

Comparaison d'algorithmes adaptatifs

en contexte non stationnaire

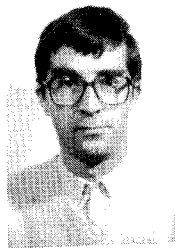
Comparison of adaptive algorithms in nonstationary context



Odile MACCHI

Laboratoire des Signaux et Systèmes (LSS), ESE, Plateau du Moulon, 91192 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 co-auteur 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 fellow des IEEE.



André GILLOIRE

CNET, Laboratoire de traitement des Signaux, route de Trégastel, 22301 Lannion.

Doctorat de 3^e cycle Electronique (1977). Entré au CNET en 1976. Recherches sur le codage de la Parole jusqu'en 1981. Depuis 1981 : travaux sur le traitement du signal à la prise de son pour les terminaux audio, particulièrement traitement adaptatifs et évaluation des performances. Participation au GRECO TDSI depuis sa création en 1988.



Christine SERVIÈRE

Institut National Polytechnique de Grenoble, Centre d'Étude des Phénomènes aléatoires et géophysiques, BP 46, 38402 ST MARTIN D'HÈRES CEDEX.

Ingénieur ENSIEG (1986). Thèse de doctorat INPG (1989). Domaines de recherche : filtrage adaptatif, soustraction de bruit.



Catherine LAMBERT-NEBOUT

Ingénieur CNES, département Traitement Bord, 18 avenue E. Belin, 31055 Toulouse Cedex.

Diplômée Ingénieur ENSEEIHT (Electronique) en 1986, elle reçoit le grade de Docteur de l'Institut National Polytechnique de Toulouse en 1989, après avoir soutenu une thèse sur l'analyse du signal basse fréquence d'un récepteur d'alignement de piste ILS. Depuis octobre 1989, elle occupe un poste d'Ingénieur au département Traitement Bord du CNES où elle participe aux projets Radar (Radar à synthèse d'ouverture, Radar à pluie) et Hermes (pour la partie compression Audio). Elle est responsable d'actions de R et T relatives à l'évaluation de composants de Traitement du Signal embarqués (Processeurs de traitement du Signal, ASICs, ...) et à la réalisation de modules embarqués de compression image optique ou radar.



Francis CASTANIE

ENSEEIH - 2, rue Camichel, 31071 Toulouse Cedex.

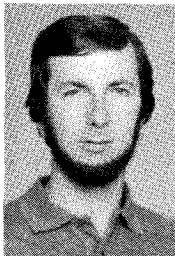
Professeur à l'INPT (ENSEEIH), Directeur du Laboratoire d'Électronique de cet établissement (LEN7). Ses activités de recherche s'effectuent au sein de l'un des groupes (GAPSE) de ce Laboratoire, qu'il anime directement. Elles sont centrées sur la Modélisation paramétrique du Signal et ses applications, telles que la Détection et la Classification - Reconnaissance automatiques.



Rocco SETTINERI

I3S/LASSY, URA n° 1376 du CNRS, Université de Nice Sophia-Antipolis, Bât. SPI n° 4, 250, avenue Albert Einstein, Sophia Antipolis, 06560 Valbonne et GDR TDSI.

Rocco Settineri, né en 1965, a obtenu le diplôme de DEA en Automatique et Traitement du Signal à l'Université de Nice-Sophia Antipolis en 1988. Il prépare actuellement sa thèse de Doctorat au Laboratoire d'Informatique, Signaux et Systèmes de Sophia Antipolis. Ses domaines de recherche concernent les algorithmes rapides en treillis et le filtrage adaptatif de signaux non stationnaires. Il participe également au projet PTAIA (Poste de Travail en Automatique et Intelligence Artificielle).



Gérard FAVIER

I3S/LASSY, URA n° 1376 du CNRS, Université de Nice Sophia-Antipolis, Bât. SPI n° 4, 250, avenue Albert Einstein, Sophia Antipolis, 06560 Valbonne et GDR TDSI.

Gérard Favier est né à Avignon en 1949. Il a obtenu le diplôme d'ingénieur de l'École Nationale Supérieure de Chronométrie et de Micromécanique de Besançon en 1973, et celui de spécialisation « Automatique Avancée » de l'École Nationale Supérieure d'Aéronautique et de l'Espace de Toulouse en 1974, ainsi que les titres de Docteur-Ingénieur (1977) et de Docteur ès Sciences (1981) de l'Université de Nice. Après avoir travaillé au Laboratoire d'Automatique et d'Analyse des Systèmes de Toulouse et au CAPCA (Direction des Constructions et Armes Navales de Toulon), il est entré au CNRS en 1976. Il est responsable de l'opération « Algorithmes et CAO pour l'Automatique et le Traitement du Signal », du Laboratoire d'Informatique, Signaux et Systèmes de Sophia Antipolis (I3S/LASSY), URA n° 1376 du CNRS, de l'Université de Nice-Sophia Antipolis. Ses activités de recherche concernent plus particulièrement l'Identification de Systèmes, le Filtrage Adaptatif, les Algorithmes Rapides, la Commande Adaptative et la CAO en Automatique et Traitement du Signal.

Jérôme BOUDY

MATRA COMMUNICATION, Division Radiocommunication, rue J. P. Timbaud, B.P. 26, 78392 Bois-d'Arcy et GDR TDSI.

Jérôme Boudy est né à Casablanca en 1960. Il a obtenu le diplôme d'ingénieur de l'Institut Catholique des Arts et Métiers de Lille en 1983, celui de DEA en Automatique et Traitement du Signal à l'Université de Nice en 1985 et le titre de Docteur Ingénieur en 1988 au Laboratoire de Signaux et Systèmes de l'Université de Nice. Il travaille actuellement à MATRA COMMUNICATION en tant qu'Ingénieur Traitement du Signal.

Philippe BRAGARD

Cephag/ENSIEG, URA 346 CNRS, B.P. 46, 38402 Saint-Martin d'Hères.

Maîtrise EEA, DEA d'électronique et traitement du signal obtenus en 1986 à l'université de Paris XI, Orsay. Prépare actuellement une thèse de doctorat dans le domaine de l'égalisation adaptative pour les transmissions de données en acoustique sous-marine, dans le cadre d'une collaboration entre la DRET et le Cephag.

Chapitre I MÉTHODOLOGIE COMPARATIVE

Odile MACCHI
Laboratoire des Signaux et Systèmes — CNRS/ESE

I.1. Introduction

Les algorithmes adaptatifs sont un puissant moyen pour parvenir à optimiser un système de structure spécifiée en fonction d'un environnement donné. Dans ce travail le système est un filtre linéaire \mathcal{H} . Le but de ce premier chapitre est de fournir un langage et des notations communes aux différents problèmes présentés dans cet article. Nous rappelons les modèles de filtrage linéaire usuels en signal et les algorithmes les plus classiques pour estimer ces modèles. Nous classons brièvement les non stationnarités qui affectent ces modèles. Enfin nous indiquons des méthodologies générales pour comparer les performances de ces algorithmes dans l'estimation de tels filtres non stationnaires. Il faut mentionner ici le rôle de pionnier joué par Widrow dès 1976 [6] dans le domaine des algorithmes adaptatifs de poursuite de non stationnarités, comme dans tout le reste des systèmes adaptatifs. Les autres contributions théoriques sur la poursuite n'ont commencé à paraître qu'à partir de 1981 [7]-[11].

Très souvent les grandeurs observées sont constituées par deux flots aléatoires.

1. Une suite $\{x(n)\}$ d'échantillons d'entrée dépendant du temps n , qui est traitée par le filtre \mathcal{H} à optimiser.

2. Une suite temporelle $\{y(n)\}$ qui constitue une référence bruitée du signal désiré à la sortie de \mathcal{H} . L'algorithme d'adaptation $\mathcal{H}_{n-1} \rightarrow \mathcal{H}_n$ du filtre \mathcal{H} est contrôlé par la suite des erreurs

$$(I.1.1) \quad e(n) = y(n) - \mathcal{H}_{n-1}(x(n)),$$

à la sortie du filtre estimé \mathcal{H}_{n-1} , dont l'état est le plus souvent calculé à l'instant $n-1$, c'est-à-dire sur la base des observations $(x(j), y(j))$ jusqu'à l'instant $j = n-1$ à l'exclusion de $j \geq n$. Une telle erreur est baptisée « a priori » car la nouvelle entrée $x(n)$ y est traitée à l'instant n par le filtre \mathcal{H}_{n-1} non encore adapté. On peut aussi utiliser l'erreur « a posteriori » recalculée selon (I.1.1) mais après avoir incrémenté \mathcal{H}_{n-1} en \mathcal{H}_n [12]. Que l'erreur soit prise a priori ou a posteriori, l'objectif est toujours d'amener $\mathcal{H}(x(n))$ à être proche de $y(n)$.

Comme il est bien connu [13], [14], il y a essentiellement deux familles d'algorithmes d'adaptation, les gradients, encore dénommés « LMS » (least mean squares) et les moindres carrés dits « récursifs » (MCR). Les premiers sont appréciés pour leur mise en œuvre plus simple et leur meilleure stabilité numérique vis-à-vis des erreurs d'arrondi, les

seconds sont connus comme ayant une vitesse de convergence supérieure (temps requis pour que \mathcal{H}_n se stabilise). Pour cette dernière raison beaucoup d'auteurs ont pensé que les MCR sont capables d'une meilleure poursuite dans un contexte non stationnaire.

Cette conjecture fait actuellement l'objet d'une controverse [3], [5], [15] que le présent article, dans tout son ensemble contribuera à tirer au clair, comme nous l'avons indiqué dans l'éditorial.

I.2. Modèles de filtrage

I.2.1. IDENTIFICATION LINÉAIRE

Dans toute la suite, l'on suppose que le signal de référence $y(n)$ est décrit par le modèle vrai

$$(I.2.1) \quad y(n) = \mathcal{F}(x(n)) + b(n)$$

qui représente une convolution, où \mathcal{F} est un filtre linéaire inconnu et où le signal $b(n)$ est un bruit de sortie non mesurable. On le suppose centré et indépendant des échantillons passés et présent $x(j)$, $j \leq n$, à l'entrée de \mathcal{F} , mais non nécessairement blanc. On appellera « signal vrai » la quantité

$$(I.2.2) \quad s(n) \triangleq \mathcal{F}(x(n))$$

par opposition au « bruit » $b(n)$. Naturellement $s(n)$ est lui-même inconnu. On appelle ρ le rapport signal sur bruit :

$$(I.2.3) \quad \rho = E(|s(n)|^2)/E(|b(n)|^2).$$

Dans le cas où $y(n)$ n'obéirait à aucun modèle linéaire du type (I.2.1), la modélisation est inadéquate et il est impropre d'essayer de s'approcher de $y(n)$ par un traitement linéaire des $x(n)$. C'est pourquoi on ne pratique le filtrage adaptatif que pour les modèles tels que (I.2.1).

On a alors affaire à un problème d'identification linéaire où le filtre estimé \mathcal{H} doit s'approcher de \mathcal{F} . Ceci est illustré par le schéma de la figure I.2.1 où l'on a noté

$$(I.2.4) \quad \hat{s}(n) \triangleq \mathcal{H}(x(n))$$

qui représente l'estimation du signal vrai $s(n)$ en (I.2.2).

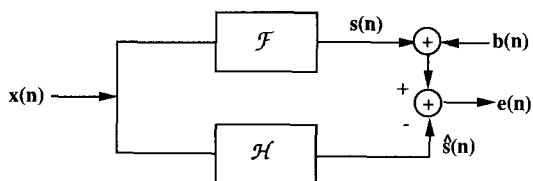


Fig. I.2.1. — Identification linéaire.

La qualité de \mathcal{H} est contrôlée par son erreur de sortie

$$(I.2.5) \quad e(n) = y(n) - \mathcal{H}(x(n)).$$

D'après (I.2.1), il apparaît que l'erreur de contrôle s'écrit

$$(I.2.6) \quad e(n) = \eta(n) + b(n),$$

à l'aide de l'erreur d'estimation sur le signal vrai

$$(I.2.7) \quad \eta(n) = s(n) - \hat{s}(n) = \mathcal{F}(x(n)) - \mathcal{H}(x(n)).$$

Ainsi donc $e(n)$ apparaît comme une estimation du bruit $b(n)$ du modèle (I.2.1). De plus comme $b(n)$ est indépendant des $x(j)$ pour $j \leq n$, en supposant que \mathcal{F} et \mathcal{H} sont des filtres causaux, les deux composantes $\eta(n)$ et $b(n)$ de (I.2.6) sont indépendantes. Ainsi, choisir l'estimation \mathcal{H} de \mathcal{F} qui minimise (en puissance) l'erreur de sortie $e(n)$, c'est aussi minimiser l'erreur d'estimation $\eta(n)$ sur le signal vrai $s(n)$.

Remarque importante. Contrairement aux apparences, les erreurs (I.1.1) et (I.2.5) ne sont pas identiques, même si nous avons cédé à l'usage en les désignant toutes deux par $e(n)$. En effet si $x(n)$ et $b(n)$ sont des suites stationnaires, l'erreur (I.2.5) l'est aussi, tandis que l'erreur (I.1.1) est non stationnaire par la variation du filtre \mathcal{H}_{n-1} .

I.2.2. TROIS PROBLÈMES TYPIQUES À RÉSOUDRE

Pour le modèle précédent d'identification linéaire, le présent article a effectué la comparaison de performance des algorithmes LMS et MCR pour trois problèmes typiques.

L'identification directe (ID), où c'est la valeur du filtre inconnu \mathcal{F} qui intéresse directement l'utilisateur. Une fois que \mathcal{F} est (bien) approché par \mathcal{H} , on en déduit la valeur (estimée) du bruit de sortie $b(n)$ grâce à la formule $\hat{b}(n) = e(n)$. Très souvent, quoique pas toujours, on suppose que la suite $b(n)$ est indépendante, identiquement distribuée (i.i.d.). Ce problème est illustré par les chapitres II.1 et II.2. Dans le premier cas, le processus d'entrée $x(n)$ est à bande large car c'est une phrase de parole. Dans le second cas il est à bande étroite, ce sont les fréquences propres d'une structure mécanique.

La prédiction (P), où il n'y a plus véritablement deux flots aléatoires $x(n)$ et $y(n)$, mais un seul flot $x(n)$ tandis que

$$(I.2.8) \quad y(n) = x(n+1)$$

est l'entrée future du filtre. Si l'on suppose que le filtre \mathcal{F} du modèle général (I.2.1) est causal, ce dernier se particularise ici selon

$$(I.2.9) \quad x(n+1) = \mathcal{F}(x(n)) + b(n),$$

qui correspond à la modélisation classique auto-régressive (AR) du signal $x(n)$ lorsque \mathcal{F} est un filtre RIF, ceci à la condition que le bruit $b(n)$ soit une suite i.i.d. Dans ce cas le « signal vrai » $s(n) = \mathcal{F}(x(n))$ est une prédiction de $x(n+1)$ puisque, au bruit près $b(n)$, \mathcal{F} fournit l'échantillon futur de x à partir de ses échantillons immédiatement antérieurs. La prédiction est donc l'identification de \mathcal{F} par un filtre \mathcal{F} causal. Quand elle est bonne elle fournit en même temps la modélisation du signal qui est généré par le filtre $\mathcal{H} = \mathcal{F}$ à partir du bruit blanc $b(n)$. Le filtre \mathcal{F} est ainsi le prédicteur optimal des échantillons x . On note :

$$(I.2.10 a) \quad \mathcal{F}(x(n)) = [\hat{x}(n+1)]_{\text{opt}}.$$

De la sorte, le bruit

$$(I.2.10 b) \quad b(n) = x(n+1) - [\hat{x}(n+1)]_{\text{opt}}$$

qui est restitué quand on identifie parfaitement \mathcal{F} par prédiction, n'est autre que le bruit générateur du signal $x(n+1)$. Pour cette raison, l'on peut dire que la prédiction se classe parmi les problèmes inverses puisqu'elle régénère l'entrée $b(n)$ du modèle de signal, à partir de l'observation du signal.

Le problème de la modélisation AR est traité dans le chapitre IV, en identifiant \mathcal{F} par un filtre \mathcal{H} en treillis (voir § I.2.3 ci-dessous).

Dans l'éditorial, nous avons aussi mentionné que si une (ou plusieurs) raies sont polluées par un bruit additif large bande, même en l'absence d'une référence corrélée avec ce bruit, la procédure de prédiction permet de séparer la (les) raie(s) du bruit additif, les valeurs $\hat{s}(n)$ et $e(n)$ fournissant respectivement les estimations de la (des) raie(s) et du bruit additif. Ce problème est illustré par les chapitres III.1 et III.2.

L'identification inverse (II) où les deux suites aléatoires $\{x(n)\}$ et $\{y(n)\}$ sont reliées par le modèle linéaire

$$(I.2.11 a) \quad x(n) = \mathcal{G}(y(n)) + b'(n),$$

qui est l'inverse du modèle (I.2.1). Là encore, le filtre \mathcal{G} et le bruit $b'(n)$ sont inconnus, les deux suites $\{y(n)\}$ et $\{b'(n)\}$ étant supposées indépendantes ; bien souvent $\{b'(n)\}$ est aussi une suite i.i.d. Il est donc naturel de se poser le problème de l'identification comme la recherche d'un filtre \mathcal{H} proche de \mathcal{G} et qui reçoit les échantillons $y(n)$. Il est clair qu'avec

$$(I.2.11 b) \quad \mathcal{F} = \mathcal{G}^{-1} \quad \text{et} \quad b(n) = -\mathcal{G}^{-1}(b'(n))$$

le modèle inverse (I.2.11 a) se réduit au modèle direct (I.2.1). Cependant $b(n)$ n'a plus la propriété i.i.d. Il devient coloré. De plus son spectre et son niveau de puissance varient avec le filtre direct

\mathcal{F} . C'est pourquoi ce problème n'est pas une extension triviale du problème direct. Il est illustré par le chapitre V.

I.2.3. STRUCTURE DU FILTRE D'ESTIMATION

Pour que l'écart entre le vrai filtre \mathcal{F} et son estimation soit le plus faible possible, la structure du filtre linéaire \mathcal{H} doit être adaptée au problème. Dans cet article deux structures seulement sont étudiées.

Structure transverse. Dans cette structure, $s(n)$ est estimé par

$$(I.2.12) \quad \hat{s}(n) = \mathcal{H}(x(n)) = \sum_{j=1}^N h_j x(n-j+1) = H^T X(n)$$

avec

$$X(n) = (x(n), \dots, x(n-N+1))^T, \\ H = (h_1, \dots, h_N)^T.$$

La structure transverse est illustrée sur la figure I.2.2.

On cherche les N coefficients du vecteur H , pour que $\hat{s}(n) - y(n)$ soit le plus petit possible ; H est directement une estimation de la réponse impulsionnelle de \mathcal{F} , qui est dit « à réponse impulsionnelle finie » (RIF). Le vrai filtre \mathcal{F} pourrait aussi être à « réponse impulsionnelle infinie » (RII) mais approché par une estimation tronquée de sa réponse.

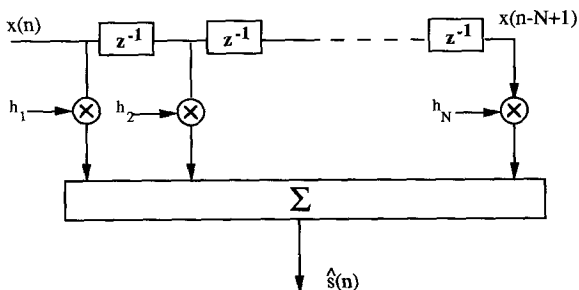


Fig. I.2.2. — La structure de filtre transverse.

Structure fréquentielle. Dans le domaine fréquentiel il est aussi possible d'estimer le filtre \mathcal{F} par son gain complexe $H' = (h'(1), \dots, h'(N))^T$, où N est le nombre de canaux fréquentiels [16]. Le signal d'entrée est alors constitué par les N suites d'échantillons $x(m, j)$ ($m = 1, \dots, N$) observés à la sortie de la voie fréquentielle numéro m . De la même manière l'observation $y(n)$ génère N suites $y(m, j)$ chacune observée dans la voie fréquentielle numéro m ($m = 1, \dots, N$).

Les $2N$ échantillons $x(m, j)$ et $y(m, j)$ résultent de la transformation de Fourier discrète à N points. Les signaux $x(n)$ et $y(n)$ sont découpés en tranches de N échantillons numérotées de 1 à j . Sur chacune sont effectuées les transformées de Fourier discrètes fournissant, à l'instant n , $2N$ suites d'échantillons

$x(m, 1), \dots, x(m, j), y(m, 1), \dots, y(m, j)$. Dans le cas particulier où les tranches sont choisies adjacentes j est relié à n par : $n = j \times N$.

Naturellement le filtrage \mathcal{H} s'écrit en fréquence comme un simple produit. Ainsi l'estimation du signal $s(m, j)$ sur la voie fréquentielle m est-elle

$$(I.2.13) \quad \hat{s}(m, j) = h'(m) x(m, j), \quad m = 1, \dots, N.$$

L'on cherche $h'(m)$ pour que l'erreur $[y(m, j) - s(m, j)]$ ou encore $[\hat{s}(m, j) - s(m, j)]$ soit la plus petite possible. Cette manière de poser le problème est plus simple car elle décompose le problème vectoriel en N sous-problèmes scalaires. La structure fréquentielle correspondante est illustrée sur la figure I.2.3.

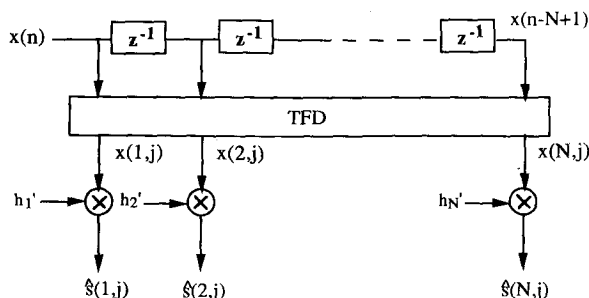


Fig. I.2.3. — La structure d'estimateur fréquentiel.

Structure treillis [17]. La structure de treillis, plus récente que la structure RIF, a l'avantage d'être modulaire car elle est faite d'une cascade d'étages identiques, et l'on montre qu'en allongeant le treillis par l'ajout d'un dernier étage, les étages antérieurs ne sont pas modifiés. Cette structure, lorsqu'elle est utilisée seule, est uniquement destinée à résoudre le problème de la prédiction, en calculant la prédiction $\hat{x}(n+1)$ et le bruit excitateur $b(n)$ indiqués dans (I.2.10 a et b).

Calcul du bruit excitateur

On définit deux familles $x_j(n)$ et $u_j(n)$ de signaux et une cascade de cellules d'ordre j selon

$$(I.2.14) \quad x_0(n) = u_0(n) \triangleq x(n).$$

La cellule d'étage j (ou d'ordre j) reçoit le couple $(u_{j-1}(n), x_{j-1}(n))$ et délivre le couple :

$$(I.2.15 a) \quad x_j(n) = x_{j-1}(n) - k_j u_{j-1}(n-1), \\ (I.2.15 b) \quad u_j(n) = -k_j x_{j-1}(n) + u_{j-1}(n-1).$$

Les k_j sont les PARCOR (« partial correlation »). Ils vérifient

$$(I.2.16) \quad |k_j| < 1.$$

Les quantités $x_j(n)$ et $u_j(n)$ sont appelées respectivement résidus (ou erreurs de prédiction) direct et rétrograde. Le bruit excitateur $b(n)$ de (I.2.10 b) est estimé par le dernier résidu direct :

$$(I.2.17) \quad \hat{b}(n) = x_N(n).$$

On cherche le vecteur

$$(I.2.18) \quad \mathbf{K} = (k_1, \dots, k_N)^T$$

des N coefficients PARCOR. Il ne donne pas directement la réponse de \mathcal{F} . La structure de treillis est illustrée sur la figure I.2.4.

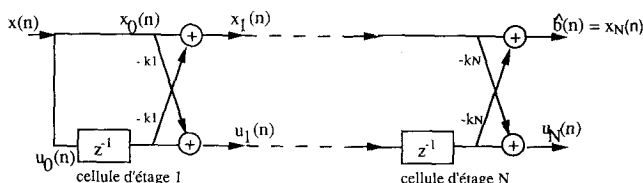


Fig. I.2.4. — La structure de treillis.

Le bon comportement numérique des filtres en treillis réside essentiellement dans leur propriété d'orthogonalisation. On peut montrer en effet [17] que le calcul récursif par rapport à l'ordre des coefficients PARCOR est effectué de façon à réaliser une orthogonalisation, étage par étage, des différents échantillons retardés du signal $x(n)$ qui entre dans le filtre. Cette orthogonalisation du type Gram-Schmidt délivre une séquence de résidus rétrogrades $\{u_j(n), j \in [0, N]\}$ qui est orthogonale :

$$(I.2.19) \quad E[u_j(n) u_k(n)] = 0 \text{ si } k \neq j.$$

Calcul du prédicteur

D'autre part la prédiction d'ordre N du signal $x(n)$ est obtenue à partir des résidus rétrogrades :

$$(I.2.20) \quad \hat{s}(n) \triangleq \hat{x}_N(n+1) = \sum_{i=1}^N k_i u_{i-1}(n).$$

En effet, en itérant N fois la formule (I.2.15 a) et en tenant compte de la condition initiale (I.2.14), nous trouvons :

$$(I.2.21) \quad x_N(n+1) = x(n+1) - \sum_{i=1}^N k_i u_{i-1}(n).$$

D'après la définition de l'erreur de prédiction avancée d'ordre N et d'après (I.2.17) il vient :

$$\hat{x}_N(n+1) = x(n+1) - x_N(n+1).$$

D'où le résultat (I.2.20).

La structure de treillis peut être utilisée comme intermédiaire pour réaliser un filtre transverse quelconque. Pour cela on réalise la structure en échelle qui est composée d'un treillis dont les sorties $u_j(n)$ des différents étages sont linéairement composées à la manière d'un filtre RIF. Elle comporte deux séries de paramètres à optimiser, ceux du treillis (vecteur \mathbf{K}) et ceux du filtre RIF (vecteur \mathbf{H}) [18].

I.3. Critères d'optimalité

Il s'agit de mesurer l'adéquation de l'estimateur \mathcal{H} au modèle vrai \mathcal{F} . Deux principaux critères sont

utilisés, les moindres carrés exacts et les moindres carrés moyens.

I.3.1. MOINDRES CARRÉS EXACTS (MCE)

Ce critère ne nécessite pas d'avoir une loi de probabilité pour les signaux $x(n)$ ou $b(n)$ [14]. A chaque instant n , on cherche le filtre \mathcal{H} donné par ses coefficients \mathbf{H} (ou par ses PARCORS dans le cas du treillis) tel que l'erreur moyenne sur le temps

$$(I.3.1) \quad J_n(\mathbf{H}) \triangleq \frac{1}{n} \sum_{j=1}^n |y(j) - \mathcal{H}(x(j))|^2$$

soit minimale. Ce filtre est fonction de n , soit $\mathbf{H}(n)$ le vecteur correspondant. Le critère J_n tient autant compte du passé que du présent. Il a une mémoire croissante du fait que la dimension de la fenêtre prise en compte dans le critère augmente. En conséquence, si \mathcal{F} varie au cours du temps n , même lentement, ce critère est inutilisable. Il donne lieu à des algorithmes d'optimisation dits « sans oubli ».

Il est à noter que d'autres formulations des Moindres Carrés Exacts sont disponibles dans la littérature. Ces différentes formulations sont obtenues soit en modifiant la fenêtre de minimisation du critère (algorithme à fenêtre glissante), soit en introduisant une pondération dans le critère. On parle dans ce dernier cas de méthodes des Moindres Carrés Pondérés dont un cas particulier important est celui des Moindres Carrés avec facteur d'oubli exposée au chapitre I.6. Nous verrons dans le chapitre IV comment ces deux techniques de modification du critère des Moindres Carrés peuvent être utilisées de façon adaptative dans le cas de signaux non stationnaires.

Dans le cas d'une estimation dans le domaine fréquentiel selon (I.2.13), le critère s'écrit dans chaque canal fréquentiel m selon

$$(I.3.2) \quad J_j^m(\mathbf{H}') = \frac{1}{j} \sum_{k=1}^j |y(m, k) - h'(m) x(m, k)|^2$$

$m = 1, \dots, N.$

Chaque minimisation fournit le gain scalaire $h'(m)$ dans la voie fréquentielle m comme fonction de j selon

$$(I.3.3) \quad h'(m, j) = \left[\sum_{k=1}^j y(m, k) x^*(m, k) \right] / \left[\sum_{k=1}^j |x(m, k)|^2 \right]$$

et par suite le vecteur de gains correspondant $\mathbf{H}'(j)$.

I.3.2. MOINDRES CARRÉS MOYENS (MCM)

Ce critère requiert une loi de probabilité pour les signaux. On cherche alors le filtre \mathcal{H} tel que l'erreur (moyenne d'ensemble ou espérance mathématique)

$$(I.3.4) \quad \mathcal{J}(\mathbf{H}) \triangleq E \{ |y(n) - \mathcal{H}(x(n))|^2 \}$$

soit minimale. Ce filtre est ou non indépendant de n selon que les suites $x(n)$ et $y(n)$ sont ou non

stationnaires. Dans le cas stationnaire on appelle \tilde{H} le vecteur correspondant. De plus, comme $b(n)$ est indépendant des $x(j)$ pour $j \leq n$ et que \mathcal{F} et \mathcal{H} sont causaux, il découle de (I.2.6), (I.2.7) que

$$(I.3.5) \quad \mathcal{J}(H) = P_b + E \{ |\mathcal{F}(x(n)) - \mathcal{H}(x(n))|^2 \}$$

où P_b est la puissance du bruit de sortie $b(n)$, de sorte que

$$(I.3.6) \quad \min_H \mathcal{J}(H) = \mathcal{J}(\tilde{H}) \triangleq \mathcal{E}_{\min} \geq P_b.$$

Si en outre \mathcal{H} possède assez de degrés de liberté pour modéliser le filtre vrai \mathcal{F} (dont on appelle F le vecteur de paramètres) alors

$$(I.3.7) \quad \tilde{H} = F; \quad \tilde{e}(n) = b(n); \quad \mathcal{E}_{\min} = P_b,$$

en notant

$$(I.3.8) \quad \tilde{e}(n) = y(n) - \tilde{H}^T X(n)$$

le bruit à la sortie du filtre optimal.

Remarque importante. Pour des signaux aléatoires $x(n)$, $b(n)$ stationnaires et ergodiques le critère (I.3.1) des (MCE) s'identifie avec le critère (I.3.4) des (MCM) lorsque n devient très grand, ceci en conséquence de la loi des grands nombres qui assimile moyennes d'ensemble et moyennes temporelles. Donc, pour n très grand, les (MCE) ont pour solution

$$(I.3.9) \quad \tilde{H}(n) = \tilde{H} (= F)$$

avec

$$(I.3.10) \quad \min_H J_n(H) = \mathcal{J}(F) = P_b.$$

1.4. Algorithmes d'acquisition

Deux familles d'algorithmes adaptatifs permettent de minimiser récursivement les deux critères précédents en recherchant le filtre \mathcal{H}_n , selon

$$(I.4.1) \quad \mathcal{H}_n = \mathcal{H}_{n-1} + \Delta(e(n), x(n), x(n-1) \dots),$$

$$(I.4.2) \quad e(n) \triangleq y(n) - \hat{s}(n) \\ = b(n) + [s(n) - \mathcal{H}_{n-1}(x(n))].$$

Ce sont des algorithmes basés principalement sur l'erreur $e(n)$ d'estimation du filtre à l'instant n . Le filtrage adaptatif qui en résulte est schématisé sur la figure I.4.1.

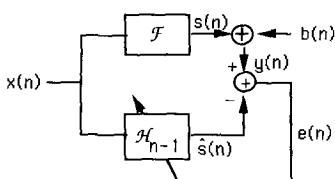


Fig. I.4.1. — Identification par filtrage adaptatif.

Structure transverse. Pour le critère (I.3.1) on peut montrer que $H(n)$ est solution de la récursion

$$(I.4.3) \quad H(n) = H(n-1) + A(n) X(n) e(n)$$

$$(I.4.4) \quad A(n) = A(n-1) - \frac{A(n-1) X(n) X(n)^T A(n-1)}{1 + X(n)^T A(n-1) X(n)} \quad (\text{MCR})$$

dénommée « moindres carrés récursifs » (MCR). La matrice $[nA(n)]$ est une estimation de l'inverse de la matrice de covariance d'ordre N du signal d'entrée $x(n)$.

Pour le critère (I.3.4), et dans le cas stationnaire, le vecteur \tilde{H} peut être obtenu comme limite, lorsque n tend vers l'infini, de la récursion

$$(I.4.5) \quad H(n) = H(n-1) + \mu(n) X(n) e(n)$$

$$(I.4.6) \quad \mu(n) > 0; \quad \sum_n \mu(n) = \infty; \quad \sum_n \mu^2(n) < \infty \quad (G)$$

dénommée « gradient stochastique » (G).

Structure fréquentielle. Dans le domaine fréquentiel, les composantes $h'(m, j)$ du vecteur $H'(j)$ ont elles aussi une écriture récursive analogue à (I.4.3) — mais avec des grandeurs complexes — selon

$$(I.4.7) \quad h'(m, j) = h'(m, j-1) + \frac{x^*(m, j) e(m, j)}{\sum_{k=1}^j |x(m, k)|^2}$$

où l'erreur sur la voie m est :

$$(I.4.8) \quad e(m, j) \triangleq y(m, j) - h'(m, j-1) x(m, j).$$

Le vecteur correspondant $H'(j)$, dont les composantes sont calculées par (I.4.7, 8) est dénommé « moindres carrés fréquentiels » (MCF).

Structure treillis ou échelle. L'algorithme des moindres carrés est un peu plus complexe pour ces structures. On pourra se reporter au livre de M. Bellanger [14]. Dans le cadre de cet article les algorithmes adaptatifs correspondants, ainsi que le gradient stochastique (qui est une approximation de la minimisation asymptotique du critère moyen) sont explicités dans le chapitre IV qui traite de la prédiction en treillis.

Nous considérons maintenant la situation non stationnaire où les signaux observés $x(n)$ et $y(n)$ ont des propriétés statistiques variables au cours du temps.

1.5. Les modèles non stationnaires

1.5.1. LES ÉLÉMENTS ÉVOLUTIFS

Dans un contexte non stationnaire, le modèle linéaire (I.2.1) devient évolutif selon

$$(I.5.1) \quad y(n) = \mathcal{F}_n(x(n)) + b(n).$$

Les changements au cours du temps peuvent affecter la statistique de $x(n)$, celle de $b(n)$ ou encore le filtre linéaire \mathcal{F} .

1. La statistique de $x(n)$: par exemple la puissance P_x du bruit devient variable, soit $P_x(n)$. Ce peut être aussi la densité spectrale $\gamma^x(f)$ de $x(n)$ qui devient évolutive : $\gamma_n^x(f)$. C'est le cas, en particulier pour les signaux de parole $x(n)$ discutés au chapitre II.1. De même pour les signaux de raie dont la fréquence varie au cours du temps (cf. ch. II.2 et III.1), ainsi que pour les signaux dont le modèle AR subit des changements (cf. ch. IV).

2. Le filtre linéaire \mathcal{F} : il devient fonction du temps n , soit \mathcal{F}_n . Par exemple, dans le chapitre II.1, \mathcal{F} est le chemin parcouru par l'écho acoustique dans ses réflexions multiples sur les murs de la pièce. Ce chemin subit des variations temporelles à cause de mouvements dans la pièce. De la même manière, au chapitre V où il s'agit d'un canal de transmission sous-marin, il y a des variations à cause (par exemple) des mouvements de la surface de l'eau.

Ce sont les variations de \mathcal{F}_n qui sont les plus critiques puisque les algorithmes adaptatifs sont précisément destinés à les poursuivre.

3. La statistique de $b(n)$: la puissance P_b ou la densité spectrale $\gamma^b(f)$ de $b(n)$ peuvent elles aussi devenir fonction du temps. Tel est bien le cas dans les chapitres II.1 et V de cet article. Dans le deuxième cas, il s'agit du problème (II). Aussi les variations du filtre \mathcal{F}_n induisent-elles des variations dans le niveau de puissance et dans le spectre du bruit de sortie, ainsi qu'il a été indiqué à la fin du § I.2.2.

Dans tous les cas, les suites aléatoires $x(n)$ et/ou $y(n)$ deviennent non stationnaires et la fonction $\mathcal{L}(H)$ en (I.3.4) devient $\mathcal{L}_n(H)$. Elle atteint son minimum pour un vecteur $\tilde{H}(n)$ qui est lui-même évolutif. Dans toute la suite on posera

$$(I.5.2) \quad T(n) \triangleq \tilde{H}(n+1) - \tilde{H}(n)$$

qui est l'incrément du filtre optimal.

Les non stationnarités peuvent être classées selon leur nature statistique, selon que les éléments $F(n)$, $P_x(n)$ etc... varient avec le temps de manière **déterministe ou aléatoire**. Le cas déterministe typique est celui de la dérive où l'un de ces éléments varie linéairement, par exemple la fréquence aux chapitres II.2 et III.1. Dans le cas aléatoire on peut supposer que les incréments $T(n)$ sont centrés, le biais fixe éventuel pouvant être inclus dans la dérive déterministe. De plus on suppose souvent que $T(n)$ est une suite i.i.d. (cf. [4], [9]), ou bien que $F(n)$ est un processus markovien (cf. [6], [18]). Le cas des incréments aléatoires n'est pas illustré dans cet article car il l'est déjà abondamment dans la littérature. Nous l'évoquons rapidement à la fin du présent chapitre. En effet il existe un autre trait caractéristique plus important pour classer les non stationnarités, c'est leur vitesse.

I.5.2. VITESSE DE LA NON-STATIONNARITÉ

On ne trouve pas dans la littérature de critère quantitatif pour distinguer entre non stationnarité lente et rapide. C'est qu'en temps continu t il est facile de distinguer les évolutions lentes où les grandeurs $\mathcal{F}(t)$, $P_x(t)$ etc... varient continûment, des ruptures qui comportent des discontinuités sur certaines de ces grandeurs. Mais cela est beaucoup moins clair en temps discret. Dans cet article nous proposons de baser le critère variations lentes/variations rapides sur l'erreur à la sortie du filtre optimal, en supposant une modélisation exacte (\mathcal{H} a autant de degrés de libertés que \mathcal{F}) pour que les équations (I.3.7) soient vérifiées, soit, en non stationnaire

$$(I.5.3) \quad \tilde{H}(n) = F(n).$$

Ceci simplifie les formules sans changer la nature du problème. Imaginons que l'on dispose du filtre optimal $\tilde{H}(n)$ à l'instant n . Si le modèle linéaire était figé, l'erreur de sortie à l'instant $(n+1)$ serait :

$$(I.5.4 a) \quad \begin{aligned} e(n+1) &= [X(n+1)^T F(n) + b(n+1)] \\ &\quad - X(n+1)^T \tilde{H}(n), \\ &= b(n+1). \end{aligned}$$

Au lieu de cela, les variations de \mathcal{F}_n font que l'erreur de sortie devient, d'après (I.5.2)

$$(I.5.4 b) \quad e(n+1) = a(n+1) + b(n+1)$$

avec

$$(I.5.5) \quad a(n+1) = X(n+1)^T T(n).$$

Ainsi l'erreur de sortie comporte le « bruit de variation » en (I.5.5) qui s'ajoute au bruit b d'observation du modèle.

On dira que les non-stationnarités sont rapides si la décision entre les deux hypothèses $e = b$ et $e = a + b$ peut se faire sans traitement supplémentaire. Il en est ainsi si la puissance de a est au moins égale à celle de b . C'est pourquoi, l'on définit le « degré de non-stationnarité » qui est fonction du temps n , selon

$$(I.5.6) \quad v(n) \triangleq \sqrt{\frac{P_a(n)}{P_b(n)}}.$$

La vitesse des non-stationnarités s'apprécie en comparant à 1 le degré de non stationnarité (I.5.6). Les variations lentes sont celles pour lesquelles, en permanence

$$(I.5.7) \quad v(n) \ll 1, \quad \forall n.$$

Pour préciser quantitativement ces notions, prenons le cas où les incréments $T(n)$ sont indépendants des entrées $X(n)$ (cas du problème (ID)) avec des échantillons d'entrée $x(n)$ centrés et indépendants. Avec le paramètre

$$(I.5.8) \quad \delta(n) \triangleq E(|T(n)|^2),$$

le degré de non stationnarité s'écrit :

$$(I.5.9) \quad v(n) = \sqrt{\frac{P_x(n) \delta(n)}{P_b(n)}}$$

Les formules (I.5.6) et (I.5.9) mettent bien en évidence que les non stationnarités ne sont pas lentes ou rapidement dans l'absolu, mais seulement en référence au bruit $b(n)$ qui entache la mesure de la sortie du modèle \mathcal{F}_n .

Il est important de remarquer que les filtres adaptatifs ne peuvent pas poursuivre des non stationnarités si rapides que $v(n)$ serait en permanence supérieur à 1. En effet, le filtre adaptatif est incapable de passer de la valeur $\tilde{H}(n)$ à la valeur $\tilde{H}(n+1)$ au cours de la seule itération $n \rightarrow n+1$. C'est pourquoi l'idée de variations rapides est implicitement restreinte au cas des **ruptures** (R) où $v(n)$ dépasse la valeur 1 mais de manière uniquement occasionnelle. On veut dire par là, qu'après une grande variation $T(n)$ pour laquelle

$$(I.5.10) \quad v(n+1) \geq 1$$

on observe ensuite des incréments très faibles pour lesquels $v(k) \ll 1$ durant une période de durée $j-n$ assez grande pour que les algorithmes adaptatifs puissent atteindre à nouveau à l'optimalité $\tilde{H}(j)$ au temps j . Durant cette période calme, on peut même supposer que $P_a(k) = 0$, soit

$$(I.5.11) \quad T(k) \equiv 0, \quad k \in]n+1, j]$$

et s'il n'en était pas ainsi, on traiterait la non stationnarité résiduelle dans le cadre dit des **variations lentes**, qui sont les non stationnarités permanentes pour lesquelles (I.5.7) est toujours vérifiée. Parmi elles on distinguera les variations permanentes déterministes (PD) et les variations permanentes aléatoires (PA) selon que $T(n)$ est déterministe ou aléatoire centré.

En résumé on rencontre, en filtrage adaptatif, trois non stationnarités, vraiment typiques selon le tableau 1 ci-après.

TABLEAU 1
Types usuels de non stationnarités

(R) ruptures	(PD) variations permanentes déterministes	(PA) variations permanentes aléatoires
occasionnellement $v(n) \geq 1$; ailleurs $v(k) \ll 1$	$v(n) \ll 1, \quad \forall n$	$E(T(n)) \equiv 0$ $v(n) \ll 1, \quad \forall n$

Il est tout à fait immédiat de généraliser la définition précédente de la vitesse des non stationnarités au cas où (I.5.3) n'est pas vérifiée (\mathcal{H} n'a pas assez de coordonnées pour modéliser \mathcal{F}). Dans ce cas $P_b(n)$ est remplacé par $\mathcal{I}(\tilde{H}(n))$ qui est plus grand.

I.6. Algorithmes de poursuite

I.6.1. TAUX D'OUBLI

Les algorithmes (MCR) en (I.4.3, 4) et (G) en (I.4.5, 6) sont aussi mal adaptés l'un que l'autre au cas non stationnaire où \mathcal{F}_n (et éventuellement $P_b(n)$, $\Gamma_x^n(k)$ etc...) varient avec n .

Pour l'algorithme (MCR) on sait en effet que

$$(I.6.1) \quad A(n) = \left[\sum_{j=1}^n X(j) X(j)^T \right]^{-1}$$

Cette matrice devient nulle pour $n \rightarrow \infty$. De même $\mu(n) \rightarrow 0$ dans (I.4.6). Ainsi pour n grand, l'incrément devient nul et $H(n)$ se bloque dans les deux algorithmes ; il ne peut poursuivre \mathcal{F}_n qui varie. D'où l'introduction d'un facteur d'oubli λ dans le critère des moindres carrés qui devient :

$$(I.6.2) \quad J_n(\lambda, H) = \sum_{j=1}^n \lambda^{n-j} |y(j) - \mathcal{H}(x(j))|^2, \quad 0 < \lambda < 1.$$

Les erreurs du passé lointain sont amorties exponentiellement. Définissons le « taux » d'oubli

$$(I.6.3) \quad \beta \triangleq 1 - \lambda.$$

Pour l'algorithme (G) du gradient, on utilise un pas d'incrément constant :

$$(I.6.4) \quad \mu(n) = \mu$$

et on appellera taux d'oubli le pas normalisé selon

$$(I.6.5) \quad \nu \triangleq \mu P_x.$$

Il en découle deux algorithmes de poursuite.

I.6.2. MOINDRES CARRÉS RÉCURSIFS AVEC OUBLI (MCRO)

Cet algorithme s'écrit :

$$(I.6.6)$$

$$a) \quad H(n) = H(n-1) + A(n) X(n) e(n),$$

$$b) \quad \lambda A(n) =$$

$$A(n-1) - \frac{A(n-1) X(n) X(n)^T A(n-1)}{\lambda + X(n)^T A(n-1) X(n)}$$

La formule (I.6.6.b) est une écriture récursive de l'inverse de la matrice

$$R(n) = \sum_{j=1}^n \lambda^{n-j} X(j) X(j)^T,$$

qui est elle-même une estimation de la matrice de covariance des entrées $x(n)$.

La quantité $C(n) = A(n) X(n)$ est dite « gain de Kalman ». L'aspect limite de cet algorithme lorsque λ tend vers 1 redonne l'algorithme (MCR) sans oubli de (I.4.4). On peut donc intégrer ce dernier dans le cadre général de l'algorithme (MCRO) avec oubli.

De même pour les algorithmes de moindres carrés fréquentiels : dans (I.4.7) la quantité $\left[\sum_{k=1}^j |x(m, k)|^2 \right]^{-1}$ devient nulle pour $j \rightarrow \infty$ et le vecteur $H'(j)$ se bloque et ne peut suivre \mathcal{F}_n . D'où l'introduction d'un facteur d'oubli dans le critère (I.3.2) qui devient :

$$(I.6.7) \quad J_j^m(\lambda, H') = \sum_{k=1}^j \lambda^{j-k} |y(m, k) - h'(m) x(m, k)|^2,$$

et qui est minimisé par le gain complexe

$$(I.6.8) \quad h'(m, j) = \frac{\left[\sum_{k=1}^j \lambda^{j-k} y(m, k) x^*(m, k) \right]}{\sum_{k=1}^j \lambda^{j-k} |x(m, k)|^2}$$

dans la voie fréquentielle m . Naturellement $h'(m, j)$ peut aussi s'exprimer récursivement suivant la formule des moindres carrés fréquentiels avec oubli (MCFO) :

$$(I.6.9) \quad h'(m, j) = h'(m, j-1) + \frac{x^*(m, j) e(m, j)}{r(m, j)}$$

$$(I.6.10) \quad r(m, j) = (1 - \lambda)r(m, j-1) + \lambda |x(m, j)|^2.$$

I.6.3. L'ALGORITHME DIT « LMS »

C'est l'algorithme (I.4.5) accompagné de (I.6.4) :

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

L'on peut aussi utiliser un pas d'incrément $\mu(n)$ dépendant du temps pourvu qu'il ne tende pas vers zéro comme c'était le cas dans (I.4.6). Ainsi l'algorithme LMS normalisé (NLMS) qui sera développé en particulier dans le chapitre II.1 de ce rapport, dans lequel

$$(I.6.11) \quad \mu(n) = \nu / \hat{P}_x(n)$$

où $\hat{P}_x(n)$ est une estimation de la puissance de x à l'instant n .

En particulier, il apparaît sur (I.6.9 et 10) que l'algorithme (MCFO) n'est rien d'autre qu'un algorithme (NLMS) scalaire, normalisé, sur chacune des voies fréquentielles qui, rappelons-le, sont décorrélées. En effet pour β petit, la quantité $r(m, j)$ donnée par (I.6.10) est très proche de $P_x(m, j)/(1 - \lambda)$, la puissance des $x(n)$ dans la voie de fréquence m , divisée par β .

Tout comme les algorithmes sans oubli (MCR) et (G), les algorithmes de poursuite (MCRO) et (LMS) donnent au filtre \mathcal{H}_{n-1} la capacité de s'approcher de la vraie valeur \mathcal{F}_n du filtre, à partir d'une valeur initiale \mathcal{H}_0 très éloignée de \mathcal{F}_1 : il s'agit de la phase transitoire d'acquisition de durée T_0 . Puis, en régime

permanent \mathcal{H}_{n-1} se met à fluctuer autour de la valeur vraie \mathcal{F}_n qui est évolutive (voir fig. I.6.1). C'est la notion de résidu étudiée au § I.7.

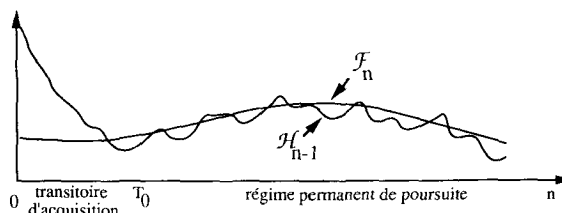


Figure I.6.1. — Algorithme adaptatif en régime non stationnaire.

I.6.4. UN FORMALISME COMMUN

Comme il a été dit au § I.5.1, en contexte non stationnaire le critère (I.3.4) des (MCM) devient une fonction $J_n(H)$ indexée par n parce que les statistiques des suites $y(n)$ et $x(n)$ dépendent de n . Le filtre optimal $\tilde{H}(n)$ varie avec n afin de poursuivre les variations du vrai modèle \mathcal{F}_n . Pour les deux familles d'algorithmes nous utiliserons les notations

$$(I.6.12) \quad V(n) \triangleq H(n) - \tilde{H}(n),$$

$$(I.6.13) \quad \tilde{e}(n) \triangleq y(n) - \tilde{H}^T(n) X(n),$$

qui sont respectivement la déviation du filtre par rapport à l'optimalité et le bruit d'observation à la sortie du filtre optimal, dont nous avons appelé \mathcal{E}_{\min} la puissance (cf. (I.3.6)). Il faut noter qu'en non stationnaire on peut avoir $\mathcal{E}_{\min} = \mathcal{E}_{\min}(n)$. On peut alors écrire les algorithmes LMS et MCRO sous la forme unique :

$$(I.6.14) \quad \boxed{V(n) = B(n) V(n-1) + A(n) X(n) \tilde{e}(n) - T(n)}$$

où B et A sont des matrices aléatoires données dans le tableau 2 ci-dessous et où $T(n)$ est l'incrément du filtre optimal, variable au cours du temps (cf. (I.5.2)).

TABEAU 2

Les matrices des algorithmes LMS et MCRO

	LMS	MCRO
$A(n)$	μI	$R(n)^{-1}$
$B(n)$	$I - \mu X(n) X^T(n)$	$I - R(n)^{-1} X(n) X^T(n)$

Il apparaît sur (I.6.14) que $V(n)$ est linéaire par rapport à $V(n-1)$. On peut alors séparer $V(n)$ en trois parties qui seront découplées, selon :

$$(I.6.15) \quad V(n) = V_i(n) + V_f(n) + V_r(n),$$

$$(I.6.16) \quad V_i(n) = B(n) V_i(n-1);$$

$$V_i(0) = V(0),$$

(I.6.17)

$$V_f(n) = B(n) V_f(n-1) + A(n) X(n) \tilde{e}(n);$$

$$V_f(0) = 0,$$

(I.6.18) $V_r(n) = B(n) V_r(n-1) - T(n);$

$$V_r(0) = 0;$$

$V_f(n)$ est la **déviatio n transitoire** qui décroît exponentiellement durant la phase d'acquisition à partir de la déviation initiale $V(0)$, indépendamment du niveau de bruit $\mathcal{E}_{\min}(n)$. La vitesse de décroissance dépend de la matrice $B(n)$. Elle est meilleure pour les MCRO que pour le LMS (transitoire T_0 plus court), voir figure I.6.1 ;

$V_f(n)$ est la **déviatio n de fluctuations** qui démarre à zéro et devient non nulle à cause du bruit de sortie $\tilde{e}(n)$, et cela même en régime permanent ($n \rightarrow \infty$) ; elle est indépendante des variations de \mathcal{F}_n ;

$V_r(n)$ est la **déviatio n de retard** qui dépend seulement des non stationnarités du contexte, c'est-à-dire du fait que $T(n)$ n'est pas nul. Elle est indépendante du bruit $\tilde{e}(n)$.

I.7. Les résidus

La présence des taux d'oubli β et ν donne aux algorithmes MCRO et LMS la capacité d'oublier le passé et de poursuivre le filtre variable \mathcal{F}_n . En contrepartie, même pour un filtre \mathcal{F} constant, le filtre adaptatif perd son exactitude : l'algorithme (MCRO) qui fournit le minimum du critère (I.6.2) ne minimise plus exactement les moindres carrés (I.3.1). On trouve pour les (MCRO) que le critère initial devient, lorsque $n \rightarrow \infty$

$$(I.7.1) J_n(H(n)) = \mathcal{E}_{\min}(n)(1 + \mathcal{M}_f^{\text{MCR}} + \mathcal{M}_r^{\text{MCR}}).$$

De même l'algorithme (LMS) fournit :

$$(I.7.2) \mathcal{J}_n(H(n)) = \mathcal{E}_{\min}(n)(1 + \mathcal{M}_f^{\text{LMS}} + \mathcal{M}_r^{\text{LMS}}).$$

Les valeurs $\mathcal{M}_f^{\text{LMS}}$ et $\mathcal{M}_f^{\text{MCR}}$, dénommées « résidus de fluctuations normalisés » sont positives. De même $\mathcal{M}_r^{\text{LMS}}$ et $\mathcal{M}_r^{\text{MCR}}$, les « résidus de retard » sont positifs. Ces résidus étant des limites pour $n \rightarrow \infty$, ils sont non nuls seulement pour des non stationnarités de type permanent c'est-à-dire (PA) ou (PD) (voir § I.5.2).

Le caractère additif des résidus de fluctuation \mathcal{M}_f et de retard \mathcal{M}_r est une conséquence du découplage qui existe entre les déviations $V_f(n)$ et $V_r(n)$. A son tour le découplage est vrai à condition que les incréments $T(n)$ (variations de \mathcal{F}_n) soient indépendants des signaux $x(j)$ à l'entrée de \mathcal{F}_n , ce qui est le plus souvent vrai. Les résidus correspondent à une erreur d'estimation supplémentaire à la sortie du filtre adaptatif. Cette erreur caractérise la précision de convergence de l'algorithme adaptatif vers le filtre optimal qui minimise les erreurs quadratiques J_n et \mathcal{J}_n . On peut évaluer cette précision, car pour un petit taux d'oubli, l'on peut montrer [9], [19] que, pour $n \rightarrow \infty$, les résidus de fluctuation sont

proportionnels au taux d'oubli et au nombre de coefficients du filtre selon

$$(I.7.3) \mathcal{M}_f^{\text{MCR}} = \frac{N}{2} \beta; \quad \mathcal{M}_f^{\text{LMS}} = \frac{N}{2} \nu$$

tandis que les résidus de retard varient en sens inverse du taux d'oubli, avec un exposant qui dépend de la statistique des incréments $T(n)$. On doit donc ici distinguer les deux cas.

Variations déterministes (PD). Il a été montré sur plusieurs exemples ([10] et aussi chapitres II.2, III.1 et V de cet article) que

$$(I.7.4) \mathcal{M}_f^{\text{MCR}} = \frac{\eta^{\text{MCR}} P_x \delta}{\mathcal{E}_{\min} \beta^2}; \quad \mathcal{M}_r^{\text{LMS}} = \frac{\eta^{\text{LMS}} P_x \delta}{\mathcal{E}_{\min} \nu^2},$$

où δ est défini comme en (I.5.8), mais à l'aide d'une moyenne temporelle

$$(I.7.5) \delta = \overline{|T(n)|^2}.$$

Les coefficients positifs η dépendent de chaque application.

Variations aléatoires (PA). Il a été montré [5], [9], [20] que les résidus de retard sont analogues à (I.7.4) mais en y remplaçant β^2 par β et ν^2 par ν aux dénominateurs, tandis que δ redevient la moyenne d'ensemble (I.5.8).

I.8. Comparaison des algorithmes

I.8.1. GÉNÉRALITÉS

La suite de cet article est une comparaison des deux familles d'algorithmes dérivées des MCRO et du LMS dans divers contextes non stationnaires. Comme nous l'avons déjà dit, l'opinion qui prévaut dans la littérature, c'est que les MCRO sont supérieurs au LMS parce que leur meilleure vitesse d'acquisition leur assurera aussi une meilleure poursuite. Cependant nous venons de montrer que la vitesse d'acquisition est une propriété transitoire concernant la déviation $V_f(n)$ du filtre adaptatif par rapport à l'optimalité, mais non les déviations permanentes que sont $V_f(n)$ et $V_r(n)$. C'est pourquoi avant de conclure à la supériorité des MCRO il faut préciser quel critère de performance est utilisé. Il est clair que pour les problèmes transitoires (acquisition), c'est bien la vitesse de convergence, mais que pour les problèmes permanents (poursuite), c'est plutôt le résidu \mathcal{M} . On voit donc que pour les non stationnarités de type (R), (PA) et (PD) les méthodologies de comparaison sont différentes.

I.8.2. CAS DES RUPTURES (R)

En présence de changements occasionnels dans le filtre vrai \mathcal{F}_n et/ou dans la statistique des $x(n)$ et des $y(n)$, le meilleur algorithme est celui pour lequel la durée d'acquisition T_0 est la plus faible (vitesse maximale). Mais celle-ci dépend de la précision (résidu \mathcal{M}) qui est atteinte par l'algorithme en régime permanent. Ainsi comparer les durées T_0 associées aux deux algorithmes en fixant arbitrai-

rement les taux d'oubli correspondants n'a pas de sens. Or on trouve souvent dans la littérature des comparaisons entre $\beta = 0$ ($\lambda = 1$, pas de capacité de poursuite pour (MCR)) et $\nu \neq 0$ ((LMS) avec capacité de poursuite). *Il faut exclure absolument une telle méthodologie.*

Une possibilité est de *comparer le temps de convergence des deux algorithmes à résidu \mathcal{M} donné en permanent.* Comme les ruptures n'ont pas de caractère permanent, on a $\mathcal{M}_r = 0$, soit $\mathcal{M} = \mathcal{M}_f$, ce qui fournit d'après (I.7.3) la relation entre les taux d'oubli des deux algorithmes :

$$(I.8.1) \quad \beta = \nu.$$

On compare alors les temps de convergence correspondants $T_0^{\text{MCR}}(\lambda = 1 - \beta)$ et $T_0^{\text{LMS}}(\mu = \beta/P_x)$. Le résultat dépend de la dispersion

$$(I.8.2) \quad d \triangleq \frac{\lambda_{\max}(\mathbf{R})}{\lambda_{\min}(\mathbf{R})}$$

des valeurs propres de la matrice de covariance \mathbf{R} des entrées. Dès que d est notablement supérieur à 1, on observe que

$$(I.8.3) \quad T_0^{\text{MCR}} < T_0^{\text{LMS}}.$$

En d'autres termes, théoriquement, **l'algorithme des moindres carrés exacts converge toujours plus vite que le gradient**, comme il a déjà été dit. Mais il s'agit là de théorie. En pratique les mises en œuvres dites rapides des MCRO se heurtent à des problèmes d'instabilité numérique surtout lorsque β est grand. En général, on exige pour cette raison (voir chap. II.1) que

$$(I.8.4) \quad \beta \leq 1/(3N),$$

ce qui limite la vitesse de l'algorithme MCRO. On peut alors trouver que le LMS est plus rapide s'il arrive par exemple que

$$(I.8.5) \quad T_0^{\text{LMS}}(\mu = 1/(NP_x)) < T_0^{\text{MCR}}(\lambda = 1 - 1/(3N)).$$

Ce peut être le cas par exemple si d est voisin de 1 (entrée large bande). Dans (I.8.5), la valeur $\nu = 1/N$ a été choisie pour le LMS car c'est elle qui maximise la vitesse de convergence au second ordre [1].

I.8.3. ÉVOLUTION PERMANENTE (P)

Dans une situation de poursuite il y a un caractère permanent dans les incréments $T(n)$ du filtre optimal. Alors, passée la phase transitoire, chacun des deux algorithmes atteint un régime permanent où la déviation transitoire $V_t(n)$ s'annule et où le résidu est

$$(I.8.6) \quad \mathcal{M} = \mathcal{M}_f + \mathcal{M}_r,$$

selon qu'on l'a indiqué en (I.7.1) et (I.7.2). Les

expressions des fluctuations \mathcal{M}_f et du retard \mathcal{M}_r sont, d'après (I.5.9), (I.7.3), (I.7.4)

$$(I.8.7) \quad \mathcal{M}^{\text{MCR}}(\beta) = \frac{N}{2} \beta + v^2 r^{\text{MCR}}(\beta),$$

$$(I.8.8) \quad \mathcal{M}^{\text{LMS}}(\nu) = \frac{N}{2} \nu + v^2 r^{\text{LMS}}(\nu),$$

où la fonction positive $r^{\text{MCR}}(\beta)$ (resp. $r^{\text{LMS}}(\nu)$) est en β^{-2} (resp. ν^{-2}) ou en β^{-1} (resp. ν^{-1}) selon que les non stationnarités sont déterministes ou aléatoires.

Pour chaque expression, les deux termes varient en sens opposé. Ainsi *chacun des algorithmes admet un taux d'oubli optimal β_{opt} ou ν_{opt} qui minimise le résidu en réalisant le compromis entre les deux résidus de fluctuations et de retard.*

Il semble que la méthodologie judicieuse pour comparer les capacités de poursuite des deux algorithmes consiste à comparer $\mathcal{M}_{\min}^{\text{MCR}} = \mathcal{M}^{\text{MCR}}(\beta_{\text{opt}})$ et $\mathcal{M}_{\min}^{\text{LMS}} = \mathcal{M}^{\text{LMS}}(\nu_{\text{opt}})$ et à **déclarer que les MCRO surclassent le LMS si et seulement si $\mathcal{M}_{\min}^{\text{MCR}} < \mathcal{M}_{\min}^{\text{LMS}}$.**

Ces équations montrent que les variations δ du modèle n'interviennent qu'après normalisation par le bruit de sortie. D'après la définition (I.5.9), le paramètre $v^2 = \frac{\delta P_x}{\mathcal{E}_{\min}}$ est le carré du degré de non

stationnarité. **Ainsi la difficulté du problème de poursuite n'est pas directement liée à la vitesse des variations, contrairement à l'intuition première. Les variations du filtre optimal \mathcal{F}_n ne sont rapides qu'en comparaison avec le bruit de mesure sur la sortie.** Si ce dernier est élevé, le problème de poursuite est plus facile car l'imprécision de la mesure limite le rendu à donner sur les variations à poursuivre.

I.9. Conclusion

Nous venons de montrer que, dans la théorie, l'algorithme MCRO est plus performant que le LMS quand les non stationnarités sont de type rupture (R), car sa phase d'acquisition est plus courte. En revanche, la comparaison théorique des deux algorithmes pour des évolutions permanentes déterministes (PD) ou aléatoires (PA) ne donne pas les MCRO toujours supérieurs au LMS en poursuite. La réponse dépend des paramètres qui interviennent dans les résidus d'erreurs de sortie (I.8.7) et (I.8.8), à savoir le nombre N de prises du filtre et le degré ν de non stationnarité. Ce dernier élément rapporte les évolutions du filtre à poursuivre au bruit qui entache la mesure de la sortie du filtre. Il peut donc exister des cas où le filtre LMS poursuit mieux que les MCRO. Nous en verrons dans la suite de cet article.

La méthodologie de comparaison que nous venons de donner peut être utilisée dans des situations non stationnaires variées. Dans la théorie, il est possible de bien classer ces situations. En revanche, dans la pratique, c'est beaucoup plus difficile. Lorsqu'on observe par exemple, les variations de la réponse acoustique d'une salle selon les mouvements dans la

salle, ou bien les variations du canal de propagation sous-marin à longue distance, il n'est pas évident de dire si ces variations sont déterministes ou aléatoires, si elles sont permanentes ou occasionnelles, etc... C'est pourquoi dans les chapitres qui suivent les

conclusions des auteurs pourront parfois apparaître comme mitigées. Pourtant, le passage par l'outil théorique est une première étape incontournable pour pouvoir se représenter et comprendre un peu ce qui se passe dans une situation pratique.

Chapitre II

IDENTIFICATION DE STRUCTURES TRANSVERSES

1. APPLICATIONS AUX RÉPONSES ACOUSTIQUES

André GILLOIRE

Centre National d'Etudes des Télécommunications

II.1.1. Introduction

Les nouveaux terminaux audio développés pour les Télécommunications (téléconférence, postes mains-libres, visiophones, radiotéléphones mains-libres de voiture...) sont caractérisés par la diffusion de la parole sur haut-parleur et la prise de son par un ou plusieurs microphones distants de l'utilisateur. De ce fait, et contrairement aux téléphones classiques équipés d'un combiné, ces nouveaux terminaux interagissent fortement avec l'environnement acoustique : l'espace (salle, véhicule) dans lequel ils sont disposés, les bruits ambiants, etc... Les problèmes nouveaux rencontrés nécessitent des traitements spécifiques à la prise de son : annulation de l'écho acoustique, débruitage et déréverbération de la parole [21]. Ces problèmes peuvent être ramenés à l'identification de canal avec entrée connue ou non, et sortie connue bruitée. L'objectif est de minimiser pour l'utilisateur la puissance des signaux perturbateurs (écho, bruit, réverbération) sans dégrader le signal utile (parole).

Pour réaliser efficacement cette identification, on doit prendre en compte les propriétés particulières des canaux acoustiques et des signaux traités.

Les canaux acoustiques (trajets des ondes sonores) ont les propriétés caractéristiques suivantes :

- réponse impulsionnelle à durée infinie ; la partie qu'il est utile d'identifier en pratique varie typiquement entre 30 ms et 250 ms suivant les applications, soit plusieurs centaines à plusieurs milliers de points aux fréquences d'échantillonnage audio (8 et 16 kHz) ; cette réponse a une structure temporelle complexe (ensemble de réflexions dépendant de la géométrie de la salle, des obstacles présents, etc...) et n'admet pas de modèle simple ayant peu de paramètres ;

- non-stationnarité due aux mouvements des personnes, aux déplacements d'objets, etc. ; l'évolution temporelle peut être rapide, mais il n'y a pas de ruptures.

Le signal à l'entrée est la parole ou un mélange bruit + parole ; sa bande passante s'étend sur plusieurs octaves (300 Hz à 3 400 Hz pour la parole téléphonique, 150 Hz à 7 000 Hz pour la parole dite à « bande élargie »), son spectre n'est pas plat (formants, pente) et très variable dans le temps.

Le bruit en sortie (bruit d'ambiance acoustique) est généralement non-stationnaire, non blanc et peut

avoir un niveau élevé (parole locale appelée « double parole », bruit de roulement dans un véhicule).

On doit donc utiliser des algorithmes d'identification adaptative (puisque les canaux acoustiques sont inconnus et évoluent au cours du temps) qui soient robustes aux perturbations en sortie.

II.1.2. Modèle pour l'annulation d'écho acoustique

L'écho acoustique a pour origine la propagation du son depuis le haut-parleur vers le microphone d'un terminal (poste « mains-libres » par exemple) ; cette propagation inclut les réflexions sur les parois de la salle et sur les obstacles présents dans cette salle. On la représente par un canal de couplage acoustique. Le microphone capte en plus la parole locale provenant du locuteur présent dans la salle (c'est le signal utile qui doit être transmis) et le bruit ambiant (fig. II.1.1).

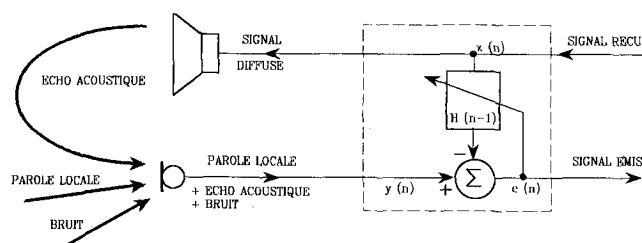


Fig. II.1.1. — Modèle de terminal audio avec écho acoustique. La partie encadrée schématise le principe de l'annulation d'écho.

L'écho acoustique est gênant pour le correspondant distant dès que le retard de transmission dans le réseau de télécommunication dépasse 30 ms (communications longue distance, transmissions par satellite) ; il doit donc être atténué le plus possible.

L'annulation d'écho représentée schématiquement figure II.1.1 consiste à retrancher au signal $y(n)$ provenant du microphone un « écho » $\hat{s}(n)$ estimé à partir du signal reçu $x(n)$ par filtrage à travers $H(n-1)$. Classiquement, H est un filtre transverse représentant directement la « partie utile » (début) de la réponse impulsionnelle ; il est obtenu par identification adaptative à partir du signal reçu comme entrée et du signal de sortie $y(n)$.

La figure II.1.2 représente une réponse impulsionnelle de couplage acoustique mesurée dans une salle de téléconférence. La durée de la « partie utile » est de l'ordre de 200 ms.

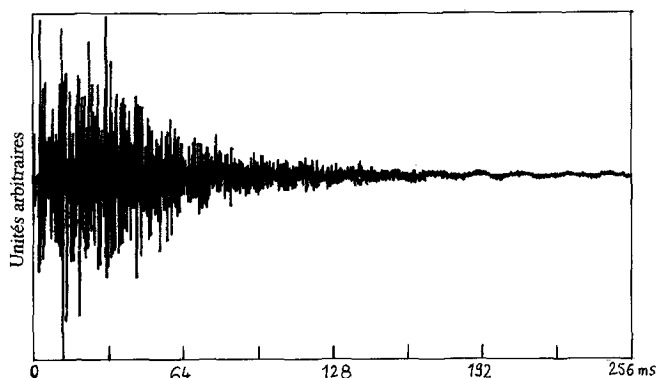


Fig. II.1.2. — Réponse impulsionnelle de canal de couplage acoustique (installation de téléconférence).

II.1.3. Comparaison des algorithmes NLMS et MCRO dans le contexte de l'annulation d'écho

Nous comparons ici la capacité de poursuite des deux algorithmes sur une variation contrôlée du système, en présence ou non de bruit en sortie, ainsi que la convergence initiale en présence de parole à l'entrée. L'identification est réalisée par un filtre transverse unique ; rappelons que diverses structures avec filtrage parallèle (sous-bandes, filtrage RIF rapide) ont été proposées pour l'annulation d'écho acoustique [22], [23].

II.1.3.1. PRÉSENTATION DES ALGORITHMES DE BASE

L'algorithme NLMS est une variante du LMS dont le gain d'adaptation est normalisé par la puissance (non-stationnaire) de l'entrée $x(n)$:

$$(II.1.1) \quad e(n) = y(n) - H^T(n-1) X(n)$$

$$(II.1.2 a) \quad \hat{P}_x(n) = \frac{1}{N} X^T(n) X(n)$$

ou :

$$(II.1.2 b) \quad \hat{P}_x(n) = (1 - \alpha) \hat{P}_x(n-1) + \alpha x^2(n)$$

$$(II.1.3) \quad H(n) = H(n-1) + \frac{\nu}{\hat{P}_x(n)} e(n) X(n)$$

avec :

$$0 < \nu \leq 2/N, \quad \alpha \approx 1/N, \quad \hat{P}_x(0) = P_0.$$

Ses avantages sont la faible complexité (mesurée en nombre de multiplications par itération : $2N$ si on utilise (II.1.2 b) pour le calcul de la puissance de l'entrée, $3N$ si on utilise (II.1.2 a) pour ce calcul) et la stabilité numérique (robustesse à la quantification du filtre et des autres paramètres).

Son principal inconvénient est que le transitoire

dépend de la dispersion spectrale du signal d'entrée ; la convergence peut être très lente dans le cas de la parole comme on le verra plus loin.

L'algorithme MCRO est une forme rapide de type Fast Transversal Filter (FTF) stabilisé numériquement par une contre-réaction des erreurs numériques sur l'erreur de prédiction retour [24], [25]. La complexité est de l'ordre de $8N$. Cet algorithme peut être décomposé comme suit :

- partie prédiction (rapide), ne dépend que du signal d'entrée $x(n)$ et non de $y(n)$
- partie filtrage, « classique » ; utilise un gain vectoriel calculé par la partie prédiction.

Le tableau II.1.1 donne toutes les équations de cet algorithme.

Tableau II.1.1. — Algorithme MCRO (FTF stabilisé)

- Choix des constantes de stabilisation : par exemple : $\mu^\gamma = \mu^\beta = \mu^b = 1$
- variables disponibles à l'instant t :

$$A(n-1), B(n-1), \tilde{C}_N(n-1),$$

$$\gamma_N(n-1), \alpha(n-1), \beta(n-1), H(n-1)$$

- Nouvelles informations : $x(n)$ et $y(n)$

- Modélisations de $x(n)$ et $x(n-N)$:

erreur de prédiction aller et variance :

$$(II.1.4) \quad \bar{e}(n) = x(n) - A(n-1)^T X(n-1)$$

$$(II.1.5) \quad \alpha(n) = \lambda \alpha(n-1) + \gamma_N(n-1) \bar{e}(n)^2$$

variable de vraisemblance et gain de Kalman dual d'ordre supérieur :

$$(II.1.6) \quad \gamma_{N+1}(n) = \frac{\lambda \alpha(n-1)}{\alpha(n)} \gamma_N(n-1)$$

$$(II.1.7) \quad \tilde{C}_{N+1}(n) = \begin{bmatrix} 0 \\ \tilde{C}_N(n-1) \end{bmatrix} - \frac{\bar{e}(n)}{\lambda \alpha(n-1)} \begin{bmatrix} 1 \\ -A(n-1) \end{bmatrix}$$

actualisation du prédicteur aller :

$$(II.1.8) \quad A(n) = A(n-1) - \bar{e}(n) \gamma_N(n-1) \tilde{C}_N(n-1)$$

erreur de prédiction retour :

$$(II.1.9) \quad \bar{r}(n) = x(n-N) - B(n-1)^T X(n)$$

stabilisation : indicateur de divergence

et erreurs de prédiction retour corrigées :

(\tilde{C}_{N+1}^{N+1}) : $N+1$ -ième composante du gain de Kalman d'ordre supérieur

$$(II.1.10) \quad \xi(n) = \bar{r}(n) + \lambda \beta(n-1) \tilde{C}_{N+1}^{N+1}(n)$$

$$(II.1.11) \quad \bar{r}^\gamma(n) = \bar{r}(n) + \mu^\gamma \xi(n)$$

$$(II.1.12) \quad \bar{r}^\beta(n) = \bar{r}(n) + \mu^\beta \xi(n)$$

$$(II.1.13) \quad \bar{r}^b(n) = \bar{r}(n) + \mu^b \xi(n)$$

variable de vraisemblance et gain de Kalman dual :

$$(II.1.14) \quad \gamma_N(n) = \frac{\gamma_{N+1}(n)}{1 + \gamma_{N+1}(n) \bar{r}^\gamma(n) \tilde{C}_{N+1}^{N+1}(n)}$$

$$(II.1.15) \quad \begin{bmatrix} \tilde{C}_N(n) \\ 0 \end{bmatrix} = \tilde{C}_{N+1}(n) - \tilde{C}_{N+1}^{N+1}(n) \begin{bmatrix} -B(n-1) \\ 1 \end{bmatrix}$$

actualisation du prédicteur retour

et variance erreur de prédiction retour :

$$(II.1.16) \quad B(n) = B(n-1) - \bar{r}^b(n) \gamma_N(n) \tilde{C}_N(n)$$

$$(II.1.17) \quad \beta(n) = \lambda \beta(n-1) + \gamma_N(n) (\bar{r}^\beta(n))^2$$

- Erreur de filtrage et actualisation du filtre transverse :

$$(II.1.18) \quad e(n) = y(n) - H(n-1)^T X(n)$$

$$(II.1.19) \quad H(n) = H(n-1) - e(n) \gamma_N(n) \tilde{C}_N(n)$$

Le principal avantage de cet algorithme est la convergence rapide, indépendante du spectre du signal d'entrée. De plus, cette version est robuste à la quantification des variables et des paramètres grâce à la technique de stabilisation employée. Son principal inconvénient est la complexité, 4 fois supérieure à celle du LMS.

Les performances en non-stationnaire (canal et signal d'entrée) de l'algorithme MCRO du tableau II.1.1 peuvent être considérablement améliorées au moyen de modifications simples décrites au paragraphe suivant.

II.1.3.2. MODIFICATIONS DE L'ALGORITHME MCRO

L'amélioration de la capacité de poursuite est obtenue par le contrôle du gain d'adaptation dans la partie filtrage transverse [26]. La partie prédiction est inchangée ; le facteur d'oubli λ est choisi de l'ordre de $1 - 1/3 N$ (soit un taux d'oubli $\beta \approx 1/3 N$) pour garantir la stabilité numérique. Dans la partie filtrage, l'équation (II.1.19) devient :

(II.1.20)

$$H(n) = H(n-1) - \frac{1}{1 - \rho \gamma_N(n)} \gamma_N(n) \tilde{C}_N(n) e(n)$$

$\rho < 1$ est un paramètre de contrôle ; en choisissant une valeur positive proche de 1, on augmente fortement le gain d'adaptation, et par conséquent la vitesse de poursuite ainsi que la vitesse de convergence sont améliorées (au détriment d'une sensibilité plus grande au bruit de sortie, comme on le verra plus loin).

Des expériences ont montré que l'algorithme MCRO perd subitement sa capacité d'adaptation en présence de signaux d'entrée non-stationnaires comme la parole [25]. Ce comportement particulier n'est pas lié à la stabilité numérique de l'algorithme, mais au caractère momentanément peu excitant du signal d'entrée.

Le maintien de la capacité d'adaptation pour les signaux d'entrée non-stationnaires est obtenu au moyen d'une procédure de réinitialisation conditionnelle simple [27]. La détection du blocage de l'algorithme est faite sur la chute vers zéro de la variable de vraisemblance $\gamma_N(n)$. On réinitialise alors les prédicteurs à zéro, les variances des erreurs de prédiction à une valeur liée à la variance à court terme du signal d'entrée et $\gamma_N(n)$ à 1. Les coefficients et la mémoire du filtre transverse sont préservés.

II.1.4. Méthodologie de comparaison adaptée à l'application

La comparaison des algorithmes dans le contexte de l'annulation d'écho acoustique n'a de sens que si l'on prend en compte les caractéristiques particulières de cette application. On distinguera deux cas typiques.

II.1.4.1. CAS DE L'ANNULATION D'ÉCHO ACOUSTIQUE EN SALLE, AMBIANCE CALME

La réponse impulsionnelle du canal est très étalée dans le temps ; en pratique, l'identification effectuée

par l'algorithme est partielle, c'est-à-dire que l'on a dans $H(n)$ la réponse impulsionnelle tronquée. Cette troncature introduit un résidu d'écho non négligeable (typiquement de l'ordre de 30 db en-dessous du niveau de l'écho). Par ailleurs, le bruit en sortie est faible (rapport signal à bruit de l'ordre de 50 dB). Les performances asymptotiques théoriques en stationnaire sont donc masquées, c'est-à-dire que M_f est négligeable devant le résidu d'écho. *Il est alors naturel de favoriser le comportement en transitoire (convergence) et la capacité de poursuite.*

II.1.4.2. CAS DE L'ANNULATION D'ÉCHO ACOUSTIQUE EN ENVIRONNEMENT « MOBILE » (VÉHICULE)

La « durée utile » de la réponse impulsionnelle est faible (50 à 100 ms) ; l'identification presque complète (non tronquée) est théoriquement possible. Le bruit en sortie est fort (rapport signal à bruit inférieur à 20 dB). Les performances asymptotiques en stationnaire sont donc visibles et importantes. Mais le canal acoustique évolue rapidement (mouvements du conducteur...), et par conséquent les performances en poursuite sont également importantes.

Un compromis vitesse-finesse est donc nécessaire (on peut de manière équivalente rechercher l'amélioration de la robustesse au bruit des algorithmes).

Les expériences présentées dans ce chapitre ont été faites avec des algorithmes optimisés pour une vitesse maximale :

— pour le NLMS :

$$(II.1.21) \quad \nu = 1/N$$

— pour le MCRO :

$$(II.1.22) \quad \beta = 1/3 N$$

Notons que ce choix, fait pour des raisons pratiques, est différent du choix $\nu = \nu_{opt}$ et $\beta = \beta_{opt}$ correspondant à la méthodologie de comparaison en poursuite proposée dans le chapitre I.

II.1.5. Poursuite d'un canal continûment variable

II.1.5.1. CONDITIONS EXPÉRIMENTALES

Une variation d'acoustique est difficile à contrôler en pratique ; il est cependant possible de simuler des variations artificielles apportant des dégradations des performances des algorithmes du même ordre de grandeur que les variations réelles. Dans les expériences présentées ici, le gain en sortie du canal (gain du microphone) a été augmenté progressivement de la valeur 1 à la valeur 2 sur un intervalle de temps de 0,7 s environ, après convergence initiale. Le canal est formé de 256 premiers points de la réponse impulsionnelle de salle de la figure II.1.2. La taille N du filtre transverse est 256.

Le signal à l'entrée du canal est un bruit stationnaire à spectre moyen de parole. Les signaux sont échantillonnés à la cadence de 16 kHz.

II.1.5.2. RÉSULTATS ET COMMENTAIRES

Les courbes sur les figures II.1.3 et II.1.4 représentent l'enveloppe de l'erreur $e(n)$ en décibels en fonction du temps, pour les 3 algorithmes : NLMS, MCRO et MCRO accéléré avec $\rho = 0,95$. Les variances des erreurs de prédiction dans les MCRO ont été initialisées de façon « optimale ».

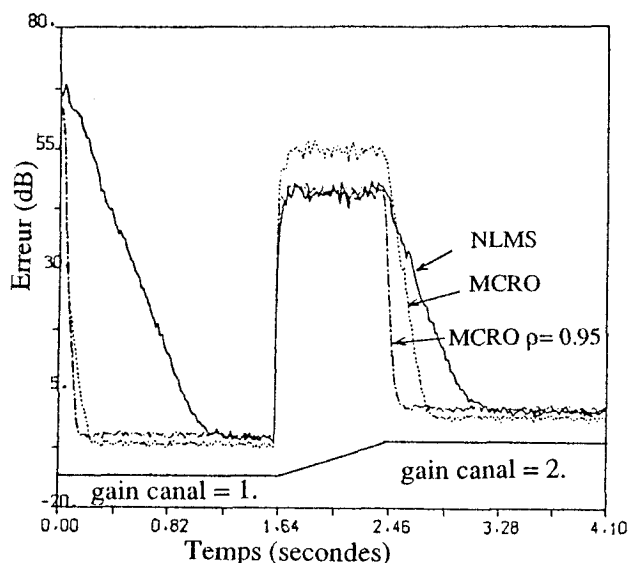


Fig. II.1.3. — Enveloppe de l'erreur $e(n)$ en fonction du temps pour trois algorithmes. Entrée : bruit à spectre de parole. Pas de bruit en sortie (1 s = 16 000 itérations).

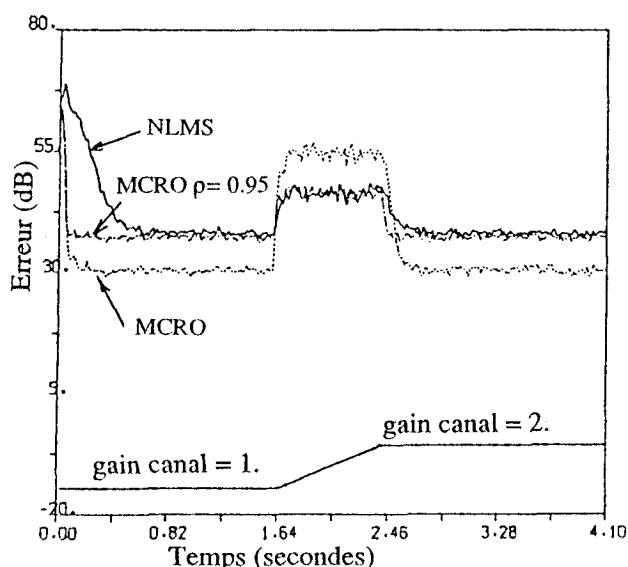


Fig. II.1.4. — Enveloppe de l'erreur $e(n)$ en fonction du temps pour trois algorithmes. Entrée : bruit à spectre de parole. RS/B en sortie : 40 dB (1 s = 16 000 itérations).

La figure II.1.3 correspond au cas où il n'y a pas de bruit ajouté en sortie ; le résidu observé (très faible) est dû à la quantification de la sortie du canal (avant gain) en entiers sur 16 bits.

La figure II.1.4 correspond à un rapport signal à

bruit de 40 dB en sortie (avant variation du canal) ; le bruit ajouté est un bruit blanc stationnaire pouvant simuler un bruit de ventilation, par exemple. L'erreur est débarrassée de sa composante de bruit, et traduit donc la performance intrinsèque de l'algorithme.

Dans les deux cas, le MCRO accéléré a la même performance que le NLMS en poursuite, mais il converge beaucoup plus vite. Dans le cas bruité, sa sensibilité au bruit est comparable à celle du NLMS, et elle est sensiblement accrue par rapport à l'algorithme MCRO non accéléré. Ceci est logique, car le gain d'adaptation — donc la capacité de poursuite mais aussi la sensibilité au bruit présent dans l'erreur — est augmenté dans le rapport 3 environ :

$$\gamma_N(n) \approx \lambda^N \text{ (asymptotiquement) .}$$

Pour $\lambda = 1 - 1/3 N$ et N grand, $\gamma_N(n) \approx 2/3$; par conséquent, le facteur multiplicateur du gain $1/(1 - \rho\gamma_{N,n})$ est voisin de 3 (ce qui est vérifié expérimentalement).

Ce facteur de 3 compense le rapport des taux d'oubli respectifs du NLMS et du MCRO, qui est précisément de 3 compte tenu des valeurs de ν et de λ choisies (équations (II.1.21) et II.1.22). Notons aussi que le rapport des erreurs de poursuite (assimilées aux erreurs globales en négligeant l'effet du bruit) pour les algorithmes MCRO et MCRO accéléré est de l'ordre de 10 dB (soit environ 3), ce qui indique un comportement du résidu de retard normalisé M_r^{MCR} en $1/\beta$.

En conclusion, le MCRO accéléré est ici globalement plus performant que le NLMS, puisqu'à erreur de poursuite et erreur de fluctuation équivalentes à celles du NLMS il offre des performances en transitoire très supérieures.

D'autres simulations dont les résultats ne figurent pas ici montrent que le modèle d'erreur complet (équations (I.2.19) et (I.2.20)) est vérifié. Une étude détaillée est présentée dans [28].

II.1.6. Comportement avec la parole à l'entrée

II.1.6.1. CONDITIONS EXPÉRIMENTALES

Le canal stationnaire est formé des 256 premiers points d'une réponse impulsionnelle de couplage acoustique mesurée dans une voiture. La taille N du filtre transverse est 256.

Le signal à l'entrée du canal est de la parole naturelle (phrases prononcées par un locuteur masculin). Les signaux sont échantillonnés à la cadence de 8 kHz.

II.1.6.2. RÉSULTATS ET COMMENTAIRES

Les courbes des figures II.1.5 et II.1.6 représentent l'enveloppe de l'erreur $e(n)$ en décibels en fonction du temps, pour les 3 algorithmes : NLMS, MCRO et MCRO accéléré avec $\rho = 0,95$.

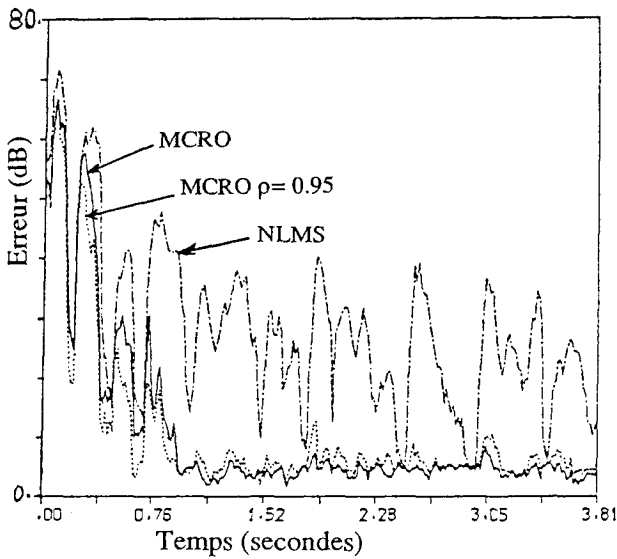


Fig. II.1.5. — Enveloppe de l'erreur $e(n)$ en fonction du temps pour trois algorithmes. Entrée : parole réelle. Pas de bruit en sortie (1 s = 8 000 itérations).

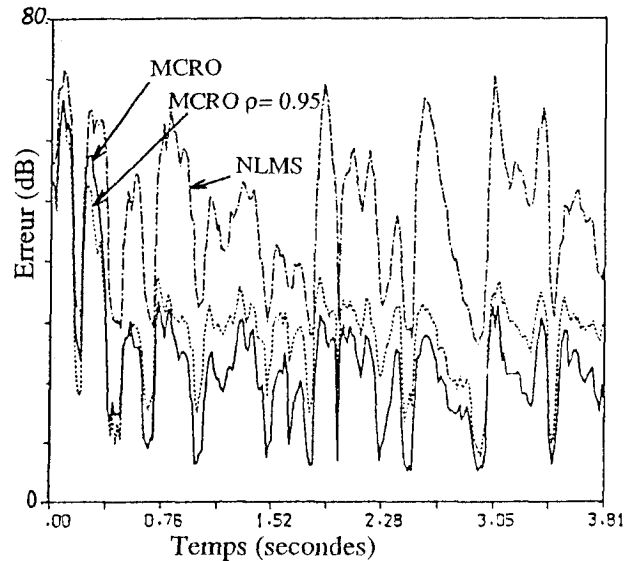


Fig. II.1.6. — Enveloppe de l'erreur $e(n)$ en fonction du temps pour trois algorithmes. Entrée : parole réelle. RS/B en sortie : 40 dB (1 s = 8 000 itérations).

Les MCRO incluent la procédure de réinitialisation sur détection de passage de la variable de vraisemblance en-dessous du seuil $\gamma_{\min} = 0,1$ décrite au § II.1.3.2. Les variances des erreurs de prédiction dans les MCRO ont été initialisées à une valeur correspondant à un niveau moyen de parole.

La figure II.1.5 correspond au cas où il n'y a pas de bruit ajouté en sortie. Le résidu asymptotique (visible seulement sur les courbes des MCRO) est dû à la quantification de la sortie du canal sur 16 bits. On vérifie que le MCRO est très supérieur au NLMS en convergence. Le MCRO accéléré est encore plus rapide ; sa performance asymptotique est légèrement inférieure à celle du MCRO non accéléré. On notera qu'à la fin de l'expérience de durée 3,81 s (soit 30 000 itérations environ) le NLMS est encore en phase transitoire.

La figure II.1.6 correspond à un rapport signal à bruit de l'ordre de 40 dB en sortie ; le bruit ajouté est un bruit blanc stationnaire. L'erreur est débarrassée de sa composante de bruit. On observe une dégradation importante de la performance asymptotique des MCRO (et, comme dans le cas étudié au II.1.5, le MCRO accéléré est plus affecté que le MCRO non accéléré), mais l'erreur asymptotique reste de l'ordre de grandeur du bruit ajouté. Pour le NLMS, la dégradation est considérable. L'algorithme est fortement perturbé ; en fait, on peut dire

qu'il ne converge pas. Ceci s'explique par la dynamique du signal de parole : pendant les périodes de faible énergie de celle-ci, le rapport signal à bruit en sortie est négatif, alors que le gain d'adaptation est très grand à cause de la normalisation automatique (II.1.3). Il en résulte une perturbation aléatoire des coefficients qui « dérègle » le filtre adaptatif. Celui-ci recommence à converger lors des périodes de forte énergie, mais cette convergence est lente et l'erreur ne peut atteindre sa valeur asymptotique avant l'arrivée des périodes de faible énergie suivantes.

Ceci montre que pour un signal non stationnaire à l'entrée comme la parole, les algorithmes MCRO sont intrinsèquement beaucoup plus robustes au bruit en sortie que le NLMS. En pratique, il est cependant possible d'améliorer sensiblement la robustesse du NLMS. Par exemple, on peut verrouiller l'adaptation dès que l'énergie du signal à l'entrée devient inférieure à un seuil faible. On peut aussi réduire le taux d'oubli ν pour le rendre plus proche de ν_{opt} (qu'il est difficile d'évaluer). Ces mesures se paient toutefois par un ralentissement de la convergence, et par un accroissement de l'erreur de retard lors des variations du canal acoustique qui peut être pénalisant subjectivement.

En conclusion, les expériences présentées montrent l'intérêt des algorithmes MCRO pour les applications de parole.

Chapitre II
IDENTIFICATION DE STRUCTURES TRANSVERSES
2. CAS DE SIGNAUX DE RAIES
APPLICATION A L'ESTIMATION D'UNE FRÉQUENCE
QUI DÉRIVE LENTEMENT

Christine SERVIÈRE

Centre d'Études des Phénomènes Aléatoires et Géophysiques

II.2.1. Introduction

Nous cherchons à estimer un signal utile $s(n)$ à partir d'une observation bruitée $y(n) = b(n) + s(n)$. Le signal $s(n)$, d'origine électrique, est constitué essentiellement de fréquences pures tandis que le bruit est large bande. Nous avons la possibilité dans cette application de capter un signal de référence $x(n)$, corrélé avec $s(n)$ et indépendant de $b(n)$.

Le principe de la méthode (§ I.1.1) est alors d'identifier le filtre physique \mathcal{F} inconnu, liant $x(n)$ et $s(n)$, et représentant ici la fonction de transfert acoustique d'une structure mécanique [par exemple l'environnement d'un sonar]. Le schéma de principe de l'identification avec voie de référence est le suivant (fig. II.2.1).

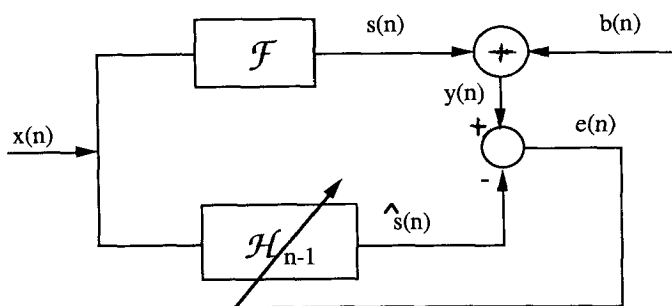


Fig. II.2.1. — Identification par filtrage adaptatif.

Le problème posé est de caractériser le filtre \mathcal{F} à l'aide des signaux $x(n)$ et $y(n)$, généralement non stationnaires.

Les deux types de non stationnarités de $x(n)$ sont déterministes (§ I.2.2 a, classe D) :

- 1) dérive lente en fréquence de raies (cas d'une modification de charge d'un moteur)
- 2) modulation en amplitude de raies.

Dans les deux cas, la vitesse des non stationnarités de $x(n)$ est faible dans le sens où la durée sur laquelle $x(n)$ peut être considéré comme localement stationnaire, c'est-à-dire sur laquelle la forme de la densité spectrale de puissance estimée de $x(n)$ varie peu (ici 1,28 s, cf. fig. II.2.2) est dix fois plus grande que la durée supposée du support du filtre \mathcal{F} .

II.2.2. Identification de \mathcal{F}

Le filtre physique \mathcal{F} à identifier est la fonction de transfert d'une structure mécanique fixe, son évolution est donc très lente par rapport à la durée précédente de 1,28 s : \mathcal{F} peut donc en pratique être considéré comme stationnaire dans l'intervalle de temps T étudié.

La structure choisie pour l'identification de \mathcal{F} est une structure transverse. Les algorithmes adaptatifs utilisés sont de type fréquentiel (Moindres Carrés Fréquentiels avec Oubli MCFO) ou de type temporel (LMS). Le premier identifie le gain complexe de \mathcal{F} par les coefficients $h'(m, j)$, pour la tranche temporelle de rang j et le canal de fréquence m (§ I.2 (I.2.16)). Le second identifie la réponse impulsionnelle de \mathcal{F} par les coefficients $H(n)$ à l'instant n (§ I.1.4 (I.1.31)). Leur choix a été déterminé en raison d'un faible coût de calcul, $\left(\frac{3}{2} \log_2 N\right) + 5$ produits et 1 division par échantillon temporel de sortie pour le MCFO, et $(3N + 1)$ produits par échantillon temporel de sortie pour le LMS [29].

Il n'est pas nécessaire ici d'employer une version normalisée du LMS, notée NLMS (§ II.1.3.1) car la puissance de la voie de référence est stationnaire. Une autre possibilité serait d'utiliser un algorithme du gradient en fréquence [16], de récurrence suivante sur les coefficients $h'(m, j)$:

$$(II.2.1) \quad h'(m, j) = h'(m, j - 1) + \mu(m) \times [x^*(m, j) e(m, j)] .$$

Cet algorithme, de faible coût de calcul $\left[\left(\frac{3}{2} \log_2 N\right) + 3 \text{ produits}\right]$ n'a d'intérêt qu'en choisissant une valeur différente de $\mu(m)$ par canal de fréquence m . De cette manière, la vitesse de convergence serait propre à chaque canal m [30]. Il nécessite cependant de faire un choix délicat des valeurs de $\mu(m)$ pour assurer la convergence de l'algorithme. C'est pourquoi nous avons préféré tester l'algorithme MCFO qui, pour un coût de calcul à peu près comparable, présente une facilité d'utilisation plus grande.

Nous comparons dans la suite les deux traitements envisagés (MCFO et LMS) dans le cas du premier

type de non stationnarité de $x(n)$: dérive en fréquence d'une raie.

II.2.3. Estimation de \mathcal{F} par son gain complexe (MCFO)

L'identification de \mathcal{F} se fait en fréquence, en estimant le gain complexe $h'(m, j)$ (I.1.19), par minimisation de (I.1.18) comme présenté au chapitre I. Cette solution est moins classique, néanmoins il existe des relations entre l'identification dans les domaines temporel et fréquentiel. Ainsi, on peut montrer que le critère classique temporel utilisé au chapitre I (I.1.17) est équivalent au critère fréquentiel suivant (II.2.1) [29]:

$$(II.2.1) \quad \mathcal{J}_n^m(H') = \frac{1}{n} \sum_{k=1}^n \left| \sum_{m=1}^N y(m, k) \exp 2 \pi i \frac{(m-1)(N-1)}{N} - \sum_{m=1}^N h'(m) x(m, k) \exp 2 \pi i \frac{(m-1)(N-1)}{N} \right|^2$$

où $y(m, k)$ est la transformée de Fourier discrète sur N points $[y(k-N+1), \dots, y(k)]$ à la fréquence m , et où le vecteur des coefficients $[h'(1), \dots, h'(N)]^T$ est la transformée de Fourier discrète du vecteur des coefficients $H = [h(1), \dots, h(N)]^T$ solution de (I.1.17) au même instant n .

Les transformées de Fourier discrètes de l'entrée $x(n)$ et de l'observation $y(n)$ sont ici effectuées sur des tranches se recouvrant de $(N-1)$ points, tandis que les coefficients $h'(m)$ sont renouvelés à chaque nouvel échantillon temporel de $x(n)$ et $y(n)$.

La solution $H' = [h'(1), \dots, h'(N)]^T$ est la suivante :

$$(II.2.2) \quad \begin{bmatrix} \alpha(1, 1) & \dots & \alpha(1, N) \\ \dots & \dots & \dots \\ \alpha(N, 1) & \dots & \alpha(N, N) \end{bmatrix} \begin{bmatrix} h'(1) \\ \dots \\ h'(N) \end{bmatrix} = \begin{bmatrix} \sum_{m=1}^N \exp 2 \pi i \frac{(m-1)(N-1)}{N} \sum_{k=1}^N y(m, k) x^*(1, k) \\ \dots \\ \sum_{m=1}^N \exp 2 \pi i \frac{(m-1)(N-1)}{N} \sum_{k=1}^N y(m, k) x^*(N, k) \end{bmatrix}$$

où

$$\alpha(u, v) = \exp 2 \pi i \frac{(m-1)(v-1)}{N} \sum_{k=1}^N y(v, k) x^*(u, k).$$

La résolution de l'équation (II.2.2) nécessite une inversion de matrice. Néanmoins, celle-ci est simplifiée si on fait l'hypothèse de canaux de fréquence

décorrélés soit :

$$\sum_{k=1}^N x(v, k) x^*(u, k) = 0 \quad \text{si } u \neq v$$

$$\sum_{k=1}^N y(v, k) x^*(u, k) = 0 \quad \text{si } u \neq v.$$

Ceci suppose que l'on travaille sur des tranches de longueur N suffisamment grandes devant le support de corrélation de $x(n)$ et $y(n)$ et que n est assez grand [16], [31], [32]. La solution obtenue H' est alors :

$$(II.2.3) \quad h'(m) = \frac{\sum_{k=1}^n y(m, k) x^*(m, k)}{\sum_{k=1}^n |x(m, k)|^2}, \quad m = 1, \dots, N.$$

Lorsque les vitesses d'évolution des phénomènes ne sont pas trop fortes, comme c'est ici le cas, il est intéressant de ne pas recalculer les coefficients $h'(m)$ à chaque instant n . Les transformées de Fourier discrètes sont alors effectuées sur des tranches de signal disjointes (ou se recouvrant plus ou moins), indicées par j , ce qui conduit aux observations $x(m, j)$ et $y(m, j)$. L'estimation sur des tranches de signal se recouvrant [38], [39], [40] est nécessaire si l'on veut ensuite reconstituer correctement le signal dans le domaine temporel. Dans notre application d'estimation de fréquence, ceci n'est pas envisagé et seule l'identification du filtre est importante. Le gain complexe est estimé par (I.2.16) sur des tranches adjacentes indicées par j , tel que $n = jN$. Pour les raisons développées au chapitre I (§ I.2), un facteur d'oubli λ est introduit dans son expression :

$$(II.2.4) \quad h'(m, j) = \frac{\sum_{k=1}^j \lambda^{j-k} y(m, k) x^*(m, k)}{\sum_{k=1}^j \lambda^{j-k} |x(m, k)|^2}, \quad m = 1, \dots, N.$$

II.2.4. Situation de poursuite

II.2.4.1. MODÉLISATION DU PROBLÈME

$x(n)$, signal de référence, est modélisé par une sinusoïde :

$$x(n) = \sqrt{P_x} \exp i \left[n\theta + \frac{n^2}{2} \psi + \phi \right]$$

où θ est la fréquence centrale, ψ le coefficient de dérive et ϕ une phase aléatoire.

$b(n)$ est modélisé en tant que bruit blanc gaussien, de moyenne nulle et de puissance P_b .

Le filtre \mathcal{F} est modélisé par un coefficient d'atténuation $\sqrt{\frac{P_s}{P_x}}$ et un retard fixe τ . Le signal $s(n)$ vaut

alors :

(II.2.5)

$$s(n) = \sqrt{P_s} \exp i \left[(n - \tau)\theta + \frac{(n - \tau)^2}{2} \psi + \phi \right].$$

Le rapport signal sur bruit de l'observation $y(n)$, avant traitement, noté ρ , est égal à :

$$\rho = \frac{P_s}{P_b}.$$

II.2.4.2. PERFORMANCES EN POURSUITE DE L'ALGORITHME MCFO

Pour comparer les capacités de poursuite des algorithmes MCFO et LMS dans le cadre de la modélisation précédente, nous calculons l'erreur résiduelle pour le canal fréquentiel d'indice m . Le signal $x(n)$ d'échantillons $x(0), \dots, x(n)$, est découpé en j tranches disjointes de N points ($n = j \cdot N$) sur lesquelles sont effectuées des transformées de Fourier discrètes $x(m, 0), \dots, x(m, j)$.

L'erreur résiduelle à la tranche d'indice j vaut (§ I) :

$$(II.2.6) \quad \mathcal{M}^{MCFO} = E \{ |\hat{s}(m, j) - s(m, j)|^2 \} \\ = E \{ |[h'(m, j - 1) - W_0(m, j)] \cdot x(m, j)|^2 \}$$

où $W_0(m, j)$ est le filtre non stationnaire optimal dépendant de l'indice de tranche j [33].

Nous cherchons à estimer parfaitement $s(n)$ constitué d'une sinusoïde dérivant lentement en fréquence. Pour la tranche d'indice j , nous nous intéresserons particulièrement à l'erreur résiduelle dans le canal fréquentiel spécifique $m(j)$ d'énergie maximale : $m(j)$ tel que $\frac{2\pi m(j)}{N} = \theta + jN\frac{\psi}{2}$.

$$(II.2.7) \quad \mathcal{M}^{MCFO} = E \{ |\hat{s}(m(j), j) - s(m(j), j)|^2 \} \\ = E \{ |[h'(m(j), j - 1) - W_0(m(j), j)] \cdot x(m(j), j)|^2 \}$$

où $W_0(m(j), j)$ est le filtre non stationnaire optimal dépendant de l'indice de tranche j .

Ce calcul, détaillé en [34], nécessite le calcul des transformées de Fourier discrètes de $x(n)$ et $s(n)$ pour le canal $m(j)$ et pour une tranche d'indice q : $0 \leq q \leq j$.

(II.2.8)

$$x(m(j), q) = \frac{\sqrt{P_x}}{N} \exp i \left(\phi + (q \cdot N)^2 \frac{\psi}{2} \right) \times \\ \times \sum_{k=0}^{N-1} \exp i \left[k(q - j) N \frac{\psi}{2} \right]$$

$$(II.2.9) \quad s(m(j), q) = \frac{\sqrt{P_s}}{N} \times \\ \times \exp i \left[\phi + (qN)^2 \frac{\psi}{2} - \tau \left(\theta + \frac{qN\psi}{2} \right) \right] \\ \times \sum_{k=0}^{N-1} \exp i \left[k(q - j) N \frac{\psi}{2} \right].$$

Ces expressions ont été obtenues, pour simplifier les calculs, en choisissant l'ordre du filtre N tel que $N\theta$ soit multiple de 2π . Ceci ne change pas l'expression finale de l'erreur résiduelle [34]. Nous avons aussi supposé la j -ième tranche au-delà du transitoire d'adaptation. Ainsi :

$$k \frac{\psi}{2} \ll jN \frac{\psi}{2} + \theta \quad \text{et} \quad (k - \tau) \frac{\psi}{2} \ll jN \frac{\psi}{2} + \theta$$

où $0 \leq k \leq N - 1$ dans (II.2.8) et (II.2.9). D'autre part, la longueur N doit être très supérieure au retard τ pour pouvoir identifier un filtre entre x et s . Si τ est trop grand, un retard Δ est introduit sur la voie de référence de manière à compenser approximativement τ .

L'expression du filtre optimal $W_0(m(j), j)$ est alors [34] :

$$(II.2.10) \quad W_0(m(j), j) = \frac{E \{ y(m(j), j) \cdot x^*(m(j), j) \}}{E \{ |x(m(j), j)|^2 \}} \\ = \frac{\sqrt{P_s}}{\sqrt{P_x}} \exp i \left[-\tau \left(\theta + \frac{jN\psi}{2} \right) \right].$$

L'erreur résiduelle \mathcal{M}^{MCFO} se décompose en deux termes, le premier dépendant du signal $s(n)$, le second du bruit $b(n)$, en détaillant l'expression des coefficients $h'(m(j), j - 1)$ (II.2.4) avec facteur d'oubli λ .

(II.2.11)

$$\mathcal{M}^{MCFO} = E \left\{ \left| \left[\frac{\sum_{k=0}^{j-1} \lambda^{j-1-k} s(m(j), k) \cdot x^*(m(j), k)}{\sum_{k=0}^{j-1} \lambda^{j-1-k} |x(m(j), k)|^2} - W_0(m(j), j) \right] \cdot x(m(j), j) \right|^2 \right\} + \\ + E \left\{ \left| \left[\frac{\sum_{k=0}^{j-1} \lambda^{j-1-k} b(m(j), k) \cdot x^*(m(j), k)}{\sum_{k=0}^{j-1} \lambda^{j-1-k} |x(m(j), k)|^2} \right] \cdot x(m(j), j) \right|^2 \right\}$$

en utilisant l'hypothèse d'indépendance entre b et x .

L'expression du premier terme, détaillé en [34], vaut :

$$(II.2.12) \quad \mathcal{M}_r^{MCFO} = P_s \cdot \left[\frac{N^2 \psi^2 \tau^2}{4 \beta^2} \right]$$

où β est le taux d'oubli introduit en (I.2.11) par $\beta = 1 - \lambda$.

L'expression (II.2.12) représente le résidu de retard, \mathcal{M}_r^{MCFO} , analogue au § I.3.2, et est une fonction décroissante de β . Cette expression est obtenue à l'aide d'hypothèses classiques [35] :

- faible dérive $N\psi \ll \beta$
- on néglige les termes transitoires en λ^j qui décroissent exponentiellement avec j (§ I.2.4) [35].

L'expression du deuxième terme, détaillé en [34], vaut :

$$(II.2.13) \quad \mathcal{M}_f^{MCFO} = \left[\frac{\beta}{2 P_b} \right].$$

Elle représente le résidu de fluctuation \mathcal{M}_f^{MCFO} , dû à la présence du bruit $b(n)$ et est proportionnelle au taux d'oubli β (§ I.3.2).

Nous utilisons pour ce calcul en plus des hypothèses ci-dessus, l'hypothèse d'indépendance entre $b(m(j), k)$ et $b(m(j), k + 1)$ [vérifiée par le fait que les transformées de Fourier sont effectuées sur des tranches disjointes] ainsi que l'hypothèse de taux d'oubli faible [développement au premier ordre en β].

L'erreur résiduelle totale est alors :

$$(II.2.14) \quad \mathcal{M}^{MCFO} = P_s \frac{N^2 \psi^2 \tau^2}{4 \beta^2} + \frac{\beta P_b}{2}.$$

Par une démarche analogue à [36], nous pouvons en déduire une valeur optimale du taux d'oubli β_{opt} qui minimise l'erreur résiduelle totale.

$$(II.2.15) \quad \beta_{opt} = [N^2 \psi^2 \tau^2 \rho]^{1/3} \text{ en posant } \rho = \frac{P_s}{P_b}.$$

La valeur minimale de l'erreur résiduelle est alors :

$$(II.2.16) \quad \mathcal{M}_{min}^{MCFO}(m(j)) = \frac{3}{4} P_s \rho^{-2/3} [N^2 \psi^2 \tau^2]^{1/3}.$$

II.2.4.3. PERFORMANCES EN POURSUITE DE L'ALGORITHME LMS

Nous calculons l'erreur résiduelle \mathcal{M}^{LMS} (§ I.3.2) d'expression :

$$(II.2.17) \quad \mathcal{M}^{LMS} = E \{ |\hat{s}(n) - s(n)|^2 \} \\ = E \{ |[H^T(n) - W_0^T(n)] \cdot X(n)|^2 \}$$

où $W_0(n)$ est le filtre non stationnaire optimal dépendant du temps [33].

$$(II.2.18)$$

$$W_0(n) = [E \{ X^*(n) X^T(n) \}]^{-1} \cdot E \{ y(n) X^*(n) \}.$$

Un calcul analogue à [33] montre que :

$$(II.2.19) \quad R_{XX}(n) = E \{ X^*(n) X^T(n) \} = \\ = P_X \underline{V}^n \underline{D}(0) [\underline{D}^T(0)]^* \underline{V}^{-n}$$

où \underline{V}^n est une matrice diagonale de terme

$$[\exp - i(n\psi\tau), \dots, \exp - i[n\psi(\tau - N + 1)]] \\ \underline{D}(0) = \exp - i\tau\theta \cdot \exp i\tau^2 \frac{\psi}{2} \times \\ \times \begin{bmatrix} 1 \\ \exp i\theta \cdot \exp - i \frac{\psi}{2} \\ \vdots \\ \exp i[(N-1) \cdot \theta] \cdot \exp - i \left[(N-1)^2 \frac{\psi}{2} \right] \end{bmatrix}$$

La matrice de corrélation $R_{XX}(n)$ n'est théoriquement pas inversible. En pratique, le capteur de référence reçoit en plus du signal $x(n)$ totalement corrélé à $s(n)$, un bruit indépendant $B(n)$ [représentant par exemple le bruit propre du capteur] de puissance P_B généralement très faible devant P_X .

Dans ces conditions, l'expression du filtre optimal a déjà été calculé en [33] :

$$(II.2.20) \quad W_0(n) = \frac{\sqrt{P_s} \sqrt{P_x}}{P_B + NP_x} \underline{V}^n \cdot \underline{D}(0).$$

Le terme de retard de l'erreur résiduelle \mathcal{M}^{LMS} a été calculé [34] en utilisant la formule classique suivante qui néglige les fluctuations faibles autour de la valeur moyenne de \mathcal{M}_r^{LMS} [37].

$$(II.2.21) \quad \mathcal{M}_r^{LMS} \approx \frac{[T_n^T]^* R_{XX}(n)^{-1} T_n P_x^2}{\nu^2}$$

où T_n est l'incrément du filtre optimal $W_0(n)$ défini par :

$$T_n = W_0(n+1) - W_0(n) \quad [37].$$

Dans notre cas :

$$(II.2.22) \quad T_n = \frac{\sqrt{P_s} \sqrt{P_x}}{P_B + NP_x} \underline{V}^n [V - I] \cdot \underline{D}(0).$$

Dans ces conditions, avec les hypothèses de dérive lente et taux d'oubli faible, et en considérant que le retard τ est en général petit devant N , l'expression du résidu de retard est [34] :

$$(II.2.23) \quad \mathcal{M}_r^{LMS} = \frac{P_s P_x^3}{\nu^2 (P_B + NP_x)^3} \left[\frac{(\tau\psi)^2}{3} N^3 \right].$$

Si P_B est très faible devant P_X , la valeur du résidu de retard tend vers :

$$(II.2.24) \quad \mathcal{M}_r^{LMS} = \frac{P_s}{\nu^2} \left[\frac{(\tau\psi)^2}{3} \right]$$

et est une fonction décroissante de ν (§ I.3.2).

Le résidu de fluctuation se calcule classiquement par :

$$(II.2.25) \quad \mathcal{M}_f^{LMS} = \frac{1}{2 P_X} \nu E \{ X^T X^* \} E \{ |b(n)|^2 \}$$

en supposant un taux d'oubli ν faible.

Dans notre cas :

$$(II.2.26) \quad \mathcal{M}_f^{LMS} \approx \frac{1}{2} \nu N \cdot P_b.$$

\mathcal{M}_f^{LMS} est une fonction croissante de ν (§ I.3.2).

L'erreur résiduelle totale vaut alors

$$(II.2.27) \quad \mathcal{M}^{LMS} = \left(\frac{1}{2} \nu N P_b \right) + \left(\frac{P_s (\tau \psi)^2}{3 \nu^2} \right).$$

Par une démarche analogue à [36], nous pouvons en déduire une valeur optimale du coefficient ν_{opt} qui minimise l'erreur résiduelle totale.

$$(II.2.28) \quad \nu_{opt} = \left(\frac{4 \rho (\tau \psi)^2}{3 N} \right).$$

La valeur minimale de l'erreur résiduelle vaut alors :

$$(II.2.29) \quad \mathcal{M}_{min}^{LMS} = \frac{3}{4} P_s \rho^{-\frac{2}{3}} (N \tau \psi)^{\frac{2}{3}} \left(\frac{4}{3} \right)^{\frac{1}{3}}.$$

II.2.4.4. COMPARAISON DES ALGORITHMES MCFO ET LMS

Dans le cas d'une référence composée uniquement d'une raie, l'erreur résiduelle n'affecte que celle-ci, le traitement étant linéaire.

Nous pouvons donc comparer, dans la mesure où $b(n)$ est un bruit blanc, les deux erreurs minimales résiduelles :

$$(II.2.29) \quad \mathcal{M}_{min}^{LMS} = \frac{3}{4} P_s \rho^{-\frac{2}{3}} (N \tau \psi)^{\frac{2}{3}} \left(\frac{4}{3} \right)^{\frac{1}{3}}$$

$$(II.2.16) \quad \mathcal{M}_{min}^{MCFO}(m(j)) = \frac{3}{4} P_s \rho^{-2/3} [N^2 \psi^2 \tau^2]^{1/3}.$$

Le rapport entre les deux erreurs résiduelles est :

$$\frac{\mathcal{M}_{min}^{MCFO}}{\mathcal{M}_{min}^{LMS}} = \left(\frac{4}{3} \right)^{-1/3}.$$

On remarque qu'il est proche de 1. Les deux algorithmes MCFO et LMS ont donc des comportements théoriques identiques dans le cas de l'estimation d'une seule raie.

II.2.5. Résultats expérimentaux

Les deux algorithmes précédents ont été appliqués à des signaux expérimentaux dans le cadre de l'estimation de raies sur des machines tournantes [environnement d'un sonar] (§ II.2.1) (fig. II.2.2).

Le signal bruité $y(n)$ observé (fig. II.2.2) est pré-

senté sous forme de 10 spectres évolutifs moyennés chacun sur 1,28 s de signal $\gamma_y(m)$. Sa largeur de bande est de 8 kHz. $y(n)$ est constitué d'un signal large bande $s(n)$ à spectre continu et d'un ensemble de raies (1) et (2) à estimer. La référence enregistrée est constituée uniquement des raies (1) et (2). Son étude a montré une forte cohérence ($\approx 0,9$) avec la raie (1) et une cohérence beaucoup plus faible (0,2 à 0,4) avec l'ensemble des raies (2) du signal bruité $y(n)$.

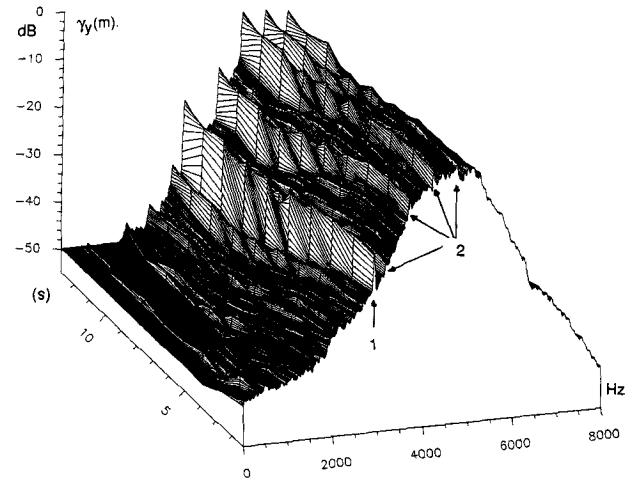


Fig. II.2.2. — Spectres évolutifs du signal bruité.

Les capacités de poursuite des deux algorithmes ont été comparées en étudiant le spectre du signal estimé $\hat{s}(n)$, noté $\gamma_s(m)$. Nous présentons les résultats obtenus pour des valeurs du taux d'oubli β_{opt} et du coefficient ν_{opt} qui minimisent cette erreur résiduelle au-delà du transitoire d'adaptation (> 5 s) (figs II.2.3 et II.2.4).

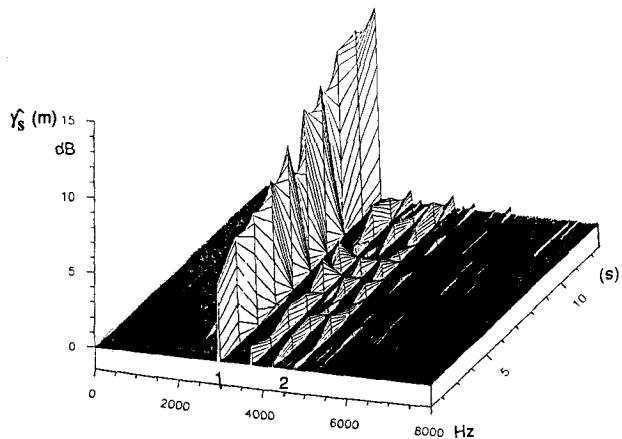


Fig. II.2.3. — Estimation par MCFO ($N = 512$ $\lambda = 0,9$).

Nous remarquons que l'erreur résiduelle est légèrement plus faible pour l'algorithme MCFO. Dans le cas de la raie (1), la différence de gain de traitement est de 1,5 à 2 dB suivant le spectre considéré. Dans le cas des raies (2), la différence de gain de traitement est de 1,5 à 3 dB suivant la raie et le

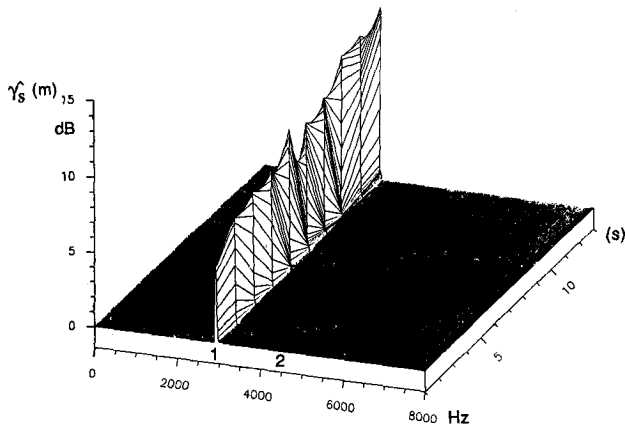


Fig. II.2.4. — Estimation par LMS ($N = 512$ $\mu = 1,3 E - 12$).

spectre considérés. Cependant, ces résultats sont très proches, ainsi que l'étude théorique l'avait montré.

Le calcul théorique de l'erreur résiduelle (§§ II.2.2.2 et II.2.2.3) a été effectué dans le cas d'une référence ne contenant qu'une raie. Si celle-ci contient plusieurs raies (2), dont l'une est prédominante sur les

autres (1) [ceci conduit en général à une cohérence faible entre $x(n)$ et $y(n)$ pour les raies (2) et une cohérence forte pour la raie (1)], il est certain que le traitement LMS ne sera efficace que pour la raie (1). Au contraire, l'intérêt d'un algorithme fréquentiel est que le traitement se fait indépendamment dans chaque bande de fréquence, ce qui est important dans le cas de signaux de raies.

II.2.3. Conclusion

Nous avons étudié deux algorithmes, de coûts de calcul faibles et comparables, minimisant un critère quadratique soit dans le domaine temporel (LMS), soit dans le domaine fréquentiel (MCFO). Le cadre de l'étude est l'estimation d'une fréquence qui dérive lentement. Le calcul théorique de l'erreur résiduelle en sortie du système montre des comportements proches entre l'algorithme MCFO et l'algorithme LMS. Cependant, dans le cas pratique d'une référence contenant plusieurs raies d'amplitude très différentes, l'algorithme LMS ne traitera que celle de plus grande amplitude, tandis que l'algorithme MCFO minimisera un critère quadratique pour chaque bande de fréquence de façon indépendante.

Chapitre III PRÉDICTION DE FRÉQUENCES

1. CAS DE LA DÉRIVE DE FRÉQUENCE

Odile MACCHI

Laboratoire des Signaux et Systèmes - CNRS/ESE

III.1.1 La prédiction linéaire

La prédiction linéaire est une méthode simple et efficace pour séparer un mélange comprenant un signal à bande étroite $s(n)$ (par la suite dénommé « signal ») et un signal à bande large $w(n)$ (par la suite dénommé « bruit »). Nous travaillerons sur des signaux complexes obtenus en passant, si besoin est, par un séparateur de phase (décomposition par filtrage de Hilbert à l'aide de composantes en quadrature).

Pour restituer une estimation $\hat{s}(n)$ de $s(n)$ à partir des observations $x(j)$, on choisit le retard entre échantillons successifs $x(n)$, $x(n-1)$, etc. de telle sorte que les $s(n)$ successifs demeurent cohérents tandis que les $w(n)$ successifs deviennent décorrélés. Ce retard est donc fonction des largeurs de bande respectives de $s(n)$ et $w(n)$. Pour simplifier, dans ce chapitre, nous prendrons un signal sinusoïdal et un bruit blanc. De la sorte le signal reçu à l'instant n est

(III.1.1)

$$x(n) = s(n) + w(n) = \sqrt{P_s} \exp j(n\theta + \varphi) + w(n)$$

où θ est la fréquence angulaire, P_s et P_w sont les puissances respectives du signal et du bruit. Par une combinaison linéaire H bien choisie des valeurs passées $x(n-j)$, $j \geq 1$, les $s(n-j)$ s'ajouteront en amplitude, tandis que les $w(n-j)$ s'ajouteront en puissance ; donc le rapport signal à bruit s'améliorera. On pose donc

$$(III.1.2) \quad \hat{s}(n) = \sum_{i=1}^N h_i x(n-i) = H^T X(n-1)$$

avec

(III.1.3)

$$X(n-1) = [x(n-1), x(n-2), \dots, x(n-N)]^T$$

L'équation (III.1.2) caractérise un prédicteur puisqu'elle n'utilise que le passé. L'erreur de prédiction est

$$(III.1.4) \quad e(n) \triangleq x(n) - \hat{s}(n)$$

Elle se décompose en deux termes :

$$(III.1.5) \quad e(n) = \eta(n) + w(n)$$

où

$$(III.1.6) \quad \eta(n) = s(n) - \hat{s}(n)$$

est l'erreur d'estimation. Quand H est bien choisi, η est faible, donc $e(n)$ est proche de $w(n)$.

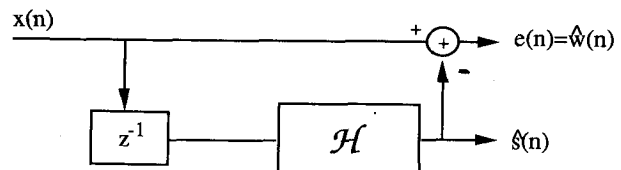


Fig. III.1.1. — Séparation d'un « signal » et d'un « bruit » par prédiction.

Le prédicteur délivre donc 2 sorties $\hat{s}(n)$ et $e(n)$ qui sont respectivement les estimations du signal et du bruit (voir fig. III.1.1). L'on doit rechercher le filtre \mathcal{H} avec ses coefficients H pour que $e(n)$ soit minimale (en puissance). Or, d'après (III.1.1 à 4)

(III.1.7)

$$e(n) = [s(n) - H^T S(n-1) - H^T W(n-1)] + w(n)$$

avec des notations semblables à (III.1.3) pour les vecteurs S et W . En supposant que la suite $w(n)$ est i.i.d., on a donc

$$(III.1.8) \quad E(|e^2(n)|) = E(|\eta^2(n)|) + P_w$$

Donc minimiser (en puissance) l'erreur d'estimation η ou l'erreur de prédiction e est équivalent.

L'on doit donc choisir \mathcal{H} pour minimiser $E(|e^2|)$. A noter que l'on pourrait raisonner sur des erreurs quadratiques exactes et non des espérances, ce qui justifie d'employer aussi bien les algorithmes de MCR que le LMS pour minimiser $e(n)$. On appelle \mathcal{F} (de paramètre F) le filtre qui minimise $E(|e^2|)$. En résumé, tout algorithme qui minimise $e(n)$ permettra de séparer $s(n)$ et $w(n)$ selon le schéma de la figure III.1.1.

En particulier, l'on considère les algorithmes avec oubli du chapitre I qui utilisent l'erreur

$$(III.1.9) \quad e(n) = x(n) - H(n-1)^T X(n-1)$$

et s'écrivent (en complexe)

$$H(n) = H(n-1) + \mu X(n-1)^* e(n) \quad (\text{LMS})$$

ou bien

$$\left. \begin{aligned} H(n) &= H(n-1) + R(n-1)^{-1} \\ &\quad \times X(n-1)^* e(n), \\ R(n) &= \sum_{j=1}^n \lambda^{n-j} X(j)^* X(j)^T. \end{aligned} \right\} \text{(MCRO)}$$

Ils convergent (en un sens probabiliste) vers le paramètre F qui minimise $E(|e^2|)$. La présence de l'oubli va permettre d'affronter une situation non stationnaire où F varie avec n , soit $F(n)$.

III.1.2. Effet d'une fréquence qui dérive

Précédemment il s'agissait d'une fréquence pure ($\theta = \text{Cte}$) de sorte que $s(n)$ et $x(n)$ étaient des signaux stationnaires. Considérons maintenant la situation non stationnaire où θ devient variable, $\theta(n)$. Le cas de la dérive correspond à une fréquence instantanée

$$(III.1.10) \quad \theta(n) = \theta_0 + n\psi, \quad n \geq 0.$$

Le coefficient ψ sera simplement appelé « dérive ». Naturellement ce modèle ne peut s'appliquer pour tout n . En pratique $\theta(n)$ varie plutôt en dent de scie et le signal correspondant s'appelle un « chirp ». Dans ce chapitre, nous supposons que la période du chirp est assez grande pour que les algorithmes atteignent leur régime permanent. L'étude de poursuite ci-dessous s'intéresse à ce régime. Elle ne considère pas la durée de convergence nécessaire à chacun des algorithmes pour l'atteindre. Ce deuxième point est traité dans les divers chapitres de cet article qui abordent le problème des ruptures.

Alors par une intégration (faite en temps continu) de la phase (III.1.10), le signal s'écrit

$$(III.1.11) \quad s(n) = \sqrt{P_s} \exp j \left(n\theta_0 + \frac{n^2}{2} \psi + \varphi \right).$$

Dans ce cas, le prédicteur optimal (ou estimateur optimal de $s(n)$) pour du bruit i.i.d. a été calculé, [33]. Naturellement il varie au cours du temps et vaut

$$(III.1.12) \quad F(n) = \frac{1}{N} \cdot \left(1 - \frac{1}{\rho} \right) \cdot \mathcal{V}^n D$$

où

$$(III.1.13) \quad \rho = 1 + N\rho_e; \quad \rho_e = P_s/P_w,$$

$$(III.1.14) \quad \mathcal{V} = \text{diag} (\exp i\psi, \dots, \exp iN\psi),$$

et où D est un vecteur fixe dont les coordonnées sont normées et dépendent de θ_0 et de ψ . Ce filtre optimal génère l'estimation (optimale)

$$(III.1.15) \quad \tilde{s}(n) \triangleq \mathcal{F}_n[X(n-1)]$$

associée à l'erreur d'estimation (minimale)

$$(III.1.16) \quad \tilde{\eta}(n) = s(n) - \tilde{s}(n).$$

On montre aisément que

$$(III.1.17) \quad P_{\tilde{\eta}} = \min E(|\eta(n)|^2) = P_s/\rho.$$

La quantité ρ est donc le rapport signal sur bruit à la sortie de l'estimateur optimal \mathcal{F}_n . A noter que ρ ne dépend pas de n . D'après (III.1.13), on peut obtenir une valeur ρ élevée en mettant un nombre N suffisant de coefficients au prédicteur. Si, à l'entrée, le rapport signal sur bruit vérifie $P_s/P_w \gg 1$, on prendra $N = 2$. Si P_s/P_w est de l'ordre de 1, il faudra prendre $N \gg 1$. Dans tous les cas on supposera qu'à la sortie ρ est très grand devant 1.

Le cas où $P_s/P_w \ll 1$ ne nous intéresse pas car il implique un nombre N de coefficients bien trop élevé, faute de quoi ρ en sortie serait inacceptable. Ce sont d'autres méthodes qu'il faut alors appliquer pour éliminer le bruit sur $s(n)$. Par exemple, en disposant d'une source auxiliaire $w'(n)$ du bruit $w(n)$.

D'après de qui précède, notre problème peut se modéliser selon le schéma de la figure III.1.2, qui est un schéma d'identification. La partie en pointillé, qui est fictive, représente le modèle linéaire vrai qui génère $x(n)$ à partir de son passé. La voie inférieure, en trait plein, est la prédiction qui identifie ce modèle.

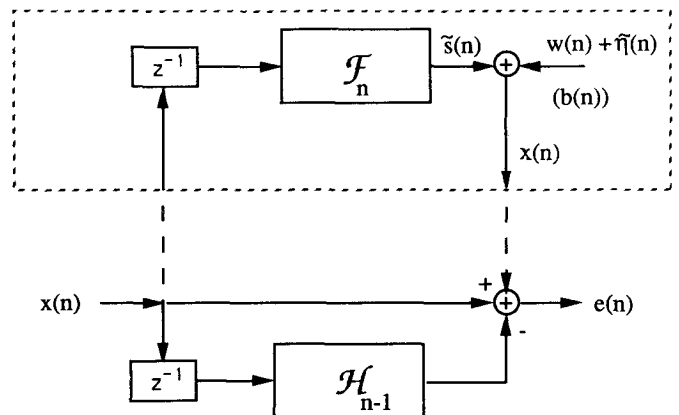


Fig. III.1.2. — Interprétation de la prédiction comme une identification. Cas non stationnaire.

L'on voit que le bruit de mesure $b(n)$ à la sortie du filtre d'identification optimal est

$$(III.1.18)$$

$$b(n) \triangleq x(n) - \mathcal{F}_n[X(n-1)] = x(n) - \tilde{s}(n),$$

$$b(n) = w(n) + \tilde{\eta}(n).$$

D'après (III.1.13) et (III.1.17)

$$(III.1.19) \quad \frac{P_{\tilde{\eta}}}{P_w} = \frac{\rho_e}{1 + N\rho_e}.$$

Donc, dès que $N \gg 1$, l'on a

$$(III.1.20) \quad P_{\tilde{\eta}} = P_w/N; \quad P_b = (1 + 1/N) P_w$$

C'est-à-dire que $\tilde{\eta}$ est négligeable devant w . Dans ces conditions, tout comme la suite $w(n)$, la suite $b(n)$ peut être considérée comme i.i.d. et indépendante des signaux $x(n-1)$, $x(n-2)$, ... à l'entrée de \mathcal{F}_n . Ce cas, caractérisé par

$$(III.1.21) \quad P_s/P_w \approx 1; \quad N \gg 1$$

est évidemment le plus intéressant : la raie et le bruit ont à peu près la même puissance et à première vue, dans $x(n)$, aucun des deux signaux ne « sort » bien. Alors que si $P_s/P_w \gg 1$, on voit déjà très bien le signal ; nul besoin de l'estimer.

Dans la suite nous nous plaçons donc dans le cas (III.1.21), avec le modèle d'identification de la figure III.1.2 et le bruit de mesure $b(n)$ supposé i.i.d.

III.1.3. Les résidus d'estimation

D'après ce qui vient d'être dit, l'on est ramené au problème d'identification non stationnaire décrit dans le chapitre I, le bruit $b(n)$ à la sortie du modèle vrai \mathcal{F}_n étant i.i.d. et indépendant des entrées $x(n-1)$, $x(n-2)$, ..., indépendant aussi des variations de \mathcal{F}_n — dues à la dérive. En régime permanent, il y a donc découplage des résidus de retard et de fluctuations, tant pour l'algorithme LMS que pour les MCRO. Il faut noter que les variations de \mathcal{F}_n sont de nature déterministe (cf. § I.5.1). D'après (III.1.12), l'incrément du filtre optimal est

$$(III.1.22) \quad \begin{aligned} T(n) &\triangleq F(n+1) - F(n) \\ &= \frac{1}{N} \left(1 - \frac{1}{\rho}\right) \mathcal{V}^n (\mathcal{V} - I) D. \end{aligned}$$

Il est facile de voir, grâce à (III.1.14), que

$$(III.1.23) \quad |T(n)| = k |\psi|, \quad k \text{ constante}$$

$$(III.1.24) \quad \text{si } |\psi| \ll 1.$$

La dernière condition est toujours satisfaite, sinon le modèle de dérive (III.1.10) n'aurait pas de sens physique. La condition (III.1.23) exprime que les variations du modèle sont d'amplitude proportionnelle à la dérive. En puissance $T(n)$ est donc en ψ^2 .

Dans ces conditions, à l'aide de l'écart à l'optimalité

$$(III.1.25) \quad V(n-1) \triangleq H(n-1) - F(n)$$

il est montré dans les articles [35] et [37] qu'en régime permanent les résidus de fluctuation

$$\mathcal{M}_f \triangleq \lim_{n \rightarrow \infty} E(|V_f(n-1)|^2) / P_b$$

et de retard

$$\mathcal{M}_r \triangleq \lim_{n \rightarrow \infty} E(|V_r(n-1)|^2) / P_b$$

sont respectivement donnés pour les algorithmes LMS et MCRO par le tableau suivant :

TABLEAU I
Les différents résidus

Algorithme	LMS	MCRO
\mathcal{M}_f (fluctuation)	$\frac{N}{2} \nu$ (III.1.26)	$\frac{N}{2} \beta$ (III.1.27)
\mathcal{M}_r (retard)	$\frac{N}{12} (1 - \gamma)(1 + \rho_e)^2 \frac{\psi^2}{\nu^2}$ (III.1.28)	$\frac{N(N+1)}{4} \rho_e \frac{\psi^2}{\beta^2}$ (III.1.29)

Dans ce tableau on a utilisé les taux d'oubli définis au chapitre I :

$$(III.1.30)$$

$$\text{LMS : } \nu = \mu (P_s + P_w); \quad \text{MCRO : } \beta = 1 - \lambda$$

et le coefficient γ vaut

$$(III.1.31) \quad \gamma = \frac{1}{N} - \frac{3(N+1)}{\rho_e N^2} < 1.$$

La première ligne de ce tableau montre que les fluctuations sont proportionnelles au taux d'oubli, tant pour les MCRO que pour le LMS. C'est là un résultat classique [11], [20]. Elles sont aussi proportionnelles à la puissance P_b du bruit de mesure et enfin au nombre N de coefficients du filtre. Tout cela est déjà connu.

La deuxième ligne du tableau montre que le retard est inversement proportionnel au carré du taux d'oubli. C'est là un comportement typique de variations déterministes. On retrouve dans notre cas particulier une forme déjà trouvée en [10], [5]. Le retard est aussi proportionnel à ψ^2 , ce qui est normal vu que les variations de $F(n)$ ont leur amplitude en ψ . Ce qui est plus remarquable, c'est que le retard du LMS soit proportionnel à N alors que celui des MCRO est proportionnel à $N(N+1)$. Cela est évidemment à l'avantage du LMS dont le retard croît moins vite avec le nombre de coefficients.

Validité du calcul

Les résultats du tableau I ont été obtenus sous deux hypothèses auxiliaires (outre l'hypothèse que $b(n)$ est i.i.d.).

Adaptation lente (AL) : le taux d'oubli est faible, ce qui s'écrit

$$(III.1.32) \quad (LMS) : N\nu \ll 2 ; \quad (MCRO) : N\beta \ll 2 .$$

Cette condition sera interprétée dans le prochain sous-paragraphe.

Variations lentes (VL) : les variations du modèle optimal sont encore plus lentes que l'adaptation, ce qui s'écrit

$$(LMS) : N\psi \ll \nu / (1 + \rho_e) ; \quad (MCRO) : N\psi \ll \beta .$$

Les simulations ont montré que cette condition n'est pas du tout critique et que les résultats numériques

tirés de l'analyse théorique restent conformes aux simulations même si (VL) n'est pas vérifiée. Dans la suite nous omettons cette condition.

III.1.4. Optimisation de la poursuite

Pour l'algorithme LMS, le tableau 1 montre que les résidus $\mathcal{M}_f^{LMS}(\nu)$ et $\mathcal{M}_r^{LMS}(\nu)$ varient en sens opposé. Il existe donc un taux d'oubli optimal ν_{opt} , qui minimise le résidu total somme des deux résidus (voir chap. I.8.3) par un compromis entre les deux contributions. De la même manière pour les MCRO, le taux d'oubli β_{opt} correspondra au minimum \mathcal{M}_{min}^{MCR} . Ces grandeurs sont faciles à calculer et données dans le tableau 2.

TABEAU 2

Le compromis entre fluctuations et retard

Algorithme	LMS	MCRO
Taux d'oubli optimal	$\nu_{opt} = \left[\frac{(1 - \gamma)(1 + \rho_e)^2 \psi^2}{3} \right]^{1/3}$ (III.1.33)	$\beta_{opt} = [(N + 1) \rho_e \psi^2]^{1/3}$ (III.1.34)
Résidu minimum \mathcal{M}_{min}	$\frac{3 N}{4} \nu_{opt}$ (III.1.35)	$\frac{3 N}{4} \beta_{opt}$ (III.1.36)

L'optimisation de la poursuite consiste donc à rechercher le taux d'oubli optimal ν_{opt} ou β_{opt} .

Les quatre quantités du tableau sont des fonctions croissantes de la dérive ψ . Elles sont en $\psi^{2/3}$ en conséquence du caractère déterministe des variations du modèle ; et l'on retrouve un tel comportement dans les chapitres II.2 et V de cet article. Pour un modèle stochastique et centré des variations $T(n)$ du filtre optimal, on n'aurait pas trouvé cet exposant $2/3$. Les variations déterministes sont en fait un type sévère de variations car la pente à l'origine de ν_{opt} et β_{opt} est infinie à l'origine : pour une faible augmentation de ψ , ν_{opt} et β_{opt} doivent être beaucoup augmentés. Et même en ce faisant, le résultat \mathcal{M}_{min} de l'optimisation augmente aussi beaucoup. A cet égard les deux algorithmes ont le même comportement.

Pour que les résultats du tableau 2 soient valables, il faut que le taux d'oubli optimum vérifie la condition (AL) en (III.1.32). En y introduisant les valeurs données par (III.1.33) et (III.1.34), on trouve une borne supérieure de la dérive, à savoir

$$(III.1.37)$$

$$LMS : |\psi| \ll \psi_{LMS} = \frac{1}{1 + \rho_e} \sqrt{\frac{24}{(1 - \gamma) N^3}}$$

$$(III.1.38)$$

$$MCRO : |\psi| \ll \psi_{MCR} = \sqrt{\frac{8}{\rho_e N^3 (N + 1)}}$$

On remarque la relation

$$(III.1.39) \quad \left(\frac{\psi_{MCR}}{\psi_{LMS}} \right)^2 = \left(\frac{\mathcal{M}_{min}^{LMS}}{\mathcal{M}_{min}^{MCR}} \right)^3$$

qui montre que pour l'algorithme le plus performant (\mathcal{M}_{min} le plus faible), le domaine de validité de la théorie est aussi le plus large (la borne supérieure de ψ est la plus grande). Cela est naturel.

Pour interpréter la condition (AL), rappelons que les résidus sont des erreurs d'estimation de sortie additionnelles, qui s'ajoutent à l'erreur d'estimation du filtre optimal. De la sorte, d'après (III.1.20)

$$(III.1.40) \quad E (|s(n) - H(n - 1)^T X(n - 1)|^2) = P_{\tilde{n}} + \mathcal{M}_{min} P_b ,$$

à condition que le taux d'oubli soit optimisé. Or l'estimateur trivial du signal $s(n)$ selon

$$(III.1.41) \quad \tilde{s}(n) = x(n)$$

fait une erreur

$$(III.1.42) \quad E (|s(n) - \tilde{s}(n)|^2) = P_w .$$

Pour que le système sophistiqué d'estimation de $s(n)$ basé sur la prédiction adaptative ait quelque intérêt, il faut évidemment que l'erreur (III.1.40) qu'il fournit en sortie soit bien plus faible que l'erreur (III.1.42) du système trivial. Sinon il n'y a

aucun intérêt à faire un système compliqué (surtout dans le cas des MCRO). Ceci implique que

$$(III.1.43) \quad P_b \mathcal{M}_{\min} \ll P_w.$$

Pour l'algorithme LMS il vient d'après (III.1.35)

$$(III.1.44) \quad \frac{3N}{4} P_b \nu_{\text{opt}} \ll P_w$$

et, en tenant compte de la valeur de P_b en (III.1.20)

$$(III.1.45) \quad \frac{3}{2} (N+1) \nu_{\text{opt}} \ll 2.$$

Cette condition ne diffère pas sensiblement de la condition d'adaptation lente ($N\nu \ll 2$). De la même manière pour l'algorithme MCRO, la condition (III.1.43) ne diffère pas sensiblement de $N\beta \ll 2$ qui est la condition d'adaptation lente.

C'est là un résultat très important : *pour le problème de poursuite, violer la condition d'adaptation lente, c'est mettre en œuvre un algorithme adaptatif compliqué qui n'apportera pas d'amélioration par rapport à l'estimation triviale $\tilde{x}(n) = x(n)$. Donc l'adaptation lente n'est pas une condition technique pour pouvoir mener les calculs à bien, mais une condition physique pour que le système adaptatif ait un intérêt pratique.*

Dans la suite on supposera donc que ψ est limité selon (III.1.37) ou (III.1.38).

III.1.5. Comparaison du LMS et des MCRO

Tout d'abord, comparons les deux algorithmes à niveau donné des résidus de fluctuations, soit, d'après (III.1.26) et (III.1.27) pour

$$(III.1.46) \quad \nu = \beta.$$

En appelant g_r le rapport des résidus-retard il vient, d'après (III.1.28) et (III.1.29)

$$(III.1.47) \quad g_r \triangleq \mathcal{M}_r^{\text{LMS}} / \mathcal{M}_r^{\text{MCR}} = \frac{1-\gamma}{3(N+1)} (1+\rho_e)(1+1/\rho_e).$$

Il est facile de voir que si le rapport signal sur bruit de sortie $\rho = N\rho_e$ est nettement supérieur à 3 et si $N \geq 2$, alors

$$(III.1.48) \quad g_r \leq 1 \Leftrightarrow \rho_e \leq (3N^2 + N + 2)/(N-1) \approx 3N.$$

Autrement dit le retard du LMS est plus faible que celui des MCRO — à fluctuations égales — si ρ_e est inférieur à $3N$. Cette situation est la plus normale, ρ_e étant par exemple voisin de 1. On peut donc dire que le LMS poursuit mieux que les MCRO.

Ce point est confirmé si l'on adopte la méthodologie proposée au chapitre I.8.3 de comparer les algorithmes après que chacun d'eux ait été optimisé, c'est-à-dire de comparer les valeurs \mathcal{M}_{\min} correspon-

dantes. On voit tout de suite que le gain du LMS sur les MCRO, défini comme

$$(III.1.49) \quad g \triangleq \mathcal{M}_{\min}^{\text{LMS}} / \mathcal{M}_{\min}^{\text{MCR}}$$

vaut

$$(III.1.50) \quad g = \left[\frac{1-\gamma}{3(N+1)} (1+\rho_e)(1+1/\rho_e) \right]^{1/3} = [g_r]^{1/3}.$$

Donc, comme g_r , g est inférieur à 1 dans une situation pratique réaliste. C'est une autre manière de voir la supériorité du LMS sur les MCRO.

Il faut remarquer que g est indépendant de la dérive ψ . Contrairement à une conviction courante, *ce n'est pas parce que les variations sont plus rapides ($|\psi|$ élevé) que les MCRO sont mieux adaptés. La conclusion sur la supériorité de l'un ou l'autre algorithme dépend uniquement du nombre N de prises et du rapport signal sur bruit ρ_e en entrée, mais non de la vitesse des variations.*

III.1.6. Résultats de simulation

Nous avons confirmé par simulation les résultats théoriques précédents, car ils avaient de quoi surprendre. Les caractéristiques de la simulation sont les suivantes :

- fréquence angulaire de base : $\theta_0 = 0,39$
- dérivé : $\psi = 0$; $1,5 \cdot 10^{-5}$; 10^{-4} ; 10^{-3} ; 10^{-2} ; 10^{-1}
- puissance du signal : $P_s = 2$
- puissance du bruit : $P_w = 2(\rho_e = 1)$
- nombre de prises : $N = 16$.

La mise en œuvre du LMS est standard. Celle des MCRO a été faite avec la formule (non rapide) (I.6.6). Ceci afin d'éviter les problèmes d'instabilité numérique. Les résidus, appelés $\mathcal{M}(n)$, sont estimés par moyennage sur une fenêtre temporelle de 50 et en itérant de 3 à 5 passages. Cette procédure convient bien pour le régime permanent qui seul nous préoccupe. La valeur moyenne \mathcal{M} s'obtient par moyennage des 500 à 1000 dernières itérations. Tous les $\mathcal{M}(n)$ sont indiqués en décibels et normalisés par rapport au bruit de sortie P_b . Donc, lorsque \mathcal{M} dépasse 0 dB, le système adaptatif est sans intérêt (voit (III.1.43)).

Parmi nos nombreuses simulations qui ont toutes confirmé la théorie, nous choisissons quelques exemples.

La figure III.1.3 (resp. III.1.4) concerne le LMS (resp. les MCRO). La figure III.1.3 montre l'existence d'un taux d'oubli optimal : quand ν varie de $\nu_{\text{opt}}/4$ à $2\nu_{\text{opt}}$ en passant par la valeur théoriquement optimale ν_{opt} , le résidu asymptotique décroît puis se remet à augmenter ; ν_{opt} est bien la valeur de compromis qui minimise $\mathcal{M}_{\min}^{\text{LMS}}$. De même pour $\mathcal{M}_{\min}^{\text{MCR}}$ sur la figure III.1.4. Enfin la figure III.1.5 illustre les courbes $\mathcal{M}^{\text{LMS}}(\nu)$ et $\mathcal{M}^{\text{MCR}}(\beta)$ avec le minimum pour $\nu = \nu_{\text{opt}}$ ou $\beta = \beta_{\text{opt}}$. Avec les valeurs choisies de ρ_e et de N , la théorie prédit 4 dB de gain sur \mathcal{M}_{\min} en faveur du LMS et c'est bien ce que l'on obtient.

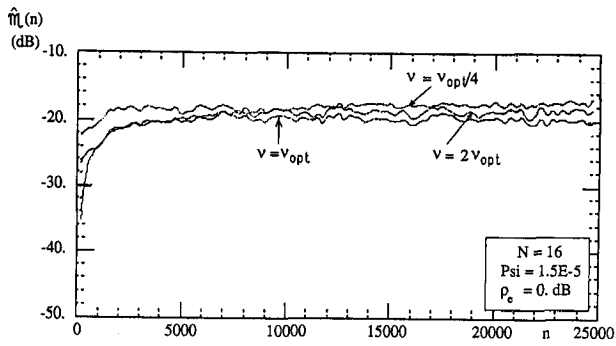


Fig. III.1.3. — Existence d'un taux d'oubli optimal en poursuite pour le LMS.

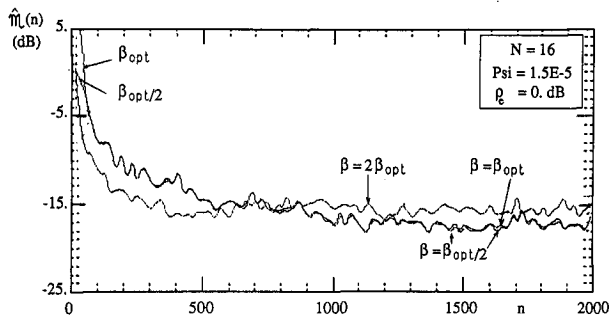


Fig. III.1.4. — Existence d'un taux d'oubli optimal en poursuite pour les MCRO.

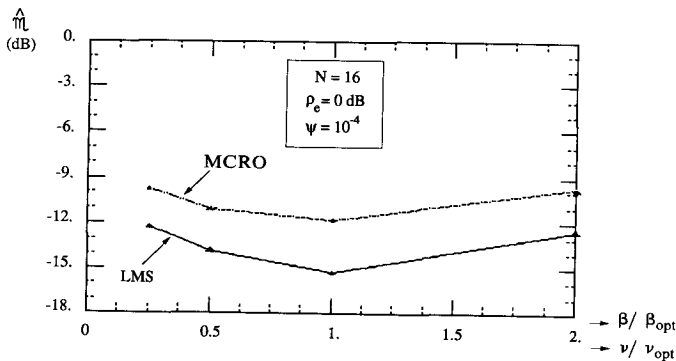


Fig. III.1.5. — Le compromis entre retard et fluctuations.

Pour toutes les autres valeurs de ψ il en a été de même : gain de 3,5 dB en faveur du LMS. Il en était encore ainsi avec $\psi = 10^{-2}$ qui commence à approcher de la limite d'adaptation lente, particulièrement pour les MCRO (voir (III.1.38)). Dans ce cas, la seule différence par rapport à la théorie est que le taux d'oubli estimé $\hat{\nu}_{opt}$ (resp. $\hat{\beta}_{opt}$) devient notablement inférieur à la valeur théorique ν_{opt} (resp. β_{opt}),

un phénomène qui s'explique aisément (voir [36]). Mais les 3,5 dB de gain du LMS sur les MCRO sont respectés.

III.1.7. Conclusion

Nous venons d'étudier un problème de poursuite par prédiction d'une sinusoïde bruitée qui dérive, où le prédicteur optimal non stationnaire a des variations déterministes. Ce problème s'apparente à de l'identification inverse. Il a seulement été considéré en régime permanent de sorte que la rapidité bien connue de la phase transitoire des MCRO n'a pas été prise en compte comme un critère de préférence sur le LMS. Le cas le plus difficile à traiter est celui où ni le bruit, ni la sinusoïde ne domine, avec un rapport signal sur bruit d'entrée ρ_e voisin de 1. Dans ce cas, si le prédicteur a assez de prises pour bien restituer la sinusoïde, nous avons démontré théoriquement et par simulations que le LMS poursuit mieux que les MCRO, avec 3 ou 4 dB de gain sur le résidu, même optimisé. Ce résultat contredit l'idée courante qui donne les MCRO meilleurs en poursuite parce qu'ils convergent plus vite que le LMS. Cette idée est fautive. Comme le LMS est aussi beaucoup plus simple à mettre en œuvre, il faut le préférer pour une application pratique. Ce résultat est indépendant du niveau de la dérive et il n'est pas vrai que pour de très fortes dérives les MCRO vont redevenir supérieurs. On peut aisément comprendre la raison de cette conclusion : puisque signal et bruit ont à peu près le même niveau, la matrice de corrélation de l'entrée est peu dispersée et l'outil très élaboré que constituent les MCRO n'a pas son utilité.

On peut alors penser que la situation se renversera si l'on augmente le rapport signal sur bruit d'entrée de sorte que la covariance matricielle de l'entrée soit fortement dispersée. Le cas a été traité dans [15], par exemple avec $\rho_e = 32$ et $N = 2$. Naturellement, le prédicteur correspondant a peu de prises puisque la sinusoïde, peu bruitée, ne nécessite pas d'être très rehaussée. Si la théorie donne alors les MCRO très légèrement supérieurs au LMS (en principe 1,8 dB de gain sur \mathcal{M}_{min}), les simulations vont à l'inverse : le LMS n'est jamais moins bon que les MCRO et il peut donner jusqu'à 3 dB de gain sur les MCRO lorsque la dérive atteint sa borne supérieure au-delà de laquelle le système adaptatif devient déraisonnable car il fait plus d'erreur que l'estimateur trivial qui estime le signal à la valeur reçue.

La conclusion est donc générale : pour ce problème de prédiction en présence de bruit blanc, en régime permanent il faut rejeter les moindres carrés, même pour des variations très rapides, poursuivent moins bien que le gradient, lequel est de surcroît beaucoup plus simple.

Chapitre III PRÉDICTION DE FRÉQUENCES

2. PRÉDICTION TRANSVERSE D'UN NOMBRE VARIABLE DE SINUSOÏDES

Catherine LAMBERT-NEBOUT, Francis CASTANIÉ

Centre National d'Etudes Spatiales (CNES), Laboratoire d'Electronique (ENSEEIH)

III.2.1. Introduction

Le système d'aide à l'atterrissage ILS (Instrument Landing System) utilise pour sa partie radioalignement de piste une bande de fréquence VHF, qui est très vulnérable aux perturbations causées notamment par les stations de radiodiffusion FM et les équipements industriels. Il est important d'avoir une connaissance sur ces brouilleurs externes au moins lors des contrôles en vol.

C'est pourquoi nous nous intéressons au signal ILS (radiolignement de piste « ou localizer ») mis en cause, observé au niveau du récepteur de bord après l'étage d'entrée (signal ILS composite).

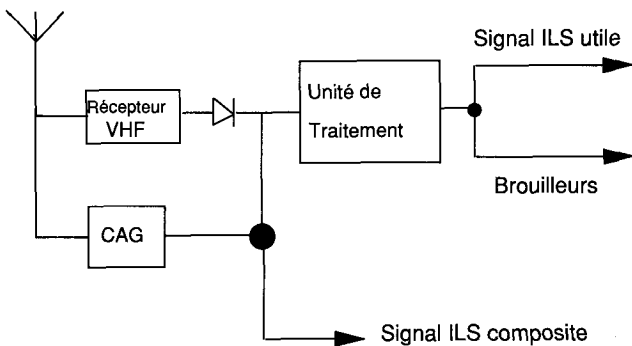


Fig. III.2.1. — Schéma de traitement.

Le signal ILS utile est composé de 2 sinusoïdes de fréquences respectives 90 Hz et 150 Hz, auxquelles se rajoute de façon intermittente (si l'indicatif de l'aéroport est émis en code morse) un troisième sinusoïde à la fréquence de 1 020 Hz. C'est la différence d'amplitude entre la sinusoïde à 90 Hz et celle à 150 Hz qui permet au pilote de savoir si l'avion se trouve à droite ou à gauche de l'axe de la piste.

L'objectif visé est de dissocier les brouilleurs éventuels du système, du signal utile (en vue d'une classification des brouilleurs). Il s'agit donc d'éliminer les 2 (ou 3) sinusoïdes le plus parfaitement possible (la fréquence d'échantillonnage est fixée à 2,5 kHz), sachant que deux problèmes différents sont posés :

— NS1 : présence de 2 sinusoïdes seulement avec dérive des amplitudes (le pilote cherche à positionner l'avion sur l'axe de la piste), cette dérive est due aux

mouvements transversaux de l'avion par rapport à l'axe ;

— NS2 : Situation de rupture, apparition ou disparition brusque d'une troisième sinusoïde (indicatif de l'aéroport).

Une structure appropriée est la structure (cf. fig. III.2.2) du filtre réjecteur adaptatif (encore nommé ALE : Adaptive Line Enhancer) [41].

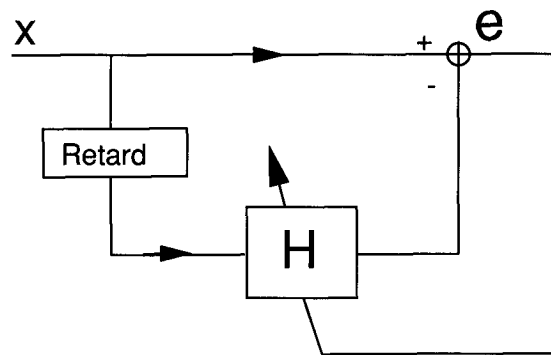


Fig. III.2.2. — Schéma bloc de l'ALE.

Cette structure est équivalente à celle introduite par un problème de prédiction (cf. fig. III.2.3) si le retard est fixé à 1 :

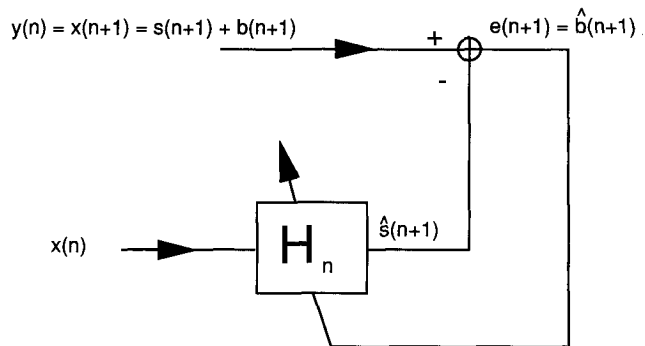


Fig. III.2.3. — Problème de prédiction.

Dans la figure ci-dessus, $x(n)$ est le signal observé, $s(n)$ représente le signal utile ILS, $b(n)$ représente le brouilleur et $e(n)$ est une estimation du brouilleur. La structure choisie pour le filtre adapté est une

structure **transverse**. Les algorithmes de poursuite implantés sont les Moindres Carrés Récursifs avec Oubli (MCRO) [42] et l'algorithme LMS [41].

Les deux types de non stationnarités (NS1 et NS2) étudiées appartiennent au même groupe de non stationnarités : $x(n)$ est non stationnaire (§ I.2.1 c) et les éléments variables avec n sont générés par un modèle aléatoire (§ I.2.2 b, classe A). Par contre, les vitesses des deux non stationnarités sont différentes : la première (NS1) est lente alors que la seconde (NS2) correspond à une rupture.

Ces deux types de non stationnarités, auxquels nous sommes confrontés feront donc l'objet des deux chapitres suivants.

III.2.2. Situation stationnaire

La situation de poursuite (NS1) peut être considérée comme localement stationnaire vue la dérive des amplitudes des sinusoïdes lors de l'approche finale de l'avion (sa vitesse moyenne est alors de 200 km/h et la fréquence d'échantillonnage est fixée à 2,5 kHz).

Comparons la variance des coefficients du filtre adaptatif en régime permanent à résidu donné :

Soit N l'ordre du filtre transversal RIF,

$$(III.2.1) \quad e(n+1) = x(n+1) - \underline{H}^T(n) \underline{X}(n)$$

$$\underline{H}(n) = (h_1(n), \dots, h_N(n))^T$$

$$\underline{X}(n) = (x(n), \dots, x(n-N+1))^T$$

Soit H^{opt} le filtre optimal de Wiener solution de :

$$(III.2.2) \quad \underline{H}^{opt} = R^{-1} \underline{P}$$

$$\underline{H}^{opt} = (h_1^{opt}, \dots, h_N^{opt})^T$$

\underline{P} vecteur d'intercorrélation de $\underline{X}(k)$ et $x(k+1)$ de dimension N

$$\underline{P} = (P_i) = (r_x(i+1))$$

R matrice d'autocorrélation de $x(k)$ de dimension (N, N)

$$R = (R_{ij}) = (r_x(i-j))$$

Nous définissons le vecteur erreur $\tilde{H}(n)$ suivant :

$$(III.2.3) \quad \tilde{H}(n) = \underline{H}(n) - \underline{H}^{opt}$$

Widrow [41] pour le LMS et Ljung [42] pour le MCR ont exprimé la matrice de covariance de $\tilde{H}(\infty)$:

	LMS	MCRO
$E[\tilde{H}(\infty) \tilde{H}^T(\infty)]$	$\frac{\mu}{2} E_{min} I \quad (1)$	$\frac{\beta}{2} E_{min} R^{-1} \quad (2)$

μ : pas d'incrément

β : « taux d'oubli », $\beta = 1 - \lambda$ (λ étant le facteur d'oubli)

I : matrice identité de dimension (N, N)

$$(III.2.4) \quad E_{min} = E[e^2(n) \text{ pour } \underline{H}(n-1) = \underline{H}^{opt}]$$

$$= E[x^2(n+1)] - \underline{H}^{optT} \underline{P}$$

μ sera choisi de façon à ce que les puissances moyennes de l'erreur résiduelle à l'infini soient équivalentes pour les deux algorithmes : LMS et MCRO.

	LMS	MCRO
$E[e^2(\infty)]$	$E_{min} \left(1 + \frac{\mu}{2} \text{tr}(R) \right)$	$E_{min} \left(1 + N \frac{\beta}{2} \right)$

$$\text{tr}(R) = \text{trace}(R) = N \times r(0)$$

d'où la valeur prise par μ :

$$(III.2.5) \quad \mu = \frac{\beta N}{\text{tr}(R)}$$

ou encore :

$$\mu P_x = \beta$$

Remarque : La valeur de la matrice de covariance des erreurs sur les coefficients du filtre dans le cas du gradient est toujours obtenue dans la littérature en faisant une hypothèse sur la « non-corrélation » des échantillons d'entrée. Dans le cas, qui est le nôtre, d'une entrée sinusoïdale, cette valeur est obtenue grâce à une hypothèse de rapport Signal sur Bruit faible. Sans cette hypothèse, la complexité des calculs est telle qu'aucune écriture utilisable de la matrice de covariance ne peut être établie.

Nous conservons néanmoins cette valeur car, sous hypothèse de fort rapport Signal sur bruit, elle semble correspondre à une borne supérieure.

Plaçons-nous dans le cas où le signal d'entrée est une sinusoïde $s(n)$ noyée dans du bruit gaussien $b(n)$ de moyenne nulle et de variance σ^2 .

$$(III.2.6) \quad s(n) = \sqrt{2} \cos(2\pi\tilde{f}_0 n + \theta)$$

$$x(n) = s(n) + b(n)$$

Le rapport Signal sur Bruit (S/B) noté γ est alors égal à :

$$\gamma = \frac{1}{\sigma^2}$$

Zeidler [48] a montré en posant $w_0 = 2\pi\tilde{f}_0$ que :

$$(III.2.7) \quad h_i^{opt} = \frac{1}{C} (\Phi \exp[jw_0 i] + \Phi^* \exp[-jw_0 i])$$

$$\Phi = B \exp[jw_0] - \exp[-jw_0(N)] \frac{\sin(Nw_0)}{\sin w_0}$$

$$C = B^2 - \frac{\sin^2(Nw_0)}{\sin^2 w_0}$$

avec

$$B = N + \frac{2}{\gamma}$$

et si de plus $Nw_0 = k\pi$ soit :

$$N = \frac{k}{2\tilde{f}_0} \quad \text{avec } k \text{ entier}$$

nous obtenons :

$$(III.2.8) \quad h_i^{opt} = \frac{\gamma}{1 + \frac{\gamma}{2} N} \cos(w_0 i)$$

Sachant que R s'écrit sous la forme :

$$R = \frac{1}{\gamma} \left(I + \frac{\gamma}{2} V^* V^T + \frac{\gamma}{2} V V^*{}^T \right)$$

avec $V = (1, \exp[jw_0], \dots, \exp[jw_0(N-1)])^T$.

En utilisant deux fois le lemme d'inversion de matrice et sachant que $\cos(2w_0 N) = 1$ nous obtenons une expression simple pour la matrice inverse de R :

$$R^{-1} = \gamma \left[I - \frac{\frac{\gamma}{2} [V V^*{}^T + V^* V^T]}{1 + \frac{\gamma}{2} N} \right]$$

Soit $(R^{-1})_{ii}$ -i^{ème} élément de la diagonale de R^{-1} , pour tout $i = 1, \dots, N-1$:

$$(III.2.9) \quad (R^{-1})_{ii} = \gamma \left(1 - \frac{\gamma}{1 + \frac{\gamma}{2} N} \right)$$

de plus $\text{trace}(R) = \left(1 + \frac{1}{\gamma}\right) \times N$ et d'après l'équation (III.2.4) [45] :

$$(III.2.10) \quad E_{\min} = \gamma^{-1} + \frac{1}{1 + \frac{\gamma}{2} N}$$

Afin d'étudier la variance engendrée par l'erreur en sortie sur les coefficients, nous supposons que la puissance de $e(n)$ est la même pour les deux algorithmes, i.e. d'après l'équation (III.2.5) :

$$(III.2.11) \quad \mu = \beta \frac{\gamma}{(1 + \gamma)}$$

il s'agit alors de comparer les termes diagonaux des 2 matrices de covariance, soit μ avec βR^{-1} , soit encore :

$$\frac{1}{1 + \gamma} \cong \frac{1 + \gamma(N/2 - 1)}{1 + \frac{\gamma}{2} N}$$

Il apparaît donc que la variance des coefficients calculés par le MCR est toujours supérieure à celle calculée par le gradient dès que $N > 2$, d'autant plus que pour de forts rapports S/B, nous avons vérifié par simulation que la valeur μE_{\min} est plus grande de 1 à 2 ordres de grandeur que la variance réelle des coefficients.

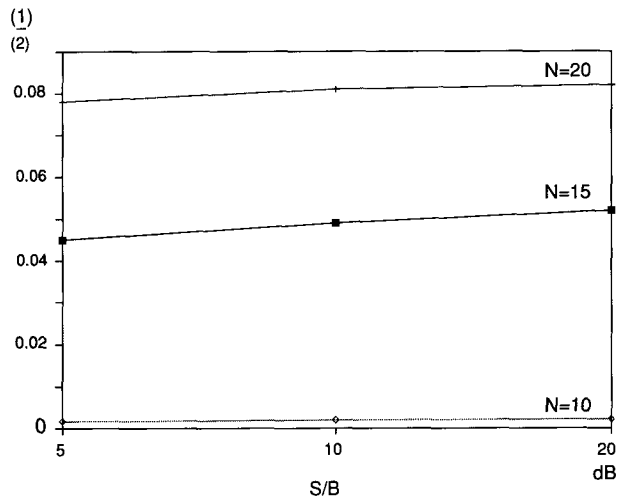


Fig. III.2.4. — Calcul du rapport de la variance des coefficients calculée par le gradient(1) sur celle calculée par les MCR(2) par simulation sur 200 réalisations de bruit. N : ordre du filtre transversal, $\tilde{f}_0 = 0,1$.

A β donné, la variance des coefficients obtenus avec l'algorithme des moindres carrés est constante pour une longueur de filtre N fixée (supérieure à 10) et pour des rapports Signal sur Bruit supérieurs à 5 dB. Ceci se vérifie en théorie à l'aide des équations (III.2.9) et (III.2.10) :

$$\beta E_{\min}(R^{-1})_{ii} = \beta \left(1 - \frac{\gamma^2}{\left(1 + \frac{\gamma}{2} N\right)^2} \right)$$

Dès que $(\gamma/2) N \gg 1$ (d'où les conditions sur la longueur du filtre et le rapport Signal sur Bruit puisque $\frac{\gamma}{2} N$ représente le rapport Signal sur Bruit en sortie du filtre prédicteur optimal), l'expression de la variance des coefficients s'approxime par :

$$\approx \beta \left(1 - \frac{4}{N^2} \right)$$

et n'est fonction que de la longueur du filtre.

Ces résultats obtenus dans un cas simplifié (une sinusoïde dont la fréquence est un multiple entier de la longueur du filtre) ont été vérifiés par simulation pour une fréquence quelconque et dans le cas d'un signal constitué de plusieurs sinusoïdes. **Ceci démontre une meilleure performance de l'algorithme du LMS lorsque les valeurs exactes des coefficients du filtre sont recherchées.**

III.2.3. Situation de rupture : application à PLS

L'utilisation du Gradient dans notre application n'est pas optimale dans la mesure où le signal est constitué soit de deux sinusoïdes, soit de trois sinusoïdes (lorsque le signal d'identification « 1 020 » de l'aéroport est émis). La troisième sinusoïde étant présente sur un maximum de 250 à 400

échantillons (pour une fréquence d'échantillonnage de 2,5 kHz), cela ne permet pas la convergence de l'algorithme du gradient.

Nous opterons donc pour un algorithme des moindres carrés rapides avec facteur d'oubli.

La présence de ce facteur d'oubli nous a confrontés au problème de divergence numérique des algorithmes MCR rapides (qui est l'un des problèmes actuels lors de l'utilisation de tels algorithmes). Plusieurs méthodes ont été testées pour combattre cette instabilité numérique :

- utilisation d'une constante de stabilisation [43] ;
- régularisation : contrôle des prédicteurs et de la variance de l'erreur de prédiction [44] ;
- stabilisation numérique de deux variables : variance de l'erreur de prédiction retour et du prédicteur retour [25].

Cette dernière méthode se distingue par une stabilisation constante sur un grand nombre d'échantillons, elle augmente de manière négligeable la complexité et ne dégrade pas le taux de réjection. Elle permet de stabiliser l'algorithme de Kalman rapide [42] ($\mathcal{O}(10 N)$ opérations par récursion), l'algorithme FTF [46] ($\mathcal{O}(7 N)$ opérations par récursion) et par conséquent l'algorithme FAEST [47] (de même complexité de calcul et ne différant que par le calcul de la variable de vraisemblance). La stabilisation est effectuée par l'introduction supplémentaire en début de récursion de seulement 3 variables.

Nous remarquons (fig. III.2.5 B et C) que le MCRO permet une convergence suffisamment rapide en présence du caractère morse à « 1 020 Hz » et donc en assure une bonne élimination, tandis que le Gradient n'a pas le temps de converger.

III.2.4. Conclusion

L'implantation d'une structure adaptative ALE pour la réjection du signal utile ILS, constitué de 2 ou 3 sinusoides bruitées par des parasites industriels ou des émetteurs de radio-diffusion en modulation de fréquence, nous a permis de montrer que l'algorithme LMS engendre une variance sur les coefficients du filtre bien inférieure à celle introduite par l'algorithme MCRO. Mais dans notre application, le phénomène de rupture lors de l'émission de l'indicatif de l'aéroport nous contraint, pour des besoins de rapidité de convergence, à opter pour un algorithme des Moindres Carrés Rapides avec Oubli et muni d'un procédé de stabilisation de faible complexité.

Fig. III.2.5. — Situation de rupture.

Fig. A : signal ILS brouillé (pas de changement d'axe sur le nombre d'échantillons visualisé).

Erreur en sortie de l'ALE ($e(n)$):

Fig. B : avec le LMS $\mu = 0,004 N = 70$

Fig. C : avec le MCRO (FTF) stabilisé $\lambda = 0,998$ (soit $\beta = 0,002 = \frac{\mu}{2}$) $N = 70$.

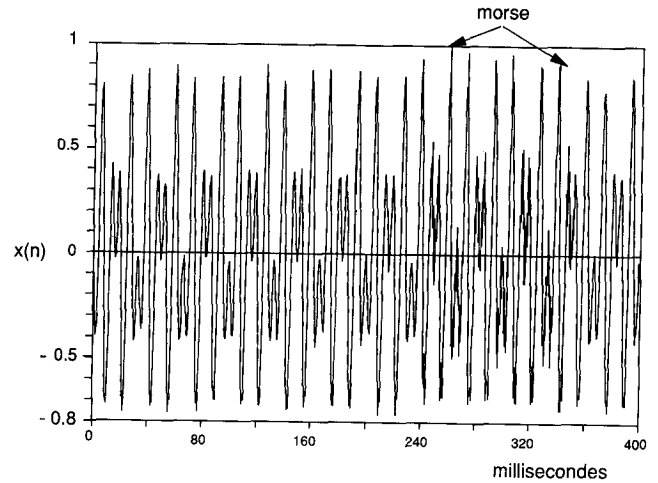


Fig. A

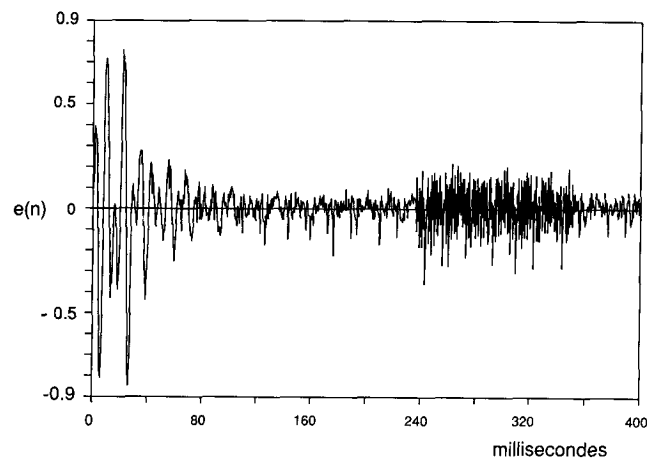


Fig. B

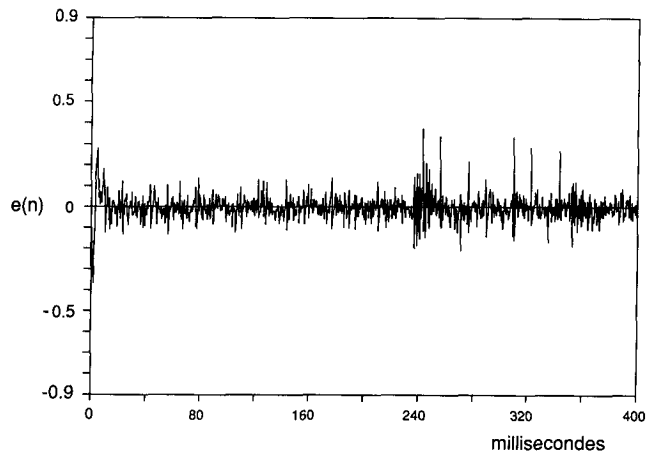


Fig. C

Chapitre IV

PRÉDICTION EN TREILLIS AVEC RUPTURES DE MODÈLES

R. SETTINERI*, G. FAVIER*, J. BOUDY**

* I3S/LASSY, URA 1376 du CNRS, Université de Nice-Sophia Antipolis

** MATRA-Communication

IV.1. Introduction

Les filtres en treillis adaptatifs sont très utilisés dans les applications de traitement du signal. Dans la plupart des situations pratiques, le signal à traiter est non stationnaire et doit être modélisé à l'aide d'un modèle variant dans le temps. En présence d'évolutions lentes des paramètres du modèle, les filtres en treillis/moindres carrés récursifs (MCR) ou treillis/gradient (LMS) permettent d'obtenir un bon compromis entre une faible sensibilité au bruit et une vitesse de convergence rapide vis-à-vis des non-stationnarités. Cependant, en présence de non-stationnarités fortes du signal, correspondant à des ruptures de modèle, les performances de ces filtres sont grandement affectées. Dans ce chapitre nous comparons deux nouvelles solutions pour le filtrage de signaux non-stationnaires qui sont caractérisés par des évolutions rapides et déterministes des coefficients du modèle générateur :

— La première solution consiste à optimiser les taux d'oubli du filtre en treillis/LMS [49], [50]. Cette optimisation est réalisée de telle sorte que l'écart par rapport à la variance (résidu de fluctuation normalisé) de la sortie optimale de chaque étage du filtre soit minimisé. Cette minimisation conduit à la résolution d'une équation du troisième degré qui dépend uniquement du degré de non-stationnarité du système, comme l'a introduit [55] dans le cas des filtres transverses.

— La deuxième solution est basée sur l'utilisation d'un détecteur de ruptures de modèles, combiné à un algorithme de covariance à fenêtre glissante normalisé (CFGN, en anglais NSWC) qui est caractérisé par une longueur de fenêtre adaptative [53], [57]. Deux longueurs de fenêtre différentes sont utilisées. Pendant les intervalles de temps où le signal est stationnaire, le filtrage est réalisé à l'aide de l'algorithme CFGN fonctionnant avec une fenêtre fongue (w_1). Quand une non-stationnarité est détectée, la longueur de la fenêtre est récursivement décrétementée jusqu'à une fenêtre courte (w_c) au moyen de l'algorithme de covariance à fenêtre décroissante normalisé (CFDN, en anglais NDMC). Ceci permet d'améliorer de façon significative la capacité de poursuite de l'algorithme des moindres carrés. Une fois la non-stationnarité terminée la longueur de la fenêtre est récursivement incrémentée jusqu'à la valeur w_1 au moyen de l'algorithme de covariance à fenêtre croissante normalisé (CFCN, en anglais NGMC).

Dans les paragraphes suivants, nous rappelons brièvement les équations des filtres en treillis/LMS optimisé et moindres carrés à fenêtre glissante adaptative. Le schéma de principe de ces deux solutions est donné. Puis une comparaison des performances obtenues à l'aide de ces deux filtres en treillis adaptatifs est réalisée à travers une analyse de type Monte Carlo.

IV.2. Filtres en treillis/LMS

La structure standard d'un filtre en treillis d'ordre N est décrite à l'aide des équations suivantes ($j = 1$ à N) :

$$(IV.1) \quad x_j(n) = x_{j-1}(n) - k_j^u(n) u_{j-1}(n-1)$$

$$(IV.2) \quad u_j(n) = u_{j-1}(n-1) - k_j^x(n) x_{j-1}(n)$$

avec :

$$(IV.3) \quad x_0(n) = u_0(n) = x(n)$$

où $x(n)$ représente l'entrée du filtre.

Les paramètres notés $k_j^u(n)$ et $k_j^x(n)$, qui caractérisent l'étage d'ordre j du filtre, sont appelés coefficients de corrélation partielle (PARCOR) direct et rétrograde respectivement. Les signaux $x_j(n)$ et $u_j(n)$ définissent les résidus (ou erreurs de prédiction) direct et rétrograde.

Une manière simple de mettre à jour récursivement vis-à-vis du temps les coefficients PARCOR consiste à utiliser l'algorithme du gradient [54], ce qui donne les équations décrites dans le tableau 1 :

TABLEAU 1

Calcul des coefficients PARCOR à l'aide de l'algorithme LMS

$$(IV.4) \quad k_j^x(n+1) = k_j^x(n) + v_j^x x_{j-1}(n) u_j(n) / S_{j-1}^x(n)$$

$$(IV.5) \quad k_j^u(n+1) = k_j^u(n) + v_j^u u_{j-1}(n-1) x_j(n) / S_{j-1}^u(n-1)$$

$$(IV.6) \quad S_{j-1}^x(n) = \lambda S_{j-1}^x(n-1) + (1-\lambda) x_{j-1}^2(n)$$

$$(IV.7) \quad S_{j-1}^u(n) = \lambda S_{j-1}^u(n-1) + (1-\lambda) u_{j-1}^2(n-1)$$

Les paramètres v_j^u et v_j^x sont appelés taux d'oubli direct et rétrograde. Ils permettent de régler la stabilité et la vitesse d'adaptation de l'algorithme. D'autre par λ est le facteur d'oubli qui intervient dans l'estimateur des variances des résidus direct et rétrograde et qui est fixé a priori par l'utilisateur.

Cet estimateur est non biaisé de telle sorte que :

$$E[S_{j-1}^x(n)] = P_{j-1}^x \quad \text{et} \quad E[S_{j-1}^u(n)] = P_{j-1}^u$$

où P_{j-1}^x et P_{j-1}^u représentent respectivement les puissances des résidus avance et rétrograde d'ordre $(j-1)$.

Nous rappelons que dans une situation stationnaire nous avons :

$$(IV.8) \quad v_j^x = v_j^u = v$$

et les conditions de convergence théoriques sont dans ce cas [51] :

- pour une convergence en moyenne : $0 \leq v \leq 2$
- pour une convergence en moyenne quadratique : $0 \leq v \leq 2/3$.

Nous allons voir maintenant comment les résultats sur les filtres transverses [55] se transposent au cas des filtres en treillis/LMS [49], [50].

Dans la suite nous considérons la cellule en treillis d'ordre j , en supposant que les cellules précédentes ont été optimisées, ce qui se traduit par les relations suivantes (nous représentons les quantités optimisées à l'aide d'une astérisque) :

$$x_{j-k}(n) \approx x_{j-k}^*(n) \quad \text{et} \quad u_{j-k}(n) \approx u_{j-k}^*(n) \\ \text{pour } k = 1, j-1.$$

Soient les résidus optimisés à l'ordre j :

$$x_j^*(n) = x_{j-1}^*(n) - [k_j^u(n)] * u_{j-1}^*(n-1) \\ u_j^*(n) = u_{j-1}^*(n-1) - [k_j^x(n)] * x_{j-1}^*(n)$$

où nous ferons l'approximation suivante :

$$(IV.9) \quad x_j^*(n) \approx x_{j-1}(n) - [k_j^u(n)] * u_{j-1}(n-1) \\ \text{et} \quad u_j^*(n) \approx u_{j-1}(n-1) - [k_j^x(n)] * x_{j-1}(n).$$

Les taux d'oubli v_j^x et v_j^u sont déterminés de façon que le résidu de fluctuation normalisé (cf. I.2.21) des sorties $u_j(n)$ et $x_j(n)$ soit minimisé, c'est-à-dire :

$$(IV.10) \quad \text{Min}_{v_j^x} (\mathcal{M}_{u_j}(v_j^x)) = \text{Min}_{v_j^x} \left(\frac{\sigma_{u_j}^2 - \sigma_{u_j^*}^2}{\sigma_{u_j^*}^2} \right)$$

où :

$$(IV.11) \quad \sigma_{u_j}^2 = E(u_j^2(n)), \quad \sigma_{u_j^*}^2 = E(u_{j^*}^2(n)).$$

De même pour l'optimisation de v_j^u .

Définissant l'incrément temporel réel :

$$(IV.12) \quad dk_j^x(n) = k_j^{x*}(n+1) - k_j^{x*}(n)$$

et :

$$(IV.13) \quad dk_j^x = \text{Sup}_n [|dk_j^x(n)|]$$

ainsi que le degré de non-stationnarité :

$$(IV.14) \quad y_j^x = \left(\frac{(dk_j^x)^2 \sigma_{x_{j-1}}^2}{4 \sigma_{u_j^*}^2} \right)^{1/2}$$

le critère à minimiser (IV.10) s'écrit :

$$(IV.15)$$

$$\mathcal{M}_{u_j}(v_j^x) = \left(\frac{v_j^x}{2} + \frac{4(y_j^x)^2 \left(1 - \frac{v_j^x}{2}\right)}{(v_j^x)^2} \right) \frac{1}{1 - \frac{3}{2}v_j^x}$$

La minimisation de ce critère par rapport à v_j^x est obtenue en annulant la dérivée partielle de $\mathcal{M}_{u_j}(v_j^x)$ par rapport à v_j^x , ce qui conduit à l'équation du troisième degré suivante :

$$(IV.16) \quad f(v_j^x) = \frac{1}{8}(v_j^x)^3 - \frac{3}{2}(y_j^x)^2(v_j^x)^2 + 5(y_j^x)^2 v_j^x - 2(y_j^x)^2 = 0.$$

Pour résoudre cette équation on utilise la méthode itérative de Newton ce qui donne :

$$(IV.17) \quad (v_j^x)_{k+1} = (v_j^x)_k - f(v_j^x)_k / f'(v_j^x)_k.$$

Un calcul identique est utilisé pour l'optimisation du taux de convergence v_j^u .

A noter qu'en pratique les incréments dk_j^x et dk_j^u sont estimés à partir des coefficients PARCOR estimés et lissés.

Le filtre en treillis/LMS ainsi optimisé est résumé à l'aide de la figure 1 :

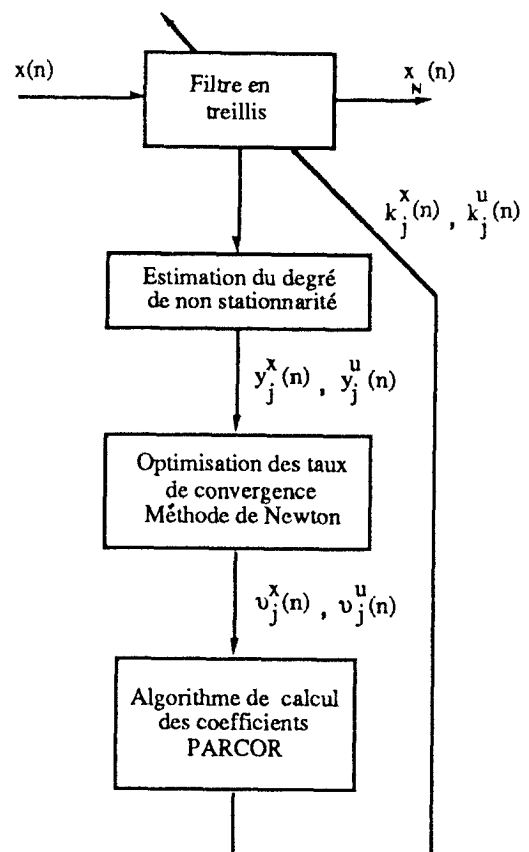


Fig. 1. — Organigramme du filtre en treillis/LMS optimisé.

IV.3. Filtres en treillis/moindres carrés

Le principe de l'algorithme de covariance à fenêtre glissante normalisé adaptatif (CFGNA) est résumé dans le tableau 2 [57] :

TABLEAU 2

Calcul des coefficients PARCOR à l'aide de l'algorithme CFGNA

<p>1. Calcul du test de détection de ruptures de modèles $J(n)$.</p> <p>2. Calcul des coefficients PARCOR du filtre en treillis :</p> <ul style="list-style-type: none"> — Si $J(n) > J_d^{(1)}$ ou $J(n) < J_d^{(2)}$ pendant l'intervalle de temps $[n - n_v, n]$, où n_v est le nombre d'itérations de validation et $J_d^{(i)}$, ($i = 1, 2$), sont les seuils de détection des deux détecteurs de ruptures que nous présentons à la fin de ce paragraphe, alors : <ol style="list-style-type: none"> 2.1. A l'instant n : utiliser l'algorithme CFDN pour décroître la longueur de fenêtre de w_1 à w_c, avec $w_c \ll w_1$. 2.2. Pendant l'intervalle de temps $[n, n + n_s - 1]$, où n_s est le temps de stabilisation : utiliser l'algorithme CFGN avec $w = w_c$. 2.3. A l'instant $n + n_s$: utiliser l'algorithme CFCN pour augmenter la longueur de fenêtre de w_c à w_1, et retourner en 1. — Autrement (pas de détection de ruptures) : utiliser l'algorithme CFGN avec une longueur de fenêtre $w = w_1$, et retourner en 1.
--

Les équations des algorithmes CFCN, CFGN et CFDN sont obtenues en utilisant l'approche géométrique basée sur les formules de réactualisation de l'opérateur de projection.

L'algorithme CFCN [56] :

Cet algorithme permet d'augmenter la longueur de la fenêtre d'une part de 0 à w_1 pendant la période d'initialisation (c'est-à-dire pendant les w_1 premières itérations de l'algorithme CFGNA) et d'autre part de w_c à w_1 après le temps de stabilisation n_s (cf. 2.3 dans le tableau 2). Les équations de cet algorithme sont données dans le tableau 3 :

où :

(IV.24)

$$F[a, b, c] = [1 - c^2]^{-1/2} [a - cb][1 - b^2]^{-1/2}$$

(IV.25) $G[a, b, c] = [1 - c^2]^{1/2} a[1 - b^2]^{1/2} + cb$

— les quantités $x_{j,w}(n)$ et $d_{j,w}(n)$ sont les composantes, aux deux bouts de la fenêtre, du vecteur de résidus avances normalisés ;

— les quantités $u_{j,w}(n-1)$ et $q_{j,w}(n-1)$ sont les composantes, aux deux bouts de la fenêtre, du vecteur de résidus rétrogrades normalisés ;

— les quantités $k_{j,w}(n)$ sont les coefficients PARCOR normalisés et $\mu_{j,w}(n)$ sont des variables auxiliaires.

L'algorithme CFGN [56] :

Cet algorithme permet de faire glisser la fenêtre de longueur constante et égale à w_c ou w_1 suivant que l'on se trouve en régime transitoire après une détection de rupture, ou en régime permanent.

TABLEAU 3

Calcul des coefficients PARCOR à l'aide de l'algorithme de covariance à fenêtre croissante normalisé (CFCN)

<p>Pour w et $\tau = 1$, $w_1 - 1$ ou $w = w_c$, $w_1 - 1$ et $\tau = n + n_s$, $n + n_s + w_1 - w_c - 1$:</p> <p>Pour $j = 0, N - 1$:</p> <p>(IV.18) $d_{j,w-j+1}(\tau) = G(d_{j,w-j}(\tau-1), x_{j,w-j+1}(\tau), \mu_{j,w-j+1}(\tau-1))$</p> <p>(IV.19) $x_{j,w-j}(\tau) = F(x_{j,w-j+1}(\tau), d_{j,w-j+1}(\tau), \mu_{j,w-j+1}(\tau-1))$</p> <p>(IV.20) $\mu_{j+1,w-j+1}(\tau) = F(\mu_{j,w-j+1}(\tau-1), x_{j,w-j+1}(\tau), d_{j,w-j+1}(\tau))$</p> <p>(IV.21) $k_{j+1,w-j}(\tau) = G(k_{j+1,w-j-1}(\tau-1), u_{j,w-j}(\tau-1), x_{j,w-j}(\tau))$</p> <p>(IV.22) $x_{j+1,w-j}(\tau) = F(x_{j,w-j}(\tau), u_{j,w-j}(\tau-1), k_{j+1,w-j}(\tau))$</p> <p>(IV.23) $u_{j+1,w-j}(\tau) = F(u_{j,w-j}(\tau-1), x_{j,w-j}(\tau), k_{j+1,w-j}(\tau))$</p> <p style="text-align: right;">fin fin</p>
--

L'algorithme CFGN est décrit dans le tableau 4 :

TABLEAU 4

Calcul des coefficients PARCOR à l'aide de l'algorithme de covariance à fenêtre glissante normalisé (CFGN)

<p>A chaque instant n et pour w fixé :</p> <p>Pour $j = 0, N - 1$:</p> <p>(IV.26) $k_{j+1,w+1}(n) = G(k_{j+1,w}(n-1), u_{j,w+1}(n-1), x_{j,w+1}(n))$</p> <p>(IV.27) $x_{j+1,w+1}(n) = F(x_{j,w+1}(n), u_{j,w+1}(n-1), k_{j+1,w+1}(n))$</p> <p>(IV.28) $u_{j+1,w+1}(n) = F(u_{j,w+1}(n-1), x_{j,w+1}(n), k_{j+1,w+1}(n))$</p> <p>(IV.29) $d_{j+1,w+1}(n) = F(d_{j,w+1}(n), q_{j,w+1}(n-1), k_{j+1,w+1}(n))$</p> <p>(IV.30) $q_{j+1,w+1}(n) = F(q_{j,w+1}(n-1), d_{j,w+1}(n), k_{j+1,w+1}(n))$</p> <p>(IV.31) $k_{j+1,w}(n) = F(k_{j+1,w+1}(n), q_{j,w+1}(n-1), d_{j,w+1}(n))$</p> <p style="text-align: right;">fin</p>
--

L'algorithme CFDN [57] :

Cet algorithme permet de diminuer récursivement la longueur de la fenêtre, de w_1 à w_c , après la validation d'une détection de rupture de modèle. L'algorithme CFDN est décrit dans le tableau 5 :

Les tests de détection de ruptures de modèle :

Pour détecter les variations brusques des coefficients du modèle, on utilise deux types de détecteurs de

TABLEAU 5
Calcul des coefficients PARCOR à l'aide de l'algorithme de covariance à fenêtre décroissante normalisé (CFDN)

Pour n fixé :	
Pour $w = w_1 - 1, w_c$	
Pour $j = 0, N - 1$	
(IV.32)	$k_{j+1,w}(n) = F(k_{j+1,w+1}(n), q_{j,w+1}(n-1), d_{j,w+1}(n))$
(IV.33)	$q_{j+1,w}(n-1) = F(q_{j+1,w+1}(n), u_{j+1,w+1}(n), \mu_{j+1,w+1}(n))$
(IV.34)	$u_{j+1,w}(n) = F(u_{j+1,w+1}(n), q_{j+1,w+1}(n), \mu_{j+1,w+1}(n))$
(IV.35)	$\mu_{j+1,w}(n) = F(\mu_{j,w}(n), u_{j,w}(n), q_{j,w}(n))$
(IV.36)	$d_{j+1,w}(n) = F(d_{j,w}(n), q_{j,w}(n-1), k_{j+1,w}(n))$
(IV.37)	$q_{j+1,w}(n) = F(q_{j,w}(n-1), d_{j,w}(n), k_{j+1,w}(n))$
fin	
fin	

ruptures. Le premier consiste à comparer des estimateurs à long terme et à court terme de la variance du résidu avancé d'ordre N [52]. Nous avons :

$$(IV.38) \quad J^{(1)}(n) = \sigma_s^2(n) / \sigma_l^2(n)$$

avec $(0 < T_c \leq T_l)$:

$$(IV.39) \quad \sigma_s^2(n) = \frac{1}{T_c} \sum_{\tau=n-T_c+1}^n x_{N,w}^2(\tau)$$

$$(IV.40) \quad \sigma_l^2(n) = \frac{1}{T_l} \sum_{\tau=n-\delta-T_c-T_l+1}^{n-\delta-T_c} x_{N,w}^2(\tau)$$

Les quantités T_c et T_l sont les longueurs de fenêtre à court terme et à long terme sur lesquelles s'effectue l'estimation de la variance. L'introduction du retard δ permet de réduire l'influence des variations des coefficients sur l'estimateur à long terme $\sigma_l^2(n)$.

Les longueurs de fenêtre T_c, T_l d'une part et w_c, w_l d'autre part sont liées à la stationnarité du signal. En pratique on pourra choisir $T_c = w_c$ et $T_l = w_l$ ainsi que $\delta = n_v$.

Le second détecteur est basé sur la variable de vraisemblance non normalisée d'ordre $j \geq N$. Ce type de détecteur a été introduit dans [58] pour la détection de pitch dans un signal de parole.

En annexe, nous développons les calculs relatifs à cette variable de vraisemblance. Le test de détection s'écrit alors :

$$(IV.41) \quad J_j^{(2)}(n) = \gamma_{j,w}(n) = 1 - [x(n) \dots x(n-j+1)] \mathcal{R}_{j,w}^{-1}(n) \begin{bmatrix} x(n) \\ \vdots \\ x(n-j+1) \end{bmatrix}$$

où $\mathcal{R}_{j,w}(n)$ est une estimée de la matrice de corrélation de la séquence $x(n)$, calculée sur une longueur de fenêtre w .

Dans le cas stationnaire, on montre en annexe que

l'espérance mathématique de $\gamma_{j,w}(n)$ vérifie la relation :

$$(IV.42) \quad E[\gamma_{j,w}(n)] \approx 1 - j/w$$

où j est l'ordre de la variable de vraisemblance et $E[\cdot]$ est l'espérance mathématique.

D'autre part, on montre expérimentalement que le test de détection $J_j^{(2)}(n)$ diminue brusquement lors des ruptures pour une valeur de j suffisamment grande. Il suffit alors de fixer un seuil de détection en dessous duquel on décide qu'il y a une rupture de modèle.

Le schéma de principe de l'algorithme en treillis/moindres carrés à fenêtre glissante adaptative est résumé à l'aide de la figure 2 :

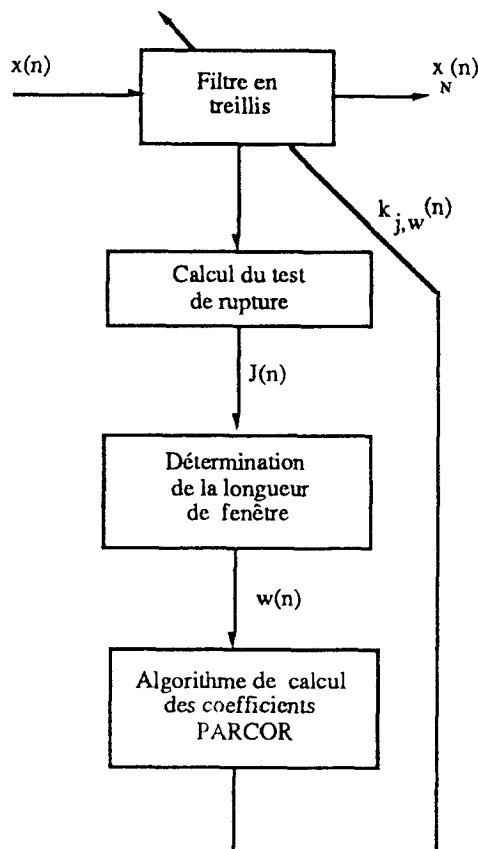


Fig. 2. — Organigramme du filtre en treillis/moindres carrés à fenêtre glissante adaptative.

En résumé, les filtres en treillis/LMS optimisé et moindres carrés à fenêtre glissante adaptative qui vont être comparés au § 4 possèdent chacun un paramètre de réglage permettant de fixer le compromis vitesse de convergence/sensibilité au bruit. Il s'agit du taux d'oubli ν pour les premiers et de la longueur de fenêtre w pour les seconds. Nous allons maintenant donner quelques relations de passage entre ces différents algorithmes.

* Relation entre MCR à facteur d'oubli λ (MCRO) et MCR à fenêtre glissante de longueur w (CFGN) [12] :

$$w \approx 1 / (1 - \lambda) = 1 / \beta$$

où : $\beta = 1 - \lambda =$ taux d'oubli.

Cette relation est obtenue en considérant que la mémoire de l'algorithme MCRO est égale à la constante de temps $1/(1 - \lambda)$ introduite par la pondération exponentielle des mesures.

* Relation entre gradient (LMS) et MCR à facteur d'oubli λ (MCRO) [49] :

$$\nu \approx 1 - \lambda = \beta .$$

Cette relation est obtenue en faisant l'analogie entre l'algorithme treillis/LMS et l'algorithme treillis/moindres carrés à facteur d'oubli.

D'où la relation entre gradient (LMS) et MCR à fenêtre glissante (CFGN) :

$$\nu \approx 1/w .$$

IV.4. Résultats de simulation

Pour chacune des méthodes décrites dans les paragraphes précédents, nous avons effectué une analyse de Monte Carlo (20 séquences de bruit différentes ont été simulées) à l'aide de deux modèles d'ordre 2 présentant une rupture à l'itération 400, dont les caractéristiques sont données dans le tableau 6 :

TABLEAU 6
Caractéristiques des modèles simulés

		Valeur du coefficient PARCOR avant rupture	Valeur du coefficient PARCOR après rupture
modèle 1	k_1	0,82	0,121
	k_2	-0,95	-0,65
modèle 2	k_1	0,121	0,82
	k_2	-0,65	-0,95

Remarque : Les coefficients PARCOR direct et rétrograde sont pris égaux, c'est-à-dire

$$k_j^x = k_j^y = k_j \text{ pour } j = 1 \text{ à } 2 .$$

Ces deux modèles ont été choisis pour mettre en évidence le comportement des algorithmes en treillis/moindres carrés en fonction, d'une part du choix du détecteur de ruptures de modèle et d'autre part du modèle lui-même (cf. annexe : comportement de la variable de vraisemblance lors des ruptures).

IV.4.1. FILTRE EN TREILLIS/MOINDRES CARRÉS

Pour les deux modèles considérés, on applique les méthodes suivantes :

* Filtre en treillis/moindres carrés à fenêtre glissante (CFGN) avec une longueur de fenêtre égale à 100 itérations.

* Filtre en treillis/moindres carrés à fenêtre glissante adaptative (CFGNA) associé au détecteur de ruptures basé sur l'erreur de prédiction.

* Filtre en treillis/moindres carrés à fenêtre glissante adaptative (CFGNA) associé au détecteur de ruptures basé sur la variable de vraisemblance.

Pour ces deux dernières méthodes, on choisit des réglages identiques pour les deux modèles et on donne les taux de non détection et de fausse alarme.

Il faut noter que les critères de détection (fig. 9, 12, 15 et 18) sont gelés après validation de la détection. Les fluctuations observées sont dues à des non détections pour certaines séquences de bruit.

a) *Algorithme CFGN avec une longueur de fenêtre égale à 100 itérations :*

Sur les figures 3 à 6 on voit que le fait de prendre une longueur de fenêtre de 100 itérations entraîne un retard important dans le suivi des paramètres, mais en contrepartie une faible sensibilité au bruit en régime permanent.

modèle 1 :

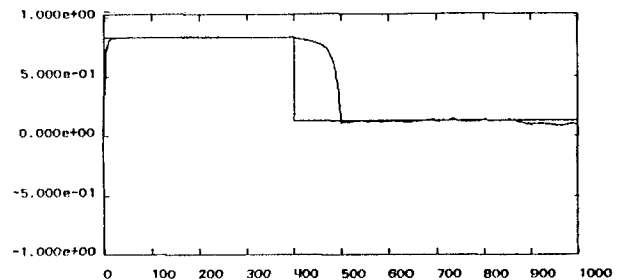


Fig. 3. — Coefficient PARCOR k_1 .

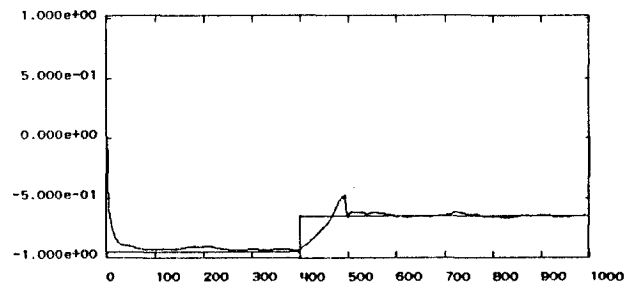


Fig. 4. — Coefficient PARCOR k_2 .

modèle 2 :

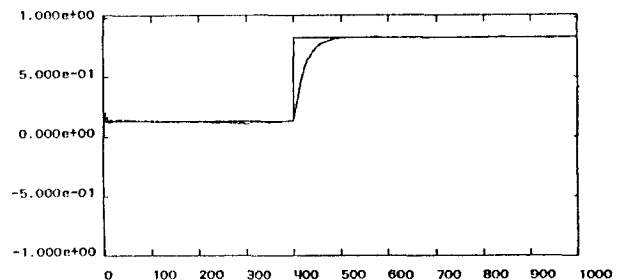


Fig. 5. — Coefficient PARCOR k_1 .

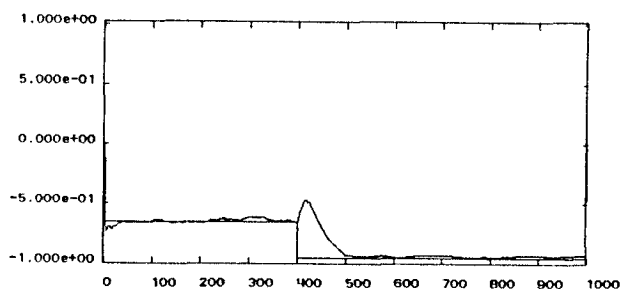


Fig. 6. — Coefficient PARCOR k_2 .

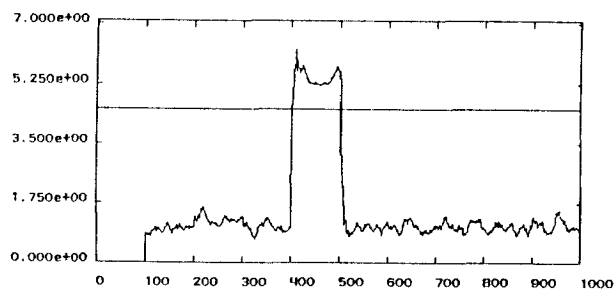


Fig. 9. — Critère de détection et seuil.

b) *Filtre en treillis/moindres carrés à fenêtre adaptative avec détection basée sur l'erreur de prédiction :*
On donne dans le tableau 7 les réglages utilisés pour les deux modèles :

TABLEAU 7
Paramètres de réglage

Longueur de la fenêtre lors des périodes stationnaires	Longueur de la fenêtre lors des ruptures	Seuil de détection
100	10	4,5

On obtient alors les figures suivantes :

modèle 1 :

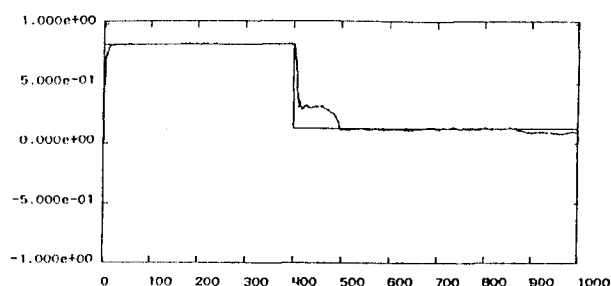


Fig. 7. — Coefficient PARCOR k_1 .

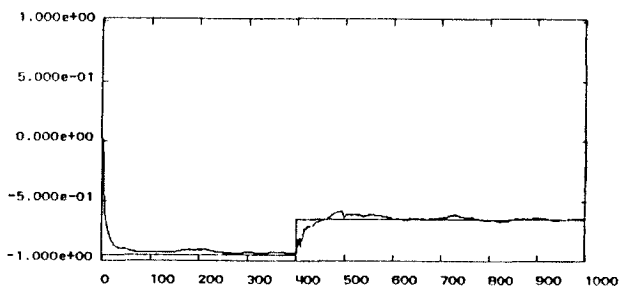


Fig. 8. — Coefficient PARCOR k_2 .

modèle 2 :

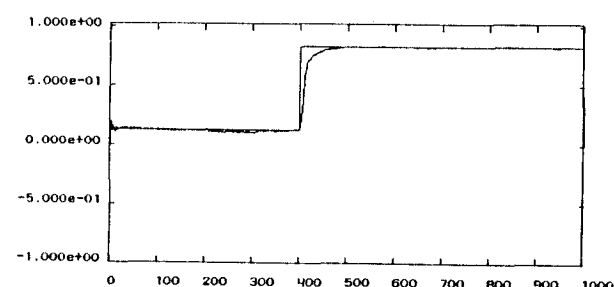


Fig. 10. — Coefficient PARCOR k_1 .

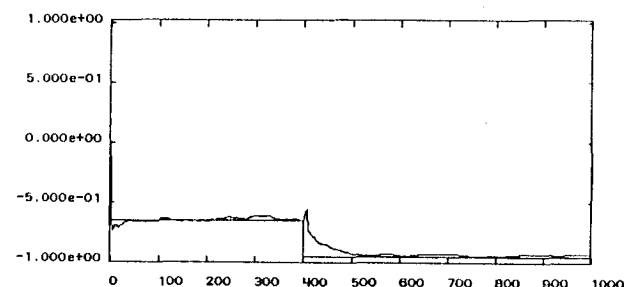


Fig. 11. — Coefficient PARCOR k_2 .

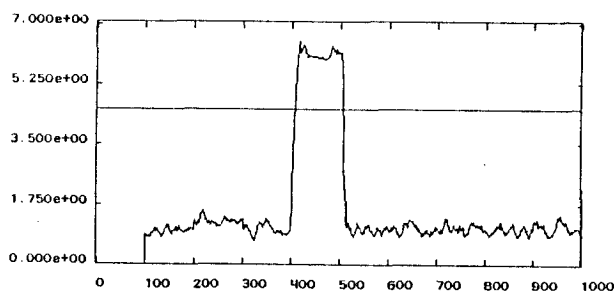


Fig. 12. — Critère de détection et seuil.

Les taux de non détection Tnd et de fausse alarme Tfa sont les suivants :

modèle 1 : *modèle 2 :*

Tnd = 25 % Tnd = 20 %
Tfa = 5 % Tfa = 5 %

c) *Filtre en treillis/moindres carrés à fenêtre adaptative avec détection basée sur la variable de vraisemblance :*

Par rapport au détecteur utilisé en b), le détecteur de ruptures basé sur la variable de vraisemblance nécessite deux réglages à savoir le seuil de détection d'une part et l'ordre de la variable de vraisemblance d'autre part.

En annexe, on montre que le fait d'utiliser un ordre élevé de la variable de vraisemblance permet d'améliorer la détection de ruptures et ceci pour un coût de calcul supplémentaire faible.

Les réglages utilisés sont donnés dans le tableau 8 :

TABLEAU 8
Paramètres de réglage

Longueur de la fenêtre lors des périodes stationnaires	Longueur de la fenêtre lors des ruptures	Ordre de la variable de vraisemblance	Seuil de détection
100	10	10	0,7

On obtient alors les figures suivantes :

modèle 1 :

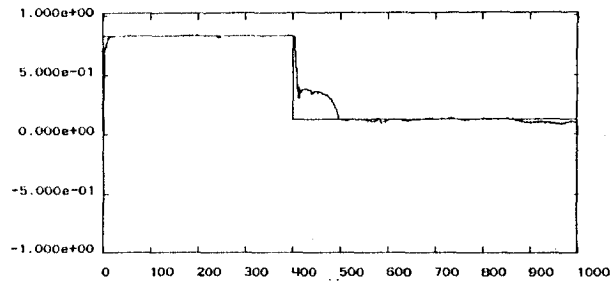


Fig. 13. — Coefficient PARCOR k_1 .

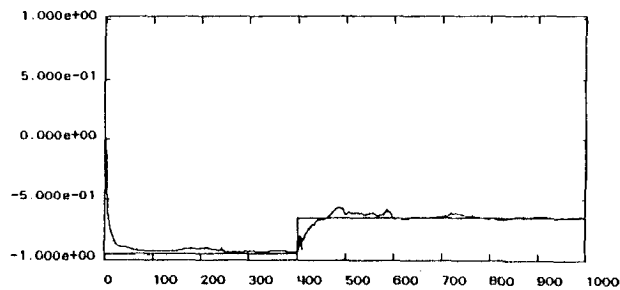


Fig. 14. — Coefficient PARCOR k_2 .

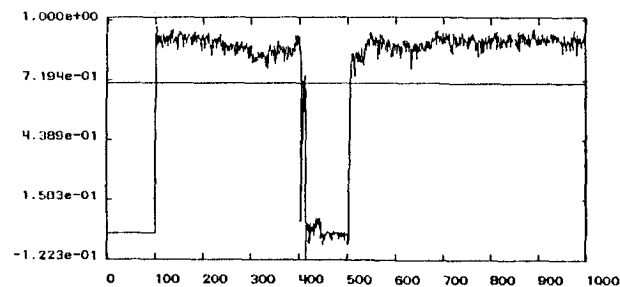


Fig. 15. — Variable de vraisemblance $\gamma_{10}(n)$ et seuil.

modèle 2 :

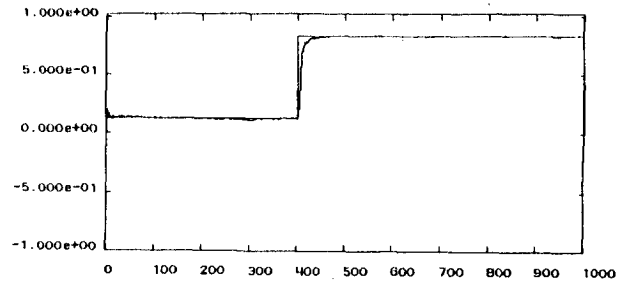


Fig. 16. — Coefficient PARCOR k_1 .

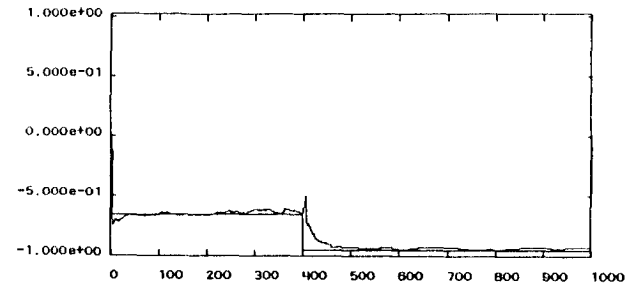


Fig. 17. — Coefficient PARCOR k_2 .

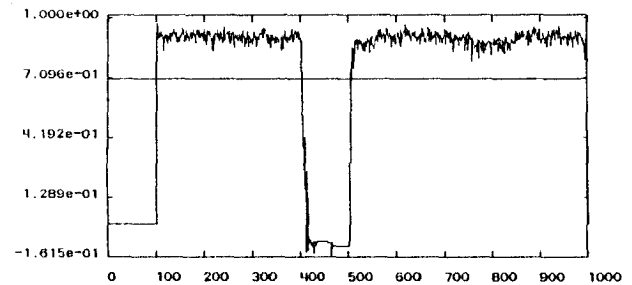


Fig. 18. — Variable de vraisemblance $\gamma_{10}(n)$ et seuil.

Les taux de non détection et de fausse alarme sont alors :

<i>modèle 1 :</i>	<i>modèle 2 :</i>
Tnd = 35 %	Tnd = 0 %
Tfa = 20 %	Tfa = 10 %

IV.4.2. — FILTRE EN TREILLIS/LMS

Pour les deux modèles considérés, on applique les méthodes suivantes :

- * Filtre en treillis/LMS standard
- * Filtre en treillis/LMS optimisé.

a) *Filtre en treillis/LMS standard*

Pour cette méthode, on choisit un réglage du taux d'oubli qui donne sensiblement la même variance de bruit sur les paramètres estimés que l'algorithme CFGN avec une longueur de fenêtre de 100 itérations, à savoir $\nu = 0,02$.

On obtient ainsi les résultats suivants :
modèle 1 :

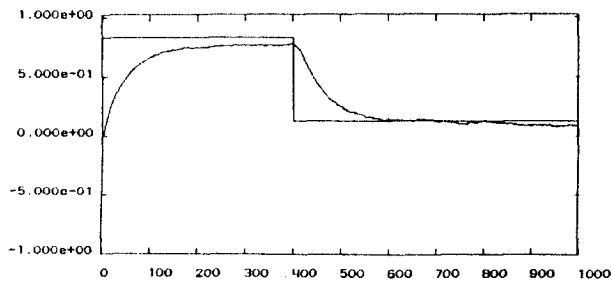


Fig. 19. — Coefficient PARCOR k_1 .

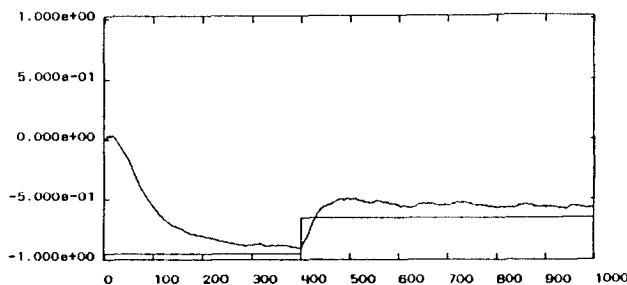


Fig. 20. — Coefficient PARCOR k_2 .

modèle 2 :

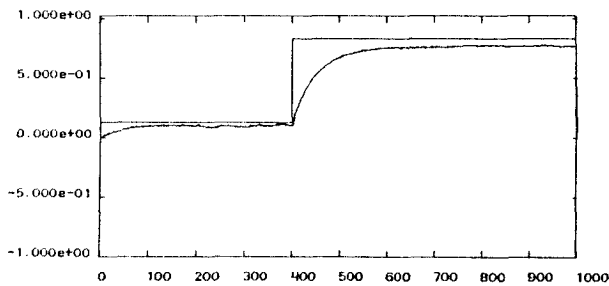


Fig. 21. — Coefficient PARCOR k_1 .

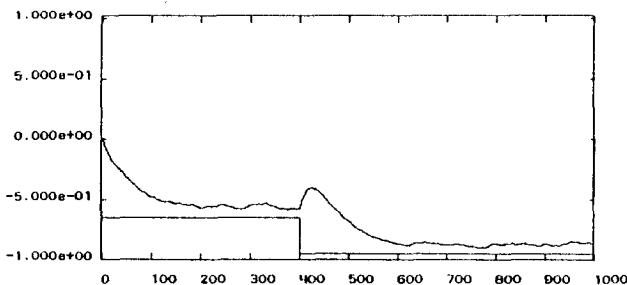


Fig. 22. — Coefficient PARCOR k_2 .

Comme il est bien connu, on constate sur les figures précédentes un temps de convergence relativement long lors de la rupture de modèle (environ 150 itérations). D'autre part, on peut noter que le deuxième coefficient PARCOR apparaît toujours moins bien estimé (biais, fluctuations). Ce type de comporte-

ment qui est moins marqué avec les algorithmes de type moindres carrés est dû à la structure en cascade du filtre en treillis.

b) *Filtre en treillis/LMS optimisé :*

Pour les deux modèles, on a choisi les mêmes réglages, à savoir :

* Facteur d'oubli pour les estimateurs des variances des résidus direct et rétrograde :

$$\lambda = 0,8 .$$

* Facteur d'oubli pour le lissage des paramètres estimés par moindres carrés récursifs :

$$\lambda' = 0,98 .$$

D'autre part, on initialise l'algorithme pendant une période de 200 itérations avec un taux d'oubli faible ($\nu = 0,02$) afin d'obtenir des paramètres estimés peu bruités avant de démarrer l'optimisation de ce taux d'oubli.

On obtient alors les figures suivantes :

modèle 1 :

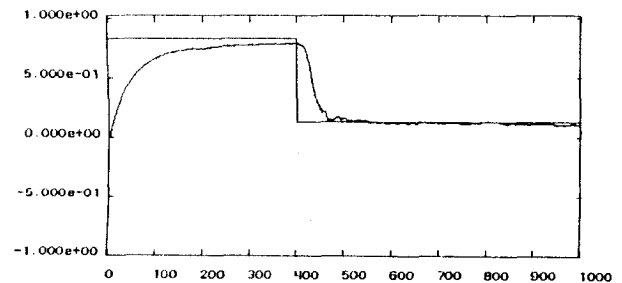


Fig. 23. — Coefficient PARCOR k_1 .

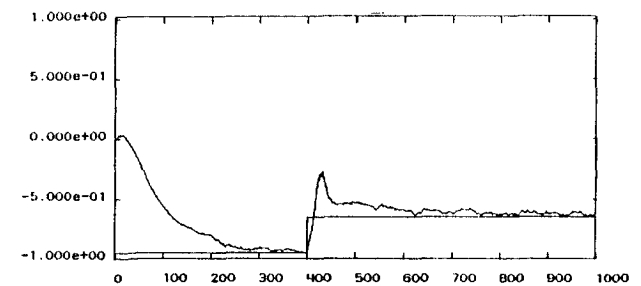


Fig. 24. — Coefficient PARCOR k_2 .

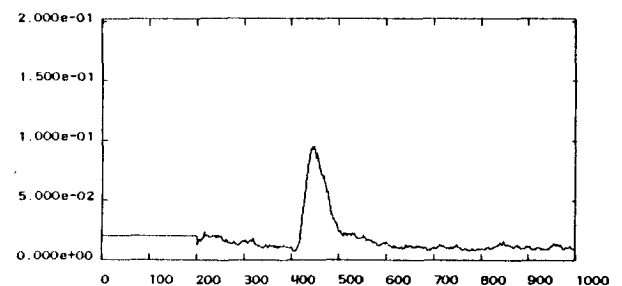


Fig. 25. — Taux d'oubli $\nu_1(n)$.

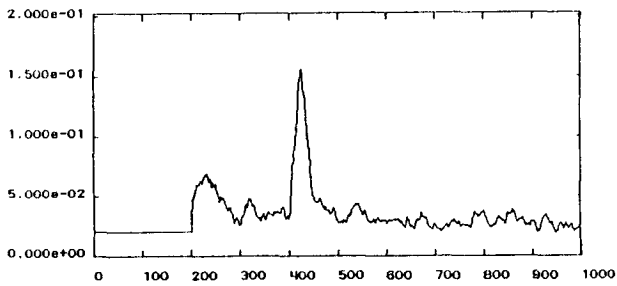


Fig. 26. — Taux d'oubli $v_2(n)$.

modèle 2 :

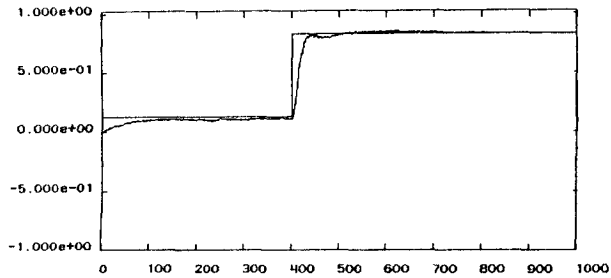


Fig. 27. — Coefficient PARCOR k_1 .

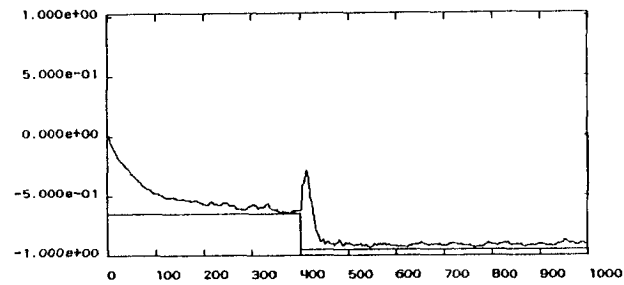


Fig. 28. — Coefficient PARCOR k_2 .

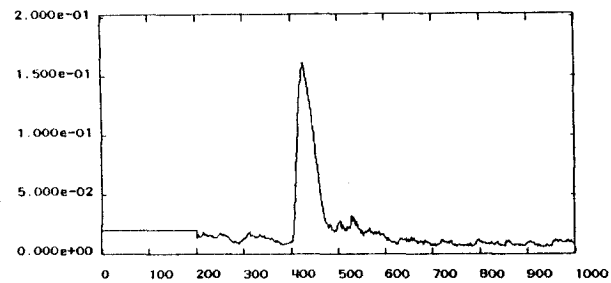


Fig. 29. — Taux d'oubli $v_1(n)$.

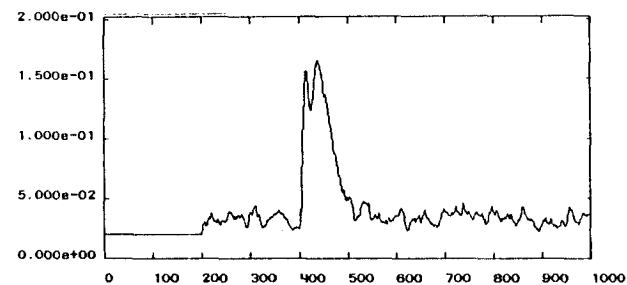


Fig. 30. — Taux d'oubli $v_2(n)$.

Pour les deux modèles on note que l'optimisation des taux d'oubli permet de diminuer de manière nette, au niveau de la rupture, le temps de convergence des paramètres estimés, tout en conservant un bon comportement en régime permanent en terme de variance de bruit sur les paramètres estimés.

Cependant, on remarque qu'il subsiste un retard dans le suivi des paramètres qui est dû au lissage des paramètres estimés, intervenant dans l'estimation du degré de non stationnarité.

IV.4.3. — COMPARAISON DES FILTRES ADAPTATIFS EN TREILLIS/GRADIENT ET TREILLIS/MOINDRES CARRÉS

La comparaison des deux filtres adaptatifs s'effectue en termes de :

- * complexité de calcul
- * complexité de réglage
- * performances

a) *Complexité de calcul :*

Au niveau du calcul des coefficients PARCOR proprement dit, il est bien connu que l'algorithme du gradient a une complexité plus faible que l'algorithme des moindres carrés. Par contre la complexité de calcul est plus importante pour le filtre en treillis/gradient *optimisé* que pour le filtre en treillis/moindres carrés à fenêtre *adaptive*, en raison du lissage des paramètres estimés et de la procédure d'optimisation.

Dans le tableau 9 ci-dessous nous avons évalué le nombre d'opérations élémentaires pour chacun de ces algorithmes (sans tenir compte du lissage des paramètres estimés pour le filtre en treillis/gradient *optimisé*).

TABLEAU 9
Comparaison de la complexité des filtres en treillis adaptatifs

	x et /	+ et -	$\sqrt{\quad}$
CFGN	40 N	18 N	12 N
CFCN	38 N	18 N	12 N
CFDN	42 N	18 N	12 N
$J^{(1)}(t)$	3	4	—
$J^{(2)}(t)$	2 N	2 N	—
LMS	14 N	6 N	—
optimisation	26 nN	8 nN	2 N

où n est le nombre d'itérations de la méthode de Newton (entre 5 et 10 en général) et N est l'ordre du filtre considéré.

La complexité de l'algorithme CFGNA est quant à elle calculée comme ci-dessous :

— Durant les périodes stationnaires ou pendant la période transitoire après une détection de rupture, la complexité est égale à celle de l'algorithme CFGN ou CFCN.

— A l'instant de validation de la détection de rupture la complexité devient $(w_1 - w_c)$ fois la complexité de l'algorithme CFDN.

b) *Complexité de réglage :*

Le paramètre de réglage déterminant pour le filtre en treillis/gradient optimisé est le facteur d'oubli λ de l'estimateur des variances des résidus direct et rétrograde. Il doit être choisi relativement faible de façon à rendre l'estimateur suffisamment « dynamique » lors des ruptures. D'autre part, ce réglage ne dépend pas du type de signal à filtrer.

Par contre, pour le filtre en treillis/moindres carrés, et pour un réglage donné — seuil de détection (détecteur de ruptures basé sur l'erreur de prédiction) et ordre de la variable de vraisemblance (détecteur de ruptures basé sur la variable de vraisemblance) — on aboutit à des taux de non détection et de fausse alarme différents en fonction du type de signal à traiter.

En définitive, on peut conclure que le filtre en treillis/gradient est plus facile à régler que le filtre en treillis/moindres carrés.

c) *Performances :*

En performance pure, le filtre en treillis/moindres carrés est plus satisfaisant notamment au niveau du temps de convergence lors de la rupture. En effet, le lissage des paramètres estimés dans le filtre en treillis/gradient entraîne un certain retard dans le suivi des paramètres (cf. fig. 23 par exemple).

Par contre pour ce même filtre, il n'y a pas de problème de fausse alarme ou de non détection, problèmes qui deviennent critiques pour le filtre en treillis/moindres carrés (cf. fig. 13 par exemple). Pour ce dernier, le réglage du seuil de détection doit être choisi de façon à satisfaire les taux de fausse alarme et de non détection fixés a priori par l'utilisateur.

En définitive, le choix de l'un des deux types de filtres dépend de la connaissance a priori que l'on a sur le modèle et d'autre part des performances que l'on veut atteindre.

Si l'on a aucune connaissance a priori sur le type de signal à filtrer, la solution la plus fiable consiste à utiliser le filtre en treillis/LMS optimisé. Il n'y a alors pas de problème de fausse alarme ou de non détection. Par contre, le trainage des paramètres estimés lors de la rupture demeure important.

IV.5. Conclusion

Nous avons présenté dans ce chapitre deux nouveaux filtres en treillis adaptatifs basés soit sur l'optimisation des taux d'oubli (treillis/LMS optimisé), soit sur le réglage automatique de la longueur de la fenêtre des Moindres Carrés à travers l'utilisation d'un test de ruptures de modèle (treillis/MC à fenêtre adaptative). Les performances de ces deux filtres ont été analysées sur des exemples simulés à l'aide d'une approche de Monte Carlo. Des améliorations peuvent être apportées aux résultats de simulation ainsi obtenus en affinant l'estimation des degrés de non

stationnarité ou encore en modifiant le détecteur de ruptures de modèle [59] ou la procédure d'ajustement de la longueur de fenêtre qui peut être effectuée par exemple en fonction de la stationnarité du signal à filtrer, comme cela a été proposé dans [60].

ANNEXE : Calcul de la variable de vraisemblance

Dans cette annexe, on définit la variable de vraisemblance puis on montre le comportement de cette variable lors de ruptures de modèle.

Définition : La variable de vraisemblance *non normalisée* d'ordre j , à l'instant $n - 1$, et calculée sur $w + 1$ mesures, s'exprime par :

$$(A.1) \quad \gamma_{j, w+1}(n-1) = 1 - [x(n-1) \dots x(n-j)] \times \mathcal{R}_{j, w+1}^{-1}(n-1) \begin{bmatrix} x(n-1) \\ \vdots \\ x(n-j) \end{bmatrix}$$

où $\mathcal{R}_{j, w+1}(n-1)$ est une estimée, à l'instant $n - 1$, de la matrice de corrélation d'ordre j du signal à filtrer $x(\cdot)$, calculée sur $w + 1$ mesures.

On montre que la réactualisation de la variable de vraisemblance s'effectue de deux façons :

— réactualisation sur l'ordre uniquement :

$$(A.2) \quad \gamma_{j+1, w+1}(n-1) = \gamma_{j, w+1}(n-1) - \bar{u}_{j, w+1}^2(n-1) R_{j, w+1}^{-u}(n-1)$$

avec :

$$(A.3) \quad R_{j, w+1}^{-u}(n-1) = [R_{j, w+1}^u(n-1)]^{-1}$$

— réactualisation sur l'ordre et sur le temps :

$$(A.4) \quad \gamma_{j+1, w+1}(n) = \gamma_{j, w+1}(n-1) - \bar{x}_{j, w+1}^2(n) R_{j, w+1}^{-x}(n)$$

où $\bar{x}_{j, w+1}(n)$ et $\bar{u}_{j, w+1}(n-1)$ représentent respectivement les résidus avance et rétrograde d'ordre j non normalisés et $R_{j, w+1}^x(n)$ et $R_{j, w+1}^u(n-1)$ les variances estimées de ces mêmes résidus.

On peut alors exprimer la variable de vraisemblance par rapport aux résidus normalisés en remarquant que :

$$(A.5) \quad x_{j, w+1}(n) = R_{j, w+1}^{-x/2}(n) \gamma_{j, w+1}^{-1/2}(n-1) \bar{x}_{j, w+1}(n)$$

et

$$(A.6) \quad u_{j, w+1}(n-1) = R_{j, w+1}^{-u/2}(n-1) \times \gamma_{j, w+1}^{-1/2}(n-1) \bar{u}_{j, w+1}(n-1).$$

On obtient :

$$(A.7) \quad \gamma_{j+1, w+1}(n) = \gamma_{j, w+1}(n-1) - x_{j, w+1}^2(n) \gamma_{j, w+1}(n-1)$$

et

$$(A.8) \quad \gamma_{j+1, w+1}(n-1) = \gamma_{j, w+1}(n-1) - u_{j, w+1}^2(n-1) \gamma_{j, w+1}(n-1).$$

Remarque : On note que pour un ordre j supérieur à N , où N est l'ordre du filtre générateur de la séquence $x(n)$, les coefficients PARCOR sont théoriquement nuls, et par suite les résidus non normalisés et leurs variances associées sont constants. En utilisant les égalités (A.7) et (A.8), nous obtenons alors les relations suivantes pour $j \geq N$:

$$(A.9) \quad \gamma_{j+1, w+1}(n) \approx (1 - (j - N + 1) x_{N, w+1}^2(n)) \gamma_{N, w+1}(n-1)$$

et

$$(A.10) \quad \gamma_{j+1, w+1}(n-1) \approx (1 - (j - N + 1) u_{N, w+1}^2(n-1)) \gamma_{N, w+1}(n-1).$$

Évaluation de la variable de vraisemblance :

On montre par récurrence que l'espérance mathématique de la variable de vraisemblance est donnée par :

$$(A.11) \quad E[\gamma_{j+1, w+1}(n-1)] \approx 1 - \frac{j+1}{w+1}.$$

Pour cela, considérons la relation (A.2) et supposons qu'à l'ordre j , on ait :

$$(A.12) \quad E[\gamma_{j, w+1}(n-1)] \approx 1 - \frac{j}{w+1}.$$

La quantité $R_{j, w+1}^u(n-1)$ constitue une estimée biaisée sur $w+1$ mesures de $E[\bar{u}_{j, w+1}^2(n-1)]$, soit :

$$(A.13) \quad R_{j, w+1}^u(n-1) \approx (w+1) E[\bar{u}_{j, w+1}^2(n-1)].$$

En combinant (A.6), (A.8) et (A.13) puis en prenant l'espérance mathématique, on obtient :

$$(A.14) \quad E[\gamma_{j+1, w+1}(n-1)] \approx E[\gamma_{j, w+1}(n-1)] - \frac{1}{w+1} = 1 - \frac{j+1}{w+1}.$$

La relation (A.12) est donc vérifiée à l'ordre $(j+1)$. D'autre part il est facile de montrer que la condition initiale de la récurrence est vérifiée à l'ordre 0 :

$$(A.15) \quad E[\gamma_{0, w+1}(n-1)] = \gamma_{0, w+1}(n-1) = 1.$$

Comportement de la variable de vraisemblance lors des ruptures :

Pour mettre en évidence le comportement de la variable de vraisemblance, on a choisi deux modèles dont la caractéristique essentielle est qu'ils présentent une forte variation au niveau de la variance du signal à filtrer avant et après la rupture. Les signaux générés par ces deux modèles sont visualisés sur les figures 31 et 32.

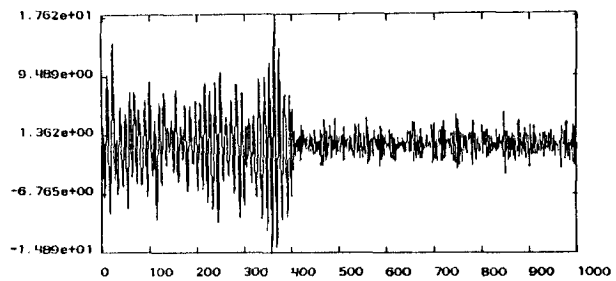


Fig. 31. — Modèle 1 : $x(n)$.

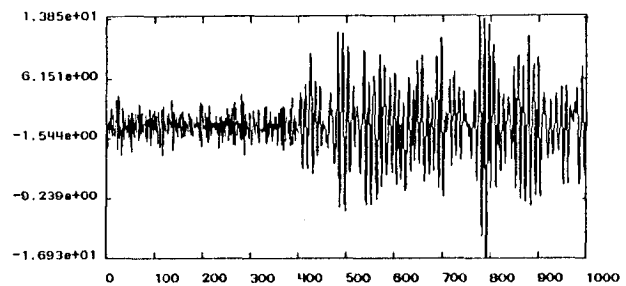


Fig. 32. — Modèle 2 : $x(n)$.

Le comportement de la variable de vraisemblance lors des ruptures de modèle est mis en évidence à l'aide des figures 33-34 pour $j = 1$ et 35-36 pour $j = 3$ (avec $w+1 = 100$). Ces figures montrent que pour une valeur de j suffisamment grande le test de détection $J_j^{(2)}(n)$ diminue brusquement lors des ruptures. Une analyse du comportement de ce type de détecteurs est effectuée dans [59].

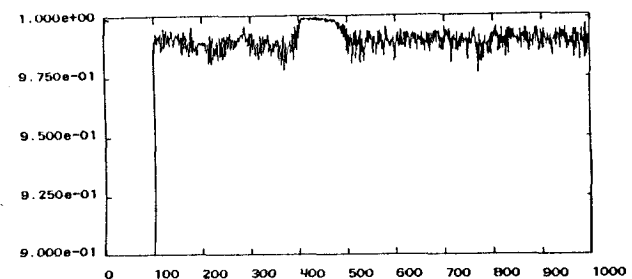


Fig. 33. — Modèle 1 : $\gamma_1(n)$.

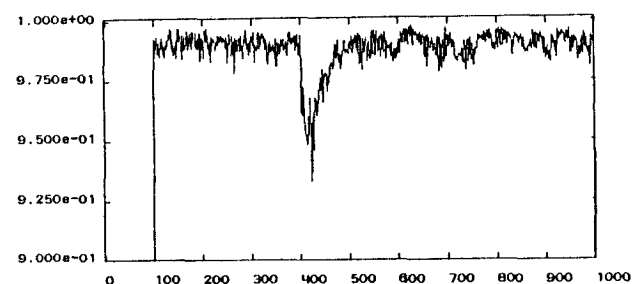
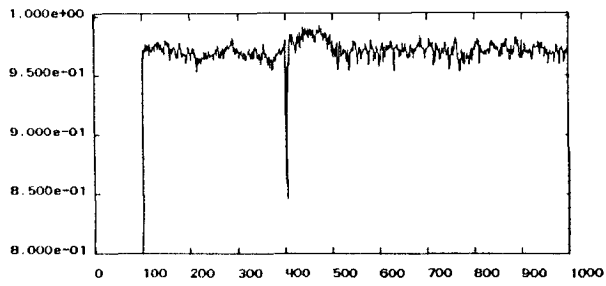
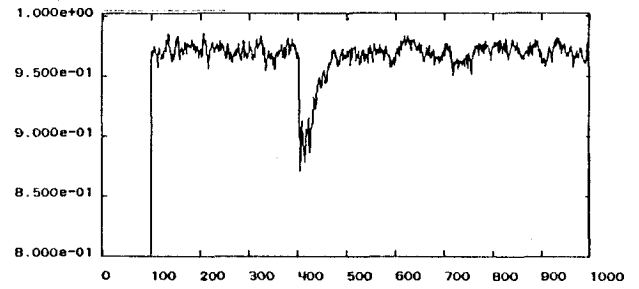


Fig. 34. — Modèle 2 : $\gamma_1(n)$.

Fig. 35. — Modèle 1 : $\gamma_3(n)$.Fig. 36. — Modèle 2 : $\gamma_3(n)$.

Chapitre V
STRUCTURE D'IDENTIFICATION INVERSE
APPLICATION A L'ÉGALISATION

P. BRAGARD
Cephag/ ENSIEG

V.1. Introduction

En acoustique sous-marine, la transmission de données est perturbée d'une part, par du bruit, et d'autre part, par le phénomène de trajets multiples dû à des réflexions et des réfractions dans le milieu sous-marin.

Lorsque le rapport signal/bruit est suffisamment bon (typiquement, supérieur à 5 dB), un moyen de lutte contre les interférences intersymboles causées par la propagation multitrajets est la mise en œuvre d'un filtre numérique « inverse » \mathcal{H} appelé égaliseur, et dont le rôle est de compenser la distorsion apportée par le canal de transmission. Nous nous limiterons ici au cas d'un égaliseur transversal d'ordre N.

Le canal de transmission sous-marin étant dans la plupart des cas non-stationnaire, les coefficients transversaux de l'égaliseur doivent être remis à jour au cours du temps à l'aide d'un algorithme adaptatif qui a pour rôle d'identifier le filtre optimal inverse $\mathcal{F}(n)$ défini par la solution de Wiener, et de poursuivre ses fluctuations lorsque celui-ci varie au cours du temps.

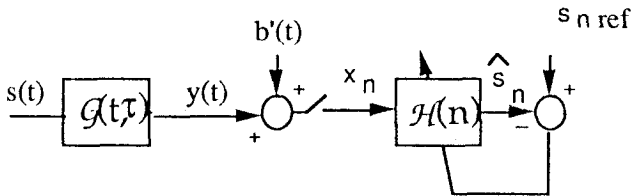


Fig. 1. — Égalisation adaptative.

Nous modélisons la réponse impulsionnelle $\mathcal{G}(t, \tau)$ du canal de transmission par une somme d'impulsions de Dirac d'amplitude complexe variable, et décalées dans le temps :

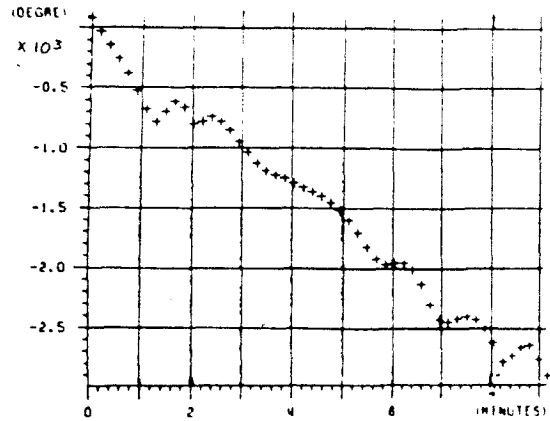
$$\mathcal{G}(t, \tau) = \sum a_i \delta(t - \tau_i) e^{j\phi_i}$$

et nous notons ρ le rapport P_y/P_b .

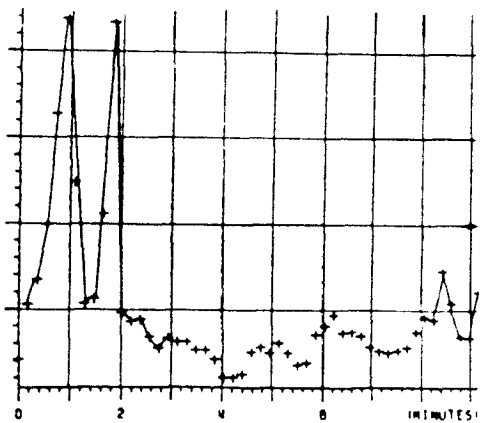
L'égaliseur adaptatif est confronté à plusieurs types de non-stationnarités que nous classons en deux groupes :

- NS1 — dérive de l'amplitude des trajets a_i
- dérive de la phase des trajets ϕ_i
- NS2 — dérive des retards τ_i
- sauts d'amplitude, de retard ou de phase des trajets.

Canal de transmission à deux trajets de propagation a_0 et a_1 ;
expérimentation en mer à 310 km de distance et par grand fond :

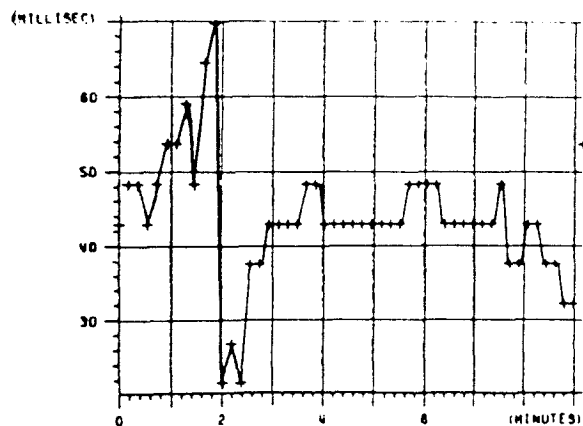


Non-stationnarité de type "dérive" : phase du premier trajet de propagation a_0 .



Non-stationnarités de type "dérives" et "ruptures" :

rapport d'amplitude entre les trajets $R = \frac{a_1}{a_0}$



Non-stationnarités de type "ruptures", avec périodes stables : retard entre les trajets τ .

Fig. 2. — Les types de non-stationnarités. ►

Suivant le cas, c'est soit la capacité de poursuite des algorithmes LMS et MCRO qui est éprouvée, soit leur aptitude à réagir rapidement à une rupture (vitesse de convergence).

Nous considérons dans cette étude le cas assez courant d'une propagation bi-trajets (souvent rencontré lorsqu'il s'agit d'une transmission à grande distance et par grands fonds), où les paramètres $a_0, a_1, \phi_0, \phi_1, \tau_0, \tau_1$ sont susceptibles de varier. L'expérience révèle alors que les paramètres importants sont le module du rapport d'amplitude entre trajets $|R| = (a_1/a_0)$, le déphasage entre trajets $\Phi = \phi_0 - \phi_1$, et le retard entre trajets noté L (en nombre d'échantillons). La figure 2 représente une estimation de quelques-uns de ces paramètres issue d'une expérimentation en mer à grande distance (310 km); on y distingue facilement les deux groupes de non-stationnarités NS1 et NS2. En pratique, c'est souvent le déphasage Φ et le module du rapport d'amplitude R qui sont variables, tandis que les retards τ_i sont fixes.

Nous proposons d'étudier le comportement des algorithmes LMS et MCRO analytiquement dans une situation de poursuite très simple (cas d'une dérive du déphasage entre trajets) et par simulations dans des cas de rupture, afin d'obtenir quelques règles d'utilisation sur l'adéquation des algorithmes MCRO en égalisation. Pour l'étude des performances en situation de poursuite, nous considérerons un cas très particulier qui permet de mener un calcul analytique jusqu'au bout, le but recherché n'étant pas de tirer une conclusion générale mais de réfuter à l'aide de cet exemple certaines idées a priori très répandues au sujet de la capacité de poursuite des algorithmes adaptatifs.

V.2. Définition et modélisation du problème

Afin d'utiliser une méthodologie commune, on modélise le problème comme un problème d'identification, où il s'agit d'identifier un filtre linéaire $\mathcal{F}(n)$ dont la sortie est entachée d'un bruit additif b_n équivalent au bruit de mesure intervenant dans le système initial de la figure 1, indépendant de l'entrée x_n .

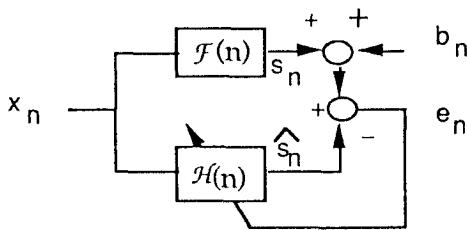


Fig. 3. — Modélisation d'un problème d'identification inverse comme un problème d'identification directe.

$$\mathcal{F}(n) = \underline{\Gamma}_{xx}^{-1}(n) \underline{\Gamma}_{sx}(n)$$

$\underline{\Gamma}_{xx}(n)$: matrice d'autocorrélation du vecteur \underline{X}_n

$\underline{\Gamma}_{sx}(n)$: vecteur d'intercorrrelation entre \underline{X}_n et la référence S_n ref

$$E(S_{n+1} \text{ ref} - \mathcal{F}^T(n) \cdot \underline{X}_{n+1})^2 = E_{\min}$$

avec $S_n \text{ ref} = S_n + b_n$.

On note

$$b_n = \mathcal{G}^{-1}(b'_n) \text{ et tel que } E(b_n)^2 = E_{\min}$$

$$\mathcal{F}(n) = \mathcal{G}^{-1}(n)$$

* : complexe conjugué
T : transposé
+ : transconjugué.

On a alors à étudier un problème d'identification classique, à la différence près que le bruit b'_n intervient non seulement dans la mesure de l'erreur d'estimation — par l'intermédiaire de b_n — mais aussi dans l'expression du filtre optimal $\mathcal{F}(n)$ (à la limite, pour $\rho \rightarrow \infty$, $\mathcal{F}(n) = 0$). Ceci a pour effet d'atténuer les variations du modèle à poursuivre par rapport à celles de $\mathcal{G}(n)$; c'est une particularité des problèmes d'identification inverse.

Pour l'étude des algorithmes LMS et MCRO en régime de poursuite, nous considérons le cas d'un canal de propagation à deux trajets, de modules fixés a_0 et a_1 , avec un déphasage entre trajets $\Phi(n) = \Phi_0 - \Phi_1(n)$ variant linéairement au cours du temps :

$$(1) \quad \Phi(n) = n \cdot \Psi + \phi_0$$

Ψ désigne l'incrément élémentaire de phase entre deux instants successifs, et ϕ_0 la phase initiale. Les retards τ_0 et τ_1 sont fixes, et espacés de L . La non-stationnarité est du type *déterministe*.

Pour $N = L + 1$, et lorsque un retard d'estimation $d = L$ est imposé, le filtre optimal \mathcal{F}_n ne comporte que deux coefficients f_0 et f_1 espacés de L , et dont l'expression est [6] :

$$(2) \quad f_0 = (R^*/a_0 e^{j\phi_0}) \cdot (g - 1)/(g^2 - |R|^2)$$

$$(3) \quad f_1 = (1/a_0 e^{j\phi_0}) \cdot (g - |R|^2)/(g^2 - |R|^2)$$

avec

$$(4) \quad g = (1 + |R|^2) \cdot (1 + \xi)$$

$$(5) \quad 1/\xi = \rho$$

$$(6) \quad R = (a_1/a_0) e^{j\Phi(n)}$$

Ce cas particulier sera utilisé pour l'étude théorique des performances en poursuite des algorithmes LMS et MCRO.

V.3. Étude de la poursuite

a) Cas de l'algorithme LMS

Soit

$$(7) \quad \underline{Y}_n = \mathcal{H}(n) - \mathcal{F}(n)$$

et

$$(8) \quad \underline{T}_n = \mathcal{F}(n+1) - \mathcal{F}(n)$$

μ est le pas d'adaptation de l'algorithme.

On montre classiquement que [37] :

$$(9) \quad \underline{V}_{n+1} = (\mathbf{I} - \mu \underline{X}_n^* \underline{X}_n^T) \underline{V}_n + \mu b_n \underline{X}_n^* - \underline{T}_n.$$

Cette récurrence est la somme de trois termes :

$$(10) \quad \underline{V}_{t_{n+1}} = (\mathbf{I} - \mu \underline{X}_n^* \underline{X}_n^T) \underline{V}_{t_n} \text{ transitoire}$$

$$(11) \quad \underline{V}_{f_{n+1}} = (\mathbf{I} - \mu \underline{X}_n^* \underline{X}_n^T) \underline{V}_{f_n} + \mu b_n \underline{X}_n^* \text{ fluctuations}$$

$$(12) \quad \underline{V}_{r_{n+1}} = (\mathbf{I} - \mu \underline{X}_n^* \underline{X}_n^T) \underline{V}_{r_n} - \underline{T}_n \text{ retard}$$

où

$$\underline{V}_t(0) = \underline{V}(0)$$

$$\underline{V}_f(0) = \underline{0}$$

$$\underline{V}_r(0) = \underline{0}$$

et on cherche à calculer $E(\underline{V}_n^T \underline{X}_n)^2$.

Hypothèse : (A) les échantillons x_n sont statistiquement indépendants des vecteurs \underline{V}_{f_n} , \underline{V}_{t_n} et \underline{V}_{r_n} , et indépendants de \underline{T}_n (c.-à-d. indépendants du phénomène physique de variations du canal de transmission, ce qui est admissible) alors,

$$(13) \quad E(x_n - \mathcal{H}^T(n) \underline{X}_n)^2 = E_{\min} + E(\underline{V}_{f_n}^T \cdot \underline{X}_n)^2 + E(\underline{V}_{r_n}^T \cdot \underline{X}_n)^2.$$

$\mathcal{M}_f \qquad \mathcal{M}_r$

L'influence du transitoire étant supposée négligeable pour n grand.

Calcul du résidu de retard \mathcal{M}_r :

On montre que la puissance d'erreur résiduelle de retard est la somme de deux termes : l'un (\mathcal{M}_r) est dû à la partie à valeur moyenne non nulle de \underline{V}_r , et l'autre (\mathcal{M}_r) à la partie à valeur moyenne nulle. Il est possible de montrer que ce dernier est d'influence négligeable. Or,

$$(14) \quad \mathcal{M}_r = \frac{\underline{T}_n^+ \underline{\mathbb{R}}_{xx}^{-1}(n) \underline{T}_n}{\mu^2}.$$

On trouve :

$$(15) \quad \mathcal{M}_r(\text{LMS}) = \frac{2}{\nu^2} P_x^2 |\mathbf{R}|^2 g \left[\frac{g-1}{g^2 - |\mathbf{R}|^2} \right]^2 \frac{1 - \cos \psi}{g^2 - |\mathbf{R}|^2}$$

avec $\nu = \mu P_x$.

Calcul de \mathcal{M}_f (LMS) :

Le calcul du résidu de fluctuations dans le cas de l'algorithme LMS est un résultat classique. Sous l'hypothèse d'indépendance de \underline{V}_n^f avec \underline{X}_n , on montre facilement que :

$$(16) \quad \mathcal{M}_f(\text{LMS}) = \frac{N}{2} \nu E_{\min}$$

E_{\min} désigne l'erreur quadratique moyenne minimale

en sortie du filtre adaptatif, dont l'expression est dans notre cas :

$$(17) \quad E_{\min} = 1 - \mathcal{F}^+ \mathcal{G} = 1 - a_0 \cdot |f_0| - a_1 \cdot |f_1|.$$

La puissance d'erreur résiduelle totale est alors :

$$(18) \quad \mathcal{M} = \mathcal{M}_r + \mathcal{M}_f = \frac{2}{\nu^2} P_x^2 |\mathbf{R}|^2 g \left[\frac{g-1}{g^2 - |\mathbf{R}|^2} \right]^2 \frac{1 - \cos \psi}{g^2 - |\mathbf{R}|^2} + \frac{N}{2} \nu E_{\min}$$

et $g = (1 + |\mathbf{R}|^2)(1 + 1/\rho)$.

La valeur de g augmente lorsque le rapport signal/bruit ρ diminue, tout comme celle de E_{\min} : la contribution du résidu de retard diminue donc lorsque ρ diminue tandis que celle du résidu de fluctuation augmente, ce qui est une particularité des problèmes d'identification inverse.

Ce résultat reflète par ailleurs le comportement caractéristique de l'algorithme LMS dans le cas d'une non-stationnarité de type déterministe ; le terme de retard décroît de $1/\nu^2$, tandis que le terme de fluctuations croît proportionnellement à ν . La valeur de ν réalisant le compromis entre fluctuations et retard se calcule en cherchant le minimum de \mathcal{M} :

$$(19) \quad \nu_{\text{opt}} = \left[\frac{8 P_x^2 (1 - \cos \psi) g |\mathbf{R}|^2 (g-1)^2}{N E_{\min} (g^2 - |\mathbf{R}|^2)^3} \right]^{1/3}.$$

b) *Cas de l'algorithme MCRO*

La méthode d'analyse est la même [35] :

$$(20) \quad \underline{V}_{n+1} = (\mathbf{I} - \underline{\mathbb{R}}_{xx}^{-1}(n+1) \underline{X}_{n+1}^* \underline{X}_{n+1}^T) \times \underline{V}_n + \underline{\mathbb{R}}_{xx}^{-1}(n+1) b_{n+1} \underline{X}_{n+1}^* - \underline{T}_{n+1}$$

où $\underline{\mathbb{R}}_{xx}$ est la matrice de covariance *estimée* de \underline{X} :

$$(21) \quad \underline{\mathbb{R}}_{xx}(n) = \sum_{p=1}^n \lambda^{n-p} \underline{X}_p^* \underline{X}_p^T$$

\underline{V}_{n+1} est la somme des trois termes :

$$(22) \quad \underline{V}_{t_{n+1}} = (\mathbf{I} - \underline{\mathbb{R}}_{xx}^{-1}(n+1) \underline{X}_{n+1}^* \underline{X}_{n+1}^T) \underline{V}_{t_n} \text{ transitoire}$$

$$(23) \quad \underline{V}_{f_{n+1}} = (\mathbf{I} - \underline{\mathbb{R}}_{xx}^{-1}(n+1) \underline{X}_{n+1}^* \underline{X}_{n+1}^T) \underline{V}_{f_n} + \underline{\mathbb{R}}_{xx}^{-1}(n+1) b_{n+1} \underline{X}_{n+1}^* \text{ fluctuations}$$

$$(24) \quad \underline{V}_{r_{n+1}} = (\mathbf{I} - \underline{\mathbb{R}}_{xx}^{-1}(n+1) \times \underline{X}_{n+1}^* \underline{X}_{n+1}^T) \underline{V}_{r_n} - \underline{T}_{n+1} \text{ retard}$$

avec

$$\underline{V}_t(0) = \underline{V}(0)$$

$$\underline{V}_f(0) = \underline{0}$$

$$\underline{V}_r(0) = \underline{0}.$$

Sous l'hypothèse (A) déjà utilisée pour l'algorithme LMS, l'erreur résiduelle de l'algorithme MCRO est la contribution des deux termes \mathcal{M}_f (MCRO) et \mathcal{M}_r (MCRO), l'influence du transitoire étant considérée nulle pour n grand.

Hypothèses

- (B) Pour tout p , les vecteurs X_p sont indépendants des échantillons b_p
- (C) Pour tout $p \leq n + 1$, les vecteurs X_p sont indépendants de X_{n+1}
- (D) $1 - \lambda \gg \psi$
- (E) La matrice $R_{xx}(n)$ est quasi-déterministe pour n grand.

Calcul de M_r (MCRO)

Comme dans le cas de l'algorithme LMS, le résidu de retard est essentiellement dû à la partie moyenne de V_r ; le calcul de M_r nécessite certaines approximations au premier ordre, qui ne sont valables que si la condition (D) est vérifiée (les variations du filtre à poursuivre sont lentes devant la capacité de poursuite de l'algorithme).

On trouve alors :

$$(25) \quad \overline{M}_r(\text{MCRO}) = \frac{2g|R|^2}{\beta^2} \left[\frac{g-1}{g^2 - |R|^2} \right]^2 (1 - \cos \psi)$$

où β est le taux d'oubli de l'algorithme : $\beta = 1 - \lambda$.

Calcul de M_f (MCRO)

En utilisant les hypothèses (A → E), on montre que le terme de fluctuations de l'algorithme MCRO est :

$$(26) \quad M_f(\text{MCRO}) = \frac{\beta}{2} NE_{\min}$$

on en déduit l'erreur résiduelle totale :

$$(27) \quad \mathcal{M}(\text{MCRO}) = \frac{2g|R|^2}{\beta^2} \times \left[\frac{g-1}{g^2 - |R|^2} \right]^2 (1 - \cos \psi) + \frac{\beta}{2} NE_{\min}$$

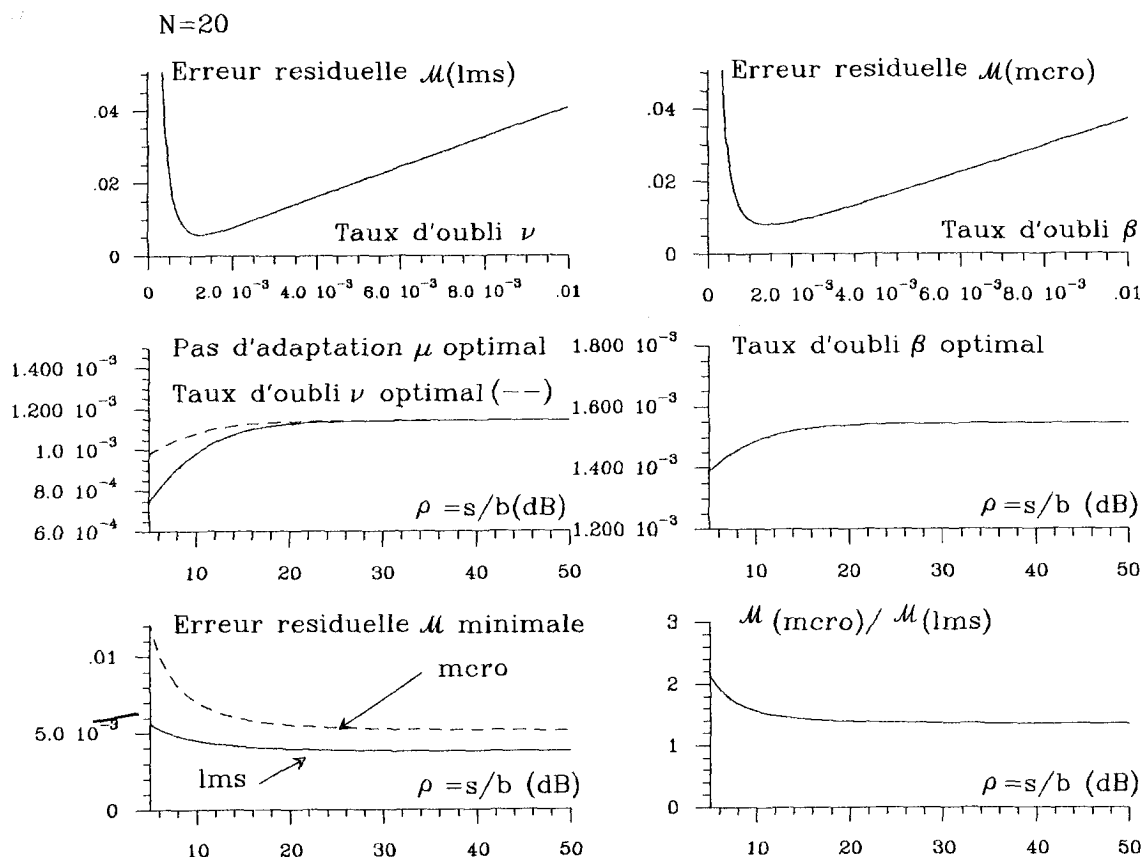
On remarquera le comportement très similaire des algorithmes LMS et MCRO décrit par les formules (18) et (27).

Comme pour l'algorithme LMS, on calcule aisément la valeur de β réalisant le compromis entre l'erreur de fluctuations et l'erreur de retard :

$$(28) \quad \beta_{\text{opt}} = \left[\frac{8g|R|^2(1 - \cos \psi)(g-1)^2}{NE_{\min}[g^2 - |R|^2]^2} \right]^{1/3}$$

c) Comparaison des algorithmes

Les valeurs optimales de ν et β étant connues, on peut comparer les performances des deux algorithmes en termes d'erreur résiduelle, pour une vitesse de variation ψ donnée. Les expressions de M (LMS) et de M (MCRO) sont très similaires. Les deux types d'algorithmes ont un comportement paraissant quasi-identique, mais si l'on trace les erreurs résiduelles optimales en fonction du rapport $\rho = P_y/P_{b'}$, on s'aperçoit que l'algorithme LMS est toujours supérieur à l'algorithme MCRO pour la poursuite de fluctuations lentes du canal de transmission (fig. 4). Par ailleurs, on constate que les valeurs



Égaliseur transversal à 20 coefficients $\Delta\phi/\Delta t = 200^\circ/\text{min}$ ($\psi = 1,9 \cdot 10^{-4}$ rd) ; Débit : 200 bauds ; S/B = 10 dB (1^{re} ligne)

Fig. 4. — Comparaison des performances en poursuite.

de v_{opt} et de β_{opt} augmentent lorsque ρ augmente, et que les plages de choix sont assez réduites.

V.4. Comportement en présence de ruptures

Un algorithme adaptatif n'est rarement confronté qu'à un seul type de non-stationnarité. Dans la plupart des cas, ruptures et non stationnarités lentes se succèdent, et il est difficile de répondre aux deux situations à la fois lorsque l'on utilise un algorithme à taux d'oubli constant. Ceci dit, il est bien connu que les algorithmes MCRO possèdent un temps de convergence plus faible que celui de l'algorithme LMS (surtout en présence de données corrélées), la comparaison se faisant à *erreur résiduelle donnée*. Il est donc naturel de penser qu'il en est de même pour les situations de rupture qui sont des périodes de convergence avec conditions initiales non nulles sur \underline{V} et sur \underline{R}_{xx} . Notons que si la convergence (apprentissage) de l'algorithme MCRO n'est que peu affectée par la valeur de λ ($\underline{R}_{xx}(0) \approx 0$), il n'en est pas de même pour une rupture où la matrice $\underline{R}_{xx}(n)$ doit être remise en question à partir de l'instant n , le retour au régime permanent étant d'autant plus rapide que la valeur de λ est faible. En acoustique sous-marine, les situations de ruptures interviennent essentiellement sur l'amplitude et la phase des trajets; les retards entre trajets sont souvent plus stables.

Les simulations de la figure 5 illustrent la supériorité des algorithmes MCRO dans ce type de situation. Les algorithmes ont le même taux d'oubli $\nu = \beta = 0,01$ dans les trois cas, avec $N = 40$; le rapport S/B est de 20 dB, et la puissance du signal est $P_x = 1$. Nous avons simulé trois cas de rupture, pour un canal de transmission à deux trajets :

- 1) Rupture sur le rapport d'amplitude des trajets $R = a_1/a_0$: la valeur de R passe brutalement de 0,2 à 0,9, tandis que les retards et les phases des trajets sont fixes ($\tau = 3 T_e$, $\phi_0 = 10^\circ$, $\phi_1 = 45^\circ$). Lorsque $R = 0,2$ ($n < 2550$), la dispersion des valeurs propres de \underline{R}_{xx} est faible, ce qui explique que les deux algorithmes se comportent de façon similaire.
- 2) Rupture sur la phase de R , $\phi = \phi_1 - \phi_0$: la valeur de ϕ passe brutalement de 35° à 135° , tandis que $R = 0,5$ et $\tau = \tau_2 - \tau_1 = 3 T_e$.
- 3) Rupture sur le retard entre trajets τ : la valeur de τ passe de $3 T_e$ à $4 T_e$, tandis que $R = 0,6$, $\phi_0 = 10^\circ$ et $\phi_1 = 45^\circ$.

Dans tous les cas, la décroissance de l'erreur quadratique moyenne après la rupture est plus rapide pour l'algorithme MCRO que pour l'algorithme LMS, ce phénomène étant d'autant plus marqué que le nombre de trajets et l'ordre de l'égaliseur est grand.

V.5. Conclusion

Les systèmes du type « identification inverse » pilotés par un algorithme adaptatif ont un comportement sensiblement différent des cas d'identification pure,

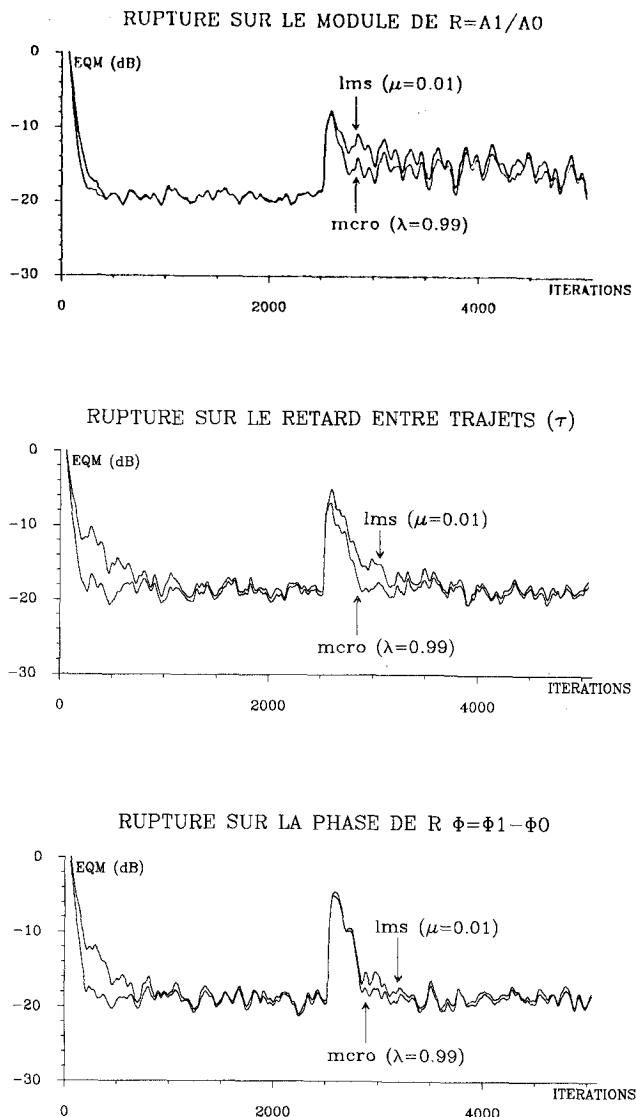


Fig. 5. — Comportement en présence de rupture.

car le bruit intervient à la fois dans la mesure de l'erreur d'estimation de l'algorithme et dans l'expression de la solution optimale de Wiener (c'est aussi le cas de la prédiction).

Nous avons vu dans un cas très particulier de poursuite que l'algorithme LMS est supérieur aux algorithmes du type MCRO en termes d'erreur résiduelle. Il faut noter que cette supériorité se traduit en pratique par des performances paraissant identiques, car le gain en erreur résiduelle n'est pas suffisant pour être discernable de l'erreur totale. Le cas que nous venons d'étudier n'est pas un cas isolé; d'autres auteurs sont parvenus à des résultats similaires [36]. En ce qui concerne les ruptures, on aboutit à un résultat connu depuis longtemps: les algorithmes du type MCRO sont supérieurs.

Pour l'utilisateur d'un algorithme adaptatif, il est donc important de savoir s'il se trouve dans l'une ou l'autre des situations, sachant que le même taux d'oubli ne pourra donner de bonnes performances

dans les deux cas. Un avantage important de l'algorithme LMS est qu'il supporte facilement un taux d'oubli variable, qui peut être ajusté de façon adaptative [62], ce qui procure de très bons résultats particulièrement pour les ruptures. Dans les versions rapides des algorithmes MCRO, il est plus difficile de faire varier le taux d'oubli β (des propriétés importantes de la matrice de covariance estimée $\underline{R}_{xx}(n)$ pouvant être détruites), de part la relative fragilité de ce type d'algorithme; une solution, proposée dans [26], est de ne faire varier que le taux d'oubli de la partie « filtrage » de l'algorithme sans modifier la partie « prédiction »; il est aussi possible d'utiliser un algorithme à fenêtre glissante dont on ajuste la taille en fonction des non-stationnarités rencontrées, ou encore, d'utiliser un facteur d'oubli directement sur les données, comme il est proposé dans [63].

BIBLIOGRAPHIE

- [1] Rapport d'activité GDR 134, *Traitement du Signal et Images*, CNRS, juillet 1989.
- [2] « Processus non stationnaires et traitements adaptatifs », Numéro spécial *Traitement du Signal*, vol. 6, n° 2, 1989.
- [3] S. McLAUGHLIN, B. MULGREW, C. COWAN, *Performance Comparison of LS and LMS as HF Channel Estimators*, Proc. ICASSP, pp. 49.2.1-4, 1987.
- [4] S. GUNNARSSON, L. LJUNG, *Frequency Domain Tracking Characteristics of Adaptive Algorithms*, IEEE Trans. ASSP, vol. 37, pp. 1072-1089, 1989.
- [5] A. BENVENISTE, *Design of Adaptive Algorithms for the Tracking of Time-Varying Systems*, Journal of Adaptive Control and Signal Processing, vol. 1, pp. 3-29, 1987.
- [6] B. WIDROW, J. M. MCCOOL, M. G. LARIMORE, C. R. JOHNSON, *Stationary and Nonstationary Learning Characteristics of the LMS Adaptive Filter*, Proc. IEEE, vol. 64, pp. 1151-1162, 1976.
- [7] D. C. FARDEN, *Tracking Properties of Adaptive Signal Processing Algorithms*, IEEE Trans. ASSP, vol. 29, pp. 439-446, 1981.
- [8] A. BENVENISTE, G. RUGET, *A Measure of the Tracking Capability of Recursive Algorithms with Constant Gains*, IEEE Trans. AC, vol. 27, n° 3, pp. 639-649, 1982.
- [9] O. MACCHI, *Optimization of Adaptive Identification for Time-Varying Filters*, Proc. 23rd CDC Conference, Las Vegas, 1984.
- [10] E. EWEDA, O. MACCHI, *Tracking Error Bounds of Adaptive Nonstationary Filtering*, Automatica, vol. 21, n° 3, pp. 293-302, 1985.
- [11] O. MACCHI, M. BELLANGER, *Le Filtrage Adaptatif Transverse*, Annales des Télécommunications, vol. 5, n° 3, pp. 115-132, 1988.
- [12] L. LJUNG, T. SÖDERSTRÖM, *Theory and Practice of Recursive Identification*, MIT Press, 1983.
- [13] B. WIDROW, S. D. STEARNS, *Adaptive Signal Processing*, Prentice Hall, 1985.
- [14] M. BELLANGER, *Analyse des Signaux et Filtrage Numérique Adaptatif*, Masson, 1989.
- [15] O. MACCHI, N. BERSHAD, M. MBOUP, *Comparison of LMS and LS for Adaptive Enhancement of a Drifting line* (Article soumis 1989).
- [16] J. C. LEE, C. K. UN, *Performance of Transform-Domain LMS Adaptive Digital Filters*, IEEE Trans. ASSP, vol. 34, n° 3, pp. 499-510, June 1986.
- [17] G. FAVIER, *Algorithmes en treillis Adaptatifs*, Annales des Télécommunications, vol. 41, n° 5, pp. 305-317, 1986.
- [18] S. ORFANDIS, *Optimal signal processing*, McMillan, 1985.
- [19] O. MACCHI, E. EWEDA, *Second Order Convergence Analysis of Stochastic Adaptive Linear Filtering*, IEEE Trans. AC, vol. 28, n° 1, pp. 76-85, 1983.
- [20] E. ELEFTHERIOU, D. D. FALCONER, *Tracking properties and Steady-State Performance of RLS Adaptive Filter Algorithms*, IEEE Trans. ASSP, vol. 34, n° 5, pp. 1097-1109, 1986.
- [21] A. GILLOIRE, J. P. JULLIEN, *L'acoustique des salles dans les Télécommunications*, l'Écho des Recherches, n° 127, pp. 43-54, 1^{er} trimestre 1987.
- [22] M. XU, Y. GRENIER, *Time Frequency Domain Adaptive Filters*, Proc. ICASSP, Glasgow, pp. 1154-1157, 1989.
- [23] P. DUHAMEL, J. BENESTY, *Une présentation unifiée du filtrage rapide fournissant tous les intermédiaires entre traitements temporels et fréquentiels*, XII^e colloque GRETSI, Juan Les Pins, pp. 37-41, juin 1989.
- [24] A. BENALLAL, A. GILLOIRE, *A New Method to stabilize Fast RLS Algorithms based on a first-order Model of Propagation of Numerical errors*, Proc. ICASSP, New York, pp. 1373-1376, 1988.
- [25] A. BENALLAL, *Étude des algorithmes des moindres carrés transversaux rapides et application à l'identification des réponses impulsionnelles acoustiques*, Thèse Université de Rennes-1, janvier 1989.
- [26] A. BENALLAL, A. GILLOIRE, *Improvement of the Convergence Speed and of the Tracking Capability of the Numerically Stable FLS Algorithms for Adaptive Filtering*, Proc. ICASSP, Glasgow, pp. 1031-1034, 1989.
- [27] A. BENALLAL, A. GILLOIRE, *Instabilité et stabilité numérique des algorithmes des moindres carrés rapides excités par la parole*, XII^e colloque GRETSI, Juan Les Pins, pp. 509-512, juin 1989.
- [28] A. GILLOIRE, T. PETILLON, *A Comparison of NLMS and Fast RLS algorithms for the identification of time-varying systems with noisy outputs. Application to acoustic echo cancellation*, EUSIPCO, Barcelone, septembre 1990.
- [29] Ch. SERVIÈRE, *Éléments de comparaison entre différentes méthodes de soustraction de bruit*, Thèse I.N.P. Grenoble, mai 1989.
- [30] Ch. SERVIÈRE, D. BAUDOIS, A. SILVENT, P. LORENZINO, *Éléments de comparaison entre différentes méthodes de soustraction de bruit*, XI^e colloque GRETSI, Nice, pp. 113-116, juin 1987.
- [31] R. M. GRAY, *On the asymptotic eigenvalue distribution of Toeplitz matrices*, IEEE Trans. Inform. Theory, vol. 18, pp. 725-730, November 1972.
- [32] J. C. LEE, S. K. MITRA, *On frequency-domain least squares adaptive algorithms*, Proc. IEEE, pp. 411-414, 1987.
- [33] N. J. BERSHAD, P. L. FEINTUCH, F. A. REED, B. FISHER, *Tracking characteristics of the LMS Adaptive Line Enhancer: Response to a Linear Chirp Signal in Noise*, IEEE Trans. ASSP, vol. 28, n° 5, pp. 504-516, October 1980.
- [34] Ch. SERVIÈRE, D. BAUDOIS, J. F. GUERRE-CHALEY, *Noise cancelling: elimination of a chirped sinusoid*, soumis à Signal Processing 1989.
- [35] O. MACCHI, N. BERSHAD, *Adaptive Recovery of a Chirped Sinusoid in noise, Part I. Performance of the RLS Algorithm*, à paraître IEEE Trans. ASSP, 1990.
- [36] N. BERSHAD, O. MACCHI, *Comparison of RLS and LMS algorithms for tracking a chirped signal*, Proc. ICASSP, Glasgow, 1989.
- [37] N. BERSHAD, O. MACCHI, *Adaptive Recovery of a Chirped Sinusoid in noise, Part II. Performance of the LMS Algorithm*, à paraître IEEE Trans. ASSP, 1990.
- [38] Ch. SERVIÈRE, D. BAUDOIS, A. SILVENT, *Comparaison de deux méthodes d'estimation du gain complexe d'un filtre optimal au sens de Wiener*, *Traitement du Signal*, vol. 6, n° 6, 1989.

- [39] D. HERTZ, D. MANSOUR, I. ENGEL, *On Least Square Frequency-Domain Adaptive Filters*, IEEE Trans. on Circuits and Systems, vol. 33, n° 3, pp. 335-337, March 1986.
- [40] E. R. FERRARA, *Fast implementation of LMS adaptive filters*, IEEE Trans. ASSP, vol. 28, pp. 474-475, August 1980.
- [41] B. WIDROW *et al.*, *Adaptive noise cancelling: Principles and applications*, Proc. IEEE, vol. 63, pp. 1692-1716, December 1975.
- [42] L. LJUNG, M. MORF, D. FALCONER, *Fast calculation of gain matrices for recursive estimation schemes*, Int. J. Control, vol. 27, pp. 1-19, January 1978.
- [43] O. MACCHI, M. BELLANGER, *Le point sur le filtrage adaptatif transverse*, XI^e colloque GRETSI, Nice, pp. 1G-14G, juin 1987.
- [44] P. FABRE, C. GUEGUEN, *Fast recursive least squares algorithms: preventing divergence*, Proc. ICASSP, pp. 1149-1152, 1985.
- [45] C. LAMBERT-NEBOUT, *Étude des moyens d'analyse du signal basse fréquence d'un récepteur d'alignement de piste ILS*, Thèse de doctorat de l'I.N.P. de Toulouse, novembre 1989.
- [46] J. CIOFFI, T. KAILATH, *Fast RLS transversal filters for adaptive filtering*, IEEE Trans. ASSP, vol. 32, pp. 304-337, April 1984.
- [47] G. CARAYANNIS, D. MANOLAKIS, N. KALOUPSIDIS, *A fast sequential algorithm for Least-Squares filtering and prediction*, IEEE Trans. ASSP, vol. 31, pp. 1394-1402, December 1983.
- [48] J. ZEIDLER *et al.*, *Adaptive enhancement of multiple sinusoids in uncorrelated noise*, IEEE Trans. ASSP, vol. 26, pp. 240-254, June 1978.
- [49] J. BOUDY, *Optimisation de filtres en treillis non stationnaires et étude comparative de filtres multidimensionnels avec application au traitement d'antenne*, Thèse de docteur-ingénieur, LASSY, Nice, 1988.
- [50] J. BOUDY, G. FAVIER, *Optimization of lattice filters for stationary and nonstationary signals*, EUSIPCO, Grenoble, pp. 591-594, September 1988.
- [51] J. BOUDY, G. FAVIER, *Propriétés et optimisation du taux de convergence des algorithmes en treillis/gradient*, XI^e colloque GRETSI, Nice, pp. 289-292, juin 1987.
- [52] G. FAVIER, *Numerically efficient adaptive identification algorithms*, APII, vol. 22, n° 1, pp. 27-52, 1988.
- [53] G. FAVIER, J. BOUDY, R. SETTINERI, *Filtres en treillis adaptatifs. Comparaison de deux approches dans le cas de signaux non stationnaires*, XII^e colloque GRETSI, Juan Les Pins, pp. 339-342, juin 1989.
- [54] L. J. GRIFFITHS, *A continuously-adaptive filter implemented as a lattice structure*, Proc. ICASSP, pp. 683-686, 1977.
- [55] O. MACCHI, *Optimization of adaptive identification for time-varying filters*, IEEE Tr. Aut. Cont., vol. AC-31, n° 3, pp. 283-287, March 1986.
- [56] B. PORAT *et al.*, *Square root covariance ladder algorithms*, IEEE Trans. Aut. Cont., vol. 27, n° 4, pp. 813-829, 1982.
- [57] R. SETTINERI, G. FAVIER, *Adaptive RLS Lattice filters for fastly non stationary signals*, Proc. ICASSP, pp. 1807-1810, Albuquerque, 1990.
- [58] D. T. LEE, M. MORF, *A novel innovation based on time domain pitch detector*, Proc. ICASSP, Denver, pp. 40-43, 1980.
- [59] M. BASSEVILLE, *The two models approach for the on line detection of changes in AR processes*, Detection of abrupt changes in signals and dynamical systems, Lecture Notes in Control and Information Sciences, Springer Verlag, vol. 77, pp. 169-215, 1986.
- [60] N. MARTIN, *An AR Spectral Analysis of non-stationary signals*, Signal Processing, vol. 10, n° 1, pp. 61-74, January 1986.
- [61] H. BOUCART, *Estimation optimale de signaux de communication transmis dans un canal à trajets multiples*, Thèse de doctorat de l'I.N.P. de Grenoble, janvier 1987.
- [62] Ph. BRAGARD, G. JOURDAIN, *A fast self-optimized LMS algorithm for non stationary identification*, Proc. ICASSP, Albuquerque, 1990.
- [63] B. TOPLIS, S. U. SUPATHY, *Tracking improvements in fast-RLS algorithms using a variable forgetting factor*, IEEE Trans. ASSP, vol. 36, pp. 206-227, February 1988.