

Le filtrage adaptatif transverse

Transversal adaptive filtering



Odile MACCHI

Laboratoire des Signaux et Systèmes, ESE, plateau du Moulon, 91190 GIF-SUR-YVETTE.

Odile Macchi est née en 1943 à Aurillac, Cantal. Ancienne élève de l'École Normale Supérieure de jeunes filles, elle est reçue septième à l'Agrégation de Mathématiques en 1966. Elle obtient son Doctorat d'État en Sciences Physiques en 1972 et depuis lors, elle se consacre à la recherche en traitement du signal, spécialement en théorie des communications. Elle est directeur de recherche au CNRS et dirige au sein du Laboratoire de Signaux et Systèmes de l'École Supérieure d'Électricité, une équipe qui travaille en communications, tout particulièrement sur les systèmes adaptatifs en transmission. Actuellement elle est adjointe au directeur du GRECO Traitement du Signal et Image. Pour la partie transmission de données, elle est coauteur du livre **Téléinformatique** qui a fait l'objet d'un très gros tirage et de trois traductions. Elle a reçu la médaille Blondel, le prix du général Kampé de Férié de l'Académie des Sciences et la distinction senior member des IEEE.



Maurice BELLANGER

Laboratoire des Signaux et Systèmes, ESE, plateau du Moulon, 91190 GIF-SUR-YVETTE.

Ingénieur ENST (1965) et Docteur ès Sciences (1981), Directeur Scientifique de la Société TRT, de 1984 à 1987 professeur associé à l'Université de Paris-Sud et chercheur au Laboratoire des Signaux et Systèmes de l'ESE. Auteur de deux ouvrages aux éditions Masson, sur le traitement Numérique du Signal en théorie et pratique d'une part et sur l'analyse des signaux et le filtrage adaptatif d'autre part. Domaines d'intérêt : Filtrage Numérique, filtres adaptatifs, analyse des signaux, réalisations du traitement numérique, logiciels et stations de travail, applications. Président de « European Association for Signal Processing » (EURASIP).

RÉSUMÉ

Cet article présente une synthèse sur les algorithmes du gradient (G) et des moindres carrés rapides (MCR) utilisés communément dans les filtres adaptatifs transverses pour estimer un modèle linéaire. Nous montrons comment les algorithmes utilisent la récursivité pour réaliser en une seule opération et le moyennage inclus dans le critère d'optimisation et la minimisation. En régime permanent où l'on peut considérer que les deux algorithmes minimisent un même critère d'erreur quadratique moyenne, ils ont des performances équivalentes : leurs fluctuations résiduelles sont les mêmes à une équivalence près entre le pas d'adaptation μ pour (G) et le taux d'oubli $(1 - w)$ pour (MCR). C'est en régime transitoire que les moindres carrés sont supérieurs de par le critère même qui privilégie la vitesse. Ceci est au prix d'une complexité environ quatre fois supérieure et de difficultés de contrôle des variables numériques, liées pour une part à la présence de divisions. Une méthodologie générale est donnée pour découpler les effets du transitoire et des fluctuations, lesquelles proviennent du bruit de mesure sur le modèle. Elle s'étend à la poursuite des variations du modèle et l'on montre la différence entre ce problème et le raccourcissement du transitoire lorsque le modèle est fixe : les deux pas d'adaptation sont nettement différents.

MOTS CLÉS

Filtrage adaptatif, gradient, moindres carrés, erreur quadratique moyenne, erreur quadratique cumulée, matrice de covariance, vecteur d'intercorrélacion, pas d'incrémentacion, régime transitoire, fluctuation, poursuite, algorithme rapide, prédiction avant, prédiction arrière, facteur de pondération.

SUMMARY

This paper presents a tutorial on the gradient (G) and recursive least squares (RLS) algorithms, both commonly used in adaptive transversal filters for estimating a linear model. It is shown how the algorithms utilize recursivity to realize at the same time the averaging involved in the optimisation criterion as well as the minimization. In steady state both algorithms can be viewed as minimizing the same mean square error criterion. Then they have equivalent performances: the same residual fluctuations up to an equivalence between the adaptation step-size μ of (G) and the forgetting rate $(1-w)$ of (RLS). Due to the criterion itself emphasizing adaptation speed, it is the transient period which exhibits the (RLS) superiority. This is essentially at the price of a four times higher complexity and of difficulties with controlling numerical effects, related (in part) to the presence of divisions that are not in (G). A general methodology for decoupling the transient effects and permanent fluctuations is given; the latter are due to the measurement noise disturbing the model. The more general problem of tracking the model variations is treated with that methodology and we emphasize the difference between the problems of tracking and of squeezing the transient period for a fixed model: the corresponding adaptation step-sizes differ significantly.

KEY WORDS

Adaptive filtering, gradient, least squares, mean square error, cumulative squared error, covariance matrix, cross-correlation vector, step-size, transient behaviour, fluctuations, tracking, fast algorithm, forward prediction, backward prediction, forgetting factor.

I. Introduction

I.1. GÉNÉRALITÉS

Le traitement adaptatif du signal est un sujet d'importance majeure par ses applications. Il s'agit d'une classe d'algorithmes d'optimisation mis en œuvre en temps réel et assurant automatiquement l'optimisation d'un système en fonction de son environnement et son adaptation lors des fluctuations de ce dernier. En pratique ces propriétés sont fondamentales car elles évitent l'intervention d'un opérateur et permettent un fonctionnement sans coupures au cours du temps.

Les techniques adaptatives s'étendent maintenant à une variété de problèmes de détection, d'estimation et d'identification, rencontrés en transmissions, acoustique sous-marine, séismologie, instrumentation biomédicale, traitement de la parole et des images, reconnaissance des formes, contrôle de procédés, etc. L'importance prise récemment par les traitements adaptatifs découle naturellement des progrès technologiques permettant des calculs très élaborés avec les microcomposants rapides. C'est dans le domaine de l'égalisation des voies de transmission de données que les premiers filtres adaptatifs (FA) ont été vulgarisés vers 1965. Antérieurement la réalisation du récepteur optimal, connu théoriquement, butait sur l'adaptation des paramètres aux caractéristiques de la voie, ce qui limitait le débit de transmission. C'est l'utilisation d'un FA numérique qui a permis de s'approcher du débit limite prévu par Shannon dans les années 40.

I.2. FILTRAGE ADAPTATIF

Parmi tous les systèmes adaptatifs, les FA numériques sont les plus utilisés à cause de l'importance pratique des traitements linéaires. Dans ces derniers la sortie $\hat{y}(n)$ du système à l'instant n est une forme linéaire de la suite $\{x(k)\}$ des entrées du système, laquelle est toujours supposée connue. Pour limiter notre propos, cet article ne considère que le cas des filtres transverses pour lesquels il existe un entier N (longueur du filtre) et un vecteur $H = (h^0, h^1, \dots, h^{N-1})^T$

de filtrage tels que la sortie s'écrive :

$$(1.2.1) \quad \hat{y}(n) = \sum_{j=0}^{N-1} h^j x(n-j) \triangleq H^T X(n)$$

avec la notation

$$(1.2.2) \quad X(n) = (x(n), x(n-1), \dots, x(n+1-N))^T$$

pour le vecteur des entrées successives présentes dans le filtre H à l'instant n (voir fig. 1).

Très souvent le filtrage est mis en œuvre pour fournir en sortie une estimation $\hat{y}(n)$ d'une suite de référence $y(n)$ et le vecteur H est choisi pour minimiser un critère quadratique par rapport à l'erreur $y(n) - \hat{y}(n)$ après un moyennage approprié. Soit $\mathcal{J}(H)$ ce critère; c'est une fonction de coût, qui dépend du filtre H , mais aussi des entrées successives $x(n), x(n-1), \dots$ et des références $y(n), y(n-1), \dots$ ou de la statistique de ces deux suites. L'optimisation du filtre consiste à déterminer le (ou les) vecteur(s) \hat{H} qui minimise(nt) $\mathcal{J}(H)$. Le problème considéré comporte donc deux aspects :

— pour un vecteur H fixé, l'aspect filtrage consiste à calculer à partir des entrées $x(n), x(n-1), \dots$ la filtrée $\hat{y}(n)$ (1.2.1) qui sera considérée comme une estimation de la référence $y(n)$. Dans cette phase la référence $y(n)$ est naturellement inconnue. Le filtre est un simple estimateur;

— pour une suite de référence $y(n), y(n-1), \dots$ connue, l'aspect optimisation consiste à calculer encore à partir des entrées $x(n), x(n-1), \dots$, le vecteur de filtrage optimal \hat{H} .

Il y a essentiellement deux manières de traiter sous ses deux aspects le problème de l'optimisation du filtrage.

Optimisation par blocs

Dans cette méthode on distingue deux phases au cours du temps. Une phase d'optimisation proprement dite, de durée τ_0 , au cours de laquelle un unique vecteur \hat{H} minimisant $\mathcal{J}(H)$, est déterminé sur la base des informations $(x(n), y(n))$, $0 \leq n < \tau_0$. La quantité τ_0 est souvent appelée temps d'intégration car elle corres-

pond essentiellement à des moyennages sur les grandeurs $x(n)$, $y(n)$ et sur l'erreur d'estimation $e(n)$. On peut par exemple calculer, pour tout H

$$(1.2.3) \quad \hat{\mathcal{J}}(H) = \frac{1}{\tau_0} \cdot \sum_{n=0}^{\tau_0-1} (y(n) - \hat{y}(n))^2$$

qui est une estimation du critère, et choisir le vecteur \hat{H} qui minimise $\hat{\mathcal{J}}(H)$. Puis une phase de filtrage proprement dit sur un intervalle consécutif $[\tau_0, \tau_0 + \tau_f]$, avec usuellement $\tau_f \gg \tau_0$, phase au cours de laquelle on utilise le vecteur de filtrage \hat{H} pour fournir les estimations $\hat{y}(n)$ de la référence inconnue $y(n)$.

Naturellement la quantité (1.2.3) est d'autant plus proche du coût réel $\mathcal{J}(H)$ que le temps d'intégration τ_0 est plus grand. Il subsiste donc un écart à l'optimalité $V = H - \hat{H}$ dans le filtre et par suite une erreur résiduelle d'estimation $\varepsilon(n)$ sur la référence $y(n)$:

$$(1.2.4) \quad \varepsilon(n) = -V^T X(n).$$

Il est clair que V et ε sont (en moyenne) fonctions décroissantes de τ_0 . Pourtant on ne peut augmenter τ_0 au-delà d'un seuil pour ne pas réduire à néant la phase τ_f de filtrage. D'autre part τ_0 est limité par la durée de stationnarité des signaux $(x(n), y(n))$, au-delà de laquelle la quantité (1.2.3) n'a plus la signification d'une moyenne. Dans cette procédure d'optimisation on doit donc refaire périodiquement des estimations de \hat{H} pour tenir compte de sa possible évolution. Ainsi les performances d'estimation sur la phase de filtrage sont-elles fondamentalement déterminées par le choix du temps d'intégration τ_0 .

Filtrage adaptatif

Dans cette méthode, au cours de la phase d'optimisation le filtre \hat{H} est approché par une suite de vecteurs $H(n)$ dont l'évolution est gouvernée par un algorithme récursif du type

$$(1.2.5) \quad H(n) = H(n-1) + \mu \Delta_n,$$

où la quantité positive μ , appelée pas d'adaptation est choisie par l'utilisateur, tandis que l'incrément Δ_n dépend des valeurs $x(n)$, $x(n-1)$, ..., $y(n)$, $y(n-1)$, ... et d'une manière spécialement critique de l'erreur

$$(1.2.6) \quad \begin{cases} e(n) = y(n) - \hat{y}(n); \\ \hat{y}(n) = H(n-1)^T X(n). \end{cases}$$

Ceci est illustré sur la figure 1. La phase d'optimisation comporte donc elle-même le filtrage pour calculer $\hat{y}(n)$. Il est clair que le pas μ contrôle la vitesse

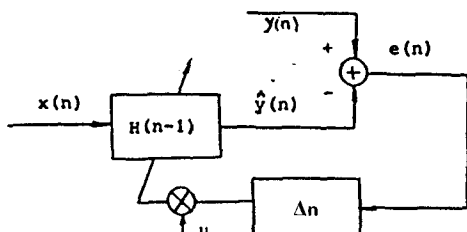


Fig. 1. - Filtrage adaptatif.

d'adaptation de sorte que μ^{-1} joue le rôle du temps d'intégration τ_0 dans une optimisation par blocs. Si μ est trop faible l'algorithme sera trop long à converger (τ_0 trop grand). Il ne sera pas non plus capable de s'adapter aux non-stationnarités des signaux $(x(n), y(n))$.

I.3. PRÉSENTATION DE L'ARTICLE

Dans ce qui suit, notre propos est une étude théorique manipulant seulement des outils simples à la portée du plus grand nombre de lecteurs. Nous considérons le FA le plus usité qui est le filtre transverse (1.2.1), avec ses deux algorithmes d'adaptation les plus connus, l'algorithme dit « du gradient » (G) et l'algorithme « des moindres carrés rapides » (MCR), selon des terminologies maintenant classiques.

Le gradient vise à minimiser un critère d'erreur quadratique moyenne (EQM)

$$(1.3.1) \quad \mathcal{J}(H) = E(e(n)^2)$$

— calculé avec un filtre H fixe — qui est une moyenne d'ensemble. Dans cette approche les signaux $x(n)$, $x(n-1)$, ..., $y(n)$, ... sont donc considérés comme aléatoires. Si le filtre se stabilise à la fin de la récursion (1.2.5) et dans un contexte ergodique, cette moyenne d'ensemble fournira en régime permanent — c'est-à-dire asymptotiquement pour n grand — la même valeur que la moyenne temporelle

$$(1.3.2) \quad \mathcal{J}_n(H) = \frac{1}{n} \sum_{j=1}^n e(j)^2$$

— calculée elle aussi avec un filtre H fixe — cette moyenne dont l'horizon est limité est justement celle que l'algorithme des moindres carrés rapides cherche à minimiser, dans un environnement qui — on le voit sur (1.3.2) — ne nécessite pas d'hypothèse sur la nature aléatoire ou déterministe des signaux traités. Puisque (1.3.1) et (1.3.2) coïncident asymptotiquement les algorithmes (G) et (MCR) donneront le même optimum en régime permanent. En conséquence les deux approches ne se distingueront que par leur transitoire, ainsi qu'il apparaîtra dans la suite. Comme l'algorithme (MCR) a une complexité de calcul environ 4 fois plus grande que (G), son intérêt essentiel concerne la phase transitoire, c'est-à-dire la vitesse d'acquisition. De par sa conception même, l'algorithme (G) est au contraire intéressant pour son régime permanent.

II. Approche moyenne : le gradient

De tous les algorithmes de FA, c'est le plus répandu car le plus simple à mettre en œuvre. On le trouve déjà dans un nombre considérable d'appareils commercialisés.

II.1. PREMIÈRE APPROCHE DU GRADIENT

Comme il a été dit en introduction on cherche par une récursion le minimum sur H de l'EQM (1.3.1):

$$(2.1.1) \quad \mathcal{J}(H) = E((y(n) - H^T X(n))^2),$$

en supposant que $y(n)$, $X(n)$ sont des variables aléatoires stationnaires. En introduisant les moyennes

$$(2.1.2) \quad \begin{cases} R = E(X(n)X(n)^T), \\ \text{matrice de covariance de l'entrée} \end{cases}$$

$$(2.1.3) \quad \begin{cases} r = E(y(n)X(n)), \\ \text{vecteur d'intercorrélation} \end{cases}$$

l'EQM s'écrit :

$$(2.1.4) \quad \mathcal{J}(H) = E(y(n)^2) - 2H^T r + H^T R H.$$

C'est une forme quadratique définie positive en H qui possède un minimum pour une unique valeur \tilde{H} de H pourvu que R soit inversible ce que nous supposons dans toute la suite.

II.1.1. Présentation et philosophie de l'algorithme

Le gradient de l'EQM (2.1.1) est :

$$(2.1.5) \quad F(H) = -2E((y(n) - H^T X(n))X(n)),$$

$$(2.1.6) \quad = -2(r - RH).$$

Il s'annule pour le minimum recherché

$$(2.1.7) \quad \tilde{H} = R^{-1}r.$$

Les fonctions $\mathcal{J}(H)$ et $F(H)$ sont déterministes car les moyennes ont été faites dans R et r . Ainsi l'algorithme du gradient

$$(2.1.8) \quad H(n) = H(n-1) - \frac{\mu}{2} F(H(n-1))$$

$$(2.1.9) \quad = H(n-1) + \mu(r - RH(n-1))$$

est-il déterministe. Cet algorithme est dit aussi « algorithme de descente suivant la ligne de plus grande pente ». Il permet de résoudre récursivement, sans inversion matricielle le problème déterministe de la minimisation de $\mathcal{J}(H)$ selon le schéma de la figure 2.1 (cas où H est bidimensionnel).

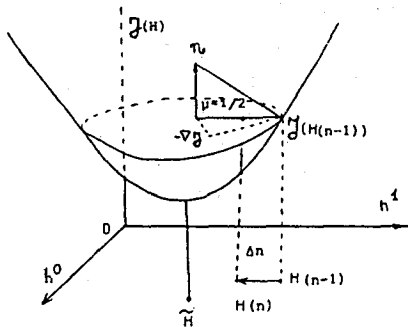


Fig. 2.1. - Minimisation récursive de l'EQM.

Il est utile de mentionner une autre structure du gradient déterministe, utilisant la notation introduite en (1.2.6) pour l'erreur d'estimation à la sortie du filtre dans son état actuel $H(n-1)$. L'intérêt de noter $H(n-1)$ le filtre qui traite le vecteur $X(n)$ provient du gradient stochastique, dans lequel $H(n-1)$ est calculé en tenant compte des informations passées jusqu'à $y(n-1)$ et $X(n-1)$ inclus. D'après (2.1.6),

l'algorithme s'écrit encore

$$(2.1.10) \quad H(n) = H(n-1) + \mu E(e(n)X(n))$$

qui est un algorithme déterministe strictement équivalent à (2.1.9). Tous deux supposent connues les moyennes d'ensemble.

Dans la pratique pourtant, on ignore le plus souvent la statistique des couples successifs $(y(n), X(n))$ et les expressions explicites de $\mathcal{J}(H)$ et de $F(H)$ en fonction de H , puisqu'elles contiennent des moyennes d'ensemble. On ne peut donc appliquer directement la procédure (2.1.9) du gradient déterministe. Il s'avère pourtant que la moyenne d'ensemble n'est pas indispensable. La suppression de cette moyenne dans (2.1.6), (2.1.9) selon

$$(2.1.11) \quad H(n) = H(n-1) + \mu(y(n) - H(n-1)^T X(n))X(n)$$

rend stochastique le vecteur $H(n)$ comme l'incrément de l'algorithme. C'est une caractéristique qui peut sembler défavorable, mais nous montrerons dans la suite que l'on peut en limiter l'impact de manière satisfaisante. Intuitivement ce sont les itérations successives qui dans l'algorithme (2.1.11) réalisent la moyenne d'ensemble et c'est ce qui fait le succès de cette procédure stochastique. La récursivité de l'algorithme « fait d'une pierre deux coups ». Par la seule addition de l'incrément, le filtrage adaptatif réalise asymptotiquement, et la moyenne d'ensemble et la minimisation. On voit en effet que l'algorithme (2.1.11) s'écrit :

$$(2.1.12) \quad H(n+p) = H(n) + p\mu \left[\frac{1}{p} \sum_{j=1}^p e(n+j)X(n+j) \right].$$

Or, sous des conditions convenables d'ergodisme, pour p assez grand

$$(2.1.13) \quad \frac{1}{p} \sum_{j=1}^p e(n+j)X(n+j) \approx E(e(n+1)X(n+1));$$

donc l'algorithme stochastique peut être approché par

$$(2.1.14) \quad H(n+p) = H(n) + p\mu E(e(n+1)X(n+1)).$$

C'est là justement l'algorithme déterministe (2.1.10) en prenant $p\mu$ comme nouveau pas d'incrément. L'algorithme stochastique prend donc les moyennes d'ensemble à sa charge; moyennage et minimisation sont simultanés.

Dans la formule (2.1.11) on voit apparaître deux phases selon

$$(F) \quad e(n) = y(n) - H(n-1)^T X(n),$$

$$(G) \quad H(n) = H(n-1) + \mu e(n)X(n).$$

La phase (F) est le filtrage proprement dit qui produit une erreur d'estimation $e(n)$ liée à l'état actuel $H(n-1)$ du filtre. Cette erreur contrôle la phase (G)

d'adaptation puisque l'incrément lui est proportionnel.

L'algorithme (FG) qui est destiné à minimiser un carré moyen s'appelle en anglais least mean squares (LMS). Il a été étudié pour la première fois par Widrow [1]. Il est très peu coûteux en calculs. L'intervention du pas d'adaptation dans (G) est insignifiante sur ce plan car μ est normalement une puissance négative de 2. En comparant l'adaptation (G) au filtrage (F), on voit donc que les deux calculs ont la même complexité arithmétique : N multiplications, N additions. C'est un prix modique à payer pour obtenir l'adaptativité du filtrage. Avec l'avantage d'une bonne distribution dans le temps des moyens de calcul requis. Le fonctionnement du filtre se fait sans à coups et sans interruption, par opposition à une optimisation par blocs, où le calcul de \tilde{H} , préalable au filtrage lui-même, exigerait de beaucoup plus gros calculs. Dans une telle procédure de réinitialisation périodique de \tilde{H} , on peut résoudre plus rapidement l'équation d'optimalité, mais la puissance de calcul nécessaire deviendra inutilisée dans la phase suivante de filtrage. Au contraire (FG) remet à jour en permanence et sans grands frais de calcul l'estimation du filtre optimal \tilde{H} .

L'algorithme (FG) possède encore deux avantages. Il est robuste vis-à-vis du contexte de son utilisation : nous verrons dans la suite que les seules véritables conditions pratiques à la convergence de H vers \tilde{H} sont le choix du gain (ni trop grand, ni trop petit) et l'inversibilité de R . Enfin, nous verrons ses bonnes propriétés de stabilité numérique : la précision de la convergence est proportionnelle au pas d'adaptation μ donc au nombre de bits de la réalisation machine.

II.1.2. Convergence du gradient déterministe

Nous commençons par étudier la convergence, plus simple, de l'algorithme déterministe. Appelons $\lambda_1, \dots, \lambda_N$ les valeurs propres de la covariance R et U_1, \dots, U_N les vecteurs propres (orthonormés) associés. À l'aide de la matrice P des U_p , R peut être diagonalisée selon

$$(2.1.15) \quad R = P^T \Lambda P; \quad P^T = P^{-1},$$

Λ étant la matrice diagonale des λ_i . En introduisant la déviation du filtre adaptatif par rapport à l'optimalité

$$(2.1.16) \quad V(n) = H(n) - \tilde{H}$$

l'algorithme (2.1.9) s'écrit

$$(2.1.17) \quad PV(n) = (I - \mu \Lambda) PV(n-1),$$

I désignant la matrice identité. Comme R est inversible, tous les λ_i sont strictement positifs. Comme le vecteur $PV(n)$ on voit donc que le vecteur $V(n)$ tend exponentiellement vers zéro si et seulement si

$$(2.1.18) \quad 0 < \mu < 2/\lambda_{\max}$$

À chaque vecteur propre U_i est associé un temps de convergence τ_i qui est le nombre d'itérations requis pour que $|V(n)| \leq \alpha |V(0)|$, α étant une atténuation

prédéterminée. Il est clair que

$$(2.1.19) \quad \tau_i \geq \log \alpha / \log(1 - \mu \lambda_i).$$

Lorsque le pas d'adaptation μ est faible, il vient

$$(2.1.20) \quad \tau_i \approx \frac{-\log \alpha}{\mu \lambda_i}.$$

Cette formule montre que la vitesse de l'algorithme est proportionnelle à μ . Mais elle n'est pas uniforme dans tout l'espace puisque les λ_i ne sont pas tous égaux. C'est là un des inconvénients du gradient. En notant

$$(2.1.21) \quad \rho = \lambda_{\max} / \lambda_{\min}$$

la dispersion de la covariance R , la borne (2.1.18) montre en particulier que le temps de convergence associé à λ_{\min} vérifie

$$(2.1.22) \quad \tau_{\max} \geq \frac{\rho}{2} (-\log \alpha).$$

Pour certaines matrices de covariance, ce temps peut être très grand.

La précédente analyse étant basée sur la connaissance de la matrice R , dans la pratique il faudra préalablement estimer R qui est inconnue.

II.1.3. Difficultés du gradient stochastique

La première difficulté de l'algorithme stochastique (FG) provient de ce que nous appelons le « bruit de sortie » défini comme

$$(2.1.23) \quad b(n) = y(n) - \tilde{H}^T X(n).$$

C'est l'écart entre la référence $y(n)$ et l'estimation $\tilde{y}(n) = \tilde{H}^T X(n)$ qui en est donnée par le filtre optimal \tilde{H} . Ce peut être par exemple le bruit de mesure de $\tilde{y}(n)$, qui est inévitable. Encore est-ce là le cas le plus favorable, car on pourra alors supposer l'indépendance entre les suites $\{X(n)\}$ et $\{b(j)\}$. De toutes façons, l'équation d'optimalité (2.1.7) rend ce bruit décorrélé des entrées selon

$$(2.1.24) \quad E(b(n) X(n)) = 0.$$

Nous noterons

$$(2.1.25) \quad E(b^2(n)) = E_{\min},$$

sa puissance qui sera supposée non nulle. Avec les notations (2.1.16), l'erreur qui contrôle l'algorithme s'écrit

$$(2.1.26) \quad e(n) = b(n) + \varepsilon_n$$

$$(2.1.27) \quad \varepsilon(n) \stackrel{\Delta}{=} -V^T(n-1) X(n),$$

ce deuxième terme étant l'erreur résiduelle qui découle de la non optimalité de $H(n-1)$. L'algorithme (FG) s'écrit

$$(2.1.28) \quad V(n) = (I - \mu X(n) X(n)^T) V(n-1) + \mu b(n) X(n),$$

et l'on voit que le bruit de sortie $b(n)$ brouille le lien entre $V(n)$ et $V(n-1)$. La convergence de l'algorithme est gênée puisque d'après (2.1.28) $V(n-1)=0$ n'implique pas $V(n)=0$. Après avoir atteint \tilde{H} le filtre adaptatif s'en écarte.

La deuxième difficulté provient de la corrélation des entrées successives $x(n-j)$ formant le vecteur $X(n)$. Les vecteurs $X(n-1)$ et $X(n)$ partagent donc $N-1$ coordonnées et sont fortement corrélés. D'après (2.1.28) il s'ensuit que $X(n)$ et $V(n-1)$ sont corrélés, ce qui rend très difficile les calculs de moments pour $V(n)$. Ainsi une théorie simple telle qu'établie dans [2] et basée sur l'indépendance de $X(n)$ avec $V(n-1)$ n'est pas du tout réaliste. Cette corrélation entre les $X(n)$ est inhérente à la dynamique du processus d'entrée. Elle peut être décrite par un schéma de corrélation à l'aide de la méthode instrumentale [25]. Mais avec l'inconvénient qu'il est très difficile voir impossible d'obtenir un bon estimateur des covariances.

Un des propos de ce travail est de surmonter ces deux obstacles avec des outils du calcul des probabilités assez simples pour que tout ingénieur les manie aisément. C'est ce que nous faisons dans la suite de ce paragraphe II.

II.2. THÉORIE DE LA CONVERGENCE

Afin d'éviter l'hypothèse classique mais irréaliste d'indépendance des $X(n)$ successifs, nous introduisons d'abord quelques hypothèses mathématiques très peu contraignantes pour une application pratique.

II.2.1. Hypothèses

Donnons ici des conditions suffisantes assurant un comportement satisfaisant de l'algorithme stochastique (FG).

M-indépendance

- (E) $\forall k$ les suites $\{(b(n), X(n)); n \leq k\}$ et $\{(b(n), X(n)), n \geq k + M\}$ sont indépendantes.

D'après cette hypothèse, il suffit d'espacer dans le temps les signaux considérés pour les rendre indépendants; l'entier M est la mémoire. En particulier les vecteurs d'entrée $X(n)$ doivent être M indépendants ce qui implique que M est au moins égal au nombre N de composantes. Dans la pratique on peut toujours trouver M , fût-ce grand, qui assure la condition (E) d'espacement.

Bornitude

- (B) $\forall n, \forall l > 0, E(|X(n)|^l) \leq B_l$
- (B') $\forall n, E(b(n)^5) \leq B'$.

D'après l'hypothèse (B), l'entrée $|X(n)|$ a tous ses moments finis. Mais ceci n'exclut pas que $|X(n)|$ prenne de grandes amplitudes pourvu que cet événement soit rare. Cette hypothèse n'est pas stricte pour la pratique où tous les systèmes ont finalement une dynamique limitée. L'hypothèse (B') qui concerne le bruit de sortie est encore moins stricte.

Les deux hypothèses suffisent pour acquérir la convergence de l'algorithme (FG). Mais la preuve est rendue plus aisée si l'on suppose encore ceci.

Bruit blanc

- (I) La suite $\{b(n)\}$ est centrée, stationnaire, indépendante et indépendante de la suite $\{X(j)\}$.

Cette hypothèse signifie que le signal à estimer $\tilde{y}(n)$ peut être modélisé par la figure 2.2 comme la sortie d'un modèle linéaire \tilde{H} (que l'on cherche à identifier récursivement) d'entrée $x(n)$ dont la sortie est observée en présence d'un bruit blanc additif inconnu $b(n)$. C'est donc un problème typique d'identification récursive, qui conserve une très bonne généralité d'applications. Il rend exactement compte par exemple de l'annulation d'écho.

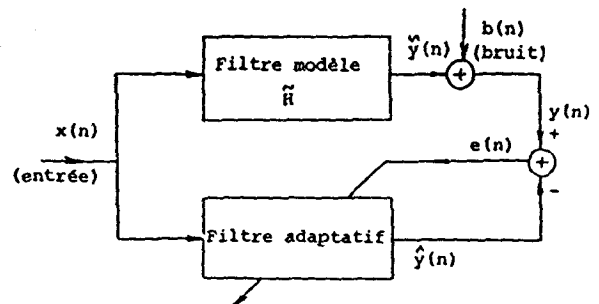


Fig. 2.2. - Identification d'un modèle linéaire par filtrage adaptatif (exemple de l'annulation d'écho).

II.2.2. Méthodologie

La méthodologie de nos démonstrations [3, 4] est de séparer la déviation $V(n)$ en une partie dite transitoire $V^t(n)$, qui dépend des conditions initiales, et une partie dite fluctuation $V^f(n)$, qui dépend du bruit de sortie, selon

$$(t) \quad V^t(n) = (I - \mu X(n)X(n)^T) V^t(n-1); \quad V^t(0) = V(0) = H(0) - \tilde{H};$$

$$(f) \quad \begin{cases} V^f(n) = (I - \mu X(n)X(n)^T) V^f(n-1) + \mu b(n)X(n); \\ V^f(0) = 0. \end{cases}$$

En comparant avec (2.1.28), on voit aussitôt que

$$(2.2.1) \quad V(n) = V^t(n) + V^f(n).$$

La composante transitoire $V^t(n)$ reflète les conditions initiales de l'algorithme, mais ne contient pas du tout de bruit $b(n)$, erreur du filtre optimal. La composante de fluctuations $V^f(n)$ est uniquement due à ce bruit mais elle ne dépend pas des conditions initiales $V(0)$. Ceci montre en particulier que la vitesse de convergence de l'algorithme devra être appréciée à partir du transitoire (t). En outre cette vitesse détermine la convergence ou la non convergence de $V^f(n)$ lui-même vu qu'en itérant (f) il vient

$$(2.2.2) \quad V^f(n) = \sum_{j=1}^n U_{j,n} \mu b(j) X(j),$$

où la matrice $U_{j,n}$ définie par

$$(2.2.3) \quad U_{j,n} = (I - \mu X(n)X(n)^T) \times (I - \mu X(n-1)X(n-1)^T) \times \dots \times (I - \mu X(j+1)X(j+1)^T); \quad n > j \\ = I; \quad n \leq j$$

est reliée au transitoire d'après

$$(2.2.4) \quad V^t(n) = U_{0,n} V(0).$$

Si le transitoire tend vers zéro, la matrice $U_{j,n}$ tend vers zéro lorsque $n-j$ tend vers l'infini et il est clair sur (2.2.2) que la série $V^t(n)$ peut alors avoir une chance de converger. Mais pour cela il faut que la décroissance de $U_{j,n}$ (donc du transitoire) soit assez rapide. C'est pourquoi dans les démonstrations de convergence une étape essentielle est de borner supérieurement [5] l'un des moments $E(\|U_{j,n}\|^m)$ ($m > 0$), où la norme matricielle considérée est $\|U\| = \sup_{(X: |X|=1)} |UX|$. Puisque la norme est non négative, un tel moment est une bonne mesure de la vitesse de convergence vers zéro du $U_{j,n}$. Dans ce qui suit nous montrons que le transitoire décroît exponentiellement.

II.2.3. Étude du transitoire

La décroissance rapide du transitoire s'énonce sous forme de théorème [4].

Théorème 1 : Si la suite $X(n)$ est stationnaire, de matrice de covariance R inversible et satisfait aux conditions (E) et (B), la déviation transitoire par rapport à l'optimalité de l'algorithme (FG) décroît exponentiellement vers zéro ainsi que tous ses moments selon

$$(2.2.5) \quad \begin{cases} \forall n \geq N_0; \quad \forall \mu \leq \mu_0, \\ \|U_{0,n}\| \leq D(1 - \delta\mu)^n \end{cases}$$

$$(2.2.6) \quad \begin{cases} \forall m > 0; \quad \forall \mu \leq \mu_m, \\ E(\|U_{0,n}\|^m) \leq C_m(1 - \delta_m\mu)^n. \end{cases}$$

Ce théorème s'appuie essentiellement sur la M-indépendance et sur la majoration suivante, liée à l'inversibilité de R et valable pour un entier P assez grand :

$$(2.2.7) \quad \forall \mu \leq \mu_m, \quad E(\|U_{n,n+P}\|^m) \leq 1 - \mu\gamma_m.$$

Dans ces inégalités $\mu_0, \mu_m, \delta, \delta_m, \gamma_m$ sont strictement positifs; $\mu_m, \delta_m, \gamma_m$ dépendent de l'entier m .

L'inégalité (2.2.5) implique la convergence presque sûre (p. s.) du transitoire vers zéro d'après (2.2.4). Il est à noter que les quantités N_0 et D sont en fait des fonctions $N_0(\omega), D(\omega)$ dépendant de l'épreuve ω tirée par le hasard. Bien que ce type de convergence soit forte par son caractère exponentiel, son efficacité est limitée par le fait qu'il n'y a pas, en général, de majorants déterministes pour les quantités aléatoires $N(\omega)$ et $D(\omega)$. D'où l'importance de la deuxième partie du théorème.

On peut donner une idée heuristique de la démonstration du théorème 1 en découpant l'intervalle $[1, n]$ en k morceaux de longueur P au moins égale à M , qui, aux effets de bord près vont rendre les vecteurs $X(n)$

indépendants sur les différents morceaux, grâce à la M-indépendance. Alors la structure produit de la matrice $U_{0,n}$ entraîne

$$(2.2.8) \quad E(\|U_{0,n}\|^m) \leq \prod_{i=0}^{k-1} E(\|U_{iP, (i+1)P}\|^m)$$

et le résultat (2.2.6) devient évident sur la base de l'inégalité (2.2.7). Le problème technique posé par les effets de bord se résoud mathématiquement par un entrelacement astucieux; voir [4]. Naturellement, grâce à la stationnarité des $X(n)$, on a le résultat plus général que

$$(2.2.9) \quad E(\|U_{j,n}\|^m) \leq C_m(1 - \delta_m\mu)^{n-j}.$$

Une conséquence intéressante de (2.2.6) est la convergence vers zéro en moyenne quadratique du transitoire. En effet d'après (2.2.4)

$$(2.2.10) \quad E(|V^t(n)|^2) \leq E(\|U_{0,n}\|^2) |V(0)|^2 \leq |V(0)|^2 C_2(1 - \delta_2\mu)^n.$$

Dans les applications, pour caractériser la convergence du filtre adaptatif on utilise souvent l'atténuation

$$(2.2.11) \quad \alpha_n = -10 \log_{10} \{E[(X(n)^T V^t(n-1))^2] / E[(X(n)^T \tilde{H})^2]\}.$$

Exprimée en décibels, c'est la puissance de l'erreur résiduelle ε_n (2.1.27), rapportée à la puissance de sortie du filtre optimal. L'atténuation caractérise bien la convergence du filtre adaptatif vers l'optimalité. D'après (2.2.10) il vient

$$(2.2.12) \quad \alpha_n \geq K_1 + K_2 n \mu$$

où K_1 et $K_2 > 0$ sont des constantes liées à la statistique des $X(n)$. L'atténuation croît linéairement avec le temps. La vitesse de convergence de l'algorithme est caractérisée par le produit $K_2\mu$. Elle est proportionnelle au pas d'adaptation μ . C'est bien ce que l'on observe dans la pratique. Ainsi la figure 2.3 présente des simulations d'annulation d'écho (cf. fig. 2.2) dans le cas particulier où le signal lointain $b(n)$ est nul (pas de bruit à la sortie du filtre optimal). Dans ces conditions la déviation de fluctuation est nulle d'après (f), et l'écart observé $V(n)$ par rapport à l'optimalité n'est autre que la déviation transitoire

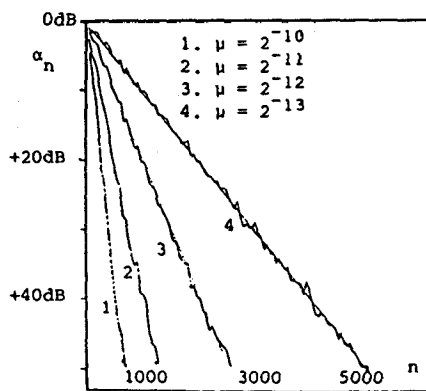


Fig. 2.3. — Atténuation de l'écho pour annuleur d'écho non bruité.

$V^f(n)$. On voit bien sur la figure que l'atténuation augmente linéairement avec les itérations n , et que la pente est effectivement proportionnelle à μ . Au bout d'un certain temps le transitoire $V^f(n)$ sera si amorti qu'il sera totalement négligeable. Alors demeurera seule l'erreur de fluctuation d'après (2.2.1).

II.2.4. Erreur de fluctuation

La fluctuation de l'algorithme est entièrement due au bruit de sortie $b(n)$ selon la formule (f) qui se développe en (2.2.2). Le calcul de la puissance de $V^f(n)$ se fait simplement grâce à l'hypothèse (I) de bruit blanc. On a

$$(2.2.13) \quad E(|V^f(n)|^2) = \mu^2 E_{\min} \sum_{j=1}^n E(|U_{j,n} X(j)|^2).$$

La stationnarité de la suite $X(n)$ entraîne aussitôt

$$(2.2.14) \quad E(|V^f(n)|^2) = \mu^2 E_{\min} \sum_{j=0}^{n-1} E(|U_{0,j} X(0)|^2).$$

Dans le théorème, on a prouvé en (2.2.6) la convergence exponentielle de tous les moments de $U_{0,j}$. Par une argumentation identique, on prouve que les moments $E(|U_{0,j} X(0)|^2)$ sont majorés par une suite géométrique du type (2.2.6). En sommant (2.2.14) il vient donc, avec des constantes C_1 et δ_1 ne dépendant que de la statistique des entrées $X(n)$ (et non de μ)

$$(2.2.15) \quad E(|V^f(n)|^2) \leq \mu E_{\min} C_1 / \delta_1.$$

La puissance des fluctuations est donc la somme d'une série positive majorée. Ainsi $E(|V^f(n)|^2)$ admet une limite $F(\mu)$ et l'on a

$$(2.2.16) \quad E(|V^f(n)|^2) \xrightarrow{n \rightarrow \infty} F(\mu) \leq \mu E_{\min} \eta$$

où η dépend seulement de la statistique des $X(n)$. C'est là un résultat fondamental. Le filtrage adaptatif fluctue proportionnellement au pas μ et aussi proportionnellement au bruit de sortie E_{\min} du filtre optimal. Ce résultat est bien connu des expérimentateurs. D'habitude on le trouve établi dans la littérature sous l'hypothèse que les $X(n)$ successifs sont indépendants [6, 7] bien que l'hypothèse soit inadmissible, puisque $X(n)$ et $X(n-1)$ ont $(N-1)$ composantes en commun. On voit que cette hypothèse n'est pas du tout nécessaire pour établir le résultat.

Erreur quadratique moyenne résiduelle (EQMR)

Pour caractériser la convergence de $V^f(n)$, on peut aussi utiliser la puissance de l'erreur résiduelle $\varepsilon^f(n) = -X^T(n) V^f(n-1)$. Grâce à l'hypothèse (I) cette erreur est indépendante du bruit $b(n)$ dans l'erreur totale (2.1.26) de sorte que

$$(2.2.17) \quad E(\varepsilon(n)^2) = E_{\min} + E((\varepsilon_n^f)^2).$$

Par une analyse identique à la précédente on montre que l'EQMR vérifie

$$(2.2.18) \quad E(\varepsilon_n^{f^2}) \xrightarrow{n \rightarrow \infty} f(\mu) \leq \mu E_{\min} \eta'$$

où η' ne dépend que de la statistique des $X(n)$.

Les deux résultats de base (2.2.16) et (2.2.18) mettent bien en évidence le fait que l'erreur de fluctuation du filtre adaptatif à pas constant ne décroît pas vers zéro. Il n'y a donc pas convergence au sens strict. Pourtant ces résultats sont satisfaisants pour l'expérimentateur qui choisira le pas μ assez petit pour que les erreurs résiduelles soient imperceptibles. Par exemple il choisira μ pour que la déviation quadratique soit inférieure à la précision numérique de son organe de calcul. Ou bien il choisira μ pour que l'EQMR soit négligeable devant le bruit additif, c'est-à-dire $\mu \eta' \ll 1$. La mise en œuvre de l'algorithme (FG) peut alors être qualifiée de « quasi » convergente en moyenne quadratique.

Pour choisir la valeur numérique de μ il est souhaitable d'obtenir la valeur exacte de $f(\mu)$. Pour cela, nous écrivons la récursion (f) sous la forme

$$(2.2.19) \quad \varepsilon_n^f = -X(n)^T V^f(n-1);$$

$$(2.2.20) \quad V^f(n) = V^f(n-1) + \mu (b(n) - \varepsilon_n^f) X(n).$$

Tenant compte du fait que $b(n)$ est un bruit blanc il vient

$$(2.2.21) \quad E(|V^f(n)|^2) = E(|V^f(n-1)|^2) - \mu E(\varepsilon_n^{f^2} (2 - \mu |X(n)|^2)) + \mu^2 E_{\min} E(|X(n)|^2).$$

Nous avons prouvé l'existence d'une valeur limite pour $E(|V^f(n)|^2)$. Le passage à la limite sur n est donc autorisé dans (2.2.21) pourvu que μ ne soit pas trop grand. On trouve, à l'aide de la puissance P_x de chaque composante $x(n-j)$

$$(2.2.22) \quad E(\varepsilon_n^{f^2} (2 - \mu |X(n)|^2)) \xrightarrow{n \rightarrow \infty} \mu E_{\min} NP_x$$

Le résultat (2.2.22) complète utilement la borne supérieure (2.2.18). Il montre que l'erreur de fluctuation est effectivement, pour μ petit, un infiniment petit du même ordre que μ . C'est un résultat que l'expérimentateur vérifie immédiatement : l'EQMR augmente de 3 dB à chaque fois que le gain est multiplié par 2. De (2.2.22) on tire la valeur numérique $f(\mu)$ de l'EQMR dans deux cas particuliers fondamentaux pour les applications.

1^{er} cas : $|X(n)| = Cte$

La puissance instantanée est constante à l'entrée du filtre, de sorte que $|X(n)|^2 = NP_x$. Alors d'après (2.2.22)

$$(2.2.23) \quad f(\mu) = \frac{\mu E_{\min} NP_x}{2 - \mu NP_x}; \quad \mu < \frac{2}{NP_x}.$$

Ce n'est pas là un cas académique. Ainsi dans l'annulation d'écho où les $x(n)$ sont des données aléatoires, ceci décrit la modulation à 4 phases où $x(n) = \pm 1 \pm j$. D'une manière générale, pourvu que le filtre soit assez long, la quantité $|X(n)|^2 = |x(n)|^2 + |x(n-1)|^2 + \dots + |x(n-N+1)|^2$ réalise une moyenne temporelle qui s'approche bien de la moyenne d'ensemble NP_x puisque les $x(n)$ sont stationnaires.

2^e cas : petit gain

Avec une forte probabilité $\mu |X(n)|^2$ est négligeable devant 2, ou encore en moyenne

$$(2.2.24) \quad \mu NP_x \ll 2.$$

Le choix d'un gain vérifiant (2.2.24) est tout à fait raisonnable. En effet l'EQMR vaut alors, d'après (2.2.22)

$$(2.2.25) \quad f(\mu) = E_{\min} \frac{\mu NP_x}{2}$$

ce qui est négligeable devant l'EQM minimale E_{\min} : le filtre adaptatif $H(n)$, s'il n'est pas convergent au sens strict, a une erreur de sortie quasiment identique à celle du filtre optimal \tilde{H} . Nous verrons au paragraphe II.3 que dans le cas plus général où \tilde{H} subit des variations au cours du temps, le choix d'un gain μ satisfaisant la condition (2.2.24) reste bien adapté pourvu que \tilde{H} varie « lentement », en un sens qui sera défini quantitativement.

C'est l'expression (2.2.25) qui est la plus connue. Elle met bien en évidence les quatre facteurs E_{\min} , μ , N et P_x qui influent sur l'EQM résiduelle. La diminution de l'EQMR grâce au choix de μ vient d'être discutée, mais il faut ajouter que diminuer μ pour réduire $f(\mu)$ diminue aussi la vitesse de convergence de l'erreur transitoire $V^t(n)$ comme il a été vu à la fin du paragraphe II.2.3 (formule 2.2.12).

Dans les applications, on mesure globalement la convergence du filtre en considérant la décroissance au cours du temps de l'EQM. A cause de (2.2.1) il faut alors inclure dans $e(n)$ une erreur transitoire $\varepsilon^t(n)$ selon

$$(2.2.26) \quad \begin{cases} e(n) = b(n) - \varepsilon^f(n) - \varepsilon^t(n); \\ \varepsilon^t(n) = X(n)^T V^t(n-1). \end{cases}$$

Tout comme $V^t(n)$ dans (2.2.6), on montre que $\varepsilon^t(n)$ a tous ses moments exponentiellement décroissants. En particulier

$$(2.2.27) \quad E(\varepsilon^t(n)^2) \leq C'(1 - \delta' \mu)^n.$$

Au début de la convergence c'est le terme $\varepsilon^t(n)$ qui est prépondérant dans l'erreur $e(n)$ car le bruit $b(n)$ est généralement faible vis-à-vis de l'effet de la déviation initiale $V(0)$ du filtre. Ainsi l'EQM diminue-t-elle exponentiellement, ce qui donne un segment de droite pour son logarithme. Sur la figure 2.4 est portée l'EQM résiduelle $R_n = E[(\varepsilon^t(n) + \varepsilon^f(n))^2]$ normalisée par le bruit E_{\min} , dans le cas où $E(|X(n)|^2) = 105$. En fin de convergence $\varepsilon^t(n) \equiv 0$ et l'on obtient le niveau $f(\mu) = E(\varepsilon^f(n)^2)$ des fluctuations. Sur ces courbes, on retrouve le compromis fondamental entre vitesse et finesse de convergence.

II.3. MAXIMISATION DE LA VITESSE DE CONVERGENCE

La constante de temps τ du FA est le nombre n d'itérations requis pour que le transitoire disparaisse. Sur les courbes de la figure 2.4, ceci correspond à l'endroit du coude de saturation vers la valeur rési-

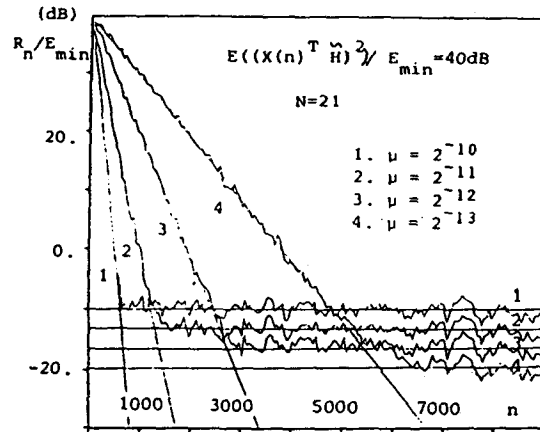


Fig. 2.4. — Convergence de l'EQM, compromis vitesse/finesse de convergence.

duelle liée aux fluctuations. Cette constante a déjà été calculée pour l'algorithme du gradient déterministe. Pour l'algorithme stochastique (FG), la théorie précédente, établie sous des hypothèses réalistes [permettant une corrélation entre $X(n)$ et $V(n-1)$] a fourni une borne supérieure du transitoire, par exemple par le calcul de l'atténuation α_n en (2.2.12). Mais elle ne donne pas les valeurs numériques des constantes de temps K_1 et K_2 . Pour cela, dans le cadre strict de ce paragraphe II.3, nous adopterons l'hypothèse que les vecteurs $V(n-1)$ et $X(n)$ sont indépendants. Certes l'hypothèse est en défaut dans le cas du filtrage, mais divers auteurs antérieurs [2, 15] l'ont légitimée en montrant que pour μ très petit les résultats asymptotiques sont corrects. On peut alors faire un calcul exact de l'EQMR et trouver le pas d'incrément μ_0 qui minimise la durée du transitoire. Pour cela nous traduisons la récurrence (2.1.28) sur le vecteur de déviation transformée $W_n = PV(n)$, où P est la matrice des vecteurs propres de R déjà utilisée pour le gradient déterministe [cf. (2.1.15)]. Après des calculs très simples et sous l'hypothèse $|X(n)|^2 = NP_x$, il vient

$$(2.2.28) \quad |V(n)|^2 = |W_n|^2 = |V(n-1)|^2 - \mu(2 - \mu NP_x) |V^t(n-1) X(n)|^2 + 2\mu b(n) V^t(n-1) (I - \mu X(n) X^T(n)) X(n) + \mu^2 NP_x b^2(n).$$

Grâce à l'indépendance entre $V(n-1)$ et $X(n)$ on trouve donc en moyennant

$$(2.2.29) \quad E(|W_n|^2) = E(|W_{n-1}|^2) - \mu(2 - \mu NP_x) E(W_{n-1}^T \Lambda W_{n-1}) + \mu^2 E_{\min} NP_x$$

qui implique que la suite $E(|V(n)|^2)$ satisfait la double inégalité

$$(2.2.30) \quad (1 - \mu(2 - \mu NP_x) \lambda_{\max}) E(|V(n-1)|^2) + \mu^2 E_{\min} NP_x \leq E(|V(n)|^2) \leq (1 - \mu(2 - \mu NP_x) \lambda_{\min}) E(|V(n-1)|^2) + \mu^2 E_{\min} NP_x$$

sous la condition de stabilité pour le pas d'adaptation

$$(2.2.31) \quad \mu < 2/NP_x.$$

Naturellement l'erreur transitoire $E(|V^t(n)|^2)$ est encadrée par les deux suites u_n et v_n ainsi définies

$$(2.2.32) \quad \begin{cases} u_n = (1 - 2\mu\lambda_{\max} + \mu^2 NP_x \lambda_{\max}) u_{n-1}, \\ v_n = (1 - 2\mu\lambda_{\min} + \mu^2 NP_x \lambda_{\min}) v_{n-1}. \end{cases}$$

Or ces deux suites ont une vitesse de convergence maximale lorsque le pas d'adaptation vaut

$$(2.2.33) \quad \boxed{\mu_0 = 1/NP_x}$$

qui est le pas optimal en ce qui concerne la durée du transitoire. La constante de temps de l'algorithme peut être définie comme la valeur τ pour laquelle

$$(2.2.34) \quad E(|V^t(\tau)|^2) = \lim_n E(|V(n)|^2).$$

En d'autres termes le transitoire est égal aux fluctuations résiduelles. Les relations (2.2.30) et (2.2.32) montrent que, malheureusement, la constante de temps dépend de la direction, comme pour l'algorithme déterministe. Pour la direction propre U_i de la matrice R , cette constante vaut, d'après ce qui précède et en adoptant la valeur μ_0 optimale

$$(2.2.35) \quad \tau_i = \frac{NP_x}{\lambda_i} \cdot \log \left(\frac{\lambda_i |\tilde{H}|^2}{E_{\min}} \right).$$

Il n'est pas étonnant de constater que la constante de temps augmente avec la longueur du filtre. Si la matrice R est très dispersée, elle peut être très grande, et elle dépend toujours de ses valeurs propres.

Pour obtenir un résultat facile à utiliser, supposons que les $x(n)$ d'entrée forment un bruit blanc de sorte que $\Lambda = P_x I$. Alors la récurrence (2.2.29) s'intègre immédiatement, et en choisissant le pas optimal $\mu_0 = 1/NP_x$, on trouve facilement l'erreur résiduelle totale (transitoire + fluctuation)

$$(2.2.36) \quad E(\varepsilon^2(n)) = \left(1 - \frac{1}{N}\right)^n P_y + E_{\min} \left(1 - \left(1 - \frac{1}{N}\right)^n\right),$$

où P_y est la puissance du signal à estimer $\tilde{y}(n)$ (voir fig. 2.2). L'hypothèse de bruit blanc $x(n)$ à l'entrée n'est pas irréaliste. Elle couvre par exemple l'annulation adaptative d'un écho de données.

L'analyse précédente ne doit pas faire oublier qu'en maximisant la vitesse de convergence on a considérablement perdu sur les fluctuations résiduelles. Dans (2.2.36), celles-ci atteignent le niveau du bruit de sortie E_{\min} , alors qu'elles peuvent être rendues arbitrairement faibles en diminuant μ bien en dessous de μ_0 .

Nous considérons maintenant la poursuite d'un filtre variable.

II.4. POURSUITE D'UN FILTRE VARIABLE

S'il n'assure pas une convergence parfaite, le filtre adaptatif a en retour la capacité fondamentale de poursuivre un filtre optimal \tilde{H} qui viendrait à varier au cours du temps. Par exemple dans la figure 2.2, le

modèle linéaire à identifier (voie du haut) devient un vecteur $\tilde{H}(n)$ variable avec n . Le bruit de sortie défini comme en (2.1.23) par

$$(2.4.1) \quad b(n) = y(n) - X(n)^T \tilde{H}(n-1),$$

différence entre la référence $y(n)$ et la sortie du filtre optimal $\tilde{H}(n-1)$, reste naturellement décorrélié – ou indépendant selon l'hypothèse (I) – par rapport aux entrées $X(n)$, selon (2.1.24). En conservant le même algorithme (FG) d'adaptation, nous cherchons comment se modifient les résultats du paragraphe II.2 à propos de l'EQM d'estimation.

Pour simplifier, nous supposons que les variations proviennent seulement du modèle $\tilde{H}(n)$, tandis que les entrées $X(n)$ et les bruits de sortie $b(n)$ demeurent stationnaires.

II.4.1. Découplage fluctuations/retard

Soient

$$(2.4.2) \quad \begin{cases} V(n) \stackrel{\Delta}{=} H(n) - \tilde{H}(n); \\ d_n \stackrel{\Delta}{=} \tilde{H}(n) - \tilde{H}(n-1) \end{cases}$$

la déviation du filtre par rapport à l'optimalité et l'incrément caractérisant les variations temporelles du filtre optimal. Avec ces notations l'erreur $e(n)$ conserve la forme (2.1.26), (2.1.27) et l'algorithme (FG) s'écrit

$$(2.4.3) \quad V(n) = (I - \mu X(n) X(n)^T) V(n-1) + \mu b(n) X(n) - d_n$$

équation qui généralise (2.1.28). Adoptant la méthodologie du II.2.2, on découpe donc $V(n)$ en 3 déviations.

Le transitoire $V^t(n)$ suit la récurrence (t). Sans changement par rapport au cas stationnaire, il s'amortit exponentiellement et nous l'omettons.

La fluctuation $V^f(n)$ suit la récurrence (f).

Enfin le retard $V^r(n)$ est défini par

$$(2.4.4) \quad \begin{aligned} (r) \quad V^r(n) &= (I - \mu X(n) X(n)^T) V^r(n-1) - d_n; \\ V^r(0) &= 0, \\ V^r(n) &= - \sum_{j=1}^n U_{j,n} d_j. \end{aligned}$$

En comparant avec le développement (2.2.2) de $V^f(n)$ on voit immédiatement que $V^f(n)$ et $V^r(n)$ sont décorrélés. Il en sera de même pour les erreurs résiduelles ε_n^f et

$$(2.4.5) \quad \varepsilon_n^r \stackrel{\Delta}{=} X(n)^T V^r(n-1)$$

pourvu que les variations $\{d_n\}$ du filtre optimal $H(n)$ soient indépendantes de la suite $\{b(j)\}$ des bruits de sortie. Il s'agit là d'une condition faible puisqu'en pratique les deux phénomènes n'ont pas la même origine. Pour l'erreur totale d'estimation

$$(2.4.6) \quad e(n) = b(n) - \varepsilon_n^f - \varepsilon_n^r$$

ceci se traduit par la formule de découplage

$$(2.4.7) \quad E(e(n)^2) = E_{\min} + E(\epsilon_n^2) + E(\epsilon_n^r)$$

le deuxième terme étant déjà calculé en (2.2.25). En effet l'incrément d_n n'apparaît pas dans l'équation (f) qui gouverne la fluctuation et tous les résultats du II.2.4 s'appliquent. Il ne reste plus qu'à calculer les moments des erreurs de retard $V^r(n)$ et ϵ_n^r .

II.4.2. Calcul de l'erreur de retard

(a) Calcul d'une borne supérieure

Faisons l'hypothèse plausible que les variations de $\tilde{H}(n)$ sont bornées en module par D. Alors d'après (2.4.4) et (2.2.9) on a

$$(2.4.8) \quad E(|V^r(n)|) \leq D \sum_{j=1}^n E(\|U_{j,n}\|) \leq \frac{DC_1}{\delta_1 \mu}$$

De la même manière on trouve des bornes en D^2/μ^2 pour les moments d'ordre 2 de $V^r(n)$ et ϵ_n^r

$$(2.4.9) \quad \begin{cases} E(|V^r(n)|^2) \leq \beta D^2/\mu^2; \\ E(\epsilon_n^r) \leq \beta' D^2/\mu^2 \end{cases}$$

Ainsi les erreurs dues aux non stationnarités sont en amplitude proportionnelles aux variations D du filtre à poursuivre et inversement proportionnelles au pas d'adaptation μ . Ce résultat est bien naturel : plus μ est grand, plus le filtre adaptatif est capable de poursuite. Des résultats du même type ont paru dans la littérature [2, 8], mais sous l'hypothèse que les $X(n)$ sont indépendants, ce qui exclut le cas du filtrage. En combinant ces résultats avec les erreurs de fluctuation il vient

$$(2.4.10) \quad E(|V(n)|^2) \leq E_{\min} \eta \mu + \beta D^2/\mu^2,$$

Erreur résiduelle :

$$(2.4.11) \quad E(\epsilon_n^2) \leq E_{\min} \eta' \mu + \beta' D^2/\mu^2.$$

Les erreurs résiduelles comportent deux termes dont les variations vis-à-vis de μ sont en sens inverses : la fluctuation croît avec μ et le retard décroît avec μ . Ainsi le pas μ contrôle le compromis entre ces deux termes, comme illustré sur la figure 2.5. Néanmoins pour déterminer la valeur optimale μ_{opt} qui minimise l'EQMR (par exemple), il faut posséder des évaluations meilleurs que de simples bornes supérieures. Pour ce faire il nous faut un modèle pour les incréments $\{d_n\}$. Du point de vue des applications pratiques il y a deux modèles complémentaires selon que les

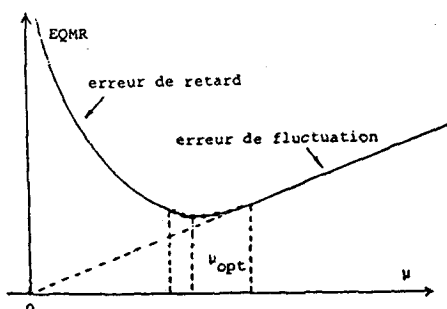


Fig. 2.5. — Le compromis entre retard et fluctuation.

variations de $\tilde{H}(n)$ sont aléatoires centrées ou bien déterministes. Le premier cas est traité plus bas. Dans le cas déterministe, le calcul exact des erreurs de retard n'est pas possible. En effet pour n grand ces quantités oscillent en permanence entre deux limites, et l'erreur ϵ_n n'a pas de distribution de probabilité limite. On s'en tient donc à des bornes supérieures telles (2.4.9), dans lesquelles on peut évaluer les constantes β et β' car elles ne dépendent que de la statistique des entrées $X(n)$. Ainsi en supposant que les composantes $x(n-j)$ de $X(n)$ sont mutuellement indépendantes, il est prouvé dans [10] que pour $\mu \ll 2/NP_x$, on a

$$(2.4.12) \quad E(\epsilon_n^2) \leq E_{\min} \frac{\mu NP_x}{2} + 3 D^2/\mu^2 P_x$$

(b) Cas du filtre optimal aléatoire

Un calcul exact de l'erreur retard est possible lorsqu'on suppose que la suite des incréments $\{d(n)\}$ est stationnaire, centrée et indépendante. Ce cas modélise bien différentes applications, telle la gigue sur la phase d'échantillonnage [10] qui peut affecter la réponse impulsionnelle $\tilde{H} = (\tilde{h}(0), \tilde{h}(T), \dots, \tilde{h}((N-1)T))^T$ du modèle optimal, telle aussi les variations temporelles du modèle liées à la gigue de phase de la fréquence porteuse dans un système de transmission [11]. Soit d^2 la puissance de $|d_n|$. Il découle de la récurrence (r) donnant le retard que

$$(2.4.13) \quad E(|V^r(n)|^2) = E(|V^r(n-1)|^2) - \mu E(\epsilon_n^r(2 - \mu |X(n)|^2)) + d^2.$$

On peut mener les calculs comme pour l'erreur de fluctuation en II.2.4. Le passage à la limite sur n est licite. Ainsi pour μ très petit on trouve respectivement pour l'erreur retard et l'erreur résiduelle totale

$$(2.4.14) \quad \lim E(\epsilon_n^r) = d^2/2\mu,$$

$$(2.4.15) \quad \lim E(\epsilon_n^2) = E_{\min} \frac{NP_x \mu}{2} + \frac{d^2}{2\mu}$$

II.4.3. Optimisation du filtre adaptatif

Il est particulièrement important pour les applications pratiques de trouver la valeur numérique μ_{opt} du pas d'adaptation qui minimise l'EQMR par un compromis entre les erreurs de fluctuation et de retard. Introduisons la quantité

$$(2.4.16) \quad y = d(NP_x/E_{\min})^{1/2}$$

que nous nommons « degré de non-stationnarité », et utilisons des paramètres normalisés de la manière usuelle :

$$(2.4.17) \quad v = NP_x \mu; \quad \eta = E(\epsilon_n^2)/E_{\min}.$$

Les résultats (2.4.12) et (2.4.15) donnés respectivement dans les cas $\tilde{H}(n)$ déterministe et aléatoire se traduisent alors par

$$(2.4.18 a) \quad \eta_a(v) = \frac{v}{2} + \frac{y^2}{2v}; \quad v_{\text{opt}}^a = y,$$

$$(2.4.18 d) \quad \eta_d(v) \leq \frac{v}{2} + \frac{3Ny^2}{v^2}; \quad v_{opt}^d = (12Ny^2)^{1/3}.$$

Ainsi l'erreur résiduelle normalisée ne dépend que de v et de y et le pas optimal ne dépend que de y . La notion de degré de non-stationnarité est donc très importante. On voit qu'elle n'est pas liée seulement à la vitesse (D ou d) des variations de $\tilde{H}(n)$. Ces dernières doivent être mesurées en proportion du bruit de sortie E_{min} [cf. (2.4.16)]. Le modèle sera dit à variations lentes lorsque ses variations – répercutées sur la sortie par l'intermédiaire du paramètre $NP_x d^2$ – sont noyées dans le bruit de sortie E_{min} , de sorte que y est très petit. Cette définition fait référence à l'ordre N du filtre : c'est là un fait connu, la poursuite de variations est d'autant moins aisée que le filtre est plus long.

L'hypothèse simplificatrice utilisée en (2.2.24) selon laquelle $v \ll 2$, peut être traduite de manière intrinsèque (indépendante du pas d'incrémenté choisi par l'utilisateur). D'après (2.4.18), l'analyse précédente est valable pourvu que

cas aléatoire :

$$(2.4.19 a) \quad y \ll 2,$$

cas déterministe :

$$(2.4.19 d) \quad y \ll (2/3N)^{1/2}.$$

La deuxième condition est toujours plus stricte que la première : un modèle à variation déterministe pose un problème de poursuite plus sévère. Sans connaissance *a priori* sur la nature des variations de $\tilde{H}(n)$, on pourra adopter la condition (2.4.19 d) comme définition d'un filtre $\tilde{H}(n)$ à variations lentes. Si elle est satisfaite, la condition $v \ll 2$ demeurera vérifiée pourvu qu'on fasse fonctionner le filtre adaptatif dans la zone d'optimalité proche de v_{opt} (voir fig. 2.5).

En conclusion, pour une mise en œuvre satisfaisante, la poursuite adaptative par l'algorithme (FG) nécessite principalement de connaître la valeur numérique du degré de non-stationnarité y . Il faut en outre insister sur la différence entre les deux notions d'optimalité μ_0 et μ_{opt} pour le pas d'adaptation

$$(2.4.20) \quad \mu_0 = \frac{1}{NP_x}; \quad \mu_{opt} = \frac{d}{\sqrt{E_{min}}} \cdot \frac{1}{\sqrt{NP_x}}.$$

La première valeur donne le plus court transitoire pour un modèle fixe; la deuxième valeur optimise la poursuite. Contrairement à une idée fort répandue, cela n'est pas du tout équivalent.

III. Approche rapide : les moindres carrés

Nous avons mentionné dans le II la faiblesse du gradient en vitesse de convergence. Comme il a été dit en introduction, le transitoire sera amélioré en prenant comme critère une moyenne temporelle d'horizon limité pour l'erreur quadratique. C'est ainsi que

l'algorithme des moindres carrés vise à déterminer récursivement à l'indice n le vecteur des coefficients $H(n)$ qui minimise l'erreur quadratique accumulée depuis l'indice de départ selon le critère (1.3.2).

III.1. L'ALGORITHME DES MOINDRES CARRÉS RAPIDES (MCR)

Pour donner à l'algorithme une capacité d'oubli donc de poursuite, on généralise le critère en adoptant la fonction de coût suivante

$$(3.0) \quad \mathcal{J}(n) = \sum_{p=1}^n w^{n-p} \left[y(p) - \sum_{i=0}^{N-1} h_i(n) x(p-i) \right]^2$$

où w est un facteur de pondération ($0 < w \leq 1$), dont on verra qu'il joue le même rôle que le pas d'adaptation du gradient, selon l'équivalence

$$(3.1) \quad \mu \propto (1-w).$$

Les signaux $x(n)$ et $y(n)$ sont supposés nuls pour $n \leq 0$.

L'annulation des dérivées de $\mathcal{J}(n)$ par rapport aux coefficients $h_i(n)$ conduit à l'équation suivante

$$(3.2) \quad R_N(n) \cdot H(n) = r(n)$$

avec les moyennes temporelles pondérées autocorrélation :

$$(3.3) \quad R_N(n) = \sum_{p=1}^n w^{n-p} X(p) X(p)^T,$$

intercorrélation :

$$(3.4) \quad r(n) = \sum_{p=1}^n w^{n-p} y(p) X(p),$$

et à l'aide du vecteur des entrées $X(p)$ défini en (1.2.2). On désigne par gain d'adaptation le vecteur

$$(3.5) \quad G(n) = R_N^{-1}(n) X(n).$$

En utilisant les relations de récurrence sur la matrice d'autocorrélation $R_N(n)$ et sur le vecteur d'intercorrélation $r(n)$, on fait apparaître la relation de récurrence suivante sur les coefficients du filtre adaptatif

$$(MCR) \quad H(n+1) = H(n) + G(n+1) [y(n+1) - X^T(n+1) H(n)].$$

Comme dans l'algorithme du gradient, l'incrément d'adaptation est proportionnel à l'erreur d'estimation $e(n+1) = y(n+1) - X^T(n+1) H(n)$ à la sortie du filtre dans son état actuel.

Les algorithmes rapides permettent de calculer $G(n+1)$ à partir de $G(n)$ avec un nombre d'opérations proportionnel à N . Ils sont basés sur la prédiction linéaire [12]-[15].

L'erreur de prédiction linéaire avant *a priori* est définie par

$$(3.6) \quad e_a(n+1) = x(n+1) - A^T(n) X(n)$$

et l'erreur de prédiction arrière *a priori* par

$$(3.7) \quad e_b(n+1) = x(n+1-N) - B^T(n) X(n+1).$$

Les coefficients de prédiction minimisent une fonction coût semblable à $\mathcal{J}(n)$ et sont obtenus par dérivation, comme les coefficients $H(n)$ du FA. En particulier l'erreur (3.6) permet l'adaptation du prédicteur avant $A(n)$ selon

$$(3.8) \quad A(n+1) = A(n) + G(n) e_a(n+1).$$

Les relations équivalentes à l'équation (3.2) sont les suivantes

$$(3.9) \quad R_N(n-1) A(n) = r_N^a(n);$$

$$(3.10) \quad R_N(n) B(n) = r_N^b(n)$$

avec

$$(3.11) \quad r_N^a(n) = \sum_{p=1}^n w^{n-p} x(p) X(p-1);$$

$$r_N^b(n) = \sum_{p=1}^n w^{n-p} x(p-N) X(p).$$

Les énergies d'erreur de prédiction sont données par :

$$(3.12) \quad E_a(n) = \sum_{p=1}^n w^{n-p} x^2(p) - A^T(n) r_N^a(n);$$

$$(3.13) \quad E_b(n) = \sum_{p=1}^n w^{n-p} x^2(p-N) - B^T(n) r_N^b(n).$$

Il est commode de regrouper les équations (3.9) et (3.12) d'une part et (3.10) et (3.13) d'autre part. On désigne par équation de la prédiction linéaire avant la relation matricielle suivante

$$(3.14) \quad \begin{bmatrix} R_1(n) & r_N^a(n)^t \\ r_N^a(n) & R_N(n-1) \end{bmatrix} \times \begin{bmatrix} 1 \\ -A(n) \end{bmatrix} = \begin{bmatrix} E_a(n) \\ 0 \end{bmatrix}$$

et par équation de la prédiction linéaire arrière :

$$(3.15) \quad \begin{bmatrix} R_N(n) & r_N^b(n) \\ r_N^b(n)^t & R_1(n-N) \end{bmatrix} \times \begin{bmatrix} -B(n) \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ E_b(n) \end{bmatrix}.$$

La matrice qui apparaît dans les équations est une estimation de la matrice $R_{N+1}(n)$, l'autocorrélation d'ordre $N+1$ du signal d'entrée $x(n)$. Il est important de noter que les relations ci-dessus font apparaître deux décompositions différentes de cette matrice, qui sont à la base des algorithmes rapides.

Le calcul de $G(n+1)$ à partir de $G(n)$ se fait par l'intermédiaire du gain d'adaptation d'ordre $N+1$. En effet, à l'ordre $N+1$ la relation de définition du gain d'adaptation à $N+1$ éléments et au temps $n+1$ s'écrit

$$(3.16) \quad R_{N+1}(n+1) G_1(n+1) = X_1(n+1) \\ = [x(n+1), x(n), \dots, x(n+1-N)]^T.$$

Comme l'indique la relation (3.14), la matrice $R_{N+1}(n+1)$ contient la matrice $R_N(n)$. Dans ces conditions il vient

$$(3.17) \quad R_{N+1}(n+1) \begin{bmatrix} 0 \\ G(n) \end{bmatrix} = \begin{bmatrix} [r_N^a(n+1)]^T G(n) \\ X(n) \end{bmatrix}.$$

Or :

$$(3.18) \quad [r_N^a(n+1)]^T G(n) \\ = [r_N^a(n+1)]^T R_N^{-1}(n) X(n) = A^T(n+1) X(n)$$

et d'après la définition de l'erreur de prédiction avant *a posteriori* [c'est-à-dire après adaptation (3.8) du vecteur $A(n)$]

$$(3.19) \quad \varepsilon_a(n+1) = x(n+1) - A^T(n+1) X(n)$$

il vient :

$$(3.20) \quad R_{N+1}(n+1) \begin{bmatrix} 0 \\ G(n) \end{bmatrix} = \begin{bmatrix} x(n+1) - \varepsilon_a(n+1) \\ X(n) \end{bmatrix} \\ = X_1(n+1) - \begin{bmatrix} \varepsilon_a(n+1) \\ 0 \end{bmatrix}.$$

Ou encore d'après la définition du gain d'adaptation

$$(3.21) \quad R_{N+1}(n+1) \left[G_1(n+1) - \begin{bmatrix} 0 \\ G(n) \end{bmatrix} \right] = \begin{bmatrix} \varepsilon_a(n+1) \\ 0 \end{bmatrix}.$$

En rapprochant cette expression de la prédiction avant (3.14), on obtient :

$$(3.22) \quad G_1(n+1) = \begin{bmatrix} 0 \\ G(n) \end{bmatrix} + \frac{\varepsilon_a(n+1)}{E_a(n+1)} \begin{bmatrix} 1 \\ -A(n+1) \end{bmatrix}.$$

Ensuite, les N premiers éléments du vecteur $G_1(n+1)$ sont notés $M(n+1)$ et séparés du dernier noté $m(n+1)$. La définition (3.16) donne alors

$$R_N(n+1) M(n+1) + r_N^b(n+1) m(n+1) \\ = X(n+1) = R_N(n+1) G(n+1).$$

Avec la prédiction linéaire arrière, il vient

$$(3.23) \quad M(n+1) + m(n+1) B(n+1) = G(n+1).$$

D'après la relation générale de récurrence des moindres carrés (MCR), les coefficients de prédiction arrière sont mis à jour par

$$(3.24) \quad B(n+1) = B(n) + G(n+1) e_b(n+1).$$

Finalement on dispose de deux relations (3.23) et (3.24) pour calculer $B(n+1)$ et $G(n+1)$. En calculant $G(n+1)$ d'abord il vient

$$(3.25) \quad G(n+1) = \frac{1}{1 - m(n+1) e_b(n+1)} \\ \times [M(n+1) + m(n+1) B(n)].$$

L'énergie d'erreur de prédiction est également mise à jour par une relation de récurrence simple mais un peu

TABLEAU 3.1

Algorithmes des moindres carrés rapides (MCR)
Quantités disponibles à l'indice n :

Vecteur des coefficients du filtre adaptatif : $H(n)$.
Vecteur des coefficients de prédiction avant : $A(n)$.
Vecteur des coefficients de prédiction arrière : $B(n)$.
Vecteur de données à N éléments : $X(n)$.
Gain d'adaptation : $G(n)$.
Erreur quadratique cumulée de préd. avant : $E_a(n)$.

Nouvelles données à l'indice $(n+1)$
Donnée reçue : $x(n+1)$.
Signal de référence : $y(n+1)$.

Calcul du nouveau gain d'adaptation :

Erreur de prédiction avant :

$$e_a(n+1) = x(n+1) - A^T(n)X(n)$$

Mise à jour des coefficients de prédiction avant :

$$A(n+1) = A(n) + G(n)e_a(n+1)$$

Erreur de prédiction avant *a posteriori* :

$$e_a(n+1) = x(n+1) - A^T(n+1)X(n)$$

Erreur quadratique de prédiction avant cumulée :

$$E_a(n+1) = wE_a(n) + e_a(n+1)e_a(n+1)$$

Gain d'adaptation intermédiaire :

$$G_1(n+1) = \begin{pmatrix} 0 \\ G(n) \end{pmatrix} + \frac{\varepsilon_a(n+1)}{E_a(n+1)} \begin{pmatrix} 1 \\ -A(n+1) \end{pmatrix} = \begin{pmatrix} M(n+1) \\ m(n+1) \end{pmatrix}$$

Erreur de prédiction arrière *a priori* :

$$e_b(n+1) = x(n+1-N) - B^T(n)X(n+1)$$

Gain d'adaptation :

$$G(n+1) = \frac{1}{1 - m(n+1)e_b(n+1)} [M(n+1) + m(n+1)B(n)]$$

Coefficients de prédiction arrière :

$$B(n+1) = B(n) + G(n+1)e_b(n+1)$$

Opérations du filtre adaptatif :

Calcul de l'erreur en sortie :

$$e(n+1) = y(n+1) - H^T(n)X(n+1)$$

Mise à jour des coefficients :

$$H(n+1) = H(n) + G(n+1)e(n+1)$$

fastidieuse à démontrer voir [15], § 6.3. L'algorithme complet est synthétisé par le tableau 3.1. Il nécessite, pour la mise à jour du gain d'adaptation $8N+2$ multiplications et 2 divisions. Pour le filtrage il faut $2N$ multiplications. Le nombre de mémoires s'élève approximativement à $6N$ au total.

III.2. RÉGIME D'ÉTABLISSEMENT ET ERREUR RÉSIDUELLE

Puisque le vecteur d'entrée $X(n)$ est nul à l'origine, le vecteur des erreurs à l'indice n s'écrit

$$(3.26) \quad \begin{bmatrix} e(1) \\ \vdots \\ e(N) \\ \vdots \\ e(n) \end{bmatrix} = \begin{bmatrix} y(1) \\ \vdots \\ y(N) \\ \vdots \\ y(n) \end{bmatrix} - \begin{bmatrix} x(1) & 0 & \dots & 0 \\ \vdots & \vdots & & \vdots \\ x(N) & x(N-1) & & x(1) \\ \vdots & \vdots & & \vdots \\ x(n) & x(n-1) & \dots & x(n+1-N) \end{bmatrix} \times \begin{bmatrix} h_0(n) \\ \vdots \\ h_{N-1}(n) \end{bmatrix}$$

Dans ces conditions, les N premières valeurs de l'erreur peuvent être annulées et les valeurs exactes des coefficients du filtre sont obtenues. Pour chaque nouvelle entrée reçue $x(n)$ avec $n \leq N$ un nouveau coefficient est déterminé. En fait, l'optimisation avec critère des moindres carrés commence réellement quand $N+1$ valeurs du signal d'entrée ont été reçues.

L'étude du régime d'établissement est faite en considérant la modélisation adaptative représentée sur la figure 2.2. Le signal de référence est donné par l'équation (2.1.23) où \tilde{H} est le filtre optimal et où $b(n)$ est un bruit blanc de puissance E_{\min} , indépendant de la suite des entrées $\{x(j)\}$. Dans ce schéma, le bruit $b(n)$ est supposé être la seule variable aléatoire. Le vecteur des coefficients $H(n)$ s'écrit alors :

$$(3.27) \quad H(n) = R_N^{-1}(n) \left[\sum_{p=1}^n w^{n-p} \times (X(p)X^T(p)\tilde{H} + X(p)b(p)) \right]$$

soit :

$$H(n) = \tilde{H} + R_N^{-1}(n) \sum_{p=1}^n w^{n-p} X(p)b(p)$$

Avec l'écart à l'optimalité $V(n)$ défini en (2.1.16), la matrice de covariance de l'écart des coefficients est donnée par

$$(3.29) \quad E[V(n)V^T(n)] = E_{\min} R_N^{-1}(n) \times \left(\sum_{p=1}^n w^{2(n-p)} X(p)X^T(p) \right) R_N^{-1}(n)$$

En raison de la présence du terme $w^{2(n-p)}$, il faut alors distinguer deux cas. D'abord on suppose $w=1$. Il vient

$$(3.30) \quad E[V(n)V^T(n)] = E_{\min} R_N^{-1}(n)$$

Le signal d'erreur *a priori* à l'indice n s'écrit, on l'a vu

$$(3.31) \quad e(n) = y(n) - H^T(n-1)X(n) = b(n) - V^T(n-1)X(n)$$

et il a pour variance :

$$(3.32) \quad E[e^2(n)] = E_{\min} + X^T(n)E[V(n-1)V^T(n-1)]X(n)$$

soit

$$(3.33) \quad E[e^2(n)] = E_{\min} [1 + X^T(n)R_N^{-1}(n-1)X(n)]$$

Si le signal $x(n)$ est tel que la matrice $R_N(n-1)$ soit approximativement proportionnelle à la matrice unité I_N , on peut écrire approximativement

$$(3.34) \quad E_{\text{MCR}}(n) - E_{\min} = E_{\min} \cdot \frac{N}{n}$$

ou encore, en considérant que l'algorithme ne fonctionne suivant les moindres carrés que quand N don-

nées ont été reçues

$$(3.35) \quad E_{\text{MCR}}(n) \simeq E_{\text{min}} \left(1 + \frac{N}{n-N} \right); \quad n > N.$$

Avec l'algorithme (MCR) on voit donc que pour $n=2N$, l'erreur totale est proche du double de E_{min} ou, si l'on préfère, l'erreur résiduelle est voisine de E_{min} .

Le second cas à étudier correspond à la présence d'un facteur de pondération $w < 1$. Alors

$$(3.36) \quad E[e^2(n)] = E_{\text{min}} \left[1 + X^T(n) R_N^{-1}(n) \times \left(\sum_{p=1}^n w^{2(n-p)} X(p) X^T(p) \right) \times R_N^{-1}(n) X(n) \right].$$

Si le signal $x(n)$ est tel que l'on puisse écrire pour de grandes valeurs de n

$$(3.37) \quad R_N(n) = \sum_{p=1}^n w^{n-p} X(p) X^T(p) \simeq P_x \frac{1-w^n}{1-w} I_N$$

où P_x représente la puissance du signal d'entrée, il vient

$$(3.38) \quad E_{\text{MCR}}(n) \approx E_{\text{min}} \left[1 + N \frac{1-w}{1+w} \frac{1+w^n}{1-w^n} \right]$$

et à la limite

$$(3.39) \quad E_{\text{MCR}}(\infty) \simeq E_{\text{min}} \left[1 + N \frac{1-w}{1+w} \right].$$

Finalement, après convergence il subsiste une erreur résiduelle

$$(3.40) \quad E_{\text{MCR}} - E_{\text{min}} = E_{\text{min}} \cdot N \frac{1-w}{1+w}.$$

Cette équation est à rapprocher de (2.2.25) qui donne l'erreur résiduelle du gradient. En régime permanent, les algorithmes (FG) et (MCR) ont donc les mêmes fluctuations à condition de choisir μ tel que

$$(3.41) \quad \mu P_x = 1 - w$$

(lorsque w est proche de 1). Pas plus que le gradient les moindres carrés ne présentent une vraie convergence. Là encore il s'agit d'une « quasi-convergence en moyenne quadratique » qui est détaillée en [16] et [17].

Il est intéressant aussi de comparer les régimes d'établissement des algorithmes MCR et gradient. En prenant pour ce dernier le pas d'adaptation (2.2.33) correspondant à la rapidité maximale, c'est-à-dire $\mu_0 = 1/NP_x$, nous avons

$$(3.42) \quad E_{(\text{FG})}(n) - E_{\text{min}} = \left(1 - \frac{1}{N} \right)^n (P_y - E_{\text{min}}) + E_{\text{min}} \\ \simeq \left(1 - \frac{1}{N} \right)^n P_y + E_{\text{min}}.$$

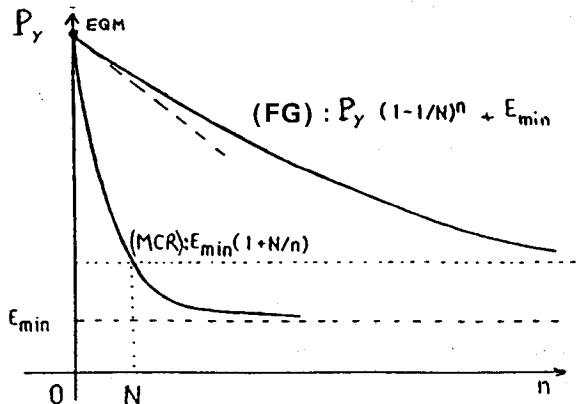


Fig. 3.2. — Erreur quadratique moyenne comparée pour les algorithmes (FG) et (MCR).

La comparaison des relations (3.34) et (3.42) fait apparaître l'avantage de l'algorithme (MCR) lorsque $w=1$, avantage illustré par la figure 3.2, et d'autant plus marqué que l'erreur minimale E_{min} est plus faible. Cet avantage en vitesse se paie par l'impossibilité de suivre des variations temporelles. En effet lorsque $w=1$ le gain $G(n)$ tend vers zéro en $1/n$, d'après (3.5), ce qui finit par stopper l'évolution du FA. Au contraire la valeur non nulle de μ_0 assure à l'algorithme (FG) une capacité de poursuite.

III.3. CONSTANTE DE TEMPS

C'est précisément l'introduction du facteur de pondération w qui permet à l'algorithme (MCR) de suivre l'évolution des signaux non stationnaires. La valeur de ce facteur est déterminée en fonction des caractéristiques de non stationnarité des signaux à traiter [18].

L'équation de mise à jour des coefficients pour l'algorithme (MCR), s'écrit, pour n grand et en approximant la matrice $R_N(n)$ par son espérance

$$(3.43) \quad H(n+1) = H(n) + (1-w) R^{-1} X(n+1) e(n+1).$$

[Ici on a supposé la suite $x(n)$ stationnaire.] En fait le terme $(1-w)$ correspond à la fenêtre d'observation du signal ou temps d'intégration τ_0 (cf. § I.2).

La relation (3.43) est à rapprocher de celle de l'algorithme du gradient (G), soit

$$(3.44) \quad H(n+1) = H(n) + \mu X(n+1) e(n+1).$$

L'apport des moindres carrés est la rotation-homothétie ou similitude qu'opère la matrice carrée symétrique R^{-1} sur le gradient.

Le comportement de l'algorithme (MCR) en présence de signaux non stationnaires est relié à la constante de temps τ_{MCR} (calculée pour des signaux stationnaires). Pour déterminer cette valeur, on peut reprendre l'étude du régime d'établissement en présence de signaux stationnaires, et écrire :

$$(3.45) \quad V(n+1) = V(n) + (1-w) R^{-1} X(n+1) e(n+1).$$

Avec les hypothèses faites au paragraphe III.2 et compte tenu de la relation (2.1.23), il vient après

quelques calculs [15]

$$(3.46) \quad E[V(n+1)V^T(n+1)] \simeq w^2 E[V(n)V^T(n)] + (1-w)^2 E_{\min} R^{-1}.$$

On observe d'abord que le facteur d'homothétie est le scalaire w^2 et non une matrice. La valeur limite est

$$(3.47) \quad E[V(\infty)V^T(\infty)] = \frac{1-w}{1+w} E_{\min} R^{-1}.$$

La relation (3.47) est à rapprocher de la relation équivalente pour l'algorithme (G) [19]. La différence principale réside dans le fait que les valeurs propres de R n'interviennent pas dans la constante de temps

$$(3.48) \quad \tau_{MCR} = 1/(1-w)$$

de l'algorithme (MCR). Au contraire nous avons vu leur importance pour déterminer la constante de temps de (FG), voir par exemple (2.2.35). C'est là un des gros avantages de l'algorithme MCR).

III.4. MISE EN ŒUVRE DE L'ALGORITHME (MCR)

L'utilisation de l'algorithme (MCR) dans les filtres adaptatifs réels implique la détermination de plusieurs paramètres, en particulier w , et aussi un certain nombre de modifications des équations théoriques, afin de tenir compte de l'effet des arrondis dans les calculs en précision finie. Les algorithmes représentant le calcul exact des moindres carrés pour tout n , une plus grande rigueur est nécessaire dans la réalisation [20] que pour le gradient qui ne correspond aux moindres carrés qu'en moyenne.

La première précaution à prendre consiste à vérifier pour tout n que l'algorithme se trouve bien dans une situation de moindres carrés. Une méthode simple et efficace consiste à comparer les erreurs de prédiction *a priori* et *a posteriori*. A l'indice $n+1$ la valeur du coût de prédiction est

$$(3.49) \quad \mathcal{J}_{pr}(n+1) = \sum_{p=1}^{n+1} w^{n+1-p} [x(p) - A^T(n+1)X(p-1)]^2$$

ou encore :

$$(3.50) \quad \mathcal{J}_{pr}(n+1) = w \sum_{p=1}^n w^{n-p} [x(p) - A^T(n+1)X(p-1)]^2 + e_a^2(n+1).$$

Soit la fonction

$$(3.51) \quad \mathcal{J}'_{pr}(n+1) = w \sum_{p=1}^n w^{n-p} [x(p) - A^T(n)X(p-1)]^2 + e_a^2(n+1) = w \mathcal{J}_{pr}(n) + e_a^2(n+1).$$

On a, par définition des coefficients $A(n)$ et $A(n+1)$ les deux inégalités suivantes, dues à l'optimalité

$$\mathcal{J}'_{pr}(n+1) \geq \mathcal{J}_{pr}(n+1);$$

$$\mathcal{J}_{pr}(n) \leq \sum_{p=1}^n w^{n-p} [x(p) - A^T(n+1)X(p-1)]^2.$$

Il en résulte

$$(3.52) \quad \varepsilon_a^2(n+1) \leq e_a^2(n+1).$$

D'après les définitions, le rapport des erreurs de prédiction est donné par

$$(3.53) \quad \Phi(n) = \frac{\varepsilon_a(n+1)}{e_a(n+1)} = 1 - X^T(n)R_N^{-1}(n)X(n),$$

$$= \frac{1}{1 + X^T(n)R_N^{-1}(n-1)X(n)}.$$

La matrice $R_N^{-1}(n-1)$ étant définie positive, $\Phi(n)$ est nécessairement positif et finalement :

$$(3.54) \quad 0 \leq \frac{\varepsilon_a(n+1)}{e_a(n+1)} \leq 1.$$

Lorsque le rapport des erreurs de prédiction sort de l'intervalle $[0, 1]$ il faut stopper l'algorithme et, par exemple, remettre à zéro les variables avant de repartir. Le rapport est facile à contrôler, puisque les deux erreurs de prédiction sont calculées dans les équations du tableau 3.1.

En théorie, le facteur de pondération w peut être choisi arbitrairement petit, car la fonction bruit $\mathcal{J}(n)$ est quadratique et il existe une solution unique pour les coefficients. En pratique, avec les erreurs d'arrondi, il n'en est plus de même et il existe une limite inférieure pour w , qui peut être estimée grossièrement en approximant $R_N(n)$ par son espérance et en considérant le cas extrême où le vecteur $X(n)$ est colinéaire au vecteur propre associé à la plus petite valeur propre λ_{\min} de la matrice R . Les relations (3.53) et (3.54) donnent alors :

$$(3.55) \quad 0 \leq 1-w \leq \lambda_{\min}/|X(n)|^2.$$

Par exemple si $x(n)$ est un bruit blanc gaussien de puissance P_x , $|X(n)|^2$ est une variable approximativement gaussienne de moyenne NP_x et de variance $2NP_x^2$, ce qui, pour un facteur de crête égal à 4, donne

$$(3.56) \quad 0 \leq 1-w \leq 1/(N+4\sqrt{2N}).$$

Les algorithmes MCR étant récursifs, il faut initialiser les variables au temps 0. En l'absence d'informations particulières les coefficients de prédiction et le gain d'adaptation sont initialisés à zéro. L'énergie d'erreur de prédiction $E_a(n)$ étant utilisée comme diviseur doit être initialisée à une valeur positive E_0 , ce qui correspond à une situation de moindres carrés avec un signal $x(n)$ constitué d'une seule impulsion au temps $-N$ et tel que :

$$(3.57) \quad w^N x^2(-N) = E_0.$$

La matrice $R_N(0)$ correspondante est une matrice diagonale et, pour $n > 0$ la matrice réellement utilisée s'écrit :

$$(3.58) \quad R_N^*(n) = R_N(n) + w^n E_0 \text{diag}[w^{N-1}, w^{N-2}, \dots, 1].$$

La valeur propre minimale correspondante peut être estimée par :

$$(3.59) \quad \lambda_{\min}^*(n) \approx \frac{1-w^n}{1-w} \lambda_{\min} + w^n E_0 w^{N-1} = \frac{\lambda_{\min}}{1-w} + w^n \left(E_0 w^{N-1} - \frac{\lambda_{\min}}{1-w} \right).$$

La condition (3.55) est alors respectée si

$$(3.60) \quad E_0 w^{N-1} - \frac{\lambda_{\min}}{1-w} \geq 0$$

ce qui fixe un minimum pour E_0 , donné en utilisant la relation (3.55)

$$(3.61) \quad E_0 \geq \frac{N P_x}{w^{N-1}}.$$

En pratique, la constante E_0 ralentit l'algorithme et il faut prendre la valeur la plus faible possible. La relation (3.61) est en fait trop pessimiste avec les précisions de calcul courantes, et l'on prend souvent $E_0 = P_x$. Des essais avec une arithmétique à 16 bits ont montré que la stabilité initiale était obtenue avec $E_0 > P_x/100$ [21].

Toujours à cause de la division par $E_a(n+1)$, il est nécessaire en pratique de lui imposer une limite inférieure, en modifiant l'équation de mise à jour comme suit :

$$(3.62) \quad E_a(n+1) = w E_a(n) + e_a(n+1) E_a(n+1) + C.$$

Il vient alors pour n grand :

$$(3.63) \quad E_a(n) \geq \frac{C}{1-w}.$$

La constante C est appelée constante de stabilisation, car, en plus de la limitation du diviseur, elle introduit un facteur de rappel à zéro [15] sur les coefficients de prédiction arrière et peut ainsi compenser leur dérive, due à l'accumulation d'erreurs d'arrondi, comme indiqué ci-dessous.

Sur les performances de l'algorithme, l'effet global de la constante C est d'augmenter le facteur de pondération et d'augmenter ainsi la constante de temps τ_{MCR} . Sa valeur est généralement liée à E_0 et un choix raisonnable est le suivant

$$(3.64) \quad C \approx (1-w) E_0.$$

Pour l'utilisation avec des signaux non stationnaires, en particulier ceux qui comportent de longues séquences de valeurs nulles, il est important de prévoir un rappel à zéro des coefficients, qui peut être obtenu notamment en introduisant, à la fin de l'ensemble d'équations du tableau 3.1, les deux équations suivantes :

$$(3.65) \quad \begin{cases} A'(n+1) = (1-\gamma) A(n+1); \\ B'(n+1) = (1-\gamma) B(n+1). \end{cases}$$

Pour ne pas introduire de dégradation excessive des performances, γ doit être maintenu à des valeurs très inférieures à $1-w$. Naturellement la constante C et le facteur γ introduisent des perturbations dans l'algorithme (MCR). Avec certains signaux critiques, par exemples des sinusoides très proches en fréquence dans du bruit, elles peuvent même conduire l'algorithme à la divergence.

Une caractéristique importante de ces algorithmes est l'accumulation des erreurs d'arrondi, en raison de rebouclages dans les calculs particulièrement pour les coefficients de prédiction arrière. Pour y pallier une approche générale consiste à introduire une variable de contrôle représentative de l'accumulation d'erreurs d'arrondi et à la minimiser par un algorithme des moindres carrés utilisant en particulier le gain d'adaptation [22]. Avec l'algorithme que nous avons décrit, une variable de contrôle peut être mise en évidence en considérant l'équation

$$R_{N+1}(n+1) \begin{bmatrix} G(n+1) \\ 0 \end{bmatrix} = \begin{bmatrix} X(n+1) \\ [r_N^b(n+1)]^T G(n+1) \end{bmatrix} = X_1(n+1) - \begin{bmatrix} 0 \\ \varepsilon_b(n+1) \end{bmatrix}.$$

En utilisant l'équation de la prédiction linéaire arrière (3.15) la définition (3.16) et la décomposition du gain $G_1(n+1)$, il vient

$$(3.67) \quad \varepsilon_b(n+1) = m(n+1) E_b(n+1).$$

Cette quantité peut être calculée directement par

$$(3.68) \quad \varepsilon_b(n+1) = x(n+1-N) - B^T(n+1) X(n+1).$$

La différence entre ces deux modes de calcul donne la variable de contrôle

$$(3.69) \quad \zeta(n+1) = [x(n+1-N) - m(n+1) E_b(n+1)] - B^T(n+1) X(n+1).$$

Une procédure simple pour empêcher l'accumulation d'erreur d'arrondi consiste à corriger les coefficients de prédiction arrière comme suit [18]

$$(3.70) \quad B'(n+1) = B(n+1) + G(n+1) \zeta(n+1).$$

Des procédures plus complètes, avec des domaines d'application plus étendus, peuvent être élaborées, avec les autres algorithmes (MCR).

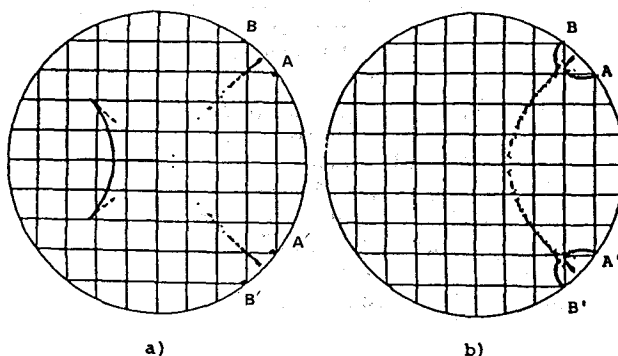


Fig. 3.3. - Détection de fréquences par prédiction adaptative.

IV. Conclusion

En manière de conclusion, nous illustrerons les possibilités de l'algorithme (MCR) et nous les comparerons aux performances du gradient (FG) sur l'exemple de la détection par prédiction adaptative de deux sinusoides proches en fréquence. Le modèle vrai correspondant est décrit par un AR d'ordre 4 de puissance $P_x=1$, bruité à -60 dB

$$(4.1) \quad x(n+1) = A^T X(n) + b(n)$$

et la position sur le cercle unité des 2 paires de pôles correspondant à ce modèle est donnée par les angles $\pm 36^\circ$, $\pm 54^\circ$ (points A, A', B, B' de la figure 3.3). Le prédicteur adaptatif $e(n) = x(n+1) - A(n)^T X(n)$ doit donc voir ses zéros tendre vers ces 4 points [24]. Sur la figure 3.3(a) est illustré le comportement du gradient pour un pas $\mu=0,01$. En 800 itérations, il est incapable de discerner les deux sinusoides. Il détecte seulement une fréquence moyenne plus du bruit. Par contre on voit sur la figure 3.3(b) que l'algorithme (MCR) détecte correctement les deux fréquences en moins de 1 000 coups. Le facteur de pondération était $w=0,99$ qui, d'après la formule (3.41) serait équivalent au point de vue fluctuations résiduelles à $\mu=0,01$ au cas où les entrées $x(n)$ seraient blanches. Naturellement il n'en est pas ainsi avec les deux sinusoides. La suite $x(n)$ est fort corrélée. Sa matrice de corrélation R est très dispersée. Comme on l'a vu dans le II, ceci ralentit beaucoup le gradient, sans toutefois changer ses performances en régime permanent. Les simulations confirment ce fait puisqu'en poursuivant les itérations l'algorithme (FG) atteint les points A, A', B, B' au bout de 100 000 coups. Il détecte donc les 2 sinusoides tout comme (MCR) mais en un temps 100 fois plus long. Avec son entrée extrêmement corrélée cet exemple a été choisi à dessein pour mettre en avant la supériorité en vitesse des moindres carrés. Naturellement ceci est au prix d'une complexité de mise en œuvre beaucoup plus grande.

Manuscrit reçu le 1^{er} juin 1987.

BIBLIOGRAPHIE

- [1] B. WIDROW et M. E. HOFF, Adaptive switching circuits, *IRE Wescon Conv. Record*, part 4, 1960, p. 96-104.
- [2] B. WIDROW, J. M. MCCOOL, M. G. LARIMORE et C. R. JOHNSON, Stationary and nonstationary learning characteristics of the LMS adaptive filter, *Proc. IEEE*, 8, 1976, p. 1151-1162.
- [3] O. MACCHI, Advances in adaptive filtering, in *Digital Communications*, G. PRATI et E. BIGLIERI éd., 1985, p. 41-57, North-Holland.
- [4] O. MACCHI et E. EWEDA, Bases théoriques pour l'égalisation adaptative en mode autodidacte, *Ann. des Télécomm.*, 41, mai 1986, p. 280-294.
- [5] O. MACCHI et E. EWEDA, Second order convergence analysis of stochastic adaptive linear filtering, *IEEE Trans. Automatic Control*, 28, n° 1, Janv. 1983, p. 76-85.
- [6] M. L. HONIG et D. G. MESSERSCHMITT, *Adaptive filters structure, Algorithms and Applications*, Kluwer Academic Publishers, 1984.
- [7] B. WIDROW et S. D. STEARNS, *Adaptive signal processing*, Prentice-Hall, 1985.
- [8] D. FARDEN, Tracking properties of adaptive signal processing algorithms, *IEEE Trans. ASSP*, 29, n° 3, 1981, p. 439-446.
- [9] A. BENVENISTE et G. RUGET, A measure of the tracking capability of recursive algorithms with constant gains, *IEEE Trans. Automatic Control*, AC-27, n° 3, 1982, p. 639-649.
- [10] S. MARCOS et O. MACCHI, Tracking capability of the LMS algorithm: application to an asynchronous echo canceller, *IEEE Trans. on ASSP-35*, n° 11, 1987, p. 1570-1578.
- [11] O. MACCHI, Optimisation du gain d'une boucle à verrouillage de phase en mode de poursuite, *Actes du 10^e Colloque GRETSI*, Nice, mai 1985, p. 545-550.
- [12] D. FALCONER et L. L. JUNG, Application of fast Kalman estimation to adaptive equalization, *IEEE Trans.*, COM-26, n° 10, Oct. 1978, p. 1439-1446.
- [13] G. CARAYANNIS, D. MANOLAKIS et N. KALOUPSIDIS, A fast sequential algorithm for least squares filtering and prediction, *IEEE Trans.*, ASSP-31, n° 6, Dec. 1983, p. 1394-1408.
- [14] J. CIOFFI et T. KAILATH, Fast recursive LS transversal filters for adaptive filtering, *IEEE Trans.*, ASSP-32, n° 2, April 1984, p. 304-337.
- [15] M. BELLANGER, *Adaptive digital filters and signal analysis*, Marcel Dekker, New York, 1987 (version française : Masson).
- [16] E. EWEDA et O. MACCHI, Convergence of the RLS and LMS adaptive filters, *IEEE Trans. Circuits and Systems*, 34, n° 7, juillet 1987, p. 799-803.
- [17] O. MACCHI et E. EWEDA, Compared speed and accuracy of the RLS and LMS algorithms with constant forgetting factors, *RAIRO-APII*, 22, n° 3, 1988, p. 255-267.
- [18] F. LING et J. PROAKIS, Non stationary learning characteristics of least squares adaptive estimation algorithms, *Proceedings of IEEE-ICASSP*, San Diego, USA, March 1984.
- [19] M. BELLANGER, Filtrage adaptatif, chap. XI dans *Traitement numérique du signal*, Masson, 1984.
- [20] P. FABRE et C. GUEGUEN, Fast recursive least squares algorithms: preventing divergence, *Proceedings of IEEE-ICASSP*, Tampa, USA, April 1985.
- [21] R. ALCANTARA, J. PRADO et C. GUEGUEN, Fixed point implementation of the fast Kalman algorithm: using a TMS 32010 microprocessor, *Signal Processing III*, Eusipco 1986, The Hague, Sept. 1986, p. 1335-1338.
- [22] J. L. BOTTO, Numerical divergence of fast recursive least-squares transversal filters, *RAIRO-APII*, 22, n° 3, 1988, p. 231-254.
- [23] G. CARAYANNIS, D. MANOLAKIS et N. KALOUPSIDIS, A unified view of parametric processing algorithms for prewindowed signals, *Signal processing*, 10, n° 4, 1986, p. 335-368.
- [24] J. M. TRAVASSOS-ROMANO et M. BELLANGER, Pôles et zéros des prédicteurs numériques, *Traitement du signal*, 3, n° 2, 1986, p. 67-77.
- [25] SODERSTRÖM et STOICA. — *Lecture notes in Control and Information Sciences*, Springer Verlag, 57, 1983.