Proceedings of the 5th WSEAS Int. Conf. on Signal Processing, Robotics and Automation, Madrid, Spain, February 15-17, 2006 (pp341-344)

Detection by the stochastic classification matched filter

JUENNARD Nicolas, BORLOZ Bruno, XERRI Bernard
Laboratoire STIC / GESSY
Université du Sud Toulon-Var BP 56
83162 La Valette du Var Cedex
FRANCE

Abstract : This paper deals with stochastic zero-mean signals detection. Second order statistical moments of signals is assumed to be known. We propose an adaptated method based on a linear filtering under constraint optimizing a criterion based on different signal to noise ratios. This optimisation leads to the choice of a 1D subspace on which the different signals are projected. We also present simulations and results obtained with different signals.

Key words : detection, classification, matched filter, subspace projection, signal to noise ratio (SNR).

1 Introduction

This paper deals with a signal classification method based on projection onto a specific subspace. All signals are supposed to be zero mean and the perturbation is additive. We suppose only known their second-order statistics only (*a priori* or experimentally estimated). The following method consists of a filtering under constraint which optimizes the Signal to Noise Ratio (SNR) in a 1D subspace. The performances of this method will be evaluated by using the classical Receiver Operating Characteristics curves (COR curves).

This document focuses on the classical signal theory problem: the detection of a specific process embedded in additive noise [4]. Most of the time are considered a signal s and a noise b, and we need to decide whether s is present or not. The usually used model is an additive superposition of a deterministic signal with a random noise. In this case, if probability density functions are known, a method like likelihood ratio [7] [5] [1] can be used. If only the second order statistics of the signals are known, the matched filter or derived methods [8] [2] can be used. When more than one noise is present, the classical approach is to consider for noise the mean of all different disturbing signals and use the classical method.
In the following, both signal and noise(s) are assumed to be zero mean realisations of independant random functions. The only information available about these signals is their second order statistics. All probabilities density functions are unknown.
The method developped in this paper is based on maximization of an energy based criterion. The interest signal s can be disturbed by one of different noises noted $\mathbf{b}_i, 1 \leq i \leq N_b$. In the case of a unique noise, the classical criterion is the SNR, but another criterion needs to be found when more that one noise can perturb the signal. This criterion will be based on SNR of each noise noted $\rho_i, 1 \leq i \leq N_b$ and in the following, the geometrical mean of each SNR : $\rho = \sqrt[N_b]{\rho_1 \rho_1 \dots \rho_{N_b}}$ will be considered.
We will determinate the linear filter which maximizes this criterion. Filtering is the same action than projecting on a 1D subspace, spanned by the filter vector.
The output magnitudes of the filtered signal are compared to a threshold and allow us to decide whether the signal is present or not.
This method is called "Stochastic classification by matched filtering".

2 Notations and classical method

The following notations will be used:

$\mathbf{A}$	:	signal covariance matrix
$\mathbf{B}_i$	:	i^{th} noise covariance matrix
$\mathbf{M}^{-1}$	:	inverse of non singular matrix $\mathbf{M}$
$tr(\mathbf{M})$	:	trace of $\mathbf{M}$
$\mathbf{I}_N$	:	N by N identity matrix
E_N	:	N dimension space
δ_{ij}	:	Kronecker symbol
SNR	:	Signal to noise ratio

All signals are sampled, represented by a N dimension vector, and zero mean. All the probability density functions are unknown and the only available information is the second-order statistics, via the covariance matrices.

The classical approach, when a signal can be disturbed by more that one noise, is to consider a single perturbing process which is the mean of all different noises. To show improvement due to the method described in this paper, it will be compared to the Stochastic Matched Filter (SMF) method.

2.1 Stochastic Matched Filter (SMF)

the SMF theory has been developpped in the specific case of a unique signal and a unique noise. The observation $\mathbf{x}$ follows one of the two hypothesis:

H_0 : $\mathbf{x}$ = noise alone
H_1 : $\mathbf{x}$ = signal + noise

The aim of this method is to find the linear filter $\mathbf{h}$ maximizing the SNR. If $\mathbf{A}$ and $\mathbf{B}$ are respectively the signal and the noise covariance matrices, $\mathbf{h}$ is a vector matched to signal and unmatched to noise.
This SNR after filtering can be written:

$$RSB = \rho = \frac{\mathbf{h}^T\mathbf{A}\mathbf{h}}{\mathbf{h}^T\mathbf{B}\mathbf{h}}$$

If normalized matrices are used ($tr(\mathbf{A}) = tr(\mathbf{B}) = 1$), ρ can be seen as a gain on the initial SNR.
If $\mathbf{B}$ is singular, the solution of this maximization problem is obtained by solving the generalized eigenvalue equation:

$$\mathbf{B}^{-1}\mathbf{A}\mathbf{h} = \lambda\mathbf{h}$$

The optimal filter $\mathbf{h}$ is the eigenvector associated to the largest eigenvalue, and this eigenvalue is the gain on the SNR.
Practically, the matrix $\mathbf{A}$ has to be estimated, and as signal and noise are supposed to be uncorrelated, we can only access to $\mathbf{A}+\mathbf{B}$. The last equation becomes $\mathbf{B}^{-1}(\mathbf{A}+\mathbf{B}) = \mathbf{B}^{-1}\mathbf{A} + \mathbf{I}_N$. The corresponding eigenvectors are the same than previously and the eigenvalues are $\lambda_i + 1$. Thus, the method is still working.

3 Stochastic classification filter

3.1 Theory

A more general case is now considered: the signal $\mathbf{s}$ can be perturbated by one of N_b different noises. All noises and signal are supposed to be uncorrelated.
We have now $2N_b$ different hypothesis:
H_{0i} : $\mathbf{x}$ = noise b_i alone
H_{1i} : $\mathbf{x} = \mathbf{s} + \delta_i \;.\; \mathbf{b}_i$

For a noise i considered, the first hypothesis is to consider the noise alone and the second is to consider the additive mixture of signal and noise. So, N noises correspond to $2N$ different hypothesis.
We are looking for a filter $\mathbf{h}$ which maximizes a specific energy based criterion. For each noise, the SNR after filtering can be written:

$$\rho_i = \frac{\mathbf{h}^T\mathbf{A}\mathbf{h}}{\mathbf{h}^T\mathbf{B}_i\mathbf{h}}$$

The criterion to optimize is the geometric mean of the ρ_i:

$$\rho_{geo} = \sqrt[N_b]{\rho_1 \cdots \rho_{N_b}}$$

The filter $\mathbf{h}$ optimising ρ is matched to the signal and unmatched to all $\mathbf{b}_i$.

To simplify calculations, the term $\rho = \prod_{i=1}^{N_b} \rho_i$ is maximized instead of $\sqrt[N_b]{\prod_{i=1}^{N_b} \rho_i}$ (because all ρ_i are positive). The following notations are used:

$$\rho = \prod_{i=1}^{N_b} \rho_i = \prod_{i=1}^{N_b} \frac{\mathbf{h}^T\mathbf{A}\mathbf{h}}{\mathbf{h}^T\mathbf{B}_i\mathbf{h}}$$

The derivative of this product is:

$$\frac{\partial\rho}{\partial\mathbf{h}} = \sum_{i=1}^{N_b} \frac{\partial\rho_i}{\partial\mathbf{h}} \prod_{j=1\neq i}^{N_b} \rho_j$$

We have

$$\frac{\partial\rho_i}{\partial\mathbf{h}} = \frac{2\left(\mathbf{A} - \rho_i\mathbf{B}_i\right)\mathbf{h}}{\mathbf{h}^T\mathbf{B}_i\mathbf{h}}$$

so

$$\sum_{i=1}^{N_b} \frac{2\left(\mathbf{A} - \rho_i\mathbf{B}_i\right)\mathbf{h}}{\mathbf{h}^T\mathbf{B}_i\mathbf{h}} \prod_{j=1\neq i}^{N_b} \rho_j = 0$$

The term $\mathbf{h}^T\mathbf{B}_i\mathbf{h}$ cannot be null, so:

$$\sum_{i=1}^{N_b} 2\left(\mathbf{A} - \rho_i\mathbf{B}_i\right)\mathbf{h} \prod_{j=1}^{N_b} \rho_j = 0$$

We definied $\rho = \prod_{j=1}^{N_b} \rho_j \neq 0$, so we finally find:

$$\sum_{i=1}^{N_b} \left(\mathbf{A} - \rho_i\mathbf{B}_i\right)\mathbf{h} = 0 \qquad (1)$$

and so

$$\mathbf{A}\mathbf{h} = \frac{1}{N_b}\sum_{i=1}^{N_b} \rho_i\mathbf{B}_i\mathbf{h} \qquad (2)$$

The filter $\mathbf{h}$ we are looking for must satisfy the equation (2) and is the vector among all associated to the largest value of ρ.

3.2 Algorithm

Now, the equation (2) must be solved. To go back to a classical eigenvalue equation, (2) can be written:

$$\left(\mathbf{A} - \frac{1}{N_b}\sum_{i=2}^{N_b} \rho_i\mathbf{B}_i\right)\mathbf{h} = \frac{1}{N_b}\rho_1\mathbf{B}_1\mathbf{h} \qquad (3)$$

The initial value of $\beta^{(0)}$ is:

$$\beta^{(0)} = (\rho_2, \ldots, \rho_{N_b})$$

Proceedings of the 5th WSEAS Int. Conf. on Signal Processing, Robotics and Automation, Madrid, Spain, February 15-17, 2006 (pp341-344)

We can take for initial values of $\rho_i, i > 2$, the largest eigenvalues of each $\mathbf{B}_i^{-1}\mathbf{A}$ matrix.
$\mathbf{h}^{(0)}$ is the eigenvector associated to the largest eigenvalue $\alpha^{(0)}$. The values of $\alpha^{(0)}$ and $\mathbf{h}^{(0)}$ are evaluated with the equation (3). The ρ_i must verify that $\prod_{i=1}^{N_b} \rho_i$ is maximum.

$\beta^{(1)}$ is estimated thanks to the vector $\mathbf{h}^{(0)}$:

$$\beta^{(1)} = \left(\frac{\mathbf{h}^{(0)T}\mathbf{A}\mathbf{h}^{(0)}}{\mathbf{h}^{(0)T}\mathbf{B}_1\mathbf{h}^{(0)}}, \ldots, \frac{\mathbf{h}^{(0)T}\mathbf{A}\mathbf{h}^{(0)}}{\mathbf{h}^{(0)T}\mathbf{B}_{N_b}\mathbf{h}^{(0)}} \right)$$

and the same operations are iterated until the value of $\Delta\rho = \rho^{(n+1)} - \rho^{(n)} < \epsilon$ with:

$$\rho^{(n)} = \prod_{i=1}^{N_b} \frac{\mathbf{h}^{(n)T}\mathbf{A}\mathbf{h}^{(n)}}{\mathbf{h}^{(n)T}\mathbf{B}_i\mathbf{h}^{(n)}}$$

3.3 Convergence

Let us note $\rho = \prod_{i=1}^{N_b} \alpha_i$, we can write for a little variation:

$$\Delta\rho = \sum_{i=1}^{N_b} \Delta\alpha_i \prod_{j=1, j\neq i}^{N_b} \alpha_j = \sum_{i=1}^{N_b} \rho \frac{\Delta\alpha_i}{\alpha_i}$$

The term $\frac{\Delta\alpha_i}{\alpha_i}$ has to be calculated. At the $n+2^{th}$ iteration, we can write:

$$\begin{aligned} \alpha_i^{(n+2)} &= \frac{\mathbf{h}^{(n+1)T}\mathbf{A}\mathbf{h}^{(n+1)}}{\mathbf{h}^{(n+1)T}\mathbf{B}_i\mathbf{h}^{(n+1)}} \\ &= \frac{\left(\mathbf{h}^{(n)} + \Delta\mathbf{h}\right)^T \mathbf{A} \left(\mathbf{h}^{(n)} + \Delta\mathbf{h}\right)}{\left(\mathbf{h}^{(n)} + \Delta\mathbf{h}\right)^T \mathbf{B}_i \left(\mathbf{h}^{(n)} + \Delta\mathbf{h}\right)} \end{aligned}$$

$$\alpha_i^{(n+2)} = \frac{\mathbf{h}^{(n)T}\mathbf{A}\mathbf{h}^{(n)}}{\mathbf{h}^{(n)T}\mathbf{B}_i\mathbf{h}^{(n)}} \frac{1 + 2\frac{\Delta\mathbf{h}^T\mathbf{A}\mathbf{h}^{(n)}}{\mathbf{h}^{(n)T}\mathbf{A}\mathbf{h}^{(n)}} + \frac{\Delta\mathbf{h}^T\mathbf{A}\Delta\mathbf{h}}{\mathbf{h}^{(n)T}\mathbf{A}\mathbf{h}^{(n)}}}{1 + 2\frac{\Delta\mathbf{h}^T\mathbf{B}_i\mathbf{h}^{(n)}}{\mathbf{h}^{(n)T}\mathbf{B}_i\mathbf{h}^{(n)}} + \frac{\Delta\mathbf{h}^T\mathbf{B}_i\Delta\mathbf{h}}{\mathbf{h}^{(n)T}\mathbf{B}_i\mathbf{h}^{(n)}}}$$

If we consider that the variation $\Delta\mathbf{h}$ is small, we can make a first order Taylor approximation and we find:

$$\alpha_i^{(n+2)} \approx \alpha_i^{(n+1)} \left(1 + 2\left(\frac{\Delta\mathbf{h}^T\mathbf{A}\mathbf{h}^{(n)}}{\mathbf{h}^{(n)T}\mathbf{A}\mathbf{h}^{(n)}} - \frac{\Delta\mathbf{h}^T\mathbf{B}_i\mathbf{h}^{(n)}}{\mathbf{h}^{(n)T}\mathbf{B}_i\mathbf{h}^{(n)}} \right) \right)$$

and we can write:

$$\frac{\alpha_i^{(n+2)} - \alpha_i^{(n+1)}}{\alpha_i^{(n+1)}} = \frac{\Delta\alpha_i^{(n+1)}}{\alpha_i^{(n+1)}}$$

$$\frac{\Delta\alpha_i^{(n+1)}}{\alpha_i^{(n+1)}} = \frac{2}{\mathbf{h}^{(n)T}\mathbf{A}\mathbf{h}^{(n)}} \left(\Delta\mathbf{h}^T\mathbf{A}\mathbf{h}^{(n)T} - \alpha_i^{n+1}\Delta\mathbf{h}^T\mathbf{B}_i\mathbf{h}^{(n)} \right)$$

but α_i checks the equation (1), and so we deduce that:

$$\frac{\Delta\alpha_i^{(n+1)}}{\alpha_i^{(n+1)}} = \frac{\Delta\rho}{\rho} = 0$$

Convergence is assured and the unicity of the solution for ρ is evident , but note that different subspaces can exist for a same ρ value.

4 Experimental results

To illustrate the interest of the method proposed, 5 stationnary signals are taken into consideration. The first signal is said to be the interesting signal and the 4 others are noises. First, the classical SMF method is used. In this case, we consider for noise the mean of all 4 previous noises. In a second time, we use the method described in this paper. Classical COR curves illustrate the improvements brought by the method.
All signals have the same power and are 88 000 points long (5500 realizations of 16 samples). Covariance matrices are evaluated on 800 realizations and their dimension is 16×16 samples. Figure (1) shows the FFT modulus of the different signals.
Figures (2) and (3) show respectively the output signals after filtering with the SMF and CSMF filters. The 5500 first samples represent the first signal filtered, the 5500 next samples represent the second signal filtered, and so on up to the last signal.

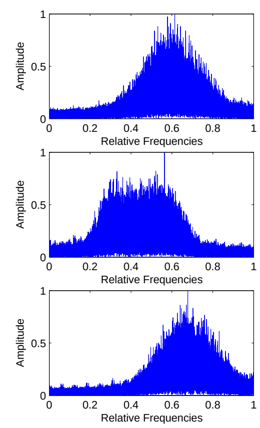

Figure 1: Signals FFT modulus

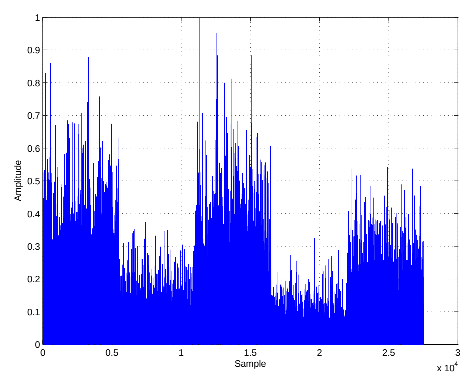

Figure 2: Filtered signals with SMF filter. Each point is the result of the filtering of 16 samples signal realization.

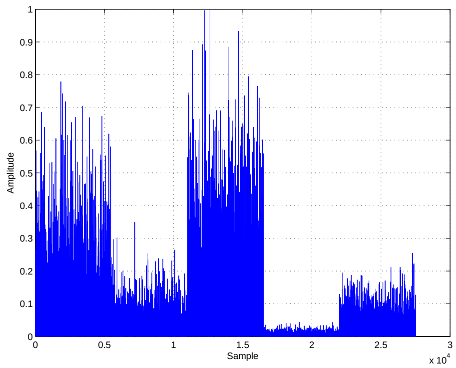

Figure 3: Filtered signals with CSMF filter. Each point is the result of the filtering of 16 samples signal realization.

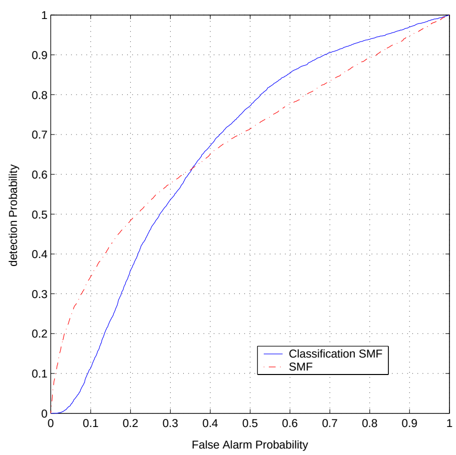

Figure 4: COR curves improvement

COR curves are presented on figure (4). For detection probability higher than 0.6, the CSMF COR curve is better than the SMF one. For applications needing a high detection level, the CSMF method is better than the SMF one, and we can get up to 0.1 in the detection probability.

5 Conclusion

In the classical detection problems, when only one signal and one noise are present and when are only known the two first order statistics, a lot of methods have been developped (SMF, ESMF, CSMF).
The method proposed in this paper deals with the case of more than one noise disturbing the interesting signal. This method is derived from the SMF method and leads to a filter matched to the signal and unmatched to all noises. The optimized criterion is the geometrical mean of all different SNR. An algorithm capable of finding this optimal filter is proposed.
One can notice that this filtering is like a projection onto a 1D subspace. A natural extension to this method finding p optimal filters consists in projecting onto the subspace spanned by these vectors. This method could be compared to Extented SMF (ESMF) or Constrained SMF (CSMF).

References

[1] Fukunaga K. Pattern classification. *Wiley Interscience ISBN 0-12-269851-7.*

[2] Lee K., Kim E.S., Doh W., and Youn D.H. Image enhancement based on signal subspace approach. *IEEE Transactions on image processing*, 8(8), august 1999. pp.1130-1134.

[3] Kim K.I., Park S.H., and Kim H.J. Kernel principal components analysis for texture classification. *IEEE Signal processing letters*, 8(2), february 2001. pp.39-41.

[4] Tuceryan M. and Jain A.K. Texture analysis in handbook of pattern recognition and computer vision, c.h. chen, l.f. pau, p.s.p. wang. *World Scientific publishing Co*, 1998. pp.207-248.

[5] Duda R.O., Hart P.E., and Stork D.G. Statistical pattern recognition. *Morgan Kaufmann ISBN 0-471-05669-3.*

[6] Verdu S. and Poor H.V. Signal selection for robust matched filtering. *IEEE Transactions on communications*, 31(5), may 1983. pp.667-670.

[7] Kailath T. and Poor H.V. Detection of stochastic processes. *IEEE Transactions on informations theory*, 44(6), october 1998. pp.2230-2259.

[8] Randen T. and Husoy J.H. Texture segmentation using filters with optimized energy separation. *IEEE Trans Image Processing*, 8(4), april 1999. pp.571-582.