# CONFIDENCE ASSESSMENT FOR SPECTRAL ESTIMATION BASED ON ESTIMATED COVARIANCES

*Johan Karlsson*[⋆]*, Per Enqvist*[⋆]*, and Ather Gattami*[†]

[⋆] KTH Royal Institute of Technology, Department of Mathematics
[†] Bitynamics Research Stockholm, Sweden

## ABSTRACT

In probability theory, time series analysis, and signal processing, many identification and estimation methods rely on covariance estimates as an intermediate statistics. Errors in estimated covariances propagate and degrade the quality of the estimation result. In particular, in large network systems where each system node of the network gather and pass on results, it is important to know the reliability of the information so that informed decisions can be made.

In this work, we design confidence regions based on covariance estimates and study how these can be used for spectral estimation. In particular, we consider three different confidence regions based on sets of unitarily invariant matrices and bound the eigenvalue distribution based on three principles: uniform bounds; arithmetic and harmonic means; and the Marčenko-Pastur Law eigenvalue distribution for random matrices. Using these methodologies we robustly bound the energy in a selected frequency band, and compare the resulting spectral bound from the respective confidence regions.

***Index Terms***— Covariance Estimation, Random Matrices, Spectral Estimation, Confidence Regions, Marčenko-Pastur Law.

## 1. INTRODUCTION

In today's society, many systems use sensors to collect a huge amount of information and thereby creating massive data sets. For the data to be processed efficiently, it needs to be handled in a decentralized fashion where the contextual information is extracted and propagated in the system. This information is then used for higher level analyses, and knowledge of the quality of this information is crucial to ensure that such analyses produce accurate and reliable results [1].

For these systems, statistical methods are key tools for estimating features and characteristics as well as for building relevant mathematical models. Often such methods are based on covariance estimates, as for example in spectral estimation, wireless channel estimation, radar, sonar [2, 3], medical imaging, and identification of network structures [4, 5]. Estimation of covariance estimates is a classical subject [6], but which has gained significant attention recently. In particular, it has gained interest in applications where a structural constraint is imposed, e.g., due to shift invariance or sensor geometries [7, 8, 9, 10]. Another such application area is analysis of network structures in dynamical systems using graphical models [11, 12]. This topic has also gained renewed interest from sparse signal processing applications where large inverse covariances are estimated from a limited set of data, see, e.g., [13].

With a few exceptions such as [14, 15], the focus on covariance estimation research has been on determining nominal covariance estimates that is optimal in some sense, e.g., the maximal likelihood solution. However, covariance estimates are inherently uncertain and errors in the covariance degrade the accuracy of the result. It therefore crucial to assess the uncertainty of the covariance estimates in order to guarantee the quality of the result. The sample covariance $\hat{R}$ is Wishart distributed under the assumption that the underlying stochastic vector is Gaussian. We use this to design confidence regions for the true covariance $R$, and to study the resulting uncertainty for spectral estimation and graphical modelling. A key observation is that the distribution of $R^{-1/2}\hat{R}R^{-1/2}$ is independent of the true covariance $R$ [15].

Section 2 describes some of the notation used. In Section 3 we introduce relevant background material and construct confidence regions. Section 4 describes the spectral estimation problem and discusses optimization problems for computing the maximum and minimum energy in an interval. In Section 5 we apply confidence regions for spectral estimation in order to compute bounds on the frequency content of a signal. In Section 6 we sum up the results and discuss further directions.

## 2. NOTATION

Let $X^*$ denote the conjugate transpose of $X$. The identity matrix is denoted by $I$. Here $\mathcal{H}_n \subset \mathbb{C}^{n\times n}$ denotes the set of Hermitian matrices, $\mathcal{H}_n^+ \subset \mathbb{C}^{n\times n}$ denotes the set of positive

This work was supported in part by Ericsson Research, ACCESS Linnaeus Centre, and NSF CCF-1218388. Ather Gattami was with Ericsson Research when the research was conducted.
Email: johan.karlsson@math.kth.se, penqvist@kth.se, atherg@gmail.com.

semidefinite Hermitian matrices, and $\mathcal{H}_n^{++} \subset \mathbb{C}^{n\times n}$ denotes the set positive definite Hermitian matrices. Further, $X \succeq 0$ means that $X \in \mathcal{H}_n^{+}$ and $X \succ 0$ means that $X \in \mathcal{H}_n^{++}$.

## 3. COVARIANCE ESTIMATION, THE WISHART DISTRIBUTION, AND CONFIDENCE REGIONS

Let $x_k \in \mathbb{C}^{n\times 1}$, for $k = 1, \ldots, N$, be zero mean, independent, circulant complex Gaussian distributed random vectors with covariance

$$R = \mathbf{E}(x_k x_k^*),$$

where we assume that $R \succ 0$ and $N \geq n$. A fundamental problem in statistics is to estimate the covariance $R$ from a sample of vector realizations of $\{x_k\}_{k=1}^N$. The sample covariance

$$\hat{R} = \frac{1}{N}\sum_{k=1}^{N} x_k x_k^*$$

is the maximum likelihood estimator and a sufficient statistic for $R$ [16], and its sample value[1] $\hat{R}_{\text{sample}}$ is therefore often taken as an estimate of $R$. The sample covariance is a sum of outer products of independent identically distributed Gaussian vectors, and hence is distributed according to the Wishart distribution $\mathcal{CW}_n(N, R/N)$ [16, 17]. The probability density of the joint distribution of $\hat{R}_{\ell\ell}$ for $1 \leq \ell \leq n$ (that is of $\mathfrak{Re}(\hat{R}_{j\ell})$ and $\mathfrak{Im}(\hat{R}_{j\ell})$ for $1 \leq j < \ell \leq n$) is

$$p_{\mathcal{CW}_n(N,R/N)}(\hat{R}) = \frac{N^{Nn}|\hat{R}|^{N-n}e^{-N\mathrm{Tr}(R^{-1}\hat{R})}}{\pi^{n(n-1)/2}\,|R|^N \prod_{j=1}^{n}(N-j)!}.$$

As observed by Lang in [15], the distribution of the random matrix $Y = R^{-\frac{1}{2}}\hat{R}R^{-\frac{1}{2}}$ is independent of the true covariance matrix $R$. In fact, the distribution of $Y$ is Wishart distributed $\mathcal{CW}_n(N, I/N)$. Consequently, events based on outcomes of $Y$ have probabilities independent of the true covariance matrix $R$, and confidence regions for $R$ may be constructed based on the random vector $Y$. In fact, such regions only depend on the parameters $n, N$, and the confidence level $p$. There are many possible ways to construct such regions, and the best choice may depend on the application at hand. We will consider a few such examples.

### 3.1. Confidence regions

Here we will consider confidence regions, being sets of covariances of the form

$$\Omega_\Delta(\hat{R}) = \{\hat{R}^{1/2} V \hat{R}^{1/2} \,|\, V \in \Delta\},$$

where $\Delta \subset \mathcal{H}_n^+$ is a fixed unitarily invariant set. By selecting such sets, the probability

$$\mathbf{Pr}(R \in \Omega_\Delta(\hat{R})) \tag{1}$$

is independent of the true covariance matrix $R$. Therefore, by designing $\Delta$ appropriately we can construct confidence regions for $R$. In addition, by letting $\Delta$ be convex and closed, $\Omega_\Delta(\hat{R})$ is convex and closed as well, which allows for using convex optimization methods to optimizing over the set $\Omega_\Delta(\hat{R})$ and finding a global optimum.

We consider three such sets. Firstly we consider sets $\Delta$ with uniform bounds on the eigenvalues, i.e., on the form

$$\mathcal{U}_{\alpha,\beta} := \{V \in \mathcal{H}_n^+ \,|\alpha I \preceq V \preceq \beta I\}$$

where $0 < \alpha < 1 < \beta$.

Secondly we consider sets with bounds on the arithmetic mean and the harmonic mean of the eigenvalues (AH), i.e., sets on the form

$$\mathcal{F}_{\alpha,\beta} = \{V \in \mathcal{H}_n^+ \,|\mathrm{Tr}(V) \leq n/\alpha, \mathrm{Tr}(V^{-1}) \leq n/\beta\}$$

where $0 < \alpha, \beta < 1$. These confidence regions were proposed in [15] where they show that this construction gives rise to convex confidence regions.

Thirdly, we consider the smallest closed convex set that contains all the matrices with eigenvalue distributions between two given distributions.

$$\mathcal{A}_{\boldsymbol{\alpha},\boldsymbol{\beta}} = \mathrm{convhull}\{V \in \mathcal{H}_n^+ \,|\, \mathrm{eig}(V) \in [\boldsymbol{\alpha}, \boldsymbol{\beta}]\}.$$

where $\boldsymbol{\alpha}, \boldsymbol{\beta} \in \mathbb{R}_+^n$ and where we with $[\boldsymbol{\alpha}, \boldsymbol{\beta}]$ with denote the set of vectors $\boldsymbol{\gamma} \in \mathbb{R}^n$ such that $\boldsymbol{\alpha}_\pi \leq \boldsymbol{\gamma} \leq \boldsymbol{\beta}_{\pi'}$ for some permutations $\boldsymbol{\alpha}_\pi$ and $\boldsymbol{\beta}_{\pi'}$ of $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$, respectively. Let $\lambda_k(V)$ denote the $k$:th largest eigenvalue of $V$.

**Proposition 1.** *Let $\boldsymbol{\alpha}, \boldsymbol{\beta} \in \mathbb{R}_+^n$ be non-ascending vectors satisfying $\boldsymbol{\alpha} \leq \boldsymbol{\beta}$. Then $\mathcal{A}_{\boldsymbol{\alpha},\boldsymbol{\beta}}$ is given by*

$$\Big\{V \in \mathcal{H}_n^+ \,|\quad \sum_{j=k}^{n}\alpha_j \leq \sum_{j=k}^{n}\lambda_j(V), \text{ and } \sum_{j=1}^{k}\lambda_j(V) \leq \sum_{j=1}^{k}\beta_j, \text{ for } k = 1\ldots n\Big\}.$$

*Proof outline.* To see this, first note that all the constraints are convex [18, Corollary 2.4] since "sum of the k largest elements of a vector" and "-sum of the $k$ smallest elements of a vector" are permutation invariant and convex functions. Next note that

$$\mathcal{A}_{\boldsymbol{\alpha},\boldsymbol{\beta}} = \{V \in \mathcal{H}_n^+ \,|\, \mathrm{eig}(V) \in \mathfrak{A}\}$$

where $\mathfrak{A}$ is the smallest permutation invariant closed convex set in $\mathbb{R}^n$ that contains $[\boldsymbol{\alpha}, \boldsymbol{\beta}]$ [18] (cf. [19]). Further, using the sepatation theorem in an analogous way as in the proof Theorem 1 a. in [20] such sets can be characterized in terms of their partial sums of the largest/smallest values, leading to the sought characterization. □

In our setting, we will select $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ to be lower and upper approximations of the asymptotic eigenvalue distribution when $N, n \to \infty$ and $N/n$ is fixed, i.e., the distribution

[1] Throughout this paper $R$ is a fixed but unknown matrix, $\hat{R}$ is a random variable, and $\hat{R}_{\text{sample}}$ is a sample from $\hat{R}$.

corresponding to the Marčenko-Pastur (MP) Law (see, e.g., [21]).

**Theorem 2** ([21]). *Consider $X \in \mathbb{C}^{n\times N}$ with independent entries $X_{j,k}$ with mean zero and variance 1. As $n, N \to \infty$ with $n/N \to d \in (0,1]$ the empirical spectrum of $Y = XX^*/N$ converges weakly and almost surely to the density $f_d$ given by*

$$f_d(\omega) = \frac{\chi_{[a,b]}}{2\pi d\omega}\sqrt{(b-\omega)(\omega-a)}$$

*where $a = (1-\sqrt{d})^2$ and $b = (1+\sqrt{d})^2$.*

To be more specific, we seek to model the distribution of $\hat{R}^{-1/2}R\hat{R}^{-1/2} = V$ in order to design $\Delta$ appropriately and so that (1) holds with a given probability. Next, we note that the eigenvalues of $\hat{R}^{-1/2}R\hat{R}^{-1/2}$ are the inverse of those of $Y = R^{-\frac{1}{2}}\hat{R}R^{-\frac{1}{2}}$, hence $\Delta$ is naturally modelled using the eigenvalue distribution corresponding to the inverse of the eigenvalues distribution of the Marčenko-Pastur Law, i.e., $f_d(1/\omega)/\omega^2$. Therefore, we let $\boldsymbol{\alpha} \in \mathbb{R}^n$ and $\boldsymbol{\beta} \in \mathbb{R}^n$ be the discrete approximations of this inverse distribution, and shifted along the frequency axis with the value $\pm\rho$, where $\rho$ is selected so that (1) is satisfied with the given confidence level for the specific $n, N$ in the problem setup. This construction is motivated the by fact that the distribution converge in probability and the asymptotic bounds in [21, Theorem 8.22].

## 4. SPECTRAL ENERGY CONFIDENCE INTERVALS

Spectral estimation is a methodology for estimating the frequency content of a signal, and here we consider the problem of estimating the energy in a given frequency band. This is a useful quantity for several applications and can be used, e.g., in radar for determining if a target is in a given area or in communication for detecting if a frequency band is used [22, 3]. The key quantities here are the maximum and the minimum possible energy in the given frequency band that is consistent with a covariance confidence region [23]. Bounds on the frequency content will be useful if sufficiently many covariances are estimated with sufficient accuracy. However, if the frequency band is too narrow or the covariance error is too large, then the uncertainty will be large and the estimate will consequently not be reliable.

Let $\Phi(\theta)$ denote the spectrum, i.e., the non-negative function[2] on $\mathbb{T} = [-\pi, \pi]$ that represents the spatial energy or frequency content of a discrete-time signal (see, e.g., [22, Chapter 2]), and let $r_k$ denote its covariances

$$r_k = \frac{1}{2\pi}\int_{-\pi}^{\pi}\Phi(\theta)e^{-ik\theta}d\theta \text{ for } k = 0, \pm1, \pm2\ldots. \quad (2)$$

Note that since the spectrum $\Phi(\theta)$ is real, $r_{-k} = \bar{r}_k$ for $k \in \mathbb{Z}$, hence we may use only positive indices.

[2]A power spectrum is in general a non-negative bounded measure and may contain, e.g., spectral lines.

The upper and lower bound of the energy in a given interval $[\theta_0, \theta_1]$ can be obtained as the optimum of the linear optimization problem

$$\begin{aligned} \inf / \sup \quad & \int_{\theta_0}^{\theta_1}\Phi(\theta)d\theta \quad (3)\\ \text{subject to} \quad & r_k = \int_{-\pi}^{\pi}\Phi(\theta)e^{-ik\theta}d\theta, \quad k = 0, 1, \ldots, n,\\ & R = [r_{k-\ell}]_{k,\ell=0}^{n} \in \Omega_\Delta(\hat{R}_{\text{sample}}), \end{aligned}$$

where $\Omega_\Delta(\hat{R}_{\text{sample}})$ is the confidence region and $\hat{R}_{\text{sample}}$ is the estimated covariance matrix.

In practice a nominal spectrum is often estimated and used as if it was the true spectrum. However, the accuracy of such estimate depends to a high extent on the accuracy of the estimated covariances, the number of estimated covariances, and the true covariance sequence. Errors in the estimated covariances clearly degrades the solution. The number of estimated covariances relate to the resolution of the estimated spectrum. However, the covariance sequence can also affect the resolution. In particular, if the Toeplitz matrix $R$ is close to singular, then the set of spectra consistent with $R$ is small (cf. [24, Corollary 4]).

The optimization problem (3) can be computed by solving the dual finite dimensional problem using semidefinite programming. However, here we simply consider the approximation obtained by discretizing the interval $[-\pi, \pi]$.

## 5. APPLICATION TO SPECTRAL ESTIMATION

Here we consider a scenario where we seek to determine the total energy in a frequency band with a given confidence level. We will study the bounds and compare the uncertainty sets using the three types of confidence regions. For the implementation we use CVX, a package for specifying and solving convex programs [25].

We consider a spectrum $\Phi(\theta)$ consisting of three spectral lines

$$\Phi(\theta) = \delta(\theta-1) + \frac{1}{2}\delta(\theta-1) + \frac{1}{4}\delta(\theta-3/4) + \epsilon,$$

where $\delta$ denotes the Dirac delta function, and we compare the two cases $\epsilon = 0.5$ and $\epsilon = 0.01$. Consider the problem of bounding the maximum and minimum energy in intervals of length $\pi/10$ given the estimated sample covariance $\hat{R}_{\text{sample}}$ with $n = 10$. Further we assume that an covariance is estimated based on 250 snapshots. The three confidence regions

$$\Omega_{\mathcal{U}_{\alpha,\beta}}(\hat{R}_{\text{sample}}),\ \Omega_{\mathcal{F}_{\alpha,\beta}}(\hat{R}_{\text{sample}}),\ \Omega_{\mathcal{A}_{\boldsymbol{\alpha},\boldsymbol{\beta}}}(\hat{R}_{\text{sample}}),$$

are created using all the methodologies in subsection 3.1. The parameters are selected so that the confidence level is 95%. Note that the parameters does not depend on the sample covariance $\hat{R}_{\text{sample}}$.

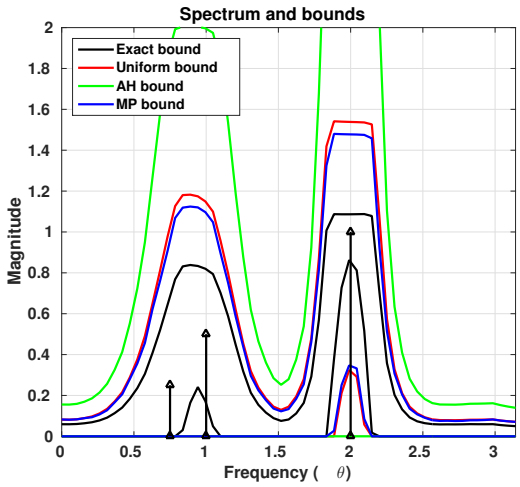


**Fig. 1**. Here the spectral lines are depicted together with upper and lower bounds on the energy in intervals of length $\pi/10$. Here $N = 250, n = 10$, Noise level: $\epsilon = 0.5$. Exact bound is the spectral uncertainty assuming the true covariance matrix is known.

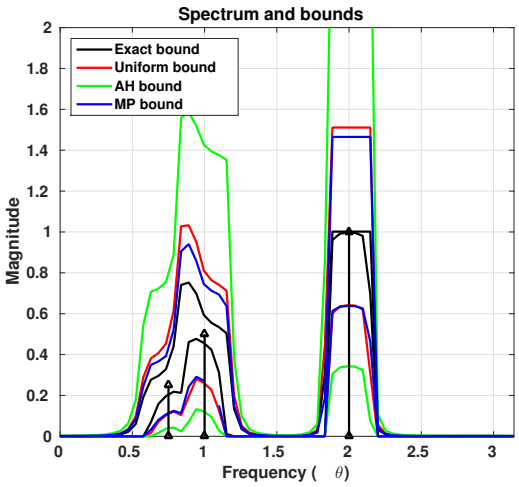


**Fig. 2**. Here the spectral lines are depicted together with upper and lower bounds on the energy in intervals of length $\pi/10$. Here $N = 250, n = 10$, Noise level: $\epsilon = 0.01$.

Figures 1 and 2 depict the energy bounds for the two cases $\epsilon = 0.5$ and $\epsilon = 0.01$, respectively. It can be seen that the uncertainty in the energy are considerably larger in the case with uncertain covariance compared to the case with known covariance. However, the localization of the energy is quite similar in the sense that the energy content is limited to the same energy band. Note that in this case, $\hat{R}_{\text{sample}}$ is not Toeplitz, and no spectrum $\Phi(\theta)$ is consistent with $\hat{R}_{\text{sample}}$. Instead the black lines denote the bounds corresponding to the true covariance $R$. In this example, the confidence region base on the Marčenko-Pastur Law gives slightly better bounds than confidence region based on uniform bounds. Both of those are substantially tighter than the bounds based on the arithmetic and harmonic means of the eigenvalues.

Here, it is important to note that with confidence level $0.95$, the energy bounds hold for all the frequencies. This is in contrast with traditional methods where the distribution of individual frequency points are derived (see, e.g., Periodogram analysis in Chapter 2.4.2 [3]).

## 6. CONCLUSIONS AND FURTHER DIRECTIONS

In this work, we quantify covariance uncertainty using confidence regions and use them for robust estimation and identification. We consider covariance uncertainty sets with a given confidence level in order to quantify and bound estimation errors. In particular, we consider spectral estimation based on covariance estimates. Uncertainty in the covariance estimates give rise to uncertainty in the respective application. We determine confidence regions for the covariance estimates, and thereby allow for uncertainty assessment.

The main contribution in this article is to propose a new type of convex confidence regions for covariance matrices based on the Marčenko-Pastur Law eigenvalue distribution for random matrices. A next step in this analysis is to consider the design in more detail and to asymptotically compute the size of the tube around the Marčenko-Pastur Law distribution in order to achieve a given confidence level for given sizes $n, N$. Even though we consider Gaussian stochastic variables in this paper, most results hold for general distribution given that the variance is constant and the samples independent (see Theorem 2 and other results from [21]).

This methodology can be used in many applications, e.g., identification of graph structures (for static graphs or dynamical systems) and in portfolio optimization using Markowitz models. We will consider such applications in a forthcoming paper. We also seek to extend this framework to cases where a time series is available, instead of independent snapshots as assumed in this analysis.

## 7. REFERENCES

[1] J. Manyika, M. Chui, B. Brown, J. Bughin, R. Dobbs, C. Roxburgh, A.H. Byers, and McKinsey Global Institute, "Big data: The next frontier for innovation, competition, and productivity," 2011.

[2] J.P. Burg, "Maximum entropy spectral analysis," *Proceedings of the 37th Meeting Society of Exploration Geophysicists*, October 1967.

[3] P. Stoica and R.L. Moses, *Spectral Analysis of Signals*, Prentice-Hall, Upper Saddle River, NJ, 2005.

[4] A.P. Dempster, "Covariance selection," *Biometrics*, vol. 28, pp. 157–75, 1972.

[5] V. Chandrasekaran, P.A. Parillo, and A.S. Willsky, "Latent variable graphical model selection via convex optimization," *The Annals of Statistics*, vol. 40, no. 4, pp. 1935–1967, 2012.

[6] A.K. Kshirsagar, *Multivariate analysis*, Dekker, 1972.

[7] J.P. Burg, D.G. Luenberger, and D.L. Wenger, "Estimation of structured covariance matrices," *Proceedings of the IEEE*, vol. 70, no. 9, pp. 963–974, September 1982.

[8] H. Li, P. Stoica, and J. Li, "Computationally efficient maximum likelihood estimation of structured covariance matrices," *IEEE Transactions on Signal Processing*, vol. 47, no. 5, pp. 1314–1323, 1999.

[9] T.T. Georgiou, "Structured covariances and related approximation questions," in *Directions in Mathematical Systems Theory and Optimization*, pp. 135–140. Springer, 2003.

[10] J. Karlsson and P. Enqvist, "Input-to-state covariances for spectral analysis: The biased estimate," in *Proc. of Int. Symp. Mathematical Theory of Network and Systems, MTNS*, 2012.

[11] E. Avventi, A.G. Lindquist, and B. Wahlberg, "ARMA identification of graphical models," *Automatic Control, IEEE Transactions on*, vol. 58, no. 5, pp. 1167–1178, 2013.

[12] M. Zorzi and R. Sepulchre, "AR identification of latent-variable graphical models," *arXiv preprint arXiv:1405.0027*, 2014.

[13] P. Ravikumar, M.J. Wainwright, G. Raskutti, and B. Yu, "High-dimensional covariance estimation by minimizing $\ell 1$-penalized log-determinant divergence," *Electronic Journal of Statistics*, vol. 5, pp. 935–980, 2011.

[14] A.B. Baggeroer, "Confidence intervals for regression (MEM) spectral estimates," *Information Theory, IEEE Transactions on*, vol. 22, no. 5, pp. 534–545, 1976.

[15] S. Lang, "Confidence regions for spectral bounds," in *Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing*, March 1984, vol. 9, pp. 577–580.

[16] N.R. Goodman, "Statistical analysis based on a certain multivariate complex gaussian distribution (an introduction)," *Annals of mathematical statistics*, pp. 152–177, 1963.

[17] R. Couillet and M. Debbah, *Random matrix methods for wireless communications*, Cambridge University Press Cambridge, MA, 2011.

[18] A. S. Lewis, "Convex analysis on the hermitian matrices," *SIAM Journal on Optimization*, vol. 6, no. 1, pp. 164–177, 1996.

[19] F. John, "On symmetric matrices whose eigenvalues satisfy linear inequalities," *Proceedings of the American Mathematical Society*, pp. 1140–1145, 1966.

[20] L Mirsky, "Results and problems in the theory of doubly-stochastic matrices," *Probability Theory and Related Fields*, vol. 1, no. 4, pp. 319–334, 1963.

[21] Z. Bai and J. W. Silverstein, *Spectral analysis of large dimensional random matrices*, vol. 20, Springer, 2010.

[22] D. H. Johnson and D. E. Dudgeon, *Array signal processing: concepts and techniques*, Prentice Hall, 1993.

[23] S. Lang and T. Marzetta, "A linear programming approach to bounding spectral power," in *IEEE International Conference on Acoustics, Speech, and Signal Processing*, 1983, vol. 8, pp. 847–850.

[24] J. Karlsson and T. Georgiou, "Uncertainty bounds for spectral estimation," *IEEE Transactions on Automatic Control*, vol. 58, no. 7, pp. 1659–1673, July 2013.

[25] CVX Research Inc., "CVX: Matlab software for disciplined convex programming, version 2.0 beta," http://cvxr.com/cvx, September 2012.