# Covariance Estimation in Elliptical Models with Convex Structure

Ilya Soloveychik and Ami Wiesel

Selim and Rachel Benin School of Computer Science and Engineering, Hebrew University of Jerusalem, Israel

ilya.soloveychik@mail.huji.ac.il, amiw@cs.huji.ac.il

***Abstract*—We develop the General Method of Moments (GMM) Approach for estimating the covariance matrices of non-Gaussian distributions with convex structure. The GMM turns out to be a non-convex optimization problem, thus making the addition of prior knowledge in form of convex structure constraints cumbersome. We propose a different approach to this estimator and show that the Tyler's estimator can be obtained as a solution of a convexly relaxed GMM problem, thus making the imposition of convex constraints easier. This new framework provides consistent solutions which outperform the standard projection methods. As an application of this method we consider Gaussian Compound samples with Toeplitz and banded covariance matrices. We provide synthetic numerical data and demonstrate the performance advantages of our method.**

***Index Terms*—Elliptical distribution, Tyler's scatter estimator, Generalized Method of Moments, non-Gaussian constrained covariance estimation.**

## I. Introduction

Covariance matrix estimation is a fundamental problem in the field of statistical signal processing. Many other algorithms for detection and inference rely on accurate covariance estimates [1, 2]. The problem is well understood in the Gaussian unstructured case. But becomes significantly harder when the underlying distribution is non-Gaussian, for example in elliptical distributions, and when there is prior knowledge on the structure. In this paper, we propose a unified framework for covariance estimation in elliptical distributions with general convex structure.

Over the last years there was a great interest in covariance estimation with known structure. The motivation to these works is that in many modern applications the dimension of the underlying distribution is large and there are not enough samples to estimate it correctly. The prior information on the structure reduces the degrees of freedom in the model and allows accurate estimation with a small number of samples. This is clearly true when the structure is exact, but also when it is approximate due to the well known bias-variance tradeoff. Prior knowledge on the structure can originate from the physics of the underlying phenomena, e.g., [3, 4, 5, 6], or from similar datasets, e.g., adjacent cells in radar systems [7]. When the structure is defined using a convex set, a natural and computationally efficient solution is to project the naive unstructured estimators onto this set.

Many covariance structures are easily represented in a convex form. Probably the most classical convex covariance structure is the Toeplitz model. It arises naturally in the analysis of stationary time series which are used in a wide range of applications in many fields including radar imaging, target detection, speech recognition, and communication systems, [3, 4, 8]. Toeplitz matrices are also used to model the correlation of cyclostationary processes in periodic time series [9]. In many applications the number of parameters can be reduced, thus making the covariance matrix sparse. A popular convex sparse model is banded covariance, which represents reduction in statistical relation between random variables [10]. Another important example of sparse structure is the SPICE estimator, which was proposed in [6] to treat high-dimensional arrays processing problems, where the covariance structure is approximated by a low-dimensional linear combination of rank one matrices. An additional kind of convex structure is low-rank covariance approximation, which is actually a regularization of non-robust models [11]. In the decades, all of these structures have been successfully considered in the Gaussian case [10, 12].

In a different line of works, there have been observed an increasing interest in robust covariance estimation for non-Gaussian distributions [13]. Significant attention is being paid to the family of elliptical distributions, which includes as particular cases generalized Gaussian distribution, compound Gaussian and many other [14]. The elliptical models are used to measure radar clutter [15], noise and interference in indoor and outdoor mobile communication channels [16] and other problems. The theory of robust covariance M-estimators was mostly developed by Maronna [17]. Following him Tyler [18] have proposed a robust scatter estimator which has become widely used in the last decades [13, 19, 20]. One of the most prominent disadvantages of these methods is that the estimator is given as a solution to a non-convex optimization problem, thus making imposition of additional constraints rather difficult. One of the options to cure this obstacle is geodesic convexity. It has been recently shown that different M-estimators are geodesically convex, which significantly simplifies their treatment [21]. But still, if one wants to impose an additional constraint on the scatter matrix it must be formulated in a form of a geodesically convex set, which is not always possible, see for example [22].

In the present work we derive COCA - COnvexly ConstrAined Covariance Matching estimator. It is based on the principle of Generalized Method of Moments (GMM) [23]. It searches for a covariance of a given convex structure that minimizes the norm of a simple moment's identity. This identity is in fact the optimality condition of the successful

978-1-4799-2893-4/14/$31.00 ©2014 IEEE

Tyler's estimate. COCA tries to simultaneously satisfy this condition while constraining the structure. Unfortunately, this requires the solution to a high dimensional non-convex minimization. Instead, we propose a relaxation and express COCA as a standard convex optimization with linear matrix inequalities which can be computed using off-the-shelf numerical solvers. Interestingly, we prove two promising results. First, in the unconstrained case, COCA is tight and identical to Tyler's estimate. This result basically "convexifies" Tyler's estimate. Second, in the structured case, COCA is asymptotically tight and hence consistent. Finally, we demonstrate the finite sample advantages of COCA over existing methods using synthetic numerical simulations.

The paper is organized in the following way. First, we formulate the problem and briefly describe the common ways of solving it: Tyler's estimator and GMM. We then propose convex relaxation of the GMM problem and show that the solution of the relaxed problem coincides with the Tyler's estimator. After this we prove that adding convex structure does not affect consistency. Finally, we provide numerical examples and applications showing the performance advantages of the proposed method.

We denote by $\mathcal{P}(p)$ the closed cone of symmetric positive semi-definite $p \times p$ matrices. We consider below the real case for simplicity. Exactly the same reasoning applies to convex case, with transposition replaced by Hermitian conjugation.

## II. Problem Formulation

Consider a $p$ dimensional, zero mean, elliptically distributed random vector $\mathbf{x}$. Such a vector can be represented as [14]

$$\mathbf{x} = r\Lambda\mathbf{u}, \tag{1}$$

where $\mathbf{u}$ is a $k$ dimensional random vector, uniformly distributed on the unit hypersphere, $r$ is a nonnegative random variable, $\Lambda \in \mathbb{R}^{p \times k}$ [14]. The random variable $r$ is called the generating variate of $\mathbf{x}$ and it is stochastically independent of $\mathbf{u}$. We assume that the distribution of this random variable is unknown.

The parameter $\mathbf{C}^{\text{True}} = \Lambda\Lambda^T$ is referred to as the dispersion or shape matrix of $\mathbf{x}$ and coincides with its covariance matrix (up to a scaling factor). In many applications, it is common to assume prior information on the structure of this matrix. In particular, we assume that it belongs to a known convex subset $\mathcal{S} \subset \mathcal{P}(p)$. Typical examples of such subsets are:

- **Toeplitz**: In stationary time series, the covariance between the $i$-th and the $j$-th components depend only on the the difference $|i-j|$. Such kind of processes is encountered very often in many engineering areas including statistical signal processing, radar imaging, target detection, speech recognition, and communications systems, [3, 4, 8, 9, 24, 25].
- **Banded**: A natural approach to covariance modeling is to formulate the reduction in statistical relation using the notion of independence or correlation, which corresponds to sparsity in the covariance matrix [10]. Assuming that $i$-th element of the random vector is uncorrelated with the $j$-th if $|i-j| > k$ leads to $k$-banded structure, also known as time varying moving average models.
- **Low rank**: One of the most common covariance models involves a low dimensional principal subspace plus white noise [26]. A typical convex representation of such models is $\mathbf{C}^{\text{True}} = \mathbf{X} + \sigma^2\mathbf{I}$ together with a bound on the nuclear norm of $\mathbf{X} \in \mathcal{P}(p)$. In this model, $\sigma^2$ is the known and fixed variance of the noise.
- **Linear parameterization**: Many interesting models can be expressed as a linear combination of known matrices. In particular, a modern approach to estimation of direction of arrivals of multiple signals involves a covariance of the form $\mathbf{C}^{\text{True}} = \sum_{i=1}^{k} p_i \mathbf{a}_i \mathbf{a}_i^T$ where $\mathbf{a}_i$ constitute a dense grid of possible directions, and $p_i$ are their corresponding coefficients. Typically, the $l_1$ norm of these sparse coefficients is constrained. See for example [6].

We can now state the problem addressed in this paper. Let $\mathbf{x}_i, i = 1, \ldots, n$ be independent and identically distributed (i.i.d) copies of $\mathbf{x}$ with $\mathbf{C}^{\text{True}} \in \mathcal{S}$. Given these realizations and knowledge of $\mathcal{S}$, we are interested in estimation of the matrix $\mathbf{C}^{\text{True}}$.

## III. Existing Solutions

### A. Sample Covariance

The classical solution to the above covariance estimation problem is the sample covariance matrix defined by

$$\mathbf{C}^{\text{Sample}} = \frac{1}{n}\sum_{i=1}^{n} \mathbf{x}_i \mathbf{x}_i^T. \tag{2}$$

The sample covariance estimator always exists and is asymptotically consistent in any distribution with bounded moments by the Law of Large Numbers. In the Gaussian case when $n \geq p$, it also maximizes the likelihood and is asymptotically efficient. In the non-Gaussian case, it has been extensively studied [27], any is generally suboptimal. Furthermore, it does not exploit any additional structure knowledge.

### B. Tyler's M-estimation approach

The most popular approach to covariance estimation in elliptical distribution is due to Tyler [18]. This estimator is defined as the fixed point solution to:

$$\mathbf{C}^{\text{Tyler}} = \frac{p}{n}\sum_{i=1}^{n} \frac{\mathbf{x}_i \mathbf{x}_i^T}{\mathbf{x}_i^T \left[\mathbf{C}^{\text{Tyler}}\right]^{-1} \mathbf{x}_i}. \tag{3}$$

Since $\mathbf{C}^{\text{Tyler}}$ is defined only up to scale, it has to be fixed by some additional constraint like $\operatorname{Tr}\left(\mathbf{C}^{\text{Tyler}}\right) = 1$. When $n > p$, it has been proven that a simple fixed point iteration converges to this unique solution [17]. This estimator is asymptotically consistent in all elliptical distributions. In fact, it is has been shown to maximize the likelihood of the normalized samples

$$\mathbf{s} = \frac{\mathbf{x}}{||\mathbf{x}||_2} = \frac{r\Lambda\mathbf{u}}{||r\Lambda\mathbf{u}||_2} = \frac{\Lambda\mathbf{u}}{||\Lambda\mathbf{u}||_2}, \tag{4}$$

which are independent of the values of the generating variate. The advantages of Tyler's estimator are its simplicity and robustness. Its drawbacks are that it does not exist if $n < p$ and does not exploit known structure. In [20] knowledge based variants of the fixed point iteration were proposed without convergence analysis. Recently, regularized and structured versions of Tyler's estimate were proposed in [19, 21, 21, 22, 28] based on the theories of concave Perron Frobenius and geodesic convexity. Unfortunately, these approaches are limited in its their modeling capabilities are cannot deal with general convex models as described above.

### C. Projection

A reasonable approach to explore the covariance structure is to use a projection. Given any estimator $\hat{\mathbf{C}}$, e.g., the sample covariance or Tyler, its projection onto $\mathcal{S}$ is defined as

$$\mathbf{C}_{\mathcal{S}}^{\text{Proj}}\left(\hat{\mathbf{C}}\right) = \underset{\mathbf{M}\in\mathcal{S}}{\operatorname{argmin}}\,||\mathbf{M}-\hat{\mathbf{C}}||, \tag{5}$$

where $\|\cdot\|$ is some norm. For simple structures as described above, the projection is a convex optimization problem which can be efficiently solved using standard numerical packages, e.g., CVX, [29, 30]. The main advantage is that, when $\mathbf{C}^{\text{True}} \in \mathcal{S}$, the projection $\mathbf{C}^{\text{Proj}}$ is usually closer to $\mathbf{C}^{\text{True}}$ than $\hat{\mathbf{C}}$ is. The disadvantage is that it requires a two-step solution which does not take into account the distribution properties and structure information and is therefore suboptimal.

## IV. COCA-ESTIMATOR

In this section, we propose a new COCA - the COnvexly ConstrAined covariance estimator for elliptical distributions. Unlike the existing solutions, COCA exploits both the elliptical nature and the structure of the underlying distribution. COCA is based on the Generalized Method of Moments [23] together with an asymptotically tight convex relaxation.

The underlying principle behind COCA is the following identity [14, 31]:

$$\mathbf{E}\left(p\frac{\mathbf{x}_i\mathbf{x}_i^T}{\mathbf{x}_i^T\left[\mathbf{C}^{\text{True}}\right]^{-1}\mathbf{x}_i}\right) = \mathbf{C}^{\text{True}}, \tag{6}$$

Indeed, Tyler's estimator is just the sample based solution that satisfies this identity. When there is an insufficient number of samples and a constraint of the structure, such a solution does not necessarily exist. Instead, the Generalized Method of Moments [23] seeks an approximate solution to the problem

$$\min_{\mathbf{C}\in\mathcal{S}_1} \left|\left|\mathbf{C} - \frac{p}{n}\sum_{i=1}^{n}\frac{\mathbf{x}_i\mathbf{x}_i^T}{\mathbf{x}_i^T\mathbf{C}^{-1}\mathbf{x}_i}\right|\right|, \tag{7}$$

where $||\cdot||$ is some norm and we ensure uniqueness by defining

$$\mathcal{S}_1 = \{\mathbf{M}\in\mathcal{S}|\operatorname{Tr}(\mathbf{M}) = 1\}. \tag{8}$$

Intuitively, this optimization tries to simultaneously solve Tyler and project it on the prior structure. By choosing the norm appropriately (with adaptive weights, e.g., [23, 32]), an optimal solution to (7) would result in an asymptotically consistent and efficient estimator. Unfortunately, the objective is non-convex and it is not clear how to find its global solution in a tractable manner.

In what follows, we propose a convex relaxation of (7) that allows a computationally efficient solution. First, let us introduce the auxiliary variables $d_i, i = 1,\ldots,n$:

$$\begin{aligned} &\min_{\mathbf{C}\in\mathcal{S}_1} && \left|\left|\mathbf{C} - \frac{1}{n}\sum_{i=1}^{n}d_i\mathbf{x}_i\mathbf{x}_i^T\right|\right| \\ &\text{subject to} && d_i = \frac{p}{\mathbf{x}_i^T\mathbf{C}^{-1}\mathbf{x}_i}, i = 1\ldots n. \end{aligned} \tag{9}$$

This problem is not convex due to the equality constraints. We suggest to relax them to the inequalities:

$$\begin{aligned} &\min_{\mathbf{C}\in\mathcal{S}_1} && \left|\left|\mathbf{C} - \frac{1}{n}\sum_{i=1}^{n}d_i\mathbf{x}_i\mathbf{x}_i^T\right|\right| \\ &\text{subject to} && d_i \leq \frac{p}{\mathbf{x}_i^T\mathbf{C}^{-1}\mathbf{x}_i}, i = 1\ldots n, \\ & && d_i \geq 0, i = 1\ldots n. \end{aligned} \tag{10}$$

This relaxed problem is a convex minimization. To see this, it is instructive to use Schur's complement formulas and express the inequalities $d_i \leq \frac{p}{\mathbf{x}_i^T\mathbf{C}^{-1}\mathbf{x}_i}, i = 1\ldots n$ as convex linear matrix inequalities (LMI):

$$\mathbf{C}^{COCA} = \arg\begin{cases} \min_{\mathbf{C}\in\mathcal{S}_1} & \left|\left|\mathbf{C} - \frac{1}{n}\sum_{i=1}^{n}d_i\mathbf{x}_i\mathbf{x}_i^T\right|\right| \\ \text{subject to} & \mathbf{C} \succeq \frac{1}{p}d_i\mathbf{x}_i\mathbf{x}_i^T, \forall i = 1\ldots n, \\ & d_i \geq 0, \forall i = 1\ldots n. \end{cases} \tag{11}$$

It can be efficiently computed by standard semi-definite programming solvers, e.g., CVX, [29, 30].

The non-relaxed version of the COCA-estimator in (7) can be considered optimal in many ways. The interesting question is how tight is the relaxation. We now provide two promising results on this tightness.

**Theorem 1.** *In the unstructured case $\mathcal{S} = \mathcal{P}(p)$ with $n \geq p$, COCA-estimator is unique up to a positive scaling factor and coincides with Tyler's estimator.*

*Proof.* In fact we know that when $n \geq p$ (11) has at least one solution which makes the target function be equal to zero. It is the Tyler's estimator itself satisfying

$$d_i = \frac{p}{\mathbf{x}_i^T\mathbf{C}^{-1}\mathbf{x}_i}, i = 1\ldots n.$$

These equalities hold up to a scaling factor.

We now show that there are no more solutions to (11). Indeed, assume there is an additional solution that attains 0 at the target function, that is $\mathbf{C} = \frac{1}{n}\sum_{i=1}^{n}d_i\mathbf{x}_i\mathbf{x}_i^T$. Multiply each inequality $d_i \leq \frac{p}{\mathbf{x}_i^T\mathbf{C}^{-1}\mathbf{x}_i}$ by the matrix $\mathbf{x}_i\mathbf{x}_i^T$ for $i = 1\ldots n$ and sum up to obtain

$$\mathbf{C} = \frac{1}{n}\sum_{i=1}^{n}d_i\mathbf{x}_i\mathbf{x}_i^T \preceq \frac{p}{n}\sum_{i=1}^{n}\frac{\mathbf{x}_i\mathbf{x}_i^T}{\mathbf{x}_i^T\mathbf{C}^{-1}\mathbf{x}_i} = f(\mathbf{C}). \tag{12}$$

The inequality (12) reads now as $\mathbf{C} \preceq f(\mathbf{C})$. As stated in the Corollary V.I from [13], this implies that $\mathbf{C}$ is the fixed point of $f$: $\mathbf{C} = f(\mathbf{C})$ (actually this result was already used by Maronna in [17], but it is not stated as a separate result there), which is exactly the Tyler's estimator (3), thus proving that this is the only solution to (11) up to a positive scaling factor. □

For a general convex set $\mathcal{S}$ the COCA-estimator is actually consistent:

**Theorem 2.** *In the structured case, COCA is an asymptotically consistent estimator of the true shape matrix* $\mathbf{C}^{True} \in \mathcal{S}$.

*Proof.* Due to space limitations, we defer this technical proof to the journal paper. In brief, the idea is to show that the inequality (12) hold asymptotically, as $n \to \infty$. □

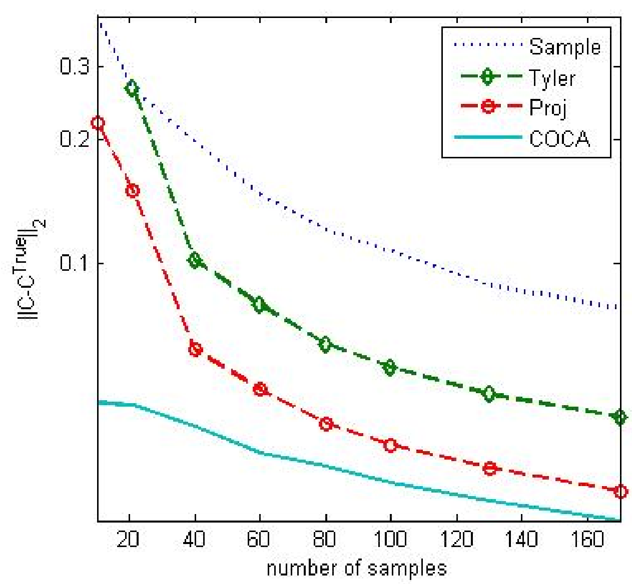
## V. Examples and Applications

In this section we choose the norm in (11) and the norm of the projector operator (5) to be the spectral norm. We we investigate the performance benefits of the COCA-estimator when the true shape matrix is Toeplitz and banded. We compare the following estimators: $\mathbf{C}^{\text{Sample}}$ (2), $\mathbf{C}^{\text{Tyler}}$ (3), $\mathbf{C}^{\text{Proj}}$ (5) and $\mathbf{C}^{COCA}$ (11). In $\mathbf{C}^{\text{Proj}}$ we project Tyler's estimator $\mathbf{C}^{\text{Proj}}(\mathbf{C}^{\text{Tyler}})$ when it exists and the sample covariance otherwise.

For each number of samples $n$ we generated $1000$ sets of independent, elliptically distributed 20-dimensional samples and calculated the empirical MSE for all the estimators. The samples were generated as Gaussian compound $\mathbf{x} = \sqrt{\tau}\mathbf{v}$, where the random variable $\tau \sim \chi^2$ and the random vector $\mathbf{v}$ was zero-mean normally distributed with covariance matrix $\mathbf{C}^{\text{True}}$.

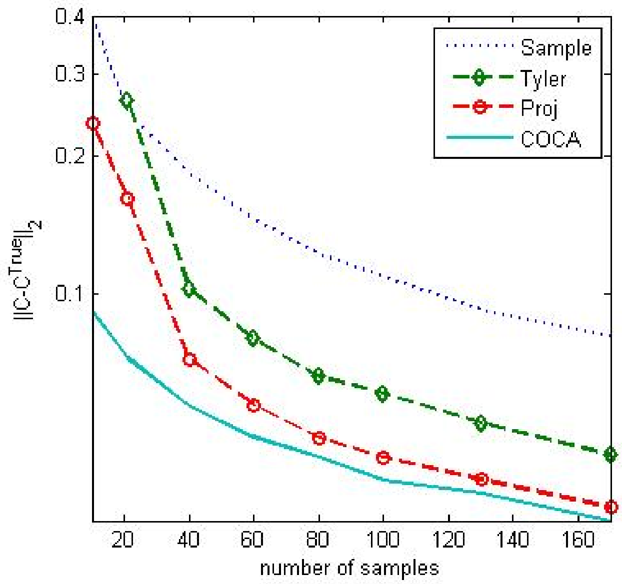
### A. Toeplitz Covariance Matrix

The $20 \times 20$ Toeplitz shape matrix was obtained as $C^{\text{True}} = FDF^T$, where $F$ is the 20-dimensional DFT matrix and $D$ - diagonal matrix with eigenvalues $1, \dots, 20$. The proof of fact that such $C^{\text{True}}$ is circulant and, in particular, Toeplitz can be found here [5]. This matrix appears to be complex Hermitian, and all the theory developed above applies to this case. You can see the numerical results in the picture.




### B. Banded Covariance Matrix

For the banded covariance matrix we took a symmetric matrix with the numbers $21, \dots, 40$ on the diagonal, $1, \dots, 19$ and $1, \dots, 18$ on the first and second sub-diagonals respectively.




### C. Discussion and Conclusions

As we can see, the performance of the proposed COCA-estimator is much better than that of the projection estimator. The most important benefits of COCA-estimator is that it is given as a convex program, thus admitting any convex structure, and it exists when $n < p$, in which case many other estimators do not exist or are rank deficient.

There are several ways of extending this result. First, it can be extended to a general M-estimator and not only the particular case of Tyler's estimator. We will treat this generalization in our future work. Secondly, it turns out in the numerical runs that this estimator is almost always better then Tyler's estimator itself or its projection. We are currently working on proving this conjectures.

## VI. Acknowledgement

This work was partially supported by the Intel Collaboration Research Institute for Computational Intelligence and Kaete Klausner Scholarship.

## References

[1] H. Krim and M. Viberg. Two decades of array signal processing research: the parametric approach. *Signal Processing Magazine, IEEE*, 13(4):67–94, 1996.

[2] E. R. Dougherty, A. Datta, and C. Sima. Research issues in genomic signal processing. *Signal Processing Magazine, IEEE*, 22(6):46–68, 2005.

[3] D. R. Fuhrmann. Application of Toeplitz covariance estimation to adaptive beamforming and detection. *IEEE Transactions on Signal Processing*, 39(10):2194–2198, 1991.

[4] W. J. Roberts and Y. Ephraim. Hidden Markov modeling of speech using Toeplitz covariance matrices. *Speech Communication*, 31(1):1–14, 2000.

[5] D. S. Pollock. Circulant matrices and time-series analysis. *International Journal of Mathematical Education in Science and Technology*, 33(2):213–230, 2002.

[6] P. Stoica, P. Babu, and J. Li. SPICE: A sparse covariance-based estimation method for array processing. *IEEE Transactions on Signal Processing*, 59(2):629–638, 2011.

[7] H. Wang and L. Cai. On adaptive spatial-temporal processing for airborne surveillance radar systems. *Aerospace and Electronic Systems, IEEE Transactions on*, 30(3):660–670, 1994.

[8] D. L. Snyder, J. A. O'Sullivan, and M. I. Miller. The use of maximum likelihood estimation for forming images of diffuse radar targets from delay-Doppler data. *IEEE Transactions on Information Theory*, 35(3):536–548, 1989.

[9] R. Dahlhaus. Efficient parameter estimation for self-similar processes. *The Annals of Statistics*, pages 1749–1766, 1989.

[10] P. J. Bickel and E. Levina. Regularized estimation of large covariance matrices. *The Annals of Statistics*, pages 199–227, 2008.

[11] A. Wiesel. Regularized covariance estimation in scaled Gaussian models. pages 309–312, 2011.

[12] T. T. Cai, Zh. Ren, and H. H. Zhou. Optimal rates of convergence for estimating Toeplitz covariance matrices. *Probability Theory and Related Fields*, pages 1–43, 2012.

[13] F. Pascal, Y. Chitour, J. P. Ovarlez, P. Forster, and P. Larzabal. Covariance structure maximum-likelihood estimates in compound Gaussian noise: Existence and algorithm analysis. *IEEE Transactions on Signal Processing*, 56(1):34–48, 2008.

[14] G. Frahm. Generalized elliptical distributions: theory and applications. 2004.

[15] E. Conte, M. Longo, and M. Lops. Modelling and simulation of non-Rayleigh radar clutter. 138(2):121–130, 1991.

[16] D. Middleton. Man-made noise in urban environments and transportation systems: Models and measurements. *IEEE Transactions on Communications*, 21(11):1232–1241, 1973.

[17] R. A. Maronna. Robust M-estimators of multivariate location and scatter. *The annals of statistics*, pages 51–67, 1976.

[18] D. E. Tyler. A distribution-free M-estimator of multivariate scatter. *The Annals of Statistics*, 15(1):234–251, 1987.

[19] Y. I. Abramovich, N. K. Spencer, and M. D. Turley. Time-varying autoregressive (TVAR) models for multiple radar observations. *Signal Processing, IEEE Transactions on*, 55(4):1298–1311, 2007.

[20] F. Bandiera, O. Besson, and G. Ricci. Knowledge-aided covariance matrix estimation and adaptive detection in compound-Gaussian noise. *Signal Processing, IEEE Transactions on*, 58(10):5391–5396, 2010.

[21] A. Wiesel. Geodesic convexity and covariance estimation. 2012.

[22] I. Soloveychik and A. Wiesel. Group symmetry and non-Gaussian covariance estimation. *arXiv preprint arXiv:1306.4103*, 2013.

[23] L. Mátyás. Generalized method of moments estimation. 5, 1999.

[24] J. P. Burg, D. G. Luenberger, and D. L. Wenger. Estimation of structured covariance matrices. *Proceedings of the IEEE*, 70(9):963–974, 1982.

[25] A. Wiesel, O. Bibi, and A. Globerson. Time varying autoregressive moving average models for covariance estimation. 2013.

[26] M. W. Ganz, R. Moses, and S. Wilson. Convergence of the SMI and the diagonally loaded SMI algorithms with weak interference [adaptive array]. *IEEE Transactions on Antennas and Propagation*, 38(3):394–399, 1990.

[27] R. Vershynin. How close is the sample covariance matrix to the actual covariance matrix? *Journal of Theoretical Probability*, 25(3):655–686, 2012.

[28] Y. Chen, A. Wiesel, and A. O. Hero. Robust shrinkage estimation of high-dimensional covariance matrices. *IEEE Transactions on Signal Processing*, 59(9):4097–4107, 2011.

[29] M. Grant and S. Boyd. CVX: Matlab software for disciplined convex programming, version 2.0 beta. http://cvxr.com/cvx, September 2013.

[30] M. Grant and S. Boyd. Graph implementations for nonsmooth convex programs. In V. Blondel, S. Boyd, and H. Kimura, editors, *Recent Advances in Learning and Control*, Lecture Notes in Control and Information Sciences, pages 95–110. Springer-Verlag Limited, 2008.

[31] G. Frahm and U. Jaekel. A generalization of Tylers M-estimators to the case of incomplete data. *Computational Statistics & Data Analysis*, 54(2):374–393, 2010.

[32] B. Ottersten, P. Stoica, and R. Roy. Covariance matching estimation techniques for array signal processing applications. *Digital Signal Processing*, 8(3):185–210, 1998.