NEUVIEME COLLOQUE SUR LE TRAITEMENT DU SIGNAL ET SES APPLICATIONS

NICE du 16 au 20 MAI 1983

PREDICTION POUR UNE SERIE TEMPORELLE EN S'APPUYANT SUR DES RESTRICTIONS PROPRES DES MATRICES D'AUTOCORRELATION

Jaume PAGES FITA

Facultat d'Informatica , Universitat Politecnica de Barcelona
c/. Jordi Girona Salgado, 31, BARCELONA (34), ESPAGNE

RESUME

Le but de cette communication est de présenter une méthode de prédiction pour des séries temporelles discrètes liée aux travaux de J.P. Burg sur les méthodes d'analyse par maximum d'entropie.

La plupart des méthodes utilisées pour estimer les caractéristiques du second ordre d'une série temporelle stationnaire de moyenne nulle ne tiennent pas compte des restrictions propres du problème et conduisent à des solutions dénuées de sens (densitées spectrales negatives, matrices d'autocorrélation qui ne sont pas semi-définies positives).

A partir de la série temporelle et par le maximum de vraisemblance on peut trouver un estimateur de la matrice d'autocorrelation dans l'ensemble des matrices de Toeplitz semi-définies positives. Le prolongement de la série temporelle se traduit par une croissance de la dimension de la matrice d'autocorrelation. Comme la caracteristique d'être de Toeplitz semi-définie positive ne peut pas se perdre, on est restreint au moment de supposer une nouvelle valeur de la fonction d'autocorrelation ; on démontre que cette nouvelle valeur doit appartenir à un certain intervalle et on a des arguments pour choisir le centre.

Une fois qu'on a l'estimateur de la matrice d'autocorrelation de dimension augmentée, le problème de la prédiction sur la série temporelle elle même se résout facilement. On tombe sur la prédiction linéaire optimale associée à la matrice des covariances de dimension augmentée mais les considérations faites sur le caractère d'être semi-definie positive la matrice des covariances conduisent à etablir l'intervalle des valeurs pour lesquelles la prédiction est compatible avec ce caractère.

SUMMARY

A method for prediction in discrete time series is presented; the method is related with the work of J.P. Burg on maximum entropy spectral analysis.

In the estimation of second order caracteristics of stationary time series some natural restrictions are usually forgoten and consequently some methods give absolutely non-sense solutions (negative spectral densities, autocorrelation matrix wich are not positive definite).

By the maximum likelihood method it is possible to estimate the autocorrelation matrix in the set of semidefinite positive Toeplitz matrices. With the prediction, the series growths and this growth is reflected in a matrix dimension increment. But the intrinsic caracteristics of being Toeplitz and semi-definite positive must be preserved. Thus the new value for the autocorrelation function is not at all arbitrary. It is shown that this new value must lie in an interval of the real line and it is natural to choose the center of the interval.

With the autocorrelation matrix of increased dimension, it is very simple to make a prediction. The least squares linear predictor associated with the autocorrelation matrix is obtained. The set of predictions compatibles with the positive-definitness of the estimated autocorrelation matrix is introduced and analyzed.

PREDICTION POUR UNE SERIE TEMPORELLE EN S'APPUYANT SUR DES RESTRICTIONS PROPRES DES MATRICES D'AUTOCORRELATION

INTRODUCTION

Beaucoup de méthodes d'estimation des caractéristiques du second ordre d'une série temporelle ne tiennent pas compte du fait que la matrice de covariances doit appartenir a l'ensemble des matrices de Toeplitz semi-définies positives.

J.P. Burg a signalé ce fait et a realisé une étude approfondie du probleme qui l'a amené a definir des coéfficients de reflexion; l'ensemble de ces coéfficients groupe les propietées du second ordre d'un processus discret, et représente une alternative a l'ensemble des valeurs de la fonction d'autocorrelation.

Il a l'avantage d'assurer le caractère défini positif la matrice que l'on obtient de Burg, en plus de la definition, a donné une méthode d'estimation par moindres carrés de ces coefficients.

La généralisation de ces coefficients au cas multivariable n'est pas simple. La recherche de cette généralisation a amené Burg a proposer une méthode d'estimation globale de la matrice d'autocorrelation en utilisant la méthode du maximum de vraisemblance.

Les travaux de Burg, visent vers l'analyse spéctrale. Une fois obtenue l'estimation de la matrice d'autocorrelation il s'occupe de chercher une fonction de densité spéctrale compatible; la méthode qui resout ce probleme et dont l'idée fondamentale est aussi de préserver le caractère défini positif est appelé méthode d'estimation spéctrale par maximum d'entropie.

Il est possible d'utiliser cette idée, centrale des travaux de Burg, pour faire l'analyse de signaux temporels sans entrer dans le domaine fréquentiel, ce qui conduit d'une façon naturelle a la méthode de prédiction présentée dans ce travail.

EQUATIONS POUR L'ESTIMATION LA PLUS VRAISEMBLABLE DE LA MATRICE DE COVARIANCE D'UNE SERIE TEMPORELLE.

Supposons qu'on a n échantillons successifs $\{x_k\}$, $k= 1,\dots,n$, d'une série temporelle gaussienne stationnaire a moyenne nulle et matrice de covariantes T. La probabilité associée a cette realisation est

$$p(x) = \frac{1}{\sqrt{(2\pi)^n\,|T|}}\; e^{-\frac{1}{2}x'\,T^{-1}x}$$

avec $x=(x_1,\dots,x_n)'$ et $|T| = \det T$.

La matrice T, dans l'ensemble $\mathfrak{T}^+$ des matrices de Toeplitz semi-définies positives, qui maximise cette expression est l'estimation la plus vraisemblable de la matrice de covariances de la série.

La maximisation de cette expression est équivalente à la maximisation de la fonction $-\log|T| - x'T^{-1}x$ qu'on appelera par la suite, fonction objectif.

Bien que $\mathfrak{T}^+$ soit contenue dans R^n il n'est pas simple dans cet espace de dériver la fonction objectif; pour celà il est plus simple considerer $\mathfrak{T}^+$ dans R^{n^2} et prendre comme produit scalaire l'habituel. Si on écrit encore comme matrices les elements de R^{n^2} le produit scalaire s'écrit $\langle A,B\rangle = tr(A'B)$. On identifie alors le gradient d'une fonction avec une matrice et il est facile d'établir grad $|T| = T^{-1}|T|$ donc grad $\log|T| = T^{-1}$ et grad $x'Tx = xx'$ donc grad $x'T^{-1}x = -T^{-1}xx'T^{-1}$.

Le gradient dans R^{n^2} de la fonction objectif est donc $G(T) = T^{-1}xx'T^{-1} - T^{-1}$. La matrice de correlation la plus vraisemblable doit être un point dans lequel $\mathfrak{T}^+$ est orthogonal a G(T).

RESOLUTION DES EQUATIONS PAR LA METHODE DE L'ITERATION INVERSE.

L'expression du gradient est non linéaire en T, mais elle est linéaire en la matrice des données D=xx'. Si on se situe dans un point $T_k \in \mathfrak{T}^+$ il sera doncs simple de trouver de changement en D appartenant a $\mathfrak{T}^+$ avec lequel le point T_k serait l'optimum. Il faut resoudre pour ΔD_k l'equation:

$$T_k^{-1}(D+\Delta D_k)\,T_k^{-1} - T_k^{-1} \quad \text{orthogonal a } \mathfrak{T}^+$$

La méthode de l'itération inverse consiste à calculer ΔD_k et prendre alors $T_{k+1} = T_k - \Delta D_k$.

On peut calculer directement T_{k+1} qui doit vérifier

$$T_k^{-1}\,D\,T_k^{-1} - T_k^{-1}\,T_{k+1}\,T_k^{-1} \quad \text{orthogonal a } \mathfrak{T}^+$$

Si on prend comme base de $\mathfrak{T}$ les matrices H_m dont les éléments sont $h^m_{ij} = 0$, si $|i-j| \neq m$ et $h^m_{ij} = 1$ si $|i-j|=m$ la matrice T_{k+1} s'exprime

$$T_{k+1} = \sum_{m=0}^{n-1} a_m H_m$$

et la condition d'orthogonalité devient l'ensemble de conditions:

$$tr\,[(T_k^{-1}\,D\,T_k^{-1} - T_k^{-1}\sum_{m=0}^{n-1} a_m H_m\, T_k^{-1})\,H_l] = 0,\quad l=0,\dots,n-1$$

c'est à dire

$$\sum_{m=0}^{n-1} a_m\, tr\,[T_k^{-1} H_m T_k^{-1} H_l] = tr\,[T_k^{-1}\,D\,T_k^{-1} H_l],\quad l=0,\dots,n-1$$

Cette dernière expression est celle d'un systeme linéaire, $Aa = b$ dont la matrice a pour éléments

$$a_{lm} = tr\;[T_k^{-1} H_m T_k^{-1} H_l]$$

et les termes indépendants sont

$$b_l = tr\,[T_k^{-1}\,D\,T_k^{-1} H_l] = (T_k^{-1}x)'\,H_l\,T_k^{-1}x$$

On démontre que la direction $-\Delta D_k$ est une direction de croissance pour la fonction objectif.

En effet, la projection du gradient de la fonction objectif au point T_k sur la direction $-\Delta D_k$ est positive

$$tr\;[(T_k^{-1}\,D\,T_k^{-1} - T_k^{-1})(-\Delta D_k)] > 0$$

puisque

$$tr\;[(T_k^{-1}\,D\,T_k^{-1} - T_k^{-1})(-\Delta D_k)] =$$

$$= tr\,[(T_k^{-1}(D+\Delta D_k)\,T_k^{-1} - T_k^{-1})(-\Delta D_k)] +$$

$$+ tr\;[T_k^{-1}\,\Delta D_k\,T_k^{-1}\,\Delta D_k]$$

$$= tr\;[T_k^{-1}\,\Delta D_k\,T_k^{-1}\,\Delta D_k]$$

T_k^{-1} est définie positive, donc peut être decomposée à la façon de Cholesky $T_k^{-1} = U^T U$ et tenant compte que $tr(AB) = tr(BA)$ on obtient

$$tr\,[T_k^{-1}\,\Delta D_k\,T_k^{-1}\,\Delta D_k] = tr\,[U^T U\,\Delta D_k\,U^T U\,\Delta D_k] =$$

$$= tr\,[(U\Delta D_k U^T)(U\Delta D_k U^T)] = \langle U\,\Delta D_k\,U^T, U\,\Delta D_k\,U^T\rangle > 0$$

On tenant compte que $U \Delta D_k U^T$ est symétrique et non nulle.

Le cas écheant où T_{k+1} ne soit pas definie il faut diminuer le pas et prendre comme T_{k+1} une matrice plus proche a T_k dans la même direction.

Pour initialiser l'algorithme ou a utilisé la matrice de covariances obtenue avec l'algorithme de Burg classique, c'est à dire celle qui resulte de l'estimation par moindres carrés sur la série temporelle des coefficients de reflexion.

INCREMENT DE LA DIMENSION D'UNE MATRICE DE TOEPLITZ DEFINIE POSITIVE.

Il est connu qu'une matrice de Toeplitz semidéfinie positive peut être toujours factorisée d'une façon semblable a celle de Cholesky.

$$\begin{bmatrix} T_0 & T_1 & \cdots & T_{n-1} \\ T_1 & T_0 & \cdots & T_{n-2} \\ \vdots & \vdots & & \vdots \\ T_{n-1} & T_{n-2} & \cdots & T_0 \end{bmatrix} =$$

$$\begin{bmatrix} 1 & a_1 & \cdots & a_{n-1} \\ 0 & 1 & \cdots & b_{n-2} \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & 1 \end{bmatrix}^{-1} \begin{bmatrix} P_{n-1} & 0 & \cdots & 0 \\ 0 & P_{n-2} & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & P_0 \end{bmatrix} \begin{bmatrix} 1 & 0 & \cdots & 0 \\ a_1 & 1 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ a_{n-1} & b_{n-2} & \cdots & 1 \end{bmatrix}^{-1}$$

avec $P_k \geqslant 0$. Le calcul effectif de la décomposition peut se faire en particulier avec l'algorithme de Levinson. Il en resulte

$$P_k = P_{k-1} \, (1 - r_k^2)$$

ou r_k sont les coéfficients de réflexion definis par Burg. Le coéfficient r_n est donné par l'expression

$$r_n = -\frac{1}{P_{n-1}} \sum_{k=0}^{n-1} T(n-k)\, a_k \qquad (a_0 = 1)$$

Il est clair que le caractère défini positif de la matrice est relié au fait que $|r_k| < 1$.

Donc, si on augmente la dimension de la matrice d'autocorrelation en ajoutant une nouvelle valeur T_n et on veut garder son caractere semidefini positif on doit soigner que $|r_n| \leqslant 1$ et pourtant T_n doit appartenir a l'intervalle

$$\left[-\sum_{k=1}^{n=1} T(n-k) a_k - P_{n-1},\ \sum_{k=1}^{n-1} T(n-k) a_k + P_{n-1} \right].$$

Si on choisit un des extrêmes de l'intervalle alors $|r_n| = 1$ et $P_n = 0$. La matrice de Toeplitz devient singulière et a partir de ce moment si on voulait encore augmenter sa dimension il n'y aurait plus de choix car les valeurs de la fonction d'autocorrelation qui conservent le caractère semi-defini positif de la matrice seraient tous fixées.

Le centre de l'intervalle est associé au choix $r_n = 0$; alors $P_n = P_{n-1}$ ce qui ne reduit absolutement par les possibilités de choix posterieurs dans le cas où la matrice devrait encore croître.

PREDICTION

Une fois qu'on a la matrice de covariances augmentée de dimension conservant son caractère défini positif, le problème de la prédiction se résout en utilisant le théoreme de calcul de probabilités d'après lequel si x,y sont les composants d'un vecteur gaussien de moyenne nulle et covariance

$$E \begin{bmatrix} x \\ y \end{bmatrix} [x', y'] = \begin{bmatrix} P_{xx} & P_{xy} \\ P_{yx} & P_{yy} \end{bmatrix}$$

l'esperance conditionelle de y connaissant x est donné par

$$E[y/x] = P_{yx} P_{xx}^{-1} x$$

Dans notre cas cette formule, s'écrit

$$\hat{x}_{n+1} = [T_n \ldots T_1] \begin{bmatrix} T_0 & \cdots & T_{n-1} \\ \vdots & & \vdots \\ T_{n-1} & \cdot & T_0 \end{bmatrix}^{-1} \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} =$$

$$= [T_n \ldots T_1] \begin{bmatrix} 1 & 0 & \cdots & 0 \\ b_1 & 1 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ b_{n-1} & c_{n-2} & \cdot & 1 \end{bmatrix} \begin{bmatrix} P_{n-1} & & \\ & \ddots & \\ & & P_0 \end{bmatrix}^{-1} \begin{bmatrix} 1 & b_1 & \cdots & b_{n-1} \\ 0 & 1 & \cdots & c_{n-2} \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} =$$

$$= -[r_n \ldots r_1] \begin{bmatrix} 1 & b_1 & \cdots & b_{n-1} \\ 0 & 1 & \cdots & c_{n-2} \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} =$$

$$= -[a_n \ldots a_1] \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}$$

qui constitue la prédiction linéaire optimale associée à la matrice de covariances de dimension augmentée.

De façon semblable on peut calculer la variance de l'estimation qui est

$$P_{yy} - P_{yx} P_{xx}^{-1} P_{xy}$$

d'où:

$$\sigma^2_{\hat{x}_n} = T_0 - [T_N \ldots T_1] \begin{bmatrix} T_0 & \cdots & T_{N-1} \\ \vdots & & \vdots \\ T_{N-1} & \cdots & T_0 \end{bmatrix}^{-1} \begin{bmatrix} T_N \\ \vdots \\ T_1 \end{bmatrix} =$$

$$= T_0 - [r_n \ldots r_1] \begin{bmatrix} P_{n-1} & r_n \\ \vdots & \\ P_0 & r_1 \end{bmatrix} =$$

$$= T_0 - (P_0 r_1^2 + P_1 r_2^2 + \ldots + P_{n-1} r_n^2) = P_n$$

Expression qui montre que parmi les valeurs possibles pour r_n au moment d'augmenter la dimension de la matrice de covariances , le choix $r_n = 0$ est le plus prudent, car il laisse la variance de la prédiction la plus grande possible.

Néanmoins il faut noter, que on obtient des prédictions diferentes selon la valeur choisie pour la fonction d'autocorrelation. L'intervalle de valeurs possibles pour T_n, d'amplitude $2P_{n-1}$, devient un intervalle de valeurs possibles pour $\hat{x}_{n+1}$ compatibles avec l'estimation faite de la matrice de covariances de dimension n. L'amplitude de cet intervalle est

$$2\sum_{i=1}^{n} x_i \; b_{i-1}$$

et chaque prédiction possible à une variance associée differente.

ALGORITHME DE PREDICTION

L'ensemble de l'algorithme de prédiction est le suivant :

1) Algorithme de Burg classique: estimation d'une matrice de covariances définie positive de dimension n à partir de n valeurs successives d'une série temporelle;
2) Resolution du systeme linéaire de dimension n pour la calcul d'une nouvelle matrice plus vraisemblable;
3) Test sur la matrice obtenue. Si elle est définie positive alors 5) sinon 4);
4) Choix d'une nouvelle matrice plus proche à celle de départ mais dans la même direction, et 3);
5) Test de convergence sur le changement de la matrice et du valeur de la fonction objectif . Si ces changements ont eté importants on va a 2) sinon a 6);
6) On a l'estimation la plus vraisemblable pour la matrice de covariances. Calcul de l'intervalle de possibilités pour la valeur de la fonction d'autocorrelation au décalage n compatible avec le caractère défini positif de la matrice de dimension augmentée.
7) Prédiction et calcul de la variance de la prédiction et de l'intervalle de valeurs pour $\hat{x}_{n+1}$.

BIBLIOGRAPHIE

- Robison, Enders A., Multichannel Time Series Analysis with digital compiter programs. Holden Day Inc., 1967.

- J.P. Burg, Maximum entropy spectral analysis. Ph.D. Thesis Geophisics Departement Stanford University, 1975.

- J.P. Burg, Estimation of structured covariance matrices: A generalization of the Burg technique. NATO ASI on Nonlinear Stochastic Problems, Algarve, Portugal, mai 1982.