

NOTES DE COURS

M2RODECO & M2 IARF OPTIMISATION ET ESTIMATION

Partie Estimation

Patrick Danès et Carine Jauberthie
LAAS-CNRS & Université Toulouse III - Paul Sabatier
`{patrick.danes,carine.jauberthie}@laas.fr`

Version du 27 août 2020
Merci de nous communiquer toute erreur que vous décèleriez.

Introduction

De nombreux problèmes de robotique, décision ou commande sont formalisés au moyen de modèles paramétriques dont il s'agit d'estimer ou d'optimiser les paramètres. Ainsi, la localisation d'un robot mobile s'appuie sur la recherche des valeurs des paramètres de situation expliquant au mieux les données. En robotique de manipulation, une tâche de positionnement peut s'exprimer comme la recherche des paramètres de configuration minimisant un critère de distance.

Dans ce cours, nous nous intéressons à l'estimation de paramètres de modèles en contexte stochastique, *i.e.*, lorsque valeurs admissibles des bruits qui affectent les observations et/ou la connaissance *a priori* sur les paramètres à estimer sont décrites par des distributions de probabilités. A l'issue de ce cours, vous devrez être capables de :

- comprendre la problématique de l'estimation en contexte stochastique ;
- caractériser les propriétés de tout estimateur ;
- synthétiser les principaux estimateurs, selon que l'on dispose ou non d'une connaissance *a priori* sur les paramètres à estimer.

Chapitre I

Problématique de l'estimation

I.1 Modélisation

Modéliser un phénomène ou un système réel consiste à établir une *représentation mathématique* de celui-ci. Cette représentation est conçue *dans un but précis*, par exemple :

- l'analyse du phénomène afin d'en approfondir sa compréhension (modèles de la physique,...),
- la prédiction de comportements,
- la commande,
- la détermination de grandeurs pour lesquelles aucun capteur n'est disponible, à partir de mesures indirectes,
- le test d'hypothèses (diagnostic médical, contrôle de sûreté de fonctionnement d'installations industrielles, ...),
- le traitement de signaux (suppression du bruit, compression de données, filtrage, interpolation,...),
- etc.

Les caractéristiques d'un modèle (sa représentativité, sa souplesse d'exploitation, sa fiabilité, sa complexité,...) sont donc adaptées au but recherché. Elles sont le fruit d'un compromis entre les moyens nécessaires à son obtention ou son exploitation (temps, calcul, expérimentations), et les retombées espérées. À titre d'exemple, les modèles pour l'analyse et la commande d'un système doivent d'une part permettre la représentation fidèle du comportement qualitatif et quantitatif de celui-ci, capturer ses propriétés cruciales sur un domaine de fonctionnement possiblement restreint, et, d'autre part, admettre une structure suffisamment simple pour permettre une procédure d'analyse relativement simple et une méthode de commande efficace.

Il existe différentes classes de modèles, parmi elles, nous trouvons :

- les modèles de connaissance (connaissance précise du processus réel : modèle basé sur les lois de la Physique,...) *vs* de représentation (on ne s'intéresse qu'à la représentation du comportement entrée/sortie sans étudier de façon approfondie les phénomènes qui en sont à l'origine),
- les modèles déterministes (on sait calculer la valeur de la sortie du modèle de façon exacte connaissant les paramètres) *vs* aléatoires (le modèle ne caractérise que certaines propriétés statistiques du système),
- les modèles à temps continu *vs* à temps discret,
- les modèles linéaires *vs* non linéaires *par rapport aux paramètres*.

Un *modèle stochastique* est un modèle de la forme

$$Z = \mathcal{M}(\theta, V) \quad (\text{I.1.1})$$

où :

- $\theta \in \mathbb{R}^M$ désigne le vecteur des paramètres « caché » à estimer ;
- $Z \in \mathbb{R}^N$ désigne le vecteur aléatoire d'observation (ou de mesure) ; pour l'expérience en cours, ce vecteur se réalise en le vecteur z des mesures effectuées[†] ;
- V est un vecteur aléatoire modélisant les bruits/perturbations ; le modèle intègre une description statistique de V .

Le problème d'estimation s'énonce alors comme suit.

Énoncé *Disposant de la réalisation z de Z , quelle information peut-on obtenir sur le vecteur de paramètres θ ?* □

Préalablement à la mise en place de tout schéma d'estimation, il convient de recenser la *connaissance a priori* dont on dispose sur le vecteur de paramètres caché θ . Il sera vu plus loin que :

- si on ne sait rien sur θ , alors ce vecteur est considéré comme déterministe inconnu ;
- si, au contraire, on dispose d'une connaissance a priori sur les valeurs admissibles de θ , alors celle-ci est exprimée au moyen de la distribution de probabilité d'un vecteur aléatoire Θ ; θ est alors considéré comme la réalisation de Θ pour l'expérience en cours[‡].

Parmi les problèmes potentiels en estimation stochastique, on peut citer :

- la sélection d'un modèle (I.1.1) inappropriate : structure ne permettant pas de rendre compte du phénomène considéré ; propriétés structurelles non satisfaites (cf. §I.2.4) ; caractérisation erronée des statistiques de Θ, V ;
- la difficulté (voire l'impossibilité) de parvenir au calcul de l'estimé de θ ;
- la dégradation des performances de l'estimateur lorsque les hypothèses formulées dans le modèle ne sont pas parfaitement respectées ; on souhaite que le schéma utilisé soit doté d'une robustesse relative par rapport au non-respect de ces hypothèses.

I.2 Formalisation mathématique d'un problème d'estimation

Dans cette section, nous supposerons que le vecteur de paramètres θ appartient à un ensemble admissible $\mathcal{U}_{ad} \subset \mathbb{R}^M$) et que le vecteur z des mesures effectuées appartient à \mathbb{R}^N .

I.2.1 Estimateur / Estimé

Un problème d'estimation stochastique consiste à analyser les propriétés, voire à synthétiser, une fonction *estimateur* définie comme suit.

Définition I.2.1 *Soit une fonction g de l'espace des observations dans l'espace des paramètres, qui, à chaque vecteur de mesure z (réalisation de Z pour l'expérience en cours) associe $\hat{\theta} = g(z)$, i.e.,*

$$\begin{aligned} g : \mathbb{R}^N &\rightarrow \mathbb{R}^M \\ z &\mapsto \hat{\theta} = g(z). \end{aligned} \quad (\text{I.2.1})$$

[†]. On peut noter $z = Z(\omega)$, où ω symbolise le résultat de l'expérience (ou « événement »).

[‡]. La densité de probabilité de Θ , ou « loi a priori », sera notée $p_\Theta(\theta)$.

$\hat{\theta}$ est appelé estimé de θ . La fonction g est appelée estimateur. \square

Dans tout le cours, la fonction g est déterministe, i.e., pour un vecteur de mesures z donné, l'image $g(z)$ de z par g est parfaitement définie, et ne dépend pas de l'expérience \dagger .

On note $\hat{\Theta} = g(Z)$, de sorte que $\hat{\Theta} \in \mathbb{R}^M$. Du fait que Z est une variable aléatoire vectorielle, $\hat{\Theta}$ est également un vecteur aléatoire, appelé *estimateur* \ddagger . Bien sûr, $\hat{\Theta}$ se réalise en $\hat{\theta}$ lorsque Z se réalise en z .

Il vient la conséquence fondamentale suivante, cf. Figure I.1.

Conséquence fondamentale *Un estimateur $\hat{\Theta}$ étant une fonction (ici, déterministe) d'une variable aléatoire (la variable aléatoire de mesure Z), il est lui-même une variable aléatoire, de sorte que ses propriétés doivent être évaluées statistiquement.* \square

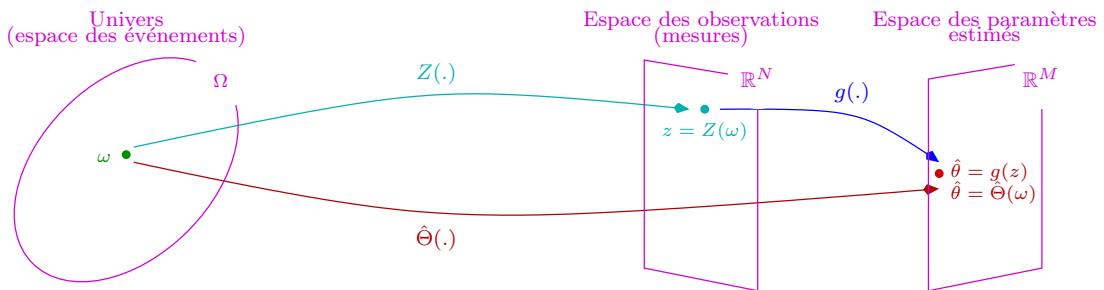


FIGURE I.1 – Principe de l'estimation stochastique

L'exemple ci-dessous permet d'illustrer cette conclusion.

Exemple 1 Soit a un niveau constant inconnu. On prélève la séquence de mesures $z[1], \dots, z[N]$, indexées par $n = 1, \dots, N$. Chaque $z[n]$ est la réalisation de la variable aléatoire $Z[n] = a + W[n]$, où W désigne du bruit. On suppose que la séquence $W[1], \dots, W[N]$ est indépendante, identiquement distribuée (i.i.d.) selon $\mathcal{N}(0, \sigma^2)$.

Disposant des observations $z[1], \dots, z[N]$, comment estimer a ?

Soient les deux estimés

$$\hat{a} = \frac{1}{N} \sum_{n=1}^N z[n] \quad \text{et} \quad \hat{a} = z[1].$$

- Sur une expérience, impossible de dire lequel est le plus proche de a .
- Si on répète l'expérience un nombre élevé de fois, alors intuitivement,
 - \hat{a} et \hat{a} renvoient sans doute « en moyenne » la valeur correcte de a (plus exactement, les estimateurs associés \hat{A} et \hat{A} sont centrés sur a);

\dagger . Il existe en effet des fonctions g aléatoires. En pratique, il s'agit d'estimateurs qui, pour calculer l'image $g(z)$ du vecteur d'observation z , reposent de manière interne sur la génération de nombres (pseudo-)aléatoires. Ces méthodes d'estimation, dites « de Monte Carlo » ne seront pas vues ici.

\ddagger . En fait, le vocable « estimateur » se référera préférentiellement à $\hat{\Theta}$. La fonction g ne sera que rarement référencée explicitement.

- grâce à un effet de « lissage » (à démontrer!), le fait que \hat{A} soit obtenu sur la base d'un nombre plus important de mesures permet qu'il se disperse de manière moins importante autour de a (plus exactement, la dispersion de \hat{A} autour de sa moyenne semble moins importante).
- En conclusion,
 - on peut avoir une idée sur la base de simulations « de Monte Carlo », mais un nombre insuffisant d'expériences peut conduire à des résultats trompeurs;
 - il est nécessaire de mener une analyse statistique de l'estimateur (examen de sa pdf, etc.) pour conclure. \square

I.2.2 Les deux familles d'estimateurs en contexte stochastique

On distingue deux familles d'estimateurs, selon la connaissance a priori disponible sur le vecteur de paramètres caché $\theta \in \mathbb{R}^M$. Les notations utilisées pour désigner des densités de probabilités sont rappelées en bas de page[†].

On ne dispose d'aucune connaissance a priori sur θ Comme indiqué plus haut, θ est alors supposé déterministe inconnu. On se situe dans le contexte des *estimateurs classiques*, ou *de Fisher*, (Figure I.2). L'estimation classique repose sur le *modèle d'observation*

$$p_{Z;\theta}(z; \theta). \quad (\text{I.2.2})$$

On commettra l'abus de notation usuel[‡] consistant à écrire (I.2.2) sous la forme

$$p_{Z|\theta}(z|\theta). \quad (\text{I.2.3})$$

On dispose d'une connaissance a priori sur θ Comme indiqué plus haut, θ est alors considéré comme la réalisation d'un vecteur aléatoire Θ . On parle d'*estimateurs bayésiens*, ou *de Bayes* (Figure I.3). L'estimation bayésienne repose sur le modèle

$$p_{Z,\Theta}(z, \theta), \quad (\text{I.2.4})$$

c'est-à-dire sur la *loi a priori*

$$p_\Theta(\theta) \quad (\text{I.2.5})$$

et le *modèle d'observation*

$$p_{Z|\Theta}(z|\theta). \quad (\text{I.2.6})$$

[†]. Les notations sont sans doute un peu lourdes, mais nous semblent nécessaires pour permettre le développement rigoureux des concepts.

- On désigne par $p_A(a)$ la densité de probabilité de la variable aléatoire A . C'est bien une fonction de la variable muette a , qui désigne une réalisation possible de A . Si on trace la représentation de $p_A(a)$, on porte en abscisse les valeurs (réalisations) possibles a de A et en ordonnée la valeur numérique de l'évaluation de $p_A(\cdot)$ en a .
- On désigne par $p_{A|B}(a|b)$ la densité de probabilité de la variable aléatoire A conditionnellement à l'événement « la variable aléatoire B s'est réalisée en la valeur b ». C'est bien une fonction de la variable muette a , qui désigne une réalisation possible de A . Si on trace la représentation de $p_{A|B}(a|b)$ pour l'événement $B = b$, on porte en abscisse les valeurs (réalisations) possibles a de A et en ordonnée la valeur numérique de l'évaluation de $p_{A|B}(\cdot|b)$ en a .

[‡]. Cet abus de notation consiste naturellement à écrire $p_{\text{variable aléatoire}|\text{paramètre déterministe}}(\cdot|.)$ au lieu de $p_{\text{variable aléatoire}|\text{variable aléatoire}}(\cdot|.)$...

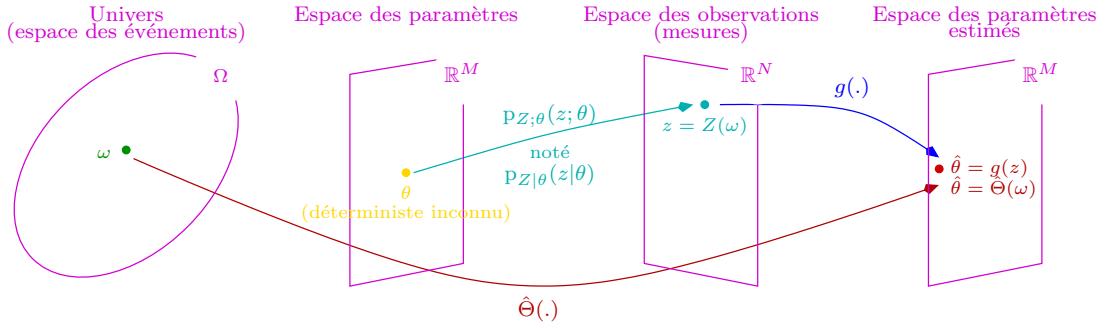


FIGURE I.2 – Principe de l'estimation stochastique classique (Fisher)

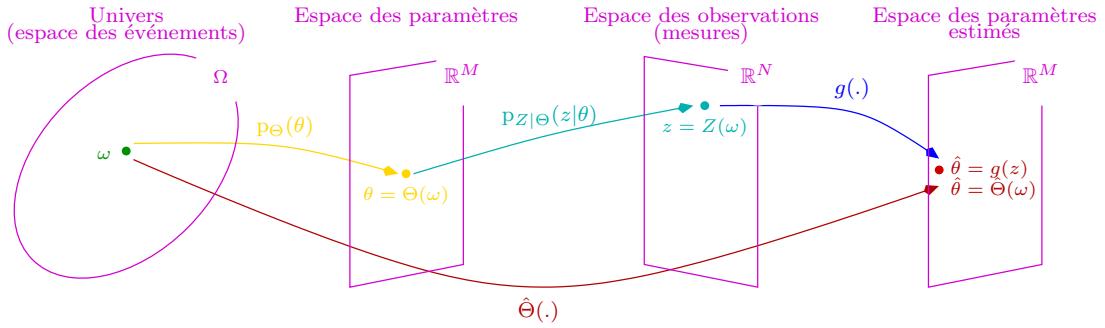


FIGURE I.3 – Principe de l'estimation stochastique bayésienne (Bayes)

Illustrons ces problématiques sur quelques exemples.

Exemple 2

1. On dispose de N échantillons

$$z[1], \dots, z[N] \xrightarrow{i.i.d.} p_{Z|(\mu, \sigma^2)}(z | (\mu, \sigma^2)) = \mathcal{N}(z; \mu, \sigma^2),$$

où $\mathcal{N}(z; \mu, \sigma^2)$ désigne la loi Gaussienne scalaire réelle de moyenne μ et de variance σ^2 . On se donne des estimés $\hat{\mu}$ et $\hat{\sigma}$ des paramètres μ et σ (supposés déterministes inconnus car pas de connaissance a priori à leur sujet). Par exemple,

$$\hat{\mu} = \frac{1}{N} \sum_{n=1}^N z[n], \quad \text{et} \quad \hat{\sigma} = \sqrt{\frac{1}{N} \sum_{n=1}^N (z[n] - \hat{\mu})^2}.$$

→ Il convient donc d'analyser les estimateurs associés.

2. Une personne immobile souhaite déterminer sa position (abscisse curviligne 1D) sur une route, sur la base de positions relevées par plusieurs capteurs (odomètre, localisation zigbee, GPS, etc.).
- Si on exploite une connaissance a priori, Bayes statique.
3. Un robot mobile souhaite se localiser sur la base de relevés relatifs à plusieurs amers, voire souhaite également construire une carte dans laquelle il estime la position de ces amers concomitamment à sa localisation : "SLAM" (Simultaneous Localization And Mapping).

↪ Bayes dynamique — Filtrage.

4. Suivi hors ligne d'animaux marins sur la base de relevés satellitaires mensuels et d'un modèle de dynamique *a priori*.

↪ Bayes dynamique — Lissage.

5. Mais aussi : radar, sonar, parole (reconnaissance de phonèmes → estimation des paramètres d'un modèle linéaire prédictif –LPC– pour déterminer l'enveloppe spectrale, qui est un attribut insensible au pitch), image (suivi d'objets / de personnes), médecine, communications, trajectographie, séismologie, météo, robotique, etc. □

I.2.3 Propriétés intuitives des estimateurs

Voyons sur l'Exemple 2-1 ci-dessus quelles peuvent être les principales propriétés des estimateurs. Avant d'effectuer la démarche d'estimation paramétrique, il est important de vérifier certaines propriétés structurelles du modèle. En particulier l'*identifiabilité*. C'est-à-dire, sous certaines hypothèses, existe-t-il un vecteur de paramètre θ , et si oui, est-il unique ? De nombreuses recherches sont menées dans ce domaine depuis les années soixante dix.

I.2.4 Deux propriétés structurelles des modèles

A 1.2.4.1 Identifiabilité

Lors de la modélisation d'un phénomène, la structure retenue pour le modèle paramétrique doit permettre l'estimation des paramètres. Ainsi, étant donné un ensemble \mathcal{U}_{ad} de paramètres admissibles, nous souhaitons savoir s'il est possible de trouver un vecteur **unique** de paramètres de \mathcal{U}_{ad} (*identifiabilité globale*) ou un **nombre localement unique** (*identifiabilité locale*), indépendamment de toute expérience. Cette démarche est nécessaire car elle permet de rejeter ou de conserver un modèle qui, a priori, semble satisfaisant : c'est la *recherche d'identifiabilité paramétrique*. Pour vérifier cette propriété, nous nous plaçons dans un cadre idéalisé, c'est-à-dire que le processus et le modèle sont supposés avoir des structures identiques (il n'y a pas d'erreur de caractérisation), les données sont considérées sans bruit et les instants de mesure peuvent être choisis de manière libre. De plus, si le modèle est contrôlé, l'entrée appliquée est supposée pouvoir être choisie librement. Différentes définitions de l'identifiabilité sont disponibles dans la littérature. Elles sont basées sur des notions d'analyse ou d'algèbre différentielle et ne sont pas toutes équivalentes.

Les définitions d'identifiabilité globale et locale présentées dans cette section utilisent le **comportement entrée-sortie** d'un système.

Définition I.2.2 Nous notons **comportement entrée-sortie** du modèle $M(\theta)$ l'application :

$$\begin{aligned} \mathcal{C}(P) : \mathcal{U} &\longrightarrow \text{appl}(\mathbb{R}^+, \mathbb{R}^N) \\ u &\longrightarrow y, \end{aligned}$$

où \mathcal{U} est une classe quelconque d'applications définies de \mathbb{R}^+ dans \mathbb{R}^m (si $u(t) \in \mathbb{R}^m$). Le comportement entrée-sortie de la structure $M(\cdot)$ est l'application \mathcal{C} qui à tout vecteur de paramètres θ associe le comportement entrée-sortie du modèle $M(\theta)$.

Pour la plupart des modèles, il existe des points atypiques dans \mathcal{U}_{ad} pour lesquels le modèle étudié n'est pas identifiable. L'identifiabilité paramétrique est en fait une propriété **structurelle** ce qui signifie qu'elle sera vraie partout sauf éventuellement sur un ensemble de mesure nulle. Dans ce contexte, E. Walter et L. Pronzato proposent dans [2] les définitions suivantes :

Définition I.2.3 *Le paramètre θ_i est dit **structurellement globalement** (respectivement **structurellement localement**) **identifiable** si pour presque tout $\bar{\theta} \in \mathcal{U}_{ad}$, (respectivement s'il existe un voisinage \mathcal{W} de $\bar{\theta}$ dans \mathcal{U}_{ad} tel que) :*

$$\theta^* \in \mathcal{U}_{ad} \text{ (resp. } \theta^* \in \mathcal{W}), \mathcal{C}(\theta^*) = \mathcal{C}(\bar{\theta}) \Rightarrow \theta_i^* = \bar{\theta}_i.$$

Les structures de modèles peuvent ainsi être **structurellement identifiables** au sens des définitions suivantes :

Définition I.2.4 *Une structure de modèle est **structurellement globalement** (respectivement **structurellement localement**) **identifiable** si tous ses paramètres le sont.*

Nous présentons ci-dessous une méthode classique pour tester l'identifiabilité de systèmes linéaires. Elle utilise la transformée de Laplace du système ou sa matrice de transfert si l'on considère les conditions initiales nulles.

Dans le cas d'une structure non linéaire, notre première démarche est de linéariser le système autour d'une trajectoire connue et d'utiliser les méthodes adaptées aux structures linéaires. Nous verrons également une méthode plus générale.

Nous considérons ci-dessous des systèmes donnés par la forme :

$$\begin{cases} \dot{x}(t, \theta) = A(\theta)x(t, \theta) + B(\theta)u(t), \\ y(t, \theta) = C(\theta)x(t, \theta), \\ x(0, \theta) = x_0(\theta), \end{cases} \quad (\text{I.2.7})$$

où $x(t, \theta)$ est un vecteur de \mathbb{R}^n .

Définition I.2.5 *Nous appelons matrice de transfert d'un système linéaire, d'entrée le vecteur m -dimensionnel u et de sortie le vecteur p -dimensionnel y , la matrice $p \times m$, $F(s, \theta)$, telle que $Y(s, \theta) = F(s, \theta)U(s)$ où s désigne la variable de Laplace, $U(s)$ (resp. $Y(s, \theta)$) désigne la transformée de Laplace de $u(t)$ (resp. $y(t, \theta)$) si elle existe.*

Nous supposons dans une première partie que $x_0(\theta) = 0$. En utilisant la transformée de Laplace appliquée au système (I.2.7), nous obtenons :

$$F(s, \theta) = C(\theta)(sI_n - A(\theta))^{-1}B(\theta),$$

où I_n représente la matrice identité $n \times n$.

Le résultat d'identifiabilité est donné en utilisant la proposition suivante :

Propriété I.2.6 *Si pour presque tout $\bar{\theta} \in \mathcal{D}_m$, nous avons :*

$$\forall s, F(s, \theta^*) = F(s, \bar{\theta}) \Rightarrow \theta^* = \bar{\theta},$$

alors le modèle est structurellement identifiable.

L'identifiabilité structurelle est globale si $\mathcal{D}_m = \mathcal{U}_{ad}$ et locale si \mathcal{D}_m est un voisinage de P dans \mathcal{U}_{ad} .

Remarque I.2.7 Si $x_0(\theta)$ est non nul, la transformée de Laplace des équations (I.2.7) s'écrit :

$$F(s, \theta) = C(\theta)[sI_n - A(\theta)]^{-1}x_0(\theta) + C(\theta)[sI_n - A(\theta)]^{-1}B(\theta)U(s).$$

L'utilisation de la proposition I.2.6 permet à nouveau d'obtenir le résultat d'identifiabilité.

Remarque I.2.8 Cas des systèmes linéaires non contrôlés :

Si, dans un cadre plus général, le système linéaire considéré n'est pas contrôlé, la proposition suivante permet de tester l'identifiabilité de ses paramètres :

Propriété I.2.9 Soit le système linéaire non contrôlé :

$$\begin{cases} \dot{x}(t, \theta) = A(\theta)x(t, \theta), \\ x(0, \theta) = x_0(P), \\ y(t, P) = C(P)x(t, P). \end{cases} \quad (\text{I.2.8})$$

La matrice de transfert du système (I.2.8) est identique à celle du système linéaire contrôlé :

$$\begin{cases} \dot{x}(t, \theta) = A(\theta)x(t, \theta) + B(\theta)u(t), \\ x(0, \theta) = 0, \\ y(t, \theta) = C(\theta)x(t, \theta), \end{cases} \quad (\text{I.2.9})$$

avec $B(\theta) = x_0(\theta)$.

(I.2.8) est identifiable si et seulement si (I.2.9) est identifiable.

Ce résultat est faux dans le cas non linéaire.

Comme dit précédemment, pour les systèmes non linéaires, la recherche d'identifiabilité peut s'effectuer en utilisant la linéarisation du système autour d'une trajectoire connue, ce qui nous permet d'utiliser les techniques adaptées aux structures linéaires. En effet, une approche classique dans la recherche d'identifiabilité consiste à linéariser le système autour d'une trajectoire connue (souvent d'équilibre) et à rechercher l'identifiabilité du modèle linéarisé. M.S. Grewal et K. Glover ont montré que l'identifiabilité du système linéarisé autour de cette trajectoire entraîne l'identifiabilité du système non linéaire. Nous verrons également comment procéder dans un cadre plus général.

Soit le système non linéaire contrôlé décrit par les équations :

$$\begin{cases} \dot{x}(t, \theta) = f(x(t, \theta), \theta, u(t)), \\ y(t, \theta) = h(x(t, \theta), \theta, u(t)), \\ x(0, \theta) = x_0(\theta). \end{cases} \quad (\text{I.2.10})$$

Le comportement d'un système non linéaire au voisinage d'une solution connue (x_e, u_e) peut être approché par un système linéaire. Nous considérons le vecteur de paramètres θ fixé et effectuons de petites perturbations sur x , u ou les deux, nous les notons δx et δu . Les fonctions f et h sont supposées posséder des dérivées partielles par rapport à x et u et la condition initiale est supposée ne pas dépendre des paramètres.

Le système (I.2.10) peut alors s'écrire sous la forme :

$$\begin{cases} \dot{x}(t, \theta) = f(x_e(t, \theta) + \delta x(t, \theta), \theta, u_e(t) + \delta u(t)), \\ y(t, \theta) = h(x_e(t, \theta) + \delta x(t, \theta), \theta, u_e(t) + \delta u(t)), \\ x(0, \theta) = x_e. \end{cases} \quad (\text{I.2.11})$$

Par la suite, δx_l et δu_l représentent les vecteurs d'état et de contrôle du système linéarisé. Ainsi les équations linéarisées peuvent se mettre sous la forme :

$$\begin{cases} \dot{\delta x}_l(t, \theta, \delta u) = \frac{\partial f}{\partial x}(x_e, \theta, u_e) \delta x_l(t, \theta, \delta u) + \frac{\partial f}{\partial u}(x_e, \theta, u_e) \delta u_l, \\ \dot{\delta y}_l(t, \theta, \delta u) = \frac{\partial h}{\partial x}(x_e, \theta, u_e) \delta x_l(t, \theta, \delta u) + \frac{\partial h}{\partial u}(x_e, \theta, u_e) \delta u_l, \\ \delta x_l(0, \theta, \delta u) = 0, \end{cases} \quad (\text{I.2.12})$$

ou encore :

$$\begin{cases} \dot{\delta x}_l(t, \theta) = A(t, \theta) \delta x_l(t, \theta) + B(t, \theta) \delta u_l(t), \\ \dot{\delta y}_l(t, \theta) = C(t, \theta) \delta x_l(t, \theta) + D(t, \theta) \delta u_l(t), \\ \delta x_l(0, \theta) = 0, \end{cases} \quad (\text{I.2.13})$$

où A, B sont les jacobiniennes de f par rapport à x et u respectivement et C, D sont celles de h en (x_e, θ, u_e) . u_e représente la position de u lorsque l'état est x_e .

Si le modèle linéarisé (I.2.13) est globalement (respectivement localement) identifiable en θ alors le modèle non linéaire (I.2.10) est globalement (resp. localement) identifiable en θ .

Remarque I.2.10 La réciproque de ce résultat est fausse en général. En effet, le système non linéaire décrit par :

$$\begin{cases} \dot{x}_1 = -k_1 x_1 - x_2 + u, \\ \dot{x}_2 = k_1 x_1 - \frac{V_m x_2}{k_m + x_2}, \\ y = x_1, \end{cases} \quad (\text{I.2.14})$$

est globalement identifiable alors que sa forme linéarisée autour d'une trajectoire connue ne l'est pas.

Remarque I.2.11 Une autre forme de linéarisation peut être envisagée en considérant des perturbations sur x et P . Elle est bien adaptée aux systèmes non linéaires, non contrôlés.

B 1.2.4.2 La discernabilité

Les résultats présentés ici sont extraits de l'ouvrage de E. Walter et L. Pronzato.

Lorsque l'on hésite entre plusieurs structures de modèles, pour la description des mêmes données, il est naturel de se demander quelle est la meilleure structure, compte tenu des mesures envisagées. C'est l'étude de la *discernabilité* des structures.

On suppose ici que le processus est de structure $M(\cdot)$ tandis que son modèle est de structure $\hat{M}(\cdot)$ différente de $M(\cdot)$. Le vecteur des paramètres de $\hat{M}(\cdot)$ est noté $\hat{\theta}$, celui des paramètres de $M(\cdot)$ est noté θ . Ils ne sont pas nécessairement de même dimension. Puisque le processus et son modèle n'ont plus la même structure, il peut devenir impossible de régler les paramètres $\hat{\theta}$ du modèle de façon à obtenir un comportement entrée-sortie identique à celui du processus. C'est cette impossibilité qui pourra permettre d'éliminer la structure $\hat{M}(\cdot)$ au profit de la structure $M(\cdot)$.

Définition I.2.12 *La structure de modèle $\hat{M}(\cdot)$ est structurellement discernable de la structure de modèle $M(\cdot)$ si, pour presque toute valeur admissible p des paramètres de $M(\cdot)$, il n'existe aucune valeur admissible $\hat{\theta}$ des paramètres de $\hat{M}(\cdot)$ telle que $\hat{M}(\hat{\theta}) = M(\theta)$.*

Remarque I.2.13 *Le fait que $\hat{M}(.)$ soit structurellement discernable de $M(.)$ n'entraîne pas que la réciproque soit vraie.*

Définition I.2.14 *Quand $\hat{M}(.)$ est structurellement discernable de $M(.)$ et $M(.)$ est structurellement discernable de $\hat{M}(.)$ alors $M(.)$ et $\hat{M}(.)$ sont dites structurellement discernables.*

Les techniques de test de discernabilité sont identiques à celles présentées dans la section portant sur l'identifiabilité. Il y a cependant deux différences dans l'exploitation des résultats. En effet, pour l'étude de la discernabilité, au lieu de considérer deux modèles de mêmes structures, on considère deux modèles de structures différentes. De plus, on espère montrer la non-existence d'une solution en $\hat{\theta}$.

Exercice 1

Pour décrire un système, on hésite entre deux représentations. La première est décrite par l'équation différentielle :

$$\frac{d^2y(t)}{dt^2} = \theta_1 \frac{dy(t)}{dt} + \theta_2 y(t) + \theta_3 u(t).$$

La deuxième est décrite par les équations :

$$\begin{cases} \frac{dy(t)}{dt} = \theta_1 y(t) + x(t), \\ \frac{dx(t)}{dt} = \theta_2 x(t) + \theta_3 u(t). \end{cases}$$

- Exprimer les fonctions de transfert $H_1(s, \theta)$ et $H_2(s, \theta)$ de ces deux modèles.
- Etudiez l'identifiabilité de chacun de ces modèles.
- Etudiez la discernabilité de chacun de ces modèles.

Exercice 2

On dispose de mesures d'un signal sinusoïdal prises aux instants t_i , avec $i = 0, 1, \dots, N - 1$. Ces données sont rangées dans un vecteur $\underline{y} = [y(t_0), y(t_1), \dots, y(t_{N-1})]^T$. On désire déterminer les paramètres de la sinusoïde de pulsation ω connue. Pour cela, on dispose de deux modèles, le premier ayant pour paramètre

$\underline{\theta}_1 = [a, b, c]^T$:

$$x_1(t) = a \cos(\omega t) + b \sin(\omega t) + c,$$

le second ayant pour paramètre $\underline{\theta}_2 = [A, \phi, c]^T$:

$$x_2(t) = A \cos(\omega t + \phi) + c.$$

- Etudiez l'identifiabilité structurelle de chacun de ces modèles.
- Etudiez la discernabilité structurelle de chacun de ces modèles.

Chapitre II

Rappels de probabilités

Le volume horaire du cours ne permettant pas de revisiter en profondeur les bases de probabilités, on ne présentera que quelques éléments utiles pour la suite. Vous êtes invités à revoir par vous mêmes certains ouvrages de référence.

II.1 Revoir dans les cours de Licence...

II.1.1 Expérience aléatoire – Événement – Probabilités – Indépendance – Probabilités conditionnelles

- Expérience aléatoire – Événements A, B, \dots – Univers Ω – Tribu T de parties de Ω (famille non vide de parties de Ω stable par complémentation et union dénombrable)
- Définition axiomatique de la probabilité, sur un espace probabilisable (Ω, T)
 $\mathbb{P} : T \rightarrow [0; 1]$ telle que
 - $\mathbb{P}(\Omega) = 1$;
 - si $\cap_i A_i = \emptyset$, alors $\mathbb{P}(\cup_i A_i) = \sum_i \mathbb{P}(A_i)$ pour des unions finies ou dénombrables.
↪ Il vient
 - $\mathbb{P}(\emptyset) = 0$;
 - $\forall A \in T, \mathbb{P}(\bar{A}) = 1 - \mathbb{P}(A)$;
 - $\forall A, B \in T, \mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)$;
 - $\forall A, B \in T, (A \subset B) \implies (\mathbb{P}(A) \leq \mathbb{P}(B))$;
 - Théorème des probabilités totales : si $\{B_i\}$ est une partition de Ω (*i.e.*, les événements B_i sont mutuellement exclusifs et exhaustifs), alors $\mathbb{P}(A) = \sum_i \mathbb{P}(A \cap B_i)$.
- Indépendance d'événements
 - Indépendance de 2 événements ;
 - Indépendance de N événements (qui implique mais n'est pas équivalente à leur indépendance 2 à 2!).
- Probabilités conditionnelles – Formule de Bayes
- ...

II.1.2 Notion de variable aléatoire

- $X \in \mathbb{R}$ variable aléatoire (VA) ; pour l'événement ω , $X(\omega)$ = réalisation x .
 - événements élémentaires $X \leq x$;
 - formation d'événements tels que $X \in I$ ou $X \in I_1 \cup I_2$ ou... par un ensemble dénombrable d'opérations union et intersection ;
 - définition des probabilités d'événements élémentaires :

$$\text{fonction de répartition (cdf)} \quad P_X(x) = \mathbb{P}(X \leq x), \quad (\text{II.1.1})$$

où $P_X(x) \geq 0$;

- définition des probabilités d'autres événements au moyen des axiomes de probabilité ;
- définition de la *densité de probabilité*

$$\text{densité de probabilité (pdf)} \quad p_X(x) = \left[\frac{dP_X(\xi)}{d\xi} \right]_{\xi=x} = \lim_{d\xi \rightarrow 0} \frac{\mathbb{P}(x < X \leq x + d\xi)}{d\xi}, \quad (\text{II.1.2})$$

où : $p_X(x) \geq 0$; $P_X(x) = \int_{-\infty}^x p_X(\xi)d\xi$; $P_X(+\infty) = 1$; $\mathbb{P}(X \in I) = \int_I p_X(\xi)d\xi$; etc.

- si $X \in \mathbb{R}^N$ vecteur aléatoire (VA), alors : $dx = dx_1 \dots dx_N$; $X \leq x$ s'interprète vectoriellement ; etc.
- X est
 - *continue* si $P_X(x)$ est continue
 - *discrete* si $P_X(x)$ est en escalier, auquel cas

$$p_X(x) = \sum_i p_i \delta(x - x_i) \quad \text{où} \quad p_i = \mathbb{P}(X = x_i), \quad (\text{II.1.3})$$

et on note

$$\text{fonction masse (pmf)} \quad \mu_X(x_i) = \mathbb{P}(X = x_i) = p_i, \quad (\text{II.1.4})$$

de sorte que $\mathbb{P}(X \in I) = \sum_{x_i \in I} \mu_X(x_i)$, etc.

II.2 Propriétés des VA

II.2.1 Moments

- Opérateur espérance mathématique

$$\mathbb{E}_X[g(X)] = \int g(x)p_X(x)dx \quad \text{ou} \quad \mathbb{E}_X[g(X)] = \sum_{x_i} g(x_i)\mu_X(x_i) = \sum_{x_i} g(x_i)p_i. \quad (\text{II.2.1})$$

→ c'est un opérateur linéaire

- Espérance/Moyenne

$$m_X = \mathbb{E}_X[X] = \int x p_X(x)dx \quad \text{ou} \quad \sum_{x_i} x_i \mu_X(x_i) = \sum_{x_i} x_i p_i. \quad (\text{II.2.2})$$

→ X centrée $\Leftrightarrow m_X = 0$.

- Moments d'ordre n – Moments centrés d'ordre n , dont

- Variance ($X \in \mathbb{R}$) : $\sigma_X^2 = \mathbb{E}_X[(X - m_X)^2]$;
 - Covariance ($X \in \mathbb{R}^N$) : $\text{Cov}_X = \mathbb{E}_X[(X - m_X)(X - m_X)^T] = \mathbb{E}_X[XX^T] - m_X m_X^T$;
- Nota : $\mathbb{E}_X[(X - m_X)^T(X - m_X)] = \text{trace}(\text{Cov}_X)$.

- Quelques exercices/propriétés à méditer...

$$\forall X \in \mathbb{R}^n \text{ va, } X = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}, \quad \mathbb{E}_X[g(X)] = \int_{\mathbb{R}^n} g(x) p_X(x) dx = \int_{\mathbb{R}^n} \left[\begin{bmatrix} g_1(x_1, \dots, x_n) \\ \vdots \\ g_n(x_1, \dots, x_n) \end{bmatrix} \right] p_{x_1, \dots, x_n}(x_1, \dots, x_n) dx_1 \dots dx_n$$

• $\mathbb{E}[\cdot]$ est linéaire $\Rightarrow \forall (\lambda, \mu) \in \mathbb{R} \times \mathbb{R}$ déterminées, $\forall (x, y) \in \mathbb{R} \times \mathbb{R}$ V.A.s, $\mathbb{E}_{x,y}[\lambda x + \mu y] = \lambda \mathbb{E}_x[x] + \mu \mathbb{E}_y[y]$
 → se démontre immédiatement en développant $\int_{\mathbb{R}^2} (\lambda x + \mu y) p_{x,y}(x,y) dx dy$; en effet,

$$= \lambda \int_{\mathbb{R}^2} x p_{x,y}(x,y) dx dy + \mu \int_{\mathbb{R}^2} y p_{x,y}(x,y) dx dy = \lambda \int_{\mathbb{R}} x \left[\int_{\mathbb{R}} p_{x,y}(x,y) dy \right] dx + \mu \int_{\mathbb{R}} y \left[\int_{\mathbb{R}} p_{x,y}(x,y) dy \right] dx = p_x(x) = p_y(y)$$

• Pour les mêmes raisons, quelle que soit les matrices (de dimensions compatibles) M et N , déterminées, et X , aléatoire,
 $\mathbb{E}[X^T] = \mathbb{E}[X]^T$, $\mathbb{E}[MXN] = M \mathbb{E}[X]N$, etc.

$$\bullet x \in \mathbb{R} \Rightarrow \bar{x} = \mathbb{E}[x] = \int_{\mathbb{R}} x p_x(x) dx \text{ et } \sigma_x^2 = \mathbb{E}[(x - \mathbb{E}[x])^2] = \int_{\mathbb{R}} (x - \bar{x})^2 p_x(x) dx$$

$$\rightarrow \text{NOTA : } \sigma_x^2 = \mathbb{E}[(x - \mathbb{E}[x])^2] = \mathbb{E}[x^2 - 2x\mathbb{E}[x] + \mathbb{E}[x]^2]$$

$$= \mathbb{E}[x^2] - \mathbb{E}[2\mathbb{E}[x]x] + \mathbb{E}[\mathbb{E}[x]^2]$$

$$= \mathbb{E}[x^2] - 2\mathbb{E}[x]\mathbb{E}[x] + \mathbb{E}[x]^2$$

$$\sigma_x^2 = \mathbb{E}[(x - \mathbb{E}[x])^2] = \mathbb{E}[x^2] - \mathbb{E}[x]^2$$

$$\bullet x \in \mathbb{R}^n \Rightarrow \bar{x} = \mathbb{E}[x] \in \mathbb{R}^n \text{ et } \text{Cov}_x = \mathbb{E}[(x - \mathbb{E}[x])(x - \mathbb{E}[x])^T] = \mathbb{E}[xx^T - x\mathbb{E}[x]^T - \mathbb{E}[x]x^T + \mathbb{E}[x]\mathbb{E}[x]^T]$$

$$= \mathbb{E}[xx^T] - \mathbb{E}[x\mathbb{E}[x]^T] - \mathbb{E}[\mathbb{E}[x]x^T] + \mathbb{E}[\mathbb{E}[x]\mathbb{E}[x]^T]$$

$$= \mathbb{E}[xx^T] - 2\mathbb{E}[x]\mathbb{E}[x]^T + \mathbb{E}[x]\mathbb{E}[x]^T$$

$$\text{Cov}_x = \mathbb{E}[(x - \mathbb{E}[x])(x - \mathbb{E}[x])^T] = \mathbb{E}[xx^T] - \mathbb{E}[x]\mathbb{E}[x]^T$$

\rightarrow NOTA : $\forall z \in \mathbb{R}^n, (z - \bar{z}) \in \mathbb{R}^n \Rightarrow (z - \bar{z})(z - \bar{z})^T \in \mathbb{R}^{nn}$ et symétrique semi-définie positive de rang 1.

\downarrow
 $\text{Gr}_x \in \mathbb{R}^{n \times n}$ est une matrice symétrique semi-définie positive (voire définie positive !)

Rappel : $\forall M \in \mathbb{R}^{n \times n}, \forall \xi \in \mathbb{R}^n$, on définit par $\xi^T M \xi$ la forme quadratique associée à M . Or, $\xi^T M \xi$ est toujours égal à $\frac{\xi^T M \xi}{2}$.
 \Rightarrow SPDG, on suppose $M = M^T$ et on dit que $M \geq 0$ (semi-définie positive) car $\forall \xi, \xi^T M \xi \geq 0$
 $M > 0$ (définie positive) car $\forall \xi \neq 0, \xi^T M \xi > 0$

⚠ Il peut être indéfinie !

$M = M^T \Rightarrow$ les v.p de M sont réelles et les T.p de M peuvent être dans orthogonaux

$\Leftrightarrow M \geq 0 \Leftrightarrow$ toutes les v.p de M sont > 0 car les t.p des matrices minores diagonales de M sont > 0

$M \geq 0 \Leftrightarrow \dots \geq 0 \Leftrightarrow \exists N$ telle que $M = NN^T$ (on sait, $\xi^T M \xi = (\xi^T N)(N \xi) = \|N\xi\|^2 \geq 0 \forall \xi$!)

$$\bullet \text{Note : } X = \begin{bmatrix} X_1 \\ \vdots \\ X_n \end{bmatrix} \rightarrow \mathbb{E}[x] = \int_{\mathbb{R}^n} x p_x(x) dx = \int_{\mathbb{R}^n} \int_{\mathbb{R}^n} \begin{bmatrix} e_1^T x \\ \vdots \\ e_n^T x \end{bmatrix} p_{X_1 \dots X_n}(x_1 \dots x_n) dx_1 \dots dx_n, \text{ où } e_i = \begin{bmatrix} 0 \\ \vdots \\ 1 \\ \vdots \\ 0 \end{bmatrix} \leftarrow \# i$$

$$\text{et bien } \int_{\mathbb{R}^n} \int_{\mathbb{R}^n} e_i^T x p_{X_1 \dots X_n}(x_1 \dots x_n) dx_1 \dots dx_n = \int_{\mathbb{R}^n} e_i \underbrace{\left[\int_{\mathbb{R}^n} \int_{\{j \neq i\}} p_{X_1 \dots X_n}(x_1 \dots x_n) \prod_{j \neq i} dx_j \right]}_{= p_{X_i}(x_i)} dx_i = \mathbb{E}[X_i]$$

II.2.2 Loi jointe d'un couple de VAs – Lois marginales – Lois conditionnelles

- Rappel : loi jointe de X_1, X_2 (cas continu)

$$p_{X_1 X_2}(x_1, x_2) = \frac{\partial P_{X_1 X_2}(x_1, x_2)}{\partial x_1 \partial x_2}. \quad (\text{II.2.3})$$

$$\hookrightarrow \mathbb{E}_{X_1 X_2}[g(X_1, X_2)] = \iint g(x_1, x_2) p_{X_1 X_2}(x_1, x_2) dx_1 dx_2; \text{ etc.}$$

- *Lois marginales*

$$p_X(x) = \int p_{XY}(x, y) dy, \quad \text{ou} \quad \mu_X(x_i) = \sum_j \mu_{XY}(x_i, y_j). \quad (\text{cf. théo. proba. totales}) \quad (\text{II.2.4})$$

- *Lois conditionnelles*

$$p_{X|Y}(x|y) = \frac{p_{XY}(x, y)}{p_Y(y)} = \frac{p_{XY}(x, y)}{\int p_{XY}(x, y) dx} \quad \text{si } p_Y(y) \neq 0, \quad (\text{Bayes pdf}) \quad (\text{II.2.5})$$

ou

$$\mu_{X|Y}(x_i|y_j) = \frac{\mu_{XY}(x_i, y_j)}{\mu_Y(y_j)} = \frac{\mu_{XY}(x_i, y_j)}{\sum_i \mu_{XY}(x_i, y_j)} \quad \text{si } \mu_Y(y_j) \neq 0. \quad (\text{Bayes proba.}) \quad (\text{II.2.6})$$

- *Indépendance – Corrélation*

- X_1, \dots, X_N indépendantes $\Leftrightarrow \forall x_1, \dots, x_N, p_{X_1 \dots X_N}(x_1, \dots, x_N) = p_{X_1}(x_1) \dots p_{X_N}(x_N)$
ou $\forall x_1, \dots, x_N, \mu_{X_1 \dots X_N}(x_1, \dots, x_N) = \mu_{X_1}(x_1) \dots \mu_{X_N}(x_N)$.
- X_1, \dots, X_N i.i.d. $\Leftrightarrow X_1, \dots, X_N$ indépendantes et $p_{X_1}(x_1) = \dots = p_{X_N}(x_N)$.
- X, Y non corrélées $\Leftrightarrow \mathbb{E}_{XY}[XY^T] = \mathbb{E}_X[X]\mathbb{E}_Y[Y]^T \Leftrightarrow C_{XY} = 0$.
 $\hookrightarrow X, Y$ indépendantes $\Rightarrow X, Y$ non corrélées (et équivalence si $(X^T, Y^T)^T$ Gaussien).
- Pour $X_1 \in \mathbb{R}, X_2 \in \mathbb{R}$, $\text{cov}(X_1, X_2)$ est liée à $\sigma_{X_1}^2, \sigma_{X_2}^2$ par $\rho_{12} = \frac{\text{cov}(X_1, X_2)}{\sigma_{X_1}\sigma_{X_2}}$, avec
 - $|\rho_{12}| \leq 1$;
 - $\rho_{12} = 0 \Leftrightarrow X_1, X_2$ non corrélées;
 - $\rho_{12} = 1 \Leftrightarrow \exists(\alpha_1, \alpha_2) \neq (0, 0), \alpha_1 X_1 + \alpha_2 X_2 = 0$, auquel cas $C_{(X_1)(X_2)}$ est singulière...

II.2.3 Quelques autres résultats

- *Densité de probabilité de la somme de deux VA indépendantes*

Si $Y = X_1 + X_2$, avec X_1, X_2 indépendantes,
alors

$$p_Y(y) = \int p_{X_1}(y - x)p_{X_2}(x)dx. \quad (\text{II.2.7})$$

- *Théorème central limite*

Si X_1, \dots, X_N indépendantes, de mêmes lois, mais avec moyennes m_i et covariances P_i possiblement distinctes ;
si $Y_N = \sum_{n=1}^N X_n$;
si $Z_N = C_{Y_N}^{-\frac{1}{2}}(Y_N - \mathbb{E}[Y_N])$, où $L^{\frac{1}{2}}$ tq $L^{\frac{1}{2}}(L^{\frac{1}{2}})^T = L$;
(Nota : $\mathbb{E}[Y_N] = \sum_i m_i$ et $C_{Y_N} = \sum_i P_i$)
alors, lorsque $N \rightarrow +\infty$, Z_N converge (faiblement, en loi) vers $\mathcal{N}(0, \mathbb{I})$.

- *Loi des espérances itérées*

$$\mathbb{E}_Y[\mathbb{E}_{X|Y}[X|y]] = \mathbb{E}_X[X]. \quad (\text{II.2.8})$$

- *Changement de variable*

Si $Y = \phi(X)$, où $\phi^{-1}(.)$ existe et $\phi(.), \phi^{-1}(.)$ sont C^1
alors

$$p_Y(y) = \left| \det \underbrace{\frac{\partial \phi^{-1}(y)}{\partial y^T}}_{\text{Jacobienne de } \phi^{-1}(.)} \right| p_X(\phi^{-1}(y)). \quad (\text{II.2.9})$$

II.2.4 Lois Gaussiennes et de chi-deux

- Soit $X \in \mathbb{R}$. $X \sim \mathcal{N}(\bar{x}, \sigma^2)$ ssi

$$p_X(x) = \mathcal{N}(x; \bar{x}, \sigma^2) \triangleq \frac{1}{\sqrt{2\pi}\sigma} \exp -\frac{(x-\bar{x})^2}{2\sigma^2}. \quad (\text{II.2.10})$$

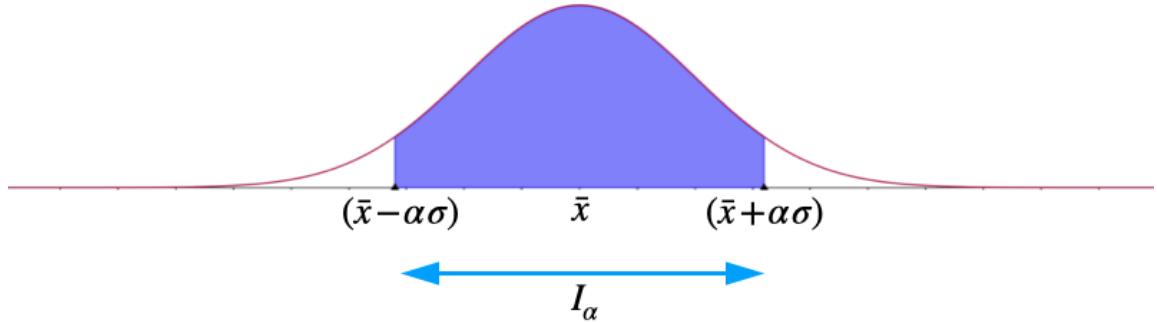
- Soit $X \in \mathbb{R}^N$. $X \sim \mathcal{N}(\bar{x}, P)$ ssi

$$p_X(x) = \mathcal{N}(x; \bar{x}, P) \triangleq \frac{1}{\sqrt{(2\pi)^N \det(P)}} \exp \left(-\frac{1}{2}(x-\bar{x})^T P^{-1}(x-\bar{x}) \right), \quad (\text{II.2.11})$$

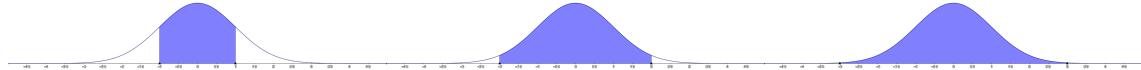
$$\triangleq \frac{1}{\sqrt{\det(2\pi P)}} \exp \left(-\frac{1}{2}(x-\bar{x})^T P^{-1}(x-\bar{x}) \right). \quad (\text{II.2.12})$$

- *Ensembles de confiance*

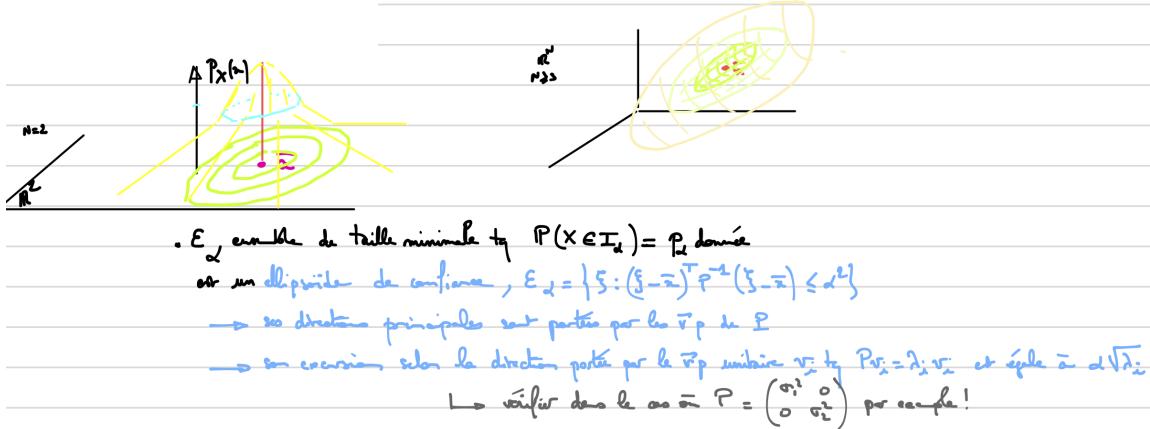
- Pour $X \in \mathbb{R}$, $X \sim \mathcal{N}(\bar{x}, \sigma^2)$, I_α de taille minimale tel que $\mathbb{P}(X \in I_\alpha) = p_\alpha$ donnée est l'intervalle de confiance $I_\alpha = [\bar{x} - \alpha\sigma; \bar{x} + \alpha\sigma] = \left\{ \xi : \frac{(\xi-\bar{x})^2}{\sigma^2} \leq \alpha^2 \right\}$



- Sous cette hypothèse, $Y = \frac{X-\bar{x}}{\sigma} \sim \mathcal{N}(0, 1)$ dont voici les intervalles de confiance I_1, I_2, I_3 (avec les probas associées $p_1 = 0.6827$, $p_2 = 0.9545$, $p_3 = 0.9973$) :



- Pour $X \in \mathbb{R}^N$, $X \sim \mathcal{N}(\bar{x}, P)$



- Fonction affine d'un VA Gaussien

Soit $X \in \mathbb{R}^N$ un VA. Soient A, b une matrice et un vecteur déterministes.

Alors,

$$X \sim \mathcal{N}(\bar{x}, P) \implies Y = AX + b \sim \mathcal{N}(A\bar{x} + b, APA^T). \quad (\text{II.2.13})$$

- Loi conditionnelle de deux VA conjointement Gaussiens (ou : Gaussiens dans leur ensemble)

Soient X et Z Gaussiens dans leur ensemble (“jointly Gaussian”).

Alors, conditionnellement à l'événement $Z = z$, la distribution de X est elle-même Gaussienne. La formule ci-dessous explicite ses moments a posteriori.

$$\begin{aligned} \binom{X}{Z} \text{ tel que } p_{X,Z}(x, z) = \mathcal{N}\left(\binom{x}{z}; \binom{m_X}{m_Z}, \begin{pmatrix} P_{XX} & P_{XZ} \\ P_{ZX} & P_{ZZ} \end{pmatrix}\right) \\ \Downarrow \\ p_{X|Z}(x|z) = \mathcal{N}(x; m_{X|Z}, P_{X|Z}) \quad (\text{II.2.14}) \end{aligned}$$

avec

$$\begin{cases} m_{X|Z} \triangleq \mathbb{E}[X|Z=z] & = m_X + P_{XZ}P_{ZZ}^{-1}(z - m_Z) \\ P_{X|Z} \triangleq \mathbb{E}[(X - m_{X|Z})(X - m_{X|Z})^T | Z=z] & = P_{XX} - P_{XZ}P_{ZZ}^{-1}P_{ZX}. \end{cases}$$

- La preuve exploite la règle de Bayes (II.2.5). On obtient un quotient de deux lois Gaussiennes (lesquelles ?) qui se « simplifie » en la loi conditionnelle dont on recherche les moments.
- Si X et Z sont non corrélés, alors les moments a posteriori $m_{X|Z}, P_{X|Z}$ de X conditionnellement à $Z = z$ sont égaux aux moments a priori m_X, P_{XX} de X .
- Si $P_{ZZ} = \infty \mathbb{I}$, alors les moments a posteriori $m_{X|Z}, P_{X|Z}$ de $X|Z=z$ sont égaux aux moments a priori m_X, P_{XX} de X .
- $P_{X|Z}$ ne dépend pas de la réalisation z de Z !

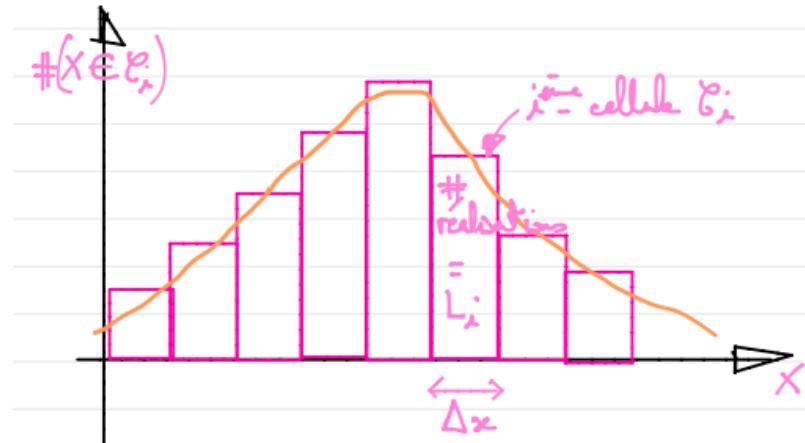
- Loi de chi-deux à N degrés de liberté

- $X_1, \dots, X_N \stackrel{\text{i.i.d.}}{\sim} \mathcal{N}(0, 1), Y = X_1^2 + \dots + X_N^2 \implies Y \sim \chi_N^2$.
- $X \in \mathbb{R}^N, X \sim \mathcal{N}(\bar{x}, P), Z = (X - \bar{x})^T P^{-1} (X - \bar{x}) \implies Z \sim \chi_N^2$.

II.2.5 Détermination empirique d'une densité de probabilité

- Soit \mathcal{H}_X un histogramme de X , avec
 - Δx la taille des cellules ;
 - M le nombre total d'échantillons (réalisations) de X ;
 - L_i le nombre d'échantillons de X dans chaque cellule no i de centre x_i .

Alors,



$$\begin{aligned} \# \text{ total échantillons } M &= \sum_i L_i \\ \Rightarrow P(X \in B_i) &= \frac{L_i}{M} \\ \Rightarrow \hat{p}(x_i \text{ centre de } B_i) &= \frac{L_i / M}{\Delta x} \end{aligned}$$

$$\text{en fait, } \hat{p}_x(x_i) = \frac{1}{\Delta x} \int_{x_i - \frac{\Delta x}{2}}^{x_i + \frac{\Delta x}{2}} p(u) du$$

Chapitre III

Propriétés des estimateurs

III.1 Biais – Covariance – Erreur quadratique moyenne

Dans tout ce qui suit, $\hat{\Theta} = g(Z)$.

III.1.1 Biais d'un estimateur

- Fisher

$$b_{\hat{\Theta}}(\theta) = \mathbb{E}_{Z|\theta}[\hat{\Theta} - \theta | \theta] = \mathbb{E}_{Z|\theta}[\hat{\Theta} | \theta] - \theta = \mathbb{E}_{Z|\theta}[g(Z) | \theta] - \theta. \quad (\text{III.1.1})$$

- $\hat{\Theta}$ non biaisé \Leftrightarrow *Quel que soit* θ , $b_{\hat{\Theta}}(\theta) = 0$.
- Ci-dessus, $\mathbb{E}_{Z|\theta}[g(Z) | \theta] = \int g(z) p_{Z|\theta}(z | \theta) dz$.

- Bayes

$$b_{\hat{\Theta}} = \mathbb{E}_{Z,\Theta}[\hat{\Theta} - \Theta] = \mathbb{E}_{Z,\Theta}[\hat{\Theta}] - \mathbb{E}_{\Theta}[\Theta] = \mathbb{E}_{Z,\Theta}[g(Z)] - \mathbb{E}_{\Theta}[\Theta]. \quad (\text{III.1.2})$$

- $\hat{\Theta}$ non biaisé $\Leftrightarrow b_{\hat{\Theta}} = 0$. $b_{\hat{\Theta}}$ est parfois appelé « biais moyen ».
- On a aussi $b_{\hat{\Theta}} = \mathbb{E}_{\Theta}[b_{\hat{\Theta}}(\theta)] = \mathbb{E}_{\Theta}\left[\mathbb{E}_{Z|\theta}[\hat{\Theta} | \theta] - \theta\right]$, car $\mathbb{E}_{Z,\Theta}[g(Z)] = \mathbb{E}_{\Theta}\left[\mathbb{E}_{Z|\Theta}[\hat{\Theta} | \theta]\right]$.
- En effet, ci-dessus, $\mathbb{E}_{Z,\Theta}[g(Z)] = \iint g(z) p_{Z,\Theta}(z, \theta) dz = \int \left[\int g(z) p_{Z|\Theta}(z | \theta) dz \right] p_{\Theta}(\theta) d\theta$.

III.1.2 Covariance d'un estimateur

- Fisher

$$\text{Cov}_{\hat{\Theta}}(\theta) = \mathbb{E}_{Z|\theta}\left[\left(\hat{\Theta} - \mathbb{E}_{Z|\theta}[\hat{\Theta} | \theta]\right)\left(\hat{\Theta} - \mathbb{E}_{Z|\theta}[\hat{\Theta} | \theta]\right)^T \middle| \theta\right]. \quad (\text{III.1.3})$$

- Bayes

$$\text{Cov}_{\hat{\Theta}} = \mathbb{E}_{Z,\Theta}\left[\left(\hat{\Theta} - \mathbb{E}_{Z,\Theta}[\hat{\Theta}]\right)\left(\hat{\Theta} - \mathbb{E}_{Z,\Theta}[\hat{\Theta}]\right)^T\right]. \quad (\text{III.1.4})$$

- Et $\text{Cov}_{\hat{\Theta}} = \mathbb{E}_{\Theta}[\text{Cov}_{\hat{\Theta}}(\theta)]$.

III.1.3 Erreur quadratique moyenne d'un estimateur

- Fisher

$$\text{EQM}_{\hat{\Theta}}(\theta) = \mathbb{E}_{Z|\theta} \left[(\hat{\Theta} - \theta)^T (\hat{\Theta} - \theta) \mid \theta \right] \quad (\text{III.1.5})$$

$$= \text{trace} \left(\text{Cov}_{\hat{\Theta}}(\theta) + b_{\hat{\Theta}}(\theta) b_{\hat{\Theta}}(\theta)^T \right). \quad (\text{III.1.6})$$

○ L'équation ci-dessus montre que l'EQM constitue un *compromis biais-variance*.

○ Preuve :

$$\text{EQM}_{\hat{\Theta}}(\theta) = \text{trace} \left(\mathbb{E}_{Z|\theta} \left[(\hat{\Theta} - \theta) (\hat{\Theta} - \theta)^T \mid \theta \right] \right);$$

$$\text{et } (\hat{\Theta} - \theta) (\hat{\Theta} - \theta)^T = (\hat{\Theta} - m_{\hat{\Theta}}(\theta) + m_{\hat{\Theta}}(\theta) - \theta) (\hat{\Theta} - m_{\hat{\Theta}}(\theta) + m_{\hat{\Theta}}(\theta) - \theta)^T,$$

$$\text{où } m_{\hat{\Theta}}(\theta) = \mathbb{E}_{Z|\theta} [\hat{\Theta} \mid \theta] \text{ et } b_{\hat{\Theta}}(\theta) = m_{\hat{\Theta}}(\theta) - \theta.$$

- Bayes

$$\text{EQM}_{\hat{\Theta}} = \mathbb{E}_{Z,\Theta} \left[(\hat{\Theta} - \Theta)^T (\hat{\Theta} - \Theta) \right]. \quad (\text{III.1.7})$$

III.1.4 Exercices : Calcul du biais et de la variance des estimateurs proposés dans les Exemples 1 et 2.1

$$z_m = a + w_m, \quad a \text{ déterminé}, \quad w_1, \dots, w_N \stackrel{iid}{\sim} \mathcal{N}(0, \sigma^2)$$

$$\hat{A} = \frac{1}{N} \sum_{m=1}^N z_m - \text{biais et (co)variance de } \hat{A}?$$

$$*\text{Calcul de } \mathbb{E}[\hat{A}] \text{ (on peut noter } \mathbb{E}_{Z|a} [\hat{A}|a]$$

$$\bullet \mathbb{E}[\hat{A}] = \frac{1}{N} \sum_{m=1}^N \mathbb{E}[z_m] = \frac{1}{N} \sum_{m=1}^N [a + \mathbb{E}[w_m]] = \frac{1}{N} \cdot [Na] = a$$

$$\Rightarrow \forall a, b_{\hat{A}}(a) = \mathbb{E}_{Z|a} [(\hat{A} - a) \mid a] = 0 \Leftrightarrow \hat{A} \text{ est non biaisé}$$

$$*\text{Calcul de } \text{Var}_{\hat{A}}(a) = \mathbb{E}_{Z|a} \left[(\hat{A} - \mathbb{E}_{Z|a} [\hat{A} \mid a])^2 \mid a \right] = \mathbb{E}[\hat{A}^2] - \mathbb{E}[\hat{A}]^2$$

$$\bullet \text{Var}_{\hat{A}}(a) = \mathbb{E}[(\hat{A} - a)^2] = \mathbb{E}[\hat{A}^2] - a^2$$

$$= \mathbb{E} \left[\left(\frac{1}{N} \sum_{m=1}^N z_m \right)^2 \right] - a^2 = \frac{1}{N^2} \mathbb{E} \left[\left(\sum_{m=1}^N z_m \right) \left(\sum_{m=1}^N z_m \right) \right] - a^2$$

$$= \frac{1}{N^2} \mathbb{E} \left[\sum_{m=1}^N z_m^2 + \sum_{m \neq n} \sum_{m=1}^N z_m z_n \right] - a^2$$

$$\bullet \text{Or, comme } z_1, \dots, z_N \stackrel{iid}{\sim} \mathcal{N}(a, \sigma^2), \text{ on a : } \mathbb{E}[z_m^2] = \sigma^2 + a^2 \text{ et } \mathbb{E}[z_m z_n] = 0 + a^2$$

$$\text{d'où } \text{Var}_{\hat{A}}(a) = \frac{1}{N^2} \left(\sum_{m=1}^N \mathbb{E}[z_m^2] + \sum_{m \neq n} \mathbb{E}[z_m z_n] \right) - a^2 = \frac{1}{N^2} \left[(N \sigma^2 + N a^2) + (N^2 - N) a^2 \right] - a^2 = \frac{\sigma^2}{N}.$$

III.1.5 Estimateur non biaisé à minimum de variance : Estimateur efficace

- Soit $\hat{\Theta} = g(Z)$ de biais nul. $\hat{\Theta}$ est un estimateur efficace si pour tout estimateur $\hat{\Theta} = f(Z)$ également non biaisé, on a $\text{Cov}_{\hat{\Theta}} \leq \text{Cov}_{\hat{\Theta}}$, i.e., $(\text{Cov}_{\hat{\Theta}} - \text{Cov}_{\hat{\Theta}})$ définie positive.
- $\hat{\Theta}$ efficace $\Rightarrow \hat{\Theta}$ non biaisé à minimum de variance.
 - cf. cas scalaire, où l'inégalité matricielle devient une inégalité scalaire...

III.1.6 Inégalité de Cramér-Rao (Cas de l'estimation classique)

Soit $\theta \in \mathbb{R}^M$ déterministe inconnu (fixe), Z une variable aléatoire liée à θ , et $p_{Z|\theta}(z|\theta)$ différentiable par rapport à θ .

- On définit la *Matrice d'Information de Fisher*

$$I(\theta) = \mathbb{E}_{Z|\theta} \left[\frac{\partial \ln p_{Z|\theta}(z|\theta)}{\partial \theta} \frac{\partial \ln p_{Z|\theta}(z|\theta)}{\partial \theta^T} \mid \theta \right] \quad (\text{III.1.8})$$

$$= -\mathbb{E}_{Z|\theta} \left[\frac{\partial^2 \ln p_{Z|\theta}(z|\theta)}{\partial \theta \partial \theta^T} \mid \theta \right]. \quad (\text{III.1.9})$$

- Soit $\hat{\Theta} = g(Z)$ un estimateur *non biaisé* quelconque de θ , alors

$$\mathbb{E}_{Z|\theta} [(\hat{\Theta} - \theta)(\hat{\Theta} - \theta)^T \mid \theta] \geq I^{-1}(\theta). \quad (\text{Inégalité de Cramér-Rao}) \quad (\text{III.1.10})$$

- Il s'agit d'une inégalité matricielle.
- Cette borne inférieure peut ne jamais être atteinte.
- Par conséquent, $(\hat{\Theta} \text{ non biaisé de variance } I^{-1}(\theta)) \Rightarrow (\hat{\Theta} \text{ efficace})$!
- Il existe aussi une borne inférieure pour tout estimateur, éventuellement biaisé.

III.2 Propriétés asymptotiques des estimateurs

Étant donné un estimateur $\hat{\Theta}$ de θ obtenu sur la base d'un échantillon de taille N , il s'agit de caractériser ses propriétés pour $N \rightarrow +\infty$.

Pour cela, on définit :

- Des notions de convergence stochastique : en probabilité ; en moyenne quadratique ; presque partout.
- La notion d'estimateur asymptotiquement non biaisé ; asymptotiquement efficace ; asymptotiquement efficace et normal.

Chapitre IV

Estimateurs de Fisher

θ est supposé déterministe inconnu. Le modèle d'observation est $p_{Z|\theta}(z|\theta)$ ou $\mu_{Z|\theta}(z|\theta)$, selon que Z est un VA continu ou discret.

IV.1 Estimateur du maximum de vraisemblance (EMV, MLE)

- Notion de *fonction de vraisemblance (likelihood)*

$$L(\theta; z) = p_{Z|\theta}(z|\theta) \text{ ou } \mu_{Z|\theta}(z|\theta). \quad (\text{IV.1.1})$$

- Anti-log-vraisemblance : $NLL(\theta; z) = -\ln L(\theta; z).$
- ATTENTION ! (Dans le cas continu, mais les résultats se transposent au cas discret)
 $\int L(\theta; z)dz = 1$, mais rien de particulier pour $\int L(\theta; z)d\theta !$

- Définition de l'*estimateur du maximum de vraisemblance* $\hat{\Theta}_{\text{MLE}}$

$$\hat{\Theta}_{\text{MLE}} \text{ tel que } \hat{\theta}_{\text{MLE}} = \arg \max_{\theta} p_{Z|\theta}(z|\theta) \quad (\text{IV.1.2})$$

$$= \arg \max_{\theta} L(\theta; z) \quad (\text{IV.1.3})$$

$$= \arg \min_{\theta} NLL(\theta; z). \quad (\text{IV.1.4})$$

- Illustration : qui de $\theta_1 = 1$ et $\theta_2 = 5$ explique le mieux le fait que $Z(\omega) = 2$ suit la loi $\mathcal{N}(0, 1)$?

- Propriétés importantes de $\hat{\Theta}_{\text{MLE}}$

- Invariance par reparamétrisation

$$(\phi = f(\theta)) \Rightarrow (\hat{\phi}_{\text{MLE}} = f(\hat{\theta}_{\text{MLE}})) \text{ si } f(\cdot) \text{ bijectif.} \quad (\text{IV.1.5})$$

- Sous quelques hypothèses de « régularité » de $p_{Z|\theta}(z|\theta)$, $\hat{\Theta}_{MLE}$ est asymptotiquement efficace et normal, *i.e.*,

$$\lim_{N \rightarrow +\infty} \hat{\Theta}_{MLE} \sim \mathcal{N}(\theta, I^{-1}(\theta)). \quad (\text{IV.1.6})$$

IV.2 Cas d'un modèle d'observation avec perturbations additives Gaussiennes

- Si $Z = h(\theta) + V$, où $Z \in \mathbb{R}^N$, $\theta \in \mathbb{R}^M$, $N \geq M$, $V \in \mathbb{R}^N$, $V \sim \mathcal{N}(0, R)$ avec $R > 0$ (*e.g.*, $R = \sigma^2 \mathbb{I}_{N \times N}$ si V_i i.i.d.), alors

$$\hat{\Theta}_{MLE} \text{ tel que } \hat{\theta}_{MLE} = \arg \min_{\theta} (z - h(\theta))^T R^{-1} (z - h(\theta)). \quad (\text{IV.2.1})$$

- Preuve : à démontrer !

IV.3 Cas d'un modèle d'observation linéaire avec perturbations additives Gaussiennes

- Si $Z = H\theta + V$, où $Z \in \mathbb{R}^N$, $\theta \in \mathbb{R}^M$, $N \geq M$, $V \in \mathbb{R}^N$, $V \sim \mathcal{N}(0, R)$ avec $R > 0$ (*e.g.*, $R = \sigma^2 \mathbb{I}_{N \times N}$ si V_i i.i.d.), H déterministe de rang plein M , alors

$$\hat{\Theta}_{MLE} \text{ tel que } \hat{\theta}_{MLE} = (H^T R^{-1} H)^{-1} H^T R^{-1} z. \quad (\text{IV.3.1})$$

- Preuve : à démontrer !
- Montrer que $\hat{\Theta}_{MLE}$ admet un biais nul, *i.e.*,

$$\forall \theta, b_{\hat{\Theta}_{MLE}}(\theta) = 0. \quad (\text{IV.3.2})$$

$Z = H\theta + V$, $V \sim \mathcal{N}(0, R)$, $\hat{\Theta}_{MLE} = (H^T R^{-1} H)^{-1} H^T R^{-1} Z$, où $H^T R^{-1} H$ inversible \Rightarrow d'après la hypothèse

* Biais $b_{\hat{\Theta}_{MLE}}(\theta)$ de $\hat{\Theta}_{MLE}$

$\bullet b_{\hat{\Theta}_{MLE}}(\theta) = E[\hat{\Theta}_{MLE}] - \theta = \left(E[(H^T R^{-1} H)^{-1} H^T R^{-1} (H\theta + V)] \right) - \theta = \theta + (H^T R^{-1} H)^{-1} H^T R^{-1} E[V] - \theta = 0$

$\Rightarrow \hat{\Theta}_{MLE}$ est non biaisé

- Montrer que la covariance de l'estimateur et l'erreur quadratique moyenne satisfont

$$\forall \theta, \text{Cov}_{\hat{\Theta}_{MLE}}(\theta) = (H^T R^{-1} H)^{-1}. \quad (\text{IV.3.3})$$

$$\forall \theta, \text{EQM}_{\hat{\Theta}_{MLE}}(\theta) = \text{trace}((H^T R^{-1} H)^{-1}). \quad (\text{IV.3.4})$$

* Covariance $C_{\hat{\theta}_{\text{ML}}}(\Theta)$ de $\hat{\theta}_{\text{ML}}$

$$\begin{aligned} \cdot C_{\hat{\theta}_{\text{ML}}}(\Theta) &= \mathbb{E}_{z|\Theta} \left[(\hat{\theta}_{\text{ML}} - \mathbb{E}_{z|\Theta}[\hat{\theta}_{\text{ML}}|\Theta]) (\Delta)^T |\Theta \right] = \mathbb{E} \left[(\hat{\theta}_{\text{ML}} - \Theta) (\Delta)^T \right] = \mathbb{E} \left[\left[(H^T R^{-1} H)^{-1} H^T R^{-1} V \right] [\Delta] \right]^T \\ &= \mathbb{E} \left[(H^T R^{-1} H)^{-1} H^T R^{-1} V V^T R^{-1} H (H^T R^{-1} H)^{-1} \right] \\ &= (H^T R^{-1} H)^{-1} H^T R^{-1} \mathbb{E}[V V^T] R^{-1} H (H^T R^{-1} H)^{-1} \\ &= (H^T R^{-1} H)^{-1} \overset{(1)}{(H^T R^{-1} H)^{-1} R^{-1} H} (H^T R^{-1} H)^{-1} = (H^T R^{-1} H)^{-1} \\ \Rightarrow \forall \Theta, \quad \text{Cov}_{\hat{\theta}_{\text{ML}}}(\Theta) &= (H^T R^{-1} H)^{-1} \end{aligned}$$

* $\hat{\theta}_{\text{ML}}$ est efficace! En effet, il est non biaisé et

$$\begin{aligned} \cdot P_{z|\Theta}(z|\Theta) &= dP(z; \Theta, R) = \frac{1}{\sqrt{(2\pi)^n \det(R)}} e^{-\frac{1}{2} (z - \Theta)^T R^{-1} (z - \Theta)} \\ \cdot -\ln P_{z|\Theta}(z|\Theta) &= C + \frac{1}{2} (z - \Theta)^T R^{-1} (z - \Theta) \\ \cdot \frac{\partial}{\partial \Theta} \left[-\ln P_{z|\Theta}(z|\Theta) \right] &= -H^T R^{-1} (z - \Theta) \\ \cdot \frac{\partial^2}{\partial \Theta \partial \Theta^T} \left[-\ln P_{z|\Theta}(z|\Theta) \right] &= H^T R^{-1} H \\ \cdot I(\Theta) &= \mathbb{E}_{z|\Theta} \left[\frac{\partial^2}{\partial \Theta \partial \Theta^T} \left[-\ln P_{z|\Theta}(z|\Theta) \right] \right] = \mathbb{E}_{z|\Theta} [H^T R^{-1} H] = H^T R^{-1} H \\ \cdot \text{CRLB} &= [I(\Theta)]^{-1} = (H^T R^{-1} H)^{-1} \Leftrightarrow \text{Cov}_{\hat{\theta}_{\text{ML}}}(\Theta) = \text{CRLB}! \quad (\text{y compris si } N < \infty!) \end{aligned}$$

Chapitre V

Estimateurs Bayésiens

On dispose d'une connaissance a priori sur θ . Le modèle sous-jacent est $p_{Z|\Theta}(z|\theta) = p_\Theta(\theta)p_{Z|\Theta}(z|\theta)$, où la *loi a priori* $p_\Theta(\theta)$ et le modèle d'observation $p_{Z|\Theta}(z|\theta)$ sont donnés. (Dans le cas de VA discrètes, on exploite les fonctions de masse en lieu et place des densités de probabilités.)

V.1 Loi a posteriori

La règle de Bayes permet d'écrire

$$p_{\Theta|Z}(\theta|z) = \frac{p_{Z|\Theta}(z|\theta)p_\Theta(\theta)}{p_Z(z)}. \quad (\text{V.1.1})$$

- La loi a posteriori $p_{\Theta|Z}(\theta|z)$ capture toute l'information sur θ caché contenue dans l'observation z !
- Nota : $p_Z(z)$ est simplement une constante de normalisation, de sorte qu'on écrit souvent $p_{\Theta|Z}(\theta|z) \propto p_{Z|\Theta}(z|\theta)p_\Theta(\theta)$.

V.2 Définition d'un estimateur Bayésien

- Le passage de $p_{\Theta|Z}(\theta|z)$ à $\hat{\Theta} = g(Z)$ s'effectue par minimisation d'un *risque de Bayes*, i.e.,

$$\hat{\Theta} = \arg \min_{\hat{\Theta}} \mathbb{E}_{Z,\Theta} \left[C(\hat{\Theta} - \Theta) \right], \text{ où } \hat{\Theta} = \hat{g}(Z), \quad (\text{V.2.1})$$

$$= g(Z), \text{ où } g(\cdot) = \arg \min_{\hat{g}(\cdot)} \iint C(\hat{g}(z) - \theta) p_{Z,\Theta}(z, \theta) d\theta dz, \quad (\text{V.2.2})$$

$$= g(Z), \text{ où } g(\cdot) = \arg \min_{\hat{g}(\cdot)} \int \underbrace{\left[\int C(\hat{g}(z) - \theta) p_{\Theta|Z}(\theta|z) d\theta \right]}_{(*)} p_Z(z) dz. \quad (\text{V.2.3})$$

Pour résoudre le problème, on recherche $\hat{g}(\cdot)$ qui minimise $(*)$ pour tout z , et il vient

$$\hat{\Theta} = g(Z), \text{ où } g(\cdot) = \arg \min_{\hat{g}(\cdot)} \mathbb{E}_{\Theta|Z} [C(\hat{g}(Z) - \Theta) | z] \text{ pour tout } z. \quad (\text{V.2.4})$$

V.3 Estimateur du minimum d'erreur quadratique moyenne (EMQ, MMSEE)

- On pose $C(\hat{\Theta} - \Theta) = (\hat{\Theta} - \Theta)^T Q(\hat{\Theta} - \Theta)$, où $Q \geq 0$ donnée.
Alors,

$$\hat{\Theta}_{\text{MMSEE}} \text{ tel que pour } Z(\omega) = z, \quad \hat{\Theta}_{\text{MMSEE}}(\omega) = \hat{\theta}_{\text{MMSEE}} = \mathbb{E}_{\Theta|Z}[\Theta|z] \quad (\text{V.3.1})$$

$$= \int \theta p_{\Theta|Z}(\theta|z) d\theta. \quad (\text{V.3.2})$$

- L'estimé $\hat{\theta}_{\text{MMSEE}}$ du minimum d'erreur quadratique moyenne est donc la *moyenne a posteriori* de Θ conditionnellement à $Z = z$.
- L'estimateur $\hat{\Theta}_{\text{MMSEE}}$ est *non biaisé* et à *minimum de variance*.
- Illustration

V.4 Estimateur du maximum a posteriori (MAP)

- Si $C(\hat{\Theta} - \Theta) = 1 - \delta(\hat{\Theta} - \Theta)$, alors

$$\hat{\Theta}_{\text{MAP}} \text{ tel que pour } Z(\omega) = z, \quad \hat{\Theta}_{\text{MAP}}(\omega) = \hat{\theta}_{\text{MAP}} = \arg \max_{\theta} p_{\Theta|Z}(\theta|z). \quad (\text{V.4.1})$$

- L'estimé $\hat{\theta}_{\text{MAP}}$ du maximum a posteriori est donc le *mode de la loi a posteriori* de Θ conditionnellement à $Z = z$.
- $\hat{\theta}_{\text{MAP}} = \arg \max_{\theta} \ln p_{\Theta|Z}(\theta|z) = \arg \max_{\theta} (\ln p_{\Theta}(\theta) + \ln p_{Z|\Theta}(z|\theta))$.
- Et si pas de connaissance a priori ($p_{\Theta}(\theta)$ uniforme), alors $\hat{\Theta}_{\text{MAP}} = \hat{\Theta}_{\text{MLE}} !!!$
- Illustration

V.5 Estimateur du meilleur estimé linéaire (BLUE)

...

V.6 Cas d'un modèle linéaire Gaussien

- Si $Z = HX + V$, où $Z \in \mathbb{R}^N$, $X \in \mathbb{R}^M$, $V \in \mathbb{R}^N$, $\begin{pmatrix} X \\ V \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} \hat{x}^- \\ 0 \end{pmatrix}, \begin{pmatrix} P^- & 0 \\ 0 & R \end{pmatrix}\right)$ avec $R > 0$, alors $p_{X|Z}(x|z) = \mathcal{N}(x; \hat{x}^+, P^+)$, et les moments $\hat{x}^+ = \mathbb{E}_{X|Z}[X|z]$ et $P^+ = \mathbb{E}_{X|Z}[(X - \hat{x}^+)(X - \hat{x}^+)^T | z]$ sont donnés par

$$\begin{cases} \hat{x}^+ = \hat{x}^- + K(z - H\hat{x}^-) \\ P^+ = P^- - KHP^- \\ K = P^- H^T (R + HP^- H^T)^{-1}. \end{cases} \quad (\text{V.6.1})$$

- Preuve : il suffit d'exprimer les moments de $\begin{pmatrix} X \\ Z \end{pmatrix}$ et d'exploiter (II.2.14).
- Remarques : on montre (!) que

$$K = P^+ H^T R^{-1}; \quad (P^+)^{-1} = (P^-)^{-1} + H^T R^{-1} H; \quad (P^+)^{-1} \hat{x}^+ = (P^-)^{-1} \hat{x}^- + H^T R^{-1} z. \quad (\text{V.6.2})$$

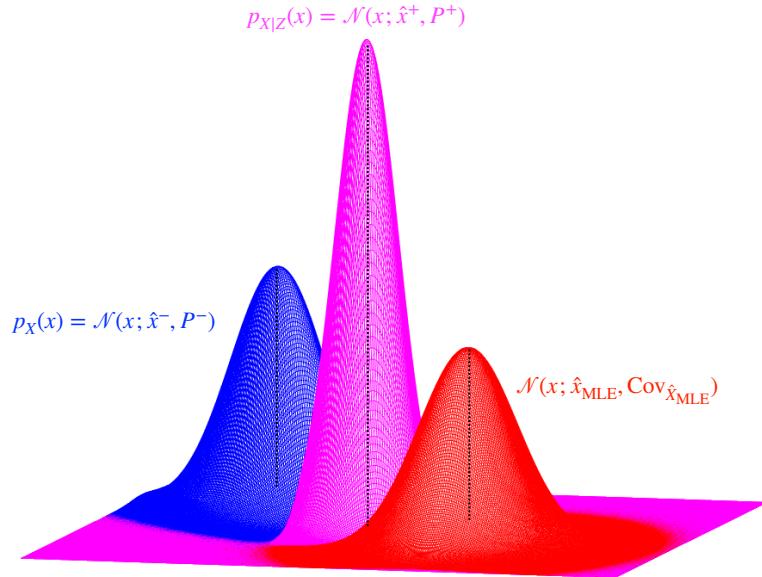
- Interprétation en terme de *fusion d'estimés*

Si $N \geq M$, H déterministe de rang plein M , alors il vient

$$\begin{cases} (P^+)^{-1} \hat{x}^+ = (P^-)^{-1} \hat{x}^- + (\text{Cov}_{\hat{X}_{\text{MLE}}})^{-1} \hat{x}_{\text{MLE}} \\ (P^+)^{-1} = (P^-)^{-1} + (\text{Cov}_{\hat{X}_{\text{MLE}}})^{-1}, \end{cases} \quad (\text{V.6.3})$$

de sorte que \hat{x}^-, P^- d'une part, et \hat{x}_{MLE} , $\text{Cov}_{\hat{X}_{\text{MLE}}}$ d'autre part, jouent un rôle symétrique dans la constitution de $\hat{x}^+, (P^+)$!

- Illustration



- Si $P^- = \infty \mathbb{I}$, on obtient une interprétation stochastique des moindres carrés récursifs...
- \hat{x}^+ est d'autant plus proche de \hat{x}^- (resp. de \hat{x}_{MLE}) que P^- (resp. $\text{Cov}_{\hat{X}_{\text{MLE}}}$) est faible.
- P^+ est à la fois plus petit que P^- et $\text{Cov}_{\hat{X}_{\text{MLE}}}$ (car apport d'information).

EXAMEN D'ESTIMATION STOCHASTIQUE – 2ASRI

1° session – Mercredi 16 Décembre 2015 – Durée 1h15

Tous documents de Cours, TD, TP autorisés – Tablettes et objets communicants interdits

les questions I/, II/, III/-8, III/-9, III/-11, III/-12 sont indépendantes.

I/ Questions de cours. Répondre en trois phrases maximum convenablement construites, sans nécessairement invoquer des formules mathématiques à chacune des questions suivantes.

1. Indiquer en quoi diffèrent les cadres théoriques de l'estimation dite « classique » et de l'estimation Bayésienne.
2. Expliquer en langage simple ce que sont le biais et la covariance d'un estimateur.
3. À quoi sert l'inégalité de Cramér-Rao ?
4. Soient Θ et Z deux variables aléatoires. Quelle signification peut-on accorder à la loi *a priori* $p_\Theta(\theta)$ de Θ et à sa loi *a posteriori* $p_{\Theta|Z}(\theta|z)$?

II/ Soit X une variable aléatoire vectorielle, et X_1, \dots, X_N ses composantes.

5. À quelle condition X_1, \dots, X_N sont-elles mutuellement indépendantes ?
6. On suppose que $X \sim \mathcal{N}(\bar{x}, P)$, c.-à-d. que X suit la loi Gaussienne réelle multidimensionnelle de moyenne \bar{x} et de covariance P . Répondre aux questions suivantes.
 - (a) Comment s'écrit la densité de probabilité $p_X(x)$ de X ?
 - (b) Si les variables aléatoires scalaires X_1, \dots, X_N sont indépendantes identiquement distribuées, comment $p_X(x)$ se réécrit-elle ?
 - (c) À quelle condition sur \bar{x} et/ou P les variables aléatoires scalaires X_1, \dots, X_N sont-elles (a) centrées ; (b) mutuellement indépendantes ; (c) corrélées sans être linéairement dépendantes ; (d) linéairement dépendantes ?

III/ Soit $\mathcal{N}(m, \sigma^2)$ la loi Gaussienne scalaire réelle de moyenne m et de variance $v = \sigma^2$. On suppose que m est une constante connue. Sur la base d'un vecteur z constitué de n échantillons z_1, \dots, z_N i.i.d. selon $\mathcal{N}(m, \sigma^2)$, on construit l'estimé

$$\hat{v} = g(z) = \frac{1}{N} \sum_{n=1}^N (z_n - m)^2 \quad (1)$$

de v . Cette question se propose d'analyser l'estimateur $\hat{V} = g(Z)$ associé, où $Z = (Z_1, \dots, Z_N)^T$ désigne la variable vectorielle aléatoire qui s'est réalisée en $z = (z_1, \dots, z_N)^T$.

7. Exprimer \hat{V} en fonction de Z_1, \dots, Z_N .
8. On souhaite montrer que \hat{V} n'est pas biaisé.
 - (a) Écrire quelle égalité mathématique doit être vérifiée pour justifier ce résultat.
 - (b) Comme indiqué ci-dessus, $\forall n, Z_n \sim \mathcal{N}(m, v)$. Comment s'écrit alors l'espérance $\mathbb{E}\{(Z_n - m)^2\}$?
 - (c) En déduire la propriété de non-biais de \hat{V} .
9. On souhaite exprimer la variance de \hat{V} . Une manière de procéder est de constater que $X = \frac{N}{\sigma^2} \hat{V}$ est équivalent à la somme des carrés de N variables Gaussiennes mutuellement indépendantes, de moyenne nulle et de variance unité.
 - (a) Quelle est la distribution de probabilité (que l'on notera $X \sim \chi_N^2$) de X ?

- (b) Exprimer le lien de proportionnalité qui unit la variance $\text{Var}(\hat{V})$ de \hat{V} à la variance $\text{Var}(X)$ de X .
- (c) Sachant que $\text{Var}(X) = 2N$ (propriété de la loi χ_N^2), en déduire $\text{Var}(\hat{V})$.
10. Déduire des propriétés démontrées ci-dessus le comportement de \hat{V} lorsque le nombre d'échantillons N croît indéfiniment.
11. On se propose de démontrer en outre que \hat{V} est l'estimateur du maximum de vraisemblance de v .
- (a) Écrire la loi $p_{Z|v}(z|v)$ et simplifier son expression en exploitant l'indépendance des composantes de Z .
- (b) On rappelle que $p_{Z|v}(z|v)$, considérée comme une fonction de v , exprime la vraisemblance $L(v; z)$ de v étant donné un vecteur d'observation z . Écrire l'anti log-vraisemblance $\text{NLL}(v; z) = -\ln L(v; z)$. Rappeler quels problèmes d'optimisation permettent l'obtention de l'estimé du maximum de vraisemblance \hat{v}_{MLE} de v à partir de $p_{Z|v}(z|v)$, puis de $\text{NLL}(v; z)$.
- (c) Écrire la condition nécessaire de stationnarité sur $\text{NLL}(v; z)$ que doit satisfaire \hat{v}_{MLE} , et la développer. En déduire l'égalité de \hat{V}_{MLE} et de \hat{V} introduit en (1).
12. \hat{V} est-il l'estimateur efficace ? Justifier soigneusement le résultat.

EXAMEN D'ESTIMATION STOCHASTIQUE – 2ASRI

2^e session – Mardi 30 Août 2016 – Durée 1h30

Tous documents de Cours, TD, TP autorisés – Tablettes et objets communicants interdits

I—Questions de cours/ Répondre en trois phrases maximum convenablement construites, sans nécessairement invoquer des formules mathématiques, à chacune des questions suivantes.

1. Indiquer en quoi diffèrent les cadres théoriques de l'estimation dite « classique » et de l'estimation Bayésienne.
2. Expliquer en langage simple ce que sont le biais et la covariance d'un estimateur.
3. À quoi sert l'inégalité de Cramér-Rao ?
4. Soient Θ et Z deux variables aléatoires. Quelle signification peut-on accorder à la loi *a priori* $p_\Theta(\theta)$ de Θ et à sa loi *a posteriori* $p_{\Theta|Z}(\theta|z)$?

II—Exercice/ Soit X une variable aléatoire vectorielle, et X_1, \dots, X_N ses composantes.

5. À quelle condition X_1, \dots, X_N sont-elles mutuellement indépendantes ?
6. On suppose que $X \sim \mathcal{N}(\bar{x}, P)$, c.-à-d. que X suit la loi Gaussienne réelle multidimensionnelle de moyenne \bar{x} et de covariance P . Répondre aux questions suivantes.
 - (a) Comment s'écrit la densité de probabilité $p_X(x)$ de X ?
 - (b) Si les variables aléatoires scalaires X_1, \dots, X_N sont indépendantes identiquement distribuées, comment $p_X(x)$ se réécrit-elle ?
 - (c) À quelle condition sur \bar{x} et/ou P les variables aléatoires scalaires X_1, \dots, X_N sont-elles (a) centrées ; (b) mutuellement indépendantes ; (c) corrélées sans être linéairement dépendantes ; (d) linéairement dépendantes ?

III—Problème/ Estimateurs du Maximum de Vraisemblance et du Maximum A Posteriori pour l'estimation d'une quantité constante

Un capteur délivre successivement N mesures z_1, \dots, z_N d'une distance $\theta \in \mathbb{R}$ constante déterministe inconnue que l'on souhaite estimer. Chaque $n^{\text{ème}}$ observation z_n est modélisée comme la réalisation pour l'expérience en cours de la variable aléatoire Z_n liée à θ par l'équation

$$Z_n = \theta + B_n, \quad n = 1, \dots, N, \tag{1}$$

où les bruits de mesure, additifs, sont représentés par les variables aléatoires scalaires réelles $\{B_n\}_{n=1,\dots,N}$ i.i.d. selon la loi Gaussienne $\mathcal{N}(0, \sigma_b^2)$ de moyenne nulle et de variance σ_b^2 . On note Z et B les vecteurs aléatoires $Z = (Z_1, \dots, Z_N)^T$ et $B = (B_1, \dots, B_N)^T$, et $z = (z_1, \dots, z_N)^T$ et $b = (b_1, \dots, b_N)^T$ leurs réalisations.

On considère dans un premier temps le problème du calcul et de la caractérisation de l'estimateur du maximum de vraisemblance.

7. Expliquer pourquoi chaque variable aléatoire Z_n , $n = 1, \dots, N$, est Gaussienne. Calculer sa moyenne et sa variance. Démontrer que les $\{Z_n\}_{n=1,\dots,N}$ sont mutuellement indépendants.
8. Écrire la loi $p_{Z|\theta}(z|\theta)$.
9. De quel problème d'optimisation l'estimé du maximum de vraisemblance $\hat{\theta}_{\text{MLE}}$ de θ pour la séquence d'observations z_1, \dots, z_N , est-il la solution ? Simplifier autant que possible l'expression du critère que minimise $\hat{\theta}_{\text{MLE}}$.

10. Comment s'écrit la solution $\hat{\theta}_{\text{MLE}}$? *L'exploitation de résultats génératifs démontrés dans le cours est tout à fait permise, dès lors que ceux-ci sont instanciés pour le problème considéré ici.*
11. Démontrer que l'estimateur $\hat{\Theta}_{\text{MLE}}$ associé admet un biais nul.

On dispose d'une connaissance a priori sur θ . Celle-ci est exprimée par le fait que θ est la réalisation de la variable aléatoire Θ Gaussienne de moyenne donnée θ_0 et de variance donnée σ_θ^2 . En outre,

$$Z_n = \Theta + B_n, \quad n = 1, \dots, N, \quad (2)$$

où Θ et B_1, \dots, B_N sont mutuellement indépendants.

12. Comment s'écrit la loi a posteriori $p_{\Theta|Z}(\theta|z)$ dans le cas général? Comment se développe-t-elle pour le problème considéré ici?
13. De quel problème d'optimisation l'estimé du maximum a posteriori $\hat{\theta}_{\text{MAP}}$ de θ pour la séquence d'observations z_1, \dots, z_N , est-il la solution? Simplifier autant que possible l'expression du critère que minimise $\hat{\theta}_{\text{MAP}}$.
14. Comment s'écrit la solution $\hat{\theta}_{\text{MAP}}$? *L'exploitation de résultats génératifs démontrés dans le cours est tout à fait permise, dès lors que ceux-ci sont instanciés pour le problème considéré ici.*
15. Que devient $\hat{\theta}_{\text{MAP}}$ si $\frac{\sigma_b^2}{\sigma_\theta^2}$ est faible? Comment interpréter ce résultat?
16. Que devient $\hat{\theta}_{\text{MAP}}$ si N est élevé? Comment interpréter ce résultat?
17. Soient $\hat{\Theta}_{\text{MLE}}$ l'estimateur associé, $\mathbb{E}[\cdot]$ l'opérateur espérance mathématique, et $e_{\hat{\Theta}_{\text{MLE}}}(\theta), e_{\hat{\Theta}_{\text{MLE}}}$ les quantités $e_{\hat{\Theta}_{\text{MLE}}}(\theta) = \mathbb{E}_{Z|\Theta}[\hat{\Theta}_{\text{MLE}}|\theta] - \theta$ et $e_{\hat{\Theta}_{\text{MLE}}} = \mathbb{E}_\Theta[e_{\hat{\Theta}_{\text{MLE}}}(\theta)]$.
 - (a) Quel est le sens physique de $e_{\hat{\Theta}_{\text{MLE}}}(\theta)$ et $e_{\hat{\Theta}_{\text{MLE}}}$?
 - (b) Quelles sont leurs valeurs?
 - (c) Quelle interprétation peut-on donner à celles-ci?

Bibliographie

- [1] A. Papoulis. *Probability, Random Variables, and Stochastic Processes*. McGraw-Hill, 1991.
- [2] E. Walter and L. Pronzato. *Identification de modèles paramétriques à partir de données expérimentales*.
- [3] S.M. Kay *Fundamentals of Statistical Signal Processing : Estimation Theory*. Prentice Hall, 1993.

Chapitre VI

Filtrage de Kalman

VI.1 Remarque importante

Bien que ce soit au prix d'une perte significative de rigueur, on adoptera désormais des notations plus légères que celles définies au bas de la page 6 pour les densités de probabilité et page 15 pour l'opérateur espérance mathématique. Celles-ci consisteront à ne plus différencier soigneusement les variables aléatoires de leurs réalisations. Il conviendra d'avoir en tête qu'elles sont abusives, et d'être très prudent dans la compréhension des résultats véhiculés par les formules mathématiques.

Ainsi,

- la densité de probabilité $p_X(x)$ de la variable aléatoire X , fonction de la variable muette x réalisation de X , sera désignée par $p(x)$;
- l'espérance mathématique $\mathbb{E}_X[g(X)]$ sera notée $\mathbb{E}[g(X)]$, voire même $\mathbb{E}[g(x)]$!
- la densité de probabilité $p_{X|Z}(x|z)$ de la variable aléatoire X conditionnellement à l'événement « la variable aléatoire Z s'est réalisée en la valeur z », fonction de la variable muette x réalisation de X , sera désignée par $p(x|Z = z)$, voire même par $p(x|z)$!
- l'espérance mathématique $\mathbb{E}_{X|Z}[g(X)|z]$ sera notée $\mathbb{E}[g(X)|Z = z]$, voire $\mathbb{E}[g(X)|z]$ ou même $\mathbb{E}[g(x)|z]$!

On conservera les notations habituelles pour toute loi Gaussienne d'espérance \bar{x} et de matrice de covariance P . Ainsi, le fait qu'un vecteur aléatoire X soit distribué selon cette loi sera noté $X \sim \mathcal{N}(\bar{x}, P)$, et la densité de probabilité correspondante $p_X(x)$ sera désignée par $\mathcal{N}(x; \bar{x}, P)$, cf. §II.2.4 page 18.

VI.2 Généralités

- Rudolf Emil Kalman : 1930–2016
 - À l'origine de contributions fondamentales en Automatique moderne (à la fois en temps continu et discret) : analyse structurelle dans l'espace d'état, analyse et synthèse par la théorie de Lyapunov, commande optimale (aspects généraux et solution du problème à critère quadratique), filtrage, etc.
 - 2008 : Charles Stark Draper Prize de la National Academy of Engineering.

- Le filtre de Kalman
 - Base pour l'*estimation de phénomènes dynamiques / la fusion de données*.
 - Sans doute le résultat fondamental d'Automatique moderne qui a le plus « diffusé » en Sciences et Ingénierie
 - Travaux de la fin des années 50. À lui seul, l'article “*A New Approach to Linear Filtering and Prediction Problems*” (Trans. of the ASME-Journal of Basic Engineering, 82 Series D : 35–45, 1960) a fait l'objet de plus de 30000 citations...
 - Omniprésent dans les systèmes
 - de guidage-navigation-contrôle (*e.g.* d'engins aéronautiques et spatiaux),
 - de localisation (*e.g.* de robots),
 - de suivi (*e.g.* de cibles manœuvrantes par radar, de personnes dans un contexte de vidéosurveillance/vidéoprotection),
 - de trajectographie (*e.g.* de migrations d'animaux, de bateaux de pêche),
 - d'analyse et prédition de séries temporelles (*e.g.* en économie), etc.
 - de détection-diagnostic
 - etc.
- Quelques ressources
 - Sites
 - Pages Wikipedia (!)
 - <http://www.cs.unc.edu/~welch/kalman/> et les références qu'il contient
 - Ouvrages (liste bien sûr non exhaustive)
 - B.D.O. Anderson, J.B. Moore. *Optimal Filtering*. Prentice Hall. 1979
 - A. Gelb. *Applied Optimal Estimation*. MIT Press. 1974
 - A.H. Jazwinski. *Stochastic Processes and Filtering Theory*. Academic Press. 1970
 - T. Kailath, A.H. Sayed, B. Hassibi. *Linear Estimation*. Prentice Hall. 2000
 - M. Labarrère, J.P. Krief, B. Gimonet. *Le filtrage et ses applications*. Cépaduès/Sup'Aéro. 1993
 - P.S. Maybeck *Stochastic Models, Estimation and Control*, Vols 1&2. Academic Press. 1979
 - cf. également les ouvrages de F. Gustafsson (filtrage adaptatif), Y. Bar-Shalom (suivi multimodèles, multicibles, association de données), etc.
- Objectifs de ce chapitre
 - Introduire les concepts les plus importants et le vocabulaire associé.
 - Poser un contexte d'étude et proposer un embryon de preuve (il existe de multiples autres approches du résultat).
 - Recenser quelques propriétés importantes et leur raisonnement sous-jacent.
 - Identifier où se posent les réels problèmes (au-delà des équations mathématiques, qui « sont ce qu'elles sont » avec leur complexité), en particulier le problème du “*tuning*”.
 - Donner quelques pistes pour des extensions au cas non linéaire.
 - Et, plus largement, de constituer un (modeste) support pour aborder plus aisément des ouvrages de référence.

VI.3 Introduction : d'un problème élémentaire vers sa formalisation mathématique

VI.3.1 Deux problèmes académiques

- En guise d'introduction, considérons les deux problèmes suivants :
 - Estimation d'un phénomène statique : localisation d'une personne immobile ;
 - Estimation d'un phénomène dynamique : localisation d'une personne en mouvement.
- Le but est, par une formulation intuitive et au détriment d'une certaine rigueur, de mieux cerner l'objectif du filtrage de Kalman et d'introduire peu à peu le vocabulaire...

A Localisation d'une personne immobile

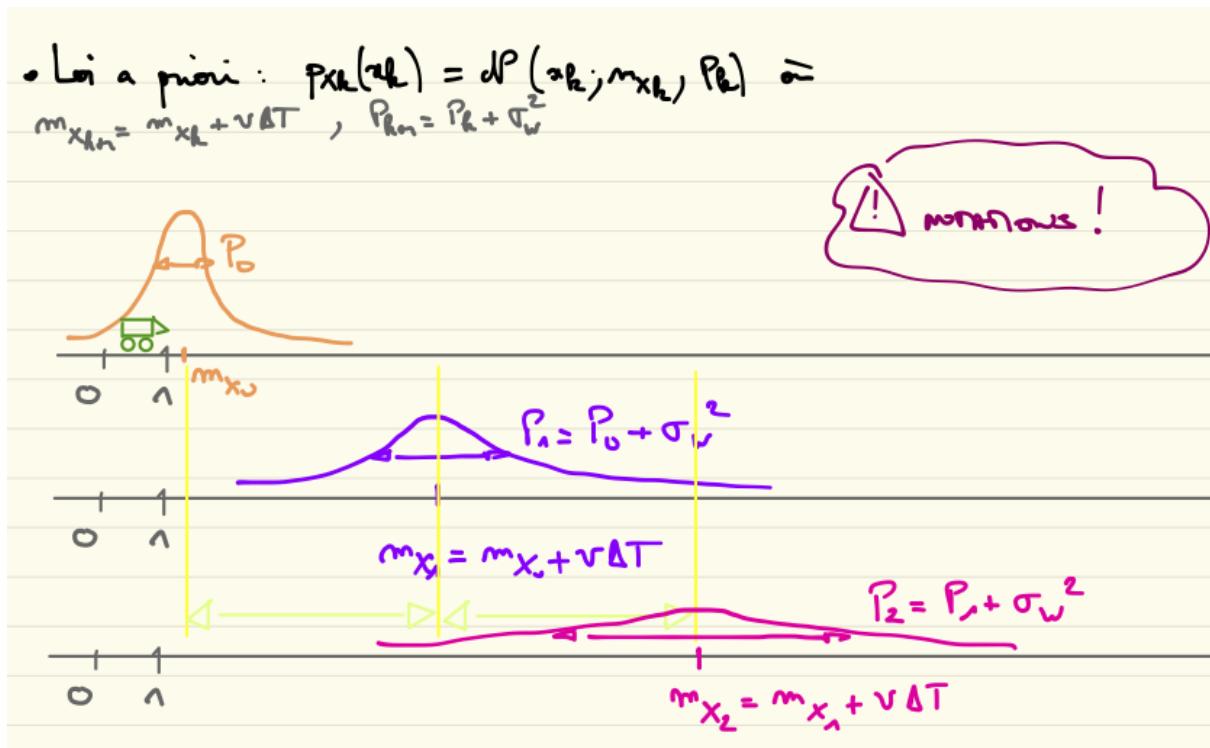
- Problème : estimer récursivement une position constante inconnue X sur la base de mesures z_1, \dots, z_k de X délivrées par un capteur, ces mesures étant entachées d'erreurs, de moyenne nulle et de (co)variances connues C_1, \dots, C_k
- Ou bien :
 - À l'instant $k = 1$, *fusionner*
 - la *connaissance a priori* $\mathcal{N}(m_{X_0}, P_0)$ sur X à l'instant $k = 0$,
 - la distribution $\mathcal{N}(z_1, C_1)$ (estimé de X et (co)variance d'erreur associée) qui peut être établie sur la base de la mesure seule à l'instant $k = 1$,

→ de façon à obtenir la loi a posteriori $\mathcal{N}(\hat{x}_1, P_1)$ à l'instant $k = 1$, où \hat{x}_1, P_1 désignent donc l'estimé a posteriori de X et la (co)variance d'erreur a posteriori.
 - À l'instant $k = 2$, fusionner $\mathcal{N}(\hat{x}_1, P_1)$ (nouvelle connaissance a priori) avec $\mathcal{N}(z_2, C_2)$ de façon à obtenir $\mathcal{N}(\hat{x}_2, P_2)$.
 - À l'instant $k = 3$, fusionner $\mathcal{N}(\hat{x}_2, P_2)$ et $\mathcal{N}(z_3, C_3)$ de façon à obtenir $\mathcal{N}(\hat{x}_3, P_3)$. Et ainsi de suite...

→ Il s'agit bien d'un *schéma récursif d'estimation d'une variable aléatoire constante*.
- Quelques considérations de bon sens : dès lors que le capteur ne délivre aucune fausse mesure (hypothèse implicite dans tout ce cours),
 - \hat{x}_k est d'autant plus proche de \hat{x}_{k-1} (resp. de z_k) que P_{k-1} (resp. C_k) est faible ;
 - P_k est à la fois plus petit que P_{k-1} et C_k (car apport d'information).

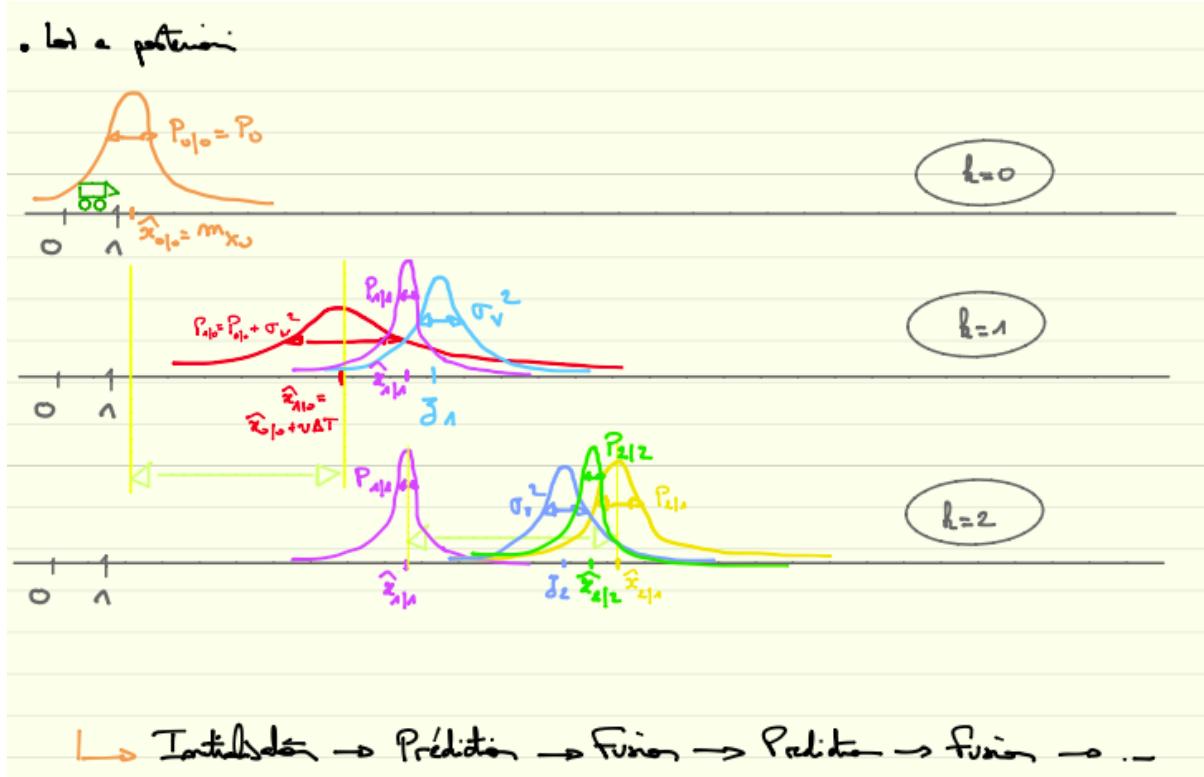
B Localisation d'une personne en mouvement

- La position inconnue varie au fil du temps. À l'instant k , elle est notée X_k . La *connaissance a priori* à son sujet se déduit de
 - la *distribution initiale* $\mathcal{N}(m_{X_0}, P_0)$ à laquelle obéit X_0 à l'instant $k = 0$;
 - la *loi de dynamique a priori* exprimant le lien statistique entre X_{k-1} et X_k (e.g. pour un mouvement uniforme, $X_k = X_{k-1} + v \cdot \Delta t + \text{bruit}$).



- Lorsqu'on n'effectue aucune mesure, la distribution de X_k à l'instant k est caractérisée par la *loi a priori* $\mathcal{N}(m_{X_k}, P_k)$:
- m_{X_k} (que l'on pourrait noter $\hat{x}_{k|0}$) désigne le comportement moyen ;
 - P_k (que l'on pourrait noter $P_{k|0}$) caractérise l'incertitude autour de celui-ci. Note : pour la dynamique a priori considérée, P_k croît avec k .

- Problème : estimer récursivement une position inconnue variante dans le temps – notée X_k à l'instant k – sur la base de mesures z_1, \dots, z_k de X_1, \dots, X_k délivrées par un capteur, ces mesures étant entachées d'erreurs de moyennes nulles et de (co)variances connues C_1, \dots, C_k



- On effectue des mesures à partir de l'instant $k = 1$
 - À l'instant $k = 1$, fusionner
 - la connaissance a priori $\mathcal{N}(\hat{x}_{1|0}, P_{1|0})$ sur X_1 , obtenue par prédition de $\mathcal{N}(m_{X_0}, P_0)$ sur la base de la dynamique a priori
 - la distribution $\mathcal{N}(z_1, C_1)$ (estimé de X_1 et covariance d'erreur associée) qui peut être établie sur la base de la mesure seule à l'instant $k = 1$,

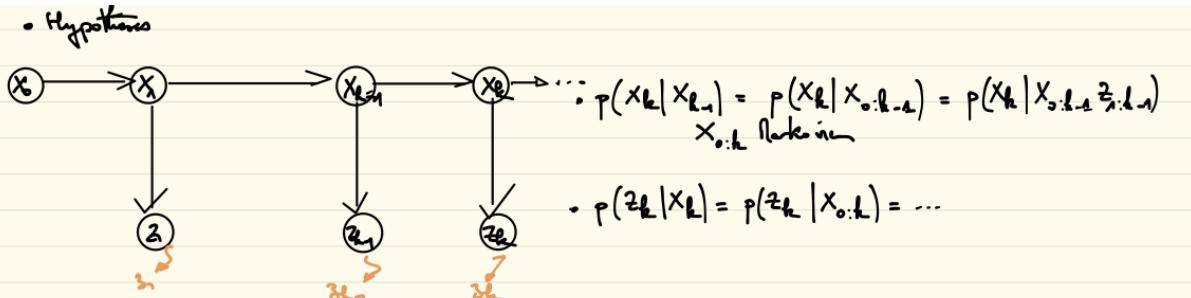
→ de façon à obtenir la *loi a posteriori* $\mathcal{N}(\hat{x}_{1|1}, P_{1|1})$, où $\hat{x}_{1|1}, P_{1|1}$ désignent donc l'estimé a posteriori de X_1 et la covariance d'erreur a posteriori.
 - À l'instant $k = 2$, fusionner $\mathcal{N}(\hat{x}_{2|1}, P_{2|1})$ – obtenue par prédition de $\mathcal{N}(\hat{x}_{1|1}, P_{1|1})$ – avec $\mathcal{N}(z_2, C_2)$ de façon à obtenir $\mathcal{N}(\hat{x}_{2|2}, P_{2|2})$.
 - À l'instant $k = 3$, fusionner $\mathcal{N}(\hat{x}_{3|2}, P_{3|2})$ – obtenue par prédition de $\mathcal{N}(\hat{x}_{2|2}, P_{2|2})$ – avec $\mathcal{N}(z_3, C_3)$ de façon à obtenir $\mathcal{N}(\hat{x}_{3|3}, P_{3|3})$.

Et ainsi de suite...
- Il s'agit bien d'un *schéma récursif d'estimation d'un processus aléatoire*.
- Dès lors que le capteur ne délivre aucune fausse mesure,
 - $\hat{x}_{k|k}$ est d'autant plus proche de $\hat{x}_{k|k-1}$ (resp. de z_k) que $P_{k|k-1}$ (resp. C_k) est faible ;
 - $P_{k|k}$ est à la fois plus petit que $P_{k|k-1}$ et C_k (car apport d'information).
- Pour la dynamique a priori considérée, $P_{k|k-1} > P_{k-1|k-1}$.
- Note : en général – contrairement à ce qui a été supposé ici – la mesure ne constitue pas une estimée de la quantité inconnue...

VI.3.2 Formalisation mathématique

A Contexte de l'estimation Bayésienne

- Formalisation
 - Variables
 - $X_{0:k} = X_0, X_1, \dots, X_k$ désigne le *processus d'état*, caché, à estimer ; $X_k \in \mathbb{R}^n$;
 - $Z_{1:k} = Z_1, \dots, Z_k$ désigne le *processus de mesure* ; $Z_k \in \mathbb{R}^p$;
la séquence de mesures $z_{1:k} = z_1, \dots, z_k$ disponible constitue une réalisation de $Z_{1:k}$ pour l'expérience en cours ;
 - Connaissance a priori - Modèles
 - *loi de dynamique a priori* $p(X_k | X_{k-1})$
 - *modèle d'observation* $p(Z_k | X_k)$ (généralement moins trivial que dans l'exemple)
 - *distribution initiale* $p(X_0)$
- FONDAMENTAL : La loi jointe a posteriori $p(X_{0:k} | Z_{1:k} = z_{1:k})$ caractérise toute l'information sur le processus d'état caché $X_{0:k}$ qui puisse être extraite de la trajectoire de mesure $z_{1:k}$.



- Bien sûr, l'évolution de X (la loi de transition $p(X_k | X_{k-1})$) peut être paramétrée par un couple u_k (commun) ou au contraire par des $p(X_k | X_{k-1}, u_{k-1})$
- Ce où X et Z sont à valeurs discrètes

Les marginales

- filtre $p(X_k | z_{1:k} = z_{1:k})$
- prédition $p(X_k | z_{1:k} = z_{1:k}), k' < k$
- lissage $p(X_k | z_{1:k'} = z_{1:k'}), k' > k$

- On s'intéressera plus particulièrement aux *lois marginales* suivantes :
 - la *loi de filtrage* $p(X_k | Z_{1:k} = z_{1:k})$;
 - les *lois de prédition* $p(X_k | Z_{1:k'} = z_{1:k'})$, où $k' < k$.
- La problématique – significativement plus complexe – du *lissage*, consistant à établir $p(X_k | Z_{1:k'} = z_{1:k'})$ pour $k' > k$ ne sera pas traitée ici.

B Simplification considérée

- Dans ce cours,
 - $X_{0:k}$ est décrit par une *équation d'état stochastique linéaire à temps discret*, avec *bruits – de dynamique – blancs, centrés, Gaussiens*
 - le lien entre X_k et Z_k est décrit par une *équation de mesure (statique) stochastique linéaire*, avec *bruits – de mesure – blancs, centrés, Gaussiens*, indépendants des bruits de dynamique
- Plus précisément, les modèles considérés sont de la forme

$$\begin{cases} X_{k+1} = F_k X_k + G_k u_k + W_k \\ Z_k = H_k X_k + V_k \end{cases} \quad \text{avec} \quad \begin{array}{l} X_0, W_k, V_k \text{ Gaussiens dans leur ensemble,} \\ W_k, V_k \text{ blancs, centrés, mutuellement} \\ \text{indépendants, indépendants de } X_0, \end{array} \quad (\text{VI.3.1})$$

ces hypothèses se résumant en

$$\forall k, k', \begin{pmatrix} X_0 \\ W_{0:k} \\ V_{0:k'} \end{pmatrix} \text{ Gaussien,} \\ \mathbb{E}\left[\begin{pmatrix} X_0 \\ W_k \\ V_k \end{pmatrix}\right] = \begin{pmatrix} m_{X_0} \\ 0 \\ 0 \end{pmatrix}, \quad \mathbb{E}\left[\begin{pmatrix} X_0 - m_{X_0} \\ W_k' \\ V_k' \end{pmatrix}\right]^T = \begin{pmatrix} P_0 & \mathbb{O} & \mathbb{O} \\ \mathbb{O} & Q_k \delta_{k,k'} & \mathbb{O} \\ \mathbb{O} & \mathbb{O} & R_k \delta_{k,k'} \end{pmatrix}, \quad (\text{VI.3.2})$$

où $m_{X_0}, P_0 \geq 0, Q_k \geq 0, R_k > 0$ sont donnés, et $\delta_{k,k'}$ désigne le symbole de Kronecker.

- Il vient alors que
 - $X_{0:k}$ et $Z_{1:k}$ est un processus aléatoire Gaussiens Markovien ;
 - $X_{0:k}$ et $Z_{1:k}$ sont conjointement Gaussiens ;
 - $p(X_k | Z_{1:k} = z_{1:k})$ est une loi Gaussienne, de moments $\hat{x}_{k|k}, P_{k|k}$; $p(X_k | Z_{1:k-1} = z_{1:k-1})$ est une loi Gaussienne, de moments $\hat{x}_{k|k-1}, P_{k|k-1}$.
 - En outre,
 - $\hat{x}_{k|k} = \mathbb{E}[X_k | Z_{1:k} = z_{1:k}]$,
 - $P_{k|k} = \mathbb{E}[(X_k - \hat{x}_{k|k})(X_k - \hat{x}_{k|k})^T | Z_{1:k} = z_{1:k}]$,
 - $\hat{x}_{k|k-1} = \mathbb{E}[X_k | Z_{1:k-1} = z_{1:k-1}]$,
 - $P_{k|k-1} = \mathbb{E}[(X_k - \hat{x}_{k|k-1})(X_k - \hat{x}_{k|k-1})^T | Z_{1:k-1} = z_{1:k-1}]$,
- sont obtenus *récursivement*, selon le schéma $(\hat{x}_{0|0}, P_{0|0}) \rightarrow (\hat{x}_{1|0}, P_{1|0}) \rightarrow (\hat{x}_{1|1}, P_{1|1}) \rightarrow \dots \rightarrow (\hat{x}_{k-1|k-1}, P_{k-1|k-1}) \rightarrow (\hat{x}_{k|k-1}, P_{k|k-1}) \rightarrow (\hat{x}_{k|k}, P_{k|k})$, et *sous forme analytique* via les équations du *filtre de Kalman*

- Nota : *Moments a priori* du processus d'état caché.

- D'après (VI.3.1)–(VI.3.2), $X_k \sim \mathcal{N}(m_{X_k}, P_k)$ et $Z_k \sim \mathcal{N}(m_{Z_k}, S_k)$, où les moments a priori $m_{X_k} = \mathbb{E}[X_k]$, $P_k = \mathbb{E}[(X_k - m_{X_k})(X_k - m_{X_k})^T]$, $m_{Z_k} = \mathbb{E}[Z_k]$, $S_k = \mathbb{E}[(Z_k - m_{Z_k})(Z_k - m_{Z_k})^T]$, se propagent selon

$$m_{X_{k+1}} = F_k m_{X_k} + G_k u_k, \quad P_{k+1} = F_k P_k F_k^T + Q_k \quad (\text{VI.3.3})$$

$$m_{Z_k} = H_k m_{X_k}, \quad S_k = R_k + H_k P_k H_k^T. \quad (\text{VI.3.4})$$

- On pourrait déterminer les moments croisés, etc.

C Quelques remarques

- On a « augmenté » une représentation d'état déterministe de façon à
 - autoriser la présence d'aléa,
 - conserver la notion d'état.
- Le deuxième point est lié à la notion de processus aléatoire Markovien. Un tel processus vérifie

$$\forall t_j < t_{j+1} < \dots < t_{i-1} < t_i, \quad p(X(t_i)|X(t_j), X(t_{j+1}), X(t_{i-1})) = p(X(t_i)|X(t_{i-1})). \quad (\text{VI.3.5})$$

- Quant au premier point, on a défini un modèle à temps discret, bien adapté à une implémentation par calculateur. Quel serait le sens de

$$\dot{X}(t) = f(X(t), u(t), W(t))? \quad (\text{VI.3.6})$$

- La réponse est complexe...

VI.4 Équations du filtre de Kalman à temps discret

VI.4.1 Rappels préliminaires

- Propriétés de la variable aléatoire $X|Z=z$ lorsque les vecteurs aléatoires X et Z sont Gaussiens dans leur ensemble
 - Cf. l'item correspondant dans §II.2.4, équation (II.2.14) page 19.
- Loi a posteriori $p(X|Z=z)$ pour un modèle statique de la forme $Z = HX + V$, avec X et Z Gaussiens dans leur ensemble
 - Cf. §V.6, équations (V.6.1) page 31.

VI.4.2 Les 7 équations du filtre de Kalman

- Rappel : modèle et hypothèses définis dans les équations (VI.3.1)–(VI.3.2).

Initialisation	$\hat{x}_{0 0} = m_{X_0}$	$P_{0 0} = P_0$	
Prédiction (Time Update)	$\hat{x}_{k+1 k} = F_k \hat{x}_{k k} + G_k u_k$	$P_{k+1 k} = F_k P_{k k} F_k^T + Q_k$	
Mise à jour (Measurement Update)		$\hat{x}_{k+1 k+1} = \hat{x}_{k+1 k} + K_{k+1} (z_{k+1} - H_{k+1} \hat{x}_{k+1 k})$ $P_{k+1 k+1} = P_{k+1 k} - K_{k+1} H_{k+1} P_{k+1 k}$ <p style="text-align: center;">où $K_{k+1} = P_{k+1 k} H_{k+1}^T (R_{k+1} + H_{k+1} P_{k+1 k} H_{k+1}^T)^{-1}$</p>	(VI.4.1)

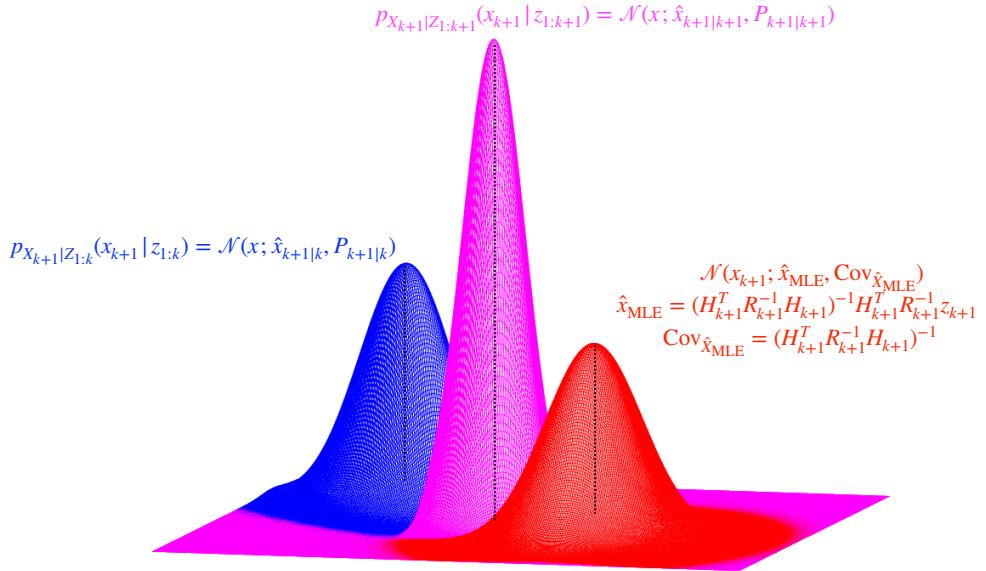
VI.4.3 Éléments de preuve

- Initialisation : évidente
 - Remarque : si \exists mesure à $k = 0$, alors on initialise avec $x_{0|-1}, P_{0|-1}$
- Prédiction : obtenue en conditionnant préalablement sur $Z_{1:k} = z_{1:k}$ l'équation d'état du modèle (VI.3.1) entre les instants k et $k + 1$ (cf. aussi (VI.3.3))
- Mise à jour : obtenue en conditionnant préalablement sur $Z_{1:k} = z_{1:k}$ l'équation de mesure du modèle (VI.3.1) à l'intant $k + 1$, puis en exploitant l'identité $p_{X_{k+1}|Z_{1:k+1}}(x_{k+1}|z_{1:k+1}) = p_{(X_{k+1}|Z_{1:k})|(Z_{k+1}|Z_{1:k})}((x_{k+1}|z_{1:k})|(z_{k+1}|z_{1:k}))$ sur la base des développements du §V.6.

VI.5 Propriétés importantes

VI.5.1 Interprétation de l'étape de mise à jour en terme de fusion d'estimés

- Dans la veine des développements effectués au §V.6 et sous des hypothèses semblables, on montre que la loi de filtrage $p_{X_{k+1}|Z_{1:k+1}}(x_{k+1}|z_{1:k+1})$ s'obtient par fusion de la loi de prédiction $p_{(X_{k+1}|Z_{1:k})}(x_{k+1}|z_{1:k})$ et d'une loi obtenue sur la base de la mesure z_{k+1} seule :



VI.5.2 Compléments de terminologie

- Soient $\phi_{\hat{x}_{k|k}} : \mathbb{R}^{kn_z} \rightarrow \mathbb{R}^{n_x}$ et $\phi_{\hat{x}_{k|k-1}} : \mathbb{R}^{(k-1)n_z} \rightarrow \mathbb{R}^{n_x}$ les fonctions *déterministes* telles que $\hat{x}_{k|k} = \phi_{\hat{x}_{k|k}}(z_{1:k})$ et $\hat{x}_{k|k-1} = \phi_{\hat{x}_{k|k-1}}(z_{1:k-1})$. On définit
 - $\hat{X}_{k|k} \triangleq \phi_{\hat{x}_{k|k}}(Z_{1:k})$ la variable aléatoire estimateur ;
 - $\tilde{X}_k \triangleq X_k - \hat{X}_{k|k}$ la variable aléatoire erreur d'estimation ;
 - $\hat{X}_{k|k-1} \triangleq \phi_{\hat{x}_{k|k-1}}(Z_{1:k-1})$ la variable aléatoire prédicteur ;
 - $\tilde{X}_k^- \triangleq X_k - \hat{X}_{k|k-1}$ la variable aléatoire erreur de prédiction ;

- $\hat{Z}_{k|k-1} \triangleq H_k \hat{X}_{k|k-1}$ la variable aléatoire prédicteur de la mesure ;
 - $\Gamma_k \triangleq Z_k - \hat{Z}_{k|k-1}$ la variable aléatoire innovation (“redisual”).
- ↪ bien sûr, pour l'événement ω tel que $Z_{1:k}(\omega) = z_{1:k}$, il vient $\hat{x}_{k|k} = \hat{X}_{k|k}(\omega)$, etc.
- ↪ en outre, $\hat{X}_{k|k} = \mathbb{E}[X_k | Z_{1:k} = .]$ et $\hat{X}_{k|k-1} = \mathbb{E}[X_k | Z_{1:k-1} = .].$
- Sur cette base, on définit également
 - $\mathbb{E}[\tilde{X}_k]$ le biais du filtre ;
 - $\text{Cov}_{\tilde{X}_k} = \mathbb{E}\left[(\tilde{X}_k - \mathbb{E}[\tilde{X}_k])(\tilde{X}_k - \mathbb{E}[\tilde{X}_k])^T\right]$ la covariance de l'erreur d'estimation ;
 - $\mathbb{E}[\tilde{X}_k^T \tilde{X}_k]$ l'erreur d'estimation quadratique moyenne ;
 - mais aussi le biais et la covariance de l'erreur de prédiction, l'erreur de prédiction quadratique moyenne, etc.

VI.5.3 Caractérisation statistique des erreurs

- *Biais de l'erreur d'estimation du vecteur d'état*
 - Par construction, le filtre est non biaisé, au sens où $\mathbb{E}[\tilde{X}_k] = 0$.
 - En effet, $\mathbb{E}_{Z,X}[\tilde{X}_k] = \mathbb{E}_X[X_k] - \mathbb{E}_{Z,X}[\hat{X}_{k|k}]$
 - $= \mathbb{E}_X[X_k] - \mathbb{E}_Z[\mathbb{E}_X[X_k | Z_{1:k} = .]]$
 - $= \mathbb{E}_X[X_k] - \mathbb{E}_X[X_k] = 0.$
- *Statistiques de l'erreur d'estimation du vecteur d'état*
 - $\tilde{X}_k \sim \mathcal{N}(0, P_{k|k})$.
 - On montre d'abord que $(\tilde{X}_k | Z_{1:k} = z_{1:k}) \sim \mathcal{N}(0, P_{k|k})$.
 - On constate ensuite que ces moments ne dépendent pas de l'événement sur lesquels ils sont conditionnés, de sorte que $p(\tilde{X}_k | Z_{1:k} = z_{1:k}) = p(\tilde{X}_k)$!
- *Statistiques de l'erreur de prédiction du vecteur d'état*
 - $\tilde{X}_k^- \sim \mathcal{N}(0, P_{k|k-1})$.
- *FONDAMENTAL : Statistiques de l'innovation Γ_k*
 - Le processus innovation est Gaussien et blanc.
 - De plus, $\Gamma_k \sim \mathcal{N}(0, S_{k|k-1})$, où $S_{k|k-1} \triangleq R_k + H_k P_{k|k-1} H_k^T$.
 - On étudie $(\Gamma_k | Z_{1:k} = z_{1:k})$, dont les moments sont indépendants de $z_{1:k}$!
 - Ceci implique que $\sum_{l=0}^{L-1} (\Gamma_{k-l}^T S_{k-l|k-l-1}^{-1} \Gamma_{k-l}) \sim \chi_{Ln_z}^2$.

VI.6 Exemple académique

Cf. pages suivantes...

- I.3 - Le cas linéaire Gaussien : le filtre de Kalman**
- Sous forme de système d'Auton / Coup de plus utilisée dans l'industrie aéronautique

Soit le système

$$\begin{cases} X_{k+1} = F_k X_k + G_k w_k + v_k \\ Z_k = H_k X_k + V_k \end{cases}$$

où $\begin{pmatrix} X_0 \\ w_k \\ V_k \end{pmatrix}$ conjointement gaussien $\forall k, k'$

$$E\left[\begin{pmatrix} X_0 \\ w_k \\ V_k \end{pmatrix}\right] = \begin{pmatrix} m_{X_0} \\ 0 \\ 0 \end{pmatrix}, m_{X_0} \text{ donné}$$

$$E\left\{\begin{pmatrix} X_0 & \cdots & X_k \\ w_k & \cdots & 0 \\ V_k & \cdots & 0 \end{pmatrix} \begin{pmatrix} X_0 & \cdots & X_k \\ w_k & \cdots & 0 \\ V_k & \cdots & 0 \end{pmatrix}^T\right\} = \begin{pmatrix} P_0 & 0 & 0 \\ 0 & P_{w,k} & 0 \\ 0 & 0 & R_{V,k} \end{pmatrix}$$

$$\text{et } P_0 \geq 0, P_k \geq 0, R_k \text{ données}$$

et $\delta_{kk'} = 1 \text{ si } k=k', 0 \text{ sinon}$

- Em d'autre formes,
- X_0 est Gaussien $X_0 \sim \mathcal{N}(m_{X_0}, P_0)$
 - w_k est Gaussien blanc, $w_k \sim \mathcal{N}(0, Q_k)$
 - V_k , ..., ..., $V_k \sim \mathcal{N}(0, R_k)$
 - X_0, w_k, V_k mutuellement indépendants

Exemple: $X, z, w, v \in \mathbb{R}$ et

$$\begin{aligned} F_k &= 1, G_k w_k = v \text{ et } A \\ Q_k &= \sigma_w^2 \text{ et } R_k = \sigma_v^2 \end{aligned}$$

L'objectif du filtre de Kalman est donc de proposer une formulation analogique pourriez des moments (moments de covariance) de la loi a posteriori (de filtrage) $P(X_k | Z_{n:k} = \bar{z}_{n:k})$ et de la loi de prédition $P(X_k | Z_{n:k-1} = \bar{z}_{n:k-1})$.

Pour le problème considéré, on peut soit Gaussien et les moments sont respectivement $\hat{x}_k | k, P_k | k$ et

$$\hat{x}_{k|k-1}, P_{k|k-1}$$

$$\hat{x}_{k|k} = E(X_k | Z_{n:k} = \bar{z}_{n:k})$$

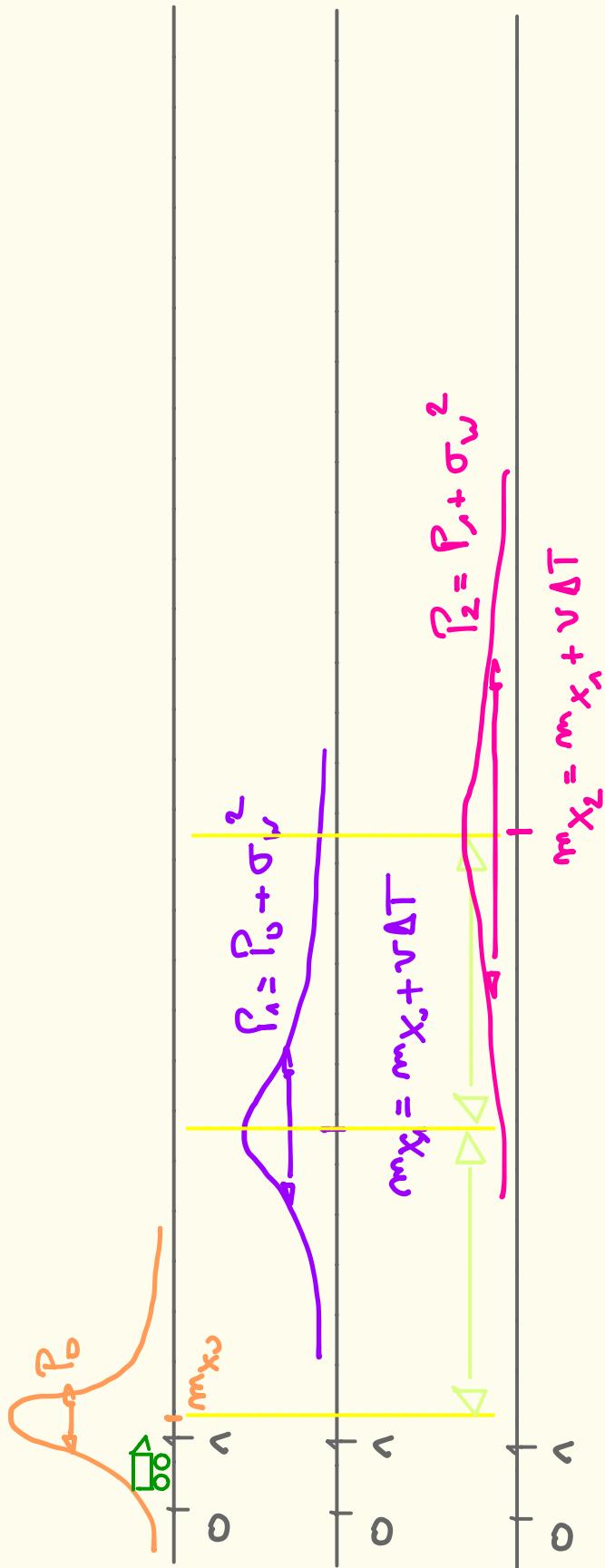
$$P_{k|k} = \text{cov}(X_k | Z_{n:k} = \bar{z}_{n:k}) = E\left\{\left(X_k - E(X_k | Z_{n:k} = \bar{z}_{n:k})\right)^T \left| Z_{n:k} = \bar{z}_{n:k}\right.\right\}$$

Parallèle : bei einer priori der X_k
 • Convient zu propagert den momenten zu X_k sur toutes les expériences?

On cherche que
 $X_k \sim \mathcal{N}(m_{X_k}, P_k)$ $\Leftrightarrow m_{X_{k+n}} = \bar{f}_k m_{X_k} + g_k w_k$ et $P_{k+n} = \bar{f}_k^T P_k \bar{f}_k + Q_k$
 \Leftrightarrow à démontrer en exercice, on applique l'équation d'Yule sur les hypothèses sur X_0, w_k .

Illustration

$$m_{X_{k+n}} = m_{X_k} + v \Delta T, \quad P_{k+n} = P_k + \sigma_w^2$$



Équations du filtre de Kalman

• Prédition (Time Update)

$$\hat{x}_{k+1|k} = f_k \hat{x}_{k|k} + g_k u_k$$

$$P_{k+1|k} = P_k$$

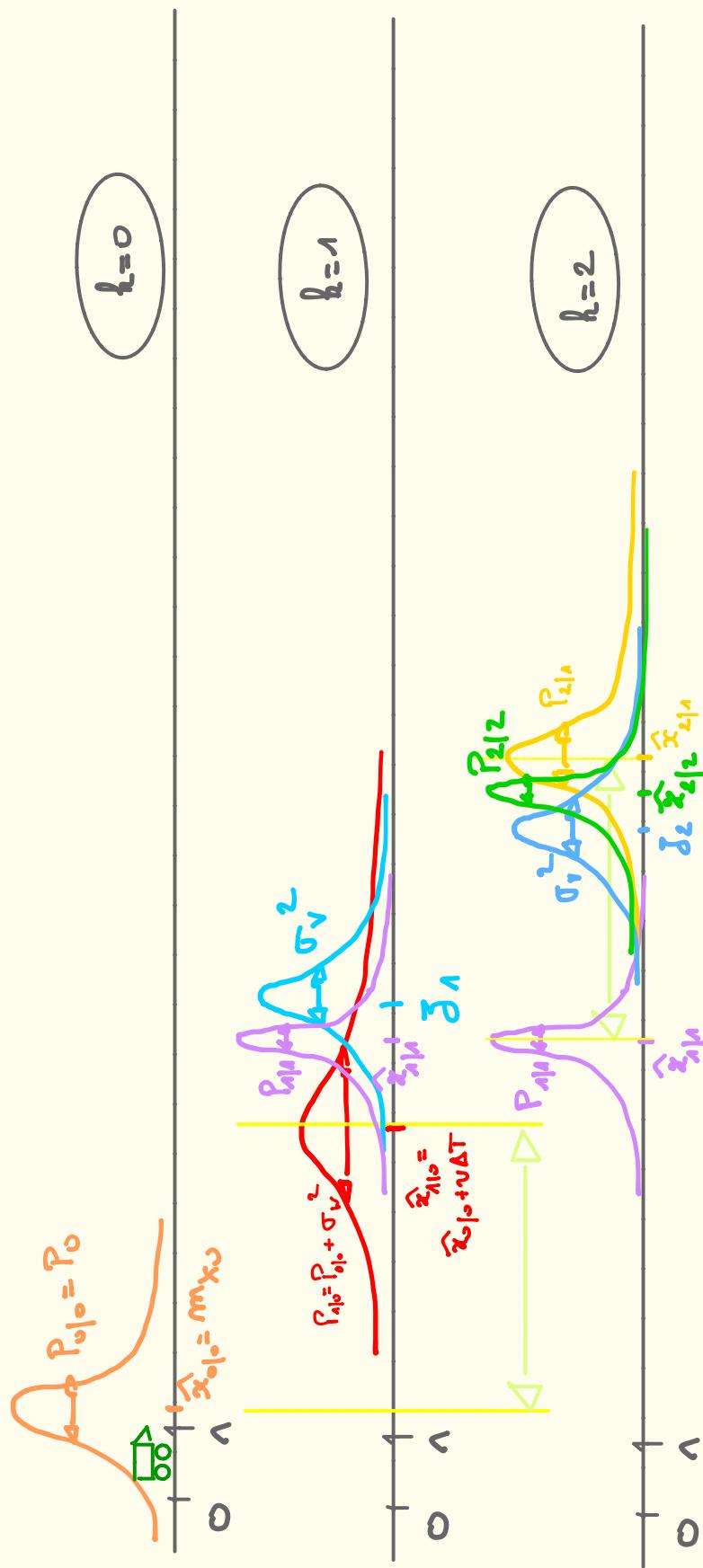
• Filtrage à jour (Measurement Update)

$$\hat{x}_{k+1|k+1} = \hat{x}_{k|k} + K_{k+1} (z_{k+1} - H_{k+1} \hat{x}_{k|k})$$

$$K_{k+1} = P_{k|k} H_{k+1}^T (H_{k+1} P_{k|k} H_{k+1}^T)^{-1}$$

Illustration

$$\hat{x}_{k+1|k} = \hat{x}_{k|k} + v \Delta T, \quad P_{k|k} = \hat{P}_{k|k} + \sigma_v^2, \quad \hat{x}_{k+1|k+1} = \frac{\sigma_v^2 \hat{x}_{k|k} + \hat{P}_{k|k} z_{k+1}}{\sigma_v^2 + \hat{P}_{k|k}}, \quad P_{k+1|k+1} = \frac{\hat{P}_{k|k}}{\sigma_v^2 + \hat{P}_{k|k}}$$



VI.7 Bric-à-brac

VI.7.1 QUELQUES AUTRES RÉSULTATS IMPORTANTS

- L'estimé $\hat{x}_{k|k}$ établi selon les équations du filtre de Kalman est également
 - l'estimé du minimum d'erreur quadratique moyenne (MMSEE = “Minimum Mean-Square Error Estimate”), par construction, *i.e.*

$$\phi_{\hat{x}_{k|k}}(\cdot) = \arg \min_{\psi(\cdot)} E_{X_k, Z_{1:k}} (X_k - \psi(Z_{1:k}))^T Q (X_k - \psi(Z_{1:k})) \quad (\text{VI.7.1})$$

quelle que soit $Q > 0$ donnée quelconque ; ou bien, de manière équivalente,

$$\phi_{\hat{x}_{k|k}}(\cdot) = \arg \min_{\psi(\cdot)} E_{X_k} \left((X_k - \psi(Z_{1:k}))^T Q (X_k - \psi(Z_{1:k})) \mid Z_{1:k} = z_{1:k} \right) \quad (\text{VI.7.2})$$

pour toute réalisation $z_{1:k}$ de $Z_{1:k}$.

- Remarque : $\phi_{\hat{x}_{k|k}}(\cdot)$ étant indépendante de Q , il est *impossible d'estimer certaines composantes de X_k mieux que d'autres*

mais aussi

- dans le contexte considéré, où toutes les distributions sont symétriques unimodales, l'estimé du maximum a posteriori (MAP), *i.e.* le mode de la loi a posteriori.
- Remarque : il est donc également le maximum de vraisemblance (MLE = “Maximum Likelihood Estimate”) si pas de connaissance a priori.

mais encore

- dans un contexte linéaire mais non Gaussien, le meilleur estimé linéaire non biaisé (BLUE = “Best Linear Unbiased Estimate”).

- Dans le cas d'un système linéaire invariant soumis à des bruits stationnaires, et sous certaines hypothèses supplémentaires (*e.g.* sous la CS observabilité de $(F, H) \& Q, R > 0$), alors le filtre admet un *comportement stationnaire*, où $P_{k|k}, P_{k|k-1}, K_k$ se stabilisent sur des constantes. Ainsi,

- $\lim_{k \rightarrow +\infty} P_{k|k-1} = P_\infty^-$, où P_∞^- est l'*unique solution semi-définie positive* de l'*équation algébrique de Riccati à temps discret*

$$P_\infty^- = FP_\infty^- F^T + Q - FP_\infty^- H^T (R + HP_\infty^- H^T)^{-1} H P_\infty^- F^T.$$
- l'équation d'état vérifiée par $\hat{x}_{k|k-1}$ admet certaines propriétés de stabilité, etc.
- $P_{0|0}$ influe sur l'amplitude du régime transitoire mais pas sur sa durée.
- possibilité d'implémenter un *filtre stationnaire* “sous-optimal”.

VI.7.2 Deux problèmes fondamentaux

- Réglage (Tuning) du filtre.
 ↳ Importance du “residual monitoring”
- Conditionnement numérique des équations.
 ↳ “Square-Root filtering”.

VI.7.3 Que faire si...

- On ne dispose pas d'information a priori, *e.g.* $P_{0|0} = \infty \mathbb{I}$?
 → filtre information.
- Le modèle est non linéaire mais une approximation Gaussienne de la loi a posteriori suffit?
 → filtre de Kalman étendu (\perp) ou filtre de Kalman “unscented” (\top) de S.J. Julier & J.K. Uhlmann.
- On est dans un contexte non linéaire non Gaussien (loi a posteriori multimodale, etc.)
 → approximation de la loi a posteriori par des mixtures de Gaussiennes (“Gaussian mixture/sum”, cf. D.L. Alpach & H.W. Sorenson)
 → filtrage particulaire (“particle filtering”, “sequential Monte-Carlo methods”, cf. A. Doucet, N. Gordon, etc.), filtrage QMC (“quasi Monte-Carlo”).
- On veut réaliser du filtrage sur la base d'un modèle à sauts Markoviens
 → filtres (GPB et) IMM (“Interacting Multiple Model”, cf. Y. Bar-Shalom).
- etc.