MODEL ORDER SELECTION IN UNKNOWN CORRELATED NOISE : A SUPERVISED APPROACH

P.Costa, J. Grouffaud, P. Larzabal and H. Clergeot

LESIR-ENS Cachan, URA CNRS D 1375,
61, av. du Pdt Wilson, 94235 CACHAN cedex France
☎ +33 1 47 40 27 09 pascale.costa@lesir.ens-cachan.fr

ABSTRACT

The purpose of this paper is to propose the design and the use of a Neural Network for model order selection. The proposed neural network learns from real life situation by constructing an imput/output mapping (for detection) which brings to mind the notion of non parametric statistical inference. Such a strategy can improve performances of traditional tests relying on linearity, stationarity and second order statistics. We focus on the case where the noise covariance matrix is unknown but is a band matrix. This paper includes simulations which show improvements obtained by supervised approach.

1. INTRODUCTION

Recently, the new class of so-called subspace methods for high resolution parametric estimation (spectral analysis or direction finding for example) has received a great deal of attention in the literature. These high resolution methods require the model order knowledge. This is a difficult task, and classical tests suffer from following limitations :

❶ a restrictive a priori hypothesis : a noise whiteness;

❷ they are efficient only in asymptotic condition;

❸ they involve an eigen-decomposition, which does not allow a real time implementation.

The main contribution of this paper is an alternative when the noise covariance matrix is unknown but is, for simulation convenience only, a band matrix. In sensor array processing, this amounts to assuming that the noise field is locally spatially correlated [1]. The robust proposed detection test will be based on a supervised approach.

In this paper, we shall focus on a Uniform Linear Antenna, in the detection case of 0, 1 or 2 sources. The number of sensors (N) and the number of snapshots (T) used to estimate the covariance matrix $\mathbf{R}_x$ are small. This practical situation is of a great interest.

2. PROBLEM FORMULATION

Antenna signal model

The conventional model of array processing is used. We consider an Uniform Linear Array (ULA) of N sensors situated in the wavefield generated by M narrow-band point sources. The distance between two adjoining sensors is an half wavelength. Then, if $\mathbf{x}$ is the observation vector (N,1), $\mathbf{s}$ the emitted vector signal (M,1) and $\mathbf{n}$ an additive noise (N,1), we have:

$$\mathbf{x}(t) = \mathbf{A}(\theta).\mathbf{s}(t) + \mathbf{n}(t) = \mathbf{y}(t) + \mathbf{n}(t) \qquad (1)$$

where the columns of $\mathbf{A}$, termed matrix of steering vectors, are the parametric response of the array of sensors to an emitter impinging from the direction θ.

The signals $\mathbf{s}$ are assumed statistically independent of the noise $\mathbf{n}$. So, from equation (1), the observation covariance matrix $\mathbf{R}_x$ can be expressed:

$$\mathbf{R}_x = \mathbf{R}_y + \mathbf{R}_n = \mathbf{A}.\mathbf{R}_s.\mathbf{A}^H + \mathbf{R}_n \qquad (2)$$

White Noise Case

If we suppose that noise components are uncorrelated from an array element to every others with equal variance σ^2, we have:

$$\mathbf{R}_x = \mathbf{R}_y + \sigma^2.\mathbf{I}_N \qquad (3)$$

where $\mathbf{I}_N$ is the N-dimensional identity matrix.

Asymptotically, the number of non zero eigenvalues of $\mathbf{R}_y$ is equal to the number of sources M, (N-M) eigenvalues being zeros. So, according to (3), $\mathbf{R}_x$ has the same eigenvectors as $\mathbf{R}_y$, with eigenvalues $\lambda_x = \lambda_y + \sigma^2$. For $\mathbf{R}_x$, σ^2 is a degenerated order (N-M) eigenvalue.

These observations are the basis of most detection schemes but are only valid for asymptotic assumption and a white noise model.

Spatially Correlated Noise

We suppose that the noise is spatially correlated over the array. We consider, for shake of simplicity, the case where the first row of noise band matrix covariance is defined by:

$$\mathbf{R}_n(k,1) = \sigma^2.\exp\left(\frac{-(k-1)}{L_c}\right) \qquad k = 1 \text{ to } N \qquad (4)$$

This model has been used for simulation convenience and clarity of presentation. It is not connected to the proposed method. Let us name L_C the correlation length:

$$\lim_{L_c \to 0} \mathbf{R}_n = \sigma^2.\mathbf{I}_n \quad \text{(white noise case)} \qquad (5)$$

The term L_C permits to appreciate the difficulty of the problem.

In underwater accoustic, this correlation is usually modelized by a Bessel Function [1].

Mesure of performance

If M is the number of sources and $\hat{M}$ its estimated value, performances are mesured by detection probability and by false alarm probability:

$$\mathrm{Pd} = \mathrm{Prob}\left[\hat{M} = M\right] \quad \mathrm{fa} = \mathrm{Prob}\left[\hat{M} > M\right] \tag{6}$$

3. CLASSICAL TESTS

Principle of statistical tests

Classical methods of statistical signal processing are founded on three basic assumptions: linearity, stationarity and second order statistics.

Model order selection is usually done from statistical tests. These tests are based either on information theory (Akaike Information Criterion (AIC) and Minimum Description Length (MDL)) [2] or on decision theory (Khi 2) [3].

The former involve the log-likelihood of the observations, expressed in terms of noise eigenvalues. The estimated order is merely the one for which the log-likelihood added to a penalty function is maximum.

The Khi-2 test relies on a recursive comparison of the generalized likelihood ratio (which is asymptotically Khi-2 distributed) to thresholds computed a priori. These ones are directly connected to the false alarm chosen. This test has been modified [3] for small size of snapshots (non asymptotic case).

These methods detect the number of "nonzero" eigenvalues with the well-known problems which occur in case of strong correlation or non equipowered sources. The performances of these methods decrease drastically when uncorrelated noise hypothesis is unaccurate.

Simulations

Let's consider a five sensors array. Covariance matrix is calculated with ten snapshots. The thresholds of the Khi-2 test have been fixed a priori for a fa of 1% when $L_c \to 0$. The first curve shows the expected value of estimated order when two sources impinge on the array, while the second curve concerns the false alarm (noise only case).

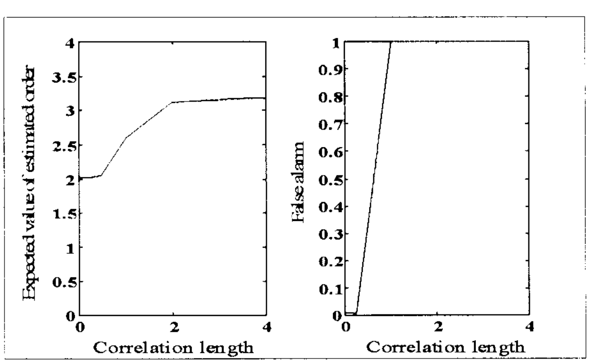

Figure 1 : Degradation of performances of Khi-2 test along the noise correlation length expressed in half wavelength.

When noise is uncorrelated, detection is accurate since we find $f_a = 1\%$. As early as correlation length becomes larger than the distance between two sensors (an half wavelength), the order of the model is overestimated and the false alarm tends to 1.

Moreover, in real life situation, we have to work with a finite sample size, irrespective to the statistical estimation procedure used. Some improvements can be gained by a supervised approach, as described in the next section.

4. SUPERVISED DETECTION

Presentation of the network

A Multilayer Perceptron with one hidden layer is used for the supervised detection (fig. 2) [4].

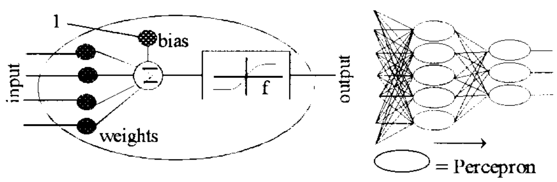

Figure 2: Perceptron and Multilayer Perceptron with one hidden layer.

If $\mathbf{X}$ is the input matrix, the outputs $\mathbf{Y}$ of the network are given by:

$$\mathbf{Y} = \mathrm{f}\left(\mathbf{W}2.\mathrm{f}\left(\mathbf{W}1.\mathbf{X} + \mathbf{b}1.\mathbf{u}^t\right) + \mathbf{b}2.\mathbf{u}^t\right) \tag{7}$$

where f is a sigmoid function (the hyperbolic tangent function in this case). The first layer is defined by the weights matrix $\mathbf{W}1$ and the biases vector $\mathbf{b}1$, and the second by $\mathbf{W}2$ and $\mathbf{b}2$. $\mathbf{u}$ is an unity vector of the same length as the training base.

Input Coding: feature extraction of R_x

In order to avoid huge training on all the correlation matrix for different scenarii, some discriminant characteristics of $\mathbf{R}_x$ will be retained. In the case of ULA with 0, 1 or 2 sources, the correlation matrix has particular properties. In fact, evolutions of modulus along sub diagonals give informations about the number of sources [5].

In this issue, according to this observation, the inputs of the neural network will be the means and standard deviations of diagonal and sub diagonals of $\mathbf{R}_x$. If N is the number of sensors, the size of $\mathbf{R}_x$ is (N,N), so the neural network requires only 2(N-1) real inputs.

The network's inputs are prescaled: a normalization avoids the saturations of neurons and the mean suppression allows to work in the linear zone of sigmoids at initialization.

Output Coding

The number of ouput neurons is the maximum number of sources, with the coding described on figure 3. After training, the number of sources is obtained by the use of a threshold on the output (fig. 4). This threshold is obtained according to the accepted false alarm probability (fa). In this work, fa=1% for $L_c \rightarrow 0$.

For training, the outputs are coded according to the exact number of sources, with level (-0.7,+0.7) instead of (-1,+1) to avoid sigmoid saturation.

	0 source	1 source	2 sources	3 sources
Output Neuron 1	-1 α_1	+1	+1	+1
Output Neuron 2	-1	-1 α_2	+1	+1
Output Neuron 3	-1	-1	-1 α_3	+1

Figure 3: Coding of number of sources (in case of M=3).

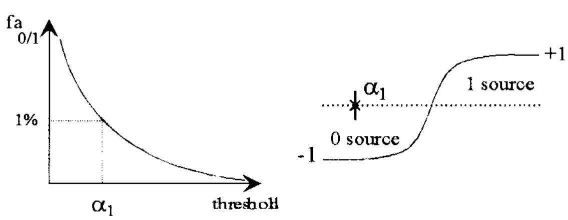

Figure 4: The threshold α on the output of neuron is given for a false alarm probability.

Training base

In operational context, training base is obtained with real life data measured under cooperative scenario.

For the simulations, the supervised learning base is obtained when the parameters (number of sources, SNR, directions θ and noise correlation length L_C) describe randomly and uniformly the domain of variation.

Training: Backpropagation

According to the previous discussion, the learning synoptic is the following:

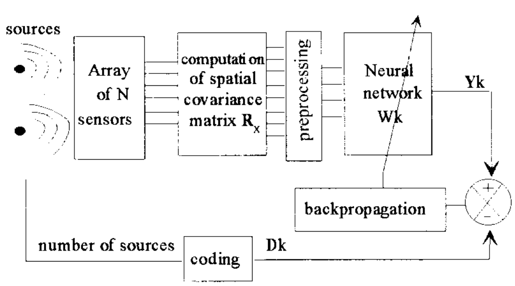

Figure 5: Learning Synoptic.

During the learning phase, weigths and biases defined by (7) move to minimize the Mean Square Error (MSE) between the desired output **Dk** and the network output **Yk**. We have at iteration k:

$$\xi_k = \left\| \mathbf{D}_k - \mathbf{Y}_k \right\|^2 \qquad (8)$$

If **w** is the weight vector which countains all the elements of **Wi** and **bi**, **w** move according to the relation [6]:

$$\mathbf{w}_{k+1} = \mathbf{w}_k + \mu_k . \mathbf{d}_k \qquad (9)$$

The most classical algorithms use a first order development of the MSE. So, the search direction $\mathbf{d}_k$ is given by the opposed of the gradient $\mathbf{g}_k$. This algorithm is slow to converge due to the nature of error surfaces. One way to guarantee a faster and more efficient convergence is to use higher order derivatives. So, we use the conjugate gradient. In fact, this method allows us to combine efficiently the steepest descent simplicity and the Newton performances: curvature of the error function is taken into account without requiring Hessian estimation and storage. The search direction is given by:

$$\mathbf{d}_{k+1} = -\mathbf{g}_{k+1} + \beta_k . \mathbf{d}_k \text{ with } \beta_k = \frac{\mathbf{g}_{k+1}^t . (\mathbf{g}_{k+1} - \mathbf{g}_k)}{\mathbf{g}_k^t . \mathbf{g}_k} \qquad (10)$$

5. SIMULATION RESULTS

In these simulations, sources are supposed in the main beam of the array pattern. This is the most difficult task of detection.

The simulations are for a Uniform Linear Antenna with 5 sensors (N=5). A surestimation of M can not be accepted here because the dimension of noise subspace will decrease drastically, increasing variance on parameters estimation: identification algorithms (MUSIC) can lack of resolution. So we have necessary to estimate M with a good accuracy.

White Noise Case

The figure 6 concerns the case of 1 source. The most used tests (AIC and MDL) are not designed for a given false alarm probability. This value can only be a posteriori computed (with no control). Notice that in this figure, the false alarm probabilities of AIC and MDL are about 20%. We show the advantages of Khi2 and Neural Network (NN) tests for which the false alarm probability is fixed a priori to 1 %. We compare the results obtained for these two tests and we observe the superiority of NN on Khi2.

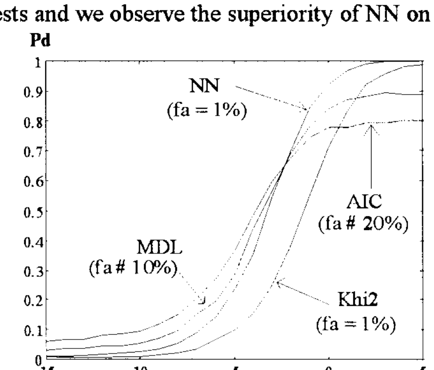

Figure 6: Comparison of the different tests (1 source, T=10 fa=1% for NN and Khi-2).

The figure 7 concerns the cases of 2 sources strongly correlated. Of course, results depend on the learning base: the results are better when the training base is important (NN2 respect NN1) and describes randomly and uniformly a large useful domain of parameters. Simulations make clear that supervised detection tests are the best.

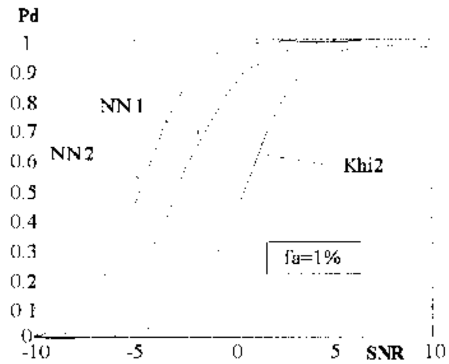

Figure 7: Comparison of the different tests
(2 equipowered and correlated sources, T=10).

Although these results are impressive, a good compromise must be found between training and generalization.

Spatially Correlated Noise

Simulations have been done when noise is spatially correlated. We have first studied influence of noise correlation length on false alarm for the supervised approach (figure 8). Although false alarm increases with correlation length, it takes reasonable values (less than 20% until $L_c = 3$) compared with thoses taken for Khi-2 test (see figure 1).

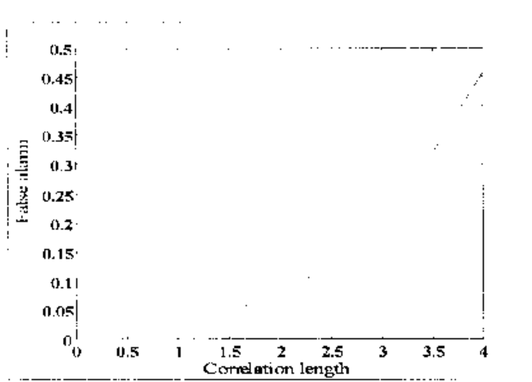

Figure 8 : Degradation of performances of supervised test along the noise correlation length.

Figure 9 shows the ROC curves of supervised detection and Khi-2 test, in correlated noise with $L_c = 2$. They are computed in the particular case of decision between two hypothesis, that is one source impinge on the array.

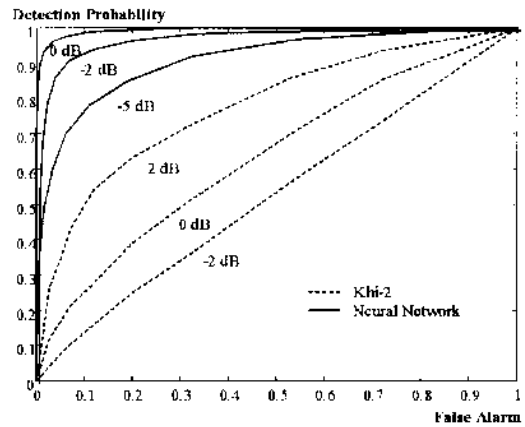

Figure 9: ROC curves of supervised test and Khi-2.

The training base is only composed of SNR equal to -15,-10,-5,0,5 dB, so the curve obtained for -2 dB proves the generalization quality of the network.

6. CONCLUSION

A new approach to the detection of the number of signals has been proposed. It allows to handle situations where the noise field is spatially correlated. Unlike almost existing tests, it does not assume the noise covariance matrix to be known. Simulations confirm expected performances in small sample size. Moreover, the proposed test does not require eigen-decomposition.

REFERENCES

[1] J.J. Fuchs, "Estimation of the Number of Signals in the Presence of Unknown Correlated Sensor Noise", *IEEE Trans. on Signal Processing*, vol. 40, N°5, May 1992.

[2] M. Wax, T. Kailath, "Detection of the number of signals by information theoretic criterion", *IEEE Trans. Acoust., Speech, Signal Processing*, vol 33, pp 387-392, 1984.

[3] A.T. James, "Test of equality of latent roots of the covariance matrix", in *Multivariate Analysis*, P. R. Krishnaiah Ed. New York: Academic, pp 205-208, 1969.

[4] D.R. Hush, B.H. Horne, "Progress in Supervised Neural Networks", *IEEE Signal Processing Magazine*, January 1993.

[5] P. Costa-Hirschauer, J. Grouffaud, P. Larzabal and H. Clergeot, "A Neural Network for model order selection in Signal Processing" *Proc. of IEEE International Conference on Neural Network*, Perth, Australia, 29 Nov-1Dec 1995.

[6] B. Widrow, M.A. Lehr, "30 Years of adaptative Neural Network : Perceptron, Madaline and Backpropagation", *Proc. of the IEEE*, vol. 78, N°9, Sept. 1990.