# –Texte–
# Filtre de Kalman-Bucy

## 1 Le modèle

Un avion se déplace entre Paris et Londres. Il suit une trajectoire théorique appelée trajectoire nominale dont les coordonnées sont connues de tous. La trajectoire de l'avion est suivie au sol par des contrôleurs aériens grâce à un radar qui reçoit un écho de l'avion à intervalles réguliers. La trajectoire effective de l'avion s'écarte de la trajectoire nominale pour de multiples raisons (météorologie, imprécision du pilote automatique, turbulences,...). On note $X_n$ l'écart entre cette trajectoire idéale et la position de l'avion au temps $n$. De plus, on note $Y_n$ la mesure donnée par le radar au temps $n$ (cette mesure est entachée d'erreurs à cause de l'imprécision du radar). Pour simplifier l'étude, on supposera que l'objet observé évolue dans un espace de dimension 1. Le problème qui se pose à l'aiguilleur est d'estimer au mieux la position de l'avion au temps $n$ au vu des observations $Y_0, \dots, Y_n$.

Grâce aux hypothèses présentées plus haut, il paraît naturel de modéliser l'évolution des suites $(X_n)_{n \geqslant 0}$ et $(Y_n)_{n \geqslant 0}$ de la manière suivante. Soit $(V_n)_{n \geqslant 0}$ et $(W_n)_{n \geqslant 0}$ des variables aléatoires indépendantes, gaussiennes, centrées telles que $(W_n)_n$ aient même variance $\sigma^2$ et $(V_n)_n$ aient même variance $\tau^2$. Soit $a$ un nombre réel. On décrit alors l'évolution de $X$ et $Y$ ainsi :

$$\begin{cases} X_0 = W_0, \\ X_n = aX_{n-1} + W_n & n \geqslant 1, \\ Y_n = X_n + V_n & n \geqslant 0. \end{cases} \tag{1}$$

Le but est ici de construire un algorithme efficace permettant de calculer $\mathbb{E}(X_n | Y_0, \dots, Y_n)$ et d'étudier les propriétés de cet objet. On pourra également s'intéresser à l'estimation des paramètres du modèle.

## 2 Quelques rappels sur les vecteurs gaussiens

Toutes les démonstrations concernant le filtre de Kalman-Bucy reposent sur la théorie des vecteurs aléatoires gaussiens. Rappelons leur définition.

**Définition 2.1.** Un vecteur aléatoire $Z$ sur $\mathbb{R}^d$ est gaussien si et seulement si, pour tout $u \in \mathbb{R}^d$ la variable aléatoire $\langle u, Z \rangle$ est gaussienne.

 

**Lemme 2.2.** *Soit $(X, Y)$ un vecteur gaussien dans $\mathbb{R}^2$ tel que $\mathcal{L}(X)$ soit la loi $\mathcal{N}(\mu, \gamma^2)$ et $\mathcal{L}(Y|X)$ soit la loi $\mathcal{N}(\alpha X + \beta, \delta^2)$ alors*

$$\mathcal{L}(X|Y) = \mathcal{N}\left(\rho^2\left(\frac{\mu}{\gamma^2} + \frac{\alpha(Y-\beta)}{\delta^2}\right), \rho^2\right) \quad \textit{avec} \quad \frac{1}{\rho^2} = \frac{1}{\gamma^2} + \frac{\alpha^2}{\delta^2}.$$

Ce résultat élémentaire se généralise à la loi conditionnelle. Nous aurons besoin dans la suite de la forme suivante.

**Proposition 2.3.** *Soit $(X, Y, Y_0, \ldots, Y_{n-1})$ un vecteur gaussien dans $\mathbb{R}^{n+1}$ tel que la loi de $X$ sachant $Y_0, \ldots, Y_{n-1}$ soit la loi $\mathcal{N}(\mu, \gamma^2)$ et la loi de $Y$ sachant $Y_0, \ldots, Y_{n-1}, X$ soit la loi $\mathcal{N}(X, \delta^2)$. Alors la loi de $X$ sachant $Y_0, \ldots, Y_{n-1}, Y$ est la loi normale de paramètres*

$$\rho^2\left(\frac{\mu}{\gamma^2} + \frac{Y}{\delta^2}\right) \quad \textit{et} \quad \rho^2 \quad \textit{où} \quad \frac{1}{\rho^2} = \frac{1}{\gamma^2} + \frac{1}{\delta^2}.$$

## 3 Le filtre optimal

La première remarque importante est que pour tout $n \geqslant 1$, le vecteur aléatoire

$$(X_0, \ldots, X_n, Y_0, \ldots, Y_n)$$

est un vecteur aléatoire gaussien. On en déduit donc que, pour tout $n \geqslant 0$, $(X_n, Y_0, \ldots, Y_n)$ est encore un vecteur gaussien. De plus, la loi de $X_n$ sachant $(Y_0, \ldots, Y_n)$ est une loi normale dont on notera la moyenne $\hat{X}_n$ et la variance $P_n$. On se propose de calculer ces quantités par récurrence.

*Remarque* 3.1. La récurrence se fait en deux étapes. L'étape de prédiction consiste à exprimer la loi $\mathcal{L}(X_n|Y_0, \ldots, Y_{n-1})$ en fonction de $\mathcal{L}(X_{n-1}|Y_0, \ldots, Y_{n-1})$. Puis, dans l'étape de filtrage, on prend en compte l'observation $Y_n$ pour exprimer $\mathcal{L}(X_n|Y_0, \ldots, Y_n)$ en fonction de $\mathcal{L}(X_n|Y_0, \ldots, Y_{n-1})$. C'est l'étape de filtrage.

**Proposition 3.2.** *La loi de $X_n$ sachant $Y_0, \ldots, Y_n$ est la loi normale $\mathcal{N}(\hat{X}_n, P_n)$ où*

$$\hat{X}_n = a\hat{X}_{n-1} + \frac{P_n}{\tau^2}(Y_n - a\hat{X}_{n-1}) \quad \textit{et} \quad P_n = \frac{a^2\tau^2 P_{n-1} + \sigma^2\tau^2}{a^2 P_{n-1} + \sigma^2 + \tau^2}.$$

*Démonstration.* Comme annoncé, on procède par récurrence.
**Initialisation.** Puisque $Y_0 = X_0 + V_0$, la loi de $Y_0$ sachant $X_0$ est tout simplement la loi $\mathcal{N}(X_0, \tau^2)$. Le lemme 2.2 assure que $\mathcal{L}(X_0|Y_0)$ est la loi $\mathcal{N}(\hat{X}_0, P_0)$ où

$$\hat{X}_0 = \frac{\sigma^2}{\sigma^2 + \tau^2} Y_0 \quad \text{et} \quad P_0 = \frac{\sigma^2\tau^2}{\sigma^2 + \tau^2}.$$

On attaque alors la récurrence.

**Prédiction.** Par définition, $\mathcal{L}(X_{n-1}|Y_0,\ldots,Y_{n-1}) = \mathcal{N}(\hat{X}_{n-1},P_{n-1})$. D'après (1), on a

$$\mathcal{L}(X_n|Y_0,\ldots,Y_{n-1}) = \mathcal{N}(a\hat{X}_{n-1},a^2P_{n-1}+\sigma^2).$$

**Filtrage.** D'après (1) à nouveau, il vient

$$\mathcal{L}(Y_n|Y_0,\ldots,Y_{n-1},X_n) = \mathcal{N}(X_n,\tau^2).$$

On applique alors la formule de Bayes conditionnelle (proposition 2.3) pour inverser le conditionnement entre $Y_n$ et $X_n$ : la loi de $X_n$ sachant $Y_0,\ldots,Y_n$ est la loi $\mathcal{N}(\hat{X}_n,P_n)$ avec

$$\frac{1}{P_n} = \frac{1}{a^2P_{n-1}+\sigma^2} + \frac{1}{\tau^2} \quad \text{et} \quad \hat{X}_n = P_n\left(\frac{a\hat{X}_{n-1}}{a^2P_{n-1}+\sigma^2} + \frac{Y_n}{\tau^2}\right),$$

ce qui achève la preuve. □

*Remarque* 3.3. On peut vérifier que

$$P_n = \mathbb{E}\left[(X_n-\hat{X}_n)^2\right] \leqslant \mathbb{E}\left[(X_n-Y_n)^2\right] = \tau^2.$$

Ceci n'est pas surprenant connaissant les propriétés de l'espérance conditionnelle mais souligne bien que l'on gagne effectivement à utiliser toutes les observations $Y_0,\ldots,Y_n$ plutôt que de se contenter de la dernière $Y_n$.

## 4 Estimation de certains paramètres du modèle

On suppose dans cette section que l'on observé les positions de l'avion sans erreurs, c'est-à-dire que l'on connaît la suite $(X_i)_{1\leqslant i\leqslant n}$. On souhaite estimer les coefficients $a$ et $\sigma$. La question n'est pas complètement évidente car les observations ne sont pas indépendantes. La vraisemblance du modèle est donnée par :

$$L(a,\sigma,X_1,\ldots,X_n) = \frac{1}{\sqrt{2\pi\sigma^2}^n}\prod_{k=1}^n \exp\left(-\frac{(X_k-aX_{k-1})^2}{2\sigma^2}\right). \tag{2}$$

**Proposition 4.1.** *Les estimateurs obtenus par la méthode du maximum de vraisemblance sont*

$$\begin{aligned}\hat{a}_n &= \frac{\sum_{k=1}^n X_{k-1}X_k}{\sum_{k=1}^n X_{k-1}^2},\\ \hat{\sigma}_n^2 &= \frac{1}{n}\sum_{k=1}^n (X_k-\hat{a}_nX_{k-1})^2.\end{aligned}$$

*Démonstration.* Les estimateurs $\hat{a}_n$ et $\hat{\sigma}_n$ sont les valeurs des paramètres $a$ et $\sigma$ rendant maximale la quantité (2) (ou son logarithme)... □

On peut montrer que ces estimateurs ont de bonnes propriétés asymptotiques. Supposons que $a$ appartienne à l'intervalle $]-1,1[$ (que cela signifie-t-il pour le pilote ?).

**Théorème 4.2.** *Les estimateurs $\hat{a}$ et $\hat{\sigma}$ sont fortement consistants, c'est-à-dire qu'ils convergent presque sûrement vers $a$ et $\sigma$ respectivement. De plus,*

$$\sqrt{n}(\hat{a}-a) \xrightarrow[n\to\infty]{\mathcal{L}} \mathcal{N}(0,1-a^2) \quad \textit{et} \quad \sqrt{n}\big(\hat{\sigma}^2-\sigma^2\big) \xrightarrow[n\to\infty]{\mathcal{L}} \mathcal{N}(0,2\sigma^4)$$

La suite du paragraphe montre comment obtenir le comportement asymptotique de l'estimateur $\hat{a}_n$ de $a$. On supposera dans la suite $\sigma$ connu et fixé. On réécrit $\hat{a}_n$ de la manière suivante :

$$\hat{a}_n = a + \frac{\sum_{k=1}^n X_{k-1}W_k}{\sum_{k=1}^n X_{k-1}^2}.$$

On définit alors la suite aléatoire $(M_n)_{n\in\mathbb{N}}$ en posant

$$M_0 = 0 \quad \text{et, pour } n \geqslant 1, \quad M_n = \sum_{k=1}^n X_{k-1}W_k.$$

La suite $(M_n)_{n\in\mathbb{N}}$ est une martingale par rapport à la filtration $(\mathcal{F}_n)_n$, où $\mathcal{F}_n$ est la tribu engendrée par les variables aléatoires $(W_k)_{1\leqslant k\leqslant n}$. On lui associe sa variation quadratique $(\langle M\rangle_n)_n$ définie de la manière suivante : c'est la seule (p.s.) suite $(A_n)_n$ (*a priori* aléatoire) telle que $M_n^2 - A_n$ soit une martingale par rapport à $(\mathcal{F}_n)$.

**Théorème 4.3.** *La variation quadratique de $M$ est donnée par*

$$\langle M\rangle_0 = 0 \quad \textit{et, pour } n \geqslant 1, \quad \langle M\rangle_n = \sigma^2 \sum_{k=1}^n X_{k-1}^2.$$

*De plus, si $|a|<1$ alors*

$$\frac{\langle M\rangle_n}{n} \xrightarrow[n\to\infty]{p.s.} \frac{\sigma^4}{(1-a^2)}, \qquad \frac{M_n}{\langle M\rangle_n} \xrightarrow[n\to\infty]{p.s.} 0 \quad \textit{et} \quad \frac{M_n}{\sqrt{\langle M\rangle_n}} \xrightarrow[n\to\infty]{\mathcal{L}} \mathcal{N}(0,1).$$

*Remarque* 4.4. Dans le cas $|a|<1$, on a en particulier $\langle M\rangle_n \xrightarrow[n\to\infty]{p.s.} +\infty$.

*Démonstration.* Par une récurrence immédiate, on a

$$X_n = a^n X_0 + \sum_{k=1}^n a^{n-k}W_k.$$

Supposons que $X_0$ soit nul. L'inégalité de Cauchy-Schwarz assure que

$$X_n^2 \leqslant \frac{1}{1-|a|}\sum_{k=1}^{n}|a|^{n-k}W_k^2.$$

Posons $S_n = \sum_{k=1}^n X_{k-1}^2$ et $L_n = \sum_{k=1}^n W_{k-1}^2$. La majoration ci-dessus montre que $S_n = O(L_n)$. D'autre part, en vertu de la loi des grands nombres, $L_n = O(n)$ et par suite $S_n = O(n)$ et $\langle M\rangle_n = O(n)$. Un résultat de martingale assure qu'alors $M_n = o(n)$.

On peut également écrire

$$S_n = a^2 S_{n-1} + 2aM_n + L_n \quad \text{et donc} \quad (1-a^2)\frac{S_n}{n} = \frac{a^2X_n^2}{n} + \frac{2aM_n}{n} + \frac{L_n}{n}.$$

Ainsi donc, la suite $S_n/n$ converge-t-elle presque sûrement vers $\sigma^2/(1-a^2)$. On a donc montré que

$$\frac{\langle M\rangle_n}{n} \xrightarrow[n\to\infty]{p.s.} \frac{\sigma^4}{(1-a^2)}.$$

La fin du théorème repose sur la loi des grands nombres et le théorème limite central pour les martingales de carré intégrable. □

# 5 Suggestions.

Pour traiter le sujet, on suggère de répondre à certaines des questions suivantes.

1. Commenter le modèle de la section 1. On commentera en particulier le choix des lois des variables aléatoires introduites et leur indépendance.
2. Expliquer l'interprétation physique du coefficient $a$ (et de sa position par rapport à 1). Pourquoi est-il légitime de supposé $|a| < 1$ s'il y a un pilote dans l'avion ? Pourquoi $a \in ]-1,0[$ correspond à un pilote en phase d'apprentissage ?
3. Démontrer le lemme 2.2 ou la proposition 2.3.
4. Démontrer la proposition 3.2 et en déduire l'algorithme de Kalman-Bucy.
5. Écrire une fonction qui prend en entrée le temps final d'observation $N$, $a$, $\sigma$ et $\tau^2$ et génère les trajectoires $(X_n)_{0\leqslant n\leqslant N}$, $(Y_n)_{0\leqslant n\leqslant N}$ et $(\hat{X}_n)_{0\leqslant n\leqslant N}$ jusqu'au temps $N$.
6. Étudier l'évolution temporelle de $P_n$ en fonction du coefficient $a$.
7. Démontrer la proposition 4.1.
8. Illustrer par la simulation le théorème 4.2.
9. Comment utiliser le théorème 4.2 pour répondre, à partir de l'observation de $X_1,\dots,X_n$ à la question suivante : le pilote est-il vraiment expérimenté ?
10. On pourra faire quelques commentaires au sujet du théorème 4.3 ($M$ est-elle vraiment une martingale ? $\langle M\rangle$ est-il bien son processus croissant ?).