Scalable and Robust PCA Approach with Random Column/Row Sampling

Mostafa Rahmani and George Atia

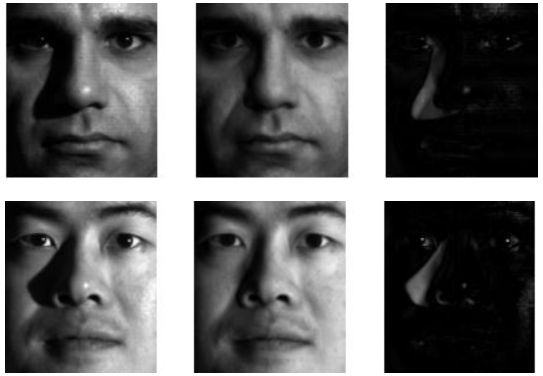
University of Central Florida

Outline

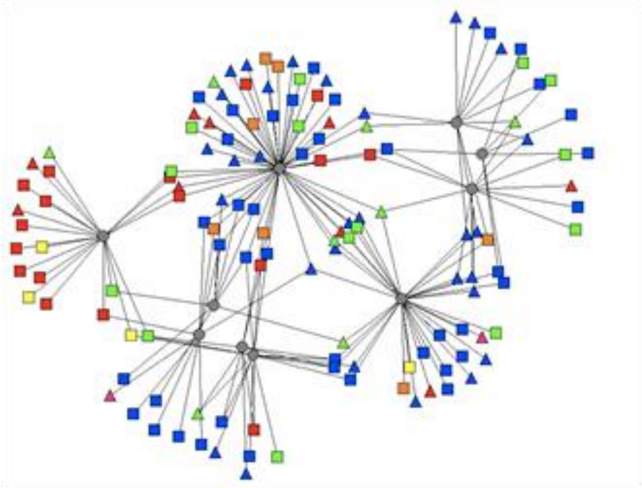
- Robust PCA problem & data corruption models
- Randomized approaches & existing results
- Proposed approach
- New result
- Numerical experiments

Applications

Background Subtraction



Removing Shadow



Network Data Analysis

Phased array systems signal processing

Robust PCA Problem

- Data Model

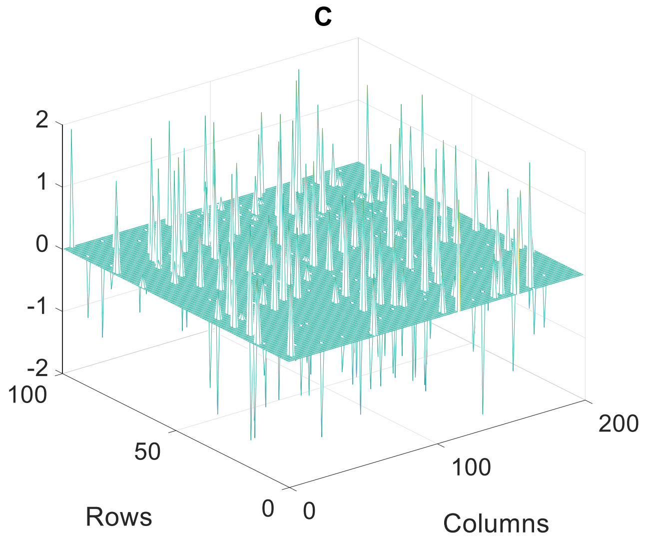
$$\mathbf{D} = \underbrace{\mathbf{L}}_{\text{Low Rank Matrix}} + \underbrace{\mathbf{C}}_{\text{Data Corruption}}$$

- The problem is defined to
  - Learn the column-space/row-space of $\mathbf{L}$
  - Decomposing matrix $\mathbf{D}$

Data Corruption Models

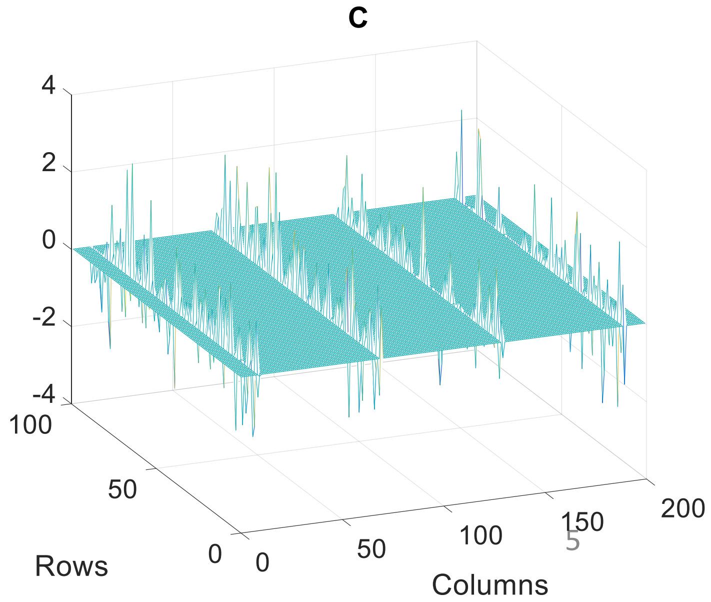
Matrix C

- Element-wise corruption model
  - Matrix $\mathbf{C}$ is a sparse matrix with arbitrary support.
  - All the columns/rows might be affected.
  - Known as low rank plus sparse matrix decomposition problem.



- Column-wise corruption model
  - A subset of the columns of $\mathbf{C}$ are non-zero columns.
  - These non-zero columns do not lie in the Column space of $\mathbf{L}$.
  - Known as subspace recovery or outlier detection problem.

Inlier-Outlier Structure



Algorithms

- Element-wise model
  - Principal Component Pursuit [Chandrasekaran et al. 2011]
  - Alternating minimization [Ke et al. 2005]

$$\min_{\hat{\mathbf{L}},\hat{\mathbf{C}}} \quad \|\hat{\mathbf{L}}\|_* + \lambda \|\hat{\mathbf{C}}\|_1$$
$$\text{subject to} \quad \hat{\mathbf{L}} + \hat{\mathbf{C}} = \mathbf{D}.$$

- Column-wise model
  - Algorithms based on column-sparsity [Xu et al. 2010, Ding et al. 2006]
  - Algorithms based on outliers linear independence [Soltanolkotabi et al, 2012]
  - Algorithm based on low coherency of outliers [Rahmani et. Al, 2016]

Complexity of Robus PCA

$$\mathbf{D} \in \mathbb{R}^{N_1 \times N_2}$$

- Computation complexity

$$\geq O(r\, N_1 N_2 T)$$

- Memory requirement

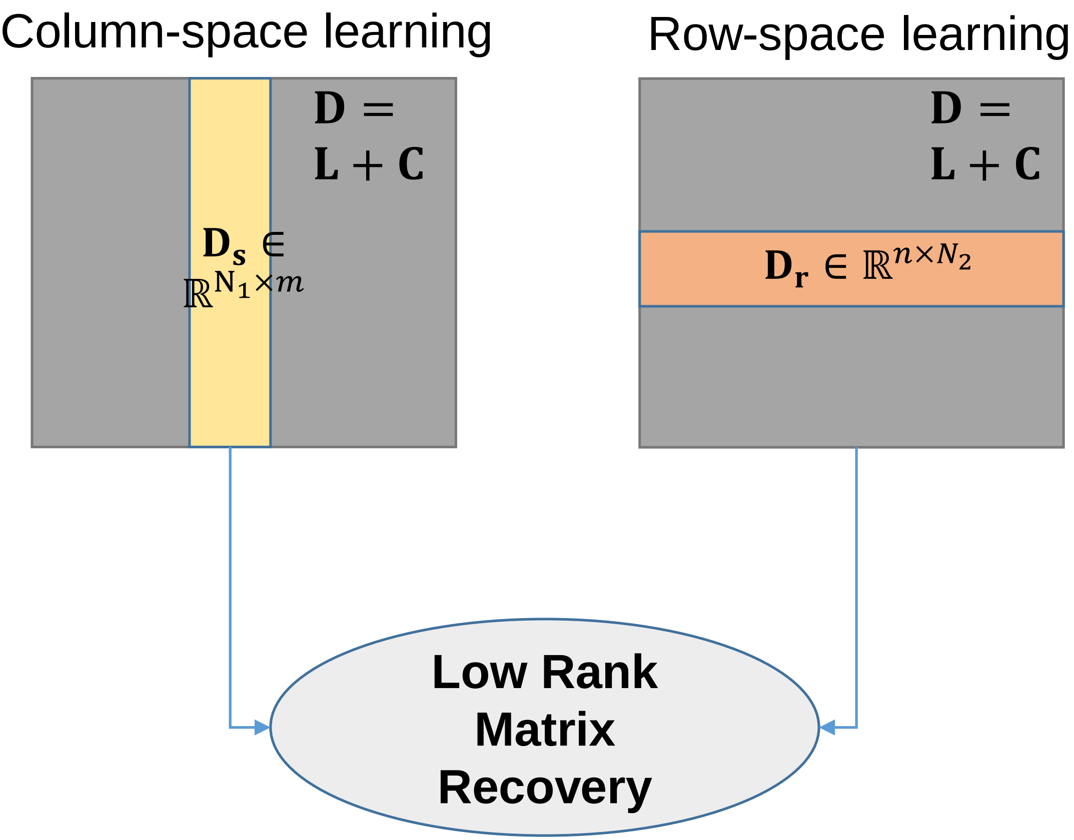
$$O(N_1 N_2)$$

Can we solve the problem with few random linear measurements?

Randomized approach

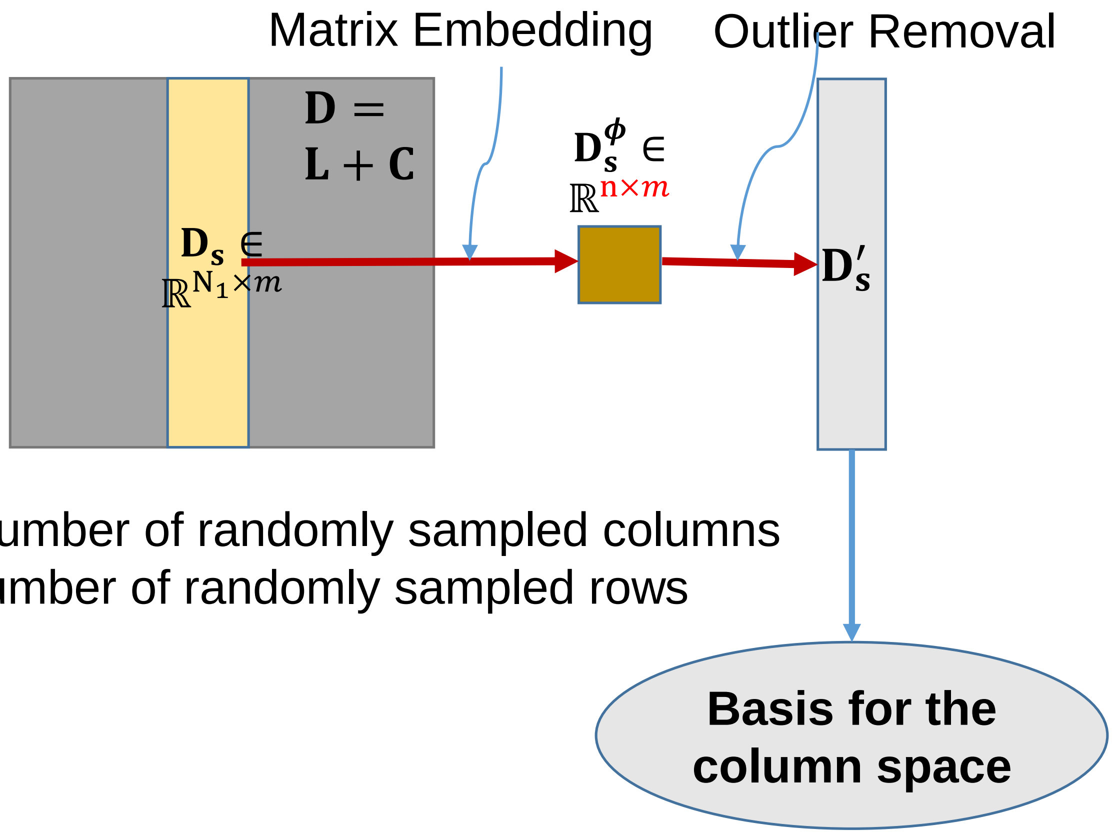
Element wise model
(Matrix Decomposition)

[Mackey et al. 2011, Rahmani et al. 2015]



Column wise model
(Subspace Recovery)

[Li et al., 2014]



m: Number of randomly sampled columns
n: Number of randomly sampled rows

Existing Results

- Elements-wise model (matrix decomposition)
  - Sample complexity $\mathbf{O(r\mu\max(N_1, N_2))}$ [Rahmani et al., 2015]
  - Computation complexity $\boldsymbol{O(r^2\mu\, max(N_1, N_2)T)}$

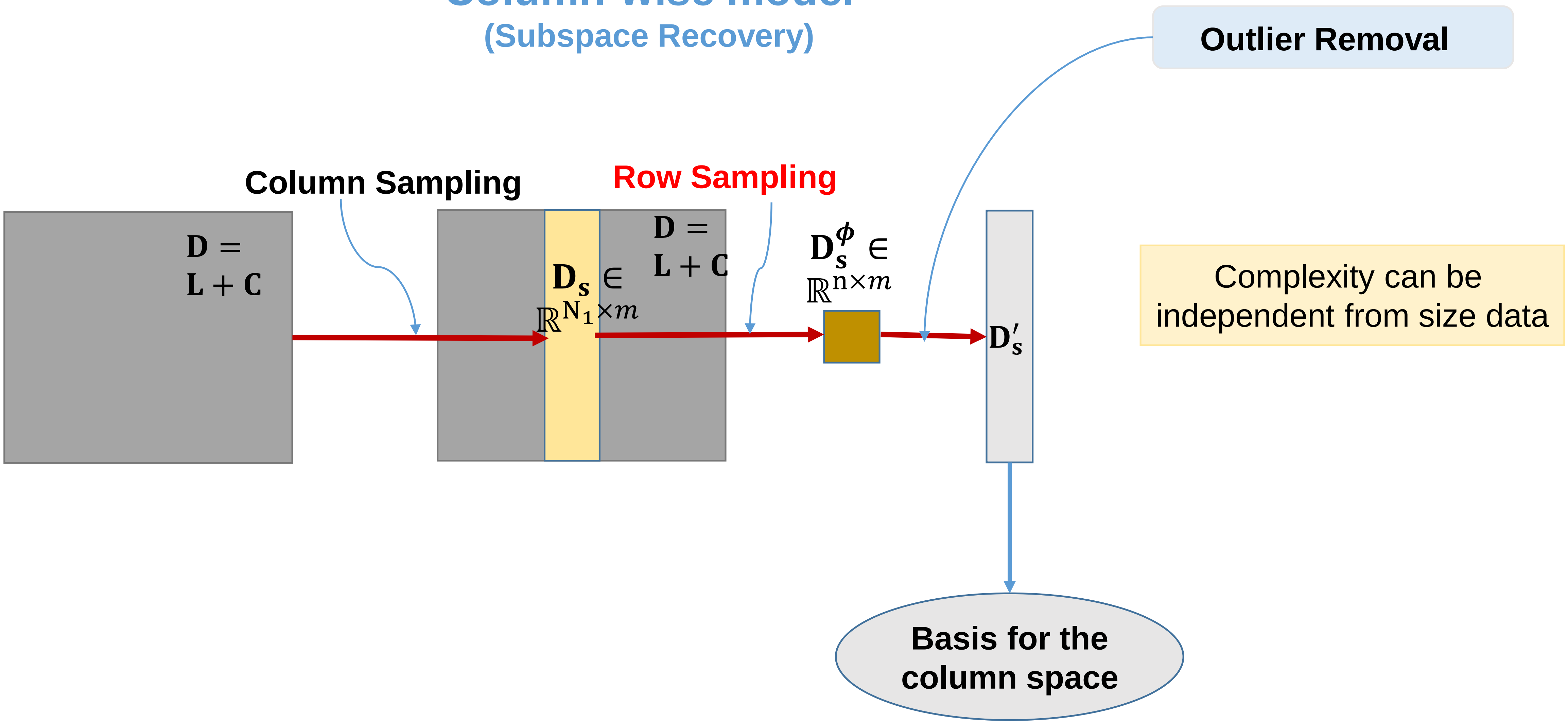
- Column-wise model (Subspace recovery)
  - Sample complexity $\boldsymbol{O(r\mathrm{N}_2)}$ [Li et al., 2015], $\boldsymbol{O(r^2\mu)}$ [Rahmani et al., 2015]
  - Computation complexity $\boldsymbol{O\left(N_1 r^2 \mu + r^3 \mu\right)}$

Motivation

In the column-wise outlier model, can we make the computation complexity of subspace recovery independent from the size of data?

Proposed Randomized Design

Column-wise model
(Subspace Recovery)



Outlier Removal

- ➢Outlier column sparsity

$$\min_{\hat{\mathbf{L}},\hat{\mathbf{C}}} \quad \|\hat{\mathbf{L}}\|_* + \lambda\|\hat{\mathbf{C}}\|_{1,2}$$
$$\text{subject to} \quad \hat{\mathbf{L}} + \hat{\mathbf{C}} = \mathbf{D}$$

- ➢Outlier linear independence

Checking if a column is linearly dependent on other columns or has sparse representation w.r.t them

Performance Guarantee

Data Model

Data Model: The given data matrix $\mathbf{D} \in \mathbb{R}^{\boldsymbol{N_1} \times \boldsymbol{N_2}}$ satisfies the following conditions

1. $\mathbf{D} = \mathbf{L} + \mathbf{C}$ and the columns of $\mathbf{D}$ are normalized.

2. Rank$(\mathbf{L}) = \mathrm{r}$.

3. Matrix $\mathbf{C}$ has K non columns. The non-zero columns of $\mathbf{C}$ are i.i.d. random vectors uniformly distributed on the unit sphere. $\longrightarrow$ $\dfrac{K}{N_2} = \dfrac{\#\ \text{outliers}}{\#\ \text{Columns}}$

Performance Guarantee

Sufficient Conditions, Outlier detection: Column sparsity

Theorem 1: If the given data follows the data model, the columns/rows are sampled randomly, and

$$\frac{K}{N_2} \leq \frac{N_2/2N_2'}{1 + 6r\mu_v(121/9)}$$

$$m \geq \max\left(12\frac{K}{N_2}\left(1+6r\mu_v(121/9)\right)^2 \log\frac{2}{\delta}\ ,\ 10r\mu_v \log\frac{2r}{\delta}\right)$$

$$n \geq \max\left[r\mu_u \max\left(c_1 \log r, c_2 \log\left(\frac{3}{\delta}\right)\right)\ ,\ r + 1 + 2\log 2K/\delta + \sqrt{8\log 2K/\delta}\right]$$

then the proposed method recovers the exact subspace with probability at least $1 - 4\delta$.

Performance Guarantee

Sufficient conditions, Outlier detection: Outlier linear independence

Theorem 2: If data follows the data model, the columns/rows are sampled randomly, and

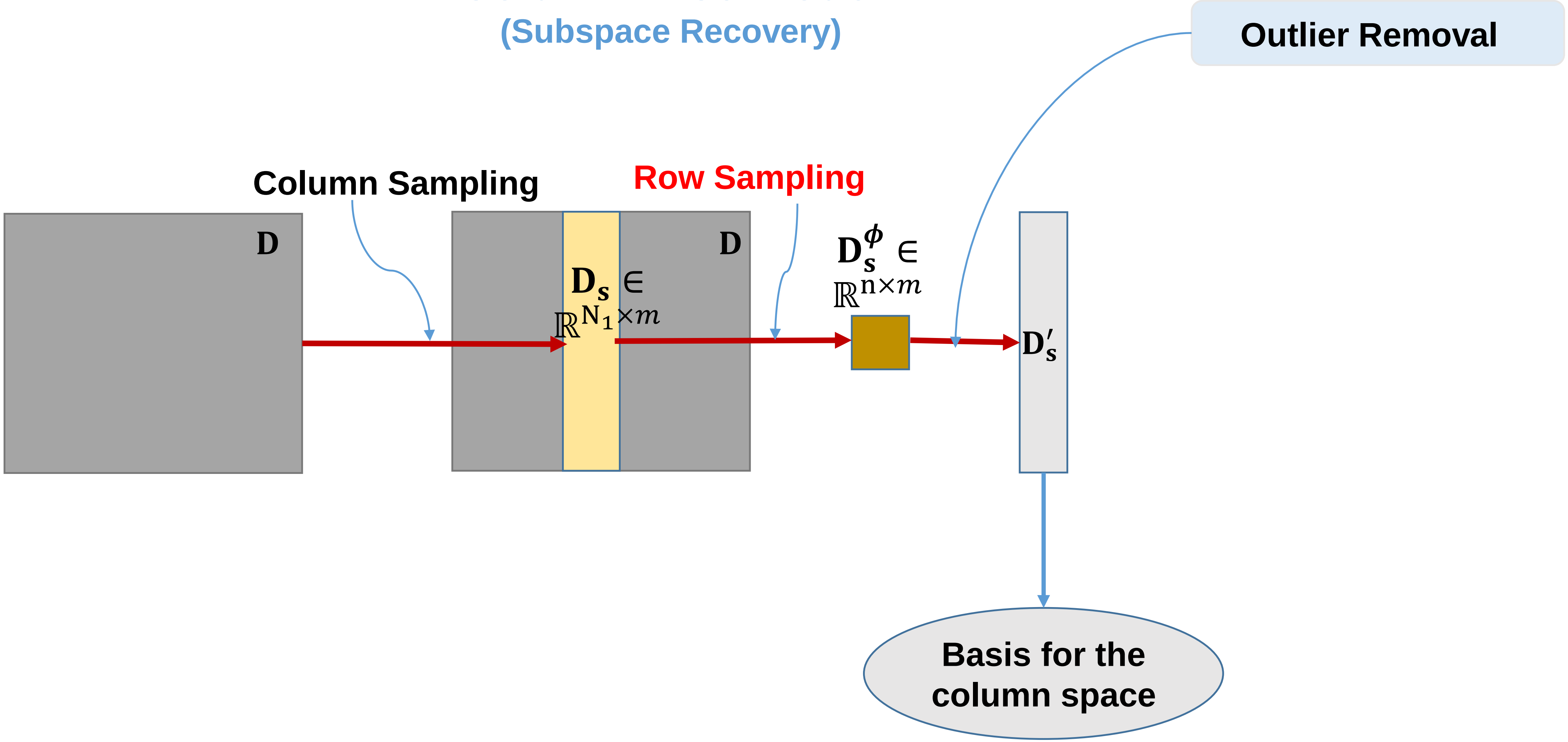
$$m_1 \geq C\,\mu_v r \log \frac{4r}{\delta}$$

$$m_2 \geq \max \left[r\mu_u \max\left(c_1 \log r, c_2 \log\left(\frac{3}{\delta}\right)\right) \,,\, r + q + 2\log\frac{2}{\delta} + \sqrt{8\,q\log\frac{K}{\delta}} \right]$$

then the proposed method recovers the exact subspace with probability at least $1 - 6\delta$.

Proposed Randomized Design

Column-wise model
(Subspace Recovery)



New Result

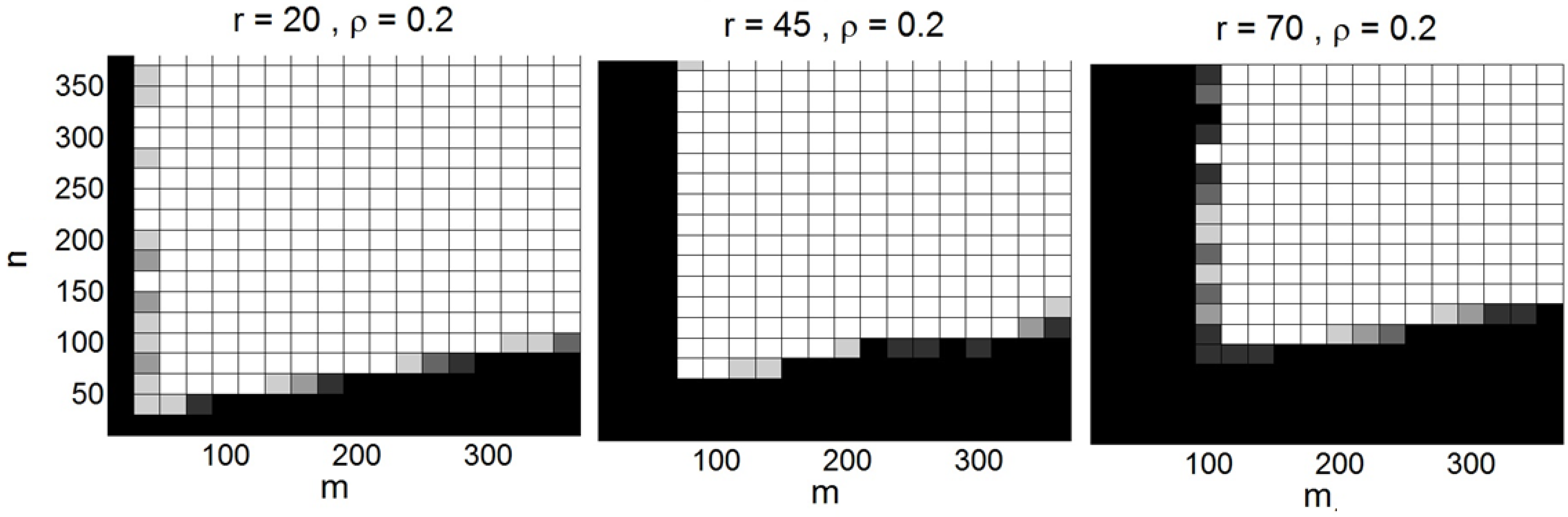
- The computation and sample complexity for exact subspace recovery is almost independent from the size of data.
  - Sample complexity
    - Column sparsity: $O(r^2\mu_u \max(\mu_v, r\mu_v^2 K/N_2))$
    - Linear independence: $O(r^2\mu_v \max(\mu_u, r\mu_v K/N_2))$
  - Computation complexity:
    - Column sparse: $O(rmnT)$
    - Linear independence: $O(rm^2n)$

Both m and n were shown to be independent from data size.

Numerical Experiment-Phase transition

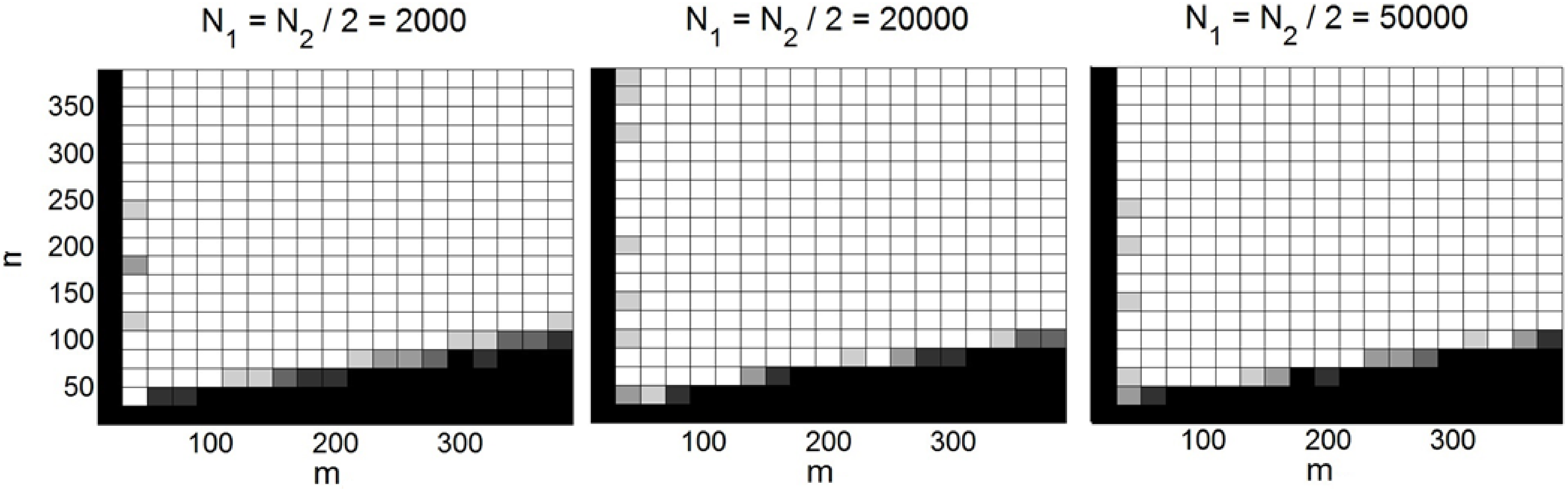
with different values for the rank of $\mathbf{L}$

$\mathbf{D} \in \mathbb{R}^{2000\times4000}$



Numerical Experiment-Phase transition

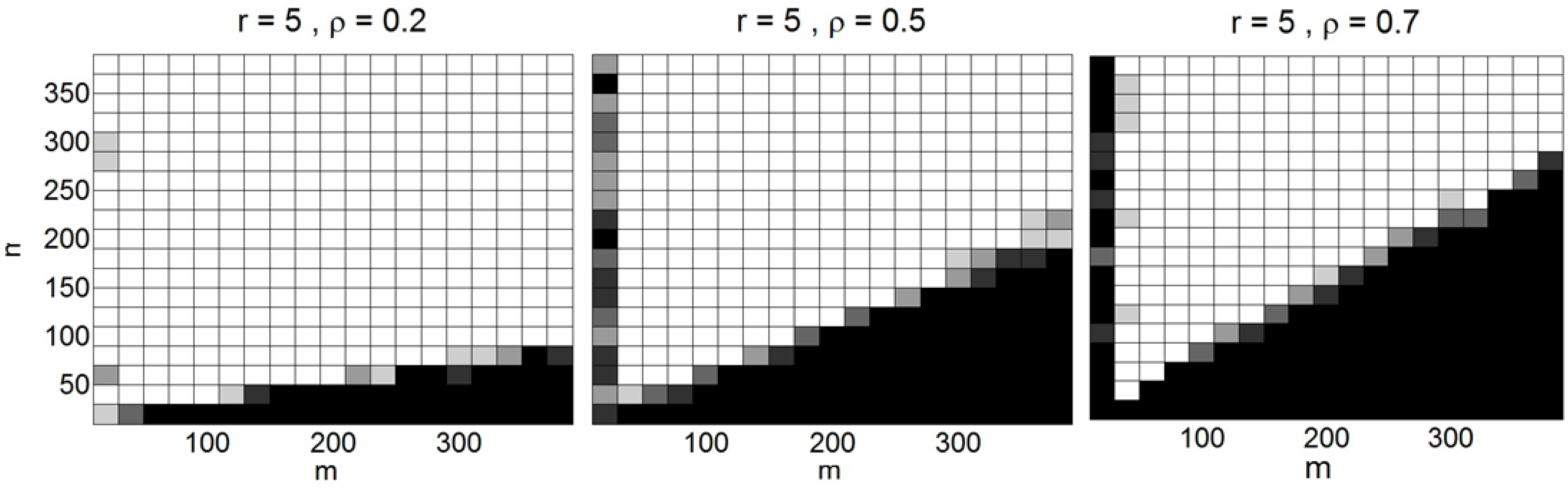
with different data dimensions



Numerical Experiment-Phase transition

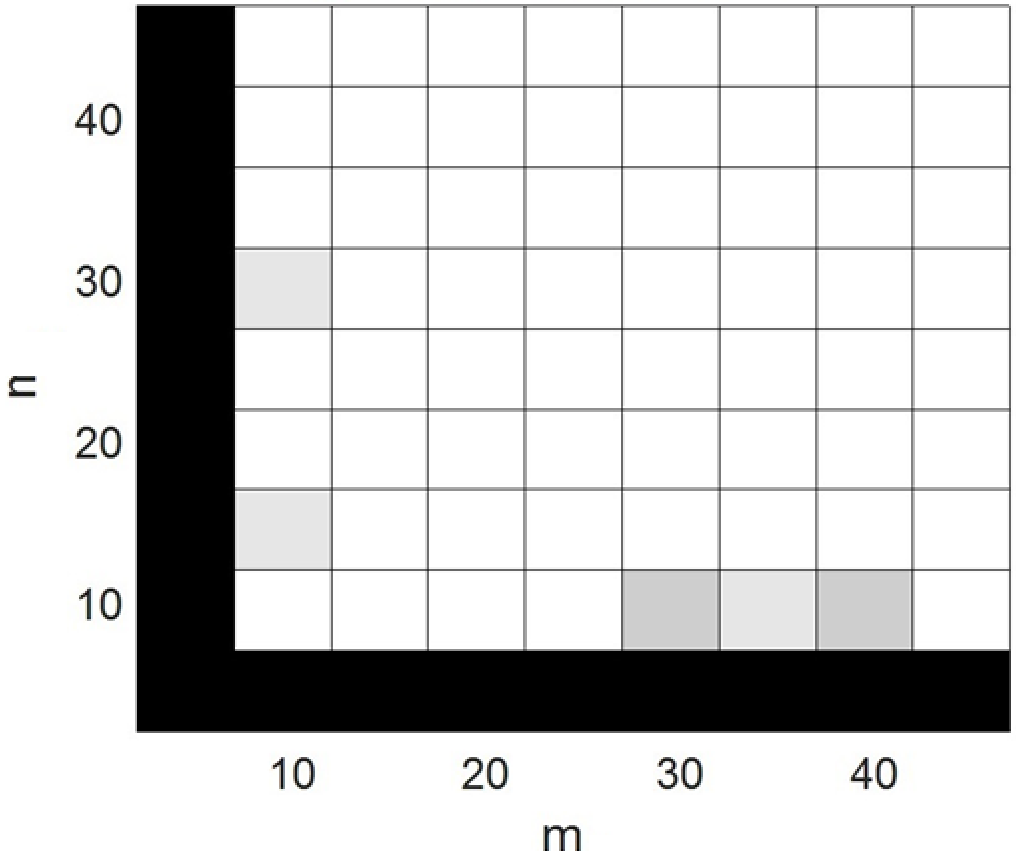
with different $\rho = \frac{K}{N_2}$

$\mathbf{D} \in \mathbb{R}^{2000\times 4000}$

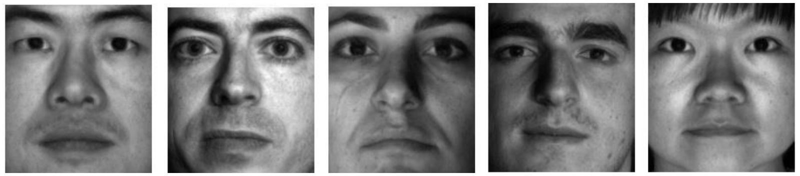


Phase Transition with Real Data

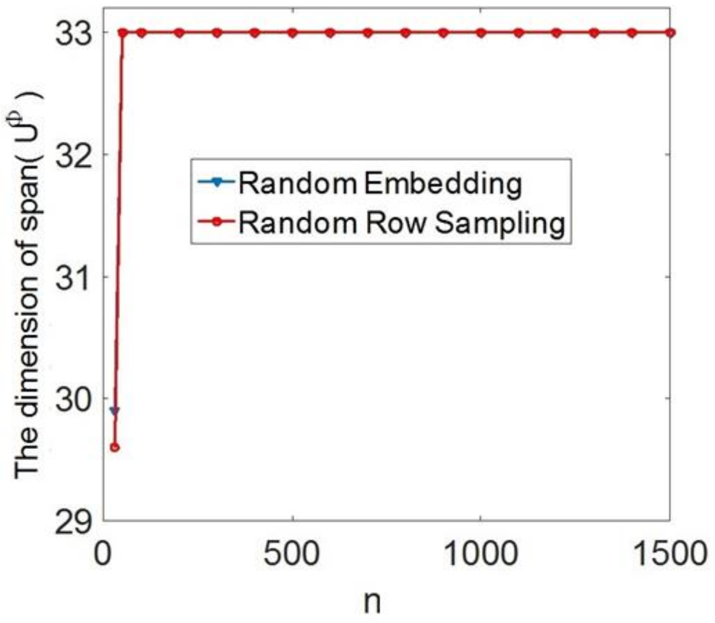
$\mathbf{D} \in \mathbb{R}^{62 \times 512}$
$\mathrm{r} \approx 3$



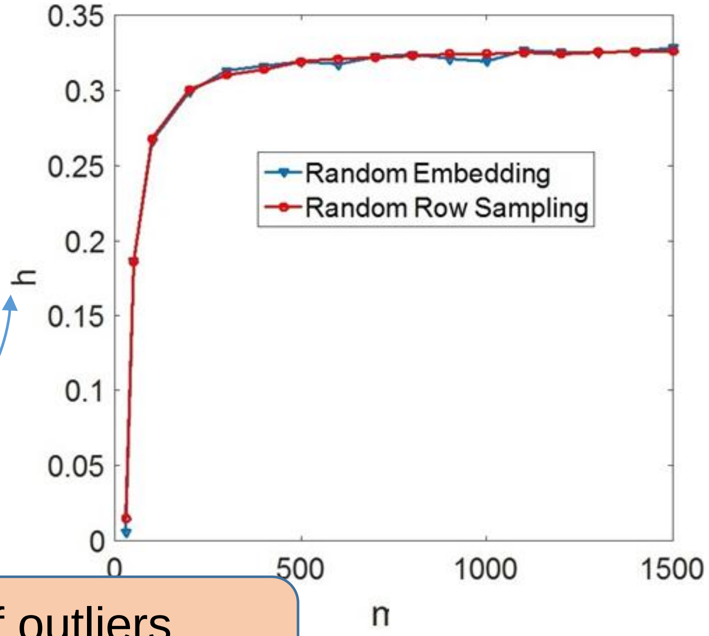
Row Sampling vs Random Embedding



Preserving the low rank component



Preserving the low rank component



The component of outliers
which does not lie in the span of inliers

Thank you.

Questions?!

New Result

- The computation and sample complexity for exact subspace recovery is almost independent from the size of data.
  - Sample complexity
    - Column sparse: $O(r^2\mu_u \max(\mu_v, r\mu_v^2 K/N_2))$
    - Linear independence: $O(r^2\mu_v \max(\mu_u, r\mu_v K/N_2))$
  - Computation complexity:
    - Column sparse: $O(rmnT)$
    - Linear independence: $O(rm^2n)$

Both m and n were shown to be independent from data size.