COLLOQUE NATIONAL SUR LE TRAITEMENT DU SIGNAL ET SES APPLICATIONS

NICE du 26 au 30 AVRIL 1977

REMARQUES SUR LES METHODES ADAPTATIVES EN ESTIMATION ET DETECTION

Bernard PICINBONO

L2S[†] - ESE - Plateau du Moulon - 91190. GIF

RESUME

Il est de plus courant d'utiliser des systèmes adaptatifs en estimation ou détection. Cet usage est imposé par les fortes variations des paramètres définissant statistiquement les signaux et les bruits. Il est malheureusement difficile de concilier les notions d'optimalité, définies dans un contexte déterminé, etd'adaptativité où ce contexte est variable. De plus l'usage des signaux numériques exige de tenir compte de la récursivité et de la simplicité des algorithmes. Après avoir présenté un cas d'optimalité adaptative on introduit différents modes d'adaptativité, en particulier à horizon fini ou non. On rappelle les solutions récursives dans le cas à horizon infini et on en donne d'autres à horizon limité. On montre enfin comment l'algorithme adaptatif de Widrow peut s'interprêter comme une approximation de ceux présentés précédemment, et même d'une certaine forme de filtrage statistique récursif.

SUMMARY

The use of adaptative systems in estimation and detection is now very common. This fact is due to the strong variations of the parameters which define statistically the signals and the noise. Unfortunately it is difficult to conceive systems which are simultaneously optimal, as in a deterministic context, and adaptative because of the variable context. Moreover with the frequent use of numerical signals it is often necessary to take into account of recursivity and simplicity of the algorithms of signal processing. After presenting a particular case of adaptive optimality we introduce several kinds of adaptive systems, in particular with finite or infinite horizon.

We recall the recursive solutions in the case of a definite horizon and we give some other solutions with finite horizon. Finally we show now the adaptive Widrow algorithm can use interpreted as an approximation of others algorithms and is also connected with the statistical recursive filtering.

REMARQUES SUR LES METHODES ADAPTATIVES EN ESTIMATION ET DETECTION

I. INTRODUCTION

La plupart des systèmes actuels d'estimation ou de détection sont réputés être adaptatifs, ce terme étant pour l'instant pris dans un sens très général. La nécessité d'utiliser des systèmes adaptatifs est devenue rapidement évidente afin de tenir compte au moins approximativement des importantes variations des paramètres qui sont supposés constants et connus dans les méthodes non adaptatives. L'usage, déjà très ancien, mais justifié plus récemment en détection [1] [2] [3], des contrôles automatiques de gain (C.A.G.), est déjà une confirmation de l'intérêt des méthodes adaptatives.

Malheureusement les théories de l'estimation et de la détection sont présentées généralement de manière non adaptative. Ces théories supposent connues un certain nombre de propriétés concernant les signaux et les bruits et en déduisent des structures optimales compte-tenu de certains critères.

On peut donc prévoir dès le départ une sorte de tension entre optimalité et adaptativité. Examinons ceci sur deux exemples très classiques.

La théorie de l'estimation linéaire en moyenne quadratique (E.L.M.Q.) permet de trouver l'estimateur optimal $\hat{y}$ d'une grandeur aléatoire y (supposée scalaire pour simplifier) à partir d'une observation $\vec{X}$. Pour calculer $\hat{y}$ il faut connaître les propriétés du second ordre de la variable aléatoire (v.a.) multidimensionnelle $(y, \vec{X})$. Comme ceci n'est pas le cas en général on peut être tenté d'estimer ces propriétés, lorsque cela est possible, et de remplacer les grandeurs inconnues par leur estimation dans une structure qui est optimale lorsque ces grandeurs sont connues. On obtient ainsi une structure qui est bien adaptative, mais dont le caractère optimal n'est pas prouvé.

Le même problème se présente en détection d'un signal déterministe dans un bruit gaussien. Le récepteur optimal est le filtre adapté qui suppose connue la covariance du bruit. Si celle-ci n'est qu'estimée on obtient un filtre adapté adaptatif mais dont l'optimalité reste à prouver. Ce problème d'optimalité est en fait très complexe en détection et a été étudié en détail par ailleurs [2] [4].

La notion d'adaptativité est elle-même trompeuse car elle recouvre, comme nous le verrons, des réalités parfois opposées.

Enfin deux autres notions viennent en général compliquer la situation, celles de simplicité et de récursivité. Il est possible de trouver dans certains cas des systèmes optimaux et adaptatifs mais dont la complexité est telle qu'on ne peut envisager de les réaliser. Cette complexité peut être d'ordre technologique ou numérique.

Enfin l'usage des suites numériques à la place des systèmes continus conduit très souvent à rechercher des solutions récursives au lieu de solutions globales nécessitant un trop grand stockage d'information.

II. OPTIMALITE

Toute la suite de cet exposé sera essentiellement consacrée à des problèmes d'estimation, ceux de détection étant généralement plus complexes et ayant été examinés par ailleurs [1] [2] [4].

2.1. Optimalité statistique

Soit y la grandeur à estimer à partir d'une observation $\vec{X}$. C'est par exemple le cas de la prédiction linéaire à horizon fini pour un processus discret x_n. Dans ce cas

$$y = x_n \quad \text{et} \quad \vec{X}^T = (x_{n-1}, x_{n-2}, \dots, x_{n-r}).$$

Si l'on connait la v.a. $(y, \vec{X})$ au second ordre, c'est-à-dire les covariances

$$\Gamma_x = E(X X^T) \quad \text{et} \quad \Gamma_{xy} = E(X y) \qquad (1)$$

l'estimateur optimal rendant minimum $E[(y-\vec{a}^T\vec{X})^2]$ est donné par la formule classique

$$\hat{y} = \vec{a}^T \vec{X} \qquad (2)$$

avec

$$\vec{a} = \Gamma_x^{-1} \Gamma_{xy} \qquad (3)$$

Cet estimateur n'est évidemment pas adaptatif. Il n'est ni simple, en raison de l'inversion de Γ_x, ni récursif, mais peut le devenir par les manipulations classiques conduisant au filtrage de Kalman.

2.2. Optimalité adaptative

Supposons qu'on ait fait une suite d'observations $(y_i, \vec{X}_i)$, $1 \leq i \leq n$, et que l'on veuille en déduire un vecteur $\vec{a}_n$ rendant minimum

$$\sum_{1}^{n} i \; (y_i - \vec{a}_n \vec{X}_i)^2. \qquad (4)$$

Cet estimateur est évidemment celui des moindres carrés, et dans le cas gaussien correspond à celui du maximum de vraisemblance. On trouve alors très simplement

$$\vec{a}_n = \Gamma_n^{-1} \vec{C}_n \qquad (5)$$

avec

$$\Gamma_n = \sum_{1}^{n} i \; X_i X_i^T \qquad (6)$$

$$\vec{C}_n = \sum_{1}^{n} i \; y_i X_i \qquad (7)$$

Bien évidemment l'inversion de Γ_n qui se présente en (6) comme une somme de projecteurs nécessite que n soit supérieur à r, dimension du vecteur $\vec{a}$. On admettra ceci dans toute la suite.

On a donc construit un estimateur qui est optimal (aux sens précisés ci-dessus) et adaptatif, puisqu'il se construit sans aucune connaissance préalable des propriétés de (y_i, X_i).

Il est également très intéressant de remarquer que cet estimateur correspond à la structure évoquée dans l'introduction. On prend en effet la solution optimale (3) dans laquelle on remplace les covariances par leur estimation données au facteur 1/n près par les Equations (6) et (7).

III. ADAPTATIVITES.

La solution (5)(6)(7) est adaptative mais peut dans la pratique ne pas être la plus intéressante.

En effet elle intègre toute l'évolution du processus de l'instant 1 à n. Certes quand n sera grand la contribution dans (6) et (7) des termes correspondants à $i \ll n$ sera relativement faible, mais apparaîtra cependant.

Si l'on a des raisons de penser que les propriétés du processus n'ont pas une stabilité très grande on peut tenter de trouver des solutions adaptatives à horizon limité. Dans ce cas il suffira dans (4) de remplacer la sommation de 1 à n par une sommation de n-N à n. On retrouvera évidemment les solutions (6) et (7) pour lesquelles les sommations vont de n-N à n.

Il y aura finalement autant de solutions adaptatives que de valeurs de N. En principe une seule est réellement optimale en régime stationnaire.

Mais un nouvel inconvénient apparait évidemment.

En régime stationnaire on voit assez aisément que lorsque $n \to \infty$, $\vec{a}_n$ converge vers la solution $\vec{a}$ donnée par (3).

Dans la solution adaptative à horizon fini ceci ne pourra plus arriver et il restera obligatoirement une fluctuation résiduelle de $\vec{a}_n, \forall n$, d'autant plus faible que N est plus grand.

Le problème que nous venons d'évoquer est bien connu dans la présentation élémentaire des CAG, puisque le choix de la constante de temps est un problème important et délicat. Le choix d'une constante de temps très grande permet en principe une bonne estimation de la puissance, mais ceci devient illusoire dès lors que les nonstationnarités que l'on veut contrôler ont un temps de corrélation trop petit. Ce type de compromis se retrouve dans la plupart des systèmes adaptatifs.

Pour spécifier les termes utilisés dans la suite on parlera de solutions adaptatives à horizon non limité (H.N.L.), applicables dans le cas stationnaire et pouvant conduire en attendant suffisamment longtemps à des erreurs nulles, opposées aux solutions à horizon limité (H.L.) qui sont les plus intéressantes dans la pratique mais comportant en général des fluctuations résiduelles.

IV. ESTIMATION ADAPTATIVE H.N.L.OPTIMALE & RECURSIVE

Il s'agit de mettre sous forme récursive les équations (5)(6)(7) donnant une solution $\vec{a}_n$ qui est par construction optimale et adaptative H.N.L.

La procédure est relativement classique et part de la formule d'inversion suivante

$$[A + b\, b^T]^{-1} = A^{-1} - \gamma\, A^{-1}\, b\, b^T\, A^{-1} \qquad (8)$$

où

$$\gamma = (1 + b^T\, A^{-1}\, b)^{-1} \qquad (9)$$

On suppose évidemment dans cette expression que A^{-1}, inverse de A, existe.

Appliquons cette formule à l'inversion de Γ_n donnée par (6) qui peut s'écrire

$$\Gamma_n = \Gamma_{n-1} + X_n\, X_n^T \qquad (10)$$

Si l'on pose

$$K_n \triangleq \Gamma_n^{-1} \qquad (11)$$

on trouve immédiatement par application de (8) et (9)

$$K_n = K_{n-1} - \frac{1}{1 + X_n^T\, K_{n-1}\, X_n} \cdot K_{n-1} X_n X_n^T\, K_{n-1} \qquad (12)$$

Cette équation permet de calculer récursivement la matrice K_n à partir des observations X_n.

Il se pose évidemment un problème d'initialisation de l'algorithme (12) puisque évidemment $X_n X_n^T$ est de rang 1 et n'a pas d'inverse. Mais le problème existe également sur (5)-(7), et on ne pourra calculer Γ_n^{-1} qu'à partir d'une certaine valeur de n, l'algorithme ne pouvant débuter qu'à ce moment. On pourrait également utiliser les pseudoinverses, mais nous n'analysons pas ce point dans la suite.

On peut maintenant donner une forme résursive à la quantité $\vec{a}_n$ définie par (5). Pour ceci on écrit $\vec{C}_n$ défini par (7) sous la forme

$$\vec{C}_n = \vec{C}_{n-1} + y_n\, X_n \qquad (13)$$

d'où l'on déduit

$$\vec{a}_n = [\Gamma_{n-1} + X_n\, X_n^T]^{-1}\, [\vec{C}_{n-1} + y_n\, X_n] \qquad (14)$$

En appliquant l'équation (8) et en notant de plus que

$$[\Gamma_{n-1} + X_n\, X_n^T]^{-1}\, X_n = \gamma\, \Gamma_{n-1}^{-1}\, X_n, \qquad (15)$$

on trouve finalement

$$\vec{a}_n = \vec{a}_{n-1} + K_n\, \vec{X}_n\, (y_n - \vec{X}_n^T\, \vec{a}_{n-1}) \qquad (16)$$

Les algorithmes (12) et (16) permettent donc de calculer de manière récursive à chaque instant n la solution $\vec{a}_n$ donnée par (5). Comme $\vec{a}_n$ est optimal et adaptatif, la solution récursive conserve ces propriétés.

Cette solution exacte n'est évidemment pas simple puisqu'elle nécessite à chaque étape n le calcul d'une matrice K_n et de divers produits de matrice. Mais il n'y a aucune inversion de matrice, à la différence de (5).

Si K_n n'est pas donné par (12), l'algorithme (16) ne donne pas à chaque instant la valeur exacte de $\vec{a}_n$. Cependant il est bien connu qu'il existe une large classe de K_n tels que $\vec{a}_n$ converge presque surement vers $\vec{a}$, limite de l'Eq. (5) quand $n \to \infty$ [5].

On a alors un algorithme stochastique qui est asymptotiquement optimal. Sa simplicité dépend beaucoup de la nature de K_n.

V. ESTIMATION ADAPTATIVE H.L. RECURSIVE

La solution du paragraphe précédent n'étant qu'une forme d'écriture particulière (récursive) de l'Eq. (5) prend en compte à chaque instant toutes les informations du passé avec un poids égal. Elle est bien adaptée à un cas stationnaire, mais il est préférable, quitte à perdre le caractère optimal, de trouver une solution ne prenant en compte le passé qu'avec un horizon fini.

Plusieurs solutions peuvent être envisagées et nous en présentons ici une de type "exponentiel". Remplaçons les Eqs. (10) et (13) par

$$\Gamma_n = \alpha\, \Gamma_{n-1} + X_n X_n^T \qquad (17)$$

$$\vec{C}_n = \beta\, \vec{C}_{n-1} + y_n X_n \qquad (18)$$

α et β étant en valeur absolue inférieurs à 1. L'introduction de ces coefficients permet de remplacer les solutions de type (6) ou (7) par d'autres où les informations $X_n X_n^T$ et $y_n X_n$ sontpondérées par des coefficients type α^k et β^k limitant l'influence du passé trop lointain.

Reprenant tout le calcul précédent on peut écrire la solution (5) sous une forme analogue à (12) et (16)

$$K_n = \frac{1}{\alpha}.\ K_{n-1} - \frac{1}{\alpha^2} \frac{1}{\frac{1}{\alpha} X_n^T K_{n-1} X_n} .K_{n-1}X_nX_n^TK_{n-1} \qquad (19)$$

$$\vec{a}_n = \frac{\beta}{\alpha}\, \vec{a}_{n-1} + K_n \vec{X}_n \left(y_n - \frac{\beta}{\alpha} X_n^T a_{n-1}\right) \qquad (20)$$

En particulier si onimpose $\alpha = \beta$, on constate que l'algorithme donnant $\vec{a}_n$ n'a pas changé de structure. Seuls les K_n ont été modifiés, rendant (20) adaptatif H.L. Bien évidemment (19) reste toujours compliqué à mettre en oeuvre.

VI. SOLUTIONS RECURSIVES APPROCHEES

La mise en oeuvre de l'algorithme (16) nécessite à chaque étape n le calcul du vecteur $K_n \vec{X}_n$, où K_n est la matrice Γ_n^{-1} calculée par (12).

On évite très souvent cette opération en remplaçant K_n par un scalaire $2\mu_n$. On obtient alors un algorithme très simple, ne donnant qu'une approximation de $\vec{a}_n$ mais pouvant être asymptotiquement optimal [5].

L'étape simplificatrice suivante consiste évidemment à remplacer $2\mu_n$ par une constante 2μ, ce qui donne alors l'algorithme classique dit de Widrow [6].

Par rapport aux algorithmes exacts donnant la solution des moindres carrés, le fait de remplacer une matrice aléatoire K_n dépendant des observations par un paramètre scalaire constant apporte une simplification considérable. Il est toutefois certain que ce que l'on gagne en simplicité doit être perdu soit en optimalité soit en vitesse de convergence.

L'algorithme de Windrow étant de loin le plus utilisé, nous allons montrer comment il peut être obtenu par différentes approximations.

6.1. Développement limite de Γ_n^{-1}

On peut écrire (10) sous la forme

$$\Gamma_n = \Gamma_{n-1} [I + \Gamma_{n-1}^{-1} X_n X_n^T] \qquad (21)$$

d'où par inversion

$$\Gamma_n^{-1} = [I + \Gamma_{n-1}^{-1} X_n X_n^T]^{-1} \Gamma_{n-1}^{-1} . \qquad (22)$$

On peut développer en série la matrice $[I + \Gamma_{n-1}^{-1} X_n X_n^T]^{-1}$ et limiter ce développement en notant que d'après (10) le terme $X_n X_n^T$ est un complément petit par rapport à Γ_{n-1} dès que n est assez grand. Il vient alors en remplaçant les Γ^{-1} par K

$$K_n \approx [I - K_{n-1} X_n X_n^T] K_{n-1} \qquad (22)$$

Dans ces conditions l'Eq. (14) devient, en notant que $\vec{a}_n = K_n \vec{C}_n$,

$$\vec{a}_n = [I - K_{n-1}X_nX_n^T]\ [\vec{a}_{n-1} + K_{n-1}y_n\vec{X}_n] \qquad (23)$$

Par une nouvelle approximation consistant à négliger le terme $K_{n-1} X_n X_n^T K_{n-1} y_n X_n$, il vient alors

$$\vec{a}_n = \vec{a}_{n-1} + K_{n-1} \vec{X}_n [y_n - \vec{X}_n^T \vec{a}_{n-1}] \qquad (24)$$

On retrouve donc ainsi très rapidement un algorithme ne différant de (16) que par le remplacement de K_n par K_{n-1}. Ceci disparaît dans l'algorithme de Windrow où K_n est remplacé par une constante 2μ.

6.2. Extension du filtrage de Kalman

Les algorithmes précédents ont une parenté évidente avec ceux rencontrés dans le filtrage de Kalman. Il est donc intéressant d'approfondir cette analogie.

Rappelons que pour ceci on considère le système représenté par les équations d'état et d'observations

$$x_{n+1} = A x_n + u_n \qquad (25)$$

$$y_n = B_n x_n + v_n \qquad (26)$$

où u_n et v_n sont des bruits blancs indépendants. Le filtrage de Kalman a pour but de calculer de manière récursive l'estimation linéaire $\hat{x}_n$ de x_n en fonction de $y_0, y_1, \dots y_{n-1}$. Le résultat s'écrit

$$\hat{x}_{n+1} = A \hat{x}_n + G_n (y_n - B_n \hat{x}_n), \qquad (27)$$

où le gain G_n est donné par l'équation

$$G_n = A P_n B_n^T [B_n P_n B_n^T + Q]^{-1}. \qquad (28)$$

Dans cette équation Q est la variance de v, qui est un scalaire comme y_n.

Par ailleurs la quantité P_n peut aussi se calculer de manière récursive par une équation du type Riccati discret qui s'écrit

$$P_{n+1} = A P_n A^T + R - A P_n B_n^T [B_n P_n B_n^T + Q]^{-1} B_n P_n A^T \qquad (29)$$

où R est la matrice variance de u_n.

Pour comparer avec les algorithmes (16) et (27), posons $\vec{\alpha}_n = \vec{a}_{n-1}$, ce qui est une simple translation dans le temps, permettant d'écrire (16) sous la forme

$$\vec{\alpha}_{n+1} = \vec{\alpha}_n + K_n \vec{X}_n (y_n - \vec{X}_n^T \vec{\alpha}_n) \qquad (30)$$

Le modèle de type (25) (26) générant (30) où $\vec{\alpha}_n = \hat{x}_n$ s'obtient en posant

$$A = I \qquad B_n = X_n^T \qquad (31)$$

En remplaçant A et B_n par ces valeurs dans (28) on obtient

$$G_n = K_n X_n \qquad (32)$$

avec

$$K_n = \frac{P_n}{Q + X_n^T P_n X_n} \qquad (33)$$

Ainsi les algorithmes (27) et (30) sont identiques, K_n étant défini par (33). L'équation (29) récursive sur P_n s'écrit aussi

$$P_{n+1} = P_n + R - \frac{1}{Q + X_n^T P_n X_n} P_n X_n X_n^T P_n \qquad (34)$$

qui doit être considérée comme l'analogue de (12).

On constate donc une identité entre les algorithmes d'estimation et une différence sur les algorithmes du gain.

Cette différence disparaît évidemment lorsque l'on prend un algorithme du type Widrow où K_n est remplacé simplement par un scalaire.

Il faut bien noter que le modèle (25)(26)(31) ne correspond pas exactement à la représentation habituelle des systèmes puisque la matrice B_n est maintenant aléatoire et fait partie des observations.

6.3. Adaptativité de l'algorithme de Widrow.

Comme nous l'avons indiqué il se déduit de (16) en remplaçant la matrice K_n par le scalaire 2μ. Il s'écrit donc

$$\vec{a}_n = \vec{a}_{n-1} + 2\mu \vec{X}_n (y_n - \vec{X}_n^T \vec{a}_n) \tag{35}$$

On peut se demander quel type d'adaptativité possède cet algorithme très simplifié par rapport à (16) ou (20). Cette question est abordée dans un autre exposé de ce colloque [7], et il apparaît que (35) possède une adaptativité H.L., l'horizon dépendant précisément de la valeur choisie pour μ. Etant H.L. cet algorithme ne convergera pas vers une valeur déterminée, mais il y aura toujours des fluctuations résiduelles de $\vec{a}_n$.

Il serait intéressant d'étudier si en remplaçant μ par μ_n déduit d'une équation du type de (19) convenablement simplifiée il ne serait pas possible d'obtenir des fluctuations résiduelles plus faibles tout en conservant le même horizon d'adaptativité.

R E F E R E N C E S

[1] G.VEZZOSI et B.PICINBONO, Détection d'un signal certain dans un bruit sphériquement invariant. Structure et Comparaison des différents récepteurs, Ann. Télécomm.27,95, 1972

[2] G.VEZZOSI, What is optimality for an adaptative detection system, NATO Advances Study Institute On Signal Processing, 1973, Academic Press.

[3] B.PICINBONO, Introduction to detection and estimation, NATO Advance Study Institute, La Spezia, 1976.

[4] G.VEZZOSI, Détection d'un signal dans un bruit de loi incertaine, Thèse de Doctorat-d'Etat, Orsay, 1976.

[5] C.MACCHI, Itération stochastique et traitements numériques adaptatifs, Thèse de Doctorat-d'Etat, Paris, 1972.

[6] B.WIDROW et al. Adaptative noise canceling : principles and applications, Pr. IEEE, 63, 1962, 1975.

[7] O.MACCHI, exposé au même Colloque.

† Laboratoire propre du CNRS et de l'ESE.