Non-convex Robust PCA

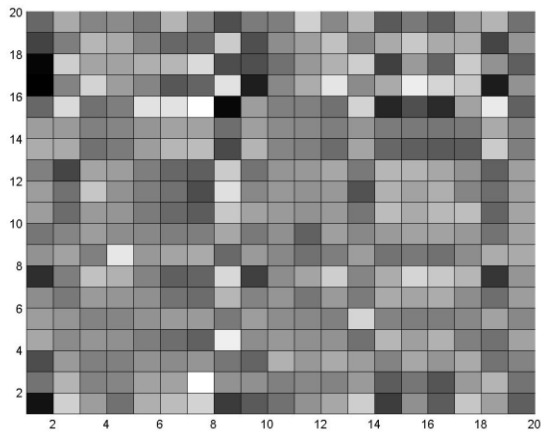
Praneeth Netrapalli

Microsoft Research

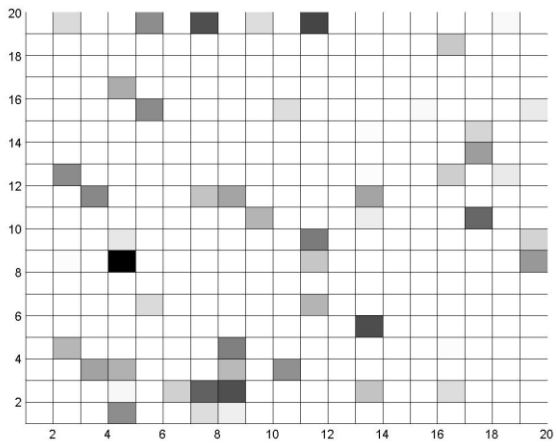
Joint w/: U. N. Niranjan, S. Sanghavi, A. Anandkumar, P. Jain

Robust PCA

$M = L^* + S^*$



Unknown Low rank Matrix



Unknown Sparse Matrix

Task: Given *M*, find *L** and *S**

Motivation: Low-rank approximation to most but not all the data …

An Application

Foreground-Background Separation



Frames in a video → Background (Low rank) + Foreground (Sparse)

Robust PCA: A brief history

Formulation: $\min \quad rank(L) + \lambda \|S\|_0$

$$s.t. \quad L + S = M$$

Convex relaxation:

$$\min \|L\|_* + \lambda \|S\|_1$$

$$s.t. \quad L + S = M$$

$$\lambda \approx \frac{1}{\sqrt{n}}$$

[Chandrasekaran, Sanghavi, Parrilo, Willsky (2009)] &

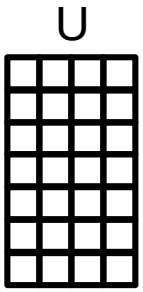
[Candes, Li, Wright, Ma (2009)]:

Certain pairs of *L,S* can be provably recovered…

Robust PCA: A brief history

(deterministic) Conditions on L*,S*:

1. Low-rank matrix is **incoherent**: if $L^* = U\Sigma V^T$ then



$$\max_i \|U_i\|_2 \leq \mu\sqrt{\frac{r}{n}} \qquad \left(1 \leq \mu \leq \sqrt{\frac{n}{r}}\right)$$

Similarly V

2. Sparse matrix is **degree bounded**:

Any row or column of *S** has at most *d* non-zero elements

[Hsu, Kakade, Zhang 2011]: *L*,S** unique optimum of convex program if

$$d \leq c\frac{n}{\mu^2 r}$$ (tight up to constants)

Robust PCA: A brief history

The problem: scale …

Special purpose convex solvers for this need

$O(m^2 n)$ complexity per iteration

$O\left(\frac{1}{\epsilon}\right)$ Iterations to get to an error of ϵ

Our Work: a **new non-convex algorithm** that takes

$O(r^2 mn)$ complexity per iteration

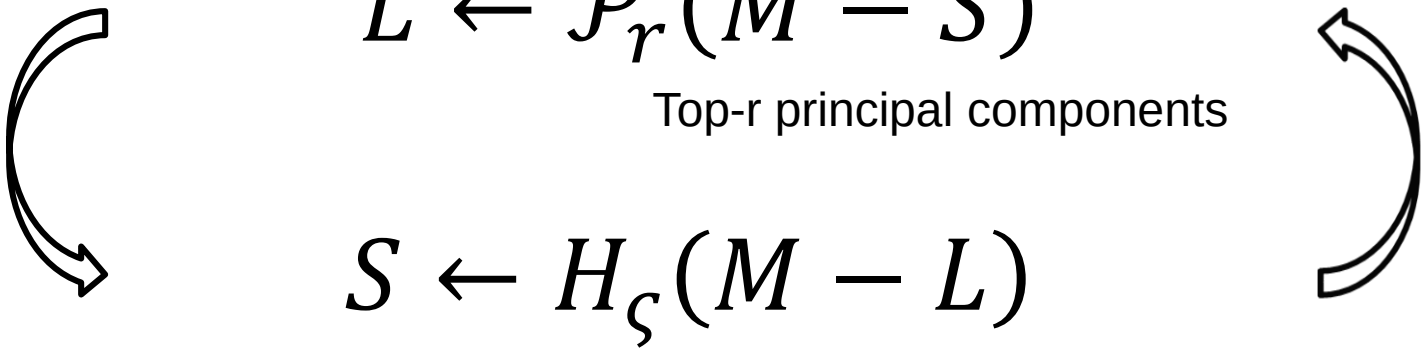
$O\left(\log \frac{1}{\epsilon}\right)$ Iterations to get to an error of ϵ

} Factor r away from the complexity of vanilla PCA …

… and **provably recovers** the same pairs of L*,S* that convex methods do.

Our Approach

Basic idea: alternating non-convex (but fast) projections



$$L \leftarrow \mathcal{P}_r(M - S)$$

Top-r principal components

$$S \leftarrow H_\varsigma(M - L)$$

Retain the largest elements

We track the errors $\|L - L^*\|_\infty$ and $\|S - S^*\|_\infty$ and show they decrease

Key Lemma

If $L = \mathcal{P}_r(L^* + S^*)$ and $\|S^*\|_\infty \leq \frac{8\mu^2 r}{n}\sigma_r$ then

$$\|L - L^*\|_\infty \leq \frac{1}{4}\|S^*\|_\infty$$

Recall: $S \leftarrow H_\varsigma(M - L)$

Once we have this lemma:

- Choose $\varsigma = \frac{1}{4}\left(\frac{8\mu^2 r}{n}\sigma_r\right)$
- This ensures $Supp(S) \subseteq Supp(S^*)$ and $\|S - S^*\|_\infty \leq \frac{1}{2}\left(\frac{8\mu^2 r}{n}\sigma_r\right)$
- Apply the lemma to $\mathcal{P}_r(M - S) = \mathcal{P}_r(L^* + (S^* - S))$

Proof Outline

If $L = \mathcal{P}_r(L^* + S^*)$ and $\|S^*\|_\infty \leq \frac{8\mu^2 r}{n}\sigma_r$ then

$$\|L - L^*\|_\infty \leq \frac{1}{4}\|S^*\|_\infty$$

- Sparsity of S^* $\Longrightarrow$ Bound on $\|S^*\|_2$
- Weyl's inequalities give : $\|L - L^*\|_2 \leq \|S^*\|_2$ and $\|L - L^*\|_F \leq \|S^*\|_F$
- Challenge : Bounds on ℓ_∞- norm much harder!

Rank-1 Case

$$\lambda u u^T = L = \mathcal{P}_1(u_* u_*^T + S^*)$$

Sparsity of S^* $\Longrightarrow$ $\|S^*\|_2 \leq 0.1$ $\Longrightarrow$ $\begin{cases} 1.1 \geq \lambda \geq 0.9 \\ u_*^T u \geq 0.9 \end{cases}$

$$(u_* u_*^T + S^*)u = \lambda u$$

$$\left(I - \frac{S^*}{\lambda}\right)u = \left(\frac{u_*^T u}{\lambda}\right)u_*$$

$$u = \left(\frac{u_*^T u}{\lambda}\right)\left(I - \frac{S^*}{\lambda}\right)^{-1} u_*$$

$$= \left(\frac{u_*^T u}{\lambda}\right)\left(I + \frac{S^*}{\lambda} + \left(\frac{S^*}{\lambda}\right)^2 + \cdots\right)u_*$$

Rank-1 Case Contd.

$$u = \left(\frac{u_*^T u}{\lambda}\right)\left(I + \frac{S^*}{\lambda} + \left(\frac{S^*}{\lambda}\right)^2 + \cdots\right)u_*$$

Sparsity of S^*
Incoherence of u_* $\Big\}$ $\Longrightarrow$ $\|(S^*)^p u_*\|_\infty \leq \left(\frac{1}{2}\right)^p \|u_*\|_\infty$

Controlling $\|u\|_\infty$, $\|u - u_*\|_\infty$ etc. very easy!

General Version

If $L = \mathcal{P}_r(L^* + S^*)$ and $\|S^*\|_\infty \leq \frac{8\mu^2 r}{n}\sigma_r$ then

$$\|L - L^*\|_\infty \leq \frac{1}{4}\|S^*\|_\infty$$

More General Lemma

If $L = \mathcal{P}_k(L^* + S^*)$ and $\|S^*\|_\infty \leq \frac{8\mu^2 r}{n}\sigma_k$ then

$$\|L - L^*\|_\infty \leq \frac{1}{4}\|S^*\|_\infty + \frac{2\mu^2 r}{n}\sigma_{k+1}$$

Our Algorithm

Do stage-wise:

For k = 1 … r

For t = 1 … $\log\left(\frac{1}{\epsilon}\right)$

$$L \leftarrow \mathcal{P}_k(M - S)$$

Top-k principal components

$$S \leftarrow H_\varsigma(M - L)$$

Retain the largest elements

end

end

Theorem

Theorem:

For a rank-r matrix L* that is μ incoherent, and *S** that has at most

$\frac{n}{512\,\mu^2 r}$ elements per row and column, we have

$$Supp(S_t) \subseteq Supp(S^*)$$

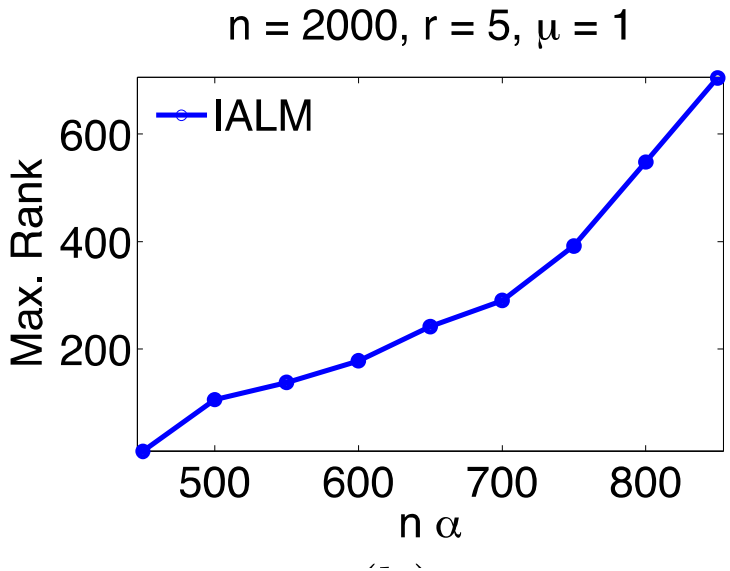
$$\|S_t - S^*\|_\infty \leq \frac{\epsilon}{n}$$

$$\|L_t - L^*\|_\infty \leq \frac{\epsilon}{n}$$

Comparison to Convex Solvers

IALM [Lin-Chen-Ma] : fastest / most popular special-purpose convex algorithm

Iteratively projects onto a nuclear norm ball
- complexity grows with the intermediate rank needed for projection

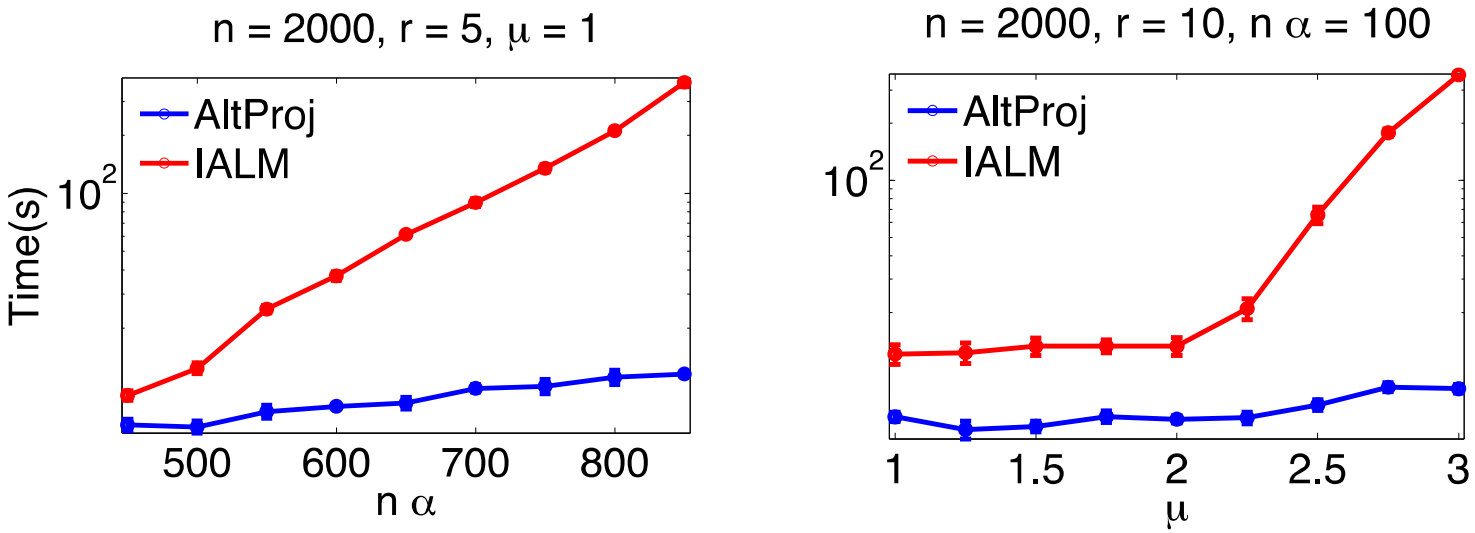


Even for maximally incoherent
low-rank matrices, high intermediate rank

- Is the primary driver of high complexity per iteration.

Comparison to Convex Solvers

Speed comparison, synthetic example. Single laptop.



Videos …

Original

Non-convex (Ours!)

Convex relaxation

Vanilla PCA

Summary

A new simple non-convex algorithm for Robust PCA

alternating non-convex projections

factor r slower than vanilla PCA

much faster than convex Robust PCA

Meta-goal: Find out to what extent convex optimization really gives **provable** gains over (often existing, popular, faster) non-convex methods in Machine learning

Thanks!