

# Estimation de l'ordre et identification des paramètres d'un processus ARMA

Nicolas MOAL et Jean-Jacques FUCHS

IRISA/Université de Rennes I  
Campus de Beaulieu - 35042 Rennes Cedex - France  
nmoal@irisa.fr - fuchs@irisa.fr

## RÉSUMÉ

Nous proposons ici une méthode permettant d'effectuer simultanément l'estimation de l'ordre et des coefficients d'un processus ARMA. Il s'agit d'une approche du type "méthode des moments". Nous faisons l'hypothèse que les observations admettent effectivement un modèle dans la classe des processus autorégressifs à moyenne glissante d'ordres finis. On décompose la suite des covariances estimées sur une base redondante. Il s'agit alors de trouver une solution parcimonieuse à un système linéaire sous déterminé. Nous cherchons la solution de norme  $\ell_1$  minimale à l'aide d'un algorithme de programmation linéaire ([2]).

## ABSTRACT

In this paper we propose a method that allows us to simultaneously estimate the order and the coefficients of an ARMA process based on a set of observations. It is based on "the method of moments". We thus suppose that the observations can be modeled as the realizations of a finite-order ARMA process. We transform the problem into a linear program by decomposing the covariance serie on a redundant basis. We search a parsimonious solution by choosing the one with the minimal  $\ell_1$  norm. This can be easily done with the simplex algorithm or the interior-point method, available in most scientific libraries.

## 1 Introduction

Les méthodes classiques d'identification d'un modèle ARMA à partir d'une série d'observations procèdent généralement en deux temps : on commence par déterminer l'ordre du filtre (en utilisant les critères d'Akaike ou de Rissanen par exemple) ou on lui attribue un ordre a priori, et ensuite seulement on estime ses coefficients ([1], [6]). Ces méthodes sont rapidement très coûteuses en calcul car il faut effectuer l'estimation des coefficients de tous les modèles ARMA possibles (jusqu'à un certain ordre maximal a priori).

Nous nous proposons ici de réaliser simultanément l'estimation de l'ordre du modèle et l'identification des paramètres à partir d'observations  $(y_n)_{n=1, \dots, T}$ . Nous faisons l'hypothèse que ces observations admettent effectivement un modèle dans la classe des processus ARMA d'ordres finis.

## 2 Spectre d'un processus ARMA

Considérons le processus ARMA(p,q) ( $p \geq q$ ) défini par l'équation suivante :

$$y_n + \sum_{i=1}^p a_i y_{n-i} = e_n + \sum_{j=1}^q c_j e_{n-j} \quad (1)$$

où  $e_n$  est un bruit blanc gaussien de variance  $\sigma_e^2$ .

Le spectre de ce processus (que l'on peut écrire de façon plus compacte  $A(z^{-1})y_n = C(z^{-1})e_n$ ) est alors donné par ([7]) :

$$\Phi(z) = \sigma_e^2 \frac{C(z) C(z^{-1})}{A(z) A(z^{-1})} \quad \text{pour } z = e^{i\omega} \quad (2)$$

expression que l'on peut aussi mettre sous la forme :

$$\Phi(z) = \frac{D(z)}{A(z)} + \frac{D(z^{-1})}{A(z^{-1})} = \tilde{\Phi}(z) + \tilde{\Phi}(z^{-1}) \quad (3)$$

Nous nous intéressons ici plutôt à la fonction d'autocovariance  $R(z)$

$$R(z) = \sum_{n=0}^{\infty} r_n z^{-n} = \tilde{\Phi}(z) + \frac{1}{2} r_0$$

où les  $r_n$  sont les covariances du processus.  $R(z)$  admet également une représentation sous la forme d'une fraction rationnelle

$$R(z) = \frac{D'(z)}{A(z)} = \frac{D(z)}{A(z)} + \frac{r_0}{2} \quad (4)$$

Nous supposons par la suite que les pôles de cette fraction sont simples, réels (les  $a_i$ ) ou complexes conjugués (les  $\rho_j e^{\pm i\theta_j}$ ). Nous pouvons alors écrire son développement en éléments simples comme la somme de deux types de termes, ceux associés aux pôles simples réels et ceux associés aux pôles complexes regroupés par paires de pôles conjugués :

$$\frac{D'(z)}{A(z)} = \sum_i \frac{\lambda_i z}{z - a_i} + \sum_j \frac{\mu_j (z^2 + \nu_j z)}{(z - \rho_j e^{i\theta_j})(z - \rho_j e^{-i\theta_j})}$$

Nous préférons à cette expression de  $\frac{D'(z)}{A(z)}$  la forme (5) :

$$\frac{D'(z)}{A(z)} = \sum_i \alpha_i \frac{z}{z - a_i} + \sum_j \beta_j \frac{\rho_j \sin(\theta_j) z}{(z - \rho_j e^{i\theta_j})(z - \rho_j e^{-i\theta_j})}$$

$$+ \sum_j \gamma_j \frac{z^2 - \rho_j \cos(\theta_j)z}{(z - \rho_j e^{i\theta_j})(z - \rho_j e^{-i\theta_j})} \quad (5)$$

dans laquelle les coefficients  $\alpha_i$ ,  $\beta_j$  et  $\gamma_j$  sont obtenus en identifiant les deux expressions.

### 3 Estimation d'un processus ARMA

#### 3.1 Présentation de la méthode

Nous nous proposons d'essayer de retrouver  $R(z)$  grâce à la décomposition (5) de  $R(z)$  en voyant chaque terme associé à un pôle réel comme la contribution d'un AR(1) et celui associé aux pôles complexes conjugués comme la contribution d'un ARMA(2,1). Cette décomposition revient ainsi à représenter le processus initial comme la somme d'un ensemble de processus AR(1) et ARMA(2,1) placés en parallèle et attaqués par autant de bruits blancs indépendants.

Ce point de vue mène naturellement à la façon suivante d'identifier un processus ARMA ([3]) : on estime dans un premier temps, un ensemble de  $K$  covariances  $(\hat{r}_k)_{k=0,\dots,K-1}$  à partir des observations, puis on modélise cette suite partielle de covariances comme la somme d'un petit nombre de suites partielles chacune associée à un AR(1) ou un ARMA(2,1). La seconde étape est cependant délicate car même si on connaît l'ordre du processus et, dans le contexte d'une telle méthode, le nombre de contributions du type AR(1) et du type ARMA(2,1), la décomposition/projection de la suite de covariance estimée en la somme de contributions élémentaires est non triviale et revient, par exemple, à résoudre un problème des moindres carrés non linéaire.

#### 3.2 Linéarisation du problème

Nous pouvons remplacer ce problème des moindres carrés par une version approchée dont la solution est plus simple à obtenir. Ainsi par exemple, pour chaque terme associé à un AR(1) il nous faudrait avec la première approche estimer la valeur de son pôle et la variance du bruit qui l'excite. Nous discrétisons la valeur de ce pôle et nous supposons qu'il ne peut plus prendre qu'un nombre fini de valeurs. Nous faisons une hypothèse analogue pour les termes associés aux ARMA(2,1) et nous obtenons ainsi une base discrète (de dimension notée  $N$ ) sur laquelle nous allons réaliser la projection de la suite de covariances du processus ARMA à retrouver, c'est-à-dire que nous aurons simplement à estimer le vecteur  $X$  des pondérations à attribuer à chacun de ces processus constituant notre base.

Nous rangeons les vecteurs formés des suites partielles des covariances des éléments de notre base dans une matrice  $A$  de dimension  $(K, N)$ . Nous devons donc résoudre le système linéaire  $AX = \hat{b}$  dans lequel le vecteur  $\hat{b}$  contient la suite des covariances partielles estimées du processus ARMA à retrouver.

Ce système ne possède pas de solution unique car  $N \gg K$ . Parmi les solutions possibles nous cherchons la plus parcimonieuse, c'est-à-dire celle dont le nombre de composantes non nulles est le plus petit. Une façon de s'en approcher est de

choisir la solution  $X$  de norme  $\ell_1$  minimale. Nous sommes ainsi ramenés à résoudre le problème de programmation linéaire suivant :

$$\min \|X\|_1 \quad \text{sous} \quad AX = \hat{b} \quad (6)$$

problème qui se résout simplement grâce aux algorithmes présents dans les bibliothèques de programmes scientifiques.

L'ordre du processus ARMA recherché est donné par le nombre de composantes non nulles de  $X$  (après regroupement des paires de colonnes voisines et élimination de contributions trop faibles), et les valeurs estimées de leurs pondérations sont justement les valeurs de ces composantes non nulles de  $X$ . Nous ne faisons pas a priori l'hypothèse que les vrais pôles appartiennent à la grille de discrétisation. Par contre, si nous avons choisi un pas de discrétisation suffisamment petit, un terme AR(1) ou ARMA(2,1) sera bien approximé par une interpolation linéaire de deux colonnes voisines de  $A$ .

#### 3.3 Introduction des erreurs d'estimation des covariances

##### Blanchiment des équations

Nous commençons par blanchir les équations de (6) en déterminant la matrice  $\hat{\Sigma}$  des covariances de  $\hat{b}$  dont les éléments sont donnés, pour un bruit blanc gaussien en entrée, par (cf. [6]) :

$$\begin{aligned} \hat{\Sigma}_{i,j} &= \text{cov}(\hat{r}_i, \hat{r}_j) \\ &= \frac{1}{T^2} \sum_{t=0}^{T-i-1} \sum_{s=0}^{T-j-1} (\hat{r}_{t-s} \hat{r}_{t-s+i-j} + \hat{r}_{t-s+i} \hat{r}_{t-s-j}) \end{aligned}$$

On définit ainsi  $\tilde{b} = \hat{\Sigma}^{-1/2} \hat{b}$  la version blanchie de  $\hat{b}$  et  $\tilde{A} = \hat{\Sigma}^{-1/2} A$ . Le système devient :

$$\min \|X\|_1 \quad \text{sous} \quad \tilde{A}X = \tilde{b} \quad (7)$$

Si l'on n'effectue pas d'autre transformation sur le système, la position de l'optimum n'est pas modifiée par cette multiplication. Par contre, les composantes du second membre sont maintenant décorréliées et de variance 1, si bien que toutes les équations ont, d'une certaine façon, la même importance.

##### Mise en place d'un tube autour de $\hat{b}$

En fait, comme  $\hat{b}$  n'est qu'une estimée du vecteur des vraies covariances  $r_k$  obtenue à partir des observations  $(y_n)_{n=1,\dots,T}$ , il est inutile, et même néfaste, d'imposer l'égalité stricte dans le problème de programmation précédent (on risque en effet de choisir des colonnes supplémentaires qui modéliseront essentiellement l'erreur d'estimation de  $\hat{b}$ ). Pour cette raison, nous introduisons une tolérance, un tube  $T(\hat{b}, r)$  de rayon  $r$ , autour de  $\hat{b}$  et nous devons alors résoudre

$$\min \|X\|_1 \quad \text{sous} \quad \|\tilde{A}X - \tilde{b}\|_\infty \leq r \quad (8)$$

$r$  est un paramètre à ajuster, choisi de l'ordre de l'écart type de l'erreur d'estimation, c'est-à-dire quelques unités pour notre système blanchi.

Au système (8) nous ajoutons certaines contraintes sur le signe des composantes du vecteur  $X$ . La pondération  $x_i$  ainsi associée à un pôle réel simple (donc à la série de covariances

d'un AR(1)) doit être positive, de même que celle d'une colonne de cosinus d'un pôle complexe (qui correspond à un ARMA(2,1) particulier). En effet, dans l'hypothèse où une seule colonne d'AR(1) ou de cosinus serait choisie il faut que soit vérifiée la condition  $x_i \cdot \tilde{r}_0 \geq 0$ .

Il faudrait également, mais cela ne semble pas possible avec l'algorithme du simplexe, pouvoir garantir que l'on ne choisira pas une colonne de sinus sans choisir aussi la colonne de cosinus correspondante. Une telle colonne isolée n'aurait en effet que peu de sens car sa première coordonnée (un  $r_0$ ) est nulle. Heureusement il semble qu'un tel cas ne se produise quasiment pas dans la pratique.

### 3.4 Reconstruction du processus ARMA

Une fois que nous avons résolu le problème de programmation linéaire (8), nous souhaitons retrouver le processus sous la forme d'une équation du type (1), c'est-à-dire passer de la décomposition additive (3) à la décomposition multiplicative (2). Pour cela, nous utilisons une représentation d'état du processus.

## 4 Simulations

Nous avons appliqué la méthode précédente à différents processus ARMA. Les paramètres à ajuster sont le nombre de composantes de  $\hat{b}$ , c'est-à-dire le nombre de covariances  $K$ , et le rayon du tube  $r$  que nous plaçons autour de ce vecteur.

### 4.1 Ajustement des valeurs de $r$ et $K$

L'algorithme présenté propose une méthode sous-optimale pour déterminer l'ordre d'un filtre ARMA et identifier ses coefficients. Les erreurs d'estimation sont prises en compte par l'introduction d'un tube dans lequel les solutions du programme linéaire peuvent se trouver. Le rayon de ce tube doit être ajusté. En effet, un rayon trop grand n'impose pratiquement plus aucune contrainte sur les composantes de  $X$  à l'optimum, qui deviennent nulles dans le cas limite où  $r$  est très grand. Au contraire, un rayon trop petit provoque l'apparition de composantes supplémentaires qui modélisent essentiellement le bruit d'estimation des covariances.

En faisant l'hypothèse qu'il existe un vecteur  $X^*$  tel que le vecteur des covariances exactes  $b$  vérifie  $AX^* = b$ , on souhaite que ce vrai  $b$  appartienne au tube  $T(\hat{b}, r)$ , et donc choisir la valeur de  $r$  en conséquence.  $\tilde{b} - \hat{b}$  suit une loi Gaussienne  $\mathcal{N}(0, I_K)$ , si bien que l'hypothèse précédente revient à choisir un  $r$  tel que le maximum de  $K$  variables aléatoires indépendantes Gaussiennes  $\mathcal{N}(0, 1)$  est supérieur à  $r$  avec une probabilité petite. On peut montrer que si l'on choisit  $r$  de la forme  $r = \sqrt{2q \ln s}$  cette probabilité est inférieure à  $K/s^q$ , si bien que le maximum étudié croît avec  $\sqrt{\ln(K)}$ . Nous pourrions donc prendre  $r$  de l'ordre de  $\sqrt{\ln(K)}$ .

En tenant compte du fait que le vrai vecteur  $b$  est un point admissible, de même que la solution  $X$  effectivement obtenue dans la résolution de (8), on montre (cf. [4]) que dans ce cas

la variance de  $AX$ , le vecteur des covariances reconstruites, et donc celle de  $X$  sont de l'ordre de  $r^2/T$ .

Le choix du nombre de covariances  $K$  que l'on met dans le vecteur  $\hat{b}$  est également important. Ce paramètre détermine notamment le nombre de pôles du processus qui pourront être identifiés. En effet, si un pôle ne tombe pas sur un noeud de la grille de discrétisation, il sera interpolé par les deux colonnes voisines dont les pondérations ne seront pas nulles. Comme dans le programme linéaire (6), génériquement  $K$  composantes de  $X$  seront non nulles, nous pourrions identifier au plus  $[K/2]$  pôles du processus ARMA. Si ce nombre de covariances  $K$  est trop petit, de nombreux processus interpoleront très bien notre jeu de covariances estimées. Il faut également augmenter la valeur de  $K$  avec le nombre de données car dans ce cas les variances des estimées  $\hat{r}_i$  d'indices élevés diminuent. Les exemples traités semblent montrer qu'un dizaine de covariances est un choix satisfaisant.

### 4.2 Identification d'un AR(1)

Nous avons commencé par un processus AR(1) de pôle réel 0.9. Nous avons pris  $T = 2000$  échantillons, et  $K = 10$  covariances, 40 valeurs possibles pour les pôles réels, 10 valeurs pour les modules des pôles complexes conjugués et 40 valeurs pour leurs phases, ce qui nous donne un  $N = 840$ . Le rayon du tube est  $r = 0.8$ . Nous avons effectué 100 simulations successives.

La résolution du problème du simplexe donne nettement une colonne non nulle dans la partie pôles réels, celle correspondant à ce pôle 0.9. Il y a par ailleurs dans la partie pôles complexes une paire de colonnes non nulles, correspondant à une paire de pôles de module 0.9 et d'argument 0.15 rad (figure 1).

Comme ils ont même module que le pôle réel, un argument petit et sont placés de manière symétrique par rapport à l'axe réel nous les regroupons avec le pôle réel pour ne plus avoir qu'un seul pôle réel d'amplitude 0.87, barycentre des trois pôles précédents, avec la pondération la somme des pondérations précédentes. Nous remplaçons la covariance  $\hat{r}_0$  estimée à partir des observations par la somme des pondérations de  $X$  (qui est un point du tube correspondant à la première inéquation de (8)). Nous obtenons finalement le processus AR(1) suivant :

$$\begin{aligned} y_n - 0.87y_{n-1} &= e_n \\ \sigma_e &= 1 \end{aligned}$$

### 4.3 Identification d'un ARMA(3,2)

Considérons maintenant le processus ARMA(3,2) suivant :

$$y_n + 1.9y_{n-1} + 1.79y_{n-2} + 0.801y_{n-3} = e_n - 1.2e_{n-1} + 0.85e_{n-2} \quad (9)$$

où  $e_n$  est un bruit blanc gaussien centré de variance unité. Ce processus possède un pôle réel,  $-0.9$ , et deux pôles complexes conjugués,  $0.9434e^{\pm 2.1294i}$ . Nous avons à nouveau pris  $T = 2000$ ,  $K = 10$ ,  $N = 840$ , et  $r = 0.8$ .

La résolution du programme linéaire nous permet, après regroupement des colonnes voisines, d'isoler le pôle réel, ainsi

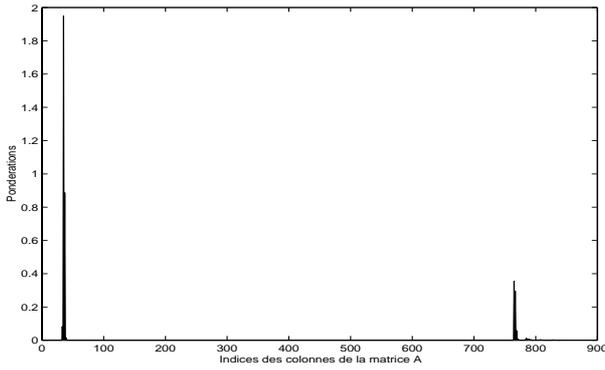


FIG. 1 — Résultats du programme linéaire pour l'identification d'un AR(1).

que la paire de pôles complexes conjugués du processus à identifier  $0.94e^{\pm 2.18i}$ , et une seconde paire de pôles conjugués,  $0.93e^{\pm 2.79i}$ , avec des pondération nettement plus faibles que les précédentes (cf. la figure 2 pour les pôles de pondération non nulle à la sortie de (8)).

Le processus reconstruit est alors l'ARMA(5,5) suivant :

$$\begin{cases} A(z) = z^5 + 3.73z^{n-1} + 6.23z^{n-2} + 5.87z^{n-3} + 3.11z^{n-4} + .73z^{n-5} \\ C(z) = z^5 + 3.09z^{n-1} + 4.22z^{n-2} + 3.20z^{n-3} + 1.35z^{n-4} + .24z^{n-5} \\ \sigma_e^2 = 47.60 \end{cases}$$

Si les pôles ont bien été identifiés, en revanche la comparaison des covariances du processus original et de celui estimé semble montrer que l'estimation de la partie MA, au travers des pondérations associées aux colonnes de  $A$ , est assez peu satisfaisante dans ce cas. En fait l'introduction d'un tube dans le programme linéaire (8) introduit un biais sur les composantes de  $X$ . Pour pallier cela, une fois que les colonnes de  $A$  intervenant dans l'identification ont été choisies et placées dans la matrice  $\bar{A}$  extraite de  $A$  nous résolvons  $\bar{A}X' = \hat{b}$  au sens des moindres carrés.

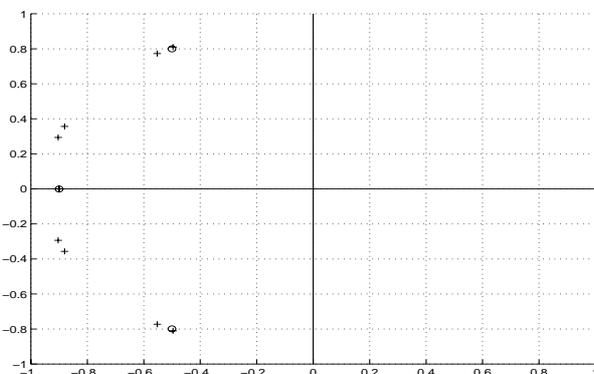


FIG. 2 — ARMA(3,2) : Pôles de pondérations non nulles à la sortie de (8) (+) et vrais pôles (o) représentés dans le plan complexe.

## 5 Conclusions

Nous avons proposé ici une méthode d'identification simultanée de l'ordre et des coefficients d'un processus ARMA à partir d'estimées de ses covariances. Pour cela nous essayons de retrouver notre spectre estimé comme une combinaison linéaire des spectres de processus AR(1) et ARMA(2,1) d'une base obtenue en discrétisant les paramètres d'AR(1) et d'ARMA(2,1). La parcimonie de la solution est assurée en minimisant la norme  $\ell_1$  du vecteur des pondérations.

Les simulations effectuées montrent que les pôles de la partie AR sont bien identifiés, qu'ils soient réels ou complexes conjugués (ainsi donc que l'ordre de la partie AR). La mise en place d'un tube autour du second membre  $\hat{b}$  introduit un biais qu'il faut corriger. La partie MA est ensuite reconstruite en utilisant une représentation d'état du processus. L'erreur sur cette partie MA semble plus importante, mais ce résultat n'est en fait pas si surprenant que cela dans la mesure où le problème de la décomposition en une série d'exponentielles armorties est connu pour être particulièrement mal conditionné (cf. [5]).

## Références

- [1] G.E.P. Box, G.M. Jenkins, *Time Series Analysis : Forecasting and Control*, Holden-Day series in time series analysis, 1970.
- [2] P.G. Ciarlet, *Introduction à l'analyse numérique matricielle et à l'optimisation*, Masson, 1982.
- [3] J.J. Fuchs, *Représentation d'état et estimation de retards. 10ème colloque Gretsi*, Nice, pp. 403-408, mai 1985. ou *State space modeling and estimation of time differences of arrival*. IEEE Trans. on ASSP, vol. 34, pp. 232-244, avril 1986.
- [4] J.J. Fuchs *Extension of the Pisarenko Method to sparse linear array*, à paraître IEEE Trans. on SP, 1997.
- [5] C. Lanczos, *Applied Analysis*, Prentice-Hall Mathematics Series, 1957.
- [6] B. Porat, *Digital Processing of Random signals*, Prentice Hall, 1994.
- [7] T. Söderström *Discrete-time Stochastic Systems, Estimation and Control*, Prentice Hall, 1994.