Geodesic Convexity and Covariance Estimation

Ami Wiesel, *Member, IEEE*

Abstract—**Geodesic convexity is a generalization of classical convexity which guarantees that all local minima of g-convex functions are globally optimal. We consider g-convex functions with positive definite matrix variables, and prove that Kronecker products, and logarithms of determinants are g-convex. We apply these results to two modern covariance estimation problems: robust estimation in scaled Gaussian distributions, and Kronecker structured models. Maximum likelihood estimation in these settings involves non-convex minimizations. We show that these problems are in fact g-convex. This leads to straight forward analysis, allows the use of standard optimization methods and paves the road to various extensions via additional g-convex regularization.**

Index Terms—**Elliptical distributions, geodesic convexity, Kronecker models, log-sum-exp, martix variate models, robust covariance estimation.**

I. INTRODUCTION

CONVEX optimization has emerged as a powerful signal processing tool with a growing variety of applications. Formulating an optimization problem in convex form is advantageous both from a theoretical perspective and from a numerical perspective. Convex optimization methods can efficiently find global solutions to large scale problems. These problems can also be easily generalized by incorporating additional convex constraints. In order to enjoy these benefits, there is an ongoing search for new classes of convex functions, and an understanding of their modeling power. In this paper, we consider a generalized form of convexity known as geodesic convexity (g-convexity) which is associated with positive definite matrix variables. Its main property is that local minimas of g-convex functions are also globally optimal. We define new classes of g-convex functions and consider their application to modern covariance estimation problems.

Two fundamental functions in convex analysis are the exponential and the logarithm of a sum of exponentials. These functions are the core ingredient in geometric programming where positive variables are modeled as exponentials, e.g., in power control in communication theory [2], [3]. We extend these results and consider Kronecker products and logarithms of determinants of positive definite matrices. We no longer use exponentials to model these matrices. Instead, we provide a brief

review to g-convexity [4]–[6], and prove that the functions are convex in this generalized notion.

Our first application for the g-convexity results is in the context of robust covariance estimation. We consider Tyler's method for scatter matrix estimation in scaled Gaussian models, including Elliptical distributions, generalized Elliptical distributions, spherically invariant random vectors and compound Gaussian processes [7]–[13]. This method has been successfully applied to different practical applications ranging from array processing to sensor networks. It has been generalized to other settings involving regularization [14]–[18] and incomplete data [19]. Additional recent contributions addressing covariance estimation in non Gaussian conditions include [20].

Maximum likelihood (ML) covariance estimation in scaled Gaussian models involves a non-convex minimization. Nonetheless, different works proved that the global solution can be efficiently found via simple fixed point iterations [7], [10]. Analysis of regularized solutions has been addressed in [16]. Recently, [17], [21] proved that the negative-log-likelihood is g-convex in the unknown covariance. Following these works, we show that the problem is in fact jointly g-convex in both the covariance and the scaling factors. These results provide more insight on the analysis and design of robust covariance estimation methods, and pave the road to numerous extensions. As an example, we demonstrate how additional prior knowledge on the covariance or the scaling can be exploited via g-convex regularization.

Our second application is in the context of structured covariance estimation. A standard approach to high dimensional covariance estimation is based on low order parametric models which may be easier to estimate. In particular, Kronecker structures, also known as separable models, transposable covariances models, or matrix-variate-normal models are typically used when dealing with random matrices (rather than random vectors) [22]–[33]. They have been successfully applied to different applications including wireless communication, bioinformatics and computer vision. ML estimation in Kronecker structures involves a non-convex optimization. It is usually addressed using the iterative Flip-Flop solution [28]. Uniqueness of the solution has been recently analyzed in [29], [32]. Based on our g-convexity results, we show that the negative-log-likelihood is in fact g-convex. This result complements the known uniqueness results and extends them to more general scenarios as multi-way Kronecker models. It allows the use of efficient numerical methods, and suggests straightforward extensions via additional g-convex regularization.

An interesting observation in our work concerns the relation between the robust and the Kronecker covariance estimation problems. These two problems originate in different applications with different goals and by different researchers. Both problems are traditionally solved by simple iterations, and sat-

The author is with the Selim and Rachel Benin School of Computer Science and Engineering, The Hebrew University of Jerusalem, Jerusalem 91904, Israel (e-mail: amiw@cs.huji.ac.il).

isfy surprising uniqueness results. Our results show that the problems are in fact g-convex minimizations. Furthermore, the derivation suggests that the models themselves are actually similar. Both use product structures for the unknown covariance. The difference is that robust methods use independent scalar scalings, whereas Kronecker techniques turn to constant matrix factors. Armed with this understanding, we propose a hybrid class of robust Kronecker models which enjoy the best of both worlds.

The paper is organized as follows. In Section II, we review and provide new results on g-convexity. In Section III, we demonstrate the use of these results in the context of robust covariance estimation. In Section IV, we consider their application to Kronecker structured covariance estimation. Next, in Section V we address hybrid robust Kronecker models. G-convex regularization schemes are discussed in Section VI, and optimization details are briefly reviewed in Section VII. Finally, simulations results are presented in Section VIII, and concluding remarks are offered in Section IX.

The following notation is used. The sets $\mathbb{R}^n$, $\mathbb{R}^{n,m}$, $\mathbb{R}_{++}^n$ and $\mathbb{S}_{++}^n$ denote the set of length n vectors, the set of size $n \times m$ matrices, the set of length n vectors with positive elements and the set of symmetric positive definite matrices, respectively. The operator $\| \cdot \|_p$ denotes the L_p norm. The superscripts $\mathbf{X}^T$ and $\mathbf{X}^{-1}$ denote the transpose and inverse operations. The superscript $\mathbf{X}^t$ where $\mathbf{X} \in \mathbb{S}_{++}^n$ outputs a matrix with the same eigenvectors as $\mathbf{X}$ and eigenvalues to the power of t. The operator $|\mathbf{X}|$ with domain $\mathbf{X} \in \mathbb{S}_{++}^n$ denotes the determinant. The vector $\mathbf{1}$ is the all ones vector. The matrix $\mathrm{diag}\{\mathbf{a}\}$ is diagonal with the elements a_j, and $\mathrm{diag}\{\mathbf{X}\}$ is a diagonal matrix with the diagonal elements of $\mathbf{X}$. For a vector $\mathbf{z}$, we use $\sqrt{\mathbf{z}}$, $\mathbf{z}^t$, $e^{\mathbf{z}}$ and $\log \mathbf{z}$ to denote vectors with the elements $\sqrt{z_j}$, z_j^t, e^{z_j} and $\log z_j$, respectively. The operators $\circ$, $\otimes$ and $\times$ denote the Hadamard element wise product, the Kronecker product and the set product, respectively. By $\mathrm{vec}\,(\mathbf{X})$ we denote the vector with the stacked columns of the matrix $\mathbf{X}$. We denote the zero mean multivariate Gaussian distribution by $\mathcal{N}(\mathbf{0}, \mathbf{\Sigma})$ where $\mathbf{\Sigma}$ is the covariance matrix. Throughout the paper, C will denote a generic constant which does not depend on the unknowns.

II. GEODESIC CONVEXITY

A. G-Convexity on $\mathcal{M}$

We begin with a brief review on general g-convexity on an arbitrary manifold $\mathcal{M}$. More details are available in [4]. Similar results can be found in [5], [34], [35].

Definition 1: For each pair $q_0, q_1 \in \mathcal{M}$ we define a geodesic[1] $q_t^{q_0, q_1} \in \mathcal{M}$ for $t \in [0, 1]$. For simplicity, we will omit the superscripts and assume q_0 and q_1 are understood from the context.

Definition 2: A real valued function f with domain $\mathcal{M}$ is g-convex if $f(q_t) \leq t f(q_1) + (1 - t) f(q_0)$ for any $q_0, q_1 \in \mathcal{M}$ and $t \in [0, 1]$. Equivalently, it is g-convex if $f(q_t)$ is convex in $t \in [0, 1]$ for any $q_0, q_1 \in \mathcal{M}$.

Proposition 1: Any local minima of a g-convex function over $\mathcal{M}$ is a global minima.

[1]The exact definition of a geodesic is given in [4]. See also [5], [34], [35] for similar approaches known as super-convexity or arcwise connectivity.

B. G-Convexity on $\mathbb{R}_{++}^m$

Throughout this paper, we will be interested in matrix variables, but it is instructive to first address the easier case of vectors.

Classical convexity is a special case of g-convexity when $\mathcal{M} = \mathbb{R}^m$ and the geodesics are defined as [36]

$$\mathbf{z}_t = t\mathbf{z}_1 + (1 - t)\mathbf{z}_0, \quad t \in [0, 1]. \tag{1}$$

It is well known that

$$e^{az_i}, \quad \log \sum_i e^{z_i} \quad \text{are convex.} \tag{2}$$

Another well known property is

$$f(e^z) \text{ is convex in } z \to f(e^{z_1 + z_2}) \text{ is convex in } z_1, z_2. \tag{3}$$

Alternatively, we can rewrite (1)–(3) in terms of g-convexity. For this purpose, we define $\mathcal{M} = \mathbb{R}_{++}^m$ with the geodesics

$$\mathbf{q}_t = \mathbf{q}_1^t \mathbf{q}_0^{1-t}, \quad t \in [0, 1]. \tag{4}$$

which is identical to (1) via a simple change of variables

$$\mathbf{z} = \log \mathbf{q}. \tag{5}$$

The analog of (2) is therefore

$$q_i^a, \quad \log \sum_i q_i \quad \text{are g-convex.} \tag{6}$$

Whereas (3) is equivalent to

$$f(q) \text{ is g-convex in } q \to f(q_1 q_2) \text{ is g-convex in } q_1, q_2. \tag{7}$$

In this case, a simple change of variables transforms g-convexity into convexity. The advantage of g-convexity is that it can be applied in more complicated cases where no such change of variables is known. In particular, we now show how to generalize these properties to $\mathbb{S}_{++}^m$.

C. G-Convexity on $\mathbb{S}_{++}^m$

Consider the manifold of positive definite matrices $\mathbb{S}_{++}^m$. With each $\mathbf{Q}_0, \mathbf{Q}_1 \in \mathbb{S}_{++}^m$ we associate the following geodesic[2] [37]–[39]

$$\mathbf{Q}_t = \mathbf{Q}_0^{\frac{1}{2}} \left(\mathbf{Q}_0^{-\frac{1}{2}} \mathbf{Q}_1 \mathbf{Q}_0^{-\frac{1}{2}} \right)^t \mathbf{Q}_0^{\frac{1}{2}}, \quad t \in [0, 1]. \tag{8}$$

Note that (8) reduces to (4) in the scalar case and in the diagonal case where $\mathbf{Q}_t = \mathrm{diag}\{\mathbf{q}_t\}$.

The matrix versions of (6) are provided in the following two lemmas.

Lemma 1: Let $\mathbf{h} \in \mathbb{R}^m$ and $a \in \pm 1$. The function

$$f(\mathbf{Q}) = \mathbf{h}^T \mathbf{Q}^a \mathbf{h} \tag{9}$$

is g-convex in $\mathbf{Q} \in \mathbb{S}_{++}^m$.

[2]This geodesic is defined in page 4 of [37] using the notation $\mathbf{X}^t = \mathrm{expm}\,(t\,\mathrm{logm}\,(\mathbf{X}))$ where expm and logm are the matrix exponential and logarithm functions. Similarly, [38], [39] characterize this geodesic using only $\mathbf{Q}_0$ and a direction parameter.

Proof: Consider the eigenvalue decomposition

$$\mathbf{Q}_0^{-\frac{1}{2}}\mathbf{Q}_1\mathbf{Q}_0^{-\frac{1}{2}} = \mathbf{U}\mathrm{diag}\,\{\mathbf{d}\}\mathbf{U}^T \tag{10}$$

where $\mathbf{U}$ is an orthogonal matrix and $\mathbf{d} \in \mathbb{R}_{++}^m$. Substituting $\mathbf{Q}_t$ for $\mathbf{Q}$ yields

$$
\begin{aligned}
f(\mathbf{Q}_t) &= \mathbf{h}^T\mathbf{Q}_0^{\frac{a}{2}}\left(\mathbf{Q}_0^{-\frac{1}{2}}\mathbf{Q}_1\mathbf{Q}_0^{-\frac{1}{2}}\right)^{at}\mathbf{Q}_0^{\frac{a}{2}}\mathbf{h} \\
&= \mathbf{h}^T\mathbf{Q}_0^{\frac{a}{2}}\mathbf{U}\mathrm{diag}\,\{\mathbf{d}^{at}\}\mathbf{U}^T\mathbf{Q}_0^{\frac{a}{2}}\mathbf{h} \\
&= \sum_{i=1}^m\left[\mathbf{U}^T\mathbf{Q}_0^{\frac{a}{2}}\mathbf{h}\right]_i^2 d_i^{at} \\
&= \sum_{i=1}^m\left[\mathbf{U}^T\mathbf{Q}_0^{\frac{a}{2}}\mathbf{h}\right]_i^2 e^{ta\log d_i}
\end{aligned}
\tag{11}
$$

which is convex in t. ∎

Lemma 2: Let $a \in \pm 1$ and $\mathbf{H}_i \in \mathbb{R}^{p,m}$ for $i = 1, \ldots, n$ be a set of matrices whose mn columns span $\mathbb{R}^p$. The function

$$\log\left|\sum_{i=1}^n\mathbf{H}_i\mathbf{Q}^a\mathbf{H}_i^T\right| \tag{12}$$

is g-convex in $\mathbf{Q} \in \mathbb{S}_{++}^m$.

Proof: Consider the eigenvalue decomposition in (10). Then,

$$
\begin{aligned}
&\log\left|\sum_{i=1}^n\mathbf{H}_i\mathbf{Q}_0^{\frac{a}{2}}\left(\mathbf{Q}_0^{-\frac{1}{2}}\mathbf{Q}_1\mathbf{Q}_0^{-\frac{1}{2}}\right)^{at}\mathbf{Q}_0^{\frac{a}{2}}\mathbf{H}_i\right| \\
&= \log\left|\sum_{i=1}^n\mathbf{H}_i\mathbf{Q}_0^{\frac{a}{2}}\mathbf{U}\mathrm{diag}\,\{\mathbf{d}^{at}\}\mathbf{U}^T\mathbf{Q}_0^{\frac{a}{2}}\mathbf{H}_i^T\right| \\
&= \log\left|\sum_{i=1}^n\sum_{j=1}^m e^{ta\log d_j}\mathbf{h}_{ij}\mathbf{h}_{ij}^T\right|
\end{aligned}
\tag{13}
$$

where

$$\mathbf{h}_{ij} = \mathbf{H}_i\mathbf{Q}_0^{\frac{a}{2}}\mathbf{U}\mathbf{e}_j. \tag{14}$$

The result then follows by noting that the log determinant of exponentially weighted positive definite matrices is convex [40] (see also the Appendix). ∎

Next, we generalize the setting and consider the manifold

$$\mathbb{S}_J = \mathbb{S}_{++}^{m_1} \times \mathbb{S}_{++}^{m_2} \times \cdots \times \mathbb{S}_{++}^{m_J}. \tag{15}$$

Each point in $\mathbb{S}_J$ is a J-tuple of positive definite matrices $\{\mathbf{Q}_j\}_{j=1}^J$. The geodesic between any two J-tuples $\{\mathbf{Q}_{j0}\}_{j=1}^J, \{\mathbf{Q}_{j1}\}_{j=1}^J \in \mathbb{S}_J$ is the J-tuple $\{\mathbf{Q}_{jt}\}_{j=1}^J \in \mathbb{S}_J$ where

$$\mathbf{Q}_{jt} = \mathbf{Q}_{j0}^{\frac{1}{2}}\left(\mathbf{Q}_{j0}^{-\frac{1}{2}}\mathbf{Q}_{j1}\mathbf{Q}_{j0}^{-\frac{1}{2}}\right)^t\mathbf{Q}_{j0}^{\frac{1}{2}}, \quad t \in [0,1] \tag{16}$$

and $j = 1, \ldots, J$. In order to emphasize the use of multiple variables, we say that a function is jointly g-convex in $\{\mathbf{Q}_j\}_{j=1}^J$ when it satisfies Definition 2 over these joint geodesics.

The following lemma is a generalization of (7) to the matrix case.

Lemma 3: Let f be a real valued and g-convex function on $\mathbb{S}_{++}^m$ with $\prod_{j=1}^J m_j = m$. Then,

$$g(\mathbf{Q}_1, \ldots, \mathbf{Q}_J) = f(\mathbf{Q}_1 \otimes \cdots \otimes \mathbf{Q}_J) \tag{17}$$

is jointly g-convex in $\{\mathbf{Q}_j\}_{ji=1}^J \in \mathbb{S}_J$.

Proof: We have

$$
\begin{aligned}
&tg(\mathbf{Q}_{11}, \ldots, \mathbf{Q}_{J1}) + (1-t)g(\mathbf{Q}_{10}, \ldots, \mathbf{Q}_{J0}) \\
&= tf(\mathbf{Q}_{11} \otimes \cdots \otimes \mathbf{Q}_{J1}) + (1-t)f(\mathbf{Q}_{10} \otimes \cdots \otimes \mathbf{Q}_{J0}) \\
&\overset{[1]}{=} tf(\mathbf{K}_1) + (1-t)f(\mathbf{K}_0) \\
&\overset{[2]}{\geq} f\left(\mathbf{K}_0^{\frac{1}{2}}\left(\mathbf{K}_0^{-\frac{1}{2}}\mathbf{K}_1\mathbf{K}_0^{-\frac{1}{2}}\right)^t\mathbf{K}_0^{\frac{1}{2}}\right) \\
&\overset{[3]}{=} f(\mathbf{Q}_{1t} \otimes \cdots \otimes \mathbf{Q}_{Jt}) \\
&= g(\mathbf{Q}_{1t}, \ldots, \mathbf{Q}_{Jt})
\end{aligned}
\tag{18}
$$

where [1] is just a definition of $\mathbf{K}_0$ and $\mathbf{K}_1$, [2] is due to g-convexity of f, and [3] is based on the properties

$$(\mathbf{A} \otimes \mathbf{B})(\mathbf{C} \otimes \mathbf{D}) = (\mathbf{AC} \otimes \mathbf{BD}) \tag{19}$$

for conforming matrices, and

$$(\mathbf{A} \otimes \mathbf{B})^t = (\mathbf{A}^t \otimes \mathbf{B}^t) \tag{20}$$

for positive definite matrices, e.g.,[41, (9.9) and Theorem 9.1].∎

III. ROBUST MODEL

In this section, we apply the results in Section II to covariance estimation in scaled Gaussian distributions [7]–[19]. We model the statistically independent observations as

$$\mathbf{x}_i \sim \mathcal{N}(\mathbf{0}, q_i\mathbf{Q}). \tag{21}$$

for $i = 1, \ldots, n$, where $\mathbf{Q} \in \mathbb{S}_{++}^p$ is the unknown covariance matrix, and $\mathbf{q} \in \mathbb{R}_{++}^n$ are deterministic factors. We assume $n > p$ and that $\mathbf{x}_i$ span $\mathbb{R}^p$. The nuisance factors allow small variations in the distributions of $\mathbf{x}_i$ and provide a robust statistical model. The negative-log-likelihood is

$$L_{\mathrm{robust}}(\mathbf{q}, \mathbf{Q}) = \sum_{i=1}^n\left[\mathbf{x}_i^T(q_i\mathbf{Q})^{-1}\mathbf{x}_i + \log|q_i\mathbf{Q}|\right]. \tag{22}$$

This objective is non-convex in the classical definitions. However, straight forward application of Lemmas 1–3 yields the following result.

Theorem 1: The function $L_{\mathrm{robust}}(\mathbf{q}, \mathbf{Q})$ in (22) is jointly g-convex $\mathbf{q}$ and $\mathbf{Q}$.

Previous works using model (21) simplified the problem and eliminated the unknown $\mathbf{q}$ [9], [11]. This can be done by minimizing (22) with respect to $\mathbf{q}$

$$q_i^*(\mathbf{Q}) = \frac{\mathbf{x}_i^T\mathbf{Q}^{-1}\mathbf{x}_i}{p}, \quad i = 1, \ldots, n \tag{23}$$

and plugging it back into (22). Alternatively, the measurements $\mathbf{x}_i$ can be normalized as [42]

$$\mathbf{s}_i = \frac{\mathbf{x}_i}{\|\mathbf{x}_i\|_2}, \quad i = 1, \ldots, n \tag{24}$$

so that they will not depend on the unknown $\mathbf{q}$. In both cases, the concentrated negative-log-likelihood of $\mathbf{x}_i$ or the negative-log-likelihood of $\mathbf{s}_i$, result in the following objective parameterized by $\mathbf{Q}$:

$$L_{\text{robust } \mathbf{Q}}(\mathbf{Q}) = p \sum_{i=1}^{n} \log \left(\mathbf{x}_i^T \mathbf{Q}^{-1} \mathbf{x}_i \right) + n \log |\mathbf{Q}| + C. \tag{25}$$

It was recently shown that this function is also g-convex.

Theorem 2 ([21]): The function $L_{\text{robust } \mathbf{Q}}(\mathbf{Q})$ in (25) is g-convex in $\mathbf{Q}$.

We now take an opposite approach and concentrate (22) with respect to $\mathbf{Q}$. The optimal solution is

$$\mathbf{Q}^*(\mathbf{q}) = \frac{1}{n} \sum_{i=1}^{n} \frac{1}{q_i} \mathbf{x}_i \mathbf{x}_i^T. \tag{26}$$

Plugging this solution back into (22) yields

$$L_{\text{robust } \mathbf{q}}(\mathbf{q}) = p \sum_{i=1}^{n} \log(q_i) + n \log \left| \sum_{i=1}^{n} \frac{1}{q_i} \mathbf{x}_i \mathbf{x}_i^T \right| + C. \tag{27}$$

Direct application of Lemma 2 yields the following result.

Theorem 3: The function $L_{\text{robust } \mathbf{q}}(\mathbf{q})$ in (27) is g-convex in $\mathbf{q}$.

In fact, using (5) the objective can be expressed as

$$L_{\text{robust } \mathbf{z}}(\mathbf{z}) = p \sum_{i=1}^{n} z_i + n \log \left| \sum_{i=1}^{n} e^{-z_i} \mathbf{x}_i \mathbf{x}_i^T \right| + C. \tag{28}$$

which is classically convex in $\mathbf{z}$ (see Lemma 4 in the Appendix).

Theorems 1–3 complement known results on robust covariance estimation in (21). Beginning with the seminal work of [7] and its numerous extensions, it was proved that the non-convex ML problem can be efficiently solved using simple fixed point iterations starting at any initial point. Next, [21] proved the concentrated negative-log-likelihood is g-convex in the covariance. Our new contributions are that the negative-log-likelihood is actually jointly g-convex in both the covariance and the scalings, and that it can be concentrated to a g-convex minimization in the scalings. Hence, the global solutions to these three problems can be easily found via simple descent algorithms. Furthermore, additional prior knowledge on both the covariance and the scalings can be exploited via g-convex regularization as will be detailed in Section VI.

IV. KRONECKER MODEL

In this section, we apply the results in Section II to ML estimation in Kronecker structured covariances [22]–[33]. We model the independent and identically distributed (i.i.d.) observations as

$$\mathbf{x}_i \sim \mathcal{N}(\mathbf{0}, \mathbf{Q}_1 \otimes \mathbf{Q}_2) \tag{29}$$

for $i = 1, \ldots, n$, where $\mathbf{Q}_1 \in \mathbb{S}_{++}^{p_1}$ and $\mathbf{Q}_2 \in \mathbb{S}_{++}^{p_2}$. Typically, this model arises when

$$\mathbf{x}_i = \text{vec}\left(\mathbf{X}_i\right) \tag{30}$$

where $\mathbf{X}_i \in \mathbb{R}^{p_2, p_1}$. In matrix terms, the distribution in (29) can be expressed as

$$\mathbf{X}_i = \mathbf{Q}_2^{\frac{1}{2}} \mathbf{W}_i \mathbf{Q}_1^{\frac{1}{2}} \tag{31}$$

where $\mathbf{W}_i \in \mathbb{R}^{p_2, p_1}$ is a matrix of i.i.d., zero mean and unit variance Gaussian random variables. The negative-log-likelihood for estimating $\mathbf{Q}_1$ and $\mathbf{Q}_2$ is

$$L_{\text{kron}}(\mathbf{Q}_1, \mathbf{Q}_2) = \sum_{i=1}^{n} \mathbf{x}_i^T (\mathbf{Q}_1 \otimes \mathbf{Q}_2)^{-1} \mathbf{x}_i \\ + \log |\mathbf{Q}_1 \otimes \mathbf{Q}_2|. \tag{32}$$

This objective is non-convex in the classical definitions. Straight forward application of Lemmas 1–3 yields the following observation.

Theorem 4: The function $L_{\text{kron}}(\mathbf{Q}_1, \mathbf{Q}_2)$ in (32) is jointly g-convex in $\mathbf{Q}_1$ and $\mathbf{Q}_2$.

Thus, the global ML estimate can be found using standard descent methods starting at any starting point. Additional prior information on $\mathbf{Q}_1$ or $\mathbf{Q}_2$ can be exploited via g-convex regularization as will be detailed in Section VI.

For simplicity, we have considered the Kronecker product of two positive definite matrices but g-convexity holds in the general case of multi-way Kronecker models where [43]

$$\mathbf{x} \sim \mathcal{N}(\mathbf{0}, \mathbf{Q}_1 \otimes \cdots \otimes \mathbf{Q}_J). \tag{33}$$

V. HYBRID ROBUST KRONECKER MODEL

An interesting byproduct of our analyses concerns the relation between the robust and the Kronecker covariance estimation problems. These originate in different applications with different goals and by different researchers. However, our derivation suggests that models (21) and (29), as well as their solutions, are actually very similar. The only difference is that (21) uses independent scalar weights, whereas (29) turns to equal matrix weights $\mathbf{Q}$. This naturally suggests a hybrid model which we now explore.

The hybrid robust-Kronecker covariance model is

$$\mathbf{x}_i \sim \mathcal{N}(\mathbf{0}, q_i[\mathbf{Q}_1 \otimes \mathbf{Q}_2]) \tag{34}$$

for $i = 1, \ldots, n$, where $\mathbf{q} \in \mathbb{R}_{++}^n$, $\mathbf{Q}_1 \in \mathbb{S}_{++}^{p_1}$ and $\mathbf{Q}_2 \in \mathbb{S}_{++}^{p_2}$. The model can also be written in matrix notations as

$$\mathbf{X}_i = \sqrt{q_i} \mathbf{Q}_2^{\frac{1}{2}} \mathbf{W}_i \mathbf{Q}_1^{\frac{1}{2}} \tag{35}$$

where $\mathbf{W}_i \in \mathbb{R}^{p_1, p_2}$ are matrices of i.i.d., zero mean and unit variance random variables. The negative-log-likelihood for estimating $\mathbf{q}$, $\mathbf{Q}_1$ and $\mathbf{Q}_2$ is

$$L_{\text{hybrid}}(\mathbf{q}, \mathbf{Q}_1, \mathbf{Q}_2) = \sum_{i=1}^{n} \mathbf{x}_i^T [q_i(\mathbf{Q}_1 \otimes \mathbf{Q}_2)]^{-1} \mathbf{x}_i \\ + \log |q_i[\mathbf{Q}_1 \otimes \mathbf{Q}_2]|. \tag{36}$$

As before, this function is non-convex, but straight forward application of Lemmas 1–3 yields the following observation.

Theorem 5: The function $L_{\text{hybrid}}(\mathbf{q}, \mathbf{Q}_1, \mathbf{Q}_2)$ in (36) is jointly g-convex in $\mathbf{q}$, $\mathbf{Q}_1$ and $\mathbf{Q}_2$.

Thus, the global ML solution to the hybrid model can be found using standard descent methods.

VI. G-CONVEX REGULARIZATION

In the previous sections we showed that the robust, Kronecker and hybrid models lead to g-convex negative-log-likelihoods which can be easily minimized. These results allow straight forward generalizations via additional g-convex regularization functions. Indeed, a common approach in statistics is to regularize ML and solve

$$\min \quad L(\,\cdot\,) + \lambda h(\,\cdot\,) \qquad (37)$$

where $L(\,\cdot\,)$ in (22), (32), or (36) is the appropriate negative-log-likelihood function, $h(\,\cdot\,)$ is a penalty function and λ is a regularization parameter. The penalties shrink the solution towards some low order parametric model in order to allow for bias-variance tradeoff. Alternatively, the penalties can exploit prior knowledge of the unknown parameters as in maximum-a-posteriori estimation. Theorems 1–5 motivate the use of g-convex penalties. These will guarantee that any local regularized ML estimate will be globally optimal.

We now list a few promising g-convex penalties:

• Scalar penalties:

$$\sum_i \log q_i, \sum_i \log q_i^{-1}, \sum_i |\log q_i|, \sum_i \log^2 q_i \qquad (38)$$

G-convexity of these penalties can be observed by changing variables to $\mathbf{z} = \log \mathbf{q}$ and noting their classical convexity in $\mathbf{z}$. Such penalties are appropriate when the scalings, also known as texture, originate from a known underlying distribution as in [44], [45]. Depending on the distribution, these penalties can be used to constrain the empirical first, second and absolute moments.

• Smoothness:

$$\sum_i (\log q_i - \log q_{i-1})^2 \qquad (39)$$

This penalty is appropriate when the scalings originate from a correlated time series, e.g., [46].

• Matrix penalties:

$$\log \text{Tr}\{\mathbf{Q}^{-1}\}, \quad \log \prod_i [\mathbf{Q}^{-1}]_{ii}, \dots \qquad (40)$$

These g-convex penalties allow improved accuracy by shrinking the covariance estimate towards a known target, typically the scaled identity matrix. Proofs and more details on matrix valued g-convex regularization are available in [17].

VII. OPTIMIZATION DETAILS

The main message in this paper is that the robust and the Kronecker negative-log-likelihoods are g-convex. Hence, each of their local minimas is also globally optimal. These objectives are smooth and well behaved. Thus, there are many numerical

methods for efficiently finding their local solutions. The details of these specific implementations are outside the scope of this paper, but we do briefly discuss them for completeness.

Efficient solutions to g-convex minimizations over the positive definite set can be obtained via optimization on manifold techniques [6]. Our experience suggests that there is no need for such sophisticated methods and that standard off-the-shelves smooth optimization packages, as Matlab's built-in fmincon function, are sufficient. When possible, we recommend changing variables and transforming the problem into a standard convex minimization as in (28). In many cases, simple approaches based on alternating minimizations also perform well. For example, the hybrid objective in (36) can be minimized by the classical Flip-Flop method which alternatively solves for $\mathbf{q}$, $\mathbf{Q}_1$ or $\mathbf{Q}_2$ while fixing the other two variables. Another promising approach relies on the majorization-minimization technique which iteratively linearizes the concave terms in the non-convex (yet g-convex) objectives as proposed in [1], [17].

Uniqueness and lack of identifiability are important issues which should be considered in the optimization. It is well known that our parametric models lack uniqueness. For example, $\{\mathbf{q}, \mathbf{Q}\}$ in (21) can be replaced by $\{\alpha\mathbf{q}, \frac{1}{\alpha}\mathbf{Q}\}$ for any $\alpha > 0$. Theoretically, this is not important as both solutions are equally optimal. However, this may cause numerical stability problems, as well as issues with the choice of regularization parameters λ in (37). For this purpose it is recommended to use normalization methods as in [14], [16], to use scale invariant penalty functions as in [17], or to impose additional constraints which will promise uniqueness as in [29]. For example, it is easy to see that (27) is invariant to a scalar multiplication of $\mathbf{q}$. Thus, without loss of generality, we recommend adding the g-convex constraint

$$\sum_i \log q_i = 0. \qquad (41)$$

VIII. SIMULATIONS

In this section, we provide a few numerical examples using Monte Carlo simulations. The purpose of these examples is to demonstrate the high level ideas in this paper, rather than a detailed practical application which is beyond the scope and will be pursued elsewhere.

In the first example, we considered robust covariance estimation with correlated scaling factors. We defined the unknown covariance matrix $\mathbf{Q}$ as a size $p = 10$ Toeplitz matrix with the (i, j) elements $0.8^{|i-j|}$. In each experiment, we generated n i.i.d. realizations of a zero mean multivariate normal of covariance $\mathbf{Q}$ with i.i.d. Chi-squared distribution with 3 degrees of freedom scaling factors. We then correlated these factors using a length 2 uniform moving average window. These realizations were used to estimate the unknown covariance. We compared the naive sample covariance, Tyler's original fixed point iteration, and our newly proposed estimator assuming correlated scaling factors. Specifically, we changed variables and used Matlab's fmincon subroutine to minimize (27) penalized by (39) with $\lambda = 10$ subject to the constraints in (41). Due to the
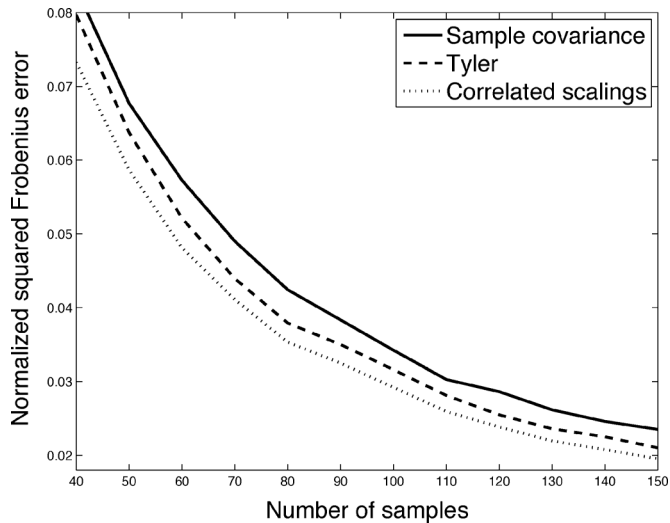
Fig. 1. Robust covariance estimation with correlated scalings.

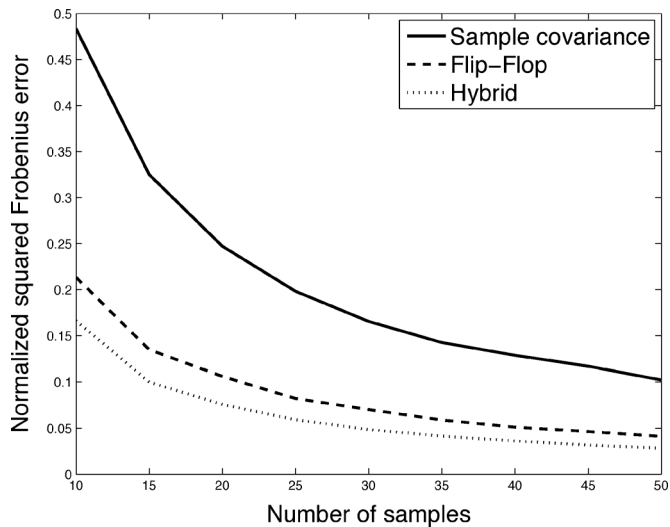Fig. 3. Hybrid estimation in a Gaussian setting.

Fig. 2. Hybrid estimation in a non-Gaussian setting.

scaling ambiguity, we normalized the true covariance and its estimates by their traces before computing the errors. In Fig. 1, we provide the average normalized squared Frobenius norm error over 1000 independent experiments. It is easy to see the performance gain of the robust estimators, and the advantage of exploiting the additional prior knowledge. Similar results (not shown) with different values of λ suggest that the solution is not too sensitive to its exact value.

In the second example, we considered the hybrid covariance model. The scalings q_i were i.i.d. scaling factors generated according to a Chi-squared distribution with 3 degrees of freedom. We defined the unknown covariance matrices $\mathbf{Q}_1$ and $\mathbf{Q}_2$ of respective sizes $p_1 = 10$ and $p_2 = 2$ as Toeplitz matrices with the (i, j) elements $0.8^{|i-j|}$. We compared the naive sample covariance, the Flip-Flop method and the hybrid ML estimator. The latter was implemented in a Flip-Flop manner by alternatively solving for the unknowns in closed form. As before, we normalized the true covariance and its estimates by their trace. In Fig. 2, we provide the average normalized squared Frobenius
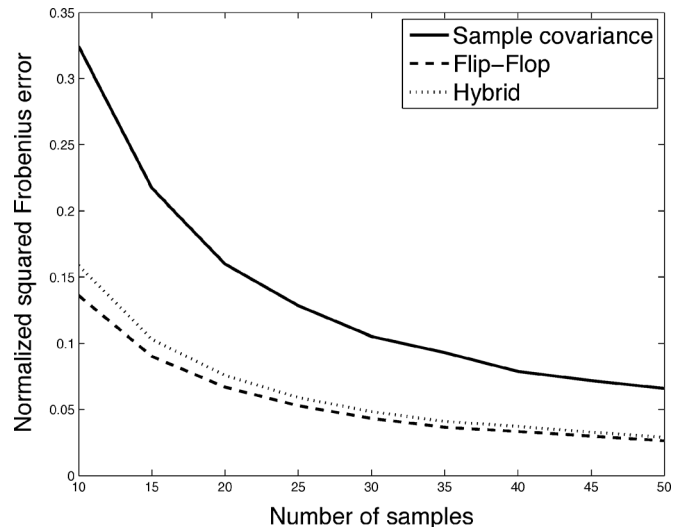
norm error over 500 independent non-Gaussian experiments. The hybrid method outperforms its competitors for all sample sizes considered. For completeness, Fig. 3 shows the results of a similar experiment with known factors $q_i = 1$, corresponding to a pure Gaussian setting. Here, as expected from a robust estimation method, the hybrid approach performs slightly worse than Tyler, but the degradation is small and vanishes as the number of samples increase.

IX. DISCUSSION

In this paper, we extended classical convexity results on exponentials and log-sum-exp functions to g-convexity in positive definite matrix variables. We then applied these results to two modern covariance estimation problems. We reformulated both of these as g-convex minimizations and noted their similarity. This analysis shed more light on these problems and paved the road to various generalizations.

APPENDIX

In this Appendix, we provide results on the convexity of log determinant functions with exponentially weighted positive semidefinite matrices. These results have been previously published in [40] using a different proof and are provided for completeness.

It is well known that the log-sum-exp function [36, pp. 74]

$$\log \sum_{i=1}^{n} e^{z_i} \tag{42}$$

is convex in $\mathbf{z} \in \mathbb{R}^n$. Indeed, its gradient and Hessian are

$$\mathbf{g} = \frac{\mathbf{y}}{\mathbf{1}^T \mathbf{y}}$$

$$\mathbf{H} = \frac{(\mathbf{1}^T \mathbf{y}) \text{diag} \{\mathbf{y}\} - \mathbf{y} \mathbf{y}^T}{(\mathbf{1}^T \mathbf{y})^2} \tag{43}$$

where

$$\mathbf{y} = e^{\mathbf{z}}. \tag{44}$$

Convexity follows since

$$\mathbf{v}^T \mathbf{H} \mathbf{v} = \frac{1}{(\mathbf{1}^T \mathbf{y})^2} \left[(\mathbf{1}^T \mathbf{y}) \sum_{i=1}^{n} v_i^2 y_i - (\mathbf{v}^T \mathbf{y})^2 \right]$$

$$= \frac{1}{(\mathbf{1}^T \mathbf{y})^2} \left[(\mathbf{u}^T \mathbf{u})(\mathbf{w}^T \mathbf{w}) - (\mathbf{u}^T \mathbf{w})^2 \right]$$

$$\geq 0, \quad \text{for all} \quad \mathbf{v} \tag{45}$$

where

$$\mathbf{u} = \sqrt{\mathbf{y}} \tag{46}$$

$$\mathbf{w} = \mathbf{v} \circ \sqrt{\mathbf{y}} \tag{47}$$

and we have used the Cauchy Schwartz inequality.

The next lemma generalizes this convexity result to log-determinant functions of exponentially weighted positive semidefinite rank one matrices.

Lemma 4 ([40]): Let $\mathbf{h}_i \in \mathbb{R}^p$ for $i = 1, \ldots, n$ be a set of vectors which span $\mathbb{R}^p$. The function

$$\log \left| \sum_{i=1}^{n} e^{z_i} \mathbf{h}_i \mathbf{h}_i^T \right| \tag{48}$$

is convex in $\mathbf{z} \in \mathbb{R}^n$.

Proof: First, we note that the argument of the determinant is positive definite when $\{\mathbf{h}_i\}_{i=1}^{n}$ span their space, and therefore inside the domain. Using the chain rule, the gradient and Hessian can be computed as

$$\mathbf{g} = \text{diag}\{\mathbf{P}\}\mathbf{1}$$
$$\mathbf{H} = \text{diag}\{\mathbf{P}\} - \mathbf{P} \circ \mathbf{P} \tag{49}$$

where

$$\mathbf{P} = \mathbf{Y}^T (\mathbf{Y}\mathbf{Y}^T)^{-1} \mathbf{Y}$$
$$\mathbf{Y} = [\mathbf{y}_1, \ldots, \mathbf{y}_n]$$
$$\mathbf{y}_i = \mathbf{h}_i e^{\frac{z_i}{2}} \tag{50}$$

and $\mathbf{P}$ is an orthogonal projection matrix. Due to Lemma 5 below, the Hessian is positive semidefinite and the function is convex. ∎

Lemma 5: Let $\mathbf{P}$ be a projection matrix, then the matrix

$$\text{diag}\{\mathbf{P}\} - \mathbf{P} \circ \mathbf{P} \tag{51}$$

is positive semidefinite.

Proof: We need to show that for any vector $\mathbf{v}$

$$\mathbf{v}^T [\text{diag}\{\mathbf{P}\} - \mathbf{P} \circ \mathbf{P}] \mathbf{v} \geq 0. \tag{52}$$

Alternatively, by expressing $\mathbf{v}$ as $\mathbf{D} = \text{diag}\{\mathbf{v}\}$, we need to show that for any diagonal matrix $\mathbf{D}$

$$\text{Tr}\{\mathbf{D}\text{diag}\{\mathbf{P}\}\mathbf{D}\} - \text{Tr}\{\mathbf{P}\mathbf{D}\mathbf{P}\mathbf{D}\} \geq 0. \tag{53}$$

This is equivalent to

$$\text{Tr}\{\mathbf{P}\mathbf{D}^2\} - \text{Tr}\{\mathbf{P}\mathbf{D}\mathbf{P}\mathbf{D}\} \geq 0 \tag{54}$$

which can also be written as

$$\text{Tr}\{\mathbf{P}[\mathbf{D}(\mathbf{I} - \mathbf{P})\mathbf{D}]\} \geq 0. \tag{55}$$

The latter holds since $\mathbf{P}$, $\mathbf{I} - \mathbf{P}$ and $\mathbf{D}(\mathbf{I} - \mathbf{P})\mathbf{D}$ are positive semidefinite, and the trace of the product of two positive semidefinite matrices is non-negative. ∎

Interestingly, Lemma 5 can be interpreted as an extension of the Cauchy Schwartz inequality from vectors to projection matrices. Indeed, specializing it to the case of a rank one projection matrix

$$\mathbf{P} = \frac{\mathbf{u}\mathbf{u}^T}{\mathbf{u}^T \mathbf{u}} \tag{56}$$

where $\mathbf{u} \neq \mathbf{0}$, yields

$$\mathbf{v}^T \left[\text{diag}\left\{ \frac{\mathbf{u}\mathbf{u}^T}{\mathbf{u}^T \mathbf{u}} \right\} - \frac{\mathbf{u}\mathbf{u}^T \circ \mathbf{u}\mathbf{u}^T}{(\mathbf{u}^T \mathbf{u})^2} \right] \mathbf{v} \geq 0. \tag{57}$$

This is equivalent to

$$\left[\sum_{i=1}^{n} u_i^2 \right] \left[\sum_{i=1}^{n} v_i^2 u_i^2 \right] \geq \sum_{i,j} v_i v_j u_i^2 u_j^2. \tag{58}$$

Defining

$$\mathbf{w} = \mathbf{v} \circ \mathbf{u} \tag{59}$$

yields

$$\left[\sum_{i=1}^{n} u_i^2 \right] \left[\sum_{i=1}^{n} w_i^2 \right] \geq \sum_{i,j} w_i w_j u_i u_j \tag{60}$$

results in the Cauchy-Schwartz inequality:

$$(\mathbf{u}^T \mathbf{u})(\mathbf{w}^T \mathbf{w}) \geq (\mathbf{u}^T \mathbf{w})^2. \tag{61}$$

REFERENCES

[1] A. Wiesel, "On the convexity in Kronecker structured covariance estimation," in *Proc. IEEE Statist. Signal Process. Workshop (SSP'12)*, Ann Arbor, MI, 2012.
[2] M. Chiang, C. W. Tan, D. P. Palomar, D. O'Neill, and D. Julian, "Power control by geometric programming," *IEEE Trans. Wireless Commun.*, vol. 6, no. 7, pp. 2640–2651, 2007.
[3] S. Boyd, S. J. Kim, L. Vandenberghe, and A. Hassibi, "A tutorial on geometric programming," *Optimiz. Eng.*, vol. 8, no. 1, pp. 67–127, 2007.
[4] T. Rapcsak, "Geodesic convexity in nonlinear optimization," *J. Optimiz. Theory Appl.*, vol. 69, no. 1, pp. 169–183, 1991.
[5] L. Liberti, "On a class of nonconvex problems where all local minima are global," *Publ. l'Instit. Math.*, vol. 76, no. 90, pp. 101–109, 2004.
[6] P.-A. Absil, R. Mahony, and R. Sepulchre, *Optimization Algorithms on Matrix Manifolds*.   Princeton, NJ: Princeton Univ. Press, 2008.
[7] D. E. Tyler, "A distribution-free M-estimator of multivariate scatter," *Ann. Statist.*, vol. 15, no. 1, pp. 234–251, 1987.
[8] J. T. Kent, "Data analysis for shapes and images," *J. Statist. Plann. Inference*, vol. 57, no. 2, pp. 181–193, 1997.

[9] E. Conte, A. De Maio, and G. Ricci, "Recursive estimation of the covariance matrix of a compound-Gaussian process and its application to adaptive CFAR detection," *IEEE Trans. Signal Process.*, vol. 50, no. 8, pp. 1908–1915, 2002.

[10] 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, Jan. 2008.

[11] F. Gini and M. Greco, "Covariance matrix estimation for CFAR detection in correlated heavy tailed clutter," *Signal Process.*, vol. 82, no. 12, pp. 1847–1859, 2002.

[12] E. Ollila and V. Koivunen, "Robust antenna array processing using m-estimators of pseudo-covariance," in *Proc. 14th IEEE Pers., Indoor and Mobile Radio Commun. (PIMRC 2003)*, 2003, vol. 3, pp. 2659–2663.

[13] E. Ollila and D. E. Tyler, "Distribution-free detection under complex elliptically symmetric clutter distribution," in *Proc. IEEE Sens. Array and Multichannel Signal Process. Workshop (SAM'12)*, Hoboken, NJ, Jun. 17–20, 2012, vol. 144.

[14] Y. I. Abramovich and N. K. Spencer, "Diagonally loaded normalised sample matrix inversion (LNSMI) for outlier-resistant adaptive filtering," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP 2007)*, 2007, vol. 3.

[15] 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, Oct. 2010.

[16] 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.

[17] A. Wiesel, "Unified framework to regularized covariance estimation in scaled Gaussian models," *IEEE Trans. Signal Process.*, vol. 60, no. 1, pp. 29–38, Jan. 2012.

[18] A. Wiesel, "Regularized covariance estimation in scaled Gaussian models," in *Proc. 4th IEEE Int. Workshop on Computat. Adv. Multi-Sens. Adapt. Process. (CAMSAP)*, 2011, pp. 309–312.

[19] G. Frahm and U. Jaekel, "A generalization of Tyler's M-estimators to the case of incomplete data," *Computat. Statist. Data Anal.*, vol. 54, no. 2, pp. 374–393, 2010.

[20] A. Aubry, A. De Maio, L. Pallotta, and A. Farina, "Maximum likelihood estimation of a structured covariance matrix with a condition number constraint," *IEEE Trans. Signal Process.*, vol. 60, no. 6, pp. 3004–3021, 2012.

[21] C. Auderset, C. Mazza, and E. A. Ruh, "Angular Gaussian and Cauchy estimation," *J. Multivar. Anal.*, vol. 93, no. 1, pp. 180–197, 2005.

[22] K. V. Mardia and C. Goodall, "Chapter Spatial-temporal analysis of multivariate environmental monitoring data," in *Multivariate Environmental Statistics*. New York: North-Holland, 1993, vol. 6.

[23] N. Lu and D. L. Zimmerman, "The likelihood ratio test for a separable covariance matrix," *Statist. Probabil. Lett.*, vol. 73, no. 4, pp. 449–457, 2005.

[24] A. K. Gupta and D. K. Nagar, *Matrix Variate Distributions*. Boca Raton, FL: Chapman & Hall/CRC, 2000, vol. 104.

[25] K. Werner, M. Jansson, and P. Stoica, "On estimation of covariance matrices with Kronecker product structure," *IEEE Trans. Signal Process.*, vol. 56, no. 2, pp. 478–491, 2008.

[26] G. I. Allen and R. Tibshirani, "Transposable regularized covariance models with an application to missing data imputation," *Ann. Appl. Statist.*, vol. 4, no. 2, pp. 764–790, 2010.

[27] Y. Zhang and J. Schneider, "Learning multiple tasks with a sparse matrix-normal penalty," *Adva. Neural Inf. Process. Syst.*, 2010.

[28] P. Dutilleul, "The MLE algorithm for the matrix normal distribution," *J. Statist. Comput. Simul.*, vol. 64, no. 2, pp. 105–123, 1999.

[29] M. Srivastava, T. von Rosen, and D. von Rosen, "Models with a Kronecker product covariance structure: Estimation and testing," *Math. Methods Statist.*, vol. 17, pp. 357–370, 2008.

[30] J. P. Kermoal, L. Schumacher, K. I. Pedersen, P. E. Mogensen, and F. Frederiksen, "A stochastic MIMO radio channel model with experimental validation," *IEEE J. Sel. Areas Commun.*, vol. 20, no. 6, pp. 1211–1226, 2002.

[31] B. Efron, "Are a set of microarrays independent of each other?," *Ann. Appl. Statist.*, vol. 3, no. 3, p. 922, 2009.

[32] C. H. Lee, P. Dutilleul, and A. Roy, "Comment on "models with a Kronecker product covariance structure: Estimation and testing" by M. Srivastava, T. von Rosen, and D. von Rosen, mathematical methods of statistics, 17 (2008), pp. 357-370," *Math. Methods of Statist.*, vol. 19, no. 1, pp. 88–90, 2010.

[33] T. Tsiligkaridis, A. O. Hero, III, and S. Zhou, "Convergence properties of Kronecker graphical lasso algorithms," [Online]. Available: http://arxiv.org/abs/1204.0585 2012

[34] A. Ben-Tal, "On generalized means and generalized convex functions," *J. Optimiz. Theory Appl.*, vol. 21, no. 1, pp. 1–13, 1977.

[35] M. Avriel and I. Zang, "Generalized arcwise-connected functions and characterizations of local-global minimum properties," *J. Optimiz. Theory Appl.*, vol. 32, no. 4, pp. 407–425, 1980.

[36] S. P. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, U.K.: Cambridge Univ. Press, 2004.

[37] S. Bonnabel and R. Sepulchre, "Riemannian metric and geometric mean for positive semidefinite matrices of fixed rank," *SIAM J. Matrix Anal. Appl.*, vol. 31, no. 3, pp. 1055–1070, 2010.

[38] M. Moakher, "A differential geometric approach to the geometric mean of symmetric positive-definite matrices," *SIAM J. Matrix Anal. Appl.*, vol. 26, no. 3, pp. 735–747, 2005.

[39] S. T. Smith, "Covariance, subspace, and intrinsic Cramér-Rao bounds," *IEEE Trans. Signal Process.*, vol. 53, no. 5, pp. 1610–1630, 2005.

[40] L. Gurvits and A. Samorodnitsky, "A deterministic algorithm for approximating the mixed discriminant and mixed volume, and a combinatorial corollary," *Discr. Computat. Geometry*, vol. 27, no. 4, pp. 531–550, 2002.

[41] T. K. Moon and W. C. Stirling, *Mathematical Methods and Algorithms for Signal Processing*. Englewood Cliffs, NJ: Prentice-Hall, 2000, vol. 204.

[42] G. Frahm, "Generalized elliptical distributions: Theory and applications," Ph.D. thesis, Univ. of Cologne, Cologne, 2004, unpublished.

[43] D. Akdemir and A. K. Gupta, "Array variate random variables with multiway Kronecker delta covariance matrix structure," *J. Algebraic Stat.*, vol. 2, no. 1, pp. 98–113, 2011.

[44] Y. Chitour and F. Pascal, "Exact maximum likelihood estimates for SIRV covariance matrix: Existence and algorithm analysis," *IEEE Trans. Signal Process.*, vol. 56, no. 10, pp. 4563–4573, Oct. 2008.

[45] J. Wang, A. Dogandzic, and A. Nehorai, "Maximum likelihood estimation of compound-Gaussian clutter and target parameters," *IEEE Trans. Signal Process.*, vol. 54, no. 10, pp. 3884–3898, 2006.

[46] T. Bucciarelli, P. Lombardo, and S. Tamburrini, "Optimum cfar detection against compound Gaussian clutter with partially correlated texture," in *Proc. IEE Radar, Sonar Navig.*, 1996, vol. 143, pp. 95–104, IET.

Ami Wiesel (M'07) 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 Computer Science, University of Michigan, Ann Arbor, during 2007–2009. Since January 2010, he has been a faculty member at the Rachel and Selim Benin School of Computer Science and Engineering, 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 Communications (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.