DETECTION DU NOMBRE DE SOURCES PAR LA METHODE DES ESPERANCES CONDITIONNELLES, CALCUL PAR NOYAUX

Denis de BRUCQ, Richard GRISEL, Christian ROBERT

Université de Rouen LACIS-ITEPEA BP 118
76134 Mont-Saint-Aignan Cedex

RESUME

Une méthode de détection du nombre de sources en sonar fondée sur la notion d'espérance conditionnelle est validée. Le calcul direct des espérances des observations conditionnellement à leur moyenne est effectuée. La démonstration montre et explique les avantages du nouvel opérateur mathématique non-linéaire découvert antérieurement. Cet opérateur s'applique dans des conditions expérimentales beaucoup plus générales que l'opérateur usuel fondé sur le rang de la matrice des covariances. L'estimation s'appuie sur un outil de statistique non-paramétrique très performant : la méthode du noyau, qui est indépendante de la loi suivie par les signaux. L'implémentation sur des <u>données réelles</u> souligne l'intérêt de cette méthode qui ne nécessite qu'un nombre relativement petit d'observations.

SUMMARY

A very important problem in sub-marine detection is to estimate the precise number of sources. Each boat generates a random pressure field. After propagation in the supposed linear random medium, the p waves plus additive noise are observed.

The number p of sources is the dimension of the vectorial space generated by the random expectations of the q observations conditionaly to their arithmetic mean. This property can be used to obtain a non linear estimator and with elementary operations the number p was obtained.

But to day, by kernel statistical technic, these conditonal expectations are estimated from simulated or experimental data and an observation time very short is needed to obtain the same precision.

1. INTRODUCTION

L'écoute passive en acoustique sous-marine, ou sonar passif, poursuit trois buts : détecter, localiser, reconnaître les sources émettrices.

Les signaux disponibles, notés Z, à la sortie des q capteurs, donnent une information vectorielle à chaque instant, sur les champs spatio-temporels de pression sous-marine.

L'aspect aléatoire des observations limite, bien sûr, les possibilités de détection. Des vecteurs gaussiens puis des vecteurs sphériquement invariants (cf C. BAKER p.154) ont été utilisés pour modéliser certains aspects des champs acoustiques sous-marins.

Le modèle retenu dans cet article (cf §2) utilise essentiellement les propriétés de <u>linéarité</u> du milieu de propagation. C'est l'hypothèse des petits mouvements en acoustique. Le modèle consiste à représenter le signal vectoriel reçu $\underline{Z}$ à l'aide des q capteurs comme une combinaison linéaire à coefficients scalaires aléatoires A des signaux vectoriels:

$\underline{X}_1, \ldots, \underline{X}_p$ issus des p sources

Les résultats obtenus s'appuient sur la notion d'espérance conditionnelle. Le nombre de sources observées par les capteurs sous des hypothèses physiques, parfaitement plausibles, est fourni.

Un algorithme élémentaire de calcul peut être utilisé (cf GRISEL) mais actuellement un procédé statistique élaboré est opérationnel (cf §5).

L'observateur ne dispose que des variables aléatoires coordonnées Z_k de l'observation vectorielle $\underline{Z}$. Dès que le nombre q de capteurs est supérieur au nombre p de sources, il est possible de déterminer le nombre p de sources à l'aide d'un résultat antérieur (cf de BRUCQ, FOLLIOT p382). La moyenne des coordonnées de $\underline{Z}$ est notée $\overline{Z}$ (cf III.2). Le calcul du rang p de la matrice des covariances des espérances conditionnelles, soit le calcul du rang de:

$$\text{(I.1)} \qquad \mathbb{E}\left(\mathbb{E}\left[\underline{Z}/\overline{Z}\right]\ \mathbb{E}\left[\underline{Z}/\overline{Z}\right]^t\right)$$

permet de trouver le nombre p de sources. Les diverses espérances conditionnelles (cf III.3) ainsi que la matrice des covariances de celles-ci peuvent être estimées de façon élémentaire par des moyennes expérimentales après seuillage et le procédé avait conduit à des résultats conforme à la théorie.

Richard GRISEL a obtenu par simulation et sur des signaux expérimentaux fournis par le laboratoire de détection sous-marine du BRUSC, la détection de sources par ce procédé. Les valeurs numériques confirmaient avec la précision attendue, le résultat énoncé.

Depuis l'estimation statistique de la dimension de l'espace vectoriel engendré par les espérances conditionnelles a fait l'objet de travaux originaux (cf §5) de Monsieur Christian ROBERT qui utilise la notion d'estimation de la densité et de la régression par la méthode des noyaux.

2.MODELE RETENU

L'expérience modélisée est la suivante:
-p sources notées S(1),..,S(j),.., S(p) dans la direction observée émettent un champ de pression acoustique
-l'observation est fournie par q capteurs: C(1),..,C(k),..,C(q).

Les ordres de grandeur sont les suivants:
- quelques dizaines de mètres pour l'antenne sur laquelle sont situés les capteurs
- quelques dizaines de kilomètres pour la distance entre chaque source et les capteurs.

La propagation acoustique en milieu marin est régie en première approximation par une équation aux dérivées partielles, linéaire. Si p ondes acoustiques sont solutions de cette équation alors toute combinaison linéaire est également solution de cette équation. Ainsi pour p solutions notées : $P_1,...,P_p$ toute combinaison linéaire:

$$(II.1) \qquad P \triangleq A_1 P_1 + \ldots + A_p P_p$$

est également solution de l'équation de propagation. Usuellement les coefficients A sont pris certains cependant le résultat reste valable si les coefficients sont aléatoires.

L'application de cet article concerne des bruits et des signaux d'origine diverse. Usuellement l'onde est considérée comme plane et issue d'une direction donnée et les bruits sont supposés gaussiens.

Nous généralisons: en plus de l'onde acoustique moyenne d'aspect sphérique et approchée bien souvent par un plan, en raison des distances relatives, s'ajoutent des fluctuations aléatoires. Une matrice Γ de covariance décrit précisemment les propriétés du second ordre de ces fluctuations. Pour chacune des p sources, nous introduisons donc la matrice Γ_j. L'hypothèse gaussienne est rajoutée maintenant puisque la distribution gaussienne est la loi probabiliste la plus naturelle en présence de petites fluctuations autour d'une moyenne.

Comme nous désirons effectuer des observations sur des durées longues afin d'éliminer les bruits par calcul de moyenne, les distances entre les sources et les récepteurs évoluent de façon aléatoire.

Or la distance relative entre une source et le récepteur est un facteur essentiel pour calculer l'énergie reçue. Pour une source j, lointaine des transducteurs, la fluctuation aléatoire de distance est la même pour tous les transducteurs. Ce phénomène physique fait apparaître un facteur unique, scalaire, aléatoire, A qui module le vecteur $\underline{X}_j$ observée.

Pour résumer les considérations antérieures décrivant les phénomènes physiques, le modèle statistique suivant est établi:

$$(II.2) \qquad Z_k = \sum_{j=1}^{p} A_j X_{j,k}$$

Nous ne faisons aucune supposition sur les propriétés statistiques des amplitudes A_j, qui sont des variables aléatoires (positives).

3.ESPERANCE CONDITIONNELLE

L'introduction effective de $\underline{s}_j$ partie certaine du vecteur gaussien $\underline{X}_j$, permet de séparer les effets aléatoires des effets déterministes.

Les matrices qxq de covariance des variables gaussiennes centrées, indépendantes $\underline{X}_1,..,\underline{X}_q$, sont définies et égales à $\Gamma_1,..,\Gamma_q$.

Le bruit blanc B gaussien, centré est de covariance le produit de la matrice identité I par un facteur certain noté σ^2.

Proposition 3.1:
Si l'observation vectorielle

$$(III.1) \qquad \underline{Z} = A_1 \, (\underline{s}_1 + \underline{X}_1) + \ldots + A_p \, (\underline{s}_p + \underline{X}_p) + \underline{B}$$

est la somme de p signaux gaussiens vectoriels:
$\underline{s}_j + \underline{X}_j$, de covariance Γ_j pondérés par une variable aléatoire scalaire aléatoire A_j et d'un bruit blanc gaussien $\underline{B}$ alors l'espérance de l'observation $\underline{Z}$ conditionnellement à la moyenne sur les capteurs:

$$(III.2) \qquad \overline{Z} \triangleq \frac{1}{q} (Z_1 + \ldots + Z_q) \qquad \text{vaut:}$$

$$(III.3) \qquad E(\underline{Z}/\overline{Z}) = \sum_{j=1}^{p} E(A_j/\overline{Z}) \, \underline{s}_j$$

$$+ \, E\left\{ \frac{\displaystyle\sum_{j=1}^{p} A_j^2 \, \Gamma_j + \sigma^2 \, I}{\displaystyle\frac{1}{q} \, \underline{1}^t \left[\sum_{j=1}^{p} A_j^2 \, \Gamma_j + \sigma^2 \, I \right] \underline{1}} / \overline{Z} \right\} \underline{1}$$

$$\left[\overline{Z} - \sum_{j=1}^{p} E(A_j/\overline{Z}) \, \underline{s}_j^t \, \frac{1}{q} \, \underline{1} \right] \text{ où } \underline{1}$$

est une colonne composée de q nombres 1.

De nombreuses situations expérimentales relèvent de cette dernière proposition. Le cas d'un signal certain dans un bruit blanc gaussien suppose que q=1, l'unique variable de pondération A_1 est certaine. De plus la variable aléatoire $\underline{X}_1$ vectorielle est nulle. Dans ce cas très particulier, l'espérance de l'observation $\underline{Z}$ conditionnellement à la moyenne $\overline{Z}$ vaut:

$$E(\underline{s}_1 + \underline{B}/\overline{Z}) = \underline{s}_1 + \underline{1} \, (\overline{Z} - \frac{1}{q} \, \underline{s}_1^t \, \underline{1})$$

L'aléatoire n'intervient que dans le terme $\overline{Z}$, facteur dans chacune des coordonnées du vecteur $\underline{1}$.

Conservons la situation d'une seule source q=1 avec le signal certain $\underline{s}_1$ multiplié par un scalaire aléatoire A_1. En ce cas, nous obtenons:

$$E(A_1\ \underline{s}_1+\underline{B}/\overline{Z}) = E(A_1/\overline{Z})\ \underline{s}_1$$

$$+E\{\frac{A_1^2\ \underline{s}_1^t\ \underline{s}_1 + \sigma^2\ I}{\frac{1}{q}\ \underline{1}^t\ [A_1^2\ \underline{s}_1^t\ \underline{s}_1+\sigma^2\ I]\ \underline{1}}/\overline{Z}\}$$

$$\underline{1}\ [\overline{Z}-E(A_1/\overline{Z})\underline{s}_1^t\ \ \frac{1}{q}\ \underline{1}]$$

Comme dernier cas particulier, prenons une seule source q=1 émettant un $\underline{X}_1$ aléatoire, gaussien, centré et de covariance Γ_1. L'espérance conditionnelle se réduit à:

$$E(A_1\ \underline{X}_1+\underline{B}/\overline{Z}) =$$

$$E\{\frac{A_1^2\ \ \Gamma_1+\sigma^2\ I}{\frac{1}{q}\ \underline{1}^t\ [A_1^2\ \ \Gamma_1+\sigma^2\ I]\ \underline{1}}/\overline{Z}\}\ \underline{1}\ \overline{Z}.$$

La situation d'un bruit blanc $\underline{B}$ négligeable conduit à la simplification du terme aléatoire de la fraction.

Le cas d'un signal aléatoire à bande étroite relève également de l'expression antérieure et la source physique S(1) en raison du caractère sinusoïdale de l'émission, doit être associée à deux sources indépendantes dans la modélisation présentée.

4. ESPACE VECTORIEL ENGENDRE:

Une démonstration directe de la Proposition 4.1 à partir de l'expression (III.3) de $E(\underline{Z}/\ \overline{Z}\ )$ peut être effectuée assez aisément. De plus l'argumentation permet de bien comprendre la généralité du résultat rappelé et de montrer que usuellement l'espace vectoriel engendrée par les coordonnées de l'observation $\underline{Z}$ conditionnellement à la moyenne $\overline{Z}$ de ces coordonnées est exactement de la dimension maximum p+1.

Proposition 4.1:

Sous les hypothèses de la Proposition 3.1 alors les espérances:

(IV.1) $\quad E(Z_1/\overline{Z}),\ldots,E(Z_q/\overline{Z})$

des coordonnées Z_j de l'observation vectorielle $\underline{Z}$ conditionnellement à la moyenne (III.2) sur les capteurs, engendrent un espace vectoriel de dimension 2p+1 au plus.

Ce résultat englobe toutes les situations expérimentales retenues à ce jour: en effet pour des covariances Γ nulles, les signaux sont certains (associés à des ondes planes ou sphériques) tout en étant multipliés par des facteurs énergétiques scalaires aléatoires de lois quelconque.

Le cas d'ondes planes certaines pour des capteurs situés dans un plan identifierait les sources situées dans une même direction. Cette situation est certainement fort exceptionnelle. Cependant cette considération théorique doit tenir compte des éventuelles imprécisions sur les mesures et le rang de la matrice estimée (I.4) pourrait être inférieure à la valeur théorique annonçée.

Le problème essentiel revient à déterminer la dimension de l'espace vectoriel engendré par les espérances (cf IV.1) des coordonnées Z_k conditionnellement à la moyenne $\overline{Z}$. La matrice de covariance (I.4) a pour rang précisémment cette dimmension. Cependant un outil statistique non trivial peut être utilisé pour estimer à partir des résultats expérimentaux les q fonctions $E(Z_k/\overline{Z})$ de $\underline{Z}$.

La loi conditionnelle à $\overline{Z}$ va être estimée par la méthode des noyaux.

5. ESTIMATION PAR LA METHODE DU NOYAU DES ESPERANCES CONDITIONNELLES

Il existe de nombreuses méthodes d'estimation de la densité d'une variable aléatoire X quelconque, la plus simple étant l'estimation par un histogramme, qui présente l'inconvénient d'approcher une fonction continue par une fonction discontinue. L'estimateur que nous utilisons ne présente pas ce défaut. Les estimateurs de la forme:

$$\text{(V.1)}\qquad f_n(x) \triangleq \frac{1}{nh_n}\sum_{m=1}^{n} K\left(\frac{x-X_m}{h_n}\right)$$

sont dits *estimateurs de la densité par le noyau*, la fonction K étant appelée *noyau* et h_n étant appelée *fenêtre de l'estimateur*.

Nous appellons *noyau d'Epanechnikov*, le cas particulier

$$\text{(V.2)}\qquad K(x) \triangleq \frac{3}{4\sqrt{5}}\ (1-\frac{x^2}{5})\ \mathbb{1}_{[-\sqrt{5};\sqrt{5}]}(x)$$

et ne considérerons que ce noyau qui, en plus de son optimalité, permet d'aboutir à des expressions analytiques plus simples des estimateurs employés.

La fenêtre que nous considérons permet d'égaler la variance empirique, s_n^2, à la variance calculée à partir de l'estimation f_n de la densité f :

Proposition 5.1:

La fenêtre optimale au sens de la variance estimée vaut:

$$\text{(V.3)}\qquad h_n^* = \sqrt{\frac{s^2}{n}}\quad \text{où}$$

$$\text{(V.4)}\qquad s_n^2 \triangleq \int_{-\infty}^{\infty} x^2 f_n(x)\ dx - \left(\int_{-\infty}^{+\infty} x f_n(x)\ dx\right)^2$$

Cette méthode du noyau s'étend de manière relativement naturelle au problème de la régression.

Nous considérons que les signaux reçus sur l'antenne s'expriment comme une suite: $Z_1(t_1),\ldots,Z_q(t_n)$ par discrétisation du temps. Après le calcul des moyennes empiriques $\overline{Z}(t_m)$ pour les différents temps t_m $(1\leqslant m\leqslant n)$, nous estimons $E(Z_j/\overline{Z}=z)$ $(1\leqslant j\leqslant q)$ par:

(V.5) $r_n^j(z) =$

$$\sum_{k=1}^{n} Z_j(t_m)\ K\left(\frac{z-\overline{Z}(t_m)}{h_n}\right)\left[\sum_{m=1}^{n} K\left(\frac{z-\overline{Z}(t_m)}{h_n}\right)\right]^{-1}$$

Comme le nombre de sources détectables est au plus (q-1), nous construisons, à partir des r_n^j $(1 \leqslant j \leqslant q)$, une matrice qxq qui donnera un minorant de p, le nombre de sources.

<u>Corollaire 5.2</u>:

Soit $I_1, \ldots, I_q$, une partition par intervalles du domaine de $\overline{Z}$. Le rang de la matrice de terme générique $E(Z_i/\overline{Z} \in I_j)$ $(1 \leqslant i,j \leqslant q)$ est au plus égal au nombre p de sources.

Pour estimer les bornes des intervalles I_j et les valeurs de $E(Z_i/\overline{Z} \in I_j)$, nous utilisons le noyau d'Epanechnikov (V.2) et la fenêtre (V.3). Le choix d'une même fenêtre pour l'estimation de la densité et de la régression permet une expression analytique simple des estimateurs obtenus. Nous avons:

<u>**Proposition 5.3:**</u>

L'estimateur F_n du noyau de la fonction F de répartition de $\overline{Z}$, associé au noyau K et à la fenêtre h_n^*, vaut:

$$F_n(x) = \frac{1}{n}\sum_{m=1}^{n} I\left(\frac{x-\overline{Z}(t_m)}{h_n^*}\right) \quad \text{où}$$

$$I(t) \triangleq 0 \text{ si } t \leqslant -\sqrt{5}$$

$$I(t) \triangleq \frac{3}{4\sqrt{5}}\left\{\frac{2}{3}\sqrt{5} + t - \frac{1}{15}t^3\right\} \text{ si } t \in [-\sqrt{5}, \sqrt{5}\,]$$

$$I(t) \triangleq 1 \text{ si } t \geqslant \sqrt{5}$$

Ce résultat nous permet d'estimer les intervalles $I_j = [b_j ; b_{j+1}]$ $(1 \leqslant j \leqslant q-1)$; en effet, nous estimons $b_j \triangleq F^{-1}(\frac{j}{q})$ par $\hat{b}_j \triangleq F_n^{-1}(\frac{q}{n})$ à l'aide d'un algorithme simple d'inversion de F_n. Enfin nous estimons $E(Z_i/\overline{Z} \in I_j)$ par: $\forall i,j \in [1,q]$

$$a_{i,j} \triangleq \int_{\hat{b}_{j-1}}^{\hat{b}_j} r_n^i(z)\ f_n(z)\ dz$$

$$= \frac{1}{n}\sum_{m=1}^{n} Z_i(t_m)\left[I\left(\frac{\hat{b}_j - \overline{Z}(t_m)}{h_n^*}\right) - I\left(\frac{\hat{b}_{j-1} - \overline{Z}(t_m)}{h_n^*}\right)\right]$$

A l'aide du corollaire 5.2, nous avons donc le nombre p de sources qui est donné par le rang p de la matrice de terme générique: à

<u>CONCLUSION:</u>

Le modèle choisi généralise les situations: signaux certains, signaux gaussiens, signaux sphériquement invariants plus bruit gaussien. La possibilité de retrouver le nombre de sources dans des situations aussi générales apparaît comme surprenante et la démonstration théorique obtenue en 1981 (cf de Brucq, Folliot) ne pouvait satisfaire l'ingénieur. Aussi de nombreuses simulations ont été effectuées.

En raison de la simplicité des opérations électroniques à mettre en oeuvre pour obtenir le nombre de sources par ce procédé, la réalisation de circuits spécialisés avait été envisagée. D'ailleurs il faut remarquer qu'un tel opérateur serait positionné en série après l'opérateur de voies préformées et l'opérateur de transformée de Fourier rapide. Cependant les impératifs de fiabilité pour la marine, dépassaient les possibilités techniques du laboratoire.

Rappelons que détecter un nouvel objet consiste au cours d'une veille où les bruits ambiants ont été répertoriés et associés à p+1 sources à constater le plus rapidement possible qu'un nouveau phénomène apparaît. La répartition spatio-temporelle de l'énergie de ce nouveau phénomène doit être distinguée le plus rapidement possible des sources déjà répertoriées.

L'obtention d'algorithmes performants permettant d'obtenir le nombre de sources à partir du plus faible nombre possible d'observations justifie l'algorithme statistique très élaboré du §5.

En conclusion la méthode statistique d'estimation de la dimmension de l'espace vectoriel engendré par les variables aléatoires:

$$E(Z_1/\overline{Z}), \ldots, E(Z_q/\overline{Z})$$

à l'aide de la matrice des covariances des espérances conditionnelles est simple mais le travail présenté dans la seconde partie montre qu'avec la même précision, là où il fallait 10000 valeurs numériques, il suffit de 200 valeurs numériques.

<u>**BIBLIOGRAPHIE.**</u>

Baker (1986) <u>Stochastic Processes in Underwater Acoustic</u>. n°85. Springer Verlag.

Bienvenu (1975) <u>Détection passive de sources ponctuelles dans un bruit isotrope</u>, cinquième colloque GRETSI, Nice, pp. 60/1-60/7.

——— (1979) <u>Principe du goniomètre passif adaptatif</u>, septième colloque GRETSI, Nice, pp. 106/1-106/10.

——— (1983) <u>Propriétés haute-résolution de la matrice de corrélation spatiale</u>, neuvième colloque GRETSI, Nice, pp. 239.

Bienvenu, Kopp (1985) <u>Méthodes haute-résolution après formation de voie</u>, dixième colloque GRETSI, Nice, pp. 325-330.

de Brucq (1981) <u>Mélange de lois statistiques sphériquement invariantes</u>, huitième colloque GRETSI, Nice, pp. 13-17.

de Brucq (1988) <u>Théorie du signal</u>, modélisations statistique,automatique et traitement , (Masson).

Epanechnikov (1969) <u>Theory of Probability and Applications</u> 14, pp.153-158.

Folliot (1980) <u>Sur les applications des structures convexes</u>, Thèse de troisième cycle. Rouen.

Grisel (1987) Détectabilité du nombre de sources en sonar. <u>Thèse de troisième cycle</u>. Rouen.