Covariance Matrix Estimation Under Low-Rank Factor Model With Nonnegative Correlations

Rui Zhou, Jiaxi Ying, and Daniel P. Palomar, *Fellow, IEEE*

***Abstract*—Inferring the covariance matrix of multivariate data is of great interest in statistics, finance, and data science. It is often carried out via the maximum likelihood estimation (MLE) principle, which seeks a covariance matrix estimator maximizing the observed data likelihood. However, such estimator is usually poor when number of samples is not sufficiently larger than the number of variables. With the assumption that a covariance matrix can be decomposed into a low-rank matrix and a diagonal matrix, factor analysis (FA) model is a popular dimensionality reduction technique in improving the estimation performance. Recently, more and more evidence shows that the covariance matrix of real-valued data may admit the nonnegative correlation structure, which has attracted a lot of interest in some areas like finance and psychometrics. There does not exist any work estimating the covariance matrix simultaneously satisfying both structures. In this paper, we propose an MLE problem formulation for covariance matrix considering jointly the low-rank FA model and nonnegative correlation structures. Since the proposed problem formulation is an intractable non-convex problem, a block coordinate descent algorithm is further proposed to solve a relaxed version of our proposed formulation. The superior performance of our proposed formulation and the algorithm are verified through numerical simulations on both synthetic data and real market data.**

***Index Terms*—Low-rank FA, nonnegative correlations, covariance matrix, block coordinate descent.**

I. Introduction

ESTIMATING the covariance matrix of the multivariate data is of common interest in various fields such as cognitive radio [1], radar detection [2], statistics [3], and finance [4]. Collecting random vector $\boldsymbol{x}_i \in \mathbb{R}^p, i = 1, \ldots, n$ as the i-th sample of p variables and assuming its mean to be zero (i.e., $\mathbb{E}[\boldsymbol{x}] = \boldsymbol{0}$), the commonly used covariance matrix estimator is the sample covariance matrix (SCM), $\boldsymbol{S} = \frac{1}{n}\sum_{i=1}^{n} \boldsymbol{x}_i \boldsymbol{x}_i^T$. Under the assumption that the data samples are independent and identically distributed (i.i.d.) under a multivariate normal distribution, the sample covariance matrix is the optimal solution to its maximum likelihood estimation (MLE) problem, i.e.,

$$\underset{\boldsymbol{\Sigma} \succ \boldsymbol{0}}{\text{minimize}} \quad \log\det\left(\boldsymbol{\Sigma}\right) + \text{tr}\left(\boldsymbol{\Sigma}^{-1}\boldsymbol{S}\right), \tag{1}$$

where $\boldsymbol{\Sigma}$ denotes the covariance matrix to be estimated. However, the sample covariance matrix usually suffers a large estimation error especially when n (number of samples) is only comparable with p [4]. For example, there may exist hundreds of assets in a financial market while only 252 historical daily prices can be recorded per year. Then the sample covariance matrix is regarded as an unreliable estimator and is rarely adopted in real-world applications. One approach to improve the covariance matrix estimation performance is incorporating prior information into the estimation procedure, typically introduced by imposing extra constraints on the parameter or via a regularization term.

Among numerous choices, the low-rank factor analysis (FA) model is one of the most popular structures [5], where the observed data are assumed to be linearly driven by a limited number of common factors (usually significantly less than the number of variables) [6], [7], i.e., $\boldsymbol{x} = \boldsymbol{\mu} + \boldsymbol{B}\boldsymbol{f} + \boldsymbol{\varepsilon}$, where $\boldsymbol{\mu} \in \mathbb{R}^p$ is a constant vector, $\boldsymbol{B} \in \mathbb{R}^{p\times r}$ ($r \ll p$) is the factor loading matrix, $\boldsymbol{f} \in \mathbb{R}^r$ is a vector of low-dimensional common factors, and $\boldsymbol{\varepsilon} \in \mathbb{R}^p$ denotes uncorrelated noise. For example, in a financial market, $\boldsymbol{x}$ can be the return of stocks, and $\boldsymbol{f}$ can be macroeconomic factors like growth rate of the GDP, inflation rate, unemployment rate, etc. The classical FA model assumes that both $\boldsymbol{f}$ and $\boldsymbol{\varepsilon}$ are uncorrelated and zero-mean, and the covariance matrix of $\boldsymbol{f}$ is the identity matrix. Then the covariance matrix of $\boldsymbol{x}$ can be written as $\boldsymbol{\Sigma} = \boldsymbol{B}\boldsymbol{B}^T + \boldsymbol{\Psi}$, where $\boldsymbol{\Psi}$ is a diagonal matrix containing the variance of noise on its diagonal.

The FA model implies that the covariance matrix consists of a positive semidefinite low-rank matrix plus a positive definite diagonal matrix. Note that, with the low-rank structure, the number of parameters of the covariance matrix is greatly reduced from $p(p+1)/2$ to $p(r+1)$. A large amount of literature has focused on estimating the covariance matrix with this low-rank structure. One popular choice is decomposing or approximating the observed covariance matrix, typically SCM, with a sum of low-rank matrix and diagonal matrix [8]. It can be regarded as a two-stage method, which first estimates the covariance matrix without extra structure and then performs the decomposition. Another choice is through directly introducing the low-rank FA structure into the corresponding MLE problem [9]–[12], which results in a challenging non-convex optimization problem.

Manuscript received 20 July 2021; revised 12 January 2022, 18 April 2022, and 16 June 2022; accepted 17 July 2022. Date of publication 22 July 2022; date of current version 16 August 2022. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Evrim Acar. This work was supported by the Hong Kong GRF 16207820 research grant. *(Corresponding author: Jiaxi Ying.)*

Rui Zhou is with the Shenzhen Research Institute of Big Data, Shenzhen, China (e-mail: rui.zhou@sribd.cn).

Jiaxi Ying is with the Department of Mathematics, Hong Kong University of Science and Technology, Clear Water Bay, Kowloon, Hong Kong (e-mail: jx.ying@connect.ust.hk).

Daniel P. Palomar is with the Department of Electronic and Computer Engineering and Department of Industrial Engineering and Decision Analytics, Hong Kong University of Science and Technology, Kowloon, Hong Kong (e-mail: palomar@ust.hk).

Digital Object Identifier 10.1109/TSP.2022.3193232

A fast and scalable algorithm has recently been proposed to solve this problem [12] and later extended to heavy tails [13].

Apart from the FA model, we are also interested in the nonnegative correlation structure of real-valued data, which means the variables are positively correlated, i.e., $\mathbf{\Sigma} \geq \mathbf{0}$ (with $\geq$ being elementwise). This structure is related to another special structure called multivariate totally positive of order 2 (MTP_2), which becomes of great interest in statistics in recent years [3], [14]. Under the MTP_2 structure, the precision matrix $\mathbf{\Theta}$ (i.e., $\mathbf{\Theta} = \mathbf{\Sigma}^{-1}$) of the data is assumed to satisfy $\Theta_{ij} \leq 0$ for all $i \neq j$, which implies $\mathbf{\Sigma} \geq \mathbf{0}$ but not vice versa. Interestingly, the MTP_2 structure is actually implied if data follow both $\mathbf{\Sigma} \geq \mathbf{0}$ and the single factor model [15]. We propose the nonnegative correlation structure based on strong evidence in some real-world applications such as financial engineering. The empirical evidence will be provided and more details will be discussed in Section II.

In this paper, we consider the estimation of the covariance matrix with a simultaneously low-rank FA and nonnegative correlation structure. To this end, we propose a Gaussian MLE problem formulation including the low-rank FA and nonnegative correlation constraints. The resulting problem formulation becomes too difficult to solve since it is a rank constrained nonconvex optimization problem. Therefore, we resort to replacing the intractable problem with a relaxed formulation, and propose an efficient algorithm based on block coordinate descent (BCD) to solve the relaxed problem. Extensive numerical experiments are performed to corroborate our claims.

This paper is organized as follows. We first provide empirical evidence of our considered low-rank FA and nonnegative correlation structure for financial data in Section II. In Section III, we pose the original problem formulation. In Section IV, we first show how to turn the original problem into a relaxed problem formulation and then present our proposed algorithm. The complexity analysis is also included. In Section V, we discuss the feasibility of our proposed algorithm on some extensions of Gaussian MLE problem formulation. Numerical experiments are given in Section VI. Finally, conclusions of this paper are summarized in Section VII.

II. Empirical Evidence of Nonnegative Correlations

In financial markets, it is usually assumed that assets' returns follow common trends, which implies that these variables should be positively correlated, i.e., the entries in the covariance matrix should be nonnegative. In this section, we provide empirical evidence for the considered nonnegative correlation structure from two perspectives, i.e., significant influence of the market and empirical correlations. The proposed approaches of identifying such nonnegative correlations can be employed on data from other fields like psychometrics [3].

A. Significant Influence of Market

Factor model is a widely used tool to model high-dimensional data with respect to low-dimensional variables, which are also known as factors [5], [6], [16]. A well-known example in finance is the capital asset pricing model (CAPM) for modeling stocks' return:

$$r_i = r_f + \beta_i (r_m - r_f) + u_i, \tag{2}$$

where r_i is the return of i-th stock, r_f is the risk-free return, β_i is its exposure to the market, r_m is the market return, and u_i is the noise term. Since β_i is typically positive for most stocks, the correlations between these stocks are usually positive [15]. We notice that the CAPM assumes a single factor, while a factor model with several factors could be more realistic. The dominating impact of the market factor can still be observed from the empirical data. We conduct an empirical check on stocks' daily returns and cryptocurrencies' hourly returns.

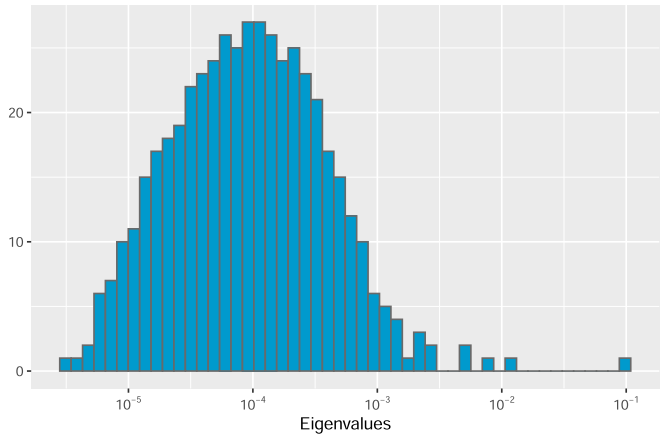

Fig. 1. Histogram of empirical eigenvalues obtained from S&P 500 stocks' data.

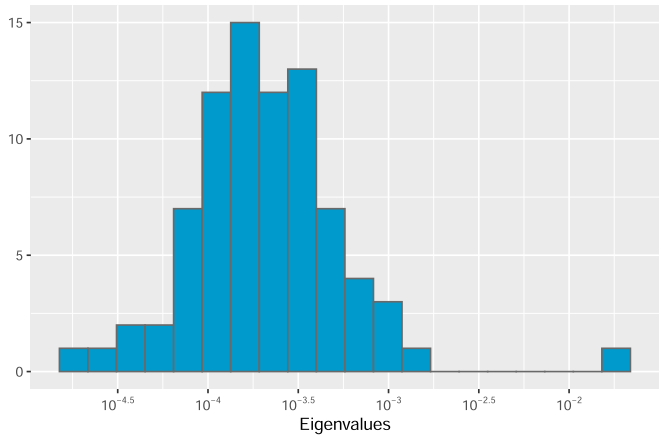

Fig. 2. Histogram of empirical eigenvalues obtained from 82 cryptocurrencies' data.

In Fig. 1, we show the histogram of empirical eigenvalues of the sample covariance matrix from S&P 500 stocks' daily returns during a period of 1000 trading days. There exist 5 eigenvalues deviating from the bulk, indicating the significant factors in the market. We also notice that the largest eigenvalue is impressively much larger than others, while the elements of the corresponding eigenvector (principal component) share the same sign, implying that positive correlation can describe the data well.

In Fig. 2, we show the histogram of empirical eigenvalues of the sample covariance matrix from 82 cryptocurrencies' hourly

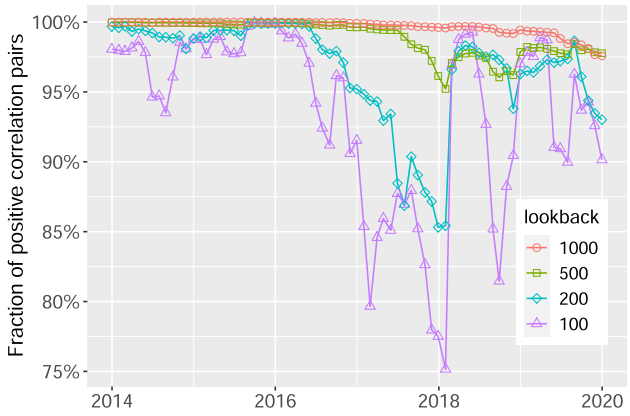

Fig. 3. Positive rate of stocks' daily return versus lookback window length in days.

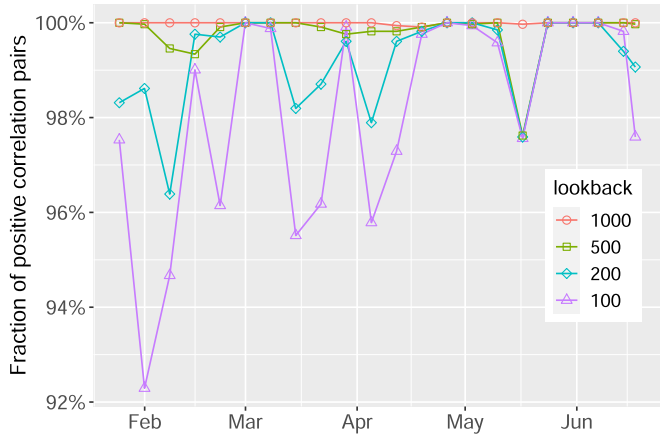

Fig. 4. Positive rate (in year 2021) of cryptocurrencies' hourly return versus lookback window length in hours.

data during a period of 5 months in year 2021. There exists an eigenvalue deviating from the bulk, showing a significant influence of the market. Elements of corresponding eigenvector also share the same sign.

B. Empirical Correlation

Apart from validating the nonnegative correlation from the perspective of the factor model, a more intuitive way is checking directly the empirical correlation of real financial data. Similarly, we perform the check using data from both stocks and cryptocurrencies.

For stocks' daily returns, the check is conducted at the end date of every month. We calculate the sample correlation of every pair of stocks using the historical price data within a certain lookback window length. Fig. 3 shows the positive rate of these stocks' correlations (fraction of positive correlation pairs) from year 2014 to 2020 for different lengths of lookback window in days. Even when the lookback window length is small, the positive rate is always larger than 75%. Some of these negative correlations may be explained as noisy estimation, competitive companies, or stocks' different reactions to external stimuli. For example, when interest rates increase, financial stocks such as insurance companies tend to get a boost [17] but the real estate companies will get hit [18]. But these external stimuli are usually temporary thus their influences on market are insignificant in the long term.

For cryptocurrencies' hourly returns, we perform a similar empirical check at the end time of every week. Fig. 4 shows the positive rate of these cryptocurrencies' correlations. It reveals that the nonnegative correlation structure is even stronger for cryptocurrency data. To sum up, it is reasonable to assume a nonnegative correlation structure in financial data.

III. Problem Formulation

In this section, we first present several existing models for covariance matrix estimation, then introduce our proposed formulation with the low-rank FA and nonnegative correlation constrains.

Suppose the data follow a multivariate Gaussian distribution under the factor model, the corresponding MLE problem for estimating the covariance matrix can be formulated as

$$\begin{aligned} \underset{\boldsymbol{\Sigma}}{\text{minimize}} \quad & \log\det\left(\boldsymbol{\Sigma}\right) + \operatorname{tr}\left(\boldsymbol{\Sigma}^{-1}\boldsymbol{S}\right) \\ \text{subject to} \quad & \boldsymbol{\Sigma} \in \mathcal{S}_{FA}, \end{aligned} \tag{3}$$

where $\boldsymbol{S}$ is the sample covariance matrix, and $\mathcal{S}_{FA}$ is the feasible set imposing the low-rank FA structure,

$$\begin{aligned} \mathcal{S}_{FA} = \{&\boldsymbol{\Sigma} = \boldsymbol{M} + \boldsymbol{\Psi},\ \operatorname{rank}\left(\boldsymbol{M}\right) \le r,\ \boldsymbol{M} \succeq \boldsymbol{0}, \\ &\boldsymbol{\Psi} = \operatorname{Diag}\left(\psi_1, \ldots, \psi_p\right) \succeq \epsilon \boldsymbol{I}\}. \end{aligned} \tag{4}$$

Problem (3) has been well studied in [12], where an efficient algorithm has been proposed. It first concentrates the problem (3) in $\boldsymbol{\Psi}$ (by finding the optimal $\boldsymbol{M}$ for fixed $\boldsymbol{\Psi}$) and then solves it via the majorization-minimization (MM) algorithm [19]. A more efficient algorithm was proposed in [20] to solve a similar MLE problem but under a more restricted low-rank FA structure.

If the data follow a multivariate Gaussian distribution under merely the nonnegative correlation structure, the corresponding MLE problem may be technically formulated as

$$\begin{aligned} \underset{\boldsymbol{\Sigma}}{\text{minimize}} \quad & \log\det\left(\boldsymbol{\Sigma}\right) + \operatorname{tr}\left(\boldsymbol{\Sigma}^{-1}\boldsymbol{S}\right) \\ \text{subject to} \quad & \boldsymbol{\Sigma} \succ \boldsymbol{0},\ \boldsymbol{\Sigma} \ge \boldsymbol{0}. \end{aligned} \tag{5}$$

One simple approach to incorporate the nonnegative correlation structure is via projection, i.e., approximating the sample covariance matrix with a satisfactory one [8]:

$$\begin{aligned} \underset{\boldsymbol{\Sigma}}{\text{minimize}} \quad & \|\boldsymbol{\Sigma} - \boldsymbol{S}\|_F^2 \\ \text{subject to} \quad & \boldsymbol{\Sigma} \succ \boldsymbol{0},\ \boldsymbol{\Sigma} \ge \boldsymbol{0}, \end{aligned} \tag{6}$$

which can be easily solved by alternating projecting $\boldsymbol{S}$ onto the convex sets $\{\boldsymbol{\Sigma}|\boldsymbol{\Sigma} \succeq \boldsymbol{0}\}$ and $\{\boldsymbol{\Sigma}|\boldsymbol{\Sigma} \ge \boldsymbol{0}\}$ [21].

As mentioned in Section I, we are interested in estimating the covariance matrix under a low-rank FA structure and nonnegative correlation (FANC) structure, and propose the formulation

as follows:

$$\begin{aligned} \underset{\boldsymbol{M},\boldsymbol{\Psi}}{\text{minimize}} \quad & \log\det\left(\boldsymbol{M}+\boldsymbol{\Psi}\right)+\text{tr}\left(\left(\boldsymbol{M}+\boldsymbol{\Psi}\right)^{-1}\boldsymbol{S}\right) \\ \text{subject to} \quad & \text{rank}\left(\boldsymbol{M}\right)\leq r,\ \boldsymbol{M}\succeq\boldsymbol{0},\ \boldsymbol{M}\geq\boldsymbol{0}, \\ & \boldsymbol{\Psi}=\text{Diag}\left(\psi_1,\ldots,\psi_p\right)\succeq\epsilon\boldsymbol{I}. \end{aligned} \tag{7}$$

Due to the nonnegative correlation constraint $\boldsymbol{M}\geq\boldsymbol{0}$, the proposed algorithm in [12] is not applicable any more. Therefore, it is necessary to develop a new framework to solve problem (7).

It should be noted that directly solving problem (7) is rather difficult due to its non-convex objective function and rank constraint. We resort to a relaxed formulation and propose an efficient algorithm based on block coordinate descent to solve this problem in next section.

IV. Solving the FANC Problem

In this section, we discuss how to solve the problem (7). Since it is challenging to directly solve it, we propose a relaxed formulation. A block coordinate descent based algorithm is further proposed to solve the relaxed problem. We will show that each sub-problem either admits a closed-from solution or can be easy to solve via our proposed projected gradient descent methods.

A. Reformulation

We define a new variable $\tilde{\boldsymbol{M}}$ as $\tilde{\boldsymbol{M}}=\boldsymbol{\Psi}^{-\frac{1}{2}}\boldsymbol{M}\boldsymbol{\Psi}^{-\frac{1}{2}}$ (similarly to [1, Algorithm 1]). Then we can rewrite the problem (7) as the problem (8). The constraints on $\tilde{\boldsymbol{M}}$ remain the same as those on $\boldsymbol{M}$ due to the constraint $\boldsymbol{\Psi}=\text{Diag}(\psi_1,\ldots,\psi_p)\succeq\epsilon\boldsymbol{I}$.

Then, in problem (8) shown at the bottom of this page, we replace the low-rank constraint $\text{rank}(\tilde{\boldsymbol{M}})\leq r$ and positive semidefinite constraint $\tilde{\boldsymbol{M}}\succeq\boldsymbol{0}$ with $\tilde{\boldsymbol{M}}=\boldsymbol{U}\boldsymbol{\Lambda}\boldsymbol{U}^T$, by introducing the new dummy variables $\boldsymbol{U}\in\mathbb{R}^{p\times r}$ satisfying $\boldsymbol{U}^T\boldsymbol{U}=\boldsymbol{I}_r$ and $\boldsymbol{\Lambda}=\text{Diag}(\lambda_1,\ldots,\lambda_r)$ with $\boldsymbol{\lambda}=[\lambda_1,\ldots,\lambda_r]^T\geq\boldsymbol{0}$. Thus the problem (8) is rewritten into problem (9).

B. Relaxation

To solve problem (9) shown at the bottom of this page, we relax the hard constraint $\tilde{\boldsymbol{M}}=\boldsymbol{U}\boldsymbol{\Lambda}\boldsymbol{U}^T$ into the regularization term $\frac{\rho}{2}\|\tilde{\boldsymbol{M}}-\boldsymbol{U}\boldsymbol{\Lambda}\boldsymbol{U}^T\|_F^2$ in the objective, obtaining the formulation (10) shown at the bottom of this page. Note that such relaxation can be made tight by choosing sufficiently large or iteratively increasing ρ. Such technique has been successfully adopted in various applications, cf. [22]–[24].

In what follows, we consider solving the problem (10) to obtain the covariance matrix estimate. Although it is still a non-convex problem, we can develop an efficient algorithm based on the block coordinate descent method [25]. The details of our proposed algorithm are provided in the next subsection.

C. Proposed Algorithm

Collecting the variables in four blocks as $(\tilde{\boldsymbol{M}},\boldsymbol{\Psi},\boldsymbol{U},\boldsymbol{\lambda})$, we solve the problem (10) by cyclically updating each block while keeping the other blocks fixed. At iteration k, the update of each block is discussed in the following.

1) Update of $\tilde{\boldsymbol{M}}$: Fixing the other three blocks in problem (10), the sub-problem with respect to $\tilde{\boldsymbol{M}}$ can be written as

$$\begin{aligned} \underset{\tilde{\boldsymbol{M}}}{\text{minimize}} \quad & \text{tr}\left(\left(\tilde{\boldsymbol{M}}+\boldsymbol{I}\right)^{-1}\tilde{\boldsymbol{S}}\right)+\frac{\rho}{2}\|\tilde{\boldsymbol{M}}-\boldsymbol{U}\boldsymbol{\Lambda}\boldsymbol{U}^T\|_F^2 \\ \text{subject to} \quad & \tilde{\boldsymbol{M}}\geq\boldsymbol{0},\ \tilde{\boldsymbol{M}}+\boldsymbol{I}\succ\boldsymbol{0}, \end{aligned} \tag{11}$$

where $\tilde{\boldsymbol{S}}=\boldsymbol{\Psi}^{-\frac{1}{2}}\boldsymbol{S}\boldsymbol{\Psi}^{-\frac{1}{2}}$. Note that the additional constraint $\tilde{\boldsymbol{M}}+\boldsymbol{I}\succ\boldsymbol{0}$ is added into sub-problem (11) to guarantee the covariance matrix, i.e., $(\boldsymbol{\Psi})^{\frac{1}{2}}(\tilde{\boldsymbol{M}}+\boldsymbol{I})(\boldsymbol{\Psi})^{\frac{1}{2}}$, to be positive definite. Throughout the paper, we assume the number of variables is not larger than the number of samples, and thus both $\boldsymbol{S}$ and $\tilde{\boldsymbol{S}}$ are positive definite. Obviously it is a convex problem and thus can be solved by several toolboxes, e.g., CVX [26], [27]. For the sake of computational reasons, we propose a practical algorithm to solve sub-problem (11). The feasible set of sub-problem (11) is the intersection of $\tilde{\boldsymbol{M}}\geq\boldsymbol{0}$ and $\tilde{\boldsymbol{M}}+\boldsymbol{I}\succ\boldsymbol{0}$. The set $\tilde{\boldsymbol{M}}\geq\boldsymbol{0}$ is closed and can be handled efficiently by projection, while the set $\tilde{\boldsymbol{M}}+\boldsymbol{I}\succ\boldsymbol{0}$ is not closed, and we guarantee this constraint through a line search procedure, which is inspired

$$\begin{aligned} \underset{\tilde{\boldsymbol{M}},\boldsymbol{\Psi}}{\text{minimize}} \quad & \log\det\left(\boldsymbol{\Psi}\right)+\log\det\left(\tilde{\boldsymbol{M}}+\boldsymbol{I}\right)+\text{tr}\left(\boldsymbol{\Psi}^{-\frac{1}{2}}\left(\tilde{\boldsymbol{M}}+\boldsymbol{I}\right)^{-1}\boldsymbol{\Psi}^{-\frac{1}{2}}\boldsymbol{S}\right) \\ \text{subject to} \quad & \text{rank}\left(\tilde{\boldsymbol{M}}\right)\leq r,\ \tilde{\boldsymbol{M}}\succeq\boldsymbol{0},\ \tilde{\boldsymbol{M}}\geq\boldsymbol{0},\boldsymbol{\Psi}=\text{Diag}\left(\psi_1,\ldots,\psi_p\right)\succeq\epsilon\boldsymbol{I} \end{aligned} \tag{8}$$

$$\begin{aligned} \underset{\tilde{\boldsymbol{M}},\boldsymbol{\Psi},\boldsymbol{U},\boldsymbol{\lambda}}{\text{minimize}} \quad & \log\det\left(\boldsymbol{\Psi}\right)+\sum_{i=1}^{r}\log\left(\lambda_i+1\right)+\text{tr}\left(\boldsymbol{\Psi}^{-\frac{1}{2}}\left(\tilde{\boldsymbol{M}}+\boldsymbol{I}\right)^{-1}\boldsymbol{\Psi}^{-\frac{1}{2}}\boldsymbol{S}\right) \\ \text{subject to} \quad & \tilde{\boldsymbol{M}}=\boldsymbol{U}\boldsymbol{\Lambda}\boldsymbol{U}^T,\ \tilde{\boldsymbol{M}}\geq\boldsymbol{0},\ \boldsymbol{U}^T\boldsymbol{U}=\boldsymbol{I}_r,\ \boldsymbol{\lambda}\geq\boldsymbol{0},\ \boldsymbol{\Psi}=\text{Diag}\left(\psi_1,\ldots,\psi_p\right)\succeq\epsilon\boldsymbol{I} \end{aligned} \tag{9}$$

$$\begin{aligned} \underset{\tilde{\boldsymbol{M}},\boldsymbol{\Psi},\boldsymbol{U},\boldsymbol{\lambda}}{\text{minimize}} \quad & \log\det\left(\boldsymbol{\Psi}\right)+\sum_{i=1}^{r}\log\left(\lambda_i+1\right)+\text{tr}\left(\boldsymbol{\Psi}^{-\frac{1}{2}}\left(\tilde{\boldsymbol{M}}+\boldsymbol{I}\right)^{-1}\boldsymbol{\Psi}^{-\frac{1}{2}}\boldsymbol{S}\right)+\frac{\rho}{2}\|\tilde{\boldsymbol{M}}-\boldsymbol{U}\boldsymbol{\Lambda}\boldsymbol{U}^T\|_F^2 \\ \text{subject to} \quad & \tilde{\boldsymbol{M}}\geq\boldsymbol{0},\ \boldsymbol{U}^T\boldsymbol{U}=\boldsymbol{I}_r,\ \boldsymbol{\lambda}\geq\boldsymbol{0},\ \boldsymbol{\Psi}=\text{Diag}\left(\psi_1,\ldots,\psi_p\right)\succeq\epsilon\boldsymbol{I} \end{aligned} \tag{10}$$

Algorithm 1: Practical Algorithm for Solving Sub-Problem (11).

1: Initialize $\tilde{\boldsymbol{M}}_0^k \in \mathcal{S}_{AB}$ and choose $\eta > 0, 0 < c < 1$.
2: **for** $t = 0, 1, 2, \ldots$ **do**
3: $\quad \tilde{\boldsymbol{M}}_{t+1}^k = [\tilde{\boldsymbol{M}}_t^k - \eta \nabla f(\tilde{\boldsymbol{M}}_t^k)]_+$;
4: $\quad$ **if** $f(\tilde{\boldsymbol{M}}_{t+1}^k) > f(\tilde{\boldsymbol{M}}_t^k) + \langle \nabla f(\tilde{\boldsymbol{M}}_t^k),\ \tilde{\boldsymbol{M}}_{t+1}^k - \tilde{\boldsymbol{M}}_t^k \rangle + \frac{1}{2\eta}\|\tilde{\boldsymbol{M}}_{t+1}^k - \tilde{\boldsymbol{M}}_t^k\|_F^2$ or $\tilde{\boldsymbol{M}}_{t+1}^k \preceq -\boldsymbol{I}$ **then**
5: $\quad\quad \eta = c\eta$;
6: $\quad\quad$ Back to step 3;
7: $\quad$ **end if**
8: $\quad$ Terminate if converges or exceeds iteration limit;
9: **end for**

by [28, Algorithm 1] and [29, Algorithm 1]. The details are given in Algorithm 1, where $\nabla f(\tilde{\boldsymbol{M}}_t^k) = \rho(\tilde{\boldsymbol{M}}_t^k - \boldsymbol{U}\boldsymbol{\Lambda}\boldsymbol{U}^T) - (\tilde{\boldsymbol{M}}_t^k + \boldsymbol{I})^{-1}\tilde{\boldsymbol{S}}(\tilde{\boldsymbol{M}}_t^k + \boldsymbol{I})^{-1}$ is the gradient of the objective of the sub-problem (11) at $\tilde{\boldsymbol{M}}_t^k$ and $[\boldsymbol{x}]_+ = \max(\boldsymbol{x}, \boldsymbol{0})$.

2) Update of $\boldsymbol{\Psi}$: Fixing the other three blocks in problem (10), the sub-problem with respect to $\boldsymbol{\Psi}$ can be written as

$$\begin{aligned} \underset{\boldsymbol{\Psi}}{\text{minimize}} \quad & \log\det(\boldsymbol{\Psi}) + \text{tr}\left(\boldsymbol{\Psi}^{-\frac{1}{2}}(\tilde{\boldsymbol{M}} + \boldsymbol{I})^{-1}\boldsymbol{\Psi}^{-\frac{1}{2}}\boldsymbol{S}\right) \\ \text{subject to} \quad & \boldsymbol{\Psi} = \text{Diag}(\psi_1, \ldots, \psi_p) \succeq \epsilon \boldsymbol{I}, \end{aligned} \tag{12}$$

which appears to be a non-convex problem. But it can be converted into a convex problem by concentrating into the newly defined variable $\boldsymbol{\alpha} = \text{diag}(\boldsymbol{\Psi}^{-\frac{1}{2}})$, i.e.,

$$\begin{aligned} \underset{\boldsymbol{\alpha}}{\text{minimize}} \quad & \boldsymbol{\alpha}^T\boldsymbol{\Gamma}\boldsymbol{\alpha} - \sum_{i=1}^{p}\log\alpha_i^2 \\ \text{subject to} \quad & \boldsymbol{0} \le \boldsymbol{\alpha} \le \epsilon^{-\frac{1}{2}}\boldsymbol{1}, \end{aligned} \tag{13}$$

where $\boldsymbol{\Gamma} = \boldsymbol{S} \odot (\tilde{\boldsymbol{M}} + \boldsymbol{I})^{-1}$ is guaranteed to be positive definite thanks to $\tilde{\boldsymbol{M}} + \boldsymbol{I} \succ \boldsymbol{0}$ and Schur product theorem [30]. This problem can be solved by state-of-the-art convex optimization tools, e.g., CVX [27]. Given concerns on the computational cost, we propose to solve it using the PGD algorithm. The details are given in Algorithm 2, where $\nabla f(\boldsymbol{\alpha}_t^k) = 2\boldsymbol{\Gamma}\boldsymbol{\alpha}_t^k - 2/\boldsymbol{\alpha}_t^k$ is the gradient of the objective of the sub-problem (13) at $\boldsymbol{\alpha}^k$ and $[\boldsymbol{x}]^u = \min(\boldsymbol{x}, \boldsymbol{1}u)$. After obtaining the optimal solution $\boldsymbol{\alpha}^\star$ via Algorithm 2, we can update $\boldsymbol{\Psi}^{k+1} = (\boldsymbol{\alpha}^\star)^{-2}$.

3) Update of $\boldsymbol{\lambda}$: Fixing the other three blocks in problem (10), the sub-problem with respect to $\boldsymbol{\lambda}$ is

$$\underset{\boldsymbol{\lambda}\ge\boldsymbol{0}}{\text{minimize}} \quad \sum_{i=1}^{r}\log(\lambda_i + 1) + \frac{\rho}{2}\|\tilde{\boldsymbol{M}} - \boldsymbol{U}\boldsymbol{\Lambda}\boldsymbol{U}^T\|_F^2 \tag{14}$$

Note that, given $\boldsymbol{U}^T\boldsymbol{U} = \boldsymbol{I}_r$, it is easy to check that $\|\tilde{\boldsymbol{M}} - \boldsymbol{U}\boldsymbol{\Lambda}\boldsymbol{U}^T\|_F^2 = \|\boldsymbol{U}^T\tilde{\boldsymbol{M}}\boldsymbol{U} - \boldsymbol{\Lambda}\|_F^2 + \|\tilde{\boldsymbol{M}}\|_F^2 - \|\boldsymbol{U}^T\tilde{\boldsymbol{M}}\boldsymbol{U}\|_F^2$. Therefore, the sub-problem (14) can be rewritten as

$$\underset{\boldsymbol{\lambda}\ge\boldsymbol{0}}{\text{minimize}} \quad \sum_{i=1}^{r}\log(\lambda_i + 1) + \frac{\rho}{2}\|\boldsymbol{U}^T\tilde{\boldsymbol{M}}\boldsymbol{U} - \boldsymbol{\Lambda}\|_F^2 \quad , \tag{15}$$

Algorithm 2: PGD Algorithm for Solving Sub-Problem (13).

1: Initialize $\boldsymbol{\alpha}_0^k$ and choose $\eta > 0, 0 < c < 1$.
2: **for** $t = 0, 1, 2, \ldots$ **do**
3: $\quad \boldsymbol{\alpha}_{t+1}^k = [\boldsymbol{\alpha}_t^k - \eta \nabla f(\boldsymbol{\alpha}_t^k)]^{\epsilon^{-\frac{1}{2}}}$;
4: $\quad$ **if** $f(\boldsymbol{\alpha}_{t+1}^k) > f(\boldsymbol{\alpha}_t^k)$ or $\exists i, \alpha_{t+1,i}^k \le 0$ **then**
5: $\quad\quad \eta = c\eta$;
6: $\quad\quad$ Back to step 3;
7: $\quad$ **end if**
8: $\quad$ Terminate if converges;
9: **end for**

where the elements of $\boldsymbol{\lambda}$ are decoupled with each other. Then, the sub-problem (15) can be further simplified as r sub-problems with each of them being

$$\underset{\lambda_i\ge 0}{\text{minimize}} \quad \log(\lambda_i + 1) + \frac{\rho}{2}(b_i - \lambda_i)^2, \tag{16}$$

where $b_i = [\boldsymbol{U}^T\tilde{\boldsymbol{M}}\boldsymbol{U}]_{ii} = \boldsymbol{U}_i^T\tilde{\boldsymbol{M}}\boldsymbol{U}_i$ with $\boldsymbol{U}_i$ being the i-th column of $\boldsymbol{U}$. We present the closed-form solution of the sub-problem (16) in the following lemma.

Lemma 1: The optimal solution of the sub-problem (16) is

$$\lambda_i^\star = \begin{cases} \underset{x\in\{0,\delta_1\}}{\text{argmin}}\ f(x) & \text{if } \triangle > 0, \delta_1 > 0, \delta_2 > 0, \\ \delta_1 & \text{if } \triangle > 0, \delta_1 > 0, \delta_2 \le 0, \\ 0 & \text{otherwise,} \end{cases} \tag{17}$$

where f is the objective function in (16), $\triangle = (b_i + 1)^2 - \frac{4}{\rho}$, $\delta_1 = \frac{b_i - 1 + \sqrt{\triangle}}{2}$, and $\delta_2 = \frac{b_i - 1 - \sqrt{\triangle}}{2}$.

Proof: See Appendix A. ■

4) Update of $\boldsymbol{U}$: Fixing the other three blocks in problem (10), the sub-problem on $\boldsymbol{U}$ is

$$\begin{aligned} \underset{\boldsymbol{U}}{\text{minimize}} \quad & \frac{\rho}{2}\|\tilde{\boldsymbol{M}} - \boldsymbol{U}\boldsymbol{\Lambda}\boldsymbol{U}^T\|_F^2 \\ \text{subject to} \quad & \boldsymbol{U}^T\boldsymbol{U} = \boldsymbol{I}_r, \end{aligned} \tag{18}$$

which can be rewritten as

$$\begin{aligned} \underset{\boldsymbol{U}}{\text{maximize}} \quad & \text{tr}\left(\boldsymbol{U}^{\text{T}}\tilde{\boldsymbol{M}}\boldsymbol{U}\boldsymbol{\Lambda}\right) \\ \text{subject to} \quad & \boldsymbol{U}^T\boldsymbol{U} = \boldsymbol{I}_r \end{aligned} \tag{19}$$

Problem (19) is an optimization problem on the orthogonal Stiefel manifold. Following from [31], [32], the closed-from solution of (19) is given in Lemma 2.

Lemma 2: The optimal solution of problem (19) is r principle eigenvectors of $\tilde{\boldsymbol{M}}$ with the corresponding eigenvalues following the same order of $\boldsymbol{\Lambda}$.

The overall FANC algorithm is summarized in Algorithm 3.

D. Complexity Analysis

In this section, we present a detailed discussion on the computational complexity of our proposed FANC algorithm. Since it is a BCD based algorithm with four blocks, we analyze the computational complexity of updating each block individually.

Algorithm 3: FANC Algorithm for Solving Problem (7).

1: Initialize $\tilde{\boldsymbol{M}}^0$, $\boldsymbol{\Psi}^0$, $\boldsymbol{U}^0$, $\boldsymbol{\lambda}^0$.
2: **for** $k = 0, 1, 2, \ldots$ **do**
3: Update $\tilde{\boldsymbol{M}}^{k+1}$ via Algorithm 1;
4: Update $\boldsymbol{\Psi}^{k+1}$ via Algorithm 2;
5: Update $\boldsymbol{U}^{k+1}$ as in Lemma 1;
6: Update $\boldsymbol{\lambda}^{k+1}$ as in Lemma 2;
7: Terminate if converges or exceeds iteration limit;
8: **end for**
9: Return $\boldsymbol{\Sigma} = (\boldsymbol{\Psi}^{k+1})^{\frac{1}{2}}(\tilde{\boldsymbol{M}}^{k+1} + \boldsymbol{I})(\boldsymbol{\Psi}^{k+1})^{\frac{1}{2}}$.

- For the update of $\tilde{\boldsymbol{M}}$, i.e., Algorithm 1, the computational cost per iteration is mainly on the gradient computation. Its computational cost is dominated by the matrix inversion of a $p \times p$ matrix and the matrix multiplication of two $p \times p$ matrices, whose computational complexities can be $O(p^{2.373})$ [33].
- For the update of $\boldsymbol{\Psi}$, i.e., Algorithm 2, the computational per iteration is still mainly on the gradient computation as $O(p^2)$.
- For the update of $\boldsymbol{\lambda}$, the computational complexity is $O(rp^2)$ on computing b_i, $i = 1, \ldots, r$ and $O(r)$ on solving problem (15), where r ($r \ll p$) is the number of factors. Then the total cost is $O(rp^2)$.
- For the update of $\boldsymbol{U}$, the total computational complexity is $O(rp^2)$ due to the eigen-decomposition of r largest eigenvalues.

V. Extensions

Assuming data to follow the Gaussian distribution may make the covariance matrix estimation fragile to potential outliers [7]. Hence, for modeling real-world phenomena and tackling the spurious effect of outliers, one may turn to heavy-tailed assumptions like the Student's t distribution [34].

Technically, assuming $\boldsymbol{x}_i$, $i = 1, \ldots, n$, follows a zero-mean Student's t distribution with a given degree of freedom ν and unknown scatter matrix $\boldsymbol{\Sigma}$, the corresponding MLE problem under the low-rank FA and nonnegative correlation constraints can be formulated as

$$\begin{aligned} \underset{\boldsymbol{\Sigma}}{\text{minimize}} \quad & \log\det(\boldsymbol{\Sigma}) + \frac{p+\nu}{n}\sum_{i=1}^{n}\log\left(1 + \frac{\boldsymbol{x}_i^T\boldsymbol{\Sigma}^{-1}\boldsymbol{x}_i}{\nu}\right) \\ \text{subject to} \quad & \boldsymbol{\Sigma} \in \mathcal{S}_{FA},\ \boldsymbol{\Sigma} \geq \boldsymbol{0} \end{aligned} \tag{20}$$

Using the same approach proposed in Sections IV-A and IV-B, we can obtain a relaxed problem formulation with its objective being (21) shown at the bottom of this page. This problem can be solved by incorporating the MM algorithm. To be specific, at each iteration, we first majorize its third item by linearizing the logarithm function at $\tilde{\boldsymbol{M}}^k$ and $\boldsymbol{\Psi}^k$ as done in [19, Section III, Example 1]. The majorized objective function is given in (22) shown at the bottom of this page, with $\hat{\boldsymbol{S}} = \sum_{i=1}^{n} a_i^k \boldsymbol{x}_i \boldsymbol{x}_i^T$, where a_i^k is a constant scale as

$$a_i^k = \frac{1}{\nu + \boldsymbol{x}_i^T\left(\boldsymbol{\Psi}^k\right)^{-\frac{1}{2}}\left(\tilde{\boldsymbol{M}}^k + \boldsymbol{I}\right)^{-1}\left(\boldsymbol{\Psi}^k\right)^{-\frac{1}{2}}\boldsymbol{x}_i} \tag{23}$$

Since the majorized problem follows the same form as problem (10), we can simply solve it via our proposed Algorithm 3. In fact, we may also choose to run Algorithm 3 for one iteration. This method is known as block majorization-minimization algorithm [19].

Our proposed problem formulation may be extended to other cases, and our proposed algorithm framework can be easily adopted to cope with the specific cases. We discuss some examples in the Appendix B.

VI. Numerical Experiments

In this section, we conduct numerical experiments to illustrate the performance of our proposed formulation and algorithm. Using synthetic data, we first show the efficiency of our proposed algorithm and compare our proposed formulation with benchmarks. Then we use the proposed covariance matrix estimator to design the portfolio and perform a backtest with real market data.

A. Synthetic Data Experiments

Here the experiments are conducted on synthetic data. We first present the procedures for generating data. Then we show the empirical convergence of our proposed algorithm, and the performance of our proposed formulation via comparing with other methods.

Data generation process is similar to that in [12] but modified to follow our FANC model. Setting $p = 200$ and $r = 20$, we generate a $p \times r$ factor loading matrix by first generating an all-zero matrix $\boldsymbol{B}$. Then, for each row of $\boldsymbol{B}$, we randomly select 5 entries and replace the numbers in these positions by drawing from a uniform distribution of the interval $[0.5, 1.5]$. The diagonal elements of the true $\boldsymbol{\Psi}$ are first drawn from a normal distribution $\mathcal{N}(\hat{\mu}, 0.1)$ and then set to be their absolute values, where $\hat{\mu}$ is the average of $\mathrm{diag}(\boldsymbol{B}\boldsymbol{B}^{\mathrm{T}})$. Then we compute the true covariance matrix by $\boldsymbol{\Sigma}^\star = \boldsymbol{B}\boldsymbol{B}^T + \boldsymbol{\Psi}$. Given the number of samples n, we generate $\boldsymbol{x}_i \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{\Sigma})$, $i = 1, \ldots, n$, and collect $\boldsymbol{X} = [\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_n]^T$. The sample covariance matrix

$$\log\det(\boldsymbol{\Psi}) + \sum_{i=1}^{r}\log(\lambda_i + 1) + \frac{p+\nu}{n}\sum_{i=1}^{n}\log\left(1 + \boldsymbol{x}_i^T\boldsymbol{\Psi}^{-\frac{1}{2}}\left(\tilde{\boldsymbol{M}} + \boldsymbol{I}\right)^{-1}\boldsymbol{\Psi}^{-\frac{1}{2}}\boldsymbol{x}_i/\nu\right) + \frac{\rho}{2}\|\tilde{\boldsymbol{M}} - \boldsymbol{U}\boldsymbol{\Lambda}\boldsymbol{U}^T\|_F^2 \tag{21}$$

$$\log\det(\boldsymbol{\Psi}) + \sum_{i=1}^{r}\log(\lambda_i + 1) + \frac{p+\nu}{n}\mathrm{tr}\left(\boldsymbol{\Psi}^{-\frac{1}{2}}\left(\tilde{\boldsymbol{M}} + \boldsymbol{I}\right)^{-1}\boldsymbol{\Psi}^{-\frac{1}{2}}\hat{\boldsymbol{S}}\right) + \frac{\rho}{2}\|\tilde{\boldsymbol{M}} - \boldsymbol{U}\boldsymbol{\Lambda}\boldsymbol{U}^{\mathrm{T}}\|_{\mathrm{F}}^2 \tag{22}$$

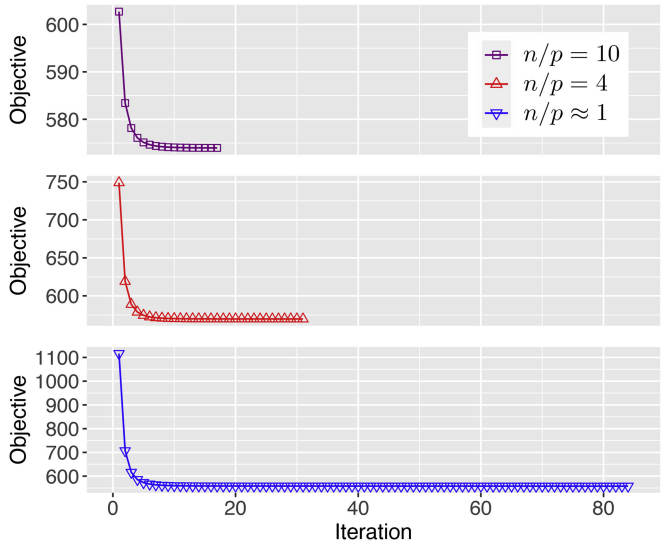

Fig. 5. Convergence of our proposed algorithm under different ratios n/p with $\rho = 100$. The algorithm terminates when the relative change of each variable is less than 10^{-6}.

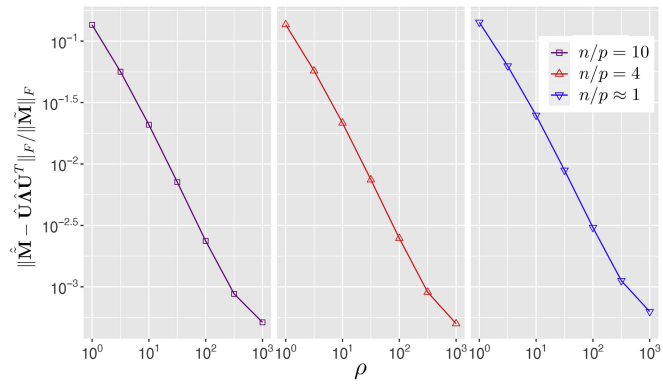

Fig. 6. Relative value of penalty term at convergence versus ρ under different ratio n/p.

is calculated by $\boldsymbol{S} = \frac{1}{n}\boldsymbol{X}^T\boldsymbol{X}$. By default, we set the number of factors $r = 0.1p$ and the convergence tolerance $\epsilon = 10^{-6}$ in all algorithms.

1) Algorithm Performance: Here we illustrate the performance of our proposed algorithm in solving problem (10). The initial point of our algorithm is set as $\tilde{\boldsymbol{M}}^0 = [\hat{\boldsymbol{\Psi}}_F^{-\frac{1}{2}}\hat{\boldsymbol{B}}_F\hat{\boldsymbol{B}}_F^T\hat{\boldsymbol{\Psi}}_F^{-\frac{1}{2}}]_+$, $\boldsymbol{\Psi}^0 = \hat{\boldsymbol{\Psi}}_F$, $\boldsymbol{\lambda}^0 = [\delta_1, \ldots \delta_r]_+^T$, and $\boldsymbol{U}^0 = [\boldsymbol{V}_1, \ldots, \boldsymbol{V}_r]$, where $(\hat{\boldsymbol{B}}_F, \hat{\boldsymbol{\Psi}}_F)$ is the solution of problem (3), and $\boldsymbol{V}\mathrm{Diag}(\boldsymbol{\delta})\boldsymbol{V}^{\mathrm{T}}$ is the eigen-decomposition of $\tilde{\boldsymbol{M}}^0$.

In Fig. 5, we compare the convergence of our proposed algorithm when $\boldsymbol{S}$ is generated with different numbers of samples. Therefore, the fed initial points and achieved final objectives are also different. Our proposed algorithm can converge in all cases within a few number of iterations. Besides, large ratio n/p benefits our proposed algorithm in the convergence speed.

In Fig. 6, we compare the relative value of the penalty term in problem (10), i.e., $\|\hat{\tilde{\boldsymbol{M}}} - \hat{\boldsymbol{U}}\hat{\boldsymbol{\Lambda}}\hat{\boldsymbol{U}}^T\|_F / \|\tilde{\boldsymbol{M}}\|_F$, where $\tilde{\boldsymbol{M}} = \boldsymbol{\Psi}^{-\frac{1}{2}}\boldsymbol{B}\boldsymbol{B}^T\boldsymbol{\Psi}^{-\frac{1}{2}}$. Clearly, the difference between estimated $\tilde{\boldsymbol{M}}$ and $\boldsymbol{U}\boldsymbol{\Lambda}\boldsymbol{U}^T$ is monotonically decreasing with the increase of ρ. When $\rho \geq 100$, the relative difference between $\tilde{\boldsymbol{M}}$ and $\boldsymbol{U}\boldsymbol{\Lambda}\boldsymbol{U}^T$ at convergence has already been very small ($\leq 0.5\%$), which implies $\tilde{\boldsymbol{M}}$ is close to a low-rank positive semidefinite matrix. Therefore, we set $\rho = 100$ in all the following experiments.

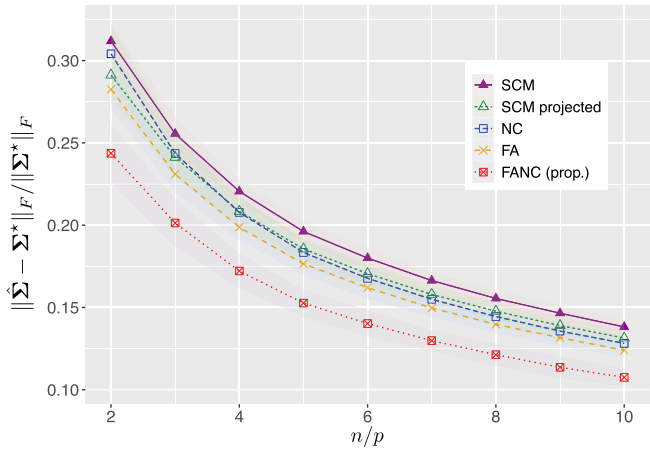

Fig. 7. RMSE of estimated covariance matrix versus ratio n/p ($r = 0.1p$).

2) Formulation Property: Here we assess the estimation performance of our proposed formulation (7). For comparison, we also estimate the covariance matrix via the following benchmark methods:

1) SCM: the sample covariance matrix;
2) SCM projected: the solution of problem (6), which uses the nonnegative correlation structure in matrix approximation fashion;
3) FA: the solution of problem (3), which only uses the low-rank FA structure;
4) NC: it is implemented via our proposed algorithm but r is set to be sufficiently large, and thus the information of the low-rank FA is approximately removed.

In Fig. 7, we compare the relative root mean square error (RMSE) of the estimated covariance matrix versus ratio n/p. The results are averaged over 100 Monte Carlo realizations. We can see that all methods obtain more accurate results with a larger ratio n/p, and our proposed formulation is able to consistently outperform the other methods. Then we test the performance of our proposed formulation with data generated from a mismatched covariance matrix, which is generated by adding noise values drawn from $\mathcal{N}(-0.5, 0.1)$ to off-diagonal entries in $\boldsymbol{\Sigma}^\star$. Now the mismatched covariance matrix has around 20% negative off-diagonal entries. In Fig. 8, we repeat experiments as conducted in Fig. 7 but with the mismatched covariance matrix. Note that the final RMSE is reported by comparing with the correct $\boldsymbol{\Sigma}^\star$. It is reasonable that all the methods are getting worse estimation performance in the mismatched setting. The SCM estimator shows the worst performance, while our proposed formulation maintains its superior performance among all the methods. It is because our proposed method makes use of the prior information from both low-rank FA model and nonnegative correlations. Besides, the SCM projected and NC methods become better than the FA method.

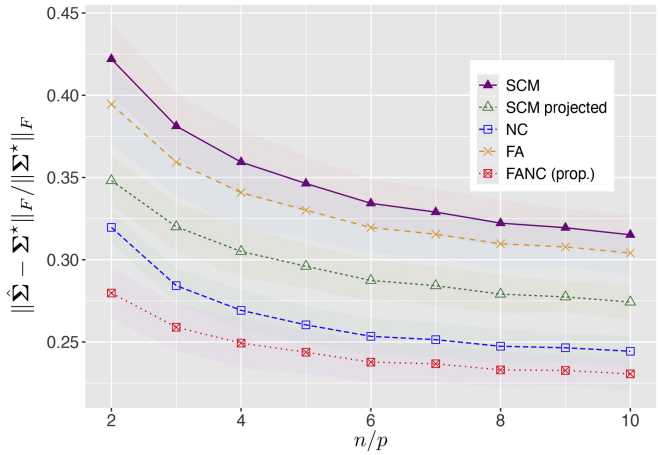

Fig. 8. RMSE of estimated covariance matrix versus ratio n/p with a mismatched $\mathbf{\Sigma}^\star$ ($r = 0.1p$).

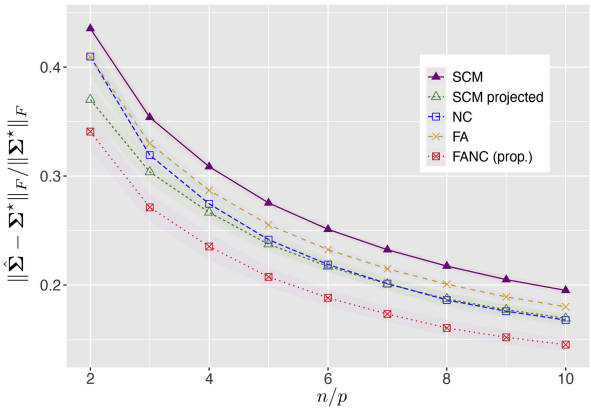

Fig. 9. RMSE of estimated covariance matrix versus ratio n/p ($r = 0.2p$).

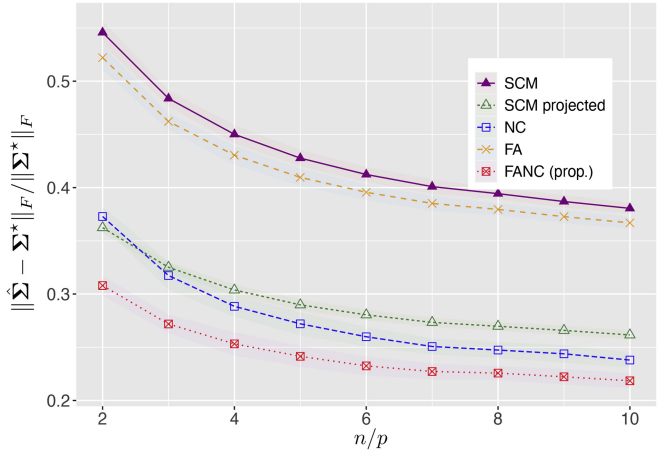

Fig. 10. RMSE of estimated covariance matrix versus ratio n/p with a mismatched $\mathbf{\Sigma}^\star$ ($r = 0.2p$).

Then, we repeat the above experiments but increase the number of factors to $r = 0.2p$. The numerical results are shown in Fig. 9 and Fig. 10. Clearly, the performance of the FA method gets worse, while our proposed FANC method still maintains its superiority. This may be explained that the correct power from the low-rank FA model becomes weaker with larger number of factors.

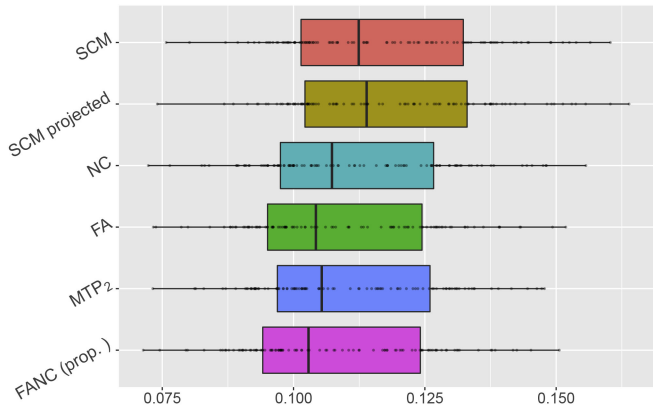

Fig. 11. Boxplot of annual volatility obtained by GMVP according to different estimators for the covariance matrix.

B. Real Financial Data Experiments

Here the experiments are conducted with the real financial data. A commonly used technique in evaluating the quality of the estimated covariance matrix $\hat{\mathbf{\Sigma}}$ with real data is through backtesting the global minimum variance portfolio (GMVP) [15]:

$$\begin{aligned} \underset{\boldsymbol{w}\in\mathbb{R}^p}{\text{minimize}} \quad & \boldsymbol{w}^T\hat{\mathbf{\Sigma}}\boldsymbol{w} \\ \text{subject to} \quad & \mathbf{1}^T\boldsymbol{w} = 1, \end{aligned} \tag{24}$$

whose closed-form solution is $\boldsymbol{w}^\star = \hat{\mathbf{\Sigma}}^{-1}\mathbf{1}/(\mathbf{1}^T\hat{\mathbf{\Sigma}}^{-1}\mathbf{1})$. The $\hat{\mathbf{\Sigma}}$ is said to have higher quality if the volatility of portfolio return is lower.

We first download the 10-year (from 2010-01-01 to 2019-12-31) historical daily price[1] data of stocks listed in the S&P 500 Index components.[2] To make sure that our results are realistic, we randomly select 100 datasets from the downloaded data with each of them containing historical daily price of 50 stocks over 500 continuous trading days. Then we perform the backtest on each selected dataset and compare the annual volatility of portfolio return. The data downloading and backtesting are both implemented via the R package portfolioBacktest [35]. The benchmark MTP$_2$ [15] is introduced in this subsection for comparison.[3]

In Fig. 11, we show the boxplot with the annual volatility obtained from backtest with a lookback window length of 150 for estimating the covariance matrix and computing the GMVP, and a window length of 20 for verifying the performance of the portfolio. When required, we pass $r = 5$ and $\epsilon = 10^{-6}$ to all algorithms. We can observe that the estimator by our proposed formulation outperforms other estimators. The SCM projected method gets a worse result than the naive SCM method. The

[1]Historical daily prices of stocks are available at https://finance.yahoo.com/.

[2]Symbols of S&P 500 Index components are available at https://en.wikipedia.org/wiki/List_of_S%26P_500_companies.

[3]We acknowledge the author of [15] for sharing their code, which is publicly accessible at https://github.com/uhlerlab/MTP2-finance.

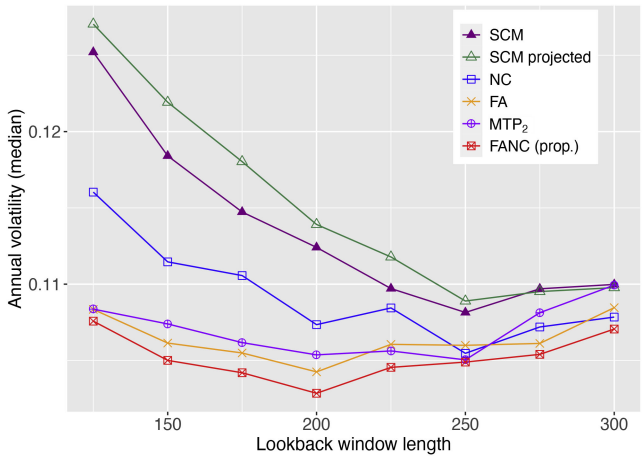

Fig. 12. Median value of annual volatility obtained by GMVP versus lookback window length according to different estimators for the covariance matrix.

reason is that such covariance matrix estimator might be singular in some cases, while the GMVP solution involves its inversion. In Fig. 12, we change the lookback window length and compare the median value of annual volatility across all datasets for each estimator. Significantly the estimator by our proposed formulation can defeat the MTP_2 and FA methods, and is consistently the best among all the estimators.

VII. Conclusion

In this paper, we have considered the covariance matrix estimation problem for factor modeled and nonnegatively correlated data. We have formulated the MLE problem under the low-rank FA and nonnegative correlation constraints. A block coordinate descent based algorithm has been proposed to solve a relaxed problem of our proposed formulation. Numerical experiments have shown the efficiency of our algorithm and the superiority of our formulation.

Appendix

A. Proof for Lemma 1

Set the gradient of the sub-problem (16) objective to zero, i.e.,

$$\begin{aligned} &1/(\lambda_i+1)+\rho(\lambda_i-b_i)\\ &=\left(\rho x^2+\rho(1-b_i)x+1-\rho b_i\right)/(\lambda_i+1)\\ &\propto \rho x^2+\rho(1-b_i)x+1-\rho b_i=0, \end{aligned} \tag{25}$$

whose real number solutions are written as $x_{i,1}$ and $x_{i,2}$. Since its objective goes infinity when $\lambda_i \to \infty$, the optimal solution of the sub-problem (16) will obviously appear either at lower bound, i.e., 0, or points where gradient is zero, i.e., $x_{i,1}$ and $x_{i,2}$. Then conditions of each case can be easily obtained thus omitted here.

B. Extensions to Other Cases

1) Sparsity: To reduce effective number of parameters and produce a sparse estimation, a regularization term may be inserted into the objective of problem (7), i.e.,

$$\begin{aligned} \underset{\boldsymbol{\Sigma}}{\text{minimize}} \quad & \log\det(\boldsymbol{\Sigma})+\operatorname{tr}\left(\boldsymbol{\Sigma}^{-1}\boldsymbol{S}\right)+\beta\sum_{\mathrm{i}\neq\mathrm{j}}\psi\left(\Sigma_{\mathrm{ij}}\right)\\ \text{subject to} \quad & \boldsymbol{\Sigma}\in\mathcal{S}_{FA},\ \boldsymbol{\Sigma}\geq\mathbf{0}, \end{aligned} \tag{26}$$

where $\psi(\cdot)$ is the penalty function encouraging the sparsity of off-diagonal entries of $\boldsymbol{\Sigma}$. It is easy to employ the similar sparsity penalty on the variable $\tilde{\boldsymbol{M}}$, i.e., add an item $\beta\sum_{i\neq j}\psi(\tilde{M}_{ij})$ to the objective of problem (10). The sparsity property is preserved in the recovered covariance matrix estimation. Then we may use the same approach proposed in Section IV-A and IV-B to obtain a relaxed problem formulation with its objective being (29), shown at the bottom of this page.

If $\psi(\cdot)$ is convex, e.g., $\psi(\tilde{M}_{ij})=|\tilde{M}_{ij}|$ as in [36], then Algorithm 3 is still applicable for solving this problem with a slight change in the gradient computation in the corresponding $\tilde{\boldsymbol{M}}$ sub-problem.

If $\psi(\cdot)$ is non-convex, e.g., some folded concave penalties such as smoothly clipped absolute deviation (SCAD) or minimax concave penalty (MCP) to avoid introducing extra bias for estimating nonzero entries with large absolute values [37]. Then at each iteration, we can first majorize such concave penalty by its linearization. For example, if $\psi(\cdot)$ is the MCP function $(\alpha,\gamma>0)$

$$\psi(x)=\begin{cases}\alpha x-\frac{x^2}{2\gamma} & \text{if } 0\leq x\leq\gamma\alpha,\\ \frac{1}{2}\gamma\alpha^2 & \text{otherwise,}\end{cases} \tag{27}$$

it can be majorized at x_k by $\psi(x_k)+a_k(x-x_k)$ with

$$a_k=\begin{cases}\alpha-x/\gamma & \text{if } 0\leq x\leq\gamma\alpha,\\ 0 & \text{otherwise.}\end{cases} \tag{28}$$

We can simply employ the above majorization approach to (29) to obtain the majorized objective function as in (30) shown at the bottom of this page. Then this problem can be solved by our proposed Algorithm 3.

2) Shrinkage: To make the estimation more robust, the regularization item may be introduced to shrink the covariance matrix

$$\log\det(\boldsymbol{\Psi})+\sum_{i=1}^{r}\log(\lambda_i+1)+\operatorname{tr}\left(\boldsymbol{\Psi}^{-\frac{1}{2}}\left(\tilde{\boldsymbol{M}}+\boldsymbol{I}\right)^{-1}\boldsymbol{\Psi}^{-\frac{1}{2}}\boldsymbol{S}\right)+\beta\sum_{\mathrm{i}\neq\mathrm{j}}\psi\left(\tilde{\mathrm{M}}_{\mathrm{ij}}\right)+\frac{\rho}{2}\|\tilde{\boldsymbol{M}}-\boldsymbol{U}\boldsymbol{\Lambda}\boldsymbol{U}^{\mathrm{T}}\|_{\mathrm{F}}^2 \tag{29}$$

$$\log\det(\boldsymbol{\Psi})+\sum_{i=1}^{r}\log(\lambda_i+1)+\operatorname{tr}\left(\boldsymbol{\Psi}^{-\frac{1}{2}}\left(\tilde{\boldsymbol{M}}+\boldsymbol{I}\right)^{-1}\boldsymbol{\Psi}^{-\frac{1}{2}}\boldsymbol{S}\right)+\beta\sum_{\mathrm{i}\neq\mathrm{j}}\mathrm{a}_{\mathrm{ij}}^{\mathrm{k}}|\tilde{\mathrm{M}}_{\mathrm{ij}}|+\frac{\rho}{2}\|\tilde{\boldsymbol{M}}-\boldsymbol{U}\boldsymbol{\Lambda}\boldsymbol{U}^{\mathrm{T}}\|_{\mathrm{F}}^2 \tag{30}$$

to a priori target, i.e.,

$$\begin{aligned} \underset{\boldsymbol{\Sigma}}{\text{minimize}} \quad & \log\det(\boldsymbol{\Sigma}) + \operatorname{tr}\left(\boldsymbol{\Sigma}^{-1}\boldsymbol{S}\right) + h_T(\boldsymbol{\Sigma}) \\ \text{subject to} \quad & \boldsymbol{\Sigma} \in \mathcal{S}_{FA},\ \boldsymbol{\Sigma} \succeq \mathbf{0}, \end{aligned} \tag{31}$$

where $h_T(\boldsymbol{\Sigma})$ is a regularized function admitting the form of

$$h_T(\boldsymbol{\Sigma}) = \alpha\left(p\log\left(\operatorname{tr}\left(\boldsymbol{\Sigma}^{-1}\boldsymbol{T}\right)\right) + \log\det(\boldsymbol{\Sigma})\right), \tag{32}$$

and achieves its minimal value at $s\boldsymbol{T}$ with $s > 0$ [38]. Similarly, the same approach in Section IV-A and IV-B can he employed here to relax the problem, and the logarithm item can be majorized by its linearization. Then we may execute the complete Algorithm 3 to solve it or run for only one iteration to update variables.

REFERENCES

[1] D. Ramírez, G. Vazquez-Vilar, R. López-Valcarce, J. Vía, and I. Santamaría, "Detection of rank-P signals in cognitive radio networks with uncalibrated multiple antennas," *IEEE Trans. Signal Process.*, vol. 59, no. 8, pp. 3764–3774, Aug. 2011.

[2] A. Aubry, A. De Maio, and L. Pallotta, "A geometric approach to covariance matrix estimation and its applications to radar problems," *IEEE Trans. Signal Process.*, vol. 66, no. 4, pp. 907–922, Feb. 2018.

[3] S. Lauritzen, C. Uhler, and P. Zwiernik, "Maximum likelihood estimation in Gaussian models under total positivity," *Ann. Statist.*, vol. 47, no. 4, pp. 1835–1863, 2019.

[4] O. Ledoit and M. Wolf, "Honey, I shrunk the sample covariance matrix," *J. Portfolio Manage.*, vol. 30, no. 4, pp. 110–119, 2004.

[5] R. S. Tsay, *Analysis of Financial Time Series*. Hoboken, NJ, USA: Wiley, 2005.

[6] W. F. Sharpe, "Capital asset prices: A theory of market equilibrium under conditions of risk," *J. Finance*, vol. 19, no. 3, pp. 425–442, 1964.

[7] Y. Feng and D. P. Palomar, *A Signal Processing Perspective on Financial Engineering*, vol. 9. Boston, MA, USA: Now Publishers, 2016.

[8] D. Bertsimas, M. S. Copenhaver, and R. Mazumder, "Certifiably optimal low rank factor analysis," *J. Mach. Learn. Res.*, vol. 18, no. 1, pp. 907–959, 2017.

[9] K. G. Jöreskog, "Some contributions to maximum likelihood factor analysis," *Psychometrika*, vol. 32, no. 4, pp. 443–482, 1967.

[10] D. Robertson and J. Symons, "Maximum likelihood factor analysis with rank-deficient sample covariance matrices," *J. Multivariate Anal.*, vol. 98, no. 4, pp. 813–828, 2007.

[11] D. B. Rubin and D. T. Thayer, "EM algorithms for ML factor analysis," *Psychometrika*, vol. 47, no. 1, pp. 69–76, 1982.

[12] K. Khamaru and R. Mazumder, "Computation of the maximum likelihood estimator in low-rank factor analysis," *Math. Program.*, vol. 176, no. 1, pp. 279–310, 2019.

[13] R. Zhou, J. Liu, S. Kumar, and D. P. Palomar, "Robust factor analysis parameter estimation," in *Proc. Comput. Aided Syst. Theory*, 2020, pp. 3–11.

[14] S. Fallat, S. Lauritzen, K. Sadeghi, C. Uhler, N. Wermuth, and P. Zwiernik, "Total positivity in Markov structures," *Ann. Statist.*, vol. 45, no. 3, pp. 1152–1184, 2017.

[15] R. Agrawal, U. Roy, and C. Uhler, "Covariance matrix estimation under total positivity for portfolio selection," *J. Financial Econometrics*, vol. 20, no. 2, pp. 367–389, 2022.

[16] E. F. Fama and K. R. French, *Common Risk Factors in the Returns on Stocks and Bonds*. Chicago, IL, USA: Univ. Chicago Press, 2021.

[17] K. Berends, R. McMenamin, T. Plestis, and R. J. Rosen, "The sensitivity of life insurance firms to interest rate changes," *Econ. Perspectives*, vol. 37, no. 2, pp. 47–78, 2013.

[18] C. Weis, R.-O. Woltering, and S. Sebastian, "Which stocks are driven by which interest rates?," *J. Property Res.*, vol. 38, no. 3, pp. 175–197, 2021.

[19] Y. Sun, P. Babu, and D. P. Palomar, "Majorization-minimization algorithms in signal processing, communications, and machine learning," *IEEE Trans. Signal Process.*, vol. 65, no. 3, pp. 794–816, Feb. 2017.

[20] A. P. Shikhaliev, L. C. Potter, and Y. Chi, "Low-rank structured covariance matrix estimation," *IEEE Signal Process. Lett.*, vol. 26, no. 5, pp. 700–704, May 2019.

[21] N. J. Higham, "Computing the nearest correlation matrix–A problem from finance," *IMA J. Numer. Anal.*, vol. 22, no. 3, pp. 329–343, 2002.

[22] A. Todeschini, F. Caron, and M. Chavent, "Probabilistic low-rank matrix completion with adaptive spectral regularization algorithms," in *Proc. Adv. Neural Inf. Process. Syst.*, 2013, vol. 26, pp. 845–853.

[23] A. Argyriou, C. A. Micchelli, M. Pontil, and Y. Ying, "A spectral regularization framework for multi-task structure learning," in *Proc. Adv. Neural Inf. Process. Syst.*, 2007, pp. 25–32.

[24] S. Kumar, J. Ying, J. V. de Miranda Cardoso, and D. P. Palomar, "A unified framework for structured graph learning via spectral constraints," *J. Mach. Learn. Res.*, vol. 21, no. 22, pp. 1–60, 2020.

[25] S. J. Wright, "Coordinate descent algorithms," *Math. Program.*, vol. 151, no. 1, pp. 3–34, 2015.

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

[27] M. Grant and S. Boyd, "CVX: Matlab software for disciplined convex programming," 2014. [Online]. Available: http://stanford.edu/boyd/cvx

[28] C.-J. Hsieh, M. A. Sustik, I. S. Dhillon, and P. Ravikumar, "QUIC: Quadratic approximation for sparse inverse covariance estimation," *J. Mach. Learn. Res.*, vol. 15, no. 83, pp. 2911–2947, 2014.

[29] D. T. Phan and M. Menickelly, "On the solution of ℓ_0-constrained sparse inverse covariance estimation problems," *INFORMS J. Comput.*, vol. 33, no. 2, pp. 531–550, 2021.

[30] F. Zhang, *The Schur Complement and Its Applications*. Berlin, Germany: Springer, 2006.

[31] P.-A. Absil, R. Mahony, and R. Sepulchre, *Optimization Algorithms on Matrix Manifolds*. Princeton, NJ, USA: Princeton Univ. Press, 2009.

[32] K. Benidis, Y. Sun, P. Babu, and D. P. Palomar, "Orthogonal sparse PCA and covariance estimation via procrustes reformulation," *IEEE Trans. Signal Process.*, vol. 64, no. 23, pp. 6211–6226, Dec. 2016.

[33] F. L. Gall, "Powers of tensors and fast matrix multiplication," in *Proc. 39th Int. Symp. Symbolic Algebr. Computation*, 2014, pp. 296–303.

[34] C. Liu and D. B. Rubin, "ML estimation of the t distribution using EM and its extensions, ECM and ECME," *Statistica Sinica*, vol. 5, no. 1, pp. 19–39, 1995.

[35] D. P. Palomar and R. Zhou, "portfolioBacktest: Automated backtesting of portfolios over multiple datasets," R Package Version 0.2.0, 2019. [Online]. Available: https://CRAN.R-project.org/package=portfolioBacktest

[36] J. Bien and R. J. Tibshirani, "Sparse estimation of a covariance matrix," *Biometrika*, vol. 98, no. 4, pp. 807–820, 2011.

[37] J. Fan, Y. Liao, and H. Liu, "An overview of the estimation of large covariance and precision matrices," *Econometrics J.*, vol. 19, no. 1, pp. C1–C32, 2016.

[38] Y. Sun, P. Babu, and D. P. Palomar, "Regularized robust estimation of mean and covariance matrix under heavy-tailed distributions," *IEEE Trans. Signal Process.*, vol. 63, no. 12, pp. 3096–3109, Jun. 2015.

Rui Zhou received the B.Eng. degree in information engineering from Southeast University, Nanjing, China, in 2017, and the Ph.D. degree from the Hong Kong University of Science and Technology, Hong Kong, in 2021. He is currently a Research Scientist with Shenzhen Research Institute of Big Data, Shenzhen, China. His research interests include optimization algorithms, statistical signal processing, machine learning, and financial engineering.

Jiaxi Ying received the Ph.D. degree from the Department of Electronic and Computer Engineering, Hong Kong University of Science and Technology, Hong Kong, in 2022. He is currently a Postdoctoral Fellow with the Department of Mathematics, Hong Kong University of Science and Technology, Hong Kong. His research interests include intersection of optimization, machine learning, and statistics. He was the recipient of the HKUST RedBird Ph.D. Scholarship Program.

Daniel P. Palomar (Fellow, IEEE) received the Electrical Engineering and Ph.D. degrees from the Technical University of Catalonia (UPC), Barcelona, Spain, in 1998 and 2003, respectively, and was a Fulbright Scholar at Princeton University during 2004–2006.

He is currently a Professor with the Department of Electronic & Computer Engineering and with the Department of Industrial Engineering & Decision Analytics, Hong Kong University of Science and Technology (HKUST), Hong Kong, which he joined in 2006. He had previously held several research appointments, namely, with King's College London (KCL), London, U.K., Stanford University, Stanford, CA, USA, Telecommunications Technological Center of Catalonia (CTTC), Barcelona, Spain, Royal Institute of Technology (KTH), Stockholm, Sweden, University of Rome La Sapienza, Rome, Italy, and Princeton University, Princeton, NJ, USA. His research interests include applications of optimization theory, graph methods, and signal processing in financial systems and big data analytics.

He was the recipient of a 2004/2006 Fulbright Research Fellowship, the 2004, 2015, and 2020 (coauthor) Young Author Best Paper Awards by the IEEE Signal Processing Society, the 2015–2016 HKUST Excellence Research Award, 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 has been a Guest Editor of the IEEE JOURNAL OF SELECTED TOPICS IN SIGNAL PROCESSING 2016 Special Issue on Financial Signal Processing and Machine Learning for Electronic Trading, an Associate Editor for the IEEE TRANSACTIONS ON INFORMATION THEORY and of 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.