

Chapter 1

Introduction

Les cours de **Compléments de Télécommunications** et **Questions spéciales en Télécommunications** sont consacrés aux problèmes d'égalisation et d'égalisation adaptative dans divers canaux non idéaux.

Dans le cours de Principes de Télécommunications, nous avons considéré, en majeure partie, le cas d'un canal à bande passante infinie, sans distorsion et à bruit blanc additif gaussien. Nous allons maintenant considérer le cas du canal à bande limitée, dans un premier invariant dans le temps avec comme cas particulier le canal linéaire ; ensuite, nous aborderons le cas du canal variant dans le temps et enfin, le cas particulier d'un bruit multiplicatif.

Nous insisterons principalement sur les méthodes optimales et sous-optimales d'égalisation / détection, et donnerons un aperçu de leurs performances.

Chapter 2

Rappels.

|| Ce chapitre rappelle ou introduit les différents types de modèle de canal que l'on peut rencontrer, ainsi que leur origine physique. La notion d'Interférence entre symboles est présentée et débouche sur le critère de Nyquist qui permet de déterminer les conditions pour éliminer cette interférence entre symboles.

2.1 Les canaux.

2.1.1 Le canal linéaire.

Un des modèles de canal les plus utilisés, et valable dans un grand nombre de cas, consiste simplement à approximer le canal par un **filtre linéaire** ayant une réponse fréquentielle équivalent passe-bas $C(f) \Leftrightarrow c(t)$. Dans ce cas, la réponse impulsionnelle équivalent passe-bas d'un signal $s(t) = \Re[v(t)e^{j2\pi f_c t}]$ sera donnée par $r(t) = \int_{-\infty}^{\infty} v(\tau)c(t-\tau)d\tau + z(t)$ où $z(t)$ représente le bruit additif.

Usuellement, pour des raisons évidentes, le canal est limité à une largeur de bande B Hz, c'est-à-dire $C(f) = 0$ pour $|f| > B$.

Si on exprime la réponse fréquentielle $C(f)$ par :

$$C(f) = |C(f)|e^{j\theta(f)} \quad (2.1)$$

où $|C(f)|$ et $\theta(f)$ sont respectivement l'amplitude et la phase de la réponse fréquentielle. On définit le *décali de groupe* par :

$$\tau(f) = -\frac{1}{2\pi} \frac{d\theta(f)}{df} \quad (2.2)$$

Un canal est dit *idéal* si sa réponse fréquentielle est constante en amplitude et linéaire en phase (dans la bande passante), c'est-à-dire que le décali de groupe est constant. En effet, si on définit un canal idéal, en temporel, comme un gain (ou une atténuation) pur et un décali, la transformée de Fourier de $A\delta(t-\tau)$ étant $Ae^{-j2\pi f\tau}$, la réponse fréquentielle doit bien avoir la caractéristique susdite. Dans le cas contraire, la forme du signal est distordue, soit en amplitude, soit en phase. La conséquence de ceci est une déformation des impulsions qui peuvent avoir une longueur plus importante qu'à l'émission et donc introduire une *interférence entre symboles*. (*ISI : InterSymbol Interference*)

La manière la plus évidente de visualiser celle-ci est de dessiner la réponse impulsionnelle du canal. Dans un cas comme celui indiqué ci-dessous, on observe directement, non seulement la

distorsion introduite, mais également la superposition des symboles successifs qui, aux instants d'échantillonnage T , introduira cette *interférence entre symboles*. Le rôle de l'*égaliseur/organe de décision*, placé derrière le canal, sera de compenser cette ISI, soit, comme dans le cas des *égaliseurs* en essayant de filtrer la sortie de manière à retrouver la forme d'impulsion d'entrée ou, en tous cas, de diminuer l'ISI présente aux instants kT , soit en identifiant le canal et en tenant compte de cette information dans le processus de *décision*.

Application 2.1 Dans le cas d'un délai de groupe $\tau(f)$ linéaire en fréquence, étant donné une impulsion de départ ne donnant pas lieu à *interférence entre symboles*, on constate que la sortie présente bien de l'ISI. La sortie de l'*égaliseur* nous donnera, idéalement, l'*impulsion de source*.

application

Une autre manière d'aborder l'égalisation est d'observer la réponse fréquentielle du canal, par son amplitude et son délai de groupe. Ensuite, on cherche à compenser l'"inidéalité" du canal par un filtre. La figure ci-dessous vous donne les caractéristiques temporelles et fréquentielles typiques d'un canal téléphonique.

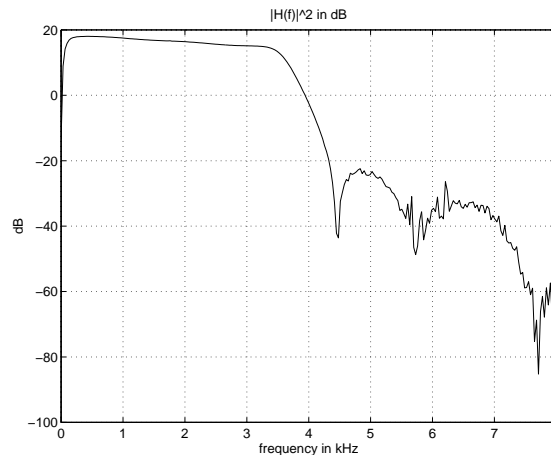


Figure 2.1: Réponse fréquentielle typique d'un canal téléphonique

D'autres types de distorsion peuvent apparaître sur les canaux :

Les distorsions non-linéaires, principalement dues aux non-linéarités des amplificateurs et des mélangeurs, elles sont très difficilement prises en compte au niveau traitement numérique du signal et doivent être réduites le plus possible à la source (d'où l'importance de la conception des amplificateurs et autres éléments électroniques du système d'émission et de réception).

Les décalages fréquentiels, présents dans les systèmes présentant des modulations de type SSB et dus à une démodulation imparfaite (porteuse résiduelle et/ou décalée). Ceux-ci peuvent être compensés soit par une boucle à verrouillage de phase, soit par un traitement numérique adéquat.

Le bruit de phase (phase jitter), qui peut être considéré, dans certains cas, comme une modulation de fréquence de faible indice de modulation. C'est souvent une spécification que les systèmes de réception doivent tenir, de manière à ne pas perturber outre mesure l'égalisation/décision.

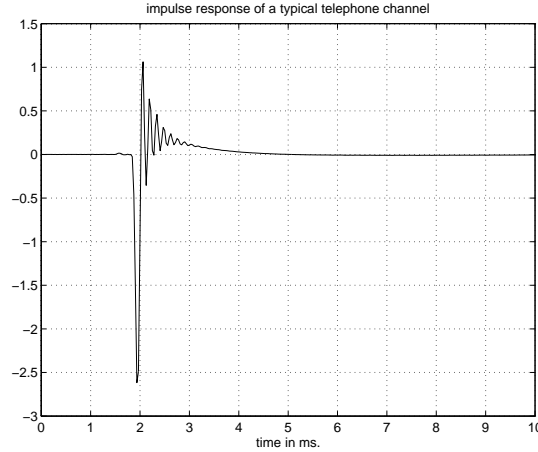


Figure 2.2: Réponse temporelle typique d'un canal téléphonique.

Le **bruit impulsionnel**, additif, qui s'ajoute éventuellement au **bruit blanc** ou coloré.

2.1.2 Le canal multichemins à évanouissement.

Bande et Temps de cohérence d'un canal, diversités.

Les notions de bande de cohérence et de temps de cohérence d'un canal de transmission permettent une mesure des espacements respectivement fréquentiel et temporel nécessaires entre différentes versions d'un même signal émis pour que celles-ci soient non corrélées ; elles constituent par conséquent des caractéristiques essentielles à connaître lors de l'utilisation des techniques de diversité.

Bande de cohérence d'un canal de transmission. Pour introduire le concept de bande de cohérence [Pro89], on considère un canal de transmission caractérisé par un nombre L de chemins conduisant de l'émetteur au récepteur. Le canal peut donc être caractérisé par une réponse impulsionnelle équivalente en bande de base de la forme :

$$h(t; \tau) = \sum_{i=1}^L \beta_i(t) e^{-i2\pi(f_c \tau_i(t) + \phi_i(t))} \delta(\tau - \tau_i(t)) \quad (2.3)$$

où $\beta_i(t)$, $\tau_i(t)$ et $\phi_i(t)$ sont respectivement les atténuation, variation de phase et délai introduits par le $i^{\text{ème}}$ chemin. Toutes ces grandeurs sont susceptibles de varier au cours du temps.

Pour caractériser le canal, nous exprimerons les corrélations entre les divers chemins d'une part et les autocorrélations des paramètres à l'intérieur d'un même chemin d'autre part. En admettant que $h(t, \tau)$ est une fonction aléatoire pour laquelle il y a ergodisme relativement à la fonction de corrélation, on peut écrire sa fonction d'autocorrélation et la transformée de Fourier de celle-ci :

$$\phi_C(\tau; \Delta t) = 0.5 E \{ h(t; \tau) h^*(t + \Delta t; \tau) \} \quad (2.4)$$

$$\Phi_C(\Delta f; \Delta t) = \int_{-\infty}^{\infty} \phi_C(\tau; \Delta t) e^{-i2\pi f \tau} d\tau \quad (2.5)$$

où nous avons fait l'hypothèse vraisemblable que le canal était faiblement stationnaire¹.

Dans la plupart des cas réels, les valeurs de $\phi_C(\tau)$, encore appelé spectre de puissance des délais et de $\Phi_C(\Delta f; \Delta t = 0) = \Phi_C(\Delta f)$ ont l'allure donnée à la figure 2.3 [Pro89]. On appelle T_m l'étalement multichemin et $(\Delta f)_c \simeq \frac{1}{T_m}$ la bande de cohérence du canal.

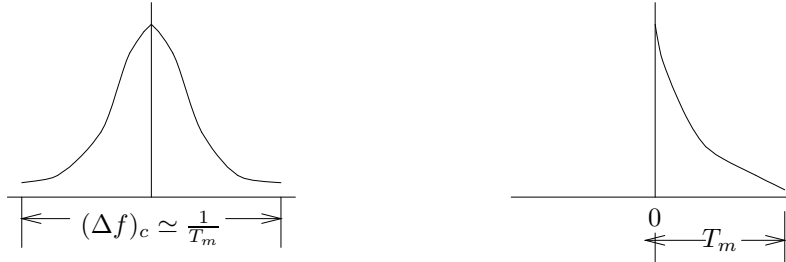


Figure 2.3: Etalement multichemin et bande de cohérence.

Temps de cohérence d'un canal de transmission. Le concept de temps de cohérence est introduit en suivant une démarche similaire. On définit dans un premier temps la fonction suivante, transformée de Fourier de $\Phi_c(\Delta f; \Delta t)$, dont l'allure est donnée en figure 2.4 [Pro89] :

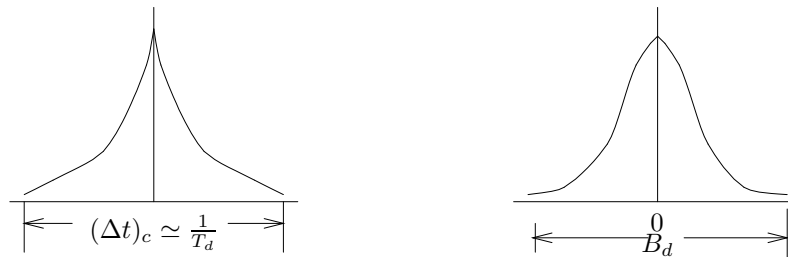


Figure 2.4: Temps de cohérence et bande d'étalement Doppler.

$$\psi_C(\Delta f; \lambda) = \int_{-\infty}^{\infty} \Phi_c(\Delta f; \Delta t) e^{-i2\pi\lambda\Delta t} d\Delta t \quad (2.6)$$

Cette fois-ci, nous nous attacherons à $\psi_C(\Delta f = 0; \lambda) = \psi_C(\lambda)$. La variable λ est appelée fréquence Doppler car elle est liée aux variations Δt qui tiennent compte des modifications de la réponse impulsionnelle du canal de transmission liée aux effets Doppler. B_d est la *bande d'étalement Doppler* du canal. Quant à $(\Delta t)_C \simeq \frac{1}{B_d}$, il est appelé *temps de cohérence* du canal de transmission. Sa signification est la suivante : si deux signaux sont émis à des instants t_1 et t_2 séparés d'une valeur supérieure à $(\Delta t)_C$, ils seront transmis dans un canal présentant deux configurations différentes puisque la réponse impulsionnelle au temps t_1 sera différente de celle

¹Dans le cas contraire, on aurait $\phi_C(\tau_1, \tau_2; \Delta t) = 0.5E\{h(t; \tau_1)h^*(t + \Delta t; \tau_2)\}$, i.e. que les atténuations et les variations de phases associées à des chemins différents sont corrélées.

au temps t_2 . Les effets du passage dans le canal de transmission de ces signaux seront alors non-corrélés.

Si la durée d'émission d'un symbole T_b d'un signal est inférieure à $(\Delta t)_C$, tous les paramètres de gains, de variations de phase et de délais pourront être considérés comme constants. On parle alors de **canal à évanouissement progressif (slowly fading channel)**.

La bande de cohérence d'un canal de transmission est quant à elle une mesure de l'espacement fréquentiel nécessaire pour que deux signaux identiques soient non-corrélés. Dès lors, si un signal émis occupe une bande de fréquence $W = \frac{1}{T_c} < B_C$, il sera seulement caractérisé par trois paramètres : un gain, une variation de phase et un délai. Nous sommes donc en présence d'un seul chemin *résolvable*. Lorsque $W < B_C$, on parle de **canal à évanouissement non sélectif**, sinon, on parle de **canal à évanouissement sélectif**. Le nombre L de chemins nécessaires à la modélisation de l'ensemble des chemins contenus dans le canal est alors donné par :

$$L = \left\lfloor \frac{W}{B_C} \right\rfloor + 1 = \left\lfloor \frac{T_m}{T_c} \right\rfloor + 1 \quad (2.7)$$

Les $L - 1$ chemins supplémentaires sont appelés chemins résolubles. Leur existence peut être mise à profit pour obtenir artificiellement plusieurs versions différentes d'un même signal ; ils permettent donc d'obtenir une forme de *diversité* fréquentielle.

2.2 Le critère de Nyquist.

2.2.1 Notion d'Interférence Entre Symboles (ISI: Inter Symbol Interference).

Dans un premier temps, nous considérons le cas d'un canal linéaire sans bruit additif. Dans le cas classique des modulations linéaires, on peut représenter les signaux de source équivalent passe-bas par :

$$\sum_{n=0}^{\infty} I_n g(t - nT) \quad (2.8)$$

où $\{I_n\}$ est la séquence de symboles d'information et $g(t)$ l'impulsion de mise en forme du signal, à bande passante limitée à B ($g(t) \Leftrightarrow G(f)$ avec $|G(f)| = 0$ pour $|f| > B$). Ce signal est transmis sur un canal ($C(f)$) ayant la même limitation de bande passante. En conséquence, on représente le signal reçu par

$$\sum_{n=0}^{\infty} I_n h(t - nT) + z(t) \quad (2.9)$$

où

$$h(t) = \int_{-\infty}^{\infty} g(t)c(t - \tau)d\tau \quad (2.10)$$

et $z(t)$ représente le bruit additif Gaussien.

La théorie de la décision nous apprend qu'il suffit d'échantillonner la sortie du canal, préalablement filtrée par un filtre adapté (de réponse fréquentielle $H^*(f)$) à un taux de $1/T$ éch./sec. On peut donc écrire

$$y(t) = \sum_{n=0}^{\infty} I_n x(t - nT) + \nu(t) \quad (2.11)$$

où $x(t)$ est la réponse impulsionnelle du filtre $h(t) \otimes h(-t)$ et $\nu(t)$ le bruit blanc filtré par $H^*(f)$. Après échantillonnage, les variables aléatoires de sortie sont :

$$y(kT + \tau_o) = y_k = \sum_{n=0}^{\infty} I_n x(kT - nT + \tau_o) + \nu(kT + \tau_o) \quad (2.12)$$

où τ_o est un délai introduit par le canal.

$$y_k = \sum_{n=0}^{\infty} I_n x_{k-n} + \nu_k \quad k = 0, 1, \dots \quad (2.13)$$

En normalisant x_o à 0, on obtient :

$$y_k = I_k + \sum_{\substack{n=0 \\ n \neq k}}^{\infty} I_n x_{k-n} + \nu_k \quad (2.14)$$

où le second terme représente l'interférence entre symboles (influence sur la sortie à l'instant k des symboles d'indice différent de k) et ν_k représente le bruit additif coloré par le filtre de réception.

2.2.2 Le critère de Nyquist.

L'objectif étant d'avoir $y_k = I_k$, c'est-à-dire d'éliminer l'interférence entre symboles, nous allons aborder le problème d'un point de vue temporel dans un premier temps.

La condition pour éliminer l'ISI dans l'équation (2.14) est :

$$x(t = kT) = x_k = \begin{cases} 1 & k = 0 \\ 0 & k \neq 0 \end{cases} \quad (2.15)$$

Le théorème d'échantillonnage des signaux à bande passante limitée nous apprend que $x(t)$ peut être exprimé par :

$$x(t) = \sum_{n=-\infty}^{\infty} x\left(\frac{n}{2B}\right) \frac{\sin 2\pi B(t - n/2B)}{2\pi B(t - n/2B)} \quad (2.16)$$

et

$$x\left(\frac{n}{2B}\right) = \int_{-B}^B X(f) e^{j2\pi n/2B} df \quad (2.17)$$

ce qui correspond à une période d'échantillonnage $T = \frac{1}{2B}$. Si on fait le choix particulier de ce débit de symboles (i.e. on envoie un symbole I_k toutes les T secondes), on obtient :

$$x(t) = \sum_{n=-\infty}^{\infty} x(nT) \frac{\sin \pi(t - nT)/T}{\pi(t - nT)/T} \quad (2.18)$$

Si l'on veut réduire l'ISI à néant, c'est-à-dire obtenir $x(nT) = 0 \forall n \neq 0$, nous devons adopter :

$$x(t) = \frac{\sin \pi t/T}{\pi t/T} \quad (2.19)$$

soit

$$X(f) = \begin{cases} T & |f| \leq \frac{1}{2T} \\ 0 & |f| > \frac{1}{2T} \end{cases} \quad (2.20)$$

Ce type de signal est de toute évidence impossible à réaliser, puisque l'impulsion $x(t)$ ainsi définie est totalement anti-causale et de durée infinie. D'autre part, la décroissance de $x(t)$ est en $1/t$ et, un échantillonnage non idéal (en un temps $kT + \tau$) auront comme effet pour l'échantillon de se voir affecter d'un terme provenant d'une somme de tous les symboles à un facteur près :

$$\sum_{\substack{n=0 \\ n \neq k}}^{\infty} I_n \frac{\sin \pi(nT + \tau)/T}{\pi(nT + \tau)/T} \quad (2.21)$$

Cette somme peut diverger (ce qui est non physique, comme l'est l'impulsion considérée!). Il est donc naturel d'adopter un débit de symboles $1/T$ inférieur à $2B$.

Dans ce cas, en se souvenant de la propriété de périodisation du spectre des signaux échantillonnés :

$$X_e(f) = \frac{1}{T} \sum_k X\left(f + \frac{k}{T}\right) \quad (2.22)$$

On obtient la relation spectrale entre la sortie et l'entrée du canal de la forme :

$$Y(f) = \frac{1}{T} \sum_k X\left(f + \frac{k}{T}\right) \quad (2.23)$$

Et la condition pour ne pas avoir d'ISI s'exprime par :

$$X_e(f) = T \quad \forall f \quad (2.24)$$

Les impulsions ci-dessous satisfont le critère de Nyquist et sont appelées *impulsions de Nyquist*.

Habituellement, on choisit le débit de symboles $B < 1/T < 2B$. Une impulsion couramment utilisée l'impulsion appelée *en cosinus surélevé*.

$$X(f) = \begin{cases} T & 0 \leq |f| \leq (1 - \beta)/2T \\ \frac{T}{2} [1 - \sin(f - 1/2T)/\beta] & (1 - \beta)/2T \leq |f| \leq (1 + \beta)/2T \end{cases} \quad (2.25)$$

où β est appelé le *paramètre de décroissance*. L'impulsion correspondante à l'allure :

$$x(t) = \frac{\sin \pi t/T}{\pi t/T} \frac{\cos \beta \pi t/T}{1 - 4\beta^2 t^2/T^2} \quad (2.26)$$

La décroissance de l'impulsion est cette fois-ci en $1/t^3$, ce qui conduit à une somme convergente pour l'ISI.

2.2.3 Lien avec le théorème d'échantillonnage.

Dans le cas du théorème d'échantillonnage, on nous apprend que si un signal (passe-bande) a une bande passante B , il faut échantillonner à un taux supérieur à $T = 1/2B$. En d'autres termes, la quantité d'information contenue dans un signal continu est équivalente à la quantité d'information contenue dans un signal échantillonné correctement.

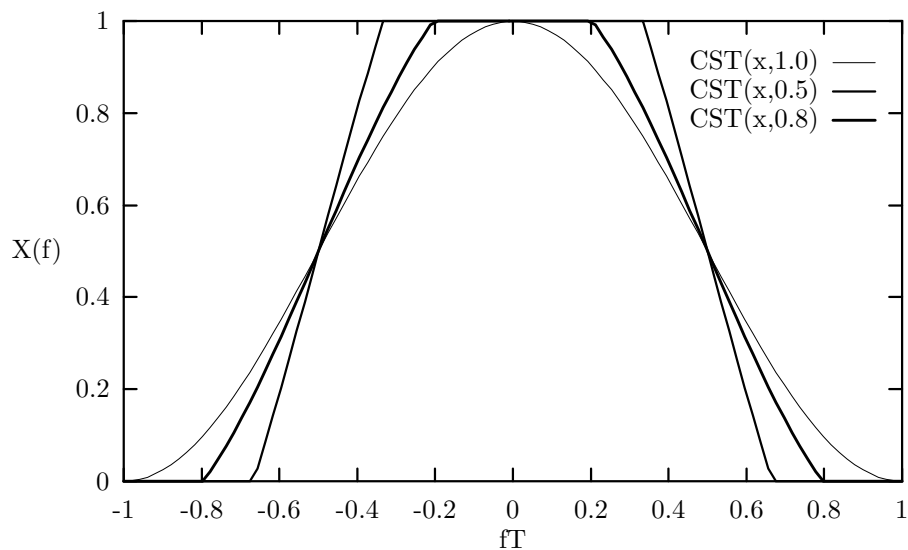


Figure 2.5: Réponse fréquentielle en cosinus surélevé

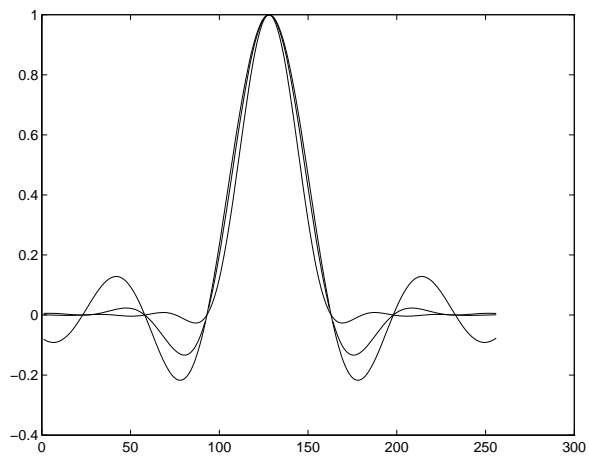


Figure 2.6: Réponse temporelle de filtres en cosinus surélevé

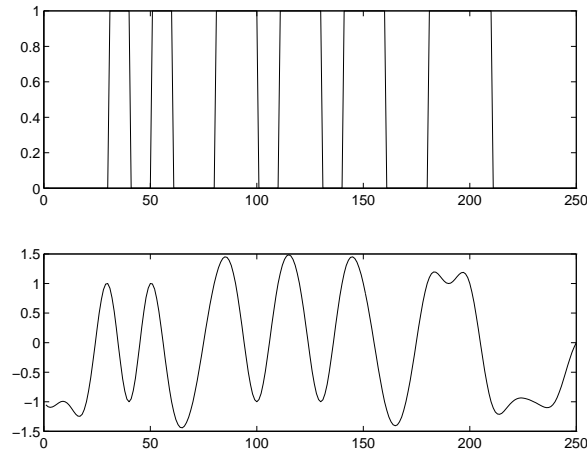


Figure 2.7: Train d'impulsions filtré par un filtre de Nyquist

Dans le cas du critère de Nyquist, on cherche un signal continu qui contienne au moins l'information du signal discret, il est donc logique que l'on doive adopter une bande passante supérieure $B = 1/2T$. Il s'agit en quelque sorte de deux théorèmes duaux.

2.2.4 Le diagramme en oeil.

Une manière classique de caractériser l'ISI est le diagramme en oeil. Il s'agit simplement d'afficher le signal continu $y(t)$ sur un oscilloscope en synchronisant la base de temps sur un multiple de T (habituellement $2T$). La figure obtenue a l'allure d'un oeil dont :

1. L'ouverture verticale reflète la résistance au bruit sous échantillonnage idéal.
2. L'ouverture horizontale reflète la sensibilité au désalignement de l'échantillonnage.
3. L'épaisseur des traits, à l'instant d'échantillonnage idéal, reflète la quantité d'ISI présente. Dans le cas où cette épaisseur est très grande, on a un oeil "fermé". Le rôle d'un égaliseur sera d'"ouvrir" le plus possible cet oeil.
4. La pente de l'oeil indique la sensibilité de l'ouverture verticale à de faibles désalignement d'échantillonnage.

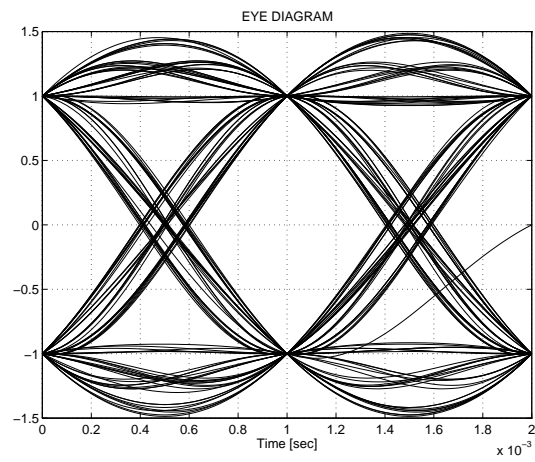


Figure 2.8: Diagrammes en oeil (binaire)

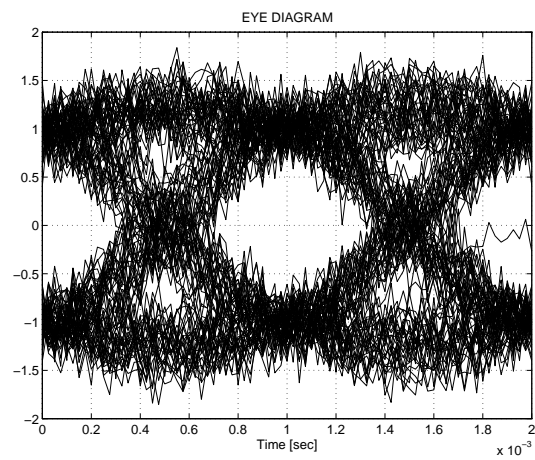


Figure 2.9: Diagrammes en oeil (binaire bruité)

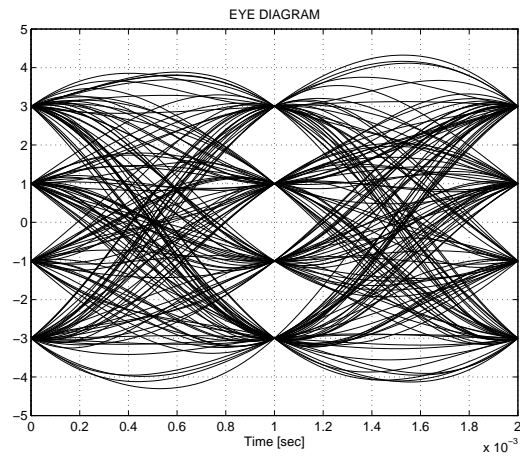


Figure 2.10: Diagrammes en oeil (PAM-4)

Chapter 3

Filtres optimaux et prédiction linéaire.

Avant d'aborder le problème de l'égalisation, il convient de mettre un certain nombre d'outils en place. Ces outils sont omniprésents en traitement de signal, leur compréhension et leur connaissance est donc très importante. Outre quelques rappels, nous abordons le principe du maximum de vraisemblance pour l'estimation de paramètres, le filtrage de Wiener et la prédiction linéaire.

Ce chapitre est inspiré de [Hay91] pour les rappels et le principe du maximum de vraisemblance, de [Slo95] et de [PRLN92] pour le filtrage de Wiener, la prédiction linéaire et les filtres en treillis.

3.1 Processus et modèles stochastiques.

3.1.1 Moyenne, autocorrélation et stationarité.

Soit un processus stochastique discret représenté par la série temporelle $u(n), u(n-1), \dots, u(n-M)$, on définit la moyenne par :

$$\mu(n) = E \{u(n)\} \quad (3.1)$$

où $E \{ \}$ est l'opérateur espérance mathématique. De même, l'autocorrélation prend la forme :

$$r(n, n-k) = E \{u(n)u^*(n-k)\}, \quad k = 0, \pm 1, \pm 2, \dots \quad (3.2)$$

où l'astérisque représente la conjugaison complexe. La fonction d'autocovariance s'écrit :

$$c(n, n-k) = E \{[u(n) - \mu(n)][u(n-k) - \mu(n-k)]^*\}, \quad k = 0, \pm 1, \pm 2, \dots \quad (3.3)$$

Un processus est dit strictement stationnaire si tous ses moments sont indépendants du temps. Un processus est dit faiblement stationnaire, ou stationnaire au sens large, si

- $\mu(n) = \mu \quad \forall n$
- $r(n, n-k) = r(k) \quad \forall n$

On notera au passage que $r(0)$ représente la *valeur quadratique moyenne* ou *puissance* de $u(n)$, tandis que $c(0) = \sigma_u^2$ représente la *variance*.

3.1.2 La matrice de corrélation.

Définissons un *vecteur d'observation* $\mathbf{u}(n)$ tel que :

$$\mathbf{u}^T(n) = [u(n), u(n-1), \dots, u(n-M+1)] \quad (3.4)$$

On définit alors la matrice de corrélation par :

$$\mathbf{R} = \mathbf{E} \left\{ \mathbf{u}(n) \mathbf{u}^H(n) \right\} \quad (3.5)$$

où $\mathbf{u}^H(n)$ représente la transposée hermitienne de $\mathbf{u}(n)$, c'est-à-dire le vecteur transposé conjugué.

En détaillant la matrice de corrélation, on obtient immédiatement :

$$\mathbf{R} = \begin{bmatrix} r(0) & r(1) & \cdots & r(M-1) \\ r(-1) & r(0) & \cdots & r(M-2) \\ \vdots & \vdots & \ddots & \vdots \\ r(-M+1) & r(-M+2) & \cdots & r(0) \end{bmatrix} \quad (3.6)$$

1. *La matrice de corrélation d'un processus stochastique discret stationnaire est hermitienne.*

Une matrice est dite hermitienne si elle est égale à sa transposée hermitienne, i.e.

$$\mathbf{R} = \mathbf{R}^H \quad (3.7)$$

Cette propriété découle directement de la définition de la matrice de corrélation. On peut d'ailleurs aisément vérifier que $r(-k) = r^*(k)$. Dans le cas d'un processus à valeurs réelles, la matrice \mathbf{R} est symétrique.

2. *La matrice de corrélation d'un processus stationnaire discret est une matrice Toeplitz carrée.*

Une matrice carrée est dite Toeplitz si tous les éléments d'une même diagonale ou sous-diagonale sont égaux. On voit directement que c'est le cas ici. D'autre part, cette propriété est directement liée à la propriété de stationnarité (au sens large) du processus. Cette propriété est importante, car elle permet dans bien des cas de simplifier les calculs algébriques.

3. *La matrice de corrélation d'un processus stationnaire discret est toujours définie non négative (et souvent définie positive).*

Soit \mathbf{x} un vecteur complexe quelconque de dimension $M \times 1$. Définissons $y = \mathbf{x}^H \mathbf{u}(n)$ (et donc $y^* = \mathbf{u}^H(n) \mathbf{x}$). La puissance de y est définie par :

$$\begin{aligned} \mathbf{E} \{|y|^2\} &= \mathbf{E} \{yy^*\} \\ &= \mathbf{E} \left\{ \mathbf{x}^H \mathbf{u}(n) \mathbf{u}^H(n) \mathbf{x} \right\} \\ &= \mathbf{x}^H \mathbf{E} \left\{ \mathbf{u}(n) \mathbf{u}^H(n) \right\} \mathbf{x} \\ &= \mathbf{x}^H \mathbf{R} \mathbf{x} \end{aligned} \quad (3.8)$$

Ce qui implique $\mathbf{x}^H \mathbf{R} \mathbf{x} \geq 0$, d'où, par définition, \mathbf{R} est définie semi-positive. En fait, la matrice \mathbf{R} sera singulière principalement si le processus est constitué de K sinusôides avec $K \leq M$.

Application 3.1 Soit un processus constitué d'une sinusôide complexe bruitée

$$u(n) = \alpha \exp(j\omega n) + v(n), \quad n = 0, 1, \dots, N-1 \quad (3.9)$$

où $v(n)$ est une réalisation du bruit blanc ($E\{v(n)v^*(n-k)\} = \sigma_v^2 \delta_k$) à moyenne nulle. La sinusôide et le bruit sont supposés indépendants, ce qui permet d'écrire :

$$\begin{aligned} r(k) &= E\{u(n)u^*(n-k)\} \\ &= \begin{cases} |\alpha|^2 + \sigma_v^2 & k = 0 \\ |\alpha|^2 \exp(j\omega k), & k \neq 0 \end{cases} \end{aligned} \quad (3.10)$$

soit

$$\mathbf{R} = |\alpha|^2 \begin{bmatrix} 1 + \frac{1}{\rho} & \exp(j\omega) & \cdots & \exp(j\omega(M-1)) \\ \exp(-j\omega) & 1 + \frac{1}{\rho} & \cdots & \exp(j\omega(M-2)) \\ \vdots & \vdots & \ddots & \vdots \\ \exp(j\omega(-M+1)) & \exp(j\omega(-M+2)) & \cdots & 1 + \frac{1}{\rho} \end{bmatrix} \quad (3.11)$$

où ρ est le *rapport signal-bruit* défini par $\rho = \frac{|\alpha|^2}{\sigma_v^2}$.

Dans le cas où ce rapport est infini (c'est-à-dire dans le cas sans bruit), \mathbf{R} peut s'écrire :

$$\mathbf{R} = \begin{bmatrix} 1 \\ e^{-j\omega} \\ e^{-2j\omega} \\ \vdots \\ e^{-(M-1)j\omega} \end{bmatrix} \cdot [1e^{j\omega} e^{2j\omega} \dots e^{(M-1)j\omega}] \quad (3.12)$$

Il s'ensuit que la matrice \mathbf{R} est de rang 1. Dans le cas de K sinusôides, nous aurons donc une matrice de rang (au plus égal) à K .

3.1.3 Les innovations.

Soit un processus stationnaire $u(n)$ de séquence d'autocorrélation $r(n)$ et de densité spectrale de puissance (dsp) $S_{uu}(f)$ définie sur $|f| \leq \frac{1}{2}$. On suppose que $S_{uu}(f)$ est réelle et continue pour tout $|f| \leq \frac{1}{2}$. On peut définir

$$S_{uu}(z) = \sum_{m=-\infty}^{\infty} r(m)z^{-m} \quad (3.13)$$

La densité spectrale étant obtenue en évaluant $S_{uu}(z)$ sur le cercle unité. Supposons que $S_{uu}(z)$ est analytique dans une région incluant le cercle unité. On peut alors écrire la série de Laurent:

$$\log S_{uu}(z) = \sum_{m=-\infty}^{\infty} \nu(m)z^{-m} \quad (3.14)$$

ce qui, sur le cercle unité, devient :

$$\log S_{uu}(f) = \sum_{m=-\infty}^{\infty} \nu(m)e^{-j2\pi fm} \quad (3.15)$$

Les coefficients $\nu(m)$ sont donc les coefficients de Fourier de la série de Fourier représentant la fonction périodique $\log S_{uu}(f)$. Donc :

$$\nu(m) = \int_{-\frac{1}{2}}^{\frac{1}{2}} \log S_{uu}(f)e^{-j2\pi fm} df, \quad m = 0; \pm 1, \dots \quad (3.16)$$

$S_{uu}(f)$ étant une fonction réelle et paire, il s'ensuit que

$$\begin{aligned} S_{uu}(z) &= \exp \left[\sum_{m=-\infty}^{\infty} \nu(m)z^{-m} \right] \\ &= \sigma_v^2 H(z)H(z^{-1}) \end{aligned} \quad (3.17)$$

où $\sigma_v^2 = e^{\nu(0)}$ et

$$H(z) = \exp \left[\sum_{m=1}^{\infty} \nu(m)z^{-m} \right], \quad |z| > r_1 \quad (3.18)$$

En évaluant $S_{uu}(z)$ sur le cercle unité, nous obtenons l'expression :

$$S_{uu}(f) = \sigma_v^2 |H(f)|^2 \quad (3.19)$$

Les coefficients $\nu(m)$ sont appelés *coefficients cepstraux* et la séquence $\nu(m)$ est appelée *cepstre* de la séquence d'autocorrélation $r(m)$.

Le filtre $H(z)$ est analytique dans la zone $|z| > r_1 < 1$. On peut donc développer $H(z)$ sous la forme causale (série de Taylor) :

$$H(z) = \sum_{m=0}^{\infty} h_m z^{-m} \quad (3.20)$$

Si on excite l'entrée de ce filtre par un bruit blanc $v(n)$ de puissance σ_v^2 , la sortie sera un processus stationnaire $u(n)$ de densité spectrale de puissance $S_{uu}(f) = \sigma_v^2 |H(f)|^2 = \sigma_v^2 H(z)H(z^{-1})|_{e^{j2\pi f}}$.

De même, si $u(n)$ est un signal stationnaire ayant cette dsp, le fait de passer ce signal dans un filtre $1/H(z)$ nous fournit un bruit blanc à la sortie. On parle de *filtre blanchissant* et la sortie $v(n)$ est appelée *processus d'innovation* associé à $u(n)$

Cette représentation est appelée *représentation de Wold*.

3.1.4 Modèles stochastiques (AR, MA, ARMA).

On restreint $S_{uu}(z)$ à être de la forme :

$$S_{uu}(z) = \sigma_v^2 \frac{B(z)B(z^{-1})}{A(z)A(z^{-1})} \quad r_1 < |z| < r_2 \quad (3.21)$$

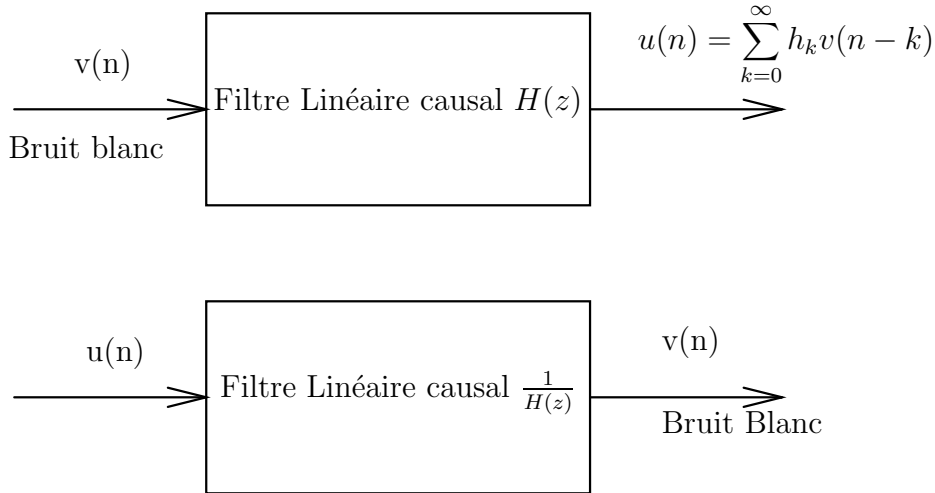


Figure 3.1: Représentation de Wold

où $B(z)$ et $A(z)$ ont leurs racines à l'intérieur du cercle unité. On peut alors écrire :

$$H(z) = \sigma_v^2 \frac{B(z)}{A(z)} = \frac{\sum_{k=0}^q b_k z^{-k}}{1 + \sum_{k=1}^p a_k z^{-k}} \quad r_1 < |z| \quad (3.22)$$

De plus, par construction, $H(z)$ est causal, stable et à phase minimale. Son inverse $1/H(z)$ est également causal, stable et à phase minimale.

On peut exprimer la relation ci-dessus par l'équation aux différences suivante:

$$u(n) + \sum_{k=1}^p a_k u(n-k) = \sum_{k=0}^q b_k v(n-k) \quad (3.23)$$

Processus autorégressif (AR)

Un processus autorégressif est caractérisé par $B(z) = 1$, et donc par l'équation aux différences :

$$u(n) + \sum_{k=1}^p a_k u(n-k) = v(n) \quad (3.24)$$

On peut également représenter ce processus par la figure suivante.

Exercice 3.1 Dans l'exemple de processus AR donnés, justifiez (par les pôles et zéros) l'allure des courbes.

Processus à moyenne mobile (MA : Moving Average)

Un processus autorégressif est caractérisé par $A(z) = 1$, et donc par l'équation aux différences :

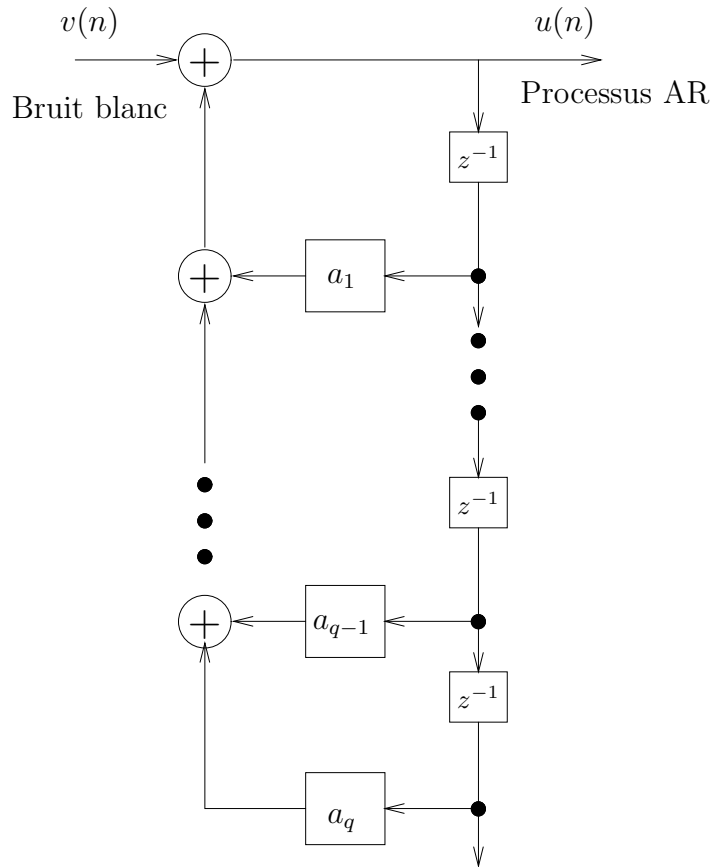
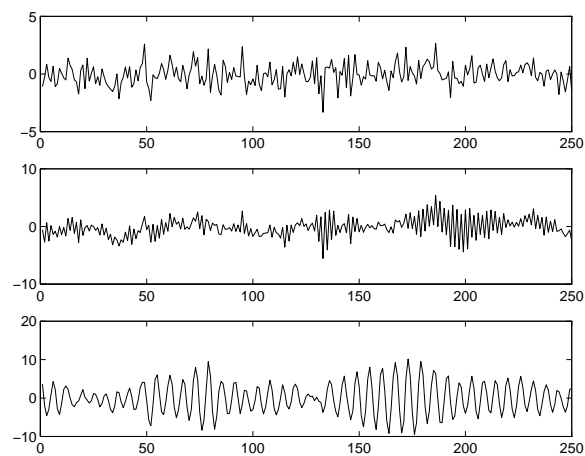


Figure 3.2: Filtre générateur de processus AR

Figure 3.3: Exemples de processus AR. Bruit blanc ; $A=[1 \ 0.1 \ -0.8]$; $A = [1 \ -0.975 \ 0.95]$

$$u(n) = \sum_{k=0}^q b_k v(n-k) \quad (3.25)$$

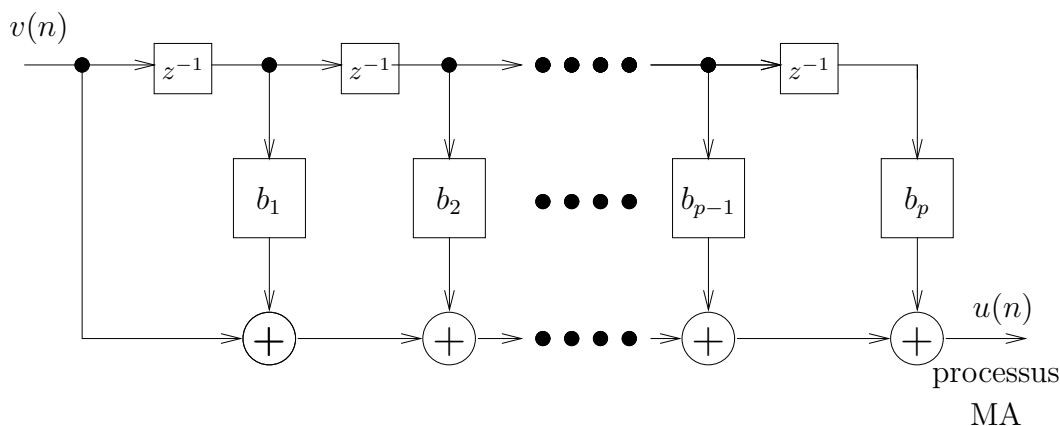
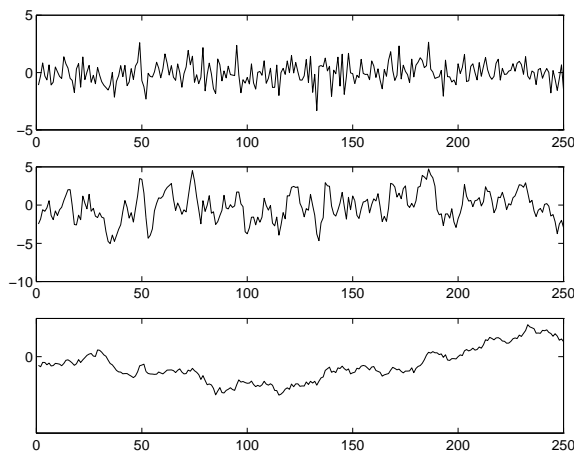


Figure 3.4: Filtre générateur de processus MA

Figure 3.5: Exemples de processus MA. Bruit blanc, $B = [1 \ 1 \ 1 \ 1]/4$; $B = [1 \ (100 \text{ éléments}) \ 1]/100$

Processus autorégressif à moyenne mobile (ARMA)

C'est le processus général décrit ci-dessus.

3.1.5 Les équations de Yule-Walker.

La représentation de Wold nous apprend qu'il y a relation biunivoque entre la densité spectrale de puissance d'un processus stationnaire $u(n)$ et sa représentation par un bruit blanc filtré (AR, MA, ARMA). D'autre part, nous savons que la dsp est reliée de la même manière à

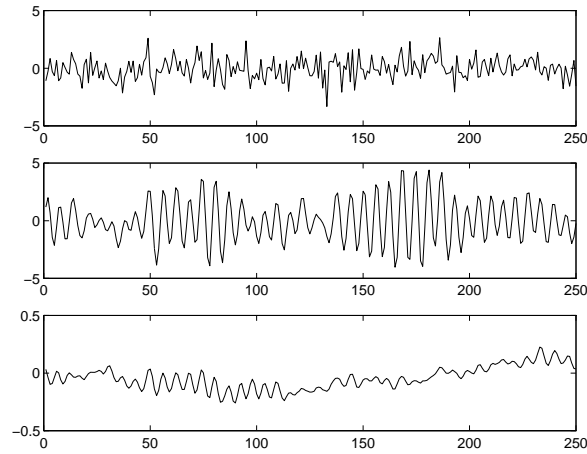


Figure 3.6: Exemples de processus AR. Bruit blanc, $A=[1 \ -0.975 \ 0.95]$ B comme ci-dessus

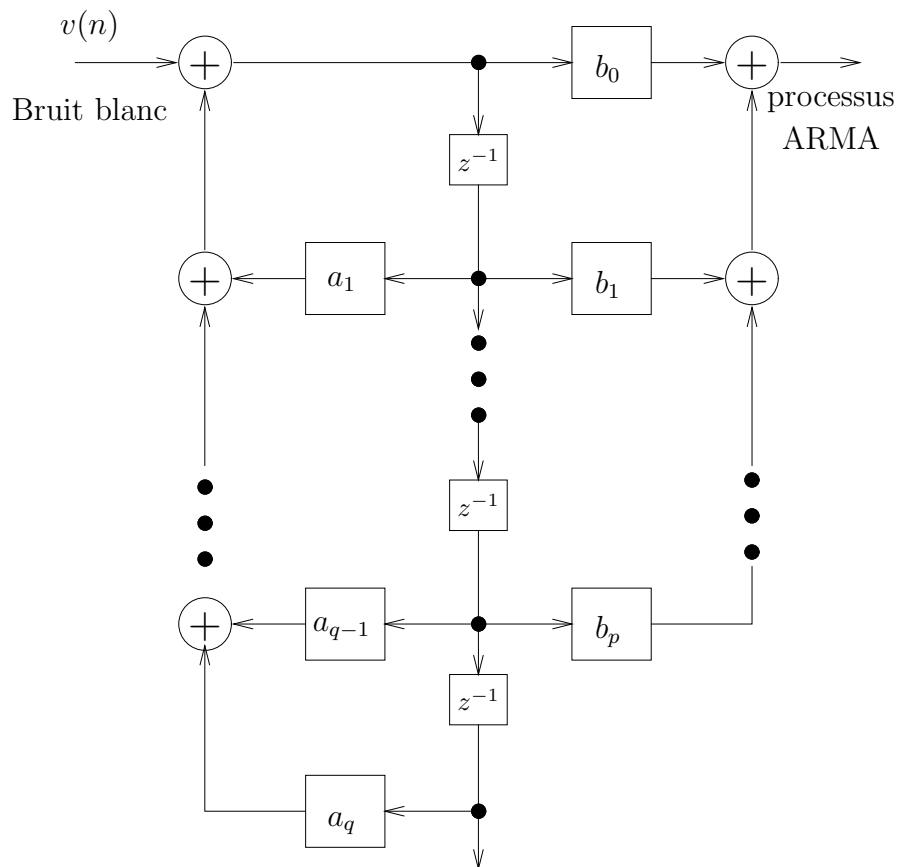


Figure 3.7: Filtre générateur de processus ARMA

la séquence d'autocorrélation. Cette section fera le lien entre les paramètres des filtres et la matrice d'autocorrélation.

Pour ce faire, il suffit d'écrire l'autocorrélation, en tenant compte de l'équation aux différences d'un processus ARMA :

$$\begin{aligned} \mathbb{E}\{u(n)u^*(n-m)\} &= -\sum_{k=1}^p a_k \mathbb{E}\{u(n-k)u^*(n-m)\} \\ &\quad + b_k \sum_{k=0}^q \mathbb{E}\{v(n-k)u^*(n-m)\} \end{aligned} \quad (3.26)$$

Soit

$$r_{uu}(m) = -\sum_{k=1}^p a_k r_{uu}(m-k) + \sum_{k=0}^q b_k r_{vu}(m-k) \quad (3.27)$$

Le terme de cross-corrélation $r_{uv}(m)$ peut s'écrire en fonction du filtre $H(z)$ par (en se rappelant que $v(n)$ est blanc) :

$$\begin{aligned} r_{vu}(m) &= \mathbb{E}\{u^*(n)v(n+m)\} \\ &= \mathbb{E}\left\{\sum_{k=0}^{\infty} h_k u^*(n-k)u(n+m)\right\} \\ &= \sigma_v^2 h_{-m} \quad m \geq 0 \end{aligned} \quad (3.28)$$

Les relations entre la séquence d'autocorrélation et les coefficients des filtres du processus ARMA peuvent donc s'écrire :

$$r_{uu}(m) = \begin{cases} -\sum_{k=1}^p a_k r_{uu}(m-k), & m > q \\ -\sum_{k=1}^p a_k r_{uu}(m-k) + \sigma_v^2 \sum_{k=0}^{q-m} h_k b_{k+m}, & 0 \leq m \leq q \\ r_{uu}^*(-m) & m < 0 \end{cases} \quad (3.29)$$

Dans le cas d'un processus AR, ces équations se simplifient comme suit :

$$r_{uu}(m) = \begin{cases} -\sum_{k=1}^p a_k r_{uu}(m-k), & m > 0 \\ -\sum_{k=1}^p a_k r_{uu}(m-k) + \sigma_v^2, & m = 0 \\ r_{uu}^*(-m) & m < 0 \end{cases} \quad (3.30)$$

Ce sont les **équations de Yule-Walker** qui s'écrivent sous forme matricielle :

$$\boxed{\begin{bmatrix} r(0) & r^*(1) & \cdots & r^*(p) \\ r(1) & r(0) & \cdots & r^*(p-1) \\ \vdots & \vdots & \ddots & \vdots \\ r(p) & r(p-1) & \cdots & r(0) \end{bmatrix} \begin{bmatrix} 1 \\ a_1 \\ \vdots \\ a_p \end{bmatrix} = \begin{bmatrix} \sigma_v^2 \\ 0 \\ \vdots \\ 0 \end{bmatrix}} \quad (3.31)$$

Nous verrons plus loin qu'il existe un algorithme rapide pour résoudre ce type de problème.

3.2 Le principe du maximum de vraisemblance.

Le problème de l'égalisation est un cas particulier de la théorie de l'estimation. Celle-ci traite de la manière d'estimer des paramètres d'un modèle à partir de données issues d'un processus stochastique. Les grandeurs importantes dans ce problème sont la moyenne du paramètre estimé. On dira qu'un estimateur est *non biaisé* si sa moyenne est égale au paramètre vrai. D'autre part, il est clair que nous désirons obtenir un estimateur dont la variance est minimale.

Dans ce cadre, il est bon d'introduire le critère du maximum de vraisemblance et de montrer en quoi il est le critère optimal. Nous introduirons également la borne de Cramer-Rao, qui donne la variance minimale d'un estimateur non-biaisé (dont la moyenne est le paramètre à estimer).

3.2.1 La fonction de vraisemblance.

Le principe du maximum de vraisemblance est assez simple : étant donné un ensemble de données, on désire trouver le paramètre le plus *probable*.

Soit $f_u(\mathbf{u}|\boldsymbol{\theta})$, la *densité de probabilité conditionnelle* du vecteur aléatoire \mathbf{U} dont le vecteur $\mathbf{u} = [u_1 u_2 \cdots u_M]$ est une réalisation et $\boldsymbol{\theta}$ est le *vecteur de paramètres* $[\theta_1 \theta_2 \cdots \theta_K]$. L'objectif étant de sélectionner les valeurs de $\boldsymbol{\theta}$ qui sont les plus probables étant donné les mesures \mathbf{u} , cela conduit tout naturellement à choisir le vecteur $\boldsymbol{\theta}$ qui maximise la densité de probabilité conditionnelle $f_u(\mathbf{u}|\boldsymbol{\theta})$.

On donne le nom de *fonction de vraisemblance* (*likelihood function*), que l'on note $l(\boldsymbol{\theta})$ à la densité de probabilité conditionnelle $f_u(\mathbf{u}|\boldsymbol{\theta})$.

Même si ces deux fonctions recouvrent la même réalité, dans le premier cas, (fonction de vraisemblance) on considère que c'est une fonction de $\boldsymbol{\theta}$, \mathbf{u} étant fixé (il s'agit bien de trouver $\boldsymbol{\theta}$ pour une observation donnée) alors que dans le second cas, $\boldsymbol{\theta}$ est un paramètre fixé et \mathbf{u} est variable.

Habituellement, on travaille avec le logarithme de la fonction de vraisemblance :

$$L(\boldsymbol{\theta}) = \ln[l(\boldsymbol{\theta})] = \ln[f_u(\mathbf{u}|\boldsymbol{\theta})] \quad (3.32)$$

Le logarithme de $l(\boldsymbol{\theta})$ étant une transformation monotone de $l(\boldsymbol{\theta})$, il est équivalent de maximiser $l(\boldsymbol{\theta})$ ou de maximiser $L(\boldsymbol{\theta})$.

La procédure est alors (théoriquement) simple, il suffit de déterminer :

$$\frac{\partial L}{\partial \boldsymbol{\theta}} = \begin{bmatrix} \frac{\partial L}{\partial \theta_1} \\ \frac{\partial L}{\partial \theta_2} \\ \vdots \\ \frac{\partial L}{\partial \theta_K} \end{bmatrix} = 0 \quad (3.33)$$

On suppose ici que la fonction de vraisemblance ne présente qu'un seul maximum, dans le cas contraire, il suffit de prendre les précautions d'usage.

On appelle $\mathbf{s}(\boldsymbol{\theta}) = \frac{\partial L}{\partial \boldsymbol{\theta}}$ la *fonction de score*.

Pour caractériser la qualité de l'estimateur, il faudrait déterminer sa moyenne et sa variance. Cependant, ce calcul n'est pas toujours évident, on introduit donc une borne inférieure pour

les estimateurs non biaisés, par rapport à laquelle tout estimateur peut être comparé. Dans le cas de l'estimateur ML (maximum-likelihood) s'en approche asymptotiquement.

3.2.2 Inégalité de Cramer-Rao

On introduit la *matrice d'information de Fisher* \mathbf{J}^1 ,

$$J(\boldsymbol{\theta}) = \mathbb{E} \left\{ \mathbf{s}(\boldsymbol{\theta}) \mathbf{s}^T(\boldsymbol{\theta}) \right\} = -\mathbb{E} \left\{ \frac{\partial}{\partial \boldsymbol{\theta}} \left(\frac{\partial}{\partial \boldsymbol{\theta}} L(\boldsymbol{\theta}) \right)^T \right\} \quad (3.34)$$

où chaque élément de cette matrice peut s'écrire :

$$J_{ij} = -\mathbb{E} \left\{ \frac{\partial^2}{\partial \theta_i \partial \theta_j} L(\boldsymbol{\theta}) \right\} \quad (3.35)$$

Alors, la covariance de l'erreur d'estimation est bornée par l'inverse de la matrice de Fisher ² :

$$\mathbf{C} = \mathbb{E} \left\{ (\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})^T \right\} \geq \mathbf{J}^{-1} \quad (3.36)$$

En particulier,

$$\mathbb{E} \left\{ (\hat{\theta}_i - \theta_i)^2 \right\} \geq (J^{-1})_{ii} \quad (3.37)$$

Si on peut trouver un estimateur non biaisé qui atteint la borne de Cramer-Rao, il est l'estimateur non biaisé de variance minimale et on dit qu'il est *efficace* (efficient estimator)..

Quelques propriétés remarquables.

- L'estimateur ML est *consistant*, c'est-à-dire que si la fonction de score est nulle, $\hat{\boldsymbol{\theta}}$ converge vers $\boldsymbol{\theta}$ en probabilité ³ si la taille du vecteur d'observation M tend vers l'infini.
- L'estimateur ML est *asymptotiquement efficace*, i.e.

$$\lim_{M \rightarrow \infty} \mathbf{C} = \mathbf{J}^{-1} \quad (3.38)$$

- L'estimateur ML est *asymptotiquement gaussien*.

3.3 Estimateur MMSE et principe d'orthogonalité.

Une autre approche de la théorie de l'estimation est l'approche de Bayes. Soit un paramètre θ à estimer, on introduit une fonction de coût $\mathcal{C}(\theta, \hat{\theta})$, où le paramètre estimé est $\hat{\theta}(\mathbf{U})$. θ et $\hat{\theta}(\mathbf{U})$ étant des variables aléatoires, nous minimiserons l'espérance mathématique de ce coût, appelé *risque de Bayes*.

$$\mathcal{R}(\hat{\theta}(\cdot)) = \mathbb{E} \left\{ \mathcal{C}(\theta, \hat{\theta}(\mathbf{U})) \right\} = \mathbb{E}_{\theta, \mathbf{U}} \left\{ \mathcal{C}(\theta, \hat{\theta}(\mathbf{U})) \right\}. \quad (3.39)$$

L'estimateur est donc la fonction $\hat{\theta}(\cdot)$ qui minimise le risque de Bayes :

¹pour plus de détails, voir Scharf, p. 223

²Une matrice \mathbf{A} est supérieure à \mathbf{B} si $\mathbf{A} - \mathbf{B} \geq 0$, c'est-à-dire si $\mathbf{A} - \mathbf{B}$ est définie semi-positive

³i.e. $\Pr_{U|\theta} \{ \lim_{M \rightarrow \infty} \hat{\boldsymbol{\theta}}(\mathbf{U}) = \boldsymbol{\theta} \} = 1, \quad \forall \boldsymbol{\theta}$

$$\hat{\theta}(\cdot) = \arg \min_{\hat{\theta}(\cdot)} \mathcal{R}(\hat{\theta}(\cdot)) \quad (3.40)$$

De plus, on peut montrer que :

$$\min_{\hat{\theta}(\cdot)} \mathcal{R}(\hat{\theta}(\cdot)) = \min_{\hat{\theta}(\mathbf{U})} \mathcal{R}(\hat{\theta}(\mathbf{U})|\mathbf{U}) \quad (3.41)$$

En d'autres mots, pour minimiser le risque global $\mathcal{R}(\hat{\theta}(\cdot))$, c'est-à-dire minimiser une fonction par rapport à la fonction $\hat{\theta}(\cdot)$, il suffit de minimiser le risque conditionnel $\mathcal{R}(\hat{\theta}(\mathbf{U})|\mathbf{U})$ par rapport à $\hat{\theta}(\mathbf{U})$, qui est un nombre.

Le critère MMSE.

Un cas particulier largement utilisé est le critère de l'erreur quadratique moyenne minimale (Minimum Mean Squared Error) où la fonction de coût est : $\mathcal{C}_{MMSE}(\tilde{\theta}) = |\tilde{\theta}|^2$ avec $\tilde{\theta} = \theta - \hat{\theta}(\mathbf{U})$.

Dans ce cas, on obtient :

$$\min_{\hat{\theta}(\mathbf{U})} \mathcal{R}_{MMSE}(\hat{\theta}(\mathbf{U})|\mathbf{U}) = \min_{\hat{\theta}(\mathbf{U})} \int_{-\infty}^{\infty} f(\theta|\mathbf{U}) |\theta - \hat{\theta}(\mathbf{U})|^2 d\theta \quad (3.42)$$

La minimisation nous donne :

$$\frac{\partial}{\partial \hat{\theta}} \mathcal{R}_{MMSE}(\hat{\theta}(\mathbf{U})|\mathbf{U}) = 2 \int_{-\infty}^{\infty} f(\theta|\mathbf{U}) (\theta - \hat{\theta}(\mathbf{U})) d\theta = 0 \quad (3.43)$$

Ce qui peut encore s'écrire :

$$\hat{\theta}(\mathbf{U}) \underbrace{\int_{-\infty}^{\infty} f(\theta|\mathbf{U}) d\theta}_{=1} = \int_{-\infty}^{\infty} \theta f(\theta|\mathbf{U}) d\theta \quad (3.44)$$

D'où

$$\hat{\theta}_{MMSE}(\mathbf{U}) = \mathbb{E} \{ \theta | \mathbf{U} \} \quad (3.45)$$

L'estimateur est donc la moyenne a posteriori de θ si \mathbf{U} . Pour s'assurer qu'il s'agit bien d'un minimum, on vérifie aisément que :

$$\frac{\partial^2}{\partial^2 \hat{\theta}} \mathcal{R}_{MMSE}(\hat{\theta}|\mathbf{U}) = 2 \int_{-\infty}^{\infty} f(\theta|\mathbf{U}) d\theta = 2 > 0 \quad (3.46)$$

De plus, cette dernière équation nous indique que nous sommes en présence d'un minimum global.

Cet estimateur peut aisément s'étendre au cas de paramètres vectoriels.

Le principe d'orthogonalité pour les estimateurs MMSE.

Ce principe est extrêmement important, car il permet d'obtenir l'estimateur MMSE autrement qu'en calculant l'espérance a posteriori du paramètre, ce qui n'est pas toujours aisé, puisque cela demande en général la connaissance de la distribution de probabilité conditionnelle de celui-ci.

Celui-ci s'exprime par :

$$\hat{\theta}(\mathbf{U}) = \mathbb{E} \{ (\theta | \mathbf{U}) \} \Leftrightarrow \mathbb{E} \{ ((\theta - \hat{\theta}(\mathbf{U}))g(\mathbf{U})) \} = 0, \quad \forall g(\cdot) \quad (3.47)$$

où $g(\cdot)$ est une fonction scalaire.

Démonstration : \Rightarrow Supposons $\hat{\theta}(\mathbf{U}) = E\{(\theta|\mathbf{U})\}$, alors, pour n'importe quel $g(\cdot)$:

$$\begin{aligned} E\{\hat{\theta}(\mathbf{U})g(\mathbf{U})\} &= \int \int f_{\theta, \mathbf{U}}(\theta, \mathbf{U})g(\mathbf{U})d\theta d\mathbf{U} \int v f_{v=\theta|\mathbf{U}}(v|\mathbf{U})dv \\ &= \int f_{\mathbf{U}}(\mathbf{U})g(\mathbf{U})d\mathbf{U} \int v f_{\theta|\mathbf{U}}(v|\mathbf{U})dv \underbrace{\int f_{\theta|\mathbf{U}}(\theta|\mathbf{U})d\theta}_{=1} \\ &= \int \int v f_{\theta, \mathbf{U}}(v, \mathbf{U})g(\mathbf{U}) = E\{\theta g(\mathbf{U})\}. \end{aligned} \quad (3.48)$$

L'autre implication se démontre d'une manière similaire.

Une interprétation géométrique sera faite dans le cadre du filtrage linéaire optimal ci-dessous.

3.4 Filtre de Wiener.

Une classe importante de problèmes peut s'énoncer de la manière suivante : soit un signal $s(n)$, corrompu par du bruit additif $v(n)$, on recherche un filtre qui permette de transformer ce signal $u(n) = s(n) + v(n)$ en un autre $y(n)$, similaire en un certain sens à un signal donné par ailleurs ($d(n)$).

Cela correspond au schéma bloc (3.8).

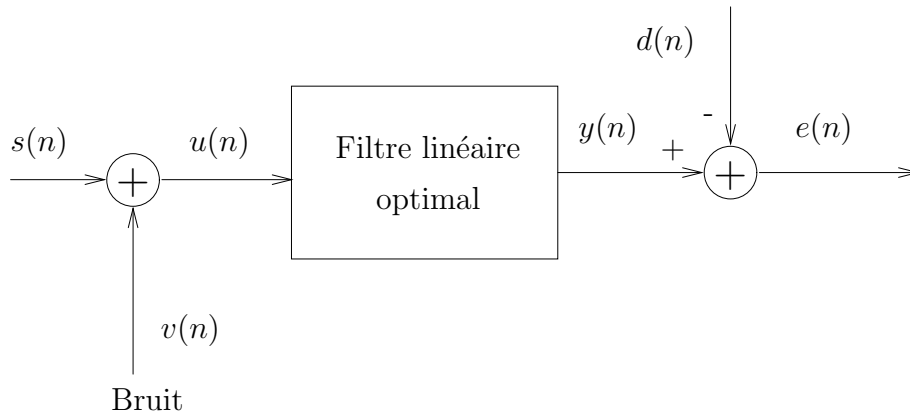


Figure 3.8: Modèle pour le filtrage linéaire optimal.

Dans le cas de l'égalisation, $u(n)$ est la sortie du canal, $v(n)$ est le bruit additif, dont on fait généralement l'hypothèse qu'il est blanc et gaussien, et $d(n)$ représente les données que l'on devrait idéalement recevoir. Il paraît évident que le filtre de Wiener sera dans ce cas une approximation de l'inverse du canal.

Le schéma bloc (3.9) illustre cette manière de voir.

Hypothèses

On suppose que les signaux $u(n)$, $d(n)$ et $v(n)$ sont stationnaires et mutuellement stationnaires au sens large, et de moyenne nulle.

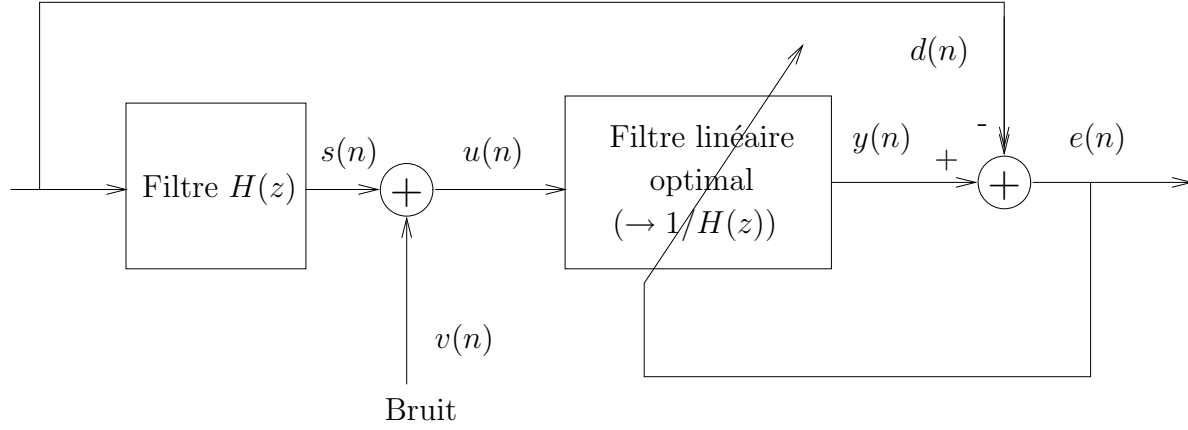


Figure 3.9: Modèle pour l'égalisation linéaire.

3.4.1 Filtre de Wiener non-causal

On adoptera le critère MMSE, de plus, comme le filtre que l'on utilise est un filtre linéaire, on parle plus précisément de LMMSE (Linear MMSE). En effet, dans le cas du critère MMSE, la fonction qui transforme $u(n)$ en $y(n)$ est généralement non linéaire (on parle de *filtrage non linéaire*).

L'équation de minimisation s'écrit donc (avec $\hat{d}(n) = y(n)$) :

$$\min_{h_{k,n}} E \left\{ |d(k) - \hat{d}(k)|^2 \right\} = \min_{h_{k,n}} E \left\{ \left| d(k) - \sum_{n=-\infty}^{\infty} h_{k,n} u(k-n) \right|^2 \right\} \quad (3.49)$$

Où $h_{k,n}$ sont les coefficients du filtre à l'instant k , rien ne nous indiquant pour l'instant que ce filtre est invariant dans le temps.

Pour obtenir l'expression du filtre, nous allons exploiter le principe d'orthogonalité des estimateurs MMSE :

$$E \left\{ (d(k) - \hat{d}(k)) u^*(k-m) \right\} = 0, \quad \forall m \quad (3.50)$$

Soit :

$$\begin{aligned} E \left\{ \hat{d}(k) u^*(k-m) \right\} &= E \left\{ \sum_{n=-\infty}^{\infty} h_{k,n} u(n) u^*(k-m) \right\} \\ &= E \left\{ d(k) u^*(k-m) \right\}, \quad \forall m \end{aligned} \quad (3.51)$$

Ce qui, accompagné des hypothèses de stationarité devient :

$$\sum_{n=-\infty}^{\infty} h_{k,n} r_{uu}(m+k-n) = r_{du}(m) \quad \forall m \quad (3.52)$$

En faisant la substitution et $n-k \rightarrow n$, on obtient :

$$\sum_{n=-\infty}^{\infty} h_{k,n+k} r_{uu}(m-n) = r_{du}(m) \quad \forall m \quad (3.53)$$

Où la dépendance par rapport au temps (à l'indice k) disparaît (i.e. la dépendance par rapport à k n'apparaît plus que dans les coefficients du filtre, donc les équations sont valable pour toute valeur de k et les valeurs des coefficients sont donc indépendantes de k), soit en posant $h_{k,n+k} = h_n$:

$$\sum_{n=-\infty}^{\infty} h_n r_{uu}(m-n) = r_{du}(m) \quad \forall m \quad (3.54)$$

Ce sont les *équations normales*.

Dans le cas tout-à-fait général que nous traitons ici, nous sommes confrontés à une infinité d'équations en une infinité d'inconnues h_n . Cependant, le membre gauche de l'équation (3.54) est un produit de convolution. En définissant :

$$S_{du}(z) = \sum_{n=-\infty}^{\infty} r_{du}(n)z^{-n}, \text{ et } H(z) = \sum_{n=-\infty}^{\infty} h_n z^{-n}. \quad (3.55)$$

l'équation (3.54) devient $H(z)S_{uu}(z) = S_{du}(z)$.

Soit :

$$H(z) = \frac{S_{du}(z)}{S_{uu}(z)} \quad (3.56)$$

En notant l'erreur quadratique minimale (le MMSE) ξ , et par le principe d'orthogonalité :

$$\begin{aligned} \mathbb{E} \left\{ (d(k) - \hat{d}(k)) \hat{d}^*(k) \right\} &= \sum_{n=-\infty}^{\infty} h_n^* \underbrace{\mathbb{E} \left\{ (d(k) - \hat{d}(k)) u^*(k-n) \right\}}_{=0} = 0 \\ \Rightarrow \mathbb{E} \left\{ d(k) \hat{d}^*(k) \right\} &= \mathbb{E} \left\{ |\hat{d}(k)|^2 \right\} \end{aligned} \quad (3.57)$$

Ce qui permet de déterminer :

$$\begin{aligned} \xi &= \mathbb{E} \left\{ |d(k) - \hat{d}(k)|^2 \right\} \\ &= \mathbb{E} \left\{ (d(k) - \hat{d}(k)) \hat{d}^*(k) \right\} - \underbrace{\mathbb{E} \left\{ (d(k) - \hat{d}(k)) \hat{d}^*(k) \right\}}_{=0} \\ &= \mathbb{E} \left\{ |d(k)|^2 \right\} - \mathbb{E} \left\{ d(k) \hat{d}^*(k) \right\} = \mathbb{E} \left\{ |d(k)|^2 \right\} - \mathbb{E} \left\{ |\hat{d}(k)|^2 \right\} \leq \mathbb{E} \left\{ |d(k)|^2 \right\} \end{aligned} \quad (3.58)$$

ou encore, pour permettre le calcul :

$$\begin{aligned} \xi &= \mathbb{E} \left\{ |d(k)|^2 \right\} - \mathbb{E} \left\{ d(k) \hat{d}^*(k) \right\} \\ &= r_{dd}(0) - \sum_{n=-\infty}^{\infty} h_n^* \mathbb{E} \left\{ u^*(k-n) d(k) \right\} \\ &= r_{dd}(0) - \sum_{n=-\infty}^{\infty} h_n^* \mathbb{E} \left\{ r_{du}(n) \right\} \end{aligned} \quad (3.59)$$

En se rappelant qu'on peut retrouver chaque élément de la séquence d'autocorrélation par la formule :

$$r_{dd}(k) = \frac{1}{2\pi j} \oint_C S_{dd}(z) z^{k-1} dz \quad (3.60)$$

où l'intégrale circulaire est calculée dans la région de convergence de $S_{dd}(z)$. Il s'ensuit que :

$$\sigma_d^2 = r_{dd}(0) = \frac{1}{2\pi j} \oint_C S_{dd}(z) z^{-1} dz \quad (3.61)$$

d'autre part, par le théorème de Parseval, on a :

$$\sum_{n=-\infty}^{\infty} h_n^* \mathbb{E} \{r_{du}(n)\} = \frac{1}{2\pi j} \oint_C H(z) S_{du}(z^{-1}) z^{-1} dz \quad (3.62)$$

En combinant ces deux équations, on obtient :

$$\xi_{nc} = \frac{1}{2\pi j} \oint_C [S_{dd}(z) - H_{nc}(z) S_{du}(z^{-1})] z^{-1} dz \quad (3.63)$$

Application 3.2 Filtre de Wiener non causal pour un processus AR(1)

Soit un signal $u(n) = s(n) + v(n)$ où $s(n)$ est un processus AR(1) décrit par l'équation

$$s(n) = 0.6s(n-1) + v(n)$$

où le processus d'innovations $v(n)$ a une puissance $\sigma_v^2 = 0.64$ et le bruit blanc additif $\sigma_v^2 = 1$. On désire obtenir le filtre de Wiener tel que, dans le sens LMMSE, $y(n) = s(n)$ (en d'autres termes, $d(n) = s(n)$).

Le filtre optimal est donné par l'équation (3.56).

On a donc

$$S_{du}(z) = S_{su}(z) \stackrel{\text{par indépendance de } s(n) \text{ et } v(n)}{=} S_{ss}(z) = \frac{\sigma_v^2}{H(z)H(z^{-1})}$$

soit

$$S_{ss}(z) = \frac{0.64}{(1 - 0.6z)(1 - 0.6z^{-1})}$$

D'autre part

$$\begin{aligned} S_{uu}(z) &= S_{ss}(z) + \sigma_v^2 (= 1) \\ &= \frac{2(1 - 0.3z^{-1} - 0.3z)}{(1 - 0.6z)(1 - 0.6z^{-1})} \end{aligned}$$

D'où

$$H_{opt_{nc}}(z) = \frac{0.3555}{(1 - 1/3z)(1 - 1/3z^{-1})}$$

On peut vérifier aisément que ce filtre est non causal.

Le MMSE peut être évalué par l'équation (3.63).

L'intégrand vaut :

$$z^{-1} S_{ss}(z) [1 - H_{opt_{nc}}(z)] = \frac{0.3555}{(z - 1/3)(1 - 1/3z)}$$

Le seul pôle à l'intérieur du cercle unité est $z = \frac{1}{3}$. Le résidu est donc donné par :

$$\frac{0.3555}{1 - 1/3z} \Big|_{z=\frac{1}{3}} = 0.40$$

L'erreur quadratique minimale atteignable est donc :

$$\xi = MMSE_{nc} = 0.40$$

3.4.2 Filtre de Wiener causal.

On se limitera cette fois à un filtre IIR causal :

$$y(n) = \sum_{k=0}^{\infty} h_k u(n-k) \quad (3.64)$$

D'une manière similaire au filtre de Wiener non-causal, on trouve :

$$\sum_{n=0}^{\infty} h_n r_{uu}(m-n) = r_{du}(m) \quad \forall m \geq 0 \quad (3.65)$$

Et l'erreur quadratique minimale est donnée par :

$$\xi = \sigma_d^2 - \sum_{n=0}^{\infty} h_{n_{opt}} E \{ r_{du}^*(n) \} \quad (3.66)$$

Où les équations de Wiener-Hopf (3.65) sont réduites aux $m \geq 0$, ce qui nous empêche d'avoir recours à la transformée en z de la même manière que ci-dessus.

Cependant, la représentation de Wold (les innovations) vient à notre secours. En effet, nous pouvons considérer $u(n)$ comme le résultat du filtrage d'un bruit blanc (innovations) $i(n)$ par un filtre $G(z)$, partie à phase minimale obtenue par la factorisation spectrale de $S_{uu}(z)$:

$$S_{uu}(z) = \sigma_i^2 G(z)G(z^{-1}) \quad (3.67)$$

Où le rayon de convergence r_1 de $G(z)$ est inférieur à 1.

En adoptant la représentation de Wold inverse, on peut considérer que le filtre de Wiener est la cascade du filtre blanchissant $1/G(z)$ appliqué à $u(n)$ suivi par un filtre $Q(z)$ dont la sortie est $y(n)$. (la cascade de deux filtres causaux ...)

$$y(n) = \sum_{k=0}^{\infty} q_k i(n-k) \quad (3.68)$$

L'application du principe d'orthogonalité à $e(n) = d(n) - y(n)$ nous donnera les équations normales :

$$\sum_{n=0}^{\infty} q_n r_{ii}(m-n) = r_{di}(m) \quad \forall m \geq 0 \quad (3.69)$$

L'avantage de partir d'un bruit blanc est que $r_{ii}(n) = \delta_n$, ce qui conduit aux solutions :

$$q_l = \frac{r_{di}(m)}{r_{ii}(0)} = \frac{r_{di}(m)}{\sigma_i^2}, \quad \forall m \geq 0 \quad (3.70)$$

D'où la transformée en z de q_l prend la forme :

$$Q(z) = \sum_{n=0}^{\infty} q_n z^{-n} = \frac{1}{\sigma_i^2} \sum_{n=0}^{\infty} r_{di}(n) z^{-n} \quad (3.71)$$

Si l'on définit la partie causale de la densité cross-spectrale de puissance $[S_{di}(z)]_+$:

$$[S_{di}(z)]_+ = \sum_{k=0}^{\infty} r_{di}(k) z^{-k} \quad (3.72)$$

Alors on remarque que :

$$Q(z) = \frac{1}{\sigma_i^2} [S_{di}(z)]_+ \quad (3.73)$$

Pour déterminer $[S_{di}(z)]_+$, il suffit d'exprimer la sortie du filtre blanchissant :

$$i(n) = \sum_{m=0}^{\infty} w_m u(n-m) \quad (3.74)$$

où w_m est la réponse impulsionnelle du filtre blanchissant :

$$\frac{1}{G(z)} = W(z) = \sum_{m=0}^{\infty} w_m z^{-m} \quad (3.75)$$

Ensuite, on exprime la cross-corrélation entre les innovations et la réponse voulue :

$$\begin{aligned} r_{di}(m) &= \mathbb{E} \{d(n) i^*(n-m)\} \\ &= \sum_{k=0}^{\infty} w_k \mathbb{E} \{d(n) u^*(n-k-m)\} \\ &= \sum_{k=0}^{\infty} w_k r_{du}(m+k) \end{aligned} \quad (3.76)$$

La transformée de $r_{di}(m)$ est alors donnée par :

$$\begin{aligned} S_{di}(z) &= \sum_{m=-\infty}^{\infty} \left[\sum_{k=0}^{\infty} w_k r_{du}(m+k) \right] z^{-m} \\ &= \sum_{k=0}^{\infty} w_k \sum_{m=-\infty}^{\infty} r_{du}(m+k) z^{-m} \\ &= \sum_{k=0}^{\infty} w_k z^k \sum_{m=-\infty}^{\infty} r_{du}(k) z^{-m} \\ &= W(z^{-1}) S_{du}(z) = \frac{S_{du}(z)}{G(z^{-1})} \end{aligned} \quad (3.77)$$

D'où :

$$Q(z) = \frac{1}{\sigma_i^2} \left[\frac{S_{du}(z)}{G(z^{-1})} \right]_+ \quad (3.78)$$

Enfin, le filtre de Wiener IIR s'exprime par :

$$H_{opt}(z) = \frac{Q(z)}{G(z)} = \frac{1}{\sigma_i^2 G(z)} \left[\frac{S_{du}(z)}{G(z^{-1})} \right]_+ \quad (3.79)$$

En résumé, pour obtenir le filtre IIR de Wiener, il faut procéder à la factorisation spectrale de $S_{uu}(z)$, pour obtenir $G(z)$ et ensuite déterminer la partie causale de $S_{du}(z)/G(z^{-1})$.

Application 3.3 *Filtre de Wiener IIR causal pour un processus AR(1)*

On écrit cette fois $S_{uu}(z)$ sous la forme :

$$S_{uu}(z) = \frac{1.8(1 - 1/3z^{-1})(1 - 1/3z)}{(1 - 0.6z)(1 - 0.6z^{-1})}$$

Ce qui permet de déterminer $\sigma_i^2 = 1.8$ et

$$G(z) = \frac{1 - 1/3z^{-1}}{1 - 0.6z^{-1}}$$

Se rappelant :

$$S_{ss}(z) = \frac{0.64}{(1 - 0.6z)(1 - 0.6z^{-1})}$$

On obtient :

$$\begin{aligned} \left[\frac{S_{du}(z)}{G(z^{-1})} \right]_+ &= \left[\frac{0.64}{(1 - 1/3z)(1 - 0.6z^{-1})} \right]_+ \\ &= \left[\frac{0.8}{1 - 0.6z^{-1}} + \frac{0.266z}{1 - 1/3z} \right]_+ \\ &= \frac{0.8}{1 - 0.6z^{-1}} \end{aligned}$$

Le filtre IIR optimal peut donc s'écrire :

$$\begin{aligned} H_{optiir} &= 1/1.8 \left(\frac{1 - 0.6z^{-1}}{1 - 1/3z^{-1}} \right) \left(\frac{0.8}{1 - 0.6z^{-1}} \right) \\ &= \frac{4/9}{1 - 1/3z^{-1}} \end{aligned}$$

et une réponse impulsionnelle

$$h_{optiir}(n) = \frac{4}{9} \left(\frac{1}{3} \right)^n, \quad n \geq 0$$

De la même manière que dans le cas non causal, on trouve le MMSE par la formule (3.63) où l'intégrand vaut cette fois :

$$\frac{0.3555}{(z - 1/3z)(1 - 0.6z)}$$

et on trouve

$$\xi = MMSE_{iir} = 0.444$$

On remarque que cette valeur est proche de celle du filtre optimal non causal.

3.4.3 Filtre de Wiener FIR

La procédure pour obtenir un filtre de Wiener est de toute évidence assez complexe et, de plus, demande la connaissance de la séquence infinie d'autocorrélation. Il paraît donc naturel de se limiter à un filtre FIR, qui nous donnera ci-dessous une procédure de calcul nettement plus simple.

On se limitera cette fois à un filtre FIR :

$$y(n) = \sum_{k=0}^{M-1} h_k u(n-k) \quad (3.80)$$

D'une manière similaire au filtre de Wiener non-causal, on trouve :

$$\sum_{n=0}^{M-1} h_n r_{uu}(m-n) = r_{du}(m) \quad m = 0, 1, \dots, M-1 \quad (3.81)$$

L'avantage est que l'on peut exprimer les équations (3.81) sous la forme matricielle :

$$\mathbf{R}\mathbf{h} = \mathbf{r}_{du} \quad (3.82)$$

où \mathbf{R} est la matrice d'autocorrélation du signal $u(n)$ et \mathbf{r}_{du} est le vecteur de crosscorrélation $\mathbf{r}_{du} = [r_{du}(0), r_{du}(1), \dots, r_{du}(M-1)]^T$.

La solution du problème est donc :

$$\mathbf{h}_{opt} = \mathbf{R}^{-1}\mathbf{r}_{du} \quad (3.83)$$

La détermination de cette solution par la méthode directe demande encore un effort de calcul de l'ordre de M^3 . Heureusement, l'algorithme de *Levison-Durbin* que nous verrons plus loin ramènera la complexité à $\mathcal{O}(M^2)$ (voire $M \log M$ par calculateur parallèle) en exploitant la structure Toeplitz de la matrice d'autocorrélation.

L'erreur quadratique minimale est donnée par :

$$\xi = \sigma_d^2 - \sum_{n=0}^{M-1} h_{n_{opt}} r_{du}^*(n) = \sigma_d^2 - \mathbf{r}_{du}^H \mathbf{R}^{-1} \mathbf{r}_{du} \quad (3.84)$$

Application 3.4 Filtre de Wiener FIR pour un processus $AR(1)$

En reprenant les exemples précédents, on détermine la densité spectrale de puissance du signal $s(n)$ par :

$$\begin{aligned} S_{ss}(f) &= \sigma_v^2 |H(f)|^2 \\ &= \frac{0.64}{|1 - 0.6e^{-j2\pi f}|^2} \\ &= \frac{0.64}{1.36 - 1.2 \cos 2\pi f} \end{aligned}$$

La séquence d'autocorrélation vaut

$$r_{ss}(m) = (0.6)^{|m|}$$

Le système d'équation est alors :

$$\begin{aligned} 2h(0) + 0.6h(1) &= 1 \\ 0.6h(0) + 2h(1) &= 0.6 \end{aligned}$$

soit :

$$h(0) = 0.451, \quad h(1) = 0.165$$

Ce qui donne l'erreur quadratique minimale :

$$MMSE_{fir} = 1 - h(0)r_{ss}(0) - h(1)r_{ss}(1) = 0.45$$

Soit une valeur très proche de celle dérivée dans le cas du filtre IIR causal.

3.5 Prédiction linéaire.

3.5.1 Prédiction linéaire avant.

Le problème posé est, étant donné un processus $y(n)$ stationnaire au sens large, de prédire l'échantillon $y(k)$ en connaissant les M échantillons précédents.

$$\hat{y}(k) = - \sum_{m=1}^M a_M(m)y(k-m) \quad (3.85)$$

Le critère que nous allons utiliser pour caractériser la qualité de la prédiction est le critère MSE :

$$\min_{a_M(m)} E \left\{ |y(k) - \hat{y}(k)|^2 \right\} = \min_{a_M(m)} E \left\{ \left| y(k) + \sum_{m=1}^M a_M(m)y(k-m) \right|^2 \right\} \quad (3.86)$$

Soit, en appelant l'erreur de prédiction d'ordre M à l'instant k : $f_M(k)$ et en posant $a_M(0) = 1$:

$$f_M(k) = y(k) - \hat{y}(k) = y(k) + \sum_{m=1}^M a_M(m)y(k-m) = \sum_{m=0}^M a_M(m)y(k-m) \quad (3.87)$$

Le filtre de coefficients $a_M(m)$ ($m = 0, 1, \dots, M$) est encore appelé *filtre d'erreur de prédiction*.

En utilisant le principe d'orthogonalité, nous interprétons la solution à ce problème de la manière suivante : le point $\hat{y}(k)$ dans le sous-espace généré par $y(k-1), \dots, y(k-M)$ est la projection orthogonale de $y(k)$ sur ce sous-espace. La figure 3.10 illustre cette interprétation.

Une expression de l'orthogonalité peut s'écrire :

$$\begin{aligned} E \{ f_M(k)y^*(k-i) \} &= \sum_{m=0}^M a_M(m) E \{ y(k-i)y^*(k-m) \} \\ &= \sum_{m=0}^M a_M(m)r_{yy}(i-m) = 0, i = 1, \dots, n \end{aligned} \quad (3.88)$$

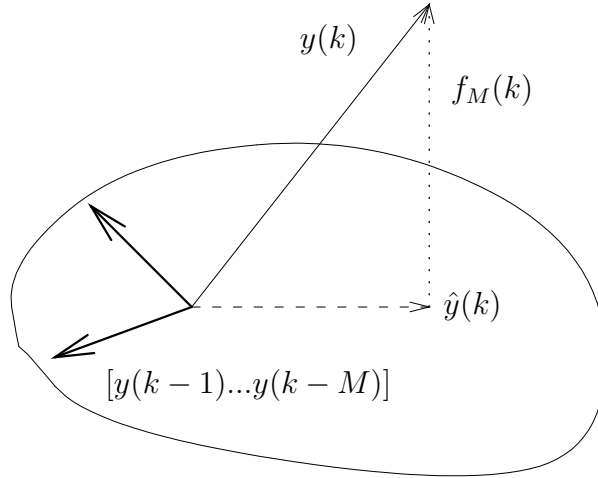


Figure 3.10: Interprétation géométrique du principe d'orthogonalité pour la prédiction linéaire

Grâce à la stationnarité de $y(n)$, les équations sont indépendantes du temps et le filtre optimal est constant dans le temps.

Le MMSE ($\xi = \sigma_{f,M}^2$) est donné par :

$$\begin{aligned}
 \sigma_{f,M}^2 &= \text{E} \{ |f_M|^2(k) \} = \text{E} \left\{ \left| y(k) + \sum_{m=1}^M a_{M_{opt}}(m) y(k-m) \right|^2 \right\} \\
 &= \text{E} \{ f_M(k) y^*(k) \} + \sum_{m=1}^M a_{M_{opt}}^*(m) \underbrace{\text{E} \{ f_M(k) y^*(k-m) \}}_{=0} \\
 &= \text{E} \{ f_M(k) y^*(k) \}
 \end{aligned} \tag{3.89}$$

En introduisant les vecteurs :

$$\mathbf{Y}_{M+1}(k) = [y(k), y(k-1), \dots, y(k-M)]^T \text{ et } \mathbf{a}_M = [1, a_M(1), \dots, a_M(M)]^T \tag{3.90}$$

l'erreur de prédiction s'écrit : $f_M(k) = \mathbf{Y}_{M+1}^T(k) \mathbf{a}_M$.

Les conditions d'orthogonalité s'écrivent alors :

$$\text{E} \{ \mathbf{Y}_{M+1}(k) f_M^*(k) \} = \begin{bmatrix} \text{E} \{ y(k) f_M^*(k) \} \\ \text{E} \{ y(k-1) f_M^*(k) \} \\ \vdots \\ \text{E} \{ y(k-M) f_M^*(k) \} \end{bmatrix} = \begin{bmatrix} \sigma_{f,M}^2 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \tag{3.91}$$

avec, en développant l'erreur de prédiction .

$$\boxed{\mathbb{E} \{ \mathbf{Y}_{M+1}(k) f_M^*(k) \} = \mathbb{E} \{ \mathbf{Y}_{M+1}(k) \mathbf{Y}_{M+1}^H(k) \} \mathbf{a}_M^* = \begin{bmatrix} \sigma_{f,M}^2 \\ 0 \\ \vdots \\ 0 \end{bmatrix} = \mathbf{R}_{M+1} \mathbf{a}_M^*} \quad (3.92)$$

Ces équations sont appelées *équations normales*. On remarque qu'on obtient exactement les équations de *Yule-Walker* que nous avons dérivées dans le cas d'un processus AR. Donc, dans le cas où $y(n)$ est un processus AR d'ordre M , nous obtiendrons comme filtre prédicteur le filtre de synthèse du processus AR, et la variance de l'erreur de prédiction $\sigma_{f,M}^2$ sera égale à la puissance σ_v^2 du processus d'innovation $\nu(n)$.

En d'autres termes, le filtre prédicteur est un filtre blanchissant qui produit la séquence d'innovations $\nu(n)$.

3.5.2 Prédiction linéaire arrière.

Quoique a priori un peu artificiel, le problème de la prédiction arrière sera très utile dans la suite des calculs et de la détermination d'un algorithme rapide de résolution des équations normales, débouchant sur la représentation des filtres en treillis, intéressante à plus d'un titre.

Le problème est simplement de déterminer une estimation d'un échantillon $y(k - M)$ en fonction des M échantillons subséquents :

$$\hat{y}(k - M) = - \sum_{m=1}^M b_M(m) y(k - M + m) \quad (3.93)$$

Le critère que nous allons utiliser pour caractériser la qualité de la prédiction est encore le critère MSE :

$$\begin{aligned} & \min_{b_M(m)} \mathbb{E} \{ |y(k - M) - \hat{y}(k - M)|^2 \} \\ & = \min_{b_M(m)} \mathbb{E} \left\{ \left| y(k - M) + \sum_{m=1}^M b_M(m) y(k - M + m) \right|^2 \right\} \end{aligned} \quad (3.94)$$

Soit, en appelant l'erreur de prédiction arrière d'ordre M à l'instant k : $g_M(k)$ et en posant $b_M(0) = 1$:

$$\begin{aligned} g_M(k) &= y(k - M) - \hat{y}(k - M) \\ &= y(k - M) + \sum_{m=1}^M a_M(m) y(k - M + m) \\ &= \sum_{m=0}^M b_M(m) y(k - M + m) \end{aligned} \quad (3.95)$$

Le filtre de coefficients $b_M(m)$ ($m = 0, 1, \dots, M$) est encore appelé *filtre d'erreur de prédiction arrière*.

En utilisant le principe d'orthogonalité, nous interprétons la solution à ce problème de la manière suivante : le point $\hat{y}(k-M)$ dans le sous-espace généré par $y(k), y(k-1), \dots, y(k-M+1)$ est la projection orthogonale de $y(k-M)$ sur ce sous-espace.

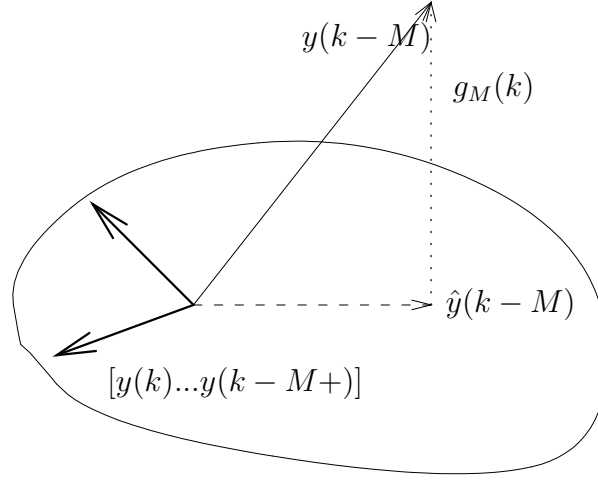


Figure 3.11: Interprétation géométrique du principe d'orthogonalité pour la prédiction linéaire arrière

Une expression de l'orthogonalité peut s'écrire :

$$\begin{aligned} \mathbb{E} \{g_M(k)y^*(k-M+i)\} &= \sum_{m=0}^M b_M(m)\mathbb{E} \{y(k-M+i)y^*(k-M+m)\} \\ &= \sum_{m=0}^M b_M(m)r_{yy}(i-m) = 0, i = 1, \dots, n \end{aligned} \quad (3.96)$$

Grâce à la stationarité de $y(n)$, ces équations sont toujours indépendantes du temps et le filtre optimal est constant dans le temps.

Le MMSE ($\xi = \sigma_{g,M}^2$) est donné par :

$$\begin{aligned} \sigma_{g,M}^2 &= \mathbb{E} \{ |g_M|^2(k) \} = \mathbb{E} \left\{ \left| y(k-M) + \sum_{m=1}^M b_{M_{opt}}(m)y(k-M+m) \right|^2 \right\} \\ &= \mathbb{E} \{ b_M(k)y^*(k-M) \} + \sum_{m=1}^M b_{M_{opt}}^*(m) \underbrace{\mathbb{E} \{ b_M(k)y^*(k-M+m) \}}_{=0} \\ &= \mathbb{E} \{ b_M(k)y^*(k-M) \} \end{aligned} \quad (3.97)$$

Les conditions d'orthogonalité s'écrivent alors :

$$\mathbb{E} \{ \mathbf{Y}_{M+1}(k)g_M^*(k) \} = \begin{bmatrix} \mathbb{E} \{ y(k)g_M^*(k) \} \\ \mathbb{E} \{ y(k-1)g_M^*(k) \} \\ \vdots \\ \mathbb{E} \{ y(k-M)g_M^*(k) \} \end{bmatrix} = \begin{bmatrix} 0 \\ \vdots \\ 0 \\ \sigma_{g,M}^2 \end{bmatrix} \quad (3.98)$$

avec, en développant l'erreur de prédiction :

$$\boxed{\mathbb{E} \{ \mathbf{Y}_{M+1}(k) g_M^*(k) \} = \mathbb{E} \{ \mathbf{Y}_{M+1}(k) \mathbf{Y}_{M+1}^H(k) \} \mathbf{b}_M^* = \begin{bmatrix} 0 \\ \vdots \\ 0 \\ \sigma_{g,M}^2 \end{bmatrix} = \mathbf{R}_{M+1} \mathbf{b}_M^*} \quad (3.99)$$

Ce sont les équations normales pour la prédiction linéaire arrière.

3.5.3 Relation entre prédiction avant et arrière

Appelons \mathbf{J} la matrice d'identité arrière :

$$\mathbf{J} = \begin{bmatrix} 0 & & 1 \\ & \ddots & \\ 1 & & 0 \end{bmatrix} \quad (3.100)$$

\mathbf{J} inverse l'ordre des lignes et des colonnes.

On voit alors clairement que :

$$\mathbf{J} \mathbf{R}_{M+1} \mathbf{J} = \mathbf{R}_{M+1}^T = \mathbf{R}_{M+1}^* \quad (3.101)$$

Des équations normales pour la prédiction arrière, on déduit alors :

$$\mathbf{R}_{M+1} \mathbf{J} \mathbf{b}_M^* = \mathbf{J} \mathbf{R}_{M+1}^* \mathbf{J} \mathbf{J} \mathbf{b}_M^* = \mathbf{J} \mathbf{R}_{M+1}^* \mathbf{b}_M^* = \mathbf{J} \begin{bmatrix} 0 \\ \vdots \\ 0 \\ \sigma_{g,M}^2 \end{bmatrix} = \begin{bmatrix} \sigma_{g,M}^2 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \quad (3.102)$$

On voit donc clairement que les équations normales pour la prédiction linéaire avant et arrière sont identiques, à l'ordre des coefficients près :

$$\mathbf{b}_M = \mathbf{J} \mathbf{a}_M, \text{ soit } b_M(i) = a_M(M-i), i = 1, \dots, M, \quad \sigma_{g,M}^2 = \sigma_{f,M}^2 \quad (3.103)$$

3.6 L'algorithme de Levinson-Durbin

Si on résout les équations de manière directe, cela implique l'inversion d'une matrice, et donc une complexité d'ordre M^3 . Un algorithme rapide a été développé par Levinson (1948) et Durbin (1959) qui permet de diminuer la complexité d'un ordre (M^2). De plus, il permet d'introduire élégamment la représentation des filtres en treillis.

L'algorithme est *récuratif dans l'ordre*, ce qui signifie que l'on détermine les prédicteurs d'ordre $m, m = 1, \dots, M$. L'étape initiale est laissée au soin du lecteur.

Admettons que nous ayons la solution pour l'ordre m , la solution à l'ordre $m+1$ peut s'écrire :

$$\mathbf{R}_{m+2} \left\{ \left[\begin{array}{c} \mathbf{a}_M^* \\ 0 \end{array} \right] + K_{m+1} \mathbf{w}_{m+1} \right\} = \left\{ \left[\begin{array}{c} \sigma_{f,m}^2 \\ 0 \\ \vdots \\ \Delta_{m+1} \end{array} \right] + K_{m+1} \mathbf{x}_{m+1} \right\} = \left[\begin{array}{c} \sigma_{f,m+1}^2 \\ 0 \\ \vdots \\ 0 \end{array} \right] \quad (3.104)$$

Pour obtenir la solution sous la forme $[\sigma_{f,m+1}^2 0 \cdots 0]$, il faut adopter \mathbf{x}_{m+1} de la forme $[*0 \cdots *]^T$ et il est clair que l'on peut choisir \mathbf{x}_{m+1} de la forme $[\Delta_{m+1} 0 \cdots \sigma_{f,m}^2]^T$, ce qui fixe simplement :

$$K_{m+1} = -\frac{\Delta_{m+1}}{\sigma_{f,m}^2} \quad (3.105)$$

Ce choix particulier, et la relation que nous avons fait entre prédiction arrière et prédiction avant permet alors d'écrire l'équation précédente sous la forme :

$$\mathbf{R}_{m+2} \left\{ \left[\begin{array}{c} \mathbf{a}_m^* \\ 0 \end{array} \right] + K_{m+1} \left[\begin{array}{c} 0 \\ \mathbf{b}_m^* \end{array} \right] \right\} = \left\{ \left[\begin{array}{c} \sigma_{f,m}^2 \\ 0 \\ \vdots \\ \Delta_{m+1} \end{array} \right] + K_{m+1} \left[\begin{array}{c} \Delta_{m+1} \\ 0 \\ \vdots \\ \sigma_{f,m}^2 \end{array} \right] \right\} = \left[\begin{array}{c} \sigma_{f,m+1}^2 \\ 0 \\ \vdots \\ 0 \end{array} \right] \quad (3.106)$$

La récursion sur la solution devient :

$$\mathbf{a}_{m+1}^* = (\mathbf{I} + K_{m+1} \mathbf{J}) \left[\begin{array}{c} \mathbf{a}_m^* \\ 0 \end{array} \right] \quad (3.107)$$

En particulier, $a_{m+1}(m+1) = K_{m+1}$. L'erreur de prédiction devient alors :

$$\sigma_{f,m+1}^2 = \sigma_{f,m}^2 + K_{m+1} \Delta_{m+1} = \sigma_{f,m}^2 (1 - K_{m+1}^2) \quad (3.108)$$

Une puissance devant être positive, l'équation précédente implique que

$$|K_{m+1}| \leq 1 \quad (3.109)$$

D'autre part, on a que

$$\sigma_{m+1}^2 \leq \sigma_m^2 \leq \cdots \leq \sigma_1^2 \leq \sigma_0^2 \quad (3.110)$$

ce qui confirme l'intuition selon laquelle l'augmentation de l'ordre de prédiction diminue l'erreur de prédiction. Dans le cas d'un processus AR(m), après l'ordre m , les variances seront égales (voir plus haut).

On peut résumer l'algorithme de Levinson comme suit :

 Algorithme de Levinson

$$\begin{aligned}
 \text{Initialisation} \quad & \mathbf{a}_0 = [1], \sigma_{f,0}^2 = r_{yy}(0) \\
 \text{récursion} \quad & \Delta_{m+1} = [r_{yy}(m+1) \cdots r_{yy}(1)] \mathbf{a}_m^* \\
 & K_{m+1} = -\frac{\Delta_{m+1}}{\sigma_{f,m}^2} \\
 & \mathbf{a}_{m+1}^* = \begin{bmatrix} \mathbf{a}_m^* \\ 0 \end{bmatrix} + K_{m+1} \begin{bmatrix} 0 \\ \mathbf{J} \mathbf{a}_m^* \end{bmatrix} \\
 & \sigma_{f,m+1}^2 = \sigma_{f,m}^2 (1 - K_{m+1}^2)
 \end{aligned}$$

Par récursion, l'algorithme de Levinson demande environ $2n$ multiplications, soit, au total :

$$\sum_{n=1}^M 2n = \frac{2M(M+1)}{2} \simeq M^2 \quad (3.111)$$

3.6.1 Interprétations des paramètres K_m et Δ_{m-1}

Les paramètres K_m sont appelés *coefficients de réflexion*, en effet, l'équation $\sigma_{f,m+1}^2 = \sigma_{f,m}^2 (1 - K_{m+1}^2)$ est similaire à celle de la théorie des lignes où K_m est le coefficient de réflexion au droit d'une discontinuité (différence d'impédance caractéristique) dans la ligne.

Le paramètre Δ_{m+1} (et donc K_{m+1} peut être interprété comme étant une cross-corrélation entre l'erreur de prédiction avant et l'erreur de prédiction arrière. On peut montrer :

$$K_{m+1} = -\frac{\mathbf{E}\{f_m(k)g_m^*(k-1)\}}{\mathbf{E}\{|f_m(k-1)|^2\}} \quad (3.112)$$

Cette dernière expression est appelée *coefficient de corrélation partielle* (**PARCOR**).

3.7 Filtres en treillis.

A partir de l'algorithme de Levinson, en écrivant les récursions sur les coefficients des filtres de prédiction, avec la relation $\mathbf{a}_m = \mathbf{J} \mathbf{b}_m$:

$$\begin{aligned}
 \mathbf{a}_{m+1} &= \begin{bmatrix} \mathbf{a}_m \\ 0 \end{bmatrix} + K_{m+1} \begin{bmatrix} 0 \\ \mathbf{b}_m \end{bmatrix} \\
 \mathbf{b}_{m+1} &= \begin{bmatrix} \mathbf{b}_m \\ 0 \end{bmatrix} + K_{m+1} \begin{bmatrix} 0 \\ \mathbf{a}_m \end{bmatrix}
 \end{aligned} \quad (3.113)$$

En écrivant ces équations sous la forme de transformées en z :

$$\begin{aligned}
 \mathcal{A}_{m+1}(z) &= \mathcal{A}_m(z) + K_{m+1} \mathcal{B}_m(z) z^{-1} \\
 \mathcal{B}_{m+1}(z) &= K_{m+1} \mathcal{A}_m(z) + \mathcal{B}_m(z) z^{-1}
 \end{aligned} \quad (3.114)$$

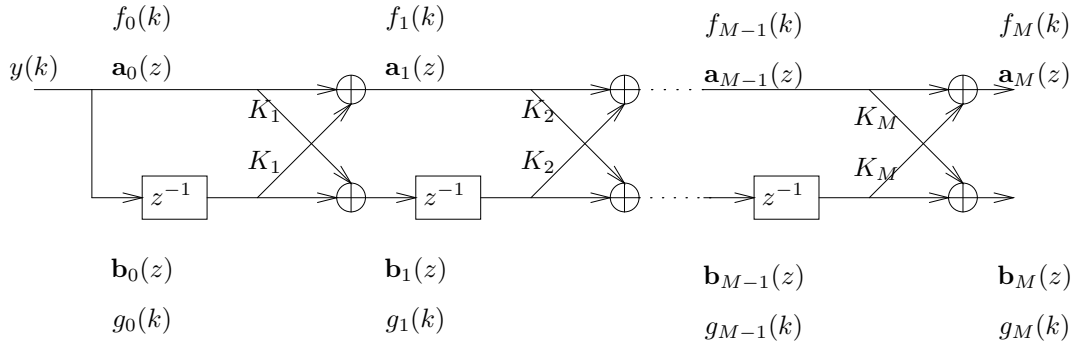


Figure 3.12: Filtre en treillis pour la prédiction linéaire.

Ces deux équations décrivent une section d'un filtre en treillis, que l'on peut représenter sous la forme de la figure 3.12.

En réécrivant les équations (3.87) et (3.95) sous la forme $f_m(k) = \mathcal{A}_m(z)y(k)$ et $b_m(k) = \mathcal{B}_m(z)y(k)$, les sorties des sections de treillis sont $f_m(k)$ et $b_m(k)$ si on excite l'entrée du premier filtre par le signal $y(k)$, ce qui justifie les annotations de la figure 3.12.

Un des premiers avantages, surtout dans des implémentations VLSI, est la modularité de ces filtres : il suffit d'ajouter des sections pour obtenir une meilleure prédiction, sans que les coefficients PARCOR précédents doivent être modifiés. En effet, dans le cas d'une implémentation directe par un filtre transversal comme en figure 3.13, l'augmentation de l'ordre total du filtre demande une modification de tous les coefficients $a_m(k)$, ce qui oblige à refaire l'entièreté du calcul.

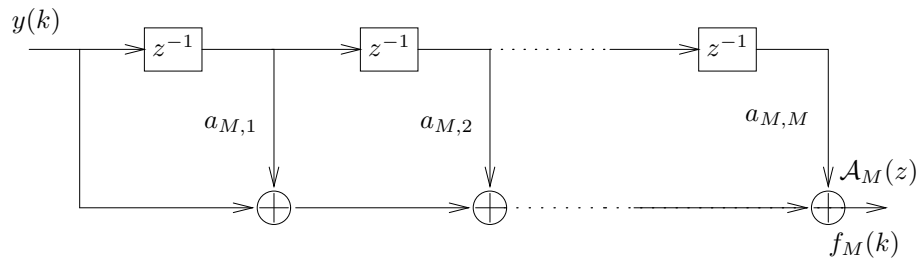


Figure 3.13: Filtre transversal pour la prédiction linéaire.

D'autre part, d'un point de vue numérique, le fait que les coefficients PARCOR sont bornés ($|K_m| < 1$) nous assure des résultats intermédiaires également bornés, ce qui est très intéressant dans les calculateurs à virgule fixe (pas d'"overflow").

3.8 La méthode des moindres carrés (LS : Least Squares)

3.8.1 Introduction

Dans le filtrage de Wiener, le critère d'optimalité est stochastique : on désire minimiser la moyenne de l'erreur au carré. Pour cela, nous avons besoin de la statistique de second ordre des processus (moyennes et covariances). Une autre approche consiste à minimiser, non plus la moyenne stochastique, mais temporelle de cette erreur, c'est la méthode des moindres carrés (L.S. : Least Squares).

En se reportant à la figure 3.8, on obtient le critère :

$$\xi(\mathbf{h}) = \sum_{k=k_1}^{k_2} |e(k)|^2 \quad (3.115)$$

Dans ce cas-ci, les coefficients \mathbf{h} du filtre optimal sont d'office constants sur la période d'observation. Si on désire tenir compte d'une non stationnarité du canal, il suffit d'opter pour le critère :

$$\xi(\mathbf{h}) = \sum_{k=k_1}^{k_2} \lambda^k |e(k)|^2 \quad 0 < \lambda \leq 1 \quad (3.116)$$

En adoptant un filtre FIR de longueur M , on a :

$$e(i) = d(i) - \sum_{k=0}^{M-1} h_k u(i-k) \quad (3.117)$$

D'autre part, pour tenir compte du caractère stochastique du problème, on modélise le processus à identifier comme étant le signal filtré auquel on a ajouté du bruit (en général blanc additif), comme indiqué à la figure 3.14. L'objectif principal de cette modélisation sera de caractériser les performances de l'estimée.

3.8.2 Fenêtrage.

Les bornes k_1 et k_2 peuvent être choisies de différentes manières. On se placera dans le cas où nous disposons des données $[u(0), u(1), \dots, u(N-1)]$

La méthode de la covariance ne fait pas d'hypothèses sur les valeurs des données en dehors de la fenêtre d'observation. On peut alors exprimer les données sous une forme matricielle

:

$$\mathbf{U} = \begin{bmatrix} u(M-1) & u(M) & \cdots & u(N-1) \\ u(M-2) & u(M-1) & \cdots & u(N-2) \\ \vdots & \vdots & \ddots & \vdots \\ u(0) & u(1) & \cdots & u(N-M) \end{bmatrix} \quad (3.118)$$

On peut alors exprimer l'équation (3.117) sous forme vectorielle,

$$\mathbf{e} = \mathbf{d} - \mathbf{U}^T \mathbf{h} \quad (3.119)$$

en notant

$$\mathbf{e} = [e(M-1)e(M)\cdots e(N-1)] \text{ et } \mathbf{d} = [d(M-1)d(M)\cdots d(N-1)]$$

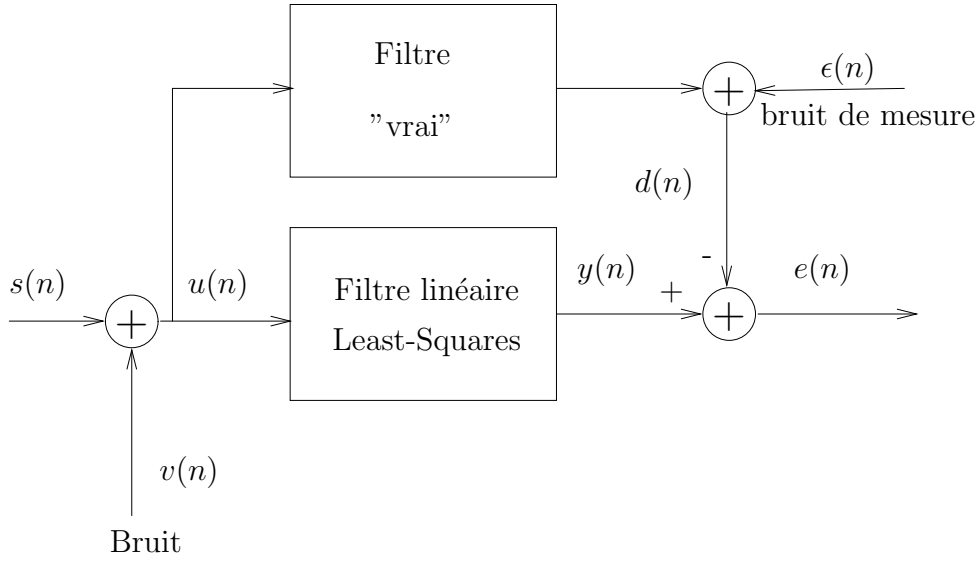


Figure 3.14: Modèle pour le filtrage par les moindres carrés.

L'expression $\mathbf{U}^T \mathbf{h}$ représentant la convolution entre les données $u(i)$ et le filtre de coefficients h_k , on appelle parfois \mathbf{U}^T la matrice de convolution.

La méthode de la corrélation fait l'hypothèse que les valeurs des données en dehors de la fenêtre d'observation sont nulles. On peut alors exprimer les données sous une forme matricielle :

$$\mathbf{U} = \begin{bmatrix} u(1) & \cdots & u(M-1) & \cdots & u(N-1) & 0 & \cdots & 0 \\ 0 & \cdots & u(M-2) & \cdots & u(N-2) & u(N-1) & \cdots & 0 \\ \vdots & \ddots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & u(0) & \cdots & u(N-M) & u(N-M+1) & \cdots & u(N-1) \end{bmatrix} \quad (3.120)$$

L'équation (3.119) reste alors valable en notant

$$\mathbf{e} = [e(0)e(1)\cdots e(N+M-1)] \text{ et } \mathbf{d} = [d(0)d(1)\cdots d(N+M-1)]$$

La méthode du préfenêtrage fait l'hypothèse que les valeurs des données avant $u(0)$ sont nulles. On peut alors exprimer les données sous une forme matricielle :

$$\mathbf{U} = \begin{bmatrix} u(1) & u(2) & \cdots & u(M-1) & u(M) & \cdots & u(N-1) \\ 0 & u(1) & \cdots & u(M-2) & u(M-1) & \cdots & u(N-2) \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & u(0) & u(1) & \cdots & u(N-M) \end{bmatrix} \quad (3.121)$$

L'équation (3.119) reste alors valable en notant

$$\mathbf{e} = [e(0)e(1)\cdots e(N-1)] \text{ et } \mathbf{d} = [d(0)d(1)\cdots d(N-1)]$$

La méthode du postfenêtrage fait l'hypothèse que les valeurs des données après $u(N-1)$ sont nulles (sans faire d'hypothèses sur les données antérieures à $u(0)$). On peut alors exprimer les données sous une forme matricielle :

$$\mathbf{U} = \begin{bmatrix} u(M-1) & u(M) & \cdots & u(N-1) & 0 & \cdots & 0 \\ u(M-2) & u(M-1) & \cdots & u(N-2) & u(N-1) & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & 0 \\ u(0) & u(1) & \cdots & u(N-M) & u(N-M+1) & \cdots & u(N-1) \end{bmatrix} \quad (3.122)$$

L'équation (3.119) reste alors valable en notant

$$\mathbf{e} = [e(M-1)e(M)\cdots e(N+M-1)] \text{ et } \mathbf{d} = [d(M-1)d(M)\cdots d(N+M-1)]$$

3.8.3 Principe d'orthogonalité pour les moindres carrés

On se placera dorénavant dans la méthode de covariance.

On définit alors le nouveau critère à minimiser sous la forme vectorielle :

$$\xi(\mathbf{h}) = \mathbf{e}^H \mathbf{e}, \text{ où } \mathbf{e} = \mathbf{d} - \mathbf{U}^T \mathbf{h} \quad (3.123)$$

On détermine alors :

$$\frac{\partial}{\partial \mathbf{h}} (\xi(\mathbf{h})) = \begin{bmatrix} \frac{\partial \mathbf{e}^H \mathbf{e}}{\partial h_0} \\ \frac{\partial \mathbf{e}^H \mathbf{e}}{\partial h_1} \\ \vdots \\ \frac{\partial \mathbf{e}^H \mathbf{e}}{\partial h_{M-1}} \end{bmatrix} \quad (3.124)$$

En utilisant les relations de l'appendice, on trouve

$$\frac{\partial}{\partial \mathbf{h}} (\xi(\mathbf{h})) = -(\mathbf{d} - \mathbf{U}^T \mathbf{h})^H \mathbf{U}^T = -\mathbf{e}^H \mathbf{U}^T \quad (3.125)$$

Cette relation, pour avoir un minimum, doit être annulée. Ensuite, on vérifie aisément que la dérivée seconde est positive.

On peut donc écrire

$$\mathbf{e}^H \mathbf{U}^T = 0 \quad (3.126)$$

qui exprime l'orthogonalité entre les erreurs et les entrées du filtre optimal. On a donc un **principe d'orthogonalité** similaire à celui du cas stochastique (filtre de Wiener).

De la même manière, toute combinaison linéaire de l'équation précédente est valable, et on obtient aisément que :

$$\mathbf{e}^H \mathbf{y} = \mathbf{e}^H \hat{\mathbf{d}} \quad (3.127)$$

que l'on interprète en disant que l'erreur doit être orthogonale aux entrées et à l'estimation.

3.8.4 Equations normales

Cette relation d'orthogonalité nous amène directement à une interprétation géométrique. Dérivons d'abord la solution du problème des moindres carrés (équation (3.125)).

En notant, pour la facilité $\mathbf{U}^T = \mathbf{X}$, on obtient les **équations normales** pour la méthode des moindres carrés :

$$(\mathbf{d} - \mathbf{X}\hat{\mathbf{h}})^H \mathbf{X} = 0 \Leftrightarrow \hat{\mathbf{h}} = (\mathbf{X}^H \mathbf{X})^{-1} \mathbf{X}^H \mathbf{d} \quad (3.128)$$

où $(\mathbf{X}^H \mathbf{X})^{-1} \mathbf{X}^H$ est appelée la **matrice pseudo-inverse** de \mathbf{X} et où on a supposé que $\mathbf{X}^H \mathbf{X}$ était invertible.

L'erreur minimale étant alors exprimée par :

$$\begin{aligned} \xi_{min} &= \hat{\mathbf{e}}^H \hat{\mathbf{e}} \\ &= (\mathbf{d} - \mathbf{X}\hat{\mathbf{h}})^H (\mathbf{d} - \mathbf{X}\hat{\mathbf{h}}) = \mathbf{d}^H \mathbf{d} - \mathbf{d}^H \mathbf{X}\hat{\mathbf{h}} - \hat{\mathbf{h}}^H \mathbf{X}^H \mathbf{d} + \hat{\mathbf{h}}^H \mathbf{X}^H \mathbf{X}\hat{\mathbf{h}} \\ &= \mathbf{d}^H \mathbf{d} - \mathbf{d}^H \mathbf{X} (\mathbf{X}^H \mathbf{X})^{-1} \mathbf{X}^H \mathbf{d} \end{aligned} \quad (3.129)$$

Ce qui exprime l'erreur et la solution directement en fonction des données et de la réponse désirée.

On définit alors, de manière analogue au cas stochastique, la matrice de corrélation :

$$\Phi = \mathbf{X}^H \mathbf{X} = \begin{bmatrix} \phi(0,0) & \phi(1,0) & \cdots & \phi(M-1,0) \\ \phi(0,1) & \phi(1,1) & \cdots & \phi(M-1,1) \\ \vdots & \vdots & \ddots & \vdots \\ \phi(0,M-1) & \phi(1,M-1) & \cdots & \phi(M-1,M-1) \end{bmatrix} \quad (3.130)$$

où on peut détailler les éléments de la matrice de corrélation que l'on appellera les *fonctions d'autocorrélation moyennées dans le temps*.

$$\phi(t, k) = \sum_{i=M-1}^{N-1} u(i-k)u^*(i-t) \quad (3.131)$$

On peut aisément vérifier que cette matrice d'autocorrélation possède essentiellement les mêmes propriétés que la matrice d'autocorrélation stochastique, à savoir qu'elle est **Hermitienne, définie non négative** et possède des **valeurs propres réelles non négatives**. D'autre part, par construction, elle est le produit de **deux matrices rectangulaires Toeplitz**.

De la même manière, on définit le vecteur de **cross-corrélation** entre l'entrée \mathbf{u} et la réponse désirée \mathbf{d} :

$$\Theta = [\theta(0), \theta(1), \dots, \theta(M-1)]^T = \mathbf{X}^H \mathbf{d} \quad (3.132)$$

On peut alors réécrire l'équation (3.128) sous la forme :

$$\Phi \hat{\mathbf{h}} = \Theta \quad (3.133)$$

qui sont les **équations normales** pour la méthode des moindres carrés. On y retrouve la relation entre la matrice d'autocorrélation et la matrice de cross-corrélation d'une manière similaire à l'équation (3.82) pour le critère MMSE dans le cas stochastique.

Le vecteur de sortie est alors exprimé par :

$$\mathbf{y} = \hat{\mathbf{d}} = \mathbf{X}\hat{\mathbf{h}} \quad (3.134)$$

L'erreur minimale s'écrit alors :

$$\xi_{min} = \mathbf{d}^H \mathbf{d} - \Theta^H \Phi^{-1} \Theta \quad (3.135)$$

3.8.5 Interprétation géométrique.

L'abstraction relative de ces équations, la notion d'orthogonalité et la représentation par des vecteurs laissent supposer que l'on peut donner une interprétation géométrique du problème des moindres carrés et de la solution.

En effet, les colonnes de la matrice de données \mathbf{X} , Φ et \mathbf{P} définissent un espace vectoriel de dimension $N - M + 1$. Le vecteur de données est de dimension N et est donc défini dans un espace de dimensions N . D'autre part, la relation d'orthogonalité entre l'erreur minimale et l'estimation des données ($\hat{\mathbf{d}}$) (on se rappellera que $\mathbf{e}\hat{\mathbf{d}} = 0$) signifie que l'on peut diviser l'espace des données en l'espace des données estimées et son complément orthogonal qui est l'espace de l'erreur. On parle souvent dans la littérature de sous-espace signal et sous-espace bruit, ce dernier étant en l'occurrence le bruit du à l'erreur.

Enfin, pour le passage de l'espace total aux sous-espaces signal et bruit, on définit l'opérateur de projection $\mathbf{P} = \mathbf{X}(\mathbf{X}^H \mathbf{X})^{-1} \mathbf{X}^H$. En effet, la relation

$$\hat{\mathbf{d}} = \mathbf{X}(\mathbf{X}^H \mathbf{X})^{-1} \mathbf{X}^H \mathbf{d} = \mathbf{P}\mathbf{d}$$

met en lumière que l'opérateur \mathbf{P} nous fait passer de \mathbf{d} à $\hat{\mathbf{d}}$. Cet opérateur peut donc être interprété comme l'opérateur de projection de l'espace total sur le sous-espace signal.

D'autre part, on définit l'opérateur de projection orthogonal

$$\mathbf{I} - \mathbf{P}$$

qui assure la projection des données sur le sous-espace bruit.

La figure 3.15 illustre cela dans le cas de $M = 1$ et $N = 3$, c'est-à-dire 4 données en entrée et un filtre de longueur 2.

Exemple 3.1 *Cas d'un filtre de longueur 2 et de 4 données*

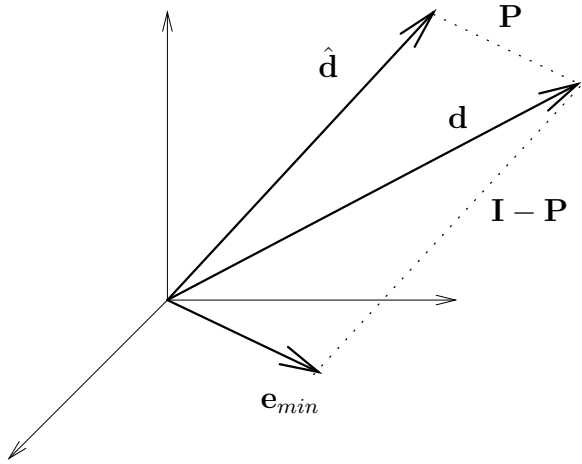


Figure 3.15: Interprétation géométrique de la méthode des moindres carrés.

Considérons le cas $N = 3, M = 1$ avec $\mathbf{u} = [5, 3, 1, 10]^T; \mathbf{d} = [7, 4, 1]^T$. On obtient immédiatement

$$\mathbf{X} = \begin{bmatrix} 3 & 5 \\ 1 & 3 \\ 10 & 1 \end{bmatrix}$$

L'opérateur de projection vaut :

$$\mathbf{P} = \mathbf{X}(\mathbf{X}^H \mathbf{X})^{-1} \mathbf{X}^T = \begin{bmatrix} 0.7257 & 0.4446 & 0.0378 \\ 0.4446 & 0.2795 & -0.0613 \\ 0.0378 & -0.0613 & 0.9948 \end{bmatrix} \quad (3.136)$$

Ce qui permet de trouver la solution :

$$\mathbf{y} = \hat{\mathbf{d}} = \begin{bmatrix} 6.8960 \\ 4.1686 \\ 1.0144 \end{bmatrix} \quad (3.137)$$

Et l'erreur d'estimation :

$$\mathbf{y} = \hat{\mathbf{d}} = \begin{bmatrix} 0.1040 \\ 0.1686 \\ 0.0144 \end{bmatrix} \quad (3.138)$$

Cet exemple est illustré par la figure 3.16. \triangleleft

3.8.6 Propriétés de l'estimation des moindres carrés.

1. L'estimée $\hat{\mathbf{h}}$ est non biaisée, si $\{\epsilon(i)\}$ est à moyenne nulle.

Notons la valeur vraie du filtre \mathbf{h}_o . Alors, de

$$\hat{\mathbf{h}} = (\mathbf{X}^H \mathbf{X})^{-1} \mathbf{X}^H \mathbf{d} \text{ et } \mathbf{d} = \mathbf{X} \mathbf{h}_o + \boldsymbol{\epsilon} \quad (3.139)$$

On déduit :

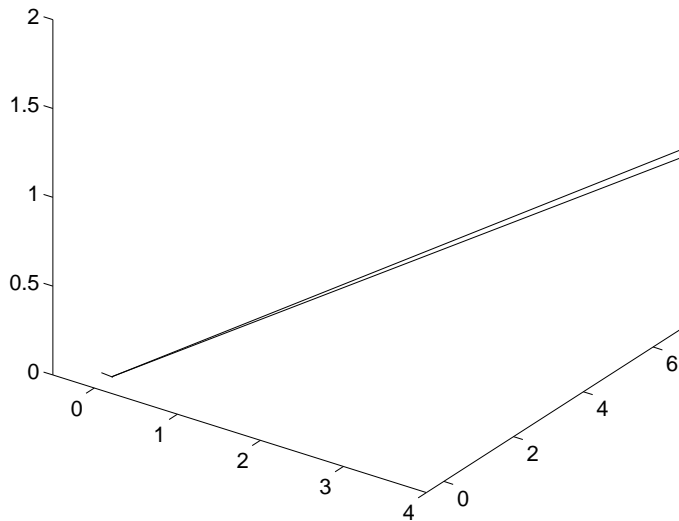


Figure 3.16: Exemple d'estimation LS

$$\begin{aligned}\hat{\mathbf{h}} &= (\mathbf{X}^H \mathbf{X})^{-1} \mathbf{X}^H \mathbf{X} \mathbf{h}_o + (\mathbf{X}^H \mathbf{X})^{-1} \mathbf{X}^H \boldsymbol{\epsilon} \\ &= \mathbf{h}_o + (\mathbf{X}^H \mathbf{X})^{-1} \mathbf{X}^H \boldsymbol{\epsilon}\end{aligned}\quad (3.140)$$

Les données représentées par \mathbf{X} étant déterministes et $\epsilon(i)$ étant à moyenne nulle, on déduit immédiatement :

$$\mathbb{E} \{ \hat{\mathbf{h}} \} = \mathbf{h}_o \quad (3.141)$$

2. Si $\epsilon(i)$ est un bruit blanc à moyenne nulle et de variance σ^2 , l'estimée $\hat{\mathbf{h}}$ a une matrice de covariance valant $\sigma^2 \boldsymbol{\Phi}^{-1}$.

Les relations suivantes nous mènent directement à ce résultat :

$$\begin{aligned}\text{cov}\{\hat{\mathbf{h}}\} &= \mathbb{E} \{ (\hat{\mathbf{h}} - \mathbf{h})(\hat{\mathbf{h}} - \mathbf{h})^H \} \\ &= \mathbb{E} \{ (\mathbf{X}^H \mathbf{X})^{-1} \mathbf{X}^H \boldsymbol{\epsilon} \boldsymbol{\epsilon}^H \mathbf{X} (\mathbf{X}^H \mathbf{X})^{-1} \} \\ &= (\mathbf{X}^H \mathbf{X})^{-1} \mathbf{X}^H \mathbb{E} \{ \boldsymbol{\epsilon} \boldsymbol{\epsilon}^H \} \mathbf{X} (\mathbf{X}^H \mathbf{X})^{-1} \\ &= (\mathbf{X}^H \mathbf{X})^{-1} \mathbf{X}^H \sigma^2 \mathbf{I} \mathbf{X} (\mathbf{X}^H \mathbf{X})^{-1} \\ &= \sigma^2 (\mathbf{X}^H \mathbf{X})^{-1} \mathbf{X}^H \mathbf{X} (\mathbf{X}^H \mathbf{X})^{-1} \\ &= \sigma^2 (\mathbf{X}^H \mathbf{X})^{-1} \\ &= \sigma^2 \boldsymbol{\Phi}^{-1}\end{aligned}\quad (3.142)$$

3. Si $\epsilon(i)$ est un bruit blanc à moyenne nulle et de variance σ^2 , l'estimée $\hat{\mathbf{h}}$ est la meilleure estimée linéaire non-biaisée (BLUE : Best Linear Unbiased Estimate).

Considérons un estimateur linéaire non biaisé

$$\tilde{\mathbf{h}} = \mathbf{B}\mathbf{d} \implies \tilde{\mathbf{h}} = \mathbf{B}\mathbf{X}\mathbf{h}_o + \mathbf{B}\epsilon \quad (3.143)$$

De même, l'hypothèse de moyenne nulle du bruit implique $E\{\tilde{\mathbf{h}}\} = \mathbf{B}\mathbf{X}\mathbf{h}_o$. Pour que l'estimée $\tilde{\mathbf{h}}$ soit non biaisée, il faut donc :

$$\mathbf{B}\mathbf{X} = \mathbf{I} \quad (3.144)$$

On peut donc écrire $\tilde{\mathbf{h}} = \mathbf{h}_o + \mathbf{B}\epsilon$ et la covariance s'exprime :

$$\begin{aligned} \text{cov}\{\tilde{\mathbf{h}}\} &= E\{(\tilde{\mathbf{h}} - \mathbf{h})(\tilde{\mathbf{h}} - \mathbf{h})^H\} \\ &= E\{\mathbf{B}\epsilon\epsilon^H\mathbf{B}^H\} \\ &= \sigma^2\mathbf{B}\mathbf{B}^H \end{aligned} \quad (3.145)$$

En définissant une matrice $\Psi = \mathbf{B} - (\mathbf{X}^H\mathbf{X})^{-1}\mathbf{X}^H$, nous avons les relations :

$$\begin{aligned} \Psi\Psi^H &= \mathbf{B}\mathbf{B}^H - \mathbf{B}\mathbf{X}(\mathbf{X}^H\mathbf{X})^{-1} - (\mathbf{X}^H\mathbf{X})^{-1}\mathbf{X}^H\mathbf{B}^H + (\mathbf{X}^H\mathbf{X})^{-1} \\ &= \mathbf{B}\mathbf{B}^H - (\mathbf{X}^H\mathbf{X})^{-1} \end{aligned} \quad (3.146)$$

or, $\Psi\Psi^H \geq 0$, puisque c'est une forme quadratique, donc $\mathbf{B}\mathbf{B}^H \geq (\mathbf{X}^H\mathbf{X})^{-1}$ et la proposition initiale est prouvée.

4. Si $\epsilon(i)$ est un bruit blanc gaussien à moyenne nulle et de variance σ^2 , l'estimée $\hat{\mathbf{h}}$ est l'estimée non-biaisée de variance minimale. (MVUE : Minimum Variance Unbiased Estimate).

La manière de démontrer cette affirmation est de comparer la covariance de l'estimée à la matrice d'information de Fisher. Nous en laissons la démonstration à titre d'exercice.

Chapter 4

Démodulateur optimal.

|| Ce (court) chapitre a comme objectif de déterminer un démodulateur optimal dans un canal présentant de l'interférence entre symboles et du bruit blanc additif gaussien. Un modèle de canal sera indiqué et nous déterminerons une méthode de calcul de complexité acceptable pour le démodulateur optimal.

4.1 Démodulateur optimal.

Cette section suit l'approche de [Pro89].

Sans entrer dans le détail du développement de Karhunen-Loève¹, il nous suffira ici de savoir qu'un signal temporel peut être exprimé dans un espace de fonctions de base orthonormales en nombre infini : en appelant $r(t)$ le signal reçu

$$r(t) = \sum_n I_n h(t - nT) + z(t) \quad (4.1)$$

$$r(t) = \lim_{N \rightarrow \infty} \sum_{k=1}^N r_k f_k(t) \quad (4.2)$$

où $\{f_k(t)\}$ forme un ensemble de fonctions orthogonales et $\{r_k\}$ représentent les projections de $r(t)$ sur la base précitée. $r(t)$ étant un processus stochastique, les r_k sont des variables aléatoires.

On peut alors montrer :

$$r_k = \sum_n I_n h_{kn} + z_k, \quad k = 1, 2, \dots \quad (4.3)$$

où h_{kn} et z_k sont respectivement les projections de $h(t - nT)$ et de $z(t)$ sur l'espace des $f_k(t)$.

La séquence des $\{z_k\}$ est une séquence gaussienne de moyenne nulle et de covariance $E\{z_m z_k^*\} = \lambda \delta_{km}$.

Il est clair que les v.a. r_k sont gaussiennes, centrées sur $\sum_n I_n h_{kn}$ et le vecteur $\mathbf{r}_N = [r_1 r_2 \dots r_N]$ a une densité de probabilité conditionnée sur $\mathbf{I}_p = [I_1 I_2 \dots I_p]$ donnée par :

$$p(\mathbf{r}_N | \mathbf{I}_p) = \frac{1}{\prod_{k=1}^N 2\pi\lambda} \exp\left(-\frac{1}{2} \sum_{k=1}^N \frac{|r_k - \sum_n I_n h_{kn}|^2}{\lambda}\right) \quad (4.4)$$

¹Voir l'appendice 4A de [Pro89]

Au passage à la limite, les sommes deviennent des intégrales et le logarithme de la densité de probabilité conditionnelle est proportionnel à :

$$\begin{aligned}\xi(\mathbf{I}_p) &= - \int_{-\infty}^{\infty} \left| r(t) - \sum_n I_n h(t - nT) \right|^2 dt \\ &= - \int_{-\infty}^{\infty} |r(t)|^2 dt + 2\Re \sum_n \left[I_n^* - \int_{-\infty}^{\infty} r(t) h^*(t - nT) dt \right] \\ &\quad - \sum_n \sum_m I_n^* I_m \int_{-\infty}^{\infty} h^*(t - nT) h(t - mT) dt\end{aligned}\quad (4.5)$$

L'objectif étant de maximiser cette expression en fonction de la séquence \mathbf{I}_p . Pour ce faire, nous remarquons que la première intégrale représente l'énergie du signal reçu et n'intervient donc pas. La deuxième intégrale nous est familière et correspond à la sortie du filtre adapté excité par le signal reçu, et échantillonné au taux $1/T$.

$$y_n \equiv y(nT) = \int_{-\infty}^{\infty} r(t) h^*(t - nT) dt \quad (4.6)$$

On peut alors réécrire l'équation précédente sous la forme :

$$\xi(\mathbf{I}_p) = 2\Re \left(\sum_n I_n^* y_n \right) - \sum_n \sum_m I_n^* I_m x_{n-m} \quad (4.7)$$

avec

$$x_n \equiv x(nT) = \int_{-\infty}^{\infty} h^*(t - nT) h(t - mT) dt \quad (4.8)$$

La séquence $\{x_n\}$ représente donc la fonction d'autocorrélation (déterministe) de la réponse impulsionnelle du canal. Celle-ci étant supposée connue, ou déterminée par une procédure d'indentification de canal, on voit clairement que les y_n forment une statistique suffisante pour déterminer les symboles émis. En d'autres mots, le passage du signal reçu dans un filtre adapté nous fournit, après échantillonnage aux instants nT , une séquence qui permet de déterminer les symboles I_k grâce au critère du maximum de vraisemblance.

La procédure directe consiste à remplacer la séquence \mathbf{I}_p dans l'équation (4.7) par toutes les valeurs possibles, c'est ce qu'on appelle la *Maximum-Likelihood Sequence Estimation (MLSE)*. Il est clair que c'est une méthode purement théorique, totalement impraticable. Nous verrons que l'algorithme de Viterbi permettra d'obtenir une estimation par le maximum de vraisemblance à moindre coût.

4.2 Un modèle de canal

En nous basant sur ce qui vient d'être développé précédemment, nous pouvons aisément introduire un modèle de canal représentant le canal en présence d'ISI et de BBAG (bruit blanc additif gaussien).

Des équations (4.1) et (4.6) on obtient :

$$y_k = \sum_n I_n x_{k-n} + \nu_k \quad (4.9)$$

où ν_k est le bruit blanc filtré par le filtre adapté :

$$\nu_k = \int_{-\infty}^{\infty} z(t)h^*(t - kT)dt \quad (4.10)$$

En normalisant $x_0 = 1$, on peut écrire :

$$y_k = I_k + \underbrace{\sum_{n \neq k} I_n x_{k-n}}_{ISI} + \nu_k \quad (4.11)$$

Un modèle équivalent de canal peut être basé sur cette équation, comme dans la figure (4.1).

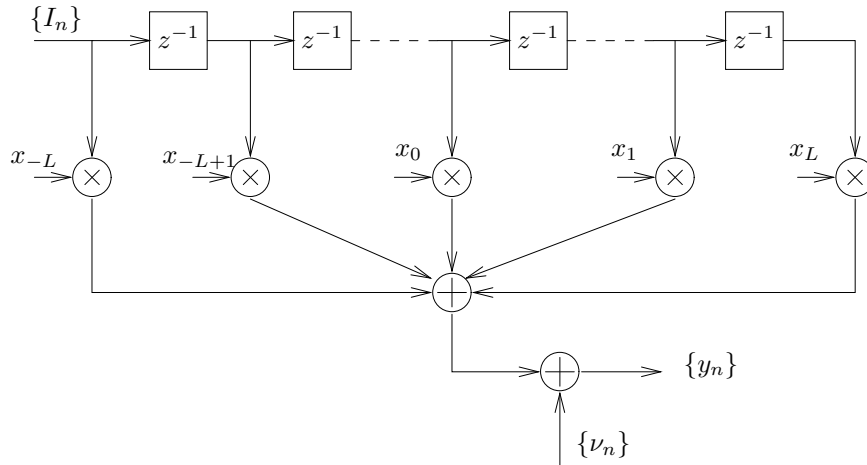


Figure 4.1: Modèle discret de canal en présence d'ISI et de BBAG

Dans ce modèle, nous avons simplement fait l'hypothèse que le canal avait une fonction d'autocorrélation de durée finie.

Le principal désavantage de ce modèle réside en l'aspect coloré du bruit additif. En effet, toutes les méthodes de détermination des performances (en termes de taux d'erreur) ont été développées dans le cas d'un bruit additif blanc. Nous allons donc modifier le modèle de manière à avoir un bruit blanc en sortie.

Pour ce faire, écrivons la transformée en z de $\{x_k\}$:

$$X(z) = \sum_{k=-L}^L x_k z^{-k} \quad (4.12)$$

On vérifie aisément que $x_k = x_{-k}^*$, donc que $X(z) = X^*(z^{-1})$. Il s'ensuit que si ρ est une racine de $X(z)$, il en est de même pour $1/\rho^*$. On peut donc écrire ² :

$$X(z) = F(z)F^*(z^{-1}) \quad (4.13)$$

On peut choisir les racines de $F^*(z^{-1})$ à l'intérieur du cercle unité de manière à obtenir un filtre $1/F^*(z^{-1})$ stable et physiquement réalisable.

²ce qui est un résultat similaire celui tiré de la représentation de Wold des signaux stationnaires

D'autre part, dans le modèle précédent, nous avons un bruit additif de caractéristiques :

$$E \{ \nu_k \nu_j^* \} = \begin{cases} N_0 x_{k-j} & |k-j| \leq L \\ 0 & \text{ailleurs} \end{cases} \quad (4.14)$$

C'est-à dire une densité spectrale de puissance $N_0 X(z)$. En multipliant ce bruit par $1/F^*(z^{-1})$, on obtient une densité spectrale de puissance ³

$$S_{\eta\eta} = N_0 X(z) \frac{1}{F^*(z^{-1})F(z)} = N_0 \quad (4.15)$$

et on obtient bien un bruit blanc en sortie.

D'autre part, cela équivaut à filtrer y_k par $F(z)$ et on a donc :

$$u_k = \sum_{n=0}^L f_n I_{k-n} + \eta_k \quad (4.16)$$

ce qui conduit à la figure 4.2. Nous appellerons ce modèle le *modèle discret équivalent à filtre blanchissant*.

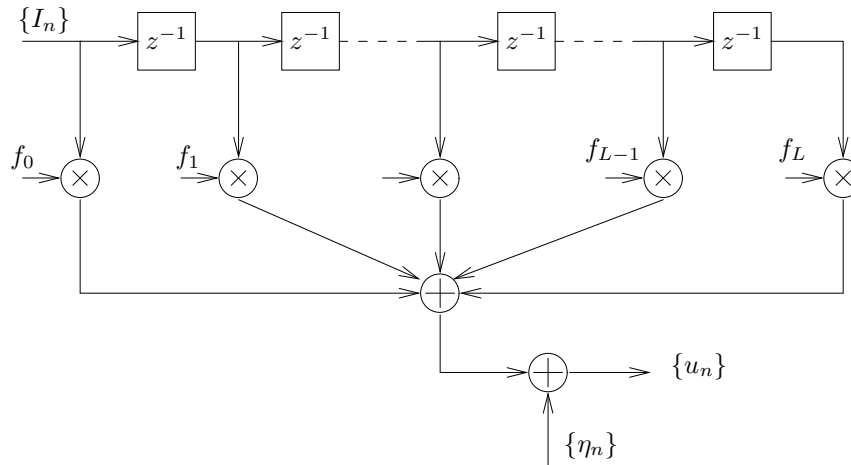


Figure 4.2: Modèle discret équivalent à filtre blanchissant

En résumé, le modèle discret équivalent à filtre blanchissant représente la cascade du canal, du filtre adapté, de l'échantillonneur et du filtre blanchissant.

³Par la relation : si

$$\begin{aligned} Y(z) &= H(z)X(z) \\ Y(e^{j\omega}) &= |H(e^{j\omega})|^2 X(e^{j\omega}) \\ &= H(e^{j\omega})H^*(e^{-j\omega})X(e^{j\omega}) \\ &= H(z)H^*(z^{-1}) \Big|_{z=e^{j\omega}} \end{aligned}$$

4.3 L'algorithme de Viterbi pour le MLSE

L'algorithme de Viterbi a été développé dans le cadre des codes convolutionnels, qui sont une classe particulière de codes correcteurs d'erreurs. Ensuite, il a été (entre autres) appliqué au problème de la décision par le maximum de vraisemblance tel qu'il a été décrit ci-dessus.

Suivant l'approche de [BBC87], nous présentons cet algorithme dans le cadre général de la minimisation d'une fonction de plusieurs variables non-indépendantes pour l'appliquer ensuite à la détection.

4.3.1 L'algorithme de Viterbi

Soit K fonctions scalaires $\lambda_0(\tau_0), \lambda_1(\tau_1), \dots, \lambda_K(\tau_{K-1})$. Le problème posé est de minimiser :

$$\mu \equiv \min_{\{\tau_i\}} \sum_{m=0}^{K-1} \lambda_m(\tau_m) \quad (4.17)$$

Dans le cas où les variables τ_i sont indépendantes (i.e. $\tau_i | \{\tau_0, \dots, \tau_{i-1}, \tau_{i+1}, \dots, \tau_{K-1}\} = \tau_i$), le problème est trivial, car il revient à effectuer K minimisations sur une variable :

$$\mu = \sum_{m=0}^{K-1} \min_{\tau_m} \lambda_m(\tau_m) \quad (4.18)$$

Le problème que nous allons tenter de résoudre, décrit mathématiquement par l'équation (4.17) est communément appelé le *problème du voyageur de commerce*.

Exemple 4.1 Problème du voyageur de commerce

A écrire \triangleleft

Une situation particulière, à laquelle notre classe de problèmes se ramènera, est le cas où :

$$\forall m = 0, \dots, K-2, \quad \{\tau_m | \tau_{n \neq m}\} = \{\tau_m | \tau_{m+1}\} \quad (4.19)$$

Dans ce cas, en minimisant μ par rapport à τ_0 , la valeur issue de cette minimisation (appelons-la μ_1) dépendra de τ_1 (et uniquement de τ_1). On obtient :

$$\mu_1(\tau_1) = \min_{\tau_0 | \tau_1} \lambda_0(\tau_0) \quad (4.20)$$

De même, à l'instant suivant, on aura :

$$\mu_2(\tau_2) = \min_{\tau_1 | \tau_2} (\lambda_1(\tau_1) + \mu_1(\tau_1)) \quad (4.21)$$

où nous avons du inclure $\mu_1(\tau_1)$ dans le processus de minimisation. En continuant ce processus jusqu'au bout, nous obtenons l'algorithme de Viterbi.

Algorithme de Viterbi

$$\begin{aligned} \mu_1(\tau_1) &= \min_{\tau_0 | \tau_1} \lambda_0(\tau_0) \\ \mu_m(\tau_m) &= \min_{\tau_{m-1} | \tau_m} (\lambda_{m-1}(\tau_{m-1}) + \mu_{m-1}(\tau_{m-1})) \\ \mu &= \min_{\tau_{K-1}} \mu_{K-1}(\tau_{K-1}) \end{aligned}$$

L'application la plus courante de l'algorithme de Viterbi est la recherche du chemin le plus court dans un graphe (orienté et métré)⁴.

Considérons une source d'information générant la séquence finie : $\mathbf{I}_P(0) = [I_0, \dots, I_{P-1}]$, ces symboles sont indépendants et peuvent prendre M valeurs distinctes. Ces symboles traversent un système (canal) dont la sortie à l'instant m dépend, de manière déterministe et invariante dans le temps, du symbole à l'entrée et des L derniers symboles :

$$u_m = g(I_m, \dots, I_{m-L}) = g(\mathbf{I}_L(m)) \quad (4.22)$$

On peut considérer la séquence u_m comme étant la sortie d'un registre à décalage, comme illustré à la figure 4.3.

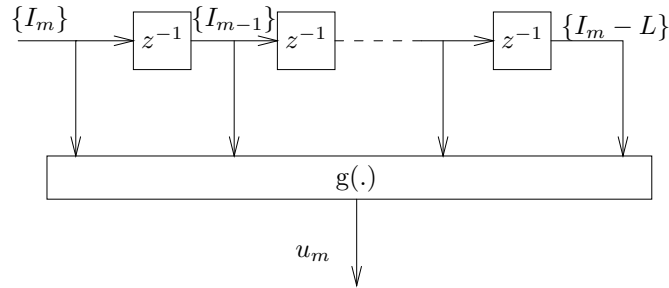


Figure 4.3: Registre à décalage équivalent au canal.

On définit alors classiquement l'état σ_m comme étant le contenu du registre à l'instant m :

$$\sigma_m \equiv [I_{m-1}, \dots, I_{m-L}] = \mathbf{I}_{L-1}(m-1) \quad (4.23)$$

et la sortie du registre s'écrit alors :

$$u_m = g(I_m, \sigma_m) \quad (4.24)$$

L'émission d'un symbole d'information peut alors se décrire comme étant le passage d'un état du registre à l'autre. On définit la transition :

$$\tau_{m+1} \equiv (\sigma_m, \sigma_{m-1}) \quad (4.25)$$

Notons que s'il on veut que l'indice m puisse parcourir les valeurs de 0 à L , il faut supposer $I_m = 0$ pour $m < 0$ et $m > K - 1$, ou alors élargir la définition de $g(\cdot)$

Nous pouvons représenter les états successifs du registre à décalage par un treillis de la manière suivante :

1. chaque point du treillis représente un état ;
2. chaque flèche représente une transition ;
3. le déroulement temporel se fait de gauche à droite.

⁴On parle alors également de l'algorithme de Bellman-Ford.

Notez que vous avez déjà rencontré ce type de problème sous la forme dans le cas équivalent des machines d'état finies.

Exemple 4.2 *Processus à deux mémoires*

Considérons un processus présentant une mémoire de longueur deux et une séquence d'information en entrée binaire de longueur 5. On peut la représenter par un registre à décalage de longueur deux, avec les paramètres : $L = 2, M = 2, I_m = \{-1, +1\}, K = 5$. Pour $2 < m < 5$, on peut avoir les états $(\pm 1, \pm 1)$. La figure 4.4 représente ce treillis. On notera l'indépendance de ce treillis par rapport à la fonction $g(\cdot)$. \triangleleft

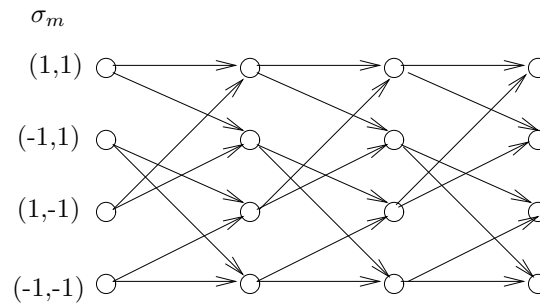


Figure 4.4: Diagramme en treillis associé au problème

Pour appliquer l'algorithme de Viterbi, il suffit d'associer une *métrique* à chaque transition. On se retrouve alors, en vertu de l'équation (4.19), en présence d'un problème de recherche du plus court chemin dans un graphe.

4.3.2 Application à la recherche du MLSE

L'objectif est de trouver la séquence \mathbf{I}_P qui maximise l'expression (4.7). En se basant sur le modèle de canal équivalent à filtre blanchissant, on peut réécrire cette équation sous la forme :

$$p(\mathbf{u}_N | \mathbf{I}_N) \equiv p(u_N, u_{N-1}, \dots, u_1 | I_N, I_{N-1}, \dots, I_1) \quad (4.26)$$

où l'on suppose avoir reçu la séquence $[u_1, u_2, \dots, u_N]$, les symboles $[I_1, I_2, \dots, I_N]$ ayant été émis.

Comme la séquence de bruit $\{\eta_k\}$ est blanche, nous pouvons exprimer :

$$p(\mathbf{u}_N | \mathbf{I}_N) = \prod_{k=1}^N p(u_k | I_k, I_{k-1}, \dots, I_{k-L}) \quad (4.27)$$

où $I_k = 0$ pour $k < 0$ et L étant la longueur du canal.

En prenant le logarithme de ces probabilités, nous nous ramenons au problème de l'équation (4.17) où le problème de minimisation est devenu un problème de maximisation.

La traduction en machine d'état peut se faire comme illustré à la figure 4.5. Les états, en reprenant l'exemple précédent sont les paires (I_m, I_{m-1}) et les transitions sont commandées par le symbole (I_{m+1}) en entrée.

Il nous reste à définir une métrique. L'approche de [BBC87] est d'utiliser l'équation (??).

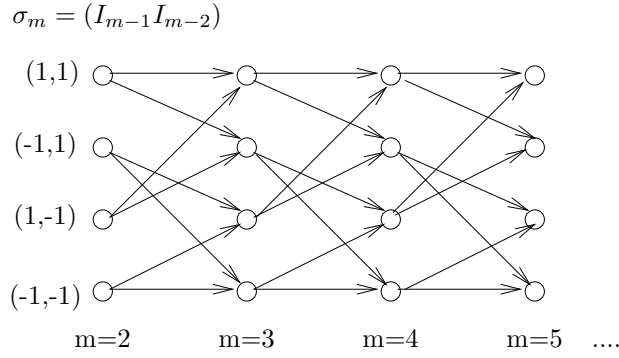


Figure 4.5: Treillis de l'exemple du canal simple.

Une approche plus conforme à ns développements précédents consiste, en adoptant les définitions :

$$\begin{aligned} \sigma_m &\equiv [I_{m-1}, \dots, I_{m-L}] = \mathbf{I}_{L-1}(m-1) \\ u_m = g(I_m, \sigma_m) &= \sum_{k=0}^L f_k I_{m-k} + \eta_m \end{aligned} \quad (4.28)$$

En faisant l'hypothèse que le bruit blanc est gaussien, l'expression de

$$\ln p(u_m | I_m, I_{m-1}, \dots, I_{m-L}) = -1/2 \ln(2\pi\sigma_{\eta\eta}^2) - \frac{1}{2\sigma_{\eta\eta}^2} \left(u_m - \sum_{k=0}^L f_k I_{m-k} \right)^2 \quad (4.29)$$

nous incite à utiliser la métrique

$$\left(u_m - \sum_{k=0}^L f_k I_{m-k} \right)^2 \quad (4.30)$$

pour le problème de minimisation (maximiser l'expression (4.29) revient à minimiser (4.30)).

Algorithme de Viterbi pour le MLSE

$$\begin{aligned} \mu_1(I_1, \sigma_1) &= \min_{\mathbf{I}_L(1)} \left(u_1 - \sum_{k=0}^L f_k I_{1-k} \right)^2 \\ \mu_m(I_m, \sigma_m) &= \min_{\mathbf{I}_L(m)} \left[\left(u_m - \sum_{k=0}^L f_k I_{m-k} \right)^2 + \mu_{m-1}(I_{m-1}, \sigma_{m-1}) \right] \\ \mu &= \min_{\sigma_K} \mu_{K-1}(\sigma_K) \end{aligned}$$

Application 4.1 *Exemple pour un canal FIR de longueur 3*

On utilise un canal FIR de longueur 3 caractérisé par les coefficients : $[0.407 \ 0.815 \ 0.407]$. La réponse fréquentielle en amplitude et en phase est donnée à la figure 4.3.2.

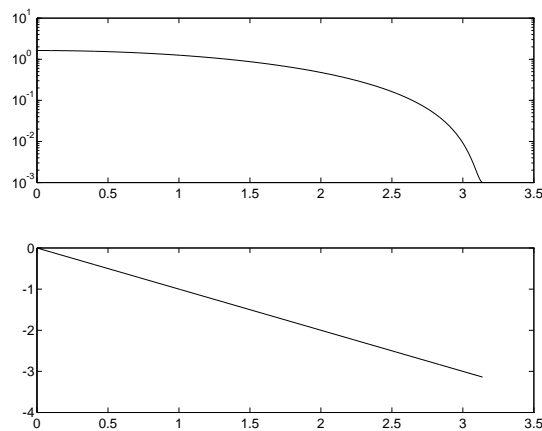


Figure 4.6: Réponse fréquentielle

En se basant sur l'algorithme décrit ci-dessus, nous pouvons définir les états possibles, décrits par les paires $[(1,1) (1,-1) (-1,1) (-1,-1)]$. Ensuite, les transitions possibles d'un état à l'autre définissent un treillis du même type que la première cellule de la figure 4.3.2. Cette figure exprime simplement les deux premières étapes de l'algorithme, le nombre du dessus exprimant l'entrée du canal (supposé sans bruit dans notre cas) et les nombres inscrits à droite la métrique obtenue à la fin de cette étape. Théoriquement, il faut attendre la fin de la transmission pour être sûr d'obtenir la séquence correcte au sens du maximum de vraisemblance. Cependant, en pratique, après un nombre fini d'étapes, on constate qu'un seul chemin survit, on peut donc prendre une décision avant la fin de la transmission de la séquence complète, quel que soit l'état de départ que l'on considère d'ailleurs (i.e., on ne doit pas nécessairement partir de l'état de métrique minimale). En pratique, on considère que si le canal est de longueur L , on peut prendre une décision après $5L$ à $7L$. Dans ce cas, à la dernière étape indiquée, on peut décider du début de séquence $[-1 \ -1 \ 1 \ 1]$.

Les graphes suivants donnent la performance de la méthode pour ce canal ainsi que le canal décrit par $\{f_k\} = \{0.227, 0.460, 0.688, 0.460, 0.2287\}$.

Pour une analyse détaillée des performances, qui permet de donner des bornes inférieures et supérieures du taux d'erreur, sans passer par la simulation, nous renvoyons le lecteur à la section 6.7 de [Pro89].

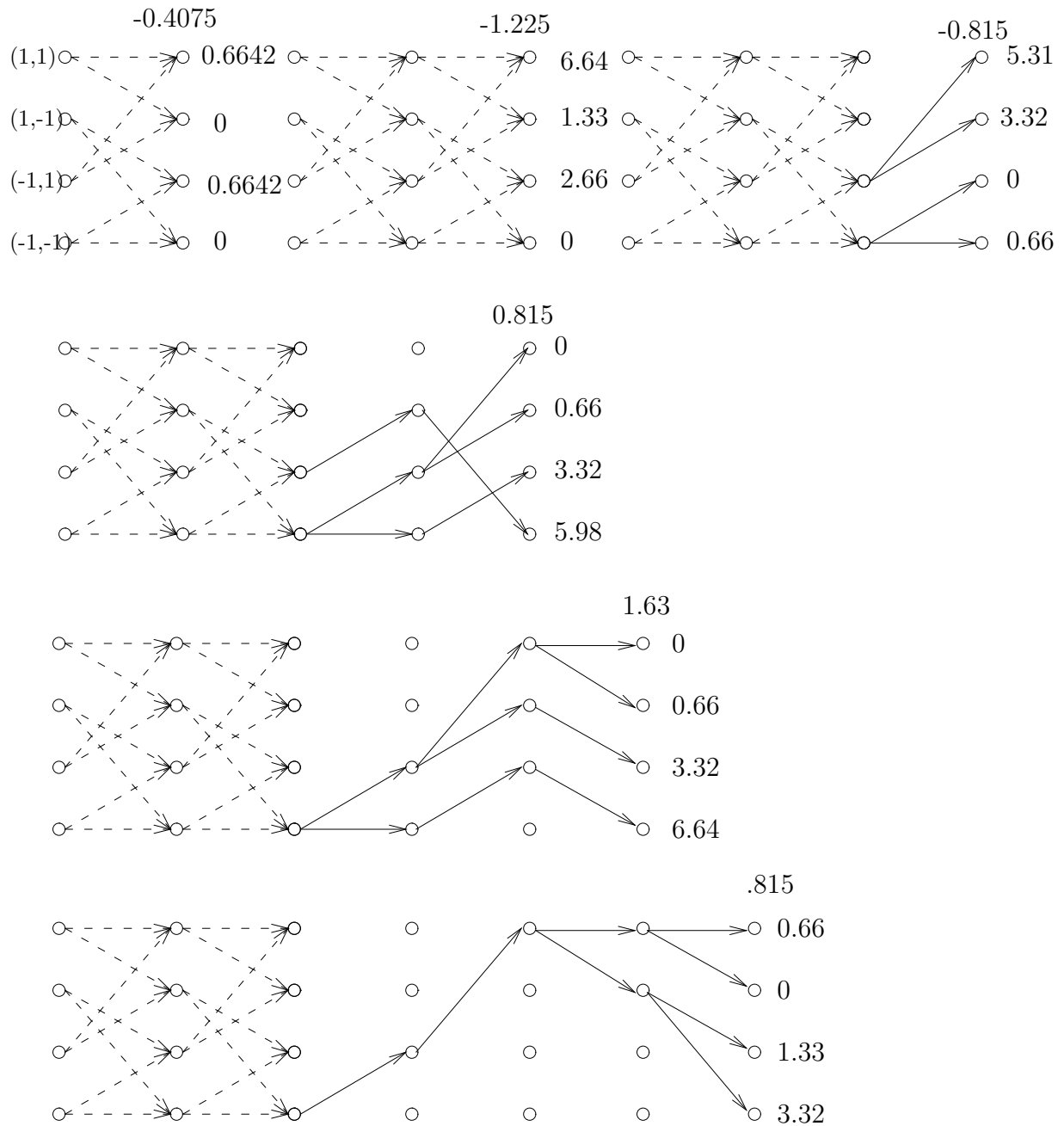


Figure 4.7: Début du treillis

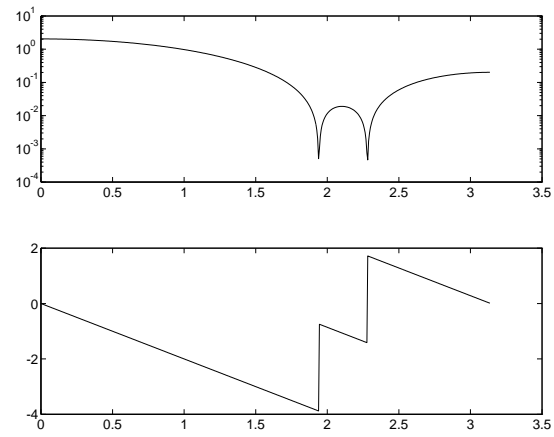


Figure 4.8: Réponse fréquentielle du second canal.

Figure 4.9: Taux d'erreurs en fonction du SNR

Chapter 5

Egalisation Linéaire à filtre adapté

Ce chapitre traitera de l'égalisation linéaire effectuée par filtrage au taux de symboles. On abordera rapidement l'égalisateur Zero-Forcing ; l'égalisateur MMSE linéaire avec ses algorithmes du gradient et LMS et enfin les solutions basées sur les moindres carrés (RLS, Lattice-RLS).

La première classe d'égaliseurs que nous allons étudier est composée d'un simple filtre linéaire. On peut le considérer comme étant un filtre optimal, avec un critère prédéfini. En outre, nous aborderons immédiatement l'égalisation adaptative, qui permet à l'égaliseur de s'adapter aux variations (non-stationarités) du canal.

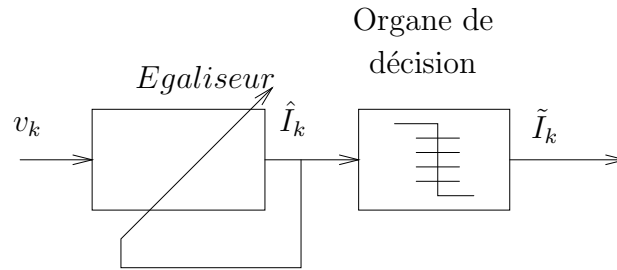


Figure 5.1: Principe de l'égaliseur linéaire

5.1 L'égalisateur Zero-Forcing (ZF)

Dans un premier temps, nous adopterons simplement comme objectif l'élimination pure et simple de l'ISI, sans se soucier du bruit additif.

La sortie de l'égaliseur vaut :

$$\hat{I}_k = \sum_{j=-K}^K c_j v_{k-j} = q_0 I_k + \sum_{n \neq k} I_n q_{k-n} + \sum_{j=-\infty}^{\infty} c_j \eta_{k-j} \quad (5.1)$$

où $\{q_n\}$ est la convolution du filtre égaliseur $\{c_j\}$ avec le modèle de canal $\{f_j\} \Rightarrow q_n = \sum_{j=-\infty}^{\infty} c_j f_{n-j}$.

Dans la mesure où les $\{I_k\}$ sont à priori aléatoires, nous allons nous fixer comme objectif de minimiser la valeur maximale possible (worst case) de l'ISI. :

$$\xi = \sum_{\substack{n=0 \\ n \neq k}}^{\infty} |q_n| = \sum_n \left| \sum_{j=-\infty}^{\infty} c_j f_{n-j} \right| \quad (5.2)$$

Cas du filtre égaliseur IIR

Dans ce cas, nous pouvons obtenir :

$$\xi = 0 \quad (\text{d'où Zero-Forcing}) \quad (5.3)$$

en adoptant

$$q_n = \sum_{j=-\infty}^{\infty} c_j f_{n-j} = \begin{cases} 1 & n = 0 \\ 0 & n \neq 0 \end{cases} \quad (5.4)$$

Soit, en z , $C(z) = \frac{1}{F(z)}$

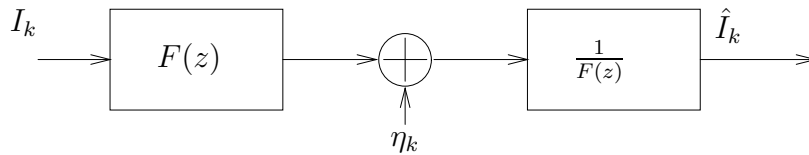


Figure 5.2: Modèle de l'égaliseur avec filtre blanchissant

En explicitant le modèle de canal complet (c'est-à-dire en faisant apparaître le filtre blanchissant de bruit), on a la figure :

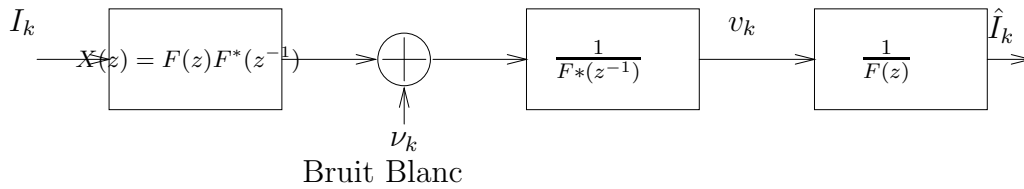


Figure 5.3: Modèle de l'égaliseur sans filtre blanchissant

Ce qui donne l'égaliseur équivalent :

$$C'(z) = \frac{1}{F(z)F^*(z^{-1})} = \frac{1}{C(z)} \quad (5.5)$$

Performances de l'égaliseur ZF-IIR en termes de SNR Adoptons, sans perte de généralité, l'hypothèse que le signal reçu à une puissance unitaire ($q_O = 1$ et $|I_k|^2 = 1$), ce qui permet de confondre le rapport $\frac{S}{N}$ avec $\frac{1}{\sigma_n^2}$ où σ_n^2 est la puissance du bruit à la sortie.

Le calcul de la puissance de bruit se fera simplement en calculant la dsp à chaque étape.

- A la sortie du canal

$$S_{\nu\nu}(\omega) = N_0|F(e^{j\omega t})|^2 = N_0X(e^{j\omega t}) \quad (5.6)$$

- A la sortie de l'égaliseur

$$S_{nn}(\omega) = N_0X(e^{j\omega t}) \frac{1}{X(e^{j\omega t})X^*(e^{-j\omega t})} = \frac{N_0}{X(e^{j\omega t})} \quad (5.7)$$

en remarquant que $X^*(e^{-j\omega t}) = X(e^{j\omega t})$ car $X(\omega) = |H(\omega)|^2$ est réel.

$$\sigma_n^2 = \frac{T}{2\pi} \int_{-\frac{\pi}{T}}^{\frac{\pi}{T}} S_{nn}(\omega) d\omega = \frac{TN_0}{2\pi} \int_{-\frac{\pi}{T}}^{\frac{\pi}{T}} \frac{d\omega}{X(e^{j\omega T})} \quad (5.8)$$

de $X(\omega) = |H(\omega)|^2$ et, par périodisation du spectre due à l'échantillonnage, on peut exprimer ce dernier :

$$X(e^{j\omega T}) = \frac{1}{T} \sum_{i=-\infty}^{\infty} |H(\omega + \frac{2\pi n}{T})|^2 \quad |\omega| \leq \frac{\pi}{T} \quad (5.9)$$

et donc, le rapport signal/bruit pour un égaliseur ZF de longueur infinie :

$$\gamma_\infty = \frac{1}{\sigma_n^2} = \left[\frac{T^2 N_0}{2\pi} \int_{-\frac{\pi}{T}}^{\frac{\pi}{T}} \frac{d\omega}{\sum_{i=-\infty}^{\infty} |H(\omega + \frac{2\pi n}{T})|^2} \right]^{-1} \quad (5.10)$$

Cette expression montre que, dans le cas où $H(\omega)$ a une valeur faible (presque zéro) en certains endroits, le SNR sera élevé (presque infini). C'est ce qu'on appelle "*noise enhancement*" et c'est la raison pour laquelle le ZF n'est quasiment jamais utilisé. Nous ne nous attarderons pas plus longtemps sur celui-ci et aborderons directement l'égaliseur linéaire MMSE.

5.2 Egaliseur MMSE linéaire

5.2.1 Dérivation de l'égaliseur MMSE-IIR

Le critère adopté sera la minimisation de l'erreur quadratique moyenne (le MSE : Mean-Squared Error) :

$$\xi = E \left\{ |I_k - \hat{I}_k|^2 \right\} \quad (5.11)$$

Outre la quadraticité du critère, on observera que le bruit est pris intrinsèquement en compte dans ce critère, par le biais de \hat{I}_k .

Application du critère d'orthogonalité.

Travaillons dans le modèle de canal à filtre blanchissant. Nous avons, après égalisation :

$$\hat{I}_k = \sum_{i=-\infty}^{\infty} c_i v_{k-i} \quad (5.12)$$

Le critère d'orthogonalité pour le MMSE nous permet d'écrire :

$$E \{ e_k v_{k-l}^* \} = 0 \quad -\infty < l < \infty \quad (5.13)$$

où $e_k = I_k - \hat{I}_k$, ce qui mène immédiatement à

$$\sum_{j=-\infty}^{\infty} c_j E \{ v_{k-j} v_{k-l}^* \} = E \{ I_k v_{k-l}^* \} \quad -\infty < l < \infty \quad (5.14)$$

qui sont les **équations normales** pour l'égaliseur linéaire MMSE. En exprimant que v_k est la sortie du canal ($v_k = \sum_{n=0}^L f_n I_{k-n} + \eta_k$). L'autocorrélation de la sortie prend la forme :

$$\begin{aligned} E \{ v_{k-j} v_{k-l}^* \} &= \sum_{n=0}^L f_n^* f_{n+l-j} + N_0 \delta_{ij} \\ &= \begin{cases} x_{l-j} + N_0 \delta_{ij} & |l-j| \leq L \\ 0 & \text{ailleurs} \end{cases} \end{aligned} \quad (5.15)$$

L'intercorrélacion s'exprimant par

$$E \{ I_k v_{k-l}^* \} = \begin{cases} f_{-l}^* + N_0 \delta_{ij} & -L \leq l \leq 0 \\ 0 & \text{ailleurs} \end{cases} \quad (5.16)$$

Soit, en exprimant les équations normales dans le domaine z :

$$C(z)[F(z)F^*(z^{-1}) + N_0] = F^*(z^{-1}) \quad (5.17)$$

Soit

$$C(z) = \frac{F^*(z^{-1})}{X(z) + N_0} \quad (5.18)$$

et donc, pour le filtre égaliseur :

$$C'(z) = \frac{1}{X(z) + N_0} \quad (5.19)$$

La différence principale par rapport à l'égaliseur ZF réside en l'introduction de la densité spectrale de bruit dans son expression. Il est clair qu'en l'absence de bruit, les deux méthodes donneront les mêmes résultats. Par contre, en présence de bruit, le filtre tentera cette fois-ci de minimiser en même temps l'influence de l'ISI et du bruit blanc additif.

5.2.2 Erreur minimale et rapport signal/bruit.

L'erreur minimale vaut :

$$\begin{aligned}
\xi_{\min} &= E\{|e_k|^2\} = E\{e_k I_k^*\} - \underbrace{E\{e_k \hat{I}_k\}}_{=0(\perp)} \\
&= E\{|I_k|^2\} - \sum_{j=-\infty}^{\infty} c_j E\{v_{k-j} I_k^*\} \\
&= 1 - \sum_{j=-\infty}^{\infty} c_j f_{-j}
\end{aligned} \tag{5.20}$$

L'expression $\sum_{j=-\infty}^{\infty} c_j f_{-j}$ est simplement le terme "0" de la convolution de $\{c_i\}$ avec $\{f_i\}$, soit, dans le domaine z .

$$\begin{aligned}
B(z) &= C(z)F(z) \\
&= \frac{F(z)F^*(z^{-1})}{F(z)F^*(z^{-1}) + N_0} \\
&= \frac{X(z)}{X(z) + N_0}
\end{aligned} \tag{5.21}$$

et donc :

$$\begin{aligned}
b_0 &= \frac{1}{2\pi j} \oint \frac{B(z)}{z} dz \\
&= \frac{1}{2\pi j} \oint \frac{X(z)}{z[X(z) + N_0]} dz
\end{aligned} \tag{5.22}$$

En passant dans le domaine de Fourier, on obtient :

$$b_0 = \frac{T}{2\pi} \int_{-\frac{\pi}{T}}^{\frac{\pi}{T}} \frac{X(e^{j\omega T})}{[X(e^{j\omega T}) + N_0]} d\omega \tag{5.23}$$

En réinjectant la valeur de b_0 dans (5.20), l'erreur quadratique minimale prend la forme :

$$\begin{aligned}
\xi_{\min} &= 1 - \frac{T}{2\pi} \int_{-\frac{\pi}{T}}^{\frac{\pi}{T}} \frac{X(e^{j\omega T})}{[X(e^{j\omega T}) + N_0]} d\omega \\
&= \frac{T}{2\pi} \int_{-\frac{\pi}{T}}^{\frac{\pi}{T}} \frac{N_0}{[X(e^{j\omega T}) + N_0]} d\omega \\
&= \frac{T}{2\pi} \int_{-\frac{\pi}{T}}^{\frac{\pi}{T}} \frac{N_0}{\frac{1}{T} \sum_{i=-\infty}^{\infty} |H(\omega + \frac{2\pi n}{T})|^2 + N_0} d\omega
\end{aligned} \tag{5.24}$$

Où l'on se rappelle de la liaison entre le spectre du signal échantillonné $X(e^{j\omega T})$, son homologue continu $X(\omega)$ et le canal $H(\omega)$.

En l'absence d'interférence entre symboles, on a

$$\xi_{\min} = \frac{N_0}{1 + N_0} \Rightarrow 0 \leq \xi_{\min} \leq 1 \tag{5.25}$$

D'autre part, ayant pris comme convention $S_0 = 1$, on a la relation

$$\gamma_\infty = \left(\frac{S}{N} \right)_\infty = \frac{1 - \xi_{\min}}{\xi_{\min}} \quad (5.26)$$

Cette relation reste valable en présence d'ISI.

On remarquera qu'en présence de bruit, la présence d'un "presque-zéro" dans le canal sera nettement moins catastrophique que dans le cas ZF.

5.2.3 Egaliseur MMSE-FIR

En adoptant un filtre de longueur $2K+1$, on peut recourir aux notations vectorielles :

$$\mathbf{C} = [c_{-K}, \dots, c_0, \dots, c_K]^T \quad (5.27)$$

$$\mathbf{v}_k = [v_{k+K}, \dots, v_k, \dots, v_{k-K}]^T \text{ ! ordre des indices} \quad (5.28)$$

On peut écrire la convolution :

$$\hat{I}_k = \sum_{j=-K}^K c_j v_{k-j} = \mathbf{C}^T \mathbf{v}_k \quad (5.29)$$

et donc le MSE

$$\xi(K) = \mathbb{E} \left\{ |I_k - \hat{I}_k|^2 \right\} = \mathbb{E} \left\{ |I_k - \mathbf{C}^T \mathbf{v}_k|^2 \right\} \quad (5.30)$$

et, par le principe d'orthogonalité, de la même manière que dans le cas IIR .

$$\sum_{j=-K}^K c_j \mathbb{E} \{ v_{k-j} v_{k-l}^* \} = \mathbb{E} \{ I_k v_{k-l}^* \} \quad -K \leq l \leq K \quad (5.31)$$

En notant

$$\mathbf{R}_{vv} = \mathbb{E} \left\{ \mathbf{v}_k \mathbf{v}_k^H \right\} = \begin{bmatrix} x_0 + N_0 & x_1 & \cdots & x_M \\ x_{-1} & x_0 + N_0 & \cdots & x_{M-1} \\ \vdots & \vdots & \ddots & \vdots \\ x_{-M} & x_{-M+1} & \cdots & x_0 + N_0 \end{bmatrix} \quad (5.32)$$

et

$$\mathbf{r}_{Iv} = [f_{-L}^*, \dots, f_0]^T = \mathbb{E} \{ I_k \mathbf{v}_k^* \} \quad (5.33)$$

On peut écrire les équations normales :

$$\mathbf{R}_{vv} \mathbf{C} = \mathbf{r}_{Iv} \quad (5.34)$$

où, pour rester cohérent avec les dimensions, \mathbf{R}_{vv} et \mathbf{r}_{Iv} sont de dimensions $(2K+1) \times (2K+1)$ et $(2K+1) \times 1$. On doit donc avoir $2K+1 \leq M$. L'égaliseur est donné par :

$$\mathbf{C}_{\text{opt}} = \mathbf{R}_{vv}^{-1} \mathbf{r}_{Iv} \quad (5.35)$$

Le MSE se détermine alors aisément :

$$\xi_{\min}(K) = 1 - \mathbf{r}_{Iv}^H \mathbf{R}_{vv}^{-1} \mathbf{r}_{Iv} \quad (5.36)$$

Egaliseur MMSE et filtrage de Wiener

On peut remarquer que le problème de l'égalisation MMSE peut se décrire par la figure 5.4.

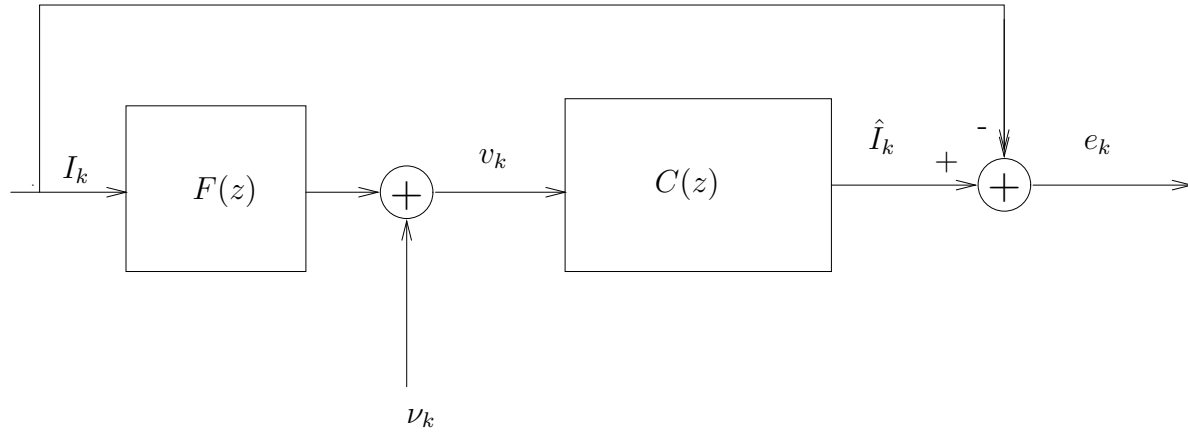


Figure 5.4: Equivalence entre égalisation MMSE et filtrage de Wiener

Ce qui correspond exactement au filtrage optimal de Wiener. Les solutions, ou algorithmes, que nous dérivons ci-dessous sont donc applicables pour les deux types de problème.

5.2.4 Algorithme de la descente de gradient

Quadraticité du critère

Dans ce paragraphe, nous allons démontrer que le critère MMSE est quadratique en l'erreur d'estimation des coefficients du filtre. Cette propriété est extrêmement intéressante, car elle assure l'existence d'un minimum global. En effet, si le critère était d'ordre plus élevé, la surface d'erreur pourrait présenter des minima locaux, et un algorithme itératif de recherche du minimum pourrait être piégé dans un minimum local.

Exprimons le MSE en fonction des corrélations :

$$\begin{aligned} \xi(K) &= E\{|e_k|^2\} \\ &= E\{(I_k - \mathbf{C}^T \mathbf{v}_k)(I_k^* - \mathbf{C}^H \mathbf{v}_k^*)\} \\ &= E\{|I_k|^2\} - E\{\mathbf{C}^T \mathbf{v}_k I_k^*\} - E\{I_k \mathbf{C}^H \mathbf{v}_k^*\} + E\{\mathbf{C}^T \mathbf{v}_k \mathbf{C}^H \mathbf{v}_k^*\} \end{aligned} \quad (5.37)$$

$\mathbf{C}^T \mathbf{v}_k$ étant scalaire, on a $\mathbf{C}^H \mathbf{v}_k^* = \mathbf{v}_k^H \mathbf{C}^*$, etc. D'autre part, on a indiqué que x_k était réel, donc \mathbf{R}_{vv} est réel et donc $\mathbf{R}_{vv} = \mathbf{R}_{vv}^*$. On peut donc continuer le développement.

$$\xi(K) = 1 - \mathbf{r}_{Iv}^H \mathbf{C}^* - \mathbf{C}^H \mathbf{r}_{Iv} + \mathbf{C}^H \mathbf{R}_{vv} \mathbf{C} \quad (5.38)$$

On peut alors écrire écrire $\xi(K)$ sous la forme quadratique

$$\xi(K) = \xi_{\min} + (\mathbf{C} - \mathbf{C}_{\text{opt}})^H \mathbf{R}_{vv} (\mathbf{C} - \mathbf{C}_{\text{opt}}) \quad (5.39)$$

à partir de cette expression, on retrouve aisément l'équation (5.38).

Nous pouvons donc utiliser un algorithme de descente du gradient qui nous permettra de trouver le minimum global du paraboloïde représentant le MSE $\xi(K)$.

Objectif

D'une part, nous voulons éviter l'inversion de \mathbf{R}_{vv} qui demanderait $O(M^3)$ opérations, où M est la dimension de \mathbf{R}_{vv} ($M = 2K + 1$). D'autre part, nous allons utiliser un algorithme itératif, qui est une première étape vers l'égalisation adaptative.

L'idée est relativement simple, on prend une estimation initiale, et on poursuit en en retranchant une quantité proportionnelle au gradient de l'erreur.

$$\mathbf{C}_{k+1} = \mathbf{C}_k - \frac{\mu}{2} \nabla_k \quad (5.40)$$

où

$$\nabla_k = \left. \frac{\partial \xi}{\partial \mathbf{C}} \right|_{\mathbf{C}=\mathbf{C}_k} = 2\mathbf{R}_{vv}\mathbf{C}_k - 2\mathbf{r}_{Iv} = -2\mathbf{R}_{vv}(\mathbf{C}_k - \mathbf{C}_{\text{opt}}) = -2\mathbf{R}_{vv}\tilde{\mathbf{C}}_k \quad (5.41)$$

où $\tilde{\mathbf{C}}_k$ représente l'erreur d'optimalité du filtre égaliseur. Ce qui mène à l'itération :

$$\mathbf{C}_{k+1} = \mathbf{C}_k - \mu(\mathbf{R}_{vv}\mathbf{C}_k - \mathbf{r}_{Iv}) = (\mathbf{I} - \mu\mathbf{R}_{vv})\mathbf{C}_k + \mu\mathbf{r}_{Iv} \quad (5.42)$$

ou encore :

$$\tilde{\mathbf{C}}_{k+1} = (\mathbf{I} - \mu\mathbf{R}_{vv})\tilde{\mathbf{C}}_k \quad (5.43)$$

Cette itération est une solution au problème de minimisation, où le gradient $0.5\nabla_k$ est le défaut d'optimalité de \mathbf{C}_k , utilisé pour se rapprocher de la solution.

Découplage des équations

L'équation (5.43) forme un système d'équations couplées, très difficile à étudier. En décomposant \mathbf{R}_{vv} en vecteurs propres : $\mathbf{R}_{vv} = \mathbf{W}\mathbf{\Lambda}\mathbf{W}^H = \sum_{i=1}^N \lambda_i \mathbf{W}_i \mathbf{W}_i^H$, avec $\mathbf{\Lambda} = \text{diag}[\lambda_1, \dots, \lambda_M]$ et $\mathbf{W} = [\mathbf{W}_1, \dots, \mathbf{W}_M]$. On peut alors utiliser le système transformé :

$$\begin{aligned} \mathbf{W}^H \tilde{\mathbf{C}}_{k+1} &= \mathbf{W}^H (\mathbf{I} - \mu\mathbf{R}_{vv}) \tilde{\mathbf{C}}_k \\ &= (\mathbf{W}^H - \mu\mathbf{W}^H \mathbf{W} \mathbf{\Lambda} \mathbf{W}^H) \tilde{\mathbf{C}}_k \\ &= (\mathbf{I} - \mu\mathbf{\Lambda}) \mathbf{W}^H \tilde{\mathbf{C}}_k \end{aligned} \quad (5.44)$$

où on a tenu compte de l'orthogonalité de la matrice des vecteurs propres : $\mathbf{W}^H \mathbf{W} = \mathbf{I} = \mathbf{W} \mathbf{W}^H$. De plus, sans perte de généralité, on peut faire l'hypothèse $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_M > 0$, la dernière inégalité supposant simplement que la matrice de corrélation est non singulière.

Nous avons maintenant un système d'équations découplées où les composantes d'erreur du système transformé sont :

$$(\mathbf{W}^H \tilde{\mathbf{C}}_k) \triangleq [w_1(k), w_2(k), \dots, w_M(k)]^T \quad (5.45)$$

Ce qui fournit les équation, pour le $i^{\text{ème}}$ mode naturel :

$$w_i(k+1) = (1 - \mu\lambda_i)w_i(k) \quad i = 1, \dots, M \quad (5.46)$$

Dont les solutions sont :

$$w_i(k+1) = (1 - \mu\lambda_i)^k w_i(0) \quad i = 1, \dots, M \quad (5.47)$$

C'est-à-dire M séries géométriques de raison $(1 - \mu\lambda_i)$, ce qui impose, pour assurer la convergence vers zéro :

$$-1 < (1 - \mu\lambda_i) < 1 \quad i = 1, \dots, M \quad (5.48)$$

Ce qui garantit la convergence de \mathbf{C}_k vers \mathbf{C}_{opt} . Avec la convention utilisée, les conditions précédentes peuvent encore s'exprimer :

$$0 < \mu < \frac{2}{\lambda_1} \leq \frac{2}{\lambda_i} \quad (5.49)$$

En général, on a pas aisément accès aux valeurs propres, alors que l'on peut facilement estimer la puissance σ_v^2 et, par les propriétés des valeurs propres :

$$\lambda_1 < \sum_{i=1}^M \lambda_i = \text{tr} \mathbf{R}_{vv} = N\sigma_v^2 \quad (5.50)$$

On obtient une borne plus simple :

$$0 < \mu < \frac{2}{N\sigma_v^2} < \frac{2}{\lambda_1} \quad (5.51)$$

En adoptant la durée d'une itération comme unité de temps, on peut associer une constante de temps τ_i à chaque mode naturel et exprimer la décroissance de la série géométrique :

$$\begin{aligned} 1 - \mu\lambda_i = e^{-\frac{1}{\tau_i}} &\Rightarrow \tau_i = -\frac{1}{\ln(1 - \mu\lambda_i)} \\ &\Rightarrow \tau_i \approx -\frac{1}{\mu\lambda_i} \quad \text{si } \mu \ll 1 \end{aligned} \quad (5.52)$$

Ce qui indique que la convergence sera d'autant plus lente que μ est faible.

A partir des relations $\mathbf{I} = \mathbf{W}\mathbf{W}^H$, on peut exprimer le filtre égaliseur en fonction des vecteurs propres de la matrice d'autocorrélation :

$$\tilde{\mathbf{C}} = \mathbf{W}\mathbf{W}^H\tilde{\mathbf{C}} = \mathbf{W}(\mathbf{W}^H\tilde{\mathbf{C}}) = \sum_{i=1}^N \mathbf{W}_i w_i \quad (5.53)$$

L'évolution du filtre égaliseur \mathbf{C}_k est alors, en fonction des modes naturels :

$$\begin{aligned} \mathbf{C}_k &= \mathbf{C}_{\text{opt}} - \tilde{\mathbf{C}}_k = \mathbf{C}_{\text{opt}} - \mathbf{W}(\mathbf{W}^H\tilde{\mathbf{C}}_k) \\ &= \mathbf{C}_{\text{opt}} - \sum_{i=1}^N \mathbf{W}_i w_i(k) \\ &= \mathbf{C}_{\text{opt}} - \sum_{i=1}^N \mathbf{W}_i w_i (0(1 - \mu\lambda_i)^k) \end{aligned} \quad (5.54)$$

La figure 5.5 illustre le processus de convergence de \mathbf{C}_k dans le référentiel des vecteurs propres de \mathbf{R}_{vv}

Tandis que l'évolution de l'erreur quadratique moyenne devient :

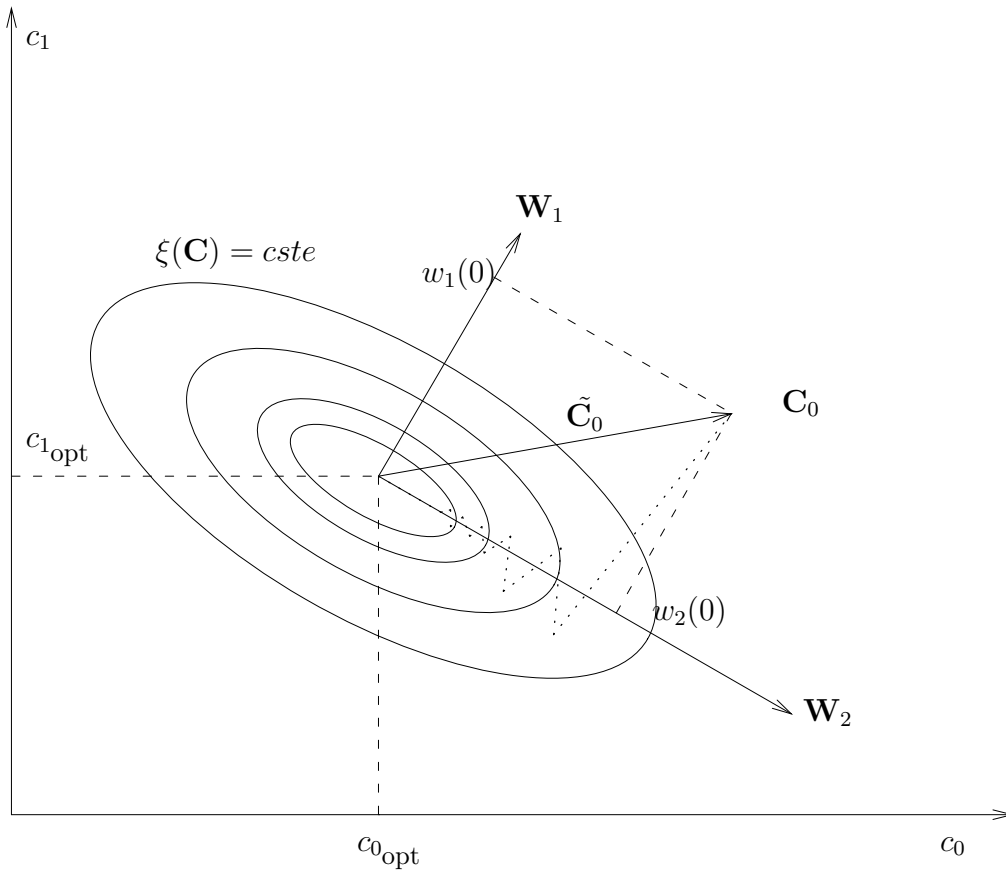


Figure 5.5: Convergence de \mathbf{C}_k dans le référentiel des vecteurs propres de \mathbf{R}_{vv}

$$\begin{aligned}
\xi(\mathbf{C}) &= \xi_{\text{opt}} + \tilde{\mathbf{C}}^H \mathbf{R}_{vv} \tilde{\mathbf{C}} \\
&= \xi_{\text{opt}} + \tilde{\mathbf{C}}^H \mathbf{W} \mathbf{\Lambda} \mathbf{W}^H \tilde{\mathbf{C}} \\
&= \xi_{\text{opt}} + [w_1^* \dots w_N^*] \begin{bmatrix} \lambda_1 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \lambda_N \end{bmatrix} \begin{bmatrix} w_1 \\ \vdots \\ w_N \end{bmatrix} \\
&= \xi_{\text{opt}} + \sum_{i=1}^N \lambda_i |w_i|^2
\end{aligned} \tag{5.55}$$

L'ensemble des points de fonction de coût constante $\xi(\mathbf{C}) = c'$ est donné par l'équation de l'ellipsoïde :

$$\sum_{i=1}^N \frac{|v_i|^2}{1/\lambda_i} = c = c' - \xi_{\text{opt}} \tag{5.56}$$

On peut encore écrire le MSE sous la forme :

$$\begin{aligned}
\xi_k \triangleq \xi(\mathbf{C}_k) &= \xi_{\text{opt}} + \sum_{i=1}^N \lambda_i |w_i(k)|^2 \\
&= \xi_{\text{opt}} + \sum_{i=1}^N \lambda_i |1 - \mu \lambda_i|^{2k} |w_i(0)|^2 \\
&\triangleq \xi_{\text{opt}} + \xi_k^e \geq \xi_{\text{opt}}
\end{aligned} \tag{5.57}$$

Le MSE est donc la somme du MMSE (ξ_{opt}) et d'un "Excess MSE" (ξ_k^e). La courbe représentant l'évolution de ξ_k en fonction du temps (de l'itération k) est appelée *courbe d'apprentissage*. En général, la courbe d'apprentissage de l'algorithme de descente du gradient consiste en une somme de N exponentielles, chacune de celles-ci correspondant à un mode naturel de l'algorithme. Si μ satisfait les conditions de convergences, le comportement asymptotique du MSE est :

$$\lim_{k \rightarrow \infty} \xi_k = \xi_{\text{opt}}, \lim_{k \rightarrow \infty} \xi_k^e = 0 \tag{5.58}$$

Dans le cas où les constantes de temps des modes naturels sont bien distincts ($\tau_1 \ll \tau_2 \ll \dots \ll \tau_N$), on peut observer l'allure de la courbe d'apprentissage sous une forme de droites (dans un repère logarithmique)

5.2.5 Algorithme du gradient stochastique (L.M.S. : Least Mean Square)

L'algorithme de la descente de gradient exige la connaissance de l'espérance mathématique, ou d'une approximation de celle-ci. L'algorithme du gradient stochastique repart de l'expression du MSE en "laissant tomber" l'opérateur espérance mathématique. La moyenne de l'erreur est donc approximée par l'erreur elle-même. Le nouveau critère est donc :

$$\xi(\mathbf{C}) = |e_k|^2 = |I_k - \mathbf{C}^H \mathbf{v}_k|^2 \tag{5.59}$$

Comme dans le cas de la descente de gradient, on définit l'estimée du gradient, ou gradient stochastique :

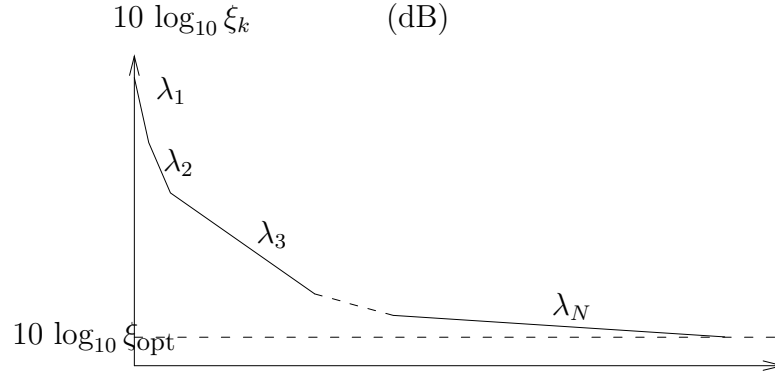


Figure 5.6: Courbe d'apprentissage

$$\hat{\nabla}_k = \left. \frac{\partial \xi_k}{\partial \mathbf{C}} \right|_{\mathbf{C}=\mathbf{C}_{k-1}} = \begin{bmatrix} \frac{\partial |e_k|^2}{\partial c_0} \\ \frac{\partial |e_k|^2}{\partial c_1} \\ \vdots \\ \frac{\partial |e_k|^2}{\partial c_{N-1}} \end{bmatrix} = 2e_k^p \begin{bmatrix} \frac{\partial e_k}{\partial c_0} \\ \frac{\partial e_k}{\partial c_1} \\ \vdots \\ \frac{\partial e_k}{\partial c_{N-1}} \end{bmatrix} = -2e_k^p \mathbf{v}_k \quad (5.60)$$

où

$$e_k^p \triangleq I_k - \mathbf{C}_{k-1}^H \mathbf{v}_k \quad (5.61)$$

c'est-à-dire l'erreur *a priori*.

Nous obtenons donc l'algorithme du type descente de gradient :

$$\begin{aligned} \mathbf{C}_k &= \mathbf{C}_{k-1} - \frac{\mu}{2} \hat{\nabla}_k^* \\ &= \mathbf{C}_{k-1} + \mu e_k^p \mathbf{v}_k^* \end{aligned} \quad (5.62)$$

Les équations (5.61) et (5.62) définissent l'algorithme LMS. Dans ce cas-ci, l'indice k , en plus de spécifier l'itération, est également lié au temps.

L'avantage principal de ce type d'algorithme est son extrême simplicité : il suffit de $2N$ opérations par échantillon, si on adopte un filtre égaliseur de longueur N . Par contre, le gradient étant stochastique, et donc bruité, on peut s'attendre à MSE bruité par rapport au MMSE et au MSE obtenu par la descente de gradient.

D'autre part, il faut se poser la question fondamentale de la convergence de l'algorithme vers le filtre égaliseur au sens MMSE.

Pour des raisons de temps, cette partie est simplement la photocopie d'une partie du cours de D. Slock.

Dérivées complexes, vecteurs

Cet appendice évoluera au cours de l'année.

.1 Dérivées complexes

Soit $f(w)$, une fonction d'une variable complexe $w = x + jy$, on définit les dérivées suivantes :

$$\frac{\partial}{\partial w} = \frac{1}{2} \left(\frac{\partial}{\partial x} - j \frac{\partial}{\partial y} \right) \quad (63)$$

$$\frac{\partial}{\partial w^*} = \frac{1}{2} \left(\frac{\partial}{\partial x} + j \frac{\partial}{\partial y} \right) \quad (64)$$

On déduit immédiatement les égalités suivantes :

$$\frac{\partial w}{\partial w} = 1 \quad ; \quad \frac{\partial w}{\partial w^*} = \frac{\partial w^*}{\partial w} = 0 \quad (65)$$

.2 Dérivées par rapport à un vecteur

On définit la dérivée par rapport à un vecteur $\mathbf{w} = [w_0 w_1 \cdots w_{N-1}]^T$ de la manière suivante :

$$\frac{\partial}{\partial \mathbf{w}} = \begin{bmatrix} \frac{\partial}{\partial w_0} \\ \frac{\partial}{\partial w_1} \\ \vdots \\ \frac{\partial}{\partial w_{N-1}} \end{bmatrix} \quad (66)$$

On peut donc définir une dérivée de vecteur ($\mathbf{a} = [a_0 a_1 \cdots a_{M-1}]^T$) par rapport à un vecteur :

$$\frac{\partial}{\partial \mathbf{w}} \mathbf{a} = \begin{bmatrix} \frac{\partial a_0}{\partial w_0} & \frac{\partial a_1}{\partial w_0} & \cdots & \frac{\partial a_{M-1}}{\partial w_0} \\ \frac{\partial a_0}{\partial w_1} & \frac{\partial a_1}{\partial w_1} & \cdots & \frac{\partial a_{M-1}}{\partial w_1} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial a_0}{\partial w_{N-1}} & \frac{\partial a_1}{\partial w_{N-1}} & \cdots & \frac{\partial a_{M-1}}{\partial w_{N-1}} \end{bmatrix} \quad (67)$$

On déduit immédiatement les égalités suivantes :

$$\frac{\partial \mathbf{w}}{\partial \mathbf{w}} = \mathbf{I} \quad ; \quad \frac{\partial \mathbf{w}}{\partial w^*} = \frac{\partial w^*}{\partial \mathbf{w}} = \mathbf{0} \quad (68)$$

Quelques égalités utiles dans le cours :

$$\frac{\partial |\mathbf{a}(\theta)|^2}{\partial \theta} = \frac{\partial \mathbf{a}^H(\theta) \mathbf{a}(\theta)}{\partial \theta} = \frac{\partial \mathbf{a}^H(\theta)}{\partial \theta} \mathbf{a}(\theta) + \mathbf{a}^H(\theta) \frac{\partial \mathbf{a}(\theta)}{\partial \theta} \quad (69)$$

$$= \mathbf{a}^T(\theta) \frac{\partial \mathbf{a}^*(\theta)}{\partial \theta} + \mathbf{a}^H(\theta) \frac{\partial \mathbf{a}(\theta)}{\partial \theta} \quad (70)$$

$$(71)$$

$$\frac{\partial |\mathbf{y} - \mathbf{X}\theta|^2}{\partial \theta} = -(\mathbf{y} - \mathbf{X}\theta)^H \mathbf{X} \quad (72)$$

Bibliography

- [BBC87] S. Benedetto, E. Biglieri, and V. Castellani. *Digital Transmission Theory*. Prentice-Hall, 1987.
- [Hay91] Simon Haykin. *Adaptive Filter Theory*. Prentice Hall, 2nd edition, 1991.
- [PRLN92] John G. Proakis, Charles M. Rader, Fuyun Ling, and C. L. Nikias. *Advanced Digital Signal Processing*. Mac Millan, 1992.
- [Pro89] John G. Proakis. *Digital Communications*. McGraw-Hill, 2nd edition, 1989.
- [Slo95] Dirk Slock. *Traitement numérique du signal*. Notes de cours, Institut EURECOM, 1995.

Contents

1	Introduction	1
2	Rappels.	3
2.1	Les canaux.	3
2.1.1	Le canal linéaire.	3
2.1.2	Le canal multichemins à évanouissement.	5
2.2	Le critère de Nyquist.	7
2.2.1	Notion d'Interférence Entre Symboles (ISI: Inter Symbol Interference). . .	7
2.2.2	Le critère de Nyquist.	8
2.2.3	Lien avec le théorème d'échantillonnage.	9
2.2.4	Le diagramme en oeil.	11
3	Filtres optimaux et prédiction linéaire.	15
3.1	Processus et modèles stochastiques.	15
3.1.1	Moyenne, autocorrélation et stationarité.	15
3.1.2	La matrice de corrélation.	16
3.1.3	Les innovations.	17
3.1.4	Modèles stochastiques (AR, MA, ARMA).	18
3.1.5	Les équations de Yule-Walker.	21
3.2	Le principe du maximum de vraisemblance.	24
3.2.1	La fonction de vraisemblance.	24
3.2.2	Inégalité de Cramer-Rao	25
3.3	Estimateur MMSE et principe d'orthogonalité.	25
3.4	Filtre de Wiener.	27
3.4.1	Filtre de Wiener non-causal	28
3.4.2	Filtre de Wiener causal.	31
3.4.3	Filtre de Wiener FIR	34
3.5	Prédiction linéaire.	35
3.5.1	Prédiction linéaire avant.	35
3.5.2	Prédiction linéaire arrière.	37
3.5.3	Relation entre prédiction avant et arrière	39
3.6	L'algorithme de Levinson-Durbin	39
3.6.1	Interprétations des paramètres K_m et Δ_{m-1}	41
3.7	Filtres en treillis.	41
3.8	La méthode des moindres carrés (LS : Least Squares)	43
3.8.1	Introduction	43
3.8.2	Fenêtrage.	43

3.8.3	Principe d'orthogonalité pour les moindres carrés	45
3.8.4	Equations normales	46
3.8.5	Interprétation géométrique.	47
3.8.6	Propriétés de l'estimation des moindres carrés.	48
4	Démodulateur optimal.	51
4.1	Démodulateur optimal.	51
4.2	Un modèle de canal	52
4.3	L'algorithme de Viterbi pour le MLSE	55
4.3.1	L'algorithme de Viterbi	55
4.3.2	Application à la recherche du MLSE	57
5	Egalisation Linéaire à filtre adapté	63
5.1	L'égalisateur Zero-Forcing (ZF)	63
5.2	Egaliseur MMSE linéaire	65
5.2.1	Dérivation de l'égaliseur MMSE-IIR	65
5.2.2	Erreur minimale et rapport signal/bruit.	66
5.2.3	Egaliseur MMSE-FIR	68
5.2.4	Algorithme de la descente de gradient	69
5.2.5	Algorithme du gradient stochastique (L.M.S. : Least Mean Square)	73
.1	Dérivées complexes	75
.2	Dérivées par rapport à un vecteur	75