîne de Markov à valeurs dans un espace E , de noyaux de transition

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

et de loi initiale

$$\mathbb{P}[X_0 \in dx] = \eta_0(dx) ,$$

et où les observations $\{Y_k\}$ vérifient l'hypothèse de *canal sans mémoire*, c'est-à-dire que

- conditionnellement aux états cachés X_0, \dots, X_n les observations Y_0, \dots, Y_n sont mutuellement indépendantes,
- pour tout $k = 0, \dots, n$, la loi conditionnelle de Y_k sachant X_0, \dots, X_n ne dépend que de X_k , avec la probabilité d'émission

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

et on définit la *fonction de vraisemblance*

$$g_k(x) = g_k(x, Y_k) ,$$

qui mesure l'adéquation d'un état quelconque $x \in \mathbb{R}^m$ avec l'observation Y_k .

En d'autres termes

$$\mathbb{P}[X_0 \in dx_0, \dots, X_n \in dx_n] = \eta_0(dx_0) \prod_{k=1}^n Q_k(x_{k-1}, dx_k) ,$$

et

$$\begin{aligned} \mathbb{P}[Y_0 \in dy_0, \dots, Y_n \in dy_n \mid X_0 = x_0, \dots, X_n = x_n] &= \prod_{k=0}^n \mathbb{P}[Y_k \in dy_k \mid X_k = x_k] \\ &= \prod_{k=0}^n g_k(x_k, y_k) dy_0 \cdots dy_n . \end{aligned}$$

On suppose en outre que pour tout instant k

- il est facile de *simuler* pour tout $x \in E$, un vecteur aléatoire selon la loi $Q_k(x, dx')$,
- il est facile d'*évaluer* pour tout $x \in E$, la fonction de vraisemblance $g_k(x)$.

5.3 Filtre bayésien

L'objectif de cette section est d'établir les équations du filtre non-linéaire optimal, pour les systèmes non-linéaires et non-gaussiens, ou plus généralement les équations du filtre bayésien optimal, pour les modèles de Markov cachés. Il s'agit donc de calculer la loi conditionnelle de la variable aléatoire X_k sachant $Y_{0:k}$, et la loi conditionnelle de la variable aléatoire X_k sachant $Y_{0:k-1}$, définies par

$$\mu_k(dx) = \mathbb{P}[X_k \in dx \mid Y_{0:k}] \quad \text{et} \quad \mu_k^-(dx) = \mathbb{P}[X_k \in dx \mid Y_{0:k-1}] ,$$

respectivement.

Théorème 5.5 *Le filtre bayésien μ_n défini par*

$$\langle \mu_n, \phi \rangle = \mathbb{E}[\phi(X_n) \mid Y_{0:n}] ,$$

admet la représentation probabiliste

$$\langle \mu_n, \phi \rangle = \frac{\langle \gamma_n, \phi \rangle}{\langle \gamma_n, 1 \rangle} \quad \text{avec} \quad \langle \gamma_n, \phi \rangle = \mathbb{E}[\phi(X_n) \prod_{k=0}^n g_k(X_k)] ,$$

pour toute fonction mesurable bornée ϕ définie sur E .

Remarque 5.6 L'espérance porte seulement sur les états cachés successifs $X_{0:n}$: pour tout $k = 0, 1, \dots, n$, la fonction de vraisemblance $g_k(x)$ est définie par abus de notation comme

$$g_k(x) = g_k(x, Y_k) ,$$

pour tout $x \in E$, et dépend implicitement de l'observation Y_k , mais celle-ci est considérée comme *fixée*.

PREUVE. D'après la formule de Bayes, et d'après la propriété de canal sans mémoire, la loi jointe des états cachés $X_{0:n} = (X_0, \dots, X_n)$ et des observations $Y_{0:n} = (Y_0, \dots, Y_n)$ s'exprime comme

$$\begin{aligned} & \mathbb{P}[X_0 \in dx_0, \dots, X_n \in dx_n, Y_0 \in dy_0, \dots, Y_n \in dy_n] \\ &= \mathbb{P}[Y_0 \in dy_0, \dots, Y_n \in dy_n \mid X_0 = x_0, \dots, X_n = x_n] \mathbb{P}[X_0 \in dx_0, \dots, X_n \in dx_n] \\ &= \mathbb{P}[X_0 \in dx_0, \dots, X_n \in dx_n] \prod_{k=0}^n g_k(x_k, y_k) dy_0 \cdots dy_n . \end{aligned}$$

En intégrant par rapport aux variables x_0, \dots, x_n , on obtient la loi jointe des observations $Y_{0:n} = (Y_0, \dots, Y_n)$, c'est-à-dire

$$\begin{aligned} & \mathbb{P}[Y_0 \in dy_0, \dots, Y_n \in dy_n] \\ &= \int_E \cdots \int_E \prod_{k=0}^n g_k(x_k, y_k) \mathbb{P}[X_0 \in dx_0, \dots, X_n \in dx_n] dy_0 \cdots dy_n \\ &= \mathbb{E}[\prod_{k=0}^n g_k(X_k, y_k)] dy_0 \cdots dy_n . \end{aligned}$$

On vérifie que

$$\begin{aligned}
& \mathbb{P}[X_0 \in dx_0, \dots, X_n \in dx_n, Y_0 \in dy_0, \dots, Y_n \in dy_n] \\
&= \mathbb{P}[X_0 \in dx_0, \dots, X_n \in dx_n] \prod_{k=0}^n g_k(x_k, y_k) dy_0 \cdots dy_n \\
&= \frac{\mathbb{P}[X_0 \in dx_0, \dots, X_n \in dx_n] \prod_{k=0}^n g_k(x_k, y_k)}{\mathbb{E}[\prod_{k=0}^n g_k(X_k, y_k)]} \mathbb{E}[\prod_{k=0}^n g_k(X_k, y_k)] dy_0 \cdots dy_n \\
&= \frac{\mathbb{P}[X_0 \in dx_0, \dots, X_n \in dx_n] \prod_{k=0}^n g_k(x_k, y_k)}{\mathbb{E}[\prod_{k=0}^n g_k(X_k, y_k)]} \mathbb{P}[Y_0 \in dy_0, \dots, Y_n \in dy_n],
\end{aligned}$$

et d'après la formule de Bayes, il vient

$$\mathbb{P}[X_0 \in dx_0, \dots, X_n \in dx_n \mid Y_0 = y_0, \dots, Y_n = y_n] = \frac{\prod_{k=0}^n g_k(x_k, y_k) \mathbb{P}[X_0 \in dx_0, \dots, X_n \in dx_n]}{\mathbb{E}[\prod_{k=0}^n g_k(X_k, y_k)]}.$$

Comme cette identité est valable pour toute suite (y_0, \dots, y_n) , on obtient

$$\mathbb{P}[X_0 \in dx_0, \dots, X_n \in dx_n \mid Y_0, \dots, Y_n] = \frac{\prod_{k=0}^n g_k(x_k) \mathbb{P}[X_0 \in dx_0, \dots, X_n \in dx_n]}{\mathbb{E}[\prod_{k=0}^n g_k(X_k)]},$$

où l'espérance porte seulement sur les états cachés successifs (X_0, \dots, X_n) : les fonctions de vraisemblance $g_0(x), \dots, g_n(x)$ dépendent implicitement des observations (Y_0, \dots, Y_n) , mais celles-ci sont considérées comme fixées dans l'expression ci-dessus. On remarque que

$$\mathbb{P}[X_0 \in dx_0, \dots, X_n \in dx_n \mid Y_0, \dots, Y_n] \propto \underbrace{\prod_{k=0}^n g_k(x_k)}_{g_{0:n}(x_{0:n})} \underbrace{\eta_0(dx_0) \prod_{k=1}^n Q_k(x_{k-1}, dx_k)}_{\eta_{0:n}(dx_{0:n})}, \quad (5.3)$$

à une constante multiplicative près, où

$$g_{0:n}(x_{0:n}) = \prod_{k=0}^n g_k(x_k),$$

et où

$$\eta_{0:n}(dx_{0:n}) = \eta_0(dx_0) \prod_{k=1}^n Q_k(x_{k-1}, dx_k) = \mathbb{P}[X_0 \in dx_0, \dots, X_n \in dx_n],$$

désigne la loi jointe des états cachés $X_{0:n} = (X_0, \dots, X_n)$. Pour toute fonction mesurable bornée ϕ définie sur E , on a

$$\int_E \dots \int_E \phi(x_n) \prod_{k=0}^n g_k(x_k) \mathbb{P}[X_0 \in dx_0, \dots, X_n \in dx_n] = \mathbb{E}[\phi(X_n) \prod_{k=0}^n g_k(X_k)],$$

de sorte que

$$\begin{aligned} \mathbb{E}[\phi(X_n) | Y_0, \dots, Y_n] &= \int_E \dots \int_E \phi(x_n) \mathbb{P}[X_0 \in dx_0, \dots, X_n \in dx_n | Y_0, \dots, Y_n] \\ &= \frac{\int_E \dots \int_E \phi(x_n) \prod_{k=0}^n g_k(x_k) \mathbb{P}[X_0 \in dx_0, \dots, X_n \in dx_n]}{\mathbb{E}[\prod_{k=0}^n g_k(X_k)]} \\ &= \frac{\mathbb{E}[\phi(X_n) \prod_{k=0}^n g_k(X_k)]}{\mathbb{E}[\prod_{k=0}^n g_k(X_k)]} = \frac{\langle \gamma_n, \phi \rangle}{\langle \gamma_n, 1 \rangle}. \quad \square \end{aligned}$$

De la même manière

$$\langle \mu_n^-, \phi \rangle = \mathbb{E}[\phi(X_n) | Y_0, \dots, Y_{n-1}] = \frac{\mathbb{E}[\phi(X_n) \prod_{k=0}^{n-1} g_k(X_k)]}{\mathbb{E}[\prod_{k=0}^{n-1} g_k(X_k)]} = \frac{\langle \gamma_n^-, \phi \rangle}{\langle \gamma_n^-, 1 \rangle},$$

où la mesure positive (non-normalisée) $\gamma_n^-(dx)$ est définie par

$$\langle \gamma_n^-, \phi \rangle = \mathbb{E}[\phi(X_n) \prod_{k=0}^{n-1} g_k(X_k)],$$

pour toute fonction mesurable bornée ϕ définie sur E .

En utilisant la factorisation (5.3), il est facile d'obtenir une équation récurrente permettant d'exprimer μ_k en fonction de μ_{k-1} .

Théorème 5.7 (Filtre bayésien optimal) *La suite $\{\mu_k\}$ vérifie l'équation récurrente suivante*

$$\mu_{k-1} \xrightarrow{\text{prédiction}} \mu_k^- = \mu_{k-1} Q_k \xrightarrow{\text{correction}} \mu_k = g_k \cdot \mu_k^-,$$

où par définition

$$\mu_{k-1} Q_k(dx') = \int_E \mu_{k-1}(dx) Q_k(x, dx')$$

désigne l'action du noyau markovien $Q_k(x, dx')$ sur la distribution de probabilité $\mu_{k-1}(dx)$, et où

$$g_k \cdot \mu_k^-(dx') = \frac{g_k(x') \mu_k^-(dx')}{\langle \mu_k^-, g_k \rangle},$$

désigne le produit projectif de la distribution de probabilité a priori $\mu_k^-(dx')$ et de la fonction de vraisemblance $g_k(x')$.

Remarque 5.8 Une autre manière de caractériser les distributions de probabilité $\mu_{k-1} Q_k$ et $g_k \cdot \mu_k^-$ est de décrire leur action sur des fonctions test : pour toute fonction mesurable bornée ϕ définie sur E

$$\begin{aligned} \langle \mu_{k-1} Q_k, \phi \rangle &= \int_E \mu_{k-1} Q_k(dx') \phi(x') \\ &= \int_E \left[\int_E \mu_{k-1}(dx) Q_k(x, dx') \right] \phi(x') = \int_E \mu_{k-1}(dx) \left[\int_E Q_k(x, dx') \phi(x') \right] \\ &= \int_E \mu_{k-1}(dx) Q_k \phi(x) = \langle \mu_{k-1}, Q_k \phi \rangle, \end{aligned}$$

et

$$\langle g_k \cdot \mu_k^-, \phi \rangle = \int_E g_k \cdot \mu_k^-(dx') \phi(x') = \frac{\int_E \mu_k^-(dx') g_k(x') \phi(x')}{\int_E \mu_k^-(dx') g_k(x')} = \frac{\langle \mu_k^-, g_k \phi \rangle}{\langle \mu_k^-, g_k \rangle}.$$

PREUVE. En utilisant la représentation probabiliste (5.3) de la loi conditionnelle des états cachés $X_{0:k}$ sachant les observations $Y_{0:k}$, on obtient la factorisation

$$\begin{aligned} \mathbb{P}[X_{0:k} \in dx_{0:k} \mid Y_{0:k}] &\propto \eta_0(dx_0) \prod_{p=1}^k Q_p(x_{p-1}, dx_p) \prod_{p=0}^k g_p(x_p) \\ &\propto g_k(x_k) Q_k(x_{k-1}, dx_k) \mathbb{P}[X_{0:k-1} \in dx_{0:k-1} \mid Y_{0:k-1}]. \end{aligned}$$

En intégrant par rapport aux variables $x_{0:k-1}$ (et dans le membre de droite, d'abord par rapport aux variables $x_{0:k-2}$ puis par rapport à la variable x_{k-1}), on obtient la loi conditionnelle de l'état caché X_k sachant les observations $Y_{0:k}$, c'est-à-dire le filtre bayésien, comme

$$\begin{aligned} \mu_k(dx_k) &= \mathbb{P}[X_k \in dx_k \mid Y_{0:k}] \\ &\propto g_k(x_k) \int_E Q_k(x_{k-1}, dx_k) \mathbb{P}[X_{k-1} \in dx_{k-1} \mid Y_{0:k-1}] \\ &\propto g_k(x_k) \int_E \mu_{k-1}(dx_{k-1}) Q_k(x_{k-1}, dx_k) \\ &\propto g_k(x_k) \mu_k^-(dx_k). \quad \square \end{aligned}$$

L'équation du filtre bayésien optimal a été obtenue très simplement, mais il est en général impossible de la résoudre, sauf dans le cas particulier des systèmes linéaires gaussiens, où elle se ramène aux équations du filtre de Kalman, présentées au Chapitre 3. Il faut donc avoir recours à une approximation numérique, et on présente ci-dessous une approximation de type Monte Carlo, appelée filtre particulaire, qui a connu un développement spectaculaire au cours des dernières années, et qui est maintenant largement répandu, en particulier dans les applications en localisation, navigation ou poursuite de mobiles, aussi bien dans le domaine militaire (aéronautique, sous-marin, bâtiment de surface, missile, drone, etc.), que dans le domaine civil, avec des applications en robotique mobile ou en communications sans-fil.

5.4 Approximation particulaire

On rappelle que la suite $\{\mu_k\}$ vérifie l'équation récurrente

$$\mu_{k-1} \xrightarrow{\text{prédiction}} \eta_k = \mu_{k-1} Q_k \xrightarrow{\text{correction}} \mu_k = g_k \cdot \eta_k ,$$

d'après le Théorème 5.7. L'idée du filtrage particulaire consiste à chercher une approximation des distributions de probabilité conditionnelle $\eta_k(dx)$ et $\mu_k(dx)$ sous la forme de distributions de probabilité empiriques pondérées

$$\eta_k \approx \eta_k^N = \sum_{i=1}^N v_k^i \delta_{\xi_k^i} \quad \text{et} \quad \mu_k \approx \mu_k^N = \sum_{i=1}^N w_k^i \delta_{\xi_k^i} \quad \text{avec} \quad \sum_{i=1}^N w_k^i = 1 ,$$

associées à une population de N particules, caractérisée par

- les positions $(\xi_k^1, \dots, \xi_k^N)$ dans l'espace d'état E ,
- et les poids positifs normalisés (v_k^1, \dots, v_k^N) et (w_k^1, \dots, w_k^N) compris entre 0 et 1.

L'algorithme d'approximation est complètement décrit par le mécanisme qui permet de mettre à jour les *positions* et les *poids* des particules. Si on applique le noyau markovien $Q_k(x, dx')$ à l'approximation

$$\mu_{k-1}^N = \sum_{i=1}^N w_{k-1}^i \delta_{\xi_{k-1}^i} ,$$

on obtient exactement

$$\begin{aligned} \langle \mu_{k-1}^N Q_k, \phi \rangle &= \int \mu_{k-1}^N(dx) \int Q_k(x, dx') \phi(x') \\ &= \sum_{i=1}^N w_{k-1}^i \int Q_k(\xi_{k-1}^i, dx') \phi(x') \\ &= \int \left[\sum_{i=1}^N w_{k-1}^i Q_k(\xi_{k-1}^i, dx') \right] \phi(x') , \end{aligned}$$

pour toute fonction mesurable bornée ϕ définie sur E , c'est-à-dire que l'approximation

$$\mu_{k-1}^N Q_k(dx') = \sum_{i=1}^N w_{k-1}^i Q_k(\xi_{k-1}^i, dx'),$$

est un *mélange fini* de distributions de probabilité, peu pratique à manipuler, et qu'on décide de remplacer par la distribution de probabilité empirique pondérée

$$\eta_k^N = \sum_{i=1}^N v_k^i \delta_{\xi_k^i},$$

associée à un N -échantillon $(\xi_k^1, \dots, \xi_k^N)$ de variables aléatoires. On peut par exemple décider de conserver les poids, c'est-à-dire que indépendamment pour tout $i = 1, \dots, N$ on pose $v_k^i = w_{k-1}^i$ et on génère une variable aléatoire ξ_k^i distribuée selon $Q_k(\xi_{k-1}^i, dx')$. Alternativement, on peut décider de générer un échantillon ayant précisément pour loi commune $\mu_{k-1}^N Q_k(dx')$. Plus généralement, indépendamment pour tout $i = 1, \dots, N$

- (i) on sélectionne un individu $\widehat{\xi}_{k-1}^i$ au sein de la population courante et en fonction des poids respectifs, ce qui peut se réaliser de nombreuses manières différentes,
- (ii) on génère une variable aléatoire ξ_k^i selon la loi $Q_k(\widehat{\xi}_{k-1}^i, dx')$, ce qui est facile par hypothèse.

Si on applique ensuite la formule de Bayes à l'approximation $\eta_k^N(dx')$, on obtient exactement

$$\langle g_k \cdot \eta_k^N, \phi \rangle = \frac{\langle \eta_k^N, g_k \phi \rangle}{\langle \eta_k^N, \phi \rangle} = \frac{\sum_{i=1}^N v_k^i g_k(\xi_k^i) \phi(\xi_k^i)}{\sum_{j=1}^N v_k^j g_k(\xi_k^j)}$$

pour toute fonction mesurable bornée ϕ définie sur E , c'est-à-dire que l'approximation

$$\mu_k^N = g_k \cdot \eta_k^N = \sum_{i=1}^N w_k^i \delta_{\xi_k^i} \quad \text{avec} \quad w_k^i = \frac{v_k^i g_k(\xi_k^i)}{\sum_{j=1}^N v_k^j g_k(\xi_k^j)} \quad \text{pour tout } i = 1, \dots, N$$

est automatiquement sous la forme recherchée.

En résumé, on obtient deux types d'algorithmes.

Algorithme SIS (sequential importance sampling) Le premier type d'algorithme peut être décrit de la façon suivante.

- pour $k = 0$, indépendamment pour tout $i = 1, \dots, N$, on simule une variable aléatoire ξ_0^i selon la distribution de probabilité η_0 , et on pose

$$w_0^i = \frac{g_0(\xi_0^i)}{\sum_{j=1}^N g_0(\xi_0^j)}.$$

- pour tout $k = 1, \dots, n$, indépendamment pour tout $i = 1, \dots, N$
 - on simule une variable aléatoire ξ_k^i selon la distribution de probabilité $Q_k(\xi_{k-1}^i, dx')$,

et on pose

$$w_k^i = \frac{w_{k-1}^i g_k(\xi_k^i)}{\sum_{j=1}^N w_{k-1}^j g_k(\xi_k^j)} .$$

Algorithme SIR (sampling with importance resampling) Le deuxième type d'algorithme, appelé aussi filtre particulière *bootstrap*, peut être décrit de la façon suivante.

- pour $k = 0$, indépendamment pour tout $i = 1, \dots, N$, on simule une variable aléatoire ξ_0^i selon la distribution de probabilité η_0 , et on pose

$$w_0^i = \frac{g_0(\xi_0^i)}{\sum_{j=1}^N g_0(\xi_0^j)} .$$

- pour tout $k = 1, \dots, n$, indépendamment pour tout $i = 1, \dots, N$
 - on sélectionne un individu $\widehat{\xi}_{k-1}^i$ au sein de la population courante et en fonction des poids respectifs,
 - on simule une variable aléatoire ξ_k^i selon la distribution de probabilité $Q_k(\widehat{\xi}_{k-1}^i, dx')$,

et on pose

$$w_k^i = \frac{g_k(\xi_k^i)}{\sum_{j=1}^N g_k(\xi_k^j)} .$$

Remarque 5.9 Etant donné une distribution de probabilité m définie comme un mélange fini

$$m = \sum_{i=1}^N w_i m_i \quad \text{avec} \quad \sum_{i=1}^N w_i = 1 ,$$

il n'est véritablement intéressant de sélectionner les composantes du mélange que si les poids (w_1, \dots, w_N) sont très déséquilibrés. Pour mesurer l'écart à l'équidistribution on peut utiliser par exemple la distance du χ^2 entre deux vecteurs de probabilité $p = (p_1, \dots, p_N)$ et $q = (q_1, \dots, q_N)$, définie par

$$\chi^2(p, q) = \sum_{i=1}^N q_i \left(\frac{p_i}{q_i} - 1 \right)^2 = \sum_{i=1}^N \frac{p_i^2}{q_i} - 1 .$$

En particulier pour $p = (w_1, \dots, w_N)$ et $q = (1/N, \dots, 1/N)$, il vient

$$0 \leq N \sum_{i=1}^N w_i^2 - 1 = \frac{N}{N_{\text{eff}}} - 1 ,$$

où N_{eff} est la taille *effective* de l'échantillon, définie par

$$1 \leq N_{\text{eff}} = 1 / \left[\sum_{i=1}^N w_i^2 \right] \leq N ,$$

et où l'égalité est atteinte à l'équidistribution, ce qui suggère de sélectionner les composantes si

$$\frac{N}{N_{\text{eff}}} - 1 \geq \chi_{\text{red}}^2 > 0 \quad \text{c'est-à-dire si} \quad N_{\text{eff}} \leq \alpha_{\text{red}} M ,$$

où le seuil $\alpha_{\text{red}} = 1/(1 + \chi_{\text{red}}^2) < 1$ reste à déterminer.

Algorithme SIR adaptatif En utilisant la Remarque 5.9, on obtient l'algorithme adaptatif suivant, qui combine les deux types d'algorithmes introduits précédemment.

- pour $k = 0$, indépendamment pour tout $i = 1, \dots, N$, on simule une variable aléatoire ξ_0^i selon la distribution de probabilité η_0 , et on pose

$$w_0^i = \frac{g_0(\xi_0^i)}{\sum_{j=1}^N g_0(\xi_0^j)} ,$$

et on définit la taille effective

$$N_{\text{eff}} = 1 / \left[\sum_{i=1}^N |w_0^i|^2 \right] .$$

- pour tout $k = 1, \dots, n$, si $N_{\text{eff}} \leq \alpha_{\text{red}} N$, alors indépendamment pour tout $i = 1, \dots, N$
 - on sélectionne un individu $\widehat{\xi}_{k-1}^i$ au sein de la population courante et en fonction des poids respectifs,
 - on simule une variable aléatoire ξ_k^i selon la distribution de probabilité $Q_k(\widehat{\xi}_{k-1}^i, dx')$,

et on pose

$$w_k^i = \frac{g_k(\xi_k^i)}{\sum_{j=1}^N g_k(\xi_k^j)} ,$$

sinon, si $N_{\text{eff}} > \alpha_{\text{red}} N$, alors indépendamment pour tout $i = 1, \dots, N$

- on simule une variable aléatoire ξ_k^i selon la distribution de probabilité $Q_k(\xi_{k-1}^i, dx')$,

et on pose

$$w_k^i = \frac{w_{k-1}^i g_k(\xi_k^i)}{\sum_{j=1}^N w_{k-1}^j g_k(\xi_k^j)} ,$$

et dans tous les cas, on définit la taille effective

$$N_{\text{eff}} = 1 / \left[\sum_{i=1}^N |w_k^i|^2 \right] .$$

Dans le cas particulier du système non-linéaire à bruits non-gaussiens décrit par (5.1) et (5.2), simuler une variable aléatoire X selon la distribution de probabilité $Q_k(x, dx')$ signifie simplement simuler une variable aléatoire W selon la distribution de probabilité $p_k^W(dw)$, et poser $X = f_k(x, W)$, et évaluer la fonction de vraisemblance $g_k(x')$ signifie simplement évaluer $q_k^V(Y_k - h_k(x'))$, d'où l'algorithme SIR suivant

- pour $k = 0$, indépendamment pour tout $i = 1, \dots, N$, on simule une variable aléatoire ξ_0^i selon la distribution de probabilité η_0 , et on pose

$$w_0^i = \frac{q_0^V(Y_0 - h_0(\xi_0^i))}{\sum_{j=1}^N q_0^V(Y_0 - h_0(\xi_0^j))} .$$

- pour tout $k = 1, \dots, n$, indépendamment pour tout $i = 1, \dots, N$
 - on sélectionne un individu $\widehat{\xi}_{k-1}^i$ au sein de la population courante et en fonction des poids respectifs,
 - on simule une variable aléatoire W_k^i selon la distribution de probabilité $p_k^W(dw)$,
 - on pose $\xi_k^i = f_k(\widehat{\xi}_{k-1}^i, W_k^i)$,

et on pose

$$w_k^i = \frac{q_k^V(Y_k - h_k(\xi_k^i))}{\sum_{j=1}^N q_k^V(Y_k - h_k(\xi_k^j))} .$$

Il s'agit d'une approximation numérique, très simple à mettre en œuvre puisqu'il suffit de savoir simuler des transitions indépendantes de la chaîne de Markov, et qui converge vers le filtre optimal lorsque le nombre N de particules utilisées pour les calculs tend vers l'infini. L'étape essentielle dans l'algorithme est l'étape de rééchantillonnage, qui sélectionne les particules ayant une forte vraisemblance, et concentre ainsi automatiquement la puissance de calcul disponible dans les régions d'intérêt de l'espace d'état E .

Chapitre 6

Borne de Cramér–Rao a posteriori

Pour évaluer la performance des algorithmes numériques de filtrage non-linéaire, y compris les nombreuses variantes du filtrage particulaire, il est utile de disposer d'une borne inférieure sur l'erreur commise par un estimateur quelconque de l'état caché. S'il s'agit d'estimer un paramètre fixe, il est bien connu que la matrice d'information de Fisher associée au modèle statistique permet d'obtenir une telle borne inférieure, sous le nom de borne de Cramér–Rao. Dans le cas du filtrage bayésien, il s'agit d'estimer un paramètre aléatoire (et dynamique), à savoir la suite des états cachés, pour lequel on dispose d'un modèle a priori : dans ce cadre bayésien, on peut utiliser la notion de borne de Cramér–Rao a posteriori, pour laquelle des algorithmes de calcul récursifs efficaces ont été obtenus [12, 11].

On considère le modèle général d'une chaîne de Markov partiellement observée, et on suppose qu'il existe

- pour $k = 0$, une densité jointe initiale

$$\mathbb{P}[X_0 \in dx, Y_0 \in dy] = r_0(x, y) dx \lambda_0^F(dy) ,$$

- pour tout $k = 1, \dots, n$, des densités de transition

$$\mathbb{P}[X_k \in dx', Y_k \in dy' \mid X_{k-1} = x, Y_{k-1} = y] = r_k(y, y', x, x') dx' \lambda_k^F(y, dy') ,$$

On peut poser dans ce cas

$$X_{0:n} = (X_0, \dots, X_n) \quad \text{et} \quad Y_{0:n} = (Y_0, \dots, Y_n) ,$$

et se ramener au problème statique considéré dans la Proposition 1.3 ci-dessus pour l'estimation du vecteur aléatoire

$$\phi(X_{0:n}) = X_n ,$$

sachant $Y_{0:n}$.

Théorème 6.1 *Sous les hypothèses de la Proposition 1.3, la matrice de corrélation de l'erreur d'estimation $(\psi(Y_{0:n}) - X_n)$ est minorée par la relation suivante*

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

et la matrice d'information de Fisher J_n peut se calculer de la façon récursive suivante

$$J_k = D_k^+ - D_k^* (J_{k-1} + D_k^-)^{-1} D_k ,$$

avec

$$D_k^- = -\mathbb{E}\left[\frac{\partial^2}{\partial x_{k-1}^2} \log r_k(Y_{k-1}, Y_k, X_{k-1}, X_k)\right] ,$$

$$D_k = -\mathbb{E}\left[\frac{\partial^2}{\partial x_{k-1} \partial x_k} \log r_k(Y_{k-1}, Y_k, X_{k-1}, X_k)\right] ,$$

$$D_k^+ = -\mathbb{E}\left[\frac{\partial^2}{\partial x_k^2} \log r_k(Y_{k-1}, Y_k, X_{k-1}, X_k)\right] .$$

PREUVE. La densité jointe des vecteurs aléatoires $X_{0:n}$ et $Y_{0:n}$ est donnée par

$$p_{0:n}(x_{0:n}, y_{0:n}) = r_0(x_0, y_0) \prod_{k=1}^n r_k(y_{k-1}, y_k, x_{k-1}, x_k) ,$$

d'où la log-densité

$$\begin{aligned} \log p_{0:n}(x_{0:n}, y_{0:n}) &= \log r_0(x_0, y_0) + \sum_{k=1}^n \log r_k(y_{k-1}, y_k, x_{k-1}, x_k) \\ &= \log p_{0:n-1}(x_{0:n-1}, y_{0:n-1}) + \log r_n(y_{n-1}, y_n, x_{n-1}, x_n) . \end{aligned}$$

On en déduit que

$$\begin{aligned}
\frac{\partial^2}{\partial x_{0:n}^2} \log p_{0:n}(x_{0:n}, y_{0:n}) &= \begin{pmatrix} \frac{\partial^2}{\partial x_{0:n-1}^2} & \frac{\partial^2}{\partial x_{0:n-1} \partial x_n} \\ \star & \frac{\partial^2}{\partial x_n^2} \end{pmatrix} \log p_{0:n}(x_{0:n}, y_{0:n}) \\
&= \begin{pmatrix} \frac{\partial^2}{\partial x_{0:n-2}^2} & \frac{\partial^2}{\partial x_{0:n-2} \partial x_{n-1}} & \frac{\partial^2}{\partial x_{0:n-2} \partial x_n} \\ \star & \frac{\partial^2}{\partial x_{n-1}^2} & \frac{\partial^2}{\partial x_{n-1} \partial x_n} \\ \star & \star & \frac{\partial^2}{\partial x_n^2} \end{pmatrix} [\log p_{0:n-1}(x_{0:n-1}, y_{0:n-1}) \\
&\quad + \log r_n(y_{n-1}, y_n, x_{n-1}, x_n)] \\
&= \begin{pmatrix} \frac{\partial^2}{\partial x_{0:n-2}^2} & \frac{\partial^2}{\partial x_{0:n-2} \partial x_{n-1}} & 0 \\ \star & \frac{\partial^2}{\partial x_{n-1}^2} & 0 \\ 0 & 0 & 0 \end{pmatrix} \log p_{0:n-1}(x_{0:n-1}, y_{0:n-1}) \\
&\quad + \begin{pmatrix} 0 & 0 & 0 \\ 0 & \frac{\partial^2}{\partial x_{n-1}^2} & \frac{\partial^2}{\partial x_{n-1} \partial x_n} \\ 0 & \star & \frac{\partial^2}{\partial x_n^2} \end{pmatrix} \log r_n(y_{n-1}, y_n, x_{n-1}, x_n)
\end{aligned}$$

de sorte que si on pose

$$J_{0:n} = -\mathbb{E}\left[\frac{\partial^2}{\partial x_{0:n}^2} \log p_{0:n}(X_{0:n}, Y_{0:n})\right] = \begin{pmatrix} A_n & B_n \\ \star & C_n \end{pmatrix}, \quad (6.1)$$

alors on a également la décomposition par bloc alternative

$$J_{0:n} = \begin{pmatrix} A_{n-1} & B_{n-1} & 0 \\ \star & C_{n-1} & 0 \\ 0 & 0 & 0 \end{pmatrix} + \begin{pmatrix} 0 & 0 & 0 \\ 0 & D_n^- & D_n \\ 0 & \star & D_n^+ \end{pmatrix} = \begin{pmatrix} A_{n-1} & B_{n-1} & 0 \\ \star & C_{n-1} + D_n^- & D_n \\ 0 & \star & D_n^+ \end{pmatrix}, \quad (6.2)$$

avec

$$\begin{aligned} D_n^- &= -\mathbb{E}\left[\frac{\partial^2}{\partial x_{n-1}^2} \log r_n(Y_{n-1}, Y_n, X_{n-1}, X_n)\right], \\ D_n &= -\mathbb{E}\left[\frac{\partial^2}{\partial x_{n-1} \partial x_n} \log r_n(Y_{n-1}, Y_n, X_{n-1}, X_n)\right], \\ D_n^+ &= -\mathbb{E}\left[\frac{\partial^2}{\partial x_n^2} \log r_n(Y_{n-1}, Y_n, X_{n-1}, X_n)\right]. \end{aligned}$$

On remarque que $\phi'(x_{0:n}) = (0 \ I) = M_n$ ne dépend pas de $x_{0:n}$, et il résulte de la Proposition 1.3 que

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

et d'après le Lemme A.3 d'inversion matricielle et la Remarque A.4, on a

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

où la matrice $J_n = C_n - B_n^* A_n^{-1} B_n$ est le complément de Schur de la matrice A_n dans la matrice–bloc $J_{0:n}$. Par comparaison des deux décompositions par bloc (6.1) et (6.2), on a les identifications suivantes

$$A_n = \begin{pmatrix} A_{n-1} & B_{n-1} \\ \star & C_{n-1} + D_n^- \end{pmatrix}, \quad B_n = \begin{pmatrix} 0 \\ D_n \end{pmatrix} \quad \text{et} \quad C_n = D_n^+,$$

et d'après le Lemme A.3 d'inversion matricielle et la Remarque A.4, on a

$$\begin{aligned} J_n &= D_n^+ - (0 \ D_n^*) \begin{pmatrix} A_{n-1} & B_{n-1} \\ B_{n-1}^* & C_{n-1} + D_n^- \end{pmatrix}^{-1} \begin{pmatrix} 0 \\ D_n \end{pmatrix} \\ &= D_n^+ - (0 \ D_n^*) \begin{pmatrix} \star & \star \\ \star & \Delta_{n-1}^{-1} \end{pmatrix} \begin{pmatrix} 0 \\ D_n \end{pmatrix} \\ &= D_n^+ - D_n^* \Delta_{n-1}^{-1} D_n, \end{aligned}$$

où la matrice $\Delta_{n-1} = C_{n-1} + D_n^- - B_{n-1}^* A_{n-1}^{-1} B_{n-1} = J_{n-1} + D_n^-$ est le complément de Schur de la matrice A_{n-1} dans la matrice–bloc A_n , de sorte que

$$J_n = D_n^+ - D_n^* (J_{n-1} + D_n^-)^{-1} D_n. \quad \square$$

Soit un modèle de Markov caché dans lequel il existe

- pour $k = 0$, une densité initiale

$$\mathbb{P}[X_0 \in dx] = p_0(x) dx,$$

- pour tout $k = 1, \dots, n$, des densités de transition

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

- et pour tout $k = 0, 1, \dots, n$, des densités d'émission

$$\mathbb{P}[Y_k \in dy \mid X_k = x] = q_k(y \mid x) \lambda_k^F(dy) .$$

Ce modèle est un cas particulier du modèle précédent, avec

$$r_0(y, x) = p_0(x) q_0(y \mid x) ,$$

et

$$r_k(y, y', x, x') = p_k(x' \mid x) q_k(y' \mid x') ,$$

et dans ce cas

$$\begin{aligned} D_k^- &= -\mathbb{E}\left[\frac{\partial^2}{\partial x_{k-1}^2} \log p_k(X_k \mid X_{k-1})\right] , \\ D_k &= -\mathbb{E}\left[\frac{\partial^2}{\partial x_{k-1} \partial x_k} \log p_k(X_k \mid X_{k-1})\right] , \\ D_k^+ &= -\mathbb{E}\left[\frac{\partial^2}{\partial x_k^2} \log p_k(X_k \mid X_{k-1})\right] - \mathbb{E}\left[\frac{\partial^2}{\partial x_k^2} \log q_k(Y_k \mid X_k)\right] . \end{aligned}$$

Exemple 6.2 Dans le cas particulier d'un système avec bruits gaussiens additifs, où

$$X_k = f_k(X_{k-1}) + W_k \quad \text{avec} \quad W_k \sim \mathcal{N}(0, Q_k^W) ,$$

et où

$$Y_k = h_k(X_k) + V_k \quad \text{avec} \quad V_k \sim \mathcal{N}(0, Q_k^V) ,$$

avec des matrices de covariance Q_k^W et Q_k^V inversibles, on obtient

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

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

$$D_k^+ = (Q_k^W)^{-1} + \mathbb{E}[[h'_k(X_k)]^* (Q_k^V)^{-1} h'_k(X_k)] .$$

Annexe A

Inversion matricielle

Lemme A.1 Soit Q et R deux matrices symétriques définies positives, de dimension m et d respectivement, et soit H une matrice $d \times m$. Alors

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

où toutes les matrices inverses sont bien définies, et de plus

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

PREUVE. On remarque d'abord que

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

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

$$H Q H^* + R \quad \text{et} \quad H^* R^{-1} H + Q^{-1}$$

sont inversibles. En développant, on vérifie que

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

et d'autre part, en multipliant à droite par H^* , on obtient

$$\begin{aligned} (H^* R^{-1} H + Q^{-1})^{-1} H^* &= Q H^* - Q H^* (H Q H^* + R)^{-1} (H Q H^* + R - R) \\ &= Q H^* (H Q H^* + R)^{-1} R . \quad \square \end{aligned}$$

Remarque A.2 Cette formule 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 $d \leq m$ en général. En particulier, dans le cas où $d = 1$, la matrice $H = h^*$ est un vecteur ligne, la matrice $R = r$ est un scalaire, et la formule devient

$$\left(\frac{h h^*}{r} + Q^{-1}\right)^{-1} = Q - \frac{Q h h^* Q}{r + h^* Q h}.$$

Lemme A.3 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, $\det M = \det \Delta \cdot \det D$ de sorte que la matrice M est inversible si et seulement si la matrice Δ est inversible, et

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

Si la matrice M est symétrique, ce qui implique en particulier que $A = A^*$, $C = B^*$ et $D = D^*$, alors le complément de Schur $\Delta = A - B D^{-1} B^*$ est également symétrique, et si en outre la matrice M est semi-définie positive, respectivement définie positive, alors la matrice Δ est également semi-définie positive, respectivement définie positive.

Remarque A.4 Si la matrice A est inversible, alors la matrice $\Delta = D - C A^{-1} B$ est appelée complément de Schur de la matrice A dans la matrice-bloc M , la matrice M est inversible si et seulement si la matrice Δ est inversible, et

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

PREUVE. En développant, on vérifie que

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

ce qui montre l'identité annoncée. On en déduit que $\det M = \det \Delta \cdot \det D$ de sorte que la matrice M est inversible si et seulement si la matrice Δ est inversible, et

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

et on remarque que

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

Si la matrice M est symétrique, on remarque que

$$(I \quad -BD^{-1}) \begin{pmatrix} A & B \\ B^* & D \end{pmatrix} \begin{pmatrix} I \\ -D^{-1}B^* \end{pmatrix} = (A - BD^{-1}B^* \quad 0) \begin{pmatrix} I \\ -D^{-1}B^* \end{pmatrix} = \Delta,$$

de sorte que

$$(u^* \quad -u^*BD^{-1}) \begin{pmatrix} A & B \\ B^* & D \end{pmatrix} \begin{pmatrix} u \\ -D^{-1}B^*u \end{pmatrix} = u^* \Delta u,$$

pour tout vecteur u , ce qui permet de conclure. \square

Bibliographie

- [1] Brian D. O. Anderson and John B. Moore. *Optimal filtering*. Prentice–Hall Information and System Sciences Series. Prentice–Hall, Englewood Cliffs, NJ, 1979.
- [2] Olivier Cappé, Éric Moulines, and Tobias Rydén. *Inference in hidden Markov models*. Springer Series in Statistics. Springer–Verlag, New York, 2005.
- [3] Randal Douc, Éric Moulines, and David S. Stoffer. *Nonlinear time series : Theory, methods and applications with R examples*. Texts in Statistical Science. Chapman & Hall / CRC Press, Boca Raton, 2014.
- [4] Arnaud Doucet, Nando de Freitas, and Neil Gordon, editors. *Sequential Monte Carlo methods in practice*. Statistics for Engineering and Information Science. Springer–Verlag, New York, 2001.
- [5] Arnaud Doucet, Simon J. Godsill, and Christophe Andrieu. On sequential Monte Carlo sampling methods for Bayesian filtering. *Statistics and Computing*, 10(3) :197–208, July 2000.
- [6] James Durbin and Siem Jan Koopman. *Time series analysis by state space methods*, volume 38 of *Oxford Statistical Science Series*. Oxford University Press, Oxford, 2nd edition, 2012.
- [7] Fredrik Gustafsson, Fredrik Gunnarsson, Niclas Bergman, Urban Forssell, Jonas Jansson, Rickard Karlsson, and Per-Johan Nordlund. Particle filters for positioning, navigation, and tracking. *IEEE Transactions on Signal Processing*, SP-50(2 (Special issue on Monte Carlo Methods for Statistical Signal Processing)) :425–437, February 2002.
- [8] Patrick Pérez, Carine Hue, Jaco Vermaak, and Michel Gangnet. Color–based probabilistic tracking. In Anders Heyden, Gunnar Sparr, Mads Nielsen, and Peter Johansen, editors, *Proceedings of the 7th European Conference on Computer Vision (ECCV’02), Copenhagen 2002*, volume 2350 of *Lecture Notes in Computer Science*, pages 661–675. Springer–Verlag, Berlin, June 2002.
- [9] Branko Ristić, M. Sanjeev Arulampalam, and Neil J. Gordon. *Beyond the Kalman filter : Particle filters for tracking applications*. Artech House, Boston, 2004.
- [10] Sebastian Thrun, Wolfram Burgard, and Dieter Fox. *Probabilistic robotics*. Intelligent Robotics and Autonomous Agents. The MIT Press, Cambridge, MA, 2005.

- [11] Petr Tichavský, Carlos H. Muravchik, and Arye Nehorai. Posterior Cramér–Rao bounds for discrete–time nonlinear filtering. *IEEE Transactions on Signal Processing*, SP–46(5) :1386–1396, May 1998.
- [12] Miroslav Šimandl, Jakub Královec, and Petr Tichavský. Filtering, predictive and smoothing Cramér–Rao bounds for discrete–time nonlinear dynamic systems. *Automatica*, 37(11) :1703–1716, November 2001.