Performance Analysis of Tyler’s Covariance Estimator
Ilya Soloveychik, Senior Member, IEEE, and Ami Wiesel

Abstract—This paper analyzes the performance of Tyler’s M-estimator of the scatter matrix in elliptical populations. We focus on the non-asymptotic setting and derive estimation error bounds depending on the number of samples $n$ and the dimension $p$. We show that under mild conditions the squared Frobenius norm of the error of the inverse estimator decays like $p^2/n$ with high probability.

Index Terms—Concentration bounds, elliptical distribution shape matrix estimation, scatter matrix M-estimators, Tyler’s scatter estimator.

I. INTRODUCTION

ESTIMATION of large covariance matrices, particularly in situations where the data dimension $p$ and the sample size $n$ are of close magnitudes, has recently attracted considerable attention. Estimators in this field can be classified based on the underlying distribution and the additional structure assumptions. Most of the research is traditionally devoted to the multivariate Gaussian setting which is currently well understood. Various algorithms based on different structures and their performance analysis have been derived for the Gaussian distributions, see e.g. [1]–[3]. This allows for higher reliability and better tuning of regularization parameters. Recently, similar challenges have appeared in the more ambitious setting of non-Gaussian and robust estimation. A prominent approach in this area is Tyler’s scatter estimator [4]. The goal of this paper is to analyze its non-asymptotic behavior. We start by reviewing the state of the art in both Gaussian and Tyler’s covariance estimation, and then introduce our result.

Most of the works on covariance estimation address the Gaussian scenario. When the number of samples is greater than the dimension, the Maximum Likelihood Estimator (MLE) of the covariance exists with probability one and coincides with the Sample Covariance Matrix (SCM). Recently, there has emerged a great amount of literature on regularized versions of this estimator, their parameter tuning and performance analysis, see e.g. [1], [3].

Manuscript received January 29, 2014; revised June 02, 2014 and September 09, 2014; accepted November 16, 2014. Date of publication December 02, 2014.

IEEE Copyright Notice © 2014 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.
\( x_i \in \mathbb{R}^p, i = 1, \ldots, n \), Tyler’s shape matrix estimator is defined as the solution to the fixed point equation

\[
T = \frac{p}{n} \sum_{i=1}^{n} \frac{x_i x_i^T}{x_i^T T^{-1} x_i}.
\]

When \( x_i \) are Generalized Elliptically (GE) distributed [22], their shape matrix \( \Theta_0 \) is positive definite and \( n > p \). Tyler’s estimator exists with probability one and is a consistent estimator of \( \Theta_0 \) up to a positive scaling factor. The GE family includes as particular cases generalized Gaussian distribution, compound Gaussian, elliptical and many others [22]. Therefore, it has been successfully used to replace the SCM in many applications such as anomaly detection in wireless sensor networks [23], antenna array processing [24] and radar detection [25]–[28].

Performance analysis of Tyler’s estimator dates back to the original work in classical robust statistics literature [4]. In the classical asymptotic regime where \( p \) is fixed and \( n \to \infty \) it was shown that, when properly scaled, Tyler’s estimator is strongly consistent with the true covariance matrix, if it exists, and is asymptotically normally distributed. Its asymptotic variance, which coincides with the Cramer-Rao bound, was analyzed in [29], [30]. RMT regime performance results were also reported. For the case \( n, p \to \infty, p/n \to 0 \) it was shown in [31] that the condition number of Tyler’s estimator multiplied by the inverse true shape matrix of the underlying distribution converges to \( 1 + O(\sqrt{p/n}) \). In [32] it was demonstrated that the empirical spectral distribution of \( \sqrt{n/p}(T - I) \) converges to the semicircle law in probability, when the underlying population is white and Tyler’s estimator is properly scaled.

In the paper [33] the authors proved that the difference between the SCM and a scaled version of Tyler’s M-estimator converges to zero in operator norm, as \( n, p \to \infty, p/n \to \zeta \in (0, 1) \), and the data samples follow the standard normal distribution. The authors extended this result to ellipsoidal distributions and proved that the empirical spectral density of Tyler’s M-estimator converges to the Marchenko-Pastur distribution. The paper [33] also quantified the non-asymptotic behavior of Tyler’s M-estimator, but in a different way which emphasizes the RMT regime, the Marchenko-Pastur law, and comparison to the SCM. The paper [34] analyzed the asymptotic behavior of regularized Tyler’s estimator in the RMT regime and allowed optimal parameter tuning. Additional RMT results on other M-estimators were developed in [35], [36].

We focus on the non-asymptotic analysis of Tyler’s estimator for moderate values of \( n \) and \( p \) based on the concentration of measure phenomenon. Note that the estimator is not given in closed form and has to be iteratively computed using the fixed point iteration (1). In order to exploit the optimization based machinery discussed above, we rely on an alternative derivation of Tyler’s estimator. In particular, the estimator can also be obtained as an MLE of normalized GE distributed vectors defined as \( \frac{x_i}{\|x_i\|_2} \) [22]. Our method is therefore based on the analysis of the negative log-likelihood function of this distribution. We present high probability error bounds on the deviation of a properly scaled Tyler’s estimator from the true underlying shape matrix. Namely, we prove that as long as \( n \) is larger than \( p \) the Frobenius norm of the error in inverse matrices decays like \( \frac{1}{\sqrt{n}} \) with high probability. We also show that our performance bounds exhibit correct asymptotic behavior in the RMT regime. This contribution basically complements the previous classical asymptotic and RMT analysis of Tyler’s method. The derivation generally follows the optimization based approach due to [3], [19].

The paper is organized as following: first we introduce notations, state the problem, the main result and provide a discussion of it. Then we outline the strategy of the proof and bring in a few auxiliary results. After this we prove the main theorem. The body of the article contains the sketch of the proof with the statements of the most significant lemmas. Finally, we provide numerical simulations illustrating the obtained results. The proofs of lemmas and the rest of the auxiliary statements are left for the Appendices.

### A. Notations

Denote by \( S(p) \) the linear space of \( p \times p \) symmetric real matrices and by \( P(p) \subset S(p) \) the closed cone of positive semi-definite matrices. \( I \) stands for the identity matrix of a proper dimension.

We endow \( S(p) \) with the scalar product \( \langle A, B \rangle = \text{Tr}(AB) \), which induces the Frobenius norm on it. \( \| \cdot \| \) will denote the Euclidean norm for vectors, \( \| \cdot \|_F \) —the Frobenius norm and \( \| \cdot \|_2 \) —the spectral norm for matrices. The linear space \( \mathbb{H}^p \) is treated as a column vector space with the standard inner product. Given a matrix \( \mathbf{A} \) we denote by \( \text{vec}(\mathbf{A}) \) a column vector obtained by stacking the columns of \( \mathbf{A} \).

For a matrix \( \mathbf{A} \in P(p) \) denote by \( \lambda_{\text{min}}(\mathbf{A}) \) and \( \lambda_{\text{max}}(\mathbf{A}) \) its minimal and maximal eigenvalues, correspondingly. When \( \lambda_{\text{min}}(\mathbf{A}) > 0 \) we write \( \mathbf{A} \succ 0 \) and denote by \( \kappa(\mathbf{A}) = \frac{\lambda_{\text{max}}(\mathbf{A})}{\lambda_{\text{min}}(\mathbf{A})} \) its condition number. \( |\mathbf{A}| \) stands for the determinant of \( \mathbf{A} \).

Let \( Q \) be a quadratic form over a finite dimensional Euclidean space \( \mathcal{V} \), define its minimal and maximal eigenvalues as

\[
\lambda_{\text{min}}(Q) = \inf_{a \in \mathcal{V}, \|a\| = 1} Q(a), \quad \lambda_{\text{max}}(Q) = \sup_{a \in \mathcal{V}, \|a\| = 1} Q(a).
\]

The norm of \( Q \) is defined as

\[
\|Q\|_2 = \max \{ |\lambda_{\text{min}}(Q)|, \lambda_{\text{max}}(Q) \}.
\]

For \( n \) instances \( a_1, \ldots, a_n \) of scalars, vectors, matrices or functions we denote by \( \bar{a} \) their arithmetic average, when the index of summation is obvious from the context.

Matrices are denoted by Capital bold letters \( \mathbf{M} \), column vectors by non-capital bold \( \mathbf{v} \), scalars by non-capital \( r \), operators and quadratic forms by Capital \( T \) letters.

### II. TYLER’S ESTIMATOR AS A MLE

We define Tyler’s estimator as an MLE of a shape matrix parameter of a specific real spherical \( p \) dimensional distribution. The likelihood function of this distribution is later used to derive error bounds of the estimator using its curvature properties.

**Definition 1:** Assume \( \Theta_0 \in P(p), \Theta_0 \succ 0 \), then the function

\[
p(x) = \frac{\Gamma(p/2)}{2\sqrt{\pi^p}} \frac{1}{\sqrt{\det(\Theta_0^{-1})}} x^T \Theta_0^{-1} x^{p/2}
\]
is a probability density function of a vector \( x \in \mathbb{R}^p \) laying on a unit sphere. This distribution is usually referred to as the Angular Central Gaussian (ACG) distribution on a sphere [37] and we denote it as \( x \sim \mathcal{U}(\Theta_0) \). The matrix \( \Theta_0 \) is referred to as a shape matrix of the distribution and is a multiple of the covariance matrix of \( x \).

The ACG distribution is closely related to the class of GE distributions, which includes Gaussian, compound Gaussian, elliptical, skew-elliptical, ACG and other distributions, [38]. An important property of the GE family is that the shape matrix of a population does not change when the vector is divided by its Euclidean norm [22], [38]. After normalization, any GE vector becomes ACG distributed. This allows us to treat all these distributions together using a single robust estimator.

Assuming \( \Theta \in \mathcal{P}(p), \Theta > 0 \) and given \( n > p \) i.i.d. copies of a vector \( x_i \sim \mathcal{U}(\Theta) \), \( i = 1, \ldots, n \) we derive the MLE estimator of the shape matrix. For this sake introduce the scaled negative log-likelihood function:

\[
\tilde{f}(\Theta, x) = \log |\Theta| + p \log(x^T \Theta^{-1} x).
\]  

The function (5) is non-convex in \( \Theta \). Nevertheless, its critical point given as the solution to (1) provides the global minima with probability one, [39]–[41].

The negative log-likelihood (5) is invariant under multiplication of the shape matrix by a positive constant, thus we are only interested in the estimation of the shape matrix up to a positive scalar factor. In order to obtain a unique MLE we fix the scale of the estimator by assuming that \( \Theta_0 \) of the true covariance matrix is known (or arbitrarily fixed). Specifically, we define Tyler’s estimator to be the solution to the program

\[
T = \arg \min_{\Theta} \frac{1}{n} \sum_{i=1}^{n} \tilde{f}(\Theta; x_i),
\]  

subject to \( \text{Tr}(\Theta^{-1}) = \text{Tr}(\Theta_0^{-1}). \)

III. THE MAIN RESULT

In this section we introduce the main result of the paper. Denote \( \Omega_0 = \Theta^{-1}_0, \Delta := \lambda_{\min}(\Theta_0) = \lambda_{\max}(\Omega_0) > 0 \) and

\[
\cos \phi_0 := \frac{\text{Tr}(\Omega_0)}{|T|_F |\Omega_0|_F} > 0.
\]  

Theorem 1: Assume we are given \( n > p \) i.i.d. copies of \( x \sim \mathcal{U}(\Theta_0) \), then for \( \theta \geq 0 \) with probability at least

\[
1 - 2 \exp \left( -\frac{\theta^2}{2 \left( 1 + 1.7 \frac{p}{\sqrt{n}} \right)} \right) - 2p^2 \exp \left( -\frac{n \cos^2 \phi_0}{77 \ln(7p) \left( 1 + \frac{2}{p} \right)} \right) \left( 1 + \frac{15 \cdot 10^3 (1 + \frac{2}{p})^4}{n^2 \cos^2 \phi_0} \right).
\]

We note that a similar bound can be derived for the SCM in Gaussian populations, suggesting that both estimators are bounded by a multiple of \( p \sqrt{n} \) with high probability.

A. Identity Covariance

The value of \( \cos \phi_0 \) is close to 1 if the condition number \( \kappa(\Theta_0) \) is close to 1 and gets close to 0 if the matrix \( \Theta_0 \) has many small eigenvalues and few large ones. As we know the estimation of the matrix becomes less stable in the latter case, as the theorem suggests. To make the statement of Theorem 1 easier to grasp let us treat the case of \( \Theta_0 = I \) in the following

Corollary 1: Assume we are given \( n > p \) i.i.d. copies of \( x \sim \mathcal{U}(I), \delta < \frac{\sqrt{n}}{4} \), then

\[
\mathbb{P} \left( \|T^{-1} - I\|_F > 10 \frac{p + 2}{\sqrt{n}} \right) < 2 \exp \left( -\frac{\theta^2}{3} \right) + 2p^2 \exp \left( -\frac{n \cos^2 \phi_0}{77 \ln(7p) \left( 1 + \frac{2}{p} \right)} \right) \left( 1 + \frac{15 \cdot 10^3 (1 + \frac{2}{p})^4}{n^2 \cos^2 \phi_0} \right).
\]

This corollary illustrates the essential behavior of the bound: as \( n \) and \( p \) get large the second probability term vanishes and we get a large deviation-type bound.

B. Choice of the Norm

Our result and the corresponding results of [3], [19] for the Gaussian settings are formulated in terms of the inverse of the estimated matrix. Actually the necessity to formulate the convergence rates in terms of inverse matrices is twofold. First, in many applications one is mostly interested in the inverse covariance matrix estimation, e.g. in regression and denoising. Second, this choice is also dictated by the properties of the negative log-likelihood function. It is more convenient for the analysis to parametrize this function by the inverse covariance matrix. In addition, the fact that the Frobenius and not the spectral norm is used in the bounds also deserves explanation. In the Gaussian case stronger bounds can be obtained with the spectral norm. In fact it can be shown that the rate of convergence of the SCM matrix to the true covariance is of order \( \sqrt{\frac{p}{n}} \) in the spectral norm, see [16] and references therein. In our case the problem of obtaining spectral norm non-asymptotic bounds is much more involved and remains an open question.

IV. PRELIMINARY RESULTS

In this section we give a few technical results which will be useful in the proof. The analysis becomes easier if the negative log-likelihood is parametrized by the inverse shape matrix. We denote \( \Omega = \Theta^{-1} \) and write

\[
f(\Omega; x) = \tilde{f}(\Omega^{-1}; x) = \log |\Omega| + p \log(x^T \Omega x).
\]
Slightly abusing the notations, below we refer to (11) as the negative log-likelihood of the ACG population.

The negative log-likelihood gradient and Hessian read as

\[
\nabla f_\Omega = -\Omega^{-1} + \frac{xx^T}{x^T \Omega x},
\]

\[
\nabla^2 f_\Omega = -\Omega^{-1} \otimes \Omega^{-1} - \frac{xx^T}{x^T \Omega x} \otimes \frac{xx^T}{x^T \Omega x}.
\]

These and their expectations can be considered as linear and quadratic forms over \(S(p)\) respectively:

\[
E[\nabla f_\Omega(U)] = -Tr(\Omega^{-1} U) + \frac{x^T U x}{x^T \Omega x},
\]

\[
E[\nabla^2 f_\Omega(U)] = Tr(\Omega^{-1} U \Omega^{-1} U) - p \left( \frac{x^T U x}{x^T \Omega x} \right)^2
\]

\[
- \left\| \Omega^{-1/2} U \Omega^{-1/2} \right\|_F^2 - p \left( \frac{x^T U x}{x^T \Omega x} \right)^2.
\]

Let us compute the expectations of the derivatives for \(x \sim U(\Theta_0)\). Denote

\[
T_\Omega = E(\nabla f_\Omega), \quad H_\Omega = E(\nabla^2 f_\Omega),
\]

then

\[
T_\Omega(U) = -Tr(\Omega^{-1} U) + pR^1(U, \Omega; \Theta_0),
\]

\[
H_\Omega(U) = Tr\left( (\Omega^{-1/2} U)^2 \right) - pR^2(U, \Omega; \Theta_0),
\]

where

\[
R^\nu(U, \Omega, \Theta_0) = E\left[ \left( \frac{x^T U x}{x^T \Omega x} \right)^\nu \right], \quad \nu \in \mathbb{N}.
\]

In particular, using the results in Appendix A, we obtain that at the true parameter value \(\Theta_0\),

\[
T_\Theta(U) = 0, \quad \forall U \in S(p),
\]

\[
H_\Theta(U) = \frac{pTr((\Theta_0 U)^2) - Tr(\Theta_0 U)^2}{p + 2}
- \frac{p \left\| \Theta_0^{1/2} U \Theta_0^{1/2} \right\|_F^2 - \left( Tr(\Theta_0^{1/2} U \Theta_0^{1/2}) \right)^2}{p + 2}.
\]

Our analysis is based on strong convexity of the Hessian. It can be easily shown that \(H_\Theta(U)\) has a one dimensional kernel spanned by \(\Omega_0\). In what follows we show that the trace constraint in (6) eliminates this direction and makes the Hessian strongly convex in the vicinity of \(\Omega_0\) with high probability.

In order to explore the convexity properties of the expected Hessian, \(H_\Theta(U)\), let us state the following simple

**Lemma 1**: Let \(V\) be a Euclidean space and \(S \subset V\) be its subspace of codimension one with normal vector \(n\), then for any \(v \in V\)

\[
\sin^2(\angle u, v) \geq \cos^2(\angle n, v) = \frac{(n, v)}{\|n\| \|v\|}, \quad \forall u \in S.
\]

**Proof**: Among all the vectors \(u \in S\) the one that minimizes the angle \(\langle n, v \rangle\) is coplanar with \(n\) and \(v\). Now the statement follows from Pythagoras theorem.

Let us now turn to the space \(S(p)\) and consider its subspace \(L\) defined by the condition \(Tr(U) = 0\). Apply to this subspace the following linear transformation:

\[
L' = \Theta_0^{1/2} L \Theta_0^{1/2}.
\]

\(\Theta_0\) is a normal vector of \(L'\), since \((\Theta_0^{1/2} U \Theta_0^{1/2}, \Theta_0) = Tr(U) = 0, \forall U \in L\). Recall that

\[
Tr(\Theta_0^{1/2} U \Theta_0^{1/2}) = \left( \Theta_0^{1/2} U \Theta_0^{1/2}, I \right)
- \sqrt{p} \left\| \Theta_0^{1/2} U \Theta_0^{1/2} \right\|_F \cos(\angle(\Theta_0^{1/2} U \Theta_0^{1/2}, I)).
\]

and set \(n = \Theta_0, v = I\) to apply Lemma 1 and get

\[
H_{\Theta_0}(U) = \frac{p}{p + 2} \sin^2(\angle(\Theta_0^{1/2} U \Theta_0^{1/2}, I)) \left\| \Theta_0^{1/2} U \Theta_0^{1/2} \right\|_F^2
\geq \frac{p}{p + 2} \sin^2(\angle(I, \Theta_0)) \left\| \Theta_0^{1/2} U \Theta_0^{1/2} \right\|_F^2, \quad \forall U \in L.
\]

For brevity denote

\[
\cos\phi_0 = \cos(\angle(I, \Theta_0)) = \frac{Tr(\Theta_0)}{\|I\|_F \|\Theta_0\|_F} = \frac{Tr(\Theta_0)}{\sqrt{p} \|\Theta_0\|_F} \geq \frac{1}{\kappa(\Theta)},
\]

as was already defined in (7). This quantity is closely related to the notion ofphericity, [42]. We see that \(\cos\phi_0\) defines the convexity properties of \(H_{\Theta_0}(U)\) restricted to \(L\) and thus plays a crucial role in the bound provided by Theorem 1.

**V. PROOF OF THE MAIN THEOREM**

We start by outlining the proof strategy, which follows a technique similar to that of [3], [19], [43]. The method of obtaining sample complexity rates proposed by these papers consists of finding the smallest ball around the true inverse covariance that contains the estimator with high probability. In its turn, this is done by considering the second order Taylor expansion of the sample average negative log-likelihood and estimating the maximal radius of the ball, for which the increase of this function on the boundary of the ball is positive with high probability.

The reasoning is as following: assume we are given a continuous function \(g\) over a finite dimensional Euclidean space which is known to have a unique local minimum. If we also know that \(g(x) < g(y)\), where \(B(x; r)\) is a closed ball of radius \(r\) around \(x\), then the unique minimum of the function belongs to the open ball \(B(x; r) \setminus \partial B(x; r)\). Indeed, \(g\) is continuous over a compact set \(B(x; r)\), thus reaches its extrema on it. The minimum of \(g\) cannot lie on the boundary since \(g(y) > g(x)\), so it is strictly inside the ball.

Strong convexity of \(g\) is a sufficient condition for it having a unique local minimum. For this purpose any notion of convexity is suitable, since the uniqueness of the minimum does not depend on the metric. The second condition \(g(x) < g(y)\) is usually demonstrated using strong local convexity of \(g\) in the
vicinity of the true parameter value. Due to the scale invariance, the Hessian of Tyler’s objective has a zero eigenvalue in the vicinity of the true parameter. The trace constraint addresses this invariance and ensures that the Hessian is positive definite.

We now provide the details of the proof. Denote
\[ \hat{f}_n = \frac{1}{n} \sum_{i=1}^{n} f(\Omega; x_i); \tag{26} \]
and let us use the Taylor polynomial formula with remainder in the Lagrange form. The expansion in the vicinity of the true inverse covariance matrix \( \Omega_0 \) reads as:
\[ \hat{f}_n - \hat{f}_{n_0} = \nabla \hat{f}_{n_0}(\Delta \Omega) + \frac{1}{2} \nabla^2 \hat{f}_{n_0}(\Delta \Omega), \tag{27} \]
where \( \Omega = \Omega_0 + \Delta \Omega, \Omega_0 = \Omega_0 + \alpha \Delta \Omega, \alpha \in [0,1] \) and \( \Delta \Omega \in \mathcal{L} \). The sample average negative log-likelihood (27) reads as:
\[ \mathbb{E} \left[ \sum_{i=1}^{n} f(\Omega; x_i) \right] \approx \frac{1}{2} \nabla^2 \hat{f}_{n_0}(\Delta \Omega), \tag{28} \]
where \( \Omega' = I + \Delta \Omega', \Omega_0' = I + \alpha \Delta \Omega', \Delta \Omega' \in \mathcal{L}' \). Since \( \epsilon' \) is a tension-compression map, the maximal and minimal distance changes are known and we can use an inequality
\[ \lambda \left\| T^{-1} - \Theta^{-1} \right\|_{F} \leq \left\| T'^{-1} - \Theta_0'^{-1} \right\|_{F}, \tag{29} \]
to establish high probability bounds for inverse Tyler’s estimator given the respective bounds on its \( \epsilon' \)-image.

The next lemmas provide high probability bounds on the gradient and Hessian.

**Lemma 2:** Uniformly over \( U' \in \mathcal{L}' \)
\[ \mathbb{P} \left( \nabla' \hat{f}(U') \geq \epsilon p \| U' \|_F \right) \leq 2 \exp \left( \frac{-nt^2}{2(1 + 1.7t)} \right), \forall t \geq 0. \tag{30} \]

**Proof:** Is provided in Appendix B. \( \square \)

We proceed by establishing concentration properties of the sample mean Hessian restricted to \( \mathcal{L}' \) at the point \( \Omega' \).

**Lemma 3:** Let \( \Omega' = I + \Delta \Omega', \Delta \Omega' \in \mathcal{L}' \) and \( \epsilon = \| \Delta \Omega' \|_2 < 1 \), then for \( \frac{1}{2 \sqrt{2 \ln(32 \sqrt{2} p^2)}} \leq \eta \leq \frac{1}{2 \sqrt{2} \eta} \) uniformly over \( U' \in \mathcal{L}' \)
\[ \mathbb{P} \left( \nabla^2 \hat{f}_n(U') \leq H_{\Omega'}(U') - \eta \left\| U' \right\|_F^2 \right) \leq 2 p^2 \exp \left( -\frac{n \eta \sqrt{2}}{8 \ln(32 \sqrt{2} \eta^2)} \right) \left( 1 + \frac{7}{n^2 \eta^4} \right). \tag{31} \]

**Proof:** Is provided in Appendix B. \( \square \)

This lemma shows that with probability depending on the parameters of the problem the sample average Hessian is not far from the expected Hessian at the same point. The next result shows the expected Hessian at a point \( \Omega' \) close to the true parameter is lower bounded by a fraction of the expected Hessian at the true parameter value.

**Lemma 4:** For \( \mathbb{E} \left[ \hat{f}_n \right] = I + \alpha \Delta \Omega', \alpha \in [0,1], \Delta \Omega' \in \mathcal{L}', \epsilon = \| \Delta \Omega' \|_2 \leq \frac{p}{6(p+2)} \cos^2(\phi_0) \) the expected Hessian \( H_{\Omega'}(\Delta \Omega') \) is bounded from below by
\[ H_{\Omega'}(\Delta \Omega') \geq \frac{p}{4(p+2)} \cos^2(\phi_0) \| \Delta \Omega' \|_F^2. \tag{32} \]

**Proof:** Is provided in Appendix B. \( \square \)

**Corollary 2:** Under the conditions of Lemma 4 for \( \tau \) satisfying \( \frac{1 + 1/p^2}{\ln(32 \sqrt{2} p^2)} \leq \tau \leq \frac{1}{p^2} \) \( \tau \cos^2(\phi_0) \leq 1 \)
\[ \mathbb{P} \left( \frac{1}{2} \nabla^2 \hat{f}_n(\Delta \Omega') \leq (1 - \tau) \frac{p}{4(p+2)} \cos^2(\phi_0) \| \Delta \Omega' \|_F^2 \right) \leq 2 p^2 \exp \left( -\frac{n \tau \cos^2(\phi_0)}{46 \ln(7p)} \left( 1 + \frac{2}{p} \right) \right) \left( 1 + \frac{2 \cdot 10^3 \left( 1 + \frac{2}{p} \right)}{n^2 \tau^4 \cos^4(\phi_0)} \right). \tag{33} \]

**Proof:** Is provided in Appendix B. \( \square \)

It now follows that with probability at least
\[ 1 - 2 \exp \left( -\frac{-nt^2}{2(1 + 1.7t)} \right), \tag{34} \]
\[ -2p^2 \exp \left( -\frac{-n \tau \cos^2(\phi_0)}{46 \ln(7p)} \left( 1 + \frac{2}{p} \right) \right) \left( 1 + \frac{2 \cdot 10^3 \left( 1 + \frac{2}{p} \right)}{n^2 \tau^4 \cos^4(\phi_0)} \right), \]
\[ \hat{f}_n - \hat{f}_1 \geq -\left\| \Delta \Omega' \right\|_F + (1 - \tau) \frac{p}{4(p+2)} \cos^2(\phi_0) \| \Delta \Omega' \|_F^2. \tag{35} \]

Demand positivity of the right-hand side to get
\[ \| \Delta \Omega' \|_F \geq \frac{4t}{1 - \tau \cos^2(\phi_0)}. \tag{36} \]

Set \( \tau = \frac{3}{4}, t = \frac{p}{\sqrt{n}} \) and use (29) to get the statement of the theorem.

**VI. NUMERICAL RESULTS**

In this section we provide numerical simulations supporting our analysis. Fig. 1 compares the behavior of Tyler’s estimator with its 0.95 and 0.5-probability bounds for \( p = 50, \Theta_0 = I \).

For a given \( n \) Tyler’s estimator error bound was obtained by
minimizing (35) with respect to \( t \) and \( \tau \) under the probability constraint given by equating (34) to 0.95 or 0.5 respectively.

Fig. 2 verifies the dependence of the performance on the dimension.

APPENDIX A

MOMENTS OF RATIOS OF QUADRATIC FORMS

Lemma 5: Let \( x' \sim \mathcal{U}(1) \), \( U' \in \mathcal{S}(p) \), \( \Omega' = I + \Delta \Omega' > 0 \) and \( \varepsilon = \| \Delta \Omega' \|_2 < 1 \), then the moments defined in (19) satisfy

\[
R^\nu(\Delta \Omega', 1 + \alpha \Delta \Omega'; 1; I) \leq R^\nu(U', \Omega'; I; I) \leq \frac{R^\nu(U', 1; I; I)}{(1 - \varepsilon)^\nu}, \quad \nu \in \mathbb{N}. \tag{36}
\]

In addition

\[
\left| R^\nu(\Delta \Omega', 1 + \alpha \Delta \Omega'; I; I) - R^\nu(\Delta \Omega', I; I) \right| + \nu \alpha R^{\nu+1}(\Delta \Omega', I; I) \\
\leq \frac{\nu(\nu + 1)\alpha^2}{2(1 - \varepsilon)^{\nu+2}} R^{\nu+2}(\Delta \Omega', I; I). \tag{47}
\]

Proof: Consider

\[
R^\nu(U', \Omega'; I) = \mathbb{E} \left[ \left( \frac{x'^T U' x'}{x'^T (I + \Delta \Omega') x'} \right)^\nu \right], \tag{37}
\]

since

\[
1 + \varepsilon \geq 1 + x'^T \Delta \Omega' x' \geq 1 - \varepsilon, \tag{38}
\]

the variable under expectation in (37) can be bounded as

\[
\left( \frac{x'^T U' x'}{1 + \varepsilon} \right)^\nu \leq \left( x'^T U' x' (1 + x'^T \Delta \Omega' x')^{-1} \right)^\nu \leq \left( \frac{x'^T U' x'}{1 - \varepsilon} \right)^\nu. \tag{39}
\]

By taking the expectations, (36) follows. We continue

\[
R^\nu(\Delta \Omega', 1 + \alpha \Delta \Omega'; I) = \mathbb{E} \left[ \left( \frac{x'^T \Delta \Omega' x'}{x'^T (I + \alpha \Delta \Omega') x'} \right)^\nu \right] \\
= \mathbb{E} \left[ (x'^T \Delta \Omega' x' (1 + \alpha x'^T \Delta \Omega' x')^{-1})^\nu \right] \\
= \mathbb{E} \left[ (x'^T \Delta \Omega' x')^\nu (1 - \nu \alpha x'^T \Delta \Omega' x' + g(\alpha \Delta \Omega', x')) \right], \tag{40}
\]

where \( g(\alpha \Delta \Omega', x') \) is bounded by

\[
|g(\alpha \Delta \Omega', x')| \leq \nu(\nu + 1)\alpha^2 (x'^T \Delta \Omega' x')^2 \frac{2}{(1 - \alpha \varepsilon)^{\nu+2}}. \tag{41}
\]

We finally obtain

\[
\left| R^\nu(\Delta \Omega', 1 + \alpha \Delta \Omega'; I) - R^\nu(\Delta \Omega', I; I) + \nu \alpha R^{\nu+1}(\Delta \Omega', I; I) \right| \\
\leq \frac{\nu(\nu + 1)\alpha^2}{2(1 - \alpha \varepsilon)^{\nu+2}} R^{\nu+2}(\Delta \Omega', I; I). \tag{42}
\]

Given \( x' \sim \mathcal{U}(1) \), we can always represent it as \( x' = \frac{x}{\sqrt{p}} \), where \( x' \) is standard normally distributed \( x' \sim \mathcal{N}(0, I) \), to obtain

\[
\frac{x'^T U' x'}{x'^T \Omega' x'} = \frac{x'^T U' x'}{x'^T \Omega' x'} = \frac{x'^T U' x'}{x'^T \Omega' x'}. \tag{43}
\]

using this identity, we develop formulae for \( R^\nu(U', I; I) \), [44]:

\[
R^1(U', I; I) = -\frac{\text{Tr}(U')}{p}; \tag{44}
\]

\[
R^2(U', I; I) = \frac{\left( \text{Tr}(U')\right)^2 + 2\text{Tr}(U'^2)}{p(p + 2)} \\
- \frac{(\text{Tr}(U')^2 + 2 \text{Tr}(U'^2))}{p(p + 2)} \\
\frac{p(p + 2)}{p(p + 2)(p + 4)}. \tag{45}
\]

\[
R^3(U', I; I) = \frac{\text{Tr}(U'^3) - 6 \text{Tr}(U'^2) \text{Tr}(U') + 3 \text{Tr}(U'^4)}{p(p + 2)(p + 4)} \tag{46}
\]

A general moment bound can also be obtained as

\[
R^\nu(U', I; I) \leq \frac{\left( \frac{p}{2} \right)! \| U' \|_p^\nu}{\sqrt{p}}, \quad \nu = 2, 4, \ldots. \tag{47}
\]
APPENDIX B

CONCENTRATION BOUNDS

Lemma 6 (Vector Bernstein Inequality, [45]): Let \( \mathbf{x}_1, \ldots, \mathbf{x}_n \in \mathbb{R}^k \) be i.i.d zero-mean random vectors and suppose there exist \( \sigma, \Lambda > 0 \) such that

\[
E \| \mathbf{x}_i \|^r \leq \frac{r!}{2} \sigma^2 \Lambda^{r-2}, r = 2,3,\ldots,
\]
then for \( t \geq 0 \)

\[
P \left( \| \mathbf{\bar{x}} \| > t \sigma \right) \leq 2 \exp \left( \frac{-nt^2}{2 \left( 1 + 1.7t \frac{\sigma}{\Lambda} \right)} \right).
\]

Lemma 7 (Matrix Bernstein Inequality, [46]): Let \( \mathbf{S}_1, \ldots, \mathbf{S}_n \in \mathbb{S}(p) \) be i.i.d zero-mean random matrices and suppose there exist \( \sigma, \Lambda > 0 \) such that

\[
E \| \mathbf{S}_1 \|^r \leq \frac{r!}{2} \sigma^2 \Lambda^{r-2}, r = 2,3,\ldots,
\]
then for \( t > \frac{\sigma^2}{4 \Lambda \ln \left( \frac{1+1/p^2}{\nu^2} \right)} \)

\[
P \left( \lambda_{\max} (\mathbf{\bar{S}}) > t \sigma \right) \leq 2p^2 \exp \left( \frac{-nt^2}{8 \Lambda \ln \left( \frac{1+1/p^2}{\nu^2} \right)} \left( 1 + \frac{6}{n^2 \nu^2 \sigma^2 \ln \left( 1 + \frac{6}{n^2 \nu^2 \sigma^2} \right)} \right) \right).
\]

Lemma 8: Let \( x'_i \sim \mathcal{U}(1), i = 1,\ldots,n \) then for \( t > 0 \)

\[
P \left( \| p x'_i x'_i^T - I \|_F > t \right) \leq 2 \exp \left( \frac{-nt^2}{2(1 + 1.7t)} \right)
\]

Proof: Define \( n \) centered random vectors

\[
\mathbf{x}_i = \text{vec} \left( x'_i x'_i^T - \frac{1}{p} \right) \in \mathbb{R}^{p^2}, i = 1,\ldots,n,
\]
and consider powers of their norms

\[
\| x_i \|^r = \left[ \text{Tr} \left( \left( x'_i x'_i^T - \frac{1}{p} \right)^2 \right) \right]^{\frac{r}{2}} = \left[ \text{Tr} \left( 1 - \frac{2}{p} x'_i x'_i^T + \frac{1}{p^2} I \right) \right]^{\frac{r}{2}} = \left( 1 - \frac{1}{p} \right)^{\frac{r}{2}},
\]
which are deterministic quantities. Set

\[
\sigma = \Lambda = 1,
\]
and apply Lemma 6 to obtain

\[
P \left( \| \mathbf{\bar{X}} \| > t \right) \leq 2 \exp \left( \frac{-nt^2}{2(1 + 1.7t)} \right).
\]

Multiply \( \mathbf{x}' \) by \( p \) to get the statement.

Proof of Lemma 2:

\[
\left| \nabla f_i (\mathbf{U}') \right| = \left| \text{Tr} \left( \left( px'_i x'_i^T - I \right) \mathbf{U}' \right) \right|.
\]

Apply Lemma 8 and the Cauchy-Schwarz inequality to get the statement.

Proof of Lemma 3: For a linear operator \( L \), its restriction \( \tilde{L} \) to a linear subspace satisfies

\[
\| \tilde{L} \|_2 \leq \| L \|_2,
\]
thus we can apply Lemma 7 to bound the deviation of \( \nabla^2 f_i \) restricted to \( \mathcal{L}' \) from its expectation \( \nabla^2 f_i \) also restricted to \( \mathcal{L}' \).

Proof of Lemma 4: From (18) we have

\[
H_{\Omega'}(\Delta \Omega') = \text{Tr} \left( \left( \Omega'^{-1} \Delta \Omega' \right)^2 \right) - p R^2 (\Delta \Omega', \Omega'; i).
\]
Using Lemma 5 we obtain
\[ H_{\hat{\Omega}^2}(\Delta \Omega'') = \text{Tr} \left( \left( (1 + \alpha \Delta \Omega'')^{-1} \Delta \Omega'' \right)^2 \right) - p R^2(\Delta \Omega', 1 + \alpha \Delta \Omega'; 1) \]
\[ = \text{Tr} (\Delta \Omega''^2 - 2 \alpha \Delta \Omega''^3 + h(\alpha \Delta \Omega'') \Delta \Omega''^2) - p \left[ R^2(\Delta \Omega', 1; 1) - 2 \alpha R^3(\Delta \Omega', 1; 1) + l(\alpha \Delta \Omega'') \right]. \quad (64) \]

Use the Lagrange remainder form for \( h(\alpha \Delta \Omega'') \) and apply Lemma 5 to bound \( l(\alpha \Delta \Omega'') \) and get
\[ h(\alpha \Delta \Omega'') \leq \frac{3(\alpha \epsilon)^2}{(1 - \alpha \epsilon)^4}, \quad (65) \]
\[ l(\alpha \Delta \Omega'') \leq \frac{3\alpha^2}{(1 - \alpha \epsilon)^4} R^4(\Delta \Omega', 1; 1). \quad (66) \]

Use formulae (45) and (46) to get
\[ H_{\hat{\Omega}^2}(\Delta \Omega'') = \| \Delta \Omega'' \|_F^2 - 2 \alpha \text{Tr}(\Delta \Omega''^3) - \frac{(\text{Tr}(\Delta \Omega'')^2 + 2 \| \Delta \Omega'' \|_F^2)}{p + 2} \]
\[ + 2 \alpha \left( \frac{\text{Tr}(\Delta \Omega'')^3 + 6 \text{Tr}(\Delta \Omega'') \| \Delta \Omega'' \|_F^2 + 8 \text{Tr}(\Delta \Omega''^3)}{(p + 2)(p + 4)} \right) \]
\[ + r(\alpha \Delta \Omega'') \| \Delta \Omega'' \|_F^2 \]
\[ = H_{\hat{\Omega}^2}(\Delta \Omega') + r(\alpha \Delta \Omega'') \| \Delta \Omega'' \|_F^2 \]
\[ - 2 \alpha \left( \frac{p^4 + 6p^2 \text{Tr}(\Delta \Omega''^3)}{(p + 2)(p + 4)} - \text{Tr}(\Delta \Omega'')^3 - 6 \text{Tr}(\Delta \Omega'') \| \Delta \Omega'' \|_F^2 \right) \]
\[ \geq \left( \frac{p^2 - 2 \alpha \| \Delta \Omega'' \|_F^2 + r(\alpha \Delta \Omega'') \| \Delta \Omega'' \|_F^2}{p + 2} \right) \| \Delta \Omega'' \|_F^2. \quad (67) \]

where due to the condition \( \| \Delta \Omega'' \|_F = \epsilon \leq \frac{1}{6} \),
\[ r(\alpha \Delta \Omega'') \leq \frac{3(\alpha \epsilon)^2}{(1 - \alpha \epsilon)^4} + \frac{3\alpha^2}{(1 - \alpha \epsilon)^2} \frac{2 \| \Delta \Omega'' \|_F^2}{p^2} \]
\[ = \frac{(\alpha \epsilon)^2 (3 + \epsilon \epsilon p^2)}{(1 - \alpha \epsilon)^4} \leq \epsilon^2 \left( \frac{3 + \epsilon^2}{1 - \epsilon^2} \right) \leq \frac{1}{6}. \quad (68) \]

Here we have applied the bound from (47) to \( R^4(\Delta \Omega', 1; 1) \). Finally,
\[ 2 \alpha \| \Delta \Omega'' \|_F \leq 2 \epsilon \leq \frac{\epsilon}{3} \quad (69) \]
\[ H_{\hat{\Omega}^2}(\Delta \Omega'') \geq \frac{p}{2(p + 2)} \cos^2 \phi_0 \| \Delta \Omega'' \|_F^2. \quad (70) \]

**Proof of Corollary 2:** Set
\[ \frac{1}{2 \sqrt{2 \ln(32 \sqrt{2})}} \leq \eta = \frac{1}{\sqrt{2}} \frac{\epsilon}{2(p + 2)} \cos^2 \phi_0 \leq \frac{1}{2 \sqrt{2}}. \quad (71) \]

**ACKNOWLEDGMENT**

The authors would like to thank the associate editor and anonymous reviewers whose input greatly improved the paper.

**REFERENCES**

[1] O. Ledoit and M. Wolf, “Improved estimation of the covariance matrix of stock returns with an application to portfolio selection,” J. Empir. Finance, vol. 10, no. 5, pp. 603–621, 2003.

[2] P. J. Bickel and E. Levina, “Regularized estimation of large covariance matrices,” Ann. Statist., pp. 199–227, 2008.

[3] A. J. Rothman, P. J. Bickel, E. Levina, and J. Zhu, “Sparse permutation invariant covariance estimation,” Electron. J. Statist., vol. 2, pp. 494–515, 2008.

[4] D. E. Tyler, “A distribution-free M-estimator of multivariate scatter,” Ann. Statist.

[5] V. A. Marčenko and L. A. Pastur, “Distribution of eigenvalues for some sets of random matrices,” Sbornik: Math., vol. 1, no. 4, pp. 457–483, 1967.

[6] C. A. Tracy and H. Widom, “On orthogonal and symplectic matrix ensembles,” Commun. Math. Phys., vol. 177, no. 3, pp. 727–754, 1996.

[7] N. El Karoui, “Tracy-Widom limit for the largest eigenvalue of a large class of complex sample covariance matrices,” Ann. Probabil., vol. 35, no. 2, pp. 663–714, 2007.

[8] Z. Bai and J.-F. Yao, “Limit theorems for sample eigenvalues in a generalized spiked population model,” 2008, arXiv preprint arXiv:0806.1141 [Online]. Available: http://arxiv.org/pdf/0806.1141v1.pdf

[9] J. Baik and J. W. Silverstein, “Eigenvalues of large sample covariance matrices of spiked population models,” Ann. Statist., vol. 35, no. 2, pp. 663–714, 2007.

[10] L. S. Cardoso, M. Debbah, P. Bianchi, and J. Najim, “Cooperative spectrum sensing using random matrix theory,” in Proc. IEEE 3rd Int. Symp. Wireless Pervas. Comput., 2008, pp. 334–338.

[11] R. Couillet and W. Bachem, “Fluctuations of spiked random matrix models and failure diagnosis in sensor networks,” IEEE Trans. Inf. Theory, vol. 59, no. 1, pp. 509–525, 2012.

[12] X. Mestre and M. A. Lagunas, “Modified subspace algorithms for DoA estimation with large arrays,” IEEE Trans. Signal Process., vol. 56, no. 2, pp. 598–614, 2008.

[13] R. Schmidt, “Multiple emitter location and signal parameter estimation,” IEEE Trans. Antennas Propag., vol. 34, no. 3, pp. 276–286, 1986.

[14] R. Couillet, J. W. Silverstein, Z. Bai, and M. Debbah, “Eigen-inference for energy estimation of multiple sources,” IEEE Trans. Inf. Theory, vol. 57, no. 4, pp. 2420–2439, 2011.

[15] P. Vallet, P. Loubaton, and X. Mestre, “Improved subspace estimation for multivariate observations of high dimension: The deterministic signals case,” IEEE Trans. Inf. Theory, vol. 58, no. 2, pp. 1043–1066, 2012.

[16] R. Vershynin, “Introduction to the non-asymptotic analysis of random matrices,” in Compressed Sensing: Theory and Applications, Y. Eldar and G. Kutyniok, Eds. Cambridge: Cambridge Univ. Press, 2012.

[17] R. Vershynin, “How close is the sample covariance matrix to the actual covariance matrix?,” J. Theor. Probab., vol. 25, no. 3, pp. 655–686, 2012.

[18] N. Srivastava and R. Vershynin, “Covariance estimation for distributions with \( 2 + \epsilon \) moments,” 2011, arXiv preprint arXiv:1106.2775 [Online]. Available: http://arxiv.org/pdf/1106.2775v4.pdf
[19] P. Ravikumar, M. J. Wainwright, G. Raskutti, and B. Yu, “High-dimen-
sional covariance estimation by minimizing 11-penalized log-determi-
nant divergence,” Electron. J. Statist., vol. 5, pp. 935–980, 2011.
[20] P. J. Huber, “Robust estimation of a location parameter,” Ann. Math.
Statist., vol. 35, no. 1, pp. 73–101, 1964.
[21] R. A. Maronna, “Robust M-estimators of multivariate location and
scatter,” Ann. Statist., pp. 51–67, 1976.
[22] G. Frahm, Generalized Elliptical Distributions: Theory and Applica-
tions Univ. Köln, 2004.
[23] Y. Chen, A. Wiesel, and A. O. Hero, “Robust shrinkage estimation of
high-dimensional covariance matrices,” IEEE Trans. Signal Process.,
vol. 59, no. 9, pp. 4097–4107, 2011.
[24] E. Ollila and V. Koivunen, “Robust antenna array processing using M-
estimators of pseudo-covariance,” in Proc. IEEE 14th Pers., Indoor
Mobile Radio Commun., 2003, vol. 3, pp. 2659–2663.
[25] Y. I. Abramovich, N. K. Spencer, and M. D. Turley, “Time-varying
autoregressive (TVAR) models for multiple radar observations,” IEEE
Trans. Signal Process., vol. 55, no. 4, pp. 1298–1311, 2007.
[26] E. Ollila, D. Tyler, V. Koivunen, and H. Poor, “Complex ellip-
tically symmetric distributions: New results and applications,” IEEE
Trans. Signal Process., vol. 60, no. 11, pp. 5597–5625, 2012.
[27] F. Bandiera, O. Besson, and G. Ricci, “Knowledge-aided covariance
matrix estimation and adaptive detection in compound Gaussian noise,”
IEEE Trans. Signal Process., vol. 58, no. 10, pp. 5391–5396, 2010.
[28] 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.
Signal Process., vol. 56, no. 1, pp. 34–48, 2008.
[29] O. Besson and Y. Abramovich, “On the Fisher information matrix for
multivariate elliptically contoured distributions,” IEEE Signal Process.
Lett., vol. 20, no. 11, pp. 1130–1133, 2013.
[30] M. Greco and F. Gini, “Cramér-Rao lower bounds on covariance ma-
trix estimation for complex elliptically symmetric distributions,” IEEE
Trans. Signal Process., vol. 61, no. 24, pp. 6401–6409, 2013.
[31] L. Dümbgen, “The asymptotic behavior of Tyler’s M-estimator of
scatter in high dimension,” Ann. Inst. Statist. Math, vol. 50, pp. 471–491, 1997.
[32] G. Frahm and K. Glombek, “Semicircle law of Tyler’s M-estimator
for scatter,” Statist. Probab. Lett., vol. 82, no. 5, pp. 959–964, 2012.
[33] T. Zhang, X. Cheng, and A. Singer, “Marchenko–Pastur law for Tyler’s
and Maronna’s M-estimators,” 2014, arXiv preprint arXiv:1401.3424
[Online]. Available: http://arxiv.org/pdf/1401.3424v3.pdf
[34] R. Couillet and M. R. McKay, “Large dimensional analysis and op-
timization of robust shrinkage covariance matrix estimators,” 2014,
arXiv preprint arXiv:1401.4083 [Online]. Available: http://arxiv.org/
pdf/1401.4083v2.pdf
[35] R. Couillet, F. Pascal, and J. W. Silverstein, “The random matrix
regime of Maronna’s M-estimator with elliptically distributed sam-
ples,” 2013, arXiv preprint arXiv:1311.7034 [Online]. Available:
http://arxiv.org/pdf/1311.7034v1.pdf
[36] R. Couillet, F. Pascal, and J. W. Silverstein, “Robust M-estimation
for array processing: A random matrix approach,” [Online]. Available:
http://www4.ncsu.edu/~jack/robust_es6t.pdf
[37] D. E. Tyler, “Statistical analysis for the angular central Gaussian distri-
bution on the sphere,” Biometrika, vol. 74, no. 3, pp. 579–589, 1987.
[38] G. Frahm and U. Jaeckel, “Tyler’s M-estimator, random matrix theory,
generalized elliptical distributions with applications to finance,” Tech.
Rep., 2007 [Online]. Available: http://ideas.repec.org/p/zbw/ucdpse/
207.html
[39] A. Wiesel, “Geodesic convexity and covariance estimation,” IEEE
Trans. Signal Process., vol. 60, no. 12, pp. 6182–6189, 2012.
[40] A. Wiesel, “Unified framework to regularized covariance estimation in
scaled Gaussian models,” IEEE Trans. Signal Process., vol. 60, no. 1,
pp. 29–38, 2012.
[41] T. Zhang, A. Wiesel, and M. S. Greco, “Multivariate generalized
Gaussian distribution: Convexity and graphical models,” IEEE Trans.
Signal Process., vol. 61, no. 16, pp. 4141–4148, 2013.
[42] O. Ledoit and M. Wolf, “Some hypothesis tests for the covariance ma-
trix when the dimension is large compared to the sample size,” Ann.
Statist., pp. 1081–1102, 2002.
[43] J. K. Bradley and C. Guestrin, “Sample complexity of composite like-
lihood,” in Proc. Int. Conf. Artif. Intell. Statist., 2012, pp. 136–160.
[44] J. G. De Gooijer, “Exact moments of the sample autocorrelations from
series generated by general ARIMA processes of order (p, d, q), d =
0 or 1,” J. Econometr., vol. 14, no. 3, pp. 365–379, 1980.
[45] V. Yurinskii, “Exponential inequalities for sums of random vectors,”
J. Multivar. Anal., vol. 6, no. 4, pp. 473–499, 1976.
[46] S. Minsker, “On some extensions of Bernstein’s inequality for self-adj-
joint operators,” 2011, arXiv preprint arXiv:1112.5448 [Online]. Avail-
able: http://arxiv.org/pdf/1112.5448v2.pdf

Ilya Soloveychik (SM’13) received the B.Sc. degree in applied mathematics and physics from the Moscow Institute of Physics and Technology, in 2007, and the M.Sc. degree in mathematics from the Hebrew Uni-
versity of Jerusalem, in 2013. He is currently a Ph.D. student with the Rachel and Selim Benin School of Computer Science and Engi-
neering, the Hebrew University of Jerusalem, Israel, under Prof. A. Wiesel. Dr. Soloveychik was the Prize recipient of the Rus-
sian Olympiad in Physics in 2003. He received the

Ami Wiesel received the B.Sc. and M.Sc. degrees in electrical engineering from Tel-Aviv University, Tel-Aviv, Israel, in 2000 and 2002, respectively, and the Ph.D. degree in electrical engineering from the Technion-Israel Institute of Technology, Haifa, Israel, in 2007. He was a postdoctoral fellow with the Department of Electrical Engineering and Com-
puter Science, University of Michigan, Ann Arbor, in 2007–2009. Since January 2010, he has been a faculty member at the Rachel and Selim Benin School of Computer Science and Engineering at the Hebrew University of Jerusalem, Israel.

Dr. Wiesel was a recipient of the Young Author Best Paper Award for a 2006 paper in the IEEE TRANSACTIONS ON SIGNAL PROCESSING and a Student Paper Award for a 2005 Workshop on Signal Processing Advances in Wireless Com-
munications (SPAWC) paper. He was awarded the Weinstein Study Prize in 2002, the Intel Award in 2005, the Viterbi Fellowship in 2005 and 2007, and the Marie Curie Fellowship in 2008.