

NEUVIEME COLLOQUE SUR LE TRAITEMENT DU SIGNAL ET SES APPLICATIONS

985



NICE du 16 au 20 MAI 1983

ESTIMATION SPECTRALE PAR LE CRITERE DU MAXIMUM DE VRAISEMBLANCE
POUR UN MODELE AR AVEC POLES VOISINS DU CERCLE UNITE

H. CLERGEOT

LABORATOIRE DES SIGNAUX ET SYSTEMES - CNRS-ESE- PLATEAU DU MOULON -
91190 GIF SUR YVETTE

RESUME

Nous rappelons le lien entre le critère d'erreur de prédiction minimale et le critère du maximum de vraisemblance. Nous discutons les différentes approximations classiques de l'erreur de prédiction (autocorrélation, covariance, méthode de Burg) et une méthode de minimisation de l'erreur de prédiction "exacte".

Plus particulièrement en vue de l'analyse de signaux de courte durée ou avec des pôles voisins du cercle unité, nous comparons ces méthodes d'erreur de prédiction minimale avec une méthode, développée indépendamment par l'auteur et S.KAY, fondée sur la minimisation récursive de la vraisemblance exacte.

Cette étude confirme le bon comportement du critère du maximum de vraisemblance et l'intérêt d'une méthode de maximum de vraisemblance exacte quand les pôles deviennent voisins du cercle unité.

SUMMARY

We recall the connexion between maximum likelihood and minimum prediction error estimation. We compare classical approximations of the prediction error (autocorrelation, covariance, forward-backward error) and the exact maximum likelihood.

For application to spectrum analysis of short signals, or signals with poles near the unit circle we give a very simple implementation of a method already proposed by the author /3/ and S.KAY /4/ found on recursive minimization of the exact likelihood.

This study gives a confirmation of the interest of exact maximum likelihood for estimation of poles near the unit circle.



INTRODUCTION

L'intérêt du critère du maximum de vraisemblance pour l'analyse spectrale paramétrique de signaux stationnaires a été largement démontré /1/,/2/,/3/. Dans le cas d'un modèle autorégressif, ce critère conduit de façon approchée à la méthode d'erreur de prédiction minimale, sous réserve que la durée T d'observation est assez longue et que les pôles ne sont pas trop voisins du cercle unité. Précisément dans le cas de pôles voisins du cercle unité les algorithmes dérivés des méthodes d'erreur de prédiction minimale conduisent à une variance importante sur le spectre et ne garantissent pas toujours que les pôles estimés ont un module inférieur à un : l'estimation est-elle améliorée si l'on revient à l'application du critère du maximum de vraisemblance exact ?

Nous montrons que la différence porte essentiellement sur l'estimation des modules des pôles, et que le maximum de vraisemblance assure que les pôles estimés restent à l'intérieur du cercle unité. Nous discutons en détail le modèle avec un seul pôle et nous montrons que la variance sur le module est diminuée. Enfin nous donnons une version d'un algorithme récursif simple pour le calcul des coefficients de réflexion par le critère du maximum de vraisemblance exact /3/,/4/.

I- DISCUSSION SUR L'APPROXIMATION DU CRITERE DU MAXIMUM DE VRAISEMBLANCE

a) Définitions

Soit le signal aléatoire stationnaire à temps discret $x(t)$, supposé représentable par un modèle autorégressif d'ordre p :

$$x(t) = \sum_{n=1}^p a_n x(t-n) + \omega(t), \quad (1)$$

où $\omega(t)$ est un bruit blanc de variance σ^2 :

$$E[\omega(t) \omega(t')] = \sigma^2 \delta_{tt'}, \quad (2)$$

soit \underline{X} le vecteur correspondant à une observation de durée N :

$$\underline{X}^t = [x(1), \dots, x(t)] \quad (3)$$

soit Γ la matrice de covariance :

$$\Gamma = E[\underline{X} \underline{X}^t] \quad (4)$$

nous poserons :

$$\gamma = \frac{1}{\sigma^2} \Gamma \quad (5)$$

Sous l'hypothèse gaussienne, la Log-vraisemblance de l'observation \underline{X} pour une valeur hypothétique $\hat{\Gamma}$ de la matrice de covariance est donnée par

$$V(X, \hat{\Gamma}) = C \underline{t} e - \frac{1}{2} \text{Log det } \hat{\Gamma} + \underline{X}^t \hat{\Gamma}^{-1} \underline{X} \quad (6)$$

soit :

$$V(X, \hat{\Gamma}) = C \underline{t} e - \frac{1}{2} \text{Log } \text{det } \hat{\gamma} + N \text{Log } \sigma^2 + \frac{1}{\sigma^2} \underline{X}^t \hat{\gamma}^{-1} \underline{X} \quad (7)$$

Nous verrons que la matrice $\hat{\gamma}$ ne dépend que des paramètres \hat{a}_i du prédicteur, indépendamment de $\hat{\sigma}$. La maximisation de (7) par rapport à $\hat{\sigma}$ conduit à :

$$\hat{\sigma}^2 = \frac{1}{N} \underline{X}^t \hat{\gamma}^{-1} \underline{X} \quad (8)$$

les paramètres \hat{a}_i étant déterminés en minimisant :

$$S(\hat{a}) = \text{Log det } \hat{\gamma} + \frac{1}{\sigma^2} \underline{X}^t \hat{\gamma}^{-1} \underline{X} \quad (9)$$

D'après (8) on voit que $\frac{1}{N} \underline{X}^t \hat{\gamma}^{-1} \underline{X}$ représente l'estimée de l'erreur de prédiction. Les méthodes

usuelles d'estimation reviennent à :

- négliger le terme en $\text{det } \hat{\gamma}$ dans (9) ; on est alors bien conduit à minimiser "l'erreur de prédiction" $\underline{X}^t \hat{\gamma}^{-1} \underline{X}$,

- utiliser une forme approchée de l'erreur de prédiction $\underline{X}^t \hat{\gamma}^{-1} \underline{X}$.

Nous allons discuter ces deux approximations.

b) Approximations de l'erreur de prédiction /3/

Nous poserons :

$$\hat{A}_p(z) = 1 - \hat{a}_1 z^{-1} - \dots - \hat{a}_p z^{-p} \quad (10)$$

Nous désignerons par $y_p(t)$ le signal filtré de l'observation $x(t)$ par le filtre $\hat{A}_p(z)$ (le signal $x(t)$ est complété par des zéros en dehors de l'intervalle d'observation) :

$$y_p(t) = x(t) - \sum_{n=1}^p \hat{a}_n x(t-n) \quad (11)$$

Nous désignerons par $\hat{y}_p(t)$ le signal filtré par le filtre retourné :

$$\hat{y}_p(t) = x(t) - \sum_{n=1}^p \hat{a}_n x(t+n) \quad (12)$$

Remarquons que pour $t \in [p, N]$, $y_p(t)$ est l'erreur de prédiction directe estimée, pour $t \in [1, N-p]$, $\hat{y}_p(t)$ est l'erreur de prédiction inverse. En dehors de ces intervalles l'erreur de prédiction est biaisée du fait que le filtrage fait intervenir des valeurs de $x(t)$ hors de l'intervalle d'observation, qui ont été remplacées par des zéros.

On montre /3/ que l'expression exacte de $\underline{X}^t \hat{\gamma}^{-1} \underline{X}$ est :

$$\underline{X}^t \hat{\gamma}^{-1} \underline{X} = \sum_{n=1}^N y_p^2(t) + \sum_{n=1}^N \hat{y}_p^2(t) - \sum_{n=1}^{N+p} y_p^2(t) \quad (13)$$

Diverses variantes s'en déduisent, compte tenu également du fait que :

$$\sum_{n=1}^{N+p} y_p^2(t) = \sum_{n=1}^N \hat{y}_p^2(t) = \frac{1}{2} \left[\sum_{n=1}^{N+p} y_p^2(t) + \sum_{n=1}^N \hat{y}_p^2(t) \right] \quad (14)$$

Les principales approximations courantes de $\hat{\sigma}^2$ sont rappelées ci-dessous; en utilisant les relations (14) on constate facilement qu'elles ne diffèrent de (13) que par les régimes transitoires sur les p points au début et à la fin de l'intervalle d'observation :

auto-corrélation : $\hat{\sigma}^2 \approx \frac{1}{N} \sum_{n=1}^{N+p} y_p^2(t) = \frac{1}{N} \sum_{n=1}^N \hat{y}_p^2(t) \quad (15)$

covariance directe : $\hat{\sigma}^2 = \frac{1}{N-p} \sum_{n=1}^N y_p^2(t) \quad (16)$

covariance directe-inverse : $\hat{\sigma}^2 = \frac{1}{2(N-p)} \left[\sum_{n=1}^{N-p} \hat{y}_p^2(t) + \sum_{n=p+1}^N y_p^2(t) \right] \quad (17)$

c) Expression de det γ /5/,/3/

Pour $N > p$ on trouve que $\text{det } \gamma$ est indépendant de N ; or d'après les résultats ci-dessus, le terme en $\underline{X}^t \hat{\gamma}^{-1} \underline{X}$ dans (9) croît proportionnellement à N , ce qui justifie en général l'approximation consistant à négliger le premier terme pour N assez grand. Nous allons donner deux expressions de $\text{det } \gamma$ qui permettront une discussion plus précise.

Une première expression est donnée par Grenander et Szegö (/5/, p.78) en fonction des zéros z_i du polynôme $A(z)$.

$$\text{Log det } \gamma = \sum_{i=1}^p \text{Log} \left[(1 - |z_i|^2) \prod_{j \neq i} |1 - z_i z_j^*| \right] \quad (18)$$

En général pour $i \neq j$ les termes $|1 - z_i z_j^*|$ gardent un module de l'ordre de 1 ; une forte valeur du gradient de $\text{Log det } \gamma$ ne peut alors provenir que des

MAXIMUM DE VRAISEMBLANCE POUR UN MODELE AR AVEC POLES VOISINS DU CERCLE UNITE

termes en $\text{Log}(1-|z_i|^2)$ si $|z_i|$ est voisin de 1. Donc en général * le terme en $\text{Log det } \gamma$ n'intervient que si le modèle possède des pôles voisins du cercle unité et ne joue que dans la détermination du module de ce pôle.

Remarquons que de toute façon, d'après (18), la Log vraisemblance tend vers $-\infty$ sur la frontière du pavé $|z_i| \geq 1, i = 1, p$ et possède forcément un maximum à l'intérieur de ce domaine : il existe une solution pour le maximum de vraisemblance exact correspondant à un modèle AR stable.

Rappelons une autre expression remarquable de $\text{det } \gamma$ qui nous sera utile pour obtenir l'algorithme récursif du paragraphe III (/3/,/5/) :

$$\text{Log det } \gamma = \sum_{n=1}^p \text{Log} \frac{\sigma_n^2}{\sigma^2} - 1, \quad (19)$$

où σ_n^2 représente l'erreur de prédiction minimale pour un prédicteur d'ordre N. En désignant par $k_n, n=1, p$ les coefficients de réflexion, rappelons que les erreurs de prédiction d'ordre croissant sont reliées par :

$$\sigma_n^2 = (1-k_n^2) \sigma_{n-1}^2 \quad (20)$$

on en déduit aussi bien :

$$\sigma^2 = \sigma_p^2 = (1-k_p^2) \dots (1-k_1^2) - \sigma_0^2 \quad (21)$$

soit en portant dans (19) :

$$\text{Log det } \gamma = - \sum_{n=1}^p n \text{Log}(1-k_n^2) \quad (22)$$

Ici encore remarquons que la Log vraisemblance tend vers $-\infty$ aux limites du pavé $|k_n| \leq 1, n=1, p$.

II-DISCUSSION POUR UN MODELE D'ORDRE 1

Nous allons discuter le cas d'un modèle AR complexe monofréquence.

$$x(t) = a, x(t-1) + e(t) \\ a = r e^{j\phi} \quad (23)$$

où $e(t)$ est un bruit blanc complexe à phase équipartie :

$$e(t) = \varepsilon(t) + j \varepsilon^*(t), \quad (24) \\ E [e(t)^2] = E[\varepsilon^*(t)^2] = \sigma^2/2$$

La variable $x(t)$ est également à phase équipartie et les relations (4), (8), (9) se généralisent simplement en remplaçant X^t par X^+ . On obtient :

$$\text{det } \hat{\gamma} = 1 - \hat{r}^2, \quad (25)$$

$$\underline{X}^+ \hat{\gamma}^{-1} \underline{X} = \sum_1^N |x(t)|^2 + \hat{r}^2 \sum_2^{n-1} |x(t)|^2 \\ - 2\hat{r} \Re e^{-j\phi} \sum_2^N x(t) \cdot x(t-1)^* \quad (26)$$

1°) Fréquence

En substituant dans (9) on trouve que la solution du maximum de vraisemblance pour ϕ ne dépend pas du terme en $\text{det } \gamma$, ce qui illustre la remarque générale du paragraphe II.C.

Plus précisément on trouve pour $\hat{\phi}$ la même solution en minimisant (9) ou les approximations (non biaisées) (16) et (17) de l'erreur de prédiction :

$$\hat{\phi} = \text{Arg} \sum_{t=2}^N x(t) \cdot x(t-1)^* \quad (27)$$

* Sinon il faudrait que deux pôles z_i et z_j aient tous les deux un module voisin de 1 et des arguments voisins. C'est le cas pour les pôles d'argument voisin de 0 ou π : $|1-z_i^2|$ est voisin de zéro si $|z_i|$ est voisin de 1.

2°) Estimation du module

On obtient aisément les valeurs \hat{r}_1 et \hat{r}_2 des estimées minimisant l'erreur directe (16) ou directe inverse (17) :

$$\hat{r}_1 = \frac{|\sum_2^N x(t) x(t-1)^*|}{\sum_2^N |x(t-1)|^2}, \quad (28)$$

$$\hat{r}_2 = \frac{2|\sum_2^N x(t) x(t-1)^*|}{\sum_2^N |x(t)|^2 + \sum_2^N |x(t-1)|^2}, \quad (29)$$

Le maximum de vraisemblance est solution du système d'équations :

$$\hat{r} = \frac{|\sum_2^N x(t) x(t-1)^*|}{\sum_2^{N-1} |x(t)|^2 + \frac{\hat{\sigma}^2}{1-\hat{r}^2}} \quad (30)$$

$$\hat{\sigma}^2 = \frac{1}{N} \left\{ \sum_1^N |x(t)|^2 + r^2 \sum_2^N |x(t)|^2 - 2\hat{r} \sum_2^N x(t) x(t-1)^* \right\}$$

Nous allons expliciter ces relations compte tenu du modèle (23). Par un changement de variable on peut se ramener au cas où a est réel : c'est ce que nous supposons pour les calculs.

On obtient :

$$\sum_2^N x(t) x(t-1)^* = r \sum_2^N |x(t-1)|^2 + \sum_2^N x(t-1)^* e(t). \quad (31)$$

On montrerait facilement que l'agrément $\Delta\phi$ à une variance $E|\Delta\phi|^2 \sim \frac{1-r^2}{(N-1)r^2}$ très petite en particulier si r est voisin de 1, ce qui permet d'approximer le module par la partie réelle :

$$\left| \sum_2^N x(t) x(t-1)^* \right| \approx r \sum_2^N |x(t-1)|^2 + \sum_2^N \Re [x(t-1)^* e(t)] \quad (32)$$

En utilisant le développement de $|x(t)|^2$ on obtient

$$2\Re [x(t-1)^* e(t)] = |x(t)|^2 - |x(t-1)|^2 + (1-r^2) |x(t-1)|^2 - |e(t)|^2 \quad (33)$$

En portant dans (32) il vient :

$$-\frac{N}{2} \sum_2^N x(t) x(t-1)^* \approx r \sum_2^N |x(t-1)|^2 + \frac{1}{2} \left\{ |x(N)|^2 - |x(1)|^2 + (1-r^2) \sum_2^N |x(t-1)|^2 - \sum_2^N |e(t)|^2 \right\} \quad (34)$$

Nous désignerons par \hat{A}^2 l'amplitude moyenne du signal sur l'échantillon :

$$\hat{A}^2 = \frac{1}{2(N-1)} \left\{ \sum_2^N |x(t)|^2 + \sum_2^N |x(t-1)|^2 \right\} \quad (35)$$

Nous admettrons que la moyenne sur l'innovation vraie $\frac{1}{(N-1)} \sum_2^N |e(t)|^2$ diffère peu de σ^2 . Compte tenu de (34) et de la relation équivalente pour le modèle inversé on obtient les interprétations suivantes de \hat{r}_1, \hat{r}_2 , et \hat{r} :

$$\hat{r}_1 = r \left\{ 1 + \frac{|x(N)|^2 - |x(1)|^2 + (1-r^2) \hat{A}^2 - A^2}{2 \hat{A}^2} \right\} \quad (36)$$



MAXIMUM DE VRAISEMBLANCE POUR UN MODELE AR AVEC POLES
VOISINS DU CERCLE UNITE

$$\hat{r}_2 = r \cdot \left\{ 1 + (1-r^2) \frac{\hat{A}^2 - A^2}{2\hat{A}^2} \right\} \quad (37)$$

$$\hat{r} = r \cdot \frac{1 + (1-r^2) \frac{\hat{A}^2 - A^2}{2\hat{A}^2}}{1 + \frac{1}{(N-1)\hat{A}^2} \left\{ \frac{\sigma^2}{1-\hat{r}^2} - \frac{|x(1)|^2 + |x(n)|^2}{2} \right\}} \quad (38)$$

Dans (36) le premier terme correctif fait apparaître le défaut de la méthode de prédiction directe: si l'amplitude moyenne tend à augmenter, l'estimateur va surévaluer le module du pôle correspondant. L'effet serait inversé pour la méthode de prédiction inverse et se trouve annulé dans (37) en utilisant la méthode directe-inverse.

La relation (37) fait apparaître un type d'erreur qui apparaît lorsqu'un pôle est voisin du cercle unité, le temps de corrélation des fluctuations d'amplitudes devient long, ce qui augmente la probabilité de trouver une amplitude moyenne \hat{A} nettement différente de son espérance A . La relation (37) montre que l'estimateur tend à surévaluer le module du pôle si l'amplitude du signal sur l'échantillon est supérieure à son espérance.

Sans rentrer dans une discussion très détaillée, pour l'estimation par le maximum de vraisemblance, le terme en $\frac{\sigma^2}{1-\hat{r}^2}$ au dénominateur de (38) a un effet stabilisant si \hat{r} se rapproche du cercle unité. Cet effet est particulièrement intéressant pour l'analyse spectrale: le spectre étant en $\text{Log}(1-r^2)$ on réduit notablement la variance sur le maximum du spectre en repoussant les fluctuations de la limite $|r|=1$.

III - ALGORITHME RECURSIF POUR LE MAXIMUM DE VRAISEMBLANCE EXACT

1°) Calcul à l'ordre $\ell \leq p$

Supposons le prédicteur optimal d'ordre $(\ell-1)$ connu; nous nous proposons de calculer le prédicteur optimal et l'erreur de prédiction à l'ordre ℓ .

Compte tenu de (22) les coefficients de réflexion étant connus jusqu'à l'ordre $\ell-1$,

$$\text{Log det } \gamma = \text{Cte} - \ell \text{Log}(1-k_\ell^2) \quad (39)$$

Pour déterminer $\underline{X}^t \gamma^{-1} \underline{X}$ on peut utiliser la relation (13) à l'ordre ℓ , en tenant compte de la relation de Levinson entre prédicteurs d'ordre ℓ et $\ell-1$:

$$\begin{aligned} y_\ell(t) &= y_{\ell-1}(t) + k_\ell \tilde{y}_{\ell-1}(t-\ell) \\ \tilde{y}_\ell(t) &= \tilde{y}_{\ell-1}(t) + k_\ell y_{\ell-1}(t+\ell) \end{aligned} \quad (40)$$

On obtient d'après (13):

$$\begin{aligned} \underline{X}^t \gamma^{-1} \underline{X} &= \sum_1^N \tilde{y}_\ell^2(t) - \frac{N+\ell}{N+1} \sum_1^N y_\ell^2(t) \\ &= A_\ell k_\ell^2 + B_\ell k_\ell + C_\ell \end{aligned} \quad (41)$$

avec

$$\begin{aligned} A_\ell &= \sum_{\ell+1}^{N+\ell} y_{\ell-1}(t)^2 - \sum_{N-\ell+1}^N \tilde{y}_{\ell-1}(t)^2 \\ B_\ell &= \sum_1^{N-\ell} \tilde{y}_{\ell-1}(t) \cdot y_{\ell-1}(t+\ell) \\ C_\ell &= \sum_1^N y_{\ell-1}(t)^2 - \sum_{N+1}^{N+\ell} \tilde{y}_{\ell-1}(t)^2 \end{aligned} \quad (42)$$

En minimisant (9) par rapport à k_ℓ on arrive pour déterminer σ_ℓ^2 et k_ℓ aux équations:

$$\sigma_\ell^2 = \frac{1}{N} (A_\ell k_\ell^2 + B_\ell k_\ell + C_\ell) \quad (43)$$

$$k_\ell \left\{ A_\ell + \frac{\ell}{N} (A_\ell k_\ell + B_\ell k_\ell + C_\ell) \right\} = -\frac{B_\ell}{2}$$

Le coefficient k_ℓ est solution d'une équation du troisième degré; sous la forme (43) cette équation pourra en général être résolue de façon itérative, compte tenu du fait que le deuxième terme dans l'accolade est petit, en initialisant à $k_\ell = -B_\ell/2A_\ell$.

2°) Calcul récursif des coefficients de réflexion

En partant de l'ordre $\ell=1$, les estimées des coefficients de réflexion jusqu'à l'ordre $\ell-1$ sont utilisées pour calculer l'estimée à l'ordre ℓ .

Les filtrées $y_\ell(t)$ et $\tilde{y}_\ell(t)$ sont calculées également de proche en proche par les relations (40), qui traduisent simplement la structure classique de filtrage en réseau.

CONCLUSION

Nous avons discuté l'intérêt d'une méthode de maximum de vraisemblance exact pour l'estimation des paramètres d'un modèle autorégressif. Il ressort que cette méthode ne modifie pratiquement pas l'estimation des fréquences par rapport aux méthodes d'erreur de prédiction minimale; elle apporte une réduction de la variance sur le spectre estimé dans le cas de pôle voisin du cercle unité.

Nous avons obtenu pour la solution un algorithme très simple à partir d'une structure de filtrage en réseau sur le signal.

BIBLIOGRAPHIE

- /1/ - P.WHITTLE, "Estimation and information in stationary time series", Arkiv für Matematik, 2,423, 1953
- /2/ - G.E.P.BOX, G.M.JENKINS, "Time series Analysis Forecasting and control", Holden-Day, San Fransisco, CA, 1976
- /3/ - H.CLERGEOT, "Estimation du spectre d'un signal aléatoire gaussien par le critère du maximum de vraisemblance", Thèse de Doctorat d'Etat, Université Paris Sud, Orsay, Oct82
- /4/ - S.KAY, "More accurate autoregressive parameter and spectral estimation for short data records", proc. ASSP, Hamilton, Aug 1981
- /5/ - U.GRENANDER, G.SZEGO, "Toeplitz forms and their applications", Univ of California Press, Berkley, 1958