A ROBUST SIGNAL DETECTION SCHEME

Jean-Jacques Fuchs

IRISA / Université de Rennes 1
Campus de Beaulieu - 35042 Rennes Cedex - France
fuchs@irisa.fr

ABSTRACT

We consider the detection of a signal in the presence of both interferences that lie in a known subspace and white noise. The signal to be detected is the product of an unknown amplitude and a known signature-vector that is itself subject to additive white Gaussian noise. We develop the maximum likelihood estimates (MLE) of the problem in order to apply the generalized maximum likelihood test (GLRT) to our detection problem. The performances of the proposed detector are illustrated by means of numerical simulations and it is compared to the standard matched subspace detector.

1. INTRODUCTION

In many applications it is desired to detect the presence of a signal s whose observation y is perturbed by interferences i and noise e:

$$y = s + i + e, \qquad s = \mu x, \qquad i = A\varphi$$

The signal $s \in R^N$ is the product the amplitude $\mu \in R$ and a signature $x \in R^N$. The interferences $i \in R^N$ lie in a known m-dimensional subspace of R^N that is spanned by the column vectors of the matrix A and $e \in R^N$ represents the additive broadband noise. Observing y the problem is to decide if μ is equal to zero or different from zero. This is one of the problems considered in [1]. We consider the case where the signature x is unperfectly known. It may be difficult to obtain or may be perturbed due to uncalibrated arrays, uncertainty in the localization, local scattering, etc. This case has been often considered, see [3, 4] and the references therein. The idea is to develop robust approaches or to identify s or x prior to detection. We believe that the model we consider in this contribution and which is described below has never been considered. It is somehow in between the two approaches alluded to above since we both identify the exact underlying x while using a robust modeling approach. The fact that we consider the interference subspace to be perfectly known is of course somehow restrictive and using our modeling approach it might be possible to remove this assumption, further investigations are needed. In section 2, we define the model we use, in section 3, we develop the maximum likelihood estimates that are used to build the generalized likelihood ratio test in section 4. Simulation results are presented in section 5 before we conclude in section 6.

2. THE MODEL

We are given two N-dimensional observation vectors y and z and we assume that they satisfy

$$y = \mu x + A\varphi + e, \qquad z = x + n \tag{1}$$

where e and n are independent Gaussian N-dimensional perturbation vectors with covariance matrices respectively $\sigma^2 I_N$ and $\alpha^2\sigma^2 I_N$. The vector $A\varphi$ represents the interferences that lie in the m-dimensional range of the known $(N,\ m)$ matrix A. The aggregated matrix $[A\ \ x]$ is assumed to be full column rank $m{+}1$. We assume $m{+}1$ to be smaller than N and indeed much smaller for the actual signal to noise ratio to be non-vanishing.

We will further assume that α and σ are known. One could of course assume more complicated but known covariance matrices for e and n without gaining in generality. One may relax the assumption on σ but not on α being known. As a matter of fact α is not identifiable even under H_1, i.e., the likelihood function can be made infinite by a proper choice of α when this parameter is assumed unknown.

3. THE MAXIMUM LIKELIHOOD ESTIMATES

From the model described in (1) and the different assumptions we made, it follows that the quantities to be estimated are μ, φ and x. The joint probability density function of y and z is easily deduced from the joint density of e and n which are independent random vectors.

The opposite of the log-likelihood function of the set $\{y,\ z\}$ is then, up to additive and multiplicative constants, equal to

$$\ell(.) = \sum_1^N (y_i - \mu x_i - a_i^T\varphi)^2 + \frac{1}{\alpha^2}(z_i - x_i)^2 \tag{2}$$

or equivalently in matrix form

$$\ell(\mu,\ \varphi,\ x) = \|y - \mu x - A\varphi\|^2 + \frac{1}{\alpha^2}\|z - x\|^2 \tag{3}$$

where $y_i,\ x_i,\ z_i$ are the i-th components of the corresponding vectors, a_i^T is the i-th row of A and $\|x\|^2 = \sum_i x_i^2$ is the square of the Euclidean norm. The maximum likelihood estimates (MLE) of μ, φ and x are obtained by minimizing $\ell(.)$. The minimum with respect to (w.r.t.) φ is attained at $\hat{\varphi} = A^+(y - \mu x)$ where $A^+ = (A^TA)^{-1}A^T$ is the pseudo-inverse of A. After substitution one has

$$\min_{\mu,\ x} \|(I - AA^+)(y - \mu x)\|^2 + \frac{1}{\alpha^2}\|z - x\|^2$$

Let us now denote $y^{\|} = AA^{+}y$ the projection of y on the range of A and $y^{\perp} = (I - AA^{+})y$ its orthogonal complement in R^N. We denote similarly $x^{\|}, x^{\perp}, z^{\|}$ and $z^{\perp}$ and rewrite the previous minimization as

$$\min_{\mu,\ x^{\|}, x^{\perp}} \|y^{\perp} - \mu x^{\perp}\|^2 + \frac{1}{\alpha^2}\|z^{\perp} - x^{\perp}\|^2 + \frac{1}{\alpha^2}\|z^{\|} - x^{\|}\|^2$$

The minimum w.r.t. $x^{\|}$ is attained for $x^{\|} = z^{\|}$ so that the last term vanishes. Let us now introduce the $(N, N-m)$ orthogonal matrix Q which is such that the projection matrix $I - AA^{+} = QQ^T$. One then has $\|x^{\perp}\|^2 = \|QQ^T x\|^2 = \|Q^T x\|^2$ and this tells us that $x^{\perp}$ has indeed only N-m degrees of freedom. The minimization problem can now be rewritten

$$\min_{\mu, \bar{x}} \|\bar{y} - \mu \bar{x}\|^2 + \frac{1}{\alpha^2}\|\bar{z} - \bar{x}\|^2$$

where $\bar{x} = Q^T x$, $\bar{y} = Q^T y$ and $\bar{z} = Q^T z$. This is a non-linear optimization problem in which one recognizes a total-least-squares (TLS) problem whose solution is known to depend upon a singular value decomposition (SVD) [5, 6, 9]. To put it into the basic TLS form we introduce a last change of notation. Let $\tilde{\mu} = \alpha\mu$, $\tilde{z} = \bar{z}/\alpha$ and $\tilde{x} = \bar{x}/\alpha$ to get

$$\min_{\tilde{\mu}, \tilde{x}} \|\bar{y} - \tilde{\mu}\tilde{x}\|^2 + \|\tilde{z} - \tilde{x}\|^2$$

setting $\rho = \tilde{z} - \tilde{x}$ and $r = \bar{y} - \tilde{\mu}\tilde{x}$, it becomes

$$\min_{\tilde{\mu}, \rho, r} \|\rho\|^2 + \|r\|^2 \ \text{ subject to } (\tilde{z} - \rho)\tilde{\mu} = \bar{y} - r$$

which can in turn be rewritten as

$$\min_{\Delta, \tilde{\mu}} \|\Delta\|_F^2 \ \text{ subject to } \ (\hat{C} - \Delta)\begin{bmatrix} \tilde{\mu} \\ -1 \end{bmatrix} = 0 \qquad (4)$$

with $\Delta = [\rho \quad r]$ and $\hat{C} = [\tilde{z} \quad \bar{y}]$ and $\|\Delta\|_F^2 = \sum_{i,j} \Delta_{i,j}^2 = \text{trace}(\Delta^T\Delta)$ is the square of the Frobenius norm of Δ. This optimization problem amounts to seek the perturbation Δ with smallest Frobenius norm that, when subtracted from $\hat{C}$ makes the matrix $C = \hat{C} - \Delta$ rank deficient. The solution of this problem is known. The optimal Δ is $\hat{s}_{\min}\hat{u}_{\min}\hat{v}_{\min}^T$ the rank-one matrix associated with the *smallest* singular triplet in the SVD of $\hat{C}$ and the value of the optimal cost is thus $\hat{s}_{\min}^2$. The MLE of $\tilde{\mu} = \alpha\mu$ is then deduced from the ratio of the two components of the 2-dimensional vector $\hat{v}_{\min}$

$$\hat{\tilde{\mu}} = -\frac{\hat{v}_{\min}(1)}{\hat{v}_{\min}(2)} \qquad (5)$$

and $\hat{\tilde{x}} = \tilde{z} - \hat{s}_{\min}\hat{u}_{\min}\hat{v}_{\min}^T(1)$. One can obtain the statistical properties of these estimates by computing the Fisher information matrix and inverting it [7] but they are indeed easier to obtain from first order matrix perturbation analysis [8, 9, 10]. We do not perform this analysis right now, we postpone it to section 4.3 where we only develop the parts that are useful for our purpose.

4. DETECTION

4.1 The generalized likelihood ratio test

In this section, we consider the detection problem associated with our model. Observing y and z we want to decide between the two hypotheses:

$$\begin{array}{lll} H_o: & y = A\varphi + e, & z = x + n \\ H_1: & y = \mu x + A\varphi + e, & z = x + n. \end{array}$$

i.e. to decide whether μ is equal to zero or $\mu \neq 0$. We propose to use the generalized likelihood ratio test (GLRT) to take the decision. Since the different noise variances are known it is convenient to work with twice the logarithmic GLR we denote $L(y, z)$. With the notations introduced in (3), one gets

$$L(y, z) = \frac{1}{\sigma^2}\{\ \ell(\hat{\mu}, \hat{\varphi}, \hat{x}|H_o) - \ell(\hat{\mu}, \hat{\varphi}, \hat{x}|H_1)\}$$

where $\ell(\hat{\mu}, \hat{\varphi}, \hat{x}|H_k)$ is the value of the minimum of $\ell(.)$ under hypothesis k, i.e., when the unknown parameters are replaced by their ML estimates.

Under H_o, when μ is known to be zero, the value of the minimum of (3) with respect to φ and x is easily found to be $\ell(0,\ \hat{\varphi},\ \hat{x}|H_o) = y^T(I - AA^{+})y = y^T QQ^T y$. Under H_1 we have seen above that the value of the minimum of $\ell(.)$ is $\ell(\hat{\mu},\ \hat{\varphi},\ \hat{x}|H_1) = \hat{s}_{\min}^2$, the square of the smallest non-zero singular value of

$$\hat{C} = [\tilde{z} \ \ \bar{y}\ ] = [\frac{\bar{z}}{\alpha} \ \ \bar{y}] = [\ \frac{Q^T z}{\alpha} \quad Q^T y]. \qquad (6)$$

which is also the smallest eigenvalue of the order two positive definite matrix $\hat{C}^T\hat{C}$. It follows that the logarithmic GLR, we will use in the test, takes the form

$$L(y,\ z) = \frac{1}{\sigma^2}\ \{\ y^T QQ^T y - s_{\min}^2(\hat{C})\} \qquad (7)$$

where we replaced $\hat{s}_{\min}^2$ by the more explicit notation $s_{\min}^2(\hat{C})$. Using (6), it is easy to show that $L(y,\ z)$ is always positive. In order to characterize the detection test which will consist in comparing $L(y,\ z)$ with a threshold η, we need to get the probability density function (p.d.f.) of $L(y,\ z)$ under both H_o and H_1. Since the hypothesis H_o is simple, the threshold η will be set to give a given false alarm rate P_{FA} and the probability of detection P_D under H_1 is then a function of the value of the parameter μ. We are in the case where the alternative is composite.

4.2 Matrix perturbation results

To evaluate the probability density function (p.d.f.) of $L(y,\ z)$ under both hypotheses we apply results from matrix perturbation theory [9, 8]. They will tell us how the noises e and n present in the observations (1) influence the parameter estimates. Though this theory is purely deterministic and is merely a first order expansion, it can be applied to stochastic signals and yield valuable information on, e.g., the characteristics of the asymptotic distributions, once it has been established by other means that it exists, e.g., that a central limit theorem holds. Here we use this theory to get the statistical properties of $s_{\min}(\hat{C})$, the smallest singular value of the (N-m,2)-dimensional matrix $\hat{C}$ (6). Indeed $\hat{C}$ can be seen as the sum of an *exact* matrix C and a perturbation Δ:

$$\hat{C} = C + \Delta = [\frac{Q^T x}{\alpha} \quad \mu Q^T x] + [\frac{Q^T n}{\alpha} \quad Q^T e]. \qquad (8)$$

Note that the scaling by α in the first column of Δ makes all the entries of this matrix have zero mean and identical standard deviation σ, a prerequisite in matrix perturbation analysis.

Since the smallest singular value of C, $s_{\min}(C)$ is zero, one expects $s_{\min}(\hat{C})$ to be quite small. Indeed since it can only increase the perturbation theory relative to $s_{\min}(C)$ is of a different nature than those of non-zero singular values. It is also more intricate. We present in the appendix a proposition that summarizes the corresponding first order perturbation results. Applied to the perturbation of $s_{\min}(C)$ the proposition says that

$$s_{\min}(\hat{C}) = \|U_2^T \Delta v_{\min}\| + O(\sigma^2) \qquad (9)$$

where U_2 is an orthogonal basis of the (N-m-1)-dimensional null-space of CC^T and $v_{\min}$ the single basis vector of the one-dimensional null-space of $C^T C$. From the definition of C above or from (4) it follows that

$$v_{\min}^T = [-\alpha\mu \quad 1]/(1 + \alpha^2\mu^2)^{1/2}. \qquad (10)$$

This result holds only asymptotically, i.e., for small enough perturbation Δ in (8).

Remark on the asymptotics

As opposed to the situation in [1] where the results are non-asymptotic and the statistics involved somehow ancillary, here because the model (1) is non-linear, the results are only valid asymptotically and it is important to characterize their domain of validity. Surprisingly using matrix perturbation analysis one can indeed be more precise than in standard statistical analysis.

Here *asymptotic* means for small enough perturbation Δ in (8). Since all the components in Δ are zero mean and have identical standard deviation σ, this holds for σ small enough or more generally for large enough signal to noise ratio, an SNR, that has yet to be defined. More precisely, the validity of (9) relies on the well-separateness of the singular values of interest. In the present case, it holds if $s_1(C) \gg s_2(C)$ with $s_1(C)$ and $s_2(C)$ the two singular values of C. Since $s_2(C) = s_{\min}(C) = 0$, the single non-zero singular value of C, $s_1(C)$ is equal to the square root of the trace of $C^T C$. But $s_1(C) \gg 0$ has no real meaning and should be replaced by $s_1(\hat{C}) \gg s_2(\hat{C})$ which can be checked on each individual $\hat{C}$-matrix. In fact $s_1(C)$ will vary by about σ while $s_2(\hat{C})$ will increase by as much as $\sigma\sqrt{N-m-1}$ as follows from (9) and will be established in section 4.3. In summary, the analysis we perform is valid if

$$\frac{x^T Q Q^T x}{\sigma^2} \left(\frac{1}{\alpha^2} + \mu^2\right) \gg N - m \qquad (11)$$

One should have a ratio of at least ten between both sides and this can be checked either on the data or a priori on the model. This kind of situation where the domain of validity is explicit and can be checked is quite rare in statistics! Note that for $\alpha = 0$, i.e., in the standard situation of [1], this condition is always satisfied.

4.3 Statistical analysis

We now evaluate the statistical properties of $L(y, z)$

$$L(y,\ z) = \frac{1}{\sigma^2}\ \{\ y^T Q Q^T y - s_{\min}^2(\hat{C})\}$$

under both H_o and H_1. Under H_o, see (8):

$$\hat{C} = C + \Delta = [\frac{Q^T x}{\alpha} \quad 0] + [\frac{Q^T n}{\alpha} \quad Q^T e].$$

and $QQ^T y = QQ^T e$, the first term in $L(.)$ becomes $e^T Q Q^T e$. To evaluate the second we observe that $v_{\min}$ in (10) becomes $v_{\min}^T = [0 \quad 1]$ and substituting $v_{\min}$ and Δ in (9) one has

$$s_{\min}(\hat{C}) \simeq \|U_2^T Q^T e\|$$

and therefore

$$L(y,\ z|H_o) \simeq \frac{e^T Q(I_{N-m} - U_2 U_2^T) Q^T e}{\sigma^2} = \frac{(u_1^T Q^T e)^2}{\sigma^2}$$

where u_1 which is such that $I_{N-m} - U_2 U_2^T = u_1 u_1^T$ is a basis of the one-dimensional range-space of C. Since $u_1^T Q^T Q u_1 = 1$, it follows that, asymptotically, $L(y,\ z|H_o)$ behaves like a chi-square random variable with one degree of freedom, i.e., the square of a standard Gaussian random variable, $L(y,\ z|H_o)\ :\ \chi_1^2$.

The analysis is similar under H_1, the matrix $\hat{C}$ is now

$$\hat{C} = [\frac{Q^T x}{\alpha} \quad \mu Q^T x] + [\frac{Q^T n}{\alpha} \quad Q^T e].$$

and $QQ^T y = \mu Q Q^T x + Q Q^T e$. To evaluate the distribution of the first term in $L(y,\ z|H_1)$ one can note that $Q^T y = \mu Q^T x + Q^T e$ where $Q^T e$ is a zero mean Gaussian vector with covariance matrix $\sigma^2 I_{N-m}$. The first term thus follows

$$\frac{y^T Q Q^T y}{\sigma^2} : \chi_{N-m}^2\left(\frac{\mu^2}{\sigma^2} x^T Q Q^T x\right)$$

a chi-square random variable with N-m degrees of freedom and non-centrality parameter $\frac{\mu^2}{\sigma^2} x^T Q Q^T x$. With $v_{\min}$ in (10) and Δ just above, one has

$$s_{\min}(\hat{C}) \simeq \frac{\|U_2^T(-\alpha\mu Q^T \frac{n}{\alpha} + Q^T e)\|}{(1 + \alpha^2\mu^2)^{1/2}}$$

which leads to, omitting the zero mean cross-products

$$s_{\min}^2(\hat{C}) \simeq \frac{\|U_2^T Q^T e\|^2 + \alpha^2\mu^2\|U_2^T Q^T \frac{n}{\alpha}\|^2}{1 + \alpha^2\mu^2}$$

since both $\|U_2^T Q^T e\|^2/\sigma^2$ and $\|U_2^T Q^T \frac{n}{\alpha}\|^2/\sigma^2$ are χ_{N-m-1}^2 random variables, it follows that $s_{\min}^2(\hat{C})/\sigma^2$

is a random variable with mean $N-m-1$ and variance $2(N-m-1)(1+\alpha^4\mu^4)/(1+\alpha^2\mu^2)^2$.

Combining these two partial results and taking into account the dependency existing between them, one has that the distribution of $L(y,\ z|H_1)$ admits no analytical expression, it is a random variable with mean $1+\lambda^2$ with $\lambda^2 = \frac{\mu^2}{\sigma^2} x^T Q Q^T x$ and variance:

$2(1+2\lambda^2) + 4(N-m-1)(\alpha^2\mu^2)/(1+\alpha^2\mu^2)^2$.

Its mean is the one of a $\chi_1^2(\lambda^2)$ random variable but its variance is larger by $4(N-m-1)(\alpha^2\mu^2)/(1+\alpha^2\mu^2)^2$ a term that vanishes with α.

5. SIMULATION RESULTS

5.1 The test procedure

In summary, to test between

$$\begin{array}{lll} H_o: & y = A\varphi + e, & z = x + n \\ H_1: & y = \mu x + A\varphi + e, & z = x + n. \end{array}$$

i.e., a simple hypothesis H_o, $\mu = 0$ versus a composite two-sided alternative $\mu \neq 0$, we propose to use the logarithmic GLR which is given by

$$L(y,\ z) = \frac{1}{\sigma^2}\,\{y^T(I - AA^+)y - s_{\min}^2(\hat{C})\}$$

where $s_{\min}(\hat{C})$ is the smallest singular value of

$$\hat{C} = [\tilde{z}\ \ \bar{y}\ ] = (I - AA^+)\ [\frac{z}{\alpha}\ \ y]$$

We have shown that $L(y,\ z)$ which is positive with probability one, is asymptotically close to central chi-squared random variable with one degree of freedom under H_o and that under H_1 it is close to a chi-squared random variable with one degree of freedom and non-centrality parameter

$$\lambda^2 = \frac{\mu^2}{\sigma^2} x^T(I - AA^+)x$$

with however a larger variance. We thus decide that H_o is true when $L(y,\ z) \leq \eta$ and that H_1 holds when $L(y,\ z) > \eta$ where the threshold η is fixed to achieve a given false alarm probability

$$P_{FA} = 1 - Pr\{\chi_1^2(0) \leq \eta\}. \qquad (12)$$

The probability of detection is then difficult to compute but depends upon μ trough λ^2 which can be seen as an SNR. The notation $Pr\{\chi_1^2(\lambda^2) \leq \eta\}$ denotes the probability that a χ_1^2 with non-centrality parameter λ^2 is less than η.

5.2 Receiver operating characteristics

The receiver operating characteristics (ROC's) for this detector are presented in Figure 1.They are obtained by simulations. For 5 different probability of false alarms (P_{FA}) we represent the probability of detection as a function of the SNR$=\frac{\mu^2}{\sigma^2}x^T(I - AA^+)x$. The thresholds are obtained using relation (12) and we have checked that as soon as condition (11) is satisfied there are no discrepancies between the expected P_{FA}'s and those observed on the simulations. In each case we further consider 4 different values of α^2, namely α^2 equal to $0,\ 1/4,\ 1/2$ and 1. Remember from (1) that α^2 is the ratio between the variance of the noise n that affects the signature and the variance of the noise e present in the observations. For $\alpha^2 = 0$, we are in what we will call the standard case considered in [1]. In this case our detector has performance similar to the one developed in [1]. Further investigations are necessary to analyze this point. One sees in Figure 1 that the probability of detection are essentially independent of the value of α^2, the four curves are almost indistinguishable for each value of P_{FA}.

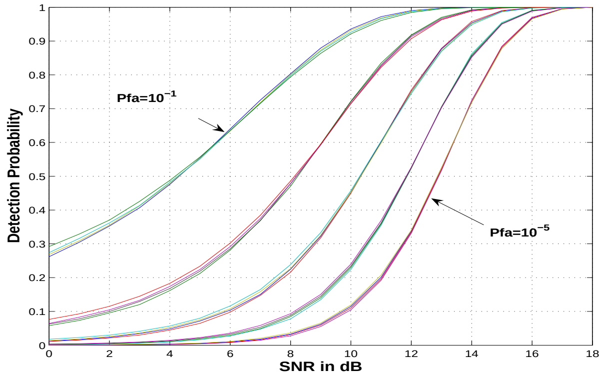

Figure 1: ROC for proposed method, Pfa equal to 10^{-k} for $k = 1$ to 5 and $\alpha^2 = 0,\ 1/4,\ 1/2,\ 1$.

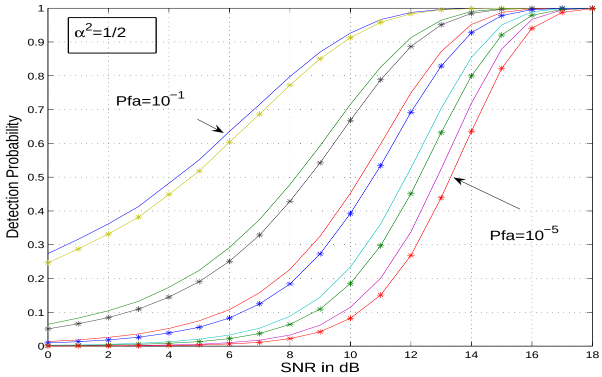

Figure 2: ROC for Pfa equal to 10^{-k} for $k = 1$ to 5 and $\alpha^2 = 1/2$. In each pair of curves: proposed detector (solid line) and standard detector (-*-line).

In Figure 2, we compare for the same 5 values of P_{FA} and the same range of SNR's, the probability of detection (P_D) of our detector and the P_D of the standard detector [1] for $\alpha^2 = 1/2$. One sees that our detector outperforms slightly the standard detector. This is quite natural since our detector is designed for the model (1) we simulated while the standard detector is optimal only if $\alpha^2 = 0$. The loss in P_D is quite small. To complete the picture we present in Figure 3, the P_D of the standard detector [1] for 3 values of P_{FA} and the 4 different values of α^2, namely α^2 equal to $0,\ 1/4,\ 1/2$ and 1. As α^2 increases, the performance degrades in a regular way which is again quite as expected.

Note that we have been careful and always respected (11) in our different simulations. This guarantees their validity.

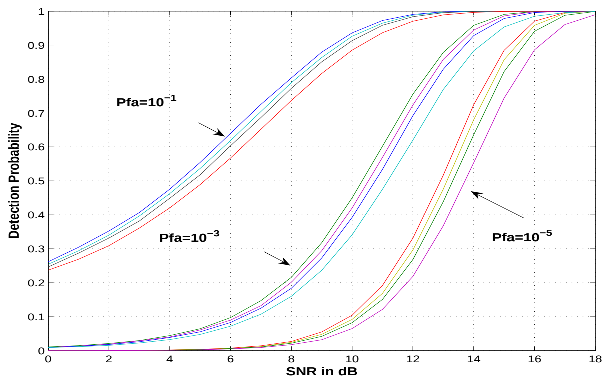

Figure 3: ROC for 3 different Pfa for standard approach and $\alpha^2 = 0,\ 1/4,\ 1/2,\ 1$.

6. CONCLUSIONS

We have proposed a new detection problem and developed the associated GLRT. The contribution can be viewed as an extension of one of the matched subspace detectors of [1]. We have only considered the case where the noise covariance matrices are known and further investigations are needed to remove or to alleviate this assumption. One can note that the presence of interferences is easily taken care of and actually introduces no additional difficulty. Another important issue that requires further analysis concerns the influence on the performance of a wrong choice of α. What happens to the performance of the proposed detector when we assume the presence of noise on x with a given level α while x is perfectly known and the standard test would be optimal? Simulations we performed indicate that there is no loss in performance but this remains to be established theoretically.

REFERENCES

[1] L.L. Scharf and B. Friedlander, "Matched subspace detectors", *IEEE Trans. Signal Processing*, vol. 42, 8, pp. 2146-2157, aug. 1994.

[2] L.L. Scharf, *Statistical signal processing, detection, estimation and time series analysis.* Addison Wesley , 1991.

[3] H.V. Trees, *Optimum array Processing.* N.Y., Wiley, 2002.

[4] O. Besson and F. Vincent, "Performance analysis of beamformers using generalized loading of the covariance matrix", *IEEE Trans. Signal Processing*, vol. 53, 2, pp. 452-459, Feb. 2005.

[5] G.H. Golub and C.F. Van Loan, "An analysis of the total least squares problem," *SIAM J. Numer. Anal.*, vol. 17, 6, 883-893, 1980.

[6] P. Dewilde and E.F. Deprettere, "Singular Value Decomposition : an introduction" *SVD and Signal Processing* pp. 3-41, North-Holland, 1988.

[7] L.J. Gleser, "Estimation in a multivariate "errors in variables" regression model: large sample results," *The Annals of Statistics*, vol. 9, 1, pp. 24-44, 1981.

[8] G.W. Stewart and J.G. Sun, *Matrix Perturbation Theory.* Academic Press, 1990.

[9] G.H. Golub and C.F. Van Loan, *Matrix Computations.* John Hopkins University Press, 1983.

[10] J.J. Fuchs, "Rectangular Pisarenko Method Applied to Source Localization," *IEEE Trans. Signal Processing*, vol. 44, 10, pp. 2377-2383, oct. 1996.

[11] J.J. Fuchs, "Estimation of the number of signals in the presence of unknown correlated sensor noise." *IEEE Trans. Signal Processing*, vol 40, 5, pp. 1053–1061, may 1992.

7. APPENDIX

We present a result from matrix perturbation analysis that concerns the singular value decomposition of a rank deficient matrix [8, 9, 10].

Let C be an $(N,m+1)$ dimensional matrix of rank m with $m+1 < N$ and singular value decomposition (SVD) $C = USV^T$. We partition U, S and V as

$$S^T = \begin{bmatrix} S_1 & 0 & 0 \\ 0 & s_{m+1} & 0 \end{bmatrix},\ U = [U_1\ \ u_{m+1}\ \ U_3],\ V = [V_1\ \ v_{m+1}]$$

with S_1 a diagonal order-m matrix, S_1= diag(s_i) and $s_1 \geq s_2 \geq .. \geq s_m > s_{m+1} = 0$. The matrix $U_2 = [u_{m+1}\ \ U_3]$ is an orthogonal basis of the null-space of CC^T and v_{m+1} is the single zero eigen-vector of C^TC. Partitioning in a similar way the elements $\hat{U}, \hat{S}$ and $\hat{V}$ of the SVD of $\hat{C} = C + \epsilon\Delta$ the following result can be shown to hold concerning the *smallest* singular triplet of $\hat{C}$ [11]

Proposition: For sufficiently small ϵ and a given perturbed matrix $\hat{C}$, there exists a $\hat{C}$-dependent orthogonal basis $\check{U}_2 = [\check{u}_{m+1}\ \ \check{U}_3]$ of the null-space of CC^T such that

$$\hat{U}_2 = \check{U}_2 + O(\epsilon)$$
$$\hat{v}_{m+1} = v_{m+1} - \epsilon C^{+}\Delta v_{m+1} + O(\epsilon^2)$$
$$\hat{s}_{m+1} = \epsilon\ \|\ U_2^T \Delta v_{m+1}\| + O(\epsilon^2) \qquad \square$$

Note that the specific perturbation dependent basis $\check{U}_2$ only intervenes in the first part of the proposition.