D. COMMENGES

Approche bayesienne en traitement de signal : estimation du signal d'entrée d'un système linéaire

Revue de statistique appliquée, tome 30, nº 4 (1982), p. 13-22

<http://www.numdam.org/item?id=RSA_1982__30_4_13_0>

APPROCHE BAYESIENNE EN TRAITEMENT DE SIGNAL : ESTIMATION DU SIGNAL D'ENTREE D'UN SYSTEME LINEAIRE

D. COMMENGES

Département Informatique (Pr. Salamon)
Université de Bordeaux II – 146 rue Léo Saignat
33076 Bordeaux – Cedex

RESUME

Nous présentons un problème de traitement de signal – l'estimation du signal d'entrée d'un système linéaire – où l'approche bayésienne semble la seule voie possible. L'information principale est portée, non par l'espérance, mais par la matrice de covariance de la loi a priori. Un estimateur du maximum de vraisemblance est proposé pour un paramètre d'échelle de cette matrice. Une étude en simulation valide cette approche.

1. INTRODUCTION

En traitement de signal, la distinction entre approche bayésienne et approche fréquentiste, n'est pas toujours nette ; la notion de modélisation permet souvent de dépasser cette opposition. Nous présentons ici, un problème dont la solution semble requérir sans équivoque une approche de type bayésien. Le problème est celui de l'estimation du signal d'entrée d'un système linéaire (au sens des automaticiens) connu, à partir du signal de sortie bruité. Ce problème d'une portée assez générale se pose en particulier dans des applications médicales [1]. L'écriture sous forme matricielle d'un tel problème, conduit à un modèle linéaire (au sens des statisticiens) dans lequel le nombre d'inconnues peut être égal ou supérieur au nombre d'équations. L'estimateur des moindres carrés est alors inutilisable, et on ne peut résoudre le problème que si l'on dispose d'une connaissance a priori sur le signal inconnu. Nous montrons que la connaissance d'une certaine "régularité" du signal peut être traduite par une loi a priori, et nous examinons le problème de l'estimation d'un paramètre de cette loi.

Notation : N(m , A) représente une loi normale de moyenne m, et de matrice de covariance A.

A' transposée de A.

Mots clés : **Approche bayésienne, Traitement de signal, estimation, modèle linéaire, estimation de la loi a priori, applications médicales.**

2. SYSTEME LINEAIRE ET MODELE LINEAIRE

L'estimation du signal d'entrée d'un système linéaire discret (qui peut être l'approximation d'un système continu) conduit à un modèle linaire. La relation entre l'entrée x (i) et la sortie y (k) s'écrit :

$$y(k) = \sum_{i=1}^{m} H(k, i)\, x(i) + v(k) \qquad k = 1, \ldots, n$$

où v(k) est un bruit gaussien centré.

Le problème que nous posons est l'estimation de $x(i)$; $i = 1, \ldots, m$. L'équation peut s'écrire matriciellement :

$$y = Hx + v$$

où y est un vecteur de dimension n,
x est un vecteur de dimension m,
H une matrice $n \times m$,
v un vecteur aléatoire de dimension n, de loi $p(v) = N(0, B)$.

Nous sommes dans le cadre classique du modèle linéaire ; cependant la solution classique – estimateur des moindres carrés ou estimateur de Markov – n'est utilisable que si m est très inférieur à n. Dans les problèmes de traitement de signaux, m peut être égal ou supérieur à n. L'approche bayésienne paraît ici nécessaire.

3. ESTIMATEUR DU MAXIMUM A POSTERIORI

Si x (k) représente l'échantillonnage d'un signal physique, on est assuré, en général, d'une certaine régularité de cette suite : cela veut dire que $x(k + 1)$ ne prendra pas une valeur très différente de x (k). Pour traduire cette connaissance, il convient de prendre une loi a priori, telle que $x(k + 1)$ et x (k) soient corrélés positivement.

On peut construire une telle loi récursivement, en utilisant un raisonnement typiquement bayésien. Si l'on connait x (k), on sait que $x(k + 1)$ prendra une valeur voisine ; on peut choisir pour loi a priori de $x(k + 1)$, connaissant x (k) :

$$p(x(k + 1)/x(k)) = N(x(k), \sigma_w^2)$$

Ceci correspond au modèle de la promenade aléatoire :

$$x(k + 1) = x(k) + w(k + 1) \; ; k = 1, \ldots, m - 1 \qquad (1)$$

où le vecteur $w = (w(1), \ldots, w(m))'$
a pour loi : $p(w) = N(0, \sigma_w^2 I)$
On peut de plus choisir $x(0) = 0$.

Pour achever de déterminer par cette méthode, la loi a priori de x, il faut choisi une valeur pour σ_w^2 : on peut, soit la fixer empiriquement, soit l'estimer, ce que nous ferons en section 4.

Supposons pour l'instant connue la loi a priori $p(x) = N(x_0, R)$. Nous pouvons utiliser l'estimateur du maximum a posteriori (MAP) qui maximise :

$$p(x/y) = p(y/x)\; p(x)/p(y)$$

On montre [3] (voir appendice A) qu'il est solution de :

$$\hat{x} = A^{-1} (H' B^{-1} y + R^{-1} x_0) \tag{2}$$

où

$$A = H' B^{-1} H + R^{-1}$$

Dans le cas particulier des systèmes invariants ($h(k, i) = h(k - i)$) et causaux ($h(k) = 0, k < 0$) le problème est appelé déconvolution. Dans ce cas H est une matrice de Toeplitz triangulaire (carrée si $n = m$) :

$$H = \begin{bmatrix} h_1 & & & & \\ h_2 & h_1 & & \bigcirc & \\ \vdots & \vdots & \ddots & & \\ \vdots & h_2 & \ddots & \ddots & \\ \vdots & \vdots & \ddots & \ddots & \ddots \\ h_n & \cdots & \cdots & h_2 & h_1 \end{bmatrix}$$

A est alors une matrice proche de Toeplitz [2] et il est possible de résoudre le système linéaire (2) par des algorithmes beaucoup plus rapides [1, 3, 4] que les algorithmes classiques. Par ailleurs nous avons proposé un algorithme performant basé sur la méthode du gradient conjugué, qui permet de minimiser la forme quadratique associé à (2) avec contrainte de positivité : $x \geqslant 0$. [3,5]. Notons qu'il est important dans le problème considéré de disposer d'algorithmes performants car (2) est souvent un système de grande dimension.

Il est à noter que la méthode bayésienne n'est pas utilisée habituellement ainsi : dans les utilisations classiques, la variance représente le plus souvent une incertitude autour d'une moyenne probable ; dans l'utilisation présente, c'est la matrice de covariance qui traduit la connaissance a priori et l'espérance x_0 est plutôt un paramètre de nuissance auquel l'estimateur est d'ailleurs peu sensible.

4. ESTIMATION D'UN PARAMETRE DE LA LOI A PRIORI

Nous avons vu qu'il est facile de trouver une matrice de covariance satisfaisante, par exemple par le modèle (1), en général à un paramètre d'échelle près.

Nous considérons la situation où la loi a priori est donnée par :

$$p(x) = N(x_0, R) \quad \text{avec} \quad R = \sigma_w^2\, R_1,\ R_1 \text{ connu, } \sigma_w^2 \text{ inconnu.}$$

Il est tentant, comme le suggèrent Lindley et Smith [6] d'estimer x et σ_w^2 par les valeurs qui maximisent la probabilité a posteriori $P(x/y) = g(x, \sigma_w^2)$. Cependant cette méthode échoue du fait de la spécificité de notre problème $(m > n)$. On peut montrer en effet que $g(x, \sigma_w^2)$ n'est pas bornée supérieurement. On a (voir appendice A) :

$$G(x, \sigma_w) = P(x/y) = P(y/x)\, P(x)/P(y)$$
$$= (2\Pi)^{-m/2} \det{}^{+1/2} A \exp \left\{ -\frac{1}{2} (x - \hat{x})' A (x - \hat{x}) \right\}$$

où $\hat{x}$ et A ont été définis dans la section précédente.

On a :

$$\max_{x, \sigma_w^2} g(x, \sigma_w^2) = \max_{\sigma_w^2} \max_{x} g(x, \sigma_w^2)$$

Or le maximum sur x est réalisé par $\hat{x}$ et vaut $(2\Pi)^{-m/2} \det^{+1/2} A$.

Donc $\hat{\sigma}_w^2$ est solution de : $\max\limits_{\sigma_w^2} f(\sigma_w^2)$

où $$f(\sigma_w) = \det A = \det \left(H' B^{-1} H + \frac{1}{\sigma_w^2} R_1^{-1} \right)$$

On voit que $f(\sigma_w^2)$ tend vers l'infini si σ_w^2 tend vers zéro. De plus la valeur $\sigma_w^2 = 0$ n'est pas satisfaisante car cela conduit à prendre pour estimateur l'espérance de la loi a priori et à ne pas tenir compte des observations.

Il est possible par contre d'estimer σ_w^2 par la méthode du maximum de vraisemblance. La densité marginale de y est (voir [6]) :

$$P(y) = N(Hx_0, HRH' + B)$$

Le logarithme de la vraisemblance de σ_w^2 s'écrit (a un facteur constant prés) :

$$L(\sigma_w) = -\frac{1}{2} [\mathrm{Log} \det C + (y - Hx_0)' C^{-1} (y - Hx_0)]$$

où $$C = HRH' + B = \sigma_w^2 HR_1H' + B$$

Cette fonction est bornée supérieurement si B est régulière. En effet det $C > \det B$ et $(y - Hx_0)' C^{-1} (y - Hx_0) > 0$ donc, $L(\sigma_w^2) < -\frac{1}{2} \mathrm{Log} \det B$

D'autre part il est facile de voir que $L(\sigma_w^2)$ tend vers $-\infty$ quand σ_w tend vers $+\infty$.

Donc $L(\sigma_w^2)$ atteint son maximum sur $[0, +\infty[$. Nous admettons ici son unicité et nous prenons pour estimateur de σ_w^2 la valeur qui réalise ce maximum. Son calcul est un problème de programmation non linéaire à un paramètre (voir [7]).

5. SIMULATIONS

Nous avons mis à l'épreuve cette méthode d'estimation sur un problème test défini dans l'appendice B, en utilisant la loi a priori issue du modèle (1). La démarche est la suivante :

– nous nous donnons analytiquement les fonctions x(t) et h(t)

– nous calculons analytiquement la fonction $z(t) = \int_0^t h(t-\tau)\, x(\tau)\, d\tau$

– nous échantillonnons z(t) et ajoutons aux valeurs obtenues un bruit blanc (voir appendice B) de matrice de covariance $B = \sigma_v^2 I$

– nous estimons m valeurs de x(t), à partir de m valeurs échantillonnées de h(t), et des n valeurs échantillonnées et bruitées de z(t).

Le programme procède en deux temps :

– estimation de σ_w^2 par maximisation de $L(\sigma_w^2)$

– résolution de (2) avec $R = \sigma_w^2 R_1$ où R_1 est la matrice de covariance du processus défini par (1). Il est facile de voir que $R_1^{-1} = F'F$

où F est la matrice d'éléments $f_{ij} = \begin{cases} 1, i = j \\ -1, i - j = 1 \\ 0, \text{sinon} \end{cases}$

Le premier temps est un problème de maximisation non linéaire, à un paramètre, mais qui exige à chaque itération l'inversion et le calcul du déterminant d'une matrice n × n.

Le deuxième temps est une résolution de système linéaire de dimension m. Nous avons abordé le problème par minimisation de la forme quadratique associée, ce qui permet d'introduire la contrainte de positivité $x \geqslant 0$. En effet, dans la plupart des applications, en particulier médicales, le signal – concentration, pression, intensité – est positif, et la méthode proposée semble la façon la plus simple d'introduire cette connaissance a priori supplémentaire. L'algorithme utilisé est celui du gradient conjugué [5].

La figure 1 montre un échantillon de 64 points de y et de x. La figure 2 illustre le résultat obtenu en prenant n = 16 et m = 64, cas où le nombre d'inconnues est très supérieur au nombre d'observations.

6. CONCLUSION

Nous avons présenté une importante application du traitement de signal où l'approche bayésienne apparait nécessaire. C'est la matrice de covariance qui porte ici l'information, ce qui contraste avec les utilisations classiques de la méthode bayésienne. La solution donnée par l'estimateur du maximum a posteriori a été testée et décrite dans d'autres articles. Nous avons montré ici que l'on peut résoudre le problème de l'estimation d'un paramètre de la loi a priori par le maximum de vraisemblance, et nous avons testé l'efficacité de cette méthode en simula-

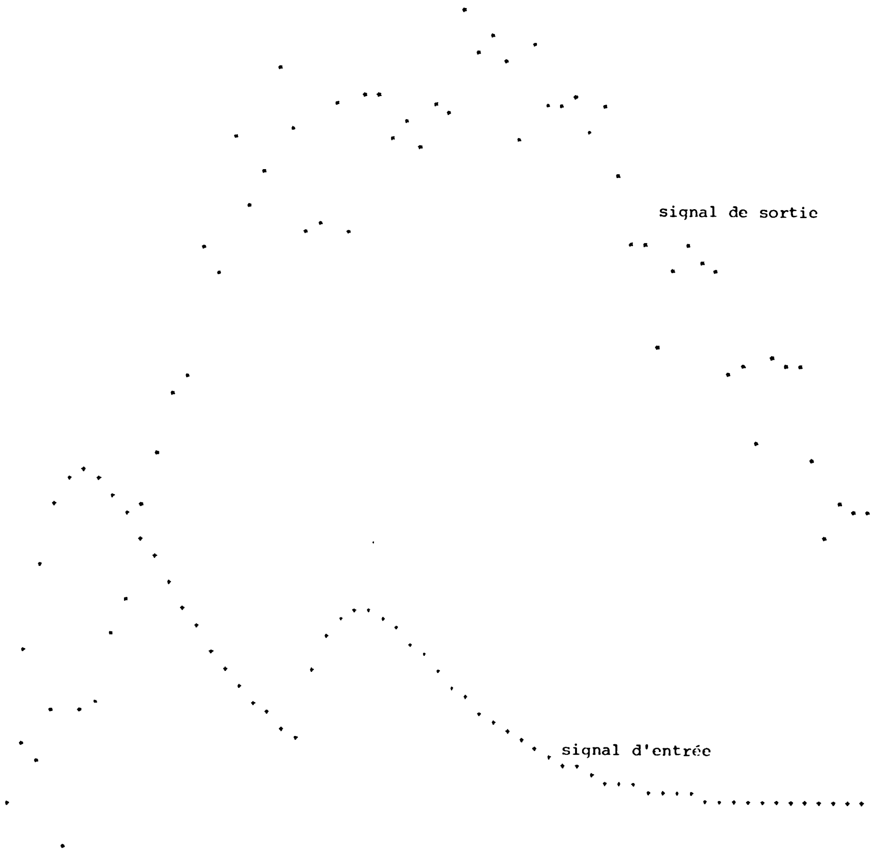

Figure 1. – Echantillonnage du signal d'entrée x(t) et du signal de sortie y(t). Seul un point sur 4 de cet échantillonnage de y a été utilisé pour l'estimation.

tion. Ce résultat nous permet d'obtenir un programme qui élimine tout empirisme dans l'obtention de la solution.

Cette méthode est utilisée dans des applications médicales, en particulier pour l'étude des systèmes vasculaires à l'aide de traceurs radioactifs [1, 3, 5, 8]. On peut estimer la distribution des temps de transit du sang a travers un organe, si on peut mesurer un signal proportionnel à la concentration du traceur, en amont et en aval de l'organe. Cette distribution est identique à la réponse impulsionnelle du système reliant les signaux amont et aval et son estimation est un problème de déconvolution.

La simplicité même de la modélisation par un système linéaire ouvre un large champ d'application à la méthode décrite ici (voir [1]).

Cette étude montre qu'elle peut être appliquée même dans le cas où le nombre d'observations est très faible ce qui permet d'aborder de nouveaux domaines où jusqu'à présent peu de résultats satisfaisants ont été obtenus (estimation de la fonction d'absorption d'un médicament pris par voie orale [9]). Enfin l'estimation d'un paramètre dela loi a priori permet une utilisation plus facile que pour les méthodes classiques.

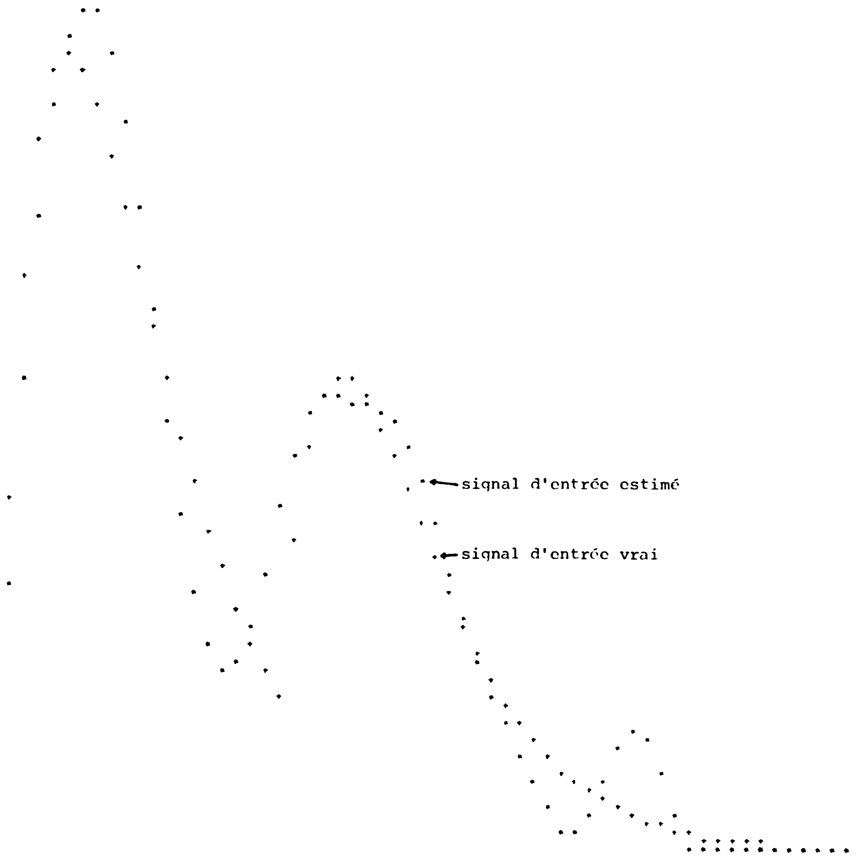

Figure 2. – Comparaison du signal d'entrée vrai $x(t)$ et du signal d'entrée estimé $\hat{x}(t)$.

APPENDICE A

Calcul de la loi a priori P(x/y)

$$P(x/y) \propto P(y/x)\,P(x)$$
$$P(y/x) = N(Hx, B)$$
$$P(x) \quad = N(x_0, R)$$

D'où :

$$P(x/y)\,\alpha \exp -\frac{1}{2}\,[(y - Hx)'\,B^{-1}(y - Hx) + (x_0 - x)'\,R^{-1}\,(x_0 - x)] \quad (1)$$

Le terme sous l'exponentielle étant une fonction quadratique des composantes de x, P(x/y) est la densité d'une loi normale. Si on désigne par $\hat{x}$ et Γ la moyenne et la matrice de covariance de cette loi on a :

$$P(x/y) = (2\Pi)^{-m/2}\,(\det \Gamma)^{-1/2} \exp -\frac{1}{2}\,[(x - \hat{x})'\,\Gamma^{-1}(x - \hat{x})] \qquad (2)$$

En identificant (1) et (2), il vient :

$$\Gamma^{-1} = H'B^{-1}\,H + R^{-1}$$
$$\hat{x} = A^{-1}\,(H'B^{-1}\,y + R^{-1}\,x_0)$$

avec
$$A = \Gamma^{-1} = H'B^{-1}H + R^{-1}$$

On a donc :

$$P(x/y) = (2\Pi)^{-m/2}\,(\det A)^{1/2} \exp -\frac{1}{2}\,[(x - \hat{x})'\,A(x - \hat{x})]$$

APPENDICE B

Problème test

$$x_1(t) = at\exp(-t) \quad \text{pour} \quad t > 0\,; \quad 0 \text{ sinon}$$
$$h(t) = \frac{1}{2}\,t\exp(-t) \quad \text{pour} \quad t > 0\,; \quad 0 \text{ sinon}$$
$$z_1(t) = \int_0^t h(t - u)\,x_1(u)\,du \qquad (4)$$

Le calcul donne :

$$z_1(t) = \frac{a}{2(1-a)^3}\,[te^{-at}\,(1 - a) - 2e^{-at} + (1 - a)\,te^{-t} + 2e^{-t}]$$

Pour obtenir un signal bimodal, nous avons utilisé :

$$x(t) = x_1(t) + bx_1(t - d\Delta t)$$

Le système étant linéaire, la ortie est :

$$z(t) = z_1(t) + bz_1(t - d\Delta t)$$

Les données fournies au programme sont :

– un échantillonnage de h(t) à une période Δt :

$$h(k\Delta t)\ ;\ k = k = 1, \ldots, m$$

– un échantillonnage bruité de z(t) comportant n valeurs, $n \leqslant m$:

$$y(k),\ k = 1, \ldots, n\ ;$$

avec :
$$y(k) = z(i_k \Delta t) + v(k)$$

où
$$i_k \in I \subset (1, \ldots, m)$$

et v(k) est une réalisation d'une variable aléatoire définie par :

$$p(v) = N(0, \sigma_v^2\, I).$$

Les valeurs choisies sont : $\Delta t = 0.1$; $a = 2$; $b = 0.5$

$$m = 64\ ;\ n = 16\ ;\ I = (1, 5, 9, \ldots, 61)$$

et
$$\sigma_v = 0.08 \sqrt{\sum_{i=1}^{n} z_i^2 / n}$$

BIBLIOGRAPHIE

[1] D. COMMENGES et A.J. BRENDEL. – Application des méthodes de déconvolution en médecine. *Innov. Tech. Biol. Med.* Vol. 2, n° 6, pp. 634-641, 1982.

[2] B. FRIEDLANDER, T. KAILATH, M. MORF. – Extended Levinson and Chandrasekhar equations for general discrete-time linear estimation problem. *IEEE Trans. on Automatic Control,* Vol. AC-23, n° 4, pp. 653-659, August 1978.

[3] D. COMMENGES. – *Etude théorique et mise en œuvre d'algorithmes de déconvolution. Application à la détermination de la distribution des temps de transits pulmonaires.* Thèse de Docteur-Ingénieur – Bordeaux 1980.

[4] M. MORF. – *Doubling algorithms for Toeplitz and related equations.* Proc. 1980 IEE ICASSP, Denver, Colo. pp. 954-959, Apr. 9-11, 1980.

[5] D. COMMENGES. – *A new deconvolution technique and its application to the study of vascular systems,* Proc. 6th IFAC Symposium on Identification and system parameter estimation, Washington, D.C., June 7-11, pp. 1 272-1 277, 1982.

[6] D.V. LINDLEY and A.F. SMITH. – Bayesian estimates for the linear model. *J.R. Statistic. Soc., B, 34, 1.* 1972.

[7] D.M. HIMMELBLAU. – *Applied non linear programming,* Mc Graw Hill, 1972.

[8] A.J. BRENDEL, D. COMMENGES, F. SAN GALLI, R. SALAMON et D. DUCASSOU. – *Déconvolution des courbes de transits pulmonaires de gamma angiocardiographies pour le calcul des shunts cardiaques : étude d'une nouvelle méthode, limitations d'ordre théorique.* Colloque de médecine nucléaire de langue française. Toulouse, Septembre 1981.

[9] P. VENG PEDERSEN. – An Algorithm and computer program for déconvolution in linear pharmacokinetics. *J. of Pharm. and Biopharmaco.*, Vol. 8, n° 5, 1980.

Je tiens à remercier J.F. BOISVIEUX et A. MALLET qui m'ont en particulier suggéré l'utilisation du modèle (1).