Chapitre 3

Filtrage de Wiener

Dans ce chapitre, on considère des signaux à temps discrets et à valeurs complexes. Le choix de travailler avec des nombres complexes est utile pour les applications en télécommunication. Pour des signaux réels, les relations se simplifieront et nous donneront fréquemment ces relations simplifiées. Le filtrage de Wiener doit son nom au scientifique Norbert Wiener (figure 3.1), né au USA le 26 novembre 1894 et décédé à Stockholm, en Suède, le 18 mars 1964.

3.1 Enoncé du problème

On dispose d'une série d'observations :

$$u(0), u(1), \ldots \tag{3.1}$$

et d'un filtre linéaire discret dont les coefficients sont :

$$w_0, w_1, \ldots \tag{3.2}$$

et qui fournit un signal de sortie $y(n)$:

$$y(n) = w_0^* u(n) + w_1^* u(n-1) + \ldots \tag{3.3}$$

L'objectif est d'estimer les paramètres w_k de sorte que $y(n)$ soit une bonne estimation d'une sortie désirée $d(n)$. On définit $e(n)$, l'erreur d'estimation à l'instant n :

$$e(n) = d(n) - y(n). \tag{3.4}$$

La solution est recherchée avec les contraintes suivantes :

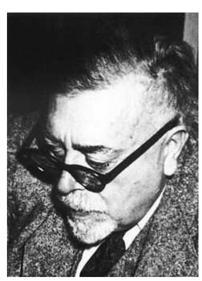

FIGURE 3.1 – Norbert Wiener, inventeur du filtrage optimal et du filtre de Wiener

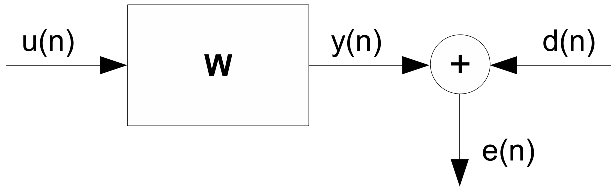

FIGURE 3.2 – Représentation schématique d'un filtre de Wiener

- le filtre W est linéaire, à temps discret,
- le critère est le minimum de l'erreur quadratique moyenne (MMSE pour *Minimum Mean Square Error*), c'est-à-dire le minimum de $E[|e(n)|^2]$.

Ce problème peut être associé au schéma de la figure 3.2.

Equations de base

Selon la figure 3.2, avec les hypothèses de linéarité et de temps discret de $w(n)$, on peut écrire :

$$y(n) = \sum_{k=0}^{+\infty} w_k^* u(n-k),\ n = 0, 1, \ldots \tag{3.5}$$

où w_k^* est le complexe conjugué de w_k. On peut voir l'équation (3.5) comme le produit scalaire dans l'espace de Hilbert de dimension infinie :

$$< \mathbf{u}(n), \mathbf{w} >= \sum_{k=0}^{+\infty} u(n-k) w_k^*, \tag{3.6}$$

avec $\mathbf{u}(n) = (u(n)u(n-1)\ldots)^T$ et $\mathbf{w} = (w_0 w_1 \ldots)$.

L'erreur d'estimation à l'instant n est égale $e(n) = d(n) - y(n)$. L'erreur quadratique moyenne, J, est la fonction de coût que l'on cherche à minimiser :

$$J = E[e(n)e(n)^*] = E[|e(n)|^2]. \tag{3.7}$$

Optimisation

Compte tenu de la nature complexe des données, on peut écrire :

$$w_k = a_k + j b_k, \quad k = 0, 1, \ldots \tag{3.8}$$

et on peut définir un opérateur de gradient complexe :

$$\nabla_k = \frac{\partial}{\partial a_k} + j \frac{\partial}{\partial b_k}, \quad k = 0, 1, \ldots \tag{3.9}$$

Appliqué à la fonction de coût J, on a :

$$\nabla_k J = \frac{\partial J}{\partial a_k} + j \frac{\partial J}{\partial b_k}, \quad k = 0, 1, \ldots \tag{3.10}$$

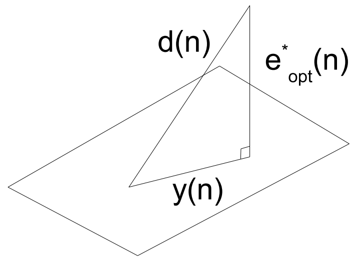

FIGURE 3.3 – Représentation géométrique du théorème de la projection orthogonale

Puisque la fonction est l'espérance d'un produit, $J = E[e(n)e^*(n)]$, en utilisant la propriété de linéarité de l'espérance et les règles de dérivation d'un produit, on a :

$$\nabla_k J = E\Big[e^*(n)\frac{\partial e(n)}{\partial a_k} + e(n)\frac{\partial e^*(n)}{\partial a_k} + je^*(n)\frac{\partial e(n)}{\partial b_k} + je(n)\frac{\partial e^*(n)}{\partial b_k}\Big]. \tag{3.11}$$

En exprimant les termes $e(n)$ et $e^*(n)$ en fonction de a_k et b_k, et en dérivant, on trouve après des calculs algébriques simples :

$$\begin{cases} \frac{\partial e(n)}{\partial a_k} &= -u(n-k) \\ \frac{\partial e^*(n)}{\partial a_k} &= -u^*(n-k) \\ \frac{\partial e(n)}{\partial b_k} &= +ju(n-k) \\ \frac{\partial e^*(n)}{\partial b_k} &= -ju^*(n-k). \end{cases} \tag{3.12}$$

En reportant les résultats (3.12) dans l'équation (3.11), on a simplement :

$$\nabla_k J = -2E[u(n-k)e^*(n)],\ k = 0, 1, \ldots \tag{3.13}$$

La solution optimale correspond à une erreur quadratique minimale, notée $e_{opt}(n)$, telle que :

$$\nabla_k J = 0,\ k = 0, 1, \ldots \tag{3.14}$$

ou encore :

$$E[u(n-k)e^*_{opt}(n)] = 0,\ k = 0, 1, \ldots \tag{3.15}$$

On remarque que l'erreur optimale est orthogonale à toutes les entrées $u(n)$, $u(n-1)$, … On retrouve ainsi le théorème de la projection orthogonale (figure 3.3).

En corrolaire à cette remarque, l'approximation optimale, notée $y_{opt}(n)$, qui s'écrit :

$$y_{opt}(n) = \sum_{k=0}^{+\infty} w^*_{opt} u(n-k), \tag{3.16}$$

est donc une combinaison linéaire de termes $u(n-k)$ qui sont tous orthogonaux à $e_{opt}(n)$. Par conséquent, $y_{opt}(n)$ est elle-même orthogonale à $e_{opt}(n)$, c'est-à-dire que l'on peut écrire :

$$E[y_{opt}(n)e^*_{opt}(n)] = 0. \tag{3.17}$$

Notations

Soit $\mathcal{U}_n$ l'espace engendré par tous les $u(k)$, $k \leq n$, on notera $\hat{d}(n/\mathcal{U}_n)$ l'estimation optimale au sens de l'erreur quadratique moyenne minimale (MMSE), obtenue à partir des données jusqu'à l'instant n inclus. D'après les résultats précédents, on peut sonc écrire :

$$y_{opt}(n) = \hat{d}(n/\mathcal{U}_n). \tag{3.18}$$

Calcul de l'erreur quadratique minimale

On rappelle que $e(n) = d(n) - y(n)$ et qu'à l'optimum, on a :

$$e_{opt}(n) = d(n) - y_{opt}(n) = d(n) - \hat{d}(n/\mathcal{U}_n). \tag{3.19}$$

La valeur minimale de l'erreur quadratique moyenne est alors :

$$J_{min} = E[|e_{opt}(n)|^2]. \tag{3.20}$$

D'après (3.19), on peut écrire $d(n) = \hat{d}(n/\mathcal{U}_n) + e_{opt}(n)$. De plus, on sait que $\hat{d}(n/\mathcal{U}_n) = y_{opt}(n) \perp e_{opt}(n)$. On en déduit donc par le théorème de Pythagore l'égalité :

$$E[|d(n)|^2] = E[\left|\hat{d}(n/\mathcal{U}_n)\right|^2] + E[|e_{opt}(n)|^2] \tag{3.21}$$

c'est-à-dire :

$$\sigma_d^2 = \sigma_{\hat{d}}^2 + J_{min}. \tag{3.22}$$

On en déduit finalement :

$$J_{min} = \sigma_d^2 - \sigma_{\hat{d}}^2. \tag{3.23}$$

En pratique, on normalise l'expression en divisant par σ_d^2, d'où l'erreur quadratique moyenne normalisée minimale :

$$\epsilon = \frac{J_{min}}{\sigma_d^2} = 1 - \left(\frac{\sigma_{\hat{d}}}{\sigma_d}\right)^2. \tag{3.24}$$

On remarque que cette erreur normalisée vérifie $0 \leq \epsilon \leq 1$. Si $\epsilon = 0$, on a alors $\sigma_{\hat{d}}^2 = \sigma_d^2$, ce qui correspond à une adéquation parfaite entre d et $\hat{d}$.

3.2 Equation de Wiener-Hopf

3.2.1 Expression générale

A partir du principe d'orthogonalité (3.15), de la définition de $e_{opt}(n)$ et de $y(n)$, l'équation (3.15) devient :

$$E[u(n-k)(d^*(n) - \sum_j w_{opt,j} u^*(n-j))] = 0,\ k = 0, 1, \ldots \tag{3.25}$$

En développant l'espérance, on arrive à :

$$E[u(n-k)(d^*(n)] = \sum_j w_{opt,j} E[u(n-k)u^*(n-j)]\ k = 0, 1, \ldots \tag{3.26}$$

Or, pour un signal $u(n)$ stationnaire, le terme $E[u(n-k)u^*(n-j)]$ n'est autre que le terme $R_{uu}(j-k)$ de la matrice d'auto-corrélation. Le terme $E[u(n-k)(d^*(n)]$ est l'inter-corrélation entre l'entrée du filtre et la sortie désirée, que l'on note $R_{ud}(-k)$. L'équation (3.26) peut donc s'écrire :

$$\sum_j w_{opt,j} R_{uu}(j-k) = R_{ud}(-k),\ k = 0, 1, \ldots \tag{3.27}$$

Cet ensemble infini d'équations porte le nom d'équations de Wiener-Hopf.

Solution pour un filtre FIR

Si on suppose que le filtre est à réponse impulsionnelle finie (RIF) de longueur M, on peut écrire les équations (3.27) sous forme d'un système de M équations :

$$\sum_{j=0}^{M} w_{opt,j} R_{uu}(j-k) = R_{ud}(-k),\ k = 0, 1, \ldots, M-1 \tag{3.28}$$

En posant les notations $\mathbf{u}(n) = (u(n)u(n-1)\ldots u(n-M+1))^T$ et $\mathbf{R}_{ud} = (R_{ud}(0)R_{ud}(-1)\ldots R_{ud}(-M+1))^T$ (on rappelle que $\mathbf{R}_{ud} = E[d^*(n)\mathbf{u}(n)]$) et en remarquant que la matrice d'auto-corrélation s'écrit :

$$\mathbf{R_{uu}} = E[\mathbf{u}(n)\mathbf{u}^H(n)] = \begin{pmatrix} R_{uu}(0) & R_{uu}(2) & \ldots & R_{uu}(M-1) \\ R^*_{uu}(1) & R_{uu}(0) & & \vdots \\ \vdots & & & \\ R^*_{uu}(M-1) & R^*_{uu}(M-2) & & R_{uu}(0) \end{pmatrix} \tag{3.29}$$

les équations de Wiener-Hopf s'écrivent sous la forme matricielle :

$$\mathbf{R_{uu}}\mathbf{w}_{opt} = \mathbf{R}_{ud}, \tag{3.30}$$

avec $\mathbf{w}_{opt} = (w_{opt,0}w_{opt,1}\ldots w_{opt,M-1})^T$. Si la matrice d'auto-corrélation $\mathbf{R_{uu}}$ est régulière, $\mathbf{R_{uu}^{-1}}$ existe et la solution optimale est donnée par :

$$\mathbf{w}_{opt} = \mathbf{R_{uu}^{-1}}\mathbf{R}_{ud}. \tag{3.31}$$

L'estimation de $\mathbf{w}_{opt}$ requiert donc la connaissance de la matrice d'auto-corrélation (c'est-à-dire des statistiques à l'ordre 2) des entrées du filtre et de l'inter-corrélation entre les entrées du filtre et la sortie désirée. Dans le cas d'un filtre d'ordre M (M coefficients), il faut donc estimer d'abord les $M(M+1)/2$ termes $\mathbf{R_{uu}}$ et les M coefficients de $\mathbf{R}_{ud}$.

Cas de signaux réels

Dans le cas de signaux réels, il suffit de remplacer les termes conjugués par les termes simples et ^{H} par ^{T} :

$$\begin{aligned} E[\mathbf{u}(n)\mathbf{u}^H(n)] &\rightarrow E[\mathbf{u}(n)\mathbf{u}^T(n)] \\ E[\mathbf{u}(n)d^*(n)] &\rightarrow E[\mathbf{u}(n)d(n)] \\ E[|e(n)|^2] &\rightarrow E[e^2(n)] \\ \mathbf{R_{uu}} &= \begin{pmatrix} R_{uu}(0) & R_{uu}(2) & \ldots & R_{uu}(M-1) \\ R^*_{uu}(1) & R_{uu}(0) & & \vdots \\ \vdots & & & \\ R^*_{uu}(M-1) & R^*_{uu}(M-2) & & R_{uu}(0) \end{pmatrix} \end{aligned} \tag{3.32}$$

3.3.3 Forme canonique

En faisant la différence (3.34) et (3.38), on a :

$$\begin{aligned} J - J_{min} &= -\mathbf{p}^H \mathbf{w} - \mathbf{w}\mathbf{p} + \mathbf{w}^H \mathbf{R_{uu}} \mathbf{w} + \mathbf{p}^H \mathbf{w}_{opt} \\ &= -\mathbf{p}^H (\mathbf{w} - \mathbf{w}_{opt}) - \mathbf{w}^H \mathbf{R_{uu}} (\mathbf{w}_{opt} - \mathbf{w}) \\ &= \mathbf{w}_{opt}^H \mathbf{R}_{\mathbf{uu}}^H (\mathbf{w} - \mathbf{w}_{opt}) - \mathbf{w}^H \mathbf{R_{uu}} (\mathbf{w}_{opt} - \mathbf{w}). \end{aligned} \tag{3.39}$$

Or, $\mathbf{R_{uu}} = \mathbf{R}_{\mathbf{uu}}^H$ en raison de la symétrie hermitienne, donc on a :

$$J = J_{min} + (\mathbf{w} - \mathbf{w}_{opt})^H \mathbf{R_{uu}} (\mathbf{w} - \mathbf{w}_{opt}). \tag{3.40}$$

Cette expression montre l'unicité de la solution optimale, puisque la forme hermitienne (ou quadratique dans le cas de signaux réels) ne s'annulle que pour $\mathbf{w} = \mathbf{w}_{opt}$.
Puisque $\mathbf{R_{uu}}$ est une matrice de corrélation à symétrie hermitienne, on peut la décomposer dans sa base de vecteurs propres et l'écrire sous la forme :

$$\mathbf{R_{uu}} = \mathbf{Q} \mathbf{\Lambda} \mathbf{Q}^H, \tag{3.41}$$

où $\mathbf{\Lambda}$ est une matrice diagonale (des valeurs propres) et $\mathbf{Q}$ est la matrice des vecteurs propres. L'équation (3.40) devient alors :

$$J = J_{min} + (\mathbf{w} - \mathbf{w}_{opt})^H \mathbf{Q} \mathbf{\Lambda} \mathbf{Q}^H (\mathbf{w} - \mathbf{w}_{opt}). \tag{3.42}$$

En posant $\mathbf{v}^H = (\mathbf{w} - \mathbf{w}_{opt})^H \mathbf{Q}$ (ce qui correspond à un changement d'origine et d'échelle), on a :

$$J = J_{min} + \mathbf{v}^H \mathbf{Q} \mathbf{\Lambda} \mathbf{v}. \tag{3.43}$$

Puisque $\mathbf{\Lambda}$ est une matrice diagonale constituée d'éléments λ_i, on a finalement :

$$J = J_{min} + \sum_i \lambda_i v_i v_i^* = J_{min} + \sum_i \lambda_i \left| v_i \right|^2. \tag{3.44}$$

Cette dernière expression met en évidence les axes principaux $\mathbf{v}$.