DIXIEME COLLOQUE SUR LE TRAITEMENT DU
SIGNAL ET SES APPLICATIONS

NICE du 20 au 24 MAI 1985

THEORIE DEDUITE DE L'INEGALITE DE CRAMER-RAO : FILTRAGE DES VECTEURS

Hubert DEBART

SINTRA - 1, Avenue Aristide Briand - 94117 ARCUEIL CEDEX

RESUME

SUMMARY

RESUME :

On se propose de définir ici une méthode simple pour le lissage, le filtrage ou la prédiction des vecteurs aléatoires observés à des instants discrets. Cette méthode repose sur la théorie des variables gaussiennes liées et l'inégalité générale de CRAMER-RAO.
On aboutit à un procédé de filtrage (ou lissage, ou prédiction) récurrent, mais dans un sens très large, c'est-à-dire à partir d'un ensemble d'observations passées, en nombre arbitraire, contemporaines ou non, indépendantes ou non. Elles s'appliquent quand la partie aléatoire des vecteurs est additive et gaussienne.

ABSTRACT :

This is an attempt to define a simple method for smoothing, filtering, or prediction of random vectors, whose observation takes place at discrete instants. Its basis is the theory of bounded gaussian variables and, on the other hand, the general inequality of CRAMER-RAO. It results in a filtering (or smoothing, or predicting) device, recurrent but a very wide sense, i.e. from an arbitrary set of past observations, whose number, times, dependancy, are arbitrary. It is of use if the random component of the vectors is additive and gaussian.

THEORIE DEDUITE DE L'INEGALITE DE CRAMER-RAO : FILTRAGE DES VECTEURS

1.- INTRODUCTION

Après avoir rappelé la théorie des variables gaussiennes dont une partie des composantes est fixée, on la généralise au cas d'une distribution de probabilités partielles données.

On rappelle ensuite l'inégalité générale de CRAMER-RAO et on définit des opérateurs linéaires optimaux pour le cas d'un vecteur d'observation à bruit gaussien additif.

On applique alors ces résultats à des processus aléatoires gaussiens observés avec erreurs ; on définit une méthode de filtrage récurrent au sens large ; l'inégalité de CRAMER-RAO permet de passer au cas plus général de l'observation indirecte d'un processus.

2.- LOI LIEE DES VARIABLES GAUSSIENNES

2.1.- Loi liée par des valeurs certaines

On considère une variable gaussienne à n dimensions : $y_1 \ldots y_n$.
Elle est déterminée par les cumulants du premier ordre $K_1(y_i)$ et du deuxième ordre $K_2(y_i, y_j)$, c'est-à-dire les moyennes et les corrélations centrées ordinaires.
On suppose que la valeur de p composantes est donnée ; les n-p = q composantes restantes formant une variable gaussienne à q dimensions dont on cherche la loi de probabilité
Pour mieux séparer les variables, on appellera
$\xi_1 \ldots \xi_p$
les variables fixées : $\xi_1 = a_1 \ldots \xi_p = a_p$
et $x_1 \ldots x_q$.
les variables restantes.
Soit la matrice $[k_2(\xi_i, \xi_j)]$ d'ordre p relative aux variables fixées et $a_2(\xi_i,\xi_j)$ l'inverse de cette matrice.

Les cumulants liés $\tilde{K}_1(X_i)$ et $\tilde{K}_2(X_i,X_j)$ sont alors calculables par des formules générales.

$$\tilde{K}_1(X_i) = K_1(X_i) + \left[K_2(X_i,\xi_1) \ldots K_2(X_i,\xi_p)\right]\left[a_2(\xi_i,\xi_j)\right]\begin{bmatrix} a_1 - K_1(\xi_1) \\ \vdots \\ a_p - K_1(\xi_p) \end{bmatrix} \quad (1)$$

$$\tilde{K}_2(X_i,X_j) = K_2(X_i,X_j) - \left[K_2(X_i,\xi_1) \ldots K_2(X_i,\xi_p)\right]\left[a_2(\xi_i,\xi_j)\right]\begin{bmatrix} K_2(X_j,\xi_1) \\ \vdots \\ K_2(X_j,\xi_p) \end{bmatrix} \quad (2)$$

2.2.- Loi liée par une distribution de probabilité

On peut généraliser les formules précédentes en supposant que l'ensemble $\xi_1 \ldots \xi_p$ est défini non par des valeurs certaines mais par une distribution gaussienne a priori $p(\xi_1 \ldots \xi_p)$.

Par exemple, soit une loi gaussienne à quatre dimensions $(\xi_1, \xi_2, X_1 X_2)$ et supposons la centrée sur toutes ses composantes.
La probabilité a priori du couple (ξ_1,ξ_2) est caractérisée par des variances $V_1 V_2$ et une corrélation C.
Soit à calculer la corrélation (ξ_1, X_1). Si ξ_1,ξ_2 sont fixés la moyenne de X_1 est de la forme :

$$\tilde{K}_1(X_1) = M\,\xi_1 + N\,\xi_2$$

d'après l'expression (1).

La corrélation cherchée est alors la moyenne pondérée.

$$\tilde{\tilde{K}}_2(\xi_1,X_1) = \int p(\xi_1,\xi_2)\,\xi_1(M\xi_1+N\xi_2)\,d\xi_1 d\xi_2 = MV_1 + NC \quad (3)$$

De même, par exemple, la variance $\tilde{\tilde{K}}_2(X_1,X_2)$
On a pu calculer $\tilde{K}_2(X_1,X_2) = P$
Si ξ_1,ξ_2 sont fixés $X_1 = M\xi_1 + N\xi_2 + \varepsilon$
ε est une variable aléatoire centrée : $\varepsilon^2 = p$
Donc :

$$\tilde{\tilde{K}}_2(X_1,X_2) = \int p(\xi_1,\xi_2)\left[(M\xi_1+N\xi_2)^2 + \varepsilon^2\right] d\xi_1\, d\xi_2 = P + M^2V_1 + 2MNC + N^2V_2 \quad (4)$$

Toutes les autres grandeurs sont calculées d'après le même principe.

3.- INEGALITE DE CRAMER-RAO

C'est un résultat très général qui permet de connaître une borne inférieure indépassable de la variance obtenue par des mesures indirectes.

3.1.- Expression générale

Si R désigne un vecteur de mesure permettant d'estimer la valeur d'un paramètre, $\hat{a}(R)$ est la valeur estimée, A la vraie valeur :

$$\mathrm{Var}\,|\hat{a}(R) - A| \geq \left[E\left[\frac{\partial \mathrm{Log}\ \mathrm{Pr}\ (a(R/A)}{\partial A}\right]^2\right]^{-1} \quad (5)$$

Si on veut mesurer simultanément plusieurs paramètres, avec un vecteur d'observation dont les erreurs sont additives, gaussiennes centrées et dépendantes, on peut exprimer cette inégalité d'une façon générale :

$X_{1o} \ldots X_{no}$ sont les composantes du vecteur d'observation,

$Y_1 \ldots Y_p$ sont les paramètres à mesurer.

S est la matrice de covariance des erreurs d'observations (nxn).

On forme la matrice M

$$M = \begin{vmatrix} \frac{\partial X_{1o}}{\partial Y_1} & \frac{\partial X_{no}}{\partial Y_1} \\ \frac{\partial X_{1o}}{\partial Y_p} & \frac{\partial X_{no}}{\partial Y_p} \end{vmatrix} \qquad (5)$$

et la matrice de FISHER $MS^{-1} M^T$.

La matrice pxp : $(MS^{-1}M^T)^{-1}$ est une matrice optimale des erreurs d'estimations de $y_1 \ldots y_p$, c'est-à-dire : la matrice est diagonalisable, avec p vecteurs propres qui représentent des paramètres indépendants, la limite de CRAMER-RAO donne les variances minimales pour ces paramètres.

3.3.- Opérateur optimal

On peut alors définir pour chaque paramètre, un opérateur linéaire sur $X_{10} \ldots X_{no}$ qui fournit l'estimation optimale de ce paramètre y_1.

Soit $a_1 X_1 + .. + a_n X_n$, cet opérateur (on simplifie $X_{io} \rightarrow X_i$).

On diagonalise la matrice S sous la forme :

$$S = |V_1 \ldots V_n| \begin{vmatrix} \lambda_1 & \\ & \lambda_n \end{vmatrix} \begin{vmatrix} V_1^T \\ V_n^T \end{vmatrix}$$

dans laquelle $\lambda_1 \ldots \lambda_n$ sont les valeurs propres de la matrice et $V_1 \ldots V_n$ ses vecteurs propres normalisés.

De même :

$$S^{-1} = |V_1 .. V_n| \begin{vmatrix} 1/\lambda_1 & \\ & 1/\lambda_n \end{vmatrix} \begin{vmatrix} V_1^T \\ V_n^T \end{vmatrix}$$

On appelle y_i le paramètre considéré : on décompose le vecteur colonne $\left|\frac{\partial X_j}{\partial y_i}\right|$ suivant les vecteurs propres de S ou S^{-1}.

$$M = \left|\frac{\partial X_j}{\partial y_i}\right| = \alpha_1 V_1 + \ldots + \alpha_n V_n$$

Comme les vecteurs propres sont orthogonaux

$$MS^{-1}M^T = \frac{\alpha_1^2}{\lambda_1} + \ldots \frac{\alpha_n^2}{\lambda_n}$$

On décompose de la même façon le vecteur colonne

$$\begin{vmatrix} a_1 \\ a_n \end{vmatrix} = \beta_1 V_1 + \ldots + \beta_n V_n$$

La variance optimale attachée au traitement linéaire ainsi défini est :

$$1/ \frac{\left(a_1 \cdot \frac{\partial X_1}{y_i} + \ldots + a_n \frac{\partial X_n}{y_i}\right)^2}{\begin{vmatrix} a_1 .. a_n \end{vmatrix} \begin{vmatrix} S \end{vmatrix} \begin{vmatrix} a_1 \\ a_n \end{vmatrix}}$$

$$= 1/ \frac{(\Sigma\, \alpha_i \beta_i)^2}{\beta_1^2 \lambda_1 + \ldots + \beta_n^2 \lambda_n}$$

On maximise l'expression dénominateur, ou encore

$$\Sigma(\alpha_i \beta_i)^2 - \mu \left[\beta_i^2 \lambda_1 + \ldots + \beta_n^2 \lambda_n\right]$$

d'après la méthode de LAGRANGE, ce qui donne les conditions :

$$2\alpha_i (\Sigma \alpha_i \beta_i) - 2\mu\, \lambda_i \beta_i = 0$$

donc :

$$\frac{\lambda_1 \beta_1}{\alpha_1} = \ldots = \frac{\lambda_n \beta_n}{\alpha_n}$$

On prend donc $\beta_i = \frac{\alpha_i}{\lambda_i}$, à un facteur près.

Le maximum de l'expression est alors égal à :

$$\varepsilon \frac{\alpha_i^2}{\lambda_i}$$

On détermine le facteur arbitraire en ajustant la variance du vecteur optimal à celle déduite de l'inégalité.

L'opération peut être effectuée pour chaque paramètre.

4.- APPLICATION AUX PROCESSUS ALEATOIRES

4.1.- Processus stationnaire gaussien simple

La première partie du paragraphe 2, donne la solution de la prédiction (ou interpolation) optimale de ce processus. On suppose que p échantillons du processus sont connus exactement, on peut alors calculer la moyenne liée de q échantillons différents du premier lot, ce qui fournit la prédiction optimale, et la matrice de covariance liée de ces q échantillons, ce qui fournit la matrice de covariance des erreurs.

EXEMPLE : On considère un processus stationnaire gaussien centré dont on mesure des échantillons à la période de 1, la loi de corrélation stationnaire du processus est en $e^{-|T|/4}$. On connait 4 échantillons consécutifs et on veut prédire les deux suivants (X_0, X_1).

Le prédicteur est alors défini par l'expression :

$$\overline{\begin{matrix} .7788 & .6065 & .4724 & .3679 \end{matrix}} \begin{bmatrix} 1 & .7788 & .6065 & .4724 \\ & 1 & .7788 & .6065 \\ & & 1 & .7788 \\ & & & 1 \end{bmatrix}^{-1} \begin{bmatrix} \xi_{-1} \\ \xi_{-2} \\ \xi_{-3} \\ \xi_{-4} \end{bmatrix}$$

Pour l'échantillon X_a par exemple :

$$\overline{\xi_{-4} \quad \xi_{-3} \quad \xi_{-2} \quad \xi_{-1} \quad X_0 \quad X_1}$$

On a alors :

$$\hat{X}_0 = 0{,}7088\xi_{-1} - 0{,}0001\xi_{-2} + 0{,}0005\xi_{-3} - 0{,}0001\xi_{-4}$$

$$\hat{X}_1 = 0{,}5518\xi_{-1} + 0{,}0003\xi_{-2} + 0{,}0003\xi_{-3} - 0{,}0001\xi_{-4}$$

avec une matrice de covariance $\tilde{K}_2(X_0, X_1)$

$$\begin{bmatrix} 0{,}3478 & 0{,}3488 \\ * & 0{,}6651 \end{bmatrix}$$

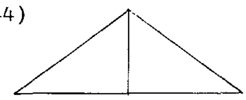
4.2.- <u>Processus non stationnaire multiple, avec erreurs d'observation</u>

Le même principe permet de résoudre le problème de prédiction, ou de filtrage, ou de lissage, d'un processus non stationnaire à composantes multiples, chaque composante étant observée avec erreur.

EXEMPLE : On considère un mouvement rectiligne dont on observe à intervalles fixes la position et la vitesse. La partie certaine du mouvement est extraite, reste un processus ainsi défini :

La vitesse est un processus gaussien centré, de variance 1, de loi de corrélation triangulaire dans la zone (-4, +4)

$C_V(t)$



-4 0 4

La distance est l'intégrale de ce processus à partir d'un instant 0.

Toutes les corrélations entre vitesses et distances sont alors calculables.

EXEMPLE : si $\overline{v = f\ (t)}$

$$\overline{V_T D_{T+\theta}} = \overline{f(T) \times \int_0^{T+\theta} f(t)dt} = \int_0^{T+\theta} \overline{f(T)f(t)}dt = \int_0^{T+\theta} C(T-t)dt$$

On veut réaliser le filtrage en un point 0, dont l'instant est 50 par rapport au début du processus, à partir des mesures de distance et de vitesses effectuées aux points 0,-1,-2,-3 (8 valeurs) entachées d'erreurs d'observation indépendantes.

On forme alors la matrice K_2 (D,V) de covariance (voir tableau 1 ci-après).

Dans la diagonale principale figurent les variances des erreurs d'observations.

Le traitement est exactement analogue au précédent et on obtient les vecteurs de filtrage :

$$\hat{D}_0 = 0{,}6796\, D_0 + 0{,}1156 D_{-1} - 0{,}00663 D_{-2} + 0{,}3346 D_{-3}$$
$$+ 0{,}1194 V_0 + 0{,}3024 V_{-1} + 0{,}3354 V_{-2} + 0{,}0892 V_{-3}$$

$$\hat{V}_0 = 0{,}0255 D_0 - 0{,}0120 D_{-1} - 0{,}0110 D_{-2} - 0{,}0022 D_{-3}$$
$$+ 0{,}5696 V_0 + 0{,}2719 V_{-1} + 0{,}0114 V_{-2} - 0{,}0407 V_{-3}$$

et la matrice de covariance $\hat{K}_2(D_0, V_0)$

$$\begin{vmatrix} 0{,}3500 & -0{,}0340 \\ -0{,}0340 & 0{,}1991 \end{vmatrix}$$

5.- <u>METHODE DE FILTRAGE RECURRENT</u>

On peut en déduire une méthode de filtrage récurrent.

Si par exemple on veut filtrer le vecteur (D_1, V_1) à partir du vecteur (D_0, V_0), le processus de filtrage, à l'instant 0, a fourni une matrice de covariance pour (D_0, V_0) c'est-à-dire, une loi de probabilité gaussienne pour ce couple de variables.

On a donc :

- la loi a priori des variables (D_0, V_0, D_1, V_1)
- la loi du couple (D_0, V_0)

THEORIE DEDUITE DE L'INEGALITE DE CRAMER-RAO :
FILTRAGE DES VECTEURS

$K_2\ (D,V) =$

D_0	D_{-1}	D_{-2}	D_{-3}	V_0	V_{-1}	V_{-2}	V_{-3}	
200,49 + 1,50	196,445	192,675	188,19	2	2,875	3,50	3,875	D_0
	196,39 + 0,75	192,345	188,575	1,125	2	2,875	3,50	D_{-1}
		192,29 + 0,90	188,245	0,50	1,125	2	2,875	D_{-2}
			188,19 + 1,10	0,125	0,50	1,125	2	D_{-3}
			*	1 +0,35	0,75	0,50	0,25	V_0
					1 + 0,10	0,75	0,50	V_{-1}
						1 + 0,05	0,75	V_{-2}
							1 + 0,40	V_{-3}

On en déduit, d'après la deuxième partie du paragraphe 1, la matrice de covariance liée de l'ensemble :

(D_0, V_0, D_1, V_1)

et on effectue le filtrage à partir de cette nouvelle matrice, qui tient compte de l'acquit des filtrages précédents.

Il importe de remarquer que le mot "récurrent" est à prendre au sens très large. On peut filtrer le vecteur à partir de n'importe quel ensemble de résultats passés.

Par exemple, pour un mouvemement de navire :

- distance à l'instant -1,
- vitesse radiales aux instants -2, -3,
- azimuts aux instants -1, -2, -3
- etc

Les concepts d'inovation et de processus MARKOVIEN étant totalement évacués.

EXEMPLE : On reprend l'exemple précédent, en appelant 0 l'instant précédemment appelé -1.

Le filtrage a fourni en 0 la matrice de covariance.

$$\begin{bmatrix} 0,5 & -0,1 \\ -0,1 & 0,2 \end{bmatrix} \text{ (distance vitesse)}$$

On construit la matrice de covariance a priori des variables (D_0, D_1, V_0, V_1).

Le but étant de filtrer le vecteur (D,V) à l'instant 1.

$K_2\ (D,V) =$

D_1	D_0	V_1	V_0	
200,49	196,445	2	2,875	D_1
	196,39	1,125	2	D_0
		1	0,75	V_1
		*	1	V_0

Calculs intermédiaires

On calcule les moyennes et les corrélations liées par des valeurs fixées de (D_0, V_0)

$\tilde{K}_1\ (D_1) = \quad 0,9912\ D_0+0,8925\ V_0 = M\ D_0+N\ V_0$

$\tilde{K}_1\ (V_1) = \quad -0,0019\ D_0+0,7339\ V_0 = M'D_0+N'V_0$

$\tilde{K}_2(D_1,D_1) = \quad 3,2013 = P$

$\tilde{K}_2(D_1,V_1) = \quad 0,2155 = Q$

$\tilde{K}_2(V_1,V_1) = \quad 0,4368 = R$

et on peut calculer alors les éléments de la nouvelle matrice de covariance.

$\tilde{\tilde{K}}_2(D_0,D_1) = M.0,5-N.0,1 = 0,4064$

$\tilde{\tilde{K}}_2(D_0,V_1) = M'.0,5-N.0,1 = -0,0763$

$\tilde{\tilde{K}}_2(V_0,D_1) = -M.0,1+N.0,2 = 0,0794$

$\tilde{\tilde{K}}_2(V_0,V_1) = -M'0,1+N'.0,2 = 0,1506$

$\tilde{\tilde{K}}_2(D_1,D_1) = P+M^2\,0,5-2MN.0,1+N^2\,0,2 = 3,6749$

$\tilde{\tilde{K}}_2(D_1,V_1) = Q+MM'0,5-(MN'+M'N)0,1+NN'0,2 = 0,2045$

$\tilde{\tilde{K}}_2(V_1,V_1) = R+M'^2\,0,5-2M'N'0,1+N'^2\,0,2 = 0,5502$

et on décrit alors la nouvelle matrice de covariance

	D_1	D_0	V_1	V_0	
	$3,6749 + \sigma^2 D$	0,4064	0,2045	0,0794	D_1
		0,5	- 0,0713	-0,1	D_0
$\tilde{\tilde{K}}_2(D,V)$		*	$0,5502 + \sigma^2 V$	0,1506	V_1
				0,2	V_0

dans laquelle, figurent ($\sigma^2 D$, $\sigma^2 V$) les erreurs d'observation au point 1 et à partir de laquelle on effectuera le filtrage.

6.- CAS DES MESURES INDIRECTES

Si on a filtré un vecteur de mesure $(X_1 \dots X_n)$, on dispose de sa matrice de covariance S. Mais on veut atteindre des paramètres $(Y_1 \dots Y_p)$. On utilise alors l'inégalité de CRAMER-RAO ainsi dans le paragraphe 3, et on définit des combinaisons linéaires optimales pour atteindre les valeurs $Y_1 \dots Y_p$, dont on a ainsi la matrice de covariance par voie de conséquence.

7.- CONCLUSION

La méthode ainsi exposée, donne la solution du traitement (lissage-filtrage-prédiction) d'un processus gaussien additif, avec des lois de corrélations quelconques. Il n'est pas nécessaire de supposer l'indépendance des variables aléatoires considérées même pas celle des erreurs d'observation. Un traitement récurrent peut être construit à partir d'un ensemble arbitraire de valeurs d'observations.

REFERENCES

[1] Topics in the theory of random noise
R.L. STRATONOVITCH WILEY 1967

[2] Detection, Estimation, and Modulation theory (Part 7)
M.L. VANTREES WILEY 1968

[3] Stockastic processus and filtring theory
JAZWINSKI ACADEMIC PRESS 1970