Régression Linéaire en Présence de Bruit: Une Solution Adaptative Non-Biaisée

David Gesbert, Pierre Duhamel*, Michael Unser**

CNET/PAB/RGF/RCN 38-40 rue du Général Leclerc, 92131 Issy-les-Mlx
*Telecom Paris/SIG 46 rue Barrault, 75634 Paris cedex 13
**National Institute of Health (USA)

Résumé

Dans cet article, nous proposons une méthode adaptative à faible complexité permettant le calcul *sans biais* d'un vecteur de régression optimal au sens de l'erreur quadratique moyenne (EQM), basé sur des données perturbées par un bruit de mesure additif blanc. La technique décrite s'applique à toute situation où les variables de régression non bruitées sont très corrélées entre elles. C'est le cas notamment de la prédiction linéaire, appliquée au signal de parole en vue du codage et, depuis peu, à l'identification aveugle de canaux de radiocommunication multicapteurs. L'algorithme est un LMS classique, auquel on adjoint une contrainte quadratique. On utilise la technique de projection du gradient, dont le principe est depuis longtemps exploité en formation de voie. Il fournit, en plus du vecteur de régression débruité, une estimée du niveau de bruit.

Abstract

This paper presents a low cost adaptive method allowing the computation of an *unbiased* regression vector, optimal in the minimum mean square error sense, in the case the data are corrupted by an additive white noise. The proposed technique applies whenever the regression variables are mutually highly correlated. This condition is met in important practical situations such as linear prediction, applied to speech encoding, or more recently to the blind identification of FIR multichannels. The algorithm is a LMS with quadratic constraint. We resort to the gradient projection technique, used in adaptive beamforming. The proposed algorithm provides estimates of the noiseless regression vector and of the noise variance.

1 Introduction

Le filtrage de Wiener, la prédiction linéaire, l'estimation linéaire relèvent d'une problématique commune: la régression linéaire (RL) de variables aléatoires. Lorsque les variables sont bruitées, la minimisation de l'EQM conduit à des paramètres de régression biaisés. Ceci pose un problème sérieux si le caractère non-biaisé de la RL est indispensable pour l'utilisation qui en est faite dans l'applications considérée. Citons par exemple le codage prédictif de parole, où le biais sur les coefficients de prédiction induit une déformation du signal restitué. Citons aussi le cas de la prédiction linéaire de signaux de radiocommunication issus d'un réseau multicapteurs, pour lesquels le prédicteur, *s'il est non-bruité* , permet l'identification des canaux de propagation [1,2].

Nous faisons ici l'hypothèse de variables de régression très corrélées. Notons que cette condition est réalisée dans les exemples cités. Nous montrons alors théoriquement qu'une contrainte bien choisie sur la norme du vecteur de régression, imposée lors de la minimisation de l'EQM, conduit au débruitage de la solution. Nous proposons un algorithme adaptatif qui effectue efficacement cette minimisation sous contrainte (LMS avec projection de gradient) et qui fournit, en plus, une estimée du niveau de bruit présent sur les données. Nous validons la méthode par des simulations dans le cadre de la prédiction linéaire de signaux multicapteurs.

2 Régression linéaire bruitée

2.1 Formulation

On considère deux processus aléatoires complexes y_n et X_n conjointement stationnaires, où y_n est scalaire et X_n est vectoriel de taille $K \times 1$. La RL au sens du minimum de l'EQM est la recherche des vecteurs $P_0 \in \mathbf{C}^K$ réalisant le minimum du coût $J_0(P) = E \mid y_n - P^+ X_n \mid^2$. E désignant ici l'espérance mathématique et $^+$ l'opérateur de trans-conjugaison. Les équations normales s'écrivent classiquement:

$$R_x P_0 = r_{xy} \tag{1}$$

où $R_x = E(X_n X_n^+)$ et $r_{xy} = E(X_n y_n^*)$ sont la matrice de covariance du processus X_n et le vecteur d'intercorrélation du couple (X_n, y_n). $*$ étant l'opérateur de conjugaison complexe. Dans tout l'article, on suppose les variables de régression très corrélées, ce qui peut s'exprimer par

- **H1** R_x est singulière, de rang $p < K$

Dans la pratique, il arrive fréquemment que les véritables variables de régression soient inaccessibles et que seules des mesures bruitées de y_n et X_n soient disponibles. Notons ces mesures $u_n = y_n + v_n$ et $Z_n = X_n + B_n$, où v_n est un bruit scalaire centré supposé décorrélé à tout instant de B_n, y_n et X_n. B_n est un bruit vectoriel centré décorrélé de y_n, X_n et pour lequel on supposera $E(B_n B_n^+) = \sigma_b^2 I$. Dans le cadre particulier de la prédiction linéaire, v_n et B_n contiennent les échantillons d'un même bruit, qui doit donc être supposé blanc. Le calcul du vecteur de régression optimal P_1 sur les données bruitées est alors biaisé. P_1 est en effet solution de

$$R_z P_1 = r_{zu} \quad \text{avec} \quad \left\{ \begin{array}{l} R_z = R_x + \sigma_b^2 I \\ r_{zu} = r_{xy} \end{array} \right. \qquad (2)$$

d'où $P_1 - P_0 = -\sigma_b^2 R_z^{-1} P_0$.

2.2 Le débruitage

Notre objectif est de compenser le biais en retrouvant *l'une des solutions* P_0 de (1). Les méthodes de débruitage existantes, qu'elles soient de type bloc (il suffit alors de retirer $\sigma_b^2 I$ sur la matrice R_z) ou adaptatives (algorithme γ-LMS [3]) nécessitent une connaissance a priori du niveau de bruit. Dans cet article, nous proposons une nouvelle méthode de débruitage adaptative, très simple, qui ne demande pas cette connaissance et qui, de plus, fournit une estimée de σ_b^2.

3 Régression sous contrainte

Pour tout vecteur $P \in \mathbf{C}^K$, notons $J(P) = E \mid u_n - P^+ Z_n \mid^2$ l'EQM définie sur les observations bruitées. Soit γ un réel positif. Dans la suite, on s'intéresse au problème de la minimisation du critère $J(P)$, sous la contrainte $\| P \| = \gamma$. Nous montrons tout d'abord le lien entre cette contrainte de norme et le débruitage. Nous proposons ensuite une implantation adaptative simple.

3.1 Points stationnaires

Lemme 1 *Soit $\mathcal{S}_\gamma$ l'ensemble des points P annulant le gradient de $J(P)$ sous contrainte $\| P \| = \gamma$. Les éléments de $\mathcal{S}_\gamma$ sont les vecteurs de norme γ, pour lesquels il existe μ réel t.q.*

$$(R_z + \mu I) P = r_{xy} \qquad (3)$$

Preuve Les points de $\mathcal{S}_\gamma$ sont aussi les points stationnaires de la fonction auxiliaire $C(P, \mu) = J(P) + \mu(P^+ P - \gamma)$, où μ est un multiplicateur de Lagrange réel. Le critère EQM bruité se développe selon:

$$J(P) = E \mid u_n \mid^2 + P^+ R_z P - P^+ r_{zu} - r_{zu}^+ P \qquad (4)$$

On trouve

$$\frac{\partial C}{\partial P} = P^+ R_z - r_{zu}^+ + \mu P^+$$

Puis, sachant que $r_{xy} = r_{zu}$, (3) est donné par $\frac{\partial C}{\partial P} = 0$.

Ce résultat indique qu'une contrainte imposée sur la norme du vecteur de régression se traduit par l'addition d'un réel μ sur les termes diagonaux de R_z. Donc, l'apport de la contrainte dans le critère EQM correspond à une augmentation ou une diminution du biais selon le signe de μ. On vérifie facilement que le biais disparait complètement en forçant $\mu = -\sigma_b^2$, puisqu'on retombe alors sur une solution de (1). Notons que les valeurs de γ conduisant au débruitage (i.e. $\mu = -\sigma_b^2$) nous sont a priori inconnues. Dans ce qui suit, nous montrons qu'il est en fait suffisant de surestimer *l'une* de ces valeurs pour résoudre le problème à l'aide d'un algorithme adaptatif de type LMS.

3.2 Stabilité

Nous supposons maintenant l'existence d'un algorithme adaptatif du type gradient stochastique, noté A_γ, minimisant $J(P)$ sous la contrainte $\| P \| = \gamma$. Les points stationnaires de A_γ sont par construction les points de $\mathcal{S}_\gamma$. En réalité, seuls les points stationnaires stables sont des points de convergence possible de A_γ. Le lemme 2 en donne une caractérisation.

Lemme 2 *Un vecteur P de norme γ, vérifiant* (3) *pour un certain réel μ, est un point stationnaire* stable *de $J(P)$ sous la contrainte $\| P \| = \gamma$, si et seulement si $\mu \geq -\sigma_b^2$.*

Preuve Considérons un déplacement ϵD autour d'un point stationnaire P, où D désigne un vecteur normé quelconque et ϵ est l'amplitude du déplacement. ϵ est aussi petite que l'on veut, et telle que le vecteur $P + \epsilon D$ reste dans la sphère de rayon γ. L'accroissement d'EQM dans la direction D s'écrit d'après (4):

$$\begin{aligned} \Delta J &\stackrel{\text{def}}{=} J(P + \epsilon D) - J(P) \\ \Delta J &= \mid \epsilon^2 \mid D^+ R_z D + 2\Re\{P^+ R_z D \epsilon\} - 2\Re\{r_{zu}^+ D \epsilon\} \end{aligned}$$

où $\Re\{\}$ est la partie réelle d'un complexe. sachant (3), nous trouvons

$$\Delta J = \mid \epsilon^2 \mid D^+ R_z D - 2\mu \Re\{P^+ D \epsilon\}$$

enfin, par contrainte, $\| P + \epsilon D \| = \| P \| = \gamma$, d'où

$$\Delta J = \mid \epsilon^2 \mid D^+ (R_z + \mu I) D \qquad (5)$$

P est stable si et seulement si ΔJ reste positif dans toutes les directions D (i.e. $R_z + \mu I \geq 0$). Sous l'hypothèse de singularité de R_x, la plus petite valeur propre de R_z est σ_b^2. La condition de stabilité devient alors $\mu \geq -\sigma_b^2$.

Lemme 3 *Soit $\gamma_0 = \| (R_z - \sigma_b^2 I)^{\#} r_{xy} \|$ où $^{\#}$ est l'inverse au sens de Moore-Penrose. Pour $\gamma > \gamma_0$, tout point stationaire stable de A_γ correspond à une solution de (1).*

Preuve Soit P une solution de (3). D'après le lemme 2, la stabilité de P n'est assurée que pour $\mu = -\sigma_b^2$ et $\mu > -\sigma_b^2$.

$\mu = -\sigma_b^2$ Si la matrice R_x est singulière, le système d'équations (3) est dégénéré, admettant une infinité de solutions débruitées de la forme $P_{part} + g$ où P_{part} est une solution particulière et g appartient au noyau $Ker(R_x)$. La norme de P peut alors prendre toute valeur dans l'intervalle $[\gamma_0; +\infty[$. γ_0 désigne donc la plus petite norme possible pour un vecteur de régression débruité.

$\mu > -\sigma_b^2$ Le système (3) n'est pas dégénéré et la solution P reste bruitée. Mais on note alors que

$\| P \| < \gamma_0$ nécessairement. En effet, si $R_z = \sum_{i=1,K} \lambda_i V_i V_i^+$, avec $\lambda_1 \geq .. > \lambda_{p+1} = .. = \lambda_K = \sigma_b^2$ est une décomposition propre orthonormale de R_z, alors $V_i^+ r_{xy} = 0$, $i > p$, et le vecteur de régression contraint P est donné par

$$(R_z + \mu I)^{-1} r_{xy} = \sum_{i=1}^{p} (\lambda_i + \mu)^{-1} V_i V_i^+ r_{xy} \qquad (6)$$

d'où, en norme,

$$\begin{aligned} \| P \|^2 &= \sum_{i=1}^{p} (\lambda_i + \mu)^{-2} \mid V_i^+ r_{xy} \mid^2 \\ \| P \|^2 &< \sum_{i=1}^{p} (\lambda_i - \sigma_b^2)^{-2} \mid V_i^+ r_{xy} \mid^2 = \gamma_0^2 \end{aligned}$$

En conséquence, un point stationnaire P de norme supérieure à γ_0 n'est stable que pour $\mu = -\sigma_b^2$ (voir fig.1).

Le lemme 3 suggère l'utilisation du critère EQM contraint, où γ est choisi assez grand pour que $\gamma \geq \gamma_0$. Cette condition n'est pas trop contraignante en pratique dans la mesure où γ_0 est indépendant du niveau de bruit. Soit P un point de convergence possible pour A_γ, avec $\gamma \geq \gamma_0$. P est alors l'une des solutions possibles du problème débruité (remarquons qu'il peut y en avoir plusieurs de même norme γ):

$$(R_z - \sigma_b^2 I) P = r_{zu} \qquad (7)$$

3.3 Estimation de σ_b^2

Le calcul de l'un des vecteurs de régression débruités permet d'estimer simplement le niveau de bruit. En effet, (7) se reécrit

$$\begin{aligned} E(Z_n(Z_n^+ P - u_n^*)) &= \sigma_b^2 P \\ \text{puis} \quad \frac{\| E(Z_n e_n^*) \|}{\gamma} &= \sigma_b^2 \qquad (8) \end{aligned}$$

où $e_n = u_n - P^+ Z_n$ est l'erreur d'estimation. (8) suggère une procédure peu coûteuse pour le calcul de σ_b^2, dans laquelle une estimée empirique de $E(Z_n e_n^*)$ est utilisée.

3.4 Implantation

La minimisation adaptative du critère $J(P)$ sous contrainte $\| P \| = \gamma$ peut être effectuée par l'algorithme de projection de gradient (APG) stochastique, classiquement utilisé en formation de voie [4]. La mise à jour de cet algorithme est la même que celle du LMS, à laquelle on ajoute une étape de projection du résultat sur l'espace de contrainte, qui est ici la sphère centrée de rayon γ. Comme tout vecteur de $\mathbf{C}^K$ est normal à la sphère de n'importe quel rayon, la projection est une simple mise à l'echelle. L'algorithme est le suivant, où P_n est le vecteur de régression débruité courant, au temps n, et α est le pas d'adaptation.

$$\begin{aligned} e_n &= u_n - P_{n-1}^+ Z_n \\ \tilde{P}_n &= P_{n-1} + \alpha Z_n e_n^* \\ P_n &= \frac{\tilde{P}_n \gamma}{\| \tilde{P}_n \|} \end{aligned}$$

3.5 Simulations

Nous donnons ici un exemple du comportement de l'algorithme proposé dans le cadre de la prédiction linéaire de signaux de communications numériques reçus au niveau d'une antenne multi-capteurs. La méthode proposée est bien adaptée à ce contexte, puisque les signaux d'entrée de l'algorithme adaptatif sont tels que leur matrice de corrélation R_x est singulière dès que les canaux de propagation sont de longueurs finies [1,2]. Les paramètres pris pour la génération des signaux sont ceux de [5]. On a $\sigma_b^2 = 0.1$ (SNR=10dB) et $K = 8$. γ_0 vaut 1.78. Le pas d'adaptation est $\alpha = 0.002$.

Les figures 2 à 4 montrent l'évolution de l'excès relatif d'EQM, défini sur les données non bruitées par $10log[(J_0(P_n) - J_0min)/J_0min]$ où P_n est le prédicteur fourni à l'itération n par l'APG et J_0min est la valeur minimum de la fonction $J_0()$.

- Dans la fig.2, on choisit $\gamma = 2 > \gamma_0$ et l'on donne pour comparaison le comportement du LMS classique (sans contrainte). Le débruitage remplit sa fonction puisque l'excès d'EQM est nettement diminué.
- Dans la fig.3, on choisit $\gamma = 4 >> \gamma_0$. On constate une augmentation de la variance, qui masque l'efficacité de la méthode, et un ralentissement de la convergence par rapport au cas $\gamma = 2$, plus favorable. La même expérience, moyennée sur 50 réalisations indépendantes, est présentée en fig.4.
- La fig.5 montre l'erreur d'estimation du niveau de bruit: $10log(\hat{\sigma}_b^2(n)/\sigma_b^2)$. $\hat{\sigma}_b^2(n)$ est calculé par la formule (8) où P est remplacé par P_n et le terme d'intercorrélation est estimé à l'aide d'un facteur d'oubli exponentiel $\lambda = 1 - 1/6K = 0.98$. On a $\gamma = 2$.

3.6 Discussion

Dans le cas de signaux non bruités très corrélés, le vecteur de régression optimal a la propriété essentielle de satisfaire un système (quasi) dégénéré d'équations normales. Nous avons montré, qu'en présence de données bruitées, on pouvait *forcer* le retour à cette propriété de singularité et donc amener au débruitage, simplement à l'aide d'une contrainte quadratique adjointe au critère de régression traditionnel. La valeur γ prise par la contrainte est théoriquement choisie de manière quelconque, supérieure à γ_0. Les simulations montrent cependant qu'il convient de ne pas trop surestimer γ_0 afin de ne pas dégrader les performances de l'algorithme adaptatif. D'où l'intérêt particulier d'une première estimation, même grossière, de γ_0.

On note aussi que l'APG converge plus lentement que le LMS non contraint. Une étude plus fine montrerait probablement que l'APG souffre du mauvais conditionnement associé à la covariance des données non bruitées, sur laquelle l'algorithme travaille de manière implicite. Une solution possible à ce problème est le recours à des algorithmes du second-ordre, moins sensibles vis-à-vis des statistiques des entrées, mais plus coûteux.

References

[1] D. Slock, "Blind fractionally-spaced equalization, perfect reconstruction filter-banks and multichannel linear prediction," in *Proc. ICASSP*, vol. 4, pp. 585-588, 1994.

[2] K. Abed Meraim, P. Duhamel, D. Gesbert, P. Loubaton, S. Mayrargue, E. Moulines, and D. Slock, "Prediction error methods for time-domain blind identification of multichannel FIR filters," in *Proc. ICASSP*, 1995.

[3] J. R. Treichler, "γ-LMS and its use in a noise compensating adaptive spectral analysis technique," in *Proc. ICASSP*, 1979.

[4] H. Cox, R. Zeskind, and M. Owen, "Robust adaptive beamforming," *IEEE Trans. on ASSP*, vol. 35, no 10, Oct. 1987.

[5] D. Gesbert, P. Duhamel, and S. Mayrargue, "Subspace-based adaptive algorithms for the blind equalization of multichannel FIR filters," in *Proc. EUSIPCO*, 1994.

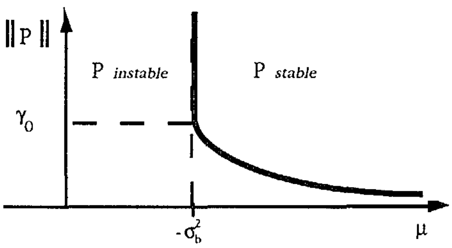

Figure 1: Normes des solutions de $(R_z + \mu I)P = r_{xy}$, pour $\mu \in [-\sigma_b^2; +\infty[$. Remarquons que $\mu = -\sigma_b^2$ rend le problème dégénéré. γ_0 correspond au vecteur de régression débruité de norme minimum.

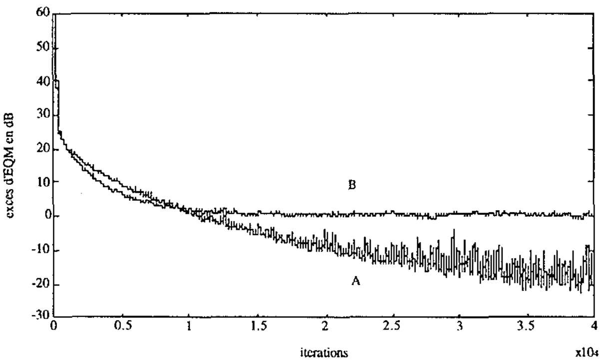

Figure 2: (A) Algorithme proposé (APG) avec $\gamma = 2$. (B) LMS classique (non contraint). $\alpha = 0.002$.

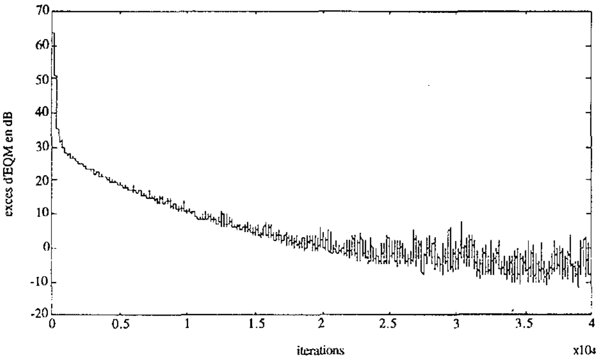

Figure 3: APG avec $\gamma = 4$. $\alpha = 0.002$.

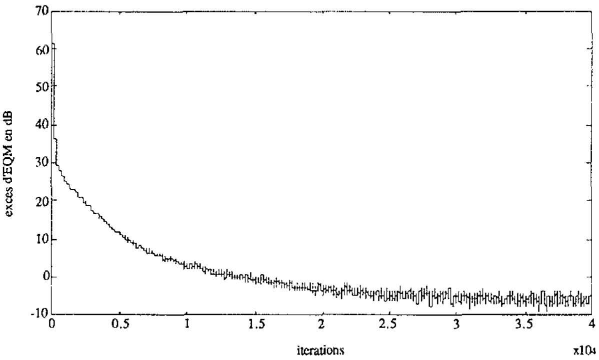

Figure 4: APG avec $\gamma = 4$, moyenné sur 50 réalisations indépendantes. $\alpha = 0.002$.

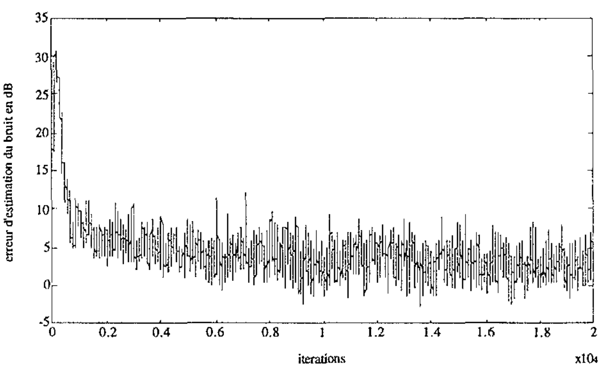

Figure 5: Apprentissage du niveau de bruit par l'APG. $\alpha = 0.002$, $\sigma_b^2 = 0.1$, $\gamma = 2$.