

Cours Signal Aléatoire

Polytech Annecy-Chambéry
IAI - Semestre 7 - EASI 742

Guillaume GINOLHAC

2017-18

Table des matières

1	Signaux aléatoires	3
1.1	Variables aléatoires	3
1.1.1	Définitions	3
1.1.2	Probabilité d'un événement	3
1.1.3	Cas des variables aléatoires discrètes	4
1.1.4	Cas des variables aléatoires continues	4
1.1.5	Exemples de densités de probabilités	4
1.1.6	Espérance mathématique	5
1.2	Définition d'un signal aléatoire	6
1.3	Propriétés statistiques d'un signal aléatoire	7
1.3.1	Lois	7
1.3.2	Indépendance	7
1.3.3	Stationnarité	7
1.3.4	Valeur moyenne et fonctions de corrélation d'un signal aléatoire	9
1.4	Propriétés temporelles d'un signal aléatoire : ergodisme	12
1.4.1	Ergodisme au premier ordre	12
1.4.2	Ergodisme au deuxième ordre	12
1.4.3	Exemples	12
1.5	Représentation fréquentielle d'un signal aléatoire stationnaire au sens large	13
1.5.1	Densité spectrale de puissance d'un signal aléatoire centré stationnaire	13
1.5.2	Propriétés de la densité spectrale de puissance	16
1.5.3	Largeur de bande	17
1.6	Opération linéaire sur signaux aléatoires	17
1.6.1	Filtrage d'un signal aléatoire stationnaire au sens large : analyse temporelle	18
1.6.2	Filtrage d'un signal aléatoire stationnaire au sens large : analyse fréquentielle	18
1.6.3	Quelques conséquences immédiates	18
1.6.4	Formule de cohérence et des interférences	20
2	Estimation	23
2.1	Définitions générales relatives à l'estimation	23
2.2	Estimation de la fonction d'autocorrélation	25
2.2.1	Estimateur non biaisé	25
2.2.2	Estimateur biaisé	26
2.3	Estimation de la DSP	27
2.3.1	Principe de l'analyse spectrale	27
2.3.2	Mise en oeuvre pratique : méthode de Welch	28
2.3.3	Qualité de l'estimation spectrale par la méthode de Welch	30

3	Filtrage Adaptatif	33
3.1	Rappels mathématiques	33
3.1.1	Algèbre matricielle	33
3.1.2	Gradient fonction à plusieurs variables	34
3.1.3	Optimisation	35
3.2	Introduction	35
3.2.1	Débruitage	35
3.2.2	Déconvolution	36
3.3	Filtre de Wiener	38
3.4	Moindres carrés exacts et pondérés	39
3.5	Moindres carrés récursifs (RLS)	41
3.6	Filtrage adaptatif par algorithme du gradient (LMS)	42

Chapitre 1

Signaux aléatoires

Les fonctions déterministes étudiées jusqu'ici permettent de décrire les signaux certains dont la forme est parfaitement connue. C'est le cas par exemple pour un signal émis en radar ou en sonar actif dont la forme est choisie pour obtenir certaines propriétés (codes permettant d'améliorer la résolution des mesures de distance, de Doppler. . .). Mais la plupart des signaux proviennent de capteurs qui mesurent des niveaux avec une certaine incertitude qui peut être liée :

- au capteur lui même : bruit de mesure, bruit "électronique". . . La tension délivrée par un micro n'est par exemple jamais vraiment nulle, même en l'absence de source sonore.
- à des variations de la grandeur physique mesurée dues à des phénomènes imprévisibles : bruit ambiant, sources parasites. . . Un micro utilisé en extérieur enregistre inévitablement un ensemble de variations de pression autre que le signal recherché : bruit de fond, vent, sources fortes dans des directions insuffisamment atténuées. . .
- au manque de connaissance sur le signal recherché : en sonar passif, le "bruit" rayonné par les bâtiments permet de les détecter bien que la forme du signal soit complètement inconnue.

Pour décrire ce type de signaux, on utilise les fonctions aléatoires. Nous étendons ici aux signaux aléatoires les notions de corrélation, de transformée de Fourier et de filtrage développées dans le cours précédent.

Mais on commence par quelques rappels sur les variables aléatoires.

1.1 Variables aléatoires

1.1.1 Définitions

Notations et terminologie :

- X : variable aléatoire discrète (nT_e) ou continue (t)
- x : événement associé à une variable aléatoire
- Ω : espace des événements
- A et B : événements disjoints

Exemples assez connus : dé, lancer de fléchette, . . .

1.1.2 Probabilité d'un événement

Définition de la probabilité d'un événement :

$$P = \lim_{N \rightarrow +\infty} \frac{\text{Nombre de cas favorables}}{N} \quad (1.1)$$

Propriétés : additivité de P , $P(\Omega) = 1$, $0 \leq P(A) \leq 1$

1.1.3 Cas des variables aléatoires discrètes

Définition de la probabilité d'un événement pour le cas discret :

$$P(X = x_i) = \lim_{N \rightarrow +\infty} \frac{N_i}{N} \quad (1.2)$$

Propriétés :

$$- P(X \in \{x_1, \dots, x_L\}) = \sum_{i=1}^L P(X = x_i)$$

$$- \sum_{i=1}^L P(X = x_i) = 1$$

1.1.4 Cas des variables aléatoires continues

Définition de la probabilité d'un événement pour le cas continu :

$$P(x < X < x + dx) = \lim_{N \rightarrow +\infty} \frac{dN_x}{N} = f(x)dx, \quad (1.3)$$

avec $f(x)$ densité de probabilité (ddp).

Propriétés :

$$- f(x) \geq 0 \text{ pour tout } x$$

$$- P(a < X < b) :$$

$$\begin{aligned} P(a < X < b) &= \lim_{\Delta x \rightarrow 0} P[(x_1 = a < X < x_1 + \Delta x) \cup \dots \cup (x_I < X < x_I + \Delta x = b)] \\ &= \lim_{\Delta x \rightarrow 0} \left(\sum_{i=1}^I P[x_i < X < x_{i-1} + \Delta x] \right) \\ &= \lim_{\Delta x \rightarrow 0} \left(\sum_{i=1}^I f(x_i) \Delta x \right) \\ &= \int_a^b f(x) dx \end{aligned} \quad (1.4)$$

$$- \int_{-\infty}^{+\infty} f(x) dx = 1$$

1.1.5 Exemples de densités de probabilités

Sur la figure 1.1(a), on voit la ddp d'une v.a. uniforme entre a et b :

$$f(x) = \Pi_{[a,b]} \cdot \frac{1}{b-a} \quad (1.5)$$

Sur la figure 1.1(b), on voit la ddp d'une v.a. gaussienne de moyenne m et d'écart type σ :

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{1}{2} \left(\frac{x-m}{\sigma}\right)^2\right) \quad (1.6)$$

L'intérêt de la v.a. aléatoire gaussienne est double : elle représente très bien de nombreux phénomènes de la vie courante (naturelle, électronique, ...) et sa forme mathématique permet de nombreux calculs.

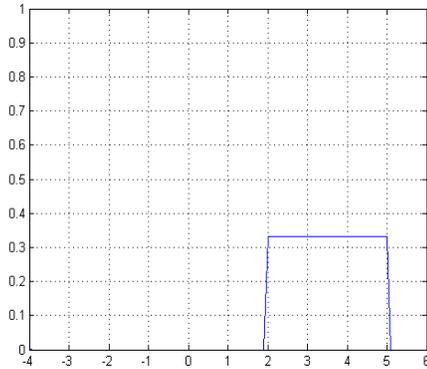
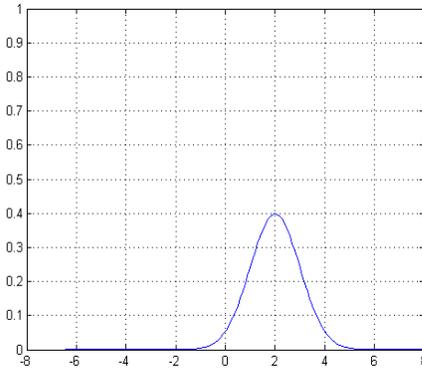
(a) ddp loi uniforme avec $a = 2$ et $b = 5$ (b) ddp loi gaussienne de moyenne $m = 2$ et d'écart type $\sigma = 1$. Le maximum est $\frac{1}{\sigma\sqrt{2\pi}}$. La largeur du pic est égal à la moitié du maximum à $2\sqrt{2\ln(2)}\sigma$

FIGURE 1.1 – ddp de deux lois très connues.

1.1.6 Espérance mathématique

Définition

Soit x des réalisations d'une variable aléatoire X (de ddp $f(x)$ en continu et prenant les valeurs $\{x_i\}_{1,N}$ en discret), alors on définit l'espérance de X comme :

$$\begin{aligned} E[X] &= \int x f(x) dx \\ E[X] &= \sum_{i=1}^N p(X = x_i) x_i \end{aligned} \quad (1.7)$$

Nous avons aussi la propriété suivante :

$$\begin{aligned} E[g(X)] &= \int g(x) f(x) dx \\ E[g(X)] &= \sum_{i=1}^N p(X = x_i) g(x_i) \end{aligned} \quad (1.8)$$

Propriétés fondamentales de l'espérance

- Soit λ une constante, alors $E[\lambda] = \lambda$
- Soit λ une constante et X une v.a., alors $E[\lambda X] = \lambda E[X]$
- Linéarité : soit X et Y deux v.a., alors $E[X + Y] = E[X] + E[Y]$

Moments

Soit X , une v.a., nous pouvons calculer ces différentes grandeurs :

- ordre 1 : $E[X] = m_X$
- ordre 2 : $E[X^2]$
- Variance : $E[(X - m_X)^2] = E[X^2] + m_X^2 - 2E[X]m_X = E[X^2] - m_X^2 = \sigma_X^2$

Remarques :

- la variance mesure la dispersion des résultats autour de la valeur moyenne. Analogue à une mesure de puissance.
- une v.a. gaussienne est entièrement caractérisée par m_X et σ_X

Exemples : 3 calculs d'espérance

- Soit X une v.a. discrète prenant comme valeur 0, 1 et -1 avec comme probabilité $p_1 = 0.5, p_2 = 0.25$ et $p_3 = 0.25$. Calculez $E[X]$ et $E[X^2]$.
- Soit Φ une v.a. continue uniforme sur $[0, 2\pi]$. Calculez $E[Y]$ et $E[Y^2]$ avec $Y = \sin(2 * \pi * f * t + \Phi)$.
- Soit X une v.a. continue gaussienne de moyenne m et de variance σ^2 . Calculez la moyenne et la variance de $Y = \frac{X-m}{\sigma}$.

Corrélation d'une v.a.

Soit X et Y deux v.a. de moyenne m_X et m_Y (on définit aussi $X_c = X - m_X$ et $Y_c = Y - m_Y$), on définit différentes grandeurs :

- corrélation : $\Gamma_{XY} = E[XY]$
- covariance : $E[X_c Y_c] = E[(X - m_X)(Y - m_Y)] = \Gamma_{XY} - m_X m_Y$
- coefficient de corrélation : $\rho = \frac{E[X_c Y_c]}{\sigma_X \sigma_Y}$. On peut montrer que (Cf TD n° 1) :

$$-1 \leq \rho \leq 1 \tag{1.9}$$

Plus $|\rho|$ est grand, plus X et Y est se ressemblent (plus exactement Y peut s'approximer par une transformation linéaire de X). Mais attention corrélation ne signifie pas causalité.

Variables décorrélées : Soit X et Y deux v.a., on dit qu'elles sont décorrélées quand $\Gamma_{XY} = E[XY] = E[X]E[Y]$, ce qui équivaut à $E[X_c Y_c] = 0$.

1.2 Définition d'un signal aléatoire

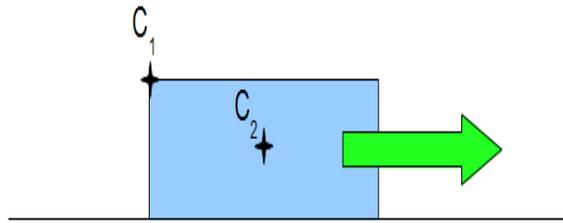
Concept d'un signal aléatoire $X(t)$: c'est un ensemble de variables aléatoires paramétrées par le temps t . Deux cas sont possibles :

- t varie continuellement : $X(t)$ v.a. à temps continu
- t est discret et ne prend que certaines valeurs : $X(t)$ v.a. à temps discret. On notera $X(n)$. Obtenus à partir d'un signal aléatoire continu $X(n) = X(nT_e)$

Exemples : on enregistre par exemple à l'aide d'accéléromètres les vibrations d'un objet en déplacement. Les signaux obtenus sont fonctions de plusieurs paramètres tels que la position du capteur, les matériaux, la vitesse de déplacement... En réalisant des mesures plusieurs fois dans les mêmes conditions expérimentales, on obtient plusieurs réalisations pour chacun des signaux aléatoires.

La figure 1.2 illustre deux signaux aléatoires $X_1(t)$ et $X_2(t)$ mesurés respectivement à une extrémité de l'objet (capteur 1) et au centre de l'objet (capteur 2) comme montré dans la figure 1.2. Le tracé de la réalisation 1 du capteur 1, $x_1(t)$, montre un signal irrégulier, sans forme ou périodicité apparente. Le tracé de la seconde réalisation, $x_1(t)$, présente la même allure d'ensemble mais ne se superpose absolument pas avec le premier tracé. Les réalisations du signal issus du capteur 2, $x_2(t)$ et $x_2(t)$, présentent également une certaine ressemblance entre elles alors qu'elles sont très différentes de celles issues du capteur 1.

La ressemblance entre les tracés obtenus dans les mêmes conditions expérimentales est due à des grandeurs caractéristiques des processus aléatoires (variance, corrélation, densité spectrale...). Ces grandeurs sont elles-mêmes liées aux paramètres physiques (tension, rugosité...) que l'on souhaite étudier. Les méthodes développées en traitement du signal permettent d'extraire des signaux aléatoires des informations très diverses pour des applications aussi variées que la détection et la classification de bateaux en acoustique sous-marine ou le réglage de la tension de courroies de transmission dans l'automobile.



1.3 Propriétés statistiques d'un signal aléatoire

1.3.1 Lois

Loi à l'ordre un

On connaît la loi d'ordre un d'un signal aléatoire $X(t)$ si on connaît la loi (ou sa ddp) de la v.a. $X(t_1)$ pour toutes les valeurs de t_1 .

Loi d'ordre N

On connaît la loi d'ordre N d'un signal aléatoire $X(t)$ si on connaît la ddp $f(X(t_1), \dots, X(t_N))$ jointe des N v.a. $X(t_1), \dots, X(t_N)$ pour tous les instants de t_1, \dots, t_N .

Loi de $X(t)$

On connaît la loi de $X(t)$ si on connaît sa loi à tous les ordres

1.3.2 Indépendance

Définition : Deux signaux aléatoires $X(t)$ et $Y(t)$ sont indépendants ssi les ensembles des variables aléatoires $(X(t_1), \dots, X(t_M))$ et $(Y(t'_1), \dots, Y(t'_N))$ sont indépendants $\forall t_1 \dots t_M, \forall t'_1, \dots, t'_N, \forall M, \forall N$. Dans ce cas on a : $f(X(t_1), \dots, X(t_M), Y(t'_1), \dots, Y(t'_N)) = f(X(t_1), \dots, X(t_M)) \cdot f(Y(t'_1), \dots, Y(t'_N))$.

Interprétation : Quand mécanismes liés à la génération de $X(t)$ et $Y(t)$ sont différents, les signaux aléatoires associés sont souvent indépendants. Par exemple, un micro enregistre le son d'une voix ainsi que le bruit du vent. Ces deux phénomènes sont distincts et donc entraînent des s.a. indépendants.

1.3.3 Stationnarité

Stationnarité du premier ordre

Un processus $X(t)$ est stationnaire au premier ordre quand sa moyenne statistique (son espérance) est indépendante du temps :

$$E[X(t)] = E(X) \quad \forall t \quad (1.10)$$

Exemple : on considère deux variables aléatoires indépendantes ω de d.d.p $f_\omega(\omega)$ et ϕ qui suit une loi

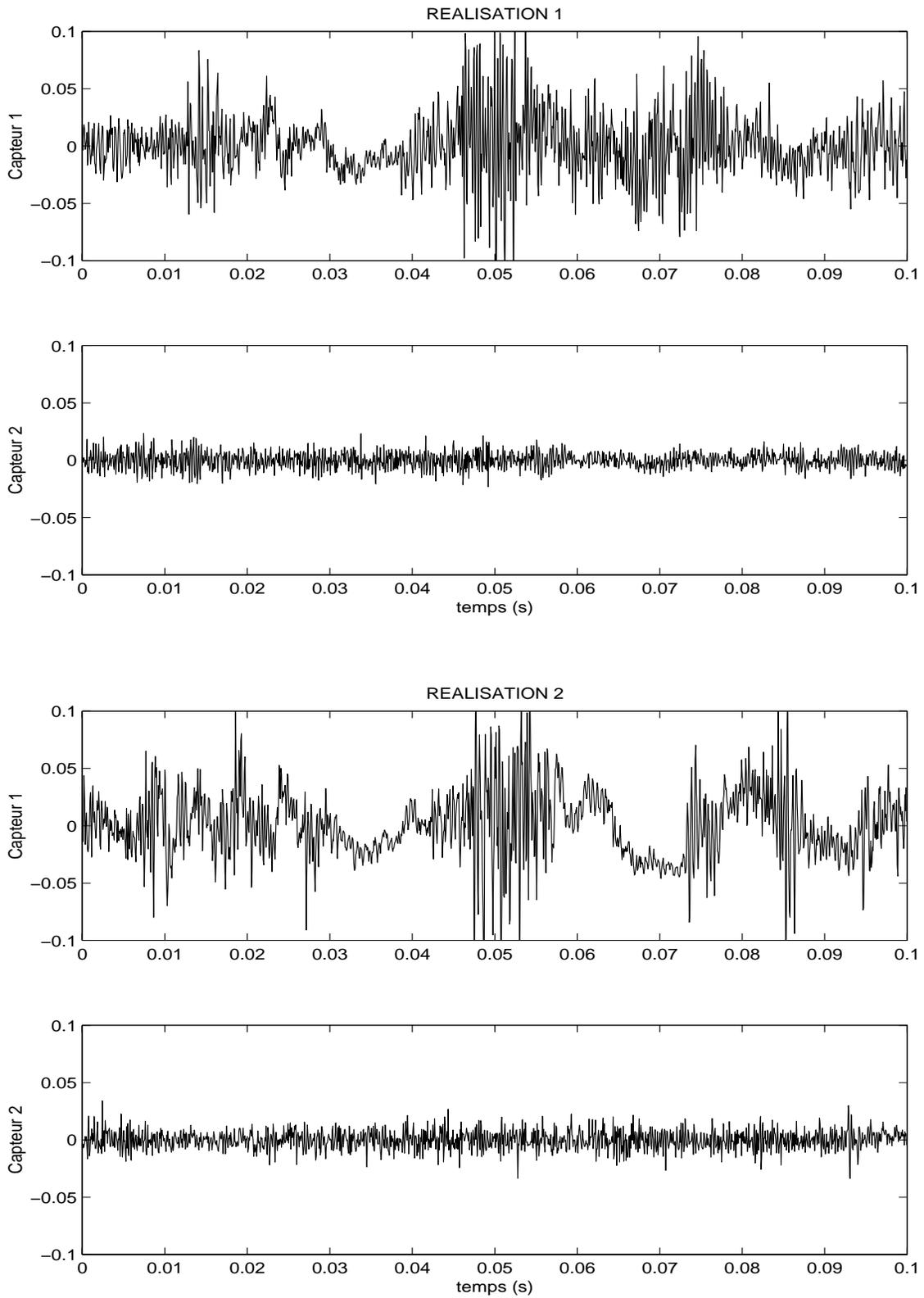


FIGURE 1.2 – Différentes réalisations de 2 fonctions aléatoires

uniforme sur $[0, 2\pi]$ et la fonction aléatoire $X(t) = \cos(\omega t + \phi)$. L'espérance à un instant t vaut :

$$\begin{aligned} E_{(\omega, \phi)}[X(t)] &= \int_{-\infty}^{+\infty} \int_0^{2\pi} \frac{1}{2\pi} f_{\omega}(\omega) \cos(\omega t + \phi) d\phi d\omega \\ &= \int_{-\infty}^{+\infty} \frac{f_{\omega}(\omega)}{2\pi} \left[\int_0^{2\pi} \cos(\omega t + \phi) d\phi \right] d\omega \\ &= 0 \end{aligned}$$

La moyenne est indépendante du temps, le processus est stationnaire d'ordre 1.

Stationnarité du second ordre

Un processus $X(t)$ est stationnaire au second ordre quand :

- il est stationnaire au premier ordre,
- sa covariance $E[X(t_1)X(t_2)]$ ne dépend que de l'écart entre les deux instants $\tau = t_2 - t_1$:

$$E[X(t_1)X(t_2)] = g(t_2 - t_1) = g(\tau) \quad \forall t_1, t_2 \quad (1.11)$$

où $g(\tau)$ est une fonction paire ($g(-\tau) = g(\tau)$).

Exemple : on vérifie que le processus précédent est également stationnaire au second ordre :

$$\begin{aligned} E_{(\omega, \phi)}[X(t_1)X(t_2)] &= \int_{-\infty}^{+\infty} \int_0^{2\pi} \frac{1}{2\pi} f_{\omega}(\omega) \cos(\omega t_1 + \phi) \cos(\omega t_2 + \phi) d\phi d\omega \\ &= \int_{-\infty}^{+\infty} \frac{f_{\omega}(\omega)}{2\pi} \frac{1}{2} \left[\int_0^{2\pi} \cos(\omega(t_2 - t_1)) + \cos(\omega(t_2 + t_1) + 2\phi) d\phi \right] d\omega \\ &= \int_{-\infty}^{+\infty} \frac{1}{2} \cos(\omega(t_2 - t_1)) f_{\omega}(\omega) d\omega \\ &= g(t_2 - t_1) \end{aligned}$$

avec $\int_0^{+2\pi} \cos(\omega(t_2 + t_1) + 2\phi) d\phi = 0$ et $\cos a \cos b = \frac{\cos(a-b) + \cos(a+b)}{2}$.

Stationnarité d'ordre N

Un processus $X(t)$ est stationnaire à l'ordre N si les lois jointes de $(X(t_1), \dots, X(t_N))$ et de $(X(t_1 + \tau), \dots, X(t_N + \tau))$ sont identiques pour tous les instants t_1, \dots, t_N et τ .

Lorsque un processus est stationnaire à tous les ordres, on dit qu'il est stationnaire au sens strict.

Conséquences : la densité de probabilité de la variable aléatoire $X(t)$ et tous ses moments (espérance, variance...) sont indépendants du temps.

Interprétation : un signal aléatoire stationnaire présente une certaine homogénéité, comme par exemple, ce signal stationnaire présentée dans la figure 1.3(a). Sur la figure 1.3(b), on peut voir un exemple de signal non stationnaire.

1.3.4 Valeur moyenne et fonctions de corrélation d'un signal aléatoire

Souvent, on n'a pas besoin de connaître la loi (de toute façon impossible d'estimer toutes les lois à l'ordre N) d'un signal aléatoire $X(t)$: sa moyenne et ses moments d'ordre 2 suffisent.

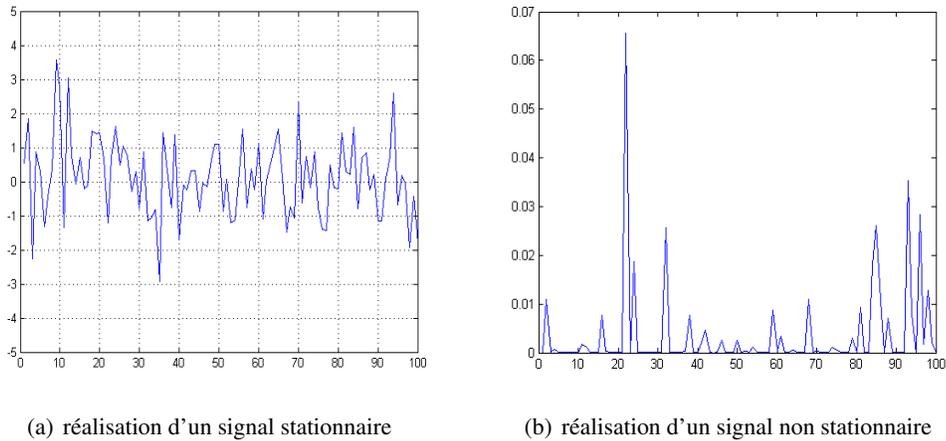


FIGURE 1.3 – Interprétation de la stationnarité à l'ordre 2.

Moyenne signal aléatoire

C'est le signal $m(t)$ défini pour tout t par :

$$m(t) = E[X(t)] \quad (1.12)$$

Fonctions d'autocorrélation et d'autocovariance

La fonction d'autocorrélation est définie par :

$$\gamma(t_1, t_2) = E[X(t_1)X(t_2)] \quad (1.13)$$

La fonction d'intercovariance est définie par :

$$r(t_1, t_2) = E[(X(t_1) - m_1)(X(t_2) - m_2)] = \gamma(t_1, t_2) - m(t_1)m(t_2) \quad (1.14)$$

Lorsqu'on a deux signaux aléatoires on peut définir des fonctions d'intercorrélations et d'intercovariance. La fonction d'intercorrélations est définie par :

$$\gamma_{XY}(t_1, t_2) = E[X(t_1)Y(t_2)] \quad (1.15)$$

La fonction d'intercovariance est définie par :

$$r_{XY}(t_1, t_2) = E[(X(t_1) - m_1)(Y(t_2) - m_2)^*] = \gamma_{XY}(t_1, t_2) - m_X(t_1)m_Y(t_2) \quad (1.16)$$

Remarque : attention dans ce cours, nous supposons que les signaux sont réels. Dans le cas de signaux complexes, les formules changent légèrement avec l'introduction d'un conjugué sur l'un des termes de l'espérance. On rencontre des signaux complexes quand ils sont démodulés comme en télécoms, Radar/Sonar, Imagerie SAR, ...

Stationnarité au sens large

La stationnarité au sens strict est très restrictive. En général, on a seulement besoin de la stationnarité des moments d'ordre 1 et 2. On parle alors de stationnarité au sens large.

Définition : $X(t)$ est stationnaire au sens large si :

- $E[X(t)] = m$ avec m une constante
- $E[X(t)X^*(t - \tau)] = \gamma(\tau)$

On a alors aussi : $r(\tau) = E[(X(t) - m)(X(t - \tau) - m)^*] = \gamma(\tau) - m^2$

De même, $X(t)$ et $Y(t)$ sont stationnaires au sens large dans leur ensemble ssi ils sont stationnaires tous les deux et si $E[X(t)Y^*(t - \tau)] = \gamma_{XY}(\tau)$.

On supposera par la suite que les signaux sont stationnaires au sens large. Cette propriété permet de résoudre la majeure partie des problématiques.

Propriétés

Theorem 1.3.1 Soit $X(t)$ un signal aléatoire stationnaire au sens large, nous avons les propriétés suivantes pour la fonction d'autocorrélation (on a des propriétés identiques pour la fonction d'autocovariance) :

- P1 : $\gamma(0) = E(|X(t)|^2)$ (puissance du signal)
- P2 : $\gamma(-\tau) = \gamma(\tau)$ (symétrie hermitienne en complexe, paire en réel)
- P3 : $|\gamma(\tau)| \leq \gamma(0)$

PROOF. Cf TD. □

Somme de deux signaux aléatoires stationnaires décorrelés

On a souvent :

$$X(t) = S(t) + B(t), \quad (1.17)$$

avec $S(t)$ signal utile et $B(t)$ bruit. Ces deux signaux aléatoires sont souvent indépendants et donc décorrelés (attention l'inverse n'est pas toujours vraie). On a alors :

- $m_X = m_S + m_B$
- $\gamma_X(\tau) = \gamma_S(\tau) + \gamma_B(\tau)$

La preuve est évidente et peut être faite en exemple.

Recherche d'une sinusoïde

On recherche une sinusoïde de fréquence inconnue, noyée dans un bruit gaussien centré stationnaire de longueur de corrélation l_c . La longueur de corrélation d'une fonction aléatoire est le décalage temporel au delà duquel on peut considérer que les v.a. $B(t)$ et $B(t + \tau)$ sont décorrelées, soit :

$$\gamma_{bb}(\tau) = E[B(t).B(t - \tau)] = 0 \quad \text{pour } |\tau| > l_c$$

Le signal reçu s'écrit sous la forme :

$$x(t) = s(t) + b(t)$$

où $b(t)$ est une réalisation du bruit et $s(t) = A \cos(2\pi f_o t)$. On calcule la fonction de corrélation

$$\gamma_{xx}(\tau) = \gamma_{ss}(\tau) + \gamma_{bb}(\tau)$$

le terme $\gamma_{xb}(\tau)$ étant nul (hypothèse de décorrélation du signal et du bruit). En se plaçant à un retard $\tau > l_c$, on obtient sous l'hypothèse de stationnarité :

$$\gamma_{xx}(\tau) = \gamma_{ss}(\tau) = \frac{A^2}{2} \cos(2\pi f_o \tau) \quad \text{pour } |\tau| > l_c$$

ce qui permet de mesurer f_o directement sur la périodicité de l'autocorrélation.

1.4 Propriétés temporelles d'un signal aléatoire : ergodisme

Il est nécessaire d'estimer la moyenne et la fonction d'autocorrélation. Deux solutions possibles :

1. la première solution consisterait à effectuer des moyennes à partir d'un grand nombre de réalisations indépendantes du processus. On parle alors de moyennes d'ensemble. Cette méthode est généralement impossible car on ne dispose que d'une seule réalisation $x(t)$ du signal aléatoire $X(t)$.
2. mais quand le signal est stationnaire, on peut faire ces estimations à l'aide de moyennes temporelles.

Et si ces moyennes temporelles tendent vers les moyennes d'ensemble, on dit que le signal aléatoire est ergodique.

1.4.1 Ergodisme au premier ordre

Définition : Soit $X(t)$ un signal aléatoire et $x(t)$ une réalisation particulière. On note, sous réserve d'existence, m la moyenne temporelle de la réalisation, $X(t)$ est ergodique au premier ordre si :

$$\frac{1}{2T} \int_{-T}^T x(t) dt \xrightarrow{T \rightarrow +\infty} m = E[X(t)] \quad (1.18)$$

1.4.2 Ergodisme au deuxième ordre

Définition : Soit $X(t)$ un signal aléatoire de moyenne m et d'autocorrélation $\gamma(\tau)$. Soit $x(t)$ une réalisation particulière. $X(t)$ est ergodique au deuxième ordre si :

- elle est ergodique au premier ordre,
- quelque soit la réalisation $x(t)$, sa fonction d'autocorrélation au sens des signaux de puissance moyenne finie existe et est indépendante de la réalisation :

$$\frac{1}{2T} \int_{-T}^T x(t)x(t-\tau) dt \xrightarrow{T \rightarrow +\infty} \gamma = E[X(t)X(t-\tau)] \quad (1.19)$$

Nous admettrons que les signaux aléatoires considérés par la suite sont ergodiques au deuxième ordre. En effet, il est pratiquement impossible de vérifier cette propriété sur des signaux réels.

Remarque : Dans le cas d'un bruit gaussien supposé ergodique et stationnaire, on accède ainsi aux paramètres de la densité de probabilité du bruit

$$f_X(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$$

en estimant la moyenne μ et la variance (statistique) σ^2 par la moyenne et la variance temporelles sur la durée d'analyse.

1.4.3 Exemples

Stationnarité et ergodisme

Pour les fonctions aléatoires suivantes, on teste la stationnarité et l'ergodisme au second ordre :

- $X(t) = u(t) + A$, où $u(t)$ est une fonction certaine et A une variable aléatoire :

$$m = E[X(t)] = u(t) + E[A] \quad (1.20)$$

m dépend du temps donc non stationnaire.

$$\frac{1}{2T} \int_{-T}^T A dt \xrightarrow{T \rightarrow +\infty} A \quad (1.21)$$

et donc pas ergodique car dépend de la réalisation.

— $X(t) = u_o + A$, où u_o est une constante et A une v.a. :

$$m = E[X(t)] = u_o + E[A] \quad (1.22)$$

m ne dépend pas du temps donc stationnaire.

$$\frac{1}{2T} \int_{-T}^T A dt \xrightarrow{T \rightarrow +\infty} A \quad (1.23)$$

et donc pas ergodique car dépend de la réalisation.

— $X(t) = \exp(iZt)$ où Z est une v.a. :

$$m = E[X(t)] = \exp(iE[Z]t) \quad (1.24)$$

donc non stationnaire mais ergodique. ($E[X(t)]$ est la fonction caractéristique de la v.a. Z , elle est a priori fonction du temps et moyenne temporelle égale à 0 quand T tend vers l'infini ?).

— $X(t) = \cos(\omega t + \phi)$, où ω est une v.a. de d.d.p. $f_\omega(\omega)$, ϕ une v.a. uniforme sur $[0, 2\pi]$

\Rightarrow stationnaire (cf. section 1.3.3), ergodique au premier ordre ($M = 0$) mais pas au second ordre :

$$\frac{1}{2T} \int_{-T}^T x(t)x(t-\tau) dt = \frac{\cos(\omega\tau)}{2} + \frac{1}{2T} \int_{-T}^T \cos(2\omega t + \phi - \omega\tau) dt \quad (1.25)$$

Le second terme s'annule en faisant tendre T vers l'infini. Comme le second dépend de la réalisation, pas d'ergodisme.

1.5 Représentation fréquentielle d'un signal aléatoire stationnaire au sens large

Nous nous intéressons dans cette section à la question : quelle bande de fréquence occupe un signal aléatoire. Cette question est fondamentale notamment en télécommunications : le signal émis est aléatoire et on a besoin de savoir la bande de fréquence de la modulation pour éviter des interférences avec d'autres émissions.

Remarque préliminaire : comme l'espérance est un opérateur linéaire, nous avons les propriétés suivantes :

$$\begin{aligned} E \left[\int_a^b X(t) dt \right] &= \int_a^b E[X(t)] dt \\ E \left[\int_a^b X(t)X(t-\tau) dt \right] &= \int_a^b E[X(t)X(t-\tau)] dt \end{aligned} \quad (1.26)$$

1.5.1 Densité spectrale de puissance d'un signal aléatoire centré stationnaire

Soit $X(t)$ un signal aléatoire centré stationnaire au sens large, ergodique à l'ordre 2, de fonction d'autocorrélation $\gamma(\tau)$ et de puissance $P = E[X(t)^2]$. Nous désirons analyser le spectre des réalisations de $X(t)$. Posons $S_T(f)$, la Transformée de Fourier (TF) finie de $X(t)$ sur un intervalle $[-T, T]$:

$$S_T(f) = \int_{-T}^T X(t) \exp(-2i\pi ft) dt \quad (1.27)$$

On remarque que $S_T(f)$ est aléatoire car $X(t)$ l'est. $S_T(f)$ dépend de la réalisation particulière de $X(t)$. La relation de Parseval nous donne :

$$\int_{-T}^T X(t)^2 dt = \int_{-\infty}^{+\infty} |S_T(f)|^2 df \quad (1.28)$$

Cela nous montre que l'énergie du signal $X(t)$ se distribue sur l'axe des fréquences. Néanmoins, elle dépend encore de la réalisation de $X(t)$. Il nous faut donc définir une énergie moyenne. Ensuite pour pouvoir calculer l'énergie au delà de l'intervalle $[-T, T]$, il faudrait faire tendre T vers l'infini. Mais cette opération n'est pas possible car :

$$\int_{-T}^T X(t)^2 dt \xrightarrow{T \rightarrow +\infty} \infty \quad (1.29)$$

Cette équation montre que les réalisations ne sont pas d'énergie finie : il faut donc étudier comment se distribue la puissance du signal en fonction de la fréquence. On réécrit la relation de Parseval sous la forme :

$$\frac{1}{2T} \int_{-T}^T X(t)^2 dt = \frac{1}{2T} \int_{-\infty}^{+\infty} |S_T(f)|^2 df \quad (1.30)$$

Prenons l'espérance des deux membres de l'équation, nous obtenons en rappelant que $P = E[X(t)^2]$:

$$\frac{1}{2T} \int_{-T}^T E[X(t)^2] dt = \int_{-\infty}^{+\infty} \frac{E[|S_T(f)|^2]}{2T} df \quad (1.31)$$

On fait tendre T vers l'infini :

$$P = \lim_{T \rightarrow +\infty} \int_{-\infty}^{+\infty} \frac{E[|S_T(f)|^2]}{2T} df \quad (1.32)$$

Le premier terme est égal à P par ergodisme. On appellera la limite de l'intégrande la Densité Spectral de Puissance (DSP) :

$$\lim_{T \rightarrow +\infty} \left(\frac{E[|S_T(f)|^2]}{2T} \right) = \Gamma(f) \quad (1.33)$$

Et l'on a donc pour la puissance une nouvelle expression :

$$P = \int_{-\infty}^{+\infty} \Gamma(f) df \quad (1.34)$$

On note que la DSP permet de mesurer la distribution de la puissance selon l'axe des fréquences.

Nous donnons la définition de la DSP pour les cas continu et discret :

Definition 1.5.1 (Définition de la DSP) — *Cas continu : soit $X(t)$ un signal aléatoire centré stationnaire au sens large, ergodique à l'ordre 2. On appelle Densité Spectral de Puissance (DSP) $\Gamma(f)$:*

$$E \left[\frac{|S_T(f)|^2}{2T} \right] \xrightarrow{T \rightarrow +\infty} \Gamma(f) \quad (1.35)$$

avec $S_T(f) = \int_{-T}^T X(t) \exp(-2i\pi ft) dt$ la TF finie de $X(t)$

— *Cas discret : soit $X(n)$ un signal aléatoire centré stationnaire au sens large, ergodique à l'ordre 2. On appelle Densité Spectral de Puissance (DSP) $\Gamma(f)$:*

$$E \left[\frac{|S_N(f)|^2}{2(N+1)} \right] \xrightarrow{N \rightarrow +\infty} \Gamma(f) \quad (1.36)$$

avec $S_N(f) = \sum_{n=-N}^N X(n) \exp(-2i\pi fn) dt$ la TF finie de $X(n)$

Le théorème suivant nous donne la relation entre DSP et fonction d'autocorrélation dans le cas continu et le cas discret :

Theorem 1.5.1 (Théorème de Wiener-Kintchine) *On considère les deux cas, continu et discret :*

— *Cas continu : soit $X(t)$ un signal aléatoire centré stationnaire au sens large, ergodique à l'ordre 2 et de fonction d'autocorrélation $\gamma(\tau)$. On a la relation suivante entre la Densité Spectral de Puissance (DSP) $\Gamma(f)$ et la transformée de Fourier de $\gamma(\tau)$:*

$$\Gamma(f) = \int_{-\infty}^{+\infty} \gamma(\tau) \exp(-2i\pi f\tau) d\tau = TF(\gamma(\tau)) \quad (1.37)$$

— *Cas discret : soit $X(n)$ un signal aléatoire centré stationnaire au sens large, ergodique à l'ordre 2 et de fonction d'autocorrélation $\gamma(m)$. On a la relation suivante entre la Densité Spectral de Puissance (DSP) $\Gamma(f)$ et la transformée de Fourier de $\gamma(m)$:*

$$\Gamma(f) = \sum_{m=-\infty}^{+\infty} \gamma(m) \exp(-2i\pi fm) = TFD(\gamma(m)) \quad (1.38)$$

Nous ne ferons que la démonstration dans le cas continu.

PROOF. A partir de la TF finie de $X(t)$, on a :

$$|S_T(f)|^2 = S_T(f) \cdot S_T(f)^* = \int_{-T}^T \int_{-T}^T X(t)X(t') \exp(-2i\pi f(t-t')) dt dt' \quad (1.39)$$

En prenant l'espérance des deux membres et en notant que $E[X(t)X(t')] = \gamma(t-t')$, nous avons :

$$E[|S_T(f)|^2] = \int_{-T}^T \int_{-T}^T \gamma(t-t') \exp(-2i\pi f(t-t')) dt dt' \quad (1.40)$$

Posons $\tau = t - t'$. τ varie entre $-2T$ et $2T$ et t varie à τ fixé entre $\max(-T, -T + \tau)$ et $\min(T, T + \tau)$ de façon à satisfaire les deux conditions $-T < t < T$ et $-T < t - \tau = t' < T$. Par conséquent :

$$\begin{aligned} E[|S_T(f)|^2] &= \int_{-2T}^{2T} \gamma(\tau) \exp(-2i\pi f\tau) \left(\int_{\max(-T, -T+\tau)}^{\min(T, T+\tau)} dt \right) d\tau \\ &= \int_{-2T}^{2T} (2T - |\tau|) \gamma(\tau) \exp(-2i\pi f\tau) d\tau \end{aligned} \quad (1.41)$$

Divisons les deux membres par $2T$ et faisons tendre T vers l'infini :

$$\lim_{T \rightarrow +\infty} \left(\frac{E[|S_T(f)|^2]}{2T} \right) = \Gamma(f) = \int_{-\infty}^{\infty} \gamma(\tau) \exp(-2i\pi f\tau) d\tau \quad (1.42)$$

car

$$\lim_{T \rightarrow +\infty} \int_{-\infty}^{\infty} \frac{(2T - |\tau|)}{2T} \gamma(\tau) \exp(-2i\pi f\tau) d\tau = \int_{-\infty}^{\infty} \gamma(\tau) \exp(-2i\pi f\tau) d\tau - \lim_{T \rightarrow +\infty} \int_{-\infty}^{\infty} \frac{|\tau|}{2T} \gamma(\tau) \exp(-2i\pi f\tau) d\tau \quad (1.43)$$

Le second terme est égal à 0. □

1.5.2 Propriétés de la densité spectrale de puissance

Nous avons le théorème suivant :

Theorem 1.5.2 Soit $\Gamma(f)$ la DSP d'un signal aléatoire $X(t)$ stationnaire, de puissance P . Nous avons les propriétés suivantes :

- P1. La DSP est positive : $\Gamma(f) \geq 0$ pour tout f
- P2. La DSP est paire : $\Gamma(f) = \Gamma(-f)$ pour tout f
- P3. $P = E[X(t)^2] = \int_{-\infty}^{\infty} \Gamma(f)df$ en continu et $P = E[X(n)^2] = \int_{-1/2}^{1/2} \Gamma(f)df$ en discret

PROOF. Comme $\Gamma(f)$ est la limite de quelque chose de positif, P1 est vérifié.

La DSP est la TF de l'autocorrélation qui est une fonction réelle paire. Par propriété de la TF, $\Gamma(f)$ est donc aussi paire.

P3 se déduit immédiatement du fait que la fonction autocorrélation est la TF inverse de $\Gamma(f)$ prise en $\tau = 0$:

$$P = \gamma(0) = \int_{-\infty}^{\infty} \Gamma(f) \exp(2i\pi f0)df \quad (1.44)$$

□

Exemple : DSP d'une somme de signaux centrés décorrelés

Considérons deux signaux aléatoires centrés décorrelés $S(t)$ et $B(t)$, de fonctions d'autocorrélation $\gamma_s(\tau)$ et $\gamma_b(\tau)$ et de DSP $\Gamma_b(f)$ et $\Gamma_s(f)$. Soit $X(t)$ la somme de ces deux signaux :

$$X(t) = S(t) + B(t) \quad (1.45)$$

Nous avons montré précédemment que $\gamma_x(\tau) = \gamma_s(\tau) + \gamma_b(\tau)$. En prenant la TF des deux membres de l'équation, on a :

$$\Gamma_x(f) = \Gamma_s(f) + \Gamma_b(f) \quad (1.46)$$

Remarque : échantillonnage de signaux aléatoires

Soit $X_a(t)$ un signal aléatoire continu et soit $X(n)$ le signal numérique obtenu par échantillonnage de $X_a(t)$ à la période $T_e = \frac{1}{F_e}$:

$$X(n) = X_a(nT_e) \quad (1.47)$$

Calculons maintenant les fonctions d'autocorrélation et de DSP de $X(m)$

Fonction autocorrélation : soit $\gamma_a(\tau)$ la fonction d'autocorrélation de $X_a(t)$ et notons $\gamma(m)$ la fonction d'autocorrélation de $X(n)$. On a :

$$\begin{aligned} \gamma(m) &= E[X(n)X(n-m)] \\ &= E[X_a(nT_e)X_a((n-m)T_e)] \\ &= \gamma_a(mT_e) \end{aligned} \quad (1.48)$$

La fonction d'autocorrélation du signal échantillonné s'obtient donc en échantillonnant la fonction d'autocorrélation du signal continu.

Densité spectrale de puissance : soit $\Gamma_a(f)$ la DSP de $X_a(t)$ et notons $\Gamma(f)$ la DSP de $X(n)$. Puisque $\gamma(m) = \gamma_a(mT_e)$, l'application du théorème d'échantillonnage donne :

$$\Gamma(f) = F_e \sum_{k=-\infty}^{+\infty} \Gamma_a(f + kF_e) \quad (1.49)$$

On note que l'échantillonnage en temps de γ entraîne la périodisation en fréquence de Γ . Néanmoins, si la DSP du signal continu est nulle en dehors de la bande de fréquence $[-F_e/2, F_e/2]$, on a simplement :

$$\Gamma(f) = F_e \Gamma_a(f) \quad \forall f \quad -F_e/2 \leq f \leq F_e/2 \quad (1.50)$$

1.5.3 Largeur de bande

Bruit blanc en continu

On dit qu'un signal aléatoire stationnaire est un bruit blanc quand sa densité spectrale de puissance est constante :

$$\Gamma(f) = N_o \quad \forall f \quad (1.51)$$

Si le signal est une tension, $\Gamma(f)$ s'exprime en Volt²/Hz.

La fonction d'autocorrélation est une impulsion de Dirac et s'obtient par la TF inverse de $\Gamma(f)$:

$$\gamma(\tau) = N_o \delta(\tau)$$

Ceci correspond au cas où le bruit est totalement décorréolé en deux instants différents, aussi proches soient-ils. Dans la pratique, un bruit n'est jamais totalement blanc (on parle de bruit coloré), le capteur du signal étant lui même nécessairement à bande limité. En revanche, si on s'intéresse à un signal utile dans une bande donnée, on fera fréquemment l'hypothèse de bruit blanc dans la bande étudiée. Dans le cas des signaux échantillonnés, l'hypothèse de blancheur implique que les échantillons successifs sont totalement décorréolés.

Bruit blanc dans une bande en continu

On dit qu'un signal aléatoire stationnaire est un bruit blanc dans une bande quand sa densité spectrale de puissance est égale à :

$$\Gamma(f) = \begin{cases} N_o & \forall -B \leq f \leq B \\ 0 & \text{sinon} \end{cases} \quad (1.52)$$

La fonction d'autocorrélation s'obtient par la TF inverse de $\Gamma(f)$:

$$\begin{aligned} \gamma(\tau) &= N_o \int_{-B}^B \exp(2i\pi f\tau) df \\ &= N_o B \frac{\exp(2i\pi\tau B) - \exp(-2i\pi\tau B)}{2i\pi\tau} \\ &= 2N_o B \frac{\sin(2\pi B\tau)}{2\pi B\tau} \\ &= 2N_o B \operatorname{sinc}(2\pi B\tau) \end{aligned} \quad (1.53)$$

La puissance du signal vaut $\gamma(0) = 2N_o B$ et premier zéro à $1/(2B)$.

1.6 Opération linéaire sur signaux aléatoires

Nous allons nous intéresser au filtrage de signaux aléatoires et leur conséquence sur le signal filtré dans les domaines temporels et fréquentiels.

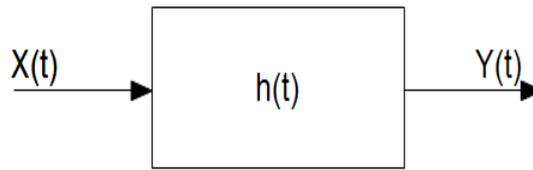


FIGURE 1.4 – Filtrage d’un signal aléatoire $X(t)$ par un filtre de réponse impulsionnelle $h(t)$.

1.6.1 Filtrage d’un signal aléatoire stationnaire au sens large : analyse temporelle

Soit $X(t)$ un signal aléatoire stationnaire au sens large de moyenne m_X et de fonction d’autocorrélation $\gamma_X(\tau)$, et $Y(t)$ le signal obtenu par filtrage de $X(t)$ par un filtre de réponse impulsionnelle $h(t)$ (réelle) (cf fig. 1.4).

$Y(t)$ s’exprime en fonction de $X(t)$ selon :

- continu : $Y(t) = h(t) * X(t) = \int h(t')X(t - t')dt'$
- numérique : $Y(n) = h(n) * X(n) = \sum_k h(k)X(n - k)$

Nous allons montrer que la moyenne et la fonction d’autocorrélation de $Y(t)$ sont liées à la moyenne et la fonction d’autocorrélation de $X(t)$ dans le théorème suivant.

Theorem 1.6.1 *Les propriétés de $Y(t) = h(t) * X(t)$ sont les suivantes :*

- $Y(t)$ a pour moyenne : $m_Y = m_X \int h(t)dt$
- $Y(t)$ a pour fonction d’autocorrélation : $\gamma_Y(\tau) = h(-\tau) * h(\tau) * \gamma_X(\tau)$

Nous avons un théorème équivalent pour le cadre numérique.

PROOF. Cf TD. □

1.6.2 Filtrage d’un signal aléatoire stationnaire au sens large : analyse fréquentielle

Nous proposons dans le théorème suivant la relation entre les DSP de $X(t)$ et $Y(t)$.

Theorem 1.6.2 *Soit $X(t)$ un signal aléatoire stationnaire au sens large de DSP $\Gamma_X(f)$. Soit $Y(t)$ le signal obtenu en filtrant $X(t)$ par un filtre de réponse impulsionnelle $h(t)$ et de gain complexe $H(f)$. Alors la DSP de $Y(t)$, $\Gamma_Y(f)$, s’exprime en fonction de $\Gamma_X(f)$ selon :*

$$\Gamma_Y(f) = |H(f)|^2 \Gamma_X(f) \tag{1.54}$$

PROOF. Cf TD. □

1.6.3 Quelques conséquences immédiates

Nous allons voir quelques conséquences immédiates de la relation (1.54). Nous reprenons les notations des sections précédentes.

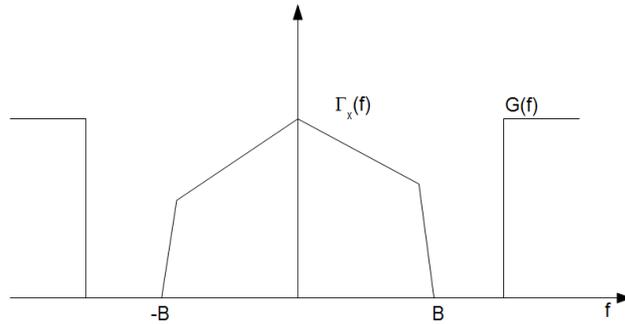


FIGURE 1.5 – Spectre d'un signal aléatoire $X(t)$ à spectre limité $[-B, B]$ et d'un filtre $H(f)$ passe haut rejetant les fréquences inférieures à B .

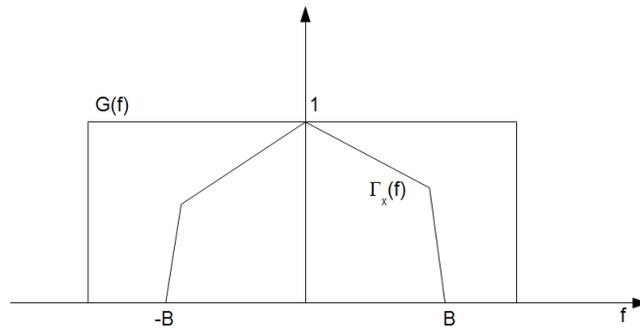


FIGURE 1.6 – Spectre d'un signal aléatoire $X(t)$ à spectre limité $[-B, B]$ et d'un filtre $H(f)$ passe tout les fréquences inférieures à B .

Débruitage

Supposons que $X(t)$ est à spectre limité $[-B, B]$ et que le filtre $H(f)$ est un filtre passe haut rejetant les fréquences inférieures à B comme le montre la figure 1.5.

Il est clair dans cette situation que $\Gamma_Y(f) = \Gamma_X(f)|H(f)|^2 = 0$. Par conséquent, le signal $Y(t)$ en sortie du filtre est de puissance nulle puisque $P = E[Y(t)^2] = \int \Gamma_Y(f)df = 0$. Et donc $Y(t)$ est nul (v.a. dont le moment d'ordre 2 est nul est elle-même nulle).

Passe-tout

Étudions le cas extrême où $H(f)$ est constant et égal à 1 dans la bande occupée par le signal $X(t)$ comme le montre la figure 1.6.

Il est clair dans cette situation que $\Gamma_Y(f) = \Gamma_X(f)|H(f)|^2 = \Gamma_X(f)$ dans la bande de fréquence $[-B, B]$.

Application au cas d'un signal bruité par un bruit bande large

Prenons le cadre d'un signal utile $s(t)$ corrompu par un bruit $b(t)$:

$$x(t) = s(t) + b(t) \tag{1.55}$$

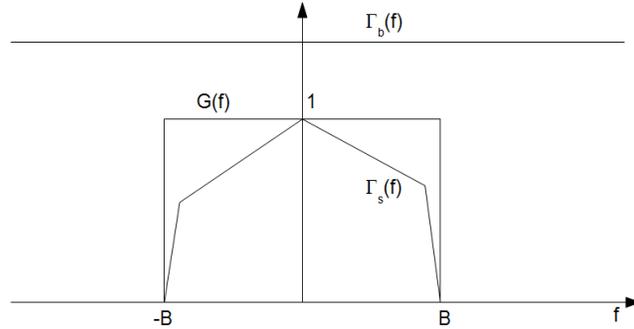


FIGURE 1.7 – Spectre d'un signal aléatoire $S(t)$ à spectre limité $[-B, B]$ + d'un bruit blanc et d'un filtre $H(f)$ passe bas idéal autour de $[-B, B]$.

Si le signal $s(t)$ est à bande limitée $[-B, B]$, un filtrage passe bas idéal, $h(t)$, de $x(t)$ dans la bande $[-B, B]$ fournit un signal $y(t)$ égal à (cf figure 1.7 pour le spectre de ces 3 signaux) :

$$\begin{aligned} y(t) &= h(t) * s(t) + h(t) * b(t) \\ &= s(t) + b'(t) \end{aligned} \quad (1.56)$$

car $S'(f) = H(f) \cdot S(f) = S(f)$ car $H(f) = 1$ dans $[-B, B]$. La puissance $P_{b'}$ du bruit $b'(t)$ après filtrage vaut :

$$P_{b'} = E[b'(t)^2] = \int_{-\infty}^{+\infty} \Gamma_b(f) |H(f)|^2 df \quad (1.57)$$

Pour un filtre passe-bas idéal,

$$H(f) = \begin{cases} 1 & \text{pour } -B \leq f \leq B \\ 0 & \text{sinon} \end{cases} \quad (1.58)$$

et la puissance du bruit après filtrage s'écrit simplement :

$$P_{b'} = \int_{-B}^{+B} \Gamma_b(f) df < \int_{-\infty}^{+\infty} \Gamma_b(f) df = P_b \quad (1.59)$$

Ainsi le filtrage permet de réduire la puissance du bruit.

1.6.4 Formule de cohérence et des interférences

On étudie le lien entre deux signaux aléatoires X et Y en calculant l'inter-spectre normalisé appelé coefficient de cohérence :

$$C_{YX}(f) = \frac{\Gamma_{YX}(f)}{\sqrt{\Gamma_{YY}(f) \cdot \Gamma_{XX}(f)}} \quad (1.60)$$

Le module $|C_{YX}(f)|$ est compris entre 0 et 1.

Enfin dans le cas général de deux filtres h_1 et h_2 ayant comme entrée deux processus stationnaires au second ordre $X_1(t)$ et $X_2(t)$, on retrouve la formule des interférences du cadre déterministe. L'interspectre des deux sorties $Y_1(t)$ et $Y_2(t)$ se calcule en fonction de l'interspectre des signaux d'entrée :

$$\Gamma_{Y_1 Y_2}(f) = H_1(f) \cdot H_2^*(f) \cdot \Gamma_{X_1 X_2}(f) \quad (1.61)$$

Si les deux filtres ont des fonctions de transfert disjointes (les spectres $H_1(f)$ et $H_2(f)$ sont non-nuls sur des intervalles disjoints), les signaux de sortie sont totalement décorrélés.

Cas particulier : lorsque X et Y sont respectivement l'entrée et la sortie d'un filtre linéaire, on en déduit l'intercorrélation entre les processus de sortie $Y(t)$ et l'entrée $X(t)$ ($Y_1(f) = H_1(f)X(f)$ et $Y_2 = X$, donc $H_2(f) = 1$) d'un filtre linéaire dans le domaine fréquentiel :

$$\Gamma_{YX}(f) = H(f) \cdot \Gamma_{XX}(f) \quad (1.62)$$

soit dans le domaine temporel :

$$\gamma_{YX}(\tau) = h(\tau) * \gamma_{XX}(\tau) \quad (1.63)$$

On déduit ensuite directement de l'équation (1.62) que le module $C_{YX}(f)$ est égal à 1 (car $\Gamma_{YX}(f) = H(f) \cdot \Gamma_{XX}(f)$). Une estimation de la cohérence permet ainsi de vérifier la linéarité d'un filtre.

Une chute de la cohérence dans une partie du spectre peut révéler une non-linéarité ou la présence d'un faible rapport signal à bruit sur l'entrée ou la sortie dans cette partie du spectre.

Chapitre 2

Estimation

Les grandeurs rencontrées dans le chapitre précédent, comme la fonction d'autocorrélation et la DSP, sont fonction de variables aléatoires. Il est donc impossible de les obtenir directement car ils sont eux-mêmes des variables aléatoires (avec donc différentes réalisations possibles). Pour utiliser ces grandeurs, il est donc nécessaire de les estimer.

Exemple : estimation de la moyenne Soit un s.a. $X(n)$ de moyenne m dont nous disposons seulement de N échantillons $x(1), \dots, x(N)$ indépendants entre eux. Un estimateur possible de la moyenne est la moyenne empirique :

$$\hat{m} = \frac{1}{N} \sum_{i=1}^N x(i) \quad (2.1)$$

2.1 Définitions générales relatives à l'estimation

Soit $X(t)$ un signal aléatoire continu (ou $X(n)$ en numérique) dont la loi de probabilité dépend d'un paramètre a .

Définition estimateur : un estimateur de a est une fonction \hat{a} de $X(t)$ (ou $X(n)$) dont on souhaite qu'elle soit la plus proche possible de a .

Remarque : \hat{a} est évidemment une variable aléatoire.

Pour mesurer la proximité de \hat{a} avec a et donc la qualité de l'estimation, on définit différentes grandeurs.

Biais de l'estimateur : le biais d'un estimateur, noté b , est la différence entre l'espérance de l'estimateur et la vraie valeur :

$$b = E[\hat{a}] - a \quad (2.2)$$

Variance de l'estimateur : la variance d'un estimateur, noté $var(\hat{a})$, mesure la variation de puissance l'estimateur autour de sa valeur moyenne :

$$var(\hat{a}) = E[(\hat{a} - E[\hat{a}])^2] \quad (2.3)$$

Erreur Quadratique Moyenne (EQM) de l'estimateur : l'EQM d'un estimateur mesure la variation de puissance l'estimateur autour de sa vraie valeur :

$$EQM(\hat{a}) = E[(\hat{a} - a)^2] \quad (2.4)$$

Nous montrons la relation suivante entre EQM et variance dans le théorème suivant.

Theorem 2.1.1 Soit \hat{a} un estimateur du paramètre a ayant un biais b . Nous avons la relation suivante entre la variance et l'EQM :

$$EQM = var(\hat{a}) + b^2 \quad (2.5)$$

PROOF. Nous avons pour l'EQM :

$$\begin{aligned} EQM(\hat{a}) &= E[(\hat{a} - E[\hat{a}] + E[\hat{a}] - a)^2] \\ &= E[(\hat{a} - E[\hat{a}])^2 + E[(E[\hat{a}] - a)^2] + 2E(\hat{a} - E[\hat{a}])E(E[\hat{a}] - a)] \\ &= var(\hat{a}) + b^2 \end{aligned} \quad (2.6)$$

En effet le dernier terme comprend : $E(\hat{a} - E[\hat{a}]) = E[\hat{a}] - E[\hat{a}] = 0$. □

Nous proposons une définition pour juger de la qualité d'un estimateur.

Définition de la consistance : soit \hat{a} un estimateur de a calculé sur un intervalle T de $X(t)$, on dit que l'estimateur \hat{a} est consistant si :

$$EQM(\hat{a}) \xrightarrow{T \rightarrow +\infty} 0 \quad (2.7)$$

Cette définition signifie que si on dispose d'un intervalle assez grand du signal (et donc de réalisations de la variable aléatoire), alors l'estimateur sera quasi égal à la vraie valeur.

Exemple : qualité de la moyenne empirique En reprenant l'exemple de la moyenne empirique, on a pour le biais :

$$E[\hat{m}] = E\left[\frac{1}{N} \sum_{i=1}^N x(i)\right] = m \quad (2.8)$$

car $E[\]$ est un opérateur linéaire. Calculons maintenant la variance, on a :

$$var(\hat{m}) = \frac{var(X(n))}{N} \quad (2.9)$$

PROOF.

$$\begin{aligned} var(\hat{m}) &= E\left[\left(\frac{1}{N} \sum_{i=1}^N x(i) - m\right)^2\right] \\ &= E\left[\left(\frac{1}{N} \sum_{i=1}^N x(i) - \frac{N}{N}m\right)^2\right] \\ &= \frac{1}{N^2} E\left[\left(\sum_{i=1}^N (x(i) - m)\right)^2\right] \\ &= \frac{1}{N^2} \sum_{i=1}^N \sum_{j=1}^N E[(x(i) - m)(x(j) - m)] \\ &= \frac{1}{N^2} \sum_{i=1}^N \sum_{j=1}^N (E[(x(i)x(j))] - m^2) \\ &= \frac{1}{N^2} \sum_{i=1}^N (E[(x(i)^2)] - m^2) \\ &= \frac{1}{N^2} \sum_{i=1}^N var(x(i)) \\ &= \frac{var(x(i))}{N} \end{aligned} \quad (2.10)$$

Le passage dans la sixième égalité suppose que les v.a. $x(i)$ sont indépendantes entre elles. □

On remarque que la variance tend vers 0 quand le nombre d'échantillons N augmente. Comme le biais est nul, l'EQM est égale à la variance. Donc la moyenne empirique est une estimée consistante de la moyenne.

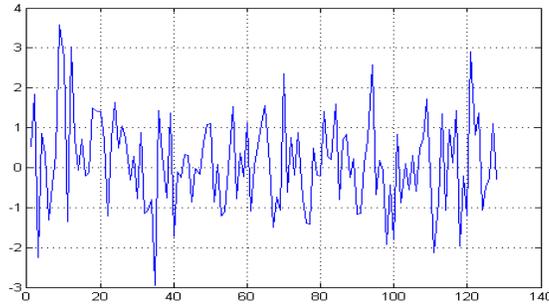


FIGURE 2.1 – Exemple d'un signal aléatoire $x(n)$ de longueur $N = 128$: bruit blanc centré de variance 1.

2.2 Estimation de la fonction d'autocorrélation

Pour cette section, nous nous plaçons dans le cadre de signaux numériques.

Soit $X(n)$ un signal aléatoire avec des réalisations notées $x(n)$. Nous rappelons sa fonction d'autocorrélation :

$$\gamma(m) = E[x(n)x(n - m)] \quad (2.11)$$

Nous supposons disposer de N échantillons $x(0), \dots, x(N - 1)$ de $x(n)$. Nous montrons un exemple sur la figure 2.1 avec $X(n)$ un bruit blanc centré de variance 1 et $N = 128$.

Nous proposons maintenant deux estimateurs de la vraie fonction d'autocorrélation, $\gamma(m)$, avec des propriétés différentes.

2.2.1 Estimateur non biaisé

Définition de l'estimateur non biaisé : Soit N échantillons $x(0), \dots, x(N - 1)$ réalisations du signal aléatoire $X(n)$, un estimateur de $\gamma(m)$ est donnée par la formule suivante :

$$\begin{aligned} \hat{\gamma}(m) &= \frac{1}{N - m} \sum_{n=m}^{N-1} x(n)x(n - m) \quad \text{pour } 0 \leq m \leq N - 1 \\ \hat{\gamma}(-m) &= \hat{\gamma}(m) \quad \text{sinon} \end{aligned} \quad (2.12)$$

Biais : on montre facilement que cet estimateur est sans biais :

$$\begin{aligned} E[\hat{\gamma}(m)] &= E \left[\frac{1}{N - m} \sum_{n=m}^{N-1} x(n)x(n - m) \right] \\ &= \frac{1}{N - m} \sum_{n=m}^{N-1} E[x(n)x(n - m)] \\ &= \gamma(m) \end{aligned} \quad (2.13)$$

Variance : la variance de cet estimateur est difficile à calculer. Nous nous contenterons d'en faire une analyse qualitative. Dans l'Eq. (2.12), on fait une moyenne sur $N - m$ termes donc en fonction de m , nous aurons des résultats différents :

- m petit : le nombre de moyennage sera important et donc la variance sera faible
- m grand : tout le contraire, faible nombre de moyennes et donc variance forte

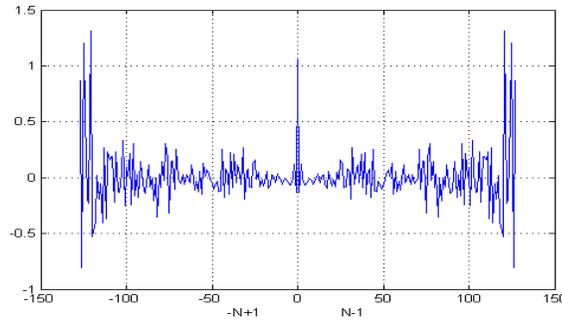


FIGURE 2.2 – Estimateur de $\gamma(m)$ sans biais de $X(n)$ un bruit blanc centré de variance 1 et $N = 128$.

Nous verrons que ce problème se rencontre souvent dans l'analyse de signaux aléatoires à durée finie : effet de bord. Nous pouvons voir sur la figure 2.2 le résultat de l'Eq. (2.12) pour le signal $X(n)$ de la figure 2.1. On remarque bien la grande variance sur les bords.

2.2.2 Estimateur biaisé

Le but est d'utiliser une fonction d'apodisation pour d'abord conserver $\hat{\gamma}(m)$ pour les faibles valeurs de m et de forcer $\hat{\gamma}(m) \approx 0$ pour m grand. Nous allons utiliser pour ça une fenêtre d'apodisation triangulaire.

Définition de l'estimateur biaisé : Soit N échantillons $x(0), \dots, x(N-1)$ réalisations du signal aléatoire $X(n)$, un estimateur de $\gamma_b(m)$ est donnée par la formule suivante :

$$\begin{aligned}\hat{\gamma}_b(m) &= \frac{N-|m|}{N} \hat{\gamma}(m) \\ &= \frac{1}{N} \sum_{n=m}^{N-1} x(n)x(n-m)\end{aligned}\quad (2.14)$$

La fonction $\frac{N-|m|}{N}$ est une fenêtre triangulaire s'annulant pour $m = N$ et $m = -N$.

Biais : on montre facilement que cet estimateur est biaisé :

$$\begin{aligned}E[\hat{\gamma}_b(m)] &= E\left[\frac{N-|m|}{N} \hat{\gamma}(m)\right] \\ &= \frac{N-|m|}{N} \gamma(m)\end{aligned}\quad (2.15)$$

Donc le biais vaut :

$$b = \frac{-|m|}{N} \gamma(m)\quad (2.16)$$

Le biais est non nul mais on remarque que sa valeur sera faible pour m faible.

Variance : on peut calculer la variance de $\hat{\gamma}_b$ en fonction de celle de $\hat{\gamma}$:

$$\text{var}(\hat{\gamma}_b) = \left(\frac{N-|m|}{N}\right)^2 \text{var}(\hat{\gamma})\quad (2.17)$$

Faisons une analyse qualitative de cette formule :

- m petit : $\text{var}(\hat{\gamma})$ est faible donc la variance de $\text{var}(\hat{\gamma}_b)$ sera aussi faible
- m grand : le terme $\left(\frac{N-|m|}{N}\right)^2$ sera très faible et donc $\text{var}(\hat{\gamma}_b)$ sera aussi faible

Nous pouvons voir sur la figure 2.3 le résultat de l'Eq. (2.14) pour le signal $X(n)$ de la figure 2.1. On remarque bien la faible variance sur les bords.

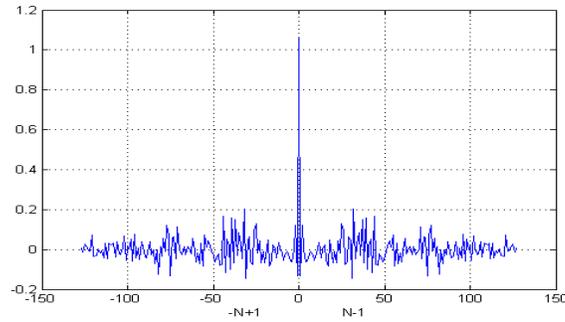


FIGURE 2.3 – Estimateur de $\gamma(m)$ avec biais de $X(n)$ un bruit blanc centré de variance 1 et $N = 128$.

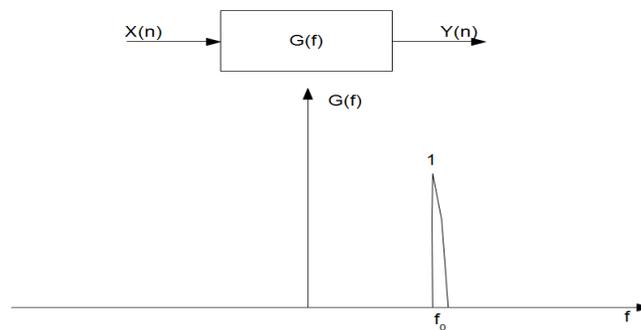


FIGURE 2.4 – Filtrage de $X(n)$ par un filtre de réponse fréquentielle $H(f)$ sélectif autour de la fréquence f_0 .

2.3 Estimation de la DSP

On rappelle la définition de la DSP d'un signal aléatoire $X(n)$ avec des réalisations notées $x(n)$. Nous supposons disposer de N échantillons $x(0), \dots, x(N-1)$ de $x(n)$, la DSP de $X(n)$ s'écrit alors :

$$\Gamma(f) = \lim_{T \rightarrow +\infty} E \left[\frac{|S_T(f)|^2}{2T} \right] \quad (2.18)$$

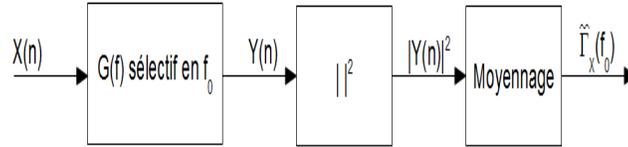
On rappelle qu'on a aussi :

$$\Gamma(f) = TF(\gamma(m)) = \sum_{n=0}^{N-1} \gamma(n) \exp(-2i\pi fnT_e) \quad (2.19)$$

L'estimation de la DSP est souvent appelée analyse spectrale.

2.3.1 Principe de l'analyse spectrale

Nous reprenons le cadre du filtrage du signal $X(n)$ par un filtre de réponse fréquentielle $H(f)$ (éventuellement complexe). Dans ce cadre, $H(f)$ est un filtre sélectif autour de la fréquence f_0 . La figure 2.4 montre le spectre $H(f)$ du filtre.


 FIGURE 2.5 – Schéma de principe de l'analyse spectrale classique autour de la fréquence f_0 .

Soit $P_y = E[Y(n)^2]$ la puissance de $Y(n)$, on a :

$$\begin{aligned}
 E[Y(n)^2] &= \frac{1}{f_e} \int_{-f_e/2}^{f_e/2} \Gamma_y(f) df \\
 &= \frac{1}{f_e} \int_{-f_e/2}^{f_e/2} \Gamma_x(f) |H(f)|^2 df \\
 &\approx \frac{1}{f_e} \int_{-f_e/2}^{f_e/2} \Gamma_x(f_0) |H(f)|^2 df \\
 &\approx \Gamma_x(f_0) \frac{1}{f_e} \int_{-f_e/2}^{f_e/2} |H(f)|^2 df
 \end{aligned} \tag{2.20}$$

En résumé :

$$P_y = E[Y(n)^2] = \alpha_g \Gamma_x(f_0) \tag{2.21}$$

avec $\alpha_h = \sum_n |h(n)|^2 = \frac{1}{f_e} \int_{-f_e/2}^{f_e/2} |H(f)|^2 df$ (Parseval). Cette dernière formule nous montre que nous avons accès à la DSP de $X(n)$ en estimant la puissance P_y . Pour cette dernière estimation, nous utilisons l'estimation classique de la puissance :

$$\hat{P}_y = \frac{1}{K} \sum_{k=1}^K |y(kM)|^2 \tag{2.22}$$

avec $M > 1$ un facteur de sous échantillonnage dans le but de garantir des sorties décorréliées. Ce qui nous donne pour l'estimation de la DSP en f_0 :

$$\hat{\Gamma}_x(f_0) = \frac{1}{\alpha_g K} \sum_{k=1}^K |y(kM)|^2 \tag{2.23}$$

Nous pouvons voir sur la figure 2.5 le synoptique de l'estimation de $\Gamma_x(f_0)$.

2.3.2 Mise en oeuvre pratique : méthode de Welch

Nous choisissons pour le filtre sélectif que la réponse impulsionnelle s'écrive de la manière suivante :

$$h(n) = \exp(2i\pi f_0 n T_e) w(n) \tag{2.24}$$

où $w(n)$ est une fenêtre d'apodisation (rectangulaire, Hanning, ...) de taille N (de 0 à $N - 1$). Les caractéristiques du fenêtre d'apodisation sont qu'elles valent 0 au bord (en 0 et $N - 1$) et qu'elles sont symétriques. La réponse en fréquence du filtre vaut donc :

$$H(f) = W(f) * \delta(f - f_0) = W(f - f_0) \tag{2.25}$$

On va montrer maintenant que $P_y = |y(n)|^2$ peut s'obtenir directement par TF finie pour $f_0 = 0, \frac{f_e}{N}, \dots, \frac{N-1}{N} f_e$. Donnons l'expression de $y(n)$:

$$\begin{aligned}
 y(n) &= \sum_{k=0}^{N-1} h(k)x(n-k) \\
 &= \sum_{k=0}^{N-1} \exp(2i\pi f_0 k T_e) w(k)x(n-k) \\
 &= \sum_{l=0}^{N-1} \exp(-2i\pi f_0 l T_e) w(l)x(n-N+1+l) \exp(2i\pi f_0 (N-1)T_e) \quad \text{avec } l = N-1-k
 \end{aligned} \tag{2.26}$$

En prenant le module au carré :

$$\begin{aligned}
 \hat{P}_y = |y(n)|^2 &= \left| \sum_{l=0}^{N-1} \exp(-2i\pi f_0 l T_e) w(l)x(n-N+1+l) \right|^2 \\
 &= |TF\{w(0)x(n-N+1), \dots, w(N-1)x(n)\}|^2
 \end{aligned} \tag{2.27}$$

Le problème ici est que \hat{P}_y est calculé sur une seule réalisation. Il va falloir donc avoir plusieurs réalisations de \hat{P}_y pour obtenir une estimation correcte de cette grandeur. En revanche, on remarque que l'avantage est de pouvoir utiliser la FFT pour $f = 0, \frac{f_e}{N}, \dots, \frac{N-1}{N} f_e$ et donc d'obtenir un résultat rapide de la DSP pour toutes les fréquences.

En résumé, la méthode de Welch se décompose en 4 étapes :

- on segmente les données de $X(n)$ en K paquets de taille N comme le montre la figure 2.6. Les données du k ème paquet sont $x((k-1)N+l)$ avec $l = 0, \dots, N-1$.
- on pondère les données de chaque paquet k par la fenêtre d'apodisation $w(0), \dots, w(N-1)$ pour obtenir l'ensemble de données $w(0)x((k-1)N+0), \dots, w(N-1)x((k-1)N+N-1)$
- on calcul par FFT la TF finie de chaque paquet k :

$$X_k(f) = \sum_{l=0}^{N-1} w(l)x((k-1)N+l) \exp(-2i\pi f l T_e) \tag{2.28}$$

pour $f = 0, \frac{f_e}{N}, \dots, \frac{N-1}{N} f_e$.

- estimation de la puissance de y en moyennant les différents blocs : $\hat{P}_y = \frac{1}{K} \sum_{k=0}^{K-1} |X_k(f)|^2$
- la dernière étape consiste à l'estimation de la DSP :

$$\hat{\Gamma}(f) = \frac{1}{\alpha_w} \hat{P}_y = \frac{1}{\alpha_w K} \sum_{k=0}^{K-1} |X_k(f)|^2 \tag{2.29}$$

avec α_w :

$$\begin{aligned}
 \alpha_w &= \sum_n |h(n)|^2 \\
 &= \sum_n |\exp(2i\pi f_0 n T_e) w(n)|^2 \\
 &= \sum_n |w(n)|^2
 \end{aligned} \tag{2.30}$$

$|X_k(f)|^2$ s'appelle le périodogramme du k ème paquet de données. La méthode de Welch s'appelle aussi la méthode de lissage temporel du périodogramme.

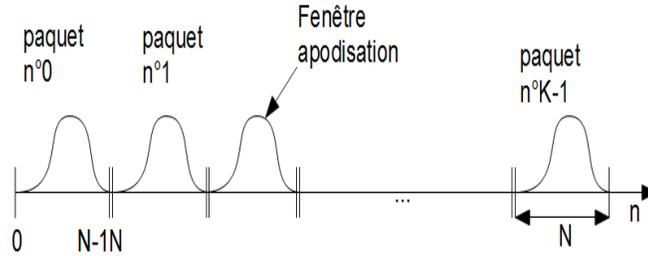


FIGURE 2.6 – Découpage et apodisation du signal $X(n)$.

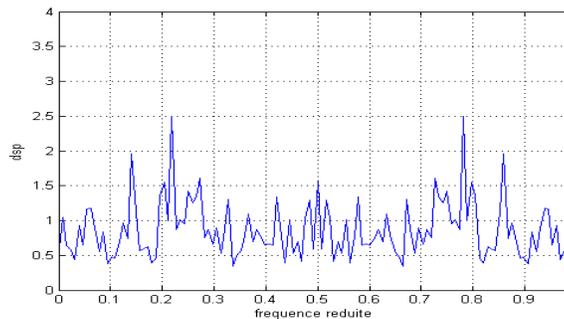


FIGURE 2.7 – Estimateur de $\Gamma(m)$ de $X(n)$ un bruit blanc centré de variance 1 en utilisant comme paramètres $N = 128$ et $K = 5$ et une fenêtre d'apodisation rectangulaire.

2.3.3 Qualité de l'estimation spectrale par la méthode de Welch

Nous ne détaillons pas les calculs qui sont assez compliqués mais juste les résultats et leurs implications. Tout d'abord, l'espérance de $\hat{\Gamma}(f)$:

$$E[\hat{\Gamma}(f)] \approx \Gamma(f) \quad (2.31)$$

On peut donc dire que l'estimateur de Welch de la DSP est pratiquement sans biais. Maintenant étudions la variance de l'estimateur :

$$\text{var}(\hat{\Gamma}(f)) \approx \frac{\Gamma(f)}{K} \quad (2.32)$$

Ce résultat signifie que plus on a de paquets plus la variance sera faible mais malheureusement, on ne peut pas prendre un nombre K trop important car les signaux étant finis, cela nous oblige à réduire trop fortement N et donc la résolution de la FFT qui est f_e/N . Il y a donc un compromis à rechercher entre variance et résolution.

Nous montrons l'exemple de l'estimation de la DSP de $X(n)$ un bruit blanc centré de variance 1 en utilisant comme paramètres $N = 128$ et $K = 5$ et une fenêtre d'apodisation rectangulaire sur la figure 2.7, en utilisant comme paramètres $N = 32$ et $K = 50$ et une fenêtre d'apodisation rectangulaire sur la figure 2.8 et en utilisant comme paramètres $N = 128$ et $K = 50$ et une fenêtre d'apodisation rectangulaire sur la figure 2.9. Nous remarquons que quand $K = 5$ la variance est très importante alors qu'elle est nettement réduite avec $K = 50$. En revanche, la valeur N influe sur la résolution.

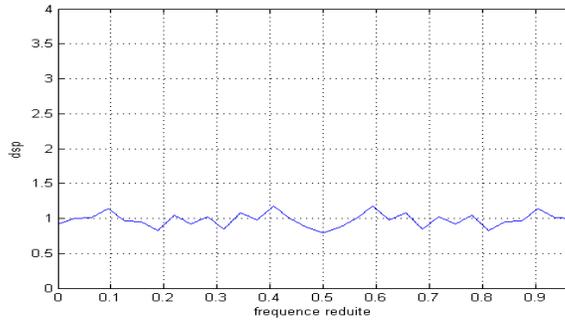


FIGURE 2.8 – Estimateur de $\Gamma(m)$ de $X(n)$ un bruit blanc centré de variance 1 en utilisant comme paramètres $N = 32$ et $K = 50$ et une fenêtre d'apodisation rectangulaire.

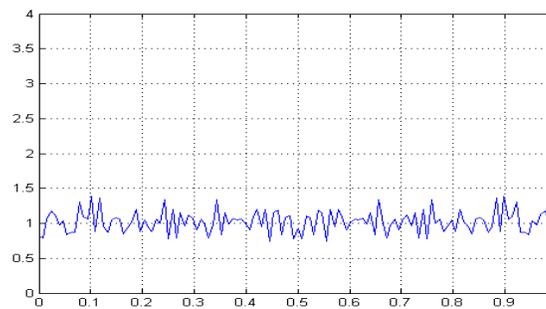


FIGURE 2.9 – Estimateur de $\Gamma(m)$ de $X(n)$ un bruit blanc centré de variance 1 en utilisant comme paramètres $N = 128$ et $K = 50$ et une fenêtre d'apodisation rectangulaire.

Chapitre 3

Filtrage Adaptatif

Nous allons voir une application très importante du traitement du signal aléatoire, le filtrage adaptatif, comme pour le débruitage ou en égalisation en télécoms mobiles.

3.1 Rappels mathématiques

3.1.1 Algèbre matricielle

Définition vecteur : un vecteur $\mathbf{x} \in \mathbb{R}^N$ est de la forme :

$$\mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_N \end{pmatrix} \quad (3.1)$$

Définition matrice : une matrice $\mathbf{A} \in \mathbb{R}^{L \times K}$ est de la forme :

$$\mathbf{A} = (a_{ij})_{1 \leq i \leq L, 1 \leq j \leq K} = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1K} \\ x_{21} & a_{22} & \dots & a_{2K} \\ \vdots & \vdots & \vdots & \vdots \\ x_{L1} & a_{L2} & \dots & a_{LK} \end{pmatrix} \quad (3.2)$$

Définition transposé : la transposé d'une matrice $\mathbf{A} \in \mathbb{R}^{L \times K}$ est une matrice $\mathbf{B} \in \mathbb{R}^{K \times L}$ (noté \mathbf{A}^T) tel que : $\mathbf{B} = (a_{ji})_{1 \leq j \leq K, 1 \leq i \leq L}$

Propriété : une matrice est carrée symétrique si son transposé est identique : $\mathbf{A} = \mathbf{A}^T$.

Définition produit scalaire : le produit scalaire entre deux vecteurs \mathbf{x} et \mathbf{y} s'écrit :

$$\mathbf{x}^T \mathbf{y} = \sum_{i=1}^N x_i y_i \quad (3.3)$$

Définition norme d'un vecteur : la norme du vecteur \mathbf{x} s'écrit :

$$\|\mathbf{x}\|^2 = \mathbf{x}^T \mathbf{x} = \sum_{i=1}^N x_i^2 \quad (3.4)$$

Produit de deux matrices : le produit de deux matrices $\mathbf{A} \in \mathbb{R}^{L \times K}$ et $\mathbf{B} \in \mathbb{R}^{K \times M}$ donne une matrice $\mathbf{C} \in \mathbb{R}^{L \times M}$ tel que :

$$c_{ij} = \sum_{k=1}^K a_{ik} b_{kj} \quad (3.5)$$

Propriété forme quadratique : soit une matrice $\mathbf{A} \in \mathbb{R}^{L \times K}$ et deux vecteurs $\mathbf{x} \in \mathbb{R}^L$ et $\mathbf{y} \in \mathbb{R}^K$, on a :

$$\mathbf{x}^T \mathbf{A} \mathbf{y} = \sum_{i=1}^L \sum_{j=1}^K x_i a_{ij} y_j \quad (3.6)$$

Propriétés diverses :

- $(\mathbf{A}^T)^T = \mathbf{A}$
- $(\mathbf{A}\mathbf{B})^T = \mathbf{B}^T \mathbf{A}^T$
- $\mathbf{x}^T \mathbf{y} = \mathbf{y}^T \mathbf{x}$
- $(\mathbf{A}\mathbf{B})^{-1} = \mathbf{B}^{-1} \mathbf{A}^{-1}$
- $(\mathbf{A}^T)^{-1} = (\mathbf{A}^{-1})^T$

3.1.2 Gradient fonction à plusieurs variables

Soit $f(x_1, \dots, x_N)$ une fonction de plusieurs variables. On pose $\mathbf{x} = (x_1, \dots, x_N)^T$. La différentielle de f est :

$$\begin{aligned} df &= f(x_1 + dx_1, \dots, x_N + dx_N) - f(x_1, \dots, x_N) \\ &= \frac{\partial f}{\partial x_1} dx_1 + \dots + \frac{\partial f}{\partial x_N} dx_N \end{aligned} \quad (3.7)$$

avec dx_1, \dots, dx_N des infiniment petits. Le vecteur

$$\frac{\partial f}{\partial \mathbf{x}} = \begin{pmatrix} \frac{\partial f}{\partial x_1} \\ \dots \\ \frac{\partial f}{\partial x_N} \end{pmatrix} \quad (3.8)$$

est le gradient de f .

Prenons quelques exemples de gradient de fonctions simples :

— $f(x_1, \dots, x_N) = a_1 x_1 + \dots + a_N x_N = \mathbf{a}^T \mathbf{x}$: alors le gradient est :

$$\frac{\partial f}{\partial \mathbf{x}} = \mathbf{a} \quad (3.9)$$

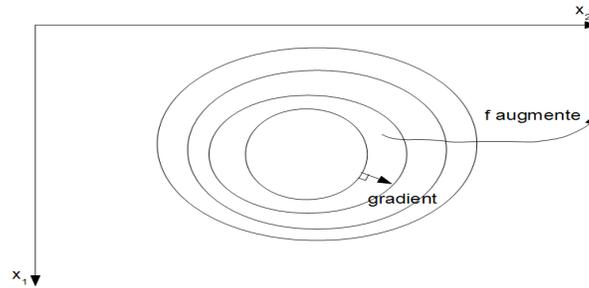
— $f(x_1, \dots, x_N) = \mathbf{x}^T \mathbf{A} \mathbf{x} = \sum_{i,j} x_i a_{i,j} x_j$: alors le gradient est :

$$\frac{\partial f}{\partial \mathbf{x}} = 2\mathbf{A}\mathbf{x} \quad (3.10)$$

car

$$\begin{aligned} \frac{\partial f}{\partial x_n} &= \sum_{i,j} a_{i,j} \delta_{i,n} x_j + \sum_{i,j} a_{i,j} \delta_{j,n} x_i \\ &= \sum_{j=1}^N a_{n,j} x_j + \sum_{i=1}^N a_{i,n} x_i \\ &= 2 \sum_{j=1}^N a_{n,j} x_j \end{aligned} \quad (3.11)$$

qui est le nième élément de $2\mathbf{A}\mathbf{x}$

FIGURE 3.1 – Gradient de $f(x_1, x_2)$.

3.1.3 Optimisation

Soit $f(x_1, \dots, x_N)$ une fonction de plusieurs variables. Beaucoup de problèmes d'optimisation se résolvent en annulant le gradient de f . Des fois on peut résoudre ce système analytiquement comme pour le MC pondérés. Mais quand on n'a pas de solutions analytiques, on a recours à des algorithmes itératifs. Ceux-ci permettent d'approcher le point $(x_{1,min}, \dots, x_{N,min})$ minimisant f . Nous utilisons dans ce cours l'algorithme du gradient. On présente le principe avec une fonction à 2 variables. Prenons l'algorithme suivant :

- Initialisation : $(x_{1,0}, x_{2,0})$
- Itérations :
 - $x_{1,n+1} = x_{1,n} + \Delta x_{1,n}$
 - $x_{2,n+1} = x_{2,n} + \Delta x_{2,n}$
 - avec $f(x_{1,n+1}, x_{2,n+1}) < f(x_{1,n}, x_{2,n})$.

Ces algorithmes ne convergent que rarement vers la bonne solution (ils s'en approchent), tombent souvent dans un minimum local. Ils peuvent converger vite si $\Delta x_{i,n}$ sont grands. On va proposer une manière de calculer ces incréments à partir du gradient de f .

On a noté dans le précédent rappel mathématiques que :

$$df = \frac{\partial f}{\partial \mathbf{x}}^T d\mathbf{x} \quad (3.12)$$

Donc $df = 0$ quand \mathbf{x} est orthogonal à $\frac{\partial f}{\partial \mathbf{x}}$. Le gradient est donc orthogonal aux courbes de niveaux. Quand $df > 0$, alors $d\mathbf{x} = \lambda \frac{\partial f}{\partial \mathbf{x}}$ ($\lambda > 0$). Le gradient est donc dirigé vers les valeurs de f croissantes comme le montre la figure 3.1.

On va donc calculer $(\Delta x_{1,n}, \Delta x_{2,n})$ en utilisant le sens opposé du gradient de f pour avoir df minimal :

$$\begin{pmatrix} \Delta x_{1,n} \\ \Delta x_{2,n} \end{pmatrix} = -\lambda \begin{pmatrix} \left(\frac{\partial f}{\partial x_1} \right)_{x_{1,n}, x_{2,n}} \\ \left(\frac{\partial f}{\partial x_2} \right)_{x_{1,n}, x_{2,n}} \end{pmatrix} \quad (3.13)$$

Il faut faire attention au choix du pas comme le montre la figure 3.2 : grand et on converge vite mais on risque de rater la solution. Il existe un pas optimal. On peut le trouver de manière itérative (pas abordé ici).

3.2 Introduction

3.2.1 Débruitage

Soit sur la figure 3.3, le problème de débruitage suivant : nous recevons un signal $d(t)$ contenant un signal utile $s(t)$ pollué par un bruit $b(t)$. Dans le but de supprimer ce bruit, nous pouvons aussi mesurer le signal $x(t)$

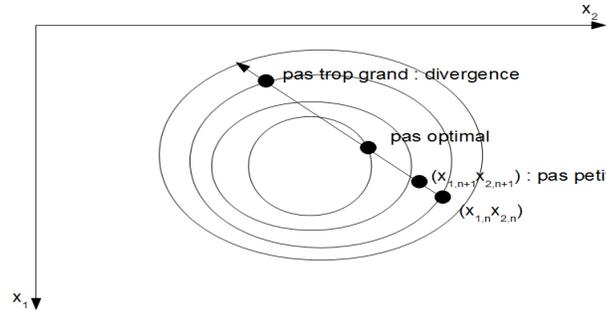


FIGURE 3.2 – Choix du pas λ dans l’algorithme du gradient.

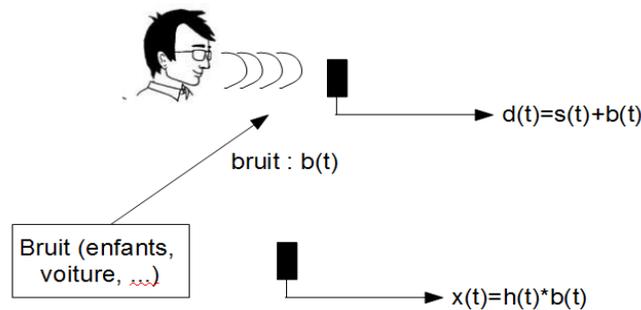


FIGURE 3.3 – Exemple de la pollution d’un signal utile pollué par un bruit filtré par un filtre inconnu.

(sans le signal utile) à l’aide d’un micro. Ce signal contient une version filtrée d’un bruit $h(t) * b(t)$. Néanmoins, est il possible de supprimer le bruit $b(t)$ dans le signal reçu $d(t)$ en disposant de $x(t) = h(t) * b(t)$? On remarque que la difficulté provient de la méconnaissance du filtre $h(t)$. Sur la figure 3.4, nous reprenons le problème en échantillonné. Nous devons donc estimer $h(n)$ pour avoir une erreur $e(n)$ la plus faible possible. De plus, nous allons supposer que ce filtre peut changer au cours du temps. Il faut donc une estimation adaptative des coefficients $h(n)$, d’où le nom de filtrage adaptatif.

La solution est présentée sur la figure 3.5 et consiste donc à déterminer automatiquement les coefficients du filtre $w(n)$ en exploitant les signaux mesurés $d(n)$ et $x(n)$. Cette solution permet de s’adapter automatiquement à l’environnement d’où le nom de filtrage adaptatif. Avec cette solution, on note que l’erreur $e(n)$ s’écrit sous la forme suivante :

$$e(n) = d(n) - w(n) * x(n) = d(n) - w(n) * h(n) * b(n) \quad (3.14)$$

Supposons que le signal utile $s(n)$ et le bruit $b(n)$ sont des signaux aléatoires stationnaires, centrés, décorrélés. On va donc choisir les coefficients du filtre $w(n) = w(0), \dots, w(M - 1)$ pour minimiser le second terme de l’erreur quadratique moyenne suivante :

$$E[e(n)^2] = E[(d(n) - \sum_{k=0}^{M-1} w(k)x(n - k))^2] \quad (3.15)$$

En pratique, les espérances seront remplacées par des moyennes empiriques.

3.2.2 Déconvolution

Ce second exemple se rencontre tout particulièrement en télécommunications mobiles. Les signaux transmis $d(n)$ en plus d’être pollué par un bruit $b(n)$ sont généralement altérés par le canal de propagation qui se comporte

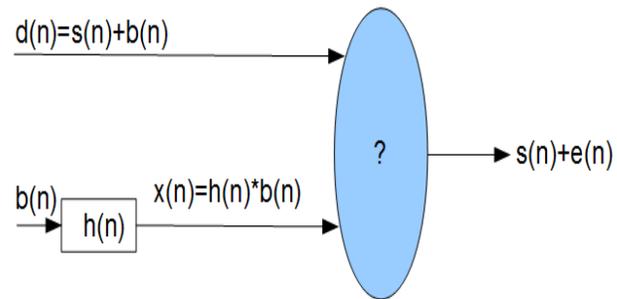


FIGURE 3.4 – Schéma du problème de la figure 3.3 en échantillonné.

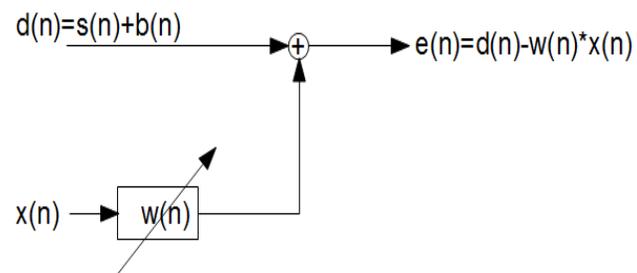


FIGURE 3.5 – Principe de débruitage en filtrage adaptatif.

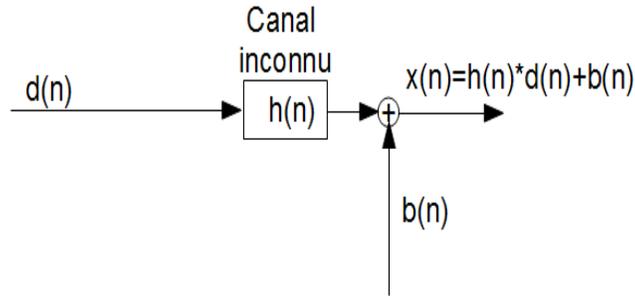


FIGURE 3.6 – Exemple de la convolution d’un signal utile par un filtre $h(n)$ inconnu.

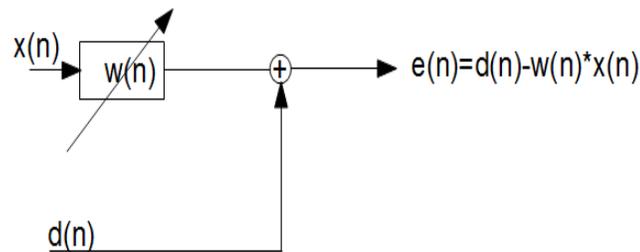


FIGURE 3.7 – Principe de la déconvolution avec séquence d’apprentissage.

alors comme un filtre. Cet exemple est résumé sur la figure 3.6. On aimerait donc récupérer les signaux $d(n)$ en compensant les effets de $h(n)$. En supposant qu’on connaît $h(n)$, on pourrait synthétiser un filtre $w(n)$ tel que $w(n) * h(n) = \delta(n)$. Dans ce cas, à la sortie du filtre, nous aurions :

$$y(n) = d(n) + w(n) * b(n) \tag{3.16}$$

Malheureusement, ce filtre $h(n)$ est très souvent inconnu en pratique et il faut donc l’estimer. Pour faire ceci, nous utilisons ce qu’on appelle une séquence d’apprentissage dans le but de s’approcher des coefficients du filtre $h(n)$: on envoie une séquence $d(n)$ parfaitement connue. Ensuite on choisit les coefficients du filtre $w(n) = w(0), \dots, w(N)$ pour minimiser l’erreur :

$$E[e(n)^2] = E\left[\left(d(n) - \sum_{k=0}^{M-1} w(k)x(n-k)\right)^2\right] \tag{3.17}$$

Le principe est montré sur la figure 3.7. Comme précédemment, les espérances seront remplacées par des moyennes empiriques.

3.3 Filtre de Wiener

Soit $x(n)$ et $d(n)$ deux signaux aléatoires stationnaires et le filtre RIF w de réponse impulsionnelle $w(0), \dots, w(M-1)$ qui doit approcher le filtre $h(n)$. Le principe pour estimer les coefficients $w(n)$ est de minimiser l’Erreur Qua-

dratique Moyenne (EQM) suivante :

$$EQM = E\left[\left(d(n) - \sum_{k=0}^{M-1} w(k)x(n-k)\right)^2\right] \quad (3.18)$$

En notant $\mathbf{w} = (w(0), \dots, w(M-1))^T$ et $\mathbf{x}_n = (x(n), x(n-1), \dots, x(n-M+1))^T$, on peut réécrire l'équation précédente sous la forme :

$$\begin{aligned} EQM &= E[(d(n) - \mathbf{w}^T \mathbf{x}_n)^2] \\ &= E[d(n)^2 - 2d(n)\mathbf{w}^T \mathbf{x}_n + \mathbf{w}^T \mathbf{x}_n \mathbf{x}_n^T \mathbf{w}] \\ &= E[d(n)^2] - 2\mathbf{w}^T E[d(n)\mathbf{x}_n] + \mathbf{w}^T E[\mathbf{x}_n \mathbf{x}_n^T] \mathbf{w} \end{aligned} \quad (3.19)$$

On pose :

$$\begin{aligned} \sigma^2 &= E[d(n)^2] \\ \mathbf{r} &= E[d(n)\mathbf{x}_n] \\ \mathbf{R} &= E[\mathbf{x}_n \mathbf{x}_n^T] \end{aligned} \quad (3.20)$$

On remarque que ces trois quantités ne dépendent pas de n car les signaux sont stationnaires. Avec ces nouvelles notations l'EQM vaut :

$$EQM = \sigma^2 - 2\mathbf{w}^T \mathbf{r} + \mathbf{w}^T \mathbf{R} \mathbf{w} \quad (3.21)$$

La valeur de \mathbf{w} qui minimise l'EQM annule son gradient :

$$\frac{\partial EQM}{\partial \mathbf{w}} = -2\mathbf{r} + 2\mathbf{R}\mathbf{w} \quad (3.22)$$

On obtient alors le filtre de Wiener :

$$\mathbf{w} = \mathbf{R}^{-1} \mathbf{r} \quad (3.23)$$

Le filtre de Wiener est le filtre optimal en environnement stationnaire. C'est une référence à laquelle nous comparons les autres filtres. En reportant ce filtre de Wiener dans l'EQM de l'Eq. (3.21), nous obtenons l'EQM minimale :

$$EQM_{min} = \sigma^2 - \mathbf{r}^T \mathbf{R}^{-1} \mathbf{r} \quad (3.24)$$

Malheureusement, ce filtre n'est pas réalisable car nous ne disposons pas de signaux à durée illimitée et donc on ne peut calculer les espérances. Nous verrons dans les prochaines sections différentes stratégies pour construire des filtres utilisables dans des contextes réalistes.

3.4 Moindres carrés exacts et pondérés

Nous remplaçons donc dans l'équation (3.19) l'espérance par une somme pondérée. Néanmoins on complexifie un peu l'estimateur de l'erreur en donnant moins de poids aux échantillons n les plus anciens : on appelle cette méthode le Moindres Carrés (MC) Pondérés. Comme les Moindres Carrés (MC) sont un cas particulier des MC pondérés dans lesquels λ vaut 1. On traite donc uniquement le cas des MC pondérés. Dans ce cas, l'erreur pour l'échantillon n s'écrit :

$$E_n = (d(n) - \mathbf{w}^T \mathbf{x}_n)^2 + \lambda(d(n-1) - \mathbf{w}^T \mathbf{x}_{n-1})^2 + \lambda^2(d(n-2) - \mathbf{w}^T \mathbf{x}_{n-2})^2 + \dots + \lambda^{n-1}(d(1) - \mathbf{w}^T \mathbf{x}_1)^2 \quad (3.25)$$

Le premier terme est le carré de l'erreur entre le signal désiré et le dernier échantillon du filtre filtré. $\lambda < 1$ est un facteur d'oubli (typiquement entre 0.9 et 0.99). Plus succinctement, on a le critère les MC pondérés :

$$\begin{aligned} E_n &= \sum_{i=1}^n \lambda^{n-i} (d(i) - \mathbf{w}^T \mathbf{x}_i)^2 \\ &= \sum_{i=1}^n \lambda^{n-i} (d(i)^2 + \mathbf{w}^T \mathbf{x}_i \mathbf{x}_i^T \mathbf{w} - 2d(i)\mathbf{w}^T \mathbf{x}_i) \end{aligned} \quad (3.26)$$

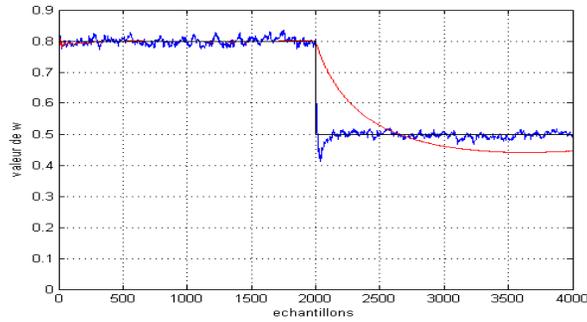


FIGURE 3.8 – Résultat de l'estimation du paramètre $w(k)$: en bleu $\lambda = 0.95$, en rouge $\lambda = 0.999$ et en noir vraie valeur $h(k)$.

En posant

$$\begin{aligned}\sigma_n^2 &= \sum_{i=1}^n \lambda^{n-i} d(i)^2 \\ \mathbf{r}_n &= \sum_{i=1}^n \lambda^{n-i} d(i) \mathbf{x}_i \\ \mathbf{R}_n &= \sum_{i=1}^n \lambda^{n-i} \mathbf{x}_i \mathbf{x}_i^T\end{aligned}\quad (3.27)$$

le critère des MC pondérés se résume à :

$$E_n = \sigma_n^2 - 2\mathbf{w}^T \mathbf{r}_n + \mathbf{w}^T \mathbf{R}_n \mathbf{w} \quad (3.28)$$

En dérivant par rapport à \mathbf{w} et en annulant cette dérivée, on obtient le filtre des MC pondérés à l'instant n :

$$\mathbf{w}_n = \mathbf{R}_n^{-1} \mathbf{r}_n \quad (3.29)$$

avec $\mathbf{w}_n = (\mathbf{w}_n(0), \dots, \mathbf{w}_n(M-1))$. On remarque donc que les coefficients du filtre varient au cours du temps.

L'algorithme 1 résume l'algorithme pour obtenir le filtre des MC pondérés.

Data : d_n et \mathbf{x}_n

Result : \mathbf{w}_n

initialisation : $\mathbf{R}_0 = \epsilon \mathbf{I}$ et $\mathbf{r}_0 = \mathbf{0}$;

for $n = M$ à $n = \text{valeur finale}$ **do**

 acquérir d_n et \mathbf{x}_n ;

$\mathbf{r}_n = d_n \mathbf{x}_n + \lambda \mathbf{r}_{n-1}$;

$\mathbf{R}_n = \mathbf{x}_n \mathbf{x}_n^T + \lambda \mathbf{R}_{n-1}$;

$\mathbf{w}_n = \mathbf{R}_n^{-1} \mathbf{r}_n$;

 estimation : $y_n = \mathbf{w}_n^T \mathbf{x}_n$;

end

Algorithme 1 : Algorithme des Moindres Carrés Pondérés

Maintenant, intéressons nous à l'influence du facteur d'oubli λ . Pour cette étude, nous utiliser un exemple de réalisation des MC pondérés dans lequel un des coefficients $h(k)$ changent de valeur. Le résultat de l'estimation du paramètre équivalent $w(k)$ est montré sur la figure 3.8.

On remarque que plus λ est grand, plus variance de l'estimation est faible (et donc plus proche de la vraie valeur) mais au détriment de l'adaptation aux changements. Si λ est relativement faible, sa variance sera importante mais la vitesse de convergence de l'algorithme sera importante ce qui est intéressant en environnement non stationnaire. Pour information, la durée du transitoire est approximativement égale à $5/(1 - \lambda)$.

La complexité de l'algorithme est très importante et nécessite $M^3/3$ multiplications et additions.

3.5 Moindres carrés récursifs (RLS)

L'objectif est de réduire la complexité de l'algorithme des MC pondérés en calculant de façon récursive $\mathbf{w}_n : \mathbf{w}_n = f(\mathbf{w}_{n-1})$. Le coeur de l'algorithme repose sur le lemme d'inversion matricielle suivant.

Theorem 3.5.1 (Lemme d'inversion matricielle) Soit $\mathbf{A}' = \mathbf{A} + \mathbf{b}\mathbf{c}^T$ avec \mathbf{A} et \mathbf{A}' des matrices inversibles. Alors on a :

$$\mathbf{A}'^{-1} = \mathbf{A}^{-1} - \frac{1}{1 + \mathbf{c}^T \mathbf{A}^{-1} \mathbf{b}} \mathbf{A}^{-1} \mathbf{b} \mathbf{c}^T \mathbf{A}^{-1} \quad (3.30)$$

L'avantage de ce lemme provient que le calcul de \mathbf{A}'^{-1} nécessite $3N^2$ multiplications au lieu de N^3 . Donc si on connaît \mathbf{A}^{-1} , ce lemme est très intéressant.

Quelques nouvelles notations sont données :

- $\mathbf{C}_n = \mathbf{R}_n^{-1}$
- $e_{n|n-1} = d(n) - \mathbf{w}_{n-1}^T \mathbf{x}_n$: erreur de prédiction

Reprenons l'algorithme des MC pondérés avec la récurrence n :

$$\begin{aligned} \mathbf{r}_n &= d(n)\mathbf{x}_n + \lambda \mathbf{r}_{n-1} \\ \mathbf{R}_n &= \mathbf{x}_n \mathbf{x}_n^T + \lambda \mathbf{R}_{n-1} = \lambda \left(\mathbf{R}_{n-1} + \frac{1}{\lambda} \mathbf{x}_n \mathbf{x}_n^T \right) \end{aligned} \quad (3.31)$$

En utilisant le lemme d'inversion matricielle, on obtient :

$$\begin{aligned} \mathbf{C}_n &= \frac{1}{\lambda} \left(\mathbf{C}_{n-1} - \frac{1}{1 + \frac{\mathbf{x}_n^T \mathbf{C}_{n-1} \mathbf{x}_n}{\lambda}} \frac{\mathbf{C}_{n-1} \mathbf{x}_n \mathbf{x}_n^T \mathbf{C}_{n-1}}{\lambda} \right) \\ &= \frac{1}{\lambda} \left(\mathbf{C}_{n-1} - \frac{1}{\lambda + \mathbf{x}_n^T \mathbf{C}_{n-1} \mathbf{x}_n} \mathbf{C}_{n-1} \mathbf{x}_n \mathbf{x}_n^T \mathbf{C}_{n-1} \right) \end{aligned} \quad (3.32)$$

On pose :

$$\begin{aligned} \mathbf{p}_n &= \mathbf{C}_{n-1} \mathbf{x}_n \\ \mu_n &= \mathbf{x}_n^T \mathbf{p}_n = \mathbf{x}_n^T \mathbf{C}_{n-1} \mathbf{x}_n \\ \mathbf{g}_n &= \frac{1}{\lambda + \mu_n} \mathbf{p}_n \end{aligned} \quad (3.33)$$

de sorte que :

$$\mathbf{C}_n = \frac{1}{\lambda} (\mathbf{C}_{n-1} - \mathbf{g}_n \mathbf{x}_n^T \mathbf{C}_{n-1}) \quad (3.34)$$

Alors en utilisant les Eqs. (3.31) et (3.34), on obtient :

$$\begin{aligned} \mathbf{w}_n &= \mathbf{C}_n \mathbf{r}_n \\ &= \frac{1}{\lambda} (\mathbf{C}_{n-1} - \mathbf{g}_n \mathbf{x}_n^T \mathbf{C}_{n-1}) (d(n)\mathbf{x}_n + \lambda \mathbf{r}_{n-1}) \\ &= \frac{1}{\lambda} d(n) \mathbf{C}_{n-1} \mathbf{x}_n + \mathbf{C}_{n-1} \mathbf{r}_{n-1} - \frac{1}{\lambda} \mathbf{g}_n \mathbf{x}_n^T \mathbf{C}_{n-1} \mathbf{x}_n d(n) - \mathbf{g}_n \mathbf{x}_n^T \mathbf{C}_{n-1} \mathbf{r}_{n-1} \\ &= \mathbf{w}_{n-1} - \mathbf{g}_n \mathbf{x}_n^T \mathbf{w}_{n-1} + \frac{1}{\lambda} d(n) \mathbf{p}_n - \frac{1}{\lambda} d(n) \mathbf{g}_n \mu_n \\ &= \mathbf{w}_{n-1} - \mathbf{g}_n \mathbf{x}_n^T \mathbf{w}_{n-1} + \frac{1}{\lambda} d(n) (\lambda + \mu_n) \mathbf{g}_n - \frac{1}{\lambda} d(n) \mathbf{g}_n \mu_n \\ &= \mathbf{w}_{n-1} - \mathbf{g}_n \mathbf{x}_n^T \mathbf{w}_{n-1} + d(n) \mathbf{g}_n \\ &= \mathbf{w}_{n-1} + (d(n) - \mathbf{w}_{n-1}^T \mathbf{x}_n) \mathbf{g}_n \\ &= \mathbf{w}_{n-1} + e_{n|n-1} \mathbf{g}_n \end{aligned} \quad (3.35)$$

L'algorithme 2 résume l'algorithme pour obtenir le filtre des MC pondérés.

Data : $d(n)$ et \mathbf{x}_n

Result : \mathbf{w}_n

initialisation : $\mathbf{C}_0 = \delta \mathbf{I}$ et $\mathbf{w}_0 = \mathbf{0}$;

for $n = M$ à $n =$ valeur finale **do**

 acquérir $d(n)$ et \mathbf{x}_n ;

$\mathbf{p}_n = \mathbf{C}_{n-1} \mathbf{x}_n$;

$\mu_n = \mathbf{x}_n^T \mathbf{p}_n$;

$\mathbf{g}_n = \frac{1}{\lambda + \mu_n} \mathbf{p}_n$;

$e_{n|n-1} = d(n) - \mathbf{w}_{n-1}^T \mathbf{x}_n$;

$\mathbf{w}_n = \mathbf{w}_{n-1} + e_{n|n-1} \mathbf{g}_n$;

$\mathbf{C}_n = \frac{1}{\lambda} (\mathbf{C}_{n-1} - \mathbf{g}_n \mathbf{p}_n^T)$;

 estimation : $y_n = \mathbf{w}_n^T \mathbf{x}_n$;

end

Algorithme 2 : Algorithme RLS forme 1

Malheureusement, pour des erreurs d'arrondi, on a $\mathbf{p}_n^T \neq \mathbf{p}_n$ ce qui entraîne une instabilité de l'algorithme. Pour remédier à ce problème on utilise la forme 2 dans l'algorithme 3.

Data : $d(n)$ et \mathbf{x}_n

Result : \mathbf{w}_n

initialisation : $\mathbf{C}_0 = \delta \mathbf{I}$ et $\mathbf{w}_0 = \mathbf{0}$;

for $n = M$ à $n =$ valeur finale **do**

 acquérir $d(n)$ et \mathbf{x}_n ;

$\mathbf{p}_n = \mathbf{C}_{n-1} \mathbf{x}_n$;

$\mathbf{q}_n = \mathbf{x}_n^T \mathbf{C}_{n-1}$;

$\mu_n = \mathbf{x}_n^T \mathbf{p}_n$;

$\mathbf{g}_n = \frac{1}{\lambda + \mu_n} \mathbf{p}_n$;

$e_{n|n-1} = d(n) - \mathbf{w}_{n-1}^T \mathbf{x}_n$;

$\mathbf{w}_n = \mathbf{w}_{n-1} + e_{n|n-1} \mathbf{g}_n$;

$\mathbf{C}_n = \frac{1}{\lambda} (\mathbf{C}_{n-1} - \mathbf{g}_n \mathbf{q}_n)$;

 estimation : $y_n = \mathbf{w}_n^T \mathbf{x}_n$;

end

Algorithme 3 : Algorithme RLS forme 2

3.6 Filtrage adaptatif par algorithme du gradient (LMS)

Nous rappelons le filtre de Wiener :

$$\mathbf{w}_{wiener} = \mathbf{R}^{-1} \mathbf{r} \quad (3.36)$$

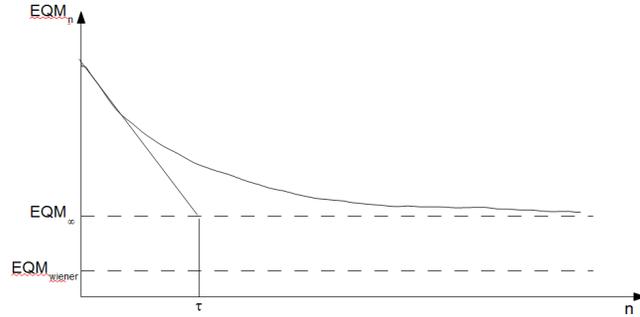
Nous avons vu que ce filtre n'est pas réalisable en pratique et nous avons utilisé différents algorithmes des MC pour estimer ce filtre. Néanmoins le temps de calcul est très important ce qui peut empêcher une implantation en temps réel. Nous allons dans cette section utiliser un algorithme de gradient pour permettre d'approcher le filtre de Wiener : le LMS.

Soit \mathbf{w}_n le filtre adaptatif à l'instant n . L'erreur quadratique d'estimation instantanée vaut :

$$e(n)^2 = (d(n) - \mathbf{w}_n^T \mathbf{x}_n)^2 = d(n)^2 - 2d(n) \mathbf{w}_n^T \mathbf{x}_n + \mathbf{w}_n^T \mathbf{x}_n \mathbf{x}_n^T \mathbf{w}_n = f(w_0, w_1, \dots, w_{M-1}) \quad (3.37)$$

Son gradient par rapport aux composantes de $\mathbf{w}_n = w_n(0), w_n(1), \dots, w_n(M-1)$ est donné par :

$$\begin{aligned} \frac{\partial e(n)^2}{\partial \mathbf{w}_n} &= -2d(n) \mathbf{x}_n + 2 \mathbf{x}_n \mathbf{x}_n^T \mathbf{w}_n^T \\ &= -2(d(n) - \mathbf{w}_n^T \mathbf{x}_n) \mathbf{x}_n \\ &= -2e(n) \mathbf{x}_n \end{aligned} \quad (3.38)$$


 FIGURE 3.9 – Evolution de l' EQM_n en fonction des échantillons n .

L'algorithme 4 résume l'algorithme du LMS.

Data : $d(n)$ et \mathbf{x}_n

Result : \mathbf{w}_n

initialisation : $\mathbf{w}_0 = \mathbf{0}$;

for $n = M$ à $n = \text{valeur finale}$ **do**

 acquérir $d(n)$ et \mathbf{x}_n ;

$e_{n|n-1} = d(n) - \mathbf{w}_{n-1}^T \mathbf{x}_n$;

$\mathbf{w}_n = \mathbf{w}_{n-1} + 2\lambda e_{n|n-1} \mathbf{x}_n$;

 estimation : $y_n = \mathbf{w}_n^T \mathbf{x}_n$;

end

Algorithme 4 : Algorithme LMS

Pour les performances du LMS, nous avons les trois théorèmes suivants.

Theorem 3.6.1 (Convergence du LMS vers le filtre de Wiener) Soit $\sigma_x^2 = E[x_n^2]$, alors on a :

$$E[\mathbf{w}_n] \xrightarrow{n \rightarrow +\infty} \mathbf{w}_{wiener} \quad \text{si } \lambda < \frac{2}{M\sigma_x^2} \quad (3.39)$$

La convergence sera d'autant plus rapide que λ est grand mais comme le montre le théorème une trop grande valeur fera diverger l'algorithme.

Theorem 3.6.2 (EQM) Soit l'EQM en sortie du LMS à l'échantillon n :

$$EQM_n = E[(d(n) - \mathbf{w}_n^T \mathbf{x}_n)^2] \quad (3.40)$$

EQM_n s'approxime par une décroissance exponentielle :

$$EQM_n \approx EQM_0 \exp\left(-\frac{n}{\tau}\right) + EQM_\infty \quad \text{si } \lambda < \frac{2}{M\sigma_x^2} \quad (3.41)$$

Ce résultat peut s'interpréter comme sur la figure 3.9. On voit que l'EQM décroît en fonction du paramètre τ et qu'il atteint un plancher EQM_∞ plus important que l'EQM donnée par le filtre de Wiener.

La différence entre EQM_∞ et EQM_{wiener} est appelée l'excès de bruit. Il est possible d'estimer le coefficient d'excès du bruit et le paramètre τ .

Theorem 3.6.3 — le coefficient d'excès de bruit peut s'approximer :

$$\begin{aligned} C &= \frac{EQM_\infty - EQM_{wiener}}{EQM_{wiener}} \\ &\approx \frac{\lambda}{2} M \sigma_x^2 \end{aligned} \quad (3.42)$$

$$- \tau = \frac{M}{4C} = \frac{1}{2\lambda\sigma_x^2}$$

Donc plus λ est grand plus la constante de temps est faible mais plus l'excès de bruit sera grand.