Improved Calibration of High-Dimensional Precision Matrices

Mengyi Zhang, *Student Member, IEEE*, Francisco Rubio, *Member, IEEE*, and Daniel P. Palomar, *Fellow, IEEE*

***Abstract*—Estimation of a precision matrix (i.e., the inverse covariance matrix) is a fundamental problem in statistical signal processing applications. When the observation dimension is of the same order of magnitude as the number of samples, the conventional estimators of covariance matrix and its inverse perform poorly. In order to obtain well-behaved estimators in high-dimensional settings, we consider a general class of estimators of covariance matrices and precision matrices based on weighted sampling and linear shrinkage. The estimation error is measured in terms of both quadratic loss and Stein's loss, and these loss functions are used to calibrate the set of parameters defining our proposed estimator. In an asymptotic setting where the observation dimension is comparable in magnitude to the number of samples, we provide estimators of the precision matrix that are as good as their oracle counterparts. We test our estimators with both synthetic data and financial market data, and Monte Carlo simulations show the advantage of our precision matrix estimator compared with well known estimators in finite sample size settings.**

***Index Terms*—Asymptotic analysis, precision matrix estimation, random matrix theory, shrinkage.**

I. Introduction

Many applications in statistical signal processing and related fields, such as financial engineering, collaborative filtering and social network analysis, require estimates of a covariance matrix and/or a precision matrix from high-dimensional data observations. If the number of samples is large compared to the observation dimension, the sample covariance matrix represents a good approximation of the true covariance matrix, and its inverse can be effectively used as an estimator of the precision matrix [1]. However, in the case of high-dimensional data, it is well-known that the eigenvalues of the sample covariance matrix considerably spread out from those of the true covariance matrix [2], [3]; see also references therein. Therefore, it is clearly not sensible to use the inverse of the sample covariance matrix as the estimator of the precision matrix, since the inversion operation can dramatically amplify the estimation error.

Manuscript received March 22, 2012; revised August 14, 2012 and November 07, 2012; accepted November 11, 2012. Date of publication December 24, 2012; date of current version February 26, 2013. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Stefano Marano. This work was supported by the Hong Kong Research Grant RPC11EG39.

M. Zhang is with the Department of Electronic and Computer Engineering, Hong Kong University of Science and Technology, Hong Kong (e-mail: zmyannie@ust.hk).

F. Rubio was with the Department of Electronic and Computer Engineering, Hong Kong University of Science and Technology, Hong Kong. He is now with Genetic Finance, Ltd., Hong Kong (e-mail: franciscoerubio@gmail.com).

D. P. Palomar is with the Hong Kong University of Science and Technology (HKUST) and the Centre Tecnològic de Telecomnicacions de Catalunya—Hong Kong (CTTC-HK), Hong Kong (e-mail: palomar@ust.hk).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TSP.2012.2236321

A. Literature Review

Many efforts have been devoted to the estimation of covariance matrices, at least since James and Stein's work [4]. For instance, Dey and Srinivasan minimized the worst-case error under Stein's loss[5]. Ledoit and Wolf considered a linear combination of the sample covariance matrix with the identity matrix under the quadratic loss[6]. Chen *et al.* improved on the Ledoit-Wolf method for Gaussian data and proposed a new estimator rooted on the Rao-Blackwell theorem [7]. In these papers, the estimation of the precision matrix is not explicitly handled. However, it is claimed for instance in [6] that an estimator of the precision matrix can be obtained by directly taking the inverse of the covariance matrix.

Another way of handling the problem of precision matrix estimation is to require additional properties on the true covariance matrix, such as sparseness or low-rank constraints. Rothman *et al.* proposed to construct a sparse estimator for the precision matrix using a penalized normal likelihood approach [8]. Fan *et al.* employed a factor model and imposed sparsity on the estimates of the covariance matrix and the precision matrix [9]. Bickel and Levina estimated the precision matrix by banding the inverse of the sample covariance matrix via its Cholesky factorization [10]. Previous work on sparse precision matrix estimation also included [11]–[14], in which l_1-based algorithm was used. These precision matrix estimators naturally lack good generalization capabilities as they are specifically focused on estimating a particularly structured precision matrix.

As usually in statistics, in all works above, the choice of a loss functions represents the central element based on which estimation accuracy is measured. Though a high estimation accuracy is arguably a main requirement, there is no general agreement on which loss functions to use depending on the particular situation. In the literature, common examples of loss functions are the Frobenius norm or quadratic loss [15], an alternative form of quadratic loss [9], Stein's loss [4] and the related normal likelihood [16], as well as the spectral norm and other kinds of operator norms [10]. We remark that optimal estimators under a given loss might be well suboptimal with respect to a different loss; for example, it is pointed out in [17] that optimal procedures for spectral and Frobenius norms can be considerably different.

B. Methodology and Contributions

In this paper, we improve on previous work in mainly two aspects: first, we do not require specific properties or structure of the true covariance matrix; second, instead of taking the inverse of the covariance matrix estimator, we obtain calibrations of the precision matrix estimator by directly optimizing the matrix loss function. In this work, in line with the mainstream literature on covariance matrix estimation, we concentrate on the quadratic loss and Stein's loss. Regarding the estimator structure, we consider a class of precision matrix parameterizations based on linear shrinkage and sample weighting. In previous works, shrinkage estimators have been extensively used to estimate covariance matrices [6], [7], [18], but have been seldom applied to the estimation of precision matrices. The reason lies in the technical impossibility of computing the expectation of the error loss of the precision matrix estimator. In order to overcome this issue, we approach the problem by studying the high-dimensional asymptotics of both quadratic and Stein's loss functions using the spectral analysis of large dimensional random matrices [19]. On a final note, we remark that weighted sampling is also a widely used technique in statistics, applied to, e.g., the bootstrap method [20].

Our main contributions can be summarized as follows: we provide precision matrix estimators that are asymptotically as good as the clairvoyant/oracle estimator; we reveal the asymptotic optimality of uniform weighting in precision matrix estimators, which coincides with intuition, but whose proof is not trivial; moreover, we apply our method to financial field and construct the global minimum variance portfolio.

C. Organization of the Paper

The rest of the paper is organized as follows. Section II is devoted to the introduction of the structure of our proposed estimators and the matrix loss functions. Important theoretical results on a type of generalized consistency (for high-dimensional observations) for precision matrix estimation are provided in Section III. In Section IV, the optimal selection of calibration parameters is discussed. Section V presents numerical results in financial application and Section VI concludes the paper. All technical details and derivations are relegated to the appendices.

Notation

In this paper, s, $\mathbf{x}$, $\mathbf{M}$ denote scalars, vectors, and matrices, respectively. A subscript can be added to the above quantities to emphasize dependence on sample dimension. The superscripts $(\cdot)^T$ and $(\cdot)^H$ denote, respectively, the transpose and conjugate transpose. The trace of $\mathbf{M}$ is denoted by $\mathsf{tr}[\mathbf{M}]$ and the mathematical expectation operator is denoted by $\mathsf{E}(\cdot)$. $\mathbb{R}$ and $\mathbb{C}$ denote the real and complex fields of dimension specified by superscripts, and $\mathcal{D}_+$ is the set of positive semidefinite diagonal matrices. We denote by $\|\mathbf{x}\|$ the Euclidean norm of a vector $\mathbf{x}$; for a matrix $\mathbf{M}$, $\|\mathbf{M}\|_2 = \sqrt{\lambda_{max}(\mathbf{M}^H\mathbf{M})}$ denotes the spectral norm;$\|\mathbf{M}\|_F = \sqrt{\mathsf{tr}[\mathbf{M}^H\mathbf{M}]}$ denotes the Frobenius norm; and $\|\mathbf{M}\|_{\mathsf{tr}} = \mathsf{tr}[(\mathbf{M}^H\mathbf{M})^{1/2}]$ denotes the trace norm. Subscripts in $\mathcal{L}_{Q,\mathbf{R}}$ and $\mathcal{L}_{S,\mathbf{R}}$ mean quadratic and Stein's error loss, respectively (abbreviated as Q and S), with respect to estimating the matrix $\mathbf{R}$. Equivalent notation is used for the estimation of the precision matrix $\mathbf{R}^{-1}$. Given two quantities a and b, $a \asymp b$ denotes they are asymptotic equivalents, i.e., $|a-b| \to 0$ almost surely. Finally, $\mathcal{L}$, $\bar{\mathcal{L}}$, and $\hat{\mathcal{L}}$ denote a certain loss, its asymptotic equivalent, and consistent estimator, respectively.

II. Estimating Precision Matrices: Problem Formulation

Let $\{\mathbf{y}_t \in \mathbb{C}^M\}_{t=1}^N$ be a collection of independent and identical distributed (i.i.d.) observations of a stochastic process with zero mean and covariance matrix $\mathbf{R}_M \in \mathbb{C}^{M\times M}$. Our target is to find an estimator of the precision matrix $\mathbf{R}_M^{-1}$ based on sample observations $\{\mathbf{y}_t\}_{t=1}^N$. For notational convenience, we denote by $\mathbf{Y}_N = [\mathbf{y}_1, \ldots, \mathbf{y}_N]$ the $M \times N$ observation matrix. In this paper, we study precision matrix estimators $\hat{\mathbf{R}}_M^{-1}$, with $\hat{\mathbf{R}}_M$ having the following structure

$$\hat{\mathbf{R}}_M = \frac{1}{N}\mathbf{Y}_N\mathbf{T}_N\mathbf{Y}_N^H + \rho_M\mathbf{I}_M \tag{1}$$

where $\mathbf{T}_N \in \mathcal{D}_+$ is a diagonal matrix with elements representing a set of nonnegative sample weights, the identity matrix $\mathbf{I}_M$ is the shrinkage target, and ρ_M is a nonnegative regularization coefficient. We notice that the choice of the identity matrix as shrinkage target is without loss of generality, since as it will become clear from our analysis, a general and arbitrary positive definite matrix can be equivalently considered.

Observe that even though the columns of $\mathbf{Y}_N$ are i.i.d., we cannot conclude that $\mathbf{T}_N$ has to be a scaled identity matrix. Indeed, i.i.d. observations imply that the problem is symmetric (i.e., the order of the diagonal elements of $\mathbf{T}_N$ can be permuted) but the lack of convexity does not support the claim of $\mathbf{T}_N$ being the identity matrix. As an illustrative example, suppose we want to maximize $x_1^2 + x_2^2$ subject to $x_1 + x_2 = 1,\ x_1 \geq 0,\ x_2 \geq 0$. This problem is also symmetric but the optimal solution is not $x_1 = x_2 = 0.5$ but $x_1 = 1, x_2 = 0$ or $x_1 = 0, x_2 = 1$.

The rationale behind our approach relies on adding some structure to the estimation problem, which can be effectively exploited to further improve the bias-variance tradeoff of the model. This is achieved by sensibly selecting the free parameters $\mathbf{T}_N$ and ρ_M as we will show in the sequel. As an example, consider the typical shrinkage covariance matrix estimation setup, given by $\mathbf{T}_N = (1-\rho_M)\mathbf{I}_N$, where $0 \leq \rho_M \leq 1$. If $\rho_M = 0$, then we obtain the sample covariance matrix (SCM) $\hat{\mathbf{R}}_{\mathsf{SCM}}$, which is an unbiased estimator of $\mathbf{R}_M$. By choosing a nonzero ρ_M, a biased estimator is obtained, though at the same time the (expected) error loss can be decreased. Thus, the bias-variance tradeoff is effectively optimized by an appropriate choice of the shrinkage intensity parameter ρ_M. Moreover, we focus on the asymptotic regime that $M,\ N \to \infty$ at the same speed. Interestingly, now the traditional intuition fails. For example, let $\mathbf{R}_M = \mathbf{I}_M$, in a traditional setting where M is fixed and $N \to \infty$, we have the reasonable result $\frac{1}{M}\|\hat{\mathbf{R}}_{\mathsf{SCM}}\|_F^2 \to \frac{1}{M}\|\mathbf{R}_M\|_F^2$. However, this intuition breaks down completely in the asymptotic regime in which case $\frac{1}{M}\|\hat{\mathbf{R}}_{\mathsf{SCM}}\|_F^2 \asymp \frac{M}{N} + \frac{1}{M}\|\mathbf{R}_M\|_F^2$.

A number of matrix loss functions can be considered in order to measure the distance between two matrices. A classical instance of matrix loss function in multivariate statistics is the quadratic loss. This is because the theory underlying quadratic losses leads to solutions based on relatively straightforward matrix manipulations. Specifically, the quadratic loss between a matrix $\mathbf{B}_M$ and a certain approximation given by $\hat{\mathbf{B}}_M$ is defined as

$$\mathcal{L}_{Q,\mathbf{B}}(\mathbf{B}_M,\ \hat{\mathbf{B}}_M) = \frac{1}{M}\|\mathbf{B}_M - \hat{\mathbf{B}}_M\|_F^2. \tag{2}$$

Another popular loss is the Stein's loss, which can be derived from the Wishart likelihood [1]. This loss is also referred to as the entropy loss [16], since it is the relative entropy [21] or Kullback-Leibler divergence between two multivariate zero-mean Gaussian distributions with covariance matrices $\mathbf{B}_M$ and $\hat{\mathbf{B}}_M$, respectively. From the viewpoint of information theory, the relative entropy is a measure of the distance between two distributions; therefore, the Stein's loss measures the inefficiency of assuming the Gaussian distribution with covariance matrix $\hat{\mathbf{B}}_M$ when the true covariance matrix is $\mathbf{B}_M$. The definition of the Stein's loss is given as

$$\mathcal{L}_{S,\mathbf{B}}(\mathbf{B}_M,\, \hat{\mathbf{B}}_M) = \frac{1}{M}\mathsf{tr}[\mathbf{B}_M^{-1}\hat{\mathbf{B}}_M] - \frac{1}{M}\mathsf{logdet}(\mathbf{B}_M^{-1}\hat{\mathbf{B}}_M) - 1, \tag{3}$$

which coincides with the divergence between multidimensional proper complex Gaussian distributions with independent components given in [22]. The above losses are normalized by the dimension of the vector observations, so that different values of M can be seamlessly handled. Although the latter is always nonnegative, we note that it does not define a distance or metric in the formal sense because the symmetry property is not satisfied.

Then, in the case of our interest regarding the estimation of precision matrices, we simply substitute $\mathbf{B}_M = \mathbf{R}_M^{-1}$ and obtain:

$$\mathcal{L}_{Q,\mathbf{R}^{-1}}(\mathbf{R}_M^{-1},\, \hat{\mathbf{R}}_M^{-1}) = \frac{1}{M}\|\mathbf{R}_M^{-1} - \hat{\mathbf{R}}_M^{-1}\|_F^2 \tag{4}$$

$$\mathcal{L}_{S,\mathbf{R}^{-1}}(\mathbf{R}_M^{-1},\, \hat{\mathbf{R}}_M^{-1}) = \frac{1}{M}\mathsf{tr}[\mathbf{R}_M\hat{\mathbf{R}}_M^{-1}] - \frac{1}{M}\mathsf{logdet}(\mathbf{R}_M\hat{\mathbf{R}}_M^{-1}) - 1. \tag{5}$$

The precision matrix estimator can be obtained by either directly estimating the precision matrix, or estimating the covariance matrix and then taking the inverse. In order to compare our precision matrix estimators with precision matrix estimators that are inverse of the estimated covariance matrix, we also consider loss functions with respect to covariance matrices by substituting $\mathbf{B}_M = \mathbf{R}_M$, i.e.,

$$\mathcal{L}_{Q,\mathbf{R}}(\mathbf{R}_M,\, \hat{\mathbf{R}}_M) = \frac{1}{M}\|\mathbf{R}_M - \hat{\mathbf{R}}_M\|_F^2 \tag{6}$$

$$\mathcal{L}_{S,\mathbf{R}}(\mathbf{R}_M,\, \hat{\mathbf{R}}_M) = \frac{1}{M}\mathsf{tr}[\mathbf{R}_M^{-1}\hat{\mathbf{R}}_M] - \frac{1}{M}\mathsf{logdet}(\mathbf{R}_M^{-1}\hat{\mathbf{R}}_M) - 1. \tag{7}$$

We define and will refer in the sequel as oracle estimators of the precision and covariance matrix, respectively, the solutions with structure (1) minimizing (4)–(5) and (6)–(7). Notice, however, that the previous optimization problem is given in terms of the unknown true covariance matrix, and therefore the term oracle. More specifically, the free oracle parameters $(\mathbf{T}_N^{\mathsf{or}}, \rho_M^{\mathsf{or}})$ are calibrated through the following optimization problem with objective functions (4)–(7), respectively:

$$\begin{aligned} &\underset{(\mathbf{T}_N,\rho_M)}{\mathsf{minimize}} && \mathcal{L}(\mathbf{T}_N, \rho_M) \\ &\mathsf{subject\ to} && \mathbf{T}_N \in \mathcal{D}_+ \quad \rho_M \geq 0. \end{aligned} \tag{8}$$

Note that the optimal parameters $(\mathbf{T}_N^{\mathsf{or}}, \rho_M^{\mathsf{or}})$ cannot be obtained in practice since the objective functions (4)–(7) depend on the unknown covariance matrix $\mathbf{R}_M$. The naive approach is based on the plug-in estimators of (4)–(7), which are given by replacing $\mathbf{R}_M$ with the sample covariance matrix $\hat{\mathbf{R}}_{\mathsf{SCM}} = \frac{1}{N}\mathbf{Y}_N\mathbf{Y}_N^H$. However, minimizing the plug-in estimators yield trivial solutions $\mathbf{T}_N = \mathbf{I}_N$ and $\rho_M = 0$ (corresponding to zero loss functions). The covariance matrix estimator becomes the sample covariance matrix, and the precision matrix estimator becomes its inverse.

If an estimator of the covariance matrix is required and (6) is used as the objective of the optimization problem (8), the following approach can be employed to obtain the optimal parameters (see, e.g., [6]). Let $\mathbf{T}_N = \alpha_M\mathbf{I}_N$ and $\hat{\mathbf{R}}_M = \alpha_M\hat{\mathbf{R}}_{\mathsf{SCM}} + \rho_M\mathbf{I}_M$, which is a special case of the structure in (1). Then, the expectation of $\mathcal{L}_{Q,\mathbf{R}}$ (cf. (2)) can be computed and minimized with respect to α_M and ρ_M. The optimal solutions are denoted by $\alpha_M^\star$ and $\rho_M^\star$. Then, consistent estimators of $\alpha_M^\star$ and $\rho_M^\star$ (denoted by $\hat{\alpha}_M$ and $\hat{\rho}_M$) are derived in the double-limit where both M and N go to infinity. Hence, the covariance matrix estimator is given by $\hat{\mathbf{R}}_M = \hat{\alpha}_M\hat{\mathbf{R}}_{\mathsf{SCM}} + \hat{\rho}_M\mathbf{I}_M$.

However, the previous approach cannot be applied for covariance matrix estimation with (7) as the objective function as well as direct precision matrix estimation with $\mathcal{L}$ (in (4)–(5)) as the objective function, simply because the expectations of (4), (5), and (7) cannot be explicitly computed. We tackle this problem by resorting to a high-dimensional asymptotic analysis of the previous quantities based on tools from random matrix theory [19]. Specifically, instead of considering the computation of their expectations, we derive asymptotic equivalents of (4), (5), and (7) (i.e., $\bar{\mathcal{L}}$) in the double-limit regime defined by both M and N going to infinity at the same rate. These asymptotic equivalents are deterministic and are given in terms of the true unknown covariance matrix. Thus, we provide consistent estimators (i.e., $\hat{\mathcal{L}}$) of these quantities or, equivalently, of (4), (5), and (7) in the double-limit regime. The optimal parameters $(\mathbf{T}_N^{\mathsf{or}}, \rho_M^{\mathsf{or}})$ in the precision matrix estimators can therefore be obtained in the double-limit regime by minimizing the consistent estimators of (4), (5), and (7), respectively. This estimation scheme will be illustrated in detail in Section IV. Before proceeding further, we remark that the double-limit regime under consideration suits realistic operation conditions in practice, where the number of samples N is comparable in magnitude to the observation dimension M.

From the optimization perspective, our method can be summarized as: we use random matrix theory to derive the double limit for given parameters $\mathbf{T}_N$ and ρ_M, and also an estimate of this limit, which are the asymptotic equivalent and consistent estimator respectively. The estimator is a function of $\mathbf{T}_N$ and ρ_M. We then optimize $\mathbf{T}_N$ and ρ_M based on that function. In the following section, we provide two theorems describing the asymptotic equivalents and consistent estimators for the quantities (4), (5) and (7). In order to unify the theories, we also provide these approximations for the quantity (6).

III. Asymptotic Equivalents and Consistent Estimators

In this section we provide the main results of this paper. We begin with some technical hypotheses and further definitions. Asymptotic equivalents and consistent estimators based on sample observations for loss functions (4)–(7) are presented in the following subsections, respectively.

A. Assumptions and Further Definitions

Let $\mathbf{X}_M$ be an $M \times N$ random matrix such that the entries of $\mathbf{X}_M$ are i.i.d. complex random variables with mean zero and variance one. The following two assumptions will be maintained throughout the paper.

(A1) The observation matrix can be expressed as $\mathbf{Y}_M = \mathbf{R}_M^{1/2}\mathbf{X}_M$, with the entries of $\mathbf{X}_M$ having moments of sufficiently high order.

(A2) The $M \times M$ positive definite matrix $\mathbf{R}_M$ and $N \times N$ diagonal matrix $\mathbf{T}_N$ have uniformly bounded spectral norm.

We will consider the limiting regime defined by $M, N \to \infty$ with $0 < \mathrm{liminf} c_M < \mathrm{limsup} c_M < \infty$, where $c_M = M/N$.

Before proceeding with the statement of the main results, we introduce some further definitions. We define $\beta_M = c_M \left(\frac{1}{M}\mathrm{tr}[\mathbf{R}_M]\right)^2$, and the asymptotic equivalent (and generalized consistent estimator in the limiting regime under consideration) given by $\hat{\beta}_M = \frac{1}{MN^2}\sum_{t=1}^N \|\mathbf{y}_t\|^4$. There are other generalized estimators of β_M, such as $c_M \left(\frac{1}{M}\mathrm{tr}[\hat{\mathbf{R}}_{\mathsf{scm}}]\right)^2$; we choose $\hat{\beta}_M$ so that we can better unify the theories supporting the following results.

Moreover, we introduce

$$\gamma_M = \frac{1}{N}\mathrm{tr}\left[(\mathbf{R}_M(\tilde{\delta}_M\mathbf{R}_M + \rho_M\mathbf{I}_M)^{-1})^2\right], \tag{9}$$

$$\tilde{\gamma}_M = \frac{1}{N}\mathrm{tr}\left[(\mathbf{T}_N(\mathbf{I}_N + \delta_M\mathbf{T}_M)^{-1})^2\right], \tag{10}$$

$$\begin{aligned}\eta_M = &\frac{1}{M}\mathrm{logdet}(\tilde{\delta}_M\mathbf{R}_M + \rho_M\mathbf{I}_M)\\ &+ \frac{1}{M}\mathrm{logdet}(\mathbf{I}_N + \delta\mathbf{T}_N) - \frac{N}{M}\delta_M\tilde{\delta}_M,\end{aligned} \tag{11}$$

where $(\delta_M, \tilde{\delta}_M)$ is the unique positive solution of the system of equations [23]:

$$\begin{cases}\delta_M = \frac{1}{N}\mathrm{tr}\left[\mathbf{R}_M(\tilde{\delta}_M\mathbf{R}_M + \rho_M\mathbf{I}_M)^{-1}\right], \\ \tilde{\delta}_M = \frac{1}{N}\mathrm{tr}\left[\mathbf{T}_N(\mathbf{I}_N + \delta_M\mathbf{T}_N)^{-1}\right].\end{cases} \tag{12}$$

Furthermore, we define the matrices:

$$\bar{\mathbf{R}}_{C,M} = \frac{1}{N}\mathrm{tr}\left[\mathbf{T}_N\right]\mathbf{R}_M + \rho_M\mathbf{I}_M, \tag{13}$$

$$\bar{\mathbf{R}}_{P,M} = \tilde{\delta}_M\mathbf{R}_M + \rho_M\mathbf{I}_M. \tag{14}$$

The above quantities will be used to describe the asymptotic deterministic equivalents we derive in the following section as well as the generalized consistent estimators that we propose in the subsequent one.

B. Asymptotic Equivalents of Loss Functions

The following theorem provides asymptotic equivalents $\bar{\mathcal{L}}$ of loss functions $\mathcal{L}$ in (4)–(7), as stated in the following theorem, $\mathcal{L} - \bar{\mathcal{L}} \to 0$ almost surely. The asymptotic equivalents $\bar{\mathcal{L}}$ are nonrandom and only depend on the true covariance matrix $\mathbf{R}_M$ and parameters $\mathbf{T}_N$ and ρ_M.

Theorem 1: Define the following deterministic quantities:

$$\begin{aligned}\bar{\mathcal{L}}_{Q,\mathbf{R}^{-1}} = &\frac{1}{M}\|\mathbf{R}_M^{-1} - \bar{\mathbf{R}}_{P,M}^{-1}\|_F^2 + \frac{\tilde{\gamma}_M c_M}{1-\gamma_M\tilde{\gamma}_M}\\ &\times\left(\frac{1}{M}\mathrm{tr}\left[\mathbf{R}_M\bar{\mathbf{R}}_{P,M}^{-2}\right]\right)^2\end{aligned} \tag{15}$$

$$\bar{\mathcal{L}}_{S,\mathbf{R}^{-1}} = \frac{1}{M}\mathrm{tr}\left[\mathbf{R}_M\bar{\mathbf{R}}_{P,M}^{-1}\right] - \frac{1}{M}\mathrm{logdet}(\mathbf{R}_M) + \eta_M - 1 \tag{16}$$

$$\bar{\mathcal{L}}_{Q,\mathbf{R}} = \frac{1}{M}\|\mathbf{R}_M - \bar{\mathbf{R}}_{C,M}\|_F^2 + \beta_M\frac{1}{N}\mathrm{tr}[\mathbf{T}_N^2] \tag{17}$$

$$\bar{\mathcal{L}}_{S,\mathbf{R}} = \frac{1}{M}\mathrm{tr}\left[\mathbf{R}_M^{-1}\bar{\mathbf{R}}_{C,M}\right] + \frac{1}{M}\mathrm{logdet}(\mathbf{R}_M) - \eta_M - 1. \tag{18}$$

Under Assumptions (A1) and (A2), it follows that $\mathcal{L}_{Q,\mathbf{R}^{-1}} \asymp \bar{\mathcal{L}}_{Q,\mathbf{R}^{-1}}$, $\mathcal{L}_{S,\mathbf{R}^{-1}} \asymp \bar{\mathcal{L}}_{S,\mathbf{R}^{-1}}$, $\mathcal{L}_{Q,\mathbf{R}} \asymp \bar{\mathcal{L}}_{Q,\mathbf{R}}$, and $\mathcal{L}_{S,\mathbf{R}} \asymp \bar{\mathcal{L}}_{S,\mathbf{R}}$.

Proof: Refer to Appendix B. ■

Theorem 1 provides asymptotic approximations of the loss functions in the double-limit regime. This allows us to analyze the consistency of the structured estimators (1) for fixed ρ_M and $\mathbf{T}_N$. Such an asymptotic analysis is of interest on its own for the purposes of evaluating the performance of a given covariance matrix estimator under predetermined scenarios described by a hypothetical true covariance matrix. From a statistical estimation perspective, we are rather interested in the calibration of the free parameters ρ_M and $\mathbf{T}_N$ based on a set of observed samples. To that effect, estimates of the loss functions (4)–(7), or, equivalently, their asymptotic equivalents (15)–(18) are required. These are provided in the following section.

C. Consistent Estimators of Loss Functions

Next, we provide consistent estimators (i.e., $\hat{\mathcal{L}}$) of loss functions $\mathcal{L}$ in (4)–(7), which only depend on the observation matrix $\mathbf{Y}_N$ and the parameters $\mathbf{T}_N$ and ρ_M. As stated in the following theorem, $\hat{\mathcal{L}} - \bar{\mathcal{L}} \to 0$ almost surely. The consistent estimators will be given in terms of corrections of the naive plug-in estimators, which for the quadratic loss function are defined as

$$\mathcal{L}_{Q,\mathbf{R}^{-1}}^{\mathsf{plug\text{-}in}} = \mathcal{L}_Q(\hat{\mathbf{R}}_{\mathsf{SCM}}^{-1}, \hat{\mathbf{R}}_M^{-1}) \tag{19}$$

$$\mathcal{L}_{Q,\mathbf{R}}^{\mathsf{plug\text{-}in}} = \mathcal{L}_Q(\hat{\mathbf{R}}_{\mathsf{SCM}}, \hat{\mathbf{R}}_M), \tag{20}$$

respectively for the precision and covariance matrices, and for Stein's loss are equivalently defined. In particular, if $\hat{\mathbf{R}}_{\mathsf{SCM}}$ is not invertible, we define

$$\begin{aligned}\mathcal{L}_{S,\mathbf{R}^{-1}}^{\mathsf{plug\text{-}in}} = &\frac{1}{M}\mathrm{tr}(\hat{\mathbf{R}}_{\mathsf{SCM}}\hat{\mathbf{R}}_M^{-1}) - \frac{1}{M}\sum_{i=1}^{\min(M,N)}\log\lambda_i(\hat{\mathbf{R}}_{\mathsf{SCM}})\\ &+ \frac{1}{M}\mathrm{logdet}(\hat{\mathbf{R}}_M) - 1,\end{aligned}$$

where $\lambda_i(\hat{\mathbf{R}}_{\mathsf{SCM}})$ denotes the ith nonzero eigenvalue of $\hat{\mathbf{R}}_{\mathsf{SCM}}$.

Moreover, the following proposition provides consistent estimators of δ_M and $\tilde{\delta}_M$, which will be later used for the main result.

Proposition 2 ([24]): Under assumptions (A1) and (A2), a consistent estimator of δ_M, denoted by $\hat{\delta}_M$, is the unique positive solution of the following equation:

$$\hat{\delta}_M \frac{1}{N}\mathrm{tr}\left[\mathbf{T}_N(\mathbf{I}_N + \hat{\delta}_M\mathbf{T}_N)^{-1}\right] = \frac{1}{N}\mathrm{tr}\left[\frac{1}{N}\mathbf{Y}_N\mathbf{T}_N\mathbf{Y}_N^H(\frac{1}{N}\mathbf{Y}_N\mathbf{T}_N\mathbf{Y}_N^H + \rho_M\mathbf{I}_M)^{-1}\right]. \quad (21)$$

Moreover, a consistent estimator of $\tilde{\delta}_M$, denoted by $\hat{\tilde{\delta}}_M$, is given by

$$\hat{\tilde{\delta}}_M = \frac{1}{N}\mathrm{tr}\left[\mathbf{T}_N(\mathbf{I}_N + \hat{\delta}_M\mathbf{T}_N)^{-1}\right]. \quad (22)$$

Theorem 3: Define the following quantities:

$$\hat{\mathcal{L}}_{Q,\mathbf{R}^{-1}} = \mathcal{L}^{\text{plug-in}}_{Q,\mathbf{R}^{-1}} + \psi_{Q,\mathbf{R}^{-1}} \quad (23)$$

$$\hat{\mathcal{L}}_{S,\mathbf{R}^{-1}} = \mathcal{L}^{\text{plug-in}}_{S,\mathbf{R}^{-1}} + \psi_{S,\mathbf{R}^{-1}} \quad (24)$$

$$\hat{\mathcal{L}}_{Q,\mathbf{R}} = \mathcal{L}^{\text{plug-in}}_{Q,\mathbf{R}} + \psi_{Q,\mathbf{R}} \quad (25)$$

$$\hat{\mathcal{L}}_{S,\mathbf{R}} = \mathcal{L}^{\text{plug-in}}_{S,\mathbf{R}} + \psi_{S,\mathbf{R}} \quad (26)$$

where we have defined the correction terms

$$\begin{aligned}\psi_{Q,\mathbf{R}^{-1}} = {} & \frac{2}{M}\mathrm{tr}[\rho_M^{-1}(\hat{\tilde{\delta}}_M\hat{\mathbf{R}}_M^{-1} - (1-c_M)\,\hat{\mathbf{R}}_{\mathsf{SCM}}^{-1}) \\ & + \hat{\mathbf{R}}_{\mathsf{SCM}}^{-1}\hat{\mathbf{R}}_M^{-1}] \\ & - (2c_M - c_M^2)\frac{1}{M}\mathrm{tr}[\hat{\mathbf{R}}_{\mathsf{SCM}}^{-2}] \\ & - (c_M - c_M^2)\left(\frac{1}{M}\mathrm{tr}[\hat{\mathbf{R}}_{\mathsf{SCM}}^{-1}]\right)^2 \quad (27)\end{aligned}$$

$$\psi_{S,\mathbf{R}^{-1}} = \frac{N}{M}\hat{\delta}_M - \frac{1}{M}\mathrm{tr}(\hat{\mathbf{R}}_{\mathsf{SCM}}\hat{\mathbf{R}}_M^{-1}) + K \quad (28)$$

$$\psi_{Q,\mathbf{R}} = \hat{\beta}_M\left(\frac{1}{N}\mathrm{tr}[\mathbf{T}_N^2] - \left(\frac{1}{N}\mathrm{tr}[\mathbf{T}_N] - 1\right)^2\right) \quad (29)$$

$$\begin{aligned}\psi_{S,\mathbf{R}} = {} & \frac{1}{N}\mathrm{tr}[\mathbf{T}_N] - \frac{1}{M}\mathrm{tr}[\hat{\mathbf{R}}_{\mathsf{SCM}}^{-1} \\ & \times(\frac{1}{N}\mathbf{Y}_N\mathbf{T}_N\mathbf{Y}_N^H + c_M\rho_M\mathbf{I}_M)] - K \quad (30)\end{aligned}$$

with

$$K = \begin{cases}(1-c_M^{-1})\log(1-c_M) - 1 & \text{if } c_M < 1 \\ -1 & \text{if } c_M = 1 \\ c_M^{-1}\log c_M & \\ +c_M^{-1}(1-c_M)\log(1-c_M^{-1}) - c_M^{-1} & \text{if } c_M > 1.\end{cases}$$

Under Assumptions (A1) and (A2), it follows that $\hat{\mathcal{L}}_{Q,\mathbf{R}^{-1}} \asymp \bar{\mathcal{L}}_{Q,\mathbf{R}^{-1}}$, $\hat{\mathcal{L}}_{S,\mathbf{R}^{-1}} \asymp \bar{\mathcal{L}}_{S,\mathbf{R}^{-1}}$, $\hat{\mathcal{L}}_{Q,\mathbf{R}} \asymp \bar{\mathcal{L}}_{Q,\mathbf{R}}$, and $\hat{\mathcal{L}}_{S,\mathbf{R}} \asymp \bar{\mathcal{L}}_{S,\mathbf{R}}$.

Proof: Refer to Appendix C. ■

Under objective $\mathcal{L}_{Q,\mathbf{R}}$, our method coincides with Ledoit-Wolf method in the double limit. To be more specific, in Ledoit-Wolf method, the expectation of $\mathcal{L}_{Q,\mathbf{R}}$ is computed first and then an estimate of the expectation is derived at the double limit. However, in our method, we first derive a deterministic equivalent of $\mathcal{L}_{Q,\mathbf{R}}$ and then an estimate of this deterministic equivalent is derived at the double limit. It can be proved that the expectation in Ledoit-Wolf paper [6] coincide with the deterministic equivalent (17) at the limit.

Now, we show how to obtain the optimal shrinkage coefficient defining the covariance matrix estimator by using Theorem 3. We use the quadratic loss of covariance matrix $\mathcal{L}_{Q,\mathbf{R}}$ as an example. In practice its consistent estimator (25) is used. Consider a simpler case $\mathbf{T}_N = (1-\rho_M)\mathbf{I}_N$, then our estimator (1) becomes a convex combination of the sample covariance matrix $\hat{\mathbf{R}}_{\mathsf{SCM}}$ and $\mathbf{I}_M$, and the loss function estimator (25) becomes $\hat{\mathcal{L}}_{Q,\mathbf{R}} = \rho_M^2\frac{1}{M}\|\hat{\mathbf{R}}_{\mathsf{SCM}} - \mathbf{I}_M\|_F^2 - 2\hat{\beta}_M\rho_M + \hat{\beta}_M$. Minimizing (25) over ρ_M with constraint $0 \le \rho_M \le 1$ yields the optimal shrinkage coefficient $\rho_M^\star = \min\left(\frac{\hat{\beta}_M}{\frac{1}{M}\|\hat{\mathbf{R}}_{\mathsf{SCM}} - \mathbf{I}_M\|_F^2}, 1\right)$.

The following section is devoted to the general problem of estimating either a covariance or a precision matrix by using the structured estimator in (1), and under one or the other loss functions considered above.

IV. Optimal Selection of Calibration Parameters

In this section and the following sections, the subscripts M and N will be omitted for clarity of presentation. The parameters $\mathbf{T}$ and ρ in the precision matrix estimator (1) represent a set of degrees-of-freedom with respect to which estimation performance can be improved. Ideally, we use consistent estimators (23)–(26) to calibrate these parameters and formulate the following set of problems:

$$\begin{aligned}&\underset{(\mathbf{T},\rho)}{\text{minimize}} && \hat{\mathcal{L}}(\mathbf{T},\rho) \\ &\text{subject to} && \mathbf{T}\in\mathcal{D}_+ \quad \rho \ge 0.\end{aligned} \quad (31)$$

In general, the set of problems (31) are nonconvex. However, we can study the structure of optimal solutions asymptotically by replacing (23)–(26) with their corresponding asymptotic equivalents (15)–(18). Following this observation, we now focus on the optimal solutions to the following set of problems:

$$\begin{aligned}&\underset{(\mathbf{T},\rho)}{\text{minimize}} && \bar{\mathcal{L}}(\mathbf{T},\rho) \\ &\text{subject to} && \mathbf{T}\in\mathcal{D}_+ \quad \rho \ge 0.\end{aligned} \quad (32)$$

The following lemma and the following proposition show asymptotic optimality of uniform weighting of problem (32) with objectives $\bar{\mathcal{L}}_{Q,\mathbf{R}^{-1}}$, $\bar{\mathcal{L}}_{Q,\mathbf{R}}$, or $\bar{\mathcal{L}}_{S,\mathbf{R}}$.

Lemma 4: Function $(\mathbf{T},\rho) \mapsto \bar{\mathcal{L}}(\mathbf{T},\rho)$ ($\bar{\mathcal{L}}$ denotes $\bar{\mathcal{L}}_{Q,\mathbf{R}^{-1}}$, $\bar{\mathcal{L}}_{Q,\mathbf{R}}$, or $\bar{\mathcal{L}}_{S,\mathbf{R}}$) is convex on $\mathcal{D}_+ \times \mathbb{R}_+$.

Proof: See Appendix D. ■

Proposition 5: There exists a global optimal solution $(\mathbf{T}^\star, \rho^\star)$ to optimization problem (32) with objective function $\bar{\mathcal{L}}(\mathbf{T},\rho)$ ($\bar{\mathcal{L}}$ denotes $\bar{\mathcal{L}}_{Q,\mathbf{R}^{-1}}$, $\bar{\mathcal{L}}_{Q,\mathbf{R}}$, or $\bar{\mathcal{L}}_{S,\mathbf{R}}$) of the form of $(\alpha^\star\mathbf{I}, \rho^\star)$.

Proof: We only provide the proof for function $\bar{\mathcal{L}}_{Q,\mathbf{R}^{-1}}$, since the other proofs follow the same line of reasoning. To start, note that $\bar{\mathcal{L}}$ (which denotes $\bar{\mathcal{L}}_{Q,\mathbf{R}^{-1}}$ for short) is invariant under permutation of diagonal elements of $\mathbf{T}$. Now suppose $\mathbf{T}$ is an optimal solution to problem (32). Let $\mathbf{T}_1, \ldots, \mathbf{T}_P$ have diagonal elements to be all possible permutations of those of $\mathbf{T}$, then $\bar{\mathcal{L}}(\mathbf{T}) = \bar{\mathcal{L}}(\mathbf{T}_1) = \cdots = \bar{\mathcal{L}}(\mathbf{T}_P)$ because of permutation invariance. Moreover, $\bar{\mathcal{L}}(\mathbf{T}) \le \bar{\mathcal{L}}(\frac{1}{P}\sum_{i=1}^P \mathbf{T}_i) \le \frac{1}{P}\sum_{i=1}^P \bar{\mathcal{L}}(\mathbf{T}_i) = \bar{\mathcal{L}}(\mathbf{T})$ where left inequality is from global optimality of $\mathbf{T}$, and the right inequality is from Lemma 4 and Jensen's inequality. Hence, $\frac{1}{P}\sum_{i=1}^P \mathbf{T}_i = \frac{1}{N}\mathrm{tr}(\mathbf{T})\mathbf{I}$ is an optimal solution which corresponds to the form of $\alpha^\star\mathbf{I}$. ■

The remaining loss function $\mathcal{L}_{S,\mathbf{R}^{-1}}$ is trickier to deal with as it is nonconvex. Consider problem (32) with objective (16), i.e., $\bar{\mathcal{L}}_{S,\mathbf{R}^{-1}}$. Recall from Theorem 1, $\bar{\mathcal{L}}_{S,\mathbf{R}^{-1}}(\mathbf{T},\rho) = \frac{N}{M}\delta(\mathbf{T},\rho) + \eta(\mathbf{T},\rho) - \frac{1}{M}\mathsf{logdet}(\mathbf{R}) - 1$ which is an alternative form of (16). We provide the following lemma to characterize the convexity-concavity properties of $\bar{\mathcal{L}}_{S,\mathbf{R}^{-1}}(\mathbf{T},\rho)$.

Lemma 6: Function $(\mathbf{T},\rho) \mapsto \delta(\mathbf{T},\rho)$ is convex while function $(\mathbf{T},\rho) \mapsto \eta(\mathbf{T},\rho)$ is concave on the set $\mathcal{D}_+ \times \mathbb{R}_+$.

Proof: The proof follows the same line of reasoning as the proof of Lemma 4. See Appendix D. ■

In order to discuss the resolution of problem (32) with objective function $\bar{\mathcal{L}}_{S,\mathbf{R}^{-1}}$, we consider the following related problem with linear approximations of $\eta(\mathbf{T},\rho)$. In the literature of minimizing a concave function over a compact convex set, successive linearization is commonly used to find a local minimum. For example, successive linearization is referred to as Frank-Wolfe algorithm and used as a local minimizer in polyhedral concave programs in [25], [26]. It is also shown in [27] that successive linearization solutions converge to a local minimum of a logdet function (interestingly, the term $\eta(\mathbf{T},\rho)$ is a logdet function in the limit). Therefore, the optimal solution of the following problem, after enough iterations, can be used as an approximation to the original problem (32):

$$\begin{aligned} &\underset{(\mathbf{T},\rho)}{\mathsf{minimize}} \quad \frac{N}{M}\delta(\mathbf{T},\rho) + \eta_{apx}^{(k)}(\mathbf{T},\rho) - \frac{1}{M}\mathsf{logdet}(\mathbf{R}) - 1 \\ &\mathsf{subject\ to} \quad \mathbf{T} \in \mathcal{D}_+ \ \rho \geq 0 \end{aligned} \tag{33}$$

where

$$\begin{aligned} \eta_{apx}^{(k)}(\mathbf{T},\rho) = &\ \eta((\mathbf{T},\rho)^{(k-1)}) + \nabla\eta((\mathbf{T},\rho)^{(k-1)}) \\ &\times (\mathsf{vec}((\mathbf{T},\rho)) - \mathsf{vec}((\mathbf{T},\rho)^{(k-1)})) \end{aligned}$$

is a linear approximation of the concave function $\eta(\mathbf{T},\rho)$ at the solution at the $(k-1)$th iteration, and $\mathsf{vec}((\mathbf{T},\rho)) = [t_1,\ldots,t_N,\rho]^T$. The following proposition shows the structure of optimal solutions for the sequence of optimization problem (33).

Proposition 7: For each of the sequence of optimization problems (33), the optimal solution $(\mathbf{T}^\star,\rho^\star)$ is in the form of $(\alpha^\star\mathbf{I},\rho^\star)$.

Proof: We first illustrate the convexity of the objective function of problem (33): $\delta(\mathbf{T},\rho)$ is convex from Lemma 6, and $\eta_{apx}^{(k)}(\mathbf{T},\rho)$ is linear by definition. The remaining proof follows the same line of reasoning as the proof of Proposition 7. ■

Note that although Proposition 7 is not as strong as Proposition 5, $(\alpha^\star\mathbf{I},\rho^\star)$ can still be used as a suboptimal solution to problem (32).

With Propositions 5 and 7, the number of variables of problem (32) is reduced from $N+1$ to 2. At this point, Proposition 2 can be simplified: Let $D = \frac{1}{N}\mathsf{tr}\left[\alpha\hat{\mathbf{R}}_{\mathsf{SCM}}(\alpha\hat{\mathbf{R}}_{\mathsf{SCM}} + \rho\mathbf{I})^{-1}\right]$, then $\hat{\delta} = \frac{D}{\alpha(1-D)}$ and $\hat{\hat{\delta}} = \alpha(1-D)$. We can come to problem (31) with variable $\mathbf{x} = [\alpha,\rho]^T$ and the feasible set $\{\mathbf{x} \mid \mathbf{x} \geq 0\}$, and exhaustive search can be used to find the optimal $\mathbf{x}$.

V. Simulations on Financial Systems

We use Monte Carlo simulations to illustrate the advantage of our estimator. We first focus on the general problem of estimating the precision matrix using the synthetic data. In financial applications, for example, the observation matrix $\mathbf{Y}$ can be regarded as the asset return data subtracting the average and $\mathbf{R}$ is the covariance matrix of the returns. Then we consider the global minimum variance portfolio (GMVP) problem: Suppose a very conservative investor wants to invest a portfolio with the least amount of risk, and does not care about the expected return. This is the portfolio with minimum variance and the solution is $\mathbf{w}^\star = \frac{\mathbf{R}^{-1}\mathbf{1}}{\mathbf{1}^T\mathbf{R}^{-1}\mathbf{1}}$ [28]. We use the real market data to estimate $\mathbf{R}^{-1}$ and evaluate the out-of-sample standard deviation.

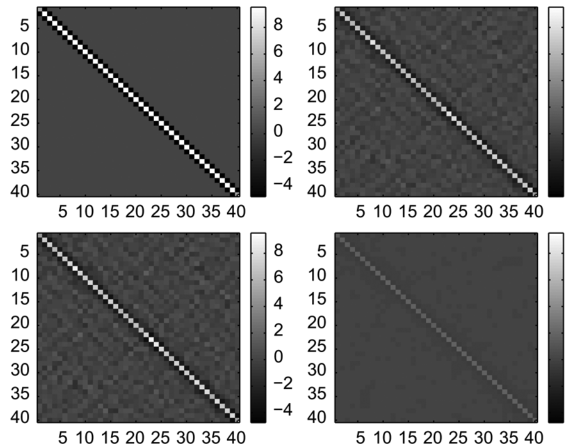

Fig. 1. Banded precision matrix and its estimators. Truth (top left), oracle (top right), SP (bottom left), CLIME (bottom right).

A. Estimating the Precision Matrix With Synthetic Data

1) Comparison With Sparse Precision Matrix Estimator: In this subsection we compare our precision matrix estimator with the CLIME estimator in [14], which is designed for sparse precision matrix estimation. We show that although our estimator is used for a general precision matrix, it can still outperform the CLIME estimator. We use the R package provided by the authors in [14] to implement the CLIME estimator. We use the covariance matrix of a Gaussian AR process, $[\mathbf{R}]_{ij} = 0.9^{|i-j|}$, as the true covariance matrices to evaluate the performances. We use a common scenario that observation dimension $M = 40$ and number of samples $N = 60$. Since in CLIME estimator, the cross-validation step is implemented under Stein's loss, we use our estimator designed for precision matrix estimation under Stein's loss, i.e., SP, for comparison purposes.

In Fig. 1 similar conclusions can be made. The difference is that the oracle is no longer the true precision matrix since it must follow the structure (1). To quantify the difference, the quadratic loss of oracle, SP, and CLIME are 23.9562, 25.9606, and 94.7450, respectively; The Stein's loss of oracle, SP, and CLIME are 0.2218, 0.2259, and 0.4076, respectively. The main reasons that our estimator can be much better than CLIME is that: i) We impose a special structure of precision matrix estimator which could restrict the range of the elements; ii) We consider both random matrix theory and optimization, while [14] only relies on optimization. iii) CLIME is to biased for Stein's loss and sacrifices the quadratic loss, then the overall performance is bad. We also comment that our method is much faster than the CLIME method. Regarding the SP estimator with ex-

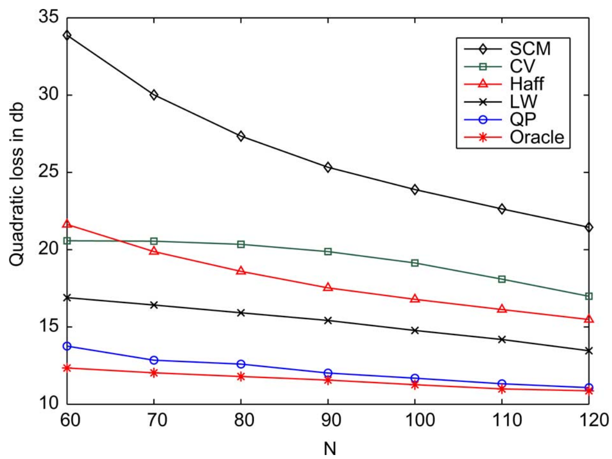

Fig. 2. Quadratic loss w.r.t. precision matrix $\mathcal{L}_{Q,\mathbf{R}^{-1}}$. Dimension $M = 40$, number of samples N varies from 60 to 120.

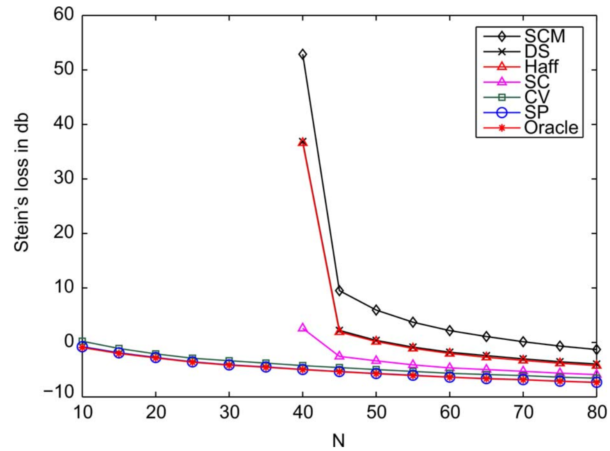

Fig. 3. Stein's loss w.r.t. precision matrix $\mathcal{L}_{S,\mathbf{R}^{-1}}$. Dimension $M = 40$, number of samples N varies from 10 to 80.

haustive search on α and ρ, it takes 48.25 seconds in the previous setting. However, to obtain the CLIME estimator, it takes 136.86 seconds and 653.13 seconds in support recovery step and cross-validation step, respectively.

2) Comparison With Classical Estimators: In our experiments, $[\mathbf{R}]_{ij} = 0.9^{|i-j|}$ and the columns of $\mathbf{Y}$ are generated from a Gaussian process. We consider 100 realizations and report plots of the average quadratic loss and Stein's loss.

First, we show the advantage of estimating a precision matrix under the quadratic loss. Our estimator is (QP). As benchmark methodologies, the following estimation methods are investigated: (SCM): Conventional sample covariance matrix $\hat{\mathbf{R}}_{\mathsf{SCM}}$; (CV): Cross Validation. We randomly partition the original sample into 5 subsamples. For each validation subsample, α and ρ are selected to minimize the quadratic loss (4) by replacing $\mathbf{R}^{-1}$ with the pseudo inverse of the sample covariance matrix in the validation subsample. Moreover, we consider classical estimator (Haff) [29] and a well known estimator (LW) [6]. We also use the performance of (Oracle) (which directly minimizes (4)) as a lower bound on the estimation error. It can be seen in Fig. 2 that (QP) outperforms all the other estimators and its performance is close to (Oracle). As the number of samples increases, all of the estimators perform better since the performance of the sample covariance matrix improves. (CV) does not improve much as N increases, this might be because it select large ρ due to the property of its quadratic loss.

In the next set of simulations, we compare our estimator (SP) with the estimators (SCM), (Haff), (CV), and (Oracle), as well as the classical estimator (DS) and (SC) which optimize (31) to obtain the covariance matrix estimator and then take the inverse. It can be seen in Fig. 3 (SP) outperforms all the other estimators and its performance is pretty close to that of oracle. Moreover, although the gap between the accuracy of (SP) and (CV) is not large, (SP) is 5 times faster than (CV) since the optimization process is repeated in all the 5 subsets in the latter method.

B. GMVP Problem With Real Data Simulation

We implement the GMVP problem in Section V using real market data. The data we use are Hang Seng Index of 45 stocks of Yahoo Finance daily close prices in the period of Jan 1, 2008 to July 31, 2011 (i.e., 720 days). We construct the matrix $\mathbf{Y}$ as the relative return of the prices.we compare our estimator (SP) with the estimators (SCM) and (LW). At a particular day t, we use the previous N days (i.e., $t-N$ to $t-1$) to construct the portfolio $\hat{\mathbf{w}}^{\star}$ and use this portfolio to compute the return in the following 100 days. The standard deviation of returns is plotted. It can be seen in Fig. 4 that our estimator (SP) outperforms the traditional (SCM) estimator as well as the (LW) estimator which is widely used in finance. Moreover, it has similar optimality as (CV) but is faster.

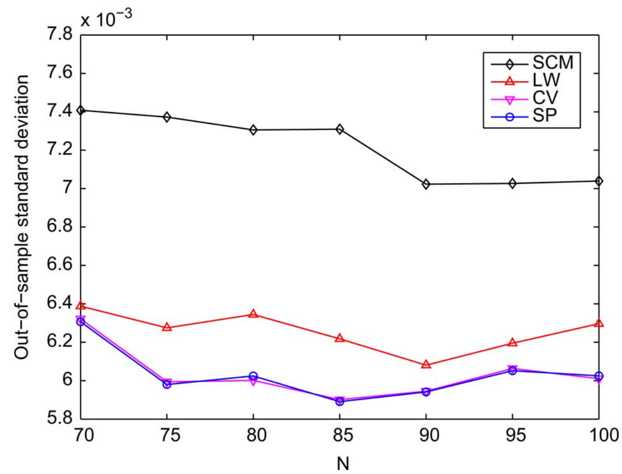

Fig. 4. Out-of-sample standard deviation of asset returns. Dimension $M = 45$, number of samples N varies from 70 to 100.

VI. Conclusions

In this paper we have studied the optimal calibration of estimators of precision matrices. We have resorted to high-dimensional asymptotics to account for the fact that the observation dimension is of the same order of magnitude as the number of samples. Moreover, the asymptotic optimality of uniform weighting has been revealed. Monte-Carlo simulations have shown the competitive advantage of our proposed methodology under realistic limited sample-size settings.

APPENDIX A
SOME USEFUL STOCHASTIC CONVERGENCE RESULTS

In this section we provide useful stochastic convergence results which will be used to prove Theorems 1 and 3. We begin with the following lemma, which can be proved by following [30].

Lemma 8: Assume that assumptions (A1) and (A2) hold, the following limit holds true almost surely:

$$\frac{1}{M}\mathrm{tr}\left[\left(\frac{1}{N}\mathbf{Y}_N\mathbf{T}_N\mathbf{Y}_N^H\right)^2\right] \asymp \frac{1}{M}\mathrm{tr}\left[\mathbf{T}_N^2\right]\left(\frac{1}{N}\mathrm{tr}\left[\mathbf{R}_M\right]\right)^2 + \frac{1}{M}\mathrm{tr}\left[\mathbf{R}_M^2\right]\left(\frac{1}{N}\mathrm{tr}\left[\mathbf{T}_N\right]\right)^2. \tag{34}$$

Now, we introduce the notation $\tilde{\mathbf{Y}}_N = \mathbf{R}_M^{1/2}\mathbf{X}_M\mathbf{T}_N^{1/2}$; additionally, in the sequel, $\mathbf{\Theta}_M$ will be a nonrandom $M \times M$ matrix whose trace norm is bounded uniformly in M. The following lemma will be instrumental in the proof of our results; see [23], [24] for a proof.

Lemma 9: Assume that Assumptions (A1) and (A2) hold. Then, for each $\rho_M > 0$ the following limits hold true almost surely:

$$\mathrm{tr}\left[\mathbf{\Theta}_M\left(\frac{1}{N}\tilde{\mathbf{Y}}_N\tilde{\mathbf{Y}}_N^H + \rho_M\mathbf{I}_M\right)^{-1}\right] \asymp \mathrm{tr}\left[\mathbf{\Theta}_M\left(\tilde{\delta}_M\mathbf{R}_M + \rho_M\mathbf{I}_M\right)^{-1}\right] \tag{35}$$

$$\frac{1}{M}\mathrm{tr}\left[\left(\frac{1}{N}\tilde{\mathbf{Y}}_N\tilde{\mathbf{Y}}_N^H + \rho_M\mathbf{I}_M\right)^{-2}\right] \asymp \frac{1}{M}\mathrm{tr}\left[\left(\tilde{\delta}_M\mathbf{R}_M + \rho_M\mathbf{I}_M\right)^{-2}\right] \frac{\tilde{\gamma}_M}{1-\gamma_M\tilde{\gamma}_M}c_M\left(\frac{1}{M}\mathrm{tr}\left[\mathbf{R}_M\left(\tilde{\delta}_M\mathbf{R}_M + \rho_M\mathbf{I}_M\right)^{-2}\right]\right)^2 \tag{36}$$

where $\{\delta_M,\, \tilde{\delta}_M\}$ and $\{\gamma_M,\, \tilde{\gamma}_M\}$ are defined as in (12), (9), and (10), respectively.

Along with Lemma 9, the following lemma will also be important for deriving our results. This lemma is originally stated in [31] (Theorem 4.1) for centered data matrix with more general variance profile. It also coincides with results in [32] for Gaussian case.

Lemma 10: Assume that Assumptions (A1) and (A2) hold. Then, for each $\rho_M > 0$ the following limit holds true almost surely:

$$\mathrm{logdet}\left(\frac{1}{N}\tilde{\mathbf{Y}}_N\tilde{\mathbf{Y}}_N^H + \rho_M\mathbf{I}_M\right) \asymp \mathrm{logdet}\left(\tilde{\delta}_M\mathbf{R}_M + \rho_M\mathbf{I}_M\right) + \mathrm{logdet}\left(\mathbf{I}_N + \delta_M\mathbf{T}_N\right) - N\delta_M\tilde{\delta}_M \tag{37}$$

where $\{\delta_M,\, \tilde{\delta}_M\}$ is defined as in (12).

APPENDIX B
PROOF OF THEOREM 1

A. The Term $\bar{\mathcal{L}}_{Q,\mathbf{R}}$

Note that

$$\mathcal{L}_{Q,\mathbf{R}} = \frac{1}{M}\mathrm{tr}\left[\left(\frac{1}{N}\mathbf{Y}_N\mathbf{T}_N\mathbf{Y}_N^H\right)^2\right] - \frac{2}{M}\mathrm{tr}\left[\left(\frac{1}{N}\mathbf{Y}_N\mathbf{T}_N\mathbf{Y}_N^H\right)\left(\mathbf{R}_M - \rho_M\mathbf{I}_M\right)\right] + \frac{1}{M}\mathrm{tr}\left[\left(\mathbf{R}_M - \rho_M\mathbf{I}_M\right)^2\right]. \tag{38}$$

The asymptotic equivalent of the first term in (38) follows directly by Lemma 8. Moreover, from the strong law of large numbers, we have the following result for the second term:

$$\frac{1}{N}\sum_{i=1}^{N} t_i\mathbf{y}_i^H\left(\mathbf{R}_M - \rho_M\mathbf{I}_M\right)\mathbf{y}_i \asymp \frac{1}{N}\mathrm{tr}\left[\mathbf{T}_N\right]\frac{1}{M}\mathrm{tr}\left[\mathbf{R}_M\left(\mathbf{R}_M - \rho_M\mathbf{I}_M\right)\right]. \tag{39}$$

Then, the result follows readily after some algebraic manipulations.

B. The Term $\bar{\mathcal{L}}_{Q,\mathbf{R}^{-1}}$

We recall that

$$\mathcal{L}_{Q,\mathbf{R}^{-1}} = \frac{1}{M}\mathrm{tr}\left[\left(\mathbf{R}_M\right)^{-2}\right] - \frac{2}{M}\mathrm{tr}\left[\mathbf{R}_M^{-1}\left(\frac{1}{N}\tilde{\mathbf{Y}}_N\tilde{\mathbf{Y}}_N^H + \rho_M\mathbf{I}_M\right)^{-1}\right] + \frac{1}{M}\mathrm{tr}\left[\left(\frac{1}{N}\tilde{\mathbf{Y}}_N\tilde{\mathbf{Y}}_N^H + \rho_M\mathbf{I}_M\right)^{-2}\right]. \tag{40}$$

The asymptotic equivalent of the second term in (40) can be obtained from Lemma 9 (35) with $\mathbf{\Theta}_M = \frac{1}{M}\mathbf{R}_M^{-1}$. Additionally, the asymptotic equivalent of the third term in (40) is presented in Lemma 9 (36).

C. The Terms $\bar{\mathcal{L}}_{S,\mathbf{R}}$ and $\bar{\mathcal{L}}_{S,\mathbf{R}^{-1}}$

Recall that

$$\mathcal{L}_{S,\mathbf{R}} = \frac{1}{M}\mathrm{tr}\left[\frac{1}{N}\mathbf{X}_M\mathbf{T}_N\mathbf{X}_M^H + \rho_M\mathbf{R}_M^{-1}\right] - \frac{1}{M}\mathrm{logdet}\left(\frac{1}{N}\tilde{\mathbf{Y}}_N\tilde{\mathbf{Y}}_N^H + \rho_M\mathbf{I}_N\right) + \frac{1}{M}\mathrm{logdet}(\mathbf{R}_M) - 1. \tag{41}$$

The asymptotic equivalent of the first term in (41) can be obtained from the strong law of large numbers, i.e.,

$$\frac{1}{M}\mathrm{tr}\left[\frac{1}{N}\mathbf{X}_M\mathbf{T}_N\mathbf{X}_M^H\right] \asymp \frac{1}{N}\mathrm{tr}[\mathbf{T}_N]. \tag{42}$$

And the asymptotic equivalent of the second term can be obtained from Lemma 10.

Moreover, notice that

$$\begin{aligned}\mathcal{L}_{S,\mathbf{R}^{-1}} = &\frac{1}{M}\mathsf{tr}\left[\mathbf{R}_M(\frac{1}{N}\tilde{\mathbf{Y}}_N\tilde{\mathbf{Y}}_N^H + \rho_M\mathbf{I}_M)^{-1}\right]\\ &+\frac{1}{M}\mathsf{logdet}\left(\frac{1}{N}\tilde{\mathbf{Y}}_N\tilde{\mathbf{Y}}_N^H + \rho_M\mathbf{I}_N\right)\\ &-\frac{1}{M}\mathsf{logdet}(\mathbf{R}_M) - 1,\end{aligned} \tag{43}$$

where the asymptotic equivalent of the first term in (43) can be obtained from Lemma 9 (a) by letting $\mathbf{\Theta}_M = \frac{1}{M}\mathbf{R}_M$, and the asymptotic equivalent of the second term in (43) can be obtained from Lemma 10.

Appendix C
Proof of Theorem 3

We first provide the following asymptotic results, which can be easily obtained by Stieltjes transform methods [19] (See, e.g., [33]) and that will be used in the following subsections:

$$\begin{aligned}&\frac{1}{M}\mathsf{tr}\left[\mathbf{R}_M\right] \asymp \frac{1}{M}\mathsf{tr}\left[\hat{\mathbf{R}}_{\mathsf{SCM}}\right]\\ &\frac{1}{M}\mathsf{tr}\left[\mathbf{R}_M^{-1}\right] \asymp (1-c_M)\frac{1}{M}\mathsf{tr}\left[\hat{\mathbf{R}}_{\mathsf{SCM}}^{-1}\right]\\ &\frac{1}{M}\mathsf{tr}\left[\mathbf{R}_M^{-2}\right] \asymp (1-c_M)^2\frac{1}{M}\mathsf{tr}\left[\hat{\mathbf{R}}_{\mathsf{SCM}}^{-2}\right]\\ &\qquad - c_M(1-c_M)\left(\frac{1}{M}\mathsf{tr}\left[\hat{\mathbf{R}}_{\mathsf{SCM}}^{-1}\right]\right)^2.\end{aligned} \tag{44}$$

A. The Term $\bar{\mathcal{L}}_{Q,\mathbf{R}}$

Note that $\bar{\mathcal{L}}_{Q,\mathbf{R}}$ can be written as

$$\begin{aligned}\bar{\mathcal{L}}_{Q,\mathbf{R}} = &2\rho_M\left(\frac{1}{N}\mathsf{tr}\left[\mathbf{T}_N\right]-1\right)\frac{1}{M}\mathsf{tr}\left[\mathbf{R}_M\right]\\ &+\frac{1}{N}\mathsf{tr}\left[\mathbf{T}_N^2\right]\frac{1}{MN}\left(\mathsf{tr}\left[\mathbf{R}_M\right]\right)^2\\ &+\left(\frac{1}{N}\mathsf{tr}\left[\mathbf{T}_N\right]-1\right)^2\frac{1}{M}\mathsf{tr}\left[\mathbf{R}_M^2\right] + \rho_M^2.\end{aligned} \tag{45}$$

Therefore, it is sufficient to estimate $\frac{1}{M}\mathsf{tr}\left[\mathbf{R}_M\right]$ and $\frac{1}{M}\mathsf{tr}\left[\mathbf{R}_M^2\right]$. Equation (44) ensures that $\frac{1}{M}\mathsf{tr}\left[\hat{\mathbf{R}}_{\mathsf{SCM}}\right]$ is a consistent estimator for $\frac{1}{M}\mathsf{tr}\left[\mathbf{R}_M\right]$. Recall that $\beta_M = \frac{1}{MN}\left(\mathsf{tr}\left[\mathbf{R}_M\right]\right)^2$, from (44), $\frac{1}{MN}\left(\mathsf{tr}\left[\hat{\mathbf{R}}_{\mathsf{SCM}}\right]\right)^2$ can be used as an estimator of β. But, a better estimator of β can be obtained from Lemma 8 by letting $\mathbf{T}_N = \mathbf{I}_N$:

$$\hat{\beta}_M = \frac{1}{MN^2}\sum_{i=1}^{N}\mathbf{y}_i^H\mathbf{y}_i\mathbf{y}_i^H\mathbf{y}_i = \frac{1}{MN^2}\sum_{i=1}^{N}\|\mathbf{y}_i\|^4. \tag{46}$$

Moreover, from Lemma 8 with $\mathbf{T}_N = \mathbf{I}_N$, $\frac{1}{M}\mathsf{tr}\left[\hat{\mathbf{R}}_{\mathsf{SCM}}^2\right] - \hat{\beta}_M \asymp \frac{1}{M}\mathsf{tr}\left[\mathbf{R}_M^2\right]$.

B. The Terms $\hat{\mathcal{L}}_{Q,\mathbf{R}^{-1}}$, $\hat{\mathcal{L}}_{S,\mathbf{R}}$, and $\hat{\mathcal{L}}_{S,\mathbf{R}^{-1}}$

Let us first consider $\hat{\mathcal{L}}_{Q,\mathbf{R}^{-1}}$, whose asymptotic equivalent is $\bar{\mathcal{L}}_{Q,\mathbf{R}^{-1}}$. Recall that

$$\begin{aligned}\bar{\mathcal{L}}_{Q,\mathbf{R}^{-1}} = &\frac{1}{M}\mathsf{tr}(\mathbf{R}_M^{-2}) + \frac{1}{M}\mathsf{tr}\left[\bar{\mathbf{R}}_{P,M}^{-2}\right]\\ &-\frac{2}{M}\mathsf{tr}\left[\mathbf{R}_M^{-1}\left(\tilde{\delta}_M\mathbf{R}_M + \rho_M\mathbf{I}_M\right)^{-1}\right]\\ &+\frac{\tilde{\gamma}_M}{1-\gamma_M\tilde{\gamma}_M}c_M\left(\frac{1}{M}\mathsf{tr}\left[\mathbf{R}_M\bar{\mathbf{R}}_{P,M}^{-2}\right]\right)^2.\end{aligned}$$

First, we provide consistent estimator of $\frac{1}{M}\mathsf{tr}\left[\mathbf{R}_M^{-1}\left(\tilde{\delta}_M\mathbf{R}_M + \rho_M\mathbf{I}_M\right)^{-1}\right]$. From matrix inversion lemma, we can decompose

$$\begin{aligned}&\mathbf{R}_M^{-1}\left(\tilde{\delta}_M\mathbf{R}_M + \rho_M\mathbf{I}_M\right)^{-1}\\ &\quad = \rho_M^{-1}\mathbf{R}_M^{-1} - \tilde{\delta}_M\rho_M^{-1}\left(\tilde{\delta}_M\mathbf{R}_M + \rho_M\mathbf{I}_M\right)^{-1}.\end{aligned} \tag{47}$$

Then, a consistent estimator of $\frac{1}{M}\mathsf{tr}\left[\rho_M^{-1}\mathbf{R}_M^{-1}\right]$ is presented by (44), and consistent estimator related to the second term of (47) can be obtained from Lemma 9 (a) by letting $\mathbf{\Theta}_M = \frac{1}{M}\mathbf{I}_M$. Then, from Lemma 9 (b), we get

$$\begin{aligned}\frac{1}{M}\mathsf{tr}\left[\hat{\mathbf{R}}_M^{-2}\right] \asymp &\frac{1}{M}\mathsf{tr}\left[\bar{\mathbf{R}}_{P,M}^{-2}\right]\\ &+\frac{\tilde{\gamma}_M}{1-\gamma_M\tilde{\gamma}_M}c_M(\frac{1}{M}\mathsf{tr}[\mathbf{R}_M\bar{\mathbf{R}}_{P,M}^{-2}])^2.\end{aligned} \tag{48}$$

And a consistent estimator of $\frac{1}{M}\mathsf{tr}\left[\mathbf{R}_M^{-2}\right]$ is presented by (44). See equations (49)–(51) at the bottom of the page.

Now we provide derivations for $\hat{\mathcal{L}}_{S,\mathbf{R}}$ and $\hat{\mathcal{L}}_{S,\mathbf{R}^{-1}}$. It is sufficient to provide consistent estimators of $\frac{1}{M}\mathsf{tr}\left[\mathbf{R}_M^{-1}\right]$, $\frac{1}{N}\mathsf{tr}\left[\mathbf{R}_M(\tilde{\delta}_M\mathbf{R}_M + \rho_M\mathbf{I}_M)^{-1}\right]$ and $\frac{1}{M}\mathsf{logdet}\left(\mathbf{R}_M\right)$. A consistent estimator of $\frac{1}{M}\mathsf{tr}\left[\mathbf{R}_M^{-1}\right]$ is (44), and consistent estimator of the second term can be obtained directly from Lemma 2

$$\mathcal{L}_{Q,\mathbf{R}^{-1}}^{(m)}(\breve{\mathbf{T}},\rho) = \frac{1}{mN}\mathsf{tr}\left[\left(\left(\frac{1}{mN}\breve{\mathbf{R}}^{1/2}\mathbf{X}_{mM\times mN}\breve{\mathbf{T}}\mathbf{X}_{mM\times mN}^H\breve{\mathbf{R}}^{1/2} + \rho\breve{\mathbf{R}}_0\right)^{-1} - \breve{\mathbf{R}}^{-1}\right)^2\right] \tag{49}$$

$$\delta^{(m)}(\breve{\mathbf{T}},\rho) = \frac{1}{mN}\mathsf{tr}[\breve{\mathbf{R}}(\frac{1}{mN}\breve{\mathbf{R}}^{1/2}\mathbf{X}_{mM\times mN}\breve{\mathbf{T}}\mathbf{X}_{mM\times mN}^H\breve{\mathbf{R}}^{1/2} + \rho\breve{\mathbf{R}}_0)^{-1}] \tag{50}$$

$$\eta^{(m)}(\breve{\mathbf{T}},\rho) = \frac{1}{mM}\mathsf{logdet}(\frac{1}{mN}\breve{\mathbf{R}}^{1/2}\mathbf{X}_{mM\times mN}\breve{\mathbf{T}}\mathbf{X}_{mM\times mN}^H\breve{\mathbf{R}}^{1/2} + \rho\breve{\mathbf{R}}_0). \tag{51}$$

by noticing that $\delta_M = \frac{1}{M}\mathsf{tr}\left[\mathbf{R}_M(\tilde{\delta}_M\mathbf{R}_M + \rho_M\mathbf{I}_M)^{-1}\right]$. Regarding the term $\frac{1}{M}\mathsf{logdet}(\mathbf{R}_M)$, from Marcenko-Pastur Law [34],

$$\frac{1}{M}\mathsf{logdet}(\mathbf{R}_M) \asymp \frac{1}{M}\sum_{i=1}^{\min(M,N)} \mathsf{log}\lambda_i(\hat{\mathbf{R}}_{\mathsf{scm},M}) - K \tag{52}$$

where $K = \frac{1}{2\pi c_M}\int_{lb}^{ub}\frac{\mathsf{log}x}{x}\sqrt{(x-lb)(ub-x)}dx$. When $c_M < 1$, the analytic solution of K is given in [35]: $K = (1 - c_M^{-1})\mathsf{log}(1 - c_M) - 1$. When $c_M > 1$, the analytic solution of K can also be computed by changing of variable $x = cy$ and using the same technique in [35], which is given by $K = c_M^{-1}\mathsf{log}c_M + c_M^{-1}(1 - c_M)\mathsf{log}(1 - c_M^{-1}) - c_M^{-1}$. When $c_M = 1$, letting $c_M \longrightarrow 1$ in either the previous two cases, we get $K = -1$.

Appendix D
Proof of Lemmas 4 and 6

We use the same technique as in [32] for the following proofs. In this section we omit the subscript M and N for clarity of presentation.

A. Proof of Lemma 4

We only provide the proof for function $\bar{\mathcal{L}}_{Q,\mathbf{R}^{-1}}(\mathbf{T},\rho)$, since the other proofs follow the same line of reasoning. Let $\otimes$ denote the Kronecker product of matrices. We introduce the matrices $\breve{\mathbf{R}} = \mathbf{I}_m \otimes \mathbf{R}$, $\breve{\mathbf{R}}_0 = \mathbf{I}_m \otimes \mathbf{R}_0$, $\breve{\mathbf{T}} = \mathbf{I}_m \otimes \mathbf{T}$ where matrices $\breve{\mathbf{R}}$ and $\breve{\mathbf{R}}_0$ are of size $mM \times mM$ and matrix $\breve{\mathbf{T}}$ is of size $mN \times mN$.

Consider the loss function (49) defined in the same way as $\mathcal{L}_{Q,\mathbf{R}^{-1}}(\mathbf{T},\rho)$, which is convex with respect to $(\breve{\mathbf{T}},\rho)$ since its Hessian is positive semidefinite [36]. Note that $\breve{\mathbf{T}} = \mathbf{I}_m \otimes \mathbf{T}$, then $\mathcal{L}^{(m)}_{Q,\mathbf{R}^{-1}}(\breve{\mathbf{T}},\rho)$ is also convex with respect to $(\mathbf{T},\rho)$.

Moreover, $\mathcal{L}^{(m)}_{Q,\mathbf{R}^{-1}}(\breve{\mathbf{T}},\rho)$ admits the asymptotic approximation $\bar{\mathcal{L}}^{(m)}_{Q,\mathbf{R}^{-1}}(\breve{\mathbf{T}},\rho)$ defined by (15), where the following substitutions are required:

$$\begin{aligned} &M \leftrightarrow mM,\ N \leftrightarrow mN \\ &\mathbf{R} \leftrightarrow \breve{\mathbf{R}},\ \mathbf{R}_0 \leftrightarrow \breve{\mathbf{R}}_0,\ \mathbf{T} \leftrightarrow \breve{\mathbf{T}}. \end{aligned} \tag{53}$$

A straightforward computation yields $\bar{\mathcal{L}}_{Q,\mathbf{R}^{-1}}(\mathbf{T},\rho) = \bar{\mathcal{L}}^{(m)}_{Q,\mathbf{R}^{-1}}(\breve{\mathbf{T}},\rho)$. Therefore, from Theorem 1,

$$\bar{\mathcal{L}}_{Q,\mathbf{R}^{-1}}(\mathbf{T},\rho) = \mathsf{lim}_{m\longrightarrow\infty}\mathcal{L}^{(m)}_{Q,\mathbf{R}^{-1}}(\breve{\mathbf{T}},\rho). \tag{54}$$

$\bar{\mathcal{L}}_{Q,\mathbf{R}}(\mathbf{T},\rho)$ is convex as a pointwise limit of convex functions.

B. Proof of Lemma 6

We denote asymptotic equivalents of $\delta^{(m)}(\breve{\mathbf{T}},\rho)$ and $\eta^{(m)}(\breve{\mathbf{T}},\rho)$ by $\bar{\delta}^{(m)}(\breve{\mathbf{T}},\rho)$ and $\bar{\eta}^{(m)}(\breve{\mathbf{T}},\rho)$ with substitution (53). Note (50) and (51), hence

$$\delta(\mathbf{T},\rho) = \bar{\delta}^{(m)}(\breve{\mathbf{T}},\rho) = \mathsf{lim}_{m\longrightarrow\infty}\delta^{(m)}(\breve{\mathbf{T}},\rho) \tag{55}$$

$$\eta(\mathbf{T},\rho) = \bar{\eta}^{(m)}(\breve{\mathbf{T}},\rho) = \mathsf{lim}_{m\longrightarrow\infty}\eta^{(m)}(\breve{\mathbf{T}},\rho). \tag{56}$$

It can easily be seen that $(\mathbf{T},\rho) \mapsto \delta^{(m)}(\breve{\mathbf{T}},\rho)$ is convex and $(\mathbf{T},\rho) \mapsto \eta^{(m)}(\breve{\mathbf{T}},\rho)$ is concave. Therefore, $\delta(\mathbf{T},\rho)$ and $\eta(\mathbf{T},\rho)$ are convex and concave as pointwise limits of convex and concave functions, respectively.

References

[1] T. Anderson, *An Introduction to Multivariate Statistical Analysis*. New York, NY, USA: Wiley, 1958.
[2] C. Stein, "Estimation of a covariance matrix," presented at the Rietz Lecture, 39th Ann. Meet. IMS, Atlanta, GA, USA, 1975.
[3] X. Mestre, "Improved estimation of eigenvalues and eigenvectors of covariance matrices using their sample estimates," *IEEE Trans. Inf. Theory*, vol. 54, no. 11, pp. 5113–5129, Nov. 2008.
[4] W. James and C. Stein, "Estimation with quadratic loss," in *Proc. 4th Berkeley Symp. Math. Statist. Probabil.*, Berkeley, CA, USA, 1961, vol. 1, pp. 361–379.
[5] D. K. Dey and C. Srinivasan, "Estimation of a covariance matrix under Stein's loss," *Ann. Statist.*, vol. 13, no. 4, pp. 1581–1591, Dec. 1985.
[6] O. Ledoit and M. Wolf, "A well-conditioned estimator for large-dimensional covariance matrices," *J. Multivar. Anal.*, vol. 88, no. 2, pp. 365–411, Feb. 2004.
[7] Y. Chen, A. Wiesel, Y. Eldar, and A. Hero, "Shrinkage algorithms for MMSE covariance estimation," *IEEE Trans. Signal Process.*, vol. 58, no. 10, pp. 5016–5029, Oct. 2010.
[8] A. Rothman, P. Bickel, E. Levina, and J. Zhu, "Sparse permutation invariant covariance estimation," *Electron. J. Statist.*, vol. 2, pp. 494–515, 2008.
[9] J. Fan, Y. Fan, and J. Lv, "High dimensional covariance matrix estimation using a factor model," *J. Econometr.*, vol. 147, no. 1, pp. 186–197, Nov. 2008.
[10] P. Bickel and E. Levina, "Regularized estimation of large covariance matrices," *Ann. Statist.*, vol. 36, no. 1, pp. 199–227, Feb. 2008.
[11] N. Meinshausen and P. Bühlmann, "High-dimensional graphs and variable selection with the lasso," *Ann. Statist.*, vol. 34, no. 3, pp. 1436–1462, 2006.
[12] C. Lam and J. Fan, "Sparsistency and rates of convergence in large covariance matrix estimation," *Ann. Statist.*, vol. 37, pp. 4254–4278, 2009.
[13] P. Ravikumar, M. Wainwright, G. Raskutti, and B. Yu, "High-dimensional covariance estimation by minimizing l1-penalized log-determinant divergence," *Electron. J. Statist.*, vol. 5, pp. 935–980, 2011.
[14] T. Cai, W. Liu, and X. Luo, "A constrained l1 minimization approach to sparse precision matrix estimation," *J. Amer. Statist. Assoc.*, vol. 106, no. 494, pp. 594–607, 2011.
[15] L. R. Haff, "Empirical Bayes estimation of the multivariate normal covariance matrix," *Ann. Statist.*, vol. 8, no. 3, pp. 586–597, May 1980.
[16] E. Levina, A. Rothman, and J. Zhu, "Sparse estimation for large covariance matrices via a nested LASSO penalty," *Ann. Appl. Statist.*, vol. 2, no. 1, pp. 245–263, Mar. 2008.
[17] T. Cai, C. Zhang, and H. Zhou, "Optimal rates of convergence for covariance matrix estimation," *Ann. Statist.*, vol. 38, no. 4, pp. 2118–2144, Aug. 2010.
[18] P. Stoica, J. Li, X. Zhu, and J. Guerci, "On using a priori knowledge in space-time adaptive processing," *IEEE Trans. Signal Process.*, vol. 56, no. 6, pp. 2598–2602, Jun. 2008.
[19] Z. Bai and J. Silverstein, *Spectral Analysis of Large Dimensional Random Matrices*. New York: Springer-Verlag, 2010.
[20] N. E. Karoui, "High-dimensionality effects in the markowitz problem and other quadratic programs with linear constraints: Risk underestimation," *Ann. Statist.*, vol. 38, no. 6, pp. 3487–3566, Dec. 2010.
[21] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. New York, NY, USA: Wiley , 2006.
[22] S. Verdú, "Spectral efficiency in the wideband regime," *IEEE Trans. Inf. Theory*, vol. 48, no. 6, pp. 1319–1343, 2002.
[23] F. Rubio and X. Mestre, "Spectral convergence for a general class of random matrices," *Statist. Probabil. Lett.*, vol. 81, no. 5, pp. 592–602, May 2011.

[24] F. Rubio, X. Mestre, and D. Palomar, "Performance analysis and optimal selection of large minimum variance portfolios under estimation risk," *IEEE J. Sel. Topics Signal Process.*, vol. 6, no. 4, pp. 337–350, Aug. 2012.
[25] M. Frank and P. Wolfe, "An algorithm for quadratic programming," *Naval Res. Logist. Quart.*, vol. 3, no. 1–2, pp. 95–110, Mar.–Jun. 1956.
[26] O. L. Mangasarian, "Machine learning via polyhedral concave minimization," Comput. Sci. Dep., Univ. Wisconsin, Math. Program., Tech. Rep. 95-20, 1995.
[27] M. Fazel, H. Hindi, and S. Boyd, "Log-det heuristic for matrix rank minimization with applications to Hankel and Euclidean distance matrices," in *Proc. IEEE Amer. Contr. Conf.*, Denver, CO, USA, Jun. 2003, vol. 3, pp. 2156–2162.
[28] G. Cornuejols and R. Tutuncu, *Optimization Methods in Finance*. Cambridge , U.K.: Cambridge Univ. Press, 2006.
[29] L. R. Haff, "Estimation of the inverse covariance matrix: Random mixtures of the inverse Wishart matrix and the identity," *Ann. Statist.*, pp. 1264–1276, 1979.
[30] X. Mestre, "Space Processing and Channel Estimation: Performance Analysis and Asymptotic Results," Ph.D., Tech. Univ. Catalonia (UPC), Catalonia, Spain, 2002.
[31] W. Hachem, P. Loubaton, and J. Najim, "Deterministic equivalents for certain functions of large random matrices," *Ann. Appl. Probabil.*, vol. 17, no. 3, pp. 875–930, Jun. 2007.
[32] W. Hachem, O. Khorunzhiy, P. Loubaton, J. Najim, and L. Pastur, "A new approach for mutual information analysis of large dimensional multi-antenna channels," *IEEE Trans. Inf. Theory*, vol. 54, no. 9, pp. 3987–4004, Sep. 2008.
[33] F. Rubio and X. Mestre, "On the estimation of the covariance eigenspectrum of array sample observations," presented at the 5th IEEE Sens. Array Multichannel Signal Process. Workshop (SAM), Darmstadt, Germany, Jul. 2008.
[34] A. Tulino and S. Verdú, *Random Matrix Theory and Wireless Communications*. New York, NY, USA: Now, 2004.
[35] Z. Bai and J. Silverstein, "CLT for linear spectral statistics of large-dimensional sample covariance matrices," *Ann. Probabil.*, vol. 32, no. 1A, pp. 553–605, Jan. 2004.
[36] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, U.K.: Cambridge Univ. Press, 2004.

Mengyi Zhang (S'09) received the B.Sc. degree in electronic information from Huazhong University of Science and Technology (HUST), Wuhan, China, in 2009.

Currently, she is pursuing the Ph.D degree with the Department of Electronic and Computer Engineering, Hong Kong University of Science and Technology (HKUST). Her research interests are mainly on statistical signal processing and optimization in wireless communication and financial systems.

Francisco Rubio (M'07) received the B.S. degree in electrical engineering from the Technical University of Madrid, Spain, the M.S. degree in information science from the Mannheim University of Applied Sciences, Germany, and the Ph.D. degree from the Technical University of Catalonia, Spain, in 2001, 2002, and 2008, respectively.

Since March 2012, he has been a Quantitative Analyst with Genetic Finance Ltd., Hong Kong. From January 2011 to March 2012, he was a Research Associate with the Hong Kong University of Science and Technology, and from June 2008 to January 2011, he was a Research Associate and the head of the Engineering Department of the Centre Tecnològic de Telecomunicacions de Catalunya, Barcelona, Spain. From July to December 2007, he held a visiting appointment as a Predoctoral Research Fellow with the Department of Electrical Engineering and Computer Science, Northwestern University (Evanston, IL). Prior to the Ph.D., he held research and professional appointments at Retevisión Móvil (now France Telecom's Orange) and Siemens, Madrid, and T-Nova Deutsche Telekom (Darmstadt, Germany), MABEL Biomedical Engineering Labs (Mannheim), and BMW AG (Munich, Germany).

Dr. Rubio has actively participated in several European research projects and industrial contracts, and is a corecipient of an academic performance award from T-Systems.

Daniel P. Palomar (S'99–M'03–SM'08–F'12) received the electrical engineering and Ph.D. degrees (both with honors) from the Technical University of Catalonia (UPC), Barcelona, Spain, in 1998 and 2003, respectively.

He is an Associate Professor with the Department of Electronic and Computer Engineering, Hong Kong University of Science and Technology (HKUST), Hong Kong, which he joined in 2006. Since 2013, he has been a Fellow of the Institute for Advanced Study (IAS) at HKUST. He had previously held several research appointments, namely, at King's College London (KCL), London, U.K.; Technical University of Catalonia (UPC), Barcelona; Stanford University, Stanford, CA; Telecommunications Technological Center of Catalonia (CTTC), Barcelona; Royal Institute of Technology (KTH), Stockholm, Sweden; University of Rome "La Sapienza," Rome, Italy; and Princeton University, Princeton, NJ. His current research interests include applications of convex optimization theory, game theory, and variational inequality theory to financial systems and communication systems.

Dr. Palomar is a recipient of a 2004/2006 Fulbright Research Fellowship, the 2004 Young Author Best Paper Award by the IEEE Signal Processing Society, the 2002/2003 best Ph.D. prize in Information Technologies and Communications by the Technical University of Catalonia (UPC), the 2002/2003 Rosina Ribalta first prize for the Best Doctoral Thesis in Information Technologies and Communications by the Epson Foundation, and the 2004 prize for the Best Doctoral Thesis in Advanced Mobile Communications by the Vodafone Foundation and COIT. He serves as an Associate Editor of the IEEE TRANSACTIONS ON INFORMATION THEORY and has been an Associate Editor of the IEEE TRANSACTIONS ON SIGNAL PROCESSING, a Guest Editor of the *IEEE Signal Processing Magazine* 2010 Special Issue on Convex Optimization for Signal Processing, the IEEE JOURNAL ON SELECTED AREAS IN COMMUNICATIONS 2008 Special Issue on Game Theory in Communication Systems, and the IEEE JOURNAL ON SELECTED AREAS IN COMMUNICATIONS 2007 Special Issue on Optimization of MIMO Transceivers for Realistic Communication Networks. He serves on the IEEE Signal Processing Society Technical Committee on Signal Processing for Communications (SPCOM).