DESIGNING GOOD ESTIMATORS FOR LOW SAMPLE SIZES: RANDOM MATRIX THEORY IN ARRAY PROCESSING APPLICATIONS

Xavier Mestre

Centre Tecnològic de Telecomunicacions de Catalunya (CTTC)
Nexus I building, c/Gran Capità, 2-4, 2a planta, Barcelona, 08034 Spain
URL: http://www.cttc.es/, E-Mail: xavier.mestre@cttc.es

ABSTRACT

Traditional signal processing architectures are usually designed to perform well in large sample size situations, i.e. when the number of observations increases to infinity while their dimension remains fixed. In practice, though, these algorithms must work with a relatively low number of samples, and this degrades their performance significantly. This paper proposes the use of general statistical analysis (a branch of random matrix theory) as a systematic approach to derive signal processing architectures that have an excellent performance even when the number of samples and their dimension have the same order of magnitude. The basic rationale is to provide estimators that are consistent when both the number of samples and their dimension increase without bound at the same rate. We demonstrate the usefulness of the approach deriving an estimator of the (asymptotically) optimum loading factor in a minimum variance beamformer for combating the finite sample size effect.

1. INTRODUCTION

Random matrix theory is a branch of multivariate statistics that deals with the asymptotic behavior of the spectrum of random matrices as their dimensions increase without bound. It is today well known that, for some random Hermitian matrix models, the empirical distribution of eigenvalues tends almost surely to a non-random function as the two dimensions of the matrix are driven to infinity at the same rate. This fact has been extensively used in the physics literature, and quite recently the theory has found its natural application to the asymptotic modelling and performance evaluation of certain large communications systems, such as CDMA systems with a large number of users and spreading factors, or the characterization of radiocommunication systems with multiple antennas at both the transmitter and the receiver.

So far, the main application of random matrix theory has been focused on the asymptotic analysis of large communications systems, assuming that two of the dimensions of the system are driven to infinity at the same rate. The rationale behind the contributions in that direction follows a two step procedure. First, the performance measure that needs to be characterized is described in terms of the eigenvalues (and sometimes also eigenvectors) of certain random matrices. Thereafter, random matrix theory results that describe the spectral behavior of such matrices are applied in order to obtain a closed-form asymptotic expression for the quantity that needs to be characterized. An interesting application of these type of analysis to statistical array processing is the performance evaluation of estimators and architectures in finite sample size situations. Assuming that the number of observations and the number of elements of the array are both large and have the same order of magnitude, one can sometimes derive closed form expressions for the asymptotic performance of certain techniques based on array observations. This is in stark contrast to traditional asymptotic performance characterization studies, that assume that the observation dimension remains fixed while the number of observations is asymptotically large. Two-dimensional limits constitute a better characterization of estimators that work on finite sample supports, where the number of available samples has the same order of magnitude as the observation dimension.

In this paper, we take a different point of view to the finite sample size problem in array processing applications. Rather than analyzing the behavior of architectures that have been designed to perform well under the infinite sample size situation, we raise the question of whether it is possible to *design* such techniques in order to perform asymptotically well when the number of observations and their dimension have the same order of magnitude. In other words, we look for a systematic way of designing estimators (and, in general, any type of array processing architecture) that are consistent, not only when the number of observations increases without bound, but also when the observation dimension is asymptotically large as well. In practice, one observes that architectures that have these property need a much lower number of samples to converge to an acceptable performance. Since the asymptotic limits are two-dimensional by nature, random matrix theory techniques seem to be the appropriate background for the development of these doubly consistent architectures. Indeed, General Statistical Analysis, a theory founded by V.L. Girko in the late 80's [1, 2] is a branch of random matrix theory that provides a general framework for deriving estimators that are consistent even when the number of estimated parameters increases at the same rate as the number of observations.

2. RANDOM MATRIX THEORY AND GENERAL STATISTICAL ANALYSIS

In this section we present the basic rationale behind general statistical analysis and its application to array processing problems. We will concentrate ourselves on the application of this theory in the design of good estimators based on the covariance matrix, denoted from now on as $\mathbf{R} \in \mathbb{C}^{M \times M}$, where M is the observation dimension (typically the number of sensors/antennas, if we are dealing with array observations). Let $\lambda_1 \geq \ldots \geq \lambda_M > 0$ and $\mathbf{e}_1, \ldots, \mathbf{e}_M$ denote the M eigenvalues and associated eigenvectors of $\mathbf{R}$, assumed positive definite. We start with the definition of a generic spectral function of the eigenvalues of the covariance matrix $\mathbf{R}$, i.e.

$$F(\lambda) = \sum_{k=1}^{M} \varphi_k \mathbb{I}_{\{\lambda_k \leq \lambda\}},$$

This work was partially funded by the Spanish Ministry of Science and Technology (MCYT) under projects PU-2002-56, TIC2002-04594-C02-02, HU2002-0032, FIT-070000-2003-257 and the European Comission IST-2001-38835 ANWIRE.

where $\mathbb{I}_{\{\lambda_k \leq \lambda\}}$ is the indicator function for the event $\{\lambda_k \leq \lambda\}$ and φ_k, $k = 1 \ldots M$, are weighting values that have certain regularity properties (cf. [2]). Note that if we fix $\varphi_k = M^{-1}$, $k = 1 \ldots M$, then $F(\lambda)$ becomes the empirical distribution of the eigenvalues of $\mathbf{R}$, i.e. the eigenvalue counting function that, for each λ, gives the percentage of eigenvalues of $\mathbf{R}$ that are lower than or equal to λ. We define the real Stieltjes transform[1] associated with the spectral function $F(\lambda)$ as

$$m(x) = \frac{1}{M} \sum_{k=1}^{M} \frac{\varphi_k}{1 + x\lambda_k}, \; x \in \mathbb{R}^+. \quad (1)$$

Observe that, if we fix $\varphi_k = M^{-1}$ for all k, the Stieltjes transform $m_M(x)$ turns out to be equal to

$$s(x) = m(x)|_{\varphi_k = M^{-1}} = \frac{1}{M} \operatorname{tr}\left[(\mathbf{I}_M + x\mathbf{R})^{-1}\right], \quad (2)$$

whereas, if we set $\varphi_k = \mathbf{a}^H \mathbf{e}_k \mathbf{e}_k^H \mathbf{b}$, where $\mathbf{a}$ and $\mathbf{b}$ are two $M \times 1$ complex vectors and $(\cdot)^H$ denotes transpose conjugate, we will have

$$t(x; \mathbf{a}, \mathbf{b}) = m(x)|_{\varphi_k = \mathbf{a}^H \mathbf{e}_k \mathbf{e}_k^H \mathbf{b}} = \mathbf{a}^H \left[\mathbf{I}_M + x\mathbf{R}\right]^{-1} \mathbf{b}. \quad (3)$$

Now, there are a lot of quantities in communications and signal processing applications that can be expressed in terms of Stieltjes transforms of the type (2) and (3). For example, the (i,j)th entry of the inverse correlation matrix $\mathbf{R}^{-1}$ can be expressed as

$$\left\{\mathbf{R}^{-1}\right\}_{i,j} = \lim_{x \to \infty} x \mathbf{u}_i^H \left[\mathbf{I}_M + x\mathbf{R}\right]^{-1} \mathbf{u}_j = \lim_{x \to \infty} x t(x; \mathbf{u}_i, \mathbf{u}_j),$$

where $\mathbf{u}_i$ is an all-zeros $M \times 1$ column vector with a 1 in the ith position. On the other hand, quadratic forms such as $\mathbf{a}^H \mathbf{R}^k \mathbf{b}$, with $k \in \mathbb{Z}$ and $\mathbf{a}$, $\mathbf{b}$ two $M \times 1$ column vectors, arise quite naturally in spectral estimation applications. These quantities can also be expressed in terms of the Stieltjes transform $t(x; \mathbf{a}, \mathbf{b})$. Indeed, if $k > 0$,

$$\mathbf{a}^H \mathbf{R}^k \mathbf{b} = \frac{(-1)^k}{k!} \left[\frac{\partial^k}{\partial x^k} \mathbf{a} \left[\mathbf{I}_M + x\mathbf{R}\right]^{-1} \mathbf{b}\right]_{x=0}.$$

Other relationships can be found for successive powers of the inverse covariance matrix (i.e. $k < 0$).

Let us assume that we want to estimate a given quantity q that depends on the covariance matrix $\mathbf{R}$ through different combinations of Stieltjes transforms such as the ones shown above. The covariance matrix $\mathbf{R}$ is generally unknown and must be estimated from the observed data, denoted by $\mathbf{x}(1), \ldots, \mathbf{x}(N)$ (all of them independent $M \times 1$ column vectors). Traditional approaches would simply replace the covariance matrix $\mathbf{R}$ with the corresponding sample estimate

$$\hat{\mathbf{R}} = \frac{1}{N} \sum_{i=1}^{N} \mathbf{x}(i) \mathbf{x}^H(i).$$

If observations are circularly symmetric, independent and identically distributed (i.i.d.) with zero mean and covariance matrix $\mathbf{R}$, one can model the sample covariance matrix as $\hat{\mathbf{R}} = \mathbf{R}^{1/2} \mathbf{U} \mathbf{U}^H \mathbf{R}^{1/2}$ with $\mathbf{R}^{1/2}$ the positive Hermitian square root of $\mathbf{R}$ and $\mathbf{U}$ an $M \times N$ matrix with i.i.d. symmetric entries having zero mean and variance $1/N$. Now, even though $\hat{\mathbf{R}}$ is a consistent estimation of $\mathbf{R}$ as the number of observations N increases without bound while their dimension M remains constant (because $\mathbf{U}\mathbf{U}^H \to \mathbf{I}_M$ in probability), that estimation might not be the best option when N and M have the same order of magnitude. For example, one can show that as $N, M \to \infty$ at the same rate[2], $\left|\mathbf{a}\hat{\mathbf{R}}^{-1}\mathbf{b} - (1-c)^{-1} \mathbf{a}\mathbf{R}^{-1}\mathbf{b}\right| \to 0$ in probability, where $c = M/N$. Hence, $\mathbf{a}\hat{\mathbf{R}}^{-1}\mathbf{b}$ is clearly an inconsistent estimator of $\mathbf{a}\mathbf{R}^{-1}\mathbf{b}$ when $N, M \to \infty$ at the same rate. Note that, instead of estimating $\mathbf{R}^{-1}$ with $\hat{\mathbf{R}}^{-1}$, one could have used $(1-c)\hat{\mathbf{R}}^{-1}$ as an estimator for $\mathbf{R}^{-1}$. This second estimator is consistent regardless of whether M scales up with N or not, and consequently will have a better performance than the traditional counterpart in finite sample size situations. This type of modification can be generalized to more complicated estimators, using the theory of General Statistical Analysis.

Assume as before that the quantity q that needs to be estimated can be expressed in terms of Stieltjes transforms of the type shown above. The basic rationale behind General Statistical Analysis is the fact that, in order to find a consistent estimator of q as $N, M \to \infty$ (number of observations and their dimension increase without bound at the same rate), one must only find a uniformly consistent estimator of the generic Stieltjes transform $m(x)$ under the same asymptotic conditions. Girko [1, 2] proved that, for the model of $\hat{\mathbf{R}}$ considered here, such an estimator exists and is given by[3]

$$\hat{m}(x) = \frac{1}{M} \sum_{k=1}^{M} \frac{\varphi_k}{1 + \theta(x) \hat{\lambda}_k}, \quad (4)$$

where $\hat{\lambda}_k$ are the eigenvalues of the sample covariance matrix $\hat{\mathbf{R}}$ and the function $\theta(x)$ is the unique positive solution to the following equation

$$\theta(x) \left[1 - c + c\frac{1}{M} \operatorname{tr}\left[\left(\mathbf{I}_M + \theta(x)\hat{\mathbf{R}}\right)^{-1}\right]\right] = x, \; x > 0. \quad (5)$$

From this basic estimator of the real generic Stieltjes transform in (1) one can construct estimators of more complicated quantities that are consistent even when the observation dimension increases to infinity with the sample size. Moreover, these estimators have the following nice properties (see further [2]):

1. They are derived without any assumption on the actual distribution of the observations (other than zero mean, bounded moments and circularity) and are only based on the inner structure of the random matrix $\hat{\mathbf{R}}$.
2. If the sample size increases and the observation dimension remains constant ($c \to 0$), the estimator reverts to its "traditional" counterpart. Note that $\theta(x) \to x$ when $c \to 0$ in (5).
3. Under some mild regularity conditions, $\hat{m}(x)$ is asymptotically (as $M, N \to \infty$) Gaussian-distributed [1].

Next, we give an application example in array signal processing that demonstrates the usefulness of general statistical analysis: the determination of the asymptotically optimum loading factor for combating the finite sample size effect in minimum variance (MV) beamformers.

[1] This definition is a bit different from the classical complex-valued Stieltjes transform usually employed in random matrix theory. This definition will be more convenient for the presentation of General Statistical Analysis.

[2] This only holds under some regularity conditions on $\mathbf{a}$, $\mathbf{b}$.

[3] This estimator is generally referred to as "G_2-estimator" when used to estimate (2), or "G_{25}-estimator" when used to estimate (3). Note that the estimator given here is valid when the sample covariance matrix has the structure presented above.

3. MV BEAMFORMING AND ASYMPTOTIC OUTPUT SINR WITH DIAGONAL LOADING

Spatial reference beamforming techniques are filtering architectures that exploit the knowledge of the angular information of the desired source in order to enhance the response towards a desired direction of arrival (DOA) while nulling out the contribution from interfering components [3]. Basically, the spatial filters are designed to minimize the output power while preserving a certain response towards de desired DOA; hence the name "minimum variance" beamformers. If $\mathbf{s}_d \in \mathbb{C}^{M\times 1}$ is a column vector containing the spatial signature of the desired signal, and $\hat{\mathbf{R}} \in \mathbb{C}^{M\times M}$ denotes the sample spatial covariance matrix obtained from N snapshots, the minimum variance beamformer is designed as $\mathbf{w} = \hat{\mathbf{R}}^{-1}\mathbf{s}_d$ (up to a constant value). We assume that the number of observations is higher than the number of elements of the array, i.e. $N > M$.

MV beamformers are known to suffer from two different impairments that might cause a severe performance degradation in practical implementations. First, an imprecise knowledge of the spatial signature of the desired signal (caused, for example, by lack of accuracy in the DOA information or small calibration errors across the array), might potentially lead to desired signal cancellation effects that corrupt the performance of the MV beamformer, sometimes making it even worse than traditional phased-array spatial filter. Second, if the number of samples available at the receiver is not sufficiently high, the response of the spatial filter can be very different from the desired one, and consequently the filter might randomly enhance and suppress the response in the wrong DOAs. The degradation due to this effect is particularly important in arrays with a large number of elements, where the number of snapshots needed to achieve convergence of the filter weights is usually too high.

Diagonal loading (i.e. adding a constant to all the elements of the diagonal of the sample covariance matrix) has long been used as a method to improve the robustness of the MV beamformer against these two types of impairments. Indeed, it has been shown that diagonal loading is the natural extension of classical MV beamformers under quadratic constraints or incorporating some degree of uncertainty of the steering vector (see [4, 5, 6, 7] and references therein). In parallel with that, diagonal loading has also been used to improve the performance of the MV beamformer under finite sample size situations [8, 9, 10]. This second application of diagonal loading is far less explored, and results on the optimum choice of the loading factor for combating the finite sample size effect are in fact very scarce [4, p.751]. This has motivated the use of rather adhoc methods for fixing the loading factor in low sample size situations. For instance, in [4, p.748] the author suggests setting the loading factor 10dB above the minimum eigenvalue of the sample correlation matrix, while in [11] the diagonal load is fixed equal to the standard deviation of the diagonal entries of the sample covariance matrix.

In a recent paper [12], an asymptotic expression of the output SINR of a diagonally loaded beamformer when both the sample size (N) and the number of antennas (M) increase without bound at the same rate was presented[4]. Note that, because the ratio M/N is constant, the asymptotic expression is a good approximation of the non-asymptotic reality. Assuming $M/N = c$ so that $0 < c < 1$, the asymptotic output SINR had the same asymptotic behavior (in probability) as the deterministic quantity $\overline{SINR} = (q(\alpha) - 1)^{-1}$, where

$$q(\alpha) = \frac{1}{1 - c\xi(\alpha)} \frac{\mathbf{s}_d^H (\mathbf{R} + \gamma\mathbf{I}_M)^{-1}\mathbf{R}(\mathbf{R} + \gamma\mathbf{I}_M)^{-1}\mathbf{s}_d}{P_s\left(\mathbf{s}_d^H(\mathbf{R} + \gamma\mathbf{I}_M)^{-1}\mathbf{s}_d\right)^2}, \quad (6)$$

$$\xi(\alpha) = \frac{1}{M}\sum_{k=1}^{M}\left(\frac{\lambda_k}{\lambda_k + \gamma}\right)^2$$

and $\gamma = \alpha[1 + cb]$, with b the unique positive solution to the following equation

$$b = \frac{1}{M}\sum_{k=1}^{M}\frac{\lambda_i[1 + cb]}{\lambda_i + \alpha[1 + cb]}, \quad (7)$$

being $\lambda_1 \ldots \lambda_N$ the eigenvalues of the true covariance matrix $\mathbf{R}$ as before. A reasonable way of fixing the loading factor would be to choose the value of α that minimizes (6). In practice, however, the covariance matrix $\mathbf{R}$ is unknown, and therefore so are $q(\alpha)$ and $\overline{SINR}$. Hence, in order to give an estimation of the asymptotically optimum loading factor α, one must first derive an estimation of either one of these functions of α. A traditional approach would be to replace $\mathbf{R}$ with $\hat{\mathbf{R}}$ and the true eigenvalues $\lambda_1 \ldots \lambda_N$ with the eigenvalues of the sample covariance matrix, denoted by $\hat{\lambda}_1 \ldots \hat{\lambda}_N$. However, since $\hat{\mathbf{R}}$ and $\hat{\lambda}_1 \ldots \hat{\lambda}_N$ are only consistent estimators of their true counterparts when $N \to \infty$ while M remains constant, the obtained estimator will not be well behaved in situations where both parameters (N and M) have the same order of magnitude (namely, the situation where diagonal loading is actually useful!). Next, we provide an estimator that is consistent even when $M, N \to \infty$ at the same rate. Because of this property, this estimator will give very good results even when M and N have the same order of magnitude.

In order to derive such an estimator, observe that one can express the function $q(\alpha)$ as an arithmetic combination of the different Stieltjes transforms and their derivatives. Using (4) as an estimator of these Stieltjes transforms, one can get to the following expression for the consistent estimation (as $M, N \to \infty$) of the asymptotically optimum loading factor α (see [15] for details):

$$\hat{\alpha} = \arg\min_{\alpha} \hat{q}(\alpha) \quad (8)$$

$$\hat{q}(\alpha) = \frac{1}{1 - c\varphi(\alpha)} \frac{\mathbf{s}_d^H\left(\alpha\mathbf{I}_M + \hat{\mathbf{R}}\right)^{-1}\hat{\mathbf{R}}\left(\alpha\mathbf{I}_M + \hat{\mathbf{R}}\right)^{-1}\mathbf{s}_d}{P_s\left(\mathbf{s}_d^H\left(\alpha\mathbf{I}_M + \hat{\mathbf{R}}\right)^{-1}\mathbf{s}_d\right)^2}$$

where

$$\varphi(\alpha) = 2 - 2\alpha\frac{1}{M}\operatorname{tr}\left[\left(\alpha\mathbf{I}_M + \hat{\mathbf{R}}\right)^{-1}\right] + \\ - c\left(\frac{1}{M}\operatorname{tr}\left[\hat{\mathbf{R}}\left(\alpha\mathbf{I}_M + \hat{\mathbf{R}}\right)^{-1}\right]\right)^2.$$

Hence, in order to obtain an appropriate estimation of the asymptotically optimum loading factor, one must first evaluate the function $\hat{q}(\alpha)$ in (8) and search for its global minimum (one-dimensional search). Despite its complicated appearance, the estimator can be expressed in very simple terms as a combination of the eigenvalues of the sample covariance matrix and the product of its eigenvectors with the spatial signature $\mathbf{s}_d$.

[4] To the best of our knowlege, neither the distribution nor the expectation of the output SINR of the diagonally load beamformer with perfect steering and finite suport have been derived. Only some asymptotic approximations as $N \to \infty$ for fixed M can be found in the literature [13, 14].

4. NUMERICAL VALIDATION

To illustrate the performance of the presented estimator, we consider here a scenario with five directional narrowband sources with received power 20dB above the noise floor, impinging on a uniform linear array of 10 elements separated half a wavelength apart. In order to create variability in the scenario, we generated the directions of arrival of the different users as independent random variables uniformly distributed on $[-60^o, 60^o]$. Figure 1 represents the cumulative distribution function of the SINR at the output of a MV beamformer for different diagonal loading methods. Method 1 fixes the diagonal load as 10 times the smallest eigenvalue of $\hat{\mathbf{R}}$ as proposed in [4, p.748], and Method 2 sets the diagonal load equal to the standard deviation of the diagonal elements of $\hat{\mathbf{R}}$ as proposed in [11]. We give simulation results for $N = 20$ and $N = 200$ snapshots. Observe that the proposed method gives significant gains in terms of output SINR with respect to both Method 1 and Method 2 in the performance region of interest. These gains are especially high (of up to 4dB) in the low sample size situation. Of course, these gains come at the expense of higher complexity in the estimation of the loading factor. Note also that there is a small region around 3dB where the proposed method is slightly outperformed by the other two; the loss is, however, insignificant compared to the high gain that can be obtained at more reasonable values for the output SINR. For the sake of clarity, we do not show in Figure 1 the performance of the estimator that fixes the loading factor by direct minimization of $q(\alpha)$ in (6), replacing the true values with their sample estimates without applying General Statistical Analysis. It can be seen [15] that such an estimator gives completely wrong values for the optimum loading factor, and is widely outperformed by the other three. This illustrates the usefulness of General Statistical Analysis in situations where the sample size and the observation dimension have the same order of magnitude.

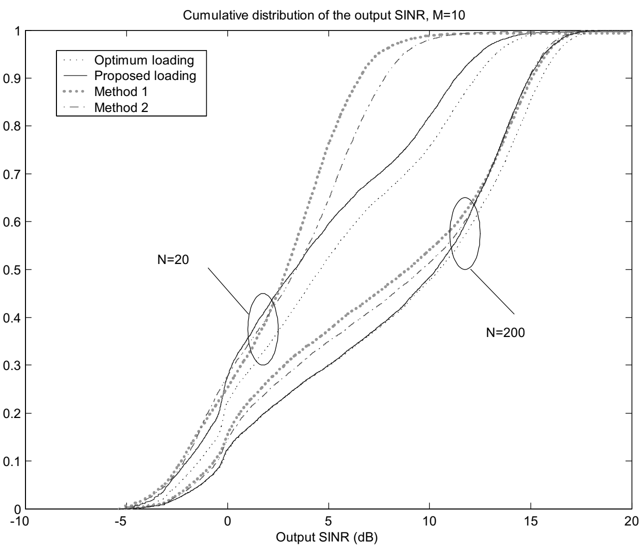

Figure 1: Distribution function of the output SINR in an array of $M = 10$ elements with different methods of diagonal loading and different values of the number of samples.

5. CONCLUSIONS

In this paper, we propose the use of General Statistical Analysis as a tool for designing array processing architectures and estimators that are consistent even when the observation dimension increases at the same rate as the simple size. This property guarantees their good behavior when the number of samples has the same order of magnitude as the number of elements of the array. To illustrate the usefulness of the approach, we have derived a new estimator of the (asymptotically) optimum loading factor in MV beamformers for perfectly known desired signal spatial signature operating under finite sample size restrictions. Simulation results illustrate high gains with respect to traditional approaches, especially in low sample size situations.

REFERENCES

[1] V. Girko, “Statistical analysis of observations of increasing dimension,” 1995.

[2] V. Girko, *An Introduction to Statistical Analysis of Random Arrays*. The Netherlands: VSP, 1998.

[3] R. Monzingo and T. Miller, *Introduction to Adaptive Arrays*. John Wiley and Sons, 1980.

[4] H. V. Trees, *Optimum Array Processing*. New York: John Wiley and Sons, 2002. Part IV of "Detection, Estimation and Modulation Theory".

[5] S. Vorobyov, A. Gershman, and L. Zhi-Quan, “Robust adaptive beamforming using worst-case performance optimization: A solution to the signal mismatch problem,” *IEEE Transactions on Signal Processing*, vol. 51, pp. 313 – 324, Feb. 2003.

[6] J. Li, P. Stoica, and Z. Wang, “On robust Capon beamforming and diagonal loading,” *IEEE Transactions on Signal Processing*, vol. 51, pp. 1702–1715, 2003.

[7] S. Shahbazpanahi, A. Gershman, Z.-Q. Luo, and K. M. Wong, “Robust adaptive beamforming for general-rank signal models,” *IEEE Transactions on Signal Processing*, vol. 51, pp. 2257 – 2269, Sept. 2003.

[8] W. Gabriel, “Using spectral estimation techniques in adaptive processing antenna systems,” *IEEE Transactions on antennas and propagation*, vol. 34, pp. 291–300, Mar. 1986.

[9] H. Cox, R. Zeskind, and M. Owen, “Robust adaptive beamforming,” *IEEE Transactions on acoustics, speech and signal processing*, vol. ASSP-35, pp. 1365–1376, Oct. 1987.

[10] B. Carlson, “Covariance matrix estimation errors and diagonal loading in adaptive arrays,” *IEEE Trans. on Aerospace and Electronic Systems*, vol. 24, pp. 397–401, July 1988.

[11] N. Ma and J. Goh, “Efficient method to determine diagonal loading value,” in *Proc. of the IEEE ICASSP 2003 (Vol. V)*, pp. 341–344.

[12] X. Mestre and M. Lagunas, “On diagonal loading for minimum variance beamformers,” in *Proc. of the IEEE International Symposium on Signal Processing and Information Technology*, (Darmstadt, Germany), 2003.

[13] M. Ganz, R. Moses, and S. Wilson, “Convergence of the SMI and the diagonally loaded SMI algorithms with weak interference,” *IEEE Transactions on AP*, vol. 38, pp. 394–399, Mar. 1990.

[14] L. Fertig, “Statistical performance of the MVDR beamformer in the presence of diagonal loading,” in *Proc. of the IEEE Sensor Array and Multichannel Signal Processing Workshop*, pp. 77 –81, Mar. 2000.

[15] X. Mestre and M. Lagunas, “Diagonal loading effect on the asymptotic output SINR of the minimum variance beamformer,” Tech. Rep. CTTC/RC/2003-01, CTTC, 2003. Available at http://www.cttc.es/drafts/cttc-rc-2003-001.pdf.