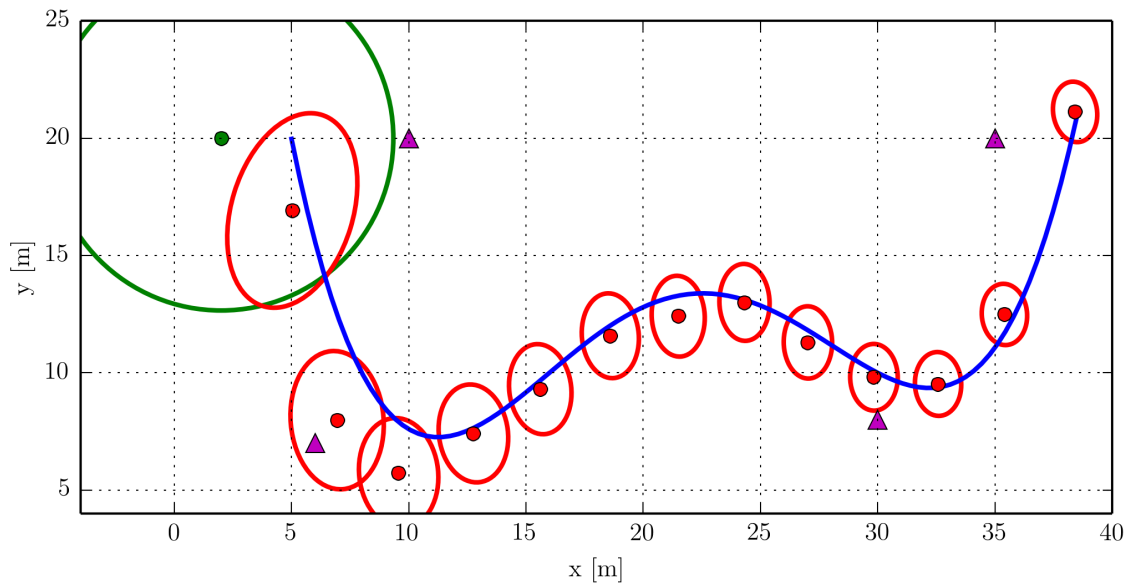


Introduction au filtre de Kalman



Pierre Bosser & Guillaume Sicot

`prenom.nom@ensta-bretagne.fr`

Avant-propos

Ce cours est une adaptation pour l'option Hydrographie - Océanographie de l'ENSTA Bretagne du cours de Luc Jaulin [[Jaulin, 2014](#)] dispensé aux étudiants de l'option SPID, et plutôt focalisé sur les aspects de robotique mobile.

L'objectif est d'introduire le filtre de Kalman ; son utilisation sera approfondie dans le cadre de l'UV de positionnement sous-marin de troisième année.

Citation : *Bosser, P. & Sicot, G. : « Introduction au filtre de Kalman », Cours de l'École Nationale Supérieure des Techniques Avancées de Bretagne, 2019.*

Version du 20 mai 2019

Table des matières

1	Vecteur aléatoire	5
1.1	Généralités sur les vecteurs aléatoires	6
1.2	Couple de vecteur aléatoire	8
1.3	Application au vecteur aléatoire gaussien	10
2	Estimation linéaire	15
2.1	Notion d'estimateur	16
2.2	Estimateur linéaire orthogonal sans biais	17
2.3	BLUE pour un modèle linéaire	20
3	Filtre de Kalman	23
3.1	Contexte	24
3.2	Équations du filtre de Kalman	24
3.3	Filtre de Kalman étendu	27
3.4	Pour aller plus loin...	30

VECTEUR ALÉATOIRE

Sommaire

1.1 Généralités sur les vecteurs aléatoires	6
1.1.1 Définition	6
1.1.2 Fonction de répartition et densité de probabilité	6
1.1.3 Moments	7
1.2 Couple de vecteur aléatoire	8
1.2.1 Indépendance	8
1.2.2 Propagation de la variance	9
1.2.3 Matrice de covariance croisée	9
1.3 Application au vecteur aléatoire gaussien	10
1.3.1 Définitions	10
1.3.2 Ellipsoïde de confiance	12

En guise de mise en bouche, nous revenons ici sur la notion de vecteur aléatoire et sur ses propriétés. L'accent est porté sur les matrices de covariance, maillon essentiel du filtre de Kalman et de l'estimation en général.

Les résultats présentés dans ce chapitre sont essentiellement tirés de [Osswald, 2015; Jaulin, 2014].

1.1 Généralités sur les vecteurs aléatoires

1.1.1 Définition

Un **vecteur aléatoire** réel X est un vecteur de \mathbb{R}^n dont les composantes $(X_i)_{i \in \llbracket 1; n \rrbracket}$ sont des variables aléatoires réelles :

$$X = \begin{pmatrix} X_1 \\ \vdots \\ X_n \end{pmatrix}$$

1.1.2 Fonction de répartition et densité de probabilité

Soit $x = (x_1, \dots, x_n)^T$ un vecteur de \mathbb{R}^n .

On définit pour le vecteur aléatoire X la **fonction de répartition** :

$$F_X : \begin{cases} \mathbb{R}^n & \longrightarrow & [0; 1] \\ x & \longmapsto & F_X(x) = F_{X_1, \dots, X_n}(x_1, \dots, x_n) \\ & & = p(X_1 \leq x_1, \dots, X_n \leq x_n) \end{cases}$$

Où p désigne la mesure de probabilité sur \mathbb{R} .

En dérivant la fonction de répartition selon chacune de ses composantes, on définit la **fonction de densité** du vecteur aléatoire :

$$f_X : \begin{cases} \mathbb{R}^n & \longrightarrow & \mathbb{R} \\ x & \longmapsto & f_X(x) = f_{X_1, \dots, X_n}(x_1, \dots, x_n) \\ & & = \frac{\partial^n F_{X_1, \dots, X_n}}{\partial x_1 \dots \partial x_n}(x_1, \dots, x_n) \end{cases}$$

Si on connaît la fonction de densité de X , il est possible de retrouver la densité de probabilité de chacune de ses composantes. On parle de **densité de probabilité marginale**. Pour une composante X_k , elle est donnée par :

$$f_{X_k} : \begin{cases} \mathbb{R} & \longrightarrow & \mathbb{R} \\ x_k & \longmapsto & f_{X_k}(x_k) = \int \dots \int_{\mathbb{R}^{n-1}} f_{X_1, \dots, X_n}(x_1, \dots, x_n) dx_1 \dots dx_{k-1} dx_{k+1} \dots dx_n \end{cases}$$

Les composantes du vecteur aléatoire X sont dites **indépendantes** si :

$$f_X(x) = \prod_k f_{X_k}(x_k)$$

Si on connaît une des composantes de X , X_k , on définit sa **densité de probabilité conditionnelle** :

$$f_{X_1, \dots, X_{k-1}, X_{k+1}, \dots, X_n | X_k = x_k}(x_1, \dots, x_{k-1}, x_{k+1}, \dots, x_n) = \frac{f_{X_1, \dots, X_n}(x_1, \dots, x_n)}{f_{X_k}(x_k)}$$

1.1.3 Moments

Espérance

L'espérance d'un vecteur aléatoire est définie par :

$$E(X) = \bar{X} = \int \dots \int_{\mathbb{R}} x^T f_X(x) dx$$

C'est un vecteur déterministe, de dimension n .

On peut également l'obtenir par le calcul de l'espérance de chacune de ses variables aléatoires :

$$\bar{X} = \begin{pmatrix} E(X_1) \\ \vdots \\ E(X_n) \end{pmatrix}$$

En langage **Python** si on dispose d'une réalisation de taille p d'un vecteur aléatoire de taille n , \tilde{X} , on peut estimer son espérance à l'aide de l'estimateur empirique de l'espérance pour chacune de ses composantes.

Python

```
# Xtilde est un np.array de taille (n,p) : p réalisations du vecteur aléatoire
# X de dimension n
Xhat = np.array( [np.mean(Xtilde, axis = 1)] ).T
```

Matrices de covariance

La **matrice de covariance** (ou **matrice de variance-covariance**) de X est la matrice définie par :

$$\Sigma_X = E \left\{ [X - \bar{X}] \cdot [X - \bar{X}]^T \right\}$$

Formulation que l'on peut développer en fonction des covariances sur les couples de variables aléatoires composant X :

$$\Sigma_X = \begin{pmatrix} \text{Cov}(X_1, X_1) & \dots & \text{Cov}(X_1, X_n) \\ \vdots & \ddots & \vdots \\ \text{Cov}(X_1, X_n) & \dots & \text{Cov}(X_n, X_n) \end{pmatrix}$$

En langage `Python` si on dispose d'une réalisation de taille p d'un vecteur aléatoire de taille n , \tilde{X} , on peut estimer sa matrice de covariance à l'aide de l'estimateur empirique de la covariance.

Python

```
# Xtilde est un np.array de taille (n,p) : p réalisations du vecteur aléatoire
# X de dimension n
Xhat = np.array( [np.mean(Xtilde, axis = 1)] ).T
deltaX = Xtilde - np.dot( Xhat, np.ones( (1,Xtilde.shape[1]) ) )
SigmaXhat = np.dot( deltaX, deltaX.T )/ deltaX.shape[1]
```

Soit Σ_X une matrice de covariance. Citons quelques-unes de ses propriétés :

- C'est une matrice **symétrique** et **semi-définie positive**.
- Elle est donc **diagonalisable** avec des **valeurs propres positives**. On peut donc la décomposer sous la forme :

$$\Sigma_X = R \times D \times R^{-1}$$

Où R est une matrice de rotation et D une matrice diagonale.

- Ses vecteurs propres sont **orthogonaux** entre eux.
- Si les différentes composantes de X , X_1, \dots, X_n sont **indépendantes**, la matrice de covariance de X est diagonale.
- Σ_X possède une unique **racine carrée** qui est également une matrice de covariance. On la note $\Sigma_X^{1/2}$:

$$\Sigma_X = \Sigma_X^{1/2} \times \Sigma_X^{1/2}$$

$\Sigma_X^{1/2}$ est donc également diagonalisable et ses valeurs propres sont les valeurs racines carrées des valeurs propres de Σ_X .

Soient Σ_X et Σ_Y deux matrices de covariance. La matrice $\Sigma = \Sigma_X + \Sigma_Y$ est également une matrice de covariance. On peut alors définir la **relation d'ordre** :

$$\Sigma_1 \leq \Sigma_2 \quad \Leftrightarrow \quad \Sigma_2 - \Sigma_1 \text{ est une matrice de covariance}$$

1.2 Couple de vecteur aléatoire

On considère dans cette section 2 vecteurs aléatoires réels X et Y de dimensions respectives n et p .

1.2.1 Indépendance

Deux vecteurs aléatoires sont **indépendants** si leur fonction de densité jointe est égale aux produits de leurs fonctions de densité marginales.

1.2.2 Propagation de la variance

On suppose les vecteurs aléatoires X , Y et Z liés par la relation linéaire :

$$Z = A \cdot X + Y + B$$

Où A est une matrice carrée (déterministe) de taille $p \times n$, B un vecteur (déterministe) de taille $n \times 1$. Les vecteurs X et Y sont supposés indépendants.

L'espérance de Z est donnée par :

$$\bar{Z} = A \cdot \bar{X} + \bar{Y} + B$$

La matrice de covariance de Z est donnée par :

$$\Sigma_Z = A \Sigma_X A^T + \Sigma_Y$$

Cette dernière relation est appelée **loi de propagation de la variance**.

1.2.3 Matrice de covariance croisée

On définit la **matrice de covariance croisée** (appelée **matrice de covariance** par abus de langage) des vecteurs aléatoires X et Y :

$$\Sigma_{XY} = E \left\{ [X - \bar{X}] \cdot [Y - \bar{Y}]^T \right\}$$

C'est une matrice de dimension $n \times p$.

On a les propriétés suivantes :

- La transposée d'une matrice de covariance croisée est donnée par :

$$\Sigma_{XY}^T = \Sigma_{YX}$$

- X et Y sont dits orthogonaux si leur matrice de covariance croisée est nulle.

En langage **Python** si on dispose de réalisations de taille p de vecteurs aléatoires de taille n et m , \tilde{X} et \tilde{Y} , on peut estimer la matrice de covariance de ces deux vecteurs à l'aide de l'estimateur empirique de la covariance.

Python

```
# Xtilde est un np.array de taille (n,p) : p réalisations du vecteur aléatoire
# X de dimension n
# Ytilde est un np.array de taille (m,p) : p réalisations du vecteur aléatoire
# Y de dimension m
Xhat = np.array( [np.mean(Xtilde, axis = 1)] ).T
Yhat = np.array( [np.mean(Ytilde, axis = 1)] ).T
deltaX = Xtilde - np.dot( Xhat, np.ones( (1,Xtilde.shape[1]) ) )
deltaY = Ytilde - np.dot( Yhat, np.ones( (1,Ytilde.shape[1]) ) )
SigmaXhatYhat = np.dot( deltaX, deltaY.T )/deltaX.shape[1]
```

1.3 Application au vecteur aléatoire gaussien

1.3.1 Définitions

Un **vecteur aléatoire** réel X est **gaussien** ou **normalement distribué** si toute combinaison linéaire de ses composantes est gaussienne.

La densité de probabilité de X est alors donnée par :

$$f_X(x) = \frac{1}{\sqrt{2^n \det(\Sigma_X)}} \exp \left[-\frac{1}{2} (x - \bar{X}) \Sigma_X^{-1} (x - \bar{X}) \right]$$

On note alors :

$$X \hookrightarrow \mathcal{N}(\bar{X}, \Sigma_X)$$

Énonçons quelques propriétés des vecteurs aléatoires gaussiens :

- Chacune des composantes de X est également gaussienne.
- Tout vecteur aléatoire peut se formuler sous la forme :

$$X = \bar{X} + \Sigma_X^{1/2} \cdot Z$$

Où Z est un vecteur aléatoire centré réduit :

$$Z \hookrightarrow \mathcal{N}(0, I)$$

Cette dernière propriété est notamment utilisée lors de la réalisation numérique d'un vecteur aléatoire gaussien.

- Tout vecteur résultant de la combinaison linéaire d'un vecteur aléatoire gaussien est également gaussien :

$$X \hookrightarrow \mathcal{N}(\bar{X}, \Sigma_X) \text{ et } Y = A \cdot X + B \implies Y \hookrightarrow \mathcal{N}(A\bar{X} + B, A\Sigma_X A^T)$$

- Deux vecteurs aléatoires gaussiens **indépendants** sont **orthogonaux** (et inversement).

EXEMPLE

Soit le vecteur aléatoire gaussien X défini par :

$$\bar{X} = \begin{pmatrix} 1 \\ 2 \end{pmatrix} \quad \Sigma_X = \begin{pmatrix} \frac{3}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{3}{2} \end{pmatrix}$$

Sous **Python**, On peut simuler 10^4 réalisations de ce vecteur à partir de réalisations d'un vecteur aléatoire centré réduit Z en utilisant les commandes suivantes :

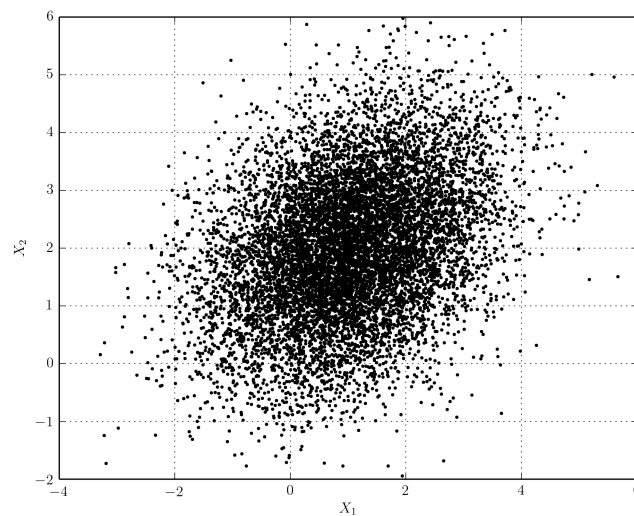
```

Python
import numpy as np
from scipy.linalg import sqrtm
from scipy.stats import norm

Xbar = np.array([[1],[2]])
SigmaX = 0.5*np.array([[3,1], [1,3]])
Z = norm.rvs(size=(2,10000))
X = np.dot(Xbar, np.ones((1,Z.shape[1]))) + np.dot(sqrtm(SigmaX),Z)

```

Si on représente dans le plan les réalisations de ce vecteur aléatoire, on obtient la figure suivante :



On définit maintenant un vecteur aléatoire Y , relié à X par :

$$Y = A \cdot X + B$$

Où :

$$A = \begin{pmatrix} 2 & 1 \\ -1 & 1 \end{pmatrix} \quad B = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$$

Y est donc un vecteur aléatoire gaussien tel que :

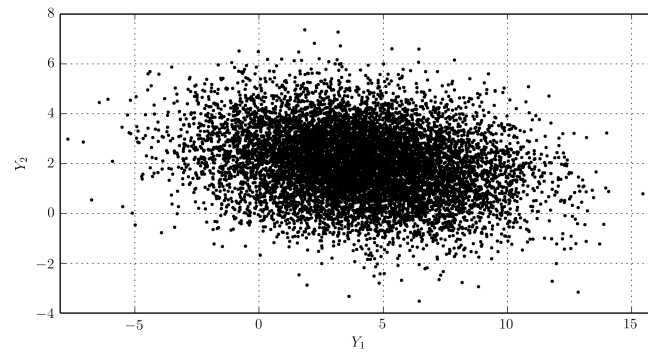
$$\bar{Y} = A\bar{X} + B = \begin{pmatrix} 4 \\ 2 \end{pmatrix} \quad \Sigma_Y = A\Sigma_X A^T = \begin{pmatrix} \frac{19}{2} & -1 \\ -1 & 2 \end{pmatrix} \quad \Sigma_{XY} = \Sigma_X A^T = \begin{pmatrix} \frac{7}{2} & -1 \\ \frac{5}{2} & 1 \end{pmatrix}$$

On déduit de la réalisation de X celle de Y :

```

Python
A = np.array([[2,1], [-1,1]])
B = np.array([[0],[1]])
Y = np.dot(B, np.ones((1,X.shape[1]))) + np.dot(A, X)

```



On estime numériquement les caractéristiques de Y :

Python

```
Yhat = np.array([np.mean(Y, axis = 1)]).T
deltaY = Y - np.dot( Yhat, np.ones( (1,Y.shape[1]) ) )
SigmaYhat = np.dot( deltaY, deltaY.T )/deltaY.shape[1]
```

Soit :

$$\hat{Y} \approx \begin{pmatrix} 3,97 \\ 1,99 \end{pmatrix} \quad \Sigma_{\hat{Y}} \approx \begin{pmatrix} 9,54 & -1,02 \\ -1,02 & 2,00 \end{pmatrix}$$

On peut également estimer la matrice de corrélation croisée :

Python

```
deltaX = X - np.dot( Xhat, np.ones( (1,Y.shape[1]) ) )
SigmaXhatYhat = np.dot( deltaX, deltaY.T )/deltaX.shape[1]
```

Soit :

$$\Sigma_{\hat{X}\hat{Y}} \approx \begin{pmatrix} 3,52 & -0,98 \\ 2,50 & 0,96 \end{pmatrix}$$

1.3.2 Ellipsoïde de confiance

On considère toujours un vecteur aléatoire X gaussien de dimension n . Soit une extraction de m éléments du vecteur X , notée X_m , de matrice de covariance Σ_{X_m} , elle-même extraite de Σ_X .

Soit $\lambda_m > 0$. On appelle **ellipsoïde de confiance** le domaine délimité par la courbe donnée par :

$$\mathcal{E}_m = \{(X_m - \bar{X}_m)^T \Sigma_{X_m}^{-1} (X_m - \bar{X}_m) = \lambda_m^2\}$$

Alors :

- Le scalaire λ_m suit une loi du χ^2 à m degrés de liberté :

$$\lambda_m \hookrightarrow \chi_m^2$$

On utilise donc les tables de fractiles pour déterminer λ_m à un niveau de confiance α .

- Les demi-axes a_i de l'ellipsoïde ainsi défini ont pour valeurs :

$$a_i = \lambda_m d_i$$

Où D , de terme générique d_i^2 , est la matrice diagonale de termes positifs, issue de la diagonalisation de Σ_{X_m} :

$$\Sigma_{X_m} = R \times D \times R^T$$

DÉMONSTRATION

Pour déterminer les limites de ce domaine, recherchons donc λ_m tel que :

$$(X_m - \bar{X}_m)^T \Sigma_{X_m}^{-1} (X_m - \bar{X}_m) = \lambda_m^2$$

Nous avons vu précédemment que la matrice Σ_{X_m} est symétrique et définie positive. On peut donc la diagonaliser sous la forme :

$$D = R^T \Sigma_{X_m} R \quad \Leftrightarrow \quad \Sigma_{X_m} = R D R^T$$

Où D est une matrice diagonale de terme générique positif $(d_i^2)_{i \in \llbracket 1; m \rrbracket}$, R est une matrice orthogonale ($R^T = R^{-1}$).

On pose $U = R^T (X_m - \bar{X}_m)$, de terme générique $(u_i)_{i \in \llbracket 1; m \rrbracket}$. On remarque que $U \hookrightarrow \mathcal{N}(0, D)$. En effet, U est une combinaison linéaire d'un vecteur gaussien et est donc gaussien. De plus :

$$\bar{U} = R^T (\bar{X}_m - \bar{X}_m) = 0 \quad \text{et} \quad \Sigma_U = R^T \Sigma_{X_m} R = D$$

On peut alors écrire :

$$\begin{aligned} (X_m - \bar{X}_m)^T \Sigma_{X_m}^{-1} (X_m - \bar{X}_m) &= \lambda_m^2 \\ \Leftrightarrow (X_m - \bar{X}_m)^T R D^{-1} R^T (X_m - \bar{X}_m) &= \lambda_m^2 \\ \Leftrightarrow U^T D^{-1} U &= \lambda_m^2 \\ \Leftrightarrow \sum_{i=1}^m \frac{u_i^2}{d_i^2} &= \lambda_m^2 \\ \Leftrightarrow \sum_{i=1}^m \frac{u_i^2}{\text{Var } u_i} &= \lambda_m^2 \end{aligned}$$

λ_m est donc défini par une somme de m variables aléatoires normales centrées réduites, soit une loi du χ^2 à m degrés de liberté :

$$\lambda_m^2 \hookrightarrow \chi_m^2$$

La forme ainsi mise en évidence est bien un ellipsoïde de dimension m dans un repère ortho-normal (donné par la diagonalisation de Σ_{X_m}). Ses demi-axes (a_i) sont donnés par :

$$a_i = \lambda_m \cdot d_i$$

Quelques propriétés des ellipsoïdes de confiance :

- En dimension 2 ou 3, R est une **matrice de rotation**.

- En dimension 2 ou 3, on retrouve respectivement l'équation d'une **ellipse** ou d'un **ellipsoïde**, d'où l'appellation.

Les termes diagonaux de la matrice D représentent alors les carrés des demi-axes de l'ellipse (ou de l'ellipsoïde), la matrice de rotation R permettant de retrouver l'orientation de l'ellipse (ou de l'ellipsoïde).

- En positionnement terrestre, il est très fréquent d'utiliser les ellipses de confiance pour caractériser l'estimation d'une position planimétrique. La composante verticale est quant à elle qualifiée à partir d'un intervalle de confiance seul.

EXEMPLE

Soit le vecteur aléatoire gaussien X défini par :

$$\bar{X} = \begin{pmatrix} 1 \\ 2 \end{pmatrix} \quad \Sigma_X = \begin{pmatrix} \frac{3}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{3}{2} \end{pmatrix}$$

En diagonalisant la matrice Σ_X , on obtient :

$$\Sigma_X = \begin{pmatrix} \frac{\sqrt{2}}{2} & -\frac{\sqrt{2}}{2} \\ \frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} \end{pmatrix} \cdot \begin{pmatrix} 2 & 0 \\ 0 & 1 \end{pmatrix} \cdot \begin{pmatrix} \frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} \\ -\frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} \end{pmatrix}$$

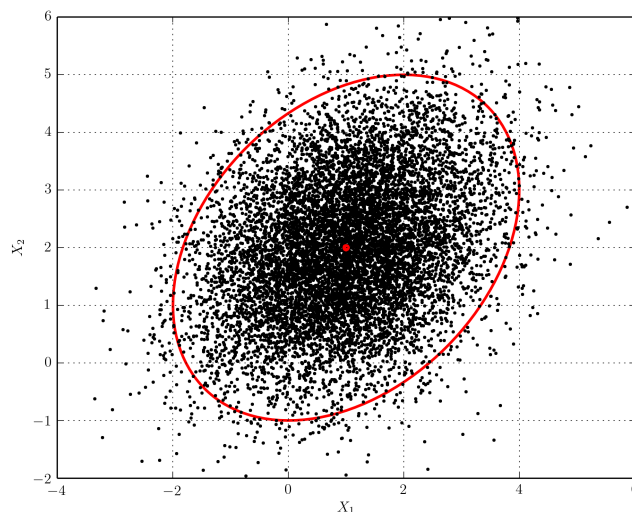
L'ellipse d'erreur à 95% sur X aura donc les caractéristiques suivantes :

- Son grand axe est orienté de 45° par rapport à l'horizontale.
- Ses demi axes auront pour valeur :

$$\begin{aligned} a_1 &\approx 3.46 \\ a_2 &\approx 2.45 \end{aligned}$$

Pour $\lambda_m \approx 2.45$.

On représente sur la figure suivant l'ellipse d'erreur à 95%. On a également représenté 10^4 réalisations du vecteur aléatoire X . On retrouve bien les résultats attendus, en terme d'orientation et de dimension de l'ellipse.



ESTIMATION LINÉAIRE

Sommaire

2.1	Notion d'estimateur	16
2.1.1	Définition	16
2.2	Estimateur linéaire orthogonal sans biais	17
2.3	BLUE pour un modèle linéaire	20

Nous revenons ici sur le principe de l'estimation linéaire et mettons en évidence la formulation du meilleur estimateur linéaire non biaisé qui constitue la base du filtre de Kalman.

Les résultats présentés dans ce chapitre sont essentiellement tirés de [\[Jaulin, 2014\]](#).

2.1 Notion d'estimateur

2.1.1 Définition

Soient X et Y deux vecteurs aléatoires réels de dimension n et p . On souhaite estimer des propriétés de X à partir de la connaissance de Y . Ce problème est un problème d'**estimation**.

On appelle **estimateur** une application de variables aléatoires (également appelée statistique), notée \hat{X} , et donc définie par $\hat{X} = \phi(Y)$.

L'objectif de l'estimateur est de se rapprocher (au moins asymptotiquement) de la grandeur à estimer. Un estimateur de l'espérance de X conditionnée par Y , est appelé espérance conditionnelle et se note $E[X/Y]$. L'expression analytique de l'espérance conditionnelle est bien souvent délicate à obtenir.

REMARQUE

L'espérance conditionnelle est elle-même une variable aléatoire.

L'**erreur d'estimation**, notée ϵ , est également une fonction de variables aléatoires et se définit comme suit :

$$\epsilon = \hat{X} - X$$

Retenons les qualifications suivantes d'un estimateur :

- Un estimateur est dit **sans biais** si $E[\epsilon] = 0$
- Un estimateur est dit **orthogonal** si $E[\epsilon(Y - \bar{Y})] = 0$
- Un estimateur est dit **linéaire** s'il peut s'écrire sous la forme suivante : $\hat{X} = K \cdot Y + B$, où K est une matrice de $\mathbb{R}^{n \times p}$ et B un vecteur de \mathbb{R}^n .

Dans la suite de ce chapitre, nous nous limitons au cas de l'estimateur linéaire.

EXEMPLE

On considère un vecteur aléatoire X de dimension 1 suivant une loi normale de moyenne 3 et de variance 4. On définit le vecteur aléatoire Y , lui aussi de dimension 1, par :

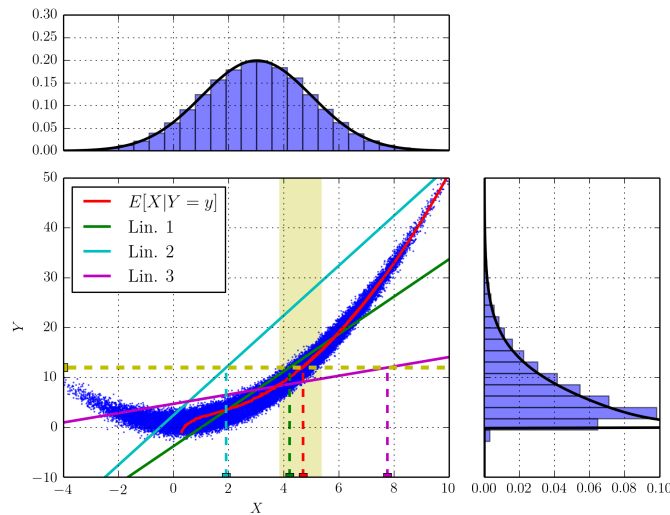
$$Y = \frac{1}{2}X^2 + 1$$

On simule des réalisations de ce vecteur X et on en déduit les réalisations correspondantes de Y .

Essayons d'interpréter graphiquement les définitions précédentes.

Sur la figure suivante, on représente :

- Les couples de réalisations (X, Y) (bleu), leurs répartitions respectives (sous forme d'histogrammes, en bleu) et densités de probabilité marginale (courbes noires).
- Un estimateur de $E[X|Y = y]$ (rouge)
- Trois estimateurs linéaires (vert, cyan et magenta, respectivement Lin. 1, Lin. 2 et Lin. 3).



On suppose disposer d'une réalisation particulière de Y , $\tilde{Y} = 12$ (tirets jaunes) ; on est amené à penser qu'une valeur possible de X connaissant cette valeur de Y est dans la zone délimitée de couleur jaune, entre 4 et 6 environ.

L'espérance conditionnelle $E[X|Y]$ suit de manière cohérente le nuage de points ; il en est de même pour l'estimateur linéaire 1. Les 2 autres estimateurs linéaires semblent ici moins pertinents.

2.2 Estimateur linéaire orthogonal sans biais

Soient X et Y deux vecteurs aléatoires réels de dimension n et m et de matrice de covariance Σ_X et Σ_Y .

THÉORÈME Il existe **un unique estimateur linéaire orthogonal sans biais** permettant l'estimation de X à partir d'une réalisation de Y (ou connaissant Y). Celui-ci est donné par :

$$\hat{X} = \bar{X} + K\delta_Y$$

Avec :

- $\delta_Y = Y - \bar{Y}$.
- $K = \Sigma_{XY}\Sigma_Y^{-1}$.

DÉMONSTRATION

Estimateur centré : calculons l'espérance de l'erreur d'estimation pour un estimateur linéaire

quelconque.

$$\begin{aligned} E(\varepsilon_X) &= E(\hat{X} - X) \\ &= E(KY + B - X) \\ &= K\bar{Y} + B - \bar{X} \end{aligned}$$

L'estimateur est sans biais si et seulement si :

$$E(\varepsilon_X) = 0 \Leftrightarrow B = \bar{X} - K\bar{Y}$$

Donc, l'estimateur est centré si et seulement si :

$$\hat{X} = \bar{X} + K\delta_Y$$

Estimateur orthogonal : Reformulons l'erreur d'estimation :

$$\begin{aligned} \varepsilon_X &= \hat{X} - X \\ &= \bar{X} + K\delta_Y - \bar{X} - \delta_X \\ &= K\delta_Y - \delta_X \end{aligned}$$

Donc :

$$\begin{aligned} E(\varepsilon_X \delta_Y^T) &= E(K\delta_Y \delta_Y^T - \delta_X \delta_Y^T) \\ &= K\Sigma_Y - \Sigma_{XY} \end{aligned}$$

L'estimateur est donc orthogonal si et seulement si :

$$E(\varepsilon_X \delta_Y^T) = 0$$

Soit encore :

$$K = \Sigma_{XY} \Sigma_Y^{-1}$$

THÉORÈME La matrice de covariance de l'estimateur linéaire orthogonal sans biais est donnée par :

$$\Sigma_{\hat{X}} = \Sigma_X - K \Sigma_{XY}^T$$

On parle aussi de matrice de covariance **a posteriori**.

DÉMONSTRATION

Calculons la covariance de l'erreur d'estimation, $\varepsilon_X = \hat{X} - X$:

$$\begin{aligned} \Sigma_{\hat{X}} &= E(\varepsilon_X \varepsilon_X^T) \\ &= E[(K\delta_Y - \delta_X)(K\delta_Y - \delta_X)^T] \\ &= K\Sigma_Y K^T - \Sigma_{XY} K^T - K\Sigma_{YX} + \Sigma_X \end{aligned}$$

Or, $K = \Sigma_{XY} \Sigma_Y^{-1}$. On a donc :

$$\Sigma_{\hat{X}} = \Sigma_X - K \Sigma_{XY}^T$$

THÉORÈME L'estimateur linéaire orthogonal sans biais est dit **optimal** dans le sens où il

est de covariance minimale.

DÉMONSTRATION

Soit une matrice $G = K + \Delta$ définissant un autre estimateur linéaire orthogonal sans biais. Calculons la matrice de covariance a posteriori associée :

$$\begin{aligned}\Sigma_{\hat{X}}(G) &= E[\varepsilon_X(G)\varepsilon_X^T(G)] \\ &= G\Sigma_Y G^T - \Sigma_{XY} G^T - G\Sigma_{YX} + \Sigma_X \\ &= (K + \Delta)(\Sigma_Y K^T + \Sigma_Y \Delta^T) - (\Sigma_{XY} K^T + \Sigma_{XY} \Delta^T) - (K\Sigma_{XY}^T + \Delta\Sigma_{XY}^T) + \Sigma_X\end{aligned}$$

On développe ce dernier membre en utilisant les relations $K = \Sigma_{XY}\Sigma_Y^{-1}$, $\Sigma_Y^{-1} = \Sigma_Y^T$ et $\Sigma_{XY} = \Sigma_{YX}^T$:

$$\begin{aligned}\Sigma_{\hat{X}}(G) &= -K\Sigma_{XY}^T + \Delta\Sigma_Y\Delta^T + \Sigma_X \\ &= \Sigma_{\hat{X}} + \Delta\Sigma_Y\Delta^T\end{aligned}$$

Donc, pour toute matrice Δ non nulle, on a :

$$\Sigma_{\hat{X}}(G) > \Sigma_{\hat{X}}$$

$\Sigma_{\hat{X}}$ est bien minimale.

On désigne fréquemment l'estimateur linéaire orthogonal sans biais par l'acronyme **BLUE** (*Best Linear Unbiased Estimator*).

REMARQUE

Si le modèle linéaire est valide et que les erreurs d'estimation sont non-corrélées, d'espérance nulle et de même variance, alors l'estimateur linéaire orthogonal sans biais est également l'estimateur des moindres carrés (Théorème de Gauss-Markov).

Par ailleurs dans le cas de variables aléatoires gaussiennes, le BLUE est également l'estimateur du maximum de vraisemblance.

EXEMPLE

On considère à nouveau les vecteurs aléatoires X et Y définis précédemment.

Le BLUE de $X|Y$ est donné par :

$$\hat{X} = \bar{X} + K\delta_Y$$

Les matrices de covariances sont déterminées à l'aide des estimateurs empiriques :

$$\begin{aligned}\Sigma_{XY} &\approx 12,17 \\ \Sigma_Y &\approx 45,86\end{aligned}$$

On en déduit la valeur de K :

$$K = 0.265$$

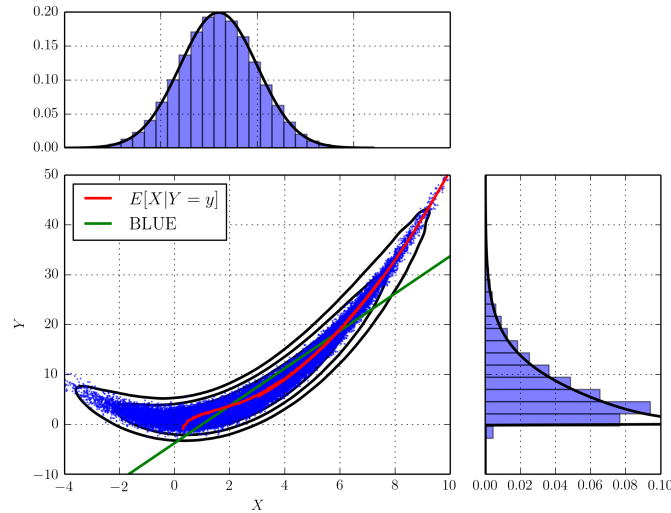
Essayons d'interpréter graphiquement sur ce cas les résultats précédents.

Sur la figure suivante, on représente :

- Les couples de réalisations (X, Y) (bleu), leurs répartitions respectives (sous forme d'histogrammes, en bleu) et densités de probabilité (courbes noires).
- Un estimateur de $X|Y$ (rouge), correspondant au maximum de vraisemblance (ou

maximum de probabilité de $X|Y$: $\max\{p(X|Y)\}$.

- Le BLUE de $X|Y$ (vert)
- Les lignes de niveau de répartition à 95, 99 et 99,9% (courbes noires).



On remarque graphiquement qu'en raison du caractère non linéaire de la relation entre X et Y , le BLUE n'est pas aussi pertinent qu'un estimateur non linéaire.

Cependant, autour des valeurs les plus vraisemblables (\bar{X} et \bar{Y}) ses performances devraient être bien meilleures en raison sa proximité de l'estimateur du maximum de vraisemblance.

2.3 BLUE pour un modèle linéaire

On considère à nouveau deux vecteurs aléatoires réels X et Y , de dimension n et p et de matrice de covariance Σ_X et Σ_Y . Ces deux vecteurs aléatoires sont reliés linéairement par :

$$Y = C \cdot X + \beta$$

Où C est une matrice déterministe de taille $p \times n$; β est un vecteur aléatoire centré ($E(\beta) = 0$) de taille p , de covariance Σ_β et orthogonal à X .

D'après la loi de propagation de la variance, les matrices de covariance de X et Y sont reliées par :

$$\Sigma_Y = C\Sigma_X C^T + \Sigma_\beta$$

THÉORÈME Au sens de l'optimalité, le meilleur estimateur linéaire orthogonal sans biais de X connaissant Y s'écrit alors :

$$\hat{X} = \bar{X} + K \cdot (Y - C\bar{X})$$

Avec :

$$K = \Sigma_X C^T \Sigma_Y^{-1}$$

DÉMONSTRATION

L'espérance étant un opérateur linéaire et β étant centré, on a tout de suite :

$$\delta_Y = Y - C\bar{X}$$

Calculons maintenant Σ_{XY} .

$$\begin{aligned}\Sigma_{XY} &= E(\delta_X \delta_Y^T) \\ &= E(\delta_X \delta_X^T C^T + \delta_X \delta_\beta^T) \quad \text{car } \delta_Y = C\delta_X + \delta_\beta \\ &= \Sigma_X C^T \quad \text{car } X \text{ et } \beta \text{ sont orthogonaux}\end{aligned}$$

Donc :

$$\begin{aligned}K &= \Sigma_{XY} \Sigma_Y^{-1} \\ &= \Sigma_X C^T \Sigma_Y^{-1}\end{aligned}$$

THÉORÈME La covariance de l'estimateur linéaire orthogonal sans biais de X est donnée par :

$$\Sigma_{\hat{X}} = \Sigma_X - K C \Sigma_X$$

DÉMONSTRATION

On a vu précédemment pour le BLUE :

$$\Sigma_{\hat{X}} = \Sigma_X - K \Sigma_{XY}^T$$

Ainsi que :

$$\Sigma_{XY} = \Sigma_X C^T$$

En remarquant que $\Sigma_X = \Sigma_X^T$, on retrouve bien la relation recherchée.

La pertinence de la matrice de covariance a posteriori ici estimée est purement formelle et dépend donc l'incertitude a priori sur l'inconnue X et sur le bruit affectant la relation linéaire.

En particulier, si le bruit affectant les mesures est bien différent de celui supposé, l'estimation s'en trouvera affectée, mais la matrice de covariance estimée sera inchangée.

En pratique, on dispose rarement d'une solution initiale avec une covariance a priori pertinente. Dans ce cas, il est usuel de choisir \bar{X} nul et une matrice covariance Σ_X « élevée ».

EXEMPLE

On dispose de 4 mesures d'une même position donnée (planimétrie uniquement), de même précision, environ 50 cm. On cherche à en déduire la position optimale à l'aide du BLUE.

Le problème se formalise sous la forme :

$$Y = C \cdot X + \beta$$

Avec :

$$Y = \begin{pmatrix} x_1 \\ \vdots \\ x_4 \\ y_1 \\ \vdots \\ y_4 \end{pmatrix} \quad C = \begin{pmatrix} 1 & 0 \\ \vdots & \vdots \\ 1 & 0 \\ 0 & 1 \\ \vdots & \vdots \\ 0 & 1 \end{pmatrix} \quad X = \begin{pmatrix} x \\ y \end{pmatrix}$$

Où les (x_i, y_i) désignent les positions mesurées, (x, y) la position recherchée. β est un vecteur aléatoire gaussien centré de variance $0,025 \text{ m}^2$.

On ne connaît pas la position recherchée. On la suppose proche de $(10, 10)$, à plus ou moins 5 m :

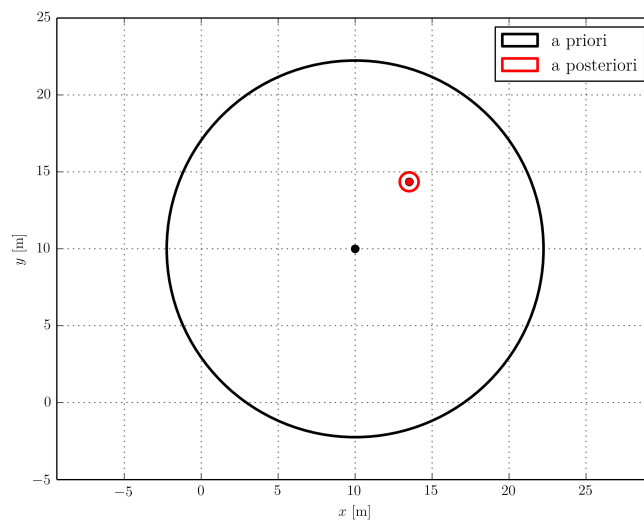
$$\bar{X} = \begin{pmatrix} 10 \\ 10 \end{pmatrix} \quad \Sigma_X = \begin{pmatrix} 25 & 0 \\ 0 & 25 \end{pmatrix}$$

β étant caractérisé, on peut calculer Σ_Y et la matrice K .

On en déduit une estimation de X , \hat{X} , et de sa matrice de covariance a posteriori, $\Sigma_{\hat{X}}$.

$$\hat{X} = \begin{pmatrix} 13,72 \\ 14,51 \end{pmatrix} \quad \Sigma_{\hat{X}} = \begin{pmatrix} 0,062 & 0 \\ 0 & 0,062 \end{pmatrix}$$

Sur la figure suivante on représente les ellipse d'erreur a priori et a posteriori déterminées à partir des matrices de covariance correspondantes.



On remarque bien sûr une forte diminution de l'ellipse d'erreur une fois la position BLUE estimée, les variances passant de 5 m pour la matrice de covariance a priori à environ $0,25 \text{ m}$ pour la matrice de covariance a posteriori.

L'allure circulaire des ellipses de confiance se justifie par le fait que les incertitudes sont identiques sur chacune des composantes.

REMARQUE

À partir de ce formalisme du BLUE dans le cas linéaire, on peut retrouver la solution au problème des moindres carrés abordé en début de semestre 3.

FILTRE DE KALMAN

Sommaire

3.1	Contexte	24
3.2	Équations du filtre de Kalman	24
3.3	Filtre de Kalman étendu	27
3.4	Pour aller plus loin...	30

Nous allons appliquer ici l'estimateur orthogonal sans biais à un système dont l'évolution temporelle du comportement est également connue. Ceci nous permet d'introduire le filtre de Kalman, méthode d'estimation très couramment utilisée.

Les résultats présentés dans ce chapitre sont essentiellement tirés de [Jaulin, 2014; Seube, 2012; Chui et Chen, 1990].

3.1 Contexte

Soit un système décrit par les équations suivantes :

$$\begin{cases} X_{k+1} &= A_k X_k + u_k + \alpha_k & (i) \\ Y_k &= C_k X_k + \beta_k & (ii) \end{cases}$$

X_k et Y_k sont des vecteurs aléatoires de dimension n et p et de matrices de covariance Σ_{X_k} et Σ_{Y_k} respectivement. L'indice k indique un changement d'état de notre système, en fonction du temps par exemple. On parlera ainsi d'**état à l'époque ou à l'instant k** (par abus de langage on pourra même simplifier en état k).

L'équation (i) est appelée **équation d'état** ; elle traduit l'évolution temporelle de notre système entre les instants discrétisés k et $k+1$. A_k est une matrice déterministe de $\mathbb{R}^{n \times n}$ et est appelée **matrice d'état**. u_k est un vecteur déterministe appelé **commande** ou **entrée**.

L'équation (ii) est appelée **équation de mesure** ou **équation d'observation**. C_k est une matrice déterministe de $\mathbb{R}^{p \times n}$ et est appelée **matrice de mesure** ou **matrice d'observation**.

α_k et β_k sont respectivement les **bruits d'état** et **de mesures** ; ce sont des vecteurs aléatoires centrés de dimension n et p et de matrices de covariances Σ_{α_k} et Σ_{β_k} , orthogonaux avec le vecteur X_k . Chacun de ces vecteurs est également orthogonal dans le temps (c'est-à-dire que pour 2 époques k_1 et k_2 , les vecteurs α_{k_1} et α_{k_2} sont orthogonaux).

L'objectif du **filtre de Kalman** est d'estimer de manière optimale à partir d'une donnée initiale, l'évolution temporelle du vecteur X_k en combinant équations d'état et équation de mesures.

Le filtre de Kalman se déroule en deux temps : une **prédiction**, utilisant l'équation d'état de notre système, et une **correction** utilisant l'équation de mesure de notre système.

Le filtre de Kalman a été initialement théorisé par Rudolph Kalman [Kalman, 1960].

3.2 Équations du filtre de Kalman

On se place à un instant $k > 0$. On connaît l'état $k-1$.

Prédiction : la prédiction utilise donc l'équation d'état pour prévoir l'état k connaissant

l'état $k-1$:

$$\begin{aligned} k|k-1 : \\ \hat{X}_{k|k-1} &= A_{k-1} \hat{X}_{k-1|k-1} + u_{k-1} \\ \Sigma_{\hat{X}_{k|k-1}} &= A_{k-1} \Sigma_{\hat{X}_{k-1|k-1}} A_{k-1}^T + \Sigma_{\alpha_{k-1}} \end{aligned}$$

DÉMONSTRATION

La première équation de la prédiction découle simplement de l'équation d'état.

Pour la seconde, on utilise la propagation de la variance et on utilise l'hypothèse selon laquelle \hat{X}_{k-1} et α_{k-1} sont orthogonaux.

Correction : la correction utilise l'équation de mesure pour corriger l'état k obtenu après prévision. Autrement dit, la prévision intervient comme une solution *a priori* à l'état k que l'on cherche à corriger au même état k :

$$\begin{aligned} k|k : \\ \hat{X}_{k|k} &= \hat{X}_{k|k-1} + K_k (Y_k - C_k \hat{X}_{k|k-1}) \\ \Sigma_{\hat{X}_{k|k}} &= (I - K_k C_k) \Sigma_{\hat{X}_{k|k-1}} \\ K_k &= \Sigma_{\hat{X}_{k|k-1}} C_k^T (C_k \Sigma_{\hat{X}_{k|k-1}} C_k^T + \Sigma_{\beta_k})^{-1} \end{aligned}$$

K_k est le **gain de Kalman**.

DÉMONSTRATION

La prévision nous donne une première estimation de l'état à k , connaissant $k-1$, $\hat{X}_{k|k-1}$. On cherche à déterminer l'état à k connaissant cette prévision, $\hat{X}_{k|k}$.

On considère l'équation d'observation (i). On retrouve la formulation d'un modèle linéaire pour lequel on a montré précédemment que l'estimateur linéaire orthogonal sans biais pouvait s'écrire :

$$\hat{X}_{k|k} = \hat{X}_{k|k-1} + K_k (Y_k - C_k \hat{X}_{k|k-1})$$

Sa covariance *a posteriori* $\Sigma_{\hat{X}_{k|k}}$ s'exprime en fonction de la covariance *a priori* $\Sigma_{\hat{X}_{k|k-1}}$:

$$\Sigma_{\hat{X}_{k|k}} = \Sigma_{\hat{X}_{k|k-1}} - K_k C_k \Sigma_{\hat{X}_{k|k-1}} = (I - K_k C_k) \Sigma_{\hat{X}_{k|k-1}}$$

Le gain de Kalman à l'époque k , K_k , s'écrit sous la forme :

$$K_k = \Sigma_{\hat{X}_{k|k-1}} C_k^T \Sigma_{Y_k}^{-1}$$

X_k et β_k étant orthogonaux, la covariance de Y_k s'écrit :

$$\Sigma_{Y_k} = C_k \Sigma_{\hat{X}_{k|k-1}} C_k^T + \Sigma_{\beta_k}$$

D'où :

$$K_k = \Sigma_{\hat{X}_{k|k-1}} C_k^T (C_k \Sigma_{\hat{X}_{k|k-1}} C_k^T + \Sigma_{\beta_k})^{-1}$$

REMARQUE

On rappelle que dans le cas gaussien, la partie correction du filtre de Kalman fournit l'estimateur du maximum de vraisemblance (voir chapitre 2). Dans ce cas, le caractère orthogonal est équivalent au caractère indépendant.

Le filtre de Kalman apparaît comme une solution récursive. En pratique, la phase d'initialisation, peut se faire en choisissant un vecteur $\hat{X}_{0|0}$ nul et une matrice covariance $\Sigma_{\hat{X}_{0|0}}$ « élevée ».

EXEMPLE

On considère le déplacement d'un mobile sur un plan. Au cours de son parcours, on mesure 4 positions de celui-ci, avec une précision de l'ordre de 5 m pour un échantillonnage régulier $\Delta t = 1$ s. On dispose également de la mesure de sa vitesse (sur chacune des composantes planimétriques), avec une précision de l'ordre de $10 \text{ cm}\cdot\text{s}^{-1}$.

L'ensemble des vecteurs aléatoires sont supposés gaussiens.

Ce problème se modélise de la forme suivante :

$$\begin{cases} X_{k+1} &= X_k + u_k + \alpha_k \\ Y_k &= C_k X_k + \beta_k \end{cases}$$

Avec :

$$X_k = \begin{pmatrix} x_k \\ y_k \end{pmatrix} \quad u_k = \begin{pmatrix} v_{x_k} \cdot \Delta t \\ v_{y_k} \cdot \Delta t \end{pmatrix} \quad Y_k = \begin{pmatrix} x_{k,1} \\ \vdots \\ x_{k,4} \\ y_{k,1} \\ \vdots \\ y_{k,4} \end{pmatrix} \quad C_k = \begin{pmatrix} 1 & 0 \\ \vdots & \vdots \\ 1 & 0 \\ 0 & 1 \\ \vdots & \vdots \\ 0 & 1 \end{pmatrix}$$

Où les $(x_{k,i}, y_{k,i})$ désignent les positions mesurées à l'époque k , (x_k, y_k) la position recherchée à l'époque k .

α_k et β_k sont des bruits centrés de variance respectives $0,01 \text{ m}^2$ et 25 m^2 :

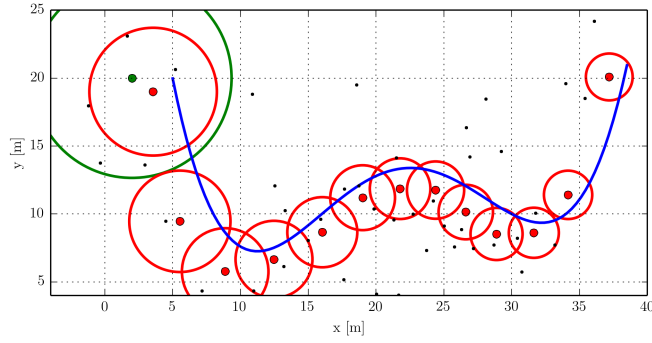
$$\Sigma_{\alpha_k} = \begin{pmatrix} 0,01 & 0 \\ 0 & 0,01 \end{pmatrix} \quad \Sigma_{\beta_k} = \begin{pmatrix} 25 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & 25 \end{pmatrix}$$

On suppose connaître de manière approchée (précision de 3 m) la position initiale du mobile :

$$\bar{X} = \begin{pmatrix} 3 \\ 20 \end{pmatrix} \quad \Sigma_X = \begin{pmatrix} 9 & 0 \\ 0 & 9 \end{pmatrix}$$

Maintenant que le problème est correctement modélisé, on peut appliquer la solution du filtre de Kalman. Le résultat du filtre est présenté sur la figure suivante :

- En noir les positions mesurées,
- En vert la solution initiale,
- En rouge les positions estimées par le filtre,
- En bleu la trajectoire attendue.



Malgré la dispersion des mesures, le filtre converge rapidement vers la trajectoire de référence. Par l'action combinée du modèle d'état et de mesure, les ellipses d'erreur diminuent au fur et à mesure de la progression du mobile.

Les ellipses de confiance conservent un aspect circulaire au cours du déplacement. Ceci est expliqué par le fait que l'on dispose de mesures de position, quantité que l'on cherche à optimiser, dont les variances sont identiques pour les 2 composantes (et de covariance nulle).

3.3 Filtre de Kalman étendu

On se place désormais dans le cas suivant :

$$\begin{cases} X_{k+1} = f_k(X_k) + \alpha_k & (i) \\ Y_k = g_k(X_k) + \beta_k & (ii) \end{cases}$$

Où f_k et g_k sont respectivement des fonctions de \mathbb{R}^n dans \mathbb{R}^n et \mathbb{R}^n dans \mathbb{R}^p .

Ici, les équation d'état (i) et d'observation (ii) ne sont plus linéaires mais dépendent de fonctions vectorielles f_k et g_k . La fonction f_k inclue une éventuelle commande. Les autres paramètres présentent les mêmes propriétés que précédemment.

Le **filtre de Kalman étendu** (noté EKF pour *Extended Kalman filter*) permet la résolution de ce problème non linéaire. Il se déroule toujours en deux temps : une **prédiction** et une **correction**.

Prédiction : la prédiction utilise donc l'équation d'état pour prévoir l'état k connaissant l'état $k-1$:

$$\begin{aligned} k|k-1 : \\ \hat{X}_{k|k-1} &= f_{k-1}(\hat{X}_{k-1|k-1}) \\ \Sigma_{\hat{X}_{k|k-1}} &= A_{k-1} \Sigma_{\hat{X}_{k-1|k-1}} A_{k-1}^T + \Sigma_{\alpha_{k-1}} \end{aligned}$$

Où A_k désigne la matrice jacobienne de la fonction f_k au point $\hat{X}_{k|k}$:

$$A_k = \left[\frac{\partial f_k}{\partial X_k}(\hat{X}_{k|k}) \right]$$

Correction : la correction utilise l'équation de mesure pour corriger l'état k obtenu après

prévision. Autrement dit, la prévision intervient comme une solution *a priori* que l'on cherche à corriger :

$$\begin{aligned}
 k|k : \\
 \hat{X}_{k|k} &= \hat{X}_{k|k-1} + K_k [Y_k - g_k(\hat{X}_{k|k-1})] \\
 \Sigma_{\hat{X}_{k|k}} &= (I - K_k C_k) \Sigma_{\hat{X}_{k|k-1}} \\
 K_k &= \Sigma_{\hat{X}_{k|k-1}} C_k^T (C_k \Sigma_{\hat{X}_{k|k-1}} C_k^T + \Sigma_{\beta_k})^{-1}
 \end{aligned}$$

Où C_k désigne la matrice jacobienne de la fonction g_k au point $\hat{X}_{k|k-1}$:

$$C_k = \left[\frac{\partial g_k}{\partial X_k} (\hat{X}_{k|k-1}) \right]$$

DÉMONSTRATION

Les équations du filtre de Kalman étendu dérivent du développement de Taylor au premier ordre des fonctions f_k et g_k au estimés. Au voisinage d'une solution \hat{X}_k on peut en effet écrire :

$$\begin{cases} f_k(X_k) &= f_k(\hat{X}_k) + A_k \cdot (X_k - \hat{X}_k) \\ g_k(X_k) &= g_k(\hat{X}_k) + C_k \cdot (X_k - \hat{X}_k) \end{cases}$$

Le problème peut ainsi être reformulé en :

$$\begin{cases} X_{k+1} &= A_k X_k + u_k + \alpha_k \\ Z_k &= C_k X_k + \beta_k \end{cases}$$

Avec :

$$\begin{cases} u_k &= f_k(\hat{X}_k) - A_k \hat{X}_k \\ Z_k &= Y_k - (g_k(\hat{X}_k) - C_k \hat{X}_k) \end{cases}$$

On dispose d'un problème linéaire que l'on sait résoudre avec le filtre de Kalman classique. On applique donc les formules vues précédemment.

L'équation de prédiction devient :

$$\begin{aligned}
 \hat{X}_{k|k-1} &= A_k \hat{X}_{k-1|k-1} + u_{k-1} \\
 &= A_k \hat{X}_{k-1|k-1} + f_k(\hat{X}_{k-1|k-1}) - A_k \hat{X}_{k-1|k-1} \\
 &= f_k(\hat{X}_{k-1|k-1})
 \end{aligned}$$

La formule de propagation de la variance nous donne :

$$\Sigma_{\hat{X}_{k|k-1}} = A_k \Sigma_{\hat{X}_{k-1|k-1}} A_k^T + \Gamma_{\alpha_k}$$

La correction est donnée par :

$$\begin{aligned}
 \hat{X}_{k|k} &= \hat{X}_{k|k-1} + K_k [Z_k - C_k \hat{X}_{k|k-1}] \\
 &= \hat{X}_{k|k-1} + K_k [Y_k - (g_k(\hat{X}_k) - C_k \hat{X}_k) - C_k \hat{X}_{k|k-1}] \\
 &= \hat{X}_{k|k-1} + K_k [Y_k - g_k(\hat{X}_{k|k-1})]
 \end{aligned}$$

Avec :

$$K_k = \Sigma_{\hat{X}_{k|k-1}} C_k^T (C_k \Sigma_{\hat{X}_{k|k-1}} C_k^T + \Sigma_{\beta_k})^{-1}$$

Z_k étant déterministe, la covariance a posteriori découle également des équations de Kalman :

$$\Sigma_{\hat{X}_{k|k}} = (I - K_k C_k) \Sigma_{\hat{X}_{k|k-1}}$$

EXEMPLE

On considère à nouveau le déplacement d'un mobile sur un plan. Au cours de son parcours, on mesure les distances à 4 balises fixes de positions connues, avec une précision de l'ordre de 2 m. On dispose également de la mesure de sa vitesse (sur chacune des composantes planimétriques), avec une précision de l'ordre de 10 cm·s⁻¹.

Le problème est toujours linéaire selon l'équation d'état mais devient non linéaire selon l'équation d'observation. Il se modélise de la forme suivante :

$$\begin{cases} X_{k+1} &= X_k + \alpha_k \\ Y_k &= g_k(X_k) + \beta_k \end{cases}$$

Avec :

$$X_k = \begin{pmatrix} x_k \\ y_k \end{pmatrix} \quad u_k = \begin{pmatrix} v_{x_k} \cdot \Delta t_k \\ v_{y_k} \cdot \Delta t_k \end{pmatrix} \quad Y_k = \begin{pmatrix} d_{k,1} \\ \vdots \\ d_{k,4} \end{pmatrix} \quad g_k(X_k) = \begin{pmatrix} \|X_k - X_1^b\| \\ \vdots \\ \|X_k - X_4^b\| \end{pmatrix}$$

Où les $d_{k,i}$ désigne la distance mesurée à l'époque k en direction de la balise i de coordonnées $X_i^b = (x_i^b, y_i^b)^T$, $(x_k, y_k)^T$ la position recherchée à l'époque k .

La jacobienne de la fonction g_k en \hat{X}_k s'écrit :

$$C_k = \begin{pmatrix} \frac{\hat{x}_k - x_1^b}{\|X_k - X_1^b\|} & \frac{\hat{y}_k - y_1^b}{\|X_k - X_1^b\|} \\ \vdots & \vdots \\ \frac{\hat{x}_k - x_4^b}{\|X_k - X_4^b\|} & \frac{\hat{y}_k - y_4^b}{\|X_k - X_4^b\|} \end{pmatrix}$$

α_k et β_k sont des bruits centrés de variance respectives 0,01 m² et 4 m² :

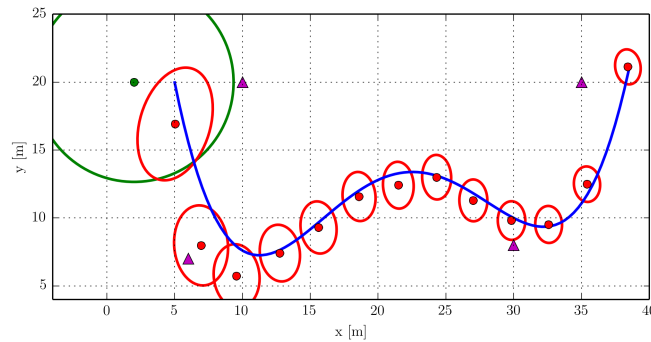
$$\Sigma_{\alpha_k} = \begin{pmatrix} 0,01 & 0 \\ 0 & 0,01 \end{pmatrix} \quad \Sigma_{\beta_k} = \begin{pmatrix} 4 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & 4 \end{pmatrix}$$

On suppose connaître de manière approchée (précision de 3 m) la position initiale du mobile :

$$\bar{X} = \begin{pmatrix} 2 \\ 30 \end{pmatrix} \quad \Sigma_X = \begin{pmatrix} 9 & 0 \\ 0 & 9 \end{pmatrix}$$

Maintenant que le problème est correctement modélisé, on peut appliquer la solution du filtre de Kalman. Le résultat du filtre est présenté sur la figure suivante :

- En magenta les balises de coordonnées connues,
- En vert la solution initiale,
- En rouge les positions estimées par le filtre,
- En bleu la trajectoire attendue.



Le filtre converge rapidement vers la trajectoire de référence. Par l'action combinée du modèle d'état et des mesures, les ellipses d'erreur diminuent au fur et à mesure de la progression du mobile.

Ici les ellipses d'erreur n'ont plus d'allure circulaire car le bruit n'affecte plus les composantes de X_k de manière symétrique et dépend de la configuration des mesures de distances (entre mobile et balises).

3.4 Pour aller plus loin...

D'autres outils d'optimisation sont dérivés du filtre de Kalman présenté jusqu'ici. Citons par exemple :

- Le **lisseur de Kalman** (*Kalman smoother*) : On dispose désormais de l'ensemble de l'information jusqu'à l'époque « finale » N . On souhaite alors déterminer de manière optimale (au sens du minimum de variance) l'ensemble des états X_k connaissant toutes les données disponibles, c'est à dire les données antérieures et postérieures à l'époque k .

La solution à ce problème consiste à lancer le problème dans le sens rétrograde et à combiner à chaque époque la solution déterminée dans le sens classique (ou vers l'avant, *forward*) et rétrograde (ou vers l'arrière, *backward*).

Pour plus de détails, on peut se référer à [Jaulin, 2014; Chui et Chen, 1990].

- Le **filtre de Kalman d'ensemble**, (EnKF, pour *Ensemble Kalman Filter*). Aussi bien pour des raisons théoriques que pour des raisons de coût de calcul, le filtre de Kalman étendu peut être inadapté pour certaines applications géophysiques, comme l'intégration de mesures dans des modèles météorologiques ou hydrodynamiques (assimilation). Pour cette variante, l'équation d'état est non-linéaire. L'approche est alors d'estimer les paramètres décrivant les erreurs du modèle.

Cette méthode procède en générant plusieurs conditions initiales, qui vont former les membres de l'ensemble. Le modèle d'évolution (et son modèle d'erreur) est utilisé pour simuler l'évolution temporelle des membres. Dès que des observations sont disponibles, chaque membre est corrigé en utilisant les équations du filtre de Kalman. Les paramètres stochastiques (moyennes et covariances) utilisés dans le filtre de Kalman sont alors estimés en utilisant les membres de l'ensemble (méthode de Monte-Carlo).

La figure 3.1 décrit le principe de fonctionnement de la méthode de Kalman d'ensemble. A chaque instant l'état du système est la moyenne des membres de l'ensemble.

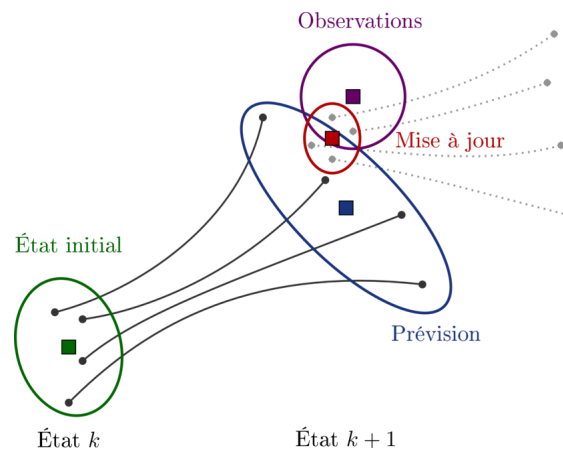


FIGURE 3.1 – Principe du filtre Kalman d'ensemble. Les cercles indiquent les évolutions de chaque membre. Les carrés indiquent les valeurs moyennes.

Pour plus de détails, on peut se référer à [Evensen, 1994, 2009].

Bibliographie

- CHUI, C. K. et CHEN (1990). *Kalman filtering with real-time applications*. Springer-Verlag.
- EVENSEN, G. (1994). Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. *Journal of Geophysical Research*, 99:10143 – 10162.
- EVENSEN, G. (2009). *Data Assimilation, The Ensemble Kalman Filter*. Springer.
- JAULIN, L. (2014). Robotique mobile. Notes de cours - Cycle Ingénieur de l'ENSTA Bretagne.
- KALMAN, R. (1960). A new approach to linear filtering and prediction problems. *Transactions of the ASME-Journal of Basic Engineering*, 82:35–45.
- OSSWALD, C. (2015). Probabilités et statistiques. Notes de cours - Cycle Ingénieur de l'ENSTA Bretagne.
- SEUBE, N. (2012). Inertial navigation and optimal estimation. Notes de cours - Cycle Ingénieur de l'ENSTA Bretagne.