MATCHED DETECTOR AND ESTIMATOR WITH SIGNATURE UNCERTAINTY.

Jean-Jacques FUCHS

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

ABSTRACT

We consider the detection and estimation of a signal, whose associated spatial signature is imperfectly known, in the presence of broad-band noise. While in most cases, the spatial signature (or steering vector) is assumed to be known, i.e., lies along a known direction, we consider the case where the signature is known up to additive white Gaussian noise. This amounts to assuming that it somehow lies in a cone, the aperture of which depends upon the level of uncertainty. We consider the multiple snapshots case and assume the unknown signature to remain constant over the whole set of snapshots. We develop the maximum likelihood estimator of the different unknowns, build the generalized likelihood ratio (GLR). We then evaluate the statistical properties of the associated test under both hypotheses and check the validity of the asymptotic analysis on simulations..

Index Terms— Signal detection, robustness, generalized likelihood ratio test, signature uncertainty

1. INTRODUCTION

It is desired to recover a signal of interest s_t in the presence of interferences and noise using an array of n sensors. More formally the basic model is the following:

at each time instant $t = 1, \dots, N$, one observes

$$\mathbf{y}_t = \mathbf{x}_t + \mathbf{i}_t + \mathbf{n}_t, \text{ with } \mathbf{x}_t = \mathbf{a}s(t) \text{ and } \mathbf{i}_t = A\mathbf{u}_t. \quad (1)$$

In this model $\mathbf{a} \in R^n$ is the known signature (or steering vector), $\mathbf{i}_t$ the interferences and $\mathbf{n}_t$ the noise. One assumes to know exactly A which defines the subspace in which the interferences are assumed to lie. This model is essentially an extension of the matched detector [1, 2, 3] to the multi-snapshot case since we consider that one observes N snapshots $\{\mathbf{y}_t\}$.

We will extend model (1) to the case where $\mathbf{a}$ is not exactly known. Since, if A is perfectly known the difficulties induced by the presence of the interferences is marginal [4, 5] we will omit this in the sequel to simplify the notation. Uncertainty in the signature $\mathbf{a}$ has been considered in [6], where it is modeled as $\mathbf{a} = H\theta$ with H a known (p, n) matrix, i.e.

the signature belongs to a known subspace. Other models for the uncertainty have already been proposed and we refer the reader to the references in [6]. We consider the case where the signature is known up to an additive noise vector, this is the model considered in the single snapshot case, in [5]. In Section 2, we define the model we propose, and in Section 3, we develop the maximum likelihood estimates (MLE) of the different unknowns that are used to build the generalized likelihood ratio test (GLRT) in Section 4. Simulation results are presented in Section 5 before we conclude in Section 6.

2. MODEL

We consider the following model for the n -dimensional received signal $\mathbf{y}_t$ and observed signature $\mathbf{b}$:

$$\mathbf{y}_t = \mathbf{a}s_t + \mathbf{n}_t, \quad t = 1, \dots, N; \quad \mathbf{b} = \mathbf{a} + \alpha\mathbf{e}, \quad (2)$$

where $\mathbf{n}_t$ and $\mathbf{e}$ are zero-mean complex valued Gaussian random vectors with covariance matrix $\sigma^2 I$. At the beginning of the transmission of the N snapshots $\mathbf{y}_t$ one observes or knows a noisy version $\mathbf{b}$ of the true signature or steering vector $\mathbf{a}$ and then, using these $N+1$ vectors, one has to decide if a signal is present, and if it is the case, one wants to estimate the deterministic scalar sequence $\{s(t)\}_{t=1}^N$. The positive real scalar α allows to tune the magnitude of the uncertainty on $\mathbf{a}$.

We will assume that α and σ are known. One could assume more complicated but known covariance matrices for $\mathbf{e}$ and $\mathbf{n}$, without gaining in generality. One may relax the assumption on σ but not on α being known since α is not identifiable.

To simplify the notation we introduce $\mathbf{Y} = [\mathbf{y}_1 \dots \mathbf{y}_N]$ and similarly define $\mathbf{s}$ and $\mathbf{N}$, the global model is then

$$\mathbf{Y} = \mathbf{a}\mathbf{s}^T + \mathbf{N}, \quad \mathbf{b} = \mathbf{a} + \alpha\mathbf{e} \quad (3)$$

with $\mathbf{Y}$ and $\mathbf{N}$, (n, N) - matrices and $\mathbf{s}^T$ an N -dimensional row vector. From a detection point of view, observing $\mathbf{Y}$ and $\mathbf{b}$, one has to decide if the waveform $\mathbf{s} = \mathbf{0}$ or not, and if one decides that the waveform is present, from an estimation point of view one has to estimate $\mathbf{s}$.

3. MAXIMUM LIKELIHOOD ESTIMATES

Since one observes both $\mathbf{Y}$ and $\mathbf{b}$, and since $\mathbf{e}$ and $\mathbf{N}$ are independent Gaussian variables with identical variances the MLE of the unknowns $\mathbf{a}$ and $\mathbf{s}$ are obtained by minimizing

$$\ell(\mathbf{a}, \mathbf{s}) = \|\mathbf{Y} - \mathbf{a}\mathbf{s}^T\|_F^2 + \left\| \frac{\mathbf{b}}{\alpha} - \frac{\mathbf{a}}{\alpha} \right\|^2 \quad (4)$$

where $\|\mathbf{Y}\|_F^2 = \text{Tr}(\mathbf{Y}^T \mathbf{Y})$ denotes the square of the Frobenius norm of $\mathbf{Y}$. Introducing new variables $\mathbf{R} = \mathbf{Y} - \mathbf{a}\mathbf{s}^T$ and $\mathbf{r} = (\mathbf{b} - \mathbf{a})/\alpha$, the optimization problem becomes, eliminating $\mathbf{a}$:

$$\min_{\mathbf{R}, \mathbf{r}, \mathbf{m}} \|\mathbf{R}\|_F^2 + \|\mathbf{r}\|^2 \quad \text{subject to} \quad \mathbf{Y} - \mathbf{R} = \left(\frac{\mathbf{b}}{\alpha} - \mathbf{r} \right) \mathbf{m}^T$$

with $\mathbf{m} = \alpha\mathbf{s}$. Thus, we have replaced an unconstrained minimization problem with $N + n$ unknowns in $\mathbf{s}$ and $\mathbf{a}$ by a equivalent constrained problem in $Nn + N + n$ unknowns with Nn linear constraints. We will further transform this problem. Introducing $\hat{C} = [\mathbf{Y} \quad \mathbf{b}/\alpha]$ and $\delta C = [\mathbf{R} \quad \mathbf{r}]$, then the constraint in the problem above becomes

$$\mathbf{Y} - \mathbf{R} = \left(\frac{\mathbf{b}}{\alpha} - \mathbf{r} \right) \mathbf{m}^T \Leftrightarrow (\hat{C} - \delta C) [I_N \quad -\mathbf{m}]^T = 0_{n,N}$$

with I_N the order N identity matrix. This says that the N first column in $\hat{C} - \delta C$ are all colinear to the last one, i.e., that $\hat{C} - \delta C$ is of rank one, hence the new formulation of the optimization problem

$$\min_{\delta C} \|\delta C\|_F^2 \quad \text{subject to} \quad \text{rank}(\hat{C} - \delta C) = 1.$$

The solution to this last problem is well known [7], it is obtained from the singular value decomposition (SVD) of $\hat{C}$, if $\hat{C} = \sum_1^n \hat{s}_k \hat{u}_k \hat{v}_k^T$ with $\hat{s}_1 > \hat{s}_2 \geq \dots \geq \hat{s}_n$, then the minimum is attained for $\delta C = \sum_2^n \hat{s}_k \hat{u}_k \hat{v}_k^T$ and the minimal value of the criterion is $\|\delta C\|_F^2 = \sum_2^n \hat{s}_k^2 = \|\hat{C}\|_F^2 - \hat{s}_1^2$. The associated MLE $\hat{\mathbf{s}}$ of $\mathbf{s} = \mathbf{m}/\alpha$ is then

$$\hat{\mathbf{s}} = -\frac{\hat{v}_1(1:N)}{\alpha \hat{v}_1(N+1)} \quad (5)$$

with $\hat{v}_1$ the maximal right singular vector of $\hat{C}$, an $N+1$ dimensional vector with components denoted $\hat{v}_1(k)$.

Indeed this result is quite logical if one observes that, say C the expected value of $\hat{C}$ is equal to $[\mathbf{a}\mathbf{s}^T \quad \mathbf{a}/\alpha] = \mathbf{a} [\mathbf{s}^T \quad 1/\alpha]$. C is thus a rank one matrix with maximal right singular vector v_1 colinear to $[\mathbf{s}^T \quad 1/\alpha]$ hence (5).

It follows also that the importance of the observation $\mathbf{b}$ depends highly on how small α is, $\mathbf{b}$ plays the same role as

an additional observation say y_0 with *known* associated signal $s_0 = 1/\alpha$.

One can establish along these lines that as $\alpha \rightarrow 0$, as the uncertainty on the signature vanishes, $\mathbf{b}$ converges towards $\mathbf{a}$ and $\hat{\mathbf{s}}$ in (5) converges to the usual estimate $\mathbf{Y}^T \mathbf{a} / \mathbf{a}^T \mathbf{a}$ [6].

4. DETECTION

4.1. Generalized Likelihood Ratio Test

We consider now the detection problem associated with our model. Observing $\mathbf{Y}$ and $\mathbf{b}$, we want to decide between:

$$H_0: \mathbf{Y} = \mathbf{N}, \quad \mathbf{b} = \mathbf{a} + \alpha \mathbf{e}$$

$$\text{and } H_1: \mathbf{Y} = \mathbf{a}\mathbf{s}^T + \mathbf{N}, \quad \mathbf{b} = \mathbf{a} + \alpha \mathbf{e}.$$

and we propose to use a generalized likelihood ratio test (GLRT) to make the decision. Since we assume to know the different noise variances it is convenient, to work with twice the logarithmic GLR which we denote $L(\mathbf{Y}, \mathbf{b})$. With the quantity $\ell(\cdot)$ introduced in (4), one gets

$$L(\mathbf{Y}, \mathbf{b}) = \frac{1}{\sigma^2} [\ell(\mathbf{a}, \mathbf{s} | H_0) - \ell(\mathbf{a}, \mathbf{s} | H_1)]$$

where $\ell(\mathbf{a}, \mathbf{s} | H_k)$ is the value of the minimum of $\ell(\cdot)$ under hypothesis k , i.e., when the unknown parameters are replaced by their ML estimates. Under H_0 , the minimum of $\ell(\mathbf{a}, 0)$ is obtained for $\mathbf{a} = \mathbf{b}$ and its value is $\|\mathbf{Y}\|_F^2 = \text{Tr}(\mathbf{Y}^T \mathbf{Y})$. Under H_1 , the minimum has been shown in Section 3 to be

$$\|\delta C\|_F^2 = \text{Tr}(\hat{C}^T \hat{C}) - \hat{s}_1^2 = \text{Tr}(\mathbf{Y}^T \mathbf{Y}) + \frac{\mathbf{b}^T \mathbf{b}}{\alpha^2} - \hat{s}_1^2$$

Hence the GLR is equal to

$$L(\mathbf{Y}, \mathbf{b}) = \frac{1}{\sigma^2} \left(\hat{s}_1^2 - \frac{\mathbf{b}^T \mathbf{b}}{\alpha^2} \right)$$

where $\hat{s}_1$ is the largest singular value of $\hat{C} = [\mathbf{Y} \quad \mathbf{b}/\alpha]$. Again it is possible to establish that as $\alpha \rightarrow 0$, one recovers the basic likelihood ratio

$$L(\mathbf{Y}, \mathbf{b}) \rightarrow L(\mathbf{Y}) = \frac{1}{\sigma^2} \frac{\mathbf{a}^T \mathbf{Y} \mathbf{Y}^T \mathbf{a}}{\mathbf{a}^T \mathbf{a}}.$$

which is itself an extension to the multi-snapshot case of the single snapshot case [1, 2, 3] for which $\mathbf{Y} = \mathbf{y}$, and the associated matched detector based on $L(y) = (1/\sigma^2) (\mathbf{a}^T \mathbf{y})^2 / (\mathbf{a}^T \mathbf{a})$.

In order to characterize the detection test which will consist in comparing $L(\mathbf{Y}, \mathbf{b})$ with a threshold η , we need to get the probability density function (p.d.f.) of $L(\mathbf{Y}, \mathbf{b})$ under both H_0 and H_1 . Since the hypothesis H_0 is simple, the threshold η will be set to obtain 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 signal to noise ratio (SNR) to be defined below.

4.2. Statistical Analysis

To evaluate the statistical properties of the LR under both H_0 and H_1 , we will use matrix perturbation results. The idea is to say that the observed matrix $\hat{C} = [\mathbf{Y} \ \mathbf{b}/\alpha]$ can be seen as the sum of a rank-1 matrix $C = [\mathbf{a}\mathbf{s}^T \ \mathbf{a}/\alpha]$ and a perturbation matrix $\Delta = [\mathbf{N} \ \mathbf{e}]$ whose components are all independent $N(0, \sigma^2)$ random variables with magnitudes of the order of σ . The largest singular value $\hat{s}_1$ of $\hat{C}$ can then be seen as a perturbation of the largest singular value s_1 of C .

To analyze $L(\mathbf{Y}, \mathbf{b})$ under H_1 , we evaluate separately the two terms in $L(\mathbf{Y}, \mathbf{b})$. Under H_1 ,

$$\hat{C} = \begin{bmatrix} \mathbf{Y} & \mathbf{b} \\ & \alpha \end{bmatrix} = C + \Delta = \begin{bmatrix} \mathbf{a}\mathbf{s}^T & \mathbf{a} \\ & \alpha \end{bmatrix} + [\mathbf{N} \ \mathbf{e}]$$

where C is a rank one matrix with SVD:

$$C = s_1 u_1 v_1^T \quad \text{with } s_1^2 = (\mathbf{a}^T \mathbf{a})(\mathbf{s}^T \mathbf{s}) + \frac{\mathbf{a}^T \mathbf{a}}{\alpha^2} \quad \text{and } u_1 = \frac{\mathbf{a}}{\|\mathbf{a}\|}.$$

We use matrix perturbation results to evaluate the statistical properties of $\hat{s}_1^2$ the maximal singular value of $\hat{C}$. We show in the Appendix that $\hat{s}_1^2/\sigma^2$ is close to a non-central chi-square random variable with $N+n$ degrees of freedom and non-centrality parameter s_1^2 and that this approximation is valid if $s_1 \gg \sigma$, which under H_1 , amounts to having $\|\mathbf{a}\|\|\mathbf{s}\| + \|\mathbf{a}\|/\alpha \gg \sqrt{N}\sigma$.

The second term in $L(\mathbf{Y}, \mathbf{b})$ is easier to analyze:

$$\frac{\mathbf{b}^T \mathbf{b}}{\alpha^2 \sigma^2} \simeq \frac{\mathbf{a}^T \mathbf{a}}{\alpha^2 \sigma^2} + \chi_n^2.$$

Combining these two partial results, it follows with some further approximation, since the two terms are correlated, that

$$L(\mathbf{Y}, \mathbf{b}|H_1) \simeq \chi_N^2(\lambda^2) \quad \text{with } \lambda^2 = (\mathbf{a}^T \mathbf{a})(\mathbf{s}^T \mathbf{s})/\sigma^2$$

To analyze $L(\mathbf{Y}, \mathbf{b}|H_0)$, we note that under H_0 ,

$$\hat{C} = \begin{bmatrix} \mathbf{N} & \mathbf{b} \\ & \alpha \end{bmatrix} = \begin{bmatrix} \mathbf{0} & \mathbf{a} \\ & \alpha \end{bmatrix} + [\mathbf{N} \ \mathbf{e}]$$

since $\mathbf{s} = \mathbf{0}$. Applying the results of the Appendix, one can indeed simultaneously analyze the two terms in $L(\mathbf{Y}, \mathbf{b})$ but it appears that a higher-order expansion is needed when α is non-zero. A thorough analysis shows that $L(\mathbf{Y}, \mathbf{b}|H_0)$ can be approximated by a central chi-square random variable with N degrees of freedom times another random variable with mean $1 + (n-1)(\alpha^2 \sigma^2)/(\mathbf{a}^T \mathbf{a})$. We will therefore consider that

$$L(\mathbf{Y}, \mathbf{b}|H_0) \simeq \beta \chi_N^2(0) \quad \text{with } \beta = 1 + (n-1) \frac{\alpha^2 \sigma^2}{\mathbf{a}^T \mathbf{a}}. \quad (6)$$

This approximation is furthermore valid if

$$\frac{\mathbf{a}^T \mathbf{a}}{\alpha^2} \gg N \sigma^2 \quad (7)$$

a condition that is more difficult to satisfy than under H_1 .

5. SIMULATION RESULTS

The aim of this section is to check the validity of the analysis above and to illustrate the detection performance. In all the simulations we take $n = 10$, $N = 10$, $\mathbf{a}^T \mathbf{a} = 20$ and $\sigma = 1$, and tune $\mathbf{s}$ to achieve the desired $\text{SNR} = \lambda^2 = (\mathbf{a}^T \mathbf{a})(\mathbf{s}^T \mathbf{s})/\sigma^2$.

5.1. Test Procedure

In summary, to test between

$$\begin{aligned} H_0: & \quad \mathbf{Y} = \mathbf{N}, & \mathbf{b} &= \mathbf{a} + \alpha \mathbf{e} \\ H_1: & \quad \mathbf{Y} = \mathbf{a}\mathbf{s}^T + \mathbf{N}, & \mathbf{b} &= \mathbf{a} + \alpha \mathbf{e} \end{aligned}$$

where $\mathbf{Y}$ is an n by N full-row rank matrix. We propose to use the GLR which is given by

$$L(\mathbf{Y}, \mathbf{b}) = \frac{1}{\sigma^2} \left[\hat{s}_1^2 - \frac{\mathbf{b}^T \mathbf{b}}{\alpha^2} \right],$$

where $\hat{s}_1$ is the largest singular value of $\hat{C} = [\mathbf{Y} \ \mathbf{b}/\alpha]$. Given the approximate p.d.f.'s we have obtained in Section 4.2 for $L(\mathbf{Y}, \mathbf{b})$ under both hypotheses, we decide that H_0 is true if $L(\mathbf{Y}, \mathbf{b}) < \eta$ and that H_1 holds otherwise, where the threshold η is fixed to achieve a given false alarm probability

$$P_{FA} = \Pr\{\beta \chi_N^2 \geq \eta\} \quad \text{with } \beta = 1 + (n-1) \frac{\alpha^2 \sigma^2}{\mathbf{a}^T \mathbf{a}} \quad (8)$$

The probability of detection P_D depends then upon λ^2 which can be seen as a SNR. The notation $\Pr\{\chi_N^2(\lambda^2) \leq \eta\}$ denotes the probability that a χ_N^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 Fig. 1. These are obtained by simulations. More precisely, the threshold is calculated using relation (8) and the probabilities of detection, obtained using an adequate number of simulations are compared in Fig. 1 to those obtained by assuming $L((y, z))$ to be a $\chi_N^2(\lambda^2)$ random variable. There are slight discrepancies between the observed and expected probabilities of detection.

Since we were careful enough and always respected condition (7) in our simulations, we made investigations to localize the origin of these discrepancies and noticed that the probability of false alarms obtained by simulations were also slightly in error, which means that the threshold given by (8) is slightly wrong and probably induces the differences observed on the probability of detection. Further analysis is now necessary to check (6).

In Fig. 2, we present the results obtained by simulations for 3 different P_{FA} 's and 3 values of α , namely $\alpha = 0, 0.5$ and 0.7 . Remember that for $\alpha = 0$, there are no uncertainties

on the signature and the proposed detector is then equivalent to the standard one [6]. The P_D 's decrease as the uncertainty on the signature increases, as might be expected.

Fig. 1. ROC for proposed method, P_{FA} equal to 10^{-k} for $k = 3$ to 5 and $\alpha^2 = 1/2$. The solid lines are the simulation results and the o' are theoretical values

Fig. 2. ROC for proposed method, P_{FA} equal to 10^{-2} , 10^{-4} and 10^{-6} and $\alpha = 0, .5$ and $.7$.

6. CONCLUSIONS

We have considered the problem of detecting and estimating a signal whose spatial signature is not perfectly known. Indeed we mainly investigated the detection part and developed the GLRT associated with our model of uncertainty on the signature. We have considered that the noise covariance matrix was known and further investigations are needed to alleviate this assumption. Another important issue that requires further analysis concerns the influence on the performance of a wrong choice of α which measures the degree of uncertainty on the signature. The simulations we performed seem to indicate that the loss of performance is negligible, but this remains to be established theoretically.

7. 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] M.N. Desai and R.S. Mangoubi, "Robust gaussian and non-gaussian matched subspace detection", *IEEE Trans. Signal Processing*, vol. 51, 12, pp. 3115-3127, Dec. 2003.
- [5] J.J. Fuchs, "A robust matched detector." submitted to *IEEE Trans. Signal Processing*, vol. 55, 11, pp. 5133-5142, Nov. 2007.
- [6] O. Besson, L.L. Scharf and F. Vincent, "Matched direction detectors and estimators for array processing with subspace steering vector uncertainties", *IEEE Trans. Signal Processing*, vol. 53, 12, pp. 4453-4463, Dec. 2005.
- [7] G.W. Stewart and J.G. Sun, *Matrix Perturbation Theory*. Academic Press, 1990.

8. APPENDIX

We develop a matrix perturbation result for a $(n, N+1)$ matrix $\hat{C}$ with $n < N + 1$, which is the sum of a rank-1 matrix C and a perturbation Δ . Let the SVD of the rank-1 matrix C be denoted as $C = U\Sigma_1V_1^T$ with U and Σ_1 square matrices of order n . We further introduce the following partitions

$$C = U\Sigma_1V_1^T = \sigma_1u_1v_1^T, \quad U = [u_1 \quad U_2], \quad V_1 = [v_1 \quad V_2]$$

From $\hat{C} = C + \Delta$, it follows

$$\begin{aligned} \hat{C}\hat{C}^T &= CC^T + C\Delta^T + \Delta C^T + \Delta\Delta^T \\ &= U\Sigma_1^2U^T + U\Sigma_1V_1^T\Delta^T + \Delta V_1\Sigma_1U^T + \Delta\Delta^T. \end{aligned}$$

We now partially and iteratively diagonalize this real symmetric matrix by pre- and post-multiplying it by orthogonal matrices. In a first step, we define $\hat{D}_1 = U^T\hat{C}\hat{C}^TU$:

$$\hat{D}_1 = \Sigma_1^2 + \Sigma_1V_1^T\Delta^TU + U^T\Delta V_1\Sigma_1 + U^T\Delta\Delta^TU$$

which we rewrite as $\hat{D}_1 = D + E_1 + E_2$ with $D = \Sigma_1^2$, which is already diagonal, E_1 representing the first-order terms, i.e., of order $\|\Delta\|$, and E_2 the second-order term. In a second step we will diagonalize E_1 , by using an orthogonal matrix U' of the form $U' = I + P + \frac{1}{2}P^2 + \dots$ with P a small (first-order) skew-hermitian matrix yet to be chosen. The matrix U' is then close to the identity matrix up to first order and $U'U'^T$ is close to the identity matrix up to third order. This transforms $\hat{D}_1$ into $\hat{D}_2 = U'^T\hat{D}_1U'$, which is equal to

$$\hat{D}_2 = D + D_1 + \left(E_2 - \frac{1}{2}PD_1 + \frac{1}{2}D_1P - \frac{1}{2}PE_1 + \frac{1}{2}E_1P \right) + \dots$$

with the first-order terms $D_1 = E_1 + DP - PD$ and the second-order terms within parenthesis. We will now choose P to *diagonalize* D_1 , the matrix $\hat{D}_2$ which is similar to $\hat{D}_1$ and $\hat{C}\hat{C}^T$ is then diagonal up to second-order terms. Let us partition the intervening matrices, similarly to

$$D = \Sigma_1^2 = \begin{bmatrix} \sigma_1^2 & 0 \\ 0 & 0 \end{bmatrix}.$$

We define

$$E_1 = \begin{bmatrix} 2\sigma_1 u_1^T \Delta v_1 & g^T \\ g & 0 \end{bmatrix}, \quad P = \begin{bmatrix} 0 & p^T \\ -p & P_2 \end{bmatrix}$$

with $g = \sigma_1 U_2^T \Delta v_1$ and $P_2^T = -P_2$. It is now possible to zero the first row and column of D_1 by taking $p = -(1/\sigma_1^2)g$ which is indeed small (of the first-order in Δ) as requested. We take P_2 , which is arbitrary so far since we only *diagonalize* the first eigenvalue $\hat{\sigma}_1^2$ equal to the zero matrix. Substituting for these values, one gets

$$\hat{D}_2 = \begin{bmatrix} \sigma_1^2 & 0 \\ 0 & 0 \end{bmatrix} + \begin{bmatrix} 2\sigma_1 u_1^T \Delta v_1 & 0 \\ 0 & 0 \end{bmatrix} + \begin{bmatrix} u_1^T \Delta \Delta^T u_1 + v_1^T \Delta^T U_2 U_2^T \Delta v_1 & u_1^T \Delta (I - v_1 v_1^T) \Delta^T U_2 \\ U_2^T \Delta (I - v_1 v_1^T) \Delta^T u_1 & U_2^T \Delta (I - v_1 v_1^T) \Delta^T U_2 \end{bmatrix} + \dots$$

One can then deduce from this relation that, up to the second-order terms, one has

$$\hat{\sigma}_1^2 \simeq \sigma_1^2 + 2\sigma_1 u_1^T \Delta v_1 + u_1^T \Delta \Delta^T u_1 + v_1^T \Delta^T U_2 U_2^T \Delta v_1$$

or equivalently

$$\hat{\sigma}_1^2 \simeq \sigma_1^2 + 2\sigma_1 u_1^T \Delta v_1 + u_1^T \Delta \Delta^T u_1 + v_1^T \Delta^T \Delta v_1 - (u_1^T \Delta v_1)^2$$

from which one deduces that $\hat{\sigma}_1^2/\sigma$ is approximatively a chi-square random variable with $N + n$ degrees of freedom and non-centrality parameter equal to σ_1^2/σ , where $\sigma_1^2 = (aa^T)(\mathbf{s}^T \mathbf{s}) + (a^T a)/\alpha^2$.