

METHODES EFFICIENTES POUR L'ESTIMATION D'UN RETARD NON STATIONNAIRE

M. POURMIR H. CLERGEOT

LESIR / ENS Cachan 61 ave. du Président-Wilson 94230 Cachan

RESUME

Nous proposons un ensemble de méthodes permettant une estimation efficace d'un retard non-stationnaire à partir de l'observation d'un signal et d'une version retardée et bruitée. Une première méthode fournit un estimateur au sens du Maximum de Vraisemblance (MV) des coefficients du retard dans une base de fonctions donnée. L'équation du MV obtenue itérativement d'après l'observation des signaux sur un horizon fixé. Une deuxième classe de méthodes, utilisant un formalisme du type "filtrage de Kalman", fournit un estimateur récursif en temps, soit des paramètres du modèle précédent, soit du retard lui-même. Nous montrons l'équivalence de ces deux méthodes. Des simulations confirment les bonnes performances obtenues en présence de retard rapidement variable et avec un niveau de bruit appréciable.

SUMMARY

We propose a consistent approach to non-stationary delay estimation in presence of noise from the observation of one signal and a noisy delayed version of itself. A first method is based on a parametric decomposition of the time varying delay on a set of K given functions. We derive THE Maximum Likelihood (ML) solution for the parameter estimation and its statistical properties. A fast algorithm is given to compute the MV parameters from the observation on a given time interval. We then demonstrate that the parametric model is equivalent to a state space model with a non linear observation equation. By linearization of this equation, a recursive solution is derived for the parameters estimates, or for the delay itself by the use of the Kalman equations. THE efficiency of this approach is illustrated by simulation results in presence of fast nonstationarity and a high level of noise.

I INTRODUCTION

Les mesures de retard représentent un domaine d'application extrêmement vaste, en radar, localisation de sources large bande, vélocimétrie, etc... Elles font l'objet d'un volume important de publications concernant le cas de retards constants, ou l'extension à des retards lentement variables par une estimation à horizon court sous forme adaptative. C'est le cas de la méthode de filtrage adaptatif, [1,2] qui présente de bonnes performances en poursuite; les meilleurs résultats semblent cependant être obtenus par les méthodes par boucle à verrouillage de retard [2,3].

La solution pour échapper à cette limitation de l'horizon est de modéliser les variations de retard, soit par une représentation paramétrique [4], soit par un modèle d'évolution temporel [2,5,6]. Dans [2],

un modèle d'état est introduit pour les coefficients du filtre adaptatif et conduit assez naturellement à une méthode à verrouillage de retard et modèle dynamique pour le retard lui-même.

Nous montrons que, en fait, l'approche paramétrique et l'approche par modèle d'état sont identiques, à un changement de repère près dans l'espace d'état. L'approche paramétrique conduit naturellement à une solution par le maximum de vraisemblance, sur un horizon donné (paragraphe II). Le modèle d'état conduit à une méthode de type Kalman, qui a l'avantage d'être récursive en temps (paragraphe III). L'introduction d'une perturbation dans le modèle d'évolution est bien entendu indispensable dans le cas d'estimation à horizon infini pour donner une possibilité d'évolution aux paramètres; ce point est discuté dans le paragraphe IV.



II MODELE PARAMETRIQUE DU RETARD

Pour alléger la discussion, nous avons fait un certain nombre d'hypothèses simplificatrices. On se limite au cas de deux signaux, le deuxième étant une réplique décalée et bruitée du premier:

$$(1) \quad \begin{cases} x(t) \text{ connu} \\ z(t) = x(t-D(t)) + w(t) \end{cases}$$

où le bruit $w(t)$ est décorrélé de $X(t)$, blanc, gaussien, de variance σ^2 . Le retard $D(t)$ est à estimer à partir d'observations échantillonnées. L'intervalle d'échantillonnage est pris comme unité de temps; l'observation est donc de la forme $Z(n) = Z_n$.

II-1 Equations du maximum de vraisemblance et bornes de Cramer Rao

On suppose que les variations du retard sont représentables par une décomposition sur une base de K fonction $\psi_k(n)$ sous la forme :

$$(2) \quad D_n = \sum_{k=1}^K \psi_{nk} \cdot d_k, \quad \psi_{nk} \stackrel{\text{(def)}}{=} \psi_k(n)$$

soit, en désignant par \underline{D} le vecteur retard, de dimension N , par \underline{d} le vecteur paramètre de dimension K et par Ψ la matrice $N \times K$ des coefficients, sous forme matricielle:

$$(3) \quad \underline{D} = \Psi \cdot \underline{d}$$

On notera $\underline{\psi}_k$ les vecteurs colonne de Ψ , de dimension N et correspondant aux fonctions de base $\psi_k(n)$, et $\underline{\psi}(n)$ les vecteurs ligne, de dimension K . Les relations (2) et (3) peuvent se réécrire :

$$(3 \text{ bis}) \quad \begin{aligned} \underline{D} &= d_1 \underline{\psi}_1 + \dots + d_K \underline{\psi}_K \\ D_n &= \underline{\psi}(n) \cdot \underline{d} \end{aligned}$$

La seule grandeur aléatoire du modèle étant le bruit $w(t)$, la Log vraisemblance $V(\underline{d})$ est donnée par :

$$(4) \quad V(\underline{d}) = -\frac{1}{\sigma^2} \sum_{n=1}^N [z_n - x(n-D_n)]^2 + C^t$$

L'estimée de \underline{d} au maximum de vraisemblance (ou aux moindres carrés!) doit rendre extrémale cette expression. Elle satisfait donc le système d'équations :

$$(5) \quad \begin{cases} \frac{\partial V}{\partial d_k} = -\frac{1}{\sigma^2} \sum_{n=1}^N [z_n - x(n-D_n)] \cdot \dot{x}(n-D_n) \cdot \psi_{nk} = 0 \\ k=1, \dots, K \end{cases}$$

Une résolution itérative sera proposée au paragraphe suivant. Auparavant l'expression peut être utilisée pour calculer les bornes de Cramer-Rao. Les bornes sont calculées dans l'hypothèse d'un biais nul : $Y(t)$ correspond exactement au modèle pour $\underline{d} = \underline{d}_0$. Les bornes se calculent à partir de la matrice de Fisher F dont les éléments sont donnés par [7] :

$$(6) \quad \begin{cases} F_{kh} = -E \left[\frac{\partial^2 V}{\partial d_k \partial d_h} \right]_{\underline{d}=\underline{d}_0} = \frac{1}{\sigma^2} \sum_{n=1}^N \dot{x}(n-D_n)^2 \psi_{nk} \psi_{nh} \\ \text{cov}[\underline{d}] = E[(\underline{d}-\underline{d}_0)(\underline{d}-\underline{d}_0)^t] = F^{-1} \end{cases}$$

Désignons par S la matrice diagonale d'éléments $\{\dots, \dot{x}(n-D_n)^2, \dots\}$, la relation (6) se réécrit :

$$(7) \quad F = \sigma^{-2} \Psi^t S \Psi$$

ce qui permet d'exprimer facilement la covariance du vecteur retard \underline{D} , plus parlante que celle de \underline{d} . D'après (3) :

$$(8) \quad \text{cov}(D) = \Psi \text{cov}(\underline{d}) \Psi^t = \sigma^2 \Psi^t \Psi^{-1} \Psi^t \Psi^t$$

Cette expression se simplifie remarquablement en utilisant comme intermédiaire la décomposition en valeur singulière de la matrice Ψ :

$$(9) \quad \begin{aligned} \Psi &\stackrel{\text{def}}{=} \sum_{i=1}^K \alpha_i U_i U_i^t \\ \text{cov}(D) &= \sigma^{-2} \sum_{i=1}^K U_i U_i^t \cdot (U_i S U_i^t)^{-1} \end{aligned}$$

On obtient donc la forme diagonale de la covariance, les U_i formant une base orthonormée pour l'espace paramétrique engendré par les fonctions de base $\underline{\psi}_k$. Pour un signal d'allure stationnaire les $\dot{x}(n)^2$ sont en moyenne constants et $U_i^t S U_i \approx \bar{x}^2$. On obtient alors pour l'erreur moyenne sur le retard D_n :

$$(10) \quad E[(\hat{D}_n - D_n)]_{\text{moyen}} = 1/N \text{Tr}[\text{cov} D] \approx (K/N) \cdot (\sigma^2 / \bar{x}^2)$$

Cette expression est particulièrement simple à interpréter: le terme σ^2 / \bar{x}^2 correspond au fait qu'un déplacement "vertical" de σ (bruit additif) est interprété comme un décalage horizontal σ / \dot{x} (décalage en temps). Le facteur $1/N$ montre l'intérêt d'un horizon long, ce qui permet le modèle à retard variable. Le facteur K au numérateur est la contre-partie: ne pas utiliser trop de paramètres, en particulier si on sait que le retard à mesurer varie peu.

II-2 Résolution des équations de MV

Le système (5) se prête à une résolution itérative par un algorithme simple à convergence très rapide s'il est initialisé par trop loin de la solution. Soit D_{no} la valeur optimale et D_n une valeur voisine du retard. Dans l'expression (5) du gradient :

$$(11) \quad \begin{aligned} z - x(n-D_n) &= r + x(n-D_{no}) - x(n-D_n) \\ &\approx r + (D_n - D_{no}) \cdot \dot{x}(n-D_n) \end{aligned}$$

dans les limites de validité de l'approximation de Taylor entre D_{no} et D_n , r_n désignant le reste après convergence quand $D_n = D_{no}$. Compte tenu de (3) et (8) la relation (5) devient :

$$(12) \quad \begin{aligned} \nabla_k &\stackrel{\text{def}}{=} -\sigma^2 \frac{\partial V}{\partial d_k} \approx \sum_{n=1}^N w_n \dot{x}(n-D_n) \psi_{nk} \\ &\quad + \sum_{n=1}^N \psi_{nk} \dot{x}(n-D_n)^2 \underline{\psi}(n) (\underline{d} - \underline{d}_0) \end{aligned}$$

En remarquant que la première somme, qui représente la valeur du gradient après convergence est nulle, on obtient, sous forme vectorielle :

$$(12) \quad \underline{\nabla} \approx \Psi^t S \Psi (\underline{d} - \underline{d}_0)$$

on en déduit le gain matriciel à donner au vecteur gradient $\underline{\nabla}$ pour converger rapidement vers l'optimum. L'algorithme itératif correspondant est :

$$(13) \quad \underline{d}(i+1) = \underline{d}(i) - [\Psi^t S(i) \Psi]^{-1} \underline{\nabla}(i)$$

La matrice de gain peut être approximée par une matrice plus simple. En particulier si les fonctions de bases sont orthonormées et si on suppose que, en moyenne, $\dot{x}(n)^2$ reste constant on peut faire l'approximation :

$$(14) \quad S(i) \approx \bar{x}^2 \cdot I_N, \quad \Psi^t S(i) \Psi \approx \bar{x}^2 \Psi^t \Psi = \bar{x}^2 \cdot I_K$$

L'inconvénient de l'algorithme est de travailler en temps différé. Nous allons envisager une autre méthode utilisant une représentation d'état pour le modèle admettant une solution récursive en temps classique par les équations de Kalman.

III REPRESENTATION PAR UN MODELE D'ETAT
(HORIZON T FINI)

III-1 Représentation directement associée au modèle paramétrique

Le principe est d'utiliser comme équations d'état associée au modèle paramétrique les relations (voir 3 bis) :

$$(15) \quad \begin{cases} \underline{d}_{n+1} = \underline{d}_n \\ \underline{z}_n = x [t_n - \underline{\psi}^t(n) \cdot \underline{d}_n] + w_n \end{cases}$$

La difficulté vient du fait que la deuxième relation, l'équation d'observation, est non linéaire. Supposons que l'on connaisse une valeur $\tilde{\underline{d}}_n$ voisine de \underline{d}_n . En effectuant un développement de Taylor de x au voisinage de $t_n + \underline{\psi}^t(n) \cdot \tilde{\underline{d}}_n$ on obtient :

$$\underline{z}_n \approx x(t_n - \underline{\psi}^t(n) \tilde{\underline{d}}_n) - \dot{x}(t_n - \underline{\psi}^t(n) \tilde{\underline{d}}_n) \cdot \underline{\psi}^t(n) (\underline{d}_n - \tilde{\underline{d}}_n)$$

ce qui donne une équation linéaire en \underline{d}_n :

$$(16) \quad \begin{cases} \underline{u}_n = \underline{H}_n^t \underline{d}_n, \text{ avec:} \\ \underline{u}_n \stackrel{\text{def}}{=} \underline{z}_n - x(t_n - \underline{\psi}^t(n) \tilde{\underline{d}}_n) + \underline{H}_n^t \tilde{\underline{d}}_n \\ \underline{H}_n \stackrel{\text{def}}{=} -\dot{x}(t_n - \underline{\psi}^t(n) \tilde{\underline{d}}_n) \cdot \underline{\psi}(n) \end{cases}$$

La résolution de ce système par les équations de Kalman donne récursivement sur l'horizon $[0, n]$ l'estimée du maximum de probabilité a posteriori, voisine du MV si la covariance initiale est grande. Pour $n = N$ et si $\tilde{\underline{d}}_n$ peut être maintenu très voisin de \underline{d}_n , on doit donc retrouver le résultat du paragraphe II.

Le choix évident pour $\tilde{\underline{d}}_n$ est de prendre l'estimée filtrée de l'itération précédente de l'algorithme de Kalman.

On obtient à l'itération n les opérations suivantes :

$$17) \quad \left\{ \begin{array}{l} \text{(a) Filtrage: } \hat{\underline{d}}_n = \hat{\underline{d}}_{n-1} \\ \text{(b) Calcul } \underline{H}_n: \underline{H}_n = -\dot{x}(n - \underline{\psi}^t(n) \hat{\underline{d}}_n) \cdot \underline{\psi}(n) \\ \text{(c) Calcul covariance: } \underline{R}_n = \underline{R}_{n-1} - \frac{\underline{R}_{n-1} \underline{H}_n^t \underline{H}_n \underline{R}_{n-1}}{\sigma^2 + \underline{H}_n^t \underline{R}_{n-1} \underline{H}_n} \\ \text{(d) Mise à jour: } \hat{\underline{d}}_n = \hat{\underline{d}}_n + \frac{\underline{z}_n - x(n - \underline{\psi}^t(n) \hat{\underline{d}}_n)}{\sigma^2} \underline{R}_n \underline{H}_n \end{array} \right.$$

Par rapport aux équations de Kalman classiques, on rajoute le calcul de \underline{H}_n fonction de $\hat{\underline{d}}_n$; on note la forme remarquable de l'équation de mise à jour, qui évoque la boucle à verrouillage de retard.

Remarquons que, au bout de N itérations, la matrice de covariance de $\hat{\underline{d}}_N$ est donnée par :

$$(18) \quad \underline{R}_{n-1}^{-1} = \underline{R}_0^{-1} + \sum_1^N \sigma^{-2} \underline{H}_n^t \underline{H}_n = \underline{R}_0^{-1} + \sigma^{-2} \underline{\Psi}^t \underline{S} \underline{\Psi}$$

on retrouve bien la matrice de Fisher du paragraphe II si $\underline{R}^{-1} \approx 0$.

Du point de vue algorithmique, l'algorithme de Kalman est le plus souvent initialisé avec $\underline{R}_0 = a^2 \underline{I}$, avec a^2 très grand. Dans ce cas, l'estimée calculée par (17) est indépendante de σ^2 qui n'a pas besoin d'être connu.

Pour le problème présent du fait de la non-linéarité de l'équation d'observation, l'algorithme doit forcément être initialisé par une valeur $\tilde{\underline{d}}_0$ voisine de \underline{d} . Il est alors souhaitable d'initialiser \underline{R}_0 par une matrice diagonale dont les éléments sont un majorant de la valeur de l'écart $|\tilde{\underline{d}}_k - \underline{d}_k|^2$ correspondant. Ceci donne dès le départ une sensibilité moins grande de l'algorithme au bruit. Il faut alors également que σ^2 ait une valeur correspondant au bruit réel.

III-2 Utilisation du retard comme vecteur d'état

Considérons le vecteur $\underline{D}_n = (\underline{D}_n, \underline{D}_{n-1}, \dots, \underline{D}_{n-k+1})^t$, dans le modèle précédent il se relie à \underline{d} par la relation suivante ($\underline{d}_n = c^{te} = \underline{d}$) :

$$(19) \quad \begin{aligned} \underline{D}_n &= \underline{\Phi}_n \underline{d} = \underline{\Phi}_n \underline{d}_n \\ \underline{\Phi}_n &= \begin{bmatrix} \underline{\psi}^t(n) \\ \dots \\ \underline{\psi}^t(n-k+1) \end{bmatrix} \end{aligned}$$

La matrice $\underline{\Phi}_n$ est un bloc $K \times K$ extrait de la matrice $\underline{\Psi}$. Si elle est régulière, la relation (19) peut s'interpréter comme un simple changement de base dans l'espace d'état. Les équations d'état précédentes se transposent en :

$$(20) \quad \begin{aligned} \underline{D}_n &= \underline{\Phi}_n \underline{\Phi}_{n-1}^{-1} \underline{D}_{n-1} = \underline{A}_n \underline{D}_{n-1} = \underline{A}_n \underline{D}_{n-1} \\ \underline{u}_n &= \underline{H}_n^t \underline{\Phi}_n^{-1} \underline{D}_n = \underline{H}_n^t \underline{D}_n \end{aligned}$$

Dans la mesure où $\underline{\Phi}_{n+1}$ a toutes ses lignes en commun avec $\underline{\Phi}_{n-1}$ sauf la première on obtient :

$$(21) \quad \underline{A}_n = \begin{bmatrix} \underline{A}_{n-1}^t & \\ \dots & \\ 1 & \dots & (0) \\ (0) & \dots & 1 & 0 \end{bmatrix}, \quad \underline{A}_n = \underline{\Phi}_{n-1}^{-1} \underline{\Psi}(n)$$

et :

$$(22) \quad \underline{H}_n = -\dot{x}(n - \underline{\psi}^t(n) \hat{\underline{d}}_n) \underline{\psi}(n) = -\dot{x}(n - \underline{\psi}^t(n) \hat{\underline{d}}_n) \cdot \begin{bmatrix} 1 \\ \vdots \\ 0 \end{bmatrix}$$

Inversement si l'on cherche directement un modèle d'état pour \underline{D}_n , on obtient nécessairement les structures (21) et (22) pour \underline{A}_n et \underline{H}_n , avec un vecteur \underline{A}_n arbitraire. Par produit des matrices \underline{A}_k de $k = 2$ à

n on obtient une relation de la forme :

$$(23) \quad \underline{D}_n = \tilde{\underline{\Phi}}_n \underline{D}_0$$

où les matrices $\tilde{\underline{\Phi}}_n$ ont nécessairement la structure (19) du fait de la relation entre vecteurs \underline{D}_n successifs; l'ensemble des vecteurs lignes peut être regroupé dans une matrice $\tilde{\underline{\Psi}}$ $N \times r$ telle que :

$$(24) \quad \underline{D} = \tilde{\underline{\Psi}} \underline{D}_0$$

ce qui constitue bien une représentation paramétrique de vecteur \underline{D}_0 .

Il y a donc une exacte équivalence entre les deux méthodes. La transposition des équations (19) est la suivante :

$$(25) \quad \begin{aligned} \text{(a) } \underline{\tilde{D}}_n &= \underline{A}_n^t \underline{\tilde{D}}_{n-1} \quad (\Leftrightarrow \underline{\tilde{D}}_n = \underline{A}_n^t \underline{\tilde{D}}_{n-1}) \\ \text{(b) calcul } \dot{x}(n - \underline{\psi}^t(n) \hat{\underline{d}}_n) &\stackrel{\text{def}}{=} \dot{x}_n \end{aligned}$$



$$(c) \tilde{R}_n = A_n R_{n-1} A_n^t$$

$$R_n = \tilde{R}_n - \frac{\tilde{x}_n^2}{\sigma^2 + \tilde{r}_n^2 \tilde{x}_n^2} \tilde{R}_n \tilde{R}_n^t$$

$$(d) \hat{D}_n = \tilde{D}_n - \frac{\tilde{x}_n^2}{\sigma^2} (z_n - x(n - \tilde{D}_n)) \tilde{R}_n$$

Dans ces relations, \tilde{R}_n désigne la première colonne de la matrice \tilde{R}_n et \tilde{r}_n^2 l'élément (1,1) de \tilde{R}_n , qui est aussi la covariance de \tilde{D}_n . Dans la relation a), seule la première composante de \tilde{D}_n doit être calculée, les autres résultant d'un simple décalage des composantes de \hat{D}_{n-1} .

On constate que le système d'équation (25) est plus simple que le système (19) : il suffit de connaître la séquence des A_n , qui peut avoir été calculée une fois pour toute. D'autre part, le terme r_n^2 donne directement la covariance du retard D_n . Par contre, la première représentation peut être plus commode comme outil d'analyse du comportement de l'algorithme, en suivant l'évolution des paramètres en cours de convergence.

Un cas particulier important est celui où le vecteur A_n est constant. Remarquons qu'il définit alors une relation de récurrence entre les retards

$$(26) \quad D_n = \sum_{k=1}^K a_k D_{n-k}$$

Les fonctions de base correspondantes sont déterminées par les racines Z_k du polynôme $A(z)$ associé :

$$(27) \quad A(z) = 1 - \sum_{k=1}^K a_k z^{-k}$$

Ce sont des exponentielles $\psi_k(n) = Z_k^n$ en cas de pôles simples, ou des exponentielles et des polynômes. Une base purement polynômiale correspond au cas où $z = 1$ est racine d'ordre K : $A(Z) = (1-Z)^K$.

Le cas où A_n est lentement variable peut s'interpréter comme une variation lente des fréquences des exponentielles de base. Ceci ouvre des perspectives de modèle adaptatif, où le vecteur A_n serait choisi récursivement d'après une évaluation des coefficients de la relation (26) par un algorithme adaptatif, où le vecteur A_n serait choisi récursivement d'après une évaluation des coefficients la relation (24) par un algorithme adaptatif.

IV SIMULATIONS

La figure (1a) représente le signal $x(t)$ utilisé pour la simulation et le signal retardé par un retard rapidement variable comprenant un retard fixe et deux retards variant sinusoïdalement avec les périodes $T/5$ et $T/10$ et les amplitudes respectives 5, 25 et 30. La figure (1b) représente le même signal $y(t)$ bruité avec un rapport signal sur bruit de 8.5 dB. La figure (2) représente la convergence de l'algorithme du MV dans les deux cas en fonction du nombre d'itérations. La figure (3) représente la convergence de l'algorithme de Kalman en fonction du temps, avec et sans bruit. La figure (4) donne la valeur correspondante de l'estimation du retard obtenue récursivement également par l'algorithme de Kalman.

V CONCLUSION

Nous avons montré les liens étroits entre l'approche paramétrique par le MV et une approche récursive utilisant le filtrage de Kalman pour la poursuite d'un retard rapidement variable.

Les simulations démontrent l'efficacité de cette méthode dans des conditions où les méthodes classiques se trouvent en défaut.

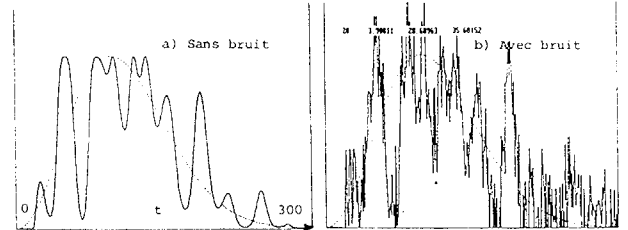


FIG 1: Signal initial et signal à retard variable

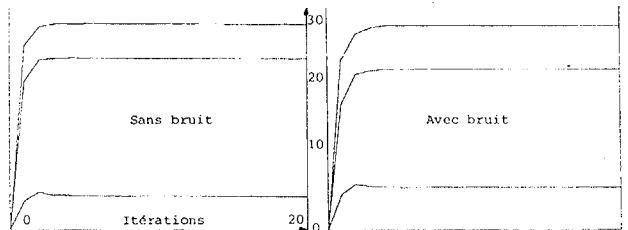


FIG 2: MV-Paramètres estimés fonction de l'itération

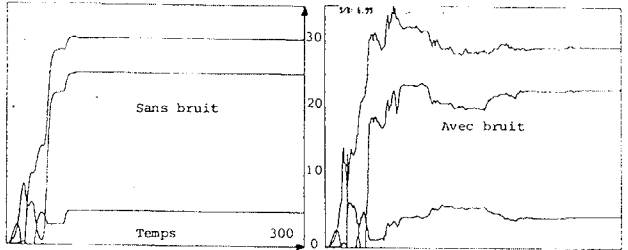


FIG 3: Kalman-Paramètres estimés fonction du temps

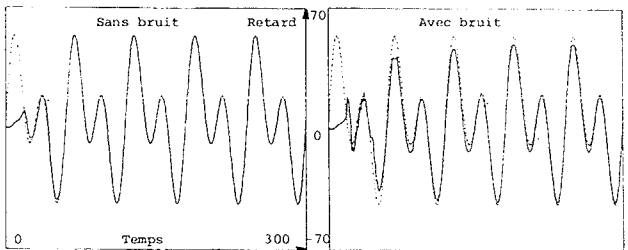


FIG 4: Retard vrai et estimé fonction du temps

REFERENCES :

- [1] P.L.FEINTUCH, N.J.BERSHAD, F.A.REED, "Time Delay Estimation Using the LMS Adaptive Filter", IEEE Trans. on ASSP, pp 571-576, June 1981
- [2] B.FENEUIL, Thèse de Doctorat de l'Université Paris-Sud, 1988.
- [3] D.U. ETTER, S.D.STEARNS, "Adaptive Estimation of Time Delays in Sampled Data Systems", IEEE Trans. on ASSP, pp 582-587, June 1981
- [4] P.M.SCHULTHEISS, E.WEINSTEIN, "Lower Bounds on the Localization Errors of a Moving Source Observed by a Passive Array", IEEE Trans. on ASSP, pp 600-607, June 1981.
- [5] L.KYRLIN, D.MOORE, R.KUBICHEK, "Improvement of Delay Measurements from Sonar Arrays via Sequential State Estimation", IEEE Trans. on ASSP, pp 514-519, June 1981.
- [6] P.M. SCHULTHEISS, "Locating a Passive Source with Array Measurements : a summary of results", in IEEE 1979 ICASSP Rec., 79CH1379.7 ASSP, pp 967-970.
- [7] H.L. Van Trees : Detection, Estimation, and Modulation Theory, Wiley 1968.
- [8] B. ANDERSON, J. MOORE : Optimal Filtering, Prentice-Hall 1979.
- [9] BUCY, JOSEPH : Filtering for Stochastic Processes with Applications to Guidance, Interscience 1968.