CALIBRATION OF HIGH-DIMENSIONAL PRECISION MATRICES UNDER QUADRATIC LOSS

Mengyi Zhang, Francisco Rubio, and Daniel P. Palomar

Department of Electronic and Computer Engineering, Hong Kong University of Science and Technology, Hong Kong

ABSTRACT

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 (i.e. the inverse covariance matrix) based on weighted sampling and linear shrinkage. The estimation error is measured in terms of the matrix quadratic loss, and the latter is used to calibrate the set of parameters defining our proposed estimator. In an asymptotic setting where the observation dimension is of the same order of magnitude as the number of samples, we provide an estimator of the precision matrix that is as good as the oracle estimator. Our research is based on recent contributions in the field of random matrix theory and Monte-Carlo simulations show the advantage of our precision matrix estimator in finite sample size settings.

1. INTRODUCTION

Many applications of statistical signal processing require an estimate of a precision matrix. If the number of samples is large compared to the observation dimension, the sample covariance matrix is consistent with the population covariance matrix and its inverse can be effectively used as an estimator of the precision matrix. However, for high-dimensional data, it is well known that the eigenvalues of the sample covariance matrix considerably spread out from those of the population covariance matrix [1, 2]; see also references therein. Therefore, it is problematic to use the inverse of the sample covariance matrix as the estimator of the precision matrix , because inverting it may amplify the estimation error dramatically.

Plenty of estimators of precision matrices have been proposed. In some of the work such as [3, 4, 5], additional properties of the population covariance matrix are required, such as sparseness or low rank (under a factor model assumption). These estimators perform well when estimating a particular structured population precision matrix.

We investigate a class of estimators of covariance matrices and precision matrices that are based on linear shrinkage estimation as well as weighted sampling. Linear shrinkage estimators are extensively used to estimate covariance matrices [6, 7, 8] but are seldom used to estimate precision matrices. The reason is that it is impossible to obtain a closed form expression of the expectation of the quadratic loss of the precision matrix estimator. To tackle this problem, we use high-dimensional asymptotics of quadratic loss function. Furthermore, weighted sampling is also a widely used technique in statistical sampling theory, e.g, nonparametric bootstrap [9].

A quadratic loss is used to quantify the estimation error. To reflect the fact that the observation dimension is of the same order of magnitude as the number of samples, we employ high-dimensional asymptotics where both the observation dimension and the number of samples go to infinity. Our first main contribution is to obtain an estimator of the precision matrix which is asymptotically as good as the oracle estimator. The second main contribution is to reveal the asymptotic optimality of uniform weighting.

The rest of the paper is organized as follows. Section II is devoted to introduce the structured estimators and quadratic loss functions. Important theoretical results for covariance matrix and precision matrix estimation are provided in Section III. In Section IV, optimal selection of calibration parameters is discussed. Section V presents numerical results and Section VI concludes this paper.

This work was supported by the Hong Kong RPC11EG39 research grant.

2. PROBLEM FORMULATION

Let $\{\mathbf{y}_i \in \mathbb{C}^M\}_{i=1}^N$ 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 $\{\mathbf{y}_i\}_{i=1}^N$. For notational convenience, we denote $\mathbf{Y}_N = [\mathbf{y}_1, ..., \mathbf{y}_N]$ the $M \times N$ observation matrix. In this paper, we study a precision matrix estimator $\hat{\mathbf{R}}_M^{-1}$ where $\hat{\mathbf{R}}_M$ has 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 $\mathbf{T}_N \in \mathcal{D}_+$ with $\mathcal{D}_+$ denoting the set of positive semidefinite diagonal matrices is a diagonal weighting matrix with diagonal elements representing a set of nonnegative sample weights, $\mathbf{I}_M$ is the shrinkage target, and ρ_M is the nonnegative shrinkage coefficient.

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. The latter is achieved by sensibly selecting the free parameters $\mathbf{T}_N$ and ρ_M as we will show in the sequel. Consider an example of $\mathbf{T}_N = (1-\rho_M)\mathbf{I}_N$ where $0 \le \rho_M \le 1$. $\hat{\mathbf{R}}_M$ is an unbiased estimator of $\mathbf{R}_M$ if $\rho_M = 0$; choosing a nonzero ρ_M will decrease the variance of $\hat{\mathbf{R}}_M$ but introduce some bias. In this particular case, we can adjust this tradeoff by different choices of ρ_M.

A quadratic loss function is used to quantify the estimation error. Given a certain unknown matrix $\mathbf{B}_M$ and its estimator $\hat{\mathbf{B}}_M$, the quadratic loss can be written as the normalized Frobenius norm of the error matrix $\mathbf{B}_M - \hat{\mathbf{B}}_M$:

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

where $||\mathbf{B}||_F = \sqrt{\mathrm{tr}[\mathbf{B}^H\mathbf{B}]}$ is the Frobenius norm. We normalize the Frobenius norm for asymptotic purpose so that $\mathcal{L}(\mathbf{I}_M) = 1$ regardless of the dimension; see [6].

In our case, the estimation error matrix with respect to covariance matrix and precision matrix are $\mathbf{R}_M - \hat{\mathbf{R}}_M$ and $\mathbf{R}_M^{-1} - \hat{\mathbf{R}}_M^{-1}$, respectively. The corresponding quadratic loss functions are

$$\mathcal{L}_C = \frac{1}{M}||\mathbf{R}_M - \hat{\mathbf{R}}_M||_F^2 \tag{2}$$

$$\mathcal{L}_P = \frac{1}{M}||\mathbf{R}_M^{-1} - \hat{\mathbf{R}}_M^{-1}||_F^2 \tag{3}$$

The oracle estimators are defined as estimators with structure (1) which have the minimum loss functions (2) and (3). In oracle estimators, we have to know the population matrix $\mathbf{R}_M$ in order to obtain the estimators. Ideally, the free parameters $(\mathbf{T}_N^{\mathsf{or}}, \rho_M^{\mathsf{or}})$ in the oracle estimators are calibrated through the following optimization problem with objective functions (2) and (3), respectively:

$$\begin{aligned} \underset{(\mathbf{T}_N, \rho_M)}{\mathsf{minimize}} \quad & \mathcal{L}(\mathbf{T}_N, \rho_M) \\ \mathsf{subject\ to} \quad & \mathbf{T}_N \in \mathcal{D}_+ \ \rho_M \geq 0. \end{aligned} \tag{4}$$

Note that the optimal parameters can not be obtained in practice since the objective functions (2) and (3) depend on the unknown covariance matrix $\mathbf{R}_M$. The naive approach is based on the plug-in estimators of (2) and (3), which is given by replacing the unknown covariance matrix $\mathbf{R}_M$ by the sample covariance matrix $\hat{\mathbf{R}}_{\mathsf{SCM}} = \frac{1}{N}\mathbf{Y}_N\mathbf{Y}_N^H$, i.e.,

$$\mathcal{L}_C^{\mathsf{plug-in}} = \frac{1}{M}||\hat{\mathbf{R}}_{\mathsf{SCM}} - \hat{\mathbf{R}}_M||_F^2 \tag{5}$$

$$\mathcal{L}_P^{\mathsf{plug-in}} = \frac{1}{M}||\hat{\mathbf{R}}_{\mathsf{SCM}}^{-1} - \hat{\mathbf{R}}_M^{-1}||_F^2. \tag{6}$$

However, minimizing (5) and (6) yield trivial solutions $\mathbf{T}_N = \mathbf{I}_N$ and $\rho_M = 0$, then $\hat{\mathbf{R}}_M$ becomes the sample covariance matrix.

If an estimate of the covariance matrix is required and (2) is used as the objective in problem (4), the following approach is employed in [6] to obtain the optimal parameters: assume $\mathbf{T}_N = \alpha_M \mathbf{I}_N$ and $\hat{\mathbf{R}}_M = \alpha_M \hat{\mathbf{R}}_{\mathsf{SCM}} + \rho_M \mathbf{I}_M$ which is a special case of structure (1), expectation of $\mathcal{L}_C$ (c.f.(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, this method can not be applied for precision matrix estimation because expectation of (3) cannot be obtained. We tackle this problem by using asymptotics as an approximation: instead of taking expectations, we first derive asymptotic equivalent of (3) in the double limit that is deterministic; the asymptotic equivalent is not observable and therefore we provide its consistent estimator, which is also consistent with (3) in the double limit. The optimal parameters $(\mathbf{T}_N^{\mathsf{or}}, \rho_M^{\mathsf{or}})$ in the precision matrix estimator can therefore be approximately obtained by minimizing the consistent estimator of (3). This estimation scheme will be illustrated in detail in Section IV.

In the following section, we provide two theorems describing the asymptotic equivalents and consistent estimators for quantities (2) and (3).

3. ASYMPTOTIC EQUIVALENTS AND CONSISTENT ESTIMATORS

In this section we provide our theoretical findings, whose proofs are included in a full version of this work [10]. In particular, Theorem 1 in [11] is instrumental in the proof. We first begin with technical hypotheses and some further definitions.

3.1. Assumptions and further definitions

The following set of assumptions will be maintained throughout the paper.

(A1) Let the population covariance matrix $\mathbf{R}_M$ be nonrandom $M \times M$ positive definite matrices, with spectral norm being uniformly bounded in M.

(A2) The sample weight matrix $\mathbf{T}_N$ is an $N \times N$ diagonal matrix with real nonnegative entries uniformly bounded in M.

(A3) Let $\mathbf{X}_M$ be an $M \times N$ random matrix such that the entries of $\mathbf{X}_M$ are i.i.d. complex Gaussian random variables with mean zero, variance one. Then the observation matrix can be expressed as $\mathbf{Y}_N = \mathbf{R}_M^{1/2}\mathbf{X}_M$.[1]

We will consider the limiting regime that $M,\ N \to \infty$ with $0 < \mathsf{liminf} c_M < \mathsf{limsup} c_M < \infty$ where $c_M = M/N$. In this limiting regime, $a \asymp b$ denotes they are asymptotic equivalents, i.e., $|a - b| \to 0$ almost surely.

Before proceeding to the main theorems in this paper, we introduce some further definitions: we define $\beta_M = c_M \left(\frac{1}{M}\mathsf{tr}[\mathbf{R}_M]\right)^2$, and its generalized consistent estimator $\hat{\beta}_M = \frac{1}{MN^2}\sum_{i=1}^N ||\mathbf{y}_i||^4$.

Moreover, we introduce

$$\gamma_M = \frac{1}{N}\mathsf{tr}\left[(\mathbf{R}_M(\tilde{\delta}_M \mathbf{R}_M + \rho_M \mathbf{I}_M)^{-1})^2\right] \tag{7}$$

$$\tilde{\gamma}_M = \frac{1}{N}\mathsf{tr}\left[(\mathbf{T}_N(\mathbf{I}_N + \delta_M \mathbf{T}_M)^{-1})^2\right]. \tag{8}$$

where $(\delta_M,\ \tilde{\delta}_M)$ is the unique positive solution of the system of equations [11]:

$$\begin{cases} \delta_M = \frac{1}{N}\mathsf{tr}\left[\mathbf{R}_M(\tilde{\delta}_M \mathbf{R}_M + \rho_M \mathbf{I}_M)^{-1}\right] \\ \tilde{\delta}_M = \frac{1}{N}\mathsf{tr}\left[\mathbf{T}_N(\mathbf{I}_N + \delta_M \mathbf{T}_N)^{-1}\right]. \end{cases} \tag{9}$$

Furthermore, we define matrices $\bar{\mathbf{R}}_{C,M}$ and $\bar{\mathbf{R}}_{P,M}$ as

$$\bar{\mathbf{R}}_{C,M} = \frac{1}{N}\mathsf{tr}\left[\mathbf{T}_N\right]\mathbf{R}_M + \rho_M \mathbf{I}_M \tag{10}$$

$$\bar{\mathbf{R}}_{P,M} = \tilde{\delta}_M \mathbf{R}_M + \rho_M \mathbf{I}_M. \tag{11}$$

The above quantities will be used to describe the asymptotic deterministic equivalents and the consistent estimators, respectively.

3.2. Asymptotic equivalents of loss functions

The following theorem shows almost sure convergence for $\mathcal{L}_C$ and $\mathcal{L}_P$.

Theorem 1. *Define the following quantities:*

$$\bar{\mathcal{L}}_C = \frac{1}{M}||\mathbf{R}_M - \bar{\mathbf{R}}_{C,M}||_F^2 + \beta_M \frac{1}{N}\mathsf{tr}[\mathbf{T}_N^2] \tag{12}$$

$$\begin{aligned} \bar{\mathcal{L}}_P = & \frac{1}{M}||\mathbf{R}_M^{-1} - \bar{\mathbf{R}}_{P,M}^{-1}||_F^2 \\ & + \frac{\tilde{\gamma}_M}{1 - \gamma_M \tilde{\gamma}_M} c_M \left(\frac{1}{M}\mathsf{tr}[\mathbf{R}_M(\bar{\mathbf{R}}_{P,M})^{-2}]\right)^2. \end{aligned} \tag{13}$$

Under Assumptions (A1)-(A3), it follows that $\mathcal{L}_C \asymp \bar{\mathcal{L}}_C$ and $\mathcal{L}_P \asymp \bar{\mathcal{L}}_P$, i.e., the random quantities (2)-(3) tend to deterministic quantities (12)-(13) in the double limit.

Theorem 1 provides asymptotic approximations of loss functions in the double-limit regime. It enables us to analyze the consistency of the structured estimators (1) with fixed parametrization, i.e., when ρ_M, $\mathbf{T}_N$ are determined.

[1] This assumption can be generalized to $\mathbf{X}_M$ with distribution-free entries.

3.3. Consistent estimators of loss functions

In the following, we provide consistent estimators of quadratic loss functions, which can be effectively used as objective functions when calibrating the parameters in covariance matrix and precision matrix estimators. We begin with the following lemma which provides consistent estimators of δ_M and $\tilde{\delta}_M$, which will be used to derive the consistent estimators of loss functions.

Lemma 1. *[12] Under assumptions **(A1)-(A3)**, a consistent estimator of δ_M, denoted by $\hat{\delta}_M$, is given by the unique positive solution of the following equation:*

$$\hat{\delta}_M \frac{1}{N}\mathsf{tr}\left[\mathbf{T}_N\left(\mathbf{I}_N+\hat{\delta}_M\mathbf{T}_N\right)^{-1}\right]=$$
$$\frac{1}{N}\mathsf{tr}\left[\frac{1}{N}\mathbf{Y}_N\mathbf{T}_N\mathbf{Y}_N^H\left(\frac{1}{N}\mathbf{Y}_N\mathbf{T}_N\mathbf{Y}_N^H+\rho_M\mathbf{I}_M\right)^{-1}\right]. \quad (14)$$

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

$$\hat{\tilde{\delta}}_M=\frac{1}{N}\mathsf{tr}\left[\mathbf{T}_N\left(\mathbf{I}_N+\hat{\delta}_M\mathbf{T}_N\right)^{-1}\right]. \quad (15)$$

Theorem 2. *Define the following quantities:*

$$\hat{\mathcal{L}}_C=\mathcal{L}_C^{\mathsf{plug-in}}+\psi_C \quad (16)$$

$$\hat{\mathcal{L}}_P=\mathcal{L}_P^{\mathsf{plug-in}}+\psi_P \quad (17)$$

where correction terms of plug-in estimators (5) and (6) are given by:

$$\psi_C=\hat{\beta}_M\left(\frac{1}{N}\mathsf{tr}[\mathbf{T}_N^2]-\left(\frac{1}{N}\mathsf{tr}[\mathbf{T}_N]-1\right)^2\right)$$
$$\psi_P=\frac{2}{M}\mathsf{tr}\left[(1-c_M)\,\rho_M^{-1}\left(\hat{\tilde{\delta}}_M\hat{\mathbf{R}}_M^{-1}-\hat{\mathbf{R}}_{\mathsf{SCM}}^{-1}\right)+\hat{\mathbf{R}}_{\mathsf{SCM}}^{-1}\hat{\mathbf{R}}_M^{-1}\right]$$
$$-(2c_M-2c_M^2)\frac{1}{M}\mathsf{tr}[\hat{\mathbf{R}}_{\mathsf{SCM}}^{-2}]-\left(c_M-c_M^2\right)\left(\frac{1}{M}\mathsf{tr}[\hat{\mathbf{R}}_{\mathsf{SCM}}^{-1}]\right)^2$$

Under Assumptions (A1)-(A3), it follows that $\hat{\mathcal{L}}_C\asymp\bar{\mathcal{L}}_C$ and $\hat{\mathcal{L}}_P\asymp\bar{\mathcal{L}}_P$, i.e., the random observable quantities (16)-(17) tend to the deterministic quantities (12)-(13) in the double limit.

Remark 1. It is required that $c_M<1$ in equation (17).

4. OPTIMAL SELECTION OF CALIBRATION PARAMETERS

In this section and the following section, the subscripts M and N will be omitted for clarity of presentation. The parameters $\mathbf{T}$ and ρ in estimator (1) represent a set of degrees-of-freedom with respect to which estimation performance can be improved. Ideally, we use consistent estimators (12) and (13) to calibrate these parameters and formulate the following set of problems:

$$\begin{array}{ll}\underset{(\mathbf{T},\rho)}{\mathsf{minimize}} & \hat{\mathcal{L}}(\mathbf{T},\rho)\\ \mathsf{subject\ to} & \mathbf{T}\in\mathcal{D}_+\ \rho\geq 0.\end{array} \quad (18)$$

In general, the set of problems (18) with objective functions (16) and (17) are nonconvex. However, we can study the structure of optimal solutions asymptotically by replacing (16) and (17) with their corresponding asymptotic equivalents (12) and (13). Now we focus on the optimal solutions to the following set of problems:

$$\begin{array}{ll}\underset{(\mathbf{T},\rho)}{\mathsf{minimize}} & \bar{\mathcal{L}}(\mathbf{T},\rho)\\ \mathsf{subject\ to} & \mathbf{T}\in\mathcal{D}_+\ \rho\geq 0.\end{array} \quad (19)$$

Proposition 1. *There exists a global optimal solution $(\mathbf{T}^\star,\rho^\star)$ to optimization problem (19) with objective function (12) and (13) in the form of $(\alpha^\star\mathbf{I},\rho^\star)$.*

Proposition 1 has revealed the asymptotic optimality of uniform weighting. With Proposition 1, the number of variables of problem (19) is reduced from $N+1$ to 2. At this point, we can return to problem (18) with variable $\mathbf{x}=[\alpha,\rho]^T$. We start with the loss function $\hat{\mathcal{L}}_C$; the expression (16) in terms of $\mathbf{x}$ is

$$\hat{\mathcal{L}}_C(\mathbf{x})=\mathbf{x}^T\mathbf{Q}\mathbf{x}-2\mathbf{p}^T\mathbf{x}-\hat{\beta}_M \quad (20)$$

where

$$\mathbf{Q}=\begin{bmatrix}\frac{1}{M}||\hat{\mathbf{R}}_{\mathsf{SCM}}||_F^2 & \frac{1}{M}\mathsf{tr}[\hat{\mathbf{R}}_{\mathsf{SCM}}]\\ \frac{1}{M}\mathsf{tr}[\hat{\mathbf{R}}_{\mathsf{SCM}}] & 1\end{bmatrix}$$
$$\mathbf{p}=\begin{bmatrix}\frac{1}{M}||\hat{\mathbf{R}}_{\mathsf{SCM}}||_F^2-\hat{\beta} & \frac{1}{M}\mathsf{tr}[\hat{\mathbf{R}}_{\mathsf{SCM}}]\end{bmatrix}^T.$$

From the Cauchy–Schwarz inequality, $\mathbf{Q}$ is positive semidefinite. Problem (18) becomes a quadratic program and can be efficiently solved.

In particular, for the relevant case $\alpha=1-\rho$, the optimal shrinkage coefficient of problem (18) is given by

$$\rho^\star=\mathsf{min}\left(\frac{\hat{\beta}}{\frac{1}{M}||\hat{\mathbf{R}}_{\mathsf{SCM}}-\mathbf{I}||_F^2},1\right). \quad (21)$$

In [6], the authors minimized the expectation of the quadratic loss function and obtained the shrinkage coefficient. Under the same asymptotic setting $M,\,N\to\infty$, the optimal shrinkage coefficient is estimated as:

$$\rho_{LW}^\star=\mathsf{min}\left(\frac{\hat{\beta}-\frac{1}{MN}||\hat{\mathbf{R}}_{\mathsf{SCM}}||_F^2}{\frac{1}{M}||\hat{\mathbf{R}}_{\mathsf{SCM}}-\frac{1}{M}\mathsf{tr}[\hat{\mathbf{R}}_{\mathsf{SCM}}]\mathbf{I}||_F^2},1\right). \quad (22)$$

If $\frac{1}{M}\mathsf{tr}[\mathbf{R}]=1$ (which can be generalized to arbitrary $\mathbf{R}$ with a general shrinkage target $\mathbf{R}_0$), it can be seen that (21) is asymptotically equivalent to (22).

We now consider the loss function $\mathcal{L}_P$; the expression (17) in terms of $\mathbf{x}$ is

$$\begin{aligned}\hat{\mathcal{L}}_P(\alpha,\rho)=&(1-c)^2\frac{1}{M}\mathsf{tr}\left[\hat{\mathbf{R}}_{\mathsf{SCM}}^{-2}\right]-c(1-c)\left(\frac{1}{M}\mathsf{tr}\left[\hat{\mathbf{R}}_{\mathsf{SCM}}^{-1}\right]\right)^2\\&+\frac{1}{M}\mathsf{tr}\left[\left(\alpha\hat{\mathbf{R}}_{\mathsf{SCM}}+\rho\mathbf{I}\right)^{-2}\right]\\&-2(1-c)\frac{1}{M}\mathsf{tr}\left[\rho^{-1}\hat{\mathbf{R}}_{\mathsf{SCM}}^{-1}\right]\\&+2(1-c)\hat{\tilde{\delta}}(\alpha,\rho)\frac{1}{M}\mathsf{tr}\left[\rho^{-1}\left(\alpha\hat{\mathbf{R}}_{\mathsf{SCM}}+\rho\mathbf{I}\right)^{-1}\right]\end{aligned} \quad (23)$$

where

$$\hat{\tilde{\delta}}(\alpha,\rho)=\alpha-\alpha^2\frac{1}{N}\mathsf{tr}\left[\hat{\mathbf{R}}_{\mathsf{SCM}}\left(\alpha\hat{\mathbf{R}}_{\mathsf{SCM}}+\rho\mathbf{I}\right)^{-1}\right].$$

Note that the objective function (23) is nonconvex in general. Exhaustive search can be used for calibrating (α, ρ) in the precision matrix estimator since there are only two variables. A convex optimization based algorithm which is more time efficient is proposed in the in a full version of this work [10], but skipped here for space reasons. With these methods, we can obtain a precision matrix estimator which is asymptotically as good as the oracle estimator in Section II.

5. MONTE CARLO SIMULATIONS

In the simulation, we show our advantage in estimating the precision matrix with finite sample size settings. We directly calibrate the parameters (α, ρ) in the precision matrix estimator $\hat{\mathbf{R}} = \alpha\hat{\mathbf{R}}_{\mathsf{SCM}} + \rho\mathbf{I}$ by optimizing problem (18) with objective function (23). Our estimator is named Quadra-Precision. Besides, the following estimation methods are investigated, in which the precision matrix estimators are obtained by estimating the covariance matrix and taking the inverse. These estimators are listed as: (1) LW [6]: $\hat{\mathbf{R}}_{LW} = (1-\rho)\hat{\mathbf{R}}_{\mathsf{SCM}} + \rho\frac{1}{M}\mathrm{tr}[\hat{\mathbf{R}}_{\mathsf{SCM}}]\mathbf{I}$ where ρ is given in (22); (2) SCM: the conventional sample covariance matrix $\hat{\mathbf{R}}_{\mathsf{SCM}}$; (3) Equally weighted SCM with the shrinkage target: i.e., $0.5\hat{\mathbf{R}}_{\mathsf{SCM}} + 0.5\mathbf{I}$. (4) Oracle: We assume covariance matrix $\mathbf{R}$ is known to us and $(\alpha,\, \rho)$ minimizes (23). The performance of Oracle is the lower bound.

We let $\mathbf{R}$ be the covariance matrix of a Gaussian autoregressive process with entries $[\mathbf{R}]_{ij} = 0.9^{|i-j|}$. The columns of $\mathbf{Y}$ are generated from a Gaussian process. The simulation is repeated 100 times and the average quadratic loss is plotted here.

In Fig.1, we omit quadratic loss of sample covariance matrix because it is over 1000. It can be seen that the performance of equally weighted SCM with the shrinkage target is the worst, since the shrinkage coefficient is fixed for each realization. The performance of Quadra-Precision is close to that of Oracle. It outperforms the LW estimator, which coincides with the intuition that it is better to estimate a precision matrix directly than estimating the covariance matrix and taking the inverse. As the number of samples increases, quadratic losses of LW estimator and Quadra-Precision estimator become closer to that of Oracle, since the sample covariance matrix behaves better with larger number of samples.

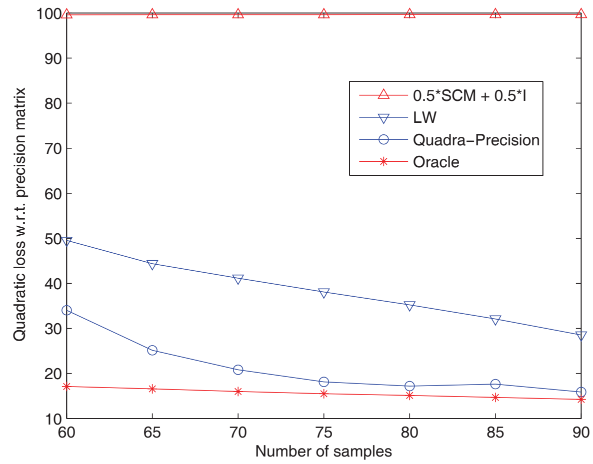

Fig. 1. Quadratic loss w.r.t. precision matrix. Dimension $M = 40$, number of samples N varies from 60 to 90.

6. CONCLUSIONS

In this paper we have studied consistency and calibration of estimators of covariance matrices and precision matrices that are based on linear shrinkage as well as weighted sampling. We have employed high-dimensional asymptotics to reflect the fact that the observation dimension is of the same order of magnitude as the number of samples. We are able to effectively use the proposed structure so as to further reduce the estimation error, as quantified by the matrix quadratic loss. This means improving the bias-variance tradeoff for covariance matrix estimation, especially the estimation of the precision matrix. Moreover, the asymptotic optimality of uniform weighting has been revealed by asymptotic deterministic equivalents of the loss functions. Monte-Carlo simulations have shown our advantage in estimating the precision matrix in finite sample size settings.

7. REFERENCES

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

[2] X. Mestre, "Improved estimation of eigenvalues and eigenvectors of covariance matrices using their sample estimates," *IEEE Transactions on Information Theory*, vol. 54, no. 11, pp. 5113 –5129, nov. 2008.

[3] A. Rothman, P. Bickel, E. Levina, and J. Zhu, "Sparse permutation invariant covariance estimation," *Electronic Journal of Statistics*, vol. 2, pp. 494–515, 2008.

[4] N. Karoui, "Spectrum estimation for large dimensional covariance matrices using random matrix theory," *The Annals of Statistics*, vol. 36, no. 6, pp. 2757–2790, 2008.

[5] J. Fan, Y. Fan, and J. Lv, "High dimensional covariance matrix estimation using a factor model," *Journal of Econometrics*, vol. 147, no. 1, pp. 186–197, 2008.

[6] O. Ledoit and M. Wolf, "A well-conditioned estimator for large-dimensional covariance matrices," *Journal of Multivariate Analysis*, vol. 88, pp. 365–411, 2004.

[7] P. Stoica, J. Li, X. Zhu, and J. Guerci, "On using a priori knowledge in space-time adaptive processing," *IEEE Transactions on Signal Processing*, vol. 56, no. 6, pp. 2598–2602, 2008.

[8] Y. Chen, A. Wiesel, Y. Eldar, and A. Hero, "Shrinkage algorithms for MMSE covariance estimation," *IEEE Transactions on Signal Processing*, vol. 58, no. 10, pp. 5016 –5029, 2010.

[9] N. El Karoui, "High-dimensionality effects in the markowitz problem and other quadratic programs with linear constraints: Risk underestimation," Tech. Rep. 6, 2010.

[10] M. Zhang, F. Rubio, and D. P. Palomar, "Estimation of high-dimensional precision matrices," Hong Kong University of Science and Technology, Tech. Rep., 2011.

[11] F. Rubio and X. Mestre, "Spectral convergence for a general class of random matrices," *Statistics and Probability Letters*, vol. 8, pp. 592–602, 2011.

[12] F. Rubio, X. Mestre, and D. Palomar, "Performance analysis and optimal selection of large mean-variance portfolios under estimation risk," *Arxiv preprint arXiv:1110.3460*, 2011.