Achieving Oracle Rate for Large Covariance Matrix Estimation From Quadratic Measurements

Wenbin Wang and Ziping Zhao
School of Information Science and Technology, ShanghaiTech University, Shanghai, China
Email: {wangwb2023, zipingzhao}@shanghaitech.edu.cn

2025 IEEE International Symposium on Information Theory (ISIT) | 979-8-3315-4399-0/25/$31.00 ©2025 IEEE | DOI: 10.1109/ISIT63088.2025.11195252

***Abstract*—The covariance matrix is a fundamental second-order statistic in signal and information processing that quantifies the linear relationships among multiple variables. When data evolves rapidly, or the acquisition devices have limited processing power and storage—particularly given the large dimensionality of modern datasets—covariance estimation becomes challenging. To address these challenges, it is desirable to estimate the covariance matrix from a single pass over the data with compressive measurements. In this paper, we study covariance matrix estimation in high dimensions based on quadratic measurements, assuming that the covariance matrix exhibits a sparse structure. We formulate the problem as a least-squares estimation with nonconvex sparsity-inducing penalties. To efficiently compute this estimator, we develop a multi-stage convex relaxation algorithm based on the majorization-minimization algorithmic framework. We comprehensively characterize the computational and statistical properties of the iterates from the algorithm and show that the estimator from the proposed method achieves the oracle statistical rate of convergence after sufficient iterations. Numerical simulations support the theoretical findings and validate the performance of the proposed estimator.**

I. Introduction

In recent decades, compressed sensing (CS) has garnered significant attention due to its remarkable compression efficiency and low computational complexity in high-dimensional signal and information processing [1]–[3]. The core principle of CS lies in exploiting the sparsity inherent in many real-world signals when represented in appropriate bases. The sparsity enables the original high-dimensional signals to be efficiently sub-sampled using random linear projections and accurately reconstructed from their low-dimensional measurements [4]–[6]. In many practical applications, however, random processes are involved. In such settings, the reconstruction of original signals itself is not meaningful [7], [8]. Instead, the focus shifts towards extracting critical quantities from signals (such as second-order or even higher-order statistics), which contain valuable features and provide reliable information about the random processes [9]–[11].

The covariance matrix, as a fundamental second-order statistical tool, is pivotal to numerous tasks in signal processing [12]–[14] and machine learning [15]. Specific applications include spectral estimation [16], adaptive filtering [17], beamformer design [18], principal component analysis [19], linear and quadratic discriminant analysis [20], among others. In practice, the covariance matrix is not directly observable and must, hence, be estimated from empirical data. Consider a zero-mean random vector $\boldsymbol{x} \in \mathbb{R}^d$ with covariance matrix $\boldsymbol{\Sigma}^\star$. A widely used estimator of $\boldsymbol{\Sigma}^\star$ is the sample covariance matrix (SCM) $\boldsymbol{S} = \frac{1}{n}\sum_{t=1}^n \boldsymbol{x}_t\boldsymbol{x}_t^\top$. The computation of SCM relies on the availability of complete data samples [21], [22]. However, obtaining a sufficient number of complete data samples for a reliable estimator is computationally expensive and resource-intensive [23], [24], especially when data evolves rapidly or sampling devices have limited processing power and storage. The problem becomes more pronounced when we are faced with modern high-dimensional datasets.

In lieu of conducting complete measurements, estimating the covariance matrix based on compressed measurements has emerged as a viable alternative. Consider a compressed measurement of $\boldsymbol{x}$ in the form $\boldsymbol{a}^\top\boldsymbol{x}$, where $\boldsymbol{a}$ is a sensing vector. The variance of $\boldsymbol{a}^\top\boldsymbol{x}$ serves as a compressed sketch of $\boldsymbol{\Sigma}^\star$, which can be effectively stored and processed. The $\boldsymbol{a}^\top\boldsymbol{\Sigma}^\star\boldsymbol{a}$ is a quadratic function of $\boldsymbol{a}$ and is hence referred to as quadratic measurements (or rank-one measurements) [25]. Our target is to efficiently reconstruct the covariance matrix from quadratic measurements [26]–[28]. A central question in this framework is whether it is possible to design the sensing vector $\boldsymbol{a}$ such that the resulting compressive sketch exhibits statistically desirable properties, enabling accurate reconstruction of the covariance matrix $\boldsymbol{\Sigma}$ from a minimal number of measurements. While $\boldsymbol{\Sigma}$ may generally not be identifiable under arbitrary sensing designs, reliable recovery is possible with cleverly designed $\boldsymbol{a}$ when $\boldsymbol{\Sigma}$ possesses certain latent structures, such as sparsity, in alignment with the principles of CS [29]–[31]. In high-dimensional covariance matrix estimation problems, where the sample size is small relative to the ambient dimension of the underlying parameter to be estimated [22], [32], the sparsity assumption is actually well-justified and widely adopted by many practical applications [25], [33], [34]. In the literature, covariance estimation from quadratic measurements has been investigated [25]–[27]. In [35], [36], the authors explored the recovery of second-order statistics of a cyclostationary signal from random linear measurements, assuming that the covariance matrix is approximately sparse. They focused on the sampling process and proposed an ℓ_1-minimization based recovery method without performance guarantees. The authors in [25] presented a unified framework for compressing and recovering sparse covariance matrices us-

The experiments of this work were supported by the core facility Platform of Computer Science and Communication, SIST, ShanghaiTech University.

ing the ℓ_1-norm heuristic. They showed that $\mathcal{O}\left(s\log\left(d^2/s\right)\right)$ measurements are sufficient for compressing a d-dimensional sparse covariance matrix with s non-zero entries. Concurrently, *Dasarathy et al.* [27] sketched the covariance matrix $\boldsymbol{\Sigma}$ as $\boldsymbol{A}\boldsymbol{\Sigma}\boldsymbol{B}^\top$, assuming $\boldsymbol{\Sigma}$ exhibits distributed sparsity, with sketching matrices $\boldsymbol{A}$ and $\boldsymbol{B}$ constructed from expander graphs. It is proved that the sample complexity for compressing distributed sparse covariance matrices is $\mathcal{O}(\sqrt{d}\log d)$. It is well-known that methods based on the ℓ_1-norm, e.g., Lasso [37], introduces a non-negligible bias into the resulting estimator [38], which compromises the estimation accuracy. To alleviate this bias effect, nonconvex penalties such as smoothly clipped absolute deviation (SCAD) penalty [38], minimax concave penalty (MCP) [39] and capped ℓ_1-regularization [40] have been proposed as alternatives. It has been demonstrated in [38]–[40] that the nonconvex penalized regression can effectively eliminate the estimation bias in Lasso and achieve improved statistical convergence rates. However, the theory of nonconvex penalty in the context of covariance estimation from compressed measurements remains underexplored.

In this paper, we study the quadratic measurement model for covariance matrix estimation. We show that for a more general class of sub-Gaussian sensing vectors than those considered in prior work [25], it is possible to derive the covariance matrix exactly and establish theoretical performance guarantees. We formulate the problem as a least-squares estimation with nonconvex sparsity-inducing penalties. However, the analysis of the theoretical performance guarantee is complicated due to the nonconvexity of the objective function. To address this, we introduce an efficient multi-stage convex relaxation algorithm based on the majorization-minimization (MM) framework, solving the original problem via a sequence of convex subproblems. We prove that the proposed estimator achieves faster statistical convergence rates compared to the conventional estimator using the ℓ_1-norm penalty. Furthermore, under mild assumptions, we rigorously prove that the proposed estimator exhibits oracle statistical properties. Numerical simulations are provided to support the theoretical findings and validate the superior performance of the proposed estimator. Due to space limitation, the proof of the theoretical results of this paper is given in [41].

II. Quadratic Model and Nonconvex Estimator

A. The Quadratic Measurement Model

Consider n independent observations $\{\boldsymbol{x}_t\}_{t=1}^n$, each drawn from a zero-mean random vector $\boldsymbol{x}$ with covariance matrix $\boldsymbol{\Sigma}^*$. Given m sensing vectors $\{\boldsymbol{a}_i\}_{i=1}^m$, the quadratic measurement measurement y_i, $i=1,\ldots,m$, is given by[1]

$$y_i = \frac{1}{n}\sum_{t=1}^{n}\left|\boldsymbol{a}_i^\top\boldsymbol{x}_t\right|^2 + \eta_i = \left\langle \boldsymbol{a}_i\boldsymbol{a}_i^\top, \boldsymbol{S}\right\rangle + \eta_i, \qquad (1)$$

where $\langle\cdot,\cdot\rangle$ denotes the inner product and $\{\eta_i\}_{i=1}^m$ are additive measurement noises. This model is simpler to implement and computationally more efficient than using full-rank measurement matrices with independent and identically distributed (i.i.d.) entries. Our goal is to recover $\boldsymbol{\Sigma}^\star$ from $\{y_i\}_{i=1}^m$.

[1]Since only finite samples are available, we have $\boldsymbol{S} = \boldsymbol{\Sigma}^\star + \boldsymbol{E}$ with $\boldsymbol{E}$ a bias term.

B. Proposed Nonconvex Estimator

Define $\boldsymbol{y} \coloneqq [y_1,\cdots,y_m]^\top$ and $\boldsymbol{\eta} \coloneqq [\eta_1,\cdots,\eta_m]^\top$. The quadratic measurement model in (1) can be compactly expressed as

$$\boldsymbol{y} = \boldsymbol{A}_\otimes \mathrm{vec}\left(\boldsymbol{S}\right) + \boldsymbol{\eta} = \mathcal{A}\left(\boldsymbol{S}\right) + \boldsymbol{\eta},$$

where $\boldsymbol{A}_\otimes = \left[\begin{array}{ccc}(\boldsymbol{a}_1\otimes\boldsymbol{a}_1) & \cdots & (\boldsymbol{a}_m\otimes\boldsymbol{a}_m)\end{array}\right]^\top$ and $\mathrm{vec}\left(\boldsymbol{S}\right)$ denotes the vectorization of $\boldsymbol{S}$ obtained by stacking its columns, and $\mathcal{A}:\mathbb{R}^{d\times d}\mapsto\mathbb{R}^m$ is a linear operator. Given a collection of m samples $\{(y_i,\boldsymbol{a}_i)\}_{i=1}^m$, we consider the following optimization problem

$$\min_{\boldsymbol{\Sigma}\succ\boldsymbol{0}} \quad \frac{1}{2m}\left\|\boldsymbol{y}-\mathcal{A}\left(\boldsymbol{\Sigma}\right)\right\|_2^2 - \tau\log\det\boldsymbol{\Sigma} + \sum_{i,j}p_\lambda\left(\left|\Sigma_{ij}\right|\right), \quad (2)$$

where the first term is an empirical error; the second term is a log-determinant barrier function with $\tau>0$ to enforce the estimator to be strictly positive definite; and the third term represents a nonconvex regularization term based on function p_λ with $\lambda>0$.

In this paper, we consider a class of nonconvex penalty functions $p_\lambda\left(\cdot\right)$ satisfying the following assumptions.

Assumption 1. *The function $p_\lambda:\mathbb{R}\to\mathbb{R}$ satisfies:*

- p_λ is symmetric around zero with $p_\lambda\left(0\right)=0$, nondecreasing on the non-negative, differentiable almost everywhere on $(0,+\infty)$, and sub-differentiable at $t=0$;
- $0\le p_\lambda'\left(t_1\right)\le p_\lambda'\left(t_2\right)\le\lambda$ for all $t_1\ge t_2\ge 0$ and $\lim_{t\to 0^+}p_\lambda'\left(t\right)=\lambda$;
- There exists an $\alpha>0$ such that $p_\lambda'\left(t\right)=0$ for $t\ge\alpha\lambda$;

Many nonconvex functions have been proven to satisfy Assumption 1, such as SCAD [38] and MCP [39].

III. Optimization Algorithm

A. The MM Algorithm Framework

The MM algorithm framework [42] is an iterative optimization framework comprising two key steps: the majorization step and the minimization step. To minimize a real-valued function $F\left(\boldsymbol{x}\right)$, the algorithm proceeds as follows:

- The majorization step: Construct a surrogate function $\bar{F}\left(\boldsymbol{x}\mid\boldsymbol{x}^{(k-1)}\right)$ that satisfies
$$\begin{cases}\bar{F}(\boldsymbol{x}\mid\boldsymbol{x}^{(k-1)})\ge F(\boldsymbol{x}),\\ \bar{F}(\boldsymbol{x}^{(k-1)}\mid\boldsymbol{x}^{(k-1)})=F(\boldsymbol{x}^{(k-1)}).\end{cases}$$
- The minimization step: Minimize the surrogate function to obtain the next iterate:
$$\boldsymbol{x}^{(k)}\in\arg\min\bar{F}\left(\boldsymbol{x}\mid\boldsymbol{x}^{(k-1)}\right).$$

This iterative process ensures that $F\left(\boldsymbol{x}^{(k)}\right)\le F\left(\boldsymbol{x}^{(k-1)}\right)$ at each step, progressively reducing the value of the objective function. The algorithm starts from an initial feasible point $\boldsymbol{x}^{(0)}$ and continues until a convergence criterion is met.

Algorithm 1: The MM Algorithm for Problem (2).

Input: $\{y_i, \boldsymbol{a}_i\}_{i=1}^m$, τ, λ;
Initialize $\tilde{\boldsymbol{\Sigma}}^{(0)} = \boldsymbol{I}$;
for $k = 1, 2, \ldots, K$ **do**
  $\Lambda_{ij}^{(k-1)} = p'_\lambda\left(\left|\tilde{\Sigma}_{ij}^{(k-1)}\right|\right)$;
  obtain $\tilde{\boldsymbol{\Sigma}}^{(k)}$ by solving problem (4);
end
Output: $\tilde{\boldsymbol{\Sigma}}^{(K)}$

B. The Multi-stage Convex Relaxation Algorithm

We now introduce a multi-stage convex relaxation algorithm based on the MM framework to solve problem (2). In each iteration, we approximate the nonconvex penalty term $\sum_{i,j} p_\lambda(\Sigma_{ij})$ using a weighted ℓ_1-norm surrogate. Specifically, for each iteration $k\,(1 \le k \le K)$, we consider the following convex problem:

$$\min_{\boldsymbol{\Sigma} \succ \mathbf{0}} \quad f(\boldsymbol{\Sigma}) + \sum_{i,j} p'_\lambda\left(\left|\hat{\Sigma}_{ij}^{(k-1)}\right|\right)|\Sigma_{ij}|, \tag{3}$$

where $f(\boldsymbol{\Sigma}) = \frac{1}{2m}\|\boldsymbol{y} - \mathcal{A}(\boldsymbol{\Sigma})\|_2^2 - \tau \log\det \boldsymbol{\Sigma}$ and $\hat{\boldsymbol{\Sigma}}^{(k)}$ denotes the optimal solution to the k-th subproblem. This subproblem can be reformulated compactly as:

$$\min_{\boldsymbol{\Sigma} \succ \mathbf{0}} \quad f(\boldsymbol{\Sigma}) + \|\boldsymbol{\Lambda} \odot \boldsymbol{\Sigma}\|_1, \tag{4}$$

where $\boldsymbol{\Lambda}$ is the regularization parameter matrix with $\Lambda_{ij} = p'_\lambda(|\hat{\Sigma}_{ij}|) \in [0, \lambda]$ and $\|\cdot\|_1$ is the ℓ_1 norm. According to the Karush-Kuhn-Tucker (KKT) conditions, the optimal solution $\hat{\boldsymbol{\Sigma}}$ to each problem (4) satisfies the first-order optimal condition:

$$\nabla f\left(\hat{\boldsymbol{\Sigma}}\right) + \boldsymbol{\Lambda} \odot \hat{\boldsymbol{\Xi}} = \mathbf{0}, \text{ with } \hat{\boldsymbol{\Xi}} \in \partial\|\hat{\boldsymbol{\Sigma}}\|_1,$$

where $\nabla f(\boldsymbol{\Sigma}) = -\frac{1}{m}\mathcal{A}^*(\boldsymbol{y} - \mathcal{A}(\boldsymbol{\Sigma})) - \tau\boldsymbol{\Sigma}^{-1}$, with $\mathcal{A}^*(\cdot)$ the conjugate operator of $\mathcal{A}(\cdot)$. Since problem (4) lacks a closed-form solution, we seek an ε-optimal solution within a pre-specified tolerance, defined in Definition 2.

Definition 2. Let ε denote the tolerance level. $\tilde{\boldsymbol{\Sigma}}^{(k)}$ is considered an ε-optimal solution for the k-th subproblem (4) if the condition $\omega(\tilde{\boldsymbol{\Sigma}}^{(k)}) \le \varepsilon$ is satisfied, with

$$\omega(\tilde{\boldsymbol{\Sigma}}^{(k)}) = \min_{\boldsymbol{\Xi} \in \partial\|\boldsymbol{\Sigma}^{(k)}\|_1} \|\nabla f(\boldsymbol{\Sigma}^{(k)}) + \boldsymbol{\Lambda}^{(k-1)} \odot \boldsymbol{\Xi}\|_{\max},$$

where $\|\cdot\|_{\max}$ denotes the maximum absolute value of all elements in the matrix and ∂ is the subgradient operator.

To obtain $\tilde{\boldsymbol{\Sigma}}^{(k)}$ by solving (4), practicals algorithms like proximal gradient [43] and proximal Newton [44] can be used. The proposed multi-stage convex relaxation algorithm is outlined in Algorithm 1. The initialization is set as $\tilde{\boldsymbol{\Sigma}}^{(0)} = \boldsymbol{I}$ for simplicity.

IV. Statistical Analysis

In this section, we present the theoretical findings. We begin with several essential assumptions.

A. Assumptions

We impose specific conditions on the sensing vectors $\{\boldsymbol{a}_i\}_{i=1}^m$ and the measurement noise $\{\eta_i\}_{i=1}^m$.

Assumption 3. *The sensing vectors $\{\boldsymbol{a}_i\}_{i=1}^m$ are i.i.d. sub-Gaussian random variables with zero mean and identity covariance.*

Assumption 3 is more general than those employed in existing studies (see, e.g., [25], [45]), as it encompasses a broader class of sub-Gaussian distributions for the sensing vectors.

Assumption 4. *The measurement noises $\{\eta_i\}_{i=1}^m$ are i.i.d. sub-exponential random variables with mean 0 and variance proxy σ^2.*

Assumption 4 generalizes the noise models typically employed in the literature (see, e.g., [25], [45]) by allowing sub-exponential rather than sub-Gaussian or bounded noise. Next, we introduce several assumptions related to $\boldsymbol{\Sigma}^\star$. Define the support set of $\boldsymbol{\Sigma}^\star$ as $\mathcal{S} = \{(i,j) \mid \Sigma_{ij}^\star \neq 0\}$, with s representing its cardinality, i.e, $s = |\mathcal{S}|$.

Assumption 5. *The true covariance matrix $\boldsymbol{\Sigma}^\star$ satisfies*

$$0 < \frac{1}{\kappa} \le \lambda_{\min}(\boldsymbol{\Sigma}^\star) \le \lambda_{\max}(\boldsymbol{\Sigma}^\star) \le \kappa < \infty,$$

for some constant $\kappa \ge 1$. Here, $\lambda_{\min}(\boldsymbol{\Sigma}^\star)$ and $\lambda_{\max}(\boldsymbol{\Sigma}^\star)$ denote the minimum and maximum eigenvalues of $\boldsymbol{\Sigma}^\star$, respectively.

Assumption 5 is a standard condition in the study of sparse covariance matrices estimation [22], [32], [46], which ensures that $\boldsymbol{\Sigma}^\star$ is well-conditioned.

Assumption 6. *There exist universal constants α and μ such that*

$$\|\boldsymbol{\Sigma}_{\mathcal{S}}^\star\|_{\min} = \min_{(i,j)\in\mathcal{S}} \left|\Sigma_{ij}^\star\right| \ge (\alpha + \mu)\lambda,$$

where α is from Assumption 1, and $\mu \in (0, \alpha)$ satisfies $p'_\lambda(\mu\lambda) \ge \frac{\lambda}{2}$.

Assumption 6, referred to as the minimum signal strength condition, is prevalent in the analysis of nonconvex penalized regression problems [38], [40], [47]. This condition is relatively mild since the tuning parameter λ is typically chosen to be on the order of $\sqrt{\frac{\log d}{mn}}$, which can be very small as the number of measurements m and sample size n increase.

We further impose two conditions on the function f over a restricted set, known respectively as restricted strong convexity (RSC) and restricted strong smoothness (RSS). Define a local cone around $\boldsymbol{\Sigma}^\star$:

$$\mathcal{B}(\boldsymbol{\Sigma}^\star, r) = \{\boldsymbol{\Sigma} \succ \mathbf{0} \mid \|\boldsymbol{\Sigma} - \boldsymbol{\Sigma}^\star\|_{\mathrm{F}} \le r\}.$$

Assumption 7 (Restricted Strong Convexity (RSC))**.** *For the function f, there exists some $\rho^- > 0$ such that, for all $\boldsymbol{\Delta} \in \mathcal{B}\left(\boldsymbol{\Sigma}^\star, \frac{\rho^-}{4\tau\kappa}\right)$,*

$$f(\boldsymbol{\Sigma}+\boldsymbol{\Delta}) \geq f(\boldsymbol{\Sigma}) + \langle \nabla f(\boldsymbol{\Sigma}), \boldsymbol{\Delta}\rangle + \frac{\rho^-}{2}\|\boldsymbol{\Delta}\|_{\mathrm{F}}.$$

Assumption 8 (Restricted Strong Smoothness (RSS))**.** *For the function f, there exists some $\infty > \rho^+ > \rho^-$ such that, for all $\boldsymbol{\Delta} \in \mathcal{B}\left(\boldsymbol{\Sigma}^\star, \frac{\rho^-}{4\tau\kappa}\right)$,*

$$f(\boldsymbol{\Sigma}+\boldsymbol{\Delta}) \leq f(\boldsymbol{\Sigma}) + \langle \nabla f(\boldsymbol{\Sigma}), \boldsymbol{\Delta}\rangle + \frac{\rho^+}{2}\|\boldsymbol{\Delta}\|_{\mathrm{F}}.$$

Assumption 3 asserts that each row of the sensing matrix $\boldsymbol{A}_\otimes$ comprises i.i.d. sub-exponential random variables. Consequently, $\boldsymbol{A}_\otimes$ exhibits both bounded maximum and strictly positive minimum sparse eigenvalues within the local cone $\mathcal{B}\left(\boldsymbol{\Sigma}^\star, \frac{\rho^-}{4\tau\kappa}\right)$. Specifically, it can be shown that provided that the number of measurements satisfies $m = \mathcal{O}\left(s\log^2(d/s)\right)$, Assumptions 7 and 8 hold with overwhelming probability at least $1 - c_1\exp\left(-c_2\sqrt{m}\right)$ for some $c_1, c_2 > 0$ over the set $\mathcal{B}\left(\boldsymbol{\Sigma}^\star, \frac{\rho^-}{4\tau\kappa}\right)$ [41], [48].

B. Statistical Guarantees and Consequences

We now present the main theorem, which establishes the contraction property of the solution sequence $\{\tilde{\boldsymbol{\Sigma}}^{(k)}\}_{k\geq 1}$.

Theorem 9 (Contraction Property)**.** *Suppose Assumptions 1 $\sim$ 8 hold. Then with probability exceeding $1 - c_1\exp\left(-c_2\sqrt{m}\right)$ for some $c_1, c_2 > 0$, the ε-optimal solution $\tilde{\boldsymbol{\Sigma}}^{(k)}$ from Algorithm 1 is bounded by:*

$$\begin{aligned}\left\|\tilde{\boldsymbol{\Sigma}}^{(k)} - \boldsymbol{\Sigma}^\star\right\|_{\mathrm{F}} \leq & \frac{1}{\rho^-}\left(\underbrace{\left\|(\nabla f(\boldsymbol{\Sigma}^\star))_{\mathcal{S}}\right\|_{\mathrm{F}}}_{\textit{oracle rate}} + \underbrace{\varepsilon\sqrt{s}}_{\textit{optimization error}}\right) \\ & + \underbrace{\delta\left\|\tilde{\boldsymbol{\Sigma}}^{(k-1)} - \boldsymbol{\Sigma}^\star\right\|_{\mathrm{F}}}_{\textit{contraction}},\end{aligned} \tag{5}$$

for $1 \leq k \leq K$, where $\delta \in (0,1)$ is the contraction factor, provided that $m = \mathcal{O}\left((s+s^\diamond)\log^2\left(d/(s+s^\diamond)\right)\right)$ with $s^\diamond \geq \beta s$ for some universal constant β.

Remark 10. Theorem 9 provides a detailed characterization of the estimation error between the ε-optimal solution $\tilde{\boldsymbol{\Sigma}}^{(k)}$ and the ground truth $\boldsymbol{\Sigma}^\star$. This Frobenius norm error is composed of three components: (i) the statistical error characterized by the oracle rate[2], (ii) the optimization error, and (iii) a contraction term.

Next, we give the explicit statistical rate of convergence under the sub-Gaussian design.

Corollary 11. *Let $\boldsymbol{x}$ be a sub-Gaussian random vector with zero mean and covariance $\boldsymbol{\Sigma}^\star$ and $\{\boldsymbol{x}_i\}_{i=1}^n$ be a collection of i.i.d. samples drawn from $\boldsymbol{x}$. Suppose Assumptions 1 $\sim$ 8 hold. If*

$$\lambda \asymp \sqrt{\frac{\log d}{mn}},\ \tau \lesssim \sqrt{\frac{1}{mn}}\left\|(\boldsymbol{\Sigma}^\star)^{-1}\right\|_{\max}^{-1},\ \varepsilon \lesssim \sqrt{\frac{1}{mn}},$$

then the ε-optimal solution $\tilde{\boldsymbol{\Sigma}}^{(1)}$ satisfies

$$\left\|\tilde{\boldsymbol{\Sigma}}^{(1)} - \boldsymbol{\Sigma}^\star\right\|_{\mathrm{F}} \lesssim \sqrt{\frac{s\log d}{mn}}$$

with high probability.

Corollary 12. *Let $\boldsymbol{x}$ be a sub-Gaussian random vector with mean zero and covariance $\boldsymbol{\Sigma}^\star$ and $\{\boldsymbol{x}_i\}_{i=1}^n$ be a collection of i.i.d. samples drawn from $\boldsymbol{x}$. Suppose Assumptions 1 $\sim$ 8 hold. If*

$$\lambda \asymp \sqrt{\frac{\log d}{mn}},\ \tau \lesssim \sqrt{\frac{1}{mn}}\left\|(\boldsymbol{\Sigma}^\star)^{-1}\right\|_{\max}^{-1},\ \varepsilon \lesssim \sqrt{\frac{1}{mn}}$$

and $K \gtrsim \log\left(\lambda\sqrt{mn}\right) \gtrsim \log\log d$, then the ε-optimal solution $\tilde{\boldsymbol{\Sigma}}^{(K)}$ satisfies

$$\left\|\tilde{\boldsymbol{\Sigma}}^{(K)} - \boldsymbol{\Sigma}^\star\right\|_{\mathrm{F}} = \mathcal{O}_p\left(\sqrt{\frac{s}{mn}}\right)$$

with high probability.

Corollary 11 and Corollary 12 directly follow from Theorem 9. These results indicate that to achieve the oracle rate, the optimization error ε must be chosen such that $\varepsilon \leq \min\left(\frac{\left\|(\nabla f(\boldsymbol{\Sigma}^\star))_{\mathcal{S}}\right\|_{\mathrm{F}}}{\sqrt{s}}, \frac{\lambda}{8}\right)$, and the parameter K must be sufficiently large. Consequently, under minimal assumptions, solving no more than approximately $\log\log d$ convex problems suffices to achieve the oracle rate $\sqrt{\frac{s}{mn}}$.

V. Numerical Experiments

In this section, we evaluate the practical performance of the proposed estimator and algorithm. We adopt the MCP, defined as

$$p_\lambda(t) \coloneqq \mathrm{sign}(t)\,\lambda\cdot\int_0^{|t|}\left(1-\frac{z}{\lambda b}\right)_+ dz,$$

with $b = 2$ across all trials. The tuning parameters λ and τ are optimized through five-fold cross-validation. $\boldsymbol{\Sigma}^\star$ is generated using the "`sprandsym`" built-in function in MATLAB with s non-zero entries. We sample n independent data points from the multivariate normal distribution $\mathcal{N}(0, \boldsymbol{\Sigma}^\star)$. η_i's are drawn from a sub-exponential distribution scaled by parameter γ, i.e., $\gamma\cdot\mathcal{N}(0,1)$. The numerical performance is measured using the Frobenius Absolute Error (FAE) $\|\hat{\boldsymbol{\Sigma}} - \boldsymbol{\Sigma}^\star\|_{\mathrm{F}}$ and the Frobenius Relative Error (FRE) $\frac{\|\hat{\boldsymbol{\Sigma}} - \boldsymbol{\Sigma}^\star\|_{\mathrm{F}}}{\|\boldsymbol{\Sigma}^\star\|_{\mathrm{F}}}$. All results are averaged on 100 Monte Carlo trials.

Fig. 1 presents the FRE by varying sparsity ($s \in \{80, 120, 200, 240\}$) with $d = 100$. It shows that increasing the number of measurements m or samples n reduces FRE. Fig. 2 depicts the oracle rate for the "`sprandsym`" matrix under different dimensions ($d \in \{80, 100\}$). The observed estimation errors grow approximately linearly with the theoretical rate, which validates our theoretical guarantee. We further examine

[2]The oracle estimator $\hat{\boldsymbol{\Sigma}}^O$ is defined with prior knowledge of the true support set $\mathcal{S}$, and is given by $\hat{\boldsymbol{\Sigma}}^O = \arg\min_{\boldsymbol{\Sigma}_{\overline{\mathcal{S}}}=\mathbf{0}} f(\boldsymbol{\Sigma})$.

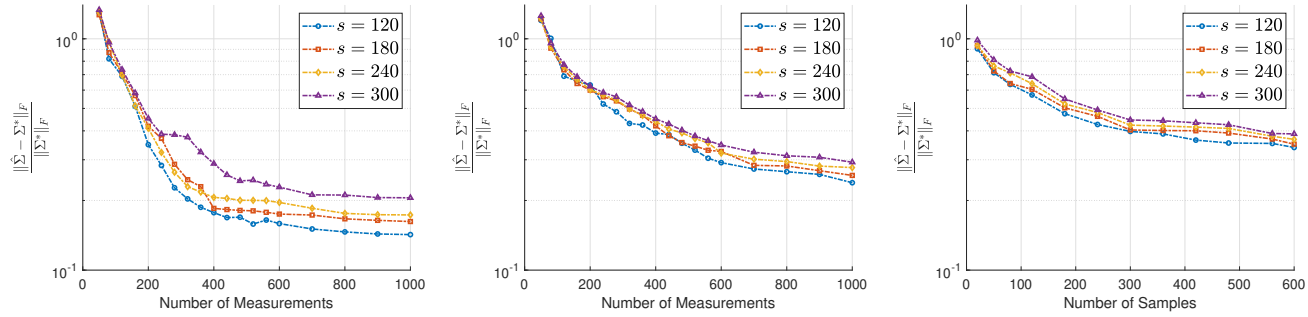

Fig. 1. The FRE of the estimated covariance matrices is examined in three distinct scenarios: (a) the true covariance without added noise; (b) the sample covariance with a noise parameter of $\gamma = 0.1$ and $n = 50$; (c) the sample covariance with a noise parameter of $\gamma = 0.1$ and $m = 300$;

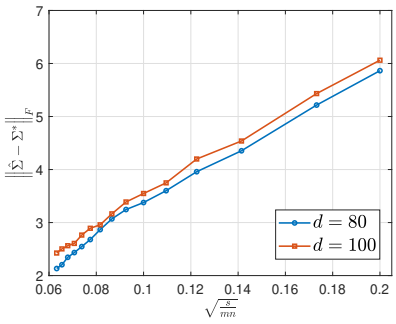

Fig. 2. The oracle rate of "sprandsym" Matrix with $s = 120$ and $\gamma = 0.1$.

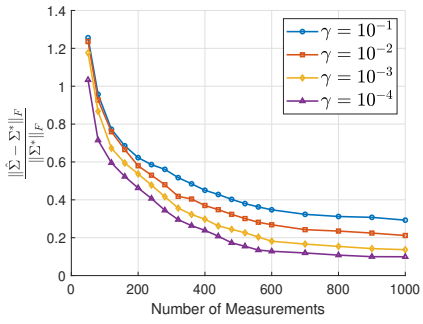

Fig. 3. The FRE of the estimated covariance matrices for different noise levels when $s = 300$.

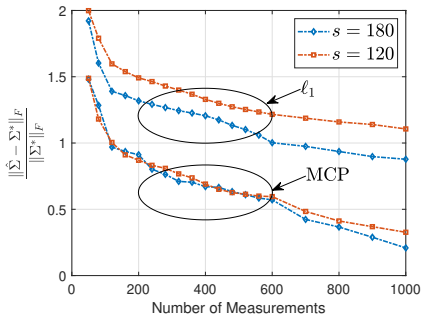

Fig. 4. The FRE of the estimated covariance matrices for different sparsity levels with noise level $\gamma = 10^{-1}$(ℓ_1 v.s. MCP).

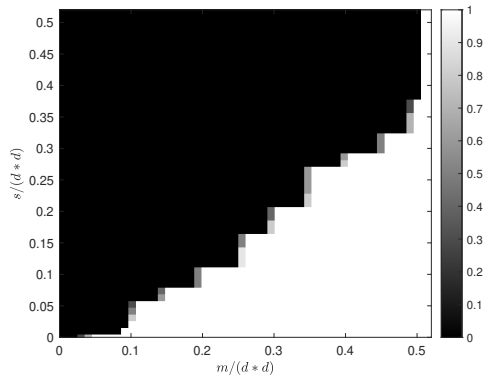

Fig. 5. The rate of successful covariance reconstruction when $d = 100$.

the impact of noise intensity by evaluating the estimator for $s = 300$ under different noise levels. As illustrated in Fig. 3, the influence of noise decreases by increasing m, resulting in improved recovery precision. Notably, the noise intensity exerts a direct and substantial effect on the performance of the oracle estimator. Furthermore, we conduct a comparative analysis between our proposed estimator and the estimator based on the ℓ_1 norm [25], as depicted in Fig. 4. The comparison reveals that the proposed estimator consistently outperforms the ℓ_1-norm-based estimator regarding the Frobenius norm error. These findings support our theoretical findings. We also evaluate the empirical probability of successful recovery using a color-coded matrix in Fig. 5. To minimize the impact of sample size n, we directly sense the true covariance matrix $\boldsymbol{\Sigma}^\star$. A recovery is considered successful if the solution $\hat{\boldsymbol{\Sigma}}$ satisfies $\frac{\|\hat{\boldsymbol{\Sigma}} - \boldsymbol{\Sigma}^\star\|_{\mathrm{F}}}{\|\boldsymbol{\Sigma}^\star\|_{\mathrm{F}}} \leq 10^{-3}$.

VI. Conclusion and Discussion

In this paper, we have investigated the problem of large sparse covariance matrix estimation based on quadratic measurements, which is relevant for scenarios with stringent processing and memory constraints, such as real-time data acquisition systems. Our results demonstrate that sparse covariance matrices can be accurately reconstructed using a minimal number of quadratic measurements, thereby significantly reducing storage requirements. We provide numerical experiments to substantiate our conclusions. Notably, our proposed estimators exhibit superior statistical convergence rates compared to existing methodologies, underscoring their practical efficacy and potential in real-world scenarios. Furthermore, it is noteworthy to mention that our approach is extendable to covariance sensing problems based on the bilinear sampling model $y_i = \langle \boldsymbol{a}_i \boldsymbol{b}_i^\top, \boldsymbol{S} \rangle + \eta_i$, where $\boldsymbol{a}_i$ and $\boldsymbol{b}_i$ are two independent sensing vectors.

REFERENCES

[1] D. L. Donoho, "Compressed sensing," *IEEE Trans. Inf. Theory*, vol. 52, no. 4, pp. 1289–1306, 2006.

[2] M. F. Duarte and Y. C. Eldar, "Structured compressed sensing: From theory to applications," *IEEE Trans. Signal Process.*, vol. 59, no. 9, pp. 4053–4085, 2011.

[3] Y. C. Eldar and G. Kutyniok, *Compressed Sensing: Theory and Applications*. Cambridge, U.K.: Cambridge Univ. Press, May 2012.

[4] E. J. Candès *et al.*, "Compressive sampling," in *Proc. Int. Congr. Math.*, vol. 3. Madrid, Spain, 2006, pp. 1433–1452.

[5] E. J. Candes, J. K. Romberg, and T. Tao, "Stable signal recovery from incomplete and inaccurate measurements," *Commun. Pure Appl. Math.*, vol. 59, no. 8, pp. 1207–1223, 2006.

[6] A. M. Bruckstein, D. L. Donoho, and M. Elad, "From sparse solutions of systems of equations to sparse modeling of signals and images," *SIAM Rev.*, vol. 51, no. 1, pp. 34–81, 2009.

[7] D. D. Arianananda and G. Leus, "Compressive wideband power spectrum estimation," *IEEE Trans. Signal Process.*, vol. 60, no. 9, pp. 4775–4789, 2012.

[8] L. Yang, J. Fang, H. Duan, and H. Li, "Fast compressed power spectrum estimation: Toward a practical solution for wideband spectrum sensing," *IEEE Trans. Wireless Commun.*, vol. 19, no. 1, pp. 520–532, 2020.

[9] W. A. Gardner, "Exploitation of spectral redundancy in cyclostationary signals," *IEEE Signal Process. Mag.*, vol. 8, no. 2, pp. 14–36, 1991.

[10] G. B. Giannakis, "Cyclostationary signal analysis," in *Digital Signal Processing Handbook*, V. K. Madisetti and D. Williams, Eds. Boca Raton, FL: CRC Press, 1998, pp. 433–464.

[11] K. Kim, I. A. Akbar, K. K. Bae, J.-S. Um, C. M. Spooner, and J. H. Reed, "Cyclostationary approaches to signal detection and classification in cognitive radio," in *Proc. IEEE Int. Symposium on New Frontiers in Dynamic Spectrum Access Networks*, 2007, pp. 212–215.

[12] M. Pourahmadi, *High-Dimensional Covariance Estimation: With High-Dimensional Data*. John Wiley & Sons, 2013, vol. 882.

[13] A. Bose and M. Bhattacharjee, *Large Covariance and Autocovariance Matrices*. Chapman and Hall/CRC, 2018.

[14] H. Tsukuma and T. Kubokawa, *Shrinkage Estimation for Mean and Covariance Matrices*. Springer, 2020.

[15] C. M. Bishop, "Pattern recognition and machine learning," *Springer google schola*, vol. 2, pp. 5–43, 2006.

[16] V. Ingle, S. Kogon, and D. Manolakis, *Statisical and Adaptive Signal Processing*. Artech, 2005.

[17] P. S. Diniz *et al.*, *Adaptive Filtering*. Springer, 1997, vol. 4.

[18] R. A. Monzingo and T. W. Miller, *Introduction to Adaptive Arrays*. Scitech publishing, 2004.

[19] I. T. Jolliffe, *Principal Component Analysis*. Springer, 2002.

[20] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Springer, 2009.

[21] R. Vershynin, "How close is the sample covariance matrix to the actual covariance matrix?" *J. Theor. Probability*, vol. 25, no. 3, pp. 655–686, 2012.

[22] Q. Wei and Z. Zhao, "Large Covariance Matrix Estimation With Oracle Statistical Rate via Majorization-Minimization," *IEEE Trans. Signal Process.*, vol. 71, pp. 3328–3342, 2023.

[23] I. M. Johnstone, "High dimensional statistical inference and random matrices," presented at the Int. Congr. of Mathematicians Madrid, Spain, 2006.

[24] N. E. Karoui, "Operator norm consistent estimation of large-dimensional sparse covariance matrices," *Ann. Stat.*, vol. 36, no. 6, pp. 2717 – 2756, 2008.

[25] Y. Chen, Y. Chi, and A. J. Goldsmith, "Exact and stable covariance estimation from quadratic sampling via convex programming," *IEEE Trans. Inf. Theory*, vol. 61, no. 7, pp. 4034–4059, 2015.

[26] D. Romero and G. Leus, "Compressive covariance sampling," in *Proc. Inf. Theory Appl. Workshop (ITA)*, 2013, pp. 1–8.

[27] G. Dasarathy, P. Shah, B. N. Bhaskar, and R. D. Nowak, "Sketching sparse matrices, covariances, and graphs via tensor products," *IEEE Trans. Inf. Theory*, vol. 61, no. 3, pp. 1373–1388, 2015.

[28] D. Romero, R. López-Valcarce, and G. Leus, "Compression limits for random vectors with linearly parameterized second-order statistics," *IEEE Trans. Inf. Theory*, vol. 61, no. 3, pp. 1410–1425, 2015.

[29] E. J. Candes and J. Romberg, "Quantitative robust uncertainty principles and optimally sparse decompositions," *Found. Comput. Math.*, vol. 6, no. 2, pp. 227–254, 2006.

[30] S. Foucart and H. Rauhut, *A Mathematical Introduction to Compressive Sensing*. Basel, Switzerland: Birkhäuser, 2013.

[31] M. Rani, S. B. Dhok, and R. B. Deshmukh, "A systematic review of compressive sensing: Concepts, implementations and applications," *IEEE Access*, vol. 6, pp. 4875–4894, 2018.

[32] C. Lam and J. Fan, "Sparsistency and rates of convergence in large covariance matrix estimation," *Ann. Stat.*, vol. 37, no. 6B, p. 4254, 2009.

[33] W. B. Wu and M. Pourahmadi, "Nonparametric estimation of large covariance matrices of longitudinal data," *Biometrika*, vol. 90, no. 4, pp. 831–844, 2003.

[34] D. Romero, D. D. Arianananda, Z. Tian, and G. Leus, "Compressive Covariance Sensing: Structure-based compressive sensing beyond sparsity," *IEEE Signal Process. Mag.*, vol. 33, no. 1, pp. 78–93, 2016.

[35] G. Leus and Z. Tian, "Recovering second-order statistics from compressive measurements," in *Proc. 4th IEEE Int. Workshop Comput. Adv. Multi-Sensor Adapt. Process. (CAMSAP)*. IEEE, Dec 2011, pp. 337–340.

[36] Z. Tian, Y. Tafesse, and B. M. Sadler, "Cyclic feature detection with sub-Nyquist sampling for wideband spectrum sensing," *IEEE J. Sel. Topics in Signal Process.*, vol. 6, no. 1, pp. 58–69, 2011.

[37] R. Tibshirani, "Regression shrinkage and selection via the lasso," *J Roy. Stat. Soc. Ser. B(Methodol.)*, vol. 58, no. 1, pp. 267–288, Jan 1996.

[38] J. Fan and R. Li, "Variable selection via nonconcave penalized likelihood and its oracle properties," *J. Am. Stat. Assoc.*, vol. 96, no. 456, pp. 1348–1360, 2001.

[39] T. Zhang, "Analysis of multi-stage convex relaxation for sparse regularization." *J. Mach. Learn. Res.*, vol. 11, no. 3, 2010.

[40] C.-H. Zhang, "Nearly unbiased variable selection under minimax concave penalty," *Ann. Stat.*, vol. 38, pp. 894–942, 2010.

[41] [Online]. Available: https://www.ncvxopt.com/pubs/WangZhao-CCS.pdf

[42] 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, 2017.

[43] L. Xiao and T. Zhang, "A proximal-gradient homotopy method for the sparse least-squares problem," *SIAM J. Optim.*, vol. 23, no. 2, pp. 1062–1091, 2013.

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

[45] E. J. Candes, T. Strohmer, and V. Voroninski, "Phaselift: Exact and stable signal recovery from magnitude measurements via convex programming," *Commun. Pure and Appl. Math.*, vol. 66, no. 8, pp. 1241–1274, 2013.

[46] A. J. Rothman, "Positive definite estimators of large covariance matrices," *Biometrika*, vol. 99, no. 3, pp. 733–740, 2012.

[47] J. Fan, H. Liu, Q. Sun, and T. Zhang, "I-LAMM for sparse learning: Simultaneous control of algorithmic complexity and statistical error," *Ann. Stat.*, vol. 46, no. 2, p. 814, 2018.

[48] R. Adamczak, A. E. Litvak, A. Pajor, and N. Tomczak-Jaegermann, "Restricted isometry property of matrices with independent columns and neighborly polytopes by random sampling," *Constructive Approx.*, vol. 34, pp. 61–88, 2011.