Robust Shrinkage Estimation of High-dimensional Covariance Matrices

[1]Yilun Chen, [2]Ami Wiesel, [1]Alfred O. Hero, III
[1] Department of EECS, University of Michigan, Ann Arbor, USA
[2] Department of CS, Hebrew University, Jerusalem, Israel
{yilun,hero}@umich.edu   amiw@cs.huji.ac.il

***Abstract*—We address high dimensional covariance estimation for elliptical distributed samples. Specifically we consider shrinkage methods that are suitable for high dimensional problems with a small number of samples (large p small n). We start from a classical robust covariance estimator [Tyler(1987)], which is distribution-free within the family of elliptical distribution but inapplicable when $n < p$. Using a shrinkage coefficient, we regularize Tyler's fixed point iteration. We derive the minimum mean-squared-error shrinkage coefficient in closed form. The closed form expression is a function of the unknown true covariance and cannot be implemented in practice. Instead, we propose a plug-in estimate to approximate it. Simulations demonstrate that the proposed method achieves low estimation error and is robust to heavy-tailed samples.**

I. Introduction

Estimating a covariance matrix (or a dispersion matrix) is a fundamental problem in statistical signal processing. Many techniques for detection and estimation rely on accurate estimation of the true covariance. In recent years, estimating a high dimensional $p \times p$ covariance matrix under small sample size n has attracted considerable attention. In these "large p small n" problems, the classical sample covariance suffers from a systematically distorted eigen-structure, and improved estimators are required.

Many efforts have been devoted to high-dimensional covariance estimation, which use Steinian shrinkage [1]–[3] or other types of regularized methods such as [4], [5]. However, most of the high-dimensional estimators assume Gaussian distributed samples. This limits their usage in many important applications involving non-Gaussian and heavy-tailed samples. One exception is the Ledoit-Wolf estimator [2], where the authors shrink the sample covariance towards a scaled identity matrix and proposed a shrinkage coefficient which is asymptotically optimal for any distribution. However, as the Ledoit-Wolf estimator operates on the sample covariance, it is inappropriate for heavy tailed non-Gaussian distributions. On the other hand, traditional robust covariance estimators [6]–[8] designed for non-Gaussian samples generally require $n \gg p$ and are not suitable for "large p small n" problems. Therefore, the goal of our work is to develop a covariance estimator for both problems that are high dimensional and non-Gaussian.

This work was partially supported by AFOSR, grant number FA9550-06-1-0324. The work of A. Wiesel was supported by a Marie Curie Outgoing International Fellowship within the 7th European Community Framework Programme.

In this paper, we model the samples using the elliptical distribution; a flexible and popular alternative that encompasses a large number of important non-Gaussian distributions in signal processing and related fields, e.g., [9], [13], [16]. A well-studied covariance estimator in this setting is the ML estimator based on normalized samples [7], [16]. The estimator is derived as the solution to a fixed point equation. It is distribution-free within the class of elliptical distributions and its performance advantages are well known in the $n \gg p$ regime. However, it is not suitable for the "large p small n" setting. Indeed, when $n < p$, the ML estimator does not even exist. To avoid this problem the authors of [10] propose an iterative regularized ML estimator that employs diagonal loading and use a heuristic for selecting the regularization parameter. They empirically demonstrated that their algorithm has superior performance in the context of a radar application.

Our approach is similar to [10] but we propose a a systematical choice of the regularization parameter. We consider a shrinkage estimator that regularizes the fixed point iterations of the ML estimator. Following Ledoit-Wolf [2], we provide a simple closed-form expression for the minimum mean-squared-error shrinkage coefficient. This clairvoyant coefficient is a function of the unknown true covariance and cannot be implemented in practice. Instead, we develop a "plug-in" estimate to approximate it. Simulation results demonstrate that the our estimator achieves superior performance for samples distributed within the elliptical family. Furthermore, for the case that the samples are truly Gaussian, we report very little performance degradation with respect to the shrinkage estimators designed specifically for Gaussian samples [3].

The paper is organized as follows. Section II provides a brief review of elliptical distributions and Tyler's covariance estimation method. The regularized estimator is introduced and derived in Section III. We provide simulations in Section IV and conclude the paper in Section V.

Notations: In the following, we depict vectors in lowercase boldface letters and matrices in uppercase boldface letters. $(\cdot)^T$ denotes the transpose operator. $\mathrm{Tr}(\cdot)$ and $\det(\cdot)$ are the trace and the determinant of a matrix, respectively.

II. ML COVARIANCE ESTIMATION FOR ELLIPTICAL DISTRIBUTIONS

A. Elliptical distribution

Let $\mathbf{x}$ be a $p \times 1$ zero-mean random vector generated by the following model

$$\mathbf{x} = \mathcal{R}\mathbf{u}, \tag{1}$$

where $\mathcal{R}$ is a positive random variable and $\mathbf{u}$ is a $p \times 1$ zero-mean, jointly Gaussian random vector with positive definite covariance $\mathbf{\Sigma}$. We assume that $\mathcal{R}$ and $\mathbf{u}$ are statistically independent. The resulting random vector $\mathbf{x}$ is elliptically distributed.

The elliptical family encompasses many useful distributions in signal processing and related fields and includes: the Gaussian distribution itself, the K distribution, the Weibull distribution and many others. Elliptically distributed samples are also referred to as Spherically Invariant Random Vectors (SIRV) or compound Gaussian vectors in signal processing and have been used in various applications such as band-limited speech signal models, radar clutter echo models [9], and wireless fading channels [13].

B. ML estimation

Let $\{\mathbf{x}_i\}_{i=1}^n$ be a set of n independent and identically distributed (i.i.d.) samples drawn according to (1). The problem is to estimate the covariance (dispersion) matrix $\mathbf{\Sigma}$ from $\{\mathbf{x}_i\}_{i=1}^n$. To remove the scale ambiguity caused by $\mathcal{R}$ we further constrain that $\mathrm{Tr}(\mathbf{\Sigma}) = p$.

The commonly used sample covariance, defined as

$$\widehat{\mathbf{S}} = \frac{1}{n}\sum_{i=1}^{n} \mathbf{x}_i \mathbf{x}_i^T, \tag{2}$$

is known to be a poor estimator of $\mathbf{\Sigma}$, especially when the samples are high-dimensional and/or heavy-tailed.

Tyler's method [7], [16] addressed this problem by working with the normalized samples:

$$\mathbf{s}_i = \frac{\mathbf{x}_i}{\|\mathbf{x}_i\|_2}, \tag{3}$$

for which the term $\mathcal{R}$ in (1) drops out. The pdf of $\mathbf{s}_i$ is given by [12]

$$p(\mathbf{s}_i; \mathbf{\Sigma}) = \frac{\Gamma(p/2)}{2\pi^{p/2}} \cdot \sqrt{\det(\mathbf{\Sigma}^{-1})} \cdot \left(\mathbf{s}_i^T \mathbf{\Sigma}^{-1} \mathbf{s}_i\right)^{-p/2}. \tag{4}$$

and the maximum likelihood estimator based on $\{\mathbf{s}_i\}_{i=1}^n$ is the solution to

$$\mathbf{\Sigma} = \frac{p}{n} \cdot \sum_{i=1}^{n} \frac{\mathbf{s}_i \mathbf{s}_i^T}{\mathbf{s}_i^T \mathbf{\Sigma}^{-1} \mathbf{s}_i}. \tag{5}$$

When $n \geq p$, the ML estimator can be found using the following fixed point iterations:

$$\widehat{\mathbf{\Sigma}}_{j+1} = \frac{p}{n} \cdot \sum_{i=1}^{n} \frac{\mathbf{s}_i \mathbf{s}_i^T}{\mathbf{s}_i^T \widehat{\mathbf{\Sigma}}_j^{-1} \mathbf{s}_i}, \tag{6}$$

where the initial $\widehat{\mathbf{\Sigma}}_0$ is usually set to the identity matrix. It can be shown [7], [16] that the iteration process in (6) converges and does not depend on the initial setting of $\widehat{\mathbf{\Sigma}}_0$. In practice a final normalization step is needed, which ensures the iteration limit $\widehat{\mathbf{\Sigma}}_\infty$ satisfies $\mathrm{Tr}(\widehat{\mathbf{\Sigma}}_\infty) = p$.

The ML estimate corresponds to the Huber-type M-estimator and has many good properties when $n \gg p$, such as asymptotic normality and strong consistency. Furthermore, it has been pointed out [7] that the ML estimate is the "most robust" covariance estimator in the class of elliptical distributions in the sense of minimizing the maximum asymptotic variance.

III. ROBUST SHRINKAGE COVARIANCE ESTIMATION

Here we extend Tyler's method to the high dimensional setting using shrinkage. It is easy to see that there is no solution to (5) when $n < p$ (its left-hand-side is full rank whereas it right-hand-side of is rank deficient). This motivates us to develop a regularized covariance estimator for elliptical samples. Following [2], [3], we propose to regularize the fixed point iterations as

$$\widetilde{\mathbf{\Sigma}}_{j+1} = (1-\rho)\frac{p}{n}\sum_{i=1}^{n} \frac{s_i^T s_i}{s_i^T \widehat{\mathbf{\Sigma}}_j^{-1} s_i} + \rho \mathbf{I} \tag{7}$$

$$\widehat{\mathbf{\Sigma}}_{j+1} = \frac{\widetilde{\mathbf{\Sigma}}_{j+1}}{\mathrm{Tr}(\widetilde{\mathbf{\Sigma}}_{j+1})/p}, \tag{8}$$

where ρ is the so-called shrinkage coefficient, which is a constant between 0 and 1. When $\rho = 0$ the proposed shrinkage estimator reduces to Tyler's unbiased method in (5) and when $\rho = 1$ the estimator reduces to the trivial estimator $\mathbf{\Sigma} = \mathbf{I}$. The term $\rho\mathbf{I}$ ensures that $\widetilde{\mathbf{\Sigma}}_{j+1}$ is always well-conditioned and thus allows continuation of the iterative process without restarts. Therefore, the proposed iteration can be applied to high dimensional estimation problems. We note that the normalization (8) is important and necessary for convergence.

We now turn to the problem of choosing a good, data-dependent, shrinkage coefficient ρ. Following Ledoit-Wolf [2], we begin by assuming we "know" the true covariance $\mathbf{\Sigma}$. The optimal ρ that minimizes the minimum mean-squared error is called the "oracle" coefficient and is

$$\rho_O = \arg\min_{\rho} E\left\{\left\|\widetilde{\mathbf{\Sigma}}(\rho) - \mathbf{\Sigma}\right\|_F^2\right\}, \tag{9}$$

where $\widetilde{\mathbf{\Sigma}}(\rho)$ is defined as

$$\widetilde{\mathbf{\Sigma}}(\rho) = (1-\rho)\frac{p}{n}\sum_{i=1}^{n} \frac{s_i s_i^T}{s_i^T \mathbf{\Sigma}^{-1} s_i} + \rho \mathbf{I}. \tag{10}$$

There is a closed-form solution to the problem (9) which is provided in the following theorem.

Theorem 1. *For i.i.d. elliptical distributed samples, the solution to (9) is*

$$\rho_O = \frac{\mathrm{Tr}^2(\mathbf{\Sigma}) + (1 - 2/p)\mathrm{Tr}(\mathbf{\Sigma}^2)}{(1 - n/p - 2n/p^2)\mathrm{Tr}^2(\mathbf{\Sigma}) + (n + 1 + 2(n-1)/p)\mathrm{Tr}(\mathbf{\Sigma}^2)}. \tag{11}$$

Proof: To ease the notation we define $\widetilde{\mathbf{C}}$ as

$$\widetilde{\mathbf{C}} = \frac{p}{n}\sum_{i=1}^{n} \frac{\mathbf{s}_i \mathbf{s}_i^T}{\mathbf{s}_i^T \mathbf{\Sigma}^{-1} \mathbf{s}_i}. \tag{12}$$

The shrinkage "estimator" in (10) is then

$$\widetilde{\boldsymbol{\Sigma}}(\rho) = (1-\rho)\widetilde{\mathbf{C}} + \rho\mathbf{I}. \tag{13}$$

By substituting (13) into (10) and taking derivatives of ρ, we obtain that

$$\begin{aligned}\rho_O &= \frac{E\left\{\mathrm{Tr}\left((\mathbf{I}-\widetilde{\mathbf{C}})(\boldsymbol{\Sigma}-\widetilde{\mathbf{C}})\right)\right\}}{E\left\{\left\|\mathbf{I}-\widetilde{\mathbf{C}}\right\|_F^2\right\}} \\ &= \frac{m_2 - m_{11} - m_{12} + p}{m_2 - 2m_{11} + p},\end{aligned} \tag{14}$$

where

$$m_2 = E\left\{\mathrm{Tr}(\widetilde{\mathbf{C}}^2)\right\}, \tag{15}$$

$$m_{11} = E\left\{\mathrm{Tr}(\widetilde{\mathbf{C}})\right\}, \tag{16}$$

and

$$m_{12} = E\left\{\mathrm{Tr}(\widetilde{\mathbf{C}}\boldsymbol{\Sigma})\right\}. \tag{17}$$

Next, we calculate the moments. We begin by eigen-decomposing $\boldsymbol{\Sigma}$ as

$$\boldsymbol{\Sigma} = \mathbf{U}\mathbf{D}\mathbf{U}^T, \tag{18}$$

and denote

$$\boldsymbol{\Lambda} = \mathbf{U}\mathbf{D}^{1/2}. \tag{19}$$

Then, we define

$$\mathbf{z}_i = \frac{\boldsymbol{\Lambda}^{-1}\mathbf{s}_i}{\|\boldsymbol{\Lambda}^{-1}\mathbf{s}_i\|_2} = \frac{\boldsymbol{\Lambda}^{-1}\mathbf{u}_i}{\|\boldsymbol{\Lambda}^{-1}\mathbf{u}_i\|_2}. \tag{20}$$

Noting that $\mathbf{u}_i$ is a Gaussian distributed random vector with covariance $\boldsymbol{\Sigma}$, it is easy to see that $\|\mathbf{z}_i\|_2 = 1$ and $\mathbf{z}_i$ and $\mathbf{z}_j$ are independent if $i \neq j$. Furthermore, $\mathbf{z}_i$ is isotropically distributed [3] and satisfies

$$E\left\{\mathbf{z}_i\mathbf{z}_i^T\right\} = \frac{1}{p}\mathbf{I}, \tag{21}$$

$$\begin{aligned}E\left\{\left(\mathbf{z}_i^T\mathbf{D}\mathbf{z}_i\right)^2\right\} &= \frac{1}{p(p+2)}\left(2\mathrm{Tr}(\mathbf{D}^2) + \mathrm{Tr}^2(\mathbf{D})\right) \\ &= \frac{1}{p(p+2)}\left(2\mathrm{Tr}(\boldsymbol{\Sigma}^2) + \mathrm{Tr}^2(\boldsymbol{\Sigma})\right),\end{aligned} \tag{22}$$

and

$$E\left\{\left(\mathbf{z}_i^T\mathbf{D}\mathbf{z}_j\right)^2\right\} = \frac{1}{p^2}\mathrm{Tr}(\mathbf{D}^2) = \frac{1}{p^2}\mathrm{Tr}(\boldsymbol{\Sigma}^2),\ i \neq j. \tag{23}$$

Expressing $\widetilde{\mathbf{C}}$ in terms of $\mathbf{z}_i$, there is

$$\widetilde{\mathbf{C}} = \frac{p}{n}\boldsymbol{\Lambda}\sum_{i=1}^{n} z_i z_i^T \boldsymbol{\Lambda}^T. \tag{24}$$

Then,

$$E\left\{\widetilde{\mathbf{C}}\right\} = \frac{p}{n}\boldsymbol{\Lambda}\sum_{i=1}^{n} E\left\{\mathbf{z}_i\mathbf{z}_i^T\right\}\boldsymbol{\Lambda}^T = \boldsymbol{\Sigma}, \tag{25}$$

and accordingly we have

$$m_{11} = E\left\{\mathrm{Tr}(\widetilde{\mathbf{C}})\right\} = \mathrm{Tr}(\boldsymbol{\Sigma}), \tag{26}$$

and

$$m_{12} = E\left\{\mathrm{Tr}(\widetilde{\mathbf{C}}\boldsymbol{\Sigma})\right\} = \mathrm{Tr}(\boldsymbol{\Sigma}^2). \tag{27}$$

For m_2 there is

$$\begin{aligned}m_2 &= \frac{p^2}{n^2}E\left\{\mathrm{Tr}\left(\boldsymbol{\Lambda}\sum_{i=1}^{n}\mathbf{z}_i\mathbf{z}_i^T\boldsymbol{\Lambda}^T\boldsymbol{\Lambda}\sum_{j=1}^{n}\mathbf{z}_j\mathbf{z}_j^T\boldsymbol{\Lambda}^T\right)\right\} \\ &= \frac{p^2}{n^2}E\left\{\mathrm{Tr}\left(\sum_{i=1}^{n}\sum_{j=1}^{n}\mathbf{z}_i\mathbf{z}_i^T\boldsymbol{\Lambda}^T\boldsymbol{\Lambda}\mathbf{z}_j\mathbf{z}_j^T\boldsymbol{\Lambda}^T\boldsymbol{\Lambda}\right)\right\} \\ &= \frac{p^2}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}E\left\{\left(\mathbf{z}_i^T\mathbf{D}\mathbf{z}_j\right)^2\right\}.\end{aligned} \tag{28}$$

Now substitute (22) and (23) to (28):

$$\begin{aligned}m_2 &= \frac{p^2}{n^2}\left(\frac{n}{p(p+2)}\left(2\mathrm{Tr}(\boldsymbol{\Sigma}^2) + \mathrm{Tr}^2(\boldsymbol{\Sigma})\right) + \frac{n(n-1)}{p^2}\mathrm{Tr}(\boldsymbol{\Sigma}^2)\right) \\ &= \left(1 - \frac{1}{n} + \frac{2}{n(1+2/p)}\right)\mathrm{Tr}(\boldsymbol{\Sigma}^2) + \frac{\mathrm{Tr}^2(\boldsymbol{\Sigma})}{n(1+2/p)}.\end{aligned} \tag{29}$$

Recalling $\mathrm{Tr}(\boldsymbol{\Sigma}) = p$, (11) is finally obtained by substituting (26), (27) and (29) into (14). ■

The oracle cannot be implemented since ρ_O is a function of the unknown true covariance $\boldsymbol{\Sigma}$. We propose to use a plug-in estimate for ρ_O determined as follows:

$$\hat{\rho} = \frac{\mathrm{Tr}^2(\widehat{\mathbf{R}}) + (1-2/p)\mathrm{Tr}(\widehat{\mathbf{R}}^2)}{(1-n/p-2n/p^2)\mathrm{Tr}^2(\widehat{\mathbf{R}}) + (n+1+2(n-1)/p)\mathrm{Tr}(\widehat{\mathbf{R}}^2)}, \tag{30}$$

where $\widehat{\mathbf{R}}$ is the normalized sample covariance:

$$\widehat{\mathbf{R}} = \frac{1}{n}\sum_{i=1}^{n}\mathbf{s}_i\mathbf{s}_i^T. \tag{31}$$

Given the plug-in estimate $\hat{\rho}$, the robust shrinkage estimator is computed using the fixed point iteration in (7) and (8).

IV. Simulations

In this section we use simulations to demonstrate the superior performance of the proposed approach. First we show that our estimator outperforms other estimators in the case of heavy-tailed samples generated by a multivariate Student-T distribution. The degree-of-freedom of this multivariate Student-T is set to 3. The dimensionality p is chosen to be 100 and an autoregressive covariance structured $\boldsymbol{\Sigma}$ is used. We let $\boldsymbol{\Sigma}$ be the covariance matrix of an AR(1) process,

$$\boldsymbol{\Sigma}(i,j) = r^{|i-j|}, \tag{32}$$

where $\boldsymbol{\Sigma}(i,j)$ denotes the entry of $\boldsymbol{\Sigma}$ in row i and column j, and r is set to 0.7 in this simulation. The sample size n varies from 5 to 225 with step size 10. All the simulations are repeated for 100 trials and the results are averaged.

For comparison, we also plot the results of the closed-form oracle in (11), the Ledoit-Wolf estimator [2], and the non-regularized solution in (6) when $n > p$. As the Ledoit-Wolf estimator operates on the sample covariance which is sensitive to heavy-tails, we also compare the Ledoit-Wolf estimator with known $\mathcal{R}$ in (1), where the samples $\mathbf{x}_i$ are firstly normalized by those known realizations $\mathcal{R}_i$, yielding truly Gaussian samples, and followed by implementation of the Ledoit-Wolf estimator. The MSE of those estimators are plotted in Fig. 1. It can be observed that the proposed method performs significantly better than the Ledoit-Wolf estimator, and the performance is very close to the ideal oracle. Even the Ledoit-Wolf with known $\mathcal{R}_i$ does not exceed the proposed estimator for small sample size. These results demonstrate the robustness of the proposed approach.

When $n > p$, we also observe a substantial improvement of the proposed method over the ML estimate, which demonstrates the power of Steinian shrinkage in reducing the MSE.

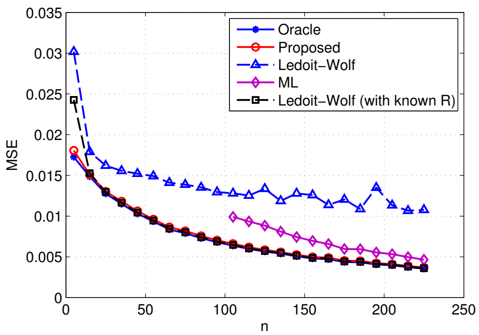

Fig. 1. Multivariate Student-T samples: Comparison of different covariance estimators when $p = 100$.

In order to asses the tradeoff between accuracy and robustness we investigate the case when the samples are truly Gaussian distributed. We use the same setting as in the previous example and the only difference is that the samples are now generated from a Gaussian distribution. The performance comparison is shown in Fig. 2, where four different methods are included: the oracle estimator derived from Gaussian setting (Gaussian oracle) [3], the iterative approximation of the Gaussian oracle (Gaussian OAS) [3], the Ledoit-Wolf estimator and the proposed method. It can be seen that for truly Gaussian samples the proposed method performs very closely to the Gaussian OAS which is specifically designed for Gaussian distributions. Indeed, for small sample size ($n < 20$), the proposed method performs even better than the Ledoit-Wolf estimator. This indicates that although the proposed method is developed for the entire elliptical family, it actually sacrifices very little performance for the case that the distribution is known to be Gaussian.

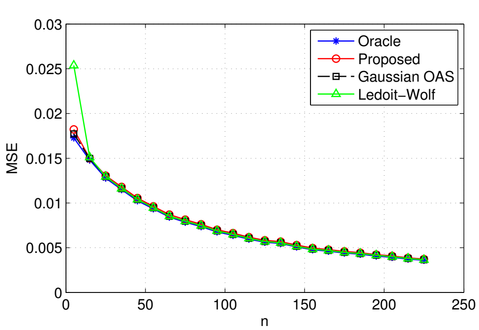

Fig. 2. Gaussian samples: Comparison of different covariance estimators when $p = 100$.

V. Conclusion

In this paper, we proposed a shrinkage covariance estimator which is robust over the class of elliptically distributed samples. The proposed estimator is obtained by fixed point iterations, and we established a systematical approach to choosing the shrinkage coefficient, which was derived using a minimum mean-squared-error framework and has a closed-form expression in terms of the unknown true covariance. This expression can be well approximated by a simple plug-in estimator. Simulations suggest that the proposed estimator is robust to heavy-tailed multivariate Student-T samples. Furthermore, we show that for the Gaussian case, the proposed estimator performs very closely to previous estimators designed expressly for Gaussian samples.

We did not give a proof that the regularized fixed point iteration (7) and (8) converges. The proof of convergence uses ideas from concave Perron-Frobenius theory [18] and further details will be provided in the full length journal version of this paper.

References

[1] C. Stein, "Estimation of a covariance matrix," In *Rietz Lecture, 39th Annual Meeting*, IMS, Atlanta, GA, 1975.

[2] O. Ledoit, M. Wolf, "A Well-Conditioned Estimator for Large-Dimensional Covariance Matrices", *Journal of Multivariate Analysis*, vol. 88, iss. 2, Feb. 2004.

[3] Y. Chen, A. Wiesel, Y. C. Eldar, A. O. Hero III, "Shrinkage Algorithms for MMSE Covariance Estimation ," to appear in *IEEE Trans. on Sig. Process.*

[4] J. Friedman, T. Hastie, R. Tibshirani, "Sparse inverse covariance estimation with the graphical lasso," *Biostatistics*, 2008.

[5] P. Bickel, E. Levina, "Regularized estimation of large covariance matrices," *The Annals of Statistics*, vol. 36, pp. 199-227, 2008.

[6] P.J. Huber, "Robust statistics," Wiley, 1981.

[7] D.E. Tyler, "A distribution-free M-estimator of multivariate scatter," *The Annals of Statistics*, 1987.

[8] P. Rousseeuw, "Multivariate estimation with high breakdown point,". *Mathematical Statistics and Applications*, Reidel, 1985.

[9] F. Gini, "Sub-optimum Coherent Radar Detection in a Mixture of Kdistributed and Gaussian Clutter, *IEE Proc. Radar, Sonar and Navigation*, vol. 144, no. 1, pp. 39-48, Feb. 1997.

[10] B. A. Johnson, Y. L. Abramovich, "Diagonally loaded normalised sample matrix inversion (LNSMI) for outlier-resistant adaptive filtering," *IEEE Intl Conf. on Acoust., Speech, and Signal Processing*, vol. 3, pp. 1105 - 1108, 2007.

[11] J.B. Billingsley, "Ground Clutter Measurements for Surface-Sited Radar, *Technical Report 780*, MIT, Feb. 1993.

[12] G. Frahm, "Generalized Elliptical Distributions: Theory and Applications," *Dissertation*, 2004.

[13] K. Yao, M.K. Simon and E. Biglieri, "A Unified Theory on Wireless Communication Fading Statistics based on SIRV, *Fifth IEEE Workshop on SP Advances in Wireless Communications*, 2004.

[14] A. L. Yuille, A. Rangarajan, "The concave-convex procedure", *NIPS* 2003.

[15] J. Wang, A. Dogandzic, A. Nehorai, "Maximum Likelihood Estimation of Compound-Gaussian Clutter and Target Parameters," *IEEE Trans. on Sig. Process.*, vol. 54, no. 10, 2006.

[16] 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 Trans. on Sig. Process.*, vol. 56, no. 1, Jan. 2008.

[17] Y. Chitour, F. Pascal, "Exact Maximum Likelihood Estimates for SIRV Covariance Matrix: Existence and Algorithm Analysis," *IEEE Trans. on Sig. Process.*, vol. 56, no. 10, Oct. 2008.

[18] U. Krause, "Concave perron-frobenius theory and applications, *Nonlinear Anal.*, vol. 47, no. 3, pp. 14571466, 2001.