Optimal Covariance Sensing via Quadratic Sampling

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

***Abstract*—The covariance matrix is a fundamental statistical quantity that captures the degree of linear correlation between multiple variables. Traditional methods for estimating the covariance matrix typically assume full access to all variable measurements, which becomes impractical when data evolves rapidly or when acquisition devices are constrained by limited processing power and storage capacity. To address these challenges, compressive covariance sensing (CCS) has emerged as a promising approach, enabling covariance matrix estimation from a reduced number of measurements, often significantly smaller than the dimension of the variables. In this paper, we focus on the quadratic sampling model for CCS under the assumption of a sparse covariance structure. We propose a least-squares estimator for the covariance matrix, regularized with positive-definiteness and non-convex sparsity-inducing penalties. To efficiently compute this estimator, we develop a multistage convex relaxation algorithm based on the majorization-minimization (MM) algorithmic framework. We further prove that the estimator from the proposed method achieves the optimal statistical rate of convergence after sufficient iterations. Numerical simulations are provided to support the theoretical findings and validate the performance of the proposed estimator.**

***Index Terms*—Compressive covariance sensing, covariance sketching, quadratic measurements, rank-one measurements, majorization-minimization, sparsity, non-convex statistical optimization.**

I. Introduction

The covariance matrix is a fundamental statistical tool used to quantify the degree of linear correlation among multiple random variables [1]–[3]. It plays a pivotal role in multivariate data processing across science and engineering, including component analysis [4], factor analysis [5], beamformer design [6], portfolio optimization [7], among others. Despite its significance, the covariance matrix is not directly observable and must be estimated from empirical data, thus requiring the development of efficient estimation techniques.

Consider n independent observations $\left\{\boldsymbol{x}_i \in \mathbb{R}^d\right\}_{i=1}^n$ drawn from a zero-mean random vector $\boldsymbol{x}$. The sample covariance matrix (SCM), $\boldsymbol{S} = \frac{1}{n}\sum_{i=1}^n \boldsymbol{x}_i\boldsymbol{x}_i^\top$, is a commonly adopted covariance estimator. However, in high-dimensional settings where the number of variables significantly exceeds the number of observations, the SCM becomes unreliable [8], [9]. To address this, methods based on sparse structural assumption—where many entries in the covariance matrix are assumed to be zero—have been widely explored. Among these methods, thresholding techniques are prominent [10], [11]. Thresholding estimators modify the values in $\boldsymbol{S}$ based on their magnitude, setting certain elements to zero to enforce sparsity. While these approaches are theoretically proven to yield asymptotically positive definite estimators, practical applications often require estimators to be positive definite even with a finite number of samples [10], [12]. To ensure both sparsity and positive definiteness, positive definite regularizers [13], [14] have been proposed as enhancements to thresholding-based covariance matrix estimation. Although considerable progress has been made in covariance estimation from full measurements [12], [15], [16], these methods become impractical in situations where data changes rapidly or the sampling equipments have limited processing and storage capabilities. It is infeasible to obtain full measurement of high-dimensional data in such scenarios.

A viable alternative to covariance estimation based on full measurements is to estimate the covariance matrix using compressed measurements, a technique known as compressive covariance sensing or compressive covariance sketching (CCS) [17]. CCS offers significant advantages by enabling the recovery of covariance matrices from data streams while substantially reducing computation and storage overhead. Compressive sensing for covariance matrices generalizes the classical compressive sensing problem for signals [18], [19], focusing on extracting second-order statistics rather than reconstructing random vectors. CCS is widely applied in signal processing, particularly in scenarios involving random processes where reconstructing the signal itself is not meaningful. This method has been used in various applications, including power spectrum estimation [20], direction of arrival estimation [21], incoherent imaging [22], among others. In biomedical engineering, CCS reduces the amount of data needed for imaging and genomic analysis, enhancing privacy protection [23]. In finance, CCS facilitates the estimation of asset covariance matrices with reduced data storage, minimizing the risk of sensitive information exposure [24].

In compressive sensing, the sparsity of signals is a common assumption. This principle of structural assumption has been extended to the estimation of covariance matrices through CCS. Compressive Spectrum Sensing, as an important application of CCS, has significantly advanced this field. In [25], the authors introduced techniques for compressive wide-band spectrum sensing in cognitive radios. This study underscores the potential of compressive sampling in scenarios not demanding full signal recovery, setting a precedent for the use of quadratic measurements in covariance estimation. Further, [26] focused on the recovery of second-order statistics from compressive measurements, particularly cyclic power spectra. Their method, which does not assume signal sparsity, validates the feasibility of structured recovery in practical signal processing scenarios. Another work [27] introduced minimal sparse ruler sampling for power spectrum estimation, which significantly reduces sampling rates. This concept of minimal sampling dovetails with our research on optimizing the sampling framework for covariance sensing.

In [28], the authors developed methods for compressive periodogram reconstruction using uniform binning, emphasizing the cyclic properties of coset correlation matrices. Their techniques for handling wideband signals without full signal reconstruction parallel our methodology in handling high-dimensional covariance structures efficiently. Moreover, the work [29] proposed to estimate sparse covariance matrix from quadratic sampling with performance guarantee. Since the conventional Restricted Isometry Property (RIP) criteria are not directly applicable to CCS problems [30], a novel variant—the mixed-norm Restricted Isometry Property (RIP-ℓ_2/ℓ_1) was designed for effective covariance recovery. A concurrent work [31] studied the CCS problem with distributed sparsity assumption. Furthermore, [17] extended the CCS problem from sparsity assumption to other assumptions, such as banded, Toeplitz, circulant matrices.

In this paper, we study the CCS problem from quadratic sampling model under the sparse assumption. Different from existing litera-

ture, we adopt a non-convex optimization formulation. We propose a multistage convex relaxation technique within the majorization-minimization (MM) framework. Further, we adopt the sparse eigenvalue for ensuring exact recovery, which is a much weaker condition than the commonly used RIP condition [32], [33]. We further prove that the estimator from the proposed method achieves the optimal statistical rate of convergence after sufficient iterations. Numerical simulations are provided to support the theoretical findings and validate the performance of the proposed estimator.

II. The Measurement Model and Design Matrix

Given a set of sensing vectors $\{\boldsymbol{a}_i\}_{i=1}^m$, where m represents the total number of measurements, a quadratic (a.k.a. rank-one) measurement model is represented as

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

where $\{y_i\}_{i=1}^m$ denotes the measurements, $\{\boldsymbol{x}_t\}_{t=1}^n$ is the observed data, $\{\eta_i\}_{i=1}^m$ indicates the measurement noises. Further, we define $\boldsymbol{y} \coloneqq [y_1,\cdots,y_m]^\top$, $\boldsymbol{\eta} \coloneqq [\eta_1,\cdots,\eta_m]^\top$, and the linear operator $\mathcal{A}:\mathbb{R}^{d\times d}\mapsto\mathbb{R}^m$. Consequently, we have

$$\boldsymbol{y} = \mathcal{A}(\boldsymbol{S}) + \boldsymbol{\eta}.$$

For the sampling model, we introduce several assumptions.

Assumption 1. *Assume that the measurement noises η_i's are independently and identically distributed (i.i.d.) from a sub-exponential distribution with mean 0 and variance proxy σ^2.*

Assumption 2. *Assume that the sensing vectors $\boldsymbol{a}_i$'s are i.i.d. sub-Gaussian random variables. Each element in $\boldsymbol{a}_i$ satisfies:*

$$\mathbb{E}\left[(\boldsymbol{a}_i)_j\right] = 0,\ \mathbb{E}\left[(\boldsymbol{a}_i)_j^2\right] = 1,\ \textit{and}\ \mathbb{E}\left[(\boldsymbol{a}_i)_j^4\right] > 1. \quad (2)$$

III. The Proposed Estimator

In this paper, we consider the following problem for CCS:

$$\min_{\boldsymbol{\Sigma}\succ\boldsymbol{0}}\ \frac{1}{2m}\left\|\boldsymbol{y}-\mathcal{A}(\boldsymbol{\Sigma})\right\|_2^2 - \tau\log\det\boldsymbol{\Sigma} + \sum_{i,j}p_\lambda\left(|\Sigma_{ij}|\right). \quad (3)$$

In (3), the first term aims to minimize the empirical errors; the log-determinant barrier function ensures positive definiteness with $\tau \geq 0$; p_λ represents a non-convex penalty function governed by $\lambda > 0$. We impose certain restrictions on p_λ.

Assumption 3. *The function $p_\lambda:\mathbb{R}\to\mathbb{R}$ satisfies:*

- $p_\lambda(t)$ is symmetric around zero with $p_\lambda(0) = 0$, nondecreasing on the nonnegative, differentiable almost everywhere on $(0,+\infty)$, and subdifferentiable at $t = 0$;
- $0 \leq p_\lambda'(t_1) \leq p_\lambda'(t_2) \leq \lambda$ for all $t_1 \geq t_2 \geq 0$ and $\lim_{t\to 0^+} p_\lambda'(t) = \lambda$;
- There exists an $\alpha > 0$ such that $p_\lambda'(t) = 0$ for $t \geq \alpha\lambda$;

Many non-convex functions have been proved to satisfy the above assumptions, such as smooth clipped absolute deviation penalty [34] and minimax concave penalty [35].

IV. The Proposed Algorithm

The MM algorithm framework is an iterative process that encompasses two main steps: the Majorization step and the Minimization step. Suppose we want to minimize a real-valued function $F(\mathbf{x})$, below is the basic procedure of this algorithm framework:

- Majorization Step: In this step, the goal is to select or construct a surrogate function $\overline{F}\left(\mathbf{x}\mid\mathbf{x}^{(k-1)}\right)$ which equals the value of the original objective function $F(\mathbf{x})$ at the current iteration point $\mathbf{x}^{(k-1)}$, but is an upper bound for $F(\mathbf{x})$ at all other points. Specifically, $\overline{F}\left(\mathbf{x}\mid\mathbf{x}^{(k-1)}\right) \geq F(\mathbf{x})$ for all $\mathbf{x}$ and $\overline{F}\left(\mathbf{x}^{(k-1)}\mid\mathbf{x}^{(k-1)}\right) = F\left(\mathbf{x}^{(k-1)}\right)$.
- Minimization Step: After the surrogate function is determined, the next step is to find the point $\mathbf{x}^{(k)}$ that minimizes $\overline{F}\left(\mathbf{x}\mid\mathbf{x}^{(k-1)}\right)$, such that $\mathbf{x}^{(k)} \in \arg\min \overline{F}\left(\mathbf{x}\mid\mathbf{x}^{(k-1)}\right)$.

These two steps are executed in an alternating fashion, with each iteration aimed at reducing or at least not increasing the value of the objective function $F(\mathbf{x})$, ensuring $F\left(\mathbf{x}^{(k)}\right) \leq F\left(\mathbf{x}^{(k-1)}\right)$. Instead of minimizing $F(\mathbf{x})$ directly, the algorithm focuses on sequentially solving a series of simple optimization problem. The process begins with a feasible initial point $\mathbf{x}^{(0)}$ and continues iteratively through $\mathbf{x}^{(1)},\mathbf{x}^{(2)},\ldots$ until a specified convergence criterion is satisfied.

We now introduce the proposed multistage convex relaxation algorithm for CCS based on the MM approach to find the stationary solution of (3). The multistage convex relaxation algorithm is described in Algorithm 1. To simplify, we define

$$f(\boldsymbol{\Sigma}) = \frac{1}{2m}\left\|\boldsymbol{y}-\mathcal{A}(\boldsymbol{\Sigma})\right\|_2^2 - \tau\log\det\boldsymbol{\Sigma}.$$

Specifically, for each $1 \leq k \leq K$, we solve the subsequent convex relaxation subproblem sequentially:

$$\min_{\boldsymbol{\Sigma}\succ\boldsymbol{0}}\ f(\boldsymbol{\Sigma}) + \sum_{i,j}p_\lambda'\left(\left|\widehat{\Sigma}_{ij}^{(k-1)}\right|\right)|\Sigma_{ij}|, \quad (4)$$

where the second term, a weighted ℓ_1-norm, serves as the surrogate function for $\sum_{i,j}p_\lambda(\Sigma_{ij})$ in (3) and $\widehat{\boldsymbol{\Sigma}}^{(k)}$ represents the optimal solution to the k-th subproblem. Each subproblem can further be reformulated as

$$\min_{\boldsymbol{\Sigma}\succ\boldsymbol{0}}\ f(\boldsymbol{\Sigma}) + \left\|\boldsymbol{\Lambda}\odot\boldsymbol{\Sigma}\right\|_1, \quad (5)$$

where $\boldsymbol{\Lambda}$ is the regularized parameter matrix defined as $\Lambda_{ij} = p_\lambda'\left(\left|\widehat{\Sigma}_{ij}\right|\right) \in [0,\lambda]$. According to the Karush-Kuhn-Tucker (KKT) conditions, the unique sparse global optimal solution $\widehat{\boldsymbol{\Sigma}}$ for each subproblem satisfies the first-order optimal condition:

$$\nabla f\left(\widehat{\boldsymbol{\Sigma}}\right) + \boldsymbol{\Lambda}\odot\widehat{\boldsymbol{\Xi}} = \boldsymbol{0},$$

where $\widehat{\boldsymbol{\Xi}} \in \partial\left\|\widehat{\boldsymbol{\Sigma}}\right\|_1$. Given that the problem (5) lacks an analytical solution, we aim for a suboptimal solution defined under a prespecified tolerance level ε, and terminate the iterations when the approximate KKT condition are satisfied.

Definition 4. For a specified tolerance level ε, the solution $\widetilde{\boldsymbol{\Sigma}}^{(k)}$ is deemed ε-optimal for the k-th subproblem 5 if the condition $\omega_{\boldsymbol{\Lambda}}\left(\widetilde{\boldsymbol{\Sigma}}^{(k)}\right) \leq \varepsilon$ is satisfied, where

$$\omega_{\boldsymbol{\Lambda}^{(k-1)}}\left(\widetilde{\boldsymbol{\Sigma}}^{(k)}\right) = \min_{\boldsymbol{\Xi}\in\partial\|\boldsymbol{\Sigma}\|_1}\left\|\nabla f\left(\boldsymbol{\Sigma}^{(k)}\right) + \boldsymbol{\Lambda}^{(k-1)}\odot\boldsymbol{\Xi}\right\|_{\max}.$$

Here, $\|\cdot\|_{\max}$ denotes the maximum absolute value of all elements in the matrix. In order to guarantee that the solution $\widetilde{\boldsymbol{\Sigma}}^{(k)}$ derived at each stage attains sufficient precision, we set $\varepsilon = \frac{c_1}{\sqrt{mn}} \leq \frac{\lambda}{8}$, where c_1 is a predetermined small constant.

To obtain $\widetilde{\boldsymbol{\Sigma}}^{(k)}$ by solving (5), several numerical methods can be chosen, including proximal gradient [36], proximal Newton [37], [38], etc.

Algorithm 1: The MM-Based Multistage Convex Relaxation Algorithm for Solving (3).

Input: $\{y_i, \boldsymbol{a}_i\}_{i=1}^m$, τ, λ;
Initialize $\widetilde{\boldsymbol{\Sigma}}^{(0)} = \boldsymbol{I}$
for $k = 1, 2, \ldots, K$ **do**
  $\Lambda_{ij}^{(k-1)} = p_\lambda' \left(\left| \widetilde{\Sigma}_{ij}^{(k-1)} \right| \right)$;
  $\widetilde{\boldsymbol{\Sigma}}^{(k)} = \arg\min f(\boldsymbol{\Sigma}) + \|\boldsymbol{\Lambda} \odot \boldsymbol{\Sigma}\|_1$;
  $k = k + 1$;
end
Output: $\widetilde{\boldsymbol{\Sigma}}^{(K)}$

V. Statistical Analysis

A. Assumptions

We present several assumptions pertinent to the true covariance matrix, denoted as $\boldsymbol{\Sigma}^*$. We define the support set of $\boldsymbol{\Sigma}^*$ as $\mathcal{S}^* = \left\{(j,k) \mid \boldsymbol{\Sigma}_{jk}^* \neq 0\right\}$, with s^* representing its size, i.e, $s^* = |\mathcal{S}^*|$.

Assumption 5. *For the true covariance matrix $\boldsymbol{\Sigma}^*$, there exists $\kappa \geq 1$ satisfies*

$$0 < \frac{1}{\kappa} \leq \lambda_{\min}(\boldsymbol{\Sigma}^*) \leq \lambda_{\max}(\boldsymbol{\Sigma}^*) \leq \kappa < \infty.$$

Assumption 5 is a common premise in the literature about the estimation of sparse covariance matrices [9], [13], [16], [39]. This assumption offers several advantages, and readers are encouraged to consult [13] for further details.

Assumption 6. *Given the true covariance matrix $\boldsymbol{\Sigma}^*$, there exist universal constants α, c_2 such that*

$$\|\boldsymbol{\Sigma}_{\mathcal{S}^*}^*\|_{\min} = \min_{(i,j)\in\mathcal{S}^*} \left|\boldsymbol{\Sigma}_{ij}^*\right| \geq (\alpha + c_2)\lambda,$$

where α is a constant introduced in Assumption 3, and $c_2 \in (0, \alpha)$ satisfies $p_\lambda'(c_2\lambda) \geq \frac{\lambda}{2}$.

Assumption 6 is referred to the minimum signal strength condition, commonly applied in the study of non-convex penalized regression problems [34], [35], [40]. Additionally, it is essential to select λ appropriately to ensure that the regularization is sufficiently large to eliminate irrelevant coordinates, thereby yielding the solution is adequately sparse.

Next, we introduce the Sparse Eigenvalues (SE) condition.

Definition 7 (Sparse Eigenvalue)**.** For a positive integer $0 < s_0 < d$, the localized sparse eigenvalues are defined as

$$\rho_{s_0,r}^+ = \sup\left\{ \frac{\boldsymbol{u}^\top \nabla^2 f(\boldsymbol{\Sigma})\, \boldsymbol{u}}{\boldsymbol{u}^\top \boldsymbol{u}} \mid \|\boldsymbol{u}\|_0 \leq s_0, \|\boldsymbol{\Sigma} - \boldsymbol{\Sigma}^*\|_{\mathsf{F}} \leq r \right\},$$
$$\rho_{s_0,r}^- = \inf\left\{ \frac{\boldsymbol{u}^\top \nabla^2 f(\boldsymbol{\Sigma})\, \boldsymbol{u}}{\boldsymbol{u}^\top \boldsymbol{u}} \mid \|\boldsymbol{u}\|_0 \leq s_0, \|\boldsymbol{\Sigma} - \boldsymbol{\Sigma}^*\|_{\mathsf{F}} \leq r \right\}.$$

The SE condition has been extensively studies in high-dimensional sparse recovery or compressed sensing [41]–[45]. Reformulating (1), we have $y_i = (\boldsymbol{a}_i \otimes \boldsymbol{a}_i)^\top \operatorname{vec}(\boldsymbol{S}) + \eta_i$. Define the design matrix $\widetilde{\boldsymbol{A}}$ as follows:

$$\widetilde{\boldsymbol{A}} = \left[\begin{array}{ccc} (\boldsymbol{a}_1 \otimes \boldsymbol{a}_1) & \cdots & (\boldsymbol{a}_m \otimes \boldsymbol{a}_m) \end{array} \right]^\top \in \mathbb{R}^{m \times d^2}.$$

The Assumption 2 indicates that each row of the design matrix $\widetilde{\boldsymbol{A}}$ is composed of i.i.d. sub-exponential random variables. And when $\widetilde{\boldsymbol{A}}$ satisfies $m = \mathcal{O}\left(s_0 \log^{\frac{3}{2}} d\right)$, the SE condition is satisfied with overwhelming probability at least $1 - \exp\left(-\frac{\epsilon^2}{2}\right)$ for some $\epsilon > 0$.

Assumption 8. *There exists a universal constant c_3 such that for $\widetilde{s} \geq c_3 s^*$, the SE property is satisfied with parameters $\rho_{2s^*+2\widetilde{s},r}^-$ and $\rho_{2s^*+2\widetilde{s},r}^+$ such that*

$$0 < \rho_{2s^*+2\widetilde{s},r}^- < \rho_{2s^*+2\widetilde{s},r}^+ < +\infty.$$

Assumption 8 indicates that the sparse eigenvalues of the Hessian matrix $\nabla^2 f(\boldsymbol{\Sigma})$ are lower and upper bounded when $\boldsymbol{\Sigma}$ is adequate sparsity and in close proximity to $\boldsymbol{\Sigma}^*$ with high probability (w.h.p.).

B. Statistical Guarantees and Consequences[1]

Theorem 9 (Contraction Property)**.** *Consider the estimator in 3 and suppose the Assumptions 2, 8, 5, and 6 hold. Then with probability exceeding $1 - \exp\left(-\frac{\epsilon^2}{2}\right)$ for some $\epsilon > 0$, the ε-optimal solution $\widetilde{\boldsymbol{\Sigma}}^{(k)}$ is bounded by:*

$$\left\|\widetilde{\boldsymbol{\Sigma}}^{(k)} - \boldsymbol{\Sigma}^*\right\|_{\mathsf{F}} \leq \frac{1}{\rho_{2s^*+2\widetilde{s}}^-} \left(\underbrace{\left\|(\nabla f(\boldsymbol{\Sigma}^*))_{\mathcal{S}^*}\right\|_{\mathsf{F}}}_{\text{oracle rate}} + \underbrace{\varepsilon\sqrt{s^*}}_{\text{optimization error}} \right) + \underbrace{\delta \left\|\widetilde{\boldsymbol{\Sigma}}^{(k-1)} - \boldsymbol{\Sigma}^*\right\|_F}_{\text{contraction}}, \tag{6}$$

for $1 \leq k \leq K$, where $\delta \in (0,1)$ is the contraction factor, provided that $m = \mathcal{O}\left((s^* + \widetilde{s}) \log^{\frac{3}{2}} d\right)$.

Remark 10. Theorem 9 elaborates the estimation discrepancy between the ε-optimal solution $\widetilde{\boldsymbol{\Sigma}}^{(k)}$ and the ground truth $\boldsymbol{\Sigma}^*$ is constrained by three primary factors: the oracle rate[2], the optimization error, and a contraction term.

Corollary 11. *Let $\mathbf{x}$ be a sub-Gaussian random vector with zero mean and covariance $\boldsymbol{\Sigma}^*$ and $\{\mathbf{x}_i\}_{i=1}^n$ be a collection of independent and identically distributed samples from $\mathbf{x}$. Suppose that Assumptions 2, 8, 5, and 6 hold, that is $\varepsilon \lesssim \sqrt{\frac{1}{mn}}$. If*

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

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

$$\left\|\widetilde{\boldsymbol{\Sigma}}^{(1)} - \boldsymbol{\Sigma}^*\right\|_{\mathsf{F}} \lesssim \sqrt{\frac{s^* \log d}{mn}}$$

with high probability (w.h.p.).

Corollary 12. *Let $\boldsymbol{x}$ be a sub-Gaussian random vector with mean zero and covariance $\boldsymbol{\Sigma}^*$ and $\{\boldsymbol{x}_i\}_{i=1}^n$ be a collection of independent and identically distributed samples from $\mathbf{x}$. Suppose that Assumptions 2, 8, 5, and 6 hold, that is $\varepsilon \lesssim \sqrt{\frac{1}{mn}}$. If*

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

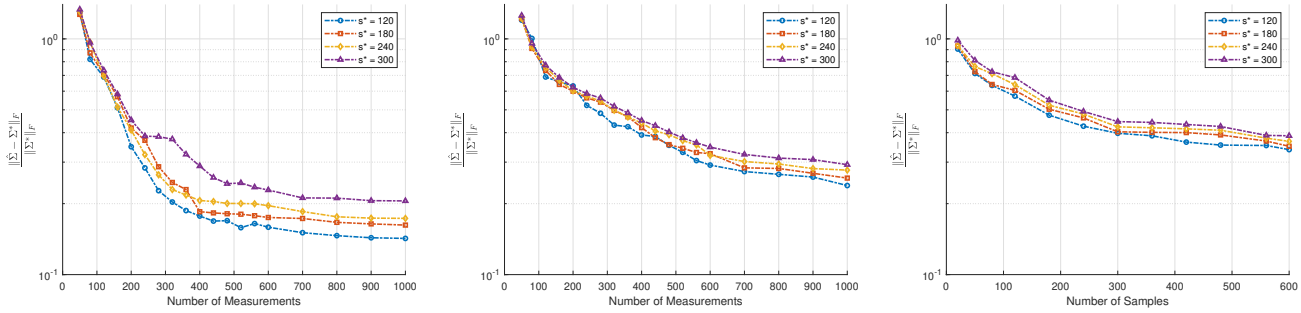
and $K \gtrsim \log(\lambda\sqrt{mn}) \gtrsim \log\log d$, then the ε-optimal solution $\widetilde{\boldsymbol{\Sigma}}^{(K)}$ satisfies

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

w.h.p.

[1]Due to space limitation, the proof to the theoretical results is given in [46].

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



(a) (b) (c)

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

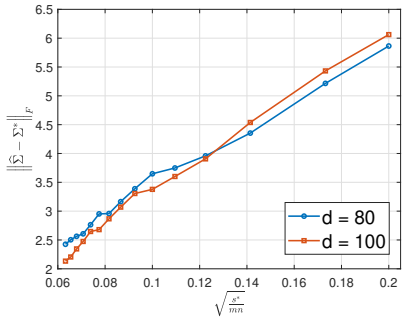


Fig. 2. The oracle rate of Sprandsym Matrix with $s^* = 120$, $n = 50$.

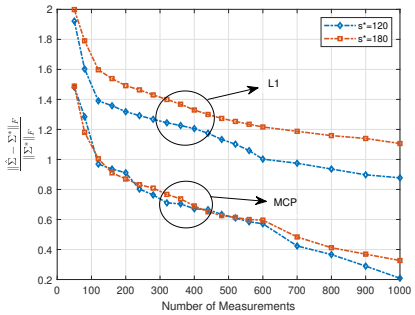


Fig. 3. ℓ_1 v.s. MCP

Corollary 11 and Corollary 12 are direct consequence of Theorem 9. It demonstrates that to achieve the oracle rate, the optimization error ε must be chosen such that $\varepsilon \leq \frac{\left\|(\nabla f(\boldsymbol{\Sigma}^*))_{\mathcal{S}^*}\right\|_{\mathsf{F}}}{\sqrt{s^*}}$, and the parameter K must be sufficiently large. The latter implies that, with minimal assumptions, solving no more than approximately $\log\log d$ convex problems is enough to achieve the oracle rate $\sqrt{\frac{s^*}{mn}}$.

VI. Numerical Experiments

We examine the practical performance of proposed estimator and provide numerical results for the proposed algorithm. We employ the MCP with a constant setting of $b = 2$ across all trials. The tuning parameters λ and τ are optimized through five-fold cross-validation. The true covariance matrix $\boldsymbol{\Sigma}^*$ is generated by the "sprandsym" built-in function in MATLAB with a sparsity value s^* and n independent data points are sampled i.i.d. from a $\mathcal{N}(0, \boldsymbol{\Sigma}^*)$ distribution. Noise η_i is generated from a sub-Gaussian distributed random variable scaled by γ, specifically $\gamma \cdot \mathcal{N}(0, 1)$. The performance is quantifies by the Frobenius absolute error (FAE) $\left\|\widehat{\boldsymbol{\Sigma}} - \boldsymbol{\Sigma}^*\right\|_{\mathsf{F}}$ and the Frobenius relative error (FRE) $\frac{\left\|\widehat{\boldsymbol{\Sigma}} - \boldsymbol{\Sigma}^*\right\|_{\mathsf{F}}}{\|\boldsymbol{\Sigma}^*\|_{\mathsf{F}}}$, where $\widehat{\boldsymbol{\Sigma}}$ [3] is the estimated covariance matrix. All results are averaged on 100 Monte Carlo trials.

Fig. 1 illustrates the FRE across various sparsity levels with $d = 100$, $s^* = \{80, 120, 200, 240\}$. These figures highlights the relationship between FRE and the number of measurements m or observed data n . It is evident the FRE diminishes as m (n resp.) increases, indicating improved estimation accuracy with more measurements (observed data, resp.). Fig. 2 illustrates the oracle rate of "Sprandsym" matrix under different configurations of d and m. As the measurement size increases, the FAE decreases. The curve distinctly presents a turning point, signifying the emergence of sparse eigenvalues once m surpasses a specific threshold. Additionally, a comparative analysis between our proposed covariance matrix sensing estimator and a conventional estimator utilizing an ℓ_1 penalty. The comparison reveals that our MCP-based estimator consistently outperforms the ℓ_1-penalty-based estimator in terms of the Frobenius norm error. This supports our theoretical proposition that using a non-convex penalty can significantly reduce the error in covariance matrix sensing.

[3]In the whole Section VI, we use $\widehat{\boldsymbol{\Sigma}}$ to denote a generic covariance estimator, which can be the estimation results obtained by various algorithms and estimators.

VII. Conclusion and Discussion

In this paper, we present a comprehensive study on covariance matrix sensing, with a particular focus on the high-dimensional sparse covariance matrices estimation through the quadratic measurements model. This model is particularly suited for scenarios with constrained processing capabilities and limited memory, such as real-time data acquisition devices. Our findings demonstrates that sparse covariance matrices can be effectively and accurately reconstructed using a minimal set of quadratic measurements with bottommost storage requirements. We provide rigorous theoretical proof and empirical evidence to support these conclusions. Moreover, our approach can be readily extended to covariance sensing problems based on bilinear sampling model [47] and involving non-zero means.

REFERENCES

[1] M. Pourahmadi, *High-dimensional covariance estimation: with high-dimensional data*. John Wiley & Sons, 2013, vol. 882.

[2] A. Bose and M. Bhattacharjee, *Large covariance and autocovariance matrices*. Chapman and Hall/CRC, 2018.

[3] H. Tsukuma and T. Kubokawa, *Shrinkage estimation for mean and covariance matrices*. Springer, 2020.

[4] H. Abdi and L. J. Williams, "Principal component analysis," *Wiley Interdiscip. Rev. Comput. Stat.*, vol. 2, no. 4, pp. 433–459, 2010.

[5] R. L. Gorsuch, *Factor analysis: Classic edition*. Routledge, 2014.

[6] R. A. Monzingo and T. W. Miller, *Introduction to adaptive arrays*. Scitech publishing, 2004.

[7] L. K. Chan, J. Karceski, and J. Lakonishok, "On portfolio optimization: Forecasting covariances and choosing the risk model," *Rev. Financ. Stud.*, vol. 12, no. 5, pp. 937–974, 1999.

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

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

[10] P. J. Bickel and E. Levina, "Covariance regularization by thresholding," *Ann. Stat.*, vol. 36, no. 6, pp. 2577 – 2604, 2008.

[11] A. J. Rothman, E. Levina, and J. Zhu, "Generalized thresholding of large covariance matrices," *J. Amer. Stat. Assoc.*, vol. 104, no. 485, pp. 177–186, 2009.

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

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

[14] L. Xue, S. Ma, and H. Zou, "Positive-definite ℓ_1-penalized estimation of large covariance matrices," *J. Amer. Stat. Assoc.*, vol. 107, no. 500, pp. 1480–1491, 2012.

[15] M. Abt and W. J. Welch, "Fisher information and maximum-likelihood estimation of covariance parameters in Gaussian stochastic processes," *Can. J. Stat.*, vol. 26, no. 1, pp. 127–137, 1998.

[16] 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.

[17] 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.

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

[19] E. J. Candes and T. Tao, "Near-optimal signal recovery from random projections: Universal encoding strategies?" *IEEE Trans. Inf. Theory*, vol. 52, no. 12, pp. 5406–5425, 2006.

[20] 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.

[21] Z.-Q. He, Z.-P. Shi, and L. Huang, "Covariance sparsity-aware doa estimation for nonuniform noise," *Digit. Signal Process.*, vol. 28, pp. 75–81, 2014.

[22] M. Tsang, "Semiparametric estimation for incoherent optical imaging," *Phys. Rev. Res.*, vol. 1, no. 3, p. 033006, 2019.

[23] A. M. Brickman, C. Habeck, E. Zarahn, J. Flynn, and Y. Stern, "Structural mri covariance patterns associated with normal aging and neuropsychological functioning," *Neurobiol. Aging*, vol. 28, no. 2, pp. 284–295, 2007.

[24] M. Loretan and P. C. Phillips, "Testing the covariance stationarity of heavy-tailed time series: An overview of the theory with applications to several financial datasets," *J. Empir. Finance*, vol. 1, no. 2, pp. 211–248, 1994.

[25] Y. L. Polo, Y. Wang, A. Pandharipande, and G. Leus, "Compressive wide-band spectrum sensing," in *Proceedings of 2009 IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, 2009, pp. 2337–2340.

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

[27] D. D. Arianananda and G. Leus, "Compressive wideband power spectrum estimation," *IEEE Transactions on Signal Processing*, vol. 60, no. 9, pp. 4775–4789, 2012.

[28] D. D. Arianananda, D. Romero, and G. Leus, "Compressive periodogram reconstruction using uniform binning," *IEEE Trans. Signal Process.*, vol. 63, no. 16, pp. 4149–4164, 2015.

[29] 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.

[30] 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.

[31] 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.

[32] E. J. Candes and T. Tao, "Decoding by linear programming," *IEEE Transactions on Information Theory*, vol. 51, no. 12, pp. 4203–4215, 2005.

[33] S. A. van de Geer and P. Bühlmann, "On the conditions used to prove oracle results for the Lasso," *Electron. J. Stat.*, vol. 3, no. none, pp. 1360 – 1392, 2009.

[34] 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.

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

[36] 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.

[37] 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.

[38] X. Li, L. Yang, J. Ge, J. Haupt, T. Zhang, and T. Zhao, "On quadratic convergence of DC proximal Newton algorithm in nonconvex sparse learning," *Advances in Neural Information Processing Systems (NeurIPS)*, vol. 30, 2017.

[39] H. Liu, L. Wang, and T. Zhao, "Sparse covariance matrix estimation with eigenvalue constraints," *J. Comput. Graph. Stat.*, vol. 23, no. 2, pp. 439–459, 2014.

[40] 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.

[41] P. J. Bickel, Y. Ritov, and A. B. Tsybakov, "Simultaneous analysis of Lasso and Dantzig selector," *Ann. Stat.*, vol. 37, no. 4, pp. 1705 – 1732, 2009.

[42] P.-L. Loh and M. J. Wainwright, "Regularized M-estimators with nonconvexity: statistical and algorithmic theory for local optima," *J. Mach. Learn. Res.*, vol. 16, no. 1, pp. 559–616, jan 2015.

[43] S. N. Negahban, P. Ravikumar, M. J. Wainwright, and B. Yu, "A unified framework for high-dimensional analysis of M-estimators with decomposable regularizers," *Stat. Sci.*, vol. 27, no. 4, pp. 538 – 557, 2012.

[44] G. Raskutti, M. J. Wainwright, and B. Yu, "Restricted eigenvalue properties for correlated gaussian designs," *J. Mach. Learn. Res.*, vol. 11, pp. 2241–2259, 2010.

[45] Z. Wang, H. Liu, and T. Zhang, "Optimal computational and statistical rates of convergence for sparse nonconvex learning problems," *Ann. Stat.*, vol. 42, no. 6, p. 2164, 2014.

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

[47] G. Dasarathy, P. Shah, B. N. Bhaskar, and R. Nowak, "Covariance sketching," in *Proceedings of 2012 50th Annual Allerton Conference on Communication, Control, and Computing (Allerton)*, 2012, pp. 1026–1033.