# Optimal Compressive Covariance Sketching via Rank-One Sampling

Wenbin Wang
*School of Information and Science Technology*
*ShanghaiTech University*
Shanghai, China
wangwb2023@shanghaitech.edu.cn

Ziping Zhao
*School of Information and Science Technology*
*ShanghaiTech University*
Shanghai, China
zipingzhao@shanghaitech.edu.cn

***Abstract*—In this paper, we study the problem of compressive covariance sketching, where the goal is to compress a high-dimensional data stream and recover its covariance matrix from a limited number of compressed measurements. This problem is particularly relevant in scenarios where the data evolves rapidly or where sensing devices are constrained by limited computational and storage resources. We consider the rank-one sampling model under the assumption that the underlying covariance matrix is sparse. To estimate the covariance matrix, we propose a regularized least-squares estimator that incorporates nonconvex sparsity-inducing penalties. To compute the estimator efficiently, we develop a multi-stage convex relaxation algorithm based on the majorization-minimization (MM) framework. Each subproblem in the MM scheme is approximately solved via a proximal Newton method, which enjoys a locally quadratic convergence rate. We establish that the proposed estimator achieves the oracle statistical convergence rate after a sufficient number of iterations. Numerical experiments corroborate our theoretical findings and demonstrate the effectiveness of the proposed approach.**

***Index Terms*—Compressive sensing, rank-one measurements, quadratic sampling, nonconvex penalty, majorization-minimization, sparsity, positive definite.**

## I. Introduction

With the rapid advancement of communication technologies, there is a growing demand for processing ultra-wideband signals, which poses new challenges in signal sampling and processing [1]–[4]. Digital signal processing typically requires the conversion of analog signals into discrete samples via sampling and quantization, a process commonly implemented using analog-to-digital converters (ADCs) [5]–[8]. To accommodate signals with higher bandwidths, increasingly high sampling rates are necessary. However, this demand leads to substantial power consumption in ADCs, which remains a fundamental bottleneck in modern systems [9], [10]. To mitigate this challenge, compressive sampling (CS) has emerged as a powerful framework that enables signal recovery from sub-Nyquist rate samples by exploiting signal sparsity [11]–[15]. While conventional CS techniques primarily aim to reconstruct the signal itself, there are many applications where the objective shifts toward estimating second-order statistics, such as the covariance matrix, from sub-Nyquist samples [16]–[19]. In this paper, we investigate the problem of estimating a covariance matrix from compressed rank-one measurements, a setting commonly referred to as compressive covariance sketching [20]. More specifically, we focus on the recovery of a sparse covariance matrix from a limited number of compressed observations. Our study is guided by two fundamental questions: 1) Can we design sketching vectors such that the resulting compressive sketches exhibit favorable statistical properties, enabling accurate estimation of the covariance matrix from a small number of measurements? 2) How can we develop an efficient algorithm for compressive covariance sketching, and what statistical guarantees can be established for the resulting estimator?

To address the first question, we adopt the sparse eigenvalue condition, which, to the best of our knowledge, is among the weakest known conditions sufficient to guarantee accurate recovery of a sparse covariance matrix. Our theoretical analysis shows that exact recovery is possible under this condition for a broad class of sub-Gaussian sketching vectors. For the second question, we propose a regularized least-squares framework that incorporates nonconvex sparsity-inducing penalties to promote structured recovery. To solve the resulting optimization problem efficiently, we develop a multi-stage convex relaxation algorithm based on the majorization-minimization (MM) framework [21], [22]. We further establish that, after sufficient iterations, the proposed estimator achieves the optimal estimation rate as if the true support were known a priori. This result highlights the statistical efficiency of the proposed method under mild regularity conditions.

## II. Rank-One Sampling

Consider a sequence of $T$ samples $\{\boldsymbol{x}_t\}_{t=1}^T$ drawn from a zero-mean distribution in $\mathbb{R}^d$ with covariance matrix $\boldsymbol{\Sigma}^\star$. Let $\{\boldsymbol{a}_i\}_{i=1}^m$ be a set of $m$ random sketching vectors. Each vector $\boldsymbol{a}_i$ is applied to all samples, yielding a collection of projected measurements $\left\{\boldsymbol{a}_i^\top \boldsymbol{x}_t\right\}_{t=1}^T$. For each sketching vector $\boldsymbol{a}_i$, the squared projections are averaged across time to produce the corresponding compressed measurement:

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

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

where $\boldsymbol{S} = \frac{1}{T}\sum_{t=1}^{T} \boldsymbol{x}_t \boldsymbol{x}_t^\top$ denotes the sample covariance matrix[1], and $\eta_i$ represents a noise term. The quantity $y_i$ thus serves as a compressed sketch of the true covariance matrix $\boldsymbol{\Sigma}^\star$. Each inner product $\left\langle \boldsymbol{a}_i \boldsymbol{a}_i^\top, \boldsymbol{S} \right\rangle$ is referred to as a rank-one measurement or a quadratic measurement. By leveraging the Kronecker product, each measurement $y_i$ can be equivalently written as

$$y_i = (\boldsymbol{a}_i \otimes \boldsymbol{a}_i)^\top \operatorname{vec}(\boldsymbol{S}) + \eta_i,$$

where $\operatorname{vec}(\boldsymbol{S})$ denotes the vectorization of the sample covariance matrix $\boldsymbol{S}$, obtained by stacking its columns into a single vector.

For notational convenience, let $\boldsymbol{y} = [y_1, \ldots, y_m]^\top$ denote the vector of measurements, and $\boldsymbol{\eta} = [\eta_1, \ldots, \eta_m]^\top$ the corresponding noise vector. Define the equivalent sketching matrix $\boldsymbol{A} = \left[\begin{array}{ccc} (\boldsymbol{a}_1 \otimes \boldsymbol{a}_1) & \cdots & (\boldsymbol{a}_m \otimes \boldsymbol{a}_m) \end{array}\right]^\top$. The measurement model can then be compactly expressed as

$$\boldsymbol{y} = \boldsymbol{A}\operatorname{vec}(\boldsymbol{S}) + \boldsymbol{\eta} = \mathcal{A}(\boldsymbol{S}) + \boldsymbol{\eta},$$

where $\mathcal{A} : \mathbb{R}^{d\times d} \to \mathbb{R}^m$ is the linear operator defined by $\mathcal{A}(\boldsymbol{S}) = \boldsymbol{A}\operatorname{vec}(\boldsymbol{S})$.

## III. Recovery via Nonconvex Optimization

### A. Proposed Nonconvex Estimator

In general, recovering the covariance matrix $\boldsymbol{\Sigma}^\star$ from $m < \frac{d(d+1)}{2}$ measurements is inherently ill-posed unless additional low-dimensional structure is imposed. A common and effective assumption is sparsity, which reduces the parameter space by encouraging most entries of $\boldsymbol{\Sigma}^\star$ to be zero. In this work, we adopt the sparsity assumption and propose to recover $\boldsymbol{\Sigma}^\star$ by solving the following optimization problem:

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

where the first term penalizes the empirical error, the second term imposes strict positive definiteness via a log-determinant barrier with parameter $\tau > 0$, and the third term introduces a sparsity-promoting regularization based on a nonconvex penalty function $p_\lambda$ with tuning parameter $\lambda > 0$. We impose the following conditions on $p_\lambda$.

**Assumption 1.** *The function $p_\lambda$ defined on $[0, +\infty)$ satisfies:*

(a) *$p_\lambda(t)$ is non-decreasing with $p_\lambda(0) = 0$, and is differentiable almost everywhere on $(0, \infty)$;*
(b) *For all $t_1 \geq t_2 \geq 0$, it holds that $0 \leq p'_\lambda(t_1) \leq p'_\lambda(t_2) \leq \lambda$ and $\lim_{t\to 0} p'_\lambda(t) = \lambda$;*
(c) *There exists an $\alpha > 0$ such that $p'_\lambda(t) = 0$ for $t \geq \alpha\lambda$.*

These conditions, which ensure sparsity and unbiasedness, are consistent with those considered in [23]–[26] and are satisfied by several nonconvex regularizers, including smoothly clipped absolute deviation (SCAD) penalty [23], minimax concave penalty (MCP) [27], and capped $\ell_1$ regularizer [28].

[1]As the number of samples increases, the sample covariance matrix $\boldsymbol{S}$ rapidly converges to the true covariance matrix $\boldsymbol{\Sigma}^\star$.

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

**Input:** $\{y_i, \boldsymbol{a}_i\}_{i=1}^m$, $\tau$, $\lambda$;
**Initialize** $\boldsymbol{\Sigma}^0 = \boldsymbol{I}$
**for** $k = 1, 2, \ldots, K$ **do**
  $\Lambda_{ij}^k = p'_\lambda(|\Sigma_{ij}^{k-1}|)$;
  obtain $\boldsymbol{\Sigma}^k$ via solving (2);
  $k = k + 1$;
**end**
**Output:** $\boldsymbol{\Sigma}^K$

### B. Optimization Algorithm

We propose a multi-stage convex relaxation algorithm based on the MM framework [21], [22] to solve problem (1). In each stage of the MM algorithm, a convex subproblem is solved using a proximal Newton method [29]–[31].

*1) The Multi-Stage Convex Relaxation Algorithm:* Define $f(\boldsymbol{\Sigma}) = \frac{1}{2m}\left\|\boldsymbol{y} - \mathcal{A}(\boldsymbol{\Sigma})\right\|_2^2 - \tau \log\det \boldsymbol{\Sigma}$. At each iteration of the MM algorithm, the nonconvex penalty term $\sum_{i,j} p_\lambda\left(|\Sigma_{ij}|\right)$ is approximated by a weighted $\ell_1$-norm, which serves as a convex surrogate. Specifically, at the $k$-th stage, we solve the following convex optimization problem:

$$\min_{\boldsymbol{\Sigma}\succ\boldsymbol{0}} f(\boldsymbol{\Sigma}) + \left\|\boldsymbol{\Lambda}^k \odot \boldsymbol{\Sigma}\right\|_1, \quad (2)$$

where $\boldsymbol{\Lambda}^k$ is a weight matrix whose entries are given by $\Lambda_{ij}^k = p'_\lambda(|\Sigma_{ij}^{k-1}|)$, with $\boldsymbol{\Sigma}^{k-1}$ denoting the solution obtained from the $(k-1)$-th stage, and $\odot$ denotes the Hadamard (elementwise) product. According to the Karush-Kuhn-Tucker conditions, the optimal solution, denoted by $\widehat{\boldsymbol{\Sigma}}^k$, to the convex subproblem in (2) satisfies

$$\nabla f(\widehat{\boldsymbol{\Sigma}}^k) + \boldsymbol{\Lambda}^k \odot \widehat{\boldsymbol{\Xi}}^k = \boldsymbol{0}, \text{ for some } \widehat{\boldsymbol{\Xi}}^k \in \partial\|\widehat{\boldsymbol{\Sigma}}^k\|_1,$$

where $\partial\left\|\cdot\right\|_1$ denotes the subdifferential of the $\ell_1$-norm. Since a closed-form solution for $\widehat{\boldsymbol{\Sigma}}^k$ is not available, we instead compute an approximate solution $\boldsymbol{\Sigma}^k$ that is *$\varepsilon$-optimal*.

**Definition 2.** For a pre-specified tolerance level $\varepsilon$, we say $\boldsymbol{\Sigma}^k$ is an $\varepsilon$-optimal solution if

$$\min_{\boldsymbol{\Xi}^k \in \partial\left\|\boldsymbol{\Sigma}^k\right\|_1} \max_{i,j} \left|\left(\nabla f(\boldsymbol{\Sigma}^k) + \boldsymbol{\Lambda}^k \odot \boldsymbol{\Xi}^k\right)_{ij}\right| \leq \varepsilon.$$

The overall optimization algorithm is summarized in Algorithm 1, where we adopt a simple initialization, specifically setting $\boldsymbol{\Sigma}^0 = \boldsymbol{I}$.

*2) Proximal Newton Algorithm:* In this work, we employ the proximal Newton algorithm to solve the convex subproblem in (2). Specifically, the algorithm computes a Newton direction that serves as a descent direction, followed by a line search to determine a suitable step size that guarantees a sufficient decrease in the objective function.

**Algorithm 2:** Proximal Newton Algorithm With Backtracking Line Search.

**Input:** $\boldsymbol{\Sigma}^{k-1}$, $\boldsymbol{\Lambda}^k$, $\varepsilon$;
**Initialize** $t=0$, $\boldsymbol{\Sigma}_t=\boldsymbol{\Sigma}^{k-1}$, $\mu=0.8$, $\alpha=0.3$;
**repeat**
  $\boldsymbol{\Sigma}_{t+\frac{1}{2}} \in \arg\min_{\boldsymbol{\Sigma}\succ\mathbf{0}} \widetilde{f}_t(\boldsymbol{\Sigma}) + \|\boldsymbol{\Lambda}\odot\boldsymbol{\Sigma}\|_1$;
  $\boldsymbol{\Delta}_t = \boldsymbol{\Sigma}_{t+\frac{1}{2}} - \boldsymbol{\Sigma}_t$;
  $\delta_t = \langle\nabla f(\boldsymbol{\Sigma}_t), \boldsymbol{\Delta}_t\rangle - \|\boldsymbol{\Lambda}^k\odot\boldsymbol{\Sigma}_t\|_1 + \|\boldsymbol{\Lambda}^k\odot(\boldsymbol{\Sigma}_t+\boldsymbol{\Delta}_t)\|_1$;
  $\beta=1$, $q=0$;
  **repeat**
    $\beta=\mu^q$, $q=q+1$;
    **if** $\boldsymbol{\Sigma}_t+\beta\boldsymbol{\Delta}_t \preceq \mathbf{0}$ **then**
      continue;
    **end**
  **until** $\bar{F}(\boldsymbol{\Sigma}_t+\beta\boldsymbol{\Delta}_t) \le \bar{F}(\boldsymbol{\Sigma}_t)+\alpha\beta\delta_t$;
  $\boldsymbol{\Sigma}_{t+1}=\boldsymbol{\Sigma}_t+\beta\boldsymbol{\Delta}_t$;
  $t=t+1$;
**until** $\max_{i,j}\left|\left(\nabla f(\boldsymbol{\Sigma}_{t+1}) + \boldsymbol{\Lambda}^k\odot\boldsymbol{\Xi}^k\right)_{ij}\right| \le \varepsilon$;
**Output:** $\boldsymbol{\Sigma}^K=\boldsymbol{\Sigma}_{t+1}$

We denote the iteration index within the $k$-th stage by $t$. For brevity, we omit the index $k$. Consider the second-order Taylor expansion of $f(\boldsymbol{\Sigma})$ around $\boldsymbol{\Sigma}_t$:

$$\begin{aligned}&\widetilde{f}_t(\boldsymbol{\Sigma})\\ =&f(\boldsymbol{\Sigma}_t)+\langle\nabla f(\boldsymbol{\Sigma}_t), \boldsymbol{\Sigma}-\boldsymbol{\Sigma}_t\rangle + \frac{1}{2}\|\boldsymbol{\Sigma}-\boldsymbol{\Sigma}_t\|^2_{\nabla^2 f(\boldsymbol{\Sigma}_t)},\end{aligned}$$

where $\|\boldsymbol{X}\|_{\boldsymbol{M}} = \sqrt{\mathrm{vec}(\boldsymbol{X})^\top \boldsymbol{M}\,\mathrm{vec}(\boldsymbol{X})}$. The proximal Newton update is derived as

$$\boldsymbol{\Sigma}_{t+\frac{1}{2}} \in \arg\min_{\boldsymbol{\Sigma}\succ\mathbf{0}} \widetilde{f}_t(\boldsymbol{\Sigma}) + \|\boldsymbol{\Lambda}\odot\boldsymbol{\Sigma}\|_1.$$

We define $\bar{F}(\boldsymbol{\Sigma}) = f(\boldsymbol{\Sigma}) + \|\boldsymbol{\Lambda}\odot\boldsymbol{\Sigma}\|_1$. The Newton direction for $\bar{F}(\boldsymbol{\Sigma})$ is computed as $\boldsymbol{\Delta}_t = \boldsymbol{\Sigma}_{t+\frac{1}{2}} - \boldsymbol{\Sigma}_t$.

Then, we perform a backtracking line search to select a step size $\beta \in (0,1]$ that ensures a sufficient decrease in $\bar{F}(\boldsymbol{\Sigma})$. Starting with a fixed constant $\mu \in (0.5,1)$ and updating $\beta=\mu^q$ from $q=0$ with a constant decrease rate, we find the smallest non-negative integer $q$ for which the Armijo condition [32] holds:

$$\bar{F}(\boldsymbol{\Sigma}_t+\beta\boldsymbol{\Delta}_t) \le \bar{F}(\boldsymbol{\Sigma}_t)+\alpha\beta\delta_t,$$

where $\alpha \in (0,0.5)$ and $\delta_t = \langle\nabla f(\boldsymbol{\Sigma}_t), \boldsymbol{\Delta}_t\rangle - \|\boldsymbol{\Lambda}\odot\boldsymbol{\Sigma}_t\|_1 + \|\boldsymbol{\Lambda}\odot(\boldsymbol{\Sigma}_t+\boldsymbol{\Delta}_t)\|_1$. Finally, we update $\boldsymbol{\Sigma}_{t+1}=\boldsymbol{\Sigma}_t+\beta\boldsymbol{\Delta}_t$. The overall proximal Newton algorithm is summarized in Algorithm 2.

## IV. Theoretical Analysis

In this section, we present the theoretical results. Define the support of $\boldsymbol{\Sigma}^\star$ as $\mathcal{S}=\{(i,j) \mid \Sigma^\star_{ij}\neq 0\}$, with $s$ representing its cardinality, i.e., $s=|\mathcal{S}|$.

### A. Assumptions

We begin by introducing some preliminaries, including key definitions and assumptions that will be used throughout the analysis.

**Assumption 3.** *There exist universal constants $\kappa$, $\alpha$, and $\xi$ such that $0<\frac{1}{\kappa}\le\lambda_{\min}(\boldsymbol{\Sigma}^\star)\le\lambda_{\max}(\boldsymbol{\Sigma}^\star)\le\kappa<\infty$.*

**Assumption 4.** *The true covariance matrix $\boldsymbol{\Sigma}^\star$ satisfies $\|\boldsymbol{\Sigma}^\star_{\mathcal{S}}\|_{\min} = \min_{(i,j)\in\mathcal{S}}\left|\Sigma^\star_{ij}\right| \ge (\alpha+\xi)\lambda$, where $\kappa\ge 1$, $\alpha$ is from Assumption 1, and $\xi\in(0,\alpha)$ satisfies $p'_\lambda(\xi\lambda)\ge\frac{\lambda}{2}$.*

**Definition 5.** 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 6.** *The sketching vectors $\{\boldsymbol{a}_i\}_{i=1}^m$ are independent and identically distributed (i.i.d.) sub-Gaussian random variables with zero mean and identity covariance; and the noise $\{\eta_i\}_{i=1}^m$ are i.i.d. sub-exponential random variables with mean $0$ and variance proxy $\sigma^2$.*

Assumption 6 implies that each row of the design matrix $\boldsymbol{A}$ consists of i.i.d. sub-exponential random variables. For $m=\mathcal{O}(s\log^2(d/s))$, within $\mathcal{B}(\boldsymbol{\Sigma}^\star, \frac{\rho^-}{4\tau\kappa})$, there exist constants $\rho^-$ and $\rho^+$ such that $0<\rho^-\le\rho^+<\infty$ with probability at least $1-c_1\exp(-c_2\sqrt{m})$ for $c_1,c_2>0$ [33].

### B. Statistical and Computational Analysis

**Theorem 7** (Contraction Property)**.** *Suppose Assumptions 1 to 6 hold. Then the $\varepsilon$-optimal solution $\boldsymbol{\Sigma}^K$ from Algorithm 1 is bounded by:*

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

*for $1\le k\le K$, where $\delta\in(0,1)$ is the contraction factor. If $\boldsymbol{x}$ be a sub-Gaussian random vector with mean zero and covariance $\boldsymbol{\Sigma}^\star$, $\{\boldsymbol{x}_i\}_{i=1}^n$ be a collection of i.i.d. samples drawn from $\boldsymbol{x}$, $\lambda\asymp\sqrt{\frac{\log d}{mn}}$, $\tau\lesssim\sqrt{\frac{1}{mn}}\left\|(\boldsymbol{\Sigma}^\star)^{-1}\right\|^{-1}_{\max}$, $\varepsilon\lesssim\sqrt{\frac{1}{mn}}$, and $K\gtrsim\log(\lambda\sqrt{mn})\gtrsim\log\log d$, then the $\varepsilon$-optimal solution $\boldsymbol{\Sigma}^K$ satisfies $\|\boldsymbol{\Sigma}^K-\boldsymbol{\Sigma}^\star\|_{\mathrm{F}}=\mathcal{O}_p\left(\sqrt{\frac{s}{mn}}\right)$ with high probability.*

Theorem 7 elaborates the estimation error between the $\varepsilon$-optimal solution $\boldsymbol{\Sigma}^k$ and the ground truth $\boldsymbol{\Sigma}^\star$ is constrained by three primary factors: the oracle rate[2], the optimization error, and a contraction term.

[2]The oracle estimator $\hat{\boldsymbol{\Sigma}}^O$ is defined as $\hat{\boldsymbol{\Sigma}}^O = \arg\min_{\boldsymbol{\Sigma}_{\mathcal{S}^c}=\mathbf{0}} f(\boldsymbol{\Sigma})$.

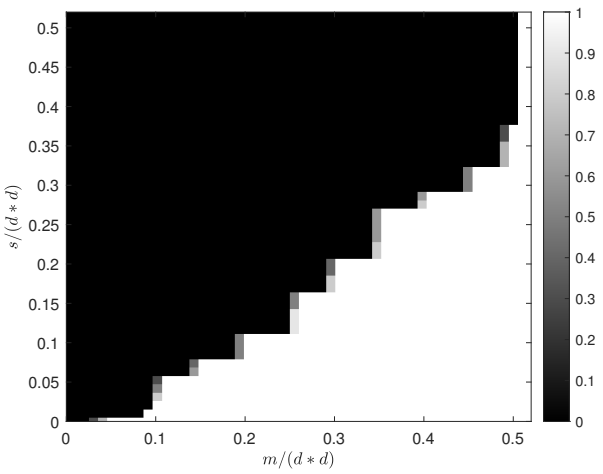


Fig. 1. The rate of successful covariance reconstruction when $d = 100$.

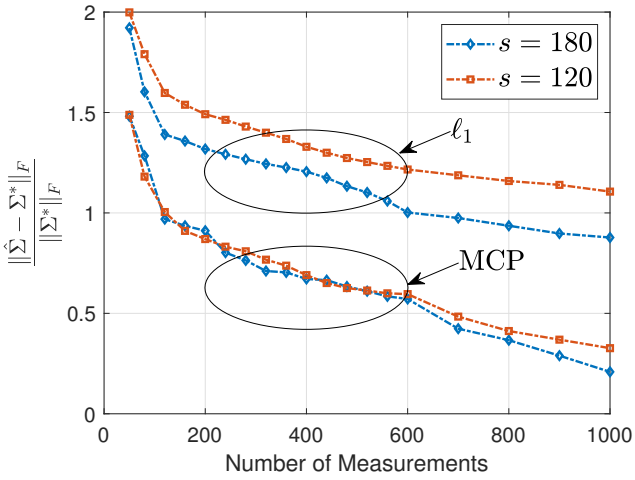


Fig. 2. The FRE of the estimated covariance matrices for different sparsity levels with $\gamma = 10^{-1}$.

## V. Simulation Results

We use the MCP penalty, defined as

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

with $b = 2$ across all trials. The regularization parameters $\tau$ and $\lambda$ are selected via five-fold cross-validation. The ground-truth covariance matrix $\boldsymbol{\Sigma}^\star$ is generated using the built-in `sprandsym` function in MATLAB with $s$ nonzero entries. We draw $n = 50$ independent samples from the multivariate normal distribution $\mathcal{N}(0, \boldsymbol{\Sigma}^\star)$, and the noise variables $\eta_i$ are sampled from a sub-exponential distribution with scale parameter $\gamma$, i.e., $\eta_i \sim \gamma \cdot \mathcal{N}(0, 1)$. To evaluate recovery performance, we measure the success probability as visualized in the color-coded matrix in Fig. 1. To reduce the impact of limited sample size, we directly apply sketching to the true covariance matrix $\boldsymbol{\Sigma}^\star$. A recovery is considered successful if the relative Frobenius error satisfies

$$\frac{\|\boldsymbol{\Sigma} - \boldsymbol{\Sigma}^\star\|_{\mathrm{F}}}{\|\boldsymbol{\Sigma}^\star\|_{\mathrm{F}}} \leq 10^{-3}.$$

Fig. 2 compares the proposed estimator with the $\ell_1$-norm-based method from [18] under a consistent noise level $\gamma = 10^{-1}$. As the number of measurements increases, the recovery error decreases, and our method consistently outperforms the $\ell_1$-based estimator.

## VI. Conclusions

In this paper, we have investigated the compressive covariance sketching problem. We have proposed a nonconvex-based estimator based on a quadratic measurement model and developed an MM-based algorithm for efficient estimation. We have shown that, for a broad class of sub-Gaussian sketching vectors, exact covariance recovery is achievable with theoretical performance guarantees.

## References


[1] L. Yang and G. B. Giannakis, "Ultra-wideband communications: an idea whose time has come," *IEEE Signal Process. Mag.*, vol. 21, no. 6, pp. 26–54, 2004.

[2] M. Ghavami, L. Michael, and R. Kohno, *Ultra wideband signals and systems in communication engineering*. John Wiley & Sons, 2007.

[3] D. D. Ariananda and G. Leus, "Wideband power spectrum sensing using sub-Nyquist sampling," in *2011 IEEE 12th International Workshop on Signal Processing Advances in Wireless Communications*. IEEE, 2011, pp. 101–105.

[4] ——, "Compressive wideband power spectrum estimation," *IEEE Trans. Signal Process.*, vol. 60, no. 9, pp. 4775–4789, 2012.

[5] A. V. Oppenheim, "Applications of digital signal processing," *Englewood Cliffs*, 1978.

[6] R. A. Roberts and C. T. Mullis, *Digital signal processing*. Addison-Wesley Longman Publishing Co., Inc., 1987.

[7] B. Mulgrew, P. Grant, and J. Thompson, "Digital signal processing: concepts and applications," 2002.

[8] P. S. Diniz, E. A. Da Silva, and S. L. Netto, *Digital signal processing: system analysis and design*. Cambridge University Press, 2010.

[9] R. H. Walden, "Analog-to-digital converter survey and analysis," *IEEE J. Sel. Areas in Commun.*, vol. 17, no. 4, pp. 539–550, 1999.

[10] A. Sabharwal, P. Schniter, D. Guo, D. W. Bliss, S. Rangarajan, and R. Wichman, "In-band full-duplex wireless: Challenges and opportunities," *IEEE J. Sel. Areas in Commun.*, vol. 32, no. 9, pp. 1637–1652, 2014.

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

[12] E. J. Candès, J. Romberg, and T. Tao, "Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information," *IEEE Trans. Inf. Theory*, vol. 52, no. 2, pp. 489–509, 2006.

[13] 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.

[14] 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.

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

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

[17] 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.

[18] 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.
[19] G. B. Giannakis, "Cyclostationary signal analysis," in *Digital Signal Processing Fundamentals*. CRC press, 2017, pp. 433–464.
[20] D. Romero, D. D. Ariananda, 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.
[21] D. R. Hunter and K. Lange, "A tutorial on MM algorithms," *Am. Stat.*, vol. 58, no. 1, pp. 30–37, 2004.
[22] 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.
[23] 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.
[24] 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, 2015.
[25] J. Fan, H. Liu, Q. Sun, and T. Zhang, "I-LAMM for sparse learning: Simultaneous control of algorithmic complexity and statistical error," *Ann. Statist.*, vol. 46, no. 2, pp. 814–841, 2018.
[26] 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.
[27] C.-H. Zhang, "Nearly unbiased variable selection under minimax concave penalty," *Ann. Statist.*, vol. 38, no. 2, pp. 894–942, 2010.
[28] ——, "Nearly unbiased variable selection under minimax concave penalty," *Ann. Stat.*, vol. 38, pp. 894–942, 2010.
[29] J. D. Lee, Y. Sun, and M. A. Saunders, "Proximal newton-type methods for minimizing composite functions," *SIAM J. Optimiz.*, vol. 24, no. 3, pp. 1420–1443, 2014.
[30] M.-C. Yue, Z. Zhou, and A. M.-C. So, "Inexact regularized proximal newton method: Provable convergence guarantees for non-smooth convex minimization without strong convexity," *arXiv: Optimization and Control*, 2016.
[31] 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," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 30, 2017, pp. 2742–2752.
[32] L. Armijo, "Minimization of functions having Lipschitz continuous first partial derivatives," *Pacific J. Math.*, vol. 16, no. 1, pp. 1–3, 1966.
[33] 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.