

Filtrage Optimal

FRÉDÉRIC ROTELLA
Ecole Nationale d'Ingénieurs de Tarbes

rotella@enit.fr

Table des matières

1	Introduction	3
2	Le filtre de Kalman discret	3
2.0.1	Equations du filtre	4
2.0.2	Formes particulières du filtre	5
2.0.3	Résultats supplémentaires	6
2.0.4	Modèle à bruits corrélés	7
2.0.5	Le filtre Information	10
2.1	Le filtre de Kalman continu	11
2.1.1	Relations formelles entre les cas continu et discret	12
2.1.2	Détermination du gain optimal	13
2.1.3	Filtre de Kalman-Bucy	14
2.2	Commentaires sur le filtre de Kalman	15
2.2.1	Cas où certaines sorties sont non bruitées	15
2.2.2	Modèle à bruits colorés	17
2.2.3	Filtre stationnaire	18
3	Mise en œuvre d'un filtre de Kalman	19
3.1	La forme de Joseph	19
3.2	Traitement séquentiel des observations	20
3.3	Amélioration des performances	22
3.3.1	Algorithmes à factorisations	22
3.3.2	Algorithmes de filtrage rapide	24
4	Applications du filtrage	24
4.1	Commande optimale stochastique	25
4.2	Lissage	26
4.2.1	Principe du lissage	26
4.2.2	Equations du filtre lisseur	27

4.3	Identification	28
4.3.1	Estimation de paramètres	29
4.3.2	Forme filtre	30
5	Annexe A : Estimation optimale	31
5.1	Estimateur du maximum de vraisemblance	31
5.2	Estimation au sens des moindres carrés	32
5.3	Estimation linéaire	33
5.4	Estimateur optimal de variables gaussiennes	35
5.5	Estimation récursive	36
6	Annexe B : Démonstration des équations du filtre de Kalman	37
6.1	Forme prédicteur-à-un-pas	38
6.2	Décomposition du filtre	39
7	Annexe C : Racine carrée	39
7.1	Lemme de factorisation matricielle	40
7.2	Décomposition de Cholewsky	40
	Références	42

1 Introduction

Le problème du filtrage consiste à déterminer des estimateurs de variables du système lorsque l'environnement présente des perturbations aléatoires. Nous allons donc étudier dans cette partie l'aspect stochastique de la notion d'observateurs. Deux points de vue peuvent être utilisés pour aborder cette question : celui de Wiener qui utilise une approche fréquentielle et celui de Kalman qui utilise l'approche temporelle. Dans tous les cas, le but est de déterminer un système (**le filtre**), optimal au sens de la minimisation de la variance d'erreur entre la variable réelle et son estimation. Nous ne regarderons ici que la deuxième approche. En effet, un filtre de Wiener est un cas particulier de filtre de Kalman et cette dernière approche permet d'appréhender directement le cas d'un système non stationnaire multivariable.

Dans ce cadre on peut également classer les problèmes d'estimation suivant la quantité d'information disponible. En effet, considérons un système dont on possède un ensemble de mesures $m(t_0, t_f)$, entre les instants t_0 (instant initial) et t_f (instant final), sur les entrées et les sorties. On peut chercher à estimer la valeur de l'état x à un instant donné τ (que l'on notera par $\hat{x}(\tau/m(t_0, t_f))$). Suivant la valeur de τ , on distingue :

- si $\tau < t_f$ il s'agit d'un problème de **lissage** ;
- si $\tau = t_f$ il s'agit d'un problème de **filtrage** ;
- si $\tau > t_f$ il s'agit d'un problème de **prédiction**.

Alors qu'un problème de prédiction peut être ramené à un problème de filtrage par une estimation de $\hat{x}_f = \hat{x}(t_f/m(t_0, t_f))$ suivie d'une prédiction par utilisation du modèle initialisé à x_f , il n'en est pas de même du lissage. En fait, ce dernier problème peut être résolu par la combinaison de deux problèmes de filtrage : un filtrage de t_0 à τ et un filtrage rétrograde de t_f à τ . Ainsi, un seul algorithme permettra de traiter ces trois problèmes.

Le filtre de Kalman est un reconstituteur d'état dans un environnement stochastique. Lorsque les variances des bruits sont connues, c'est un estimateur linéaire minimisant la variance de l'erreur d'estimation. Les algorithmes donnant la solution de ce problème ont été déterminés initialement par [KALMAN, 1960] dans le cas discret et [KALMAN, BUCY, 1961] dans le cas continu. Nous établirons, dans un premier temps, les équations du filtre de Kalman discret puis, celles du filtre de Kalman continu par passage à la limite.

2 Le filtre de Kalman discret

Etant donné un système linéaire stochastique dont l'évolution dynamique est modélisée à l'aide de l'équation d'état :

$$\begin{aligned} x_{k+1} &= A_k x_k + B_k u_k + G_k w_k, \\ y_k &= C_k x_k + v_k, \end{aligned} \tag{1}$$

où $k \geq 0$ représente les instants successifs du temps, x_k , l'état du système de dimension n , y_k , la sortie (mesure ou observation) de dimension m , u_k , l'entrée certaine de dimension l , w_k , le bruit d'entrée (ou de dynamique) de dimension l , v_k , le bruit de mesure de dimension m . Les

matrices certaines A_k, B_k, G_k, C_k sont de dimensions convenables. Pour éviter toute confusion, nous représenterons dans la suite, la matrice identité d'ordre n , simplement par I .

Ce modèle peut être considéré comme représentatif d'un système à temps discret ou plus généralement être obtenu à partir de la discrétisation d'un modèle représentatif d'un système à temps continu. Les séquences de bruit $\{w_k\}$ et $\{v_k\}$ sont des séquences indépendantes de bruits blancs centrés et l'état initial x_0 est également une variable aléatoire indépendante des séquences $\{w_k\}$ et $\{v_k\}$. Leurs propriétés aux premier et second ordres sont données par :

$$\begin{aligned} E\{w_k\} = 0, \quad E\{v_k\} = 0, \quad E\{x_0\} = \bar{x}_0, \\ E \left\{ \begin{bmatrix} v_k \\ w_k \\ \tilde{x}_0 \end{bmatrix} [v_l^T \ w_l^T \ \tilde{x}_0^T] \right\} = \begin{bmatrix} R_k \delta_{kl} & & \\ & Q_k \delta_{kl} & \\ & & P_0 \end{bmatrix}, \end{aligned} \quad (2)$$

où $E\{\cdot\}$ représente l'espérance mathématique, $\tilde{x}_0 = x_0 - \bar{x}_0$, R_k, Q_k, P_0 sont des matrices symétriques définies positives, et δ_{kl} est le symbole de Kroknecker.

2.0.1 Equations du filtre

Le problème du filtrage, au sens de Kalman, est de trouver, pour le système dynamique (??), la meilleure estimation \hat{x} de l'état x à l'instant k , à partir d'observations effectuées jusqu'à l'instant discret j , au sens du critère de la variance conditionnelle minimum. Cela signifie que l'estimé \hat{x} est tel que :

$$E\{\|x_k - \hat{x}\|^2 / \{y_0, y_1, \dots, y_j\}\} \leq E\{\|x_k - z\|^2 / \{y_0, y_1, \dots, y_j\}\},$$

pour tout vecteur z fonction des observations $\{y_0, y_1, \dots, y_j\}$. Nous noterons $\hat{x}_{k/j}$, cet estimateur optimal, et $\tilde{x}_{k/j}$ et $\tilde{y}_{k/j}$ les erreurs d'estimations :

$$\begin{aligned} \tilde{x}_{k/j} &= x_k - \hat{x}_{k/j}, \\ \tilde{y}_{k/j} &= y_k - C_k \hat{x}_{k/j}. \end{aligned}$$

Suivant que $k < j, k = j$ ou $k > j$, on dit que $\hat{x}_{k/j}$ est une valeur lissée, filtrée ou prédite de x_k . Contrairement au problème de prédiction qui utilise, à partir d'une valeur filtrée le modèle (??) non bruité, le problème de lissage nécessitera un traitement plus complexe (on doit tenir compte d'informations futures) qui sera étudié dans un paragraphe particulier. Pour simplifier les notations, nous poserons désormais :

$$\begin{aligned} \text{cov}(z) &= E\{zz^T\}, \\ P_{k/t} &= \text{cov}(\tilde{x}_{k/t}). \end{aligned}$$

Le problème de filtrage au sens de Kalman est résolu en utilisant les principes de base de l'estimation, simple puis récursive, d'une variable aléatoire que nous décrivons dans l'annexe A. Dans l'annexe B, on établit que les équations de fonctionnement d'un filtre de Kalman discret se décomposent en deux étapes :

– une étape de prédiction :

$$\begin{aligned}\hat{x}_{k+1/k} &= A_k \hat{x}_{k/k} + B_k u_k, \\ P_{k+1/k} &= A_k P_{k/k} A_k^T + G_k Q_k G_k^T;\end{aligned}\tag{3}$$

– une étape de correction :

$$\begin{aligned}\hat{x}_{k/k} &= \hat{x}_{k/k-1} + K_k (y_k - C_k \hat{x}_{k/k-1}), \\ P_{k/k} &= (I - K_k C_k) P_{k/k-1},\end{aligned}\tag{4}$$

où K_k est le gain optimal du filtre, donné par :

$$\begin{aligned}K_k &= P_{k/k-1} C_k^T \Sigma_k^{-1}, \\ \Sigma_k &= R_k + C_k P_{k/k-1} C_k^T.\end{aligned}\tag{5}$$

La construction de cet algorithme montre qu'il permet de répondre à deux objectifs différents :

- c'est un filtre linéaire minimisant la variance *a priori* de l'erreur d'estimation. Dans ces conditions, les bruits peuvent ne pas être gaussiens ;
- c'est un filtre maximisant la probabilité *a posteriori* des grandeurs à estimer. Cela n'est alors applicable que dans l'hypothèse de bruits gaussiens.

2.0.2 Formes particulières du filtre

Le filtre de Kalman discret est composé de l'ensemble des relations (3) à (5), mais, suivant que l'étape de prédiction suit ou précède l'étape de correction, on peut réaliser un filtre estimateur ou un filtre prédicteur-à-un-pas.

On obtient alors les formes suivantes :

- $(\hat{x}_{k/k}, P_{k/k}) \longrightarrow (\hat{x}_{k+1/k+1}, P_{k+1/k+1})$, on réalise alors un filtre estimateur :

$$\begin{aligned}\hat{x}_{k+1/k+1} &= (I - K_{k+1} C_{k+1}) A_k \hat{x}_{k/k} \\ &\quad + (I - K_{k+1} C_{k+1}) B_k u_k + K_{k+1} y_{k+1}, \\ P_{k+1/k+1} &= (I - K_{k+1} C_{k+1}) (A_k P_{k/k} A_k^T + G_k Q_k G_k^T), \\ K_{k+1} &= (A_k P_{k/k} A_k^T + G_k Q_k G_k^T) C_{k+1}^T \Sigma_{k+1}^{-1}, \\ \Sigma_{k+1} &= R_{k+1} + C_{k+1} (A_k P_{k/k} A_k^T + G_k Q_k G_k^T) C_{k+1}^T;\end{aligned}\tag{6}$$

- $(\hat{x}_{k/k-1}, P_{k,k-1}) \longrightarrow (\hat{x}_{k+1/k}, P_{k+1/k})$, on réalise alors un filtre prédicteur-à-un-pas :

$$\begin{aligned}\hat{x}_{k+1/k} &= A_k (I - K_k C_k) \hat{x}_{k/k-1} + B_k u_k + A_k K_k y_k, \\ P_{k+1/k} &= A_k (I - K_k C_k) P_{k/k-1} A_k^T + G_k Q_k G_k^T,\end{aligned}\tag{7}$$

où K_k est définie en (5).

Bien que ces expressions n'apportent aucune simplification, il est nécessaire de remarquer que, dans le filtre de Kalman, le filtre estimateur et le filtre prédicteur-à-un-pas, les expressions permettant de calculer le gain K_k et les matrices de covariance $P_{k/k}$ et $P_{k+1/k}$ sont indépendantes des mesures y_k et des valeurs de $\hat{x}_{k/k}$ et $\hat{x}_{k+1/k}$. Elles peuvent donc être calculées à l'avance. On aboutit, à partir de (5), (4), et (7), à l'équation de récurrence fournissant $P_{k+1/k}$:

$$P_{k+1/k} = A_k P_{k/k-1} A_k^T + G_k Q_k G_k^T - A_k P_{k/k-1} C_k^T (R_k + C_k P_{k/k-1} C_k^T)^{-1} C_k P_{k/k-1} A_k^T, \quad (8)$$

qui est une équation de Riccati permettant de calculer $P_{k/k}$ et K_k par (4) et (5). Une fois connues ces matrices, les filtres peuvent être réalisés par (6) ou (7) suivant que l'on dispose ou non de la sortie à l'instant $k + 1$.

2.0.3 Résultats supplémentaires

Les matrices intervenant dans le filtre de Kalman ont les propriétés particulières suivantes :

1. L'élimination de C_k ou Σ_k dans (4) à l'aide de (5) conduit directement aux expressions :

$$\begin{aligned} P_{k/k} &= P_{k/k-1} - K_k \Sigma_k K_k^T, \\ &= P_{k/k-1} - P_{k/k-1} C_k^T (R_k + C_k P_{k/k-1} C_k^T)^{-1} C_k P_{k/k-1}. \end{aligned} \quad (9)$$

2. L'utilisation du lemme matriciel :

$$(A + BCD)^{-1} = A^{-1} - A^{-1}B(C^{-1} + DA^{-1}B)^{-1}DA^{-1}, \quad (10)$$

conduit directement, à partir de (9), à :

$$P_{k/k}^{-1} = P_{k/k-1}^{-1} + C_k^T R_k^{-1} C_k. \quad (11)$$

3. L'utilisation du lemme matriciel sur Σ_k conduit à :

$$\Sigma_k^{-1} = R_k^{-1} - R_k^{-1} C_k \underbrace{(P_{k/k-1}^{-1} + C_k^T R_k^{-1} C_k)^{-1}}_{P_{k/k}} C_k^T R_k^{-1}, \quad (12)$$

soit, d'après (5) :

$$K_k = P_{k/k-1} C_k^T R_k^{-1} - P_{k/k-1} \underbrace{C_k^T R_k^{-1} C_k}_{P_{k/k}^{-1} - P_{k/k-1}^{-1}} P_{k/k} C_k^T R_k^{-1}.$$

On obtient finalement :

$$K_k = P_{k/k} C_k^T R_k^{-1}. \quad (13)$$

Cette relation est d'interprétation plus facile que la formule initiale (5) qui doit être envisagée lorsque R_k est singulière. En effet, K_k est le gain de correction (4) et l'on peut faire le raisonnement qualitatif suivant :

- confiance dans les précédentes estimations ($P_{k/k-1}$ faible) et doute dans les mesures (R_k élevé) doivent impliquer un gain K_k faible ;
 - doute sur les précédentes estimations et confiance dans les mesures actuelles doit entraîner un gain de correction élevé,
- ce qui est vérifié par la forme (13).

4. Comme $\hat{y}_{k/k-1} = C_k \hat{x}_{k/k-1}$, il vient :

$$\tilde{y}_{k/k-1} = y_k - \hat{y}_{k/k-1} = C_k \tilde{x}_{k/k-1} + v_k. \quad (14)$$

La matrice de covariance de l'erreur de prédiction sur la sortie est alors donnée par :

$$\text{cov}(\tilde{y}_{k/k-1}) = C_k P_{k/k-1} C_k^T + R_k = \Sigma_k,$$

ce qui permet de donner une interprétation évidente de la matrice Σ_k .

5. De même que précédemment, on a :

$$\begin{aligned} \tilde{y}_{k/k} &= y_k - \hat{y}_{k/k}, \\ &= y_k - C_k \hat{x}_{k/k}, \\ &= C_k \tilde{x}_{k/k} + v_k. \end{aligned}$$

La matrice de covariance de l'erreur d'estimation sur la sortie est alors donnée par :

$$\text{cov}(\tilde{y}_{k/k}) = C_k P_{k/k} C_k^T + R_k.$$

6. D'après les relations précédentes :

$$\tilde{y}_{k/k-1} - \tilde{y}_{k/k} = C_k (\hat{x}_{k/k} - \hat{x}_{k/k-1}),$$

ce qui, d'après (3) et (14), conduit à :

$$\tilde{y}_{k/k-1} - \tilde{y}_{k/k} = C_k K_k (C_k \tilde{x}_{k/k-1} + v_k).$$

La matrice de covariance de l'écart estimation-prédiction, en utilisant (9), est donnée par :

$$\text{cov}(\tilde{y}_{k/k-1} - \tilde{y}_{k/k}) = C_k K_k \Sigma_k K_k^T C_k^T = C_k (P_{k/k-1} - P_{k/k}) C_k^T.$$

2.0.4 Modèle à bruits corrélés

Dans le cas où les bruits de dynamique et de mesure de (??) sont corrélés, il est possible de reprendre toute l'étude précédente. Une façon plus rapide d'établir les équations du filtre de Kalman correspondant consiste à se ramener au cas d'un modèle à bruits décorrélés par la construction d'un modèle équivalent.

En effet, supposons que l'on ait :

$$\text{E}\{v_k w_l^T\} = S_k^T \delta_{kl}.$$

On introduit alors le nouveau vecteur de bruit de dynamique :

$$\bar{w}_k = w_k - S_k R_k^{-1} v_k,$$

pour lequel il est facile de vérifier que $E\{v_k \bar{w}_l^T\} = 0$.

L'utilisation de ce bruit, dans les équations du modèle initial (??), conduit directement au modèle équivalent cherché :

$$\begin{aligned} x_{k+1} &= \bar{A}_k x_k + B_k u_k + \bar{G}_k y_k + G_k \bar{w}_k, \\ y_k &= C_k x_k + v_k, \end{aligned} \quad (15)$$

avec $\bar{A}_k = A_k - G_k S_k R_k^{-1} C_k$, $\bar{G}_k = G_k S_k R_k^{-1}$.

Les propriétés statistiques des bruits de ce modèle sont telles que :

$$\begin{aligned} E\{v_k\} &= 0, \quad E\{\bar{w}_k\} = 0, \\ E\left\{\begin{bmatrix} v_k \\ \bar{w}_k \end{bmatrix} \begin{bmatrix} v_l^T \\ \bar{w}_l^T \end{bmatrix}\right\} &= \begin{bmatrix} R_k & 0 \\ 0 & \bar{Q}_k \end{bmatrix} \delta_{kl}, \end{aligned}$$

où $\bar{Q}_k = Q_k - S_k R_k^{-1} S_k^T$.

Il suffit alors d'appliquer, sur le système (15), le filtre de Kalman discret défini dans le cas d'un modèle à bruits non corrélés. Les étapes de correction (4) et de calcul de gain (5) sont conservées. Seule est modifiée l'étape de prédiction qui devient :

$$\begin{aligned} \hat{x}_{k+1/k} &= \bar{A}_k \hat{x}_{k/k} + B_k u_k + \bar{G}_k y_k, \\ P_{k+1/k} &= \bar{A}_k P_{k/k} \bar{A}_k^T + G_k \bar{Q}_k G_k^T. \end{aligned} \quad (16)$$

En tenant compte des expressions de \bar{A}_k et \bar{G}_k , il vient :

$$\hat{x}_{k+1/k} = A_k \hat{x}_{k/k} + B_k u_k + G_k S_k R_k^{-1} (y_k - C_k \hat{x}_{k/k}),$$

et l'utilisation de (3) conduit à la forme :

$$\hat{x}_{k+1/k} = A_k \hat{x}_{k/k} + B_k u_k + G_k S_k R_k^{-1} (I - C_k K_k) (y_k - C_k \hat{x}_{k/k-1}).$$

Or, d'après (5), on a la relation suivante :

$$\begin{aligned} R_k^{-1} (I - C_k K_k) &= R_k^{-1} (I - C_k P_{k/k-1} C_k^T \Sigma_k^{-1}), \\ &= R_k^{-1} \underbrace{(\Sigma_k - C_k P_{k/k-1} C_k^T)}_{R_k} \Sigma_k^{-1}. \end{aligned}$$

L'étape de prédiction (5) est donc modifiée par ajout d'un terme correcteur :

$$\hat{x}_{k+1/k} = \underbrace{A_k \hat{x}_{k/k} + B_k u_k}_{\text{partie pour } S_k = 0} + \bar{K}_k (y_k - C_k \hat{x}_{k/k-1}), \quad (17)$$

où $\bar{K}_k = G_k S_k \Sigma_k^{-1}$.

De même, en développant la deuxième expression de (16), on obtient :

$$\begin{aligned} P_{k+1/k} &= A_k P_{k/k}^T A_k^T + G_k Q_k G_k^T - G_k S_k R_k^{-1} S_k^T G_k^T \\ &\quad + G_k S_k R_k^{-1} C_k P_{k/k} C_k^T R_k^{-1} S_k^T G_k^T \\ &\quad - G_k S_k R_k^{-1} C_k P_{k/k} A_k^T - A_k P_{k/k} C_k^T R_k^{-1} S_k^T G_k^T. \end{aligned}$$

L'utilisation des relations (12), (13) et (17), nous permet d'écrire :

$$\begin{aligned} P_{k+1/k} &= A_k P_{k/k} A_k^T + G_k Q_k G_k^T - \bar{K}_k \Sigma_k \bar{K}_k^T \\ &\quad - \bar{K}_k \Sigma_k K_k^T A_k^T - A_k K_k \Sigma_k \bar{K}_k^T. \end{aligned}$$

La détermination de la matrice de covariance de l'étape de prédiction du filtre de Kalman pour un système à bruits corrélés est donc déduite de celle obtenue dans le cas de bruits non corrélés par ajout d'un terme correcteur. On a alors :

$$P_{k+1/k} = \underbrace{A_k P_{k/k} A_k^T + G_k Q_k G_k^T}_{\text{partie pour } S_k = 0} + A_k K_k \Sigma_k K_k^T A_k^T - L_k \Sigma_k L_k^T, \quad (18)$$

où $L_k = \bar{K}_k + A_k K_k$.

En résumé, pour un système à bruits corrélés, il suffit de remplacer les relations (3) par les relations réajustées (17) et (18) pour obtenir le filtre de Kalman discret correspondant.

Lorsque l'on s'intéresse à la réalisation d'un filtre prédictif-à-un-pas pour un système à bruits corrélés, l'élimination de $\hat{x}_{k/k}$ et $P_{k/k}$ entre (17) et (18) conduit directement à :

$$\begin{aligned} \hat{x}_{k+1/k} &= (A_k - L_k C_k) \hat{x}_{k/k-1} + B_k u_k + L_k y_k, \\ P_{k+1/k} &= A_k P_{k/k-1} A_k^T + G_k Q_k G_k^T - L_k \Sigma_k L_k^T, \\ L_k &= (A_k P_{k/k-1} C_k^T + G_k S_k) (R_k + C_k P_{k/k-1} C_k^T)^{-1}, \end{aligned}$$

où les matrices $P_{k/k-1}$ et L_k peuvent être calculées à l'avance par la résolution de l'équation de Riccati :

$$\begin{aligned} P_{k+1/k} &= A_k P_{k/k-1} A_k^T + G_k Q_k G_k^T \\ &\quad - (A_k P_{k/k-1} C_k^T + G_k S_k) (R_k + C_k P_{k/k-1} C_k^T)^{-1} \\ &\quad (A_k P_{k/k-1} C_k^T + G_k S_k)^T. \end{aligned}$$

A l'aide du modèle équivalent à bruits non corrélés, on a donc déterminé le filtre de Kalman dans le cas général. Un autre avantage de ce modèle est de permettre la simulation de tout système caractérisé par des bruits corrélés. En effet, les programmes de génération de bruits blancs pseudo-aléatoires ne permettent pas de générer des bruits corrélés. L'utilisation du modèle équivalent avec bruits non corrélés permet d'utiliser des bruits non corrélés \bar{w}_k et v_k .

2.0.5 Le filtre Information

Le filtre Information est une formulation du filtre de Kalman en termes d'inverses de matrices de covariance (appelées matrices d'informations). Nous le présenterons dans le cas d'un système à bruits non corrélés, mais il va de soi que le cas d'un système à bruits corrélés serait traité de la même manière que précédemment par l'intermédiaire du modèle équivalent (15). Soient les estimations définies par les transformations :

$$\begin{aligned}\hat{z}_{k/k-1} &= P_{k/k-1}^{-1} \hat{x}_{k/k-1}, \\ \hat{z}_{k/k} &= P_{k/k}^{-1} \hat{x}_{k/k}.\end{aligned}\quad (19)$$

Le filtre Information est constitué des algorithmes de prédiction permettant de passer de $(\hat{z}_{k-1/k-1}, P_{k-1/k-1}^{-1})$ à $(\hat{z}_{k/k-1}, P_{k/k-1}^{-1})$ et des algorithmes de correction permettant de passer de $(\hat{z}_{k/k-1}, P_{k/k-1}^{-1})$ à $(\hat{z}_{k/k}, P_{k/k}^{-1})$.

La relation (11), que nous rappelons ici, donne le passage de $P_{k/k-1}^{-1}$ à $P_{k/k}^{-1}$:

$$P_{k/k}^{-1} = P_{k/k-1}^{-1} + C_k^T R_k^{-1} C_k. \quad (20)$$

L'utilisation du lemme matriciel d'inversion (10) sur la relation (3) conduit à :

$$P_{k+1/k}^{-1} = (I - L_k G_k^T) M_k,$$

où :

$$\begin{aligned}M_k &= A_k^{-T} P_{k/k}^{-1} A_k^{-1}, \\ L_k &= M_k G_k \Lambda_k^{-1}, \\ \Lambda_k &= Q_k^{-1} + G_k^T M_k G_k.\end{aligned}\quad (21)$$

Une dualité apparaît donc entre l'étape de prédiction (resp. correction) du filtre de Kalman et l'étape de correction (resp. prédiction) du filtre Information. Par analogie avec (4), la matrice L_k est la matrice de gain de correction du filtre Information. Ces calculs ne dépendant pas des observations, la même remarque que pour le filtre de Kalman s'applique donc ici, les matrices d'informations et le gain du filtre peuvent être calculés à l'avance.

On passe des estimations du filtre de Kalman à celles données par le filtre Information par les transformations (19). Cela fournit donc, d'après (4) et (13) :

$$P_{k/k}^{-1} \hat{x}_{k/k} = \underbrace{(P_{k/k}^{-1} - C_k^T R_k^{-1} C_k)}_{P_{k/k-1}^{-1}} \hat{x}_{k/k-1} + C_k^T R_k^{-1} y_k,$$

soit, pour l'étape de correction :

$$\hat{z}_{k/k} = \hat{z}_{k/k-1} + C_k^T R_k^{-1} y_k. \quad (22)$$

De même, à partir de (3), on obtient :

$$P_{k+1/k}^{-1} \hat{x}_{k+1/k} = P_{k+1/k}^{-1} A_k P_{k/k} \hat{z}_{k/k} + P_{k+1/k}^{-1} B_k u_k,$$

ce qui se transforme, par (21), en :

$$\hat{z}_{k+1/k} = (I - L_k G_k^T) A_k^{-T} \hat{z}_{k/k} + P_{k+1/k}^{-1} B_k u_k. \quad (23)$$

La dualité entre le filtre de Kalman et le filtre Information existe aussi à ce niveau mais est moins évidente car l'influence des mesures et des entrées reste respectivement sur les étapes de correction et de prédiction, alors que le gain passe de l'une à l'autre des étapes. Les relations (20) à (23) constituent les formules de base du filtre Information.

Bien que les filtres de Kalman et Information soient équivalents, ce dernier sera préféré au filtre de Kalman dans les cas suivants :

- lorsque les connaissances statistiques sur l'état initial sont mal connues (ce qui peut conduire à une valeur propre de P_0 très grande), le filtre Information permet alors de démarrer l'algorithme de filtrage avec des matrices finies jusqu'à ce que $P_{k/k}^{-1}$ et $P_{k+1/k}^{-1}$ soient de rang plein. A partir de ce moment on peut reprendre le filtre de Kalman ;
- lorsque la dimension du bruit de dynamique est inférieure à la dimension du bruit de mesure, par exemple lorsque le bruit de dynamique est un bruit d'entrée et que l'on en a moins d'entrées que de mesures. La matrice Λ_k a alors une taille plus petite que Σ_k , le calcul du gain dans le filtre Information est donc plus rapide que dans le filtre de Kalman.

2.1 Le filtre de Kalman continu

Le filtre de Kalman-Bucy [13] résout le problème de l'estimation de l'état d'un système continu défini par l'équation d'état :

$$\begin{aligned} \dot{x}(t) &= A(t)x(t) + B(t)u(t) + G(t)w(t), \\ y(t) &= C(t)x(t) + v(t), \end{aligned} \quad (24)$$

où t représente le temps, $x(t)$, l'état de dimension n , $y(t)$, la mesure de dimension m , $u(t)$, l'entrée certaine de dimension l , $w(t)$, le bruit d'entrée (ou de dynamique) de dimension l' , et $v(t)$, le bruit de mesure de dimension m . On suppose que ces bruits sont blancs, gaussiens, et connus par leurs matrices de covariance :

$$\begin{aligned} \mathbb{E}\{w(t)w^T(t')\} &= Q(t)\delta_{t-t'}, \quad \mathbb{E}\{v(t)v^T(t')\} = R(t)_{t-t'}, \\ \mathbb{E}\{v(t)w^T(t')\} &= 0, \quad \mathbb{E}\{v(t)\tilde{x}^T(0)\} = 0, \quad \mathbb{E}\{w(t)\tilde{x}^T(0)\} = 0, \\ \mathbb{E}\{\tilde{x}(0)\tilde{x}^T(0)\} &= P_0, \end{aligned} \quad (25)$$

où δ_t est l'impulsion de Dirac en t , et en considérant $x(0)$ comme une variable aléatoire d'espérance m_0 , $\tilde{x}(0) = x(0) - m_0$.

Il serait possible en utilisant les méthodes d'estimation de variables aléatoires, détaillées dans l'annexe A, de déterminer directement la structure du système linéaire fournissant la meilleure estimation $\hat{x}(t/\tau)$, au sens de la variance d'erreur minimale, de l'état de (24) à t à partir de la connaissance de la mesure y depuis l'instant initial jusqu'à l'instant τ (la démonstration est analogue au cas des systèmes discrets). Nous déterminerons ici le filtre de Kalman dans le cas

continu, par passage à la limite à partir des équations du filtre de Kalman dans le cas discret, pour un problème de filtrage ($\tau = t$).

En toute rigueur, pour éviter tout problème de définition en ce qui concerne les matrices de covariance, il serait préférable d'utiliser la représentation différentielle d'Ito ([2, 21]) d'un système continu stochastique :

$$\begin{aligned} dx_t &= A_t x_t dt + B_t u_t dt + G_t d\beta_t, \\ dy_t &= C_t x_t dt + d\eta_t, \end{aligned}$$

où $d\beta_t$ et $d\eta_t$ sont des mouvements browniens centrés tels que :

$$\begin{aligned} \lim_{dt \rightarrow 0} \frac{1}{dt} E\{d\beta_t d\beta_t^T\} &= Q(t), \\ \lim_{dt \rightarrow 0} \frac{1}{dt} E\{d\eta_t d\eta_t^T\} &= R(t). \end{aligned}$$

Nous ne le ferons pas, car cette représentation, ainsi que sa justification, fait appel à des notions complexes en calcul stochastique. Bien que l'assimilation de $w(t)$ avec $d\beta_t^T/dt$ et $v(t)$ avec $d\eta_t^T/dt$ soit abusive en ce sens que β_t et η_t n'admettent pas de dérivées, les deux types de représentation sont *formellement* équivalentes ([6, 7]) et nous utiliserons le modèle (24), en conservant à l'esprit que le bruit blanc est une limite qui n'existe pas.

2.1.1 Relations formelles entre les cas continu et discret

Le passage du modèle continu au modèle discret se fait en deux étapes, d'une part par discrétisation du modèle (24) avec une période d'échantillonnage constante et d'autre part, par passage à la limite en faisant tendre la période d'échantillonnage vers 0. Nous allons ici établir les relations nécessaires entre les modèles discret (??) et continu (24) imposées par ces transformations.

Soit Δ la période de discrétisation, la relation suivante (formelle) entre le symbole de Kronecker et l'impulsion de Dirac :

$$\lim_{\Delta \rightarrow 0} \frac{\delta_{ij}}{\Delta} = \delta_{(i-j)\Delta},$$

impose les relations entre les matrices de covariances discrètes et continues :

$$\begin{aligned} \lim_{\Delta \rightarrow 0} Q_k \Delta &= Q(t_k), \\ \lim_{\Delta \rightarrow 0} R_k \hat{E} \Delta &= R(t_k). \end{aligned}$$

D'autre part, la discrétisation du modèle continu, par la méthode simple d'approximation d'Euler, conduit à, lorsque $\Delta \rightarrow 0$:

$$\begin{aligned} x_{k+1} &= (I + A_k \Delta) x_k + \Delta B_k u_k + \Delta G_k w_k, \\ y_k &= C_k x_k + v_k, \end{aligned}$$

où l'indice k représente une évaluation pour en $k\Delta$.

L'utilisation du filtre de Kalman sur ce système donne, d'après l'équation du prédicteur-à-un-pas (7) :

$$\hat{x}_{k+1/k} = (I + A_k \Delta) \hat{x}_{k/k-1} + \Delta B_k u_k + (I + A_k \Delta) K_k (y_k - C_k \hat{x}_{k/k-1}),$$

ce qui peut s'écrire sous la forme :

$$\frac{\hat{x}_{k+1/k} - \hat{x}_{k/k-1}}{\Delta} = A_k \hat{x}_{k/k-1} + B_k u_k + (I + A_k \Delta) \frac{K_k}{\Delta} (y_k - C_k \hat{x}_{k/k-1}). \quad (26)$$

En considérant :

$$\begin{aligned} \lim_{\Delta \rightarrow 0} \frac{\hat{x}_{k+1/k} - \hat{x}_{k/k-1}}{\Delta} &= \dot{\hat{x}}(t), \\ \lim_{\Delta \rightarrow 0} \frac{K_k}{\Delta} &= K(t), \\ \lim_{\Delta \rightarrow 0} \hat{x}_{k/k-1} &= \hat{x}(t), \end{aligned} \quad (27)$$

on obtient à partir de (26), l'équation de l'estimateur linéaire continu du système (24) :

$$\dot{\hat{x}}(t) = A(t) \hat{x}(t) + B(t) u(t) + K(t) (y(t) - C(t) \hat{x}(t)). \quad (28)$$

2.1.2 Détermination du gain optimal

Le gain optimal de l'estimateur est déterminé également par passage à la limite à partir des équations discrètes. D'après (27) et (5), on a :

$$\begin{aligned} K(t) &= \lim_{\Delta \rightarrow 0} \frac{K_k}{\Delta}, \\ &= \lim_{\Delta \rightarrow 0} P_{k/k-1} C_k^T \underbrace{(C_k P_{k/k-1} C_k^T \Delta + R_k \Delta)}_{\rightarrow 0}^{-1} \underbrace{R_k \Delta}_{\rightarrow R(t)}. \end{aligned}$$

Soit, en notant $\lim_{\Delta \rightarrow 0} P_{k/k-1} = P(t)$, il vient l'expression du gain optimal du filtre (28) :

$$K(t) = P(t) C^T(t) R(t)^{-1}.$$

En écrivant (4), sous la forme :

$$P_{k/k} = P_{k/k-1} - \underbrace{\frac{K_k}{\Delta}}_{\rightarrow K(t)} \underbrace{C_k P_{k/k-1} \Delta}_{\rightarrow 0}.$$

on arrive à :

$$\lim_{\Delta \rightarrow 0} P_{k/k} = \lim_{\Delta \rightarrow 0} P_{k/k-1} = P(t).$$

L'équation de Riccati récurrente (8) s'écrit ici :

$$\begin{aligned} P_{k+1/k} &= (I + A_k \Delta) \\ &[P_{k/k-1} - P_{k/k-1} C_k^T (R_k + C_k P_{k/k-1} C_k^T)^{-1} C_k P_{k/k-1}] \\ &(I + A_k^T \Delta) + \Delta G_k Q_k G_k^T \Delta, \end{aligned}$$

ce qui se met sous la forme :

$$\begin{aligned} \frac{P_{k+1/k} - P_{k/k-1}}{\Delta} &= G_k \underbrace{Q_k \Delta}_{\rightarrow Q(t)} G_k^T + A_k P_{k/k-1} + P_{k/k-1} A_k^T \\ &\quad - (I + A_k \Delta) P_{k/k-1} C_k^T \left(\underbrace{\Delta R_k}_{\rightarrow R(t)} + \underbrace{\Delta C_k P_{k/k-1} C_k^T}_{\rightarrow 0} \right)^{-1} \\ &\quad C_k P_{k/k-1} (I + A_k^T \Delta) + \Delta A_k P_{k/k-1} A_k^T. \end{aligned}$$

En prenant la limite, quand $\Delta \rightarrow 0$, de chacun des deux termes de cette égalité on aboutit à l'équation différentielle de Riccati définissant l'évolution de la matrice de covariance de l'erreur d'estimation :

$$\dot{P}(t) = G(t)Q(t)G^T(t) + A(t)P(t) + P(t)A^T(t) - P(t)C^T(t)R(t)^{-1}C(t)P(t). \quad (29)$$

2.1.3 Filtre de Kalman-Bucy

En résumé, le filtre linéaire optimal du système continu stochastique a pour structure :

$$\dot{\hat{x}}(t) = A(t)\hat{x}(t) + B(t)u(t) + K(t)[y(t) - C(t)\hat{x}(t)], \quad (30)$$

avec $\hat{x}(0) = m_0$ et $K(t) = P(t)C^T(t)R(t)^{-1}$, où $P(t)$ est la solution de l'équation de Riccati (29) telle que $P(0) = P_0$. On peut noter ici l'analogie entre les expressions de gains optimaux discret (13) et continu (30).

Dans le cas où les bruits de mesure et de dynamique sont corrélés :

$$E\{w(t)v^T(t')\} = S(t)\delta_{t-t'},$$

on peut construire, comme dans le cas des systèmes discrets, un filtre de Kalman-Bucy, en faisant appel au modèle équivalent à bruits non corrélés. Ce modèle est défini par :

$$\begin{aligned} \dot{x}(t) &= \bar{A}(t)x(t) + B(t)u(t) + G(t)\bar{w}(t) + \bar{G}(t)y(t), \\ y(t) &= C(t)x(t) + v(t), \end{aligned}$$

où :

$$\begin{aligned} \bar{A}(t) &= A(t) - G(t)S(t)R^{-1}(t)C(t), \\ \bar{G}(t) &= G(t)S(t)R^{-1}(t), \\ \bar{w}(t) &= w(t) - S(t)R^{-1}(t)v(t). \end{aligned} \quad (31)$$

Dans ce modèle équivalent, $\bar{w}(t)$ est un bruit de dynamique décorrélé de $v(t)$, dont la variance est donnée par :

$$E\{\bar{w}(t)\bar{w}^T(t')\} = \bar{Q}(t)\delta_{t-t'}, \quad \bar{Q}(t) = Q(t) - S(t)R^{-1}(t)S^T(t). \quad (32)$$

L'utilisation de l'estimateur (30), sur ce système, conduit au filtre :

$$\dot{\hat{x}}(t) = \bar{A}(t)\hat{x}(t) + B(t)u(t) + \bar{G}(t)y(t) + \bar{K}(t)[y(t) - C(t)\hat{x}(t)],$$

avec $\bar{K}(t) = \bar{P}(t)C^T(t)R^{-1}(t)$, où $\bar{P}(t)$ est la solution de l'équation différentielle de Riccati :

$$\dot{\bar{P}}(t) = G(t)\bar{Q}(t)G^T(t) + \bar{A}(t)\bar{P}(t) + \bar{P}(t)\bar{A}^T(t) - \bar{P}(t)C^T(t)R^{-1}(t)C(t)\bar{P}(t),$$

telle que $\bar{P}(0) = P_0$.

Ce qui, compte tenu des relations (31) et (32), se met sous la forme :

$$\dot{\hat{x}}(t) = A(t)\hat{x}(t) + B(t)u(t) + L(t)[y(t) - C(t)\hat{x}(t)],$$

où :

$$L(t) = \bar{K}(t) + G(t)S(t)R^{-1}(t) = [\bar{P}(t)C^T(t) + G(t)S(t)]R^{-1}(t),$$

et $\bar{P}(t)$ est la solution de :

$$\begin{aligned} \dot{\bar{P}}(t) = & G(t)Q(t)G^T(t) + A(t)\bar{P}(t) + \bar{P}(t)A^T(t) \\ & - [\bar{P}(t)C^T(t) + G(t)S(t)]R^{-1}(t)[\bar{P}(t)C^T(t) + G(t)S(t)]^T, \end{aligned}$$

telle que $\bar{P}(0) = P_0$.

Ces relations correspondent au filtre de Kalman-Bucy dans le cas d'un système à bruits corrélés.

2.2 Commentaires sur le filtre de Kalman

2.2.1 Cas où certaines sorties sont non bruitées

L'existence de mesures non bruitées implique la singularité de la matrice de covariance des bruits de mesure. Il s'ensuit une difficulté d'application des filtres de Kalman (particulièrement dans le cas continu). Nous allons montrer que l'application de la méthode de simplification, utilisée dans le cas des observateurs déterministes [20], permet de contourner cette difficulté et conduit à une réduction notable du volume des calculs. Par raison de simplicité nous établirons les relations dans le cas des systèmes continus mais elles peuvent bien sûr être déterminées pour le modèle discret.

Considérons le système continu stationnaire :

$$\begin{aligned} \dot{x}(t) &= Ax(t) + Bu(t) + Gw(t), \\ y_1(t) &= C_1x(t), \\ y_2(t) &= C_2x(t) + v_2(t), \end{aligned}$$

où $x(t) \in \mathbb{R}^n$, $y_1(t) \in \mathbb{R}^{m_1}$, représente les mesures non bruitées, $y_2(t) \in \mathbb{R}^{m_2}$, les mesures bruitées, $w(t)$ et $v_2(t)$ sont des bruits centrés, non corrélés, connus par leurs matrices de covariance :

$$\begin{aligned} E\{w(t)w^T(t')\} &= Q\delta_{t-t'}, \\ E\{v_2(t)v_2^T(t')\} &= R_2\delta_{t-t'}. \end{aligned}$$

En supposant C_1 de rang plein (ce qui est toujours possible par élimination de mesures redondantes), un changement de variables élémentaire permet de se ramener à la forme :

$$\begin{aligned}\dot{X}_1(t) &= A_{11}X_1(t) + A_{12}X_2(t) + B_1u(t) + G_1w(t), \\ \dot{X}_2(t) &= A_{21}X_1(t) + A_{22}X_2(t) + B_2u(t) + G_2w(t), \\ y_1(t) &= X_1(t), \\ y_2(t) &= C_{21}X_1(t) + C_{22}X_2(t) + v_2(t).\end{aligned}$$

La mesure $y_1(t)$ donnant une estimation certaine de $X_1(t)$, il reste à construire un estimateur fournissant une estimation de $X_2(t)$. A partir des nouvelles mesures :

$$\begin{aligned}y'_2(t) &= y_2(t) - C_{21}X_1(t), \\ z(t) &= \dot{X}_1(t) - A_{11}X_1(t) - B_1u(t),\end{aligned}$$

les équations donnant $X_2(t)$ se ramènent à :

$$\begin{aligned}\dot{X}_2(t) &= A_{22}X_2(t) + A_{21}y_1(t) + B_2u(t) + G_2w(t), \\ Z(t) &= CX_2(t) + V(t), \\ Z(t) &= \begin{bmatrix} z(t) \\ y'_2(t) \end{bmatrix} \quad C = \begin{bmatrix} A_{12} \\ C_{22} \end{bmatrix}, \quad V(t) = \begin{bmatrix} G_1w(t) \\ v_2(t) \end{bmatrix}.\end{aligned}$$

Ces équations font donc apparaître un modèle à bruits de dynamique et de sortie corrélés, dont les matrices de covariance sont :

$$\begin{aligned}E\{w(t)w^T(t')\} &= Q\delta_{t-t'}, \\ E\{v(t)v^T(t')\} &= \begin{bmatrix} G_1QG_1^T & O \\ O & R_2 \end{bmatrix} \delta_{t-t'} = \bar{R}\delta_{t-t'}, \\ E\{v(t)w^T(t')\} &= \begin{bmatrix} G_1Q \\ 0 \end{bmatrix} \delta_{t-t'} = \bar{S}\delta_{t-t'}.\end{aligned}$$

On construit alors le filtre de Kalman sur ce système par la méthode développée précédemment sur un système à bruits corrélés. Celà donne la structure :

$$\dot{\hat{X}}_2(t) = A_{22}\hat{X}_2(t) + A_{21}y_1(t) + B_2u(t) + K(Z(t) - C\hat{X}_2(t)),$$

où K est le gain optimal.

De la même façon que dans le cadre des systèmes déterministes, ce filtre implique (pour la connaissance de $z(t)$, donc de $Z(t)$) la dérivation de la sortie $y_1(t)$. Pour contourner cette dernière difficulté, on introduit la variable :

$$Y(t) = \hat{X}_2(t) - K_1y_1(t),$$

où $K = [K_1, K_2]$, dont l'équation d'évolution se met sous la forme :

$$\begin{aligned}\dot{Y}(t) &= [A_{22} - KC]Y(t) + [B_2 - K_1B_1]u(t) + K_2y_2(t) \\ &\quad + [A_{21} - K_1A_{11} - K_2C_{21} + (A_{22} - KC)K_1]y_1(t).\end{aligned}$$

L'estimation optimale de $X_2(t)$ est alors fournie par la variable :

$$\hat{X}_2(t) = Y(t) + K_1 y_1(t).$$

Rappelons que dans le cas d'un système discret, une démarche analogue peut être proposée et conduit également à un estimateur d'ordre réduit.

2.2.2 Modèle à bruits colorés

Le cas que l'on vient d'étudier n'intervient pas seulement lorsque certaines mesures sont parfaites mais également lorsque l'hypothèse de bruits de dynamique et de sortie blancs n'est plus vérifiée. Supposons en effet, que les matrices de covariance de ces bruits sont de la forme (par exemple, dans le cas discret) :

$$\begin{aligned} E\{w_k w_\ell^T\} &= Q_{k-\ell}, \\ E\{v_k v_\ell^T\} &= R_{k-\ell}. \end{aligned} \tag{33}$$

Comme la transmission à travers un filtre linéaire stationnaire d'un bruit blanc génère, pour les variances de l'état ou de la sortie, de telles séquences de matrices de covariance. Ces séquences peuvent donc être considérées comme des réponses impulsionnelles. On cherche ainsi deux systèmes, appelés filtres formateurs, dont l'entrée est une séquence blanche $\{\nu_k\}$, telle que $E\{\nu_k^T \nu_l\} = N_k \delta_{kl}$, et dont la sortie fournit les bruits w_k et v_k .

L'application d'une méthode de réalisation conduit à modéliser les systèmes générant $\{w_k\}$ et $\{v_k\}$ à partir de $\{\nu_k\}$ sous la forme équations d'état :

$$\begin{cases} d_{k+1} = A^d d_k + G^d \nu_k, \\ w_k = C^d d_k + L^d \nu_k \\ m_{k+1} = A^m m_k + G^m \nu_k, \\ v_k = C^m m_k + L^m \nu_k \end{cases}$$

L'application du filtre de Kalman se fait alors sur le système à état-augmenté, soit à partir de (??) :

$$\begin{aligned} \begin{bmatrix} x_{k+1} \\ d_{k+1} \\ m_{k+1} \end{bmatrix} &= \begin{bmatrix} A_k & G_k C^d & 0 \\ 0 & A^d & 0 \\ 0 & 0 & A^m \end{bmatrix} \begin{bmatrix} x_k \\ d_k \\ m_k \end{bmatrix} + \begin{bmatrix} G_k L^d \\ G^d \\ G^m \end{bmatrix} \nu_k + \begin{bmatrix} B_k \\ 0 \\ 0 \end{bmatrix} u_k, \\ y_k &= [C_k \quad 0 \quad C^m] \begin{bmatrix} x_k \\ d_k \\ m_k \end{bmatrix} + L^m \nu_k, \end{aligned}$$

qui est un système où les bruits sont blancs mais les bruits de sortie ou de mesure sont corrélés ou, si $\det(L^m N_k L^{mT}) = 0$, la covariance des bruits de mesure est singulière.

Nous verrons, dans la partie consacrée aux applications, que le filtre de Kalman est un filtre formateur particulier.

2.2.3 Filtre stationnaire

Dans le cas de systèmes stationnaires, c'est-à-dire lorsque les paramètres ne dépendent pas du temps, les gains optimaux des filtres de Kalman continus et discrets sont donnés à partir de la solution des équations de Riccati suivantes :

– cas continu :

$$\begin{aligned}\dot{P}(t) &= GQG^T + AP(t) + P(t)A^T - P(t)C^T R^{-1}CP(t), \\ P(0) &= P_0;\end{aligned}\tag{34}$$

– cas discret :

$$\begin{aligned}P(k+1) &= AP(k)A^T + GQG^T \\ &\quad - AP(k)C^T[R + CP(k)C^T]^{-1}CP(k)A^T, \\ P(0) &= P_0;\end{aligned}\tag{35}$$

où les matrices A , G , C , Q et R sont des matrices constantes définissant l'équation d'état du processus considéré. Nous n'entrerons pas dans l'analyse de la stabilité de ces équations, mais on a le résultat suivant :

Théorème 1 *Si (A, C) est une paire observable, le filtre optimal est asymptotiquement stable, quelles que soient les matrices GQG^T et R , définies positives et quelle que soit l'initialisation P_0 . La matrice de covariance de l'erreur d'estimation tend vers la solution unique et définie positive des équations algébriques de Riccati :*

– pour le cas continu :

$$0 = GQG^T + AP + PA^T - PC^T R^{-1}CP;\tag{36}$$

– pour le cas discret :

$$P = APA^T + GQG^T - APC^T[R + CPC^T]^{-1}CPA^T.\tag{37}$$

On a alors intérêt, dans un but de simplification, à remplacer le gain optimal $K(t)$ (resp. K_k) du filtre de Kalman, calculé à partir de $P(t)$ (resp. P_k), par (34) (resp. (35)), par le gain optimal constant donné par :

– pour le cas continu :

$$K_c = PC^T R^{-1};$$

– pour le cas discret :

$$K_d = PC^T[R + CPC^T]^{-1};$$

où P est solution de l'équation continue (36) ou discrète (37) suivant le cas.

La différence entre le filtre optimal et le filtre stationnaire sous-optimal ne se fera sentir qu'en début de fonctionnement (on perd en effet des informations initiales) mais le gain en temps de calcul peut être, en contrepartie, appréciable. D'autre part, on peut montrer [9] que le filtre de Kalman stationnaire ainsi déterminé est identique au filtre de Wiener déterminé à partir d'une réalisation du signal aléatoire $x(t)$ [14]. Des méthodes numériques de résolution de ces équations de Riccati, qui interviennent également en commande optimale [4, 16] ou en commande robuste [25] sont décrites dans [3, 15].

3 Mise en œuvre d'un filtre de Kalman

Concernant l'implantation numérique d'un filtre de Kalman, en particulier dans le cas discret, on peut faire quelques remarques qui vont permettre d'améliorer le comportement numérique des équations. En effet les algorithmes à mettre en œuvre contiennent des équations matricielles de récurrence et il y a lieu de conserver certaines propriétés et ceci indépendamment des erreurs de troncature. D'autre part, le volume d'information à traiter simultanément et la résolution numérique des équations de Riccati ne doivent pas conduire à des temps de calcul trop longs. Nous décrirons, ici succinctement, des méthodes d'amélioration de chacun de ces points en renvoyant à [10] pour des explications plus détaillées.

3.1 La forme de Joseph

La forme (4) ne garantit pas, lors du passage de l'étape de prédiction à l'étape de correction, la conservation des propriétés de symétrie des matrices de covariances tout au long du fonctionnement du filtre. Afin de pallier à cet inconvénient, on peut utiliser une relation pour le calcul de $P_{k/k}$, dite "forme de Joseph". L'écriture de l'étape de correction sous la forme :

$$\hat{x}_{k/k} = (I - K_k C_k) \hat{x}_{k/k-1} + K_k y_k,$$

permet d'obtenir, pour l'erreur d'estimation :

$$\tilde{x}_{k/k} = (I - K_k C_k) \tilde{x}_{k/k-1} - K_k v_k.$$

En prenant la covariance de ce vecteur, on obtient une expression de $P_{k/k}$ sous "forme de Joseph" :

$$P_{k/k} = (I - K_k C_k) P_{k/k-1} (I - K_k C_k)^T + K_k R_k K_k^T, \quad (38)$$

qui a l'avantage de faire intervenir la somme de deux matrices définies positives ou semi-définies positives. Par conséquent, la symétrie et la définie positivité de $P_{k/k}$ sont numériquement mieux préservées que par (4). D'autre part, la "forme de Joseph" est moins sensible à de petites erreurs sur K_k . En effet, si l'on désigne par δK_k une erreur commise lors du calcul de K_k , l'utilisation de (4) conduit à une erreur $\delta P_{k/k}$ du premier ordre en δK_k :

$$\delta P_{k/k} = -(\delta K_k) C_k P_{k/k-1},$$

alors que (38) conduit à :

$$\begin{aligned} \delta P_{k/k} &= -(\delta K_k) C_k P_{k/k-1} (I - K_k C_k)^T \\ &\quad - (I - K_k C_k) P_{k/k-1} C_k^T (\delta K_k)^T \\ &\quad + (\delta K_k) R_k K_k^T + K_k R_k (\delta K_k)^T + o((\delta K_k)^2), \end{aligned}$$

ce qui, après simplification à l'aide de (5), donne :

$$\delta P_{k/k} = o((\delta K_k)^2).$$

La “forme de Joseph” entraîne donc une erreur du second ordre en δK_k . A l’aide de (3), (38) conduit à la “forme de Joseph” pour $P_{k/k-1}$:

$$P_{k+1/k} = A_k(I - K_k C_k)P_{k/k-1}(I - K_k C_k)^T A_k^T + A_k K_k R_k K_k^T A_k^T + G_k Q_k G_k^T,$$

qui implique également une erreur $\delta P_{k+1/k}$ du deuxième ordre en δK_k .

En ce qui concerne le filtre Information, on peut également déterminer d’autres relations, numériquement mieux adaptées, permettant le calcul de $P_{k+1/k}^{-1}$.

D’une part, il est en effet facile de vérifier que la “forme de Joseph” de $P_{k+1/k}^{-1}$ s’écrit :

$$P_{k+1/k}^{-1} = (I - L_k G_k^T)M_k(I - L_k G_k^T)^T + L_k Q_k^{-1} L_k^T.$$

D’autre part, l’élimination de L_k et de Λ_k dans (21) conduit à :

$$P_{k+1/k}^{-1} = M_k - M_k G_k (Q_k + G_k^T M_k G_k)^{-1} G_k^T M_k.$$

Or, d’après (20) et (21), M_{k+1} et $P_{k+1/k}^{-1}$ sont reliées par :

$$M_{k+1} = A_{k+1}^{-T} [P_{k+1/k}^{-1} + C_{k+1}^T R_{k+1}^{-1} C_{k+1}] A_{k+1}^{-1}.$$

La suite des matrices M_k qui permettent de calculer les matrices d’informations par (21) et $P_{k/k}^{-1} = A_k^T M_k A_k$, peut être déterminée par la récurrence de Riccati :

$$\begin{aligned} M_{k+1} = & A_{k+1}^{-T} [C_{k+1}^T R_{k+1}^{-1} C_{k+1} + M_k \\ & - M_k G_k (Q_k + G_k^T M_k G_k)^{-1} G_k^T M_k] A_{k+1}^{-1}. \end{aligned}$$

3.2 Traitement séquentiel des observations

Dans le cas où les mesures proviennent de sources statistiquement indépendantes, la matrice de covariance des bruits de sortie est diagonale, soit :

$$R_k = \text{diag}_{i=1}^{\nu} \{R_k^i\}$$

et, en notant C_k^i la i -ième ligne de C_k , l’on peut partitionner l’équation de sortie de (??) sous la forme :

$$\begin{bmatrix} y_k^1 \\ \dots \\ \vdots \\ \dots \\ y_k^\nu \end{bmatrix} = \begin{bmatrix} C_k^1 \\ \dots \\ \vdots \\ \dots \\ C_k^\nu \end{bmatrix} x_k + \begin{bmatrix} v_k^1 \\ \dots \\ \vdots \\ \dots \\ v_k^\nu \end{bmatrix}.$$

Le traitement séquentiel des observations consiste à traiter séquentiellement les composantes y_k^1, \dots, y_k^ν lors de l’étape de correction, à l’aide de l’algorithme suivant :

– initialisation :

$$\hat{x}_k^0 = \hat{x}_{k/k-1}, \quad P_k^0 = P_{k/k-1};$$

– itération pour $i = 1$ à ν :

$$\begin{aligned}
\Sigma_k^i &= C_k^i P_k^{i-1} C_k^{iT} + R_k^i, \\
K_k^i &= P_k^{i-1} C_k^{iT} (\Sigma_k^i)^{-1}, \\
\hat{x}_k^i &= \hat{x}_k^{i-1} + K_k^i (y_k^i - C_k^i \hat{x}_k^{i-1}), \\
P_k^i &= [I - K_k^i C_k^i] P_k^{i-1} [I - K_k^i C_k^i]^T + K_k^i R_k^i K_k^{iT};
\end{aligned} \tag{39}$$

– estimation et covariance corrigées :

$$\hat{x}_{k/k} = \hat{x}_k^\nu, \quad P_{k/k} = P_k^\nu.$$

La démonstration de ces relations est immédiate à partir de la relation (11). Il vient, compte tenu du partitionnement des matrices C_k et R_k :

$$P_{k/k}^{-1} = P_{k/k-1}^{-1} + \sum_{i=1}^{\nu} C_k^{iT} (R_k^i)^{-1} C_k^i,$$

d'où l'on peut tirer la formule de récurrence :

$$(P_k^i)^{-1} = (P_k^{i-1})^{-1} + C_k^{iT} (R_k^i)^{-1} C_k^i, \quad i = \{1, \dots, \nu\},$$

avec $P_k^0 = P_{k/k-1}$ et $P_k^\nu = P_{k/k}$.

L'utilisation du lemme d'inversion matricielle conduit à la forme analogue à (9) :

$$P_k^i = P_k^{i-1} - P_k^{i-1} C_k^{iT} (\Sigma_k^i)^{-1} C_k^i P_k^{i-1}, \tag{40}$$

où Σ_k^i est définie en (39). La dernière relation de (39) correspond à la forme de Joseph de (40).

Cet algorithme permet une réduction notable du volume des calculs (notamment par l'inversion de matrices de dimensions plus petites). Une économie supplémentaire en temps de calcul peut être réalisée dans le cas où R_k est régulière. En effet, nous avons vu que, dans ce cas, la valeur corrigée est donnée par :

$$\hat{x}_{k/k} = \hat{x}_{k/k-1} + P_{k/k} C_k^T R_k^{-1} (y_k - C_k \hat{x}_{k/k-1}),$$

et on peut éliminer de l'algorithme de traitement séquentiel toutes les estimations intermédiaires $\hat{x}_k^i, i = \{1, \dots, \nu\}$.

D'autre part, dans le cas où R_k n'est pas diagonale, cet algorithme est utilisable moyennant un changement de variables de mesures. Dans le cas où R_k est régulière, l'algorithme de Cholewski (annexe C) permet de la factoriser sous la forme :

$$R_k = T_k T_k^T,$$

où T_k est une matrice triangulaire inférieure. Le changement de variables sur la sortie :

$$y_k^* = T_k^{-1} y_k,$$

conduit à l'équation de sortie :

$$y_k^* = C_k^* x_k + v_k^*,$$

où $C_k^* = T_k^{-1} C_k$ et $v_k^* = T_k^{-1} v_k$. La covariance des bruits sur cette sortie devient :

$$R_k^* = E\{v_k^* v_k^{*T}\} = T_k^{-1} R_k (T_k^T)^{-1} = I,$$

qui est une matrice diagonale.

Nous venons de montrer que l'on peut se ramener à une matrice diagonale, on peut donc toujours aboutir à un traitement séquentiel ne nécessitant que l'inversion de scalaires.

3.3 Amélioration des performances

L'une des principales difficultés de la mise en œuvre du filtre de Kalman réside dans la résolution de l'équation de Riccati donnant les matrices de covariance. Ce traitement implique de nombreuses opérations arithmétiques qui peuvent entraîner, par erreur d'arrondis, la non positivité ou la non symétrie de ces matrices. L'amélioration de ces calculs peut se faire en déterminant des formes factorisées des matrices de covariance ou bien en diminuant le nombre de calculs. Le premier point de vue conduit aux algorithmes à factorisations et le deuxième aux algorithmes de filtrage rapide. Dans cette partie, nous allons regarder successivement les principes de base de chacun de ces deux types de méthode en notant qu'elles ont pour objet, essentiellement, l'amélioration de la résolution d'une équation de Riccati discrète.

3.3.1 Algorithmes à factorisations

Parmi cette classe de méthodes, dont le principe de base consiste à décomposer la matrice de covariance de l'erreur d'estimation, on peut distinguer les algorithmes de type RC (racine carrée) et ceux de type UD (basés sur la factorisation UDU^T d'une matrice symétrique [22]). Nous n'aborderons ici que le principe des algorithmes RC ceux de type UD pouvant être traités de manière similaire.

Les matrices de covariance $P_{k/k}$ et $P_{k+1/k}$ sont factorisées sous la forme racine carrée :

$$\begin{aligned} P_{k/k} &= P_{k/k}^{1/2} P_{k/k}^{T/2}, \\ P_{k+1/k} &= P_{k+1/k}^{1/2} P_{k+1/k}^{T/2}, \end{aligned}$$

où $P^{T/2}$ représente $(P^{1/2})^T$ et $P^{1/2}$ une racine carrée de P , dont une méthode de calcul est donnée dans l'annexe C. Le filtre de Kalman sera alors réalisé à partir des évolutions des matrices $P_{k/k}^{1/2}$ et $P_{k+1/k}^{1/2}$. Il s'agit donc de remplacer l'équation de correction de la covariance :

$$\begin{aligned} P_{k/k} &= P_{k/k-1} - P_{k/k-1} C_k^T \Sigma_k^{-1} C_k P_{k/k-1}, \\ \Sigma_k &= R_k + C_k P_{k/k-1} C_k^T, \end{aligned} \tag{41}$$

par un algorithme ne faisant intervenir que $P_{k/k}^{1/2}$ et $P_{k+1/k}^{1/2}$.

Soient les matrices :

$$\begin{aligned} F_k &= P_{k+1/k}^{T/2} C_k^T, \\ \Sigma_k^{1/2} &= [R_k + F_k^T F_k]^{1/2}, \end{aligned} \quad (42)$$

la relation (41) se met sous la forme :

$$P_{k/k}^{1/2} P_{k/k}^{T/2} = P_{k/k-1}^{1/2} [I - F_k \Sigma_k^{-T/2} \Sigma_k^{-1/2} F_k^T] P_{k/k-1}^{T/2}. \quad (43)$$

L'utilisation du lemme de factorisation matricielle (10) permet d'écrire :

$$P_{k/k}^{1/2} = P_{k/k-1}^{1/2} [I - F_k \Sigma_k^{-T/2} \psi_k^{-1} \Sigma_k^{-1/2} F_k^T],$$

où :

$$\psi_k = I + (I - \Sigma_k^{-1/2} F_k^T F_k \Sigma_k^{-T/2})^{1/2}.$$

En tenant compte de la relation (42), il vient :

$$\psi_k = I + (\Sigma_k^{-1/2} R_k \Sigma_k^{-T/2})^{1/2} = \Sigma_k^{-1/2} [\Sigma_k^{1/2} + R_k^{1/2}],$$

ce qui fournit l'étape de correction sous forme RC :

$$P_{k/k}^{1/2} = P_{k/k-1}^{1/2} [I - F_k \Sigma_k^{-T/2} (\Sigma_k^{1/2} + R_k^{1/2})^{-1} F_k^T]. \quad (44)$$

L'étape de prédiction :

$$P_{k/k+1} = A_k P_{k/k} A_k^T + G_k Q_k G_k^T, \quad (45)$$

peut également être mise sous la même forme. En introduisant les matrices suivantes :

$$\begin{aligned} M_k^{1/2} &= A_k P_{k/k}^{1/2}, \\ B_k &= M_k^{-1/2} G_k Q_k^{-1/2}, \end{aligned} \quad (46)$$

l'équation (45) s'écrit sous la forme :

$$P_{k/k+1} = M_k^{1/2} [I + B_k B_k^T] M_k^{T/2}.$$

Par l'application du lemme de factorisation matricielle (annexe C), on obtient directement l'étape de prédiction sous la forme RC :

$$P_{k/k+1}^{1/2} = M_k^{1/2} [I + B_k \psi_k^{-1} B_k^T], \quad (47)$$

où $\psi_k = I + (I + B_k^T B_k)^{1/2}$.

Les algorithmes basés sur les formules (44) et (47) sont respectivement appelés algorithmes d'Andrews et de Dyer-Mc Reynolds et demandent la décomposition de Cholewski (annexe C) des matrices R_k , $R_k + F_k^T F_k$, Q_k , $I + B_k^T B_k$ et l'inversion des matrices triangulaires $\Sigma_k^{1/2}$, $[\Sigma_k^{1/2} + R_k^{1/2}]$ et des matrices $M_k^{1/2}$, et ψ_k .

3.3.2 Algorithmes de filtrage rapide

Les équations de filtrage rapide sont obtenues en exprimant les équations du filtre de Kalman, que nous rappelons ici :

$$\begin{aligned} K_k &= P_{k/k-1} C_k^T \Sigma_k^{-1}, \\ \Sigma_k &= R_k + C_k P_{k/k-1} C_k^T, \\ P_{k+1/k} &= A_k P_{k/k-1} A_k^T + G_k Q_k G_k^T \\ &\quad - A_k P_{k/k-1} C_k^T (R_k + C_k P_{k/k-1} C_k^T)^{-1} C_k P_{k/k-1} A_k^T, \end{aligned}$$

en fonction des incréments de la matrice de covariance :

$$\delta P_k = P_{k+1/k} - P_{k/k-1}.$$

Les équations obtenues, que nous ne démontrerons pas, sont dites équations de Chandrasekhar et peuvent être résumées sous la forme de l'algorithme suivant :

- condition initiale : à partir de la factorisation de δP_0 sous la forme :

$$\delta P_0 = L_0 M_0 L_0^T,$$

où $L_0 \in \mathbb{R}^{q \times \alpha}$, $M_0 \in \mathbb{R}^{\alpha \times \alpha}$, $\alpha = \text{rang } \delta P_0$;

- pour $k \in \mathbb{N}$, la décomposition $\delta P_k = L_k M_k L_k^T$ est mise à jour par les étapes :
- calcul de :

$$\begin{aligned} \Sigma_{k+1} &= \Sigma_k + C_k \delta P_k C_k^T, \\ K_{k+1} &= [K_k \Sigma_k + A_k \delta P_k C_k^T] \Sigma_{k+1}^{-1}; \end{aligned}$$

- calcul de L_{k+1} et M_{k+1} à l'aide de deux types d'expressions équivalentes :

$$\begin{cases} M_{k+1} = M_k - M_k L_k^T C_k^T \Sigma_{k+1}^{-1} C_k L_k M_k, \\ L_{k+1} = [A_k - K_k C_k] L_k, \end{cases} \quad \begin{cases} M_{k+1} = M_k - M_k L_k^T C_k^T \Sigma_k^{-1} C_k L_k M_k, \\ L_{k+1} = [A_k - K_{k+1} C_k] L_k. \end{cases}$$

Cet algorithme permet donc de calculer, par récurrence, l'incrément à apporter à chaque itération d'un filtre de Kalman. Le nombre des équations à résoudre, malgré leur non linéarité, est nettement moins élevé que un algorithme de filtrage usuel.

4 Applications du filtrage

Le filtre de Kalman (continu ou discret) est utilisé fréquemment dans de nombreuses applications pratiques. Nous allons détailler, dans cette partie, quelques principes d'application du filtre de Kalman en commande optimale, identification ou lissage de processus. D'autres exemples d'applications sont proposés dans [8, 14].

4.1 Commande optimale stochastique

L'estimation statistique de l'état d'un système est généralement effectuée dans le but de réaliser une commande par retour d'état. Dans un cadre déterministe, la notion de régulateur-observateur permet de générer une commande à partir de l'état reconstruit [16]. Dans un cadre stochastique, le principe de séparation indique que cette approche est également valide. En effet, le résultat suivant, que nous ne démontrerons pas, montre que la commande optimale d'un système stochastique est obtenue en construisant la commande optimale obtenue sur le système déterministe associé à l'aide de l'état estimé à partir d'un filtre de Kalman [17].

Considérons le système continu stochastique :

$$\begin{aligned}\dot{x}(t) &= A(t)x(t) + B(t)u(t) + G(t)w(t), \\ y(t) &= C(t)x(t) + v(t),\end{aligned}$$

où les notations sont définies en (24). Rappelons que $w(t)$ et $v(t)$ sont des vecteurs aléatoires centrés, de matrices de covariance :

$$\begin{aligned}\mathbb{E}\{w(t)w^T(t')\} &= Q(t)\delta_{t-t'}, \\ \mathbb{E}\{v(t)v^T(t')\} &= R(t)\delta_{t-t'}, \\ \mathbb{E}\{v(t)v^T(t')\} &= 0.\end{aligned}$$

Le problème d'optimisation stochastique consiste à chercher la commande optimale minimisant le critère :

$$J = \mathbb{E} \left\{ x^T(t_f)S_{t_f}x(t_f) + \int_{t_o}^{t_f} (x^T(t)M(t)x(t) + u^T(t)N(t)u(t))dt \right\}.$$

Dans le cas d'un système où l'état est complètement accessible, la commande optimale a la forme usuelle :

$$u^*(t) = -N^{-1}(t)B^T(t)P(t)x(t),$$

où $P(t)$ est solution de l'équation de Riccati :

$$\dot{P}(t) = -A^T(t)P(t) - P(t)A(t) - M(t) + P(t)B(t)N^{-1}(t)B^T(t)P(t), \quad (48)$$

avec $P(t_f) = -S_{t_f}$. Par contre, dans le cas où seule la sortie est accessible (systèmes à état non complètement accessible), la commande optimale s'écrit sous la forme :

$$\hat{u}^*(t) = -N^{-1}(t)B^T(t)P(t)\hat{x}(t),$$

où $P(t)$ est définie par l'équation de Riccati (48) et $\hat{x}(t)$ est l'estimation optimale de $x(t)$ obtenue à l'aide du filtre de Kalman continu :

$$\dot{\hat{x}}(t) = A(t)\hat{x}(t) + B(t)\hat{u}^*(t) + \Sigma(t)C^T R(t)^{-1}[y(t) - C(t)\hat{x}(t)],$$

où $\Sigma(t)$ est solution de l'équation de Riccati :

$$\dot{\Sigma}(t) = A(t)\Sigma(t) + \Sigma(t)A^T(t) + G(t)Q(t)G^T(t) - \Sigma(t)C^T(t)R(t)^{-1}C(t)\Sigma(t),$$

avec $\Sigma(t_0) = E\{(x(t_0) - m_0)(x(t_0) - m_0)^T\}$ et $m_0 = E\{x(t_0)\}$.

Ces différentes relations [4] constituent le principe de séparation : si la commande et l'observateur sont calculés séparément, mais de façon optimale, alors l'ensemble, réuni dans une structure de commande de type régulateur-observateur, sera également optimal. Il faudra cependant faire attention à la robustesse du contrôleur réalisé qui n'est pas garantie [5]. Les relations ont été établies dans un cadre continu mais peuvent bien sûr l'être dans un cadre discret.

4.2 Lissage

4.2.1 Principe du lissage

A partir d'un ensemble de mesures sur $[t_0, t_F]$, il s'agit d'estimer l'état x d'un système à l'instant t , $t_0 < t < t_F$. Le principe du lissage [11] consiste à déterminer deux estimations de la même grandeur : $\hat{x}_A(t)$, à partir des informations passées (mesures réalisées avant t) et $\hat{x}_R(t)$, à partir des informations futures (mesures réalisées après t). Puis on réalise une moyenne $\hat{x}_L(t)$ entre ces deux estimations, sous la forme :

$$\hat{x}_L(t) = P_L(P_A^{-1}\hat{x}_A(t) + P_R^{-1}\hat{x}_R(t)), \quad (49)$$

où $P_L = (P_A^{-1} + P_R^{-1})^{-1}$, P_A est la matrice de covariance d'erreur d'estimation obtenue par le filtrage "Aller" à t , et P_R est celle obtenue par le filtrage "Retour". La formule (49) peut être interprétée immédiatement. En effet, les matrices de covariance d'erreur représentant la précision avec laquelle est connue l'estimation (lorsque $\|P\|$ diminue, la précision augmente), et il est naturel de prendre comme valeur lissée une pondération, à l'aide des matrices de covariance, de chacune des estimations obtenues. Ce qu'il faut noter ici, c'est que cette pondération est optimale. En effet, on cherche l'estimation d'une variable $x(t)$ à partir :

- d'une condition initiale $\hat{x}_A(t)$ de variance P_A ;
- d'une mesure $\hat{x}_R(t)$ de variance P_R .

Comme tout se déroule au même instant, on peut considérer que $x(t)$ est un système formel indépendant du temps dont le fonctionnement est défini par les matrices suivantes :

$$A_k = I, \quad B_k = 0, \quad C_k = I, \quad R_k = P_R, \quad P_{k/k-1} = P_A,$$

et l'application d'un filtre de Kalman sur ce système conduit, suivant les formules (4), (11) et (13) aux expressions :

- covariance de l'erreur d'estimation ($P_{k/k} = P_L$) :

$$P_L^{-1} = P_A^{-1} + P_R^{-1};$$

- gain du filtre optimal :

$$K = P_L P_R^{-1};$$

- estimation optimale :

$$\begin{aligned} \hat{x}_L(t) &= \hat{x}_A(t) + K(\hat{x}_R(t) - \hat{x}_A(t)), \\ &= [I - P_L P_R^{-1}]\hat{x}_A(t) + P_L P_R^{-1}\hat{x}_R(t), \\ &= P_L(P_A^{-1}\hat{x}_A(t) + P_R^{-1}\hat{x}_R(t)), \end{aligned}$$

qui justifient la formule de lissage (49).

4.2.2 Equations du filtre lisseur

Nous établirons, pour être plus brefs, les équations du filtre lisseur dans le cas d'un système continu, la transposition au cas d'un système discret étant immédiate. Considérons le système continu :

$$\begin{aligned} \dot{x}(t) &= Ax(t) + Bu(t) + Gw(t), \\ y(t) &= Cx(t) + v(t), \end{aligned} \quad (50)$$

où les notations sont définies en (24). Nous rappelons simplement ici que le vecteur initial $x(0)$ a pour moyenne *a priori* m_0 et pour covariance centrée *a priori* P_0 . Les bruits $v(t)$ et $w(t)$ sont centrés, non corrélés et de covariances respectives R et Q . Les matrices peuvent dépendre éventuellement du temps mais cette extension n'offre pas de difficulté particulière et l'argument t est ici omis. On suppose donc que l'on connaît $y(t)$ pour $t \in [0, T]$ et on désire construire un filtre lisseur décrivant l'évolution de l'estimation lissée $\hat{x}_L(t)$ et la matrice de covariance d'erreur de lissage $P_L(t)$, pour $t \in [0, T]$.

Filtres “Aller” et “Retour” Le filtre “Aller” est le filtre de Kalman continu associé au système (50). Ce filtre décrit l'évolution de $\hat{x}_A(t)$ et $P_A(t)$ sous la forme des équations :

$$\begin{aligned} \dot{\hat{x}}_A(t) &= A\hat{x}_A(t) + Bu(t) + P_A(t)C^T R^{-1}[y(t) - C\hat{x}_A(t)], \\ \dot{P}_A(t) &= AP_A(t) + P_A(t)A^T + GQG^T - P_A(t)C^T R^{-1}CP_A(t), \end{aligned} \quad (51)$$

à partir des conditions initiales, $\hat{x}_A(0) = m_0$ et $P_A(0) = P_0$.

Pour obtenir le filtre “Retour”, le changement de variable, $t' = T - t$, conduit à l'équation d'évolution rétrograde du système (50) :

$$\begin{aligned} \frac{dx(t')}{dt'} &= -Ax(t') - Bu(t) - Gw(t'), \\ y(t') &= Cx(t') + v(t'). \end{aligned} \quad (52)$$

Comme la covariance des bruits est inchangée par cette transformation, le filtre de Kalman continu sur (52) conduit aux équations d'évolution de $\hat{x}_R(t')$ et $P_R(t')$:

$$\begin{aligned} \frac{d\hat{x}_R(t')}{dt'} &= -A\hat{x}_R(t') - Bu(t') + P_R(t')C^T R^{-1}[y(t') - C\hat{x}_R(t')], \\ \frac{dP_R(t')}{dt'} &= -AP_R(t') - P_R(t')A^T + GQG^T - P_R(t')C^T R^{-1}CP_R(t'). \end{aligned}$$

Le filtre “Retour” est obtenu en revenant à la variable t , ce qui donne :

$$\begin{aligned} \dot{\hat{x}}_R(t) &= A\hat{x}_R(t) + Bu(t) - P_R(t)C^T R^{-1}[y(t) - C\hat{x}_R(t)], \\ \dot{P}_R(t) &= AP_R(t) + P_R(t)A^T - GQG^T + P_R(t)C^T R^{-1}CP_R(t), \end{aligned} \quad (53)$$

qui doivent être intégrées à rebours, à partir des conditions finales $\hat{x}_R(T)$ et $P_R(T)$.

Evolution de la covariance d'erreur de lissage L'utilisation de la relation :

$$(\dot{P}^{-1}) = -P^{-1}\dot{P}P^{-1}, \quad (54)$$

permet d'écrire, à partir de (51) et (53) :

$$\begin{aligned} (\dot{P}_A^{-1}(t)) &= -P_A^{-1}(t)A - A^T P_A^{-1}(t) - P_A^{-1}(t)GQG^T P_A^{-1}(t) + C^T R^{-1}C, \\ (\dot{P}_R^{-1}(t)) &= -P_R^{-1}(t)A - A^T P_R^{-1}(t) + P_R^{-1}(t)GQG^T P_R^{-1}(t) - C^T R^{-1}C. \end{aligned} \quad (55)$$

On obtient donc, d'après (55) :

$$\begin{aligned} (\dot{P}_L^{-1}(t)) &= -P_L^{-1}(t)[A + GQG^T P_A^{-1}(t)] \\ &\quad - [A + GQG^T P_A^{-1}(t)]^T P_L^{-1}(t) + P_L^{-1}(t)GQG^T P_L^{-1}(t). \end{aligned}$$

$P_L(t)$ est donc défini par l'équation différentielle linéaire :

$$\begin{aligned} (\dot{P}_L(t)) &= [A + GQG^T P_A^{-1}(t)]P_L(t) \\ &\quad + P_L(t)[A + GQG^T P_A^{-1}(t)]^T - GQG^T, \end{aligned}$$

avec $P_L(T) = P_A(T)$ comme condition terminale.

Evolution de l'état lissé A partir de la relation (49), on obtient :

$$\begin{aligned} \dot{\hat{x}}_L(t) &= \dot{P}_L(t)[P_A^{-1}(t)\hat{x}_A(t) + P_R^{-1}(t)\hat{x}_R(t)] \\ &\quad + P_L(t)[(P_A^{-1}(t)\dot{\hat{x}}_A(t) + (P_R^{-1}(t)\dot{\hat{x}}_R(t) + P_A^{-1}(t)\dot{\hat{x}}_A(t) + P_R^{-1}(t)\dot{\hat{x}}_R(t)]. \end{aligned}$$

En utilisant les relations précédentes, cela s'écrit, après quelques simplifications, sous la forme :

$$\dot{\hat{x}}_L(t) = A\hat{x}_L(t) + Bu(t) + GQG^T P_A^{-1}(t)[\hat{x}_L(t) - \hat{x}_A(t)], \quad (56)$$

avec $x_L(T) = x_A(T)$.

Ainsi le lissage, dans sa version continue, peut être obtenu par :

- un filtrage "Aller" à partir de la condition initiale m_0, P_0 . On en déduit alors $\hat{x}_A(t), P_A(t)$ pour tout t dans $[0, T]$, en particulier les valeurs finales $\hat{x}_A(T)$ et $P_A(T)$;
- un lissage "Retour" à partir de ces valeurs finales, sur tout l'intervalle $[0, T]$ par intégration à rebours du système (56).

La version discrète, basée sur des équations de récurrence, conduit à une formulation analogue.

4.3 Identification

Le filtrage permettant d'estimer des variables dynamiques à partir d'un ensemble de mesures, il est naturel qu'un des principaux domaines d'application du filtre de Kalman soit aussi

l'estimation de paramètres du système. Ainsi le filtre de Kalman peut être utilisé en identification ou en commande adaptative. En tant que méthode d'identification nous verrons que l'utilisation du filtre de Kalman peut se faire directement à partir d'une régression linéaire ou bien sous la forme filtre en utilisant la notion de processus d'innovation. Dans ce paragraphe nous ne ferons qu'ébaucher ce type d'applications en renvoyant à [18, 23] pour des plus amples détails.

4.3.1 Estimation de paramètres

De nombreuses méthodes d'identification conduisent à exprimer la relation entre les paramètres inconnus du système et les mesures, sous la forme d'une régression linéaire :

$$y_k = C_k \Theta_k + l_k, \quad (57)$$

où Θ_k est le vecteur des paramètres à estimer, y_k et C_k sont un vecteur et une matrice issus des mesures et l_k est un bruit (ou erreur) d'équation.

Si l'on suppose que la séquence $\{l_k\}$ est la réalisation d'une variable aléatoire centrée dont on connaît la variance Σ , on peut, en considérant les paramètres à déterminer comme constants, résoudre ce problème de façon récurrente par un filtre de Kalman. En effet, les paramètres constants sont décrits par l'équation d'état :

$$\Theta_{k+1} = \Theta_k, \quad (58)$$

ce qui, associé à l'équation de mesure (57) conduit au filtre estimateur optimal des paramètres :

$$\begin{aligned} \hat{\Theta}_{k+1} &= \hat{\Theta}_k + K_k (y_k - C_k \hat{\Theta}_k), \\ K_k &= P_k C_k^T (C_k P_k C_k^T + \Sigma)^{-1}, \\ P_{k+1} &= P_k - K_k C_k P_k. \end{aligned}$$

dont les expressions sont à rapprocher des formules de moindres carrés récurrents.

Cette méthode peut également être employée (contrairement aux moindres carrés récursifs) dans des cas plus généraux comme, par exemple, les systèmes non stationnaires où les paramètres dépendent du temps. L'équation (58) est alors remplacée par l'équation d'évolution :

$$\Theta_{k+1} = \Theta_k + w_k,$$

où w_k est un bruit blanc.

Dans le cas où lon ne peut pas se ramener à la forme (57) (par exemple dans le cas modèle à régression non linéaire ou lorsque l'on procède à l'estimation simultanée de paramètres et de l'état d'un système), on utilise le filtre de Kalman étendu qui est une extension aux systèmes non-linéaires du filtre de Kalman. Le principe du filtre étendu consiste à appliquer un filtrage linéaire sur le système non-linéaire, linéarisé autour d'un point de fonctionnement (avec toutes les réserves sur les limites de validité que cela comporte).

Considérons le système défini par l'équation d'état :

$$\begin{aligned} x_{k+1} &= A(\Theta)x_k + B(\Theta)u_k + Gw_k, \\ y_k &= C(\Theta)x_k + v_k, \end{aligned} \quad (59)$$

où A, B, C sont des matrices dépendant de paramètres regroupés dans le vecteur Θ , et l'on veut, dans un même filtre, estimer simultanément l'état x_k et Θ . On construit le vecteur d'état augmenté :

$$X_k = \begin{bmatrix} x_k \\ \Theta_k \end{bmatrix},$$

où Θ_k est défini par l'équation (58). Il est évident que le système décrivant X_k est non-linéaire. Après linéarisation, un filtre de Kalman peut permettre d'estimer correctement toutes les variables.

4.3.2 Forme filtre

L'innovation est la différence, non prévisible, entre la mesure réalisée à un instant donné et la prédiction optimale de cette mesure, compte tenu des informations passées. Elle a pour expression, suivant les formes obtenues dans le filtre de Kalman :

– pour le cas continu :

$$\tilde{y}(t) = y(t) - C\hat{x}(t),$$

– pour le cas discret :

$$\tilde{y}_k = y_k - C\hat{x}_{k/k-1}, \quad (60)$$

Dans le cas discret stationnaire, l'évolution de la valeur filtrée est donnée par :

$$\hat{x}_{k+1/k} = A\hat{x}_{k/k-1} + Bu_k + K\tilde{y}_k, \quad (61)$$

où K est la valeur limite du gain optimal.

Si le filtre est optimal, \tilde{y}_k possède la propriété d'être une séquence blanche, en effet, $\tilde{x}_{k/k-1}$ et w_k sont orthogonaux à toutes les mesures de y_0 à y_{k-1} . Suivant les équations (60) et (61), que l'on peut mettre sous la forme (en posant $\hat{x}_{k+1/k} = X_{k+1}$) :

$$\begin{aligned} X_{k+1} &= AX_k + Bu_k + K\tilde{y}_k, \\ y_k &= CX_k + \tilde{y}_k, \end{aligned} \quad (62)$$

on montre que la sortie du système initial, sur lequel intervenait un bruit d'entrée et un bruit de mesure, peut être considérée comme la sortie d'un système d'état X_k perturbé par une seule séquence blanche de dimension inférieure aux séquences aléatoires $\{w_k^T, v_k^T\}$ initiales. On a donc intérêt à chercher directement une modélisation d'un système sous la forme filtre (62) ou processus d'innovation. L'autre intérêt de considérer cette forme est de pouvoir déterminer directement K , gain limite optimal du filtre de Kalman du système.

Si l'on prend la transformée en z du processus d'innovation (62), il vient :

$$X(z) = (zI - A)^{-1}[BU(z) + K\tilde{Y}(z)],$$

soit :

$$Y(z) = [C(zI - A)^{-1}B]U(z) + [I + C(zI - A)^{-1}K]\tilde{Y}(z).$$

On reconnaît alors un modèle ARMAX de la forme :

$$\mathcal{A}(z)Y(z) = \mathcal{B}(z)U(z) + \mathcal{C}(z)\tilde{Y}(z), \quad (63)$$

dont on peut déterminer les paramètres par des méthodes usuelles d'identification [18].

5 Annexe A : Estimation optimale

Nous décrivons, dans cette partie, quelques principes de base de la théorie de l'estimation d'une variable aléatoire par la mesure ou l'observation d'une autre variable aléatoire. De façon générale, les variables considérées sont vectorielles et nous distinguerons la variable aléatoire (en lettre capitale) de sa réalisation (en lettre minuscule).

D'autre part pour une variable aléatoire X , nous noterons par $f_X(x)$ sa densité de probabilité, et pour tout couple de variables aléatoires (X, Y) , $f_{X/Y}(x, y)$ et $f_{X,Y}(x, y)$ désigneront respectivement les densités de probabilité conditionnelles et conjointes de X et Y . Ces densités sont reliées par le théorème de Bayes :

$$f_{X,Y}(x, y) = f_{X/Y}(x, y)f_Y(y) = f_{Y/X}(x, y)f_X(x).$$

L'estimation optimale consiste à réaliser une estimation ponctuelle par l'optimisation d'un critère. Nous allons en décrire ici essentiellement deux : l'estimation au maximum de vraisemblance et l'estimation au sens des moindres carrés, cette dernière conduisant à l'estimation linéaire sous la forme du filtre de Kalman dont une démonstration est proposée à l'annexe B.

5.1 Estimateur du maximum de vraisemblance

Pour une variable aléatoire X dont la densité de probabilité, $f_X(x)$, est unimodale, la valeur la plus probable (c'est-à-dire vraisemblable), x_p , est celle correspondant au maximum de cette densité, soit :

$$x_{MV} = \arg_x \left\{ \frac{d f_X(x)}{d x} = 0 \right\}.$$

Lorsque l'on dispose de données mesurées qui dépendent d'une variable θ que l'on cherche à estimer :

$$z = \phi(\theta),$$

l'utilisation du principe précédent indique que la valeur la plus probable de θ , soit $\hat{\theta}_{MV}$, est celle qui possède le plus de probabilité d'avoir produit le z que l'on observe. Ainsi, à partir de la densité de probabilité conditionnelle $f_{Z/\Theta}(z, \theta)$, appelée fonction de vraisemblance lorsqu'elle est considérée comme une fonction de θ , on a :

$$\hat{\theta}_{MV} = \arg_{\theta} \max_{\theta} f_{Z/\Theta}(z, \theta).$$

Quand la fonction de vraisemblance est différentiable et que le maximum est à l'intérieur du domaine admissible, l'expression précédente peut être réécrite sous la forme :

$$\hat{\theta}_{MV} = \arg_{\theta} \left\{ \frac{\partial f_{Z/\Theta}(z, \theta)}{\partial \theta} = 0 \right\}.$$

Comme la fonction logarithme est strictement croissante et qu'une densité de probabilité est strictement positive, il est équivalent de maximiser $f_{Z/\Theta}(z, \theta)$ ou $\log f_{Z/\Theta}(z, \theta)$, appelée fonction de log-vraisemblance. On obtient dans ce cas :

$$\hat{\theta}_{MV} = \arg_{\theta} \left\{ \frac{\partial \log f_{Z/\Theta}(z, \theta)}{\partial \theta} = 0 \right\}. \quad (64)$$

L'utilité de considérer la fonction de log-vraisemblance réside dans le fait que de nombreuses variables aléatoires, notamment des variables gaussiennes, ont des densités de probabilité, donc des fonctions de vraisemblance potentielles, qui mettent en œuvre des exponentielles.

Remarquons enfin qu'à l'aide du théorème de Bayes on peut écrire :

$$f_{Z/\Theta}(z, \theta) = \frac{f_{Z,\Theta}(z, \theta)}{f_{\Theta}(\theta)},$$

où $f_{Z,\Theta}(z, \theta)$ est la densité de probabilité conjointe des variables aléatoires Z et Θ et $f_{\Theta}(\theta)$ est la densité de probabilité marginale de Θ . On obtient ainsi :

$$\frac{\partial \log f_{Z/\Theta}(z, \theta)}{\partial \theta} = \frac{1}{f_{Z,\Theta}(z, \theta)} \frac{\partial f_{Z,\Theta}(z, \theta)}{\partial \theta} - \frac{1}{f_{\Theta}(\theta)} \frac{\partial f_{\Theta}(\theta)}{\partial \theta},$$

ce qui permet d'écrire que, lorsque l'écriture (64) est valide, $\hat{\theta}_{MV}$ est solution de l'équation :

$$\frac{\partial f_{Z,\Theta}(z, \theta)}{\partial \theta} = f_{Z/\Theta}(z, \theta) \frac{\partial f_{\Theta}(\theta)}{\partial \theta}.$$

5.2 Estimation au sens des moindres carrés

Soient, X et Y des vecteurs aléatoires liés. On peut s'attendre à ce que le fait d'observer une valeur y pour Y puisse apporter une information sur la valeur correspondante (mais non mesurée) x de X . De façon plus précise, en sachant que Y prend la valeur y , l'estimation au sens des moindres carrés \hat{x} est celle qui minimisera, parmi tous les vecteurs z de même dimension, l'espérance conditionnelle :

$$\mathbb{E}\{\|X - z\|^2 / Y = y\} = \mathbb{E}\{[X - z]^T [X - z] / Y = y\}. \quad (65)$$

Cette estimation \hat{x} , qui minimise une mesure d'incertitude, est également appelée estimation au sens du minimum de variance. L'utilisation de la propriété de linéarité de $\mathbb{E}\{\cdot\}$ conduit à écrire le critère à minimiser sous la forme :

$$\begin{aligned} \mathbb{E}\{\|X - z\|^2 / Y = y\} &= \mathbb{E}\{X^T X - 2z^T X + z^T z / Y = y\}, \\ &= \mathbb{E}\{X^T X / Y = y\} - 2z^T \mathbb{E}\{X / Y = y\} + z^T z, \\ &= \mathbb{E}\{X^T X / Y = y\} - \|\mathbb{E}\{X / Y = y\}\|^2 + \|z - \mathbb{E}\{X / Y = y\}\|^2. \end{aligned}$$

Comme seul le dernier terme dépend de z , le minimum de cette expression est obtenu pour :

$$z = \hat{x} = \mathbb{E}\{X / Y = y\}. \quad (66)$$

Ainsi l'estimation au sens des moindres carrés est l'espérance conditionnelle de X sachant que Y prend la valeur y , et elle est donnée par :

$$\hat{x} = \int_{-\infty}^{+\infty} x f_{X/Y}(x, y) dx = \int_{-\infty}^{+\infty} x \frac{f_{X,Y}(x, y)}{f_Y(y)} dx.$$

Il est nécessaire de remarquer que l'estimation optimale \hat{x} (66) dépend de la valeur de y , ainsi \hat{x} peut être considérée comme la réalisation d'une variable aléatoire pour une valeur y donnée, sous la forme $\hat{x} = \hat{X}(y)$. L'estimateur au sens des moindres carrés est donc le vecteur aléatoire $\hat{X} = E\{X/Y\}$, qui réalise cette opération. Par construction, pour toute fonction g du vecteur aléatoire Y , on a :

$$E\{\|X - \hat{X}(Y)\|^2/Y = y\} \leq E\{\|X - g(Y)\|^2/Y = y\}. \quad (67)$$

Si l'on tient compte de l'égalité suivante :

$$\int_{-\infty}^{+\infty} E\{./Y\} f_Y(y) dy = E\{.\}, \quad (68)$$

la relation (67) s'écrit sous la forme :

$$E\{\|X - \hat{X}(Y)\|^2\} \leq E\{\|X - g(Y)\|^2\}. \quad (69)$$

Ainsi, \hat{X} possède la propriété de minimiser $E\{\|X - g(Y)\|^2\}$ pour toute fonction g de Y . Dans le cas sans contrainte on peut construire \hat{X} en déterminant \hat{x} pour chaque y . Lorsque l'on impose des contraintes sur la structure de \hat{X} , par exemple être linéaire, cette démarche n'est plus valable et il peut exister des cas où l'inégalité (69) n'est plus vérifiée. Nous verrons dans la suite la construction d'un estimateur optimal linéaire et que dans le cas de vecteurs aléatoires gaussiens l'estimateur optimal est linéaire.

5.3 Estimation linéaire

Supposons connues les propriétés statistiques des vecteurs aléatoires X et Y :

– moyennes :

$$E\{X\} = m_X, \quad E\{Y\} = m_Y; \quad (70)$$

– matrices de covariance :

$$\begin{aligned} E\{(X - m_X)(X - m_X)^T\} &= P_{XX}, \\ E\{(X - m_X)(Y - m_Y)^T\} &= P_{XY}, \\ E\{(Y - m_Y)(Y - m_Y)^T\} &= P_{YY}. \end{aligned} \quad (71)$$

Dans ces relations, $X - m_X = X_C$ et $Y - m_Y = Y_C$ représentent les vecteurs aléatoires centrés. On cherche ici à construire l'estimateur linéaire optimal :

$$\hat{X} = A_o Y + B_o, \quad (72)$$

minimisant la variance de l'erreur d'estimation :

$$E\{\|X - AY - B\|^2\}.$$

En développant le deuxième terme de l'identité :

$$E\{\|X - AY - B\|^2\} = \text{trace } E\{[X_C - AY_C - B_C] [X_C - AY_C - B_C]^T\},$$

où $B_C = B - m_x + Am_Y$, il vient :

$$\mathbb{E}\{\|X - AY - B\|^2\} = \text{trace}[P_{XX} - P_{XY}A^T - AP_{XY}^T + AP_{YY}A^T] + \|B_C\|^2. \quad (73)$$

Lorsque P_{YY} est régulière, l'utilisation de l'égalité suivante :

$$\begin{aligned} P_{XX} - P_{XY}A^T - AP_{XY}^T + AP_{YY}A^T &= \\ P_{XX} - P_{XY}P_{YY}^{-1}P_{XY}^T &+ [A - P_{XY}P_{YY}^{-1}]P_{YY}[A - P_{XY}P_{YY}^{-1}]^T, \end{aligned}$$

conduit, à écrire (73) sous la forme :

$$\begin{aligned} \mathbb{E}\{\|X - AY - B\|^2\} &= \text{trace}[P_{XX} - P_{XY}P_{YY}^{-1}P_{XY}^T] + \|B_C\|^2 + \\ &\text{trace}([A - P_{XY}P_{YY}^{-1}]P_{YY}[A - P_{XY}P_{YY}^{-1}]^T). \end{aligned} \quad (74)$$

Cette expression est minimale, si l'on a :

$$\begin{aligned} A &= A_o = P_{XY}P_{YY}^{-1}, \\ B &= B_o = m_X - P_{XY}P_{YY}^{-1}m_Y. \end{aligned}$$

Dans ce cas, l'estimateur linéaire optimal se met sous la forme :

$$\hat{X} = m_X + P_{XY}P_{YY}^{-1}(Y - m_Y), \quad (75)$$

où P_{XY} et P_{YY} sont les matrices de covariance des variables centrées définies en (71). La matrice de covariance de l'erreur de d'estimation est alors, d'après (74) :

$$\mathbb{E}\{[X - \hat{X}][X - \hat{X}]^T\} = P_{XX} - P_{XY}P_{YY}^{-1}P_{XY}^T. \quad (76)$$

L'estimateur linéaire (75), qui minimise *a priori* la variance d'erreur, possède les propriétés suivantes :

- il est sans biais. (75) conduit directement à $\mathbb{E}\{X - \hat{X}\} = 0$. Il est à noter que ceci reste vrai pour tout estimateur de la forme $m_X + A(Y - m_Y)$ avec A quelconque ;
- l'erreur d'estimation et la mesure sont non corrélées (cela constitue le principe d'orthogonalité) :

$$\begin{aligned} \mathbb{E}\{(X - \hat{X})Y^T\} &= \mathbb{E}\{(X - \hat{X})(Y - m_Y)^T\}, \\ &= \mathbb{E}\{[X_C - P_{XY}P_{YY}^{-1}Y_C]Y_C^T\}, \\ &= P_{XY} - P_{XY}P_{YY}^{-1}P_{YY} = 0. \end{aligned} \quad (77)$$

De façon plus générale, l'erreur d'estimation est non corrélée avec toute fonction linéaire de Y , donc en particulier \hat{X} .

Il est possible de montrer également que si un estimateur linéaire vérifie ces deux propriétés alors il minimise la variance d'erreur et on retrouve (75). Les relations que l'on vient d'obtenir seront utilisées pour établir les expressions du filtrage linéaire.

5.4 Estimateur optimal de variables gaussiennes

Nous allons montrer que, dans le cas particulier important de vecteurs, X et Y , aléatoires gaussiens, l'estimateur optimal est linéaire. Il sera donc de la forme (75).

Un vecteur aléatoire Z de dimension n est gaussien (ou normal) si sa densité de probabilité est de la forme :

$$f_Z(z) = \frac{1}{(2\pi)^{n/2}|P|^{1/2}} \exp\left[-\frac{1}{2}(z-m)^T P^{-1}(z-m)\right], \quad (78)$$

où m et P sont des vecteur et matrice de dimensions convenables et $|P|$ désigné le déterminant de P . On peut montrer que m représente la moyenne de Z , $E\{Z\} = m$, et P la matrice de covariance de la variable aléatoire centrée, $Z-m$, $E\{(Z-m)(Z-m)^T\} = P$. Dans ce qui suit, la notation $\mathcal{N}(m, P)$ indiquera une variable aléatoire gaussienne qui suit une loi normale dont la densité de probabilité est de la forme (78). L'importance des variables gaussiennes réside dans le théorème central limite qui montre qu'elles permettent de modéliser de nombreux phénomènes aléatoires naturels.

Dans le cas où X et Y sont gaussiens, nous allons déterminer la forme de l'estimateur $\hat{X} = E\{X/Y\}$ minimisant la variance d'erreur *a posteriori* (65).

Posons les notations suivantes :

$$Z = \begin{bmatrix} X \\ Y \end{bmatrix}, \quad M = \begin{bmatrix} m_X \\ m_Y \end{bmatrix}, \quad P = \begin{bmatrix} P_{XX} & P_{XY} \\ P_{XY}^T & P_{YY} \end{bmatrix},$$

où m_X , m_Y , P_{XX} , P_{XY} , et P_{YY} sont définies en (70) et (71), et calculons la densité de probabilité $f_{X/Y}(x, y)$. Suivant la loi de Bayes étendue aux vecteurs aléatoires, on a :

$$f_{X/Y}(x, y) = f_Z(z)[f_Y(y)]^{-1}.$$

Or Z et Y sont de la forme $\mathcal{N}(m, P)$ et $\mathcal{N}(m_Y, P_{YY})$, ce qui donne, avec $p = \dim X$ et P^{-1} décomposée sous la forme :

$$P^{-1} = \begin{bmatrix} S_{XX} & S_{XY} \\ S_{XY}^T & S_{YY} \end{bmatrix},$$

la densité de probabilité :

$$f_{X/Y}(x, y) = \frac{|P_{YY}|^{1/2}}{(2\pi)^{p/2}|P|^{1/2}} \exp\left\{-\frac{1}{2}(z-m)^T \begin{bmatrix} S_{XX} & S_{XY} \\ S_{XY}^T & S_{YY} - P_{YY}^{-1} \end{bmatrix} (z-m)\right\}. \quad (79)$$

L'utilisation de la formule d'inversion d'une matrice par blocs, conduit à :

$$\begin{aligned} S_{XX} &= \Delta^{-1}, \\ S_{XY} &= -\Delta^{-1}P_{XY}P_{YY}^{-1}, \\ S_{YY} &= P_{YY}^{-1} + P_{YY}^{-1}P_{XY}^T\Delta^{-1}P_{XY}P_{YY}^{-1}, \\ \Delta &= P_{XX} - P_{XY}P_{YY}^{-1}P_{XY}^T. \end{aligned}$$

Ainsi, l'argument de l'exponentielle dans (79) se met sous la forme :

$$-\frac{1}{2}[x - m_X - P_{XY}P_{YY}^{-1}(y - m_Y)]^T \Delta^{-1}[x - m_X - P_{XY}P_{YY}^{-1}(y - m_Y)], \quad (80)$$

et, comme de plus $|P| = |P_{YY}||\Delta|$, on tire, de (79) et (80), la conclusion que $E(X/Y)$ est un vecteur aléatoire gaussien $\mathcal{N}(m_x + P_{XY}P_{YY}^{-1}(y - m_Y), \Delta)$. L'estimateur optimal $\hat{X} = E(X/Y)$ est donc linéaire et l'on retrouve (75) sous la forme :

$$\hat{X} = P_{XY}P_{YY}^{-1}Y + [m_X - P_{XY}P_{YY}^{-1}m_Y],$$

associé à la variance d'erreur minimale *a posteriori* (76).

5.5 Estimation réursive

Soient Y et Z , deux vecteurs aléatoires. L'estimateur linéaire optimal d'un vecteur aléatoire X , à partir de Y et Z , s'écrit d'après (75) :

$$E\{X/W\} = m_X + P_{XW}P_{WW}^{-1}(W - m_W), \quad (81)$$

où $W^T = [Y^T, Z^T]$ et $m_W^T = [m_Y^T, m_Z^T]$. Dans le cas où Y et Z sont non corrélés, on a :

$$P_{XW} = [P_{XY}, P_{XZ}],$$

$$P_{WW} = \begin{bmatrix} P_{YY} & 0 \\ 0 & P_{ZZ} \end{bmatrix},$$

et il vient :

$$E\{X/W\} = m_X + P_{XY}P_{YY}^{-1}(Y - m_Y) + P_{XZ}P_{ZZ}^{-1}(Z - m_Z).$$

Soient :

$$\hat{X}_{/Y} = E\{X/Y\} = m_X + P_{XY}P_{YY}^{-1}(Y - m_Y),$$

$$\tilde{X}_{/Y} = X - \hat{X}_{/Y},$$

alors :

$$E\{\tilde{X}_{/Y}/Z\} = E\{X/Z\} - m_X - P_{XY}P_{YY}^{-1}E\{(Y - m_Y)/Z\}.$$

Or, Y et Z sont supposées non corrélés, donc l'estimateur (81) devient :

$$\hat{X}_{/Y,Z} = \hat{X}_{/Y} + E\{\tilde{X}_{/Y}/z\}.$$

La matrice de covariance d'erreur est donnée, d'après (76), par :

$$P_{XX} - P_{XY}P_{YY}^{-1}P_{XY}^T - P_{XZ}P_{ZZ}^{-1}P_{XZ}^T = P_{\tilde{X}_{/Y}, \tilde{X}_{/Y}} - P_{\tilde{X}_{/Y}, Z}P_{ZZ}^{-1}P_{\tilde{X}_{/Y}, Z}^T,$$

où $P_{\tilde{X}_{/Y}, \tilde{X}_{/Y}} = E\{\tilde{X}_{/Y}\tilde{X}_{/Y}^T\}$, et $P_{\tilde{X}_{/Y}, Z} = E\{\tilde{X}_{/Y}Z^T\}$.

Dans le cas où Y et Z sont corrélées on se ramène au cas précédent par l'intermédiaire de la variable :

$$\tilde{Z}_{/Y} = Z - E\{Z/Y\},$$

qui est non corrélée avec Y (cf. (77)), et l'on a :

$$E\{X/W\} = E\{X/Y, \tilde{Z}/Y\}.$$

Ainsi l'estimateur linéaire optimal prend la forme :

$$E\{X/W\} = \hat{X}/Y + E\{\tilde{X}/Y/\tilde{Z}/Y\},$$

où $\hat{X}/Y = E\{X/Y\}$, et $\tilde{X}/Y = X - \hat{X}/Y$.

La matrice de covariance de l'erreur de l'estimation est alors, dans ce cas, égale à :

$$P_{\tilde{X}/Y, \tilde{X}/Y} - P_{\tilde{X}/Y, \tilde{Z}/Y} P_{\tilde{Z}/Y, \tilde{Z}/Y}^{-1} P_{\tilde{Z}/Y, \tilde{X}/Y}^T. \quad (82)$$

De ces considérations on peut tirer la détermination de l'estimateur linéaire optimal récursif d'un vecteur aléatoire X à partir d'une suite de vecteurs aléatoires Y_1, Y_2, \dots, Y_{k+1} . Cet estimateur noté $\hat{X}/_{k+1}$ est donné d'après ce qui précède par :

$$\hat{X}/_{k+1} = \hat{X}/_k + E\{\tilde{X}/_k/\tilde{Y}_{k+1/k}\}, \quad (83)$$

où :

$$\begin{aligned} \tilde{X}/_k &= X - \hat{X}/_k = X - E\{X/Y_1, \dots, Y_k\}, \\ \tilde{Y}_{k+1/k} &= Y_{k+1} - E\{Y_{k+1}/Y_1, \dots, Y_k\}. \end{aligned}$$

D'après (82), la matrice de covariance de l'erreur d'estimation $\tilde{X}/_{k+1}$ est donnée par :

$$P_{\tilde{X}/_{k+1}, \tilde{X}/_{k+1}} = P_{\tilde{X}/_k, \tilde{X}/_k} - P_{\tilde{X}/_k, \tilde{Y}_{k+1/k}} P_{\tilde{Y}_{k+1/k}, \tilde{Y}_{k+1/k}}^{-1} P_{\tilde{Y}_{k+1/k}, \tilde{X}/_k}^T. \quad (84)$$

6 Annexe B : Démonstration des équations du filtre de Kalman

Dans toute la suite, pour ne pas allourdir les notations, nous confondrons les variables aléatoires et leurs réalisations.

Pour le système stochastique (dont nous rappelons les équations pour plus de clarté) :

$$\begin{aligned} x_{k+1} &= A_k x_k + B_k u_k + G_k w_k, \\ y_k &= C_k x_k + v_k, \end{aligned}$$

l'entrée u_k , ainsi que les matrices A_k, B_k, G_k, C_k sont des grandeurs certaines. L'état initial x_0 est non corrélé avec les bruits de sortie (v_k) et de dynamique (w_k) qui sont connus par :

$$\begin{aligned} E\{x_0\} &= m_0, \quad E\{(x_0 - m_0)(x_0 - m_0)^T\} = P_0, \\ E\{w_k\} &= 0, \quad E\{v_k\} = 0, \quad E\{w_k v_j^T\} = 0, \\ E\{w_k w_j^T\} &= Q_k \delta_{kj}, \quad E\{v_k v_j^T\} = R_k \delta_{kj}. \end{aligned}$$

Le problème du filtre de Kalman est de déterminer l'équation récurrente de l'estimateur optimal $\hat{x}_{k+1/k}$ de x_{k+1} à partir de la séquence de sortie $\mathcal{Y}_k = \{y_0, y_1, \dots, y_k\}$. En utilisant les principes d'estimation récursive d'une variable aléatoire par un ensemble de variables aléatoires nous allons établir simplement les relations d'un filtre prédicteur-à-un-pas.

6.1 Forme prédicteur-à-un-pas

Supposons connus $\hat{x}_{k/k-1} = E\{x_k/\mathcal{Y}_{k-1}\}$ et la matrice de covariance $P_{k/k-1}$ de l'erreur d'estimation $\tilde{x}_{k/k-1} = x_k - \hat{x}_{k/k-1}$. L'utilisation de la formule d'estimation récursive (83) conduit à écrire :

$$\hat{x}_{k+1/k} = \hat{x}_{k+1/k-1} + E\{\tilde{x}_{k+1/k-1}/\tilde{y}_{k/k-1}\},$$

où :

$$\begin{aligned} \tilde{x}_{k+1/k-1} &= x_{k+1} - \hat{x}_{k+1/k-1}, \\ \tilde{y}_{k/k-1} &= y_k - \hat{y}_{k/k-1}, \\ &= y_k - C_k \hat{x}_{k/k-1}, \\ &= C_k \tilde{x}_{k/k-1} + v_k. \end{aligned} \tag{85}$$

D'après les principes d'estimation linéaire optimale, on a :

$$E\{\tilde{x}_{k+1/k-1}/\tilde{y}_{k/k-1}\} = E\{\tilde{x}_{k+1/k-1}\tilde{y}_{k/k-1}^T\}[E\{\tilde{y}_{k/k-1}\tilde{y}_{k/k-1}^T\}]^{-1}(\tilde{y}_{k/k-1}),$$

car $\tilde{x}_{k+1/k-1}$ et $\tilde{y}_{k/k-1}$ sont des variables centrées. D'autre part, la matrice $P_{k+1/k}$ de covariance de l'erreur d'estimation $\tilde{x}_{k+1/k} = x_{k+1} - \hat{x}_{k+1/k}$ est donnée, suivant (84), par :

$$\begin{aligned} P_{k+1/k} &= E\{\tilde{x}_{k+1/k-1}\tilde{x}_{k+1/k-1}^T\} - E\{\tilde{x}_{k+1/k-1}\tilde{y}_{k/k-1}^T\} \\ &\quad \times [E\{\tilde{y}_{k/k-1}\tilde{y}_{k/k-1}^T\}]^{-1}[E\{\tilde{x}_{k+1/k-1}\tilde{y}_{k/k-1}^T\}]^T. \end{aligned}$$

Or, par linéarité de l'estimateur linéaire optimal, il vient :

$$\hat{x}_{k+1/k-1} = A_k \hat{x}_{k/k-1} + B_k u_k,$$

car y_{k-1} ne dépend (linéairement) que de x_0, w_0, \dots, w_{k-2} et v_{k-1} , il est donc non corrélé à w_k . Ainsi (71) et (72) conduisent à :

$$\tilde{x}_{k+1/k-1} = A_k \tilde{x}_{k/k-1} + G_k w_k. \tag{86}$$

De (85) et (86), on tire les matrices de covariance :

$$\begin{aligned} E\{\tilde{x}_{k+1/k-1}\tilde{x}_{k+1/k-1}^T\} &= A_k P_{k/k-1} A_k^T + G_k Q_k G_k^T, \\ E\{\tilde{x}_{k+1/k-1}\tilde{y}_{k/k-1}^T\} &= A_k P_{k/k-1} C_k^T, \\ E\{\tilde{y}_{k/k-1}\tilde{y}_{k/k-1}^T\} &= C_k P_{k/k-1} C_k^T + R_k. \end{aligned} \tag{87}$$

La structure récursive du filtre de Kalman prédicteur-à-un pas est donc, en combinant les relations précédentes :

$$\hat{x}_{k+1/k} = A_k \hat{x}_{k/k-1} + B_k u_k + K_k (y_k - C_k \hat{x}_{k/k-1}), \tag{88}$$

où :

$$\begin{aligned} K_k &= A_k P_{k/k-1} C_k^T (C_k P_{k/k-1} C_k^T + R_k)^{-1}, \\ P_{k+1/k} &= A_k P_{k/k-1} A_k^T + G_k Q_k G_k^T - K_k C_k P_{k/k-1} A_k^T. \end{aligned} \tag{89}$$

L'initialisation de ces récurrences se fait, pour $k = 0$, par $\hat{x}_{0/-1} = m_0$ et $P_{0/-1} = P_0$. En effet, en l'absence de toute information, il est naturel de prendre comme meilleures estimations, les données statistiques que l'on possède sur l'état initial.

6.2 Décomposition du filtre

Si l'on pose :

$$\begin{aligned}\Sigma_k &= C_k P_{k/k-1} C_k^T + R_k, \\ K_k &= P_{k/k-1} C_k^T \Sigma_k^{-1}, \\ \hat{x}_{k/k} &= \hat{x}_{k/k-1} + K_k (y_k - C_k \hat{x}_{k/k-1}).\end{aligned}\tag{90}$$

la dernière étape apparaît bien comme une étape de correction compte-tenu des informations reçues à l'instant k . Si l'on note $P_{k/k}$ la matrice de covariance de l'erreur d'estimation après correction, on obtient :

$$\begin{aligned}P_{k/k} &= P_{k/k-1} - \mathbb{E}\{\tilde{x}_{k/k-1} \tilde{y}_{k/k-1}^T\} K_k^T \\ &\quad - K_k [\mathbb{E}\{\tilde{x}_{k/k-1} \tilde{y}_{k/k-1}^T\}]^T + K_k \Sigma_k K_k^T.\end{aligned}$$

Or l'utilisation de (86), (87) et la non corrélation des bruits de sortie et de dynamique permet de montrer que :

$$\mathbb{E}\{\tilde{x}_{k/k-1} \tilde{y}_{k/k-1}^T\} = P_{k/k-1} C_k^T,$$

ce qui conduit à :

$$\begin{aligned}P_{k/k} &= P_{k/k-1} - K_k \Sigma_k K_k^T, \\ &= P_{k/k-1} - P_{k/k-1} C_k^T \Sigma_k^{-1} C_k P_{k/k-1}, \\ &= [I - K_k C_k] P_{k/k-1},\end{aligned}\tag{91}$$

qui termine l'étape de correction.

L'utilisation de (90) et (91) permet de mettre (89) sous la forme d'une étape de prédiction :

$$\begin{aligned}\hat{x}_{k+1/k} &= A_k \hat{x}_{k/k} + B_k u_k, \\ P_{k+1/k} &= A_k P_{k/k} A_k^T + G_k Q_k G_k^T.\end{aligned}$$

L'application de ce filtre en deux parties peut se faire en commençant par l'étape de prédiction ou l'étape de correction. Dans le premier cas l'initialisation se fera par $\hat{x}_{0/0} = m_0$, et $P_{0/0} = P_0$, alors que dans le second cas les conditions initiales seront les conditions initiales du filtre prédictif-à-un-pas (88).

7 Annexe C : Racine carrée

On désigne par racine carrée d'une matrice $A(n \times n)$, toute matrice carrée S de même taille, telle que $A = S S^T$. S est souvent notée $A^{1/2}$, S^{-1} et S^T sont notées respectivement $A^{-1/2}$, et $A^{T/2}$. D'autres formules de calcul matriciel et factorisations de matrices utiles peuvent être trouvées dans [22].

7.1 Lemme de factorisation matricielle

Théorème 2 *Pour toute matrice $\Lambda(n, p)$, on peut factoriser la matrice $(I_n - \Lambda\Lambda^T)$ sous la forme :*

$$(I_n - \Lambda\Lambda^T) = (I_n - \Lambda\Psi^{-1}\Lambda^T)(I_n - \Lambda\Psi^{-1}\Lambda^T)^T, \quad (92)$$

avec $\Psi = I_p + (I_p - \Lambda^T\Lambda)^{1/2}$.

Pour montrer ce résultat, il suffit de remarquer que le développement de (92) donne :

$$\Lambda\Lambda^T = \Lambda[\Psi^{-1} + \Psi^{-T} - \Psi^{-1}\Lambda^T\Lambda\Psi^{-T}]\Lambda^T. \quad (93)$$

Soient $D = (I_p - \Lambda^T\Lambda)^{1/2}$ et $\Psi = I_p + D$, on obtient les relations suivantes :

$$DD^T = I_p - \Lambda^T\Lambda,$$

$$\begin{aligned} \Psi + \Psi^T - \Lambda^T\Lambda &= I_p + D + D^T + DD^T, \\ &= (I_p + D)(I_p + D)^T, \end{aligned}$$

ce qui conduit à :

$$\Psi^{-1} + \Psi^{-T} - \Psi^{-1}\Lambda^T\Lambda\Psi^{-T} = I_p,$$

et l'identité (93) est vérifiée.

Une forme équivalente de ce théorème est :

$$I_n + \Lambda\Lambda^T = (I_n + \Lambda\Psi^{-1}\Lambda^T)(I_n + \Lambda\Psi^{-1}\Lambda^T)^T,$$

avec $\Psi = I_p + (I_p + \Lambda^T\Lambda)^{1/2}$.

7.2 Décomposition de Cholewsky

La décomposition de Cholewsky permet de factoriser une matrice symétrique définie positive $P(q \times q)$ sous la forme :

$$P = T_I T_I^T,$$

où T_I est une matrice triangulaire inférieure d'ordre q .

Cette décomposition permet donc de déterminer une racine carrée d'une matrice symétrique, sous forme triangulaire inférieure. Un algorithme dual permettrait également de trouver une racine carrée sous la forme triangulaire supérieure. Les coefficients t_{ij} de T_I sont obtenus en écrivant les relations :

$$\begin{aligned} \forall i \in \{1, \dots, q\}, p_{ii} &= \sum_{k=1}^i t_{ik}^2, \\ \forall i \in \{1, \dots, q\}, \forall j \in \{1, \dots, i-1\}, p_{ij} &= \sum_{k=1}^j t_{ik}t_{jk}, \end{aligned} \quad (94)$$

où : $P = [p_{ij}]_{\substack{i \downarrow 1, \dots, q \\ j \rightarrow 1, \dots, q}}, T_I = [t_{ij}]_{\substack{i \downarrow 1, \dots, q \\ j \rightarrow 1, \dots, q}}$, avec $t_{ij} = 0$ pour $j > i$.

Les égalités (94) permettent donc la détermination des coefficients non nuls de T_I par l'algorithme suivant :

- pour i de 1 à q : $t_{i1} = \frac{p_{i1}}{\sqrt{p_{11}}}$;
- pour j de 2 à $i - 1$: $t_{ij} = \frac{1}{t_{jj}} [p_{ij} - \sum_{k=1}^{j-1} t_{ik} t_{jk}]$;
- $t_{ii} = \sqrt{p_{ii} - \sum_{k=1}^{i-1} t_{ik}^2}$.

Le calcul des coefficients t_{ij} de T_I est alors effectué ligne par ligne, de la gauche vers la droite et de haut en bas.

Références

- [1] Arnold, L., *Stochastic differential equations : theory and applications*, John Wiley & Sons, 1974.
- [2] Aström, K.J., *Introduction to stochastic control theory*, Academic Press, 1970.
- [3] Bittanti, S., Laub, A.J., Willems, J.C., Eds, *The Riccati equation*, Springer-Verlag, 1991.
- [4] Borne, P., Dauphin-Tanguy, D., Richard, J.P., Rotella, F., Zambettakis, I., *Commande et optimisation des processus*, Technip, 1992.
- [5] Borne, P., Rotella, F., “Commande optimale”, *Techniques de l’ingénieur*, T.8 : *Mesures et contrôle*, R-7427, Dunod, 1996.
- [6] Boudarel, R., Delmas, J., Guichet, P., *Commande optimale des processus*, tome 1, Dunod, 1967.
- [7] Bozzo, C.A., *Le filtrage optimal*, tome 1, Technique et Documentation, 1982.
- [8] Bozzo, C.A., *Le filtrage optimal*, tome 2, Technique et Documentation, 1983.
- [9] Faure, P., Clerget, M., Germain, F., *Opérateurs rationnels positifs*, Modèles mathématiques pour l’Informatique, vol. 8, Dunod, 1979.
- [10] Favier, G., *Filtrage, modélisation et identification de systèmes linéaires stochastiques à temps discret*, Editions du CNRS, 1982.
- [11] Fraser, D.C., Potter, J.E., “The optimum linear smoother as a combination of two optimum linear filters”, *IEEE Trans. Aut. Control*, vol. AC-14, pp. 387–390, 1969.
- [12] Kalman, R.E., “A new approach to linear filtering and prediction problems”, *Trans. ASME ser. D, J. Basic Engineering*, vol. 82, pp. 35–45, 1960.
- [13] Kalman, R.E., Bucy, R.S., “New results in linear filtering and prediction problems”, *Trans. ASME ser. D, J. Basic Engineering*, vol. 83, pp. 95–108, 1961.
- [14] Kamen, E.W., Su, J.K., *Introduction to optimal estimation*, Springer, 1999.
- [15] Kučera, V., “Riccati equations and their solution”, *The control handbook*, Levine, W., Ed., CRC Press, 1996.
- [16] Lewis, F.L., Syrmos, V.L., *Optimal control*, John Wiley & Sons, 1995.
- [17] Lewis, F.L., *Optimal estimation with an introduction to stochastic control*, John Wiley & Sons, 1986.
- [18] Ljung, L., *System identification : theory for the user*, Prentice Hall, 1987.
- [19] Mc Kean, H.P., Jr, *Stochastic integrals*, Academic Press, 1969.
- [20] O’Reilly, J., *Observers for linear systems*, Academic Press, 1983.
- [21] Papoulis, A., *Probability, random variables and stochastic processes*, Mc-Graw Hill, 1965.
- [22] Rotella, F., Borne, P., *Théorie et pratique du calcul matriciel*, Technip, 1995.
- [23] Walter, E., Pronzato, L., *Identification de modèles paramétriques*, Masson, 1994.
- [24] Wiener, N., *Extrapolation, interpolation ou smoothing of stationary time series*, MIT Press, 1949.
- [25] Zhou, K, Doyle, J., Glover, *Robust and optimal control*, Prentice Hall, 1995.