Covariance Sketching

Parikshit Shah

Joint work with Gautam Dasarathy, Badri Bhaskar, Rob Nowak

Motivation: Covariance Estimation

Covariance matrix: $\Sigma = \mathbb{E}[XX^T]$. Captures statistical dependencies.

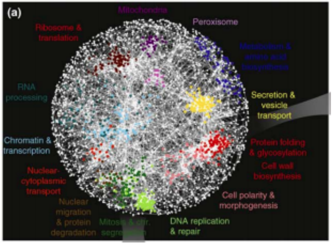


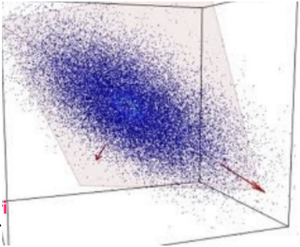
Image: Boone et al.

Covariance estimation: ubiquitious problem in science and engineering. Particularly relevant to model interaction between variates in "high-dimensional data"

However: estimation in high-dimensional regime hard (MLE is useless).

Meaningful structure often present:

1. Sparsity in covariance matrices mean few statistical interactions.
2. Low rankitude mean few "principal components".



Structure exploitation: $\log p$ samples sufficient for sparse Σ! [Bickel,Levina]

Motivation: Sketching

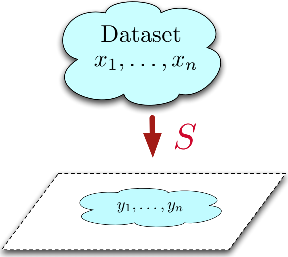
In some applications: hard to access/store all the variates.
For example, some variates may be latent.

Basic idea in CS to deal with "big data": sketch.

Less storage, efficient to communicate.

Today: A method for sketching/pooling information for estimating Σ.

Applications: "pooling" in biological experiments, efficient storage/communication

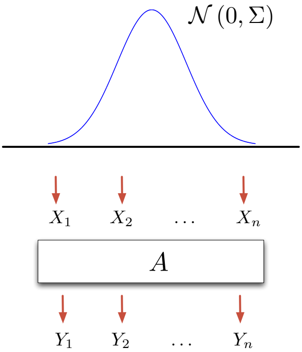


Covariance Sketching Problem

Statistical setup: $X \sim \mathcal{N}(0, \Sigma)$, $X \in \mathbb{R}^p$.
Samples not directly accessible: $X_1, \ldots, X_n$.

Sampling: sketch via fixed $A \in \mathbb{R}^{m \times p}$:

$$Y_i = AX_i, \qquad i = 1, \ldots, n.$$



Q1: How do we sketch? randomly pool.
Good combinatorial reasons: graph expansion, etc.

Q2: Can we reconstruct Σ? In general, no.
When model is structurally constrained, ℓ_1 minimization.

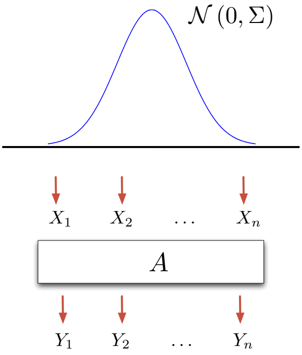
Notice that sketching completely destroys the independence structure.

Covariance Sketching Problem

Statistical setup: $X \sim \mathcal{N}(0, \Sigma)$, $X \in \mathbb{R}^p$.
Samples not directly accessible: $X_1, \ldots, X_n$.

Sampling: sketch via fixed $A \in \mathbb{R}^{m \times p}$:

$$Y_i = AX_i, \qquad i = 1, \ldots, n.$$



Q1: How do we sketch? randomly pool.
Good combinatorial reasons: graph expansion, etc.

Q2: Can we reconstruct Σ? In general, no.
When model is structurally constrained, ℓ_1 minimization.

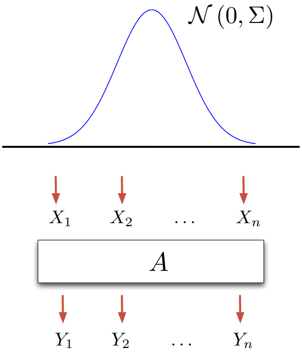
Notice that sketching completely destroys the independence structure.

Covariance Sketching Problem

Statistical setup: $X \sim \mathcal{N}(0, \Sigma)$, $X \in \mathbb{R}^p$.
Samples not directly accessible: $X_1, \ldots, X_n$.

Sampling: sketch via fixed $A \in \mathbb{R}^{m \times p}$:

$$Y_i = AX_i, \qquad i = 1, \ldots, n.$$



Q1: How do we sketch? randomly pool.
Good combinatorial reasons: graph expansion, etc.

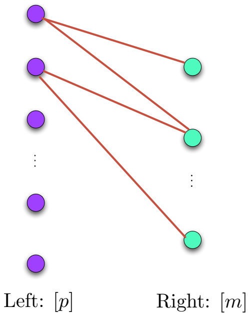
Q2: Can we reconstruct Σ? In general, no.
When model is structurally constrained, ℓ_1 minimization.

Notice that sketching completely destroys the independence structure.

Idealized model

$$A = \begin{array}{c} \\ \end{array}\begin{array}{ccccc} 1 & 2 & 3 & \dots & p \end{array}$$

$$A = \left[\begin{array}{cccc} 1 & 1 & 0 & \dots \\ 0 & 1 & 1 & \dots \\ \vdots & & \ddots & \end{array}\right]\begin{array}{c} 1 \\ 2 \\ \vdots \\ m \end{array}$$



Given sample sketches: $Y_1, \dots, Y_n$.

Empirical covariance matrix:
$\hat{\Sigma}_Y = \frac{1}{n}\sum_{i=1}^{n} Y_i Y_i^T = A\left(\frac{1}{n}\sum_{i=1}^{n} X_i X_i^T\right)A^T$.

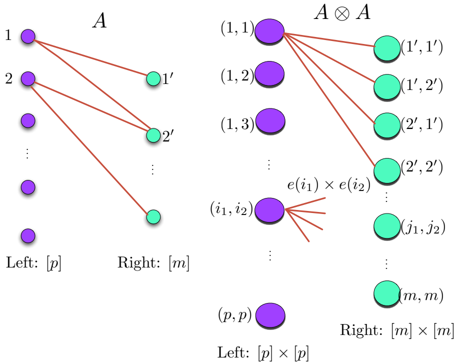
Idealization: Given $\Sigma_Y = A\Sigma A^T$.

Given Σ_Y, recover Σ.

Sensing matrix

Note that $A\Sigma A^T = (A \otimes A)\text{vec}(\Sigma)$.

Denote sparsity of Σ by Ω.

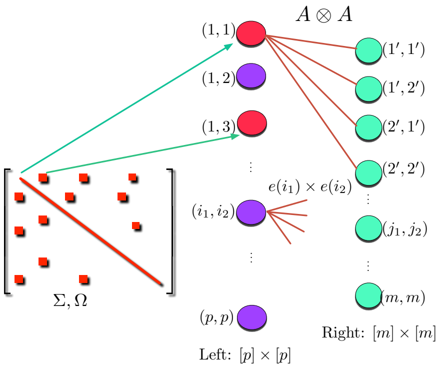


In general, $A \otimes A$ is a bad sensing matrix.
Standard CS analysis breaks down.

Sensing matrix

Note that $A\Sigma A^T = (A \otimes A)\text{vec}(\Sigma)$.

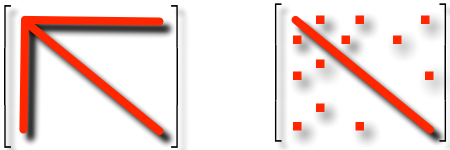
Denote sparsity of Σ by Ω.



In general, $A \otimes A$ is a bad sensing matrix.
Standard CS analysis breaks down.

Identifiability

Intuition: “sparse” models should be identifiable.



"Arrow" matrix

Distributed d-sparse matrix
Support: Ω

Theorem

If Σ is d-distributed sparse, $A \in \mathbb{R}^{m \times p}$ with $m > \sqrt{2dp}$, and every m columns of A are linearly independent, then model is identifiable.

Non-trivialty: Sensing operator is $A \otimes A$, $\text{vec}(X) \in \mathbb{R}^{p^2}$, dp-sparse, and structurally constrained.

ℓ_1 Recovery

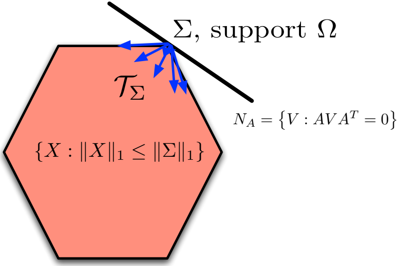
Recovery algorithm: solve

$$\begin{array}{ll} \underset{X}{\text{Minimize}} & \|X\|_1 \\ \text{subject to} & AXA^T = \Sigma_Y. \end{array}$$

Theorem

Suppose Σ is d-distributed sparse. If $A \in \mathbb{R}^{m \times p}$ is a random δ-left-regular bipartite graph, $\delta = O(\log^3(p))$, and $m = O(\sqrt{dp}\log^3(p))$ then w.h.p. $X^* = \Sigma$.

Proof ideas



Transverse intersection: $N_A \cap \mathcal{T}_\Sigma = 0$

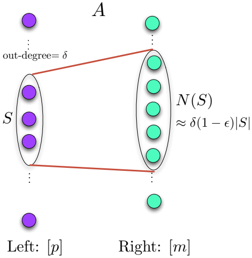
Sufficient to show the Nullspace Property with respect to Ω, A:
For all V such that $AVA^T = 0$ we have

$$\|V_\Omega\|_1 \le \frac{\epsilon}{1-\epsilon}\|V\|_1 .$$

Satisfied w.h.p. because $A \otimes A$ with left set Ω has some special combinatorial properties.

Graph Properties: Expansion

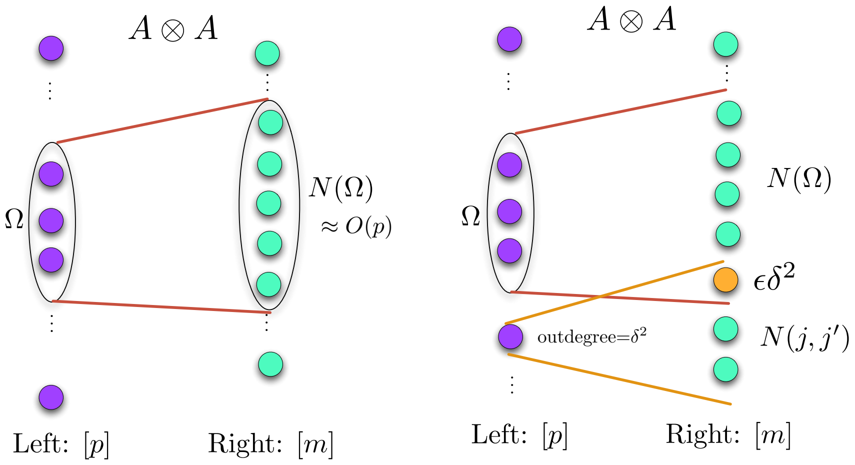
Random bipartite graphs are expanders.
(Plentiful applications, e.g. distributed routing, storage, ECCs)



Nullspace property implied by expansion + small collisions.
[Berinde et al.]

What properties does $A \otimes A$ have?
General expansion fails ...

Distributed Expansion and Small Crossover

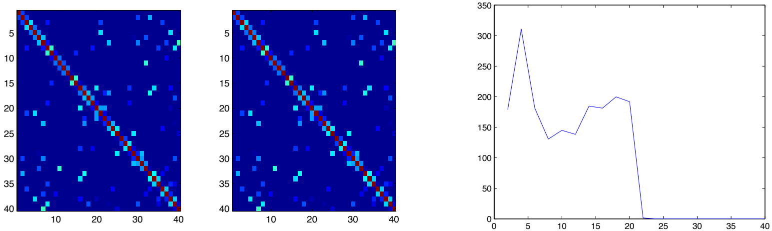


Ω distributed d-sparse

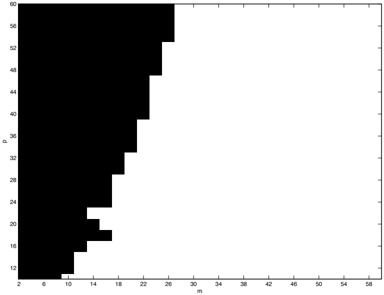
Failure of “standard approaches”

- If A is random Gaussian, then entries of $A \otimes A$ are not i.i.d., thus RIP seems difficult.
- For instance, $\delta_k^{\mathrm{RIP}}(A \otimes A) \geq \delta_k^{\mathrm{RIP}}(A)$ [Jokar’10], can only recover $\sqrt{dp}$-sparse matrices.
- Related: even though A is i.i.d., the entries of $A \Sigma A^T$ are all dependent.
- Gaussian width: $\mathrm{Ker}(A \otimes A)$ is not a uniformly random subspace.
- Similar obstructions for frame-based arguments, coherence, ...

Experiments



Exact Recovery: $\approx$ 50% compression of Σ by sketching.



Current Efforts

- Recovery from $\hat{\Sigma}_Y$: robustness to Wishart noise.
- Analysis for Gaussian A?
- Tensor recovery from tensor product sensing matrices?
- Extensions to low-rank case, sparse $+$ low-rank, etc...
- Daydream: recover Markov structure, graphical models.