Sketching Sparse Matrices

Geethu Joseph

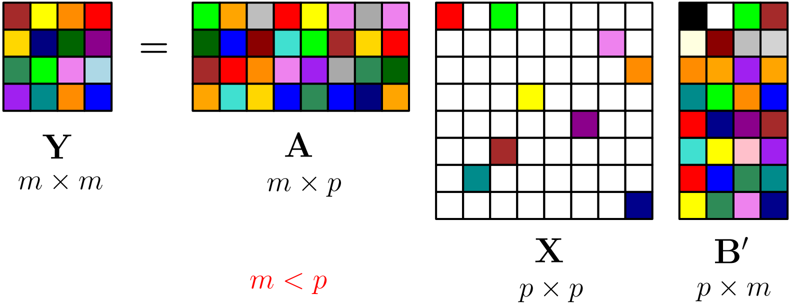
SPC Lab, IISc

September 12, 2015

Paper Reviewed

- **Title**: Sketching Sparse Matrices, Covariances, and Graphs via Tensor Products
- **Authors**: Gautam Dasarathy, Parikshit Shah, Badri Narayan Bhaskar, and Robert D. Nowak
- **Publication date**: March 2015
- **Journal name**: IEEE Transactions on Information Theory, vol. 61, no. 3, pp. 1373-1388

Sparse Matrix Recovery Problem

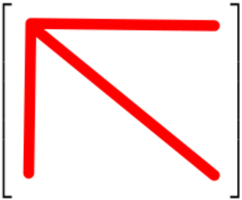


- **Goal**: Recover distributed-sparse matrix $\boldsymbol{X}$ from $\boldsymbol{Y} = \boldsymbol{A}\boldsymbol{X}\boldsymbol{B}^{\mathsf{T}}$
- Distributed sparsity: Every row and every column of $\boldsymbol{X}$ has only a few non-zeros

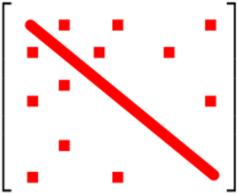
Motivation

1. Source localization: basis expansion model approach in brain using EEG signals
2. Covariance matrices: Only a few variables are correlated to each other
3. Multi-dimensional signals: Natural images are sparse in the gradient domain

Why Distributed Sparsity?



"Arrow" matrix



Distributed sparse matrix

- Arrow matrix: Impossible to recover $\boldsymbol{X}$ even if non-zero pattern is known
  - Matrix with $\boldsymbol{v} \in \ker(\boldsymbol{A})$ added to first column of $\boldsymbol{X}$ is also a potential solution
- Our focus: Size of the sketch $\boldsymbol{Y}$ to recover distributed sparse matrix $\boldsymbol{X}$

Definition: Distributed Sparsity

- $\boldsymbol{\Omega} \subset [p] \times [p]$ is d-distributed subset if for all $k \in [p]$
  1. $(k, k) \in \boldsymbol{\Omega}$
  2. $|\{(k, i) \in \boldsymbol{\Omega}\}| \leq d$
  3. $|\{(i, k) \in \boldsymbol{\Omega}\}| \leq d$
- $\boldsymbol{X}$ is d-distributed sparse matrix if

$$\operatorname{supp}(\boldsymbol{X}) \subset \boldsymbol{\Omega}$$

  - # off-diagonal nonzeros of every row and column $\leq d - 1$

Convex Relaxation using l_1

- Solve underdetermined linear system

$$\boldsymbol{Y} = \boldsymbol{A}\boldsymbol{X}\boldsymbol{B}^{\mathsf{T}}$$

  - Known matrices: $\boldsymbol{A}, \boldsymbol{B} \in \mathbb{R}^{m \times p}$ with $m < p$

- Using tensor product notations

$$\operatorname{vec}(\boldsymbol{Y}) = \boldsymbol{B} \otimes \boldsymbol{A} \operatorname{vec}(\boldsymbol{X})$$

- l_1 based recovery:

$$\min_{\boldsymbol{X}} \quad \|\operatorname{vec}(\boldsymbol{X})\|_1 \quad \text{subjected to} \quad \boldsymbol{Y} = \boldsymbol{A}\boldsymbol{X}\boldsymbol{B}^{\mathsf{T}}$$

Results from Compressed Sensing

- Guarantees on solution based on restricted isometric properties of $\boldsymbol{B} \otimes \boldsymbol{A}$

$$\delta_r(\boldsymbol{A}) \triangleq \inf \left[\delta : (1-\delta(\boldsymbol{A}))\,\|\boldsymbol{x}\|_2^2 \leq \|\boldsymbol{A}\boldsymbol{x}\|_2^2 \leq (1-\delta(\boldsymbol{A}))\,\|\boldsymbol{x}\|_2^2 \; \|\boldsymbol{x}\|_0 \leq r\right]$$

- RIC of $\boldsymbol{B} \otimes \boldsymbol{A}$ is higher than that of $\boldsymbol{A}$ and $\boldsymbol{B}$
- Proof works when $\boldsymbol{X}$ is very sparse: sparsity $k = \Theta(1)$
- Our focus: Guarantees on recovery when $k = \mathcal{O}(p)$ and $\boldsymbol{X}$ has distributed sparsity

Uniformly Random δ-Left Regular Bipartite Ensemble

- **Bipartite graph**: $G = ([p], [m], E)$
- **Uniform random δ-left regular bipartite graph**: $\forall i \in [p]$ one chooses δ vertices uniformly and independently at random (with replacement) from $[m]$
- **Uniformly random δ-left regular bipartite ensemble**: Adjacency matrix of a uniform random δ-left regular bipartite graph

Main Result

l_1 based recovery:

$$\begin{aligned} \boldsymbol{X}^* = \quad & \underset{\boldsymbol{X}}{\arg\min} && \|\mathrm{vec}\,(\boldsymbol{X})\|_1 \\ & \text{subjected to} && \boldsymbol{Y} = \boldsymbol{A}\boldsymbol{X}\boldsymbol{B}^{\mathsf{T}} \end{aligned}$$

Under following conditions:

- $\boldsymbol{X}$: d-distributed $p \times p$ sparse matrix
- $\boldsymbol{A}, \boldsymbol{B} \in \{0,1\}^{m\times p}$ are drawn independently and uniformly from the δ-random bipartite ensemble
- $\delta = \mathcal{O}(\log p)$

there exists a $c > 0$ such that $\boldsymbol{X}^* = \boldsymbol{X}$ with probability exceeding $1 - p^{-c}$ when

$$m = \mathcal{O}(\sqrt{dp}\log p)$$

- Furthermore, this holds even if $\boldsymbol{A} = \boldsymbol{B}$.

Implications

- Constraint of distributed sparsity need not be factored into the optimization problem

$$\boldsymbol{X}^* = \arg\min_{\boldsymbol{X}} \|\text{vec}\,(\boldsymbol{X})\|_1 \text{ subjected to } \boldsymbol{Y} = \boldsymbol{A}\boldsymbol{X}\boldsymbol{B}^{\mathsf{T}}$$

- Near optimal bound $m = \mathcal{O}(\sqrt{dp}\log p)$
  - logarithm away from the trivial lower bound $\mathcal{O}(\sqrt{dp})$

Distributed Matrix + Perturbation

Under following conditions:

- $\boldsymbol{X}$: arbitrary $p \times p$ matrix
- $\boldsymbol{A}, \boldsymbol{B} \in \{0,1\}^{m \times p}$ are drawn independently and uniformly from the δ-random bipartite ensemble
- $\delta = \mathcal{O}(\log p)$

there exists a $c > 0$ and $\epsilon \in (0, 1/4)$ such that

$$\|\boldsymbol{X}^* - \boldsymbol{X}\|_1 \leq \frac{2-4\epsilon}{1-4\epsilon} \left(\min_{\{\boldsymbol{X}_\Omega : d-\text{distributed}\}} \|\boldsymbol{X} - \boldsymbol{X}_\Omega\|_1 \right)$$

with probability exceeding $1 - p^{-c}$ when $m = \mathcal{O}(\sqrt{dp} \log p)$

- Furthermore, this holds even if $\boldsymbol{A} = \boldsymbol{B}$.

Rectangular Case

- Problem: $\boldsymbol{Y} = \boldsymbol{A}\boldsymbol{X}\boldsymbol{B}^{\mathsf{T}}$
  - Rectangular sparse matrix: $\boldsymbol{X} \in \mathbb{R}^{p_1 \times p_2}$
  - Square measurement matrix: $\boldsymbol{Y} \in \mathbb{R}^{m \times m}$
- Extend previous theorem by padding zeros to sparse matrix to make it square
- The size of sketch is

$$m = \mathcal{O}(\sqrt{dp} \log p)$$

  - $p = \max \{p_1, p_2\}$
  - $d = \max \{\text{row degree, column degree}\}$
- Weak result when $p_2 = 1$

Noisy Measurements

- Model:

$$\boldsymbol{Y} = \boldsymbol{A}\boldsymbol{X}\boldsymbol{B}^{\mathsf{T}} + \boldsymbol{W}$$

where $\boldsymbol{W}_{ij}$ ∼iid zero mean Gaussian noise

- Optimization problem

$$\boldsymbol{X}^* = \arg\min_{\boldsymbol{X}} \left\| \boldsymbol{Y} - \boldsymbol{A}\boldsymbol{X}\boldsymbol{B}^{\mathsf{T}} \right\|_2^2 + \lambda \left\| \mathrm{vec}\left(\boldsymbol{X}\right) \right\|_1$$

- Analysis is an open problem!

Summary

- Notion of distributed sparsity
- A distributed sparse matrix can be recovered from linear model $\boldsymbol{Y} = \boldsymbol{A}\boldsymbol{X}\boldsymbol{B}^{\mathrm{T}}$ via l_1 minimization when sensing matrices are suitable random binary matrices
- Recovery procedure is robust to distributed matrix plus a perturbation