Robust and Universal Covariance Estimation from Quadratic Measurements via Convex Programming

Yuxin Chen
EE, Stanford University
Email: yxchen@stanford.edu

Yuejie Chi
ECE, Ohio State University
Email: chi.97@osu.edu

Andrea J. Goldsmith
EE, Stanford University
Email: andrea@wsl.stanford.edu

***Abstract*—This paper considers the problem of recovering the covariance matrix of a stream of high-dimensional data instances from a minimal number of stored measurements. We develop a quadratic random sampling method based on rank-one measurements of the covariance matrix, which serves as an efficient *covariance sketching* scheme for processing data streams. This also allows modeling of phaseless measurements that arise in high-frequency wireless communication and signal processing applications.**

We propose to recover the covariance matrix from the above quadratic measurements via convex relaxation with respect to the presumed parsimonious covariance structure. We show that in the absence of noise, exact and universal recovery of low-rank or Toeplitz low-rank covariance matrices can be achieved as soon as the number of stored measurements exceeds the fundamental sampling limit. The convex programs are also robust to noise and imperfect structural assumptions. Our analysis is established upon a novel notion called the mixed-norm restricted isometry property (RIP-ℓ_2/ℓ_1), as well as the conventional RIP-ℓ_2/ℓ_2 for near-isotropic and bounded measurements. Our results improve upon best-known phase retrieval performance guarantees with a significantly simpler approach. Numerical results are provided to demonstrate the practical applicability of our technique.

***Index Terms*—covariance estimation, sketching, quadratic measurements, convex programming, phaseless measurements**

I. Introduction

Accurate estimation of second-order statistics of data streams and random processes is of ever-growing importance to various applications that exhibit high dimensionality. Covariance estimation is the cornerstone of modern statistical analysis and information processing, as the covariance matrix constitutes the sufficient statistic to many signal processing tasks, and is particularly crucial for extracting reduced-dimension representation of the data of interest.

In resource-constrained environments, there might be limited memory and computation power available at the data acquisition devices to take a complete snapshot of the system or to store the entire data for problems of high dimensionality. Therefore it is highly desirable to estimate the covariance matrix of the data with minimal storage and low computational complexity. This is in general not possible unless appropriate structural assumptions are incorporated. Fortunately, a broad class of high-dimensional objects indeed possesses low-dimensional structures, and the intrinsic dimension of the covariance matrix can be far smaller than its ambient dimension. For different types of data, the covariance matrix may exhibit different structures; two of the most widely considered structures, studied in this paper, are listed below.

- *Low Rank:* The covariance matrix is low-rank, which occurs when a small number of components accounts for most of the variability in the data [1]. Low-rank covariance matrices arise in applications such as traffic data monitoring, collaborative filtering, and metric learning.
- *Toeplitz and Low Rank:* The covariance matrix is simultaneously low-rank and Toeplitz, which arises when each instance is drawn from a stationary process. Recovery of Toeplitz low-rank covariance matrices, often related to spectral estimation, is crucial in many fields including wireless communications (e.g. detecting spectral holes in cognitive radio networks) and array signal processing [2].

In this paper, we wish to recover an unknown covariance matrix $\mathbf{\Sigma} \in \mathbb{R}^{n\times n}$ with the above covariance structures from the following rank-one measurements:

$$y_i = \boldsymbol{a}_i^{\top} \mathbf{\Sigma} \boldsymbol{a}_i + \eta_i, \quad i = 1, \ldots, m, \tag{1}$$

where $\boldsymbol{y} := \{y_i\}_{i=1}^m$ denotes the measurements, $\boldsymbol{a}_i \in \mathbb{R}^n$ represents the sensing vector, $\boldsymbol{\eta} := \{\eta_i\}_{i=1}^m$ denotes the noise term, and m is the number of rank-one measurements. The measurements $\boldsymbol{a}_i^T \mathbf{\Sigma} \boldsymbol{a}_i$'s are quadratic in data instances and are therefore referred to as *quadratic measurements*. In practice, the number of measurements is constrained by the *limited available storage*, which is often far smaller than the ambient dimension of the covariance matrix. This sampling scheme brings various advantages compared with other types of measurements, as detailed in the rest of the paper.

A. Motivation

The quadratic measurement scheme finds applications in many practical scenarios. Due to space limitations, we restrict our discussion to the following two scenarios.

1) Covariance Sketching for Data Streams: A high-dimensional data stream model represents *real-time* data that arrives sequentially, where each data instance is itself high-dimensional. In many resource-constrained applications, the available memory and processing power at the data acquisition devices are severely limited compared with the volume of the data [3]. Therefore it is desirable to extract the covariance matrix from inputs on the fly without storing the whole stream. Interestingly, quadratic sampling can be leveraged as an effective data stream processing method to extract the covariance from real-time data with low memory and computational cost.

Specifically, consider an input data stream $\{\boldsymbol{x}_t\}_{t=1}^{\infty}$ that arrives sequentially, where each $\boldsymbol{x}_t \in \mathbb{R}^n$ itself is a high-dimensional instance generated at time t. The goal is to estimate its covariance matrix $\mathbf{\Sigma} = \mathbb{E}[\boldsymbol{x}_t \boldsymbol{x}_t^{\top}]$ from as small a memory as possible[1]. We propose to pool the data stream $\{\boldsymbol{x}_t\}_{t=1}^{\infty}$ into a small set of measurements in an easy-to-update fashion with a collection of sketching vectors $\{\boldsymbol{a}_i\}_{i=1}^m$. Our covariance sketching method is outlined below:

(i) At each time t, we randomly choose a sketch vector indexed by $\ell_t \in \{1, \dots, m\}$, and obtain a single linear *sketch* $\boldsymbol{a}_{\ell_t}^\top \boldsymbol{x}_t$.

(ii) All sketches employing the same sketching vector $\boldsymbol{a}_i$ are squared, aggregated and normalized, which converge *rapidly* to a *measurement*[2]

$$y_i = \mathbb{E}[(\boldsymbol{a}_i^\top \boldsymbol{x}_t)^2] + \boldsymbol{\eta}_i = \boldsymbol{a}_i^\top \boldsymbol{\Sigma} \boldsymbol{a}_i + \boldsymbol{\eta}_i, \qquad (2)$$

where $\boldsymbol{\eta} := \{\eta_i\}_{i=1}^m$ denotes the inaccuracy term.

(iii) Recover $\boldsymbol{\Sigma}$ from m stored measurements $\boldsymbol{y} := \{y_i\}_{i=1}^m$.

This covariance sketching method has low sensing and memory complexity. The computational cost for sketching each instance is linear with respect to the data dimension n. Each measurement y_i is aggregated from a different set of instances, allowing the sketching scheme to be performed over multiple machines in a distributed manner. The required memory complexity of this scheme, denoted by m, approaches the fundamental sampling limit for near-optimal covariance estimation at the sensing stage, as will be shown in this paper.

2) Spectral Estimation from Energy Measurements: A large class of wireless communication tasks in stochastic environments rely on reliable estimation of the spectral characteristics of random processes [4]. When communication takes place in the high-frequency regime, energy measurement are often more reliable than phase measurements. One potential application in multi-antenna communication problems involves estimating the covariance of signals across all receive antennas, where only the power measurements of certain linear combination of receive signals are obtainable.

Specifically, for a random process $\{\boldsymbol{x}_t\}$, if we employ a random sampling vector $\boldsymbol{a}_i$ and observe the average energy measurements of $\boldsymbol{a}_i^\top \boldsymbol{x}_t$ over N instances $\{\boldsymbol{x}_t\}_{1\le t\le N}$, the energy measurement can be expressed as

$$y_i = \frac{1}{N}\sum_{t=1}^{N} |\boldsymbol{a}_i^\top \boldsymbol{x}_t|^2 = \boldsymbol{a}_i^\top \boldsymbol{\Sigma}_N \boldsymbol{a}_i, \quad i = 1, \dots, m \qquad (3)$$

where $\boldsymbol{\Sigma}_N := \frac{1}{N}\sum_{t=1}^N \boldsymbol{x}_t \boldsymbol{x}_t^\top$ denotes the sample covariance matrix. This leads to the quadratic-form observations.

B. Contributions

We develop tractable algorithms to recover the covariance matrix from quadratic measurements (1). The estimation algorithms are based on convex relaxation with respect to the presumed low-dimensional structures. For a broad class of sub-Gaussian sensing vectors, we derive theoretical performance guarantees that allow exact and universal recovery, i.e. once the sensing vectors are selected, all covariance matrices satisfying the presumed structure (i.e. low-rank or Toeplitz low-rank) can be recovered exactly in the absence of noise. We also establish that the algorithms enable recovery of the covariance matrix with high accuracy even in the presence of imperfect structural assumptions; additionally, if the measurements are corrupted by bounded noise, the estimate deviates from the ground truth by at most a constant multiple of the noise level. Finally, the algorithms succeed as soon as the number of measurements exceeds the theoretic sampling limits, implying a minimal storage complexity at the sensing stage.

The analysis for general low-rank matrices is established upon a novel mixed-norm restricted isometry property, termed RIP-ℓ_2/ℓ_1. This key metric allows us to significantly simplify the proof, and bypass the need to construct dual certificates [5], [6], which is often the most complicated step in existing approaches to certify optimality of a convex program.

On the other hand, we show that a truncated combination of quadratic measurements satisfies the conventional restricted isometry property (RIP-ℓ_2/ℓ_2) when restricted to *Toeplitz* low-rank matrices. We also establish RIP-ℓ_2/ℓ_2 for bounded and near-isotropic operators, enabling universal and stable low-rank matrix recovery for a broad class of operators including Fourier-type measurements. This strengthens existing results [7] and might be of independent interest.

C. Related Work

In most existing work, the covariance matrix is estimated from a collection of *full* data samples, and performance guarantees have been derived on how many samples are sufficient to approximate the ground truth. These schemes are no longer useful when acquisition of full data samples is infeasible due to power consumption or other constraints. For example, in real-time monitoring environments using battery-powered sensors, the battery life is typically extended by taking only a few samples for each object of interest. In contrast, this paper is motivated by the success of compressed sensing (e.g. [8]), which asserts that compression can be achieved at the same time as sensing without compromising the relevant information of the signal. Our covariance sketching scheme can be regarded as covariance estimation from compressed data, which is a memory-efficient scheme in preserving the covariance information.

When the covariance matrix is assumed to be sparse, recent work [9] proposed to estimate a covariance matrix from measurements of the form $\boldsymbol{Y} = \boldsymbol{A}\boldsymbol{\Sigma}\boldsymbol{A}^\top$, where $\boldsymbol{A} \in \mathbb{R}^{d\times n}$ denotes the sketching matrix generated from an expander graph with $d \ll n$. Nevertheless, this scheme cannot be applied to low-rank covariance matrix recovery due to the non-empty null space of $\boldsymbol{A}$.

Our method is motivated by recent developments in phase retrieval [5], [6] (in particular PhaseLift), which is equivalent to recovering rank-one covariance matrices from quadratic samples. When specializing our result to this case, we recover and improve upon the best-known theoretical guarantee for a much larger class of sub-Gaussian measurements with a much simpler proof. Our recovery algorithm is also related to matrix recovery from random sampling [10], [11], but quadratic samples do not satisfy the presumed condition (RIP-ℓ_2/ℓ_2) therein, as pointed out by Candes et. al. [5]. Finally, after submitting our paper, we become aware of an independent work that also studies low-rank matrix recovery under rank-one quadratic measurements [12], via an interesting analytical framework.

[1] The scenario we consider is quite general. The only assumption we impose is that the covariance of a random substream of $\{\boldsymbol{x}_t\}_{t=1}^{\infty}$ converges to the same covariance as $\boldsymbol{\Sigma}$.

[2] Note that we might only obtain measurements for empirical covariance matrices instead of $\boldsymbol{\Sigma}$, but this inaccuracy can be absorbed into the term $\boldsymbol{\eta}$.

In comparison, our results accommodate a larger class of covariance structure including Toeplitz low-rank matrices.

D. Notation

Before proceeding, we provide a brief summary of the notation used throughout this paper. We use $\|\boldsymbol{X}\|$, $\|\boldsymbol{X}\|_{\mathrm{F}}$, and $\|\boldsymbol{X}\|_*$ to denote, respectively, the spectral norm, the Frobenius norm, and the nuclear norm of $\boldsymbol{X}$. The Euclidean inner product between $\boldsymbol{X}$ and $\boldsymbol{Y}$ is defined as $\langle \boldsymbol{X}, \boldsymbol{Y} \rangle = \mathrm{Tr}(\boldsymbol{X}^T \boldsymbol{Y})$. Besides, we denote by $\mathcal{T}$ the orthogonal projection operator onto Toeplitz matrices, and by $\mathcal{T}^\perp$ its orthogonal complement.

II. Convex Relaxation and Main Results

In this paper, we restrict our attention to the following random sampling model, where the sensing vectors are composed of i.i.d. *sub-Gaussian* entries. In particular, we assume $\boldsymbol{a}_i$'s ($1 \le i \le m$) are i.i.d. copies of $\boldsymbol{z} = [z_1, \cdots, z_n]^T$, where each z_i is i.i.d. satisfying

$$\mathbb{E}[z_i] = 0, \quad \mathbb{E}[z_i^2] = 1, \quad \text{and} \quad \mu_4 := \mathbb{E} z_i^4 > 1. \tag{4}$$

We assume that the noise $\boldsymbol{\eta}$, which is possibly adversarial, is bounded in either ℓ_1 norm or ℓ_2 norm as specified in the theoretical guarantees in Theorem 1 and Theorem 2. For notational simplicity, let $\boldsymbol{A}_i := \boldsymbol{a}_i \boldsymbol{a}_i^T$ represent the equivalent rank-one sensing matrix. We also define the operator $\mathcal{A}(\boldsymbol{M}) : \mathbb{R}^{n\times n} \mapsto \mathbb{R}^m$ that maps a matrix $\boldsymbol{M} \in \mathbb{R}^{n\times n}$ to $\{\langle \boldsymbol{M}, \boldsymbol{A}_i \rangle\}_{i=1}^m$, which allows us to express (1) by $\boldsymbol{y} = \mathcal{A}(\boldsymbol{\Sigma}) + \boldsymbol{\eta}$.

A. Recovery of Low-Rank Covariance Matrices

Suppose that $\boldsymbol{\Sigma}$ is low-rank such that its rank is much smaller than the ambient dimension n. A natural heuristic is to perform rank minimization to encourage the low-rank structure

$$\hat{\boldsymbol{\Sigma}} = \mathrm{argmin}_{\boldsymbol{M}}\, \mathrm{rank}(\boldsymbol{M}) \quad \text{s.t.} \quad \boldsymbol{M} \succeq 0, \quad \|\boldsymbol{y} - \mathcal{A}(\boldsymbol{M})\|_1 \le \epsilon_1, \tag{5}$$

where ϵ_1 is an upper bound on[3] $\|\boldsymbol{\eta}\|_1$, and $\boldsymbol{M} \succeq 0$ denotes the positive semidefinite (PSD) constraint. Since rank minimization is NP-hard, we propose instead to minimize the trace norm:

$$\hat{\boldsymbol{\Sigma}} = \mathrm{argmin}_{\boldsymbol{M}}\, \mathrm{Tr}(\boldsymbol{M}) \quad \text{s.t.} \quad \boldsymbol{M} \succeq 0, \quad \|\boldsymbol{y} - \mathcal{A}(\boldsymbol{M})\|_1 \le \epsilon_1. \tag{6}$$

The trace norm forms a convex surrogate for the rank function, as motivated by PhaseLift [5]. Encouragingly, the trace minimization (6) returns faithful estimates even when $\boldsymbol{\Sigma}$ is approximately low rank and/or when the samples are noisy.

Theorem 1. *Consider the sub-Gaussian sampling model in* (4)*, then with probability exceeding* $1 - \exp(-c_1 m)$*, the solution* $\hat{\boldsymbol{\Sigma}}$ *to* (6) *satisfies*

$$\|\hat{\boldsymbol{\Sigma}} - \boldsymbol{\Sigma}\|_{\mathrm{F}} \le C_1 \|\boldsymbol{\Sigma} - \boldsymbol{\Sigma}_r\|_* / \sqrt{r} + C_2 \epsilon_1 / m, \tag{7}$$

for all covariance matrices $\boldsymbol{\Sigma} \in \mathbb{R}^{n\times n}$*, provided that* $m > c_0 nr$*. Here,* $\boldsymbol{\Sigma}_r := \arg\min_{\boldsymbol{M}:\mathit{rank}(\boldsymbol{M})=r} \|\boldsymbol{\Sigma} - \boldsymbol{M}\|_{\mathrm{F}}$*, and* c_0*,* c_1*,* C_1*,* C_2 *are universal constants.*

[3]Note that ϵ_1 is allowed to scale with the memory complexity m, and can be easily modified to accommodate stochastic noise, as discussed in [13]. In many applications, $\frac{\epsilon_1}{m}$ is bounded with high probability, which is sufficient for our results to be meaningful.

The main implications of Theorem 1 and its associated performance bound (7) are as follows.

1) **Exact Recovery from Noiseless Measurements.** Consider the case where $\mathrm{rank}(\boldsymbol{\Sigma}) = r$. In the absence of noise, one can see from (7) that the trace minimization (6) (with $\epsilon = 0$) allows perfect covariance recovery with exponentially high probability, provided that $m = \Omega(nr)$. Since PSD rank-r matrices have $nr - \frac{r(r-1)}{2}$ degrees of freedom, our algorithm allows order-optimal memory complexity.

2) **Near-Optimal Universal Recovery**. Our trace minimization (6) enables universal recovery, in the sense that once the sensing vectors are chosen, *all* low-rank covariance matrices can be perfectly recovered from noiseless measurements. This universality feature arises as soon as the memory complexity exceeds the theoretic limit. When specialized to the rank-one case, this recovers the best known guarantee for Phase Retrieval [6], and extends the optimality results to a large class of sub-Gaussian measurements beyond the Gaussian model.

3) **Robust Recovery for Approximately Low-Rank Matrices**. If $\boldsymbol{\Sigma}$ is approximately low-rank, then from (7) the recovery inaccuracy from noiseless samples is at most $\|\hat{\boldsymbol{\Sigma}} - \boldsymbol{\Sigma}\|_{\mathrm{F}} \le O\left(\|\boldsymbol{\Sigma} - \boldsymbol{\Sigma}_r\|_* / \sqrt{r}\right)$. This asserts that the trace minimization returns an almost accurate estimate in a manner that requires no prior knowledge of the ground truth (other than the power law decay that is natural for a broad class of data).

4) **Stable Recovery from Noisy Samples**. When $\boldsymbol{\Sigma}$ is exactly of rank r and the noise is bounded $\|\boldsymbol{\eta}\|_1 \le \epsilon_1$, Theorem 1 asserts that the reconstruction inaccuracy of (6) can be bounded above by $\|\hat{\boldsymbol{\Sigma}} - \boldsymbol{\Sigma}\|_{\mathrm{F}} \le C_2 \epsilon_1 / m$. This reveals that the trace minimization recovers an unknown object with an error at most a constant multiple of the noise level, which makes it practically appealing.

B. Recovery of Toeplitz Low-Rank Covariance Matrices

Suppose that $\boldsymbol{\Sigma}$ is low-rank and represents the covariance matrix of n-dimensional *stationary* data instances. Similar to recovery in the general low-rank model, we propose to perform trace-norm minimization. Since $\boldsymbol{x}_i$ is drawn from a stationary process, we further impose a Toeplitz constraint to enforce stationarity conditions, which yields the following estimate

$$\hat{\boldsymbol{\Sigma}} = \arg\min_{\boldsymbol{M}} \mathrm{Tr}(\boldsymbol{M}) \quad \text{s.t. } \boldsymbol{M} \text{ is Toeplitz, } \boldsymbol{M} \succeq 0, \quad \|\boldsymbol{y} - \mathcal{A}(\boldsymbol{M})\|_2 \le \epsilon_2, \tag{8}$$

where ϵ_2 is an upper bound of $\|\boldsymbol{\eta}\|_2$.

Encouragingly, the semidefinite relaxation (8) is exact under noise-free measurements and provides stable recovery from noisy measurements, as asserted in the following theorem.

Theorem 2. *Consider the sub-Gaussian sampling model in* (4)*, and assume that*[4] $\mu_4 \le 3$ *and* $\|\boldsymbol{\eta}\|_2 \le \epsilon_2$*. Then with probability exceeding* $1 - 1/n^2$*,*

$$\|\hat{\boldsymbol{\Sigma}} - \boldsymbol{\Sigma}\|_{\mathrm{F}} \le C\epsilon_2 / \sqrt{m} \tag{9}$$

for all Toeplitz covariance matrices $\boldsymbol{\Sigma}$ *of rank at most* r*, provided that* $m > cr\log^{10} n$*. Here,* c *and* C *are constants.*

[4]We are only able to prove the theorem when $\mu_4 \le 3$, which roughly requires that the tails of the distributions are no heavier than for the Gaussian measure (e.g. $\mu_4 = 3$ for Gaussian and $\mu_4 = 1$ for Bernoulli distribution).

Once we obtain accurate recovery of $\boldsymbol{\Sigma}$, the spectrum can be identified by conventional harmonic retrieval methods, e.g. ESPRIT. Some implications of Theorem 2 are as follows:

1) **Exact Recovery without Noise**. By Theorem 2, exact recovery of stationary covariance matrices occurs as soon as the number of measurements is on the theoretic sampling limit $\Omega(r)$ up to some poly-logarithmic factor. Note that this sampling theoretic limit is n times smaller than that for general low-rank matrices, and is about n/r times lower than the degrees of freedom for general Toeplitz matrices.

2) **Stable and Universal Recovery from Noisy Measurements**. The proposed algorithm (8) returns faithful estimates in the presence of noise, as revealed by Theorem 2. This feature is universal: if $\mathcal{A}$ is randomly sampled, then with high probability, the error bounds (9) hold simultaneously for all Toeplitz low-rank matrices.

III. Approximate Isometry

Prevailing wisdom asserts that stable recovery from minimal measurements is possible if the sampling mechanism preserves the signal strength when acting on matrices of interest [10]. This property is often demonstrated in terms of the restricted isometry property (RIP), which arises if the sampling output preserves the input strength under certain metrics. The most commonly used one is RIP-ℓ_2/ℓ_2, for which the signal strength before and after sampling are both measured in terms of the ℓ_2 norm (see definition in [10], [11]). However, RIP-ℓ_2/ℓ_2 does not hold under the quadratic operator $\mathcal{A}$ for either low-rank matrices or Toeplitz low-rank matrices. This motivates us to investigate other RIP metrics and/or construct new operators.

A. Mixed-Norm RIP (RIP-ℓ_2/ℓ_1) for Low-rank Matrices

It has been argued in [5] that RIP-ℓ_2/ℓ_2 cannot be guaranteed from $O(nr)$ samples even for rank-one matrices. Consequently, prior analysis of PhaseLift [5], [6] operates upon RIP-ℓ_1/ℓ_1, for which the strength for both the input and output are measured in terms of the ℓ_1 norms. Nevertheless, RIP-ℓ_1/ℓ_1 no longer holds for general low-rank matrices beyond matrices of constant rank. Moreover, the proof based on RIP-ℓ_1/ℓ_1 typically relies on a delicate dual construction [5], [6], which is often mathematically complicated.

The key and novel ingredient in our analysis for low-rank structure is a mixed-norm approximate isometry termed RIP-ℓ_2/ℓ_1, which measures the input and output in terms of the Frobenius norm and the ℓ_1 norm, respectively.

Definition 1 (RIP-ℓ_2/ℓ_1 for low-rank matrices). For the set of rank-r matrices, we define the RIP-ℓ_2/ℓ_1 constants δ_r^{lb} and δ_r^{ub} with respect to an operator $\mathcal{B}$ as the smallest numbers such that for all $\boldsymbol{X}$ of rank at most r:

$$\left(1-\delta_r^{\text{lb}}\right)\|\boldsymbol{X}\|_{\text{F}} \leq \frac{1}{m}\left\|\mathcal{B}\left(\boldsymbol{X}\right)\right\|_1 \leq \left(1+\delta_r^{\text{ub}}\right)\|\boldsymbol{X}\|_{\text{F}}.$$

Unfortunately, the original sampling operator $\mathcal{A}$ does *not* satisfy RIP-ℓ_2/ℓ_1. This occurs primarily because each $\boldsymbol{A}_i$ has *non-zero mean*, which biases the output samples. In order to get rid of this undesired effect, we introduce a "debiased" set of auxiliary measurement matrices as follows

$$\boldsymbol{B}_i := \boldsymbol{A}_{2i-1} - \boldsymbol{A}_{2i}. \tag{10}$$

Let $\mathcal{B}\left(\boldsymbol{X}\right)$ represent the linear transformation that maps $\boldsymbol{X}$ to $\{\langle \boldsymbol{B}_i, \boldsymbol{X}\rangle\}_{i=1}^m$, then $\mathcal{B}$ satisfies RIP-ℓ_2/ℓ_1 as stated below.

Lemma 1. *Consider the sub-Gaussian sampling model in* (4)*. There exist universal constants $c_1, c_2, c_3, c_4 > 0$ such that with probability exceeding $1-\exp\left(-c_3 m\right)$, $\mathcal{B}$ satisfies RIP-ℓ_2/ℓ_1 for all matrices $\boldsymbol{X}$ of rank at most r, and obeys*

$$1-\delta_r^{\text{lb}} \geq c_1, \quad 1+\delta_r^{\text{ub}} \leq c_2, \tag{11}$$

provided that $m > c_4 nr$.

The RIP-ℓ_2/ℓ_1 of $\mathcal{B}$ in turn leads to the establishment of Theorem 1. Interested readers are referred to [13] for proof.

B. Constructing RIP-ℓ_2/ℓ_2 Operators for Toeplitz Low-Rank Matrices

While quadratic measurements in general do not exhibit RIP-ℓ_2/ℓ_2 with respect to general low-rank matrices, a truncated combination of them can indeed satisfy RIP-ℓ_2/ℓ_2 when restricted to *Toeplitz* low-rank matrices. Before proceeding to the Toeplitz case, we characterize RIP-ℓ_2/ℓ_2 of near-isotropic and bounded operators for general low-rank manifold as follows.

Theorem 3. *Suppose that for all $1 \leq i \leq m$,*

$$\|\boldsymbol{B}_i\| \leq K, \quad \textit{and} \quad \|\mathbb{E}\mathcal{B}_i^*\mathcal{B}_i - \mathcal{I}\| \leq \frac{c_5}{n} \tag{12}$$

hold for some quantity $K \leq n^2$. If $m > c_0 r K^2 \log^7 n$, then with probability at least $1-1/n^2$, $\mathcal{B}$ satisfies RIP-ℓ_2/ℓ_2 with respect to all matrices of rank at most r.

In fact, the bound on $\|\boldsymbol{B}_i\|$ can be as small as $\Theta\left(\sqrt{n}\right)$, and we say a measurement matrix $\boldsymbol{B}_i$ is *well-bounded* if $K = O\left(\sqrt{n}\text{poly}\log\left(n\right)\right)$, which subsumes the Fourier-type basis as discussed in [7]. Theorem 3 asserts that simultaneously well-bounded and near-isotropic operators satisfy RIP-ℓ_2/ℓ_2 when $m = \Omega\left(nr\text{poly}\log\left(n\right)\right)$, which in turn ensures universal and stable recovery. This strengthens the prior work in [7] which does not yield universal recovery.

Unfortunately, $\mathcal{A}$ is neither isotropic nor well-bounded. In order to apply Theorem 3, we construct an auxiliary set of measurement matrices $\tilde{\boldsymbol{B}}_i$ through the following procedure.

(i) Generate M matrices independently such that

$$\hat{\boldsymbol{B}}_i := \begin{cases} \sqrt{n}\mathcal{T}\left(\alpha\boldsymbol{A}_{3i} + \beta\boldsymbol{A}_{3i-1} + \gamma\boldsymbol{A}_{3i-2}\right), & \text{w.p. } \frac{1}{n}, \\ \sqrt{\frac{n}{n-1}}\mathcal{T}^{\perp}\left(\boldsymbol{G}_i\right), & \text{else,} \end{cases}$$

where $\boldsymbol{G}_i$ is an i.i.d. standard Gaussian matrix, and α, β, γ are specified in [13].

(ii) Define a *truncated* version $\tilde{\boldsymbol{B}}_i$ of $\hat{\boldsymbol{B}}_i$ as follows

$$\tilde{\boldsymbol{B}}_i := \hat{\boldsymbol{B}}_i \mathbf{1}_{\left\{\|\hat{\boldsymbol{B}}_i\| \leq c_{10}\log^{3/2} n\right\}}, \quad 1 \leq i \leq M. \tag{13}$$

One can demonstrate that the $\tilde{\boldsymbol{B}}_i$'s are nearly-isotropic and well-bounded, and hence by Theorem 3 the associated operator $\tilde{\mathcal{B}}$ enables exact and stable recovery for all rank-r matrices when M exceeds $nr\text{poly}\log(n)$. This in turn establishes Theorem 2 through an equivalence argument; details can be found in [13].

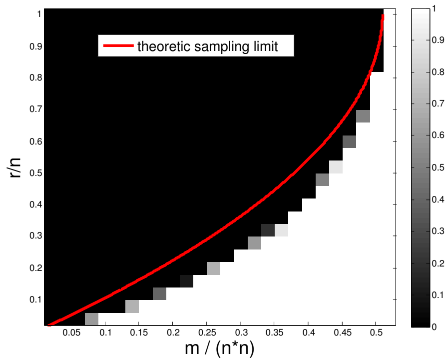

Figure 1. Phase transition plots for low-rank covariance matrix recovery from quadratic measurements when $n = 50$.

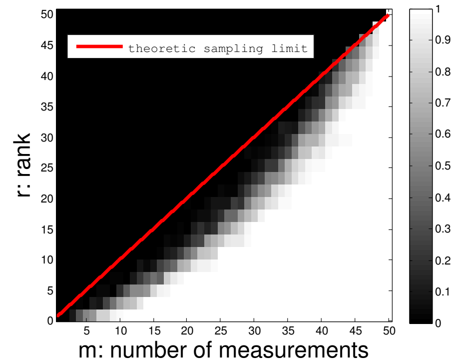

Figure 2. Phase transition plots for Toeplitz low-rank covariance matrix recovery where frequency locations are randomly generated and $n = 50$.

IV. Numerical Examples

To demonstrate the practicality of convex relaxation under quadratic measurements, we now consider several numerical examples by solving the programs via SDPT3. Numerically, we declare $\boldsymbol{\Sigma}$ to be accurately recovered if the solution $\hat{\boldsymbol{\Sigma}}$ satisfies $\|\hat{\boldsymbol{\Sigma}} - \boldsymbol{\Sigma}\|_{\mathrm{F}} / \|\boldsymbol{\Sigma}\|_{\mathrm{F}} \leq \epsilon_0 = 10^{-3}$. Note that ϵ_0 can be selected to be much smaller than 10^{-3} without affecting the phase transition diagram.

For the recovery of low-rank matrices using (6), we conduct the following series of Monte Carlo trials. Specifically, we choose $n = 50$, and for each (m, r) pair, we repeat the following experiments 20 times. We generate $\boldsymbol{\Sigma}$, an $n \times n$ PSD matrix via $\boldsymbol{\Sigma} = \boldsymbol{L}\boldsymbol{L}^{\top}$, where $\boldsymbol{L}$ is a randomly generated $n \times r$ matrix with i.i.d. Gaussian entries. The sensing vectors are generated as i.i.d. Gaussian vectors, and we obtain noiseless quadratic measurements $\boldsymbol{y}$. Figure 1 illustrates the empirical probability of success recovery, which is reflected through the color of each cell. It turns out that the practical phase transition curve is very close to the theoretic limit, which confirms the optimality of our algorithm.

For the recovery of Toeplitz low-rank matrices using (8), we perform a series of experiments for Toeplitz low-rank matrices when $n = 50$. By Caratheodory's theorem, each PSD Toeplitz matrix can be uniquely decomposed into a linear combination of line spectra. In the real-valued situation, the underlying spectral spikes occur in conjugate pairs (i.e. $(f_1, -f_1), (f_2, -f_2), \cdots$). We independently generate $r/2$ frequency pairs within the unit disk uniformly at random, and the amplitudes are generated as i.i.d. $\chi^2(1)$ random variables. Figure 2 illustrates the phase transition diagram for varied choices of (m, r). The empirical success rate is calculated by averaging over 50 trials. While there are r degrees of freedom, our algorithm exhibits a near-linear phase transition boundary, which justifies our theoretical prediction.

V. Conclusions and Future Work

We have proposed and analyzed a technique to obtain estimation from quadratic measurements. This sampling model acts as an effective method for processing real-time data under constraints on memory and computational complexity, arising for example in high-frequency signal processing tasks using energy measurements. Covariance recovery from quadratic measurements can be achieved via efficient convex programming as soon as the number of measurements exceeds the fundamental sampling theoretic limit of the parsimonious covariance structures. Our results highlight the stability of the convex program in the presence of noise and imperfect structural assumptions. It remains to see whether the proposed sensing scheme can be used to recover other types of low-dimensional covariance structures, such as a sparse inverse covariance matrix. It will also be interesting to explore general types of sampling models that satisfy RIP-ℓ_2/ℓ_1 such as structured random measurements.

Acknowledgment

The authors would like to thank Emmanuel Candes, Xiaodong Li and Yihong Wu for stimulating discussions. Y. Chen and A. J. Goldsmith are supported by the NSF Center for Science of Information (CSoI) under grant NSF-CCF-0939370 and by AFOSR Grant FA9550-09-1-0643. Y. Chi is supported in part by a Google Faculty Research Award.

References

[1] I. Jolliffe, *Principal component analysis*. Wiley Online Library, 2005.
[2] J. Burg, D. Luenberger, and D. Wenger, "Estimation of structured covariance matrices," *Proc. of the IEEE*, vol. 70, pp. 963–974, 1982.
[3] S. Muthukrishnan, *Data streams: Algorithms and applications*, 2005.
[4] A. Jung, G. Taubock, and F. Hlawatsch, "Compressive spectral estimation for nonstationary random processes," *IEEE Transactions on Information Theory*, vol. 59, no. 5, May 2013.
[5] E. J. Candes, T. Strohmer, and V. Voroninski, "Phaselift: Exact and stable signal recovery from magnitude measurements via convex programming," *Communications on Pure and Applied Mathematics*, 2012.
[6] E. J. Candes and X. Li, "Solving quadratic equations via PhaseLift when there are about as many equations as unknowns," *Foundations of Computational Mathematics*, 2013.
[7] D. Gross, "Recovering low-rank matrices from few coefficients in any basis," *IEEE Trans on Info Theory*, vol. 57, pp. 1548–1566, Mar. 2011.
[8] E. J. Candes and T. Tao, "Near-optimal signal recovery from random projections: Universal encoding strategies?" *IEEE Transactions on Information Theory*, vol. 52, no. 12, pp. 5406–5425, Dec. 2006.
[9] G. Dasarathy, P. Shah, B. N. Bhaskar, and R. Nowak, "Sketching sparse matrices," *arXiv preprint arXiv:1303.6544*, 2013.
[10] B. Recht, M. Fazel, and P. A. Parrilo, "Guaranteed minimum-rank solutions of linear matrix equations via nuclear norm minimization," *SIAM Review*, vol. 52, no. 3, pp. 471–501, 2010.
[11] E. Candes and Y. Plan, "Tight oracle inequalities for low-rank matrix recovery from a minimal number of noisy random measurements," *IEEE Trans on Information Theory*, vol. 57, no. 4, pp. 2342–2359, 2011.
[12] T. Cai and A. Zhang, "ROP: Matrix recovery via rank-one projections," *arXiv preprint arXiv:1310.5791*, 2013.
[13] Y. Chen, Y. Chi, and A. Goldsmith, "Exact and stable covariance estimation from quadratic sampling via convex programming," October 2013. [Online]. Available: http://arxiv.org/abs/1310.0807