

Processus Aléatoires

COMPLEMENTS COURS RASS - M. POURMIR

Chapitre 1

Initiation à la simulation aléatoire

1.1 Dans cette partie on s'initiera à quelques notions importantes par la pratique de la simulation numérique. Les références sont : [1] Lasota & Mackey, *Probabilistic Properties of Deterministic Systems*, Springer ; [2] B. Oksendal, *Stochastic Differential Equations*, Springer ; [3] P. Diaconis, *The Markov Chain Monte Carlo Revolution*, Bull. AMS, April 2009.

1.2 CHAOS ET APPLICATIONS ITEREES

On peut générer des phénomènes évolutifs avec des degrés d'irrégularité variables par l'itération d'une transformation (application) S d'un ensemble Ω dans lui-même. Ce sont alors des modèles simples et intéressants (appelés *systèmes dynamiques*) pour l'étude des processus aléatoires.

1.2.1 TRANSFORMATION QUADRATIQUE

On considère le système dynamique donné par $\Omega = [0, 1]$, $S(x) = \alpha x(1 - x)$, $x \in \Omega$, $\alpha > 0$. Pour $x_0 \in \Omega$ on étudie l'évolution de $x_1 = S(x_0)$, $x_2 = S^2(x_0) = S(S(x_0))$, ... L'ensemble Ω est l'espace des états et $\{x_n\}_{n \geq 0}$ est la trajectoire du système initialisé en x_0 .

SIMULATION Pour différentes valeurs de α et x_0 tracer la trajectoire $\{x_n\}_{n \geq 0}$. Dans la suite on pose $\alpha = 4$. S possède alors un point fixe que l'on calculera. Identifier par la simulation les domaines de x_0 qui conduisent à une évolution selon un des 3 types suivants : (a) Convergence vers le point

fixe, (b) évolution périodique, (c) évolution irrégulière ou chaotique. Dans ce dernier cas on étudiera la sensibilité des trajectoires à l'égard des petites variations de x_0 .

1.2.2 ANALYSE PROBABILISTE

Un outil puissant pour l'étude d'un processus chaotique est l'opérateur de Perron-Frobenius défini comme suit. Supposons donnée une densité de probabilité 'initiale' $p_0(\cdot)$ sur Ω . Lorsqu'on applique S à tous les points, les points transformés se répartissent suivant une nouvelle densité $p_1(\cdot)$ liée à $p_0(\cdot)$ par la relation suivante valable pour toute partie mesurable Δ :

$$\int_{\Delta} p_1(x) dx = \int_{S^{-1}(\Delta)} p_0(x) dx$$

On peut itérer cette relation pour obtenir une suite $p_n(\cdot)$, $n \geq 0$. Plus généralement, pour une densité quelconque $f(\cdot)$ sa transformée de Perron-Frobenius est définie implicitement par $f \mapsto \Pi f$ tel que :

$$\int_{\Delta} \Pi f(x) dx = \int_{S^{-1}(\Delta)} f(x) dx, \Delta \subseteq \Omega$$

On montre que $f_n = \Pi^{(n)} f_0$ a une limite f_{∞} pour $n \rightarrow \infty$ qui est, sous certaines conditions, indépendante de f_0 . f_{∞} vérifie évidemment $f_{\infty} = \Pi f_{\infty}(x)$ et de ce fait est appelée *densité invariante* du système dynamique. Intuitivement, un ensemble de points répartis suivant la densité f_{∞} auquel on applique Ω changeront de position sans que leur répartition d'ensemble change. Par exemple, pour la transformation quadratique précédente on a :

$$f_{\infty}(x) = \frac{1}{\pi \sqrt{x(1-x)}}$$

On peut maintenant donner à l'idée d'une évolution irrégulière ou chaotique un sens numérique. En gros, le degré d'irrégularité d'une évolution se mesure au fait qu'elle transforme asymptotiquement une densité initiale en une densité proche de l'uniforme. Soit Ω un espace d'états (partie fermée bornée de \mathbb{R}^n) et une transformation $S : \Omega \rightarrow \Omega$. Les 3 notions suivantes formalisent à des degrés croissants l'irrégularité des trajectoires :

(a) S est *ergodique* si pour toute densité f , $\Pi^{(n)} f$ converge en moyenne de Cesaro vers la densité uniforme, i.e :

$$\lim_{N \rightarrow \infty} \frac{1}{N} \sum_{n=0}^N \Pi^{(n)} f(x) = \frac{1}{Vol(\Omega)} \quad (\forall x)$$

(b) S est *mélangeant* (mixing) si $\Pi^{(n)}f$ converge faiblement vers la densité uniforme, i.e. pour toute fonction continue g :

$$\lim_{N \rightarrow \infty} \int_{\Omega} g(x) \cdot \Pi^{(n)}f(x) dx = \frac{1}{Vol(\Omega)} \int_{\Omega} g(x) dx$$

(c) S est *exact* si $\Pi^{(n)}f$ converge en tout point vers la densité uniforme.

La simulation suivante tente de rendre ces notions plus intuitives.

SIMULATION (Transformation d'Anosov) On prend $\Omega = [0, 1] \times [0, 1]$ et $I = [0, 0.01] \times [0, 0.01]$. Choisir un ensemble de $N = 1000$ points $\Omega_0 = \{x_{0,1}, \dots, x_{0,N}\}$ dans I et itérer S 10 fois sur Ω_0 . Enregistrer et afficher successivement les ensembles images $\Omega_0, S(\Omega_0), \dots, S^{(10)}(\Omega_0)$. Pour S on appliquera successivement les trois algorithmes suivants :

Ergodique : $S_a(x, y) = (\sqrt{2} + x, \sqrt{3} + y) \bmod 1$.

Mixing : $S_b(x, y) = (x + y, x + 2y) \bmod 1$.

Exact : $(3x + y, x + 3y) \bmod 1$.

Commentez vos résultats.

1.3 MARCHE ALEATOIRE ET POTENTIEL

1.3.1 On considère le réseau de points défini par $\Omega = \{(x, y) = (n\alpha, m\alpha); n, m \in \mathbb{Z}\}$, $\alpha > 0$ est le pas du réseau et D un domaine régulier fermé et borné de \mathbb{R}^2 de bord $B = \partial D$. On considère également une suite aléatoire *indépendante* $\{U(n), n \geq 0\}$ à valeurs $(\pm 1, \pm 1)$ équiprobables. Une marche aléatoire simple sur Ω est une suite aléatoire $\{X(n), n \geq 0\}$ à valeurs dans Ω telle que : $X(0) \in D$ et pour $n \geq 0$: $X(n+1) = X(n) + \alpha U(n)$.

SIMULATION Tracer quelques exemples de trajectoires de X .

1.3.2 On va illustrer une correspondance intéressante entre le phénomène apparemment terne décrit ci-dessus et la notion physique de potentiel. Considérons une fonction Φ harmonique sur D i.e. qui vérifie l'équation de Laplace avec les conditions au bord :

$$(a) \frac{\partial^2 \Phi}{\partial x^2} + \frac{\partial^2 \Phi}{\partial y^2} = 0 \quad \forall (x, y) \in \text{int}(D)$$

$$(b) \Phi(x, y) = f(x, y) \quad \forall (x, y) \in \partial D$$

$\Phi(x, y)$ est le potentiel induit en (x, y) par le potentiel fixé sur le bord par la fonction continue f . Soit $X(0) = (x, y)$ un point de $D \cap \Omega$ et $X(\tau_D)$ la position de X au temps de 1er franchissement du bord ∂D i.e. le premier n tel que $X(n) \in \overline{D}$. Soit $\overline{X}(\tau_D)$ le point d'intersection de la trajectoire avec ∂D . Alors on peut montrer que pour un pas de réseau assez petit :

$$\Phi(x, y) \approx \mathbb{E} \{ f(\overline{X}(\tau_D)) | X(0) = (x, y) \}$$

Cette propriété est un cas particulier de la formule de Feynman-Kac valable en toute rigueur pour le mouvement Brownien. Elle exprime le fait que le potentiel en (x, y) peut être estimé en émettant *un grand nombre* de trajectoires depuis ce point et en moyennant les valeurs obtenues pour $f(X(n))$ lorsque les trajectoires croisent le bord. C'est un exemple de méthode Monte Carlo consistant à résoudre un problème numérique par la simulation aléatoire

SIMULATION Prendre pour D le disque unité, et $\alpha = 0.01$. Définir une fonction f continue sur $[-\pi, \pi]$. Alors Φ est calculable en coordonnées polaires par la formule de Poisson :

$$\Phi(r, \theta) = \frac{1 - r^2}{2\pi} \int_{-\pi}^{\pi} \frac{f(\varphi)}{1 - 2r \cos(\theta - \varphi) + r^2} d\varphi$$

qui, pour tout θ , peut être approchée par une somme de Riemann sur une subdivision assez fine de $[-\pi, \pi]$. Choisir quelques points sur $D \cap \Omega$ et pour chaque point calculer $\Phi(x, y)$ donnée par la formule de Poisson et par la formule approchée de Feynman-Kac. Comparer et commenter le résultat.

1.4 MARCHE ALÉATOIRE ET OPTIMISATION

On reprend le réseau Ω et la marche aléatoire simple définis en 1.3. Considérons maintenant une fonction non négative $F(., .)$ définie sur \mathbb{R}^2 avec un minimum absolu en $\omega = (x^*, y^*)$. Le but est de construire une marche aléatoire modifiée $\{Z(n), n \geq 0\}$ telle que ses trajectoires tendent vers un voisinage suffisamment étroit de ω . Cette construction est appelée (à tort) algorithme de Metropolis et fonctionne comme suit :

- (a) Choisir au hasard $Z(0) = (x_0, y_0)$
- (b) Choisir T paramètre réel positif (appelé température).

- (c) Pour $n \geq 0$ calculer une réalisation de $Z^*(n) = Z(n) + U(n)$ où $U(n)$ est défini comme en 1.3.
- (d) Calculer $\Delta F = F(Z^*(n)) - F(Z(n))$.
- (e) Si $\Delta F \leq 0$ alors $Z(n+1) = Z^*(n)$.
- (f) Si $\Delta F > 0$ (i) calculer $\rho = \exp(-\Delta F/T)$; (ii) choisir $Z(n+1) = Z(n)$ avec la probabilité $1 - \rho$ et choisir $Z(n+1) = Z^*(n)$ avec la probabilité ρ .
- (g) Diminuer *lentement* T et itérer depuis (c).

Intuitivement, lorsque la température T n'est pas trop basse, les fluctuations de $Z(n)$ calculées en (e) et (f) permettent au processus de franchir les minima locaux de F avec une probabilité positive et d'approcher un voisinage de ω . Lorsque la température T est assez basse les fluctuations de Z se resserrent autour de ω . Pour T fixe on montre que $Z(n)$ est asymptotiquement stationnaire et distribué approximativement suivant la loi de Gibbs-Boltzmann :

$$P(Z(n) = (x, y)) \sim \exp - \frac{F(x, y)}{T}$$

On voit facilement que pour $T \searrow 0$ cette loi se concentre autour du minimum absolu ω et donc que la moyenne de $Z(n)$ en phase stationnaire est une bonne estimation de ω . L'avantage de cette méthode d'optimisation est sa simplicité et sa relative indépendance de la dimension (on peut généraliser à $n > 2$ sans ajout de complexité notable). En revanche le choix du protocole de refroidissement $T \searrow 0$ influe notablement sur la fiabilité et la vitesse de cette méthode. De même il est recommandé en pratique d'itérer plusieurs processus en parallèle (avec différentes initialisations) pour pallier ces défauts.

SIMULATION Programmer une fonction continue positive $F(.,.)$ de votre choix avec un profil irrégulier (oscillatoire) et avec un minimum global en $\omega = (0, 0)$. Programmer l'algorithme de Metropolis et étudier son comportement de convergence avec plusieurs initialisations et des profils de température du type $T(n) = a/(b + cn)$ et $T(n) = a/(b + \log n)$. Tracer l'évolution de $Z(n)$ et analyser les résultats.

Chapitre 2

Chaines de Markov et optimisation

2.1 LE MODELE

Les chaînes de Markov sont en un sens les modèles dynamiques aléatoires les plus simples que l'on puisse considérer et pour cette raison leurs applications sont innombrables. En un sens large, les processus étudiés dans le chapitre 1 sont tous des exemples de processus Markoviens. Formellement, une chaîne de Markov (homogène, i.e. invariante dans le temps et finie) est une séquence de variables aléatoires $X(n), n \geq 0$ à valeurs dans un ensemble fini $S = \{s_1, \dots, s_K\}$ appelé *espace des états* ou *alphabet* et caractérisé par la donnée d'un vecteur-ligne stochastique (avec K composantes non négatives et de somme unité) $\mathbf{p}(0)^T$, appelé *distribution des états initiaux*, et d'une K -matrice stochastique (dont les lignes sont des vecteurs stochastiques) $\mathbf{\Pi}$ appelée *matrice de transition*. On montre qu'il existe un processus $X(n), n \geq 0$ de loi \mathbb{P} tel que $\mathbb{P}[X(0) = s_i] = \mathbf{p}_i(0)$ et

$$\mathbb{P}[X(n+1) = s_{i_{n+1}} | X(0) = s_{i_0}, \dots, X(n) = s_{i_n}] = \mathbb{P}[X(n+1) = s_{i_{n+1}} | X(n) = s_{i_n}] \quad (2.1)$$

$X(n)$ est appelé *l'état du processus à l'instant n* et la propriété fondamentale (1) est dite *de Markov*. La propriété de Markov ramène donc la loi d'évolution du processus aux probabilités de transition entre instants continus.

Dans un espace d'états fini, la matrice de transition joue un rôle analogue à l'opérateur de Perron-Frobenius introduit en 1.2 sur des espaces plus larges, notamment dans la définition des probabilités invariantes. Brièvement, dans les cas intéressants, le vecteur de probabilité défini par $\mathbf{p}_\infty^T = \lim_{n \rightarrow \infty} \mathbf{p}^T(0) \mathbf{\Pi}^n$ existe et sa valeur est indépendante de la probabilité initiale $\mathbf{p}^T(0)$. Ce cas se produit notamment lorsque $\mathbf{\Pi}$ a toutes ses composantes positives. Dans ce cas \mathbf{p}_∞ est le vecteur des probabilités d'état invariant de la chaîne de Markov. Asymptotiquement les états seront distribués suivant cette loi de probabilité.

SIMULATION Générer des exemples de $\mathbf{\Pi}$ de dimension 5×5 et vérifier la convergence vers la loi invariante avec plusieurs choix de $\mathbf{p}(0)$. La convergence est en général exponentielle. Vérifier ce point.

2.2 FILTRAGE ET ALGORITHME DE VITERBI

2.2.1 Dans les applications pratiques la situation suivante se produit fréquemment. L'état $X(n)$, $n = 0, \dots, N$ est rarement directement observable. En pratique on a souvent accès à une autre séquence aléatoire $Y(n)$ qui est statistiquement dépendant de $X(n)$. Le problème du filtrage est le suivant : soit $y(0), \dots, y(N)$ une réalisation de $Y(0), \dots, Y(N)$. On veut déterminer la suite d'états la plus probable *a posteriori*. On supposera que la loi conjointe des $X(n)$ et $Y(n)$ est une loi à densité. Cela revient à chercher les valeurs $x(0) = \hat{x}(0), \dots, x(N) = \hat{x}(N)$ qui maximise la densité de probabilité définie pour tout N par

$$H_N(x(0), \dots, x(N)) = p[X(0) = x(0), \dots, X(N) = x(N); Y(0) = y(0), \dots, Y(N) = y(N)]$$

Dans la suite on se placera dans le cas où *conditionnellement* à $X(0), \dots, X(N)$ la séquence $Y(0), \dots, Y(N)$ est *indépendante*. Ceci se produit p. ex. si les $Y(n)$ sont fonction des $X(n)$ plus un 'bruit blanc'. On peut dans ce cas résoudre ce problème avec un algorithme numériquement efficace.

2.2.2 EXERCICE Ayant fixé $y(0), \dots, y(n)$ on définit les fonctions

$$B(s, n) = p[Y(n) = y(n) | X(n) = s], \quad A(s', s, n) = \mathbf{\Pi}_{s's} B(s, n)$$

$s, s' \in S$. En utilisant les propriétés de Markov (1) et d'indépendance conditionnelle des Y montrer que H_N se factorise comme suit :

$$H_N(x(0), \dots, x(N)) = p_{x(0)}(0) B(x(0), 0) A(x(0), x(1), 1) \dots A(x(N-1), x(N), N)$$

2.2.3 EXERCICE On définit pour tout N

$$Q(s, N) = \max_{x(0), \dots, x(N-1)} H_N(x(0), \dots, x(N-1), x(N) = s), \quad s \in S$$

Montrer que

$$Q(s, N) = \max_{s' \in S} [Q(s', N-1)A(s', s, N)]$$

2.2.4 L'ALGORITHME DE VITERBI Le résultat précédent permet de construire récursivement la suite $x(0) = \hat{x}(0), \dots, x(N) = \hat{x}(N)$ qui maximise H_N comme suit (par commodité on identifie lorsque nécessaire un état s_j avec son indice j)

(a) Initialisation : définir le vecteur de composantes $Q(j, 0) = p_j(0)B(s_j, 0)$, $j = 1, \dots, K$.

(b) Pour $n = 1, \dots, N$ construire la suite de vecteurs de composantes données par la récurrence :

$$Q(j, n) = \max_{i=1, \dots, K} [Q(i, n-1)A(s_i, s_j, n)], \quad j = 1, \dots, K$$

De même on calcule

$$\varphi(j, n) = \arg \max_{i=1, \dots, K} [Q(i, n-1)A(s_i, s_j, n)], \quad j = 1, \dots, K$$

($\arg \max f(x)$ est la valeur de x qui maximise une fonction f)

(c) Déterminer $\hat{x}(N) = \arg \max_{j=1, \dots, K} Q(j, N)$

(d) Par récurrence inverse déterminer $\hat{x}(n-1) = \varphi(\hat{x}(n), n)$, $n = N, N-1, \dots, 1$.

2.2.5 Pour préparer l'exercice de simulation on appliquera les spécifications suivantes. L'espace des états est $S = \{1, 2, 3, 4\}$. On définira la séquence d'observations par $Y(n) = h(X(n), n) + W(n)$ où $h(\cdot, n)$ sont des fonctions réelles connues, $W(n)$ est une suite de variables aléatoires indépendantes normales, centrées et de variance $b > 0$. On a donc $B(s, n) = \frac{1}{\sqrt{2\pi b}} \exp -\frac{1}{2b}(y(n) - h(s))^2$.

2.2.6 SIMULATION Prendre $N = 100$, $h(s, n) = s$. Générer une suite $x(0), \dots, x(N)$ avec un modèle de Markov de distribution initiale uniforme et Π fixé. Générer la séquence $y(0), \dots, y(n)$ obtenu avec le modèle 1.4.5. Programmer l'algorithme de Viterbi et l'appliquer à la suite $y(0), \dots, y(n)$. Comparer la séquence filtrée (estimée) obtenue $\hat{x}(0), \dots, \hat{x}(N)$ à la séquence

'réelle' $x(0), \dots, x(N)$ pour différentes valeurs de b .

2.2.7 REMARQUE En utilisant l'algorithme 1.4.4 il peut parfois s'avérer que l'on manipule des nombres très petits, ce qui peut poser des problèmes de précision de calcul. Dans ce cas on aura intérêt à modifier légèrement la procédure d'optimisation précédente en cherchant le maximum de $\overline{H} = \log H$ au lieu de H . En surmontant d'une barre les symboles de variables transformées en leur logarithme, on réécrira aisément les procédures (a), (b), (c) ci-dessus modifiées en conséquence. L'étape (b) p.ex. se transformera en $\overline{Q}(j, n) = \max_{i=1, \dots, K} [\overline{Q}(i, n-1) + \overline{A}(s_i, s_j, n)]$, $j = 1, \dots, K$, etc.

2.3 APPLICATION A L'EGALISATION

2.3.1 Le modèle de Markov et l'algorithme de Viterbi peuvent être utilisés pour résoudre un problème crucial en télécommunications. Nous en verrons ici une version simplifiée. Un signal de communication consiste en une suite binaire aléatoire i.e. une séquence de variables aléatoires indépendantes $Z(n), n \geq 0$ à valeurs dans $\{-1, 1\}$ avec $\mathbb{P}[Z(n) = 1] = 1/2$. Un *message* est la donnée d'un bloc $z(0), \dots, z(N)$ de réalisations. Le signal reçu par un récepteur est de la forme

$$Y(0) = h_1 Z(0), Y(n) = h_0 Z(n) + h_1 Z(n-1) + W(n), n \geq 1$$

Dans cette dernière relation les 2 premiers termes mesurent l'effet du canal de transmission qui résulte en une interférence entre bits contigus (ici limitée à 2) dans le temps. $\mathbf{h}^T = [h_0, h_1]$ est la réponse impulsionnelle du canal et $W(n)$ est une séquence normale centrée indépendante de variance $b > 0$ qui modèle l'effet du bruit.

2.3.2 EXERCICE On définit $\mathbf{X}(0) = [Z(0), -1]$, $\mathbf{X}(n) = [Z(n), Z(n-1)]$, $n \geq 1$. Montrer que $\mathbf{X}(n)$ est une chaîne de Markov dont on précisera l'espace des états, la distribution initiale et la matrice de transition. Utiliser 1.2.3 pour vérifier si cette chaîne est irréductible (et donc ergodique).

2.3.3 SIMULATION Ecrire un sous-programme pour générer des réalisations $z(0), \dots, z(N)$ et $y(0), \dots, y(N)$ respectivement de $Z(0), \dots, Z(N)$ et $Y(0), \dots, Y(N)$. Appliquer l'algorithme de Viterbi à ce nouveau modèle pour obtenir une estimation des valeurs de $\hat{z}(0), \dots, \hat{z}(N)$ maximisant la probabilité a posteriori. Etudier l'influence de la réponse du canal \mathbf{h} et de la variance b du bruit sur le taux d'erreur d'estimation par bit transmis.

2.4 APPLICATION AU DECODAGE

2.4.1 Une bactérie exotique code et transmet sa séquence génétique sur 2 bits, ce qui nous permet de l'assimiler à un processus indépendant et à valeurs équiprobables $Z(n) \in \{0, 1\}$, $n \geq 0$. Au cours de la reproduction, ce signal, transmis aux descendants, est susceptible d'être dégradé par des erreurs de transcription; c'est pourquoi la bactérie parent l'immunise au préalable avec un codage primitif comme suit. La séquence originale $Z(n)$ est d'abord transformée en le processus à 4 états $\mathbf{X}(n) = [Z(n), Z(n+1)]$ qui fonctionne comme un registre à décalage et introduit de la redondance. Ensuite, le signal $\mathbf{X}(n)$, $n \geq 0$ est transformé en le processus $\mathbf{V}(n) = h(\mathbf{X}(n))$, $n \geq 0$ où l'application injective $h : \mathbb{C}^2 \rightarrow \mathbb{C}^3$, applique le cube unité de \mathbb{R}^2 dans son homologue de \mathbb{R}^3 suivant le tableau de codage :

$$\begin{array}{ll} [0, 0] & \mapsto [0, 0, 0] \\ [0, 1] & \mapsto [0, 1, 1] \\ [1, 0] & \mapsto [1, 0, 0] \\ [1, 1] & \mapsto [1, 1, 1] \end{array}$$

L'effet du bruit (erreurs de transcription) sur le signal transmis $\mathbf{V}(n)$ est modélisé par le fait que le signal reçu est le processus $\mathbf{Y}(n)$ obtenu de $\mathbf{V}(n)$ en inversant de manière aléatoire et indépendante la valeur du bit de chaque composante de $\mathbf{V}(n)$ avec une probabilité $0 < q < 1$ connue.

2.4.2 EXERCICE Dessiner \mathbb{C}^3 et représenter le domaine image de $h(\cdot)$. Quel est l'avantage de transformer $\mathbf{X}(n)$ en $\mathbf{V}(n)$? Déterminer la matrice de transition Π de la chaîne de Markov $\mathbf{X}(n)$. Calculer la loi conditionnelle $\mathbb{P}(\mathbf{Y}(n) = y | \mathbf{X}(n) = x)$ pour tous les couples $(x, y) \in \mathbb{C}^2 \times \mathbb{C}^3$. On stockera le résultat de ce calcul dans une matrice aux dimensions appropriées pour la suite. Le but étant maintenant d'estimer la séquence originale $Z(n)$ à partir de $\mathbf{Y}(n)$, il faut apporter quelques changements mineurs et évidentes à l'algorithme de Viterbi pour tenir compte du fait que $\mathbf{Y}(n)$ étant à valeurs discrètes, la grandeur maximisée H_N n'est plus la densité conjointe, mais la loi conjointe $\mathbb{P}(\mathbf{X}(n) = x_n, n = 0, \dots, N; \mathbf{Y}(n) = y_n, n = 0, \dots, N)$. On définira donc notamment $\mathbf{B}(s, n) = \mathbb{P}[\mathbf{Y}(n) = y(n) | \mathbf{X}(n) = s]$.

2.4.3 SIMULATION Ecrire un sous-programme pour générer des réalisations $z(0), \dots, z(N)$ et $y(0), \dots, y(N)$ respectivement de $Z(0), \dots, Z(N)$ et $\mathbf{Y}(0), \dots, \mathbf{Y}(N)$. Appliquer l'algorithme de Viterbi à ce nouveau modèle pour obtenir une estimation des valeurs de $\hat{z}(0), \dots, \hat{z}(N)$ maximisant la probabilité a posteriori. Etudier l'influence de la probabilité d'inversion de bit q sur

le taux d'erreur d'estimation par bit transmis.

Chapitre 3

Corrélation et filtrage

Le but de cette partie est de montrer sur 2 exemples l'application des propriétés de corrélation des séquences aléatoires au problème du filtrage.

3.1 RAPPELS SUR L'ESTIMATION LINEAIRE

3.1.1 Dans toute la suite on travaillera dans l'espace des variables aléatoires (v.a.) réelles, centrées et de variance finie associées à un modèle aléatoire implicite. Considérons deux séquences de v.a. X_1, \dots, X_m et Y_1, \dots, Y_n . En définissant les vecteurs aléatoires $\mathbf{X} = [X_1, \dots, X_m]^T$ et $\mathbf{Y} = [Y_1, \dots, Y_n]^T$ l'estimation linéaire de \mathbf{X} connaissant \mathbf{Y} (dite aussi régression linéaire de \mathbf{X} sur \mathbf{Y}), noté $\mathbf{L}^{\mathbf{Y}}\mathbf{X}$, est le vecteur $\hat{\mathbf{X}} = \mathbf{P}\mathbf{Y}$ où la matrice \mathbf{P} est la valeur de l'opérateur \mathbf{A} minimisant l'écart quadratique $\mathbf{Q} = \text{Tr Cov}(\mathbf{X} - \mathbf{A}\mathbf{Y}) = \text{Tr } \mathbb{E}(\mathbf{X} - \mathbf{A}\mathbf{Y})(\mathbf{X} - \mathbf{A}\mathbf{Y})^T$. \mathbf{P} est donnée par la formule explicite (avec les hypothèses d'inversibilité adaptées)

$$\mathbf{P} = \mathbb{E}(\mathbf{X}\mathbf{Y}^T)\mathbb{E}(\mathbf{Y}\mathbf{Y}^T)^{-1} \quad (3.1)$$

Le vecteur erreur d'estimation $\mathbf{X}^* = \mathbf{X} - \mathbf{P}\mathbf{Y}$ est non corrélé avec \mathbf{Y} et de covariance

$$\text{Cov } \mathbf{X}^* = \mathbb{E}(\mathbf{X}\mathbf{X}^T) - \mathbb{E}(\mathbf{X}\mathbf{Y}^T)\mathbb{E}(\mathbf{Y}\mathbf{Y}^T)^{-1}\mathbb{E}(\mathbf{Y}\mathbf{X}^T) \quad (3.2)$$

3.1.2 EXERCICE En utilisant (1) vérifier que si \mathbf{H} est une matrice de dimensions adaptées on a :

$$\mathbf{L}^{\mathbf{Y}}\mathbf{H}\mathbf{X} = \mathbf{H}\mathbf{L}^{\mathbf{Y}}\mathbf{X} = \mathbf{H}\mathbf{P}\mathbf{Y} \quad (3.3)$$

De même vérifier que pour un opérateur \mathbf{A} inversible

$$\mathbf{L}^{\mathbf{A}\mathbf{Y}}\mathbf{X} = \mathbf{L}^{\mathbf{Y}}\mathbf{X} \quad (3.4)$$

Montrer que si $Y_n = Y^*$ est une v.a. non corrélée avec Y_1, \dots, Y_{n-1} alors

$$\mathbb{L}^{Y_1, \dots, Y_{n-1}, Y^*} \mathbf{X} = \mathbb{L}^{Y_1, \dots, Y_{n-1}} \mathbf{X} + \mathbb{L}^{Y^*} \mathbf{X} \quad (3.5)$$

3.2 UN MODELE DE TRAITEMENT D'ANTENNE

3.2.1 LE SIGNAL RECU EN UN POINT Le modèle du signal délivré par un capteur placé en un point de l'espace est

$$\mathbf{Y} = \mathbf{X} + \mathbf{W} = \mathbf{S}\boldsymbol{\theta} + \mathbf{W} \quad (3.6)$$

où $\mathbf{S} = [\mathbf{S}_1, \dots, \mathbf{S}_m]$ est une matrice $n \times m$ non aléatoire dont les colonnes linéairement indépendantes sont appelées *sources*, $\boldsymbol{\theta} = [\theta_1, \dots, \theta_m]^T$ est un vecteur (*message*) aléatoire centré et de covariance \mathbf{C} , \mathbf{W} est le vecteur bruit non corrélé avec $\boldsymbol{\theta}$ et de loi $\mathcal{N}(0, \beta \mathbf{I}_n)$, $\beta > 0$.

EXERCICE Avec ces données, vérifier les expressions

$$\mathbf{P} = \mathbf{S}\mathbf{C}\mathbf{S}^T \left[\mathbf{S}\mathbf{C}\mathbf{S}^T + \beta \mathbf{I}_n \right]^{-1} \quad (3.7)$$

et

$$\mathbf{P} = \mathbf{I}_n - \beta \Gamma_{\mathbf{Y}}^{-1} \quad (3.8)$$

avec $\Gamma_{\mathbf{Y}} = \text{Cov} \mathbf{Y}$.

3.2.2 En pratique, l'applicabilité de la formule (7) est limitée du fait que \mathbf{S} et \mathbf{C} sont généralement inconnues. La relation (8) montre cependant que \mathbf{P} ne dépend en réalité que de $\Gamma_{\mathbf{Y}}$ et β qui peuvent être estimés à partir des observations comme on va le voir. A cet effet considérons M copies du vecteur aléatoire \mathbf{Y} , c.à.d des vecteurs $\mathbf{Y}(1), \dots, \mathbf{Y}(M)$, $M > n$, de même loi que \mathbf{Y} et indépendants. Ces vecteurs sont supposés fournis par M capteurs formant une *antenne*. Pour M assez grand, d'après la loi forte des grands nombres

$$\hat{\Gamma}_{\mathbf{Y}} = \frac{\mathbf{Y}(1)\mathbf{Y}(1)^T + \dots + \mathbf{Y}(M)\mathbf{Y}(M)^T}{M} \approx \Gamma_{\mathbf{Y}} \quad (3.9)$$

Dans les simulations que l'on effectuera on supposera que cette approximation est suffisamment précise pour pouvoir plaquer sans trop d'erreur sur $\hat{\Gamma}_{\mathbf{Y}}$ la structure de $\Gamma_{\mathbf{Y}}$ (et de \mathbf{P}) que l'on découvrira dans l'exercice suivant.

EXERCICE Soit \mathcal{S} le sous espace de \mathbb{R}^n engendré par les colonnes de \mathbf{S} et Π le projecteur orthogonal sur \mathcal{S} . Montrer que $\Pi = \mathbf{S} [\mathbf{S}^T \mathbf{S}]^{-1} \mathbf{S}^T$. On peut

avoir une idée de la structure de \mathbf{P} en se plaçant dans le cas où $\mathbf{C} = \mathbf{I}_m$. A cet effet combiner (7) et l'identité générale $[\mathbf{I} + \mathbf{A}\mathbf{A}^T]^{-1} = \mathbf{I} - \mathbf{A}[\mathbf{I} + \mathbf{A}^T\mathbf{A}]^{-1}\mathbf{A}^T$ pour montrer que

$$\mathbf{P} = \mathbf{\Pi} + o\left(\beta/\text{Tr}(\mathbf{S}^T\mathbf{S})\right) \quad (3.10)$$

Ainsi, pour les rapports signal/bruit forts et des messages non corrélés l'estimateur optimal est proche du projecteur orthogonal sur l'espace source \mathcal{S} .

En développant l'expression $\mathbf{V}^T\mathbf{\Gamma}_Y\mathbf{V}$ pour un vecteur propre montrer que $\lambda \geq \beta$. En conclusion, β est valeur propre minimale de $\mathbf{\Gamma}_Y$ et donc les valeurs propres peuvent être classées comme $\lambda_1 \geq \dots \geq \lambda_m > \lambda_{m+1} = \dots = \lambda_n = \beta$ avec les vecteurs propres orthonormés associés $\mathbf{U}_1, \dots, \mathbf{U}_m, \mathbf{U}_{m+1}, \dots, \mathbf{U}_n$. Montrer que $\mathbf{U}_1, \dots, \mathbf{U}_m$ engendre \mathcal{S} et

$$\mathbf{P} = \sum_{i=1}^m \left(1 - \frac{\beta}{\lambda_i}\right) \mathbf{U}_i \mathbf{U}_i^T \quad (3.11)$$

Retrouver (10) avec cette formule et montrer que la variance d'erreur $Q_{\min} = \text{Tr Cov}(\mathbf{X} - \mathbf{P}\mathbf{Y}) = m\beta \left(1 - \frac{1}{m} \sum_{i=1}^m \frac{\beta}{\lambda_i}\right)$. Développer au 1er ordre cette quantité pour $\beta \rightarrow 0$ et $\beta \rightarrow \infty$ (attention : les λ_i ne sont pas indépendants de β !)

3.2.3 SIMULATION On simulera d'abord le signal délivré par une antenne de $M = 400$ capteurs obéissant chacun au modèle (6) avec des $\boldsymbol{\theta}$ et \mathbf{W} indépendants. On choisira $m = 2$, $n = 100$, $0 < \beta < 1$, les composantes de $\boldsymbol{\theta}$ sont deux v.a. indépendantes uniformément distribuées dans $[-1, 1]$, les signatures \mathbf{S}_1 et \mathbf{S}_2 sont des vecteurs de composantes $\sin(2\pi\nu k)$, $k = 0, \dots, n-1$ pour deux valeurs différentes de $\nu \in]0, 1/2[$. Donner une estimation empirique $\hat{\mathbf{\Gamma}}_Y$ de $\mathbf{\Gamma}_Y$ par la formule (9). Utiliser la fonction `svd` de Matlab pour tracer la courbe des valeurs propres $\hat{\lambda}_i$ de $\hat{\mathbf{\Gamma}}_Y$ dans l'ordre décroissant. On notera que $\hat{\beta} = \frac{\lambda_{m+1} + \dots + \lambda_n}{n-m}$ donne une bonne estimation de β . On estimera ensuite \mathbf{P} par $\hat{\mathbf{P}}$ obtenu de la formule (11) en remplaçant les grandeurs exactes par celles estimées à partir de $\hat{\mathbf{\Gamma}}_Y$. Calculer et tracer $\hat{\mathbf{X}} = \hat{\mathbf{P}}\mathbf{Y}$ pour des valeurs croissantes de β et les comparer aux réalisations du vecteur \mathbf{X} correspondants.

3.3. MODELE MARKOVIAN LINEAIRE ET FILTRE DE KALMAN

3.3.1 On considère deux séquences aléatoires centrées globalement indépendantes $\mathbf{U}(n)$ et $\mathbf{W}(n)$, $n \geq 0$ à valeurs resp. dans \mathbb{R}^K et \mathbb{R}^M et de matrices de covariance resp. $\mathbf{Q}_U(n)$ et $\mathbf{Q}_W(n)$. Avec ces 2 processus on construit les

processus $\mathbf{X}(n)$ et $\mathbf{Y}(n)$ à valeurs dans \mathbb{R}^K et \mathbb{R}^M et obéissant aux statistiques initiales : $\mathbb{E}\mathbf{X}(0) = 0$; $\text{Cov}\mathbf{X}(0) = \mathbf{Q}_X(0)$ et pour $n \geq 0$ aux équations

$$\mathbf{X}(n+1) = \mathbf{A}(n)\mathbf{X}(n) + \mathbf{U}(n) ; \mathbf{Y}(n) = \mathbf{C}(n)\mathbf{X}(n) + \mathbf{W}(n) \quad (3.12)$$

où $\mathbf{A}(n)$ et $\mathbf{C}(n)$ sont deux suites de matrices non aléatoires réelles de dimensions appropriées. Dans ce contexte, $\mathbf{X}(n)$ et $\mathbf{Y}(n)$ sont resp. appelées séquences d'état et d'observation et $\mathbf{U}(n)$ et $\mathbf{W}(n)$ sont resp. les bruits d'état et d'observation. Le modèle ainsi posé donne une description simplifiée du comportement de systèmes dynamiques soumis à des sollicitations aléatoires et des observations bruitées.

3.3.2 EXERCICE Pour $n \geq 1$ calculer $\mathbf{Q}_X(n)$.

3.3.4 EVOLUTION DE L'ESTIMATEUR LINEAIRE Définissons $\mathbf{X}^+(n) = \mathbf{L}^{\mathbf{Y}_0, \dots, \mathbf{Y}_n} \mathbf{X}(n)$. Au temps $n+1$, lorsqu'une nouvelle observation $\mathbf{Y}(n+1)$ est enregistrée, le calcul de $\mathbf{X}^+(n+1) = \mathbf{L}^{\mathbf{Y}_0, \dots, \mathbf{Y}_{n+1}} \mathbf{X}(n+1)$ est en principe à refaire. En fait, nous allons voir que l'estimateur $\mathbf{X}^+(n)$ obéit lui-même à une récurrence linéaire, rendant sa programmation attrayante et efficace. Dans la suite on utilisera les résultats de 2.1.1 et 2.1.2 pour découvrir la structure de cette récurrence.

EXERCICE On définit $\mathbf{X}^-(0) = 0$ et pour $n \geq 1$ $\mathbf{X}^-(n) = \mathbf{L}^{\mathbf{Y}_0, \dots, \mathbf{Y}_{n-1}} \mathbf{X}(n)$, $\mathbf{X}^*(n) = \mathbf{X}(n) - \mathbf{X}^-(n)$, et $\mathbf{Y}^*(n) = \mathbf{Y}(n) - \mathbf{L}^{\mathbf{Y}_0, \dots, \mathbf{Y}_{n-1}} \mathbf{Y}(n)$. La transformation

$$\begin{bmatrix} \mathbf{Y}_0 \\ \dots \\ \mathbf{Y}_{n-1} \\ \mathbf{Y}_n \end{bmatrix} \mapsto \begin{bmatrix} \mathbf{Y}_0 \\ \dots \\ \mathbf{Y}_{n-1} \\ \mathbf{Y}_n^* \end{bmatrix}$$

étant linéaire et inversible, appliquer les formules (4), (5) et (1) pour montrer les relations : $\mathbf{X}^+(n) = \mathbf{X}^-(n) + \mathbf{L}^{\mathbf{Y}^*} \mathbf{X}(n)$ et $\mathbf{L}^{\mathbf{Y}^*} \mathbf{X}(n) = \mathbf{K}(n) \mathbf{Y}^*(n)$ avec $\mathbf{K}(n) = \mathbb{E} [\mathbf{X}(n) \mathbf{Y}^{*T}(n)] \mathbb{E} [\mathbf{Y}^*(n) \mathbf{Y}^{*T}(n)]^{-1}$. En appliquant (3) montrer que $\mathbf{Y}^*(n) = \mathbf{Y}(n) - \mathbf{C}(n) \mathbf{X}^-(n) = \mathbf{C}(n) \mathbf{X}^*(n) + \mathbf{W}(n)$ et en déduire d'après (2) les relations $\mathbf{X}^+(n) = \mathbf{X}^-(n) + \mathbf{K}(n) [\mathbf{Y}(n) - \mathbf{C}(n) \mathbf{X}^-(n)]$ et $\mathbf{K}(n) = \mathbf{P}^-(n) \mathbf{C}(n)^T [\mathbf{C}(n) \mathbf{P}^-(n) \mathbf{C}(n)^T + \mathbf{Q}_W(n)]^{-1}$ avec $\mathbf{P}^-(n) = \text{Cov} \mathbf{X}^*(n)$, $\mathbf{P}^-(0) = \mathbf{Q}_X(0)$.

Il reste à établir des récurrences sur $\mathbf{X}^-(n)$ et $\mathbf{P}^-(n)$. Montrer que $\mathbf{X}^-(n+1) = \mathbf{A}(n) \mathbf{X}^+(n)$. En déduire $\mathbf{X}^*(n) = \mathbf{A}(n) [\mathbf{X}(n) - \mathbf{X}^+(n)] + \mathbf{U}(n)$, ce qui donne : $\mathbf{P}^-(n+1) = \mathbf{A}(n) \mathbf{P}^+(n) \mathbf{A}(n)^T + \mathbf{Q}_U(n)$ où $\mathbf{P}^+(n) = \text{Cov} [\mathbf{X}(n) - \mathbf{X}^+(n)]$. Il reste à voir que $\mathbf{X}(n) - \mathbf{X}^+(n) = [\mathbf{I} - \mathbf{K}(n) \mathbf{C}(n)] \mathbf{X}^*(n) - \mathbf{K}(n) \mathbf{W}(n)$, ce qui

donne $\mathbf{P}^+(n) = [\mathbf{I} - \mathbf{K}(n)\mathbf{C}(n)] \mathbf{P}^-(n) [\mathbf{I} - \mathbf{K}(n)\mathbf{C}(n)]^T + \mathbf{K}(n)\mathbf{Q}_W(n)\mathbf{K}(n)^T$.

3.3.5 RECAPITULATION En rassemblant les résultats précédents, l'évolution de l'estimateur $\mathbf{X}^+(n)$ de $\mathbf{X}(n)$ obéissant au modèle 2.3.1 se décrit ainsi :

(i) Initialisation : $\mathbf{X}^-(0) = 0$, $\mathbf{P}^-(0) = \mathbf{Q}_X(0)$.

(ii) Pour $n \geq 0$

$$\mathbf{K}(n) = \mathbf{P}^-(n)\mathbf{C}(n)^T [\mathbf{C}(n)\mathbf{P}^-(n)\mathbf{C}(n)^T + \mathbf{Q}_W(n)]^{-1}$$

$$\mathbf{X}^+(n) = \mathbf{X}^-(n) + \mathbf{K}(n) [\mathbf{Y}(n) - \mathbf{C}(n)\mathbf{X}^-(n)]$$

$$\mathbf{X}^-(n+1) = \mathbf{A}(n)\mathbf{X}^+(n)$$

$$\mathbf{P}^+(n) = [\mathbf{I} - \mathbf{K}(n)\mathbf{C}(n)] \mathbf{P}^-(n) [\mathbf{I} - \mathbf{K}(n)\mathbf{C}(n)]^T + \mathbf{K}(n)\mathbf{Q}_W(n)\mathbf{K}(n)^T$$

$$\mathbf{P}^-(n+1) = \mathbf{A}(n)\mathbf{P}^+(n)\mathbf{A}(n)^T + \mathbf{Q}_U(n)$$

(iii) $n = n + 1$ et retour à (ii).

3.3.6 SIMULATION Pour simuler le modèle 2.3.1 on choisira $K = 2$, $M = 1$ et deux constantes de structure $0 < \nu < 1/2$ et $\tau \geq 10$. On posera $\mathbf{A}(n) = \begin{bmatrix} 0 & 1 \\ -e^{-2/\tau} & 2e^{-1/\tau} \cos 2\pi\nu \end{bmatrix}$, $\mathbf{C}(n) = [0, 1]$, $\mathbf{U}(n) = \begin{bmatrix} 0 \\ U(n) \end{bmatrix}$ avec $U(n) \sim \mathcal{N}(0, \beta_U)$, $\mathbf{W}(n) \sim \mathcal{N}(0, \beta_W)$. Les composantes de $\mathbf{X}(n)$ obéissent à une récurrence linéaire du 2nd ordre avec $U(n)$ comme 2ème membre que l'on explicitera. C'est un exemple de processus auto-regressif que l'on étudiera plus en détail dans l'étude suivante. On initialisera les composantes de $\mathbf{X}(0)$ non corrélées et de loi $\mathcal{N}(0, \beta_{X(0)})$. Calculer $\mathbf{Q}_X(0)$.

Programmer le modèle $(\mathbf{X}(n), \mathbf{Y}(n))$ avec les données ci-dessus. Programmer l'estimateur récursif de Kalman donné en 2.3.5 et comparer l'évolution de $\mathbf{X}(n)$ et $\mathbf{X}^+(n)$ pour plusieurs choix de constantes de structure et avec des valeurs de β_W croissantes.

Chapitre 4

Analyse spectrale

4.1 On se focalisera dans cette étude sur les processus stationnaires du second ordre. Pour ces signaux, un lien particulier entre la structure dynamique et le contenu spectral existe. Estimer le modèle dynamique d'un tel signal à partir d'une réalisation finie permet de 'compresser' l'information qu'il contient i.e. de remplacer la réalisation observée par un codage plus économique qui simplifie les opérations de reconnaissance automatique, stockage, transmission etc.

4.2 INNOVATION ET PREDICTION LINEAIRE

4.2.1 On se replace dans le cadre de l'estimation linéaire résumé au début de l'étude précédente. Toutes les v.a. rencontrées seront des éléments de L^2 i.e. centrées et avec des moments du 2nd ordre finis. Les propriétés déjà énumérées peuvent se résumer comme suit. La régression linéaire d'une v.a. centrée du second ordre sur un ensemble dénombrable de v.a. $\{Y_1, Y_2, \dots\}$ est la projection orthogonale $\hat{X} = \mathbb{L}^H X$ de X sur le sous-espace de L^2 engendré $H = \text{span}(Y_1, Y_2, \dots)$. Cette projection s'entend au sens de la minimisation de la variance de l'écart de X à une combinaison linéaire de Y_1, Y_2, \dots . A un processus $(X_n)_{n \in \mathbb{Z}}$ on peut alors associer les constructions suivantes : $H_n = \text{span}(X_n, X_{n-1}, \dots)$, $H_{n,k} = \text{span}(X_n, \dots, X_{n-k})$; $k \geq 0$. On définit de même les projections $X_n^- = \mathbb{L}^{H_{n-1}} X_n$, $X_{n,k}^- = \mathbb{L}^{H_{n-1,k}} X_n$, resp. prédiction sur le passé infini et prédiction sur le passé de longueur k de X_n , la première étant aussi la limite en moyenne quadratique de la seconde pour $k \rightarrow \infty$ (par continuité de la projection), et

$$X_n - X_n^- = X_n^* \quad (4.1)$$

qui est l'erreur de prédiction, orthogonale (non corrélée) avec H_{n-1} . On a

évidemment les inclusions $H_n \supseteq H_{n-1} \supseteq \dots$. Soit $H_{-\infty} = \bigcap_{j=0}^{\infty} H_{n-j}$. De même on définira H_{∞} comme l'espace vectoriel de toutes les v.a. de la forme $Z = \sum_{n \in \mathbb{Z}} c_n X_n$ avec $\sum_{n \in \mathbb{Z}} |c_n|^2 < \infty$. On a donc la hiérarchie : $H_{-\infty} \subseteq \dots \subseteq H_{n-1} \subseteq H_n \subseteq \dots \subseteq H_{\infty}$. Posons $X_n^s = \mathbb{L}^{H_{-\infty}} X_n$ et $X_n^r = X_n - X_n^s$. (X_n^s) et (X_n^r) sont resp. les parties singulière et régulière du processus (X_n) qui sont, par construction, non corrélées. Un *processus régulier* est tel que $H_{-\infty} = \{0\}$ et donc $X_n^s = 0$.

EXERCICE En projetant X_n sur un sous-espace approprié montrer que pour un processus régulier $\lim_{k \rightarrow \infty} \Gamma(k) = 0$ où $\Gamma(k) = \mathbb{E}[X_n X_{n-k}]$ est la fonction de corrélation de (X_n) .

On montre que si (X_n) est stationnaire, (X_n^s) et (X_n^r) le sont aussi. Supposons maintenant que (X_n) soit régulier. Le processus $X_n^* = X_n - X_n^-$ appelé *innovation* dans ce contexte est tel que, par construction : $H_n = \mathbb{R}.X_n^* \oplus H_{n-1}$ où \oplus désigne la somme orthogonale. En itérant : $H_n = \mathbb{R}.X_n^* \oplus \dots \oplus \mathbb{R}.X_{n-j}^* \oplus H_{n-j-1}$. Ceci veut dire que pour n fixé existent des b_k , $0 \leq k \leq j$ tel que $X_n = b_0 X_n^* + \dots + b_j X_{n-j}^* + \mathbb{L}^{H_{n-j-1}} X_n$. En faisant $j \rightarrow \infty$, et puisque $\mathbb{L}^{H_{-\infty}} = 0$ pour un processus régulier :

$$X_n = \sum_{k=0}^{\infty} b_k X_{n-k}^* \quad (4.2)$$

avec $b_0 = 1$ et $\sum_{k=0}^{\infty} |b_k|^2 < \infty$. On écrira aussi dans le langage des séries formelles : $\tilde{X}(z) = B(z) \tilde{X}^*(z)$ avec $B(z) = b_0 + b_1 z^{-1} + b_2 z^{-2} + \dots$ et où z^{-1} est le symbole retard unité. De plus la stationnarité de (X_n) entraîne que les b_k sont indépendants de n . En conclusion un processus régulier obéit à un modèle du type (1) dit de moyenne mobile ou MA (moving average) où le signal est la sortie d'un filtre de réponse (b_k) avec le processus d'innovation comme entrée.

4.2.2 Les deux représentations (1) et (2) sont à la base de l'estimation spectrale paramétrique (par modèle) d'un processus régulier. Pour le voir p.ex. sur (2), en appliquant la formule de transformation linéaire du spectre on obtient pour la densité spectrale de (X_n) :

$$\gamma(\lambda) = \beta^* \left| \sum_{k=0}^{\infty} b_k e^{-i2\pi k \lambda} \right|^2 ; 0 \leq \lambda < \frac{1}{2} \quad (4.3)$$

où $\beta^* = E |X_n^*|^2$ est la variance de l'innovation. En principe donc estimer la densité spectrale et estimer les coefficients (b_k) du modèle MA sont

équivalents. De la relation (2) on peut encore tirer le système d'équations non linéaire en les paramètres β^* et b_k :

$$\Gamma(k) = \beta^* \sum_{j=k}^{\infty} b_j b_{j-k}; \quad k = 0, 1, 2, \dots \quad (4.4)$$

EXERCICE Prouver la relation (4).

Une autre représentation dite autorégressive ou AR pour (X_n) régulier est directement issu de (1) si on écrit $X_n^- = \mathbf{L}^{H_{n-1}} X_n = \sum_{k=1}^{\infty} a_k X_{n-k}$ où les a_k vérifient $\sum_{k=1}^{\infty} |a_k|^2 < \infty$ et par stationnarité ne dépendent pas de n . On a donc

$$X_n = \sum_{k=1}^{\infty} a_k X_{n-k} + X_n^* \quad (4.5)$$

ou encore symboliquement : $\tilde{X}(z) = [1 - A(z)]^{-1} \tilde{X}^*(z)$ où $A(z) = \sum_{k=1}^{\infty} a_k z^{-k}$. En comparant avec la relation analogue obtenue pour la représentation MA on obtient l'équation suivante entre les séries formelles associés aux coefficients MA et AR :

$$1 + b_1 z^{-1} + b_2 z^{-2} + \dots = \frac{1}{1 - A(z)} \quad (4.6)$$

EXERCICE Dédurre de (5) le système infini d'équations pour la fonction de corrélation :

$$\Gamma(k) = \sum_{j=1}^{\infty} a_k \Gamma(k-j); \quad k = 0, 1, 2, \dots \quad (4.7)$$

et la formule suivante pour la densité spectrale :

$$\gamma(\lambda) = \frac{\beta^*}{|1 - \sum_{k=1}^{\infty} a_k e^{-i2\pi k\lambda}|^2}; \quad 0 \leq \lambda < \frac{1}{2} \quad (4.8)$$

EXERCICE Le but de cet exercice est de trouver une formule asymptotique pour β^* . Sa solution n'est pas indispensable pour la compréhension de la suite mais est utile dans l'estimation des paramètres par le maximum de vraisemblance (non traité dans cette étude). En reprenant la régression sur le passé fini défini en 2.3.1 posons $X_{n,k}^* = X_n - X_{n,k}^-$, $\beta_k = E |X_{n,k}^*|^2$,

$\mathbf{g}_k^T = [\Gamma(1), \dots, \Gamma(k)]$ et $\Gamma_k = \text{Cov}(X_{n-1}, \dots, X_{n-k}) = [\Gamma(i-j)]_{1 \leq i, j \leq k}$. En développant le déterminant de la matrice $\Gamma_{k+1} = \begin{bmatrix} \Gamma(0) & \mathbf{g}_k^T \\ \mathbf{g}_k & \Gamma_k \end{bmatrix}$ suivant sa première ligne montrer que :

$$\det \Gamma_{k+1} = [\Gamma(0) - \mathbf{g}_k^T \Gamma_k^{-1} \mathbf{g}_k] \det \Gamma_k$$

Montrer que le terme entre crochets n'est autre que β_k et donc

$$\beta_k = \frac{\det \Gamma_{k+1}}{\det \Gamma_k} \quad (4.9)$$

On veut maintenant prouver $\beta^* = \lim_{k \rightarrow \infty} \beta_k$. Soit $F = \text{span}(X_{n-1}, \dots, X_{n-k})$ et le sous-espace G tel que $H_{n-1} = G \oplus H_{n-1-k}$. Les vecteurs (v.a.) dans F ont a priori des projections non nulles dans G et dans H_{n-1-k} , c'est pourquoi on écrira $F = G \oplus H$ où $H \subseteq H_{n-1-k}$ et de même $\mathbf{f} = \mathbf{g} + \mathbf{h}$ où \mathbf{f} , \mathbf{g} , \mathbf{h} sont resp. les projections sur F , G , H . Montrer que $X_{n,k}^- = \mathbf{g}X_n + \mathbf{h}L^{H_{n-1-k}}X_n$ et $X_{n,k}^* = X_n^* + [\mathbf{I} - \mathbf{h}]L^{H_{n-1-k}}X_n$. En déduire $\beta_k = \beta^* + \mathbb{E} |[\mathbf{I} - \mathbf{h}]L^{H_{n-1-k}}X_n|^2$ et (borner le 2e terme) $\beta^* = \lim_{k \rightarrow \infty} \beta_k$. Finalement, en utilisant le fait que dans une suite convergente les moyennes arithmétiques convergent vers la limite de la suite on déduit de (8) la formule

$$\log \beta^* = \lim_{k \rightarrow \infty} \frac{\log \beta_1 + \dots + \log \beta_k}{k} = \lim_{k \rightarrow \infty} \frac{1}{k} \log \det \Gamma_{k+1} \quad (4.10)$$

4.3 ESTIMATION EMPIRIQUE DES PARAMETRES AR

4.3.1 Pour ajuster un modèle AR à une réalisation finie $\{x_1, \dots, x_N\}$ d'un processus régulier on peut d'abord partir du système (6) et remplacer brutalement les $\Gamma(k)$ par leur estimation $\hat{\Gamma}(k) = \frac{1}{N-k} \sum_{j=k+1}^N x_j x_{j-k}$ à partir des données disponibles. Celles-ci étant en nombre fini N , on doit tronquer ce système à un ordre p ce qui revient à postuler que $a_{j \geq p+1} = 0$ et le problème devient celui de l'estimation d'un modèle d'ordre fini AR- p (dans la pratique on aura intérêt à fixer p petit devant N). En posant $\mathbf{a}^T = [a_1, \dots, a_p]^T$, $\hat{\mathbf{g}}_p^T = [\hat{\Gamma}(1), \dots, \hat{\Gamma}(p)]$ et $\hat{\Gamma}_p = [\hat{\Gamma}(i-j)]_{0 \leq i, j \leq p-1}$ on aboutit au système $\hat{\Gamma}_p \mathbf{a} = \hat{\mathbf{g}}_p$ dont la solution $\mathbf{a} = \hat{\mathbf{a}}_p$ est l'estimateur recherché.

Une autre méthode pour arriver à un résultat similaire consiste à partir de la définition des a_k comme les coefficients de la régression de X_n sur

$\mathbf{X}_{n-1}^T = [X_{n-1}, \dots, X_{n-p}]$. La version 'empirique' de cette régression est le vecteur $\mathbf{a} = \hat{\mathbf{a}}_p$ qui minimise le critère quadratique :

$$\varepsilon(\mathbf{a}) = \frac{1}{N-p} \sum_{n=p+1}^N |x_n - \mathbf{x}_{n-1}^T \mathbf{a}|^2 \quad (4.11)$$

où $\mathbf{x}_{n-1}^T = [x_{n-1}, \dots, x_{n-p}]$.

3.3.2 EXERCICE Montrer que $\mathbf{a} = \hat{\mathbf{a}}_p$ minimisant $\varepsilon(\mathbf{a})$ est solution de l'équation

$$\mathbf{R}_p \mathbf{a} = \mathbf{s}_p \quad (4.12)$$

avec $\mathbf{R}_p = \frac{1}{N-p} \sum_{n=p+1}^N \mathbf{x}_{n-1} \mathbf{x}_{n-1}^T$ et $\mathbf{s}_p = \frac{1}{N-p} \sum_{n=p+1}^N x_n \mathbf{x}_{n-1}$. Comparer les termes de cette équation pour \mathbf{a} avec le précédent. Montrer que $\varepsilon_{\min} = \varepsilon(\hat{\mathbf{a}}_p) = \frac{1}{N-p} \sum_{n=p+1}^N |x_n|^2 - \mathbf{s}_p^T \mathbf{R}_p^{-1} \mathbf{s}_p$. En revenant à la formule de l'erreur d'estimation (cf. 2.1.1) montrer que $\hat{\beta}^* = \varepsilon_{\min}$ est une 'approximation raisonnable' de β^* .

4.3.3 SIMULATION On testera l'algorithme de l'exercice 3.3.2 sur une réalisation d'un processus autorégressif d'ordre $m \geq 2$. Pour cela il faudra générer N échantillons x_1, \dots, x_N du processus $X_n = a_1 X_{n-1} + \dots + a_m X_{n-m} + W_n$, $m \geq 1$ avec $X_{-1} = \dots = X_{-m} = 0$, W_n bruit blanc $\mathcal{N}(0, \beta)$. Pour assurer la stabilité et la stationnarité asymptotiques les coefficients a_1, \dots, a_m doivent être choisis de sorte que le polynôme réel $P(z) = z^m - a_1 z^{m-1} - \dots - a_m$ ait toutes ses racines incluses dans l'intérieur du disque unité de \mathbb{C} . Pour cela on spécifiera d'abord une factorisation $P(z) = (z - z_1) \dots (z - z_m)$ avec $|z_i| < 1$ et en choisissant les racines complexes par paires conjuguées. On en déduira a_1, \dots, a_m . Ayant obtenu x_1, \dots, x_N , programmer l'algorithme des moindres carrés en choisissant un ordre $p \geq m$. On comparera les coefficients $\hat{a}_1, \dots, \hat{a}_p$ à a_1, \dots, a_m et $\hat{\beta}^* = \varepsilon_{\min}$ à β en étudiant notamment l'influence de p et N . Finalement on réalisera un tracé de la densité spectrale estimée à l'aide de la formule $\hat{\gamma}(\lambda) = \frac{\hat{\beta}^*}{|1 - \sum_{k=1}^p \hat{a}_k e^{-i2\pi k\lambda}|^2}$; $0 \leq \lambda < \frac{1}{2}$ qu'on comparera au tracé du spectre exact calculé avec β et les coefficients a_1, \dots, a_m .

Dans une 2ème partie on testera cet algorithme sur le signal $X_n = U_n + \sin(2\pi\lambda_1 n + \varphi_1) + \sin(2\pi\lambda_2 n + \varphi_2)$ avec $0 < \lambda_1, \lambda_2 < \frac{1}{2}$; φ_1, φ_2 , U_n aléatoires indépendants, les phases φ_1 et φ_2 sont équiréparties sur $[0, 2\pi[$ et le bruit blanc U_n est $\mathcal{N}(0, \sigma^2)$. Ce processus n'est évidemment pas régulier du fait de la composante singulière périodique. Néanmoins on peut tester la capacité de l'algorithme 3.3.2 à en fournir une approximation spectrale. Pour

cela on ajustera à une réalisation x_1, \dots, x_N de ce signal un modèle AR- p et on étudiera sur l'estimation du spectre $\hat{\gamma}(\lambda)$ la précision de la localisation des fréquences des sinusoides en fonction de l'ordre p choisi, de la variance du bruit et du nombre de données N .

3.4 ESTIMATION DES PARAMETRES MA

4.4.1 Dans la méthode classique (Box & Jenkins) on part du système d'équations (4) tronqué à un ordre m postulé et où l'on remplace la fonction de corrélation $\Gamma(k)$ par son estimation $\hat{\Gamma}(k)$ (obtenue à partir de x_1, \dots, x_N comme en 3.3.1). On aboutit alors à chercher numériquement (p.ex. avec l'algorithme itératif de Newton) la solution en β^* et en (b_j) de $\hat{\Gamma}(k) = \beta^* \sum_{j=k}^m b_j b_{j-k}$, $k = 0, \dots, m$. Le polynôme $P(z) = 1 + \hat{b}_1 z^{-1} + \dots + \hat{b}_m z^{-m}$ doit encore satisfaire à la contrainte d'avoir ses zéros à l'extérieur du cercle unité ce qui implique un autre calcul non linéaire sur les racines.

La méthode suivante est a priori plus simple. On commence par ajuster avec l'algorithme de 3.3.2 un modèle AR- p aux données. Ensuite on applique la relation (6) avec $A(z) = \hat{a}_1 z^{-1} + \dots + \hat{a}_p z^{-p}$. En développant le terme de droite en série entière on a :

$$1 + \hat{b}_1 z^{-1} + \hat{b}_2 z^{-2} + \dots = 1 + A(z) + A^2(z) + \dots$$

Finalement on obtient les coefficients estimés \hat{b}_k du modèle MA à un ordre m en égalisant les coefficients des z^{-k} de part et d'autre de cette relation et en retenant les m premiers termes, ce qui donne : $\hat{b}_1 = \hat{a}_1$, $\hat{b}_2 = \hat{a}_1^2 + \hat{a}_2$, $\hat{b}_3 = \hat{a}_1^3 + 2\hat{a}_1\hat{a}_2 + \hat{a}_3$ etc.

4.4.2 SIMULATION On étudiera l'application de la méthode décrite sur l'exemple suivant. On génère N échantillons de la séquence aléatoire $X_n = W_n + b_1 W_{n-1} + \dots + b_m W_{n-m}$ où (W_n) est un bruit blanc $\mathcal{N}(0, 1)$ et $b_k = 1 - \frac{k}{m+1}$ pour $k = 1, \dots, m$. Appliquer la technique décrite ci-dessus en ajustant aux données un modèle AR- p et en augmentant progressivement $p \geq 1$. Etudier simultanément l'influence de N sur la précision de l'estimation des b_k .

REMARQUES Une adaptation facile des méthodes décrites ci-dessus permet d'estimer les paramètres d'un modèle ARMA qui est une equation composite du type $X_n = a_1 X_{n-1} + \dots + a_p X_{n-p} + W_n + b_1 W_{n-1} + \dots + b_m W_{n-m}$. Dans le cas régulier ces modèles décrivent tous les processus à spectre rationnel. Mais le concept fondamental reste la prédiction linéaire. On

a éludé partout la question essentielle du choix de l'ordre du modèle. En fait les méthodes décrites en 3.3 et 3.4 servent d'étape d'initialisation pour des procédures itératives plus précises et beaucoup plus complexes pour estimer simultanément l'ordre et les paramètres (le maximum de vraisemblance et ses avatars).