ENSICA - Deuxième année **1999-2000**

Introduction à la théorie de l'estimation

Notes de cours
Version 1.2

D. Arzelier

Avertissement : Ce document est constitué de notes de cours et ne prétend donc ni à l'exhaustivité ni à l'originalité. Ces notes doivent en effet beaucoup aux emprunts faits aux ouvrages référencés en bibliographie.

Notations

- $\mathbb{R}$: corps des nombres réels.
- A': matrice transposée de la matrice A.
- $A > 0$: A matrice définie positive.
- $\| \bullet \|$: norme Euclidienne pour un vecteur et induite par la norme Euclidienne pour une matrice.
- $P[\bullet]$: probabilité simple.
- $P[\bullet/\bullet]$ probabilité conditionnelle.
- V.A. : vecteur aléatoire.
- $E[\bullet]$: espérance mathématique.
- $E[x/y]$: espérance conditionnelle de x sachant y.
- $p_x(\alpha)$: densité de probabilité de la v.a. x.
- $p_{x_1,\cdots,x_n}(\alpha_1,\cdots,\alpha_n)$: densité de probabilité conjointe.
- $p_{x/y}(\alpha/\beta)$: densité de probabilité conditionnelle.
- $Z \sim \mathcal{N}(m_Z, Q_Z)$: vecteur gaussien de moyenne m_Z et de matrice de covariance Q_Z.
- $R(t,\tau)$: matrice d'autocorrélation.
- $P(t,\tau)$: matrice d'autocovariance.
- δ_{kl}: symbole de Kronecker.
- $\delta(t)$: impulsion de Dirac.
- $\mathbf{1}_n$: matrice identité de dimension n.
- $Arg[\bullet]$: argument du problème d'optimisation.
- $\hat{x}_{k/k}$: estimé optimal à l'instant k sachant k mesures.
- $P_{k/k}$: matrice de covariance à l'instant k sachant k mesures.

Table des matières

Chapitre I

La théorie de l'estimation

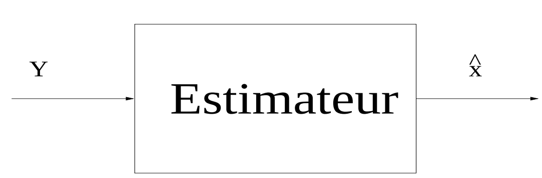
I.1 Introduction

Ces notes de cours constituent une introduction à la théorie de l'estimation. Par théorie de l'estimation, nous entendons l'ensemble des méthodes permettant de calculer "un estimé optimal" à partir de données bruitées. Les origines véritables de la théorie de l'estimation remontent aux travaux de C.F. Gauss, (1795) et A.V. Legendre, (1810), qui développèrent presque simultanément **la méthode des moindres carrés** et l'appliquèrent à l'estimation des paramètres de l'orbite de la planète Cérès. La méthode des moindres carrés fut ainsi la première méthode d'estimation optimale fournissant de plus un lien entre sciences théoriques et sciences expérimentales: elle a donné aux praticiens une méthode pratique afin d'estimer les paramètres inconnus de modèles théoriques.
Historiquement, depuis les travaux de Gauss, les avancées majeures ont été faites durant les 60 premières années de ce siècle. Peu de progrès furent faits après Gauss et Legendre jusqu'à ce que R.A. Fischer, (1912), introduise l'approche du **maximum de vraisemblance** qui fut principalement utilisée par les statisticiens.
A partir de la fin des années trente, les progrès effectués en **théorie des probabilités et des processus stochastiques**, avec les travaux de A.N. Kolmogorov, permirent de considérer la théorie de l'estimation sous un angle nouveau. Kolmogorov réétablit ainsi les fondations de **la théorie de la mesure**, ce qui reste encore de nos jours la base mathématique de la théorie des probabilités et des processus stochastiques. Kolmogorov et N. Wiener sont tous deux crédités de la naissance de la théorie de la prédiction, du filtrage des processus de Markov et de la théorie des processus ergodiques. Ce dernier, dans les années 40, résolut le problème du filtrage linéaire continu au sens des moindres carrés pour les processus stochastiques stationnaires, donnant naissance au **filtre de Wiener**. Il utilisa, pour se faire, les fonctions de corrélation et de densité spectrale conduisant à la nécessité de la résolution des équations intégrales de Wiener-Hopf afin de réaliser l'estimateur. Ces résultats furent étendus durant les années 50-60 à travers des techniques temporelles

et fréquentielles, avec une prédilection pour ces dernières. Les travaux dont l'orientation fut donnée par les résultats de Wiener ont ainsi clairement dominé la période 1940-1960 jusqu'à l'apparition des travaux de R. E. Kalman qui ont dominé la période 1960-1975. Ils permirent de donner une solution pratique à de nombreux problèmes techniques et particulièrement dans le domaine aérospatial où ils sont appliqués pour la détermination d'orbites ou pour la navigation. Spécifiquement, Kalman utilise les techniques d'espace d'état afin d'obtenir des algorithmes numériquement efficaces, (décrivant **le filtre de Kalman**), mais également les conditions techniques, commandabilité-observabilité, pour lesquelles le filtre optimal est stable et robuste. Ces résultats ont conduit à une somme considérable de travaux complémentaires dans les années 60-70.
Alors que les problèmes d'estimation linéaire recevaient une attention soutenue et des réponses plus ou moins complètes, dans le même temps, des études sur les problèmes d'estimation non linéaires étaient lancées. Il était en effet nécessaire de disposer de méthodes d'estimation toujours plus sophistiquées afin de résoudre des problèmes pratiques toujours plus difficiles. Les procédures d'estimation linéaire ne conduisent effectivement pas à des résultats satisfaisants, (divergence des algorithmes), quand elles sont appliquées à des modèles de systèmes intrinsèquement non linéaires. On peut citer comme principaux acteurs de cette évolution, R.L. Stratonovich, H.J. Kushner, R. Bellman, T. Kailath ... Les méthodes développées dans ce cadre sont largement au delà des limites que nous nous sommes fixés dans ce cours et n'y seront donc pas présentées.
Après un chapitre introductif sur la théorie de l'estimation au sens des moindres carrés et la présentation de la méthode des moindres carrés, nous n'aborderons que le problème du filtrage linéaire sous les hypothèses gaussiennes et par une approche probabiliste dans le cadre défini par Kalman. Finalement, le problème de la commande optimale des systèmes linéaires bruités sera abordé et **le principe de séparation** donnant lieu à **la commande LQG** sera énoncé.

I.2 Le problème d'estimation

La théorie de l'estimation est un champ thématique de la théorie des statistiques qui regroupe les techniques permettant l'estimation de la valeur d'un ensemble de paramètres inconnus, regroupés dans un vecteur x, à partir d'un ensemble d'observations, regroupées dans un vecteur Y, contenant de l'information sur x. Un estimé $\hat{x}$ est déterminé à partir de Y en utilisant une loi d'estimation appelée **estimateur**, qui peut être optimal au sens de la minimisation d'un certain critère. Il est à noter qu'un estimateur est une fonction $e(Y)$



des observations. Afin de décider de la valeur relative des estimateurs possibles pour un problème d'estimation donné, on associe usuellement un problème d'optimisation pour lequel une fonction coût $\mathcal{L}(x, \hat{x})$ doit être définie. Un exemple classique est celui de la fonction coût quadratique de l'erreur, donnée par $\|x - \hat{x}\|$. Une fois le choix de la fonction coût effectué, il est nécessaire de définir un critère d'optimisation. Le choix a priori du critère définissant l'optimalité de l'estimé est difficile et peut faire partie intégrante du problème d'optimisation associé au problème d'estimation. Dans un cadre probabiliste, le critère défini comme l'espérance mathématique de la fonction coût, appelé **critère de Bayes**, définit une classe de problèmes d'estimation désormais classique.

Déf. I.2.1 *: estimateur de Bayes*
Un estimateur optimal au sens du critère de Bayes est appelé **estimateur de Bayes**.

Déf. I.2.2 *: estimateurs du minimum de variance de l'erreur*
Un estimateur de Bayes opérant sur des fonctions coût quadratiques de l'erreur sont appelés **estimateurs du minimum de variance de l'erreur**.

Le processus d'estimation est généralement exécuté par un système appelé filtre et ceci en temps réel. Cela signifie que le filtre d'estimation agit sur un ensemble d'entrées $y(t)$, contenant l'information sur un certain signal inconnu x, et fournit un estimé $\hat{x}(t+\alpha|t)$ pour la valeur de x en $t+\alpha$.

Déf. I.2.3 *:*
Le filtre d'estimation est appelé

- $\alpha > 0$: **filtre prédicteur.**

- $\alpha < 0$: **filtre de lissage.**

- $\alpha = 0$: **filtre.**

En temps que système, il peut être caractérisé par un certain nombre de propriétés liées à la nature de la relation entrées-sorties: linéarité, invariance vis à vis du temps... . Il est souvent intéressant, dans les problèmes d'estimation, de restreindre la classe des estimateurs recherchés à une classe prédéfinie telle que, par exemple, la classe des filtres estimateurs linéaires. Le problème d'estimation peut alors s'écrire:

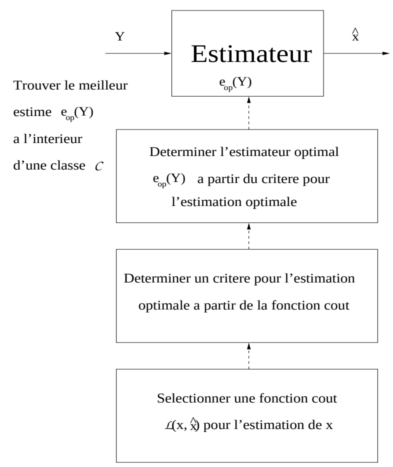
$$\min_{e(Y) \in \mathcal{C}} \quad Cr(\mathcal{L}(x, e(Y))$$

Schématiquement, un problème d'estimation peut donc être défini de la façon suivante: Les éléments à définir dans un problème d'estimation sont les suivants:

- Un vecteur de paramètre inconnu x appartenant à un espace des paramètres X, dont la valeur est à estimer.

- Un vecteur d'observation Y appartenant à un espace des observations $\mathcal{Y}$, contenant de l'information sur x. Y peut être un vecteur de paramètres, un vecteur de signaux aléatoires en temps continu ou en temps discret.

- Une loi d'estimation, (un estimateur), $e(Y)$ qui fournit un estimé de x à partir des observations Y. Un estimateur est une fonction de l'ensemble des observations $\mathcal{Y}$ à valeurs dans l'ensemble des paramètres X. Il fait correspondre à tout vecteur Y un vecteur estimé $\hat{x}$.



- Un vecteur estimé $\hat{x}$, sortie du filtre estimateur $e(Y)$.
- Une fonction coût $\mathcal{L}(x, \hat{x})$ qui fournit le coût associé à la différence entre l'estimé $\hat{x}$ et x.
- Un critère pour l'estimation optimale, utilisé conjointement avec la fonction coût afin de déterminer un estimateur optimal.

I.3 Estimation linéaire optimale au sens du minimum de variance

Ainsi que nous l'avons présenté dans la section précédente, un estimateur linéaire optimal au sens du minimum de variance est un estimateur optimal de Bayes, dont la fonction coût associée est une fonction quadratique de l'erreur et à choisir dans la classe des estimateurs linéaires. Chaque point de cette définition est maintenant précisé.
En théorie de l'estimation Bayesienne, le vecteur des paramètres x à estimer et le vecteur des observations Y sont considérés comme des vecteurs aléatoires décrits par une fonction densité de probabilité conjointe $p_{x,Y}(\alpha, \beta)$ et le critère choisi est du type $E[\mathcal{L}(x, e(Y))]$. Le choix de l'estimateur est donc guidé par le choix de la fonction coût. Un estimateur de Bayes $e(Y)$ au sens du minimum de variance est un estimateur qui minimise la fonction:

$$E[\|x - e(Y)\|^2] = E[(x - e(Y))'(x - e(Y))]$$

Nous rappellons que la matrice de covariance d'un estimateur $e(Y)$ est définie par $E[(x - e(Y)(x - e(Y))']$ et que nous avons la relation suivante:

$$E[\|x - e(Y)\|^2] = Trace(E[(x - e(Y))(x - e(Y))'])$$

Le problème d'estimation que l'on se pose dans cette section est donc le suivant:

$$\min_{e(Y) \in \mathcal{C}} \quad E[\|x - e(Y)\|^2]$$

On peut alors montrer que ce problème d'estimation possède une solution simple donnée dans le théorème suivant:

Théorème I.3.1 *:*
L'estimateur optimal Bayesien au sens du minimum de variance est la moyenne conditionnelle.

$$e_{opt.}(Y) = E[x/Y]$$

Ce résultat, très fort, pose toutefois le problème du calcul de l'estimé optimal qui n'est pas toujours possible à partir de la formule $E[x/Y]$. On ne peut pas toujours obtenir une formule analytique pour l'estimé optimal. Dans les applications pratiques, des estimés de x ne sont pas déterminés à partir de cette formulation mais en recherchant l'estimateur dans une classe précise définissant des contraintes sur l'estimateur à choisir. La linéarité mais aussi le fait qu'il soit sans biais sont deux caractéristiques primordiales définissant une classe importante d'estimateurs.

Déf. I.3.1 *:*
Un estimateur est **un estimateur linéaire** *si la fonction $e(Y)$ est une fonction linéaire.*

Déf. I.3.2 *:*
Un estimateur est **sans biais** *si l'espérance mathématique de l'estimé $\hat{x}$ est égale à l'espérance mathématique de x:*

$$E[e(Y)] = E[x]$$

On recherche donc l'estimateur optimal de Bayes au sens des moindres carrés dans la classe des estimateurs de la forme:

$$e(Y) = a + BY$$

Si le vecteur à estimé $x \in \mathbb{R}^n$ et le vecteur des mesures $Y \in \mathbb{R}^m$ alors les paramètres $(a, B) \in \mathbb{R}^n \times \mathbb{R}^{n\times m}$. Le problème d'estimation se réduit alors au problème de détermination des paramètres a et B.

$$\min_{e(Y) \in \mathcal{C}_{lin.}} E[\|x - e(Y)\|^2] \equiv \min_{e(Y) = a + BY} E[\|x - e(Y)\|^2]$$

L'étude des relations entre l'estimateur optimal de Bayes au sens du minimum de variance. (noté MV), et l'estimateur linéaire optimal de Bayes au sens du minimum de variance, (noté LMV), montre que l'estimateur LMV fournit généralement une fonction coût optimale plus élevée que l'estimateur MV. Toutefois, du fait de sa simplicité, il est plus souvent utilisé dans les problèmes pratiques d'estimation optimale. D'autant plus qu'un résultat important permet d'établir que sous des hypothèses supplémentaires, ces deux estimateurs sont identiques.

Théorème I.3.2 *:*
Sous l'hypothèse que les vecteurs aléatoires X et Y sont conjointement normalement distribués alors l'estimateur LMV est égal à la moyenne conditionnelle $E[x/Y]$.

$$e(Y) = a + BY = E[x/Y]$$

I.4 La méthode des moindres carrés et moindres carrés récursifs

Tous les développements précédents ont été faits dans un cadre Bayesien ou probabiliste. Dans cette section, nous étudions un problème d'estimation paramétrique sans hypothèses probabilistes. Une solution à ce problème est donnée à travers la méthode des moindres carrés. Cette méthode très féconde est un outil servant à l'identification des systèmes mais également un outil standard pour les méthodes de régression linéaire en statistique mathématique. Cette méthode fournit une solution dans la recherche de paramètres minimisant une fonction quadratique de l'erreur. Sa puissance réside d'une part dans la formulation quadratique du critère mais également dans la forme linéaire de la relation à identifier. Le vecteur de paramètres recherché s'exprime alors comme la solution d'un système d'équations linéaires. Hors du cadre probabiliste défini dans la section précédente, le choix d'une fonction quadratique peut sembler arbitraire. Outre le fait que l'on peut interpréter ce critère quadratique comme l'énergie de l'erreur, nous verrons dans cette section une interprétation probabiliste de la méthode des moindres carrés faisant clairement apparaitre le lien avec les développements précédents concernant la théorie de l'estimation mais également avec la théorie du filtre de Kalman développée ultérieurement dans le prochain chapitre.
Nous présenterons la forme globale de la méthode des moindres carrés mais également la forme récursive aussi bien dans le cas monovariable, (estimation d'un paramètre unique) que dans le cas multivariable, (estimation d'un vecteur de paramètres inconnus).

I.4.1 Cas monovariable

Soit à estimer un paramètre a intervenant dans une relation linéaire liant deux vecteurs de même dimension.

$$y = ax$$

Cette estimation est faite à partir des mesures y_i correspondant aux points x_i. Ces mesures peuvent être bruitées, (approche probabiliste) ou entachées d'erreur. Cela est modélisé par l'équation de mesure:

$$y_i = ax_i + b_i$$

où b_i est le bruit affectant chaque mesure y_i. La méthode des moindres carrés consiste à déterminer l'estimé du paramètre a, noté $\overline{a}$ qui minimise un critère quadratique défini par la somme des carrés des erreurs b_i sur l'ensemble des mesures.

$$\sum_{i=1}^{n}(y_i - ax_i) = J$$

Soit n le nombre total de mesures. Le problème d'estimation se pose comme le problème d'optimisation:

$$\boxed{\overline{a} = arg\left[\min_a(\sum_{i=1}^{n}(y_i - ax_i)^2)\right]}$$

Dans ce problème d'optimisation, la fonction coût possède une propriété importante de stricte convexité par rapport à a. Cela signifie que le problème d'optimisation associé possède un minimum globale unique et que l'on dispose de conditions nécessaires et suffisantes d'optimalité caractérisant le paramètre optimal.

Théorème I.4.1 *:*
Le paramètre $\overline{a}$ est optimal au sens de la minimisation du critère J si et seulement si

$$[\frac{\partial J}{\partial a}]_{|a=\overline{a}} = 0$$

La condition nécessaire et suffisante d'optimalité précédente conduit immédiatement au calcul de l'estimé optimal.

Théorème I.4.2 *:*
L'estimé optimal au sens des moindres carrés pour un n mesures est donné par:

$$\boxed{\overline{a} = \frac{\sum_{i=1}^{n} y_i x_i}{\sum_{i=1}^{n} x_i^2}}$$

Ce calcul est peu adapté au cas souvent réaliste en pratique où les mesures sont effectuées répétitivement, conduisant à une croissance infinie de l'ensemble de mesures, $(n \to \infty)$. Dans ce cas, une solution récursive est préférable. La récursivité consiste à trouver une relation de récurrence liant $\overline{a}_{n+1}$, l'estimé optimal avec $n+1$ mesures à $\overline{a}_n$ estimé optimal avec n mesures. le principe consiste à poser:

$$\overline{a}_{n+1} = P_{n+1}.Q_{n+1}$$

avec:

$$\left\{ \begin{array}{l} P_{n+1} = 1/\sum_{i=1}^{n+1} x_i^2 \\ Q_{n+1} = \sum_{i=1}^{n+1} x_i y_i \end{array} \right.$$

or

$$P_{n+1}^{-1} = \sum_{i=1}^{n+1} x_i^2 = P_n^{-1} + x_{n+1}^2$$

$$Q_{n+1} = Q_n + x_{n+1} y_{n+1}$$

soit

$$P_n = P_{n+1} + P_{n+1} x_{n+1}^2 P_n$$

$$P_{n+1} = P_n(1 + x_{n+1}^2 P_n)^{-1} = P_n - x_{n+1}^2 P_n^2 (1 + x_{n+1}^2 P_n)^{-1}$$

donc

$$\overline{a}_{n+1} = [P_n - x_{n+1}^2 P_n^2 (1 + x_{n+1}^2 P_n)^{-1}][Q_n + x_{n+1} y_{n+1}]$$

$$\overline{a}_{n+1} = \overline{a}_n + \frac{P_n x_{n+1} y_{n+1} - x_{n+1}^2 P_n \overline{a}_n}{1 + x_{n+1}^2 P_n}$$

On obtient donc les formules de récurrence:

$$\boxed{\begin{aligned} &\overline{a}_{n+1} = \overline{a}_n + K_{n+1}(y_{n+1} - \overline{a}_n x_{n+1}) \\ &K_{n+1} = P_n x_{n+1} (1 + x_{n+1}^2 P_n)^{-1} \\ &P_{n+1} = P_n - K_{n+1} x_{n+1} P_n \end{aligned}}$$

L'estimé à l'instant $(n+1)$ est égal à l'estimé à l'instant n plus une correction qui dépend de l'erreur faite à l'estimation précédente $[y_{n+1} - \overline{a}_n x_{n+1}]$, et d'un gain K_{n+1} qui peut également s'exprimer comme $K_{n+1} = P_{n+1} x_{n+1}$.
Ainsi, la correction d'estimation à chaque pas s'écrit:

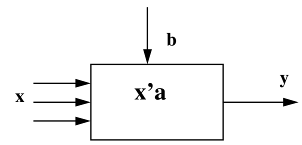
$$P_{n+1} x_{n+1} (y_{n+1} - \overline{a}_n x_{n+1})$$

Dans cette formule, le terme $y_{n+1} - \overline{a}_n x_{n+1}$ est l'information apportée par la mesure y_{n+1} alors que P_{n+1} est la pondération associée à $x_{n+1}(y_{n+1} - \overline{a}_n x_{n+1})$ valeur instantannée du gradient du critère à l'instant $n+1$. Cette pondération est une fonction des estimations précédentes et décroit continument quand le nombre de mesures n augmente. Il est clair que plus il est fait d'estimations, meilleures elles sont et moins l'on attachera d'importance aux écarts entre mesures et prédictions. Les écarts ne sont alors plus considérés que comme des bruits de mesure.
L'algorithme récursif nécessite, *a priori*, une initialisation. Elle peut être faite à partir de n_0 observations utilisant la formulation globale. Mais il est possible de montrer que si l'on choisit une condition initiale pour P_n suffisamment grande, alors l'estimateur converge vers une valeur qui est indépendante de la condition initiale choisie. Ceci se comprend aisément par le fait qu'avec une forte condition initiale pour P_n un grand nombre de corrections significatives seront faites pour l'estimateur, (prise en compte d'un grand nombre de mesures), alors qu'une faible condition initiale pour P_n traduit le fait que l'algorithme a presque atteint sa valeur limite.
Noter que l'évolution de P_n ne dépend pas du bruit de mesure.

I.4.2 Cas multivariable

Dans ce cas, le but est d'estimer un vecteur de paramètres a. L'équation de mesure s'écrit alors:

$$y_i = x_i' a + b_i$$



b et y scalaires

où a et x_i sont respectivement le vecteur des paramètres à estimer et le vecteur des entrées à l'instant i selon le schéma-bloc:

Le problème reste celui de trouver le meilleur estimé du vecteur paramétrique a au sens d'un critère quadratique portant sur l'erreur d'estimation:

$$\boxed{\overline{a} = arg\left[\min_a (J = \sum_{i=1}^{n} (y_i - x_i' a)^2)\right]}$$

On applique de nouveau la condition nécessaire et suffisante d'optimalité. Le gradient de J par rapport à a s'écrit:

$$2\left\{\left[\sum_{i=1}^{n} x_i x_i'\right] a - \sum_{i=1}^{n} y_i x_i\right\}$$

d'où la solution dans le cas où la matrice $\left[\sum_{i=1}^{n} x_i x_i'\right]$ est inversible, ce qui est équivalent à l'hypothèse de convexité stricte de J par rapport à a.

Théorème I.4.3 *:*

Le vecteur estimé optimal au sens des moindres carrés est donné par:

$$\boxed{\overline{a} = \left[\sum_{i=1}^{n} x_i x_i'\right]^{-1} \sum_{i=1}^{n} y_i x_i}$$

En notation matricielle où l'on définit:

- *Y_n le vecteur $[y_1\ y_2 \cdots y_n]'$*
- *X_n la matrice* $\begin{bmatrix} x_1' \\ x_2' \\ \vdots \\ x_n' \end{bmatrix}$
- *B_n le vecteur bruit $[b_1\ b_2 \cdots b_n]'$*

Soit une équation de mesures pour n mesures:

$$Y_n = X_n a + B_n$$

et le critère devient:

$$J_n = (Y_n - X_n a)'(Y_n - X_n a)$$

$X_n' X_n$ étant supposée inversible, le vecteur estimé optimal est donné par:

$$\boxed{\overline{a}_n = [X_n' X_n]^{-1} X_n' Y_n}$$

Comme précédemment, il est intéressant d'établir une formulation récursive, relation de récurrence liant l'estimé optimal $\overline{a}_{n+1}$ obtenu avec $n+1$ mesures à celui obtenu avec n mesures, $\overline{a}_n$. On pose de nouveau:

$$\overline{a}_{n+1} = P_{n+1} Q_{n+1}$$

où:

$$\begin{cases} P_{n+1}^{-1} = X'_{n+1} X_{n+1} = \displaystyle\sum_{i=1}^{n+1} x_i x'_i = P_n^{-1} + x_{n+1} x'_{n+1} \\ Q_{n+1} = Q_n + y_{n+1} x_{n+1} \end{cases}$$

Utilisant le lemme d'inversion matriciel, il vient

$$P_{n+1} = P_n - P_n x_{n+1} [1 + x'_{n+1} P_n x_{n+1}]^{-1} x'_{n+1} P_n$$

Un calcul similaire au cas scalaire conduit à la formulation récursive:

$$\boxed{\begin{array}{c} \overline{a}_{n+1} = \overline{a}_n + K_{n+1}(y_{n+1} - x'_{n+1}\overline{a}_n) \\ \\ K_{n+1} = P_n x_{n+1}(1 + x'_{n+1} P_n x_{n+1})^{-1} \\ \\ P_{n+1} = P_n - K_{n+1} x'_{n+1} P_n \end{array}}$$

Les mêmes remarques que celles faites précédemment dans le cas scalaire peuvent être répétées concernant la structure de l'estimateur et son initialisation. On peut remarquer que contrairement à la formulation globale, l'estimateur récursif ne nécessite aucune inversion matricielle puisque $(x'_{n+1} P_n x_{n+1})$ est un scalaire.

I.4.3 Interprétation probabiliste des moindres carrés

La dérivation de la méthode des moindres carrés et les formules s'y rattachant, a été conduite sans faire aucune hypothèse probabiliste sur la nature des bruits affectant l'équation de mesure. On s'intéresse maintenant à la formulation matricielle en supposant que B_n est un vecteur aléatoire Gaussien de moyenne nulle dont les éléments ont même variance et ne sont pas corrélés entre eux. En termes de processus stochastiques, les b_i forment une séquence blanche Gaussienne définie par:

$$\begin{cases} E[B_n] = 0 \\ \\ E[B_n B'_n] = \sigma^2 I_n \end{cases} \Longleftrightarrow B_n \sim \mathcal{N}[0, \sigma^2 \mathbf{1}_n]$$

On désire caractériser statistiquement le vecteur estimé $\overline{a}_n$. Cela peut être fait en calculant la moyenne et la variance de l'erreur d'estimation $\epsilon_n = a - \overline{a}_n$. On a alors:

$$\epsilon_n = a - \overline{a}_n = a - [X'_n X_n]^{-1} X'_n Y_n = [X'_n X_n]^{-1} X'_n [X_n a - Y_n] = -[X'_n X_n]^{-1} X'_n B_n$$

Ce qui conduit à:

$$E[\epsilon_n] = -[X'_n X_n]^{-1} X'_n E[B_n] = 0$$

Cela montre que l'estimateur optimal au sens des moindres carrés, (MC), $\overline{a}$ est sans biais. D'autre part,

$$E[\epsilon_n \epsilon_n'] = [X_n' X_n]^{-1} X_n' E[B_n B_n'] X_n [X_n' X_n]^{-1} = \sigma^2 \left[X_n' X_n\right]^{-1} = \sigma^2 P_n$$

La matrice P_n qui intervient à chaque itération dans le calcul des estimés est, à un coefficient prés, la matrice de covariance de l'erreur d'estimation. Cette matrice de covariance tend donc vers 0 lorsque le nombre d'observations augmente, on dit alors que l'estimateur est **consistant**.

I.4.4 La méthode des moindres carrés et le facteur d'oubli

Les moindres carrés développés fournissent le meilleur estimé de paramètres constants sur toute la durée de l'essai. Si on considère des systèmes avec paramètres lentement variables, il faut adapter la méthode de manière à ce que les mesures trop "anciennes" soient filtrées, pour ne pas avoir une incidence trop importante sur l'estimé dans un futur lointain où les paramètres vrais peuvent avoir évolué significativement. Ceci peut être fait par pondération exponentielle des erreurs de prédiction intervenant dans le critère. Soit

$$J_n = \sum_{i=1}^{n} \lambda^{n-i} (y_i - x_i' a)^2 \qquad 0 < \lambda < 1$$

En appliquant la méthode de la même manière que précédemment, on obtient l'estimé optimal par la formulation globale:

$$\boxed{\overline{a}_n = [\sum_{i=1}^{n} \lambda^{n-i} x_i x_i']^{-1} [\sum_{i=1}^{n} \lambda^{n-i} y_i x_i]}$$

et pour la formulation récursive, on définit:

$$\overline{a}_{n+1} = P_{n+1} Q_{n+1}$$

où

$$P_{n+1}^{-1} = \sum_{i=1}^{n+1} \lambda^{n-i} x_i x_i' = \lambda P_n^{-1} + x_{n+1} x_{n+1}'$$

$$Q_{n+1} = \lambda Q_n + y_{n+1} x_{n+1}$$

ce qui conduit à:

$$\boxed{\begin{aligned} &\overline{a}_{n+1} = \overline{a}_n + K_{n+1}[y_{n+1} - x_{n+1}' \overline{a}_n] \\ &K_{n+1} = P_n x_{n+1} [\lambda + x_{n+1}' P_n x_n]^{-1} \\ &P_{n+1} = 1/\lambda [P_n - K_{n+1} x_{n+1}' P_n] \end{aligned}}$$

Chapitre II

Filtrage optimal linéaire

II.1 Introduction

La fonction filtrage a pour but essentiel la mise en forme d'un signal donné, c'est à dire l'élimination ou l'atténuation de bruits superposés au signal utile. La mise en oeuvre d'une telle fonction a longtemps et essentiellement été l'apanage des électronitiens. Toutefois, le développement de l'Automatique et des calculateurs numériques a élargi le champ des applications où la fonction filtrage est essentielle. On peut citer pour mémoire et de manière non exhaustive, la réception et la discrimination des signaux radios, la détection des signaux radar, l'utilisation du GPS, (Global Positionning System), la transmission de données numériques sur les lignes téléphoniques, le traitement d'images transmises par satellites, l'analyse de données, (sismiques, issues d'examens médicaux...), la reconnaissance des formes, l'élaboration des lois de commande de systèmes industrielles ...
Etant donnée la diversité de ces applications, il n'est pas surprenant de constater que l'on peut définir différentes classes de problèmes de filtrage et les techniques appropriées.

- Mise en forme de spectre, (filtrage avant codage, élimination de signaux parasites, extraction de composantes parasites). L'approche utilisée pour ces applications est **l'approche fréquentielle déterministe** reposant sur des méthodes analogiques ou numériques pour le développement de filtres spécifiques, (Butterworth, Bessel...).

- Détermination de paramètres pour l'identification des filtres et des systèmes. Dans ce cas, **une approche optimale déterministe** est utilisée que ce soit par l'utilisation de la méthode des moindres carrés que par celle de techniques de P.N.L..

- Réduction du bruit d'un message en vue de l'extraction de messages et la synthèse d'observateurs d'état optimaux. **L'approche optimale statistique** permet alors de développer la théorie de Wiener ou sa généralisation due à Kalman en vue de l'élimination opimale du bruit.

Dans le cadre de ce cours, seule la dernière approche sera étudiée, l'approche par la méthode des moindres carrés est vue en détail dans le cours d'identification des systèmes. Bien entendu, ces différentes classes ne sont pas totalement imperméables les unes aux autres et différents liens peuvent être mis en lumière.
Le problème du filtrage linéaire a été la première fois étudié par Wiener, (1949), et Kolmogorov, (1941), qui ont formulé le problème de la synthèse du filtre dans le domaine fréquentiel en utilisant la théorie des probabilités. Dans le cas de processus stationnaires, ayant un spectre rationnel, le filtre de Wiener fournit une solution analytique au problème du filtrage linéaire optimal. Utilisant la notion d'état, le filtre de Kalman se présente sous la forme d'un ensemble d'équations différentielles ou récurrentes plus faciles à résoudre par calculateur que la célèbre équation de Wiener-Hopf. Le filtre de Kalman généralise ainsi le filtrage optimal aux systèmes non stationnaires en présence de conditions initiales non nulles et à l'aide d'outils déterministes.

II.2 Le problème d'estimation linéaire optimale de l'état

Dans ce chapitre, nous allons étudier le problème de l'estimation de l'état d'un système dynamique linéaire, en temps continu et en temps discret, soumis à des entrées aléatoires et ceci, à partir d'un ensemble de mesures bruitées. Sous les hypothèses que les différents bruits intervenant dans la modélisation du problème sont blancs et gaussiens, ce problème trouve une solution à travers la théorie du filtre de Kalman, (modèles en temps discret), ou du filtre de Kalman-Bucy, (modèles en temps continu). Le filtre de Kalman est principalement un outil mathématique reposant sur un corps théorique important. S'il ne nécessite pas, pour son application, la connaissance complète de sa théorie, il est toutefois indispensable d'en connaitre les idées majeures présidant à son utilisation et à son fonctionnement. Il se présente comme un modèle mathématique particulièrement adapté à la programmation sur calculateur numérique. Cela n'empèche pas que de nombreux problèmes numériques, (précision finie $\neq$ précision infinie), doivent être étudiés avec précaution et rigueur avant toute implémentation pratique.
Notre objectif dans ce chapitre est d'aborder les différentes facettes du filtre de Kalman. Ainsi, nous présenterons le plus simplement possible la théorie, parfois difficile, du filtre, tout en rappelant en fin de chapitre, quelques considérations pratiques indispensables. La dérivation des équations du filtre de Kalman et de Kalman-Bucy a été repoussée en annexe afin de favoriser la lisibilité de ces notes de cours.

II.2.1 Modèles dynamiques et hypothèses

Les modèles dynamiques et les hypothèses utilisées dans ce chapitre sont résumés dans le tableau de la page suivante. En partant de ces modèles linéaires et en supposant la distribution de probabilité de la condition initiale x_0 connue, le problème du filtrage est défini comme suit.

Déf. II.2.1 *: problème du filtrage*
Le problème du filtrage *est celui de la détermination d'un estimé $\hat{x}_k$, $(\hat{x}(t))$ de l'état x_k, $(x(t))$, à partir de la connaissance de l'ensemble de mesures Y_k, $(Y(t))$.*

Deux problèmes connexes à celui-ci peuvent être définis:

- **Le problème de prédiction** qui consiste en la détermination d'un estimé $\hat{x}_l$, $(\hat{x}(\tau))$ à partir de Y_k, $(Y(t))$ avec $k < l$, $(t < \tau)$.

- **Le problème de lissage** qui consiste en la détermination d'un estimé $\hat{x}_l$, $(\hat{x}(\tau))$ à partir de Y_k, $(Y(t))$ avec $k > l$, $(t > \tau)$.

Théoriquement, il existe deux approches fondamentales au problème de l'estimation de l'état d'un système dynamique stochastique à partir d'observations bruitées de l'état.

- L'approche variationnelle, (des moindres carrés) ou encore statistique,

- l'approche probabiliste, (baysienne).

En fait, les méthodes statistiques, si on ne les interprète pas de manière probabiliste, n'impliquent pas l'utilisation **des équations différentielles stochastiques**. On travaille dans un cadre purement déterministe. Les bruits sont considérés comme des erreurs déterministes sur lesquelles on ne fait pas d'hypothèses probabilistes. Une fois l'interprétation probabiliste de ces résultats faite, on retrouve des approches formellement proches des méthodes probabilistes, sans le fondement mathématique rigoureux.

Systèmes en temps discret	**Systèmes en temps continu**
Equation de la dynamique	**Equation de la dynamique**
$x_{k+1} = F_k x_k + G_k w_k \quad , \quad x_0$	$\dot{x}(t) = F(t)x(t) + G(t)w(t) \quad , \quad x_0$
$x_0 :\ V.A.G. \sim \mathcal{N}(m_0, P_0)$	$x_0 :\ V.A.G. \sim \mathcal{N}(m_0, P_0)$
$\{w_k\}_k$ séquence blanche de V.A.G. $\sim \mathcal{N}(0, Q_k)$	$w(t)$ bruit blanc gaussien $\sim \mathcal{N}(0, Q(t))$
Equation de mesure	**Equation de mesure**
$y_k = H_k x_k + v_k$	$y(t) = H(t)x(t) + v(t)$
$\{v_k\}_k$ séquence blanche gaussienne $\sim \mathcal{N}(0, R_k)$	$v(t)$ bruit blanc gaussien $\sim \mathcal{N}(0, R(t))$
Hypothèse d'indépendance	**Hypothèse d'indépendance**
x_0, $\{w_k\}_k$ et $\{v_k\}_k$ sont indépendants	x_0, $w(t)$ et $v(t)$ sont indépendants
Ensemble de mesures	**Ensemble de mesures**
$Y_k = \{y_1, y_2, \cdots, y_k\}$	$Y(t) = \{y(\tau)\ ,\ \tau \leq t\}$

II.2.2 Approche probabiliste du problème d'estimation de l'état

Dans cette section, nous ne considérons que le cas des modèles en temps discret, l'extension au cas des modèles en temps continu pouvant se faire formellement. Pour plus de concision, il est fait abandon de la notation $p_x(\alpha)$ au profit de p_x pour la fonction densité de probabilité.
La fonction densité de probabilité conditionnelle p_{x_k/Y_k} contient toute l'information statistique de x_k fournies par les observations disponibles et par p_{x_0}. Sa connaissance fournit donc la solution complète du problème du filtrage. Comme nous l'avons vu par ailleurs au chapitre 3, pour des modèles dynamiques linéaires, p_{x_k/Y_k} est gaussienne et est donc complètement caractérisée par la moyenne conditionnelle et la matrice de covariance conditionnelle. La densité de probabilité étant caractérisée, il reste à choisir un *estimé* $\hat{x}_k$ de x_k. Il est donc nécessaire de se fixer un critère permettant de comparer différents estimés et de choisir *l'estimé optimal* au sens de ce critère. Pour cela, définissons l'erreur d'estimation par:

$$\epsilon_k = x_k - \hat{x}_k$$

Un bon estimé doit fournir *une mesure statistique* faible de l'erreur. Cette mesure statistique peut être définie comme l'espérance d'une fonction coût de l'erreur d'estimation.

Déf. II.2.2 *: fonction coût*
On appelle **fonction coût** *de l'erreur d'estimation, toute fonction de l'erreur d'estimation $\mathcal{L}(\epsilon_k)$ telle que:*

$$\mathcal{L}(0) = 0 \quad , \quad ||\epsilon^2|| \geq ||\epsilon^1|| \;\Rightarrow\; \mathcal{L}(\epsilon^2) \geq \mathcal{L}(\epsilon^1) \geq 0$$

Une classe particulièrement importante de fonctions coût est définie par la classe des fonctions quadratiques déjâ présentées au prenier chapitre.

Déf. II.2.3 *: fonction coût quadratique*
On définit **les fonctions coût quadratiques** *par:*

$$\mathcal{L}(\epsilon_k) = \epsilon_k' S \epsilon_k$$

où $S = S' > 0$, $S \in \mathbb{R}^{nn}$ afin d'assurer la propriété de croissance monotone vis à vis de la distance $|| \bullet ||$.

Ayant défini une certaine classe de fonctions coût ainsi que les caractéristiques d'un estimateur, le problème du filtrage linéaire optimal de l'état se redéfinit comme suit.

Déf. II.2.4 *: problème du filtrage optimal*
Etant donnés les modèles dynamiques et de mesure linéaires définis précédemment et l'ensemble de mesure Y_k, **le problème du filtrage optimal linéaire** *est de déterminer l'estimé optimal $\hat{x}_k$ tel que:*

$$\hat{x}_k = Arg\left[\min_{\hat{x}_k} \; E[\epsilon_k' S \epsilon_k]\right] \qquad S = S' > 0$$

Déf. II.2.5 *: minimum de variance*
L'estimé optimal ainsi défini est appelé **estimé du minimum de variance**.

Nous présentons maintenant le résultat clé de cette section qui nous permet de déterminer l'estimé du minimum de variance. Ce résultat peut se déduire d'un résultat de S. Sherman qui est de portée plus générale comme nous l'avons vu dans le premier chapitre. Afin d'alléger ce cours, nous le citons sans démonstration.

Théorème II.2.1 *;*
Soit une fonction coût quadratique $\mathcal{L}(\epsilon_k) = \epsilon_k' S \epsilon_k$ alors l'estimé du minimum de variance est la moyenne conditionnelle

$$\boxed{\hat{x}_{k/k} = E[x_k/Y_k]}$$

Ce résultat très important est renforcé par le résultat suivant.

Théorème II.2.2 *:*
Soit $\mathcal{L}[\epsilon_k] = \epsilon_k' S \epsilon_k$, alors la moyenne conditionnelle $\hat{x}_{k/k} = E[x_k/Y_k]$ minimise $E[\mathcal{L}(\epsilon_k)]$ $\forall\ S = S' > 0,\ S \in \mathbb{R}^{nn}$.

La moyenne conditionnelle fournit donc l'estimé optimal pour toute la classe des fonctions coût quadratiques. De manière générale, même si nous nous limiterons dans ce cours à cette classe, il est possible de montrer que la moyenne conditionnelle fournit l'estimé optimal pour toute la classe des fonctions coût de la définition 7, sous les hypothèses définies précédemment. De plus, cet estimateur possède la propriété fondamentale d'être sans biais.

Théorème II.2.3 *:*
L'estimateur du minimum de variance est un estimateur sans biais.

Démonstration:

$$E[x_k - \hat{x}_{k/k}] = E[x_k] - E[E[x_k/Y_k]]$$

Ainsi, d'après les résultats précédents, afin de résoudre le problème du filtrage linéaire optimal, il est nécessaire et suffisant de connaitre la densité de probabilité conditionnelle p_{x_k/Y_k}, ce qui dans le cas gaussien implique de connaitre les deux premiers moments conditionnels:

- la moyenne conditionnelle

$$\boxed{\hat{x}_{k/k} = E[x_k/Y_k]}$$

- la matrice de covariance conditionnelle de l'erreur d'estimation

$$\boxed{P_{k/k} = E[(x_k - \hat{x}_{k/k})(x_k - \hat{x}_{k/k})'/Y_k]}$$

Le premier n'est rien d'autre que l'estimé optimal appelé estimé du minimum de variance alors que le second fournit une mesure de la validité de l'estimé.
En effet, on peut montrer que $E[\epsilon_k' \epsilon_k] = Trace[E[P_{k/k}]]$. Le but est donc de rechercher les équations d'évolution de ces deux moments, équations qui constituent les équations du filtre optimal. En temps discret, on parlera de filtre de Kalman alors qu'en temps continu, on parlera de filtre de Kalman-Bucy.

II.3 Le filtre de Kalman

II.3.1 Rappel du problème

Les équations de la dynamique et de la mesure sont données par:

$$\begin{cases} x_{k+1} = F_k x_k + B_k u_k + G_k w_k \quad , \quad x_0 \\ \\ y_k = H_k x_k + v_k \end{cases}$$

sous les hypothèses:

Hypothèses II.3.1 *:*

- $x_0 \sim \mathcal{N}(m_0, P_0)$.
- $\{w_k\}_k \sim s.b.g. \sim \mathcal{N}(0, Q_k)$.
- $\{v_k\}_k \sim s.b.g. \sim \mathcal{N}(0, R_k)$.
- *Indépendance mutuelle.*

On a ajouté dans ce modèle une commande déterministe supplémentaire $B_k u_k$. Le problème du filtrage optimal linéaire est de déterminer $\hat{x}_{k/k} = E[x_k/Y_k]$, étant donné l'ensemble $Y_k = \{y_l \quad , \quad l \leq k\}$. On note comme précédemment l'estimé du minimum de variance, la moyenne conditionnelle, par $\hat{x}_{k/k} = E[x_k/Y_k]$ et la matrice de covariance conditionnelle par $P_{k/k} = E[(x_k - \hat{x}_{k/k})(x_k - \hat{x}_{k/k})'/Y_k]$.

II.3.2 Equations du filtre

Les équations du filtre peuvent se décomposer en deux étapes:

Etape de prédiction à un pas:

Elles correspondent à la propagation de l'estimé optimal en t_k par les équations dynamiques.

$$\boxed{\begin{cases} \hat{x}_{k+1/k} = F_k \hat{x}_{k/k} + B_k u_k \\[2ex] P_{k+1/k} = F_k P_{k/k} F'_k + G_k Q_k G'_k \end{cases}}$$

Etape de réactualisation

C'est la réactualisation de l'estimé optimal intervenant à l'arrivée de la mesure y_{k+1}.

$$\boxed{\begin{cases} \hat{x}_{k+1/k+1} = \hat{x}_{k+1/k} + K_{k+1}(y_{k+1} - H_{k+1}\hat{x}_{k+1/k}) \\[2ex] K_{k+1} = P_{k+1/k} H'_{k+1}[R_{k+1} + H_{k+1} P_{k+1/k} H'_{k+1}]^{-1} \\[2ex] P_{k+1/k+1} = P_{k+1/k} - K_{k+1} H_{k+1} P_{k+1/k} \end{cases}}$$

Déf. II.3.1 *: gain du filtre*
K_{k+1} *est appelé* **le gain du filtre de Kalman**.

Déf. II.3.2 *: matrice de covariance de l'erreur a priori et a posteriori*
$P_{k+1/k}$ *est appelée* **la matrice de covariance de l'erreur a priori** *alors que* $P_{k+1/k+1}$ *est appelée* **la matrice de covariance de l'erreur a posteriori**.

De même,

Déf. II.3.3 *: estimé a priori et a posteriori*
$\hat{x}_{k+1/k}$ *est appelé* **estimé a priori** *alors que* $\hat{x}_{k+1/k+1}$ *est* **l'estimé a posteriori**.

Il est important de remarquer que l'équation d'évolution de la matrice de covariance est indépendante de y_k. Cette matrice peut être déterminée hors ligne, a priori, de même que le gain du filtre.
Le schéma temporel de la propagation des calculs du filtre de Kalman est le suivant.

1- Calcul de $P_{k+1/k}$ connaissant $P_{k/k},\ F_k,\ Q_k$.

2- Calcul de K_{k+1} connaissant $P_{k+1/k},\ H_{k+1},\ R_{k+1}$.

3- Calcul de $P_{k+1/k+1}$ connaissant $K_{k+1},\ P_{k+1/k}$.

4- Calcul de $\hat{x}_{k+1/k+1}$.

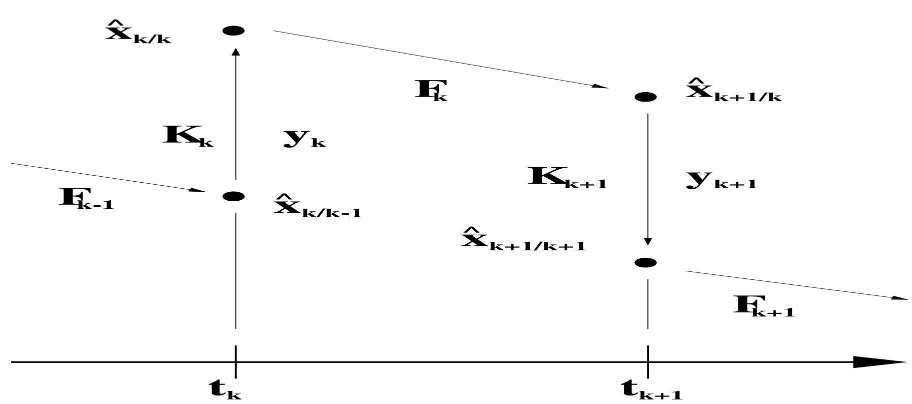


FIG. II.1 – *Schéma temporel*

Exemple:
Soit le système donné par les équations:

$$\begin{cases} x_k = x_{k-1} + w_{k-1} \\ z_k = x_k + v_k \end{cases}$$

sous les les hypothèses:

$$\begin{array}{l} E[v_k] = E[w_k] = 0 \\ E[v_{k_1} v_{k_2}] = 2\delta_{k_2-k_1} \\ E[w_{k_1} w_{k_2}] = \delta_{k_2-k_1} \\ z_1 = 2 \\ z_2 = 3 \\ z_3 = 1 \\ E[x_0] = \bar{x}_0 \\ E[(x_0 - \bar{x}_0)(x_0 - \bar{x}_0)'] = P_0 \end{array}$$

L'objectif est de déterminer la récurrence définissant l'estimé optimal et la matrice de covariance. En appliquant les différentes équations du filtre de Kalman, on obtient pour la récurrence sur la covariance:

$$P_{k+1/k} = P_{k/k} + 1$$

$$P_{k+1/k+1} = \frac{2(P_{k/k}+1)}{P_{k/k}+3}$$

et pour la récurrence sur l'estimé optimal:

$$\hat{x}_{k+1/k+1} = \hat{x}_{k/k} + K_{k+1}(z_k - \hat{x}_{k/k})$$

où:

$$K_{k+1} = \frac{P_{k/k} + 1}{P_{k/k} + 3}$$

II.3.3 Observateur optimal et filtre de Kalman

Par une simple réécriture des équations d'évolution de l'estimé optimal, on obtient:

$$\hat{x}_{k+1/k+1} = \underbrace{F_k\hat{x}_{k/k} + B_k u_k}_{modèle} + K_{k+1}\underbrace{(y_{k+1} - H_{k+1}F_k\hat{x}_{k/k})}_{commande}$$

Cela montre que le modèle dynamique d'état du filtre de Kalman a la structure d'un observateur d'ordre plein. Il consiste en effet en un modèle du système plus une commande supplémentaire proportionnelle à la différence $y_{k+1} - \hat{y}_{k+1}$. En fait, le filtre de Kalman est **un observateur en temps discret optimal**, au sens où il minimise le "carré" de l'erreur de reconstruction puisqu'il minimise tout critère de la forme $E[\epsilon_k' S \epsilon_k]$, $\quad S > 0$, pour lequel nous avons vu que $E[\epsilon_k' S \epsilon_k] = Trace[E[P_{k/k}]S]$.
Il est à noter que l'erreur de reconstruction, $\epsilon_k = x_k - x_{k/k}$, dans notre cas, vérifie l'équation d'évolution donnée par:

$$\epsilon_{k+1} = (\mathbf{1} - K_{k+1}H_{k+1})F_{k+1}\epsilon_k - (\mathbf{1} - K_{k+1}H_{k+1})G_k w_k - K_{k+1}v_{k+1} - K_{k+1}H_{k+1}B_k u_k$$

Le schéma fonctionnel suivant correspond aux équations du filtre. Il a bien la structure d'un observateur.

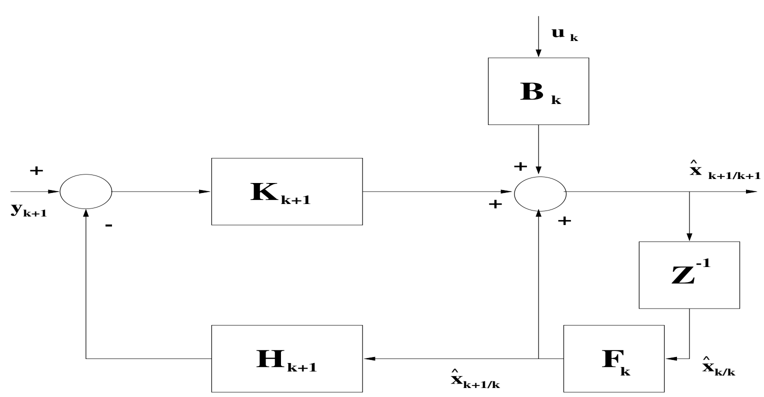


FIG. II.2 – *Schéma fonctionnel*

II.3.4 Le filtre de Kalman stationnaire

Dans ce cas, le modèle dynamique et de mesure est invariant dans le temps et donné par:

$$\begin{cases} x_{k+1} = Fx_k + Bu_k + Gw_k \\ \\ y_k = Hx_k + v_k \end{cases}$$

Il est nécessaire d'introduire des hypothèses de stationnarité sur les bruits affectant le modèle ainsi que des hypothèses de détectabilité et de stabilisabilité supplémentaires.

Hypothèses II.3.2 *:*

1- $\{w_k\}_k$ *s.b.g. stationnaire* $\sim \mathcal{N}(0, Q)$.

2- $\{v_k\}_k$ *s.b.g. stationnaire* $\sim \mathcal{N}(0, R)$.

3- (F, H) *est une paire détectable.*

4- (F, G) *est une paire stabilisable.*

Théorème II.3.1 *:*
Sous les hypothèses 12,

- *Le filtre de Kalman tend vers un régime permanent unique défini par les équations suivantes:*

$$\boxed{\begin{cases} \hat{x}_{k+1/k+1} = F\hat{x}_{k/k} + Bu_k + K_\infty(y_{k+1} - HF\hat{x}_{k/k}) \\ \\ K_\infty = P_\infty H'[R + HP_\infty H']^{-1} \\ \\ FP_\infty F' - P_\infty - FP_\infty H'(R + HP_\infty H')^{-1}HP_\infty F' + GQG' = 0 \end{cases}}$$

La matrice de covariance a priori $P_{k+1/k}$ *tend vers* P_∞*, matrice constante semi-définie positive solution de l'équation matricielle et le gain du filtre prend également la forme d'un gain constant. L'équation algébrique matricielle permettant de déterminer la matrice de covariance constante* P_∞ *est appelée* **équation algébrique de Riccati discrète**.

- *La valeur minimale moyenne de l'erreur de reconstruction est:*

$$\lim_{k \to \infty} E[\epsilon_k' S \epsilon_k] = Trace(P_\infty S)$$

- *L'erreur de reconstruction* $\epsilon_k = x_k - \hat{x}_{k/k}$ *vérifie l'équation d'évolution à matrices constantes:*

$$\epsilon_{k+1} = (\mathbf{1} - K_\infty H)F\epsilon_k + (\mathbf{1} - K_\infty H)Gw_k - K_\infty v_{k+1} - K_\infty HBu_k$$

où $(\mathbf{1} - K_\infty H)F$ *est une matrice stable.*

Exemple: suite

De l'exemple précédent on déduit que les hypothèse nécessaires sont vérifiées, ce qui conduit à:

$$P_{k+1/k+1} = P_{k/k} = P = \frac{2(P+1)}{P+3}$$

soit $P = 1$, ce qui permet de calculer:

$$\hat{x}_{1/1} = \bar{x}_0 + \frac{P_0 + 1}{P_0 + 3}(2 - \bar{x}_0)$$

Il suffit alors de continuer le processus itératif.

II.4 Le cas continu-discret

Dans ce paragraphe, sont fournies les équations du filtre, (estimé du minimum de variance), d'un processus stochastique défini par une équation dynamique continue et bruitée, sur lequel sont effectuées des mesures par échantillonnage discret sur la sortie.

II.4.1 Présentation du problème

- équation de la dynamique:

$$\dot{x}(t) = F(t)x(t) + G(t)w(t) \quad , \quad x_0$$

où:

- x_0 variable aléatoire gaussienne $\sim \mathcal{N}[m_0, P_0]$
- $w(t)$ bruit blanc gaussien $\sim \mathcal{N}[0, Q(t)]$

- équation de mesure:

$$y(t_k) = H_k x(t_k) + v_k$$

$\{v_k\}_k$ séquence blanche gaussienne $\sim \mathcal{N}[0, R_k]$

et les hypothèses usuelles, x_0 et $w(t)$ indépendants de même que $\{v_k\}_k$ et $w(t)$.
L'ensemble de mesures est donné par:

$$Y(t) = \{y(t_0),\ y(t_1), \cdots,\ y(t_n) \quad ; \quad t_n \leq t\}$$

On rappelle que le problème posé est celui de la détermination d'un estimé de $x(t)$ à partir de $Y(t)$, noté $x_{t/t}$ qui minimise $E[(x(t) - \hat{x}(t))'S(x(t) - \hat{x}(t))] \quad \forall\ S > 0$.
Il s'agit de la moyenne conditionnelle:

$$x_{t/t} = E[x(t)/Y(t)]$$

appelée estimé du minimum de variance.
Comme précédemment, il reste à déterminer l'équation d'évolution de la moyenne conditionnelle, ce qui nécéssite la connaissance de l'équation d'évolution de la matrice de covariance conditionnelle.

$$P_{t/t} = E[(x(t) - x_{t/t})(x(t) - x_{t/t})'/Y(t)]$$

De la même façon que pour le filtre discret, on établit les équations du filtre en deux étapes:

- équations du prédicteur à un pas:

$$x_{t_k/t_k} \longrightarrow x_{t_{k+1}^-/t_{k+1}^-}$$

$$P_{t_k/t_k} \longrightarrow P_{t_{k+1}^-/t_{k+1}^-}$$

Ces équations correspondent à une évolution "libre" entre deux mesures, ($t \in [t_k\ ,\ t_{k+1}[$), et sont établies comme au chapitre 3 dans le cas continu.

$$\boxed{\begin{cases} \dot{x}_{t/t} = F(t)x_{t/t} \quad , \quad x_{t_k/t_k} \qquad t \in [t_k\ ,\ t_{k+1}[\\ \\ \dot{P}_{t/t} = F(t)P_{t/t} + P_{t/t}F(t)' + G(t)Q(t)G(t)' \quad , \quad P_{t_k/t_k} \end{cases}}$$

- Réactualisation à l'arrivée d'une mesure:

  Ce sont les mêmes équations que celles dérivées dans le cas discret qui s'écrivent avec les notations retenues:

$$\boxed{\begin{cases} x_{t_{k+1}/t_{k+1}} = x_{t^-_{k+1}/t^-_{k+1}} + K_{k+1}(y(t_{k+1}) - H_{k+1}x_{t^-_{k+1}/t^-_{k+1}}) \\ \\ K_{k+1} = P_{t^-_{k+1}/t^-_{k+1}} H'_{k+1}(R_{k+1} + H_{k+1}P_{t^-_{k+1}/t^-_{k+1}} H'_{k+1})^{-1} \\ \\ P_{t_{k+1}/t_{k+1}} = P_{t^-_{k+1}/t^-_{k+1}} - K_{k+1}H_{k+1}P_{t^-_{k+1}/t^-_{k+1}} \end{cases}}$$

L'ensemble des équations encadrées précédentes constitue le filtre continu-discret.

II.5 Le filtre de Kalman-Bucy

II.5.1 Rappel du problème

Les équations de la dynamique et de mesure sont données par:

$$\begin{cases} \dot{x}(t) = F(t)x(t) + B(t)u(t) + G(t)w(t) \quad , \quad x_0 \\ \\ y(t) = H(t)x(t) + v(t) \end{cases}$$

sous les hypothèses:

Hypothèses II.5.1 *:*

- $x_0 \sim \mathcal{N}(m_0, P_0)$.
- $w(t) \sim b.b.g. \sim \mathcal{N}(0, Q(t))$.
- $v(t) \sim b.b.g. \sim \mathcal{N}(0, R(t))$.
- *Indépendance mutuelle.*

On a ajouté dans ce modèle une commande déterministe supplémentaire $B(t)u(t)$. Le problème du filtrage optimal linéaire est de déterminer $x_{t/t} = E[x(t)/Y(t)]$, étant donné l'ensemble $Y(t) = \{y(\tau) \quad , \quad \tau \leq t\}$. On note comme précédemment l'estimé du minimum de variance, la moyenne conditionnelle, par:

$$x_{t/t} = E[x(t)/Y(t)]$$

et la matrice de covariance conditionnelle par:

$$P_{t/t} = E[(x(t) - x_{t/t})(x(t) - x_{t/t})'/Y(t)]$$

II.5.2 Equation du filtre

Nous donnons ici directement les équations du filtre sans détailler leur dérivation. Celles-ci peuvent être derivées plus ou moins rigoureusement par passage à la limite du filtre discret, méthode qui est présentée en annexe, ou par une méthode variationnelle que nous ne présenterons pas dans ce document. En toute rigueur, il serait nécessaire de faire appel au calcul intégral de Itô et calcul différentiel stochastique afin de dériver ces équations. Cette théorie mathématique difficile est largement au-delà des objectifs de ce cours.

$$\boxed{\begin{array}{l} \dot{x}_{t/t} = F(t)x_{t/t} + B(t)u(t) + K(t)\underbrace{(y(t) - H(t)x_{t/t})}_{processus\ d'innovation} \quad , \quad m_0 \\[2ex] K(t) = P_{t/t}H(t)'R^{-1}(t) \\[2ex] \dot{P}_{t/t} = F(t)P_{t/t} + P_{t/t}F(t)' - P_{t/t}H(t)'R^{-1}(t)H(t)P_{t/t} + G(t)Q(t)G(t)' \quad , \quad P_0 \end{array}}$$

La dernière de ces équations, définissant l'évolution de la matrice de covariance, est appelée **équation de Riccati différentielle continue**.

Déf. II.5.1 *: gain du filtre*
$K(t)$ est appelé le **gain du filtre** *et est une fonction continue du temps.*

II.5.3 Observateur optimal et filtre de Kalman-Bucy

De la même manière que dans le cas discret, les équations du filtre de Kalman définissent un modèle dynamique d'état qui a une structure **d'observateur optimal d'ordre plein** où $K(t)$, le gain du filtre, est le gain de l'observateur. L'équation dynamique que vérifie l'erreur de reconstruction $x(t) - x_{t/t}$ s'écrit alors:

$$\dot{\epsilon}(t) = (F(t) - K(t)H(t))\epsilon(t) + G(t)w(t) - K(t)v(t)$$

Ces équations correspondent au schéma fonctionnel suivant, qui est le schéma fonctionnel d'un observateur:

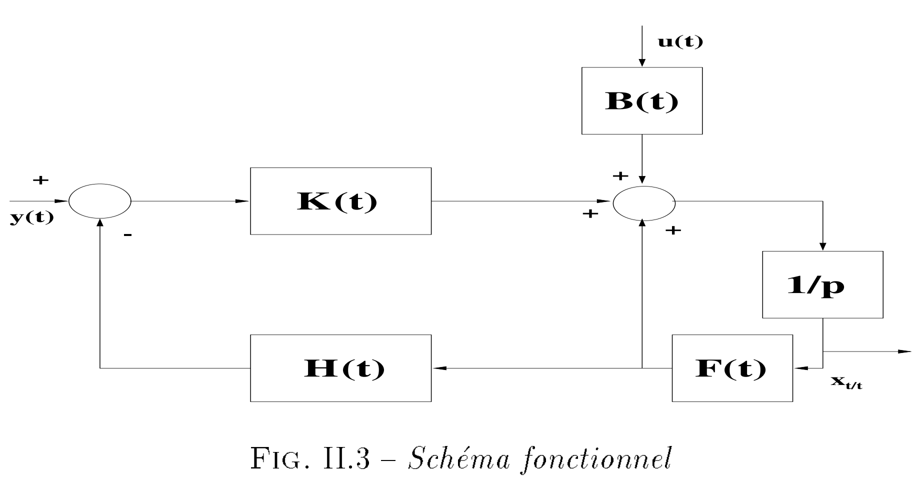


FIG. II.3 – *Schéma fonctionnel*

II.5.4 Le filtre de Kalman-Bucy stationnaire

Dans ce cas, le modèle dynamique et de mesure est invariant dans le temps et donné par:

$$\begin{cases} \dot{x}(t) = Fx(t) + Bu(t) + Gw(t) \quad , \quad x_0 \\ \\ y(t) = Hx(t) + v(t) \end{cases}$$

Les hypothèses de stationnarité sur les bruits ainsi que les hypothèses supplémentaires sont données par:

Hypothèses II.5.2 *:*

- $w(t) \sim$ *b.b.g. stationnaire* $\sim \mathcal{N}(0, Q)$.
- $v(t) \sim$ *b.b.g. stationnaire* $\sim \mathcal{N}(0, R)$.
- *Indépendance mutuelle.*
- *La paire* (F, H) *est détectable.*
- *La paire* (F, G) *est stabilisable.*

Théorème II.5.1 *:*
Sous ces hypothèses,

- *Le filtre de Kalman-Bucy tend vers un régime permanent unique défini par les équations suivantes:*

$$\boxed{\begin{array}{l} \dot{x}_{t/t} = Fx_{t/t} + Bu(t) + K_\infty \underbrace{(y(t) - Hx_{t/t})}_{processus\ d'innovation} \quad , \quad m_0 \\ \\ K_\infty = P_\infty H' R^{-1} \\ \\ FP_\infty + P_\infty F' - P_\infty H' R^{-1} H P_\infty + GQG' = 0 \end{array}}$$

  La solution de l'équation de Riccati différentielle continue tend vers une matrice constante, $P_{t/t} \to P_\infty$*, qui est solution semi-définie positive de* **l'équation algébrique de Riccati continue**.

- *La valeur minimale moyenne de l'erreur de reconstruction quadratique est donnée par:*

$$\lim_{t \to \infty} E[\epsilon'(t) S \epsilon(t)] = Trace(P_\infty S)$$

- *Le filtre de Kalman-bucy stationnaire est un observateur optimal asymptotiquement stable, l'erreur de reconstruction vérifie l'équation d'évolution:*

$$\dot{\epsilon}(t) = (F - K_\infty H)\epsilon(t) + Gw(t) - K_\infty v(t)$$

  où la matrice $F - K_\infty H$ *est asymptotiquement stable.*

II.6 Quelques remarques intuitives

II.6.1 La matrice de covariance de l'erreur

Si l'on étudie attentivement les équations décrivant le comportement de la matrice de covariance de l'erreur, quelques observations intéressantes peuvent être faites.

- Effet des perturbations dynamiques:
  Plus les paramètres statistiques des perturbations dynamiques reflétés par "la taille" de la matrice Q sont importants et plus les effets des perturbations reflétés par "la taille" de la matrice G sont prononcés, plus rapidement la covariance de l'erreur augmente.

- Effet du bruit de mesure:
  On rappelle dans le cas discret que

$$P^{-1}_{k+1/k+1} = P^{-1}_{k+1/k} + H'_{k+1} R^{-1}_{k+1} H_{k+1}$$

  Un bruit de mesure important, (R_k^{-1} faible), conduit à une faible décroissance de la covariance de l'erreur: les mesures associées contribuent faiblement à la réduction de l'erreur d'estimation. Réciproquement, de faibles erreurs de mesure, (R_k^{-1} grand), entraine une grande décroissance de la covariance de l'erreur.

II.6.2 La matrice de gain

On rappelle que:

$$K_{k+1} = P_{k+1/k+1} H'_{k+1} R^{-1}_{k+1} \qquad K(t) = P_{t/t} H(t)' R^{-1}(t)$$

Si l'on suppose que $H = \mathbf{1}$ et R^{-1} diagonale, chaque élément de K est le rapport entre les mesures statistiques de l'incertitude contenue dans l'état estimé et l'incertitude contenue dans la mesure. Si l'erreur de mesure est grande et l'erreur dans l'etat estimé est faible, cela résulte en une action faible du gain. Réciproquement, si l'erreur de mesure est faible et l'erreur dans l'état estimé grande, cela nécessite une forte correction par le gain.

II.6.3 Stabilité du filtre

Le problème de la stabilité du filtre de Kalman est principalement celui du comportement de l'estimé de l'état quand des mesures sont supprimées. Par exemple, dans le cas continu, cela revient à étudier l'équation dynamique du filtre:

$$\dot{x}_{t/t} = [F(t) - K(t)H(t)]x_{t/t}$$

c'est à dire à assurer que $x_{t/t} \to 0$ quand $t \to +\infty$ pour toute condition initiale x_0.
Du point de vue théorique, des conditions existent assurant la stabilité asymptotique du filtre. Du point de vue pratique, l'étude de la stabilité du filtre est associée aux problèmes d'erreurs de modélisation et à celui de l'implémentation que nous allons étudier plus en détails dans la prochaine section.

II.7 Quelques considérations pratiques

II.7.1 Divergence entre comportement pratique et théorique

Le filtre de Kalman fournit une solution élégante au problème de filtrage linéaire optimal. Toutefois, quand on applique les équations du filtre de Kalman à des problèmes pratiques, il n'est pas rare de rencontrer des difficultés. En effet, le filtre réellement optimal doit modéliser toutes les sources d'erreur dans le système. Cela résulte généralement en une charge de calcul trop importante pour le matériel disponible. De plus, il est supposé qu'une description exacte des dynamiques du système, des paramètres statistiques associés aux erreurs dynamiques et de mesure est bien connue. Du fait de la puissance machine généralement disponible, des simplifications sont introduites, conduisant à la synthèse d'un filtre sous-optimal. De ce fait, il est généralement recommandé d'effectuer **une analyse de sensibilité** du filtre obtenu. Cela consiste principalement en l'étude des effets sur la matrice de covariance de l'errreur d'une implémentation incorrecte des paramètres dynamiques et de mesure ou du calcul simplifié du gain à partir de données dynamiques et de mesure parfaites. Nous ne développerons pas ici les techniques spécifiques de ce type d'analyse pour nous intéresser plus particulièrement aux problèmes pratiques d'implémentation. Le lecteur intéressé peut se reporter à l'ouvrage de Gelb référencé.
On rappelle que les séquences de calcul nécessaires à l'implémentation du filtre de Kalman sont les suivantes:

1- Calcul de la matrice de covariance a priori $P_{k/k-1}$, étant données les matrices $P_{k-1/k-1},\ F_{k-1},\ Q_{k-1},\ G_{k-1}$.

$$\boxed{P_{k/k-1} = F_{k-1}P_{k-1/k-1}F'_{k-1} + G_{k-1}Q_{k-1}G'_{k-1}}$$

Cela entraine un test de symétrie et de définie positivité sur la matrice solution.

2- Calcul du gain du filtre K_k, étant données $P_{k/k-1},\ H_k,\ R_k$.

$$\boxed{K_k = P_{k/k-1}H'_k(R_k + H_kP_{k/k-1}H'_k)^{-1}}$$

Ce qui entraine l'inversion d'une matrice.

3- Calcul de la matrice de covariance a posteriori $P_{k/k}$, étant données $K_k,\ P_{k/k-1},\ H_k$.

$$\boxed{P_{k/k} = P_{k/k-1} - K_kH_kP_{k/k-1} = (\mathbf{1} - K_kH_k)P_{k/k-1}}$$

Cela entraine un test de symétrie et de définie positivité sur la matrice solution.

4- Calcul de l'estimé optimal $\hat{x}_{k/k}$, étant données $y_k,\ F_{k-1},\ \hat{x}_{k-1/k-1},\ K_k,\ H_k$.

$$\boxed{\hat{x}_{k/k} = F_{k-1}\hat{x}_{k-1/k-1} + K_k(y_k - H_kF_{k-1}\hat{x}_{k-1/k-1})}$$

Il est à noter que la forme de l'équation permettant le calcul de la matrice de covariance a posteriori à l'étape 3, n'est pas très adaptée et peut entrainer des problèmes dans les tests de symétrie ou de définie positivité. En pratique, cette équation est réécrite sous **la forme de Joseph**:

$$P_{k/k} = [\mathbf{1} - K_k H_k] P_{k/k-1} [\mathbf{1} - K_k H_k]' + K_k R_k K_k'$$

qui permet par sa forme, somme de deux termes symétriques et définis positifs, de garantir le signe et la symétrie de $P_{k/k}$.
Jusqu'à présent, nous n'avons présenté que les performances théoriques du filtre de Kalman, performances théoriques caractérisées par la matrice de covariance de l'erreur d'estimation, calculée comme la solution d'une équation différentielle dans le cas continu ou d'une équation aux différences de Riccati dans le cas discret. Toutefois, à la lumière de l'expérience acquise lors d'implémentation pratique, un certain nombre de phénomènes non prévus par cette analyse théorique préliminaire apparaissent. Entre autres,

- Les variances des erreurs d'estimation observées divergent de leur valeur théorique.
- L'erreur d'estimation est plus grande que celle prédite par la matrice de covariance.
- Les solutions des équations de Riccati ne sont pas définies positives.

Nous sommes alors confrontés à des problèmes de **divergence ou de non convergence** du filtre linéaire. Différentes variables peuvent avoir un comportement numérique divergent.

- x_k, la séquence d'état réels du système peut diverger si le système dynamique est lui-même instable.
- $\hat{x}_{k/k}$, l'état estimé peut diverger de la valeur réelle de l'état.
- $\epsilon_{k/k-1}$, l'erreur a priori d'estimation.
- $\epsilon_{k/k}$, l'erreur a posteriori d'estimation.
- $P_{k/k-1}$, la matrice de covariance a priori.
- $P_{k/k}$, la matrice de covariance a posteriori.

Tous ces comportements divergents peuvent être analysés et certains peuvent être prévus par une analyse théorique plus poussée. Nous présentons ici un certain nombre de causes de divergence que l'on a illustrées par des exemples quand cela a été possible.

II.7.2 Quelques causes de divergence pratique du filtre de Kalman

Les problèmes de modélisation

Il va sans dire que le calcul du filtre de Kalman est effectué à partir d'un modèle qui n'est au mieux qu'une bonne approximation de la réalité et qu'il est ensuite appliqué au système réel. Cela pose évidemment des problèmes de divergence pratique que l'on peut classer en *divergence apparente* et *divergence réelle*. Le premier cas correspond à des erreurs d'estimation réelles plus grandes mais bornées que celles prévues et intervient quand il y a des erreurs de modélisation conduisant à un filtre sous optimal. Le second type de divergence correspond à une véritable divergence de l'erreur d'estimation, (variables non bornées). La cause en est l'instabilité du filtre ou la non modélisation de certains états du système. Les causes de divergence issues de problèmes de modélisation peuvent être diverses.

- Variables d'état non modélisées.
- Bruit sur la dynamique non modélisé.
- Erreurs dans les coefficients définissant la dynamique.

Exemple: variable d'état non modélisée
On considère les modèles suivants:

$$\begin{array}{cc}
Modèle\ "réel" & Modèle\ de\ filtrage \\
\dot{x}_1(t) = 0 & \\
\dot{x}_2(t) = x_1(t) & \dot{x}_2(t) = 0 \\
z(t) = x_2(t) + v(t) & z(t) = x_2(t) + v(t) \\
x_2(t) = x_2(0) + x_1(0)t & x_2(t) = x_2(0)
\end{array}$$

Si l'on utilise le modèle de filtrage, on aboutit à un gain du filtre de Kalman donné par:

$$K(t) = \frac{P_0}{r(P_0 t + 1)} \qquad \lim_{t \to +\infty} K(t) = 0$$

Tenant compte du modèle "réel", on s'aperçoit que le filtre calculé ne peut fournir l'action correctrice de l'erreur dans l'estimé et l'erreur d'estimation $\hat{x}_2(t) - x_2(t)$ est ainsi divergente. Cette cause de divergence est difficile à mettre directement en évidence. L'analyse de Fourier du processus d'innovation peut être utilisée pour caractériser fréquentiellement les dynamiques non modélisées. Dans le cas d'un filtre correctement modélisé, les innovations doivent être non corrélées avec une densité spectrale de puissance plate. Des pics dans la densité spectrale de puissance indiquent les fréquences caractéristiques des dynamiques non modélisées. Un moyen de corriger ce type de comportement est d'ajouter un bruit fictif sur la dynamique du modèle servant au calcul du filtre de Kalman.

Exemple: suite, addition d'un bruit fictif
Le modèle de calcul du filtre de Kalman est alors:

$$\begin{array}{l}
\dot{x}_2(t) = w(t) \\
z(t) = x_2(t)
\end{array}$$

Ce modèle est invariant dans le temps et conduit à l'écriture d'une équation différentielle de Riccati scalaire:

$$\dot{P}_{t/t} = \dot{P} = \frac{-P^2}{R} + Q$$

qui a une solution constante $P(\infty) = \sqrt{RQ}$ permettant de calculer un gain constant

$$K(\infty) = \sqrt{\frac{Q}{R}}$$

Le modèle d'état invariant du filtre de Kalman est alors:

$$\dot{\hat{x}}_2(t) = -K_\infty \hat{x}_2(t) + K_\infty z(t)$$

dont la fonction de transfert est donnée par:

$$\frac{\hat{x}_2(p)}{z(p)} = \frac{K_\infty}{p + K_\infty}$$

Si l'on reprend le modèle réel, l'erreur d'estimation tend alors vers une constante, (même si l'on supprime le bruit de mesure), qui constitue un biais. En effet,

$$\begin{aligned}
z(t) &= x_2(t) = x_2(0) + x_1(0)t \\
z(p) &= x_2(p) = \tfrac{x_2(0)}{p} + \tfrac{x_1(0)}{p^2} \\
\epsilon_2(p) &= \hat{x}_2(p) - z(p) = \left(\tfrac{-p}{p+K_\infty}\right) x_2(p) \\
\epsilon_2(\infty) &= \hat{x}_2(\infty) - x_2(\infty) = \lim_{p \to 0} p(\hat{x}_2(p) - x_2(p)) \\
&= \lim_{p \to 0} p \left(\frac{-p}{p + K_\infty}\right) \left(\frac{x_2(0)}{p} + \frac{x_1(0)}{p^2}\right) = \frac{-x_1(0)}{K_\infty}
\end{aligned}$$

Cette artifice permet de corriger en partie le problème de modélisation incorrecte mais si l'on souhaite aller plus loin et supprimer le biais, il peut être nécessaire d'ajouter des variables d'état au modèle de calcul du filtre.

Une mauvaise modélisation des paramètres intervenant dans la matrice dynamique F, dans la matrice de transition Φ ou dans la matrice de sortie conduit généralement à des problèmes de convergence. cela ne résulte généralement pas dans une matrice de covariance ou un vecteur estimé particuliers mais ce comportement est détectable à partir du vecteur des innovations $z(t) - H\hat{x}(t)$.

Certains types de divergence du filtre peuvent être prédits par l'examen de la solution de l'équation de Riccati. C'est le cas quand les mesures disponibles ne permettent pas de garantir l'hypothèse d'observabilité.

Exemple:

Soit le système:

$$\begin{cases} \dot{x}_1(t) = w(t) & cov(w) = Q \\ \dot{x}_2(t) = 0 & \\ z(t) = x_2(t) + v(t) & cov(v) = R \end{cases}$$

avec les conditions initiales:

$$P_0 = \begin{pmatrix} \sigma_1^2 & 0 \\ 0 & \sigma_2^2 \end{pmatrix} = \begin{pmatrix} p_{11}(0) & 0 \\ 0 & p_{22}(0) \end{pmatrix}$$

L'équation différentielle de Riccati du filtre de Kalman-Bucy est donnée par:

$$\dot{P} = FP + PF' + GQG' - PH'R^{-1}HP$$

soit:

$$\begin{cases} \dot{p}_{11} = Q - \frac{p_{12}^2}{R} \\ \dot{p}_{12} = -\frac{p_{12}p_{22}}{R} \\ \dot{p}_{22} = -\frac{p_{22}^2}{R} \end{cases}$$

soit les solutions:

$$p_{22}(t) = \frac{Rp_{22}(0)}{p_{22}(0)t+R}$$

$$p_{12}(t) = 0$$

$$p_{11}(t) = Qt + p_{11}(0)$$

L'élément p_{11} divergera donc, ce qui peut s'expliquer par la matrice d'observabilité:

$$\begin{pmatrix} H \\ HF \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}$$

Une solution peut être alors d'ajouter des mesures pour que l'on retrouve la propriété d'observabilité.

D'autres causes de divergence sont plus difficiles à prédire; on peut citer par exemple une mauvaise initialisation de l'algorithme due à un mauvais choix de l'estimé initial.

$$|\hat{x}(0) - x|^2 >> Trace(P_0)$$

Les contraintes imposées par le calculateur

Généralement, les calculateurs disponibles pour une application donnée sont contraints, (en termes de place mémoire, puissance de calcul), par des considérations autres que celles issues de la complexité des équations du filtre. Il est donc souvent nécessaire de réduire autant que possible la complexité des calculs intervenant dans l'implémentation du filtre. Pour cela, il est possible d'utiliser des techniques de suppression d'états, de découplage des équations ou de préfiltrage. Toutefois, même après réduction de la complexité, il est clair qu'un calculateur numérique ne peut résoudre exactement des équations analytiques. L'approximation des opérations d'intégration et de dérivation conduit à des erreurs de troncature et l'arithmétique en longueur finie conduit à des erreurs d'arrondis. La nature de ces erreurs doit être étudiée avant toute implémentation pratique et si possible leur effet minimisé à l'aide de techniques appropriées telles que l'élimination de données redondantes, la réduction de la complexité des produits matriciels par des techniques spécifiques... .

II.8 Applications du filtre de Kalman à des problèmes d'estimation non linéaire

Dans le développement des équations du filtre de Kalman, les hypothèses de linéarité des équations, celles liées au critère quadratique et à la normalité des statistiques sont essentielles. Toutefois, les systèmes réels auxquels sont appliqués les équations du filtre de Kalman, sont rarement linéaires même s'ils peuvent être correctement approximés par un modèle linéaire. Du fait de la simplicité des équations du filtre de Kalman, il est naturel de vouloir étendre le champs d'application de ces techniques à des problèmes d'estimation non linéaire. L'objet de cette section est de présenter les extensions les plus simples qui ont été proposées à partir de deux techniques différentes de linéarisation.

II.8.1 Définition du problème

On suppose dans cette partie que le système dynamique a un modèle en temps continu ou en temps discret non linéaire donné dans la table ci-dessous. Les fonctions f, h sont supposées continûment différentiable par rapport au vecteur d'état x. Les bruits ne sont pas supposés normaux alors que les hypothèses sur l'état initial sont identiques au cas linéaire.

Systèmes en temps discret	**Systèmes en temps continu**
Equation de la dynamique:	**Equation de la dynamique:**
$x_k = f(x_{k-1}, k-1) + w_{k-1} \quad , \quad x_0$	$\dot{x}(t) = f(x(t), t) + G(t)w(t) \quad , \quad x_0$
$x_0 :\ V.A.G. \sim \mathcal{N}(m_0, P_0)$	$x_0 :\ V.A.G. \sim \mathcal{N}(m_0, P_0)$
$E[w_k] = 0 \quad E[w_k w_i'] = Q_k \delta_{ki}$	$E[w(t)] = 0 \quad E[w(t)w'(\tau)] = Q(t)\delta(t-\tau)$
Equation de mesure:	**Equation de mesure:**
$y_k = h(x_k, k) + v_k$	$y(t) = h(x(t), t) + v(t)$
$E[v_k] = 0 \quad E[v_k v_i'] = R_k \delta_{ki}$	$E[v(t)] = 0 \quad E[v(t)v'(\tau)] = R(t)\delta(t-\tau)$

Etant donnés les modèles de la table, l'objectif est d'estimer x_k ou (t) minimisant la variance de l'erreur d'estimation de manière identique au cas linéaire. Seuls les modèles sont différents.

II.8.2 Les méthodes de linéarisation

Nous présentons des techniques de linéarisation qui nécessitent l'hypothèse que f, h sont deux fonctions continûment différentiables deux fois.

Linéarisation autour d'une trajectoire nominale

Les développements sont présentés dans le cas de modèles en temps discret à extension immédiate pour les modèles en temps continu. Seul un résumé des équations dans le cas continu sera fourni.

Déf. II.8.1 *: trajectoire nominale*

Une trajectoire *est une solution particulière d'un système stochastique, pour une réalisation particulière des vecteurs aléatoires concernés. Elle est définie par une séquence de vecteurs* $\{x_k \mid k = 0,\ 1, \cdots\}$. **La trajectoire nominale** *est la trajectoire obtenue pour des valeurs espérées des vecteurs aléatoires concernés.*

Exemple:

La séquence $\{x_k^{nom.}\}$ solution de:

$$x_k^{nom.} = f(x_{k-1}^{nom.}, k-1) \qquad x_0^{nom.} = E[x_0]$$

où il n'y a pas de bruit sur la dynamique pourrait être une trajectoire nominale possible.

Si la fonction f est suffisamment continûment différentiable alors x_k est le résultat de perturbations faibles, dues aux valeurs hors moyenne des vecteurs aléatoires, autour de sa valeur nominale. L'influence des perturbations peut être représentée par un développement en série de Taylor autour de la trajectoire nominale. De plus, si ces perturbations sont suffisamment faibles, il est possible d'obtenir une bonne approximation en supprimant les termes d'ordre élevé. On définit donc:

$$\begin{aligned} \delta x_k &= x_k - x_k^{nom.} \\ \delta y_k &= y_k - h(x_k, k) \end{aligned}$$

Le développement de Taylor au premier ordre est alors:

$$\delta x_k \sim \left[\frac{\partial f(x, k-1)}{\partial x} |_{x = x_{k-1}^{nom.}} \right] \delta x_{k-1}$$

où:

$$\left[\frac{\partial f(x, k-1)}{\partial x} |_{x = x_{k-1}^{nom.}} \right] = \Phi_{k-1}^{[1]} = \begin{pmatrix} \frac{\partial f_1}{\partial x_1}|_{x=x_{k-1}^{nom.}} & \cdots & \frac{\partial f_1}{\partial x_n}|_{x=x_{k-1}^{nom.}} \\ \vdots & & \vdots \\ \frac{\partial f_n}{\partial x_1}|_{x=x_{k-1}^{nom.}} & \cdots & \frac{\partial f_n}{\partial x_n}|_{x=x_{k-1}^{nom.}} \end{pmatrix}$$

est la matrice Jacobienne de f évaluée en $x_{k-1}^{nom.}$. De même,

$$\delta y_k \sim \left[\frac{\partial h(x, k)}{\partial x} |_{x = x_k^{nom.}} \right] \delta x_k$$

où:

$$\left[\frac{\partial h(x, k)}{\partial x} |_{x = x_k^{nom.}} \right] = H_k^{[1]} = \begin{pmatrix} \frac{\partial h_1}{\partial x_1}|_{x=x_k^{nom.}} & \cdots & \frac{\partial h_1}{\partial x_n}|_{x=x_k^{nom.}} \\ \vdots & & \vdots \\ \frac{\partial h_m}{\partial x_1}|_{x=x_k^{nom.}} & \cdots & \frac{\partial h_m}{\partial x_n}|_{x=x_k^{nom.}} \end{pmatrix}$$

est la matrice Jacobienne de h évaluée en $x_k^{nom.}$. Ici, le vecteur de sortie $y_k \in \mathbb{R}^m$.
Dans le cas d'un modèle en temps continu, on définit:

$$F^{[1]} = \left[\frac{\partial f(x(t),t)}{\partial x(t)} |_{x(t)=x^{nom.}} \right]$$

$$H^{[1]} = \left[\frac{\partial h(x(t),t)}{\partial x(t)} |_{x(t)=x^{nom.}} \right]$$

En résumé, les équations linéarisées sont:

Systèmes en temps discret	**Systèmes en temps continu**
Equation de la dynamique	**Equation de la dynamique**
$\delta x_k = \Phi^{[1]}(x^{nom.}, k-1)\delta x_{k-1} + w_{k-1} \quad , \quad x_0$	$\delta \dot{x}(t) = F^{[1]}\delta x(t) + G(t)w(t) \quad , \quad x_0$
Equation de mesure	**Equation de mesure**
$\delta y_k = H^{[1]}(x^{nom.}, k)\delta x_k + v_k$	$\delta y(t) = H^{[1]}\delta x(t) + v(t)$

Le principal inconvénient de ce type de linéarisation tient au fait qu'au cours du temps, la déviation entre la trajectoire nominale et la trajectoire réelle tend à croitre et les termes négligés dans l'approximation de Taylor également. Une autre approche possible consiste à linéariser autour de la trajectoire estimée.

Linéarisation autour de la trajectoire estimée

Définissant:

$$\begin{aligned} \Phi^{[1]}(\hat{x}, k) &= \left[\frac{\partial f(x,k)}{\partial x} |_{x=\hat{x}_{k/k-1}} \right] \\ H^{[1]}(\hat{x}, k) &= \left[\frac{\partial h(x,k)}{\partial x} |_{x=\hat{x}_{k/k-1}} \right] \end{aligned}$$

dans le cas de modèles en temps discret et:

$$\begin{aligned} F^{[1]}(t) &= \left[\frac{\partial f(x(t),t)}{\partial x(t)} |_{x=x_{t/t}} \right] \\ H^{[1]}(t) &= \left[\frac{\partial h(x(t),t)}{\partial x(t)} |_{x=x_{t/t}} \right] \end{aligned}$$

dans le cas de modèles en temps continu. Les équations de modélisation sont alors:

Systèmes en temps discret	**Systèmes en temps continu**
Equation de la dynamique	**Equation de la dynamique**
$\delta x_k = \Phi^{[1]}(\hat{x}, k-1)\delta x_{k-1} + w_{k-1} \quad , \quad x_0$	$\delta \dot{x}(t) = F^{[1]}(t)\delta x(t) + G(t)w(t) \quad , \quad x_0$
Equation de mesure	**Equation de mesure**
$\delta y_k = H^{[1]}(\hat{x}, k)\delta x_k + v_k$	$\delta y(t) = H^{[1]}(t)\delta x(t) + v(t)$

Les deux modélisations linéaires présentées dans cette partie conduisent à des implémentations des filtres très différentes. L'approche par linéarisation autour d'une trajectoire nominale est appelée **approche de filtrage linéarisé** et a généralement une meilleure efficacité du point de vue de l'implémentation en temps réel. Elle est toutefois moins robuste vis à vis des erreurs d'approximation non linéaires que la deuxième approche connue sous le nom de **filtrage de Kalman étendu**.

II.8.3 Le filtre de Kalman linéarisé

Nous donnons, sans les établir, (cela peut faire un excellent exercice), les équations du filtre de Kalman linéarisé dans le cas continu et dans le cas discret.

Modèle en temps discret

Trajectoire nominale non linéaire

$$\boxed{x_k^{nom.} = f(x_{k-1}^{nom.}, k-1)}$$

Trajectoire perturbée linéarisée

$$\boxed{\begin{array}{l} \delta x = x - x^{nom.} \\ \\ \delta x_k \sim \Phi^{[1]}(x^{nom.}, k-1)\delta x_{k-1} + w_{k-1} \\ \\ w_k \sim \mathcal{N}(0, Q_k) \end{array}}$$

Modèle de mesure non linéaire

$$\boxed{\begin{array}{l} y_k = h(x_k, k) + v_k \\ \\ v_k \sim \mathcal{N}(0, R_k) \end{array}}$$

Equations d'approximation linéarisées

prédiction :

$$\hat{\delta x}_{k/k-1} = \Phi^{[1]}_{k-1} \hat{\delta x}_{k-1/k-1}$$

$$\Phi^{[1]}_{k-1} \sim \left[\frac{\partial f(x,k-1)}{\partial x} |_{x=x^{nom.}_{k-1}} \right]$$

Réactualisation :

$$\hat{\delta x}_{k/k} = \hat{\delta x}_{k/k-1} + K_k \left[y_k - h(x^{nom.}_k, k) - H^{[1]}_k \hat{\delta x}_{k/k-1} \right]$$

$$H^{[1]}_k \sim \left[\frac{\partial h(x,k)}{\partial x} |_{x=x^{nom.}_k} \right]$$

Calcul de la matrice de covariance a priori :

$$P_{k/k-1} = \Phi^{[1]}_{k-1} P_{k-1/k-1} \Phi^{[1]'}_{k-1} + Q_{k-1}$$

Calcul du gain de Kalman :

$$K_k = P_{k/k-1} H^{[1]'}_k \left[H^{[1]}_k P_{k/k-1} H^{[1]'}_k + R_k \right]^{-1}$$

Calcul de la matrice de covariance a posteriori :

$$P_{k/k} = \left[\mathbf{1} - K_k H^{[1]}_k \right] P_{k/k-1}$$

Modèle en temps continu

Trajectoire nominale non linéaire

$$\boxed{\dot{x}^{nom.}(t) = f(x^{nom.}(t), t)}$$

Trajectoire perturbée linéarisée

$$\delta x(t) = x(t) - x^{nom.}$$

$$\delta \dot{x}(t) = F^{[1]} \delta x(t) + G(t) w(t)$$

$$w(t) \sim \mathcal{N}(0, Q(t))$$

Modèle de mesure non linéaire

$$\boxed{\begin{array}{l} y(t) = h(x(t), t) + v(t) \\ \\ v(t) \sim \mathcal{N}(0, R(t)) \end{array}}$$

Equations d'approximation linéarisées

$$\boxed{\begin{array}{l} \underline{Equation\ de\ l'estimé:} \\ \\ \dot{\hat{\delta x}}_{t/t} = F^{[1]}\hat{\delta x}_{t/t} + K(t)\left[y(t) - h(x^{nom.}, t) - H^{[1]}\hat{\delta x}_{t/t}\right] \\ \\ H^{[1]} \sim \left[\frac{\partial h(x(t),t)}{\partial x(t)}|_{x=x^{nom.}}\right] \\ \\ F^{[1]} \sim \left[\frac{\partial f(x(t),t)}{\partial x(t)}|_{x=x^{nom.}}\right] \\ \\ \underline{Equations\ du\ gain\ de\ Kalman:} \\ \\ \dot{P}(t) = F^{[1]}P(t) + P(t)F^{[1]'} + G(t)Q(t)G(t)' - K(t)R(t)K'(t) \\ \\ K(t) = P(t)H^{[1]'}R^{-1}(t) \end{array}}$$

II.8.4 Le filtre de Kalman étendu

Nous donnons, sans les établir, (cela peut faire un excellent exercice), les équations du filtre de Kalman étendu dans le cas continu et dans le cas discret.

Modèle en temps discret

Modèle dynamique non linéaire

$$\boxed{\begin{array}{l} x_k = f(x_{k-1}, k-1) + w_{k-1} \\ \\ w_k \sim \mathcal{N}(0, Q_k) \end{array}}$$

Modèle de mesure non linéaire

$$\boxed{\begin{array}{l} y_k = h(x_k, k) + v_k \\ \\ v_k \sim \mathcal{N}(0, R_k) \end{array}}$$

Equations dynamiques de l'estimé

$$\boxed{\begin{array}{l} \underline{Calcul\ de\ l'estimé\ prédit:} \\ \\ \hat{x}_{k/k-1} = f(\hat{x}_{k-1/k-1}, k-1) \\ \\ \underline{Calcul\ de\ la\ mesure\ prédite:} \\ \\ \hat{y}_k = h(\hat{x}_{k/k-1}, k) \end{array}}$$

Equations d'approximation linéaires

$$\Phi_{k-1}^{[1]} \sim \left[\frac{\partial f(x,k-1)}{\partial x}\Big|_{x=\hat{x}_{k/k-1}}\right]$$

$$\hat{x}_{k/k} = \hat{x}_{k/k-1} + K_k(y_k - \hat{y}_k)$$

$$H_k^{[1]} \sim \left[\frac{\partial h(x,k)}{\partial x}\Big|_{x=\hat{x}_{k/k-1}}\right]$$

Calcul de la matrice de covariance a priori :

$$P_{k/k-1} = \Phi_{k-1}^{[1]} P_{k-1/k-1} \Phi_{k-1}^{[1]'} + Q_{k-1}$$

Calcul du gain de Kalman :

$$K_k = P_{k/k-1} H_k^{[1]'} \left[H_k^{[1]} P_{k/k-1} H_k^{[1]'} + R_k\right]^{-1}$$

Calcul de la matrice de covariance a posteriori :

$$P_{k/k} = \left[\mathbf{1} - K_k H_k^{[1]}\right] P_{k/k-1}$$

Modèle en temps continu

Modèle dynamique non linéaire

$$\boxed{\begin{array}{l} \dot{x}(t) = f(x(t),t) + G(t)w(t) \\ w(t) \sim \mathcal{N}(0, Q(t)) \end{array}}$$

Modèle de mesure non linéaire

$$\boxed{\begin{array}{l} y(t) = h(x(t),t) + v(t) \\ v(t) \sim \mathcal{N}(0, R(t)) \end{array}}$$

Equations d'implémentation

$$\boxed{\dot{x}_{t/t} = f(x_{t/t},t) + K(t)\left[y(t) - h(x_{t/t},t)\right]}$$

Equations d'approximation linéarisées

$$F^{[1]}(t) \sim \left[\frac{\partial f(x(t),t)}{\partial x}|_{x=x_{t/t}}\right]$$

$$H^{[1]}(t) \sim \left[\frac{\partial h(x(t),t)}{\partial x}|_{x=x_{t/t}}\right]$$

Equations du gain de Kalman :

$$\dot{P}(t) = F^{[1]}(t)P(t) + P(t)F^{[1]'}(t) + G(t)Q(t)G(t)' - K(t)R(t)K'(t)$$

$$K(t) = P(t)H^{[1]'}(t)R^{-1}(t)$$

Alors que dans le premier cas, le filtre de Kalman linéarisé a la forme d'un modèle linéaire sur les petites perturbations, le filtre de Kalman étendu est clairement non linéaire.

Exemple:

On considère le modèle non linéaire en temps discret:

$$x_k = x_{k-1}^2 + w_k \qquad E[w_k] = 0 \quad E[w_{k_1} w_{k_2}] = \delta_{k_1 k_2}$$

$$y_k = x_k^3 + v_k \qquad E[v_k] = 0 \quad E[v_{k_1} v_{k_2}] = \delta_{k_1 k_2}$$

$$E[x(0)] = \hat{x}_0 = 2 \quad P_0 = 1$$

$$x_k^{nom.} = 2$$

Alors, on calcule tout d'abord:

$$\begin{aligned} \Phi^{[1]}(x_k^{nom.}) &= \frac{\partial [x^2]}{\partial x}|_{x=x^{nom.}} &= 4 \\ H^{[1]}(x_k^{nom.}) &= \frac{\partial [x^3]}{\partial x}|_{x=x^{nom.}} &= 12 \end{aligned}$$

Cela conduit à un filtre de Kalman linéarisé:

$$\begin{aligned} \hat{x}_{k/k} &= \hat{\delta} x_{k/k} + 2 \\ \hat{\delta} x_{k/k} &= 4\hat{\delta} x_{k-1/k-1} + K_k \left[y_k - 8 - 48\hat{\delta} x_{k-1/k-1}\right] \\ P_{k/k-1} &= 16 P_{k-1/k-1} + 1 \\ K_k &= \frac{12 P_{k/k-1}}{144 P_{k/k-1} + 2} \\ P_{k/k} &= [1 - 12K_k] P_{k/k-1} \end{aligned}$$

A partir des mesures $y_k,\ k = 1,\ 2,\ 3, \cdots$, les valeurs de $P_{k/k-1},\ K_k,\ P_{k/k}$ et $\hat{x}_{k/k}$ peuvent être calculées. Pour k grand avec Q et R très petits, la différence $\hat{x}_k - x_k^{nom.}$ ne reste pas

faible et les résultats obtenus par cette méthode ne sont plus justifiés.
Un filtre de Kalman étendu peut être également calculé.

$$\begin{aligned}
\hat{x}_{k/k} &= \hat{x}^2_{k-1/k-1} + K_k(y_k - \hat{x}^3_{k-1/k-1}) \\
P_{k/k-1} &= 4P_{k-1/k-1}\hat{x}^2_{k-1/k-2} + 1 \\
\Phi^{[1]}_{k-1} &= 2\hat{x}_{k-1/k-2} \\
H^{[1]}_k &= 3\hat{x}^2_{k/k-1} \\
P_{k/k} &= [1 - 3K_k\hat{x}^2_{k/k-1}]P_{k/k-1} \\
K_k &= \frac{3\hat{x}^2_{k/k-1}P_{k/k-1}}{9\hat{x}^4_{k/k-1}P_{k/k-1}+2}
\end{aligned}$$

Ces équations sont plus complexes que les premières mais plus robustes vis à vis d'éventuelles erreurs.

Chapitre III

Commande optimale stochastique - Principe de séparation

III.1 Introduction et rappels de commande optimale

III.1.1 Introduction

Une des approches les plus utilisées en théorie de la commande moderne est l'approche L.Q.G., (Linear Quadratic Gaussian). Cette approche repose essentiellement sur l'utilisation d'un retour d'état optimal calculé à l'aide des techniques de commande optimale et sur l'utilisation d'un estimé optimal produit d'un filtre de Kalman. Dans ce chapitre est développée la théorie de la synthèse L.Q.G. qui utilise un résultat fondamental de la commande optimale des systèmes linéaires stochastiques: **le principe de séparation.** La commande optimale obtenue par minimisation de l'espérance mathématique d'un critère quadratique en l'état et la commande, est définie par un retour comprenant:

- le filtre de Kalman, fournissant un estimé de l'état à partir de mesures bruitées,

- le gain optimal calculé pour le système non bruité, (retour d'état), appliqué à l'estimé fourni par le filtre de Kalman.

Avant d'établir le principe de séparation, nous rappelons tout d'abord quelques résultats de la commande optimale des systèmes linéaires déterministes avec critère quadratique et retour d'état.

III.1.2 Le régulateur linéaire quadratique, (L.Q.R.)

Soit un système linéaire décrit par son équation d'état:

$$\dot{x}(t) = F(t)x(t) + B(t)u(t) \quad , \quad x_0$$

Problème III.1.1 *: L.Q. non stationnaire*

Il s'agit de déterminer la commande $u(t)$ qui minimise le critère quadratique:

$$J = \frac{1}{2}x(t_1)'Sx(t_1) + \int_{t_0}^{t_1}(x'L(t)x + u'M(t)u)dt$$

sous les hypothèses:

Hypothèses III.1.1 *:*

- $S = S' \geq 0$ *et* $L(t) = L'(t) \geq 0$.
- $M(t) = M'(t) > 0$.

La solution du problème est un retour d'état linéaire et qui est donné par:

$$\boxed{u(t) = -K(t)x(t) \qquad K(t) = M^{-1}(t)B'(t)P(t)}$$

où $P(t)$ est la matrice solution définie-positive de **l'équation différentielle matricielle de Riccati**:

$$\boxed{-\dot{P}(t) = F(t)'P(t) + P(t)F(t) - P(t)B(t)M^{-1}(t)B(t)'P(t) + L(t) \qquad P(t_1) = S}$$

Dans le cas stationnaire, le modèle est alors donné par:

$$\dot{x}(t) = Fx(t) + Bu(t) \quad , \quad x_0$$

où F et B sont des matrices constantes.

Problème III.1.2 *: L.Q. stationnaire*

Il s'agit de déterminer la commande $u(t)$ qui minimise le critère quadratique de performance avec un horizon d'optimisation infini, $(t_1 \to \infty)$:

$$J = \int_{t_0}^{\infty}(x'Lx + u'Mu)dt$$

Théorème III.1.1 *:*

Sous les hypothèses suivantes

- **Hypothèses III.1.2** *:*
  - $M = M' > 0$ *et* $L = L' \geq 0$.
  - *La paire (F, B) est stabilisable et la paire $(F, Q^{1/2})$ est détectable.*
- *La solution du problème est un retour d'état linéaire à gain constant et qui est donné par:*

$$\boxed{u(t) = -K_c x(t) \qquad K_c = M^{-1}B'P_c}$$

où P_c est la matrice solution de **l'équation algébrique de Riccati***:*

$$\boxed{F'P + PF - PBM^{-1}B'P + L = 0}$$

- *L'équation algébrique de Riccati:*

$$F'P + PF - PBM^{-1}B'P + L = 0$$

a une solution symétrique semi-définie positive unique P_c. De plus, si la paire $(F, Q^{1/2})$ est observable alors P_c est définie positive et il existe un nombre fini d'autres solutions.

- *La valeur minimale du critère de performance est*

$$J^* = x'(0)P_c x(0)$$

- *Le système en boucle fermée*

$$\dot{x}(t) = (F - BK_c)x(t)$$

est stable.

La commande par retour d'état ainsi réalisée est appelée **régulateur linéaire quadratique, L.Q.**. Ce résultat peut être obtenu par utilisation du principe du maximum de Pontryagin ou encore par utilisation du calcul des variations.

Propriété III.1.1 *:*
Un système bouclé par un régulateur linéaire quadratique possède une marge de phase de plus ou moins 60 *degrés et une marge de gain supérieure à* 0.5*.*

$$\begin{array}{ccccc} 0.5 & \leq M_g & \leq & +\infty \\ -60 & \leq & M_\phi & \leq & 60 \end{array}$$

Si l'on introduit une perturbation complexe $diag\{k_i e^{j\theta_i}\}$, elle ne conduira pas à l'instabilité si

i- $\theta_i = 0$ et $0.5 \leq k_i \leq +\infty$

ii- $k_i = 1$ et $|\theta_i| \leq 60$

III.2 La synthèse L.Q.G.

La synthèse L.Q.G. repose essentiellement sur l'utilisation conjointe du régulateur linéaire quadratique et du filtre de Kalman afin de réaliser la synthèse d'un compensateur dynamique. La structure du compensateur ainsi synthétisé est prédéfinie et est très proche de celle que l'on peut mettre en évidence à travers la théorie des observateurs.
La synthèse L.Q.G. peut être conduite également pour des systèmes décrits par des modèles en temps continu ou en temps discret. Toutefois, nous insisterons plus particulièrement sur le cas continu et nous donnerons uniquement un formulaire pour le cas discret. De même, la synthèse L.Q.G. peut s'appliquer aux modèles variants dans le temps bien que nous ne traitions que le cas stationnaire.

III.2.1 Formulation du problème de synthèse L.Q.G.

Soit le système de modèle linéaire à dynamique et mesures bruitées, donné par:

$$\begin{cases} \dot{x}(t) = Fx(t) + Bu(t) + Gw(t) \quad , \quad x_0 \\ \\ y(t) = Hx(t) + v(t) \end{cases}$$

avec les hypothèses habituelles:

Hypothèses III.2.1 *:*

- $x_0 \sim \mathcal{N}[m_0, P_0]$
- $w \sim \mathcal{N}[0, Q]$
- $v \sim \mathcal{N}[0, R]$
- *Indépendance mutuelle.*
- $L \geq 0$ *et* $M > 0$

Le problème est alors le suivant:

Problème III.2.1 *:*

Il s'agit de déterminer une commande $u(t)$ qui minimise l'espérance mathématique d'un critère quadratique:

$$u^*(t) = Arg\left\{\min_{u(t)}\left(E\left[\int_{t_0}^{\infty}(x'(t)Lx(t) + u'(t)Mu(t))dt\right]\right)\right\}$$

en utilisant l'information contenue dans les mesures y. Soit $Y(t) = \{y(\tau) \quad \tau \leq t\}$.

III.2.2 Le principe de séparation

La solution du problème L.Q.G. réside dans le principe de séparation qui établit le résultat suivant.

Théorème III.2.1 *:*

La commande optimale est obtenue par la procédure suivante:

1- *Calculer l'estimé optimal solution du problème de filtrage linéaire associé au modèle, $x_{t/t}$,*

2- *Utiliser cet estimé optimal comme une mesure exacte de l'état afin de résoudre le problème du régulateur linéaire quadratique déterministe associé.*

La démonstration de ce résultat est difficile et va bien au dela des objectifs de ce cours. Nous donnons en suivant quelques idées intuitives permetant d'expliquer de résultat.

- Considérons le même problème mais sans le terme de bruit affectant la dynamique, $w(t)$ et en considérant que tout l'état est accessible. Nous devons résoudre le problème L.Q. associé et la commande optimale est donnée par

$$u^*(t) = -K_c x(t) = -M^{-1}B'P_c x(t)$$

- Si l'on introduit le bruit $w(t)$ qui affecte la dynamique du système, alors l'état du système ne peut être ramené à son point d'équilibre 0 et l'intégrale définissant le critère de performance ne va pas converger. En fait, dans ce cas, la loi de commande optimale $u^*(t)$ calculée pour le problème L.Q. minimise:

$$\lim_{T\to\infty} \int_{t_0}^{T} E[x'(t)Lx(t) + u'(t)Mu(t)]dt$$

ce qui est en fait équivalent à:

$$\lim_{t\to\infty} E[x'(t)Lx(t) + u'(t)Mu(t)]dt$$

Le retour d'état optimal $u^*(t)$ minimise donc la valeur finale de l'erreur quadratique moyenne.

- L'état n'est évidemment pas complètement accessible. Il doit donc être estimé de manière optimale à l'aide d'un filtre de Kalman et l'état doit être remplacé dans l'expression de la commande par l'état estimé optimal $x_{t/t}$ issu du filtre de Kalman.

III.2.3 Le régulateur L.Q.G. en temps continu

Synthèse du filtre de Kalman

Le fitre de Kalman a la structure d'un observateur optimal:

$$\dot{x}_{t/t} = Fx_{t/t} + Bu(t) + K_f(y(t) - Hx_{t/t})$$

Le choix optimal de K_f minimisant $E[(x - x_{t/t})'S(x - x_{t/t})]$ est donné par:

$$\boxed{K_f = P_f H' R^{-1}}$$

où $P_f = P_f' \geq 0$ est la solution de l'équation algébrique de Riccati:

$$\boxed{FP_f + P_f F' - P_f H' R^{-1} H P_f + GQG' = 0}$$

Synthèse du régulateur linéaire quadratique

La solution optimale du problème linéaire quadratique associé est:

$$\boxed{K_c = M^{-1} B' P_c}$$

où $P_c = P_c' \geq 0$ est la solution de l'équation de Riccati algébrique:

$$\boxed{F' P_c + P_c F - P_c B M^{-1} B' P_c + L = 0}$$

Le régulateur dynamique L.Q.G.

Le régulateur L.Q.G. est un compenesateur dynamique de même ordre que l'ordre su système.

$$\boxed{\begin{array}{l} \dot{x}_{t/t} = (F - BK_c - K_f H)x_{t/t} + K_f y(t) \quad , \quad m_0 \\[2ex] u(t) = -K_c x_{t/t} \end{array}}$$

Le régulateur L.Q.G. a donc la structure suivante de la figure 5.1.

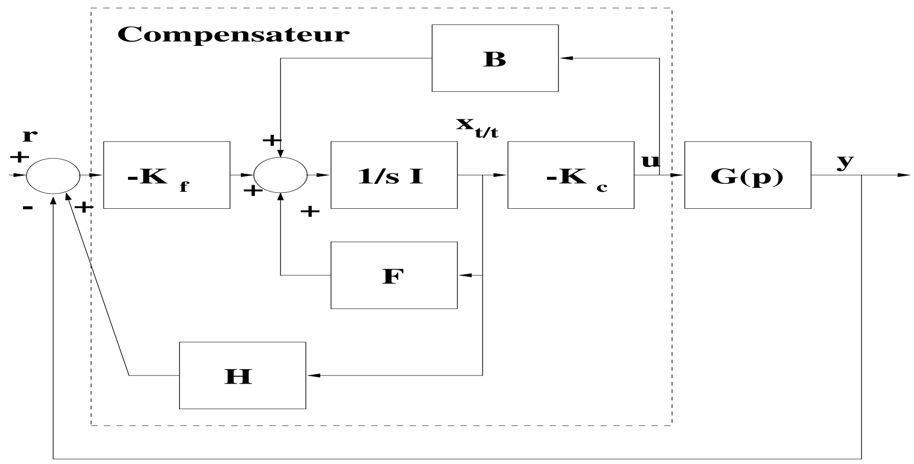


FIG. III.1 – *Schéma fonctionnel*

III.2.4 Le régulateur L.Q.G. en temps discret

Les résultats de la synthèse L.Q.G. pour les modèles en temps continu peuvent être étendus aux modèles en temps discret et sont présentés synthétiquement ci-dessous.

Synthèse du filtre de Kalman

$$\boxed{\begin{array}{l} K_f = FP_fH'(R + HP_fH')^{-1} \\ \\ FP_fF' - P_f - FP_fH'(R + HP_fH')^{-1}HP_fF' + GQG' = 0 \end{array}}$$

Synthèse du régulateur linéaire quadratique

$$\boxed{\begin{array}{l} K_c = (M + B'P_cB)^{-1}B'P_cF \\ \\ F'P_cF - P_c - F'P_cB(M + B'P_cB)^{-1}B'P_cF + L = 0 \end{array}}$$

Le régulateur dynamique L.Q.G.

$$\boxed{\begin{array}{l} x_{k+1/k+1} = Fx_{k/k} + Bu_k + K_f(y_k - H(Fx_{k/k} + Bu_k)) \quad , \quad m_0 \\ \\ u_k = -K_cx_{k/k} \end{array}}$$

III.2.5 Calcul du régulateur

Le calcul du régulateur se déduit principalement du calcul des deux matrices de gain K_c et K_f. Ces gains sont essentiellement déterminés par les paramètres de synthèse que sont les matrices de pondération M et L pour K_c définissant le critère de performance et les matrices de covariance Q et R définissant les bruits pour K_f.
Les matrices de pondération du critère de performance dans la théorie classique du régulateur L.Q. sont choisies généralement à partir de considérations physiques et par essais successifs. On cherche l'obtention de réponses temporelles en boucle fermée satisfaisantes,

(en termes de temps de réponse, de dépassement...) et le respect de contraintes sur l'allure de certaines réponses fréquentielles. La pluspart du temps, de tels objectifs peuvent être atteints en choisissant des pondérations diagonales:

$$L = \begin{pmatrix} l_1 & \mathbf{0} & \\ & \vdots & \\ & \mathbf{0} & l_n \end{pmatrix} \quad M = \rho\mathbf{1}$$

Ce choix de M implique que toutes les commandes sont pondérées à l'identique dans le critère quadratique. Pour cela, il est nécessaire de mettre les différentes commandes à l'echelle.
Il est possible de montrer, à partir d'un simple contre-exemple monovariable, que les propriétés de robustesse, (marge de gain et marge de phase) du régulateur L.Q. ne sont plus garanties si l'on introduit un observateur dans la boucle. Cela provient essentiellement du fait qu'avec l'introduction de l'observateur, les transferts en boucle ouverte du régulateur L.Q. et L.Q.G. sont différents et il n'y a aucune raison pour qu'ils aient les mêmes propriétés. Des techniques existent afin d'incorporer un observateur ne dégradant pas les propriétés du régulateur L.Q. ou plutôt permettant de les recouvrer.

Exemple:

On considère un système de positionnement angulaire d'un antenne qui est décrit par les équations d'état:

$$\dot{x}(t) = \begin{pmatrix} 0 & 1 \\ 0 & -\alpha \end{pmatrix} x(t) + \begin{pmatrix} 0 \\ k \end{pmatrix} \mu(t) + \begin{pmatrix} 0 \\ \gamma \end{pmatrix} \tau_d(t)$$

où $x(t) = [\theta(t) \quad \dot{\theta}(t)]'$, respectivement déplacement et vitesse angulaire de l'antenne, $\tau_d(t)$ couple perturbateur, modélisé comme un bruit blanc Gaussien de moyenne nulle et d'intensité constante V_d. L'équation d'observation est:

$$\eta(t) = (1\ 0)x(t) + v_m(t)$$

où $v_m(t)$ est un bruit blanc Gaussien de moyenne nulle et d'intensité constante V_m. Le système est au repos à $t = 0$. Les hypothèses d'indépendance usuelles sont vérifiées. On souhaite déterminer le régulateur L.Q.G. minimisant le critère quadratique:

$$E\left\{\int_0^\infty [\theta^2(t) + \rho\mu^2(t)]dt\right\} \quad \rho > 0$$

En appliquant le principe de séparation, le filtre de Kalman et le régulateur optimal linéaire quadratique peuvent être calculés indépendamment et dans n'importe quel ordre. Les équations du filtre de Kalman sont données par:

$$\begin{cases} \dot{\hat{x}}(t) = F\hat{x}(t) + K_f(t)(\eta(t) - H\hat{x}(t)) + B\mu(t) \quad , \quad 0 \\ K_f(t) = P_f(t)H'V_m^{-1} \\ \dot{P}_f(t) = FP_f(t) + P(t)F' - P(t)H'V_m^{-1}HP(t) + GV_dG' \quad , \quad 0 \end{cases}$$

L'équation différentielle de Riccati s'ecrit:

$$\left\{\begin{array}{l} \dot{p}_{11}(t) = 2p_{12}(t) - V_m^{-1}p_{11}^2(t) \\ \\ \dot{p}_{12}(t) = p_{22}(t) - \alpha p_{12}(t) - V_m^{-1}p_{11}(t)p_{12}(t) \\ \\ \dot{p}_{22}(t) = -2\alpha p_{22}(t) - V_m^{-1}p_{12}^2(t) + \gamma^2 V_d \end{array}\right.$$

et le gain de Kalman est:

$$K_f(t) = \left(\begin{array}{c} \frac{P_{11}(t)}{V_m} \\ \\ \frac{p_{12}(t)}{V_m} \end{array}\right)$$

Soit un filtre dont les équations d'évolution sont:

$$\left\{\begin{array}{l} \frac{d}{dt}\hat{\Theta}(t) = \dot{\hat{\Theta}}(t) + \frac{p_{11}}{V_m}\left[\eta(t) - \hat{\Theta}(t)\right] \\ \\ \frac{d}{dt}\dot{\hat{\Theta}}(t) = -\alpha\dot{\hat{\Theta}}(t) + \frac{p_{12}}{V_m}\left[\eta(t) - \gamma\dot{\hat{\Theta}}(t)\right] + k\mu(t) \end{array}\right.$$

Ici, les matrices définissant le modèle sont invariantes dans le temps, de plus les bruits sont stationnaires et la paire (F, H) est observable alors que la paire (F, G) est commandable. Le filtre de Kalman admet donc un régime permanent unique décrit par les équations:

$$\left\{\begin{array}{l} 2p_{12} - V_m^{-1}p_{11}^2 = 0 \\ \\ p_{22} - \alpha p_{12} - V_m^{-1}p_{11}P_{12} = 0 \\ \\ -2\alpha p_{22} + \gamma^2 V_d - V_m^{-1}p_{12}^2 = 0 \end{array}\right.$$

La solution définie positive est alors:

$$P_f = V_m \left(\begin{array}{cc} -\alpha + \sqrt{\alpha^2 + 2\beta} & \alpha^2 + \beta - \alpha\sqrt{\alpha^2 + 2\beta} \\ \alpha^2 + \beta - \alpha\sqrt{\alpha^2 + 2\beta} & -\alpha^3 - 2\alpha\beta + (\alpha^2 + \beta)\sqrt{\alpha^2 + 2\beta} \end{array}\right)$$

où $\beta = \gamma\sqrt{\frac{V_d}{V_m}}$. Le gain du filtre est alors:

$$K_f = \left(\begin{array}{c} -\alpha + \sqrt{\alpha^2 + 2\beta} \\ \\ \alpha^2 + \beta - \alpha\sqrt{\alpha^2 + 2\beta} \end{array}\right)$$

L'application numérique avec:

$$k = 0.787 rad/(Vs^2) \quad \alpha = 4.6 s^{-1} \quad \gamma = 0.1 kg^{-1}m^{-2} \quad V_d = 1 N^2m^2s \quad V_m = 10^{-7} rad^2 s$$

donne:

$$P_f = \left(\begin{array}{cc} 2.09\ 10^{-6} & 2.2\ 10^{-5} \\ 2.2\ 10^{-5} & 0.56\ 10^{-3} \end{array}\right) \qquad K_f = \left(\begin{array}{c} 20.96 \\ 219.78 \end{array}\right)$$

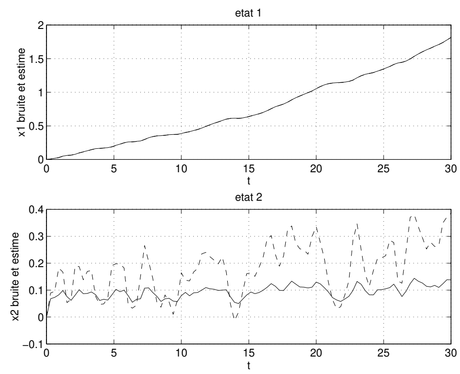


FIG. III.2 – *Résultats du filtrage*

La figure suivante donne les états bruités et les états estimés par le filtre de Kalman. La divergence de l'état 1 est naturelle puisque le système en boucle ouverte a un pole instable. Le calcul du régulateur linéaire quadratique optimal est donné par la résolution des équations algébriques suivantes:

$$\begin{cases} 1 - \frac{k^2}{\rho} q_{12}^2 = 0 \\ \\ q_{11} - \alpha q_{12} = \frac{k^2}{\rho} q_{12} q_{22} \\ \\ 2q_{12} - 2\alpha q_{22} = \frac{k^2}{\rho} q_{22}^2 \end{cases}$$

et le gain est donné par:

$$K_c = \left[\frac{1}{\sqrt{\rho}} \quad \frac{1}{k} \left[-\alpha + \sqrt{\alpha^2 + 2\frac{k}{\sqrt{\rho}}} \right] \right]$$

On obtient donc les équations du régulateur L.Q.G.

$$\begin{cases} \dot{\hat{x}}(t) = F\hat{x}(t) + K_f \left[\eta(t) - H\hat{x}(t) \right] \\ \\ u(t) = -K_c \hat{x}(t) \end{cases}$$

On voit très bien sur cet exemple comment intervient le choix des matrices de pondération du critère quadratique pour définir une classe de compensateurs dynamiques stabilisant. Cela est illustré par les figures suivantes où l'on a représenté les états du système bouclé par le compensateur L.Q.G. pour différentes valeurs de ρ. On remarque très nettement les différences de transitoires.

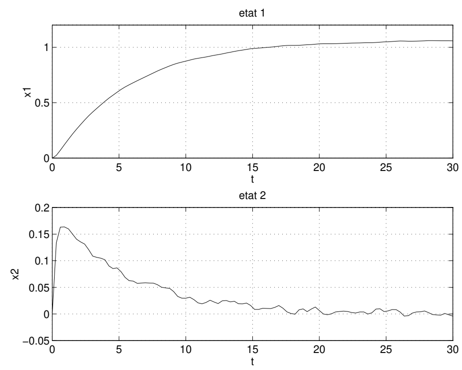


FIG. III.3 – $\rho = 1$

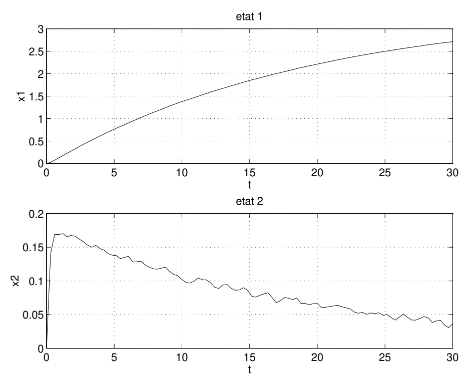


FIG. III.4 – $\rho = 10$

III.3 Propriétés du régulateur L.Q.G.

Une conséquence intéressante du principe de séparation est que la stabilité asymptotique de la boucle fermée est garantie sous les conditions de détectabilité et de stabilisabilité. En effet, définissant l'erreur d'estimation par $\epsilon(t) + x(t) = x_{t/t}$, on sait que:

$$\dot{\epsilon}(t) = (F - K_f H)\epsilon(t) + Gw(t) - K_f v(t)$$

d'autre part:

$$\dot{x}(t) = (F - BK_c)x(t) + BK_c\epsilon(t) + Gw(t)$$

ce qui permet de réecrire le modèle augmenté:

$$\begin{cases} \frac{d}{dt}\begin{pmatrix} x \\ \epsilon \end{pmatrix} \begin{pmatrix} F - BK_c & -BK_c \\ 0 & F - K_f H \end{pmatrix} \begin{pmatrix} x \\ \epsilon \end{pmatrix} + \begin{pmatrix} G \\ -G \end{pmatrix} w(t) + \begin{pmatrix} 0 \\ K_f \end{pmatrix} v(t) \\ \\ y(t) = [H \quad 0] \begin{pmatrix} x \\ \epsilon \end{pmatrix} + v(t) \end{cases}$$

Ces équations décrivent complètement le comportement dynamique en boucle fermée. Les pôles en boucle fermée sont les valeurs propres de $F - BK_c$ plus les valeurs propres de $F - K_f H$, puisque la matrice dynamique du système augmenté est bloc triangulaire. Ces valeurs propres sont stables sous les hypothèses de stabilisabilité pour les paires (F, G) et (F, B), et de détectabilité pour la paire (F, H).

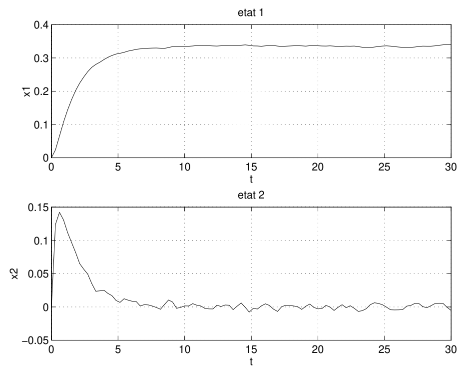


FIG. III.5 – $\rho = 0.1$

Annexe A

Détermination des équations du filtre de Kalman

A.1 Le filtre de Kalman

Le problème posé est celui de l'obtention d'un estimé de l'état. L'estimé recherché est celui qui minimise l'espérance mathématique d'un coût quadratique sur l'erreur d'estimation:

$$\min_{\hat{x}_{k/k}} E[(x_k - \hat{x}_{k/k})' S (x_k - \hat{x}_{k/k})]$$

On montre que cet estimé du minimum de variance est la moyenne conditionnelle $E[x_k/Y_k]$, qui également coincide, dans le cas gaussien, avec l'estimé du maximum de vraissemblance, c'est à dire celui obtenu par $\max_x p[x/Y_k]$.
On note $\hat{x}_{k/k} = E[x_k/Y_k]$, l'estimé a posteriori et $\hat{x}_{k/k-1} = E[x_k/Y_{k-1}]$, l'estimé a priori. Le filtre récursif est déterminé par l'écriture de la relation de récurrence liant $\hat{x}_{k+1/k+1}$ à l'estimé $\hat{x}_{k/k}$, à l'instant précédent et à la mesure collectée à l'instant $k+1$, y_{k+1}.
Nous procédons en deux étapes:

$$\hat{x}_{k/k} \rightarrow \hat{x}_{k+1/k} \rightarrow \hat{x}_{k+1/k+1}$$

A l'instant k sont supposés connus $\hat{x}_{k/k}$ ainsi que $P_{k/k}$, (covariance conditionnelle), ou de manière équivalente, (propriété des gaussiennes), $p[x_k/Y_k]$.
Etape de prédiction: Cette étape consiste en la propagation de l'estimé $\hat{x}_{k/k}$. Pour cela, on applique les résultats concernant la réponse des systèmes linéaires à des entrées bruitées.

$$\hat{x}_{k+1/k} = E[x_{k+1}/Y_k] = E[F_k x_k + B_k u_k + G_k w_k / Y_k]$$

soit:

$$\hat{x}_{k+1/k} = F_k \hat{x}_{k/k} + B_k u_k \quad , \quad m_0 \tag{A.1}$$

Pour une caractérisation complète de la loi de distribution conditionnelle de x_k sachant Y_k, il est nécessaire de déterminer également la matrice de covariance conditionnelle, via son équation d'évolution:

$$P_{k/k} = E[(x_k - \hat{x}_{k/k})(x_k - \hat{x}_{k/k})'/Y_k]$$

$$P_{k+1/k} = E[(x_{k+1} - \hat{x}_{k/k})(x_{k+1} - \hat{x}_{k/k})'/Y_k]$$

Or dans le cas gaussien, il a été vu que, X et Y étant conjointement gaussiens, $(X - E(X/Y))$ est indépendant de Y, donc, les matrices précédentes s'écrivent simplement,

$$\begin{aligned} P_{k/k} &= E[(x_k - \hat{x}_{k/k})(x_k - \hat{x}_{k/k})'] \\ P_{k+1/k} &= E[\underbrace{(x_{k+1} - \hat{x}_{k/k}}_{erreur\ a\ priori\ \epsilon^-_{k+1}})(x_{k+1} - \hat{x}_{k/k})'] \end{aligned} \tag{A.2}$$

d'où,

$$P_{k+1/k} = F_k P_{k/k} F'_k + G_k Q_k G'_k \quad , \quad P_0 \tag{A.3}$$

<u>Etape de réactualisation:</u>

Il reste à établir les relations liant $x_{k+1/k+1}$ à l'estimé a priori $x_{k+1/k}$. On recherche un estimé linéaire:

$$\hat{x}_{k+1/k+1} = K^1_{k+1} + K^2_{k+1} y_{k+1}$$

où les constantes $K^1_{k+1},\ K^2_{k+1}$ sont à déterminer sachant que l'on cherche à résoudre:

$$\min_{(K^1_{k+1},\ K^2_{k+1})} E[\underbrace{\epsilon'_{k+1}}_{erreur\ a\ posteriori} \epsilon_{k+1}]$$

Soit $J = E[\epsilon'_{k+1}\epsilon_{k+1}]$ alors

$$\begin{aligned} J &= E\left[trace[(x_{k+1} - \hat{x}_{k+1/k+1})(x_{k+1} - \hat{x}_{k+1/k+1})']\right] \\ J &= trace\left[E\left[((x_{k+1} - \hat{x}_{k+1/k}) - (\hat{x}_{k+1/k+1} - \hat{x}_{k+1/k}))((x_{k+1} - \hat{x}_{k+1/k}) - (\hat{x}_{k+1/k+1} - \hat{x}_{k+1/k}))'\right]\right] \\ J &= trace\left[P_{k+1/k} + K^2_{k+1}(P_{y_{k+1}} + \overline{y}_{k+1}\overline{y}'_{k+1})K^{2'}_{k+1} + (K^1_{k+1} - \hat{x}_{k+1/k})(K^1_{k+1} - \hat{x}_{k+1/k})' + 2K^2_{k+1}\overline{y}_{k+1}(K^1_{k+1} - \hat{x}_{k+1/k})' \right. \\ &\quad \left. -2K^2_{k+1}P_{y_{k+1}x_{k+1}}\right] \end{aligned}$$

où:

$$\overline{y}_{k+1} = H_{k+1}\overline{x}_{k+1/k}$$

$$P_{y_{k+1}x_{k+1}} = E\left[(y_{k+1} - \overline{y}_{k+1})(x_{k+1} - \hat{x}_{k+1/k})'\right]$$

$$P_{y_{k+1}} = E\left[(y_{k+1} - \overline{y}_{k+1})(y_{k+1} - \overline{y}_{k+1})'\right]$$

On écrit alors les conditions nécessaires d'optimalité du critère J:

$$\frac{\partial J}{\partial K^1_{k+1}} = 2(K^1_{k+1} - \hat{x}_{k+1/k}) + 2K^2_{k+1}\overline{y}_{k+1} = 0$$

$$\frac{\partial J}{\partial K^2_{k+1}} = 2K^1_{k+1}(P_{y_{k+1}} + \overline{y}_{k+1}\overline{y}'_{k+1}) - 2P_{x_{k+1}y_{k+1}} + 2(K^1_{k+1} - \hat{x}_{k+1/k})\overline{y}'_{k+1}$$

Cela permet d'obtenir

$$\begin{aligned} K^1_{k+1} &= \hat{x}_{k+1/k} - K^2_{k+1}\overline{y}_{k+1} = (\mathbf{1} - K^2_{k+1})H_{k+1}\hat{x}_{k+1/k} \\ K^2_{k+1} &= P_{x_{k+1}y_{k+1}}P^{-1}_{y_{k+1}} = P_{k+1/k}H'_{k+1}(H_{k+1}P_{x_{k+1/k}}H'_{k+1} + R_{k+1})^{-1} \end{aligned}$$

K^2_{k+1} est le gain du filtre de Kalman. Soit, finalement:

$$\hat{x}_{k+1/k+1} = \hat{x}_{k+1/k} - P_{k+1/k}H'_{k+1}(H_{k+1}P_{k+1/k}H'_{k+1} + R_{k+1})^{-1}$$

On a d'autre part pour la covariance a posteriori:

$$\begin{aligned} P_{k+1/k+1} &= E\left[\epsilon_{k+1}\epsilon'_{k+1}\right] \\ P_{k+1/k+1} &= P_{k+1/k} - P_{x_{k+1}y_{k+1}}P^{-1}_{y_{k+1}}P_{y_{k+1}x_{k+1}} \end{aligned}$$

d'où l'équation récurrente de Riccati:

$$P_{k+1/k+1} = P_{k+1/k} - P_{k+1/k}H'_{k+1}(H_{k+1}P_{k+1/k}H'_{k+1} + R_{k+1})^{-1}H_{k+1}P_{k+1/k}$$

A.2 Détermination du filtre de Kalman-Bucy

Les équations du filtre continu sont obtenues comme limites de celles du filtre discret, lorsque la période d'échantillonnage T tend vers zero.
Ainsi, prenant:

$$t_k = kT \quad , \quad x_k = x(kT) \quad , \quad F_k = F(kT) \cdots w_k = w(kT)$$

on écrit le système discret:

- équation de la dynamique:

$$x_{k+1} = (\mathbf{1} + F_kT)x_k + TG_kw_k \quad , \quad x_0 \tag{A.4}$$

  - x_0, w_k indépendants,
  - w_k séquence blanche gaussienne $\sim \mathcal{N}[0, \frac{Q_k}{T}]$

- équation de mesure, (observation):

$$y_k = H_kx_k + v_k \tag{A.5}$$

  v_k séquence blanche gaussienne $\sim \mathcal{N}[0, \frac{R_k}{T}]$

Ainsi, reprenant les équations du filtre discret, il vient

- pour le gain:

$$K_{k+1} = P_{k+1/k} H'_{k+1} \left(\frac{R_{k+1}}{T} + H_{k+1} P_{k+1/k} H'_{k+1} \right)^{-1}$$

$$K_{k+1} = T P_{k+1/k} H'_{k+1} \left(R_{k+1} + T H_{k+1} P_{k+1/k} H'_{k+1} \right)^{-1}$$

$$\cong T P_{k+1/k} H'_{k+1} R^{-1}_{k+1}$$

- d'où pour l'estimé:

$$x_{k+1/k+1} = [\mathbf{1} + F_k T] x_{k/k} + T P_{k+1/k} H'_{k+1} R^{-1}_{k+1} (y_{k+1} - H_{k+1} x_{k+1/k})$$

et, par passage à la limite,

$$\left\{ \begin{array}{l} \dot{x}_{t/t} = F(t) x_{t/t} + K(t)(y(t) - H(t) x_{t/t}) \quad , \quad m_0 \\ \textit{où } K(t) = P_{t/t} H(t)' R(t)^{-1} \end{array} \right. \tag{A.6}$$

Enfin, pour la matrice de covariance, en se limitant aux termes au premier ordre en T, il vient:

$$P_{k+1/k+1} = P_{k/k} + T \left[F_k P_{k/k} + P_{k/k} F'_k + G_k Q_k G'_k - P_{k/k} H'_{k+1} R^{-1}_{k+1} H_{k+1} P_{k/k} \right]$$

d'où, par passage à la limite:

$$\dot{P}_{t/t} = F(t) P_{t/t} + P_{t/t} F(t)' - P_{t/t} H(t)' R(t)^{-1} H(t) P_{t/t} + G(t) Q(t) G(t)' \quad , \quad P_0 \tag{A.7}$$

(A.6), (A.7) constituent les équations du filtre de Kalman continu.

Annexe B

Rappels de la théorie des probabilités

La théorie des probabilités est l'étude des lois régissant les phénomènes aléa-

toires de masse. Il ne faut pas réduire la théorie à l'étude des méthodes de traitement des résultats des épreuves aléatoires et d'extraction des données nécessaires est appelée **statistique mathématique**. Un phénomène aléatoire est un phénomène dans lequel le hasard intervient. L'exemple le plus simple de phénomène aléatoire est donné par les erreurs de mesure. Si l'on répète à de multiples reprises la mesure d'une même grandeur sous les mêmes conditions, les résultats obtenus ne seront jamais identiques. Le résultat de chaque mesure est entaché d'une erreur aléatoire et les résultats des différentes mesures contiennent différentes erreurs. On a ainsi une dispersion des résultats de mesure. Un autre exemple est fourni par la dispersion des obus tirés par un même canon visant la même cible. L'impossibilité de connaitre avec une précision absolue les paramètres définissant l'atmosphère en tous points suivis par la trajectoire de l'obus, conduit à cette dispersion aléatoire. Les pannes dans un dispositif technique quelconque sont à l'origine d'un autre type d'exemple de phénomènes aléatoires. En effet, il est impossible de prévoir si une panne aura lieu et dans l'affirmative, à quel instant précis. Dans la suite de ce chapitre, nous allons rappeler quelques fondements de la théorie des probabilités en introduisant tout d'abord les notions et le vocabulaire élémentaire à l'aide d'une approche empirique. Partant de ces notions empiriques de probabilité, variable aléatoire..., nous déduirons en séquence, les notions abstraites par axiomatisation.

B.1 Approche empirique des phénomènes aléatoires

Pendant l'observation répétée de phénomènes aléatoires, certaines lois régissant ces phénomènes aléatoires peuvent être mises en évidence. L'étude de ces lois permet de "maitriser" le phénomène aléatoire étudié et d'en prévoir en quelque sorte les produits. Ces lois ne peuvent évidemment être déduites que si le phénomène aléatoire est observé un nombre suffisant de fois. Les phénomènes aléatoires pouvant être observés suffisamment souvent sont appelés **phénomènes aléatoires de masse**. La théorie des probabilités est donc la théorie des lois régissant les phénomènes aléatoires de masse. Son point de départ est constitué par des faits expérimentaux sur la base desquels seront formulés des concepts abstraits généralisant ces faits. Avant de définir ces concepts abstraits, il est nécessaire de détailler le vocabulaire que nous allons employer.

Déf. B.1.1 *: épreuve*
L'observation d'un phénomène au cours de la réalisation d'un certain nombre de conditions et d'actions est appelée **épreuve**.

Une épreuve peut être caractérisée qualitativement ou quantitativement. Ainsi, les résultats produits par une épreuve peuvent ou non vérifier certaines propriétés.

Déf. B.1.2 *: évènement*
Un évènement *est le fait que les résultats d'une épreuve vérifie une certaine propriété.*

Par exemple, le fait d'atteindre ou non une cible après un tir d'obus est un évènement. Quantitativement, il peut être souhaitable de déterminer certaines grandeurs obtenues au cours du déroulement de l'épreuve.

Déf. B.1.3 *: variable aléatoire*
Une variable aléatoire *est une grandeur choisie ou obtenue suite à une épreuve et pouvant prendre différentes valeurs, (***ses réalisations***), non prévisibles.*

Des exemples simples de variables aléatoires sont donnés par le résultat d'une mesure ou le temps de fonctionnement d'un appareil quelconque sans panne. Un exemple typique d'évènement associé tout particulièrement à une variable aléatoire est défini par l'appartenance d'une variable aléatoire à un ensemble donné. Un autre élément quantitatif intéressant dans l'étude d'un phénomène aléatoire est la fréquence d'occurence d'un évènement lors de la répétition d'une épreuve donnée.

Déf. B.1.4 *: fréquence d'un évènement*
La fréquence *d'un évènement est définie par le quotient du nombre de réalisations d'un évènement par le nombre d'épreuves réalisées.*

Bien évidemment, afin de pouvoir faire des comparaisons entre différentes fréquences, il est nécessaire de supposer qu'une épreuve donnée peut être répétée indéfiniment. Par extension directe, il peut être nécessaire de calculer une fréquence d'un évènement A sous la condition qu'un évènement B a au préalable été réalisé.

Déf. B.1.5 *: fréquence conditionnelle*
La fréquence conditionnelle *d'un évènement est le quotient du nombre de réalisations d'un évènement A, sachant qu'un évènement B s'est réalisé, par le nombre d'épreuves réalisées.*

Les propriétés des fréquences sont facilement déduites de leur définition.

Propriété B.1.1 *:*

- $0 \leq f \leq 1$.

- *La fréquence d'un évènement impossible est* 0.

- *La fréquence d'un évènement certain est* 1.
- $f(A \text{ ou } B) = f(A) + f(B)$ *si* A *et* B *sont incompatibles.*
- $f(A \text{ et } B) = f(A)f(B/A)$.

Expérimentalement, il est remarquable de constater que la valeur d'une fréquence d'un évènement se stabilise quand le nombre d'épreuves augmente. Il semble donc que la fréquence n'est plus un résultat aléatoire dépendant du nombre d'épreuves réalisées. Cette stabilité de la fréquence des évènements permet de supposer que chaque évènement est associé à un certain nombre, **la probabilité** de l'évènement. Cette probabilité est une caractéristique objective d'un évènement dans une épreuve donnée. Associée à la fréquence conditionnelle, on peut également définir la notion abstraite de **probabilité conditionnelle**. Une fois définie cette notion, il est important, afin d'approfondir l'étude quantitative, de définir des moyens d'étudier la position des valeurs d'une variable aléatoire ainsi que leur dispersion lors de la tenue de n épreuves.

Déf. B.1.6 *: échantillon*

Un échantillon *est l'ensemble des valeurs prises par une variable aléatoire dans une suite d'épreuves.*

Afin de caractériser la position des points expérimentaux d'un échantillon, on utilise habituellement la moyenne arithmétique des valeurs de la variable aléatoire.

Déf. B.1.7 *: moyenne*

Supposons qu'une variable aléatoire X *a pris au cours de* n *épreuves les valeurs* $\alpha_1, \cdots, \alpha_n$ *alors* **sa moyenne** *est définie par:*

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} \alpha_i$$

De même, afin de caractériser la dispersion des valeurs d'une variable aléatoire, on définit la moyenne arithmétique des carrés des écarts des valeurs expérimentales de la variable aléatoire à la moyenne.

Déf. B.1.8 *: variance*

Supposons qu'une variable aléatoire X *a pris au cours de* n *épreuves les valeurs* $\alpha_1, \cdots, \alpha_n$ *alors* **la variance** *est donnée par:*

$$v_x = \frac{1}{n}\sum_{i=1}^{n} (\alpha_i - \bar{x})^2$$

L'inconvénient majeur de la variance est qu'elle a la dimension au carré de la valeur de la variable aléatoire. Pour des raisons d'homogénéité, on adopte comme mesure de dispersion l'écart quadratique moyen défini comme suit.

Déf. B.1.9 *: écart-type*

$$\sigma_x = \sqrt{v_x} = \sqrt{\frac{1}{n}\sum_{i=1}^{n} (\alpha_i - \bar{x})^2}$$

Lors de l'étude de plusieurs variables aléatoires liées au même évènement, il est également intéressant d'étudier la dépendance existant entre elles.

Déf. B.1.10 *: covariance*

Soient deux variables aléatoires X et Y prenant respectivement au cours de n épreuves les valeurs $\alpha_1, \cdots, \alpha_n$ et $\beta_1, \cdots, \beta_n$. On définit **la covariance** *de X et Y par:*

$$C_{xy} = \frac{1}{n}\sum_{i=1}^{n}(\alpha_i - \bar{x})(\beta_i - \bar{y})$$

Cette grandeur entre en fait dans le calcul de la meilleur dépendance linéaire entre X et Y. Associé à la covariance, une mesure de dépendance entre X et Y est définie par la grandeur sans dimension suivante:

Déf. B.1.11 *: coefficient de corrélation*

$$\rho_{xy} = \frac{C_{xy}}{\sigma_x \sigma_y}$$

Nous sommes parvenus à la notion de probabilité en introduisant celle de fréquence. Toutefois, les propriétés des fréquences ne peuvent être démontrées une fois étendues aux probabilités. Il est donc nécessaire de les poser comme des axiomes définissant le concept abstrait de probabilité. On constatera dans le paragraphe suivant que nombres de notions définies empiriquement ont leur contrepartie abstraite servant à l'élaboration de la théorie des probabilités.

B.2 Théorie axiomatique des probabilités

B.2.1 Espace probabilisé

Il est nécessaire de définir en premier lieu le concept *d'espace probabilisé* associé à une expérience donnée et qui est le fondement de la théorie. Pour cela, on doit tout d'abord se donner un ensemble Ω, ensemble des résultats possibles à une expérience donnée, dont les éléments sont notés ω et sont exclusifs. Un *évènement*, la réalisation ou la non réalisation d'un phénomène, est défini comme un sous-ensemble de Ω. Alors, Ω est *l'évènement sûr* et $\emptyset$ est *l'évènement impossible*.
On note $\mathcal{B}$ une classe d'ensembles d'éléments ω, (un ensemble d'évènements), munie de la structure algébrique de σ-algèbre, (ou de tribu Borélienne).

Déf. B.2.1 *: σ-algèbre*

$\mathcal{B}$ est une **σ-algèbre** *construite sur Ω si les propriétés suivantes sont satisfaites:*

Propriété B.2.1 *:*

$$\begin{cases} \bigcup_{i=1}^{\infty} A_i \in \mathcal{B} \quad \forall\ A_i \in \mathcal{B} \\ \bigcap_{i=1}^{\infty} A_i \in \mathcal{B} \quad \forall\ A_i \in \mathcal{B} \\ (\Omega - A) \in \mathcal{B} \quad \forall\ A \in \mathcal{B} \\ \emptyset \in \mathcal{B},\ \Omega \in \mathcal{B} \end{cases} \tag{B.1}$$

Déf. B.2.2 *: mesure de probabilité*

Pour compléter la structure de l'espace probabiliste, on définit **une mesure de probabilité P**, *qui n'est rien d'autre qu'une fonction définie de $\mathcal{B} \rightarrow [0 \quad 1]$, satisfaisant les axiomes de probabilité suivants:*

Axiomes B.2.1 *:*

$$\begin{cases} P[A] \geq 0 \quad \forall\ A \in \mathcal{B} \\ P[\Omega] = 1 \\ Si\ A_i \bigcap A_j = \emptyset, \quad i \neq j \quad P[A_i \bigcup A_j] = P[A_i] + P[A_j] \end{cases} \tag{B.2}$$

Le dernier axiome s'étend à une séquence d'évènements A_i, mutuellement exclusifs deux à deux:

$$P[\bigcup_{i=1}^{\infty} A_i] = \sum_{i=1}^{\infty} P[A_i]$$

A partir de ces axiomes, on déduit les résultats, (de base), suivants:

Propriété B.2.2 *:*

$$\begin{cases} P[\emptyset] = 0 \\ P[A] = 1 - P[\Omega - A] \\ P[A_i \bigcup A_j] = P[A_i] + P[A_j] - P[A_i \bigcap A_j] \\ A_1 \subset A_2 \Rightarrow P[A_1] \leq P[A_2] \end{cases} \tag{B.3}$$

Déf. B.2.3 *: espace probabilisé*

Le triplet $(\Omega, \mathcal{B}, P)$ est appelé un **espace probabilisé**, *(ou encore probabiliste).*

Exemple 1: Le jeu de dé

$$\Omega = \{1, 2, 3, 4, 5, 6\}$$

ω, entier compris entre 1 et 6, (correspond au tirage de la face ω). Le tirage d'un nombre pair correspond à l'évènement $\{2, 4, 6\}$. Ainsi,

$$\mathcal{B}_1 = \{\emptyset, \{1, 3, 5\}, \{2, 4, 6\}, \Omega\}$$

est une σ-algèbre alors que:

$$\mathcal{B}_2 = \{\emptyset, \{1\}, \{1, 3, 5\}, \{2, 4, 6\}, \Omega\}$$

n'est pas une σ-algèbre car:

$$[1] \bigcup [2, 4, 6] \notin \mathcal{B}_2$$

De plus, il est aisé de voir que $P[\{2, 4, 6\}] = 0.5$.

B.2.2 Probabilités conditionnelles

A partir de l'espace probabilisé $(\Omega, \mathcal{B}, P[\bullet])$, on peut définir une nouvelle loi de probabilité sur la même tribu $P[\bullet/A]$, qui est, par définition, la probabilité d'évènements sachant que l'évènement A est réalisé.

Déf. B.2.4 *: probabilité conditionnelle*
$P[\bullet/A]$, appelée **probabilité conditionnelle***, est une application de $\mathcal{B}$ dans $[0 \quad 1]$ définie par:*

$$P[B/A] = \frac{P[B \bigcap A]}{P[A]} \qquad \forall\ A,\ B \in \mathcal{B} \tag{B.4}$$

Exemple 2:

On suppose que la population d'une ville est masculine à 40% et féminine à 60%. On suppose de plus que 50% des hommes sont des fumeurs alors que 30% de femmes fument. Trouver la probabilité qu'un fumeur soit un homme.

Soit H l'évènement correspondant à la sélection d'un homme et F l'évènement correspondant à la sélection d'une femme. De même, soit S l'évènement correspondant à la sélection d'un fumeur et N à celui d'un non fumeur. Alors:

$$P[S/H] = 0.5 \quad P[S/F] = 0.3$$

$$P[H] = 0.4 \qquad P[F] = 0.6$$

On recherche $P[H/S]$. Par définition,

$$P[H/S] = \frac{P[H \bigcap S]}{P[S]}$$

De plus,

$$P[H \bigcap S] = P[H]P[S/H] = 0.2$$

Or,

$$S = (S \bigcap H) \bigcup (S \bigcap F)$$

soit

$$P[S] = P[S \bigcap H] + P[S \bigcap F]$$

et

$$P[S \bigcap F] = P[F][S/F] = 0.18$$

d'où

$$P[S] = 0.2 + 0.18 = 0.38 \qquad P[H/S] = 0.2/0.38 = 0.53$$

De la définition de la probabilité conditionnelle et de l'exemple précédent, on déduit aisément **la formule de Bayes**:

$$\boxed{P[B/A] = P[A/B]\frac{P[B]}{P[A]}}$$

La probabilité conditionnelle permet d'introduire la notion d'indépendance statistique.

Déf. B.2.5 *: indépendance*
A et B sont dits **indépendants** *si $P[B/A] = P[B]$ alors $P[B \bigcap A] = P[A]P[B]$.*

Remarques:

Deux évènements mutuellement exclusifs ne sont pas indépendants puisque:

$$P[B/A] = \frac{P[B \bigcap A]}{P[A]} = 0$$

B.2.3 Variables aléatoires discrètes et continues

Soit l'ensemble Ω, sur lequel on a défini une tribu $\mathcal{B}$ et une mesure de probabilité $P[\bullet]$, définissant ainsi un espace probabiliste $(\Omega, \mathcal{B}, P[\bullet])$. On définit également un ensemble discret qui peut être fini ou infini et dénombrable $\mathcal{C} = \{\alpha_1, \alpha_2, \cdots\}$.

Déf. B.2.6 *: variable aléatoire réelle continue, (discrète)*
A tout point $\omega \in \Omega$, on associe un nombre réel $x(\omega)$, par la fonction, appelée **variable aléatoire réelle, v.a.r.**, *$x \; : \; \Omega \rightarrow \mathbb{R}$, $(\mathcal{C})$ et possédant la propriété:*

Propriété B.2.3 *:*
L'ensemble des points $\omega \in \Omega$, tels que $x(\omega) \leq \alpha$, pour $\alpha \in \mathbb{R}$, est un élément de la tribu $\mathcal{B}$.

L'ensemble $\{\omega \in \Omega \ : \ x(\omega) \leq \alpha\}$ est un évènement pour lequel il est possible de définir une valeur de la probabilité associée.

Déf. B.2.7 *: fonction de répartition*
La fonction $F_x(\alpha) \ : \ I\!R \rightarrow [0 \quad 1]$ *telle que:*

$$\boxed{F_x(\alpha) = P[\{\omega \in \Omega \ : \ x(\omega) \leq \alpha\}]}$$

s'appelle **la fonction de répartition** *de la v.a.* x*.*

Propriété B.2.4 *:*

- $F_x(\alpha) \geq 0 \quad \forall \ \alpha \in I\!R$
- $F_x(\alpha)$ *est une fonction non décroissante.*
- $F_x(-\infty) = 0 \qquad F_x(+\infty) = 1$
- $P[\alpha_1 < x(\omega) \leq \alpha_2] = F(\alpha_2) - F(\alpha_1)$

Exemple: cas discret
On définit:

$$\Omega = \{1, 2, 3, 4, 5, 6\}$$

et une v.a.:

$$x \ : \ \Omega \rightarrow I\!R$$
$$x(i) = i \qquad \forall \ i \in \Omega$$

alors la fonction de répartition $F_x(\alpha)$ a l'allure suivante:

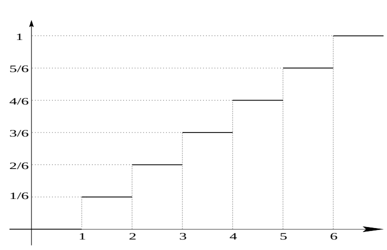


FIG. B.1 – *Exemple de fonction de répartition*

La v.a. x de fonction de répartition $F_x(\alpha)$ est ici discrète car elle prend ses valeurs dans un ensemble discret.

Exemple: cas continu

Soit l'expérience de lancer d'une fléchette sur une cible circulaire de rayon R et centré en 0. On suppose que le lanceur est suffisamment habile pour être toujours dans la cible. Soit x la v.a. associée à la distance de la fléchette à l'origine de la cible. On a alors

$$\begin{cases} \alpha < 0 & P[x \leq \alpha] = F_x(\alpha) = 0 \\ 0 \leq \alpha \leq R & P[x \leq \alpha] = F_x(\alpha) = \frac{\pi\alpha^2}{\pi R^2} = \frac{\alpha^2}{R^2} \\ R < \alpha & P[x \leq \alpha] = F_x(\alpha) = 1 \end{cases}$$

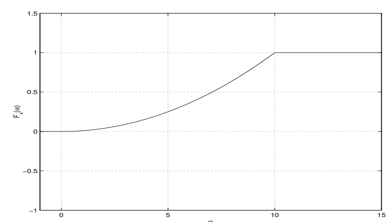


FIG. B.2 – *Exemple de fonction de répartition continue* $(R = 10)$

Nota:

Une v.a. continue est associée le plus souvent à des mesures de quantités physiques telles que coordonnées spatiales, poids, temps, température, tension, ... alors que les v.a. discrètes permettent plutôt de dénombrer des objets ou des évènements.

Déf. B.2.8 *:*

Une v.a.x est continue si sa fonction de répartition $F_x(\alpha)$ est une fonction continue. De plus,

$$P[x = \alpha] = 0 \qquad -\infty < \alpha < +\infty$$

Déf. B.2.9 *:*

Etant donnée une v.a. discrète $x(\omega)\ : \Omega \to \mathcal{C}$, la fonction:

$$\begin{array}{l} p_x\ :\ \Omega \to \mathcal{C} \\ p_x(\alpha) = P[\{x(\omega) = \alpha\}] \end{array}$$

est appelée **fonction densité de probabilité discrète de x**

Propriété B.2.5 *:*

- $p_x(\alpha) \geq 0 \quad \forall\ \alpha \in I\!R$.
- $\{\alpha\ :\ p_x(\alpha) \neq 0\}$ *est un sous ensemble fini ou infini dénombrable de* $I\!R$.
- $\sum_i p_x(\alpha_i) = 1$.

- $F_x(t) = \sum_{\alpha \leq t} p_x(\alpha) \quad -\infty < t < +\infty.$

Exemple: suite du jeu de dés

La fonction densité de probabilité de la v.a. discrète associée au jeu de dé précédent est représentée à la figure ci-dessous.

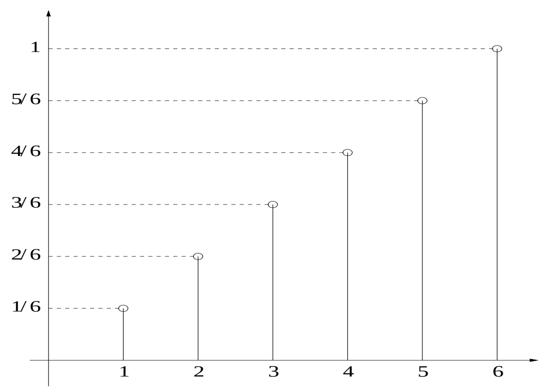


FIG. B.3 – *Exemple de fonction densité discrète*

Déf. B.2.10 *:*

Etant donnée une v.a. continue $x(\omega) : \Omega \rightarrow \mathbb{R}$, la fonction $p_x(\alpha)$ telle que:

$$\boxed{\int_{-\infty}^{+\infty} p_x(\xi) d\xi = 1}$$

est appelée **fonction densité de probabilité**.

On obtient immédiatement:

Propriété B.2.6 *:*

- $F_x(\alpha) = \int_{-\infty}^{\alpha} p_x(\xi) d\xi.$

- *En tout point où $F_x(\alpha)$ est dérivable, il vient $p_x(\alpha) = \frac{d}{d\alpha} F_x(\alpha)$.*

- $P[\alpha \leq x \leq \alpha + d\alpha] = p_x(\alpha) d\alpha.$

Exemple: suite du jeu de fléchette

La fonction densité de probabilité de la v.a. continue associée au jeu de fléchette précédent est représentée à la figure ci-dessous.

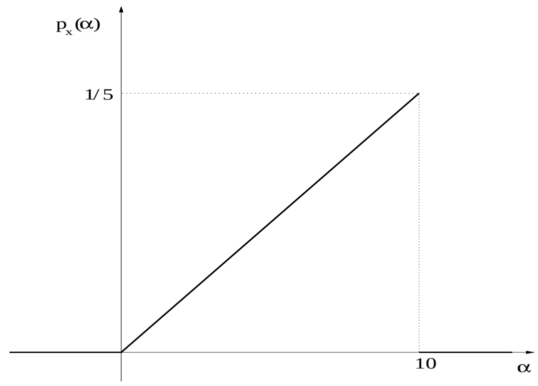


FIG. B.4 – *Exemple de fonction densité*

B.2.4 Propriétés statistiques des v.a.

Par souci de concision et sauf nécessité, il est fait abandon de la notation $x(\omega)$ au profit de x pour désigner la variable aléatoire.

Habituellement, lorsque l'on cherche la valeur moyenne d'une quantité aléatoire, on effectue N mesures et si l'on trouve n_i fois la quantité α_i, on calcule la moyenne par:

$$M = \frac{1}{N}\sum_i n_i x_i$$

Déf. B.2.11 *: moyenne*
Ainsi, par analogie, remplaçant la fréquence $\frac{n_i}{N}$ par $P_i = P_i[x = \alpha_i]$, on définira la moyenne ou moment du premier ordre par:

- *Cas discret*

$$\boxed{m_1 = \sum_i P_i \alpha_i}$$

- *Cas continu*

$$\boxed{m_1 = E[x] = \int_{-\infty}^{+\infty} \alpha p_x(\alpha) d\alpha}$$

L'opérateur $E[\bullet]$ est appelé **opérateur espérance mathématique**. Il représente une moyenne pondérée des valeurs possibles de la variable aléatoire; le poids associé à la i^{eme} valeur est sa probabilité P_i.

Propriété B.2.7 *:*
Soit y, variable aléatoire définie par $y = f(x)$. Sa moyenne peut être calculée de deux manières:

- *Cas discret:*

$$E[y] = \sum_\alpha f(\alpha) p_x(\alpha)$$

- *Cas continu:*

$$E[y] = \int_{-\infty}^{+\infty} \beta p_y(\beta) d\beta = \int_{-\infty}^{+\infty} f(\alpha) p_x(\alpha) d\alpha$$

- *Linéarité de l'opérateur espérance mathématique:*

$$\forall\ a, b \in I\!R \qquad E[ag(x) + bf(x)] = aE[g(x)] + bE[f(x)]$$

- *Soient x, y deux variables aléatoires indépendantes.*

$$E[xy] = E[x]E[y]$$

Déf. B.2.12 *: moment d'ordre k*
Le moment d'ordre k *est:*

- *Cas discret:*

$$\boxed{m_k = E[x^k] = \sum_{\alpha} \alpha^k p_x(\alpha)}$$

- *Cas continu:*

$$\boxed{m_k = E[x^k] = \int_{-\infty}^{+\infty} \alpha^k p_x(\alpha) d\alpha}$$

Déf. B.2.13 *: moment centré d'ordre k*
Le moment centré d'ordre k *est:*

- *Cas discret:*

$$\boxed{\mu_k = E[(x - E[x])^k] = \sum_{\alpha} (\alpha - E[x])^k p_x(\alpha)}$$

- *Cas continu:*

$$\boxed{\mu_k = E[(x - E[x])^k] = \int_{-\infty}^{+\infty} (\alpha - E[x])^k p_x(\alpha) d\alpha}$$

Noter que $E[x]$ est une constante, moyenne de la variable aléatoire x. En général, la connaissance du plus grand nombre de moments entraine une meilleure information sur la distribution de la variable aléatoire.

Déf. B.2.14 *: la variance*
Le moment centré d'ordre 2 s'appelle **la variance** *qui est définie par:*

$$\boxed{var(x) = E[(x - E[x])^2] = \sigma(x)^2}$$

où $\sigma(x)$ *est appelé* **l'écart-type**.

La variance et l'écart-type donnent une information sur la dispersion de la variable aléatoire autour de sa moyenne. Noter que $var(x) = E[x^2] - E^2[x]$.

Déf. B.2.15 *: la covariance*
Etant données deux variables aléatoires x et y, la quantité

$$cov(x, y) = E[(x - E[x])(y - E[y])]$$

s'appelle **la covariance** *de x et y. De plus,*

$$\boxed{var(x + y) = var(x) + var(y) + 2cov(x, y)}$$

Propriété B.2.8 *:*
Etant données deux variables aléatoires x et y indépendantes alors

$$cov(x, y) = 0 \qquad var(x + y) = var(x) + var(y)$$

Déf. B.2.16 *: coefficient de corrélation*
Etant données deux variables aléatoires x et y, la quantité

$$\boxed{\rho(x, y) = \frac{cov(x,y)}{\sqrt{var(x)var(y)}}}$$

s'appelle **le coefficient de corrélation** *de x et y.*

C'est une mesure du degré de dépendance entre x et y. Toutefois, $\rho(x, y)$ indique que les deux v.a. sont **décorrélées** mais cela ne signifie pas obligatoirement qu'elles sont indépendantes.

B.2.5 Exemples de lois de distribution

Dans cette section, nous présentons un exemple typique de loi de distribution qui sera particulièrement utilisée lors de ce cours.

Déf. B.2.17 *: loi normale*
La variable aléatoire associée x est dite **normale**, **gaussienne**, *ou* **normalement distribuée** *si sa fonction densité de probabilité s'écrit:*

$$\boxed{p_x(\alpha) = \frac{1}{\sqrt{2\pi\sigma^2}} exp\left[-\frac{1}{2}\frac{(\alpha-m)^2}{\sigma^2}\right]}$$

Elle dépend de deux paramètres m et σ^2 qui sont, en fait, respectivement la moyenne et la variance de la variable aléatoire:

$$\begin{cases} E[x] = m \\ var(x) = E[(x - m)^2] = \sigma^2 \end{cases}$$

Ces deux paramètres définissent complètement la loi de distribution.

Remarques:

- Pour $m = 0$, la loi normale est dite centrée.

- Pour $m = 0$, $\sigma = 1$, la loi normale est dite réduite, (standard).

$$\boxed{p_x(\alpha) = \frac{1}{\sqrt{2\pi}} e^{-\frac{\alpha^2}{2}}}$$

La fonction de répartition d'une loi normale ne peut être calculée facilement et doit donc être évaluée numériquement. Elle est donnée par:

$$F_x(\alpha) = P(x < \alpha) = \int_{-\infty}^{\alpha} \frac{1}{\sigma\sqrt{2\pi}} exp[-\frac{(t-m)^2}{2\sigma^2}]dt$$

qui par le changement de variables: $u = \frac{(t-m)}{\sigma}$ se ramène à:

$$F_x(x) = \int_{-\infty}^{\frac{(\alpha-m)}{\sigma}} \frac{1}{\sqrt{2\pi}} exp[-\frac{u^2}{2}]du$$

qui peut, elle-même, se ramener à l'intégrale évaluée à l'aide de tables:

$$Erf(\frac{x}{\sqrt{2}}) = \frac{1}{\sqrt{2\pi}} \int_{-x}^{+x} e^{-\frac{u^2}{2}} du$$

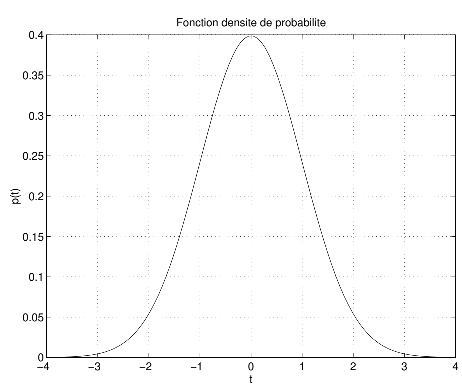


FIG. B.5 – *Loi normale standard*

La loi de distribution normale est en pratique très importante. Elle permet de modéliser un grand nombre de phénomènes aléatoires pratiques. Toutefois, dans la plus part des applications, les variables aléatoires ont une loi de distribution approximativement normale. De plus, le théorème de la limite centrale que nous ne présenterons pas dans ce cours établit que de nombreux phénomènes aléatoires complexes tendent à être bien représentés par une loi normale.
D'autres exemples de lois continues sont résumées dans le tableau de la page suivante.

Nom	expression de la densité
Lognormal	$\frac{1}{\alpha} exp[-ln^2(x\alpha/2)]$
Rayleigh	$\alpha exp[-\alpha^2/2]$
Maxwell	$\alpha^2 exp[-\alpha^2/2]$
Beta	$\alpha^b(1-\alpha)^c$
Gamma	$\alpha^n exp[-\alpha]$
Laplace	$exp[-\lvert\alpha\rvert]$
Cauchy	$\frac{1}{1+\alpha^2}$

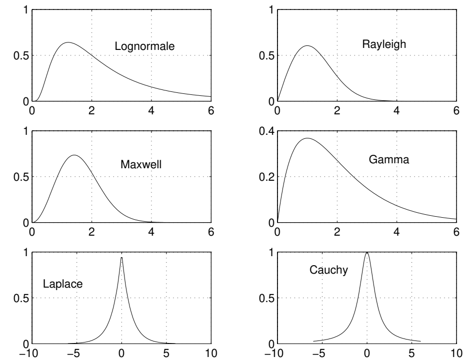


FIG. B.6 – *Quelques lois de distribution*

Déf. B.2.18 *: Loi de Poisson*

Une v.a. x est distribuée de manière poissonienne avec un paramètre a si elle prend les valeurs $0, 1, \cdots, n$ avec:

$$P[x = k] = e^{-a} \frac{a^k}{k!} \qquad k = 0, 1, \cdots$$

La densité de probabilité est alors donnée par:

$$p_x(\alpha) = e^{-a} \sum_{k=0}^{\infty} \frac{a^k}{k!} \delta(x - k)$$

Il est à noter que:

$$\frac{P[x = k - 1]}{P[x = k]} = \frac{k}{a}$$

B.3 Vecteurs aléatoires

Nous ne présenterons dans cette partie que le cas des vecteurs aléatoires continus, le cas discret pouvant s'en déduire aisément.

B.4 Loi de distribution conjointe

Déf. B.4.1 *: loi de distribution conjointe*
Des variables aléatoires $x_1, \ x_2, \ \cdots, x_n$ *ont* **une loi de distribution conjointe** *si elles sont définies par rapport au même univers* Ω*.*

Elles peuvent être caractérisées via:

- leur fonction de répartition conjointe:

$$\boxed{F_{x_1,\cdots,x_n}(\alpha_1,\cdots,\alpha_n) = P[x_1 < \alpha_1, \cdots, x_n < \alpha_n]}$$

qui est la probabilité de l'évènement défini comme l'ensemble des éléments ω qui satisfont:

$$x_1(\omega) < \alpha_1 \ et \ \cdots \ et \ x_n(\omega) < \alpha_n \Leftrightarrow \omega \in \{x_1(\omega) < \alpha_1\} \bigcap \cdots \bigcap \{x_n(\omega) < \alpha_n\}$$

- leur fonction densité de probabilité conjointe:

$$p_{x_1,\cdots,x_n}(\alpha_1,\cdots,\alpha_n)$$

définie par:

$$\boxed{F_{x_1,\cdots,x_n}(\alpha_1,\cdots,\alpha_n) = \int_{-\infty}^{\alpha_1} \cdots \int_{-\infty}^{\alpha_n} p_{x_1,\cdots,x_n}(\xi_1,\cdots,\xi_n) d\xi_1 \cdots d\xi_n}$$

c'est à dire en tous les points où la fonction de répartition est dérivable par rapport à tous ses arguments:

$$p_{x_1,\cdots,x_n}(\alpha_1,\cdots,\alpha_n) = \frac{\partial^n F_{x_1,\cdots,x_n}(\alpha_1,\cdots,\alpha_n)}{\partial\alpha_1 \cdots \partial\alpha_n}$$

En dimension 2, ces concepts sont facilement illustrés. La fonction de répartition est donc la probabilité pour que le point de coordonnées $(x_1(\omega), \ x_2(\omega))$ soit à l'intérieur du rectangle défini par $(\alpha_1, \ \alpha_2)$.
<u>Exemple:</u>
Une loi dont l'utilisation est particulièrement répandue est la loi uniforme.

Déf. B.4.2 *: loi uniforme*

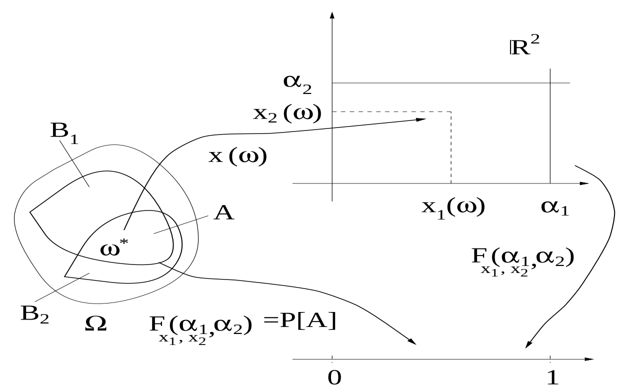


FIG. B.7 – *Distribution conjointe en deux dimensions*

On appelle **loi uniforme** *sur* $[a \ \ b]$, *une loi de probabilité définie par:*

$$F_x(\alpha) = \begin{cases} 0 \ pour \ \alpha < a \\ \\ \frac{\alpha-a}{b-a} \ pour \ a \leq \alpha < b \\ \\ 1 \ pour \ \alpha \geq b \end{cases}$$

ou

$$p_x(\alpha) = \begin{cases} \frac{1}{b-a} \ pour \ a \leq \alpha < b \\ \\ 0 \ ailleurs \end{cases}$$

Soient x et y deux variables aléatoires uniformément distribuées sur $[0 \ \ T] \times [0 \ \ T]$, alors:

$$P[x < T/2, y < T/2] = 1/4$$

$$p_{x,y}(\alpha, \beta) = \begin{cases} 1/T^2 \ \ sur \ [0 \ \ T] \times [0 \ \ T] \\ 0 \ \ ailleurs \end{cases}$$

Propriété B.4.1 *:*

- *La fonction de répartition est une fonction non décroissante de chacun de ses arguments:*

$$p_{x_1,\cdots,x_n}(\alpha_1, \cdots, \alpha_n) \geq 0$$

- *Ce qui suit est exprimé pour* $n = 2$ *mais à extension immédiate pour* $n > 2$.

  -

$$F(-\infty, \alpha_2) = F(\alpha_1, -\infty) = F(-\infty, -\infty) = 0$$

  -

$$F(+\infty, +\infty) = 1 \Leftrightarrow \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} p_{x_1,x_2}(\alpha_1, \alpha_2) d\alpha_1 d\alpha_2 = 1$$

  -

$$F_{x_1,x_2}(\alpha_{11}, \alpha_2) - F_{x_1,x_2}(\alpha_{12}, \alpha_2) = P[\alpha_{12} \leq x_1 < \alpha_{11}, x_2 < \alpha_2]$$

B.4.1 Loi de distribution marginale

Si l'on s'intéresse à un sous-ensemble $x_1, \cdots, x_m$, $m < n$ des variables aléatoires précédentes distribuées conjointement, on définit **la loi de distribution marginale** à partir de la distribution conjointe:

- la fonction de répartition marginale:

$$\boxed{F_{x_1,\cdots,x_m}(\alpha_1,\cdots,\alpha_m) = F_{x_1,\cdots,x_m,\cdots,x_n}(\alpha_1,\cdots,\alpha_m,+\infty,\cdots,+\infty)}$$

- La fonction densité de probabilité marginale:

$$\boxed{p_{x_1,\cdots,x_m}(\alpha_1,\cdots,\alpha_m) = \int_{-\infty}^{+\infty}\cdots\int_{-\infty}^{+\infty} p_{x_1,\cdots,x_n}(\alpha_1,\cdots,\alpha_m,\alpha_{m+1},\cdots,\alpha_n)d\alpha_{m+1}\cdots d\alpha_n}$$

- De plus,

$$p_x(\alpha) = \frac{\partial F(\alpha,\infty)}{\partial\alpha} \qquad p_y(\beta) = \frac{\partial F(\infty,\beta)}{\partial\beta}$$

<u>Exemple:</u>
Toujours sur le même exemple, les densités marginales sont:

$$p_x(\alpha) = \int_{-\infty}^{+\infty} p_{x,y}(\alpha,\beta)d\beta = \int_0^T (1/T^2)d\beta = 1/T \quad pour\ \alpha \in [0\ \ T]$$

$$p_x(\alpha) = 0 \quad pour\ \alpha \notin [0\ \ T]$$

On montre de la même manière que:

$$p_y(\beta) = 1/T \quad pour\ \beta \in [0\ \ T]$$

$$p_y(\beta) = 0 \quad pour\ \beta \notin [0\ \ T]$$

et donc, $p_{x,y}(\alpha,\beta) = p_x(\alpha)p_y(\beta)$, ce qui implique que x et y sont indépendantes.

B.4.2 Propriétés statistiques

Les moments sont bien sûr définis à partir des lois de distribution conjointes ou marginales précédentes et des fonctions associées.

- au premier ordre, (moyenne des variables aléatoires x_i):

$$\boxed{E[x_i] = \int_{-\infty}^{+\infty}\cdots\int_{-\infty}^{+\infty} \alpha_i p_{x_1,\cdots,x_n}(\alpha_1,\cdots,\alpha_n)d\alpha_1\cdots d\alpha_n \qquad i = 1,\cdots,n}$$

- ordre quelconque:

$$\boxed{E[x_1^{k_1}\cdots x_n^{k_n}] = \int_{-\infty}^{+\infty}\cdots\int_{-\infty}^{+\infty} \alpha_1^{k_1}\cdots\alpha_n^{k_n} p_{x_1,\cdots,x_n}(\alpha_1,\cdots,\alpha_n)d\alpha_1\cdots d\alpha_n}$$

- et les moments centrés correspondants:

$$\boxed{E[(x_1-m_1)^{k_1}\cdots(x_n-m_n)^{k_n}] = \int_{-\infty}^{+\infty}\cdots\int_{-\infty}^{+\infty} (\alpha_1-m_1)^{k_1}\cdots(\alpha_n-m_n)^{k_n} p_{x_1,\cdots x_n}(\alpha_1,\cdots,\alpha_n)d\alpha_1\cdots d\alpha_n}$$

Exemple: suite de l'exemple sur la loi uniforme

$$E[x] = E[y] = T/2$$

Notation matricielle:
Quand on traite des vecteurs aléatoires, il est plus pratique d'utiliser une notation matricielle permettant d'obtenir des écritures plus concises. Ainsi, notant:

$$X = [x_1\ x_2\ \cdots\ x_n]' \qquad \alpha = [\alpha_1\ \alpha_2\ \cdots\ \alpha_n]'$$

On définit alors:

$$F_{x_1,\cdots,x_n}(\alpha_1,\cdots,\alpha_n) = F_X(\alpha)$$

et

$$p_{x_1,\cdots,x_n}(\alpha_1,\cdots,\alpha_n) = p_X(\alpha)$$

De plus,

$$E[X] = [E[x_1]\ E[x_2]\ \cdots\ E[x_n]]' = [m_1\ m_2\ \cdots\ m_n]' = M_X$$

Déf. B.4.3 *: matrice de covariance*
La matrice de covariance Q_X *est la matrice dont les éléments* q_{ij} *sont égaux à* $cov(x_i, x_j)$*, soit:*

$$Q_X = E[(X - M_X)(X - M_X)']$$

Noter que Q_X est symétrique, semi-définie positive.

Déf. B.4.4 *: matrice de corrélation*
La matrice $C_X = E[XX']$ *est dite* **matrice de corrélation***, (matrice des moments d'ordre deux non centrés)*

C'est également une matrice symétrique, semi-définie positive et l'on a la relation suivante:

$$\boxed{Q_X = C_X - M_X M_X'}$$

Exemple:
Soient les deux variables aléatoires x, y définies par:

$$x = sin(2\pi t) \qquad y = cos(2\pi t)$$

où, t est uniformément distribuée sur $[0\ \ 1]$. On peut calculer:

$$E[x] = E[y] = E[xy] = 0$$

donc $cov(x, y) = 0$.

B.4.3 Un théorème important

Soit $X = [x_1, \cdots, x_n]'$ un vecteur aléatoire dont la fonction densité de probabilité conjointe est connue:

$$p_{x_1,x_2,\cdots,x_n}(\alpha_1, \alpha_2, \cdots, \alpha_n) = p_X(\alpha)$$

Soit $Y = f(X)$ où f est une application de $\mathbb{R}^n \to \mathbb{R}^n$ inversible et continûment différentiable. Le théorème suivant permet d'exprimer la fonction densité de probabilité de Y en fonction de celle de X.

Théorème B.4.1 *:*

Sous les hypothèses précédentes concernant f et f^{-1}

$$\boxed{p_Y(\beta) = p_X[f^{-1}(\beta)]|det(J)|}$$

où J est la matrice jacobienne de f^{-1} :

$$J = [\frac{\partial f^{-1}(\beta)}{\partial \beta}]$$

<u>Exemple:</u>

Soit $y = x + z$ où $x,\ z$ sont deux variables aléatoires indépendantes. On cherche la densité de probabilité de y connaissant celles de x et z. Posant:

$$Y = \begin{bmatrix} y \\ z \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} X$$

où $X = \begin{bmatrix} x \\ z \end{bmatrix}$, alors $A^{-1} = \begin{bmatrix} 1 & -1 \\ 0 & 1 \end{bmatrix}$ et l'on peut écrire:

$$p_{y,z}(\alpha, \beta) = p_{x,z}(\alpha - \beta, \beta)$$

d'où:

$$p_y(\alpha) = \int_{-\infty}^{+\infty} p_x(\alpha - \beta) p_z(\beta) d\beta$$

B.5 Distributions conditionnelles

B.5.1 Définition

Il a été vu précédemment qu'étant donnés deux évènements $A_1,\ A_2$ définis en relation avec une expérience aléatoire caractérisée par le triplet $\{\Omega, \mathcal{B}, P[\bullet]\}$, on détermine la probabilité conditionnelle de A_1 par rapport à A_2 par:

$$P[A_1/A_2] = \frac{P[A1 \bigcap A_2]}{P[A_2]}$$

$P[A_1/A_2]$ satisfait tous les axiomes des probabilités de telle sorte que l'on peut caractériser une expérience aléatoire par $(\Omega, \mathcal{B}, P[\bullet/A_2])$ où l'on s'intéresse tout particulièrement aux évènements pour lesquels A_2 est réalisé.

Exemple:
Reprenons l'exemple du jeu de dé avec $\Omega = \{1,2,3,4,5,6\}$.
Soit l'évènement $A = \{2,4,6\} = \{Tirage\ pair\}$, alors nous pouvons calculer les probabilités conditionnelles suivantes:

$$P[1/A] = P[3/A] = P[5/A] = 0 \qquad P[2/A] = P[4/A] = P[6/A] = 1/3$$

Soient deux variables aléatoires x, y distribuées conjointement sur l'univers Ω. Etant données la distribution conjointe et la densité de x et y, on recherche la distribution et la densité de x étant donnée une réalisation β de y.
Soient les évènements:

$$A_1 = \{\omega \in \Omega \ : \ x(\omega) < \alpha\} \qquad A_2 = \{\omega \in \Omega \ : \ \beta \leq y(\omega) < \beta + d\beta\}$$

ainsi $P[A_1/A_2]$ représente la fonction de répartition de x conditionnellement à A_2 et notée:

$$F_{x/y}(\alpha/\beta) = \frac{P[x < \alpha\ ,\ \beta \leq y < \beta + d\beta]}{P[\beta \leq y < \beta + d\beta]} = \frac{F_{x,y}[\alpha\ ,\ \beta + d\beta] - F_{x,y}[\alpha\ ,\ \beta]}{F_y[\beta + d\beta] - F_y[\beta]}$$

Par passage à la limite, ($d\beta \to 0$), et sous l'hypothèse de dérivabilité des fonctions de répartition au point considéré, il vient:

$$\boxed{F_{x/y}(\alpha/\beta) = \frac{\partial F_{x,y}(\alpha,\beta)/\partial\beta}{\partial F_y(\beta)/\partial\beta} = \frac{\displaystyle\int_{-\infty}^{\alpha} p_{x,y}(\xi,\beta)d\xi}{p_y(\beta)}}$$

qui est donc la fonction de répartition de la variable aléatoire x conditionnellement à la réalisation β de la variable aléatoire y. La dérivée par rapport à α, fournit la fonction densité de probabilité conditionnelle:

$$\boxed{p_{x/y}(\alpha/\beta) = \frac{p_{x,y}(\alpha,\beta)}{p_y(\beta)}}$$

Ces notions de distributions conditionnelles sont très importantes car elles introduisent la notion de dépendance stochastique entre deux variables aléatoires.

B.5.2 La règle de Bayes

Ainsi, par raison de symétrie,

$$p_{x,y}(\alpha,\beta) = p_{x/y}(\alpha/\beta)p_y(\beta) = p_{y/x}(\beta/\alpha)p_x(\alpha)$$

d'où l'on tire la règle de Bayes:

$$\boxed{p_{x/y}(\alpha/\beta) = \frac{p_{y/x}(\beta/\alpha)p_x(\alpha)}{p_y(\beta)}}$$

B.5.3 Propriétés statistiques

L'opérateur $E[x/y]$ est l'espérance mathématique conditionnelle définie, classiquement, à partir de la fonction densité de probabilité conditionnelle par:

$$\boxed{E[x/y] = \int_{-\infty}^{+\infty} \alpha p_{x/y}(\alpha/\beta) d\alpha}$$

L'opérateur espérance mathématique conditionnelle permet la définition des moments conditionnels et notamment:

- l'espérance conditionnelle:

$$\boxed{m_{x/y} = E[x/y]}$$

- la covariance conditionnelle:

$$\boxed{cov(x_i, x_j/\beta) = E[(x_i - m_{x_i/y})(x_j - m_{x_j/y})/\beta]}$$

Il est à noter que $p_{x/y}(\alpha/\beta)$ dépend des réalisations de y; c'est donc elle-même une variable aléatoire. Ceci est également vrai pour la moyenne conditionnelle et tout moment conditionnel. On peut ainsi appliquer à ces diverses variables aléatoires, l'opérateur espérance mathématique qui consiste à calculer la moyenne de ces variables aléatoires sur toutes les réalisations β de la variable aléatoire y.
Ainsi:

$$\boxed{\begin{array}{c} E[p_{x/y}(\alpha/\beta)] = p_x(\alpha) \\ \\ E[E[x/y]] = E(x) \end{array}}$$

Développés pour le cas scalaire, les résultats et définitions ci-dessus s'étendent directement au cas multidimensionnel, où x, y sont des vecteurs.

B.5.4 Propriété d'indépendance

Des variables aléatoires, (ou vecteurs), à distribution conjointe, sont dites indépendantes si l'une quelconque des propriétés suivantes est satisfaite:

$$\boxed{\begin{array}{l} F_{x,y}(\alpha, \beta) = F_x(\alpha) F_y(\beta) \\ \\ p_{x,y}(\alpha, \beta) = p_x(\alpha) p_y(\beta) \\ \\ E[f(x)g(y)] = E[f(x)]E[g(y)] \end{array}}$$

Dans ce cas,

$$\boxed{p_{x/y}(\alpha/\beta) = p_x(\alpha)}$$

<u>Nota:</u> Deux variables aléatoires indépendantes sont décorrélées alors que l'inverse n'est pas forcément vrai.

$$\boxed{\text{Indépendance} \Rightarrow \text{décorrélation.}}$$

B.6 Vecteurs gaussiens

B.6.1 Définition

Soit Z un vecteur de variables aléatoires décomposable en deux sous-vecteurs $X \in \mathbb{R}^n$, $Y \in \mathbb{R}^m$, $Z' = [X' \quad Y']$. Z est gaussien, ses composantes z_i, $i = 1, 2, \cdots n + m$ sont conjointement distribuées suivant une loi normale ou gaussienne.
Définissant les deux premiers moments m_Z et Q_Z, la fonction densité de probabilité s'écrit:

$$\boxed{p_Z(\gamma) = \left[(2\pi)^{\frac{m+n}{2}}(det(Q_Z))^{1/2}\right]^{-1} e^{\left[-1/2(\gamma - m_Z)' Q_Z^{-1}(\gamma - m_Z)\right]}}$$

Correspondant à la partition de Z en X et Y, m_X et m_Y désignent les vecteurs espérance mathématique, (moyenne), de X et Y respectivement, la matrice de covariance Q_Z se partitionne en:

$$Q_Z = \begin{bmatrix} Q_X & Q_{XY} \\ Q_{YX} & Q_Y \end{bmatrix} \qquad Q_Z > 0$$

où:

$$Q_X = E[(X - m_X)(X - m_X)'] \quad Q_Y = E[(Y - m_Y)(Y - m_Y)']$$

$$Q_{XY} = E[(X - m_X)(Y - m_Y)']$$

Important: Les deux premiers moments définissent entièrement la loi de distribution normale ou gaussienne.
Notation:

$$Z \sim \mathcal{N}[m_Z, Q_Z]$$

B.6.2 Propriétés

Théorème B.6.1 *:*
Si Z est gaussien alors X et Y sont également gaussiens.

Démonstration:
Si X et Y sont décorrélés $\Leftrightarrow$ $cov(X, Y) = E[(X - m_X)(Y - m_Y)'] = 0$ alors, il vient aisément:

$$p_{X,Y}(\alpha, \beta) = p_X(\alpha) p_Y(\beta)$$

c'est à dire, X et Y sont indépendantes.

Théorème B.6.2 *:*
Deux vecteurs gaussiens non corrélés sont indépendants.

De plus,

Théorème B.6.3 *:*
Si Z est gaussien, $W = CZ + A$ est également gaussien tel que:

$$\boxed{W \sim \mathcal{N}[A + Cm_Z, CQ_ZC']}$$

A et C sont respectivement un vecteur et une matrice de dimensions appropriées et d'éléments déterministes.

Théorème B.6.4 *:*
Si X et Y sont deux vecteurs conjointement normalement distribués, alors la fonction densité de probabilité conditionnelle $p_{X/Y}$ est elle-même normale et définie par les deux paramètres

- *moyenne conditionnelle* $E[X/Y] = m_X + Q_{XY}Q_Y^{-1}(\beta - m_Y)$
- *covariance conditionnelle* $Q_{X/Y} = Q_X - Q_{XY}Q_Y^{-1}Q_{YX}$

Il est à remarquer que la matrice de covariance conditionnelle ne dépend pas de β, ceci est une propriété importante du cas gaussien, propriété qui se concrétisera par la suite dans le problème du filtrage des systèmes linéaires excités par des bruits gaussiens.
Un autre résultat important est contenu dans la proposition suivante.

Proposition B.6.1 *:*
$V = X - E[X/Y]$ *est un vecteur aléatoire gaussien indépendant de* Y*.*

Démonstration: V apparait comme une combinaison linéaire de vecteurs gaussiens, c'est donc un vecteur gaussien. Pour l'indépendance vis à vis de Y, il suffit de démontrer que V et Y sont décorrélés :

$$\begin{aligned} cov(V,Y) &= E[(X - E[X/Y])(Y - E[Y])'] \\ &= E[(X - E[X] - Q_{XY}Q_Y^{-1}(Y - E[Y])(Y - E[Y])'] \\ &= Q_{XY} - Q_{XY}Q_Y^{-1}Q_Y = 0 \end{aligned}$$

B.7 Introduction à l'étude des processus stochastiques

B.8 Introduction et définition

La notion de processus stochastique est une notion fondamentale qui va servir d'une part à la définition de modèles pour les perturbations, les phénomènes de bruit mais également à la définition des équations aux différences stochastiques et équations différentielles stochastiques. Les solutions de ces équations stochastiques définiront également des processus stochastiques possédant des propriétés particulières.

Déf. B.8.1 *: processus stochastique*

Un processus stochastique, *(scalaire ou vectoriel), est défini comme une famille de variables ou vecteurs aléatoires indexée par un ensemble de paramètres $t \in T$, que l'on va considérer dans la suite du cours comme étant le temps. La notation est $\{x_t(\omega) \mid t \in T\}$. T peut être un ensemble discret ou continu.*

C'est donc une fonction de deux paramètres: le temps et ω paramètre aléatoire, (résultat d'une expérience).

- Pour chaque t, $x_t(\bullet)$ est une variable aléatoire égale à l'état du processus considéré à l'instant t.

- Pour ω fixé , $x_\bullet(\omega)$ est **une réalisation du processus** qui est une fonction du temps.

- Pour t et ω fixés, $x_t(\omega)$ est un nombre.

Si la variable, (vecteur), aléatoire prend ses valeurs dans un espace discret, on dira que c'est **une chaine**, ou **un processus stochastique à état discret**, qui peut être défini en temps discret ou en temps continu. Dans le cas contraire, le processus sera défini comme **un processus stochastique à état continu** en temps discret ou continu. Dans le cadre de ce cours, nous nous intéresserons exclusivement aux processus à état continu qui, en temps discret sont appelés **séquences aléatoires** et dans le cas continu plus simplement **processus stochastiques** si aucune confusion n'est possible. La notation est alors simplifiée en $x(t)$ pour un processus en temps continu et x_t pour une séquence.

Exemple:

Soit un satellite en orbite périodique autour de la terre pour lequel on ne connait pas toutes les actions définissant son mouvement que l'on modélise de manière probabiliste. Il est possible de définir le mouvement du satellite par une équation différentielle stochastique. L'altitude $x(t)$ du satellite, qui est une partie de la solution de cette équation différentielle est un processus stochastique à état continu et en temps continu.

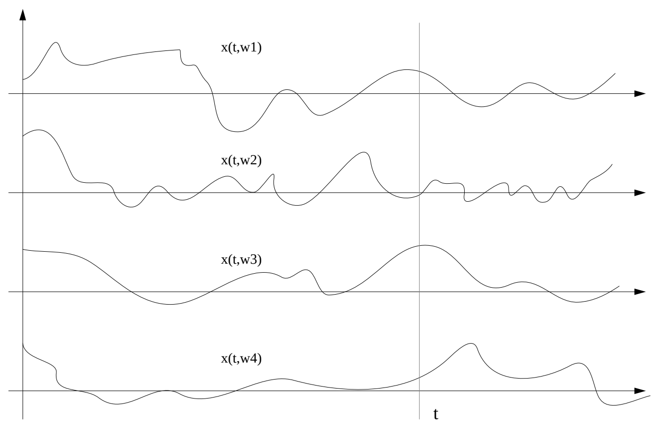


FIG. B.8 – *réalisations de l'altitude d'un satellite*

B.9 Caractérisations probabilistes des processus stochastiques

B.9.1 Loi de distribution

Soit $\{x_t(\omega) \mid t \in T\}$ un processus stochastique. T peut être continu ou discret. Pour n'importe quelle partition $\{t_1, \cdots, t_n\}$ de T, qui peut être obtenue par discrétisation dans le cas continu, la loi de distribution conjointe des variables, (vecteurs), aléatoires $x_{t_1}, \cdots, x_{t_n}$ est appelée **loi de distribution finie** du processus. Le processus stochastique peut être caractérisé par la définition de la loi de distribution finie pour tous les ensembles finis $\{t_i\} \in T$, donnée de manière équivalente par:

- La fonction de répartition conjointe:

$$F_{x(t_1),\cdots,x(t_n)}(\alpha_1, \cdots, \alpha_n)$$

- La fonction densité de probabilité:

$$p_{x(t_1),\cdots,x(t_n)}(\alpha_1, \cdots, \alpha_n)$$

pour toutes les partitions $\{t_i\} \in T$.
Spécifier l'une des deux fonctions précédentes $\forall$ $\{t_i\} \in T$, consiste à définir **la loi de probabilité** du processus stochastique. Dans le cas discret, il y a un nombre fini de lois de distribution finies alors que dans le cas continu, il y en a un nombre infini.

Exemple:
Un processus stochastique $x(t)$ peut être défini par:

$$x(t) = a + bt$$

où a, b sont des variables aléatoires de distribution connue. On suppose de plus que a, b sont deux variables aléatoires conjointement normalement distribuées. Puisque l'on peut écrire:

$$\begin{bmatrix} x(t_1) \\ \vdots \\ x(t_n) \end{bmatrix} = \begin{bmatrix} 1 & t_1 \\ \vdots & \vdots \\ 1 & t_n \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix}$$

on peut dire que le vecteur aléatoire $[x(t_1), \cdots, x(t_n)]'$ est un vecteur gaussien dont la moyenne et la covariance peuvent être données à partir de la moyenne et de la covariance de $[a \quad b]'$.

B.9.2 Processus du second ordre

Soit un processus stochastique $\{x_t(\omega) \mid t \in T\}$. Bien que non suffisantes dans le cas le plus général, la spécification des densités de probabilité du **premier ordre** et du **deuxième ordre** est souvent intéressante:

- Densité du premier ordre:

$$\boxed{p_{x(t)}(\alpha) = p(\alpha, t)}$$

- Densité du deuxième ordre:

$$\boxed{p_{x(t),x(\tau)}(\alpha, \beta) = p(\alpha, \beta, t, \tau)}$$

Ces fonctions ne suffisent, en général, pas à la caractérisation complète des processus stochastiques mais il existe certaines classes de processsus, (gaussiens - markoviens), pour lesquels la loi de distribution est entièrement déterminée par la connaissance des fonctions densité de probabilité, au second ordre.

Déf. B.9.1 *: processus du second ordre*
Un processus stochastique caractérisé entièrement par ses lois de distributions au premier et deuxième ordre est appelé **processus du second ordre.**

B.9.3 Caractéristiques statistiques

A partir de ces fonctions densité de probabilité, peuvent être définis les moments du processus stochastique.

Déf. B.9.2 *: moyenne*
Soit un processus stochastique $\{x_t(\omega) \mid t \in T\}$ noté $x(t)$, alors **la moyenne de** $x(t)$ *est définie par:*

$$\boxed{m_x(t) = E[x(t)] = \int_{-\infty}^{+\infty} \alpha p(\alpha, t) d\alpha}$$

C'est une fonction du paramètre t.

Déf. B.9.3 *: matrice d'autocorrélation et d'autocovariance*
On peut définir **la matrice d'autocorrélation**

$$\boxed{R(t,\tau) = E[x(t)x(\tau)']}$$

dont les éléments sont:

$$r_{ij}(t,\tau) = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} \alpha_i\beta_j p[\alpha,t,\beta,\tau]d\alpha d\beta$$

et la matrice d'autocovariance:

$$\boxed{P(t,\tau) = E[(x(t)-m_x(t))(x(\tau)-m_x(\tau))']}$$

On vérifie aisément les propriétés suivantes:

Propriété B.9.1 *:*

- $P(t,\tau) = P(\tau,t) \quad \forall\ t,\tau$
- $R(t,\tau) = R(\tau,t) \quad \forall\ t,\tau$
- $R(t,t) \geq 0 \quad \forall\ t$
- $Q(t) = P(t,t) \geq 0 \quad \forall\ t$
- $P(t,\tau) = R(t,\tau) - m_x(t)m_x(\tau)' \quad \forall\ t,\tau$

Le préfixe "auto", souvent omis, n'est utilisé que lorsqu'il peut y avoir confusion avec l'intercorrélation et l'intercovariance, définies, par rapport à deux processus stochastiques distincts $x(t)$ et $y(t)$ par:

$$\Gamma_{xy}(t,\tau) = E[x(t)y(\tau)']$$

pour l'intercorrélation, et:

$$C_{xy}(t,\tau) = E[(x(t)-m_x(t))(y(\tau)-m_y(\tau))']$$

pour l'intercovariance.
La moyenne et la matrice de corrélation sont généralement plus faciles à manipuler que la loi de probabilité et n'en donne pas moins de nombreuses informations sur le processus.

<u>Exemple:</u>
Soit le processus stochastique:

$$x(t) = a + bt$$

de l'exemple précédent. Les moments de ce processus sont donnés par:

$$m_x(t) = E[a] + tE[b]$$

$$r(t,\tau) = E[a^2] + (t+\tau)E[ab] + t\tau E[b^2]$$

$$p(t,\tau) = var(a) + cov(a,b)(t+\tau) + var(b)t\tau$$

On peut noter que ces moments ne dépendent que des moments de a et b.

B.10 Caractéristiques importantes

La stationnarité fixe la propriété d'invariance par rapport au temps t des lois de probabilité qui caractérisent le processus stochastique.

B.10.1 Stationnarité au sens strict

Déf. B.10.1 *: stationnarité au sens strict*
$x(t)$ est dit **stationnaire au sens strict**, *sur T intervalle de définition, si pour toute partition $\{t_1, t_2, \cdots, t_n\}$ de T:*

$$\forall\ n,\ \forall\ \theta,\quad p_{x(t_1),\cdots,x(t_n)}(\alpha_1, \cdots, \alpha_n) = p_{x(t_1+\theta),\cdots,x(t_n+\theta)}(\alpha_1, \cdots, \alpha_n)$$

B.10.2 Stationnarité au sens large

La stationnarité au sens large, est une notion plus pratique qui est définie à l'ordre deux et, généralement, à partir des deux premiers moments du processus.

Déf. B.10.2 *: stationnarité au sens large*
$x(t)$ est un processus stochastique **stationnaire au sens large** *si*

$$\begin{cases} m_x(t) = m_x \quad (\textit{indépendant du temps}) \\ \\ P(t, \tau) = P(t - \tau) \end{cases}$$

Dans la condition précédente, la covariance peut être de façon équivalente remplacée par la corrélation.

Exemple:

Le processus

$$x(t) = acos(2\pi t) + bsin(2\pi t)$$

où $a,\ b$ sont des variables aléatoires non corrélées de moyenne nulle et de variance unité est stationnaire au sens large:

$$m(t) = 0$$

$$r(t + \tau, t) = cos(2\pi\tau)$$

B.10.3 Ergodicité

On a défini précédemment des moyennes statistiques, définies à des instants donnés. Il est également intéressant de définir des moyennes temporelles.

Déf. B.10.3 *: moyenne temporelle*
On définit la **moyenne temporelle** *d'un échantillon du processus par la relation:*

$$< x(t) >= \lim_{T \to +\infty} \frac{1}{2T} \int_{-T}^{T} x(t)dt$$

lorsque cette limite existe.

Nota: D'autres moyennes temporelles peuvent être définies.

$$< x^2(t) >= \lim_{T\to+\infty} \frac{1}{2T} \int_{-T}^{T} x^2(t)dt$$

La limite ainsi définie peut ne pas exister pour certains échantillons ou dépendre de l'échantillon, mais elle ne dépend pas du temps. Toutefois, il existe des cas où la moyenne temporelle et la moyenne statistique sont égales.

Déf. B.10.4 *: ergodicité*

Un processus stochastique $x(t)$ est **ergodique** *si toutes les moyennes temporelles existent et ont même valeur quelque soit l'échantillon.*

$$\forall f, \quad \lim_{t_1\to\infty} [\frac{1}{t_1 - t_0} \int_{t_0}^{t_1} f[x(t)]dt] = E[f(x(t))] \tag{B.1}$$

Cette notion d'ergodicité est très importante du fait que pratiquement, pour évaluer les moyennes statistiques, l'on ne dispose généralement que d'un échantillon sur lequel on fait une moyenne temporelle. Ce procédé n'a de valeur que si le processus stochastique est stationnaire et ergodique.

B.10.4 La densité spectrale de puissance

En théorie du signal, les spectres sont associés à la transformée de Fourier. Dans le cas de signaux déterministes, c'est une façon de représenter le signaux comme une superposition d'exponentielles. Pour les signaux aléatoires, il s'agit principalement d'étudier la transformée de moyennes.

Déf. B.10.5 *: puissance*

On définit **la puissance moyenne** *d'un processus stochastique stationnaire au sens large, par:*

- *Cas continu $x(t)$:*

$$\boxed{P = E[x'(t)x(t)]}$$

- *Cas discret x_N:*

$$\boxed{P = \lim_{N\to+\infty} \frac{1}{2N+1} \sum_{-N}^{N} x_N^2}$$

Propriété B.10.1 *:*

$$P = Trace[R_x(0)]$$

Soit un processus stochastique $x(t)$ stationnaire au sens large et possédant une matrice de covariance $P_x(t-\tau)$.

Déf. B.10.6 *: densité spectrale de puissance*

La densité spectrale de puissance $\Psi_x(\omega)$ *du processus stochastique* $x(t)$ *est définie comme la transformation de Fourier, si elle existe, de la matrice de covariance* $P_x(t-\tau)$*:*

$$\boxed{\Psi_x(\omega) = \int_{-\infty}^{\infty} e^{-j\omega\tau} P_x(\tau) d\tau}$$

Réciproquement,

$$\boxed{P_x(\tau) = \tfrac{1}{2\pi} \int_{-\infty}^{\infty} e^{-j\omega\tau} \Psi_x(\omega) d\omega}$$

$\Psi_x(\omega)$ est une matrices complexe possédant les propriétés:

Propriété B.10.2 *:*

- $\Psi_x(-\omega) = \Psi_x(\omega)' \quad \forall\ \omega$
- $\Psi_x^*(\omega) = \Psi_x(\omega) \quad \forall\ \omega$
- $\Psi_x(\omega) \geq 0 \quad \forall\ \omega$

Exemple:

Soit le processus stochastique $v(t)$ stationnaire au sens large de matrice de covariance:

$$P_v(t-\tau) = \sigma^2 exp[-\frac{|t-\tau|}{\theta}] \qquad \theta > 0$$

Il est facile de montrer par le calcul que:

$$\Psi_v(\omega) = \frac{2\sigma^2\theta}{1+\omega^2\theta^2}$$

Nous donnons quelques exemples de densité spectrale de puissance de processus stochatiques définis par leur matrice de corrélation.

$R_x(\tau)$	$\Psi_x(\omega)$
$\delta(\tau)$	$1 - 2\pi m_x^2 \delta(\omega)$
$e^{j\beta\tau}$	$2\pi(\delta(\omega-\beta) - m_x^2\delta(\omega))$
$e^{-\alpha\|\tau\|}$	$\frac{2\alpha}{\alpha^2+\omega^2} - 2\pi m_x^2\delta(\omega)$
$e^{-\alpha\|\tau\|}cos(\beta\tau)$	$\frac{\alpha}{\alpha^2+(\omega-\beta)^2} + \frac{\alpha}{\alpha^2+(\omega+\beta)^2} - 2\pi m_x^2\delta(\omega)$
1	$2\pi(1-m_x^2)\delta(\omega)$
$cos(\beta\tau)$	$\pi\delta(\omega+\beta) + \pi\delta(\omega-\beta) - 2\pi m_x^2\delta(\omega)$

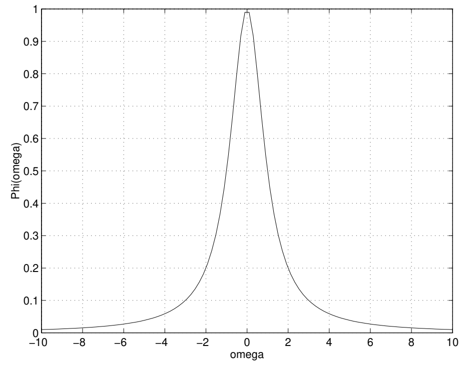


FIG. B.9 – *Spectre de $v(t)$*

L'introduction de la densité spectrale de puissance peut se justifier à travers le résultat suivant.

Théorème B.10.1 *:*
Soit le processus stochastique $x(t)$ alors pour toute matrice symétrique $W(t)$:

$$E[x(t)'W(t)x(t)] = Trace[W(t)R_x(t,t)]$$

Si, de plus, $x(t)$ est stationnaire au sens large de moyenne nulle et W est constante alors:

$$E[x(t)'Wx(t)] = Trace[WP_x(0)] = (1/2\pi)Trace[\int_{-\infty}^{\infty} W\Psi_x(\omega)d\omega]$$

Ce résultat permet de donner une interprétation au terme densité spectrale de puissance. En effet, la "puissance" totale $E[x(t)'Wx(t)]$ d'un processus stochastique de moyenne nulle et stationnaire au sens large est obtenue par intégration de $Trace[W\Psi_x(\omega)]$ sur toutes les fréquences. Ainsi, $trace[W\Psi_x(\omega)]$ peut être considérée comme une mesure de la densité de puissance à la fréquence ω. la matrice W est une matrice de pondération qui détermine la contribution des divers composantes de $x(t)$ à la puissance. La densité spectrale de puissance représente la répartition harmonique de la puissance moyenne P de $x(t)$.

B.10.5 Indépendance et décorrélation

Il est possible d'étendre les notions d'indépendance et de décorrélation aux processus stochastiques.

Déf. B.10.7 *: indépendance*
Soient $v_1(t)$, $v_2(t)$ deux processus stochastiques. Alors $v_1(t)$, $v_2(t)$ sont dits **indépendants** *si $\{v_1(t_1), v_1(t_2), \cdots, v_1(t_l)\}$ et $\{v_2(t'_1), v_2(t'_2), \cdots, v_2(t_m)\}$ sont des ensembles indépendants de vecteurs aléatoires pour tous $t_1,\ t_2, \cdots, t_l, t'_1, \cdots, t_m$ et pour tous m et l.*

De même,

Déf. B.10.8 *: décorrélation*
$v_1(t)$, $v_2(t)$ sont dits **décorrélés** *si $v_1(t_1)$ et $v_2(t_2)$ sont des vecteurs aléatoires décorrélés pour tous t_1, t_2.*

Soient deux processus stochastiques $x(t)$ et $y(t)$.

Déf. B.10.9 *: Matrice d'intercorrélation et d'intercovariance*
On définit

- **la matrice d'intercorrélation** *par:*

$$\boxed{R_{xy}(t_1, t_2) = E[x(t_1)y'(t_2)]}$$

- **la matrice d'intercovariance** *par:*

$$\boxed{P_{xy}(t_1, t_2) = R_{xy}(t_1, t_2) - m_x(t_1)m'_y(t_2)}$$

Une notion importante liée à ces concepts est l'orthogonalité.

Déf. B.10.10 *: orthogonalité*
$x(t)$, $y(t)$ *sont deux processus stochastiques* **orthogonaux** *si*

$$R_{xy}(t_1, t_2) = 0 \qquad \forall\ t_1,\ t_2$$

De plus,

Déf. B.10.11 *: décorrélation*
$x(t)$, $y(t)$ *sont deux processus stochastiques* **décorrélés** *si*

$$P_{xy}(t_1, t_2) = 0 \qquad \forall\ t_1,\ t_2$$

B.11 Processus stochastiques remarquables

B.11.1 Processus gaussien

Les processus gaussiens sont des processus très importants du fait qu'on les rencontre très souvent en pratique et du fait que de nombreux processus physiques sont approximativement gaussiens. Ceci est la conséquence du théorème de la limite centrale. Sans entrer dans les détails, le théorème de la limite centrale établit que si l'on considère la somme:

$$x = x_1 + \cdots + x_n$$

où x_i sont des variables aléatoires indépendantes, alors sous certaines conditions, la distribution de probabilité de x approche la distribution gaussienne quand $n \to \infty$.

Déf. B.11.1 *: processus gaussien*
Le processus stochastique $x(t)$ *est* **gaussien** *si pour toute partition* $\{t_1, t_2, \cdots, t_n\}$ *de* T*, le vecteur des variables aléatoires* $[x(t_1), \cdots, x(t_n)]$ *est gaussien.*

La loi de distribution d'un vecteur gaussien est entièrement déterminée par la connaissance des moments du premier ordre, (moyenne et covariance), qui, eux-mêmes, ne nécessitent pour leur calcul que les fonctions densité de probabilité du premier et du second ordre $p_x(\alpha, t)$, $p_x(\alpha, t, \beta, \tau)$. Il s'agit donc bien d'un processus du second ordre. Noter que la fonction du premier ordre peut être calculée à partir de celle du second ordre par la fonction marginale:

$$p_x(\alpha, t) = \int_{-\infty}^{+\infty} p_x(\alpha, t, \beta, \tau) d\beta$$

Propriété B.11.1 *:*
Un processus gaussien stationnaire au sens large l'est au sens strict.

Exemple:
Soit le processus défini par:

$$x(t) = a + bt$$

où a, b sont des variables aléatoires gaussiennes, non corrélées, de moyenne nulle et de variance unité. Alors, réécrivant

$$\begin{bmatrix} x(t_1) \\ \vdots \\ x(t_n) \end{bmatrix} = \begin{bmatrix} 1 & t_1 \\ \vdots & \vdots \\ 1 & t_n \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix}$$

on peut trouver que:

$$E[x(t_1), \cdots, x(t_n)]' = 0$$

$$P = \begin{bmatrix} 1 + t_1^2 & \cdots & 1 + t_1 t_n \\ \vdots & \vdots & \vdots \\ 1 + t_n t_1 & \cdots & 1 + t_n^2 \end{bmatrix}$$

Le processus est de plus gaussien.

De par le théorème de la limite centrale dont nous avons parlé en début de section, un processus gaussien est souvent un bon modèle pour le bruit dans les circuits physiques tels que les transmetteurs et récepteurs radios, les radars etc.

B.11.2 Processus markovien

Cette classe de processus stochastiques est importante du fait qu'elle généralise à l'univers stochastique une propriété fondamentale des équations différentielles ordinaires. En effet, la solution $x(t_2) = g(t_2, x(t_1), t_1)$ de l'équation différentielle $\dot{x}(t) = f(x(t))$ est une fonction de $x(t_1)$ et ne dépend pas de $x(\tau)$, $\tau < t_1$.

Déf. B.11.2 *: processus markovien*

Le processus stochastique $x(t)$ est **markovien** *si* $\forall\ \{t_1, t_2, \cdots, t_n\}$ *de* T*:*

$$P[x(t_n) < \alpha_n / x(t_{n-1}), \cdots, x(t_1)] = P[x(t_n) < \alpha_n / x(t_{n-1})]$$

soit, en termes de fonction densité de probabilité:

$$p_{x(t_n)/x(t_{n-1}),\cdots,x(t_1)}(\alpha_n/\alpha_{n-1}, \cdots, \alpha_1) = p_{x(t_n)/x(t_{n-1})}(\alpha_n/\alpha_{n-1})$$

Toute l'information sur le passé est concentrée dans le dernier état "observé". Un processus de Markov est donc un processus stochastique dont le passé n'a pas d'influence sur le futur si le présent est connu.

Déf. B.11.3 *: densité de probabilité de transition*
$p_{x(t_n)/x(t_{n-1})}(\alpha_n/\alpha_{n-1})$ *est appelée* **fonction densité de probabilité de transition**.

Propriété B.11.2 *:*
Le processus markovien est un processus du second ordre.

En effet, la fonction densité de probabilité conjointe de $(x(t_1), \cdots, x(t_n))$ s'exprime ainsi :

$$p_{x(t_n),x(t_{n-1}),\cdots,x(t_1)}(\alpha_n/\alpha_{n-1}, \cdots, \alpha_1) = p_{x(t_n)/x(t_{n-1})}(\alpha_n/\alpha_{n-1}) \bullet$$
$$p_{x(t_{n-1})/x(t_{n-2})}(\alpha_{n-1}/\alpha_{n-2}) \cdots p_{x(t_2)/x(t_1)}(\alpha_2/\alpha_1) p_{x(t_1)}(\alpha_1)$$

qui se détermine donc à partir de la connaissance de $p_{x(t)}(\alpha) = p_x(\alpha, t)$, la densité de probabilité simple, et de $p_{x(t)/x(t_0)}(\alpha/x_0)$, la densité de probabilité de transition. On rappelle que:

- $p_x(\alpha, t) = \frac{\partial F(\alpha,t)}{\partial \alpha} = \frac{\partial P[x(t) \leq \alpha]}{\partial \alpha}$
- $p_{x(t)/x(t_0)}(\alpha/x_0) = p_{x(t)/x(t_0)}(\alpha/x(t_0) = x_0)$
- $p_x(\alpha, t) = \displaystyle\int_{-\infty}^{+\infty} p_x(x_0, t_0) p_{x(t)/x(t_0)}(\alpha/x_0) dx_0$
- $p_{x(t)/x(t_0)} \xrightarrow{t \to t_0} \delta(\alpha - x_0)$

<u>Exemple:</u>
On considère l'équation aux différences:

$$x_{n+1} = x_n + w_n \quad n = 0, 1, \cdots,$$

où les variables w_i sont des variables aléatoires indépendantes gaussiennes et la condition initiale x_0 est gaussienne et indépendante de w_i, $\forall i$. La séquence $\{x_n\}_n$ est clairement markovienne puisque pour x_n donnée, x_{n+1} dépend uniquement de w_n qui est indépendant de $x_{n-1}, \cdots, x_0$. De plus, elle est également gaussienne puisque combinaison linéaire de variables gaussiennes.

Comme nous le verrons dans la suite de ce cours, les processus de Markov ont une grande importance dans l'étude des équations aux différences stochastiques, (systèmes dynamiques en temps discret) et dans l'étude des équations différentielles stochastiques, (systèmes dynamiques en temps continu).

B.11.3 La marche aléatoire

La marche aléatoire est le nom donné à un processus stochastique en temps discret et à état discret que l'on peut décrire de la façon suivante.
On lance une pièce non truquée toutes les T secondes et suivant que l'obtient face ou pile, on effectue un pas à droite ou à gauche de longueur s. Le processus débute pour $t = 0$ et la position du marcheur aléatoire à l'instant t suit une fonction en escalier où les discontinuités sont aux points $t = nT$. On suppose maintenant que l'on a effectué n

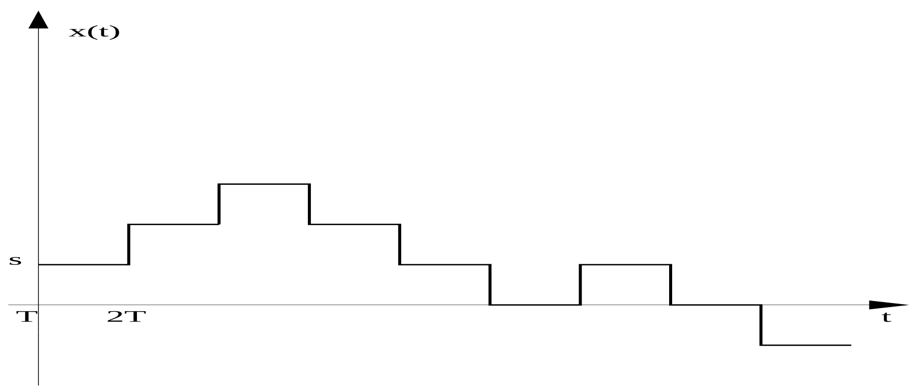


FIG. B.10 – *Exemple de marche aléatoire*

tirages de la pièce, pour lesquels on a observé k fois pile et donc $n-k$ fois face. La position du marcheur aléatoire à l'instant $t = nT$ est:

$$x(nT) = ks - (n-k)s = ms \quad m = 2k - n$$

De plus, la probabilité d'avoir k fois pile dans n tirages est alors:

$$P[x(nT) = ms] = \frac{\mathcal{C}_n^k}{2^n} \qquad k = \frac{m+n}{2}$$

En remarquant, que:

$$x(nT) = x_1 + x_2 + \cdots + x_n$$

où x_i est la taille du i^{eme} pas et les x_i sont indépendantes, prenant la valeur $\pm s$ avec $E[x_i] = 0$, $E[x_i^2] = s^2$, on obtient:

$$\boxed{E[x(nT)] = 0 \qquad E[x^2(nT)] = ns^2}$$

Quand le nombre de tirages augmente, il est possible de montrer, en utilisant le théorème de DeMoivre-Laplace que:

$$\mathcal{C}_n^k\, p^k q^{n-k} \simeq \frac{1}{\sqrt{2\pi npq}} e^{-(k-np)^2/2npq}$$

pour k dans un $\sqrt{npq}$ voisinage de np. Ici, $p = q = 0.5$ et $m = 2k - n$, ce qui conduit à:

$$P[x(nT) = ms] \simeq \frac{1}{\sqrt{n\pi/2}} e^{-m^2/2n}$$

pour m de l'ordre de $\sqrt{n}$.

B.11.4 Processus de Wiener

Le processus suivant a une importance théorique toute particulière dans l'étude générale des processus stochastiques. Avant de le définir rigoureusement, il est nécessaire d'introduire une définition intermédiaire.

Déf. B.11.4 *: incréments stationnairement indépendants*

Un processus stochastique $\{x_t(\omega)\ ,\ t \in T\}$ *noté* $x(t)$ *a des* **incréments stationnairement indépendants** *si pour tout les ensembles finis* $\{t_i :\ t_i < t_{i+1}\} \in T$*:*

- $x(t_2) - x(t_1), \cdots, x(t_n) - x(t_{n-1})$ *sont des vecteurs aléatoires indépendants.*
- $x(t+h) - x(\tau + h)$ *a la même distribution que* $x(t) - x(\tau)\ \ \forall\, t > \tau \in T,\ h > 0.$

Déf. B.11.5 *: processus de Wiener*

Un processus de Wiener *est un processus stochastique continu* $\{x_t,\ \ t \geq 0\}$ *si:*

- *Il est à incréments stationnairement indépendants.*
- *Il est normalement distribué.*
- *Sa moyenne est nulle.*
- $P[\{x_0 = 0\}] = 1.$

Il est remarquable de noter que pour la spécification de la loi de probabilité d'un processus à incréments indépendants, il suffit de spécifier la distribution de $x(t)$ et la distribution de $x(t) - x(\tau),\ \ \forall\, t > \tau \in T$. Du fait des propriétés vues dans le chapitre 1, alors, $x(t) - x(\tau)$ est aussi normalement distribué pour tout $t,\ \tau \geq 0$. Ainsi, afin de spécifier la loi de probabilité d'un processus de Wiener, il reste à spécifier la distribution de $x(t) - x(\tau)$ pour $t > \tau \geq 0$ qui est gaussienne et est donc spécifiée par sa moyenne et sa matrice de covariance. On a clairement:

$$E[x(t) - x(\tau)] = 0$$

De plus, pour $t \leq \tau$:

$$\begin{aligned} R(t,\tau) = E[x(t)x(\tau)'] &= E[((x(t) - x(t_0))(x(t) - x(\tau) + x(\tau) - x(t_0))'] \\ &= E[(x(t) - x(t_0))(x(t) - x(t_0))'] = E[x(t)x(t)'] \\ &= Q(t) \end{aligned}$$

De même, pour $t \geq \tau$:

$$R(t,\tau) = Q(\tau)$$

donc:

$$\boxed{R(t,\tau) = Q(min(t,\tau))}$$

On peut montrer également que la matrice $Q(t) = E[x(t)x(t)']$ est une fonction non décroissante monotone de t, soit:

$$Q(\tau) \geq Q(t) \qquad \forall\ \tau \geq t$$

Si l'on suppose de plus que $Q(t)$ est absolument continue alors il est possible d'écrire:

$$Q(t) = E[x(t)x(t)'] = \int_0^t V(s)ds$$

où $V(s) = V(s)' \geq 0$. Cela permet de réécrire alors:

$$R(t,\tau) = \int_0^{min(t,\tau)} V(s)ds \qquad \forall\ t,\ \tau > 0$$

$$P(t,\tau) = E[((x(t) - x(\tau))(x(t) - x(\tau))'] = Q(t) - Q(\tau) = \int_\tau^t V(s)ds$$

Il est à noter qu'un processus de Wiener est un processus gaussien-markovien. Un processus de Wiener est en fait le processus limite quand $T \to 0$ de la marche aléatoire.

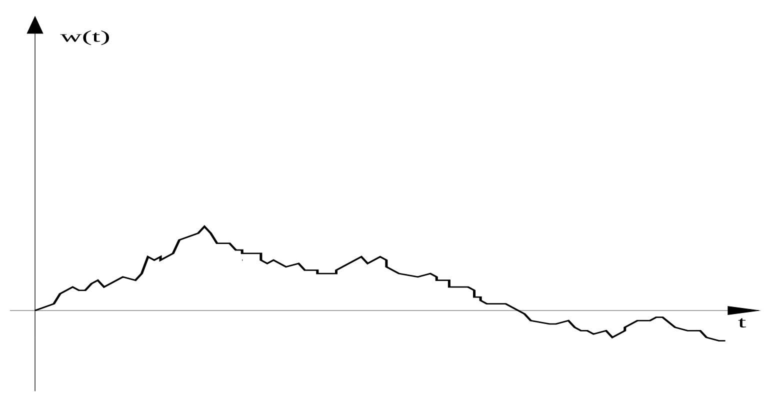


FIG. B.11 – *Processus de Wiener*

B.11.5 Bruit blanc

Le bruit blanc est un processus stochastique utilisé afin de modéliser les bruits intervenant dans toute modélisation de systèmes dynamiques.

Déf. B.11.6 *: séquence blanche*

Une séquence aléatoire blanche $\{x_n\ ,\ n = 1, 2, \cdots\}$ *est une séquence de markov telle que:*

$$p_{x_k/x_l}(\alpha_k/\alpha_l) = p_{x_k}(\alpha_k) \qquad (k > l)$$

Cela signifie que les x_k sont mutuellement indépendants. Connaitre la realisation de x_l n'aide pas à connaitre celle de x_k. Une séquence blanche est totalement imprévisible.
Si les x_k sont normalement distribués alors c'est **une séquence blanche gaussienne** définie par sa moyenne et sa covariance:

$$E[x_n] = m_n$$

$$P(n,k) = V_n \delta_{nk} \quad \delta_{nk} = \begin{cases} 1 & n = k \\ 0 & n \neq k \end{cases}, \; V_n \geq 0 \; Intensit\acute{e}$$

En général, le bruit du à la superposition d'un grand nombre de faibles effets aléatoires indépendants est gaussien, (ceci provenant de l'application du théorème de la limite centrale).
Puisque une séquence blanche gaussienne est un bon modèle pour le bruit affectant un système dynamique décrit par un modèle en temps discret, on souhaite disposer d'un équivalent pour les bruits affectant les modèles dynamiques en temps continu. Dans ce cas, le bruit blanc peut être défini *formellement* par extension du cas discret.

Déf. B.11.7 *: bruit blanc*
$x(t)$ *est* **un bruit blanc** *s'il vérifie:*

$$p_{x(t)/x(\tau)}(\alpha/\beta) = p_{x(t)}(\alpha) \quad t > \tau$$

Cette définition formelle quoique possédant les propriétés indispensables caractérisant un bruit, n'est pas d'une utilisation très facile.
Nous définissons donc formellement un bruit blanc gaussien par sa moyenne et sa covariance:

Déf. B.11.8 *: bruit blanc gaussien*
Un bruit blanc gaussien *est un processus stochastique gaussien défini par*

$$E[x(t)] = m_x(t)$$

$$P(t,\tau) = E[(x(t) - m_x(t))(x(\tau) - m_x(\tau))'] = V(t)\delta(t - \tau)$$

où:

- $V(t) \geq 0$ *intensité du bruit blanc.*
- $\delta(\bullet)$ *impulsion de Dirac.*

Dans le cas où l'intensité du bruit blanc est constante, le processus est stationnaire au sens large et il est possible d'introduire sa densité spectrale de puissance qui est donnée par:

$$\Psi_x(\omega) = V$$

Cela montre qu'un bruit blanc gaussien stationnaire a une densité de puissance identique à toutes les fréquences. Cela justifie la dénomination de bruit blanc par analogie avec la lumière blanche. Toutefois, si l'on calcule la puissance totale d'un bruit blanc en utilisant le résultat du théorème 6, nous obtenons une valeur infinie qui montre que ce type de processus n'existe pas dans le monde physique. De plus, du point de vue strictement mathématique, le bruit blanc n'est pas rigoureusement défini. En fait, on peut montrer qu'un bruit blanc gaussien est le processus "dérivé" d'un processus de Wiener bien que le processus de Wiener ne soit pas dérivable au sens classique. Dans le cadre de ce cours, nous nous contenterons de sa définition formelle du fait qu'un traitement rigoureux irait au delà des objectifs de ce cours.

B.12 Systèmes linéaires et processus gaussiens-markoviens

Dans ce chapitre, nous étudions le processus stochastique défini comme l'état d'un système dynamique en temps discret ou en temps continu, perturbé par des entrées aléatoires. On s'intéressera exclusivement à établir les équations d'évolution des deux premiers moments, (moyenne et covariance), caractérisant ces processus. Une attention toute particulière sera portée au cas où les entrées aléatoires perturbatrices sont des bruits blancs gaussiens.

B.13 Systèmes discrets

Soit le système récurrent:

$$x_{k+1} = F_k x_k + G_k w_k \quad , \quad x_0 \tag{B.1}$$

où $x_k \in \mathbb{R}^n$ est le vecteur d'état du système à un instant t_k repéré par l'indice k.
$F_k \in \mathbb{R}^{nn}$, $G_k \in \mathbb{R}^{nm}$ sont les matrices dynamique et d'entrée du système.x_0 est la condition initiale, vecteur aléatoire de moyenne m_0 et de covariance P_0.
Enfin, $w_k \in \mathbb{R}^m$, $k = 0, 1, \cdots$ est une séquence de bruits stochastiques dont les deux premiers moments sont définis par μ_k, la moyenne et Q_{kl} la matrice de covariance.
Le système (**??**) est appelé **système dynamique stochastique en temps discret**. Si l'entrée aléatoire w_k n'est pas présente dans l'équation (**??**), on retrouve une équation récurrente classique qui a pour solution x_k. Dans le cas contraire, il est nécessaire de connaitre p_{x_k} à partir de laquelle on calcule $E[x_k]$ et P_{kl} qui donnent de l'information sur le comportement statistique de x_k.

Hypothèses B.13.1 *:*
x_0 et $\{w_k, \ \forall \ k\}$ sont indépendants.

La solution de l'équation (**??**) s'écrit:

$$x_k = \Phi(k, 0)x_0 + \sum_{i=0}^{k-1} \Phi(k, i+1)G_i w_i \quad , \quad k > 1 \tag{B.2}$$

où la matrice $\Phi(k, l)$ est la matrice de transition entre les instants l et k, définie par:

$$\begin{cases} \Phi(k, l) = F_{k-1}F_{k-2} \cdots F_{l+1}F_l \\ \\ \Phi(l, l) = \mathbf{1_n} \end{cases}$$

A partir de l'expression de x_k qui apparait comme une combinaison linéaire des variables aléatoires x_0 et $(w_i, \ i = 0, \cdots, k-1)$, il est possible de déterminer les moments de ce vecteur aléatoire.

B.13.1 La moyenne ou espérance mathématique

A partir de (**??**), les matrices Φ et G étant à éléments déterministes et du fait de la linéarité de l'opérateur espérance, il vient:

$$E[x_k] = m_k = \Phi(k,0)m_0 + \sum_{i=0}^{k-1} \Phi(k,i+1)G_i\mu_i \tag{B.3}$$

Formellement équivalente à (**??**), cette expression montre que la moyenne du processus stochastique discret généré par (**??**) obéit à l'équation dynamique récurrente équivalente à (**??**).

$$\boxed{m_{k+1} = F_k m_k + G_k\mu_k \quad , \quad m_0}$$

Son obtention peut être effectuée directement à partir de (**??**), en appliquant l'opérateur espérance mathématique à cette expression.

B.13.2 La matrice de covariance

Soit:

$$P_{kl} = E[(x_k - m_k)(x_l - m_l)'] \tag{B.4}$$

Portant les expressions (**??**) et (**??**) dans (**??**), développant et utilisant le fait que x_0 et $(w_i,\ i = 0, \cdots, k-1)$ sont indépendants, (annulation des termes croisés puisque $E[(x_0 - m_0)(w_i - \mu_i)'] = 0$), il vient:

$$P_{kl} = \Phi(k,0)P_0\Phi'(l,0) + \sum_{i=0}^{k-1}\sum_{j=0}^{l-1} \Phi(k,i+1)G_iQ_{ij}G_j'\Phi'(l,j+1) \tag{B.5}$$

qui est l'expression, dans le cas général, de la matrice de covariance. Cette expression ne permet pas de rechercher une équation d'évolution récurrente "simple". La raison se trouve dans le second terme du second membre de (**??**) et notamment le fait qu'il y a un couplage dans le temps au niveau de la perturbation w, (a priori $Q_{ij} \neq 0 \ \ \forall\, i,j$).
Pour poursuivre, on introduit une hypothèse supplémentaire.

Hypothèses B.13.2 *:*
$\{w_k\}_k$ est une séquence blanche, c'est à dire:

$$E[(w_k - \mu_k)(w_l - \mu_l)'] = Q_k\delta_{kl} \tag{B.6}$$

où δ_{kl} est le symbole de Kronecker

$$\delta_{kl} = \begin{cases} 0 \ \textit{pour} \ k \neq l \\ 1 \ \textit{pour} \ k = l \end{cases}$$

Avec cette hypothèse, (importante et intéressante du point de vue pratique), (**??**) se transforme en:

$$P_{kl} = \Phi(k,0)P_0\Phi'(l,0) + \sum_{i=0}^{min(k-1,l-1)} \Phi(k,i+1)G_iQ_iG_i'\Phi'(l,i+1) \qquad \text{(B.7)}$$

et considérant la matrice de covariance P_k:

$$P_k = P_{kk} = E[(x_k - m_k)(x_k - m_k)']$$

on obtient:

$$P_k = \Phi(k,0)P_0\Phi'(k,0) + \sum_{i=0}^{k-1} \Phi(k,i+1)G_iQ_iG_i'\Phi'(k,i+1)$$

On vérifie qu'elle obéit à **l'équation récurrente de Lyapunov**, d'évolution suivante:

$$\boxed{P_{k+1} = F_kP_kF_k' + G_kQ_kG_k' \quad , \quad P_0}$$

On pourra vérifier que l'expression précédente peut être obtenue bien plus simplement à partir de (**??**). La dérivation faite ici a fait apparaitre la nécéssité de l'hypothèse 2 (**??**), nécessité qui peut passer inaperçue en utilisant l'approche directe.
Avec l'hypothèse 2 et utilisant l'expression de P_k, il est possible d'établir que:

$$\boxed{P_{kl} = \begin{cases} \Phi(k,l)P_l \; pour \; k > l \\ \\ P_k\Phi(l,k)' \; pour \; k < l \end{cases}}$$

B.13.3 Nature du processus stochastique x_k

Jusqu'ici, nous nous sommes intéressés aux lois de variation des deux premiers moments associés au processus x_k sans connaitre a priori la nature du processus stochastique. Dans cette section, sous des hypothèse données, nous précisons cette nature et le cas où la connaissance des deux premiers moments est suffisante.

Hypothèses B.13.3 *:*

- *$\{w_k\}_k$ est une séquence blanche, (indépendance de w_k et w_l , $\forall\ k \neq l$).*
- *x_0 et $\{w_k,\ k = 0,1,\cdots\}$ sont indépendants*

Sous ces hypothèses,

$$\boxed{\text{Le processus stochastique } x_k \text{ est } \textbf{un processsus Markovien}.}$$

Démonstration:

En effet, à partir de (**??**), pour x_k donné, x_{k+1} ne dépend que de w_k qui, étant indépendant de w_l, $l < k$ et x_0, est donc indépendant de $x_{k-1}, x_{k-2}, \cdots, x_1$, donc

$$p_{x_{k+1}/x_k,x_{k-1},\cdots,x_1,x_0}(\alpha_{k+1}/\alpha_k,\cdots,\alpha_0) = p_{x_{k+1}/x_k}(\alpha_{k+1}/\alpha_k)$$

Les équations récurrentes définies précédemment permettent de déterminer les deux premiers moments du processus stochastique x_k. Bien que permettant de répondre à nombre de questions pratiques, ces éléments ne suffisent pas seuls à caractériser complètement le processus stochastique, ce qui est le cas si on rajoute en plus l'hypothèse gaussienne.

Hypothèses B.13.4 *:*

- $\{w_k\}_k$ *est une séquence blanche gaussienne.*

- x_0 *est une variable aléatoire gaussienne.*

Sous ces hypothèses,

Le processus stochastique est **un processus gaussien-markovien**.

Démonstration:

Pour la propriété markovienne, elle est établie ci-dessus. Pour la propriété gaussienne, elle découle très simplement du fait que toute combinaison linéaire de variables aléatoires gaussiennes est elle-même gaussienne, (théorème 3). En effet,

$$x_1 = F_0 x_0 + G_0 w_0$$

x_1 dépend linéairement des vecteurs gaussiens x_0 et w_0, x_1 est donc un vecteur gaussien. Par induction, il est établi que x_k est gaussien.
Pour ce cas, la connaissance des deux premiers moments définis par les équations récurrentes revêt une importance toute particulière car elle permet de caractériser entièrement le processus stochastique du fait de son caractère gaussien.

B.13.4 Processus stationnaires

Dans le cas d'un système dynamique invariant dans le temps, les équations (**??**) sont données par:

$$x_{k+1} = F x_k + G w_k \quad , \quad x_0 \tag{B.8}$$

Deux hypothèses, intéressantes en pratique, sont faites.

Hypothèses B.13.5 *:*

- *Le système (**??**) est asymptotiquement stable, (les valeurs propres de F ont un module strictement inférieur à 1),*

- *la séquence blanche gaussienne $\{w_k\}_k$ est stationnaire, ce qui signifie que moyenne et covariance Q_k sont indépendantes du temps*

$$w_k \sim \mathcal{N}[\mu, Q]$$

Les équations d'évolution de la moyenne m_k et de la covariance P_k du processus stochastique x_k s'écrivent:

$$\begin{cases} m_{k+1} = Fm_k + G\mu \quad , \quad m_0 \\ \\ P_{k+1} = FP_kF' + GQG' \quad , \quad P_0 \end{cases} \tag{B.9}$$

Alors, lorsque k croit indéfiniment, le processus stochastique x_k tend vers un processus gaussien de distribution stationnaire définie par:

$$m_k \to m = (\mathbf{1} - F)^{-1}G\mu$$

$$P_k \to P$$

où P est la solution symétrique définie positive de **l'équation matricielle algébrique de Lyapunov**:

$$\boxed{FPF' - P + GQG' = 0}$$

L'hypothèse de stabilité asymptotique sur F assure l'unicité de la solution et son caractère non-négatif. Ainsi, $P_{kl} = F^{k-l}P_l$, $(k > l)$ est uniquement fonction de l'intervalle $(k - l)$.

B.13.5 Fonction densité de probabilité de transition

Pour un processus markovien, la loi de distribution est entièrement caractérisée par p_{x_0}, (*la distribution initiale*) et p_{x_{k+1}/x_k}, (*la distribution de transition*). Dans le cas gaussien, (x_{k+1}, x_k sont conjointement gaussiens), p_{x_{k+1}/x_k} définit également une distribution gaussienne. Pour sa caractérisation complète, il suffit de déterminer les deux premiers moments associés, à savoir:

- la moyenne conditionnelle

$$\boxed{E[x_{k+1}/x_k] = m_{k+1/k}}$$

- la covariance conditionnelle

$$\boxed{E[(x_{k+1} - m_{k+1/k})(x_{k+1} - m_{k+1/k})'/x_k] = P_{k+1/k}}$$

A partir de (**??**), il vient:

$$m_{k+1/k} = F_kx_k + G_k\mu_k \quad , \quad x_0$$

car

$$E[x_k/x_k] = m_{k/k} = x_k \quad et \quad E[w_k/x_k] = E[w_k] = \mu_k$$

De plus, $x_{k+1} - m_{k+1/k} = G_k(w_k - \mu_k)$ d'où,

$$P_{k+1/k} = E[G_k(w_k - \mu_k)(w_k - \mu_k)'G_k'] = G_kQ_kG_k'$$

Dans le cas où cette matrice de covariance conditionnelle est inversible, l'expression analytique de la fonction densité de probabilité de transition peut être fournie:

$$\boxed{p_{x_{k+1}/x_k}(\alpha_{k+1}/\alpha_k) = \frac{1}{(2\pi)^{n/2}det^{1/2}[P_{k+1/k}]}exp[-1/2(\alpha_{k+1} - m_{k+1/k})'P_{k+1/k}^{-1}(\alpha_{k+1} - m_{k+1/k})]}$$

B.14 Systèmes continus

Dans le cas des systèmes dynamiques stochastiques à temps continu, il serait nécessaire afin de justifier tous les résultats, de développer la théorie des équations différentielles stochastiques. Les processus stochastiques ne sont pas des fonctions intégrables au sens courant de Riemann. L'intégrale de Riemann du calcul classique doit être modifiée afin de rentrer dans le cadre du calcul de Itô. Nous nous contenterons, dans le cadre de ce cours d'énoncer des résultats formellement étendus à partir de ceux du cas discret. Soit le système dynamique à temps continu décrit par:

$$\dot{x}(t) = F(t)x(t) + G(t)w(t) \quad , \quad x(t_0) = x_0 \tag{B.10}$$

où $x \in \mathbb{R}^n, \; F \in \mathbb{R}^{nn}, \; G \in \mathbb{R}^{nm}$ sont définis classiquement.

Hypothèses B.14.1 *:*

- x_0, *condition initiale, est un vecteur aléatoire de moyenne* m_0 *et de matrice de covariance* P_0.
- $w(t)$ *est un bruit, processus stochastique tel que:*

$$E[w(t)] = \mu(t) \text{ et } E[(w(t) - \mu(t))(w(\tau) - \mu(\tau))'] = Q(t, \tau)$$

- x_0 *et* $w(t)$ *sont indépendants.*

Pour une réalisation du bruit $w(t)$, la solution de (**??**) s'écrit:

$$x(t) = \Phi(t, t_0)x_0 + \int_{t_0}^{t} \Phi(t, \tau)G(\tau)w(\tau)d\tau \tag{B.11}$$

où $\Phi(t, t_0)$ est la matrice de transition du système, solution de l'équation différentielle matricielle:

$$\frac{d\Phi(t, t_0)}{dt} = F(t)\Phi(t, t_0) \quad , \quad \Phi(t_0, t_0) = \mathbf{1}$$

B.14.1 Espérance mathématique

Appliquant l'opérateur espérance mathématique à (**??**), permuttant les opérateurs espérance et intégration, Φ et G étant des matrices à éléments déterministes, il vient:

$$E[x(t)] = m(t) = \Phi(t, t_0)m_0 + \int_{t_0}^{t} \Phi(t, \tau)G(\tau)\mu(\tau)d\tau \tag{B.12}$$

(**??**) est formellement identique à (**??**) montrant à l'évidence que la moyenne $m(t)$ obéit au système differentiel ordinaire:

$$\boxed{\dot{m}(t) = F(t)m(t) + G(t)\mu(t) \quad , \quad x_0}$$

Cette expression peut être immédiatement obtenue à partir de (**??**) par application de l'opérateur espérance mathématique en admettant que $E[\frac{dx}{dt}] = \frac{d}{dt}E[x]$.

B.14.2 Matrice de covariance

Par définition,

$$P(t, \tau) = E[(x(t) - m(t))(x(\tau) - m(\tau))']$$

Utilisant la solution (**??**) et du fait que x_0 et $w(t)$ sont indépendants, il vient:

$$P(t, \tau) = \Phi(t, t_0)P_0\Phi(\tau, t_0)' + \int_{t_0}^{t}\int_{t_0}^{\tau} \Phi(t, \theta_1)G(\theta_1)Q(\theta_1, \theta_2)G'(\theta_2)\Phi'(\theta_1, \theta_2)d\theta_1 d\theta_2$$

Expression générale complexe qui, comme dans le cas des systèmes discrets, est développée sous l'hypothèse:

Hypothèses B.14.2 *:*
$w(t)$ est un bruit blanc:

$$Q(t, \tau) = Q(t)\delta(t - \tau)$$

alors:

$$P(t,\tau) = \Phi(t,t_0)P_0\Phi'(\tau,t_0) + \int_{t_0}^{min(t,\tau)} \Phi(t,\theta)G(\theta)Q(\theta)G'(\theta)\Phi'(t,\theta)d\theta \qquad \text{(B.13)}$$

Considérons tout d'abord la matrice de covariance de $x(t)$:

$$P(t) = P(t,t) = E[(x(t)-m(t))(x(t)-m(t))']$$

$$P(t) = \Phi(t,t_0)P_0\Phi'(t,t_0) + \int_{t_0}^{t} \Phi(t,\theta)G(\theta)Q(\theta)G'(\theta)\Phi'(t,\theta)d\theta$$

Dérivant par rapport à t il vient:

$$\frac{dP(t)}{dt} = F(t)\left[\Phi(t,t_0)P_0\Phi'(t,t_0) + \int_{t_0}^{t} \Phi(t,\theta)G(\theta)Q(\theta)G'(\theta)\Phi'(t,\theta)d\theta\right] +$$
$$\left[\Phi(t,t_0)P_0\Phi'(\tau,t_0) + \int_{t_0}^{t} \Phi(t,\theta)G(\theta)Q(\theta)G'(\theta)\Phi'(t,\theta)d\theta\right] F'(t) + G(t)Q(t)G'(t)$$

soit **l'équation différentielle de Lyapunov**:

$$\boxed{\dot{P}(t) = F(t)P(t) + P(t)F'(t) + G(t)Q(t)G'(t) \quad , \quad P(t_0) = P_0}$$

On établit alors de façon similaire au cas discret:

$$\boxed{P(t,\tau) = \begin{cases} \Phi(t,\tau)P(\tau) & t \geq \tau \\ P(t)\Phi'(t,\tau) & t \leq \tau \end{cases}}$$

B.14.3 Nature du processus stochastique $x(t)$

On énonce sans les démontrer, les résultats similaires au cas discret.

Hypothèses B.14.3 *:*
$w(t)$ est un bruit blanc indépendant de x_0.

Sous cette hypothèse,

$$\boxed{\text{Le processus stochastique } x(t) \text{ est } \textbf{un processus markovien}.}$$

Hypothèses B.14.4 *:*

- *$w(t)$ est un bruit blanc gaussien.*
- *x_0 est un vecteur aléatoire gaussien.*
- *x_0 et $w(t)$ sont indépendants.*

$$\boxed{\text{Le processus stochastique } x(t) \text{ est } \textbf{un processus gaussien - markovien.}}$$

B.14.4 Processus stationnaires

On se place ici dans le cas de systèmes dynamiques invariants.

$$\dot{x}(t) = Fx(t) + Gw(t) \quad , \quad x_0 \tag{B.14}$$

Hypothèses B.14.5 *:*

- *F et G sont des matrices constantes.*
- *F est stable asymptotiquement, les valeurs prores de F ont une partie réelle strictement négative.*
- *$w(t)$ est un bruit blanc stationnaire décrit par:*

$$\mu(t) = \mu \quad Q(t) = Q$$

Les équations dynamiques des deux premiers moments deviennent:

$$\boxed{\dot{m}(t) = Fm(t) + G\mu \quad , \quad m_0}$$

$$\boxed{\dot{P}(t) = FP(t) + P(t)F' + GQG' \quad , \quad P_0}$$

ainsi:

$$\lim_{t\to\infty} m(t) = F^{-1}G\mu$$

$$\lim_{t\to\infty} P(t) = P$$

où P est la solution définie positive de **l'équation algébrique matricielle de Lyapunov**:

$$\boxed{FP + PF' + GQG' = 0}$$

De plus $P(t,\tau) = e^{F(t-\tau)}P(\tau) = P(t-\tau)$, $(t > \tau)$ Donc, $x(t)$ tend asymptotiquement vers un processus stationnaire.

B.15 Processus générateur - représentation de markov

B.15.1 Définition

Etant donné un processus $s(t)$, (s_n) de moyenne nulle, on peut se poser le problème général de rechercher deux fonctions F et H et un processus $v(t)$, (v_n), de moyenne nulle tels que $v(t_1)$, $v(t_2)$, (v_{n_1}, v_{n_2}) soient non corrélés et tels que:

- Cas continu:

$$\begin{cases} \dot{x}(t) = F(t)x(t) + v(t) \\ \\ y(t) = H(t)x(t) \end{cases}$$

- Cas discret:

$$\begin{cases} x_{n+1} = F_n x_n + v_n \\ \\ y_n = H_n x_n \end{cases}$$

où $P_y(t,\tau) = P_s(t,\tau)$, $(P_y(k,l) = P_s(k,l))$.

Déf. B.15.1 *: processus générateur*

Le modèle d'état associé au processus $s(t)$, (s_n), est appelé **processus générateur** *du processus stochastique.*

La recherche d'un processus générateur dans le cas général est un problème complexe qui fait appel, pour sa résolution, à des techniques d'analyse harmonique ou spectrale qui sont hors du cadre de ce cours. Dans le cas où $s(t)$, (s_n), est un processus gaussien-markovien stationnaire de moyenne nulle, il est aisé de montrer que le processus générateur est caractérisé par:

- Cas continu:

$$F(t) = P_x(t,\tau) P_x^{-1}(t,\tau)$$

- Cas discret:

$$F_n = P_x(n+1,n) P_x^{-1}(n,n)$$

Déf. B.15.2 *: représentation markovienne*

Si un processus générateur existe alors on dit que $s(t)$, (s_n), est à **représentation markovienne**.

Si la densité spectrale fournit une représentation des signaux aléatoires stationnaires, le processus générateur associé à la représentation gaussienne-markovienne permet d'étudier les processus non stationnaires.

B.15.2 Quelques exemples

Dans cette section, nous donnons quelques exemples de processus générateurs associés à des processus aléatoires classiques. Une version continue et discrète est donnée.

B.15.3 Constante aléatoire

La constante aléatoire est une variable non dynamique d'amplitude aléatoire.

- Cas continu:

$$\dot{x}(t) = 0$$

- Cas discret:

$$x_{k+1} = x_k$$

La constante aléatoire peut être vue comme la sortie d'un intégrateur sans entrée mais avec une condition initiale aléatoire non nulle.

B.15.4 La marche aléatoire

Ce processus a été défini au chapitre 2.

- Cas continu:

$$\dot{x}(t) = w \qquad E[w(t)w(\tau)] = V(t)\delta(t-\tau)$$

- Cas discret:

$$x_{k+1} = x_x + w_k \qquad E[w_k w_l] = V_k \delta_{kl}$$

B.15.5 Variable aléatoire exponentiellement corrélée

Une variable aléatoire dont la fonction d'autocorrélation est une exponentielle décroissante, est dite exponentiellement corrélée.

$$R_s(\tau) = \sigma^2 e^{-\beta|\tau|}$$

Ce type de processus est utilisé comme représentation des perturbations affectant les systèmes dynamiques. Cela fournit également une bonne approximation des signaux à bande limitée dont la densité spectrale de puissance est plate sur une bande passante finie.

- Cas continu:

$$\dot{x}(t) = -\beta x(t) + w(t)$$

où $w(t)$ est non corrélé, (bruit blanc par exemple).

- Cas discret:

$$x_{k+1} = e^{-\beta(t_{k+1}-t_k)} x_k + w_k$$

Références bibliographiques

[1] J. Bernussou, *Processus stochastique - filtrage*, Cours ENSICA, Toulouse, 1994.

[2] D.R. Cox, H.D. Miller, *The theory of Stochastic Processes*, Chapamn et Hall, 1965.

[3] G. Duc, *Robustesse des systèmes linéaires multivariables*, Polycopié de cours SUPELEC, 1994.

[4] A. Gelb, *Applied Optimal Estimation*, The MIT Press, 1994.

[5] M.S. Grewal, A.P. Andrews, *Kalman filtering*, Prentice hall information and system siences series, Englewood Cliffs, New jersey, 1993.

[6] P.G. Hoel, S.C. Port, C.J. Stone, *Introduction to Probability Theory*, Haughton Milfilin Company, Boston, 1971.

[7] H. Kwakernaak, R. Sivan, *Linear optimal control systems*, Wiley interscience, New York, 1976.

[8] A.H. Jazwinski, *Stochastic processes and filtering*, Academic press, New York, 1970.

[9] M. Labarrere, J.P. Krief, B. Gimonet, *Le fitrage et ses applications*, Cepadues editions, 1978.

[10] F.L. Lewis, *Applied optimal control and estimation*, Prentice hall, Englewood Cliffs, New Jersey, 1992.