TD de filtrage adaptatif

1 Introduction : filtrage de Wiener

On désire reconstruire un *signal d'intérêt*, non directement observable issu d'un processus aléatoire $\{d(n)\}_{n\geq 0}$ à partir d'un *signal de référence* issu d'un processus aléatoire $\{x(n)\}_{n\geq 0}$ corrélé avec $\{d(n)\}_{n\geq 0}$. On dispose par ailleurs d'une seconde observation, issue du processus aléatoire $\{y(n)\}_{n\geq 0}$ lui-même défini comme la somme de $\{d(n)\}_{n\geq 0}$ et d'un bruit perturbateur $\{b(n)\}_{n\geq 0}$:

$$y(n) = d(n) + b(n) \tag{1.1}$$

On se propose par ailleurs d'estimer le signal d'intérêt par filtrage FIR du signal de référence. Cela revient à considérer que pour tout entier naturel n, l'estimée de $d(n)$ peut s'écrire sous la forme suivante :

$$\hat{d}(n) = \sum_{m=0}^{N-1} h(m)x(n-m) = \boldsymbol{h}^{\mathrm{T}}\boldsymbol{x}(n) \tag{1.2}$$

où $\boldsymbol{h}$ et $\boldsymbol{x}(n)$ sont deux vecteurs de longueur N définis par :

$$\boldsymbol{h} = \begin{bmatrix} h(0) & \cdots & h(N-1) \end{bmatrix}^{\mathrm{T}} \qquad \boldsymbol{x}(n) = \begin{bmatrix} x(n) & \cdots & h(n-N+1) \end{bmatrix}^{\mathrm{T}} \tag{1.3}$$

L'Erreur Quadratique Moyenne (EQM) de reconstruction en sortie du filtre est donnée par la quantité :

$$E\left[d(n) - \hat{d}(n)\right] \tag{1.4}$$

Il est possible de montrer que minimiser cette dernière est équivalent à minimiser l'EQM suivante :

$$\overline{\varepsilon_d^2}(\boldsymbol{h}) = E\left[y(n) - \hat{d}(n)\right] = \sigma_y^2 + \boldsymbol{h}^{\mathrm{T}}\boldsymbol{\Gamma}_x\boldsymbol{h} - 2\boldsymbol{\Gamma}_{x,y}^{\mathrm{T}}\boldsymbol{h} \tag{1.5}$$

où $\boldsymbol{\Gamma_x}$=E[$\boldsymbol{x}(n)\boldsymbol{x}(n)^{\mathrm{T}}$] est la matrice d'autocorrélation du vecteur $\boldsymbol{x}(n)$, et $\boldsymbol{\Gamma_{x,y}}$ =E[$\boldsymbol{x}(n)y(n)$] le vecteur d'intercorrélation entre le vecteur $\boldsymbol{x}(n)$ et l'observation bruitée $y(n)$. Notons que les précités processus aléatoires sont supposés être stationnaires au sens large à l'ordre 2.

Le gradient de l'EQM vis-à-vis du vecteur $\boldsymbol{h}$ est donné par la relation suivante :

$$\nabla\overline{\varepsilon_d^2}(\boldsymbol{h}) = 2\boldsymbol{\Gamma}_x\boldsymbol{h} - 2\boldsymbol{\Gamma}_{x,y} \tag{1.6}$$

Le filtre optimal au sens de l'EQM minimale, dit filtre de Wiener, noté $\boldsymbol{h_o}$ est alors obtenu en minimisant et par la même occasion annulant ce gradient, ce qui mène à :

$$\boldsymbol{h_o} = \boldsymbol{\Gamma}_x^{-1}\boldsymbol{\Gamma}_{x,y} \tag{1.7}$$

Remarque : le second signal d'observation issu du processus aléatoire $\{y(n)\}_{n\geq 0}$ n'est utilisé ici que pour l'apprentissage de la matrice d'intercorrélation $\boldsymbol{\Gamma_{x,y}}$. Néanmoins, une fois cette dernière connue, seul le signal de référence est nécessaire pour estimer le signal d'intérêt. Une manière plus optimale de procéder serait d'exploiter conjointement le second signal d'observation issu de $\{y(n)\}_{n\geq 0}$ et le signal de référence, ce qui est plus connu sous le nom de filtrage de Wiener vectoriel.

2 Algorithme du gradient ou "de la plus grande pente" (gradient déterministe)

L'algorithme du gradient à pas fixe est défini par :

$$\begin{cases} z(0) = (\cdots) & \text{(initialisation)} \\ z(n+1) = z(n) - \mu \nabla f\left(z(n)\right) \end{cases} \tag{1.8}$$

où $f : z \rightarrow f(z)$ est une fonction à minimiser.

> Donner l'expression de l'algorithme du gradient permettant d'estimer le vecteur $\boldsymbol{h}$ en prenant comme fonction f l'EQM définie par l'équation (1.5).

Par définition, l'algorithme $\{\hat{\boldsymbol{h}}(n+1) = F(\hat{\boldsymbol{h}}(n)), n=0,1,\ldots\}$ est dit stable (convergent) si pour tout vecteur $\hat{\boldsymbol{h}}(0)$, $\hat{\boldsymbol{h}}(n)$ tend vers le vecteur $\boldsymbol{h}_o$ quand n tend vers l'infini.

> Mettre l'algorithme précédent sous la forme $\tilde{\boldsymbol{h}}(n+1) = \boldsymbol{A}\,\tilde{\boldsymbol{h}}(n)$ où $\tilde{\boldsymbol{h}}(n) = \hat{\boldsymbol{h}}(n) - \boldsymbol{h}_o$, et expliquer comment se traduit sur $\tilde{\boldsymbol{h}}(n)$ l'hypothèse de stabilité.

On suppose que la matrice d'autocorrélation $\boldsymbol{\Gamma_x}$ est diagonale, autrement dit $\boldsymbol{\Gamma_x} = \text{diag}([\lambda_1, \lambda_2, \ldots, \lambda_N])$.

> Expliquer pourquoi les composantes diagonales λ_i de $\boldsymbol{\Gamma_x}$ sont toutes positives. Puis, pour un vecteur initial $\hat{\boldsymbol{h}}(0)$ arbitraire, donner l'expression de $\tilde{h}_i(n)$, $i^{\text{ème}}$ composante de $\tilde{\boldsymbol{h}}(n)$, en fonction de n, de $h_{o,i}$ et de λ_i.

On considère maintenant que $\boldsymbol{\Gamma_x}$ est inversible mais non diagonale. $\boldsymbol{\Gamma_x}$ étant symétrique elle est diagonalisable, c'est-à-dire, elle admet la décomposition $\boldsymbol{\Gamma_x} = \boldsymbol{P\Lambda P}^{-1}$ où $\boldsymbol{\Lambda}$ est la matrice diagonale des valeurs propres $\{\lambda_i\}_{i=1,\ldots,N}$ de $\boldsymbol{\Gamma_x}$ et $\boldsymbol{P}$ la matrice des vecteurs propres (colonnes) associés. D'autre part, le caractère symétrique de $\boldsymbol{\Gamma_x}$ nous assure qu'il est possible de trouver une matrice de passage $\boldsymbol{P}$ *orthonormée* (ou *unitaire*), c'est-à-dire, vérifiant la propriété $\boldsymbol{PP}^{\mathrm{T}} = \boldsymbol{P}^{\mathrm{T}}\boldsymbol{P} = \mathbf{I}_N$. Il en résulte l'égalité $\boldsymbol{P}^{\mathrm{T}} = \boldsymbol{P}^{-1}$.

> En utilisant le changement de variable $\boldsymbol{\hbar}(n) = \boldsymbol{P}^{\mathrm{T}}\,\tilde{\boldsymbol{h}}(n)$, montrer que l'on obtient $\boldsymbol{\hbar}(n+1) = \boldsymbol{B}\,\boldsymbol{\hbar}(n)$, où $\boldsymbol{B}$ est une matrice diagonale. Donner alors la condition de stabilité sur μ en fonction des valeurs propres de $\boldsymbol{\Gamma_x}$.

Le changement de variable précédent, nous permet d'écrire l'égalité $\tilde{\boldsymbol{h}}(n) = \boldsymbol{P}\,\boldsymbol{\hbar}(n)$, et de ce fait d'écrire la $i^{\text{ème}}$ composante de $\tilde{\boldsymbol{h}}(n)$ en fonction d'éléments de $\boldsymbol{P}$, $\boldsymbol{B}$ et $\tilde{\boldsymbol{h}}(0)$.

> Montrer qu'il est possible d'écrire $\tilde{h}_i(n)$ sous la forme $\tilde{h}_i(n) = \sum_{j=1}^{N} \left(1 - 2\mu\lambda_j\right)^n \alpha_{ij}$. Vérifier ainsi que la stabilité de l'algorithme dépend explicitement de la quantité $\max_i\{|1-2\mu\lambda_i|\}$.

3 Etude de l'algorithme du gradient stochastique (LMS)

On considère l'erreur quadratique instantanée suivante :

$$\varepsilon_d^2(\boldsymbol{h})(n) = \left(y(n) - \boldsymbol{h}^{\mathrm{T}}\boldsymbol{x}(n)\right)^2 \qquad (1.9)$$

> Développer l'erreur quadratique précédente de manière analogue à l'équation (1.5), puis calculer son gradient par rapport au vecteur $\boldsymbol{h}$ pris au point $\hat{\boldsymbol{h}}(n)$. Donner alors l'expression de l'algorithme correspondant. Pourquoi peut-on qualifier cet algorithme de gradient stochastique ?

Supposons que $\{\hat{\boldsymbol{h}}(n)\}_{n\geq 0}$ soit un processus aléatoire vectoriel indépendant du processus vectoriel $\{\boldsymbol{x}(n)\}_{n\geq 0}$.

> Quelle est la relation mathématique liant l'erreur quadratique précédente à celle définie par l'équation (1.5) ?

L'algorithme du gradient stochastique se déduit de l'algorithme du gradient déterministe en substituant dans l'équation (1.8) la quantité $\nabla f\left(\hat{\boldsymbol{h}}(n)\right)$ par le gradient de l'équation (1.11) pris au point $\hat{\boldsymbol{h}}(n)$ et en remplaçant les quantités aléatoires $y(n)$ et $\boldsymbol{x}(n)$ par les réalisations déterministes correspondantes.

> Ecrire le système d'équations définissant la méthode du gradient stochastique nommé également algorithme LMS (Least Mean Square). Pour quelle raison qualifie-t-on cette méthode de stochastique ? Montrer par ailleurs que l'étude de convergence de la quantité $\mathrm{E}[\hat{\boldsymbol{h}}(n)]$ où les vecteurs $\hat{\boldsymbol{h}}(n)$ sont obtenus à l'aide de la méthode LMS, se ramène à celle effectuée dans la section précédente.