ANNALES DE L'I. H. P., SECTION B

ODILE MACCHI

Résolution adaptative de l'équation de Wiener-Hopf. Cas d'un canal de données affecté de gigue

Annales de l'I. H. P., section B, tome 14, n° 3 (1978), p. 355-377

<http://www.numdam.org/item?id=AIHPB_1978__14_3_355_0>

Ann. Inst. Henri Poincaré, Section B :
Vol. XIV, n° 3, 1978, p. 355-377. *Calcul des Probabilités et Statistique.*

Résolution adaptative de l'équation de Wiener-Hopf. Cas d'un canal de données affecté de gigue

par

Odile MACCHI (*)

Laboratoire des Signaux et Systèmes C. N. R. S.-E. S. E.,
Plateau du Moulon, 91190 Gif-sur-Yvette

SOMMAIRE ([1]). — On considère la convergence et on discute les propriétés d'adaptativité au contexte statistique, d'un algorithme d'itération stochastique très fréquemment utilisé dans la pratique pour optimiser des estimateurs linéaires, et on généralise les démonstrations aux grandeurs complexes. L'algorithme d'adaptation utilise l'écart entre le message vrai et son estimation actuelle. Il peut être à pas décroissant ou constant. Lorsque le pas est constant, on discute le choix du pas optimal.

Ces résultats sont appliqués à la transmission sur un canal filtrant affecté de bruit additif et d'une erreur de restitution de phase (gigue de phase ou dérive de fréquence) comme c'est le cas en transmission de données. On discute la validité de la théorie dans ce cas pratique.

SUMMARY. — A stochastic approximation algorithm is theoretically analyzed. It is often used in practice for optimizing linear estimation. Its convergence is proved and its adaptivity is discussed. The proofs are extended to the case of complex quantities. The algorithm is based on the difference between the actual message and its present estimate. The step size is either fixed or decreasing. In the former case, we derive the optimal step size.

(*) Une partie de cet article a fait l'objet d'une conférence [*10*] au 6e Colloque National sur le Traitement du Signal et ses Applications, Nice, 26 au 30 avril 1977.

These results are applied to communication over a noisy filtering channel affected by an imperfect phase recovery (phase jitter or frequency drift), as in data transmission. The validity of the algorithm is then discussed.

I. INTRODUCTION

Lorsqu'on veut estimer un message a par l'intermédiaire d'un signal reçu x, dépendant de a, la méthode la plus usuelle est de choisir une estimation y qui soit linéaire par rapport à x. On a

$$y = hx \tag{1}$$

où l'opérateur h qui caractérise l'estimateur est supposé de type intégral. Pour optimiser cet estimateur, on choisit habituellement le critère de l'erreur quadratique moyenne minimum. On doit alors résoudre une équation de Wiener-Hopf du type

$$\mathrm{R}_{xx}h = \mathrm{R}_{xa} \tag{2}$$

où R_{xx} et R_{xa} sont les covariances croisées de x et de a. Il existe de nombreuses techniques pour résoudre cette équation, mais la plupart sont compliquées à mettre en œuvre, même dans le cas de signaux discrets. L'une d'elle cependant est très simple. C'est un algorithme qui résout itérativement l'équation (2) à l'aide d'une suite d'estimateurs h_k, de messages et de signaux observés :

$$h_{k+1} = h_k + \mu(a_k - y_k)x_k \tag{3}$$

où μ est un nombre positif appelé pas d'incrémentation.

Dans la pratique, l'algorithme (3) atteint l'optimalité en quelques dizaines d'itérations. De ce fait, et vu son extrême simplicité, il est de plus en plus utilisé dans les systèmes modernes numériques de communications (radar, sonar, télécommunications). Or l'algorithme (3) est un algorithme d'itération stochastique. Des théorèmes généraux existent dans ce domaine [*1*]-[*4*]. Mais ils comportent un grand nombre d'hypothèses, dont il est en général très difficile de vérifier la validité sur des exemples particuliers. Dans le cas de l'algorithme (3), la convergence n'a jamais été établie. Dans les articles qui utilisent cet algorithme, les auteurs se contentent généralement de dire que la convergence a été vérifiée sur des simulations.

Il existe aussi une version à pas décroissants μ_k de l'algorithme (3) :

$$h_{k+1} = h_k + \mu_k(a_k - y_k)x_k, \tag{4}$$

dont la convergence est plus fine.

Enfin les algorithmes (3) et (4) sont depuis peu utilisés sous une forme complexe (x^* désignant le complexe conjugué de x) :

$$h_{k+1} = h_k + \mu_k(a_k - y_k)x_k^* . \tag{5}$$

En effet les systèmes modernes de communications émettent fréquemment non plus un mais deux messages réels a_1 et a_2 sur deux porteuses en quadrature [5], et la réception comporte alors deux signaux réels en quadrature x_1 et x_2. Dans ce cas, message et signal deviennent des grandeurs complexes.

$$a = a^1 + ia^2 ; \qquad x = x^1 + ix^2 , \tag{6}$$

ainsi que l'estimateur h et l'estimation y.

Il était donc utile d'établir les conditions de convergence des algorithmes (3), (4) et de la forme complexe (5) et de fournir la preuve théorique de leur convergence.

Dans ce qui suit, nous montrons que les algorithmes (3) et (4) convergent sous les mêmes conditions mais sur des modes différents. En contrepartie de sa convergence plus fine, l'algorithme (4) a une moins bonne capacité d'adaptation aux variations de statistique du signal x et du message a, que l'algorithme (3).

Dans le cas de l'algorithme (3), nous discutons le choix du pas d'incrémentation μ optimal. Il est bien connu des auteurs qui utilisent ce type d'algorithme que μ doit être d'autant plus petit que l'estimateur utilise un plus grand nombre d'échantillons du signal reçu. Mais il était utile d'établir une relation plus précise entre ces deux grandeurs, d'autant plus qu'on peut trouver dans la littérature des formules fausses à ce propos.

Enfin nous appliquons ces résultats à un exemple concret ; c'est celui des transmissions de données sur un canal filtrant (par exemple une ligne téléphonique, un câble coaxial, un système de faisceaux hertziens). Le système de transmission utilise la modulation d'amplitude de deux porteuses en quadrature (MAQ). Par suite d'une erreur de phase $\theta(t)$ entre cette porteuse à la réception et l'oscillateur de réception chargé de régénérer la porteuse, le signal reçu comporte un bruit multiplicatif complexe en $e^{i\theta(t)}$, outre le bruit additif toujours présent.

Pour cet exemple, nous considérons les conditions que doit vérifier le canal, l'erreur de phase, le bruit additif, pour que les algorithmes de l'estimateur optimal des données converge, et nous discutons de leur validité pratique.

II. POSITION DU PROBLÈME

Nous considérons des signaux échantillonnés de nature complexe. Soit $\ldots, a_k, \ldots$ la suite des messages à estimer. Pour estimer a_k on dispose d'un vecteur $\vec{X}_k$, à P coordonnées, qui regroupe tous les signaux reçus contenant une information sur a_k ; $\vec{X}_k$ est appelé le vecteur observation. Les suites a_k et $\vec{X}_k$ sont supposées stationnaires dans leur ensemble.

L'estimateur linéaire h est alors caractérisé par un vecteur $\vec{H}$ à P coordonnées et l'estimée de a_k vaut :

$$y_k = \vec{X}_k^T \vec{H} . \tag{7}$$

L'optimisation de l'estimateur au sens de l'erreur quadratique moyenne consiste à rechercher le vecteur $\vec{H}$ tel que :

$$\varepsilon(\vec{H}) = E(|\, e_k(\vec{H})\,|^2) \quad \text{minimum} \tag{8}$$

où

$$e_k(\vec{H}) = a_k - y_k = a_k - \vec{X}_k^T \vec{H} . \tag{9}$$

Or d'après (8) et (9)

$$\varepsilon(\vec{H}) = \vec{H}^{T*} \underset{\sim}{R} \vec{H} - 2\mathscr{R}(\vec{H}^{T*} \vec{T}) + E(|\, a_k\,|^2) . \tag{10}$$

où la matrice $\underset{\sim}{R}$ et le vecteur $\vec{T}$ sont les moments du second ordre suivants

$$\underset{\sim}{R} = E(\vec{X}_k^* \vec{X}_k^T), \qquad \vec{T} = E(a_k \vec{X}_k^*) . \tag{11}$$

La matrice $\underset{\sim}{R}$ est définie non-négative et pour la suite nous la supposons inversible. Introduisons alors le vecteur $\vec{H}_0$ solution de l'équation de Wiener-Hopf

$$\underset{\sim}{R} \vec{H}_0 = \vec{T} \tag{12}$$

(qui n'est autre que l'équation (2)), et posons :

$$\vec{H} = \vec{H}_0 + \vec{V} . \tag{13}$$

Il vient

$$\varepsilon(\vec{H}) = \varepsilon(\vec{H}_0) + \vec{V}^{T*} \underset{\sim}{R} \vec{V} . \tag{14}$$

Cette équation montre que $\vec{H}_0$ est l'estimateur optimal cherché. Pour

---

Nota : $\mathscr{R}$ désigne la partie réelle.

le construire, les méthodes usuelles, même celle de Kalman-Bucy, commencent par évaluer les grandeurs $\underset{\sim}{R}$ et $\vec{T}$ à partir d'une suite d'observations et d'une suite de messages (dits messages tests). Ensuite plusieurs techniques sont possibles pour résoudre l'équation de Wiener-Hopf.

Les algorithmes stochastiques d'apprentissage utilisés dans la pratique réalisent ces deux opérations en un seul temps, et de manière itérative. C'est pourquoi ils sont préférés par les utilisateurs. En particulier si un changement se manifeste dans la statistique des signaux, ils ne nécessitent pas une procédure de réinitialisation pour estimer les nouvelles valeurs de $\underset{\sim}{R}$ et de $\vec{T}$. Ils dérivent de l'algorithme du gradient

$$\vec{H}_{k+1} = \vec{H}_k - \mu_k \nabla_H \varepsilon(\vec{H}_k) \tag{15}$$

destiné à fournir itérativement la solution de (8) (Remarquons que $\nabla_{\vec{H}}$ est un vecteur complexe qui regroupe le gradient par rapport aux coordonnées réelles et imaginaires de $\vec{H}$ en un seul vecteur). D'après (10), (11), (15) s'écrit :

$$\vec{H}_{k+1} = \vec{H}_k + \mu_k(-\mathrm{E}(\vec{X}_k^* \vec{X}_k^{\mathrm{T}})\vec{H}_k + \mathrm{E}(a_k \vec{X}_k^*)). \tag{16}$$

Comme les moyennes d'ensemble qui figurent dans (16) ne sont pas connues dans la pratique — et peuvent en outre être lentement variables — on supprime ces moyennes dans l'algorithme stochastique, avec l'espoir que les itérations successives réaliseront elles-mêmes la moyenne nécessaire. Il vient :

$$\vec{H}_{k+1} = \vec{H}_k + \mu_k \vec{Y}_k(\vec{H}_k), \tag{17 a}$$

où la fonction aléatoire $Y_k(\vec{H})$, appelé incrément, s'écrit

$$\vec{Y}_k(\vec{H}) = \vec{X}_k^*(a_k - \vec{X}_k^{\mathrm{T}}\vec{H}). \tag{17 b}$$

Ce n'est rien d'autre que l'algorithme (3), puisque $\vec{X}_k^{\mathrm{T}}\vec{H}_k$ est l'estimée y_k que donne du message a_k l'estimateur à sa $k^{\text{ième}}$ itération.

Quant au vecteur $\vec{X}_k$, c'est l'observation à l'instant k.

C'est donc l'un des algorithmes les plus simples que l'on puisse imaginer, commandé par l'erreur d'estimation $a_k - y_k$. On l'utilise avec des coefficients μ_k non négatifs, tendant vers zéro ou constants suivant le cas

$$\mu_k > 0 \qquad \sum_k \mu_k = \infty \qquad \sum_k \mu_k^2 < \infty ; \tag{18}$$

$$\mu_k = \mu > 0 . \tag{19}$$

Nous discutons maintenant la convergence de ces algorithmes.

III. CONVERGENCE DES ALGORITHMES

THÉORÈME 1 (algorithme à pas décroissant). — Si la matrice $\underset{\sim}{R}$ est inversible et si les couples successifs de variables aléatoires $(a_k, \vec{X}_k)$ sont indépendants, l'algorithme (17 *a-b*), (18) converge presque sûrement et en moyenne quadratique vers l'estimateur optimal $\vec{H}_0$, pourvu que l'incrément (17 *b*) soit du second ordre.

THÉORÈME 2 (algorithme à pas constant). — Dans les hypothèses du théorème 1 et si

$$\mu < \frac{2}{\lambda} \tag{20}$$

l'algorithme (17 *a-b*), (19) définit un vecteur $\vec{H}_k$ dont la moyenne d'ensemble converge vers l'estimateur optimal $\vec{H}_0$ et qui vérifie

$$\lim_{k \to \infty} \sup \, E(\| \vec{H}_k - \vec{H}_0 \|^2) < \alpha\mu , \tag{21}$$

où α est un nombre positif indépendant de μ et où λ est défini dans (23).

Indépendance. — Dans la pratique, le vecteur $\vec{X}_k$ est souvent constitué par P échantillons successifs, régulièrement espacés dans le temps, d'un signal continu $x(t)$. Si Δ est le pas, on a donc :

$$\vec{X}_k = (x(\tau_k), \ldots, x(\tau_k + (P - 1)\Delta)) . \tag{22}$$

Pour assurer la condition d'indépendance, on peut espacer ces vecteurs dans le temps par un choix convenable des τ_k. Naturellement cela ralentit la vitesse de convergence du facteur correspondant. C'est pourquoi, dans la pratique, l'espacement choisi n'est pas suffisant pour assurer l'indépendance. Pourtant on observe quand même la convergence des algorithmes. Il semble que récemment deux auteurs [6] [7] aient démontré la convergence en remplaçant l'indépendance par des conditions de décroissance sur des moments d'ordre élevé des a_k et des $\vec{X}_k$.

NOTATIONS. — Dans cet article, nous utiliserons les notations suivantes :

$$\lambda = \text{Sup} \{ \text{ valeurs propres de } \underset{\sim}{R} \} , \tag{23}$$

$$\nu = \text{Inf} \{ \text{ valeurs propres de } \underset{\sim}{R} \} , \tag{24}$$

$$\vec{V}_k = \vec{H}_k - \vec{H}_0 , \tag{25}$$

$$\sigma_k^2 = E \{ \| \vec{H}_k - \vec{H}_0 \|^2 \} = E \{ \vec{V}_k^{*T} \vec{V}_k \} . \tag{26}$$

Nous avons aussi besoin de définir la matrice

$$\underset{\sim}{T}(\mu) = E\{(\underset{\sim}{I} - \mu\vec{X}_k^* \vec{X}_k^T)^2\}, \tag{27}$$

ainsi que les valeurs propres :

$$t(\mu) = \text{Sup}\{\text{ valeurs propres de } \underset{\sim}{T}(\mu)\}, \tag{28}$$

$$u(\mu) = \text{Inf}\{\text{ valeurs propres de } \underset{\sim}{T}(\mu)\} \tag{29}$$

et la fonction $\eta_t(\mu)$ définie par

$$2\nu(1 + \eta_t(\mu)) = \frac{1 - t(\mu)}{\mu}, \tag{30}$$

dont il est facile de voir que

$$\eta_t(\mu) \to 0 \quad \text{si} \quad \mu \to 0. \tag{31}$$

Enfin, principalement pour les applications décrites au paragraphe VII, nous définirons les moments :

$$\underset{\sim}{A} = E\{(\vec{X}_k^* \vec{X}_k^T)^2\}, \tag{32}$$

$$\vec{B} = E\{a_k \vec{X}_k^* \vec{X}_k^T \vec{X}_k^*\}, \tag{33}$$

$$\rho_0 = E\{|a_k|^2 \, \|\vec{X}_k\|^2\}, \tag{34}$$

$$\gamma = E\{\|\vec{X}_k\|^2 . \|\vec{X}_k^T \vec{H}_0 - a_k\|^2\}. \tag{35}$$

La démonstration de la convergence presque sûre dans le théorème 1 est une généralisation aisée au cas des vecteurs $\vec{X}_k$ et $\vec{H}_k$ complexes, d'un théorème établi par C. Macchi dans le cas réel (cf. [4], théorème 1.2), lorsque la valeur moyenne de l'incrément $\vec{Y}_k(\vec{H})$ de l'algorithme (17) a la forme :

$$E(\vec{Y}(\vec{H})) = -\underset{\sim}{R'}\vec{H} + \vec{T}', \tag{36}$$

où $\underset{\sim}{R'}$ est une matrice définie-positive.

Pour établir la convergence en moyenne quadratique dans le théorème 1, on peut utiliser un résultat établi dans [3] pour les fonctions vectorielles réelles et le généraliser de la même manière au cas complexe. Notons que l'énoncé de [3] (hypothèse A-1) est assez ambigu et pourrait laisser croire qu'il est affranchi de l'hypothèse d'indépendance. En fait, mis à part le cas où les restes $\vec{Y}_k(\vec{H})$ sont des fonctions aléatoires indépendantes, on n'imagine pas de cas simple où l'hypothèse (A-1) soit vérifiée.

Pour la démonstration du théorème 2, on peut se référer à un article antérieur de C. Macchi, J. P. Jouannaud, O. Macchi (cf. [5] Annexe E), où il est montré le résultat suivant, sous les hypothèses du théorème 1 :

$$\limsup_{k\to\infty} \sigma_k^2 \leqslant \mu \frac{\gamma}{2\nu(1+\eta_t(\mu))}. \tag{37}$$

D'après ce théorème, la convergence de l'algorithme à pas constant est une pseudo-convergence en moyenne quadratique. D'après (37), la finesse de convergence est d'autant meilleure que le pas μ est plus petit. On choisira celui-ci en fonction de la qualité de convergence désirée (par exemple liée à la précision des mesures).

Au contraire l'algorithme à pas décroissant converge strictement. Il est donc supérieur en ce sens à l'algorithme à pas constant. Nous verrons plus loin que cette supériorité est au prix d'une moindre qualité d'adaptativité.

IV. CHOIX DU PAS DE L'ALGORITHME. CAS STATIONNAIRE

Bien que le théorème 2 garantisse le mode de convergence de l'algorithme à pas constant, le résultat (21) est insuffisant dans la pratique. En effet il est nécessaire de déterminer la valeur de μ qui assure la finesse de convergence souhaitable. C'est ce que nous faisons dans ce paragraphe.

D'après l'équation (14), et du fait de l'indépendance entre le couple $(a_k, \vec{X}_k)$ et le vecteur $\vec{V}_k$, on a

$$\mathrm{E}(\|\vec{H}_k^T \vec{X}_k - a_k\|^2) = \varepsilon(\vec{H}_0) + \mathrm{E}(\vec{V}_k^T * \underset{\sim}{R}\, \vec{V}_k). \tag{38}$$

Or $\varepsilon(\vec{H}_0)$ est l'erreur quadratique moyenne minimale correspondant à l'estimateur optimal $\vec{H}_0$. Le deuxième terme est l'erreur quadratique additionnelle qui provient de la non convergence de $\vec{H}_k$ (fluctuations résiduelles dues au pas μ qui est non nul). Dans la pratique il est souhaitable que cette erreur additionnelle soit du même ordre que l'erreur minimale $\varepsilon(\vec{H}_0)$.

D'après les définitions (23), (24)

$$\nu\sigma_k^2 \leqslant \mathrm{E}(\vec{V}_k^T * \underset{\sim}{R}\, \vec{V}_k) \leqslant \lambda\sigma_k^2. \tag{39}$$

En conjuguant (37) avec l'inégalité de droite il vient

$$\limsup_{k\to\infty} \mathrm{E}(\vec{V}_k^T * \underset{\sim}{R}\, \vec{V}_k) \leqslant \frac{\lambda\mu\gamma}{2\nu(1+\eta_t(\mu))}. \tag{40}$$

On peut faire un calcul analogue avec la limite inférieure de σ_k^2, faisant intervenir la plus petite valeur propre $u(\mu)$ de la matrice $\underset{\sim}{T}(\mu)$ pour laquelle

$$u(\mu) = 1 - 2\lambda\mu(1 + \eta_u(\mu)), \qquad \text{avec} \quad \eta_u(\mu) \to 0 \quad \text{si} \quad \mu \to 0. \tag{41}$$

On trouve une borne inférieure :

$$\frac{\nu\mu\gamma}{2\lambda(1 + \eta_u(\mu))} \leqslant \liminf_{k\to\infty} E(\vec{V}_k^T * \underset{\sim}{R}\, \vec{V}_k). \tag{42}$$

L'erreur additionnelle $E(\vec{V}_k^T * \underset{\sim}{R}\, \vec{V}_k)$ va fluctuer autour de la valeur $\varepsilon(\vec{H}_0)$ par exemple si l'intervalle délimité par les bornes (40) et (42) est centré autour de $\varepsilon(\vec{H}_0)$. Ceci s'écrit :

$$\mu\left[\frac{\lambda}{\nu}\cdot\frac{1}{1+\eta_t(\mu)} + \frac{\nu}{\lambda}\cdot\frac{1}{1+\eta_u(\mu)}\right] = \frac{4E\{\|\vec{H}_0^T\vec{X}_k - a_k\|^2\}}{E\{\|\vec{X}_k\|^2\,\|\vec{H}_0^T\vec{X}_k - a_k\|^2\}} \tag{43}$$

En général, cette équation se simplifie parce que les fonctions $\eta_u(\mu)$ et $\eta_t(\mu)$ tendent vers zéro avec μ. Ainsi lorsque la matrice $\mu\underset{\sim}{A}$ est négligeable devant $2\underset{\sim}{R}$, c'est-à-dire lorsque

$$\mu \ll 2\nu(\text{sup. valeurs propres de } \underset{\sim}{A})^{-1} \tag{44}$$

(43) se simplifie en

$$\mu = 4\frac{\lambda/\nu}{1 + \lambda^2/\nu^2}\cdot\frac{E\{\|\vec{H}_0^T\vec{X}_k - a_k\|^2\}}{E\{\|\vec{X}_k\|^2 . \|\vec{H}_0\vec{X}_k - a_k\|^2\}}. \tag{45}$$

Remarquons ici que la condition (44) implique la condition (20) du théorème 2.

En effet il est facile de voir sur les définitions des matrices $\underset{\sim}{R}$ et $\underset{\sim}{A}$ que

$$\underset{\sim}{A} = (\underset{\sim}{R})^2 + \underset{\sim}{B} \tag{46}$$

où $\underset{\sim}{B}$ est une matrice définie non négative, de sorte que

$$(\text{sup. valeurs propres de } \underset{\sim}{A}) \geqslant \lambda^2$$

et que

$$2\nu(\text{sup. valeurs propres de } \underset{\sim}{A})^{-1} \leqslant \frac{2}{\lambda}.$$

Remarquons encore que le résultat (45) est en contradiction avec les résultats d'Ungerboeck [8] obtenus dans le cas particulier du canal filtrant

réel (cf. paragraphe VII). Celui-ci trouve comme pas souhaitable pour l'algorithme

$$\mu = \frac{1}{\mathrm{E}(\| \vec{\mathrm{X}}_k \|^2)}. \tag{47}$$

Cette formule n'est très proche de notre résultat que s'il est vrai que

$$\mathrm{E}\{ \| \vec{\mathrm{X}}_k \|^2 \| \vec{\mathrm{H}}_0^{\mathrm{T}} \vec{\mathrm{X}}_k - a_k \|^2 \} = \mathrm{E}\{ \| \vec{\mathrm{X}}_k \|^2 \} \mathrm{E}\{ \| \vec{\mathrm{H}}_0^{\mathrm{T}} \vec{\mathrm{X}}_k - a_k \|^2 \}. \tag{48}$$

Or l'équation (48) sous-entend une indépendance entre $\| \vec{\mathrm{X}}_k \|$ et $\| \vec{\mathrm{H}}_0 \vec{\mathrm{X}}_k - a_k \|$, qui n'a pas nécessairement lieu, malgré la non corrélation effective entre ces deux grandeurs

$$(\mathrm{E}\{ \vec{\mathrm{X}}_k^* (\vec{\mathrm{X}}_k^{\mathrm{T}} \vec{\mathrm{H}}_0 - a_k) \} = \underset{\sim}{\mathrm{R}} \vec{\mathrm{H}}_0 - \vec{\mathrm{T}} = 0).$$

V. ADAPTATIVITÉ DES ALGORITHMES

Pour expliciter la différence entre l'algorithme à pas constant et l'algorithme à pas décroissant, nous allons étudier le temps d'oubli τ de chacun d'eux. Celui-ci sera défini comme le nombre d'itérations $(k - p)$ au bout desquelles la valeur de l'estimateur $\vec{\mathrm{H}}_{k+1}$ ne dépend plus du couple $(a_p, \vec{\mathrm{X}}_p)$, de manière significative (la dépendance étant prise au sens fonctionnel et non statistique).

En itérant les équations (17) de l'algorithme, on obtient

$$\vec{\mathrm{H}}_{k+1} = \sum_{l=1}^{k} \underset{\sim}{\mathrm{D}}_l^k a_l \vec{\mathrm{X}}_l^* + \prod_{l=1}^{k} (\underset{\sim}{\mathrm{I}} - \mu_l \vec{\mathrm{X}}_l^* \vec{\mathrm{X}}_l^{\mathrm{T}}) \vec{\mathrm{H}}_l \tag{49}$$

où $\underset{\sim}{\mathrm{D}}_l^k$ est la matrice aléatoire

$$\underset{\sim}{\mathrm{D}}_l^k = \mu_l \prod_{j=l+1}^{k} (\underset{\sim}{\mathrm{I}} - \mu_j \vec{\mathrm{X}}_j^* \vec{\mathrm{X}}_j^{\mathrm{T}}), \qquad l = 1, \ldots, k. \tag{50}$$

Ainsi $\vec{\mathrm{H}}_{k+1}$ est une fonction linéaire (aléatoire) des vecteurs successifs $a_l \vec{\mathrm{X}}_l^*$, $l = 1, \ldots, k$ et les coefficients de cette fonction linéaire sont les matrices (50). Si nous supposons, comme dans le théorème, que les couples successifs $(a_l, \vec{\mathrm{X}}_l)$ sont indépendants, la matrice $\underset{\sim}{\mathrm{D}}_l^k$ et le vecteur $a_l \vec{\mathrm{X}}_l^*$ sont indépendants. Le temps d'oubli est donc l'intervalle de temps $k - p$ nécessaire pour que $\underset{\sim}{\mathrm{D}}_l^k$ devienne négligeable, pour tout $l \leqslant p$. Cela implique

que $\underset{\sim}{D}_l^k$ soit de moyenne négligeable. Or d'après (50) et l'indépendance

$$\mathrm{E}(\underset{\sim}{D}_l^k) = \mu_l \prod_{j=l+1}^{k} (\underset{\sim}{I} - \mu_j \underset{\sim}{R}), \tag{51}$$

matrice dont il faut considérer le comportement asymptotique pour $k \to \infty$, et l borné par p.

A. Algorithme à pas décroissant

Supposons, pour que le calcul asymptotique de (51) soit possible, que μ_k est de la forme

$$\mu_k = \frac{\rho}{k+1}, \qquad \rho > 0 \tag{52}$$

qui satisfait à la condition (18). Si en outre

$$\rho \leqslant 1/\lambda, \tag{53}$$

toutes les matrices $(\underset{\sim}{I} - \mu_j \underset{\sim}{R})$ sont définies non négatives. Il est facile de voir qu'alors toutes les valeurs propres d_l^k de la matrice (51) sont telles que

$$\frac{\rho}{k+1} \leqslant d_l^k \leqslant \frac{\rho}{l+1} \prod_{j=l+1}^{k} \left(1 - \frac{\rho v}{j+1}\right). \tag{54}$$

Le terme de droite est lié au développement en produit infini de la fonction Γ :

$$\Gamma^{-1}(-\rho v) = \lim_{n \to \infty} (-\rho v) n^{\rho v} (1 - \rho v)\left(1 - \frac{\rho v}{2}\right) \cdots \left(1 - \frac{\rho v}{n}\right)$$

valable pourvu que ρv ne soit pas un entier positif ; ceci est assuré lorsque $\rho v < 1$, en particulier lorsque l'inégalité (53) est stricte (on rappelle que $\Gamma(x) < 0$ pour $-1 < x < 0$, de sorte que $-\Gamma(-\rho v)$ est positif).

Si l est majoré par p, l'écriture asymptotique de (54) pour k tendant vers l'infini est

$$\frac{\rho}{k+1} \leqslant d_l^k \leqslant \frac{1}{(-v.\Gamma(-\rho v)) \prod_{j=1}^{p} \left(1 - \frac{\rho v}{j+1}\right)} \cdot \frac{1}{(k+1)^{\rho v}}. \tag{55}$$

D'où des inégalités entre matrices définies positives

$$\frac{\rho}{k+1}\underset{\sim}{\mathrm{I}} \leqslant \mathrm{E}(\underset{\sim}{\mathrm{D}}_l^k) \leqslant \alpha_p \frac{1}{(k+1)^{\rho v}}\underset{\sim}{\mathrm{I}}, \qquad l \leqslant p. \tag{56}$$

L'inégalité de gauche montre que le vecteur $\vec{\mathrm{H}}_{k+1}$ est fonction linéaire des vecteurs successifs $a_l \vec{\mathrm{X}}_l^*$ à coefficients successifs d'égale importance. Autrement dit les observations les plus anciennes interviennent autant que les observations les plus récentes dans le calcul de l'estimateur. L'inégalité de droite montre que tous ces coefficients tendent (en moyenne) conjointement mais lentement vers zéro lorsque $(k - l)$ tend vers l'infini, en $(k+1)^{-\rho v}$. Pour définir le temps d'oubli, on se donnera un niveau ε petit devant un, et pour p donné, on cherchera l'indice k_0 tel que

$$\mathrm{E}(\underset{\sim}{\mathrm{D}}_l^k) \leqslant \varepsilon \underset{\sim}{\mathrm{I}}, \qquad \forall k \geqslant k_0, \quad \forall l \leqslant p. \tag{57}$$

D'après (56) et (57) l'indice k_0 est (asymptotiquement) solution de l'équation

$$\frac{\alpha_p}{(k_0+1)^{\rho v}} = \varepsilon, \tag{58}$$

et on en déduit le temps d'oubli $k_0 - p$. Si p lui-même est assez grand, l'équation (58) prend, d'après (55), la forme

$$k_0 + 1 = (p+1).\left(\frac{\rho}{\varepsilon}\right)^{1/\rho v} \tag{59}$$

qui montre que le temps d'oubli tend à devenir proportionnel à $(p+1)$:

$$\frac{\tau_p}{p+1} \to \left(\frac{\rho}{\varepsilon}\right)^{1/\rho v} - 1, \qquad \text{si} \quad p \to \infty. \tag{60}$$

Ainsi si les propriétés statistiques du couple $(a_l, \vec{\mathrm{X}}_l)$ changent à l'itération p, l'estimateur oublie la statistique passée au bout d'un temps τ_p qui tend vers l'infini. En conséquence, une variation tardive de statistique ne pourra pratiquement pas être prise en considération par l'estimateur à pas décroissant. C'est pourquoi l'algorithme à pas décroissant n'est pas utilisé dans les applications où l'on peut prévoir des variations de statistique des signaux traités.

La vitesse de convergence de l'algorithme à pas décroissant en $\frac{\rho}{k+1}$ a

été étudiée par plusieurs auteurs (cf. par exemple [3]) qui ont montré que

$$E(\| \vec{H}_k - \vec{H}_0 \|^2) = 0\left(\frac{1}{k+1}\right)$$

Ainsi l'écart quadratique moyen est un infiniment petit du même ordre que μ_k.

B. Algorithme à pas constant

Dans ce cas

$$E(\underset{\sim}{D}_l^k) = \mu(\underset{\sim}{I} - \mu\underset{\sim}{R})^{k-l}, \tag{61}$$

matrice définie positive et qui tend vers zéro lorsque $k - l$ tend vers l'infini sous la condition (20). L'algorithme effectue une combinaison linéaire à coefficients décroissant exponentiellement des vecteurs passés $a_l \bar{X}_l^*$, les observations les plus anciennes intervenant moins que les plus récentes, comme il est naturel, dans le calcul de l'estimateur. Le temps d'oubli τ est indépendant de l'instant p auquel survient un changement de statistique et il est donné par

$$\tau = \frac{\text{Log } \varepsilon/\mu}{\text{Log } (1 - \mu\nu)}. \tag{62}$$

Si $\mu\nu \ll 1$ (de telle sorte que l'erreur quadratique moyenne additionnelle, due à la fluctuation résiduelle de $\vec{H}_k$, soit faible — cf. (38) et (40) —) on a asymptotiquement

$$\mathcal{C} = \frac{\text{Log } \mu/\varepsilon}{\mu/\varepsilon} \cdot \frac{\varepsilon}{\nu}. \tag{63}$$

Le temps d'oubli est d'autant plus grand que μ est plus faible, c'est-à-dire, d'après (40), que la finesse de convergence est bonne. Comme on a vu que les qualités d'adaptativité d'un algorithme sont directement liées à la petitesse du temps d'oubli, il apparaît qu'adaptativité et finesse de convergence sont deux qualités contradictoires pour un même algorithme.

Remarquons enfin la parfaite similitude entre les temps d'oubli (60) et (63), si l'on tient compte dans (60) de la relation (52). Nous résumons ces résultats dans un tableau comparatif

Algorithme à pas décroissant	Algorithme à pas constant
$E\{\| \vec{H}_k - \vec{H}_0 \|^2\} \to 0$ en μ_k Temps d'oubli $\to \infty$ en $\frac{1}{\mu_k}$	$E\{\| \vec{H}_k - \vec{H}_0 \|^2\} \simeq \alpha\mu$ Temps d'oubli $\simeq \beta . \frac{\text{Log } \mu}{\mu}$

VI. CAS DE STATIONNARITÉ LOCALE

On dit que le couple $(a_k, \vec{X}_k)$ présente une stationnarité locale si les propriétés statistiques sont invariantes au cours de tranches successives de K itérations. Au contraire, d'une tranche à l'autre, les propriétés statistiques de $(a_k, \vec{X}_k)$ peuvent changer; K s'appelle le temps de stationnarité.

Naturellement cette définition n'est qu'une schématisation de la réalité, dans laquelle en général les propriétés statistiques varient de manière continue au cours du temps.

Nous avons vu qu'à cause des changements de statistique, l'algorithme à pas décroissant n'est pas utilisable, pour construire l'estimateur optimal $\vec{H}_0$, qui naturellement varie au cours du temps à intervalles de durée T.

L'algorithme à pas constant est utilisable à condition de trouver un pas μ réalisant une bonne finesse de convergence $\alpha\mu$ et conjointement un temps d'oubli $\beta . \dfrac{\text{Log } \mu}{\mu}$ plus petit que K. Ces exigences amènent souvent à une impossibilité. On peut trouver dans [9], [11] une manière de résoudre ce problème en utilisant en même temps deux algorithmes, ne portant pas sur le même nombre d'observations pour le vecteur $\vec{X}_k$ et n'ayant pas le même pas d'incrémentation (μ).

VII. EXEMPLE : LE CANAL FILTRANT AFFECTÉ DE GIGUE DE PHASE

Nous donnons maintenant un exemple d'utilisation d'estimateur adaptatif en transmission de données. Comme nous l'avons déjà mentionné, les auteurs se contentent le plus souvent dans ce domaine, de vérifier la convergence sur des simulations. Il est donc utile de vérifier que les théorèmes 1 et 2 s'appliquent effectivement dans ce cas. On peut trouver une présentation des systèmes de transmission de données dans différents ouvrages (voir par exemple [5] et [12]). Les données sont les messages à estimer, et le canal de transmission se comporte comme un filtre linéaire pour le signal transmis. Lorsque la modulation utilisée est linéaire (et c'est le cas de la modulation de deux porteuses en quadrature), on peut montrer que le signal après transmission et démodulation vaut

$$S(t) = \sum_k m(k\Delta)s(t - k\Delta) \tag{64}$$

où Δ est l'intervalle de rythme et $m(k\Delta)$ est la donnée à l'instant $k\Delta$, où $s(t)$ et $m(k\Delta)$ sont des grandeurs complexes ; $s(t)$ est appelée réponse du canal.

Ce signal est en outre affecté d'un bruit multiplicatif $e^{i\theta(t)}$, où $\theta(t)$, réel, est un déphasage lié au canal et à la démodulation, que l'on appelle souvent gigue de phase. Naturellement, ce déphasage est aléatoire. Enfin un bruit additif complexe $b(t)$ affecte le signal. Ainsi le signal reçu s'écrit

$$x(t) = \left(\sum_k m(k\Delta) s(t - k\Delta) \right) e^{i\theta(t)} + b(t) . \tag{65}$$

Regroupons ici les hypothèses sur la statistique des divers éléments aléatoires.

Les données

$$\begin{cases} \mathrm{E}(m(k\Delta)) = \mathrm{E}(m^2(k\Delta)) = 0 \ ; \quad \mathrm{E}(\,|\, m(k\Delta)\,|^2) = a^2 < \infty \ ; \\ \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad \mathrm{E}(\,|\, m(k\Delta)\,|^4) - a^4 = b^2 < \infty , \\ \text{les } m(k\Delta) \text{ sont indépendants et de même loi.} \end{cases} \tag{66}$$

Notons que dans le cas de données, les $m(k\Delta)$ prennent seulement un nombre fini de valeurs, et les conditions de finitude des moments sont automatiquement remplies. En outre les données émises sont toujours centrées (1re condition), et en général non corrélées entre les deux voies, et de même statistique (2e condition).

Le bruit

$$\begin{cases} \mathrm{E}(b(t)) = 0 \ ; \quad \mathrm{E}(\,|\, b^2(t)\,|\,) = \sigma_b^2 < \infty \ ; \quad \mathrm{E}(\,|\, b^4(t)\,|\,) = m_4 < \infty \\ b(t) \text{ est stationnaire.} \end{cases} \tag{67}$$

La gigue de phase

$$e^{i\theta(t)} \text{ est stationnaire} \tag{68}$$

(69) Les éléments $\{ \ldots, m(k\Delta), \ldots \}$, $b(t)$, $e^{i\theta(t)}$ sont statistiquement indépendants.

Le canal filtrant

$$\sum_k |\, s^2(k\Delta)\,| = \mathrm{E} < \infty . \tag{70}$$

Cette condition est vérifiée si la réponse $s(t)$ du canal est d'énergie finie, au moins lorsque $|\, s(t)\,|$ est monotone en dehors d'un compact, ou encore si la transformée de Fourier de $s(t)$ est continue et à support borné. Dans la pratique cette condition est toujours satisfaite. Le signal reçu (65) est

échantillonné à intervalles réguliers Δ ; le vecteur observation $\vec{X}_k$ est formé suivant (22), avec un espacement convenable pour assurer l'indépendance. On suppose en outre que τ_k est un multiple entier de Δ. Le message a_k à estimer sur la base de l'observation $\vec{X}_k$ est alors la valeur à l'instant τ_k de $m(t)$:

$$a_k = m(\tau_k). \tag{71}$$

Puis l'estimateur $\vec{H}_k$ du message a_k est construit suivant l'algorithme (17), (18) [ou (19)] et nous vérifions l'une après l'autre les hypothèses des théorèmes 1 et 2. Par commodité, introduisons les notations suivantes :

$$\vec{S}_j = (s(j\Delta), \ldots, s([j + (P - 1)]\Delta), \tag{72}$$

$$\underset{\sim}{\mathscr{S}} = \sum_k \vec{S}_k^* \vec{S}_k^T, \tag{73}$$

$$\begin{cases} \vec{B}_k^T = (b(\tau_k), \ldots, b(\tau_k + (P - 1)\Delta)) \\ \vec{R}_B = E(\vec{B}_k^* \vec{B}_k^T) \end{cases} \tag{74}$$

(les éléments de ces matrices sont bornés en module, respectivement par E et par σ_b^2).

$$\underset{\sim}{D}_k = \underset{\sim}{D}_k^T = \text{Diag}\,(e^{i\theta(\tau_k)}, \ldots, e^{i\theta(\tau_k + (P-1)\Delta)}), \tag{75}$$

$$\underset{\sim}{\mathscr{S}}' = E\{\underset{\sim}{D}_k^* \underset{\sim}{\mathscr{S}} \underset{\sim}{D}_k^T\}, \tag{76}$$

$$G_2(l - j) = E\{\exp i(\theta(j\Delta) - \theta(l\Delta))\}. \tag{77}$$

Si s_{lj} et s'_{lj} sont les éléments génériques respectifs de $\underset{\sim}{\mathscr{S}}$ et $\underset{\sim}{\mathscr{S}}'$, on a

$$s'_{lj} = s_{lj} G_2(l - j) \tag{78}$$

et en particulier

$$\text{trace}\ \underset{\sim}{\mathscr{S}}' = \text{trace}\ \underset{\sim}{\mathscr{S}}. \tag{79}$$

Inversibilité de $\underset{\sim}{R}$

Après échantillonnage, (65) devient

$$\vec{X}_k = \underset{\sim}{D}_k \sum_l m(\tau_k - l\Delta)\vec{S}_l + \vec{B}_k. \tag{80}$$

Le calcul de la matrice de covariance utilise les hypothèses (66) à (70). Il vient

$$\underset{\sim}{R} = a^2 \underset{\sim}{\mathscr{S}}' + \underset{\sim}{R}_B. \tag{81}$$

Il est facile de voir sur (73) que $\underset{\sim}{\mathscr{S}}$ est définie non négative. Or d'après (76), pour tout vecteur $\vec{H}$

$$\vec{H}^{*T}\underset{\sim}{\mathscr{S}}'\vec{H} = E\{(\underset{\sim}{D}_k^T\vec{H})^{*T}\underset{\sim}{\mathscr{S}}(\underset{\sim}{D}_k^T\vec{H})\} \tag{82}$$

Ainsi $\underset{\sim}{\mathscr{S}}'$ est elle-même définie non négative. En outre, puisque $(\underset{\sim}{D}_k^T\vec{H})^{*T}\underset{\sim}{\mathscr{S}}(\underset{\sim}{D}_k^T\vec{H})$ est non négative, sa moyenne (82) n'est nulle que si cette variable est presque sûrement nulle. Puisque la matrice $\underset{\sim}{D}_k^T$ est toujours inversible, il en résulte que $\underset{\sim}{\mathscr{S}}$ et $\underset{\sim}{\mathscr{S}}'$ sont inversibles en même temps. Un critère d'inversibilité pour $\underset{\sim}{\mathscr{S}}$ est donné dans [4] p. 258. Il concerne le « spectre replié du canal », soit $S^{\Delta}(\nu)$, défini comme la série de Fourier des échantillons $s(k\Delta)$. $\underset{\sim}{\mathscr{S}}$ est inversible dès que $|S^{\Delta}(\nu)|$ est continu par morceaux et non nul sur l'intervalle $\left[-\frac{1}{2\Delta}, \frac{1}{2\Delta}\right]$. La dénomination de « spectre replié » est liée à la formule de Poisson

$$S^{\Delta}(\nu) = \frac{1}{\Delta}\sum_k S\left(\nu + \frac{k}{\Delta}\right), \qquad \nu \in \left[-\frac{1}{2\Delta}, \frac{1}{2\Delta}\right] \tag{83}$$

où $S(\nu)$ (gain du canal) est la transformée de Fourier de $s(t)$. Dans la pratique cette propriété est toujours vérifiée, tout au moins lorsque l'instant d'échantillonnage initial est convenablement choisi.

La matrice $\underset{\sim}{R}_B$ est elle aussi définie non négative ; son inversibilité est liée au spectre du bruit replié dans la bande d'échantillonnage.

$$\gamma^{\Delta}(\nu) = \frac{1}{\Delta}\sum_k \gamma\left(\nu + \frac{k}{\Delta}\right), \tag{84}$$

où $\gamma(\nu)$ est la densité spectrale de puissance de $b(t)$; $\underset{\sim}{R}_B$ est inversible dès que $\gamma^{\Delta}(\nu)$ est continu par morceaux et non nul dans l'intervalle $\left[-\frac{1}{2\Delta}, \frac{1}{2\Delta}\right]$ ([4] p. 58), condition qui est pratiquement toujours satisfaite dans les applications.

D'après (81), $\underset{\sim}{R}$ est inversible dès que l'une ou l'autre des matrices $\underset{\sim}{\mathscr{S}}$ et $\underset{\sim}{R}_B$ le sont. Cependant, dans les applications l'inversibilité de $\underset{\sim}{\mathscr{S}}$ est prépondérante à cause de la très faible valeur de $\frac{\sigma_b^2}{a^2}$. Pour la suite de ce paragraphe, supposons donc que

$$\underset{\sim}{R} = a^2\underset{\sim}{\mathscr{S}}'. \tag{85}$$

On peut préciser un peu l'étude de l'inversibilité de $\underset{\sim}{R}$ (ou de $\mathscr{S}'$) en chiffrant approximativement la dispersion $\frac{\lambda}{v}$ entre ses valeurs propres. On sait en effet que l'algorithme (16), qui est la version déterministe des algorithmes étudiés ici, converge d'autant plus vite que $\frac{\lambda}{v}$ est plus voisin de 1, la valeur limite $\frac{\lambda}{v} = \infty$ correspondant à une matrice $\mathscr{S}'$ non inversible.

En outre, on a vu au paragraphe IV, que pour l'algorithme à pas constant et pour une erreur quadratique additionnelle égale à l'erreur quadratique minimale, le pas d'incrémentation a la forme (45). C'est donc une fonction décroissante de la dispersion. Naturellement, le temps nécessaire à la convergence de l'algorithme est une fonction décroissante de μ et par conséquent une fonction croissante de la dispersion. C'est pourquoi il est utile de trouver un majorant de cette dispersion.

Disons d'abord que l'on peut se ramener à la dispersion de la matrice $\mathscr{S}$. En effet, si $\lambda' = \lambda/a^2$(resp $v' = v/a^2$) est la plus grande (resp. petite) valeur propre de $\mathscr{S}'$, et si λ_1 et v_1 sont les mêmes quantités relatives à $\mathscr{S}$, l'égalité (82), et le caractère orthonormé de $\underset{\sim}{D}_k^T$ entraînent les inégalités

$$v_1 \leqslant v' \leqslant \lambda' \leqslant \lambda_1 , \tag{86}$$

d'où

$$\frac{\lambda}{v} = \frac{\lambda'}{v'} \leqslant \frac{\lambda_1}{v_1} . \tag{87}$$

Il est montré dans [4] que $\mathscr{S}$ est la matrice de corrélation d'un processus aléatoire centré dont le spectre replié (cf. (84)) est $|S^{\Delta}(v)|^2$.

Pour de telles matrices on connaît [13] des bornes des valeurs propres. Supposant que la fonction (83) est continue par morceaux, on a

$$\inf_v |S^{\Delta}(v)|^2 \leqslant v_1 \leqslant \lambda_1 \leqslant \sup_v |S^{\Delta}(v)|^2 . \tag{88}$$

On en déduit la majoration

$$\frac{\lambda}{v} \leqslant r = \frac{\max_v |S^{\Delta}(v)|^2}{\min_v |S^{\Delta}(v)|^2} . \tag{89}$$

Si la fonction $S(v)$ ne dépend de l'instant initial que par sa phase on voit sur (83) que le spectre replié du canal peut au contraire en dépendre beaucoup. Ceci est illustré sur la figure 1. Dans la pratique, on doit donc choisir très soigneusement cet instant initial, de manière à minimiser le rapport r.

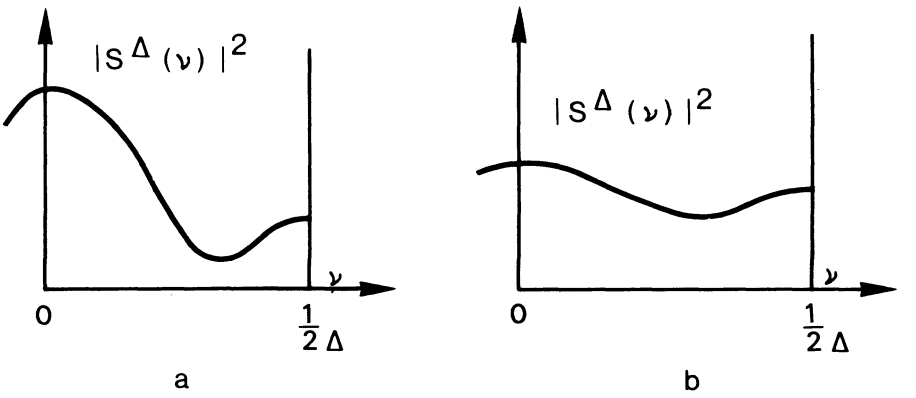

FIG. 1. — Spectre replié d'un même canal ($s(t)$). pour un instant initial d'échantillonnage *a*) mal choisi *b*) bien choisi.

En transmission de données sur lignes téléphoniques, on pourra par exemple admettre que r est inférieur à 3. On aura alors

$$\frac{3}{10}v^2 \leqslant \frac{\lambda v}{\lambda^2 + v^2} \leqslant \frac{v^2}{2}. \tag{90}$$

L'introduction d'un résultat de ce type dans la formule (45) permet d'optimiser pratiquement le choix du pas d'incrémentation μ.

Étude de l'estimateur optimal

L'estimateur cherché est solution de l'équation (12). On calcule le vecteur $\vec{T}$ à l'aide de (65) et des conditions (66) à (70). Il vient

$$\vec{T} = a^2 \mathrm{E}(e^{-i\theta(t)})\vec{S}_0^* . \tag{91}$$

On doit s'assurer que $\vec{T}$ est non nul pour que l'estimateur cherché soit utile. Or l'instant d'échantillonnage est toujours choisi tel que $|s(0)|$ soit grand. Donc $\vec{S}_0$ est non nul. Quant au bruit multiplicatif, sa moyenne n'est pas nulle, au moins si $\theta(t)$ est gaussien stationnaire, ou formé de composantes spectrales de phases aléatoires équi-réparties et indépendantes. Ces deux modèles suffisent pour décrire la gigue de phase sur les canaux réels.

On vérifie en annexe que l'incrément $\vec{X}_k^*(a_k - \vec{X}_k^{\mathrm{T}}\vec{H})$ est du second ordre, pour le signal (65) et sous les conditions (66)-(70).

Nous avons donc prouvé que l'algorithme (17) converge vers l'estimateur optimal dans le cas du canal filtrant affecté de gigue.

ANNEXE

Nous montrons que les moments (32) à (34) sont finis, sous les hypothèses (65) à (70), de sorte que la fonction $\vec{Y}_k(\vec{H}) = \vec{X}_k^*(a_k - \vec{X}_k^T\vec{H})$ est du second ordre.

1) Étude de ρ_0

D'après (32)-(34), (66)-(70), (80)

$$\rho_0 = E\left(|a_k|^2 \sum_{i,j} m(\tau_k - i\Delta)m^*(\tau_k - j\Delta)\vec{S}_i^T\vec{S}_j^* \right) + a^2 E(||\vec{B}_k||^2). \tag{a-1}$$

Dans le premier terme, il n'y a pas de contribution pour $i \neq j$, de sorte que, d'après (71)

$$\rho_0 = a^4 \sum_i ||\vec{S}_i||^2 + b^2 ||\vec{S}_0||^2 + a^2 P\sigma_b^2. \tag{a-2}$$

Or

$$\sum_i ||\vec{S}_i||^2 = PE. \tag{a-3}$$

D'où

$$\rho_0 \leqslant a^2 P(a^2 E + \sigma_b^2) + b^2 E < \infty. \tag{a-4}$$

2) Étude de $\vec{B}$

En introduisant (80) dans (33), développant et prenant l'espérance, on trouve 8 termes dont 5 sont directement nuls du fait des hypothèses (66)-(70). Les termes restant sont tous du premier ordre en $\underset{\sim}{D}_k$. Or

$$E(\underset{\sim}{D}_k) = G.\underset{\sim}{I} \triangleq E\{ e^{i\theta(t)} \}.\underset{\sim}{I}. \tag{a-5}$$

Ainsi, au facteur G ou G* près, la gigue disparaît dans la moyenne. Restent les termes

$$\vec{B}_1 = G^*E\left\{ a_k \sum_{ijl} m^*(\tau_k - i\Delta)m(\tau_k - j\Delta)m^*(\tau_k - l\Delta)\vec{S}_i^*\vec{S}_j^T\vec{S}_l^* \right\},$$

$$\vec{B}_2 = G^*E\left\{ a_k \sum_i m^*(\tau_k - i\Delta)\vec{S}_i^*\vec{B}_k^T\vec{B}_k^* \right\},$$

$$\vec{B}_3 = G^*E\left\{ \vec{B}_k^*\vec{B}_k^T a_k \sum_i m^*(\tau_k - i\Delta)\vec{S}_i^* \right\}.$$

Dans le calcul de B_1, on peut distinguer 3 contributions non nulles suivant que $i = j = l = 0$, $i = 0$ et $j = l \neq 0$, ou $l = 0$ et $i = j \neq 0$. Cela donne

$$\vec{B}_1 = G^*\left[\left\{ (a^4 + b^2)||\vec{S}_0||^2 + a^4 \sum_{j\neq 0} ||\vec{S}_j||^2 \right\} \underset{\sim}{I} + a^4 \sum_{j\neq 0} \vec{S}_j^*\vec{S}_j^T \right] \vec{S}_0^*,$$

$$\vec{B}_1 = G^*[\{ (b^2 - a^4)||\vec{S}_0||^2 + a^4 PE \}\underset{\sim}{I} + a^4\underset{\sim}{\mathscr{I}}]\vec{S}_0^*. \tag{a-6}$$

En outre

$$\vec{B}_2 = G^* a^2 P \sigma_b^2 \vec{S}_0^* , \tag{a-7}$$

$$\vec{B}_3 = G^* a^2 \underset{\sim}{R}_B \vec{S}_0^* . \tag{a-8}$$

D'après les 3 dernières équations, $\| \vec{B} \|$ est borné.
Si l'on désire calculer une borne, on peut utiliser la trace des matrices $\underset{\sim}{\mathscr{L}}$ et $\underset{\sim}{R}_B$ comme majorant de leurs valeurs propres. Il vient

$$\| \vec{B} \| \leqslant | G | a^2 P \sqrt{E} \left(\frac{| b^2 - a^4 | E}{a^2 P} + 2a^2 E + 2\sigma_b^2 \right) < \infty . \tag{a-9}$$

3) Étude de $\underset{\sim}{A}$

Le développement obtenu en introduisant (80) dans (32) fait apparaître 16 termes, dont 8 sont nuls parce que du premier degré soit en $\vec{B}_k$, soit en $m(\tau_k)$. Deux autres sont nuls parce qu'ils contiennent des moments du type $E(m^2(k\Delta))$ (Hypothèse (66)). Nous considérons les 6 termes restants en effectuant la moyenne d'ensemble sur la gigue qui ne fait apparaître que des moments d'ordre 2. Il vient d'abord

$$\underset{\sim}{A}_1 = E \left\{ \sum_{ij} m^*(\tau_k - i\Delta) m(\tau_k - j\Delta) \vec{S}_i^{*T} \vec{S}_j \underset{\sim}{D}_k^* \sum_{lm} m^*(\tau_k - l\Delta) m(\tau_k - m\Delta) \vec{S}_l^* \vec{S}_m^T \underset{\sim}{D}_k \right\} . \tag{a-10}$$

Il est facile de voir que la forme quadratique associée à cette matrice vaut

$$\vec{H}^{*T} \underset{\sim}{A}_1 \vec{H} = E \left\{ \left\| \sum_i m(\tau_k - i\Delta) \vec{S}_i \right\|^2 \left| \sum_l m(\tau_k - l\Delta) \vec{S}_l^T \underset{\sim}{D}_k \vec{H} \right|^2 \right\} \geq 0 \tag{a-11}$$

de sorte que $\underset{\sim}{A}_1$ est définie non négative. Pour vérifier que cette matrice est bornée, on peut donc montrer que sa trace est finie. Or

$$\text{trace } \underset{\sim}{A}_1 = E \left\{ \sum_{\alpha=0}^{P-1} \sum_{ijlm} m^*(\tau_k - i\Delta) m(\tau_k - j\Delta) m^*(\tau_k - l\Delta) m(\tau_k - m\Delta) \vec{S}_i^{*T} \vec{S}_j s_{l+\alpha}^* s_{m+\alpha} \right\} . \tag{a-12}$$

Comme au paragraphe précédent, on calcule les contributions suivant les valeurs respectives de i, j, l, m ; 3 seulement sont non nulles ($i = j = l = m$), ($i = j$, $l = m \neq i$) ($i = m$, $j = l \neq i$). Cela donne

$$\text{trace } \underset{\sim}{A}_1 = (b^2 - a^4) \sum_{\alpha=0}^{P-1} \sum_i \| \vec{S}_i \|^2 s_{i+\alpha}^* s_{i+\alpha} + a^4 \sum_{\alpha=0}^{P-1} \sum_{i,l} \| \vec{S}_i \|^2 | s_{l+\alpha} |^2 + a^4 \sum_{\alpha=0}^{P-1} \sum_{i,j} \vec{S}_i^{*T} \vec{S}_j s_{j+\alpha}^* s_{i+\alpha} \tag{a-13}$$

$$= (b^2 - a^4) \sum_i \| \vec{S}_i \|^4 + a^4 P^2 E^2 + a^4 \text{ trace } (\underset{\sim}{\mathscr{L}}^2) . \tag{a-14}$$

La série $\sum_i \| \vec{S}_i \|^4$ converge, comme la série $\sum_i \| \vec{S}_i \|^2$:

$$E_2 \triangleq \sum_i \| \vec{S}_i \|^4 < \infty . \tag{a-15}$$

En outre $\mathscr{L}^2$ est définie non négative et bornée comme $\mathscr{L}$ et

$$\text{trace } \mathscr{L}^2 \leqslant (\text{trace } \mathscr{L})^2 . \tag{a-16}$$

Il découle de (a-14), (a-16) que $\underset{\sim}{A}_1$ est bornée et satisfait

$$\text{Trace } \underset{\sim}{A}_1 \leqslant | b^2 - a^4 | E_2 + 2a^4 P^2 E^2 . \tag{a-17}$$

Il est facile de calculer les autres termes du développement de la matrice $\underset{\sim}{A}$. On trouve

$$\underset{\sim}{A}_2 = E\left\{ \sum_{ij} m^*(\tau_k - i\Delta) m(\tau_k - j\Delta) \vec{S}_i^{*T} \vec{S}_j \vec{B}_k^* \vec{B}_k^T \right\} = a^2 P E \underset{\sim}{R}_B , \tag{a-18}$$

$$\underset{\sim}{A}_3 = E\left\{ || \vec{B}_k ||^2 \sum_{ij} m^*(\tau_k - i\Delta) m(\tau_k - j\Delta) \underset{\sim}{D}_k^* \vec{S}_i^* \vec{S}_j^T \underset{\sim}{D}_k \right\} = a^2 P \sigma_b^2 \mathscr{L}' , \tag{a-19}$$

$\underset{\sim}{A}_4 = E\{ || \vec{B}_k ||^2 \vec{B}_k^* \vec{B}_k^T \}$, matrice définie non négative dont la trace, d'après l'inégalité de Schwarz vérifie (a-20)

$$\text{trace } \underset{\sim}{A}_4 = \sum_{\alpha,\beta=0}^{P-1} E\{ | b_{k+\alpha} |^2 | b_{k+\beta} |^2 \} \leqslant P^2 E(| b_{k+\alpha} |^4) . \tag{a-21}$$

Elle est finie puisque le bruit a son 4e moment borné

$$\text{trace } \underset{\sim}{A}_4 \leqslant P^2 m_4 . \tag{a-22}$$

Les deux dernières composantes de $\underset{\sim}{A}$ sont des matrices adjointes satisfaisant

$$\underset{\sim}{A}_5 = \underset{\sim}{A}_6^{T*} = a^2 \underset{\sim}{R}_B \mathscr{L}' . \tag{a-23}$$

Elles sont définies non négatives et leurs valeurs propres sont bornées par $P^2 \sigma_b^2 E$.
Il découle de ce résultat et des équations (a-17, 18, 19, 22) que

$$\text{trace } \underset{\sim}{A} \leqslant | b^2 - a^4 | E_2 + P^2 a^2 E^2 \left[2a^2 + 4 \frac{\sigma_b^2}{E} + \frac{m_4}{a^2 E^2} \right] < \infty . \tag{a-24}$$

Ainsi les moments (32)-(34) sont finis.

Remarquons qu'il n'était pas indispensable de calculer explicitement des bornes des moments (32)-(34) pour montrer leur finitude. Cependant les bornes (a-4), (a-9), (a-24) sont très utiles car dans la pratique on en a besoin pour évaluer le pas d'incrémentation μ (cf. formule (45)).

On utilise donc ces résultats pour optimiser l'algorithme (17).

RÉFÉRENCES

[1] L. SCHEMETTERER, Stochastic Approximation. *Fourth Berkeley Symposium*, 1961, p. 587-609.

[2] A. DVORETSKY, On stochastic approximation. *Proceedings of the third Berkeley Symposium on Mathematical Statistics and Probability*, University of California Press, Vol. **1**, 1956, p. 39-55.

[3] J. M. MENDEL, K. S. FU, *Adaptive, learning and pattern recognition systems*. Academic Press, New York, 1970.

[4] C. MACCHI, Itération Stochastique et Traitements Numériques Adaptatifs. *Thèse d'État*, Paris, 1972.

[5] C. MACCHI, J. P. JOUANNAUD, O. MACCHI, Récepteurs adaptatifs pour transmission de données à grande vitesse. *Annales Télécom.*, t. **30**, n° 9-10, 1975, p. 311-330.

[6] D. C. FARDEN, L. L. SCHARF, Convergence results for adaptive filtering with correlated data. Report of the Colorado State University soumis aux *I. E. E. E. Trans. on Information Theory*, 1975.

[7] D. C. FARDEN, Stochastic Approximation with correlated data. *Ph. D. Dissertation*, Department of Electrical Engineering Colorado State University, 1975.

[8] G. UNGERBOECK, Theory of the Speed of Convergence in Adaptive Equalizers for Digital Communication. *IBM J. Res. Develop.*, Nov. 1972, p. 546-555.

[9] M. LEVY, O. MACCHI, Égaliseur de gigue. *6^e Colloque National sur le Traitement du Signal et ses Applications*, Nice, avril 1977, p. 82/1-82/8.

[10] O. MACCHI, Estimation Linéaire Adaptative. Applications aux Transmissions de Données. *6^e Colloque National sur le Traitement du Signal et ses Applications*, Nice, avril 1977, p. 85/1-85/7.

[11] M. LEVY, Élimination conjointe des interférences intersymboles et des écarts de phase dans les systèmes de transmission de données. *Thèse de Docteur-Ingénieur*, Orsay, juillet 1977.

[12] O. MACCHI, J. F. GUILBERT, Réseaux Téléinformatiques Éditeurs, chapitre 3. *A paraître*, Dunod 1978.

[13] GRENANDER, SZEGO, *Toeplitz forms and their application*. University of California Press, 1958 (chap. 5).

(Manuscrit révisé, reçu le 26 avril 1978)