MULTICHANNEL DETECTION AND SPATIAL SIGNATURE ESTIMATION WITH UNCALIBRATED RECEIVERS

Amir Leshem[1,2] *and Alle-Jan van der Veen*[1]

[1] Delft University of Technology, Dept. Electrical Engineering/DIMES, 2628 CD Delft, The Netherlands
[2] Metalink Broadband Access, Yakum Business Park 60972, Israel

Abstract A problem occuring in radio astronomy is the detection and cancellation of spatially correlated interfering signals entering via the sidelobes of the telescopes in an array. A complicating factor is that the noise powers can be different at each telescope. For the case that the sensors are uncalibrated, we formulate the detection problem as a test on the covariance structure, state the GLRT for this problem, and relate it to a simpler ad-hoc detector. We derive algorithms to estimate the noise powers and the subspace of interferer signature vectors. Once the subspace is estimated, the interference can be projected out. We compare this method to the conventional multichannel subspace detector and show its robustness to non-identical channels on data collected with the Westerbork radio telescope.

1. INTRODUCTION

In this paper we study the detection and suppression of spatially correlated signals impinging on an array of uncalibrated non-identical sensors, in the presence of spatially uncorrelated noise. The noise covariance matrix is diagonal but otherwise unknown.

The motivation for this study comes from an application in radio astronomy, where we wish to detect and suppress man-made interfering sources impinging on an array of telescopes. The output of the receiver after processing is essentially a sequence of short-term (~10 second) sample correlation matrices, composed of the contributions of astronomical sources in the pointing direction, the additive receiver noise, and the interference. The receiver noise is largely independent among the sensors, but the receiver gains are not identical, with differences of up to a few dB. Until now, calibration of this has been done separately and taken into account off-line. An interfering source is usually in the near field and received through the side-lobes of the parabolic dishes, hence the received signals are correlated but with arbitrary unknown gains. Our aim is to detect and cancel the interference online; this requires online calibration processing as well.

Two types of interference play a role: intermittent signals (e.g., TDMA signals as in the GSM system, certain radar signals) and continuously present signals (e.g., television signals, GPS). Our approach for intermittent signals is to detect their presence on-line on milli-second periods, and discard those periods which are deemed contaminated (temporal excision) [1]. For continuous interference, we also wish to estimate the signature (direction) vector, so that we can project out that dimension from the data. This is more ambitious, and also requires modifications to the way the astronomical data is processed after recording [2]. Note that the astronomical signals of interest are much weaker than the receiver noise and hence it is necessary to detect interference even if it is much below the noise power. The astronomical signals themselves are too weak to be detected at these short time scales.

When the interferers are weaker than the system noise and the receivers are non-identical, the change in eigenstructure of the sample covariance matrix is not detectable unless one of two steps is taken. The first is pre-calibration and whitening. The second which is easier to implement on-line is to use a different model where the noise covariance matrix is assumed diagonal but not necessarily equal to $\sigma^2\mathbf{I}$, and to detect deviation from this nominal model. This is the approach taken here. The Generalized Likelihood Ratio Test (GLRT) for this problem turns out to be the determinant of the sample correlation matrix, a fact which is not very well known in signal processing but has been used for a long time in certain other disciplines.

We demonstrate the results of the excision using the GLRT detector and compare it to a detector which assumes identical receivers. We also demonstrate the improvement in the estimate of the spatial signature as compared to the usual eigendecomposition technique.

2. PROBLEM FORMULATION

Assume that we have a set of q narrow-band Gaussian signals impinging on an array of p sensors. The received signal can be described in complex envelope form by

$$\mathbf{x}(k) \;=\; \sum_{i=1}^{q}\mathbf{a}_i s_i(k)+\mathbf{n}(k) \;=\; \mathbf{A}\mathbf{s}(k)+\mathbf{n}(k) \qquad (1)$$

where $\mathbf{x}(k)=[x_1(k),\cdots,x_p(k)]^T$ is a $p\times 1$ vector of received signals at sample times k, $\mathbf{A}=[\mathbf{a}_1,\cdots,\mathbf{a}_q]$, where $\mathbf{a}_i$ is the array response vector for the i'th signal, $\mathbf{s}(k)=[s_1(k),\cdots,s_q(k)]^T$ is a $q\times 1$ vector of gaussian source signals at sample times k with covariance matrix $\mathbf{R}_s=E(\mathbf{s}\mathbf{s}^{\mathrm{H}})$, $\mathbf{n}(k)$ is the $p\times 1$ additive noise vector, which is assumed to have independent gaussian entries with unknown diagonal covariance matrix $\mathbf{R}_n=\mathrm{diag}\{\nu_1,\ldots,\nu_p\}$.

We would like to detect the presence of signals satisfying the above model, i.e., given data vectors $\mathbf{x}(1),\ldots,\mathbf{x}(N)$ decide whether $q=0$ or $q>0$. Secondly, if $q>0$, we would like to detect q and estimate the interfering subspace, i.e., $\mathrm{span}(\mathbf{A})$, so that we can project out this subspace from the data. We do not assume parametric knowledge of the array manifold (since the interferers enter in the side lobes) or a calibration of the noise power in each channel. Under these assumptions the only way to distinguish between signal and noise is to use the fact that the noise is spatially uncorrelated, hence has a diagonal covariance matrix.

The detection problem is thus given by a collection of hypotheses ($\mathcal{CN}(0,\mathbf{R})$ denotes the zero-mean complex normal distribution with covariance $\mathbf{R}$)

$$\begin{array}{ll}\mathcal{H}_q:\ \mathbf{x}(k)\sim\mathcal{CN}(0,\mathbf{R}_q) & \\ \mathcal{H}':\ \mathbf{x}(k)\sim\mathcal{CN}(0,\mathbf{R}'), & q=1,2,\cdots\end{array} \qquad (2)$$

where $\mathbf{R}_q$ is the covariance matrix of the model with q interferers,

$$\mathbf{R}_q = \mathbf{A}\mathbf{A}^{\mathrm{H}} + \mathbf{D}\,, \quad \text{where} \quad \mathbf{A}: \; p \times q, \quad \mathbf{D} \text{ diagonal}$$

and $\mathcal{H}'$ corresponds to a default hypothesis of an arbitrary (unstructured) positive definite matrix $\mathbf{R}'$. (Without loss of generality, we absorbed the interferer covariance matrix $\mathbf{R}_s$ in $\mathbf{A}$.)

As it turns out, this problem has been studied in the psychometrics, biometrics and statistics literature since the 1930s under the heading of *factor analysis* (but usually for real-valued matrices) [3, 4]. The problem has received much less attention in the signal processing literature. Related recent work includes e.g. direction estimation using two subarrays with mutually uncorrelated noise [5, 6].

3. THE GLRT DETECTOR

In this section we give a short derivation of the GLRT for the detection problem $\mathcal{H}_q$ versus $\mathcal{H}'$. Note that both hypotheses are composite and we have to derive maximum likelihood estimates of the parameters for each of the hypotheses. Under $\mathcal{H}_q$, the likelihood function is given by

$$L(\mathbf{X}|\mathcal{H}_q) \;\equiv\; L(\mathbf{X}|\mathbf{R}_q) \;=\; \left(\frac{1}{|\mathbf{R}_q|} e^{-\mathrm{tr}\left(\mathbf{R}_q^{-1}\hat{\mathbf{R}}\right)} \right)^N ,$$

where $\mathbf{X} = [\mathbf{x}(1), \ldots, \mathbf{x}(N)]$ and $\hat{\mathbf{R}} = \frac{1}{N}\sum_{k=1}^{N} \mathbf{x}(k)\mathbf{x}(k)^H$ is the sample covariance matrix, $|\cdot|$ denotes the determinant and $\mathrm{tr}(\cdot)$ the trace operator.

The ML estimate of $\mathbf{R}_q$ is found by maximizing $L(\mathbf{X}|\mathbf{R}_q)$ over the parameters of the model $\mathbf{R}_q = \mathbf{A}\mathbf{A}^{\mathrm{H}} + \mathbf{D}$, or equivalently the log-likelihood function

$$\mathcal{L}(\mathbf{X}|\mathbf{R}_q) \;=\; N\left(-\ln|\mathbf{R}_q| - \mathrm{tr}(\mathbf{R}_q^{-1}\hat{\mathbf{R}}) \right) .$$

Denote the estimate by $\hat{\mathbf{R}}_q = \hat{\mathbf{A}}\hat{\mathbf{A}}^{\mathrm{H}} + \hat{\mathbf{D}}$. Under $\mathcal{H}'$ we obtain that the ML estimate of $\mathbf{R}'$ is given by $\hat{\mathbf{R}}$, the sample covariance matrix. The log-likelihood GLRT test statistic is thus given by

$$\ln \frac{L(\mathbf{X}|\mathcal{H}_q)}{L(\mathbf{X}|\mathcal{H}')} \;=\; -N\big(\,\mathrm{tr}(\hat{\mathbf{R}}_q^{-1}\hat{\mathbf{R}}) - \ln|\hat{\mathbf{R}}_q^{-1}\hat{\mathbf{R}}| - p\, \big) .$$

A further result is that the ML estimate of $\hat{\mathbf{R}}_q$ is such that $\mathrm{tr}(\hat{\mathbf{R}}_q^{-1}\hat{\mathbf{R}}) = p$ so that we can base the test on

$$T_q(\mathbf{X}) \;:=\; N \ln|\hat{\mathbf{R}}_q^{-1}\hat{\mathbf{R}}| . \tag{3}$$

If we generalize the results in [3,4] to complex data, we obtain the following.

Lemma 3.1 *If $\mathcal{H}_q$ is true and N is moderately large (say $N - q \geq 50$), then $2T_q(\mathbf{X})$ has approximately a χ^2_ν distribution with $\nu = (p-q)^2 - p$ degrees of freedom.*

In view of results of Box and Bartlett, a better fit is obtained by replacing N in (3) by [3]

$$N' = N - \frac{1}{6}(2p+5) - \frac{2}{3}q .$$

This provides a threshold for a test of $\mathcal{H}_q$ versus $\mathcal{H}'$ corresponding to a desired probability of false alarm P_{FA}. The test replaces the more familiar eigenvalue test on the rank of $\hat{\mathbf{R}}$ in the case of white noise, $\mathbf{D} = \sigma^2\mathbf{I}$. Note that before we can perform the test, we need to compute the ML estimates of $\mathbf{A}: p \times q$ and $\mathbf{D}$ (see section 5).

4. TEST FOR DIAGONALITY

Under $\mathcal{H}_0$ we can make the test more explicit. To estimate $\hat{\mathbf{R}}_0 = \hat{\mathbf{D}}$, we set the derivative of $\mathcal{L}$ with respect to the parameters of $\mathbf{D}$ to zero, which immediately gives $\hat{\mathbf{D}} = \mathrm{diag}(\hat{\mathbf{R}})$. Therefore the GLRT test statistic is given by

$$\frac{L(\mathbf{X}|\mathcal{H}_0)}{L(\mathbf{X}|\mathcal{H}_1)} = \frac{|\hat{\mathbf{R}}|^N}{\prod_{i=1}^{p} \hat{\mathbf{R}}_{ii}^N} = |\hat{\mathbf{C}}|^N , \tag{4}$$

where $\hat{\mathbf{C}}$ is the sample correlation matrix given by $\hat{\mathbf{C}} = \mathbf{W}\hat{\mathbf{R}}\mathbf{W}$ and $\mathbf{W} = \mathrm{diag}\{\hat{r}_{11}, \ldots, \hat{r}_{pp}\}^{-1/2}$. Note that $0 \leq |\hat{\mathbf{C}}| \leq 1$, where equality to 1 is obtained asymptotically for $N \to \infty$ if $q = 0$. Thus, for a certain threshold $\gamma = \gamma(N)$ between 0 and 1, the GLRT is

$$T_1 \;\equiv\; |\hat{\mathbf{C}}| \quad \underset{\mathcal{H}_1}{\overset{\mathcal{H}_0}{\gtrless}} \quad \gamma \tag{5}$$

This result is identical to that in the real-valued case (see [4, p.137]). The expression is rather satisfactory since in the absence of sensor calibration data all the spatial information exists in the spatial correlation coefficients between the different sensors, and the GLRT suggests a proper way of combining these different correlations. It is also quite easy to implement and does not involve any eigenstructure computations. From lemma 3.1, under $\mathcal{H}_0$ we know that $-2N\ln|\hat{\mathbf{C}}|$ has asymptotically a chi-square distribution with $p^2 - p$ degrees of freedom. Again, a better asymptotic fit is obtained by replacing N by $N' = N - \frac{1}{6}(2p+11)$.

A related ad-hoc detector to which we can compare is based on the Frobenius-norm of the off-diagonal entries of $\hat{\mathbf{C}}$. Since the diagonal entries are equal to 1, it is equivalent to take the norm of $\hat{\mathbf{C}}$ itself, i.e.,

$$T_2 \;\equiv\; \|\,\hat{\mathbf{C}}\,\|_F \quad \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\gtrless}} \quad \gamma' \tag{6}$$

In fact, it is straightforward to prove that, for weak signals, the performance of this detector must be approximately equal to that of the GLRT. Indeed, for weak signals, the eigenvalues of $\hat{\mathbf{C}}$ are equal to $\lambda_i = 1 + \varepsilon_i$, for small ε_i. Note that $\mathrm{tr}(\hat{\mathbf{C}}) = p \;\Rightarrow\; \sum_i \lambda_i = p \;\Rightarrow\; \sum_i \varepsilon_i = 0$. We can write

$$\begin{aligned} & T_1 \;=\; \textstyle\prod_i \lambda_i \;=\; e^{\sum_i \ln\lambda_i} \\ \Rightarrow \quad & \ln(T_1) \;=\; \textstyle\sum_i \ln\lambda_i \;=\; \sum_i \varepsilon_i - \frac{1}{2}\varepsilon_i^2 + \mathcal{O}(\varepsilon_i^3) \\ & \qquad\;\; =\; -\textstyle\sum_i \frac{1}{2}\varepsilon_i^2 + \mathcal{O}(\varepsilon_i^3) \end{aligned}$$

whereas

$$\begin{aligned} & T_2^2 \;=\; \|\,\hat{\mathbf{C}}\,\|_F^2 \;=\; \textstyle\sum_i \lambda_i^2 \;=\; \sum_i 1 + 2\varepsilon_i + \varepsilon_i^2 \\ \Rightarrow \quad & -\tfrac{1}{2}(T_2^2 - p) \;=\; -\textstyle\sum_i \frac{1}{2}\varepsilon_i^2 \end{aligned}$$

Since a monotonic transformation of a test statistic does not change the outcome of the test if the threshold is modified accordingly,[1] the two detectors are equivalent up to third order. Computing the Frobenius-norm requires only $\mathcal{O}(p^2)$ operations, versus $\mathcal{O}(p^3)$ for the determinant test (implemented via a Cholesky factorization of $\hat{\mathbf{C}}$).

[1] Note that the decisions in (5) and (6) are opposite, hence the change of sign in the second transformation.

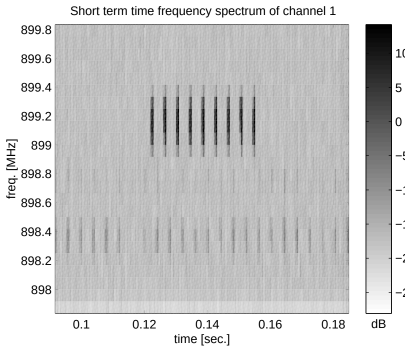

Figure 1. Time-frequency spectrum of channel 1, showing GSM interference

5. PARAMETER ESTIMATION

To enable the GLRT, we have to find ML estimates of the factors $\mathbf{A}: p \times q$ and $\mathbf{D}$, both dependent on the choice of q. The largest permissible value of q is that for which the number of degrees of freedom $v = (p-q)^2 - p \geq 0$, or $q \leq p - \sqrt{p}$. For larger q, there is no identifiability of $\mathbf{A}$ and $\mathbf{D}$: any sample covariance matrix $\hat{\mathbf{R}}$ can be fitted. Even for smaller q, $\mathbf{A}$ can be identified only up to a $q \times q$ unitary transformation at the right, i.e., we can identify $\text{span}(\mathbf{A})$. This generalizes the white noise case (where $\text{span}(\mathbf{A})$ would be given by the eigenvectors of $\hat{\mathbf{R}}$), and is sufficient for our application of interference cancellation.

For $q > 0$, there is no closed form solution to the estimation of the factors $\mathbf{A}$ and $\mathbf{D}$ in the ML estimation of $\hat{\mathbf{R}}_q = \hat{\mathbf{A}}\hat{\mathbf{A}}^{\text{H}} + \hat{\mathbf{D}}$. There are several approaches for this:

– Suppose that the optimal ML-estimate $\hat{\mathbf{D}}$ has been found. We can then whiten $\hat{\mathbf{R}}$ to $\tilde{\mathbf{R}} = \hat{\mathbf{D}}^{-1/2}\hat{\mathbf{R}}\hat{\mathbf{D}}^{-1/2}$, and similarly the model, giving $\tilde{\mathbf{R}}_q = \tilde{\mathbf{A}}\tilde{\mathbf{A}}^{\text{H}} + \mathbf{I}$. Note that $|\hat{\mathbf{R}}_q^{-1}\hat{\mathbf{R}}| = |\tilde{\mathbf{R}}_q^{-1}\tilde{\mathbf{R}}|$, which is the usual problem for white noise, solved via an eigenvalue decomposition of $\tilde{\mathbf{R}}$. This is equivalent to solving $\min \|\tilde{\mathbf{R}} - (\tilde{\mathbf{A}}\tilde{\mathbf{A}}^{\text{H}} + \mathbf{I})\|_F^2$. Since $\hat{\mathbf{D}}$ is not known, this leads to an iteration where $\tilde{\mathbf{A}}$ is plugged back, $\hat{\mathbf{D}}$ is estimated, etc.

  A related technique is alternating least squares, where we alternatingly minimize $\|\hat{\mathbf{R}} - \mathbf{A}\mathbf{A}^{\text{H}} + \mathbf{D}\|_F^2$ over $\mathbf{A}$ keeping $\mathbf{D}$ fixed, and over $\mathbf{D}$ keeping $\mathbf{A}$ fixed. (This is not equivalent to the determinant cost function unless a weighting by $\mathbf{D}^{-1/2}$ is used.) Both iterative techniques tend to be very slow.

– Gauss-Newton iterations on the original (determinant) cost function, or on the (weighted) least squares cost. This require an accurate starting point.

– Ad-hoc techniques for solving the least squares problem, possibly followed by a Gauss-Newton iteration. These techniques try to modify the diagonal of $\hat{\mathbf{R}}$ such that the modified matrix is low-rank q, hence can be factored as $\mathbf{A}\mathbf{A}^{\text{H}}$. For this we can exploit the fact that submatrices away from the main diagonal with $q+1$ columns have rank q. See [7] for an example with $q = 1$.

More details on estimation algorithms will appear in an extended version of this paper.

6. APPLICATION TO RADIO ASTRONOMY

The main motivation for the detection and subspace estimation problem stems from applications to interference mitigation in radio astronomy. We give two illustrations.

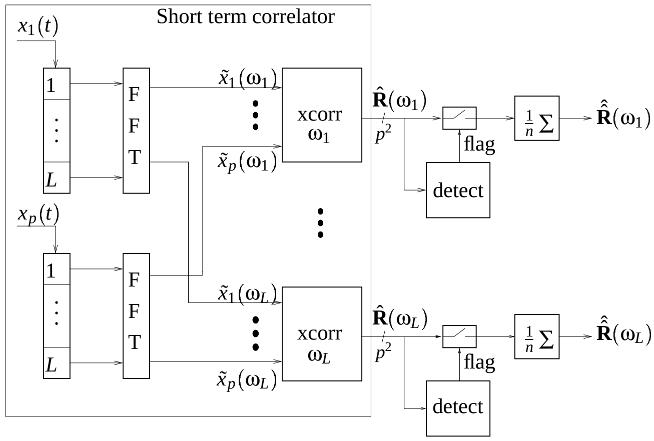

Figure 2. Computational structure of the blanking process

We first apply the detector for $\mathcal{H}_0$ to sample data collected with the Westerbork radio telescope. The data was recorded using the 8-channel NOEMI project data recorder [1]. We selected a bandwidth of 2 MHz, around 899 MHz, with a duration of 3 seconds. This band is contaminated with various GSM mobile telephony signals. Such signals are intermittent, occupying time slots of length 0.577 ms in frames of 4.6 ms. A segment of the data is shown in figure 1. The received data channels were split into subbands of 83 kHz by means of windowing and short-term FFTs, and subsequently correlated per frequency bin. Each covariance matrix is an average based on 21 samples and covers a period of 0.24 ms.

Our aim is to test for the presence of interference in each covariance matrix. Only if no interference is detected, the block is passed to a long-term correlator. Two detectors have been applied. The first is the detector of (4), and the other one is given by

$$T_3 \equiv \frac{|\hat{\mathbf{R}}|}{[\frac{1}{p}\text{tr}(\hat{\mathbf{R}})]^p}. \tag{7}$$

This detector is a GLRT assuming identical channels (or $\mathbf{D} = \sigma^2\mathbf{I}$) [4].

Since $N = 21$ is small, we have not used the theoretical thresholds. Instead, we have excised the worst 10 percent of the data at each frequency channel and generated spectral estimates by further averaging the covariance matrices of the remaining 90 percent of the data. The processing structure is shown in figure 2.

Figure 3 shows the power spectrum of channel 1 and the cross-spectrum of channels 1 and 3, respectively, before and after blanking. Without excision, we can see that several interfering signals are present, most weak but one rather strong. We can clearly see that while both detectors excised properly the strong interference, the detector based on the $\mathbf{D} = \sigma^2\mathbf{I}$ assumption failed to excise the weak features of the interference.

In a second application, we wish to spatially filter out continuously present interference. The approach is to estimate $\text{span}(\mathbf{A})$, and to apply a projector $\mathbf{P}_{\hat{\mathbf{A}}}^{\perp}$ onto the orthogonal complement of the span. Here, we describe only a limited-scope simulation on synthetic data, where we estimate a rank-1 subspace (i) using factor analysis, and for comparison (ii) using eigendecomposition assuming that $\mathbf{D} = \sigma^2\mathbf{I}$, or (iii) using eigendecomposition after whitening by $\mathbf{D}^{-1/2}$, assuming the true $\mathbf{D}$ is known from calibration. The algorithm used for factor analysis is a non-iterative ad hoc technique

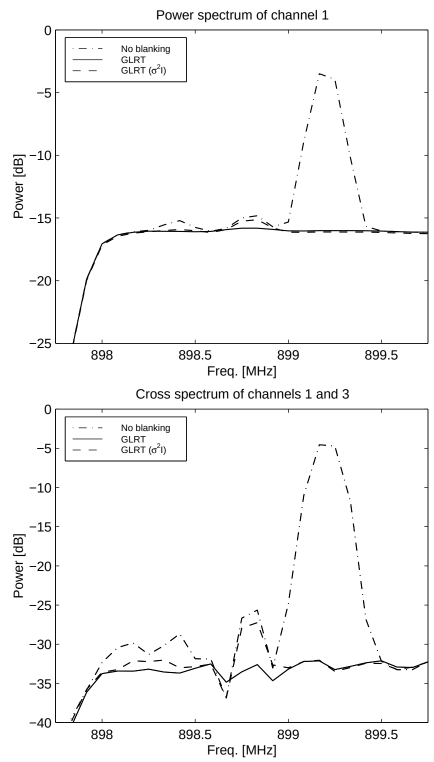

Figure 3. Power spectra and cross-spectra of channels 1 and 3, before and after interference excision

used to obtain a consistent initial estimate, followed by a Gauss-Newton optimization of the weighted least squares cost function (3 iterations). The weighting is by $\hat{\mathbf{D}}^{-1/2}$ as obtained from the ad hoc technique. We have generated covariance matrices based on the model (1) with $q = 1$, and show the residual interference power after projection, i.e., $||\mathbf{P}_{\hat{\mathbf{a}}}^{\perp}\mathbf{a}||$ as a function of number of samples N, mean noise power, and deviation in noise power. The noise powers are randomly generated at the beginning of the simulation, uniformly in an interval. Legends in the graphs indicate the nominal noise power and the maximal deviation. All simulations use $p = 8$ sensors and $q = 1$ interferer, and a nominal interference to noise ratio per channel of 0 dB.

The results are shown in figure 4. The first graph shows the residual interference power for varying maximal deviations, the second graph shows the residual for varying number of samples N, and a maximal deviation of 3 dB of the noise powers. The figures indicate that already for small deviations of the noise powers it is essential to take this into account. Furthermore, the estimates from the factor analysis are nearly as good as can be obtained via whitening with known noise powers.

Acknowledgement

We are grateful to L.L. Scharf for pointing us to the topic of factor analysis and sharing his notes on this.

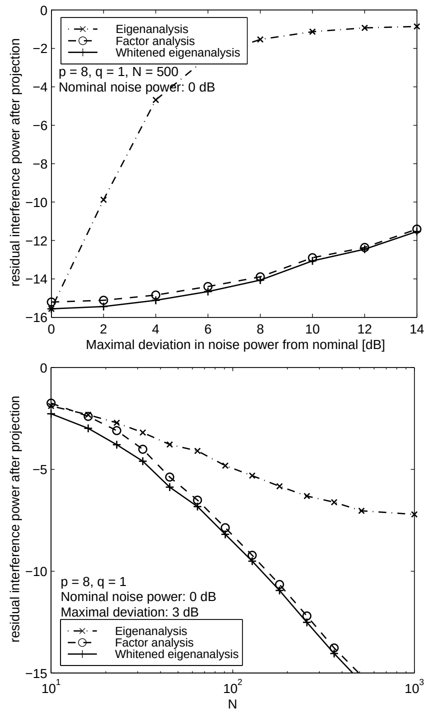

Figure 4. Residual interference power after projections

REFERENCES

[1] A. Leshem, A-J. van der Veen, and A-J. Boonstra, "Multichannel interference mitigation techniques in radio astronomy," *The Astrophysical Journal Supplements*, Nov. 2000.

[2] A. Leshem and A-J. van der Veen, "Radio-astronomical imaging in the presence of strong radio- interference," *IEEE Trans. Informat. Th.*, pp. 1730–1747, August 2000.

[3] D.N. Lawley and A.E. Maxwell, *Factor Analysis as a Statistical Method.* Butterworth and Co, 1963.

[4] K.V. Mardia, J.T. Kent, and J.M. Bibby, *Multivariate Analysis*. Academic Press, 1979.

[5] Q. Wu and K.M. Wong, "UN-MUSIC and UN-CLE: An application of generalized correlation analysis to the estimation of the direction of arrival of signals in unknown correlated noise," *IEEE Trans. Signal processing*, vol. 42, pp. 2331–2343, Sept. 1994.

[6] P. Stoica and M. Cedervall, "Detection tests for array processing in unknown correlated noise fields," *IEEE Trans. Signal Processing*, vol. 45, pp. 2351–2362, Sept. 1997.

[7] A.J. Boonstra and A.J. van der Veen, "Gain decomposition methods for radio astronomy," in *Submitted to Proc. Workshop IEEE Stat. Signal Proc.*, (Singapore), Aug. 2001.