Exact and Stable Covariance Estimation from Quadratic Sampling via Convex Programming

Yuxin Chen†, Yuejie Chi*, Andrea J. Goldsmith†

Stanford University†, Ohio State University*
• Data Stream / Stochastic Processes
  ○ Each data instance can be high-dimensional
  ○ We’re interested in information in the data rather than the data themselves

• Covariance Estimation
  ○ second-order statistics $\Sigma \in \mathbb{R}^{n \times n}$
  ○ cornerstone of many information processing tasks
What are Quadratic Measurements?

- Quadratic Measurements
  - obtain \( m \) measurements of \( \Sigma \) taking the form
    \[
y_i \approx a_i^\top \Sigma a_i \quad (1 \leq i \leq m)
    \]
  - rank-1 measurements!
Example: Applications in Spectral Estimation

- **High-frequency wireless and signal processing** (Energy Measurements)
  - Spectral estimation of **stationary processes** *(possibly sparse)*
Example: Applications in Spectral Estimation

- **High-frequency wireless and signal processing** (Energy Measurements)
  - Spectral estimation of stationary processes (*possibly sparse*)
  - Channel Estimation in MIMO Channels
Example: Applications in Optics

- **Phase Space Tomography**
  - measure correlation functions of a wave field

Fig credit: Chi et al
Example: Applications in Optics

- **Phase Space Tomography**
  - measure correlation functions of a wave field

- **Phase Retrieval**
  - signal recovery from magnitude measurements
Example: Applications in Data Streams

- **Covariance Sketching**
  - data stream: real-time data \( \{x_t\}_{t=1}^{\infty} \) arriving sequentially at a high rate...

- **Challenges**
  - limited memory
  - computational efficiency
  - hopefully a single pass over the data
Proposed Quadratic Sketching Method

1) Sketching:
   ○ at each time $t$, obtain a quadratic sketch $(a_i^\top x_t)^2$
     — $a_i$: sketching vector
Proposed Quadratic Sketching Method

1) Sketching:
   - at each time $t$, obtain a quadratic sketch $\left( a_i^T x_t \right)^2$
     — $a_i$: sketching vector

2) Aggregation:
   - all sketches are aggregated into $m$ measurements
     \[
y_i = a_i^T \left( \frac{1}{T} \sum_{t=1}^{T} x_t x_t^T \right) a_i \approx a_i^T \Sigma a_i \quad (1 \leq i \leq m)
\]
Proposed Quadratic Sketching Method

1) Sketching:
○ at each time $t$, obtain a quadratic sketch $(a_i^T x_t)^2$
  — $a_i$: sketching vector

2) Aggregation:
○ all sketches are aggregated into $m$ measurements
  
  $$y_i = a_i^T \left( \frac{1}{T} \sum_{t=1}^{T} x_t x_t^T \right) a_i \approx a_i^T \Sigma a_i \quad (1 \leq i \leq m)$$

• Benefits:
  ○ one pass
  ○ minimal storage *(as will be shown)*
Given: \( m (\ll n^2) \) quadratic measurements \( y = \{y_i\}_{i=1}^m \)

\[
y_i = a_i^\top \Sigma a_i + \eta_i, \quad i = 1, \ldots, m,\]

\( a_i \): sampling vectors
\( \eta = \{\eta_i\}_{i=1}^m \): noise terms

more concise operator form:

\[
y = A(\Sigma) + \eta\]

Goal: recover \( \Sigma \in \mathbb{R}^{n \times n} \).

Sampling model

sub-Gaussian i.i.d. sampling vectors
Geometry of Covariance Structure

- **# unknown > # stored measurements**
  - exploit low-dimensional structures!

- **Structures considered in this talk:**
  - low rank
  - Toeplitz low rank
  - simultaneously sparse and low-rank

1) low rank  
2) Toeplitz low rank  
3) jointly sparse and low rank
Low Rank

- **Low-Rank Structure:**
  - A few components explain most of the data variability
  - metric learning, array signal processing, collaborative filtering ... 

- \( \text{rank}(\Sigma) = r \ll n. \)
Trace Minimization for Low-Rank Structure

- **Trace Minimization**

\[(\text{TraceMin}) \quad \text{minimize}_{M} \quad \underbrace{\text{trace}(M)}_{\text{low rank}}\]

\[\text{s.t.} \quad \|\mathcal{A}(M) - y\|_1 \leq \epsilon, \quad \text{noise bound}\]

\[M \succeq 0.\]

○ inspired by *Candes et. al.* for phase retrieval
Theorem 1 (Low Rank). With high prob, for all $\Sigma$ with $\text{rank}(\Sigma) \leq r$, the solution $\hat{\Sigma}$ to TraceMin obeys

$$
\|\hat{\Sigma} - \Sigma\|_F \lesssim \frac{\|\Sigma - \Sigma_r\|_*}{\sqrt{r}} + \frac{\epsilon}{m},
$$

due to imperfect structure

due to noise

provided that $m \gtrsim rn$. ($\Sigma_r$: rank-$r$ approx of $\Sigma$)

- **Exact recovery** in the noiseless case
- **Universal recovery**: simultaneously works for all low-rank matrices
- **Robust recovery** when $\Sigma$ is *approximately* low-rank
- **Stable recovery** against bounded noise
empirical success probability of Monte Carlo trials: $n = 50$

- **Near-Optimal** Storage Complexity!
  - degrees of freedom $\approx rn$
Toeplitz Low Rank

- **Toeplitz Low-Rank Structure:**
  - **Spectral sparsity!**
    - possibly *off-the-grid* frequency spikes (Vandemonde decomposition)
  - wireless communication, array signal processing ...

- \( \text{rank}(\Sigma) = r \ll n. \)
Trace Minimization for Toeplitz Low-Rank Structure

• Trace Minimization

\[(\text{ToepTraceMin}) \quad \text{minimize}_M \quad \text{trace}(M) \quad \text{low rank}\]

\[\text{s.t.} \quad \| \mathcal{A}(M) - y \|_2 \leq \varepsilon_2 \quad \text{noise bound},\]

\[M \succeq 0,\]

\[M \text{ is Toeplitz}.\]
minimize \( \text{tr}(M) \) \ s.t. \( \|A(M) - y\|_2 \leq \epsilon_2, \quad M \succeq 0, \quad M \text{ is Toeplitz} \)

**Theorem 2 (Toeplitz Low Rank).** With high prob, for all Toeplitz \( \Sigma \) with \( \text{rank}(\Sigma) \leq r \), the solution \( \hat{\Sigma} \) to ToepTraceMin obeys

\[
\|\hat{\Sigma} - \Sigma\|_F \lesssim \frac{\epsilon_2}{\sqrt{m}},
\]

due to noise

provided that \( m \gtrsim r \text{poly log}(n) \).

- **Exact recovery** in the absence of noise
- **Universal recovery**: simultaneously works for all Toeplitz low-rank matrices
- **Stable recovery** against bounded noise
empirical success probability of Monte Carlo trials: $n = 50$

- **Near-Optimal** Storage Complexity!
  - degrees of freedom $\approx r$
Simultaneous Structure

- **Joint Structure:** $\Sigma$ is *simultaneously* sparse and low-rank.
  - **rank:** $r$
  - **sparsity:** $k$

  \[
  \Sigma = U \Lambda U^\top, \quad \text{where } U = [u_1, \ldots, u_r]
  \]
Convex Relaxation for Simultaneous Structure

- Convex Relaxation

\[
\begin{align*}
\text{minimize}_{M} & \quad \text{trace}(M) + \lambda \|M\|_1 \\
\text{subject to} & \quad \|A(M) - y\|_1 \leq \epsilon, \\
& \quad M \succeq 0.
\end{align*}
\]

- \L_i and Voroninski for rank-1 cases
minimize $\text{tr}(M) + \lambda \|M\|_1$ \quad \text{s.t.} \quad A(M) = y, \quad M \succeq 0

**Theorem 3 (Simultaneous Structure).** SDP with $\lambda \in \left[ \frac{1}{n}, \frac{1}{N_\Sigma} \right]$ is exact with high probability, provided that

$$m \gtrsim \frac{r \log n}{\lambda^2} \quad (1)$$

where $N_\Sigma := \max \left\{ \| \text{sign} (\Sigma_\Omega) \|, \sqrt{\frac{k \sum_{i=1}^r \| u_i \|^2}{r}} \right\}$.

- Exact recovery with appropriate regularization parameters
- Question: how good is the storage complexity (1)?
**Definition (Compressible Matrices)**

- *non-zero entries* of $u_i$ exhibit *power-law decays*
  - $\|u_i\|_1 = O(\text{poly log}(n))$. 
Compressible Covariance Matrices: Near-Optimal Recovery

Definition (Compressible Matrices)

- non-zero entries of $u_i$ exhibit power-law decays
  - $\|u_i\|_1 = O(\text{poly log}(n))$.

Corollary 1 (Compressible Case). For compressible covariance matrices, SDP with $\lambda \approx \frac{1}{\sqrt{k}}$ is exact w.h.p., provided that

$$m \gtrsim kr \cdot \text{poly log}(n).$$

- Near-Minimal Measurements!
  - degree-of-freedom: $\Theta(kr)$
Stability and Robustness

- **noise**: $\|\eta\|_1 \leq \epsilon$

- **imperfect structural assumption**: $\Sigma = \Sigma_\Omega + \Sigma_c$
  
  simultaneous sparse and low-rank residuals
Stability and Robustness

- **noise**: $\|\eta\|_1 \leq \epsilon$

- **imperfect structural assumption**: $\Sigma = \Sigma_\Omega + \Sigma_c$
  \[\text{simultaneous sparse and low-rank residuals}\]

**Theorem 4.** Under the same $\lambda$ as in Theorem 1 or Corollary 1,

$$\left\| \hat{\Sigma} - \Sigma_\Omega \right\|_F \lesssim \frac{1}{\sqrt{r}} \left( \|\Sigma_c\|_* + \lambda \|\Sigma_c\|_1 \right) + \frac{\epsilon}{m}$$

- stable against bounded noise
- robust against imperfect structural assumptions
• **Restricted Isometry Property**: a powerful notion for compressed sensing

\[ \forall \mathbf{X} \text{ in some class : } \| \mathcal{B}(\mathbf{X}) \|_2 \approx \| \mathbf{X} \|_F. \]

◦ *unfortunately, it does *NOT* hold for quadratic models*
Mixed-Norm RIP (for Low-Rank and Joint Structure)

- **Restricted Isometry Property**: a powerful notion for compressed sensing

  \[ \forall X \text{ in some class} : \quad \| B(X) \|_2 \approx \| X \|_F. \]

  - unfortunately, it does **NOT** hold for quadratic models

- **A Mixed-norm Variant**: **RIP-\(\ell^2/\ell^1\)**

  \[ \forall X \text{ in some class} : \quad \| B(X) \|_1 \approx \| X \|_F. \]
• **Restricted Isometry Property:** a powerful notion for compressed sensing

\[ \forall X \text{ in some class}: \quad \| B(X) \|_2 \approx \| X \|_F. \]

- unfortunately, it does **NOT** hold for quadratic models

• **A Mixed-norm Variant:** **RIP-\(\ell_2/\ell_1\)**

\[ \forall X \text{ in some class}: \quad \| B(X) \|_1 \approx \| X \|_F. \]

- does **NOT** hold for \(A\), but hold after \(A\) is **debiased**
- A very simple proof for PhaseLift!
Concluding Remarks

- **Our approach / analysis works for other structural models**
  - Sparse covariance matrix
  - Low-Rank plus Sparse matrix

- **The way ahead**
  - Sparse *inverse* covariance matrix
  - Beyond sub-Gaussian sampling
  - Online recovery algorithms
Full-length version available at arXiv:

Exact and Stable Covariance Estimation from Quadratic Sampling via Convex Programming


Thank You! Questions?