Nearly Optimal Robust Matrix Completion

Yeshwanth Cherapanamjeri, Kartik Gupta and Prateek Jain

Microsoft Research India

Outline

- Robust Matrix Completion
  - Problem Definition
  - Motivating Applications
- Algorithm
- Theoretical Results
- Empirical Results
- Summary & Future Work

Robust Matrix Completion

1		3			5			5		4	
		5	4			4			2	1	3
2	4		1	2		3		4	3	5	
	2	4		5			4			2	
		4	3	4	2					2	5
1		3		3			2			4	

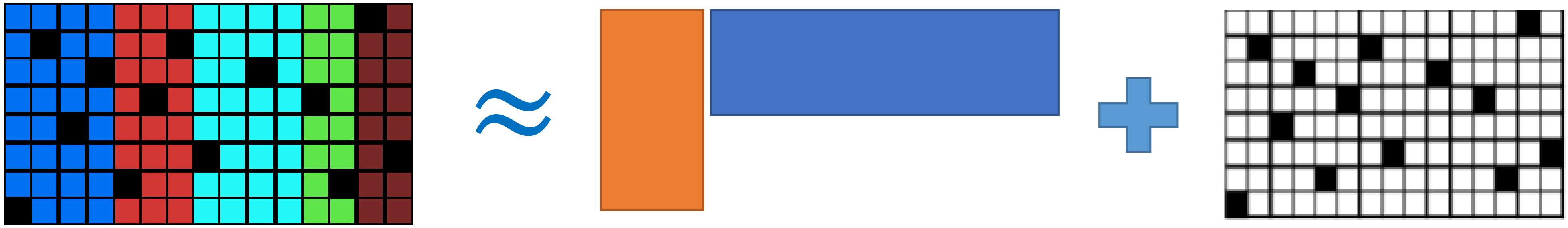
$P_\Omega(M)$

$M \in R^{m \times n}$

$P_\Omega(M) \in R^{m \times n}$

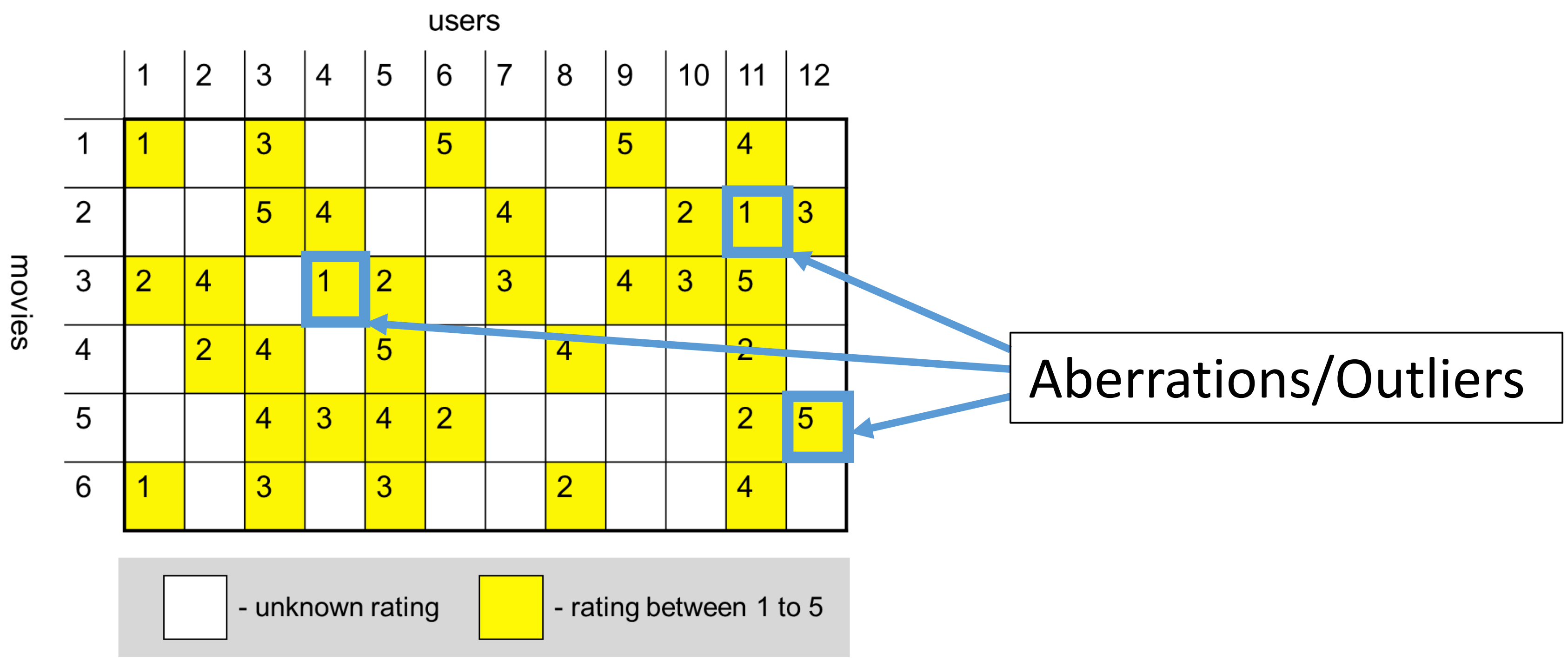
- $\left(P_\Omega(M)\right)_{ij} = M_{ij}, \ (i,j) \in \Omega$
- $\left(P_\Omega(M)\right)_{ij} = 0, \quad (i,j) \notin \Omega$

GOAL: Decompose M as a sum of a low-rank (L^*) and a sparse matrix (S^*)



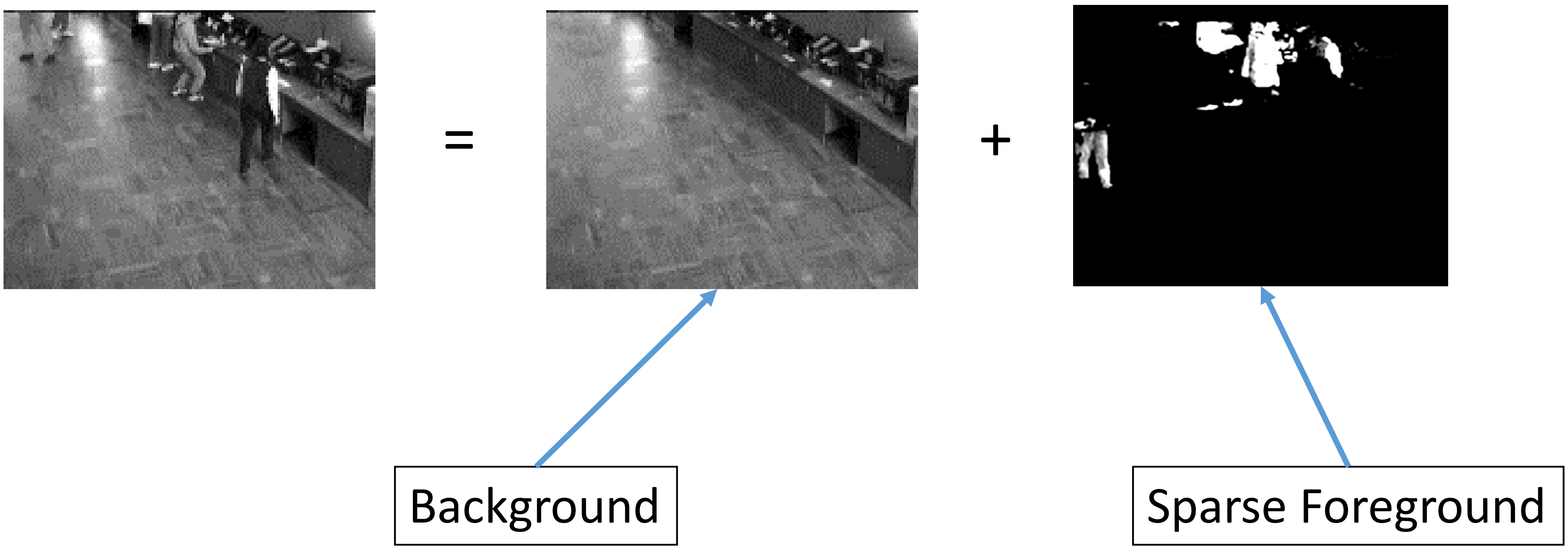
$M \quad \approx \quad L^* \quad + \quad S^*$

Motivating Application 1: Robust Recommender System

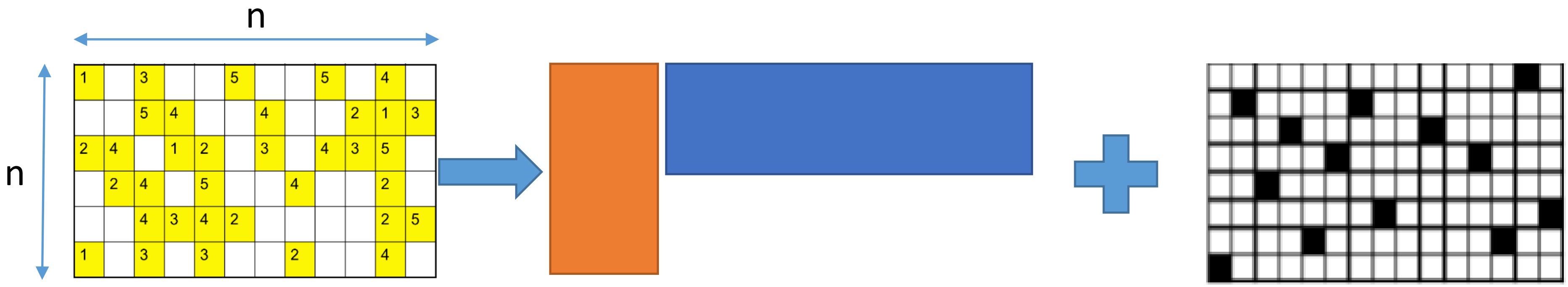


- Random aberrations
- Adversarial outliers: merchant tries to bias ratings towards his products

Motivating Application 2: Background Extraction



Robust Matrix Completion



$Given\colon\ P_\Omega(M)$

$$M \quad \approx \quad L^* \quad + \quad S^*$$

- $\left(P_\Omega(M)\right)_{ij} = M_{ij},\ \ (i,j) \in \Omega$
- $\left(P_\Omega(M)\right)_{ij} = 0, \quad (i,j) \notin \Omega$
- Naturally, $S^* = P_\Omega(S^*)$

$$\min_{L,S} ||P_\Omega(M) - P_\Omega(L+S)||_F^2$$
$$s.t.\ \ rank(L) \le r, |supp(S)| \le s$$

Non-convex Sets

Existing Method

$$\min_{L,\,S} ||P_\Omega(M) - P_\Omega(L+S)||_F^2$$
$$s.t.\ rank(L) \le r, |supp(S)| \le s$$

- $||L||_* = \sum_i \sigma_i(L)$, $||S||_1 = \sum$ $|S_{ij}|$
- Assumption (|… L^*
  - |Ω| is "l…
  - $||S^*||_0$
- R…
- Issues:
  - Convex progra… …g t… (n^3)
  - Requirement o… $|\Omega|, ||S^*||_0$ significantly worse than "optimal" (info. theoretically)

Robust Matrix Completion in:

a) nearly linear time (in n) ?

b) With nearly weakest assumptions?

Our Contribution

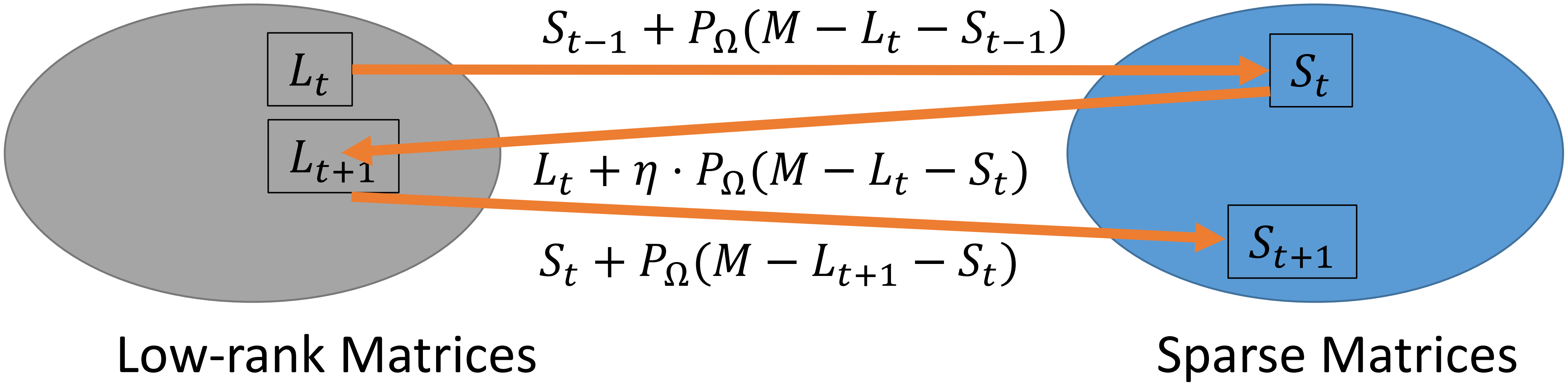
- Simple algorithm that runs in nearly optimal time
  - Based on simple proj. grad. method with projection on non-convex set
  - Scales well to very large problems

- Provable convergence to Global Optima in Linear time
  - Requires:
    - No. of observed entries is “large” enough
    - No. of corruptions is “small” enough
  - Randomness in observations
  - But allows adversarial corruptions

Our Approach: Projected Gradient Descent

$$\min_{L,S} ||P_\Omega(M - L - S)||_F^2$$
$$s.t.\ \ rank(L) \le r, ||S||_0 \le s$$

- Block projected gradient (L, S)
- Gradient: $\mathrm{G} = -P_\Omega(M - L_t - S_{t-1})$
  - Given: $L_t,\ S_{t-1}$



Non-convex RMC

- $L_0 \rightarrow 0$
- $\zeta = \mu^2 r/n$
- For t=1, 2, … T
  - $\zeta = \frac{1}{4} \cdot \zeta$
  - $S_t = HT_\zeta(P_\Omega(M) - P_\Omega(L_t))$
  - $L_{t+1} = P_r(L_t + P_\Omega(M - L_t - S_t))$
- Output, L_T, S_T

Time Complexity

$+ nr^2)$

Linear in the number of observations!

Main Result

Assumptions

- L^*: low-rank matrix
  - $rank(L^*) \le r, L^* \in R^{n \times n}$
  - L^* is μ-incoherent, $\kappa = \sigma_{max}(L^*)/\sigma_{min}(L^*)$
- Ω: set of known entries, random entries
  - $|\Omega| \ge n\mu^4 r^2 \log^3 \frac{n}{\epsilon}$ (optimal: $|\Omega| \ge n\mu^2 r \log n$)
- S^*: corruption matrix
  - $||S_i^*||_0 \le \frac{n}{\mu^2 r}$ (optimal $\left|\left|S_i^*\right|\right|_0 \le \frac{n}{\mu^2 r}$)

Result [CGJ-ICML'17]

- $T = \log\left(\frac{1}{\epsilon}\right)$

$$||L_T - L^*||_\infty \le \epsilon$$

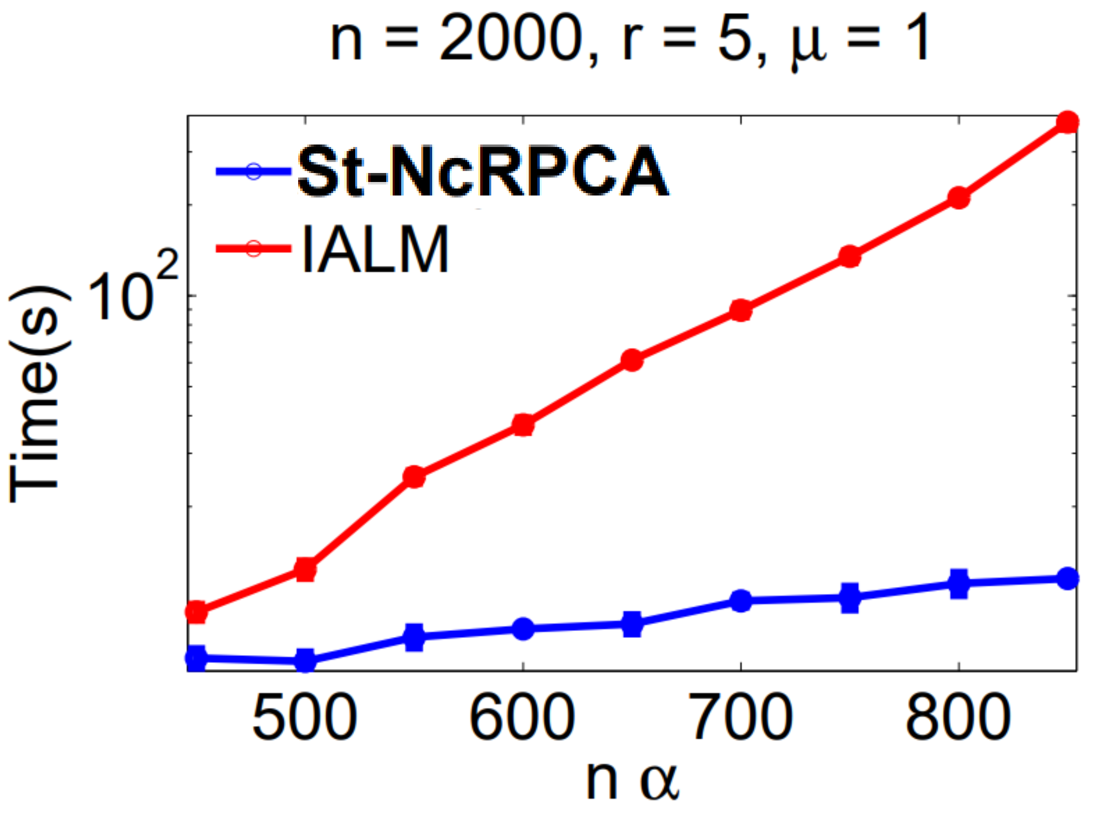
- Total running time:

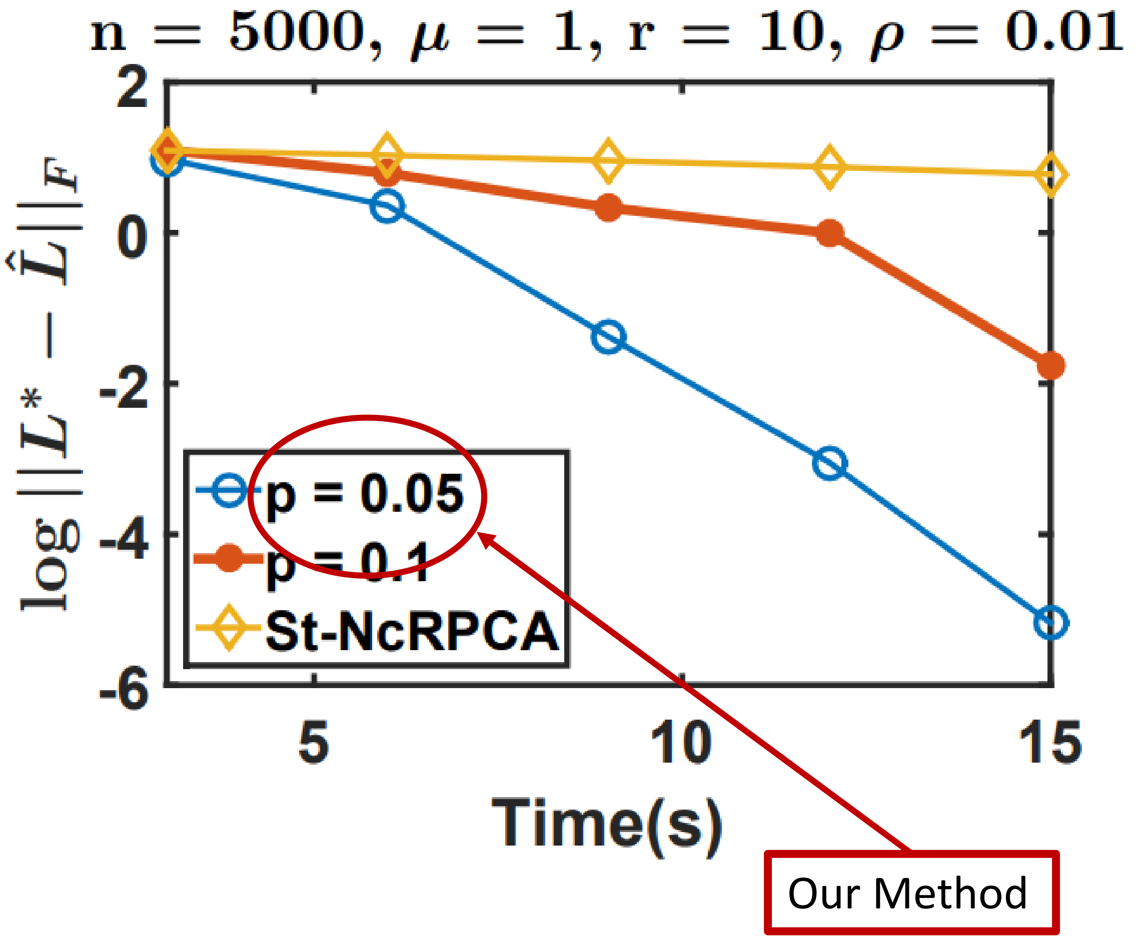
$$O\left(\left(|\Omega| \cdot \mathrm{r} + \mathrm{nr}^2\right) \cdot \log\frac{1}{\epsilon} \cdot \log \kappa\right)$$

Independent result [Yi et al'2016]:

$$|\Omega| \ge c \cdot n\kappa^4 \mu^2 r^2 \log n\,, \qquad ||S_i^*||_0 \le n \cdot min\left\{\frac{1}{\mu\sqrt{\kappa r}^3}, \frac{1}{\mu\kappa^2 r}\right\}$$

Empirical Results (Synthetic Datasets)





Empirical Results

Convex Method Runtime: 1700 sec


=

+

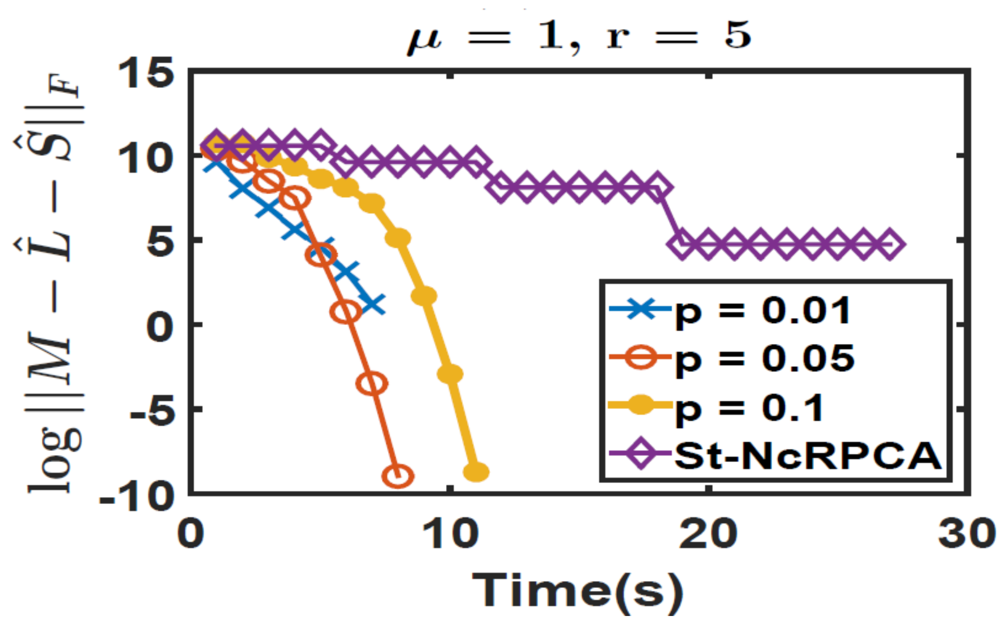

Non-Convex Method [NUSAJ-NIPS’14] Runtime: 28 sec


=

+


Empirical Results

New Non-Convex Method. Runtime: 8 sec



Foreground Extraction methods from Open CV
Runtime: 6 sec, 4 sec

Summary

- Robust Matrix Completion
    - Low-rank+Sparse Decomposition with Missing Entries
- Projected Gradient Based Method to solve Robust Matrix Completion
- Almost linear convergence rate under standard assumption
- Near optimal run-time, sample complexity, number of corruptions

Future Directions

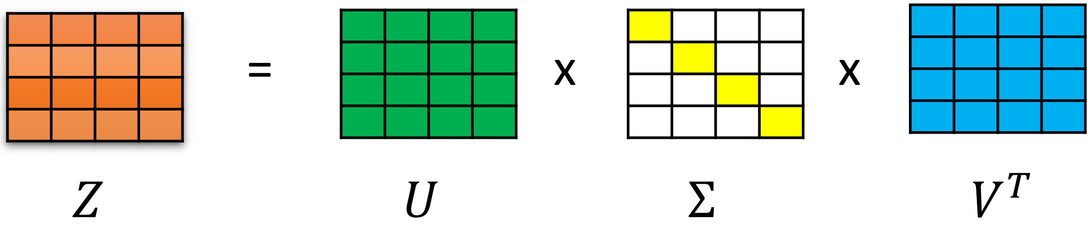
- Exploit more information about the corruption
    - Corruptions are zero mean with random support?
- General Problem: learning in high-dimensions
    - Design algorithms that are robust to outliers/corruptions
    - General theory of provable non-convex optimization

Questions?

Projection onto Low-rank Matrices

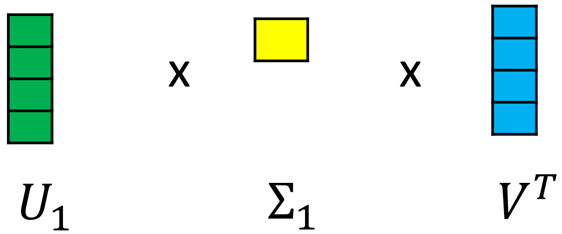
- Non-convex projections: NP-hard in general
- But $P_r(Z)$ can be computed efficiently:

$$Z = U\Sigma V^T$$



- $P_r(Z) = U_r \Sigma_r V_r^T$

$\boldsymbol{P_1(Z)}$ =



- Time complexity: $O(n^2 r)$
  - Using special structure in our updates: $O(|\Omega| r + n r^2)$

Projection onto Sparse Matrices

- Non-convex projection
- $HT_\zeta(Z)$: removes all elements with magnitude smaller than ζ

1	0.1	0.22
0.1	0.01	.9
0.11	0.02	0.12

$$\xrightarrow{HT_{0.5}}$$

1	0	0
0	0	.9
0	0	0