Matrix vs. Tensor Methods: Robustness to Block Sparse Perturbations

Anima Anandkumar

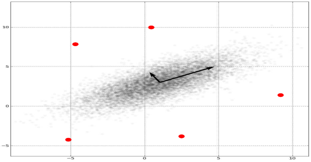
U.C. Irvine

Joint work with Prateek Jain, Yang Shi, and U.N. Niranjan.

Denoising Big Data

Denoising: Remove noise to reveal hidden structures in data.

Classical analysis: PCA



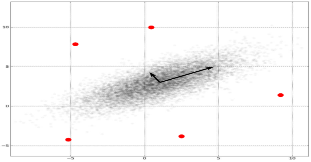
- Pros: Efficient computation, requires only pairwise correlation.
- Cons: Not robust to even a few outliers

Denoising Big Data

Denoising: Remove noise to reveal hidden structures in data.

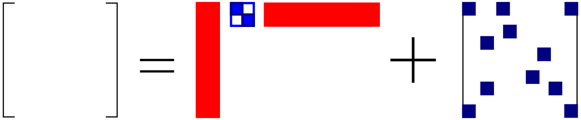
Classical analysis: PCA

- Pros: Efficient computation, requires only pairwise correlation.
- Cons: Not robust to even a few outliers



Robust PCA

- Find low rank structure after removing sparse corruptions.
- Decompose input matrix as low rank + sparse matrices.



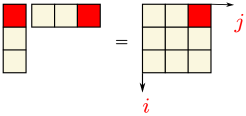
M L^* S^*

- Applications in computer vision, community modeling, ...

From Matrices to Tensors

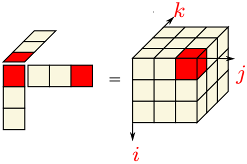
Matrix: Two dimensional

- $M \in \mathbb{R}^{d \times d}$.
- Rank 1: $M = a \otimes b$, $M_{i,j} = a_i b_j$.

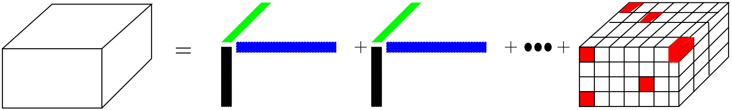


Tensor: Multi dimensional

- $T \in \mathbb{R}^{d \times d \times d}$.
- Rank 1: $T = a \otimes b \otimes c$, $T_{i,j,k} = a_i b_j c_k$.



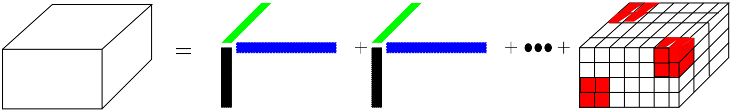
Robust Tensor PCA



Given T, find L, S such that:

$$T = L + S, \quad L = \sum_{i=1}^{r} \sigma_i^* u_i \otimes u_i \otimes u_i, \|S\|_0 \leq s$$

Why is Robust PCA difficult?



Natural constraints for identifiability?

- Low rank tensor is NOT sparse and viceversa.
- Incoherent low rank tensor: $\|u_i\|_\infty \leq \frac{\mu}{\sqrt{n}}$.
- Sparse tensor: Block sparsity pattern of size d with B blocks and overlap fraction η, ie,

$$\text{supp}(S) = \sum_{i=1}^{B} \psi_i \otimes \psi_i \otimes \psi_i,$$

$$\|\psi_i\|_0 \leq d, \max_{i \neq j} \langle \psi_i, \psi_j \rangle \leq \eta d, \; \psi_i(j) = 0 \text{ or } 1$$

Summary of Results

- Convex relaxation of tensor CP rank is NP-hard.
- Propose an efficient non-convex method based on alternating projections.
- Prove convergence to globally optimal solution.
- Prove that tensor method is more robust compared to matrix robust PCA under block sparsity.

Outline

1. Introduction
2. Alternating Projections Algorithm
3. Analysis
4. Experiments
5. Conclusion

Proposal for Non-convex Robust PCA

$$\boxed{T = L + S, \quad \text{Rank}(L) = r, \|S\|_0 \leq s}$$

Proposal for Non-convex Robust PCA

$$\boxed{T = L + S, \quad \mathrm{Rank}(L) = r, \|S\|_0 \le s}$$

A non-convex heuristic (AltProj)

- Initialize $L, S = 0$ and iterate:
- $\boxed{L \leftarrow P_r(T - S)}$ and $\boxed{S \leftarrow H_\zeta(T - L)}$.
- $P_r(\cdot)$: rank-r projection. $H_\zeta(\cdot)$: thresholding with ζ.

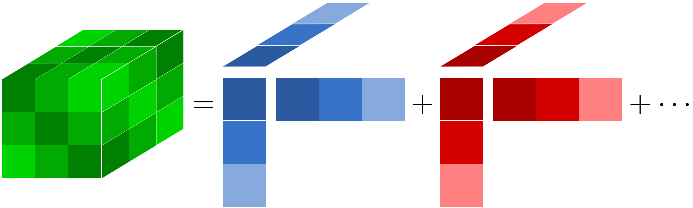
Proposal for Non-convex Robust PCA

$$\boxed{T = L + S, \quad \mathrm{Rank}(L) = r, \|S\|_0 \leq s}$$

A non-convex heuristic (AltProj)

- Initialize $L, S = 0$ and iterate:
- $\boxed{L \leftarrow P_r(T - S)}$ and $\boxed{S \leftarrow H_\zeta(T - L)}$.
- $P_r(\cdot)$: rank-r projection. $H_\zeta(\cdot)$: thresholding with ζ.
- how to find rank-r projections for tensors?

Tensor Power Iterations



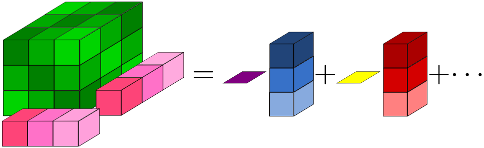
$$L = u_1^{\otimes 3} + u_2^{\otimes 3} + \cdots ,$$

A., R. Ge, D. Hsu, S. Kakade, M. Telgarsky, "Tensor Decompositions for Learning Latent Variable Models," JMLR 2014.

Tensor Power Iterations

Tensor Power Method

$$\boxed{v \mapsto \frac{L(v,v,\cdot)}{\|L(v,v,\cdot)\|}.}$$



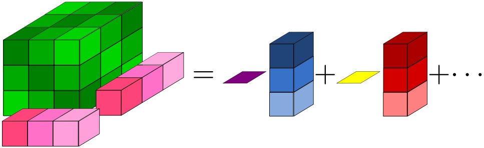
$$L(v,v,\cdot) = \langle v, u_1\rangle^2 u_1 + \langle v, u_2\rangle^2 u_2$$

A., R. Ge, D. Hsu, S. Kakade, M. Telgarsky, "Tensor Decompositions for Learning Latent Variable Models," JMLR 2014.

Tensor Power Iterations

Tensor Power Method

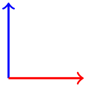
$$v \mapsto \frac{L(v,v,\cdot)}{\|L(v,v,\cdot)\|}.$$



$$L(v,v,\cdot) = \langle v, u_1\rangle^2 u_1 + \langle v, u_2\rangle^2 u_2$$

Orthogonal Tensors

- $\vec{u}_1 \perp \vec{u}_2$.
- $L(u_1, u_1, \cdot) = \lambda_1 u_1$.

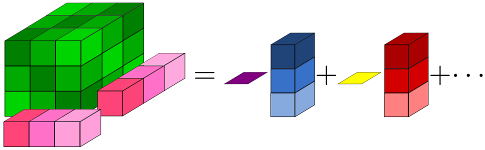


A., R. Ge, D. Hsu, S. Kakade, M. Telgarsky, "Tensor Decompositions for Learning Latent Variable Models," JMLR 2014.

Tensor Power Iterations

Tensor Power Method

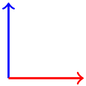
$$v \mapsto \frac{L(v,v,\cdot)}{\|L(v,v,\cdot)\|}.$$

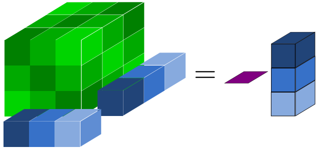


$$L(v,v,\cdot) = \langle v, u_1 \rangle^2 u_1 + \langle v, u_2 \rangle^2 u_2$$

Orthogonal Tensors

- $\vec{u}_1 \perp \vec{u}_2$.
- $L(u_1, u_1, \cdot) = \lambda_1 u_1$.



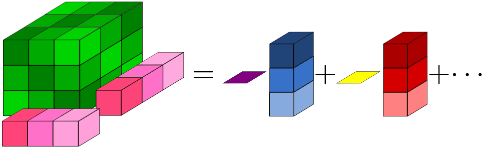


A., R. Ge, D. Hsu, S. Kakade, M. Telgarsky, "Tensor Decompositions for Learning Latent Variable Models," JMLR 2014.

Tensor Power Iterations

Tensor Power Method

$$v \mapsto \frac{L(v,v,\cdot)}{\|L(v,v,\cdot)\|}.$$



$$L(v,v,\cdot) = \langle v, u_1\rangle^2 u_1 + \langle v, u_2\rangle^2 u_2$$

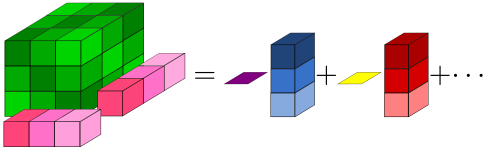
Exponential no. of stationary points for power method:
$L(v,v,\cdot) = \lambda v$

A., R. Ge, D. Hsu, S. Kakade, M. Telgarsky, "Tensor Decompositions for Learning Latent Variable Models," JMLR 2014.

Tensor Power Iterations

Tensor Power Method

$$v \mapsto \frac{L(v,v,\cdot)}{\|L(v,v,\cdot)\|}.$$

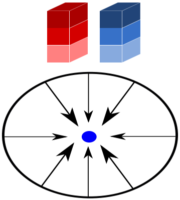


$$L(v,v,\cdot) = \langle v, u_1\rangle^2 u_1 + \langle v, u_2\rangle^2 u_2$$

Exponential no. of stationary points for power method:

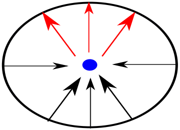
$L(v,v,\cdot) = \lambda v$

Stable



Unstable

Other statitionary points

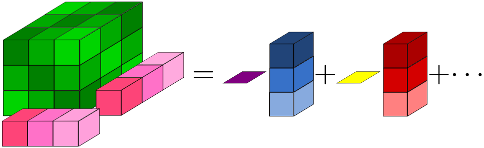


A., R. Ge, D. Hsu, S. Kakade, M. Telgarsky, "Tensor Decompositions for Learning Latent Variable Models," JMLR 2014.

Tensor Power Iterations

Tensor Power Method

$$v \mapsto \frac{L(v, v, \cdot)}{\|L(v, v, \cdot)\|}.$$



$$L(v, v, \cdot) = \langle v, u_1 \rangle^2 u_1 + \langle v, u_2 \rangle^2 u_2$$

Exponential no. of stationary points for power method:
$L(v, v, \cdot) = \lambda v$

For power method on orthogonal tensor, no spurious stable points

- For Robust PCA: do gradient ascent after power iteration to ensure convergence under noise

A., R. Ge, D. Hsu, S. Kakade, M. Telgarsky, "Tensor Decompositions for Learning Latent Variable Models," JMLR 2014.

Main Theorem

- Find $T = L + S$, where $L = \sum_i \sigma_i u_i \otimes u_i \otimes u_i$ and S is sparse.
- Incoherent low rank tensor: $\|u_i\|_\infty \leq \frac{\mu}{\sqrt{n}}$.
- Sparse tensor: Block sparsity pattern of size d with B blocks and overlap fraction η, ie,

$$\text{supp}(S) = \sum_{i=1}^{B} \psi_i^{\otimes 3}, \;\; \|\psi_i\|_0 \leq d, \max_{i \neq j} \langle \psi_i, \psi_j \rangle \leq \eta d, \; \psi_i(j) = 0 \text{ or } 1$$

$$\boxed{d = O(n/r^{2/3}\mu^2), \; B = O(\min(n^{2/3} r^{1/3}, \; \eta^{-1.5}))}$$

Theorem

Under above conditions, the proposed RTD algorithm converges to globally optimal solution L^* and S^*.

Implications

Superiority over matrix Robust PCA

- Applying matrix RPCA: flatten or slice the tensor to obtain a matrix.
- Tensor method can handle much higher levels of perturbation.

Implications

Superiority over matrix Robust PCA

- Applying matrix RPCA: flatten or slice the tensor to obtain a matrix.
- Tensor method can handle much higher levels of perturbation.

Demonstration under Random Block Sparsity

- Randomly draw ψ_i's as d-sparse vectors.
- D_{RTD}: total sparsity tolerated by proposed tensor method.
- D_{mat}: total sparsity tolerated by matrix RPCA applied on flattened tensor.

$$\frac{D_{RTD}}{D_{\mathrm{mat}}} = \begin{cases} \Omega\left(n^{1/6} r^{4/3}\right), r < n^{0.25} \\ \Omega\left(n^{5/12} r^{1/3}\right), o.w. \end{cases}$$

- Thus, we can handle more gross corruptions than matrix methods.

Outline

Toy example: Rank-1 matrix RPCA

$$\boxed{M = L^* + S^*, \quad L^* = u^*(u^*)^\top}$$

Non-convex method (AltProj)

- Initialize $L, S = 0$ and iterate:
- $\boxed{L \leftarrow P_1(M - S)}$ and $\boxed{S \leftarrow H_\zeta(M - L)}$.
- $P_1(\cdot)$: rank-1 projection. $H_\zeta(\cdot)$: thresholding.

Toy example: Rank-1 matrix RPCA

$$\boxed{M = L^* + S^*, \quad L^* = u^*(u^*)^\top}$$

Non-convex method (AltProj)

- Initialize $L, S = 0$ and iterate:
- $\boxed{L \leftarrow P_1(M - S)}$ and $\boxed{S \leftarrow H_\zeta(M - L)}$.
- $P_1(\cdot)$: rank-1 projection. $H_\zeta(\cdot)$: thresholding.

Immediate Observations

- First PCA: $L \leftarrow P_1(M)$.

Toy example: Rank-1 matrix RPCA

$$\boxed{M = L^* + S^*, \quad L^* = u^*(u^*)^\top}$$

Non-convex method (AltProj)

- Initialize $L, S = 0$ and iterate:
- $\boxed{L \leftarrow P_1(M - S)}$ and $\boxed{S \leftarrow H_\zeta(M - L)}$.
- $P_1(\cdot)$: rank-1 projection. $H_\zeta(\cdot)$: thresholding.

Immediate Observations

- First PCA: $L \leftarrow P_1(M)$.
- Matrix perturbation bound: $\|M - L\|_2 \leq O(\|S^*\|)$

Toy example: Rank-1 matrix RPCA

$$\boxed{M = L^* + S^*, \quad L^* = u^*(u^*)^\top}$$

Non-convex method (AltProj)

- Initialize $L, S = 0$ and iterate:
- $\boxed{L \leftarrow P_1(M - S)}$ and $\boxed{S \leftarrow H_\zeta(M - L)}$.
- $P_1(\cdot)$: rank-1 projection. $H_\zeta(\cdot)$: thresholding.

Immediate Observations

- First PCA: $L \leftarrow P_1(M)$.
- Matrix perturbation bound: $\|M - L\|_2 \leq O(\|S^*\|)$
- If $\|S^*\| \gg 1$, no progress!

Toy example: Rank-1 matrix RPCA

$$\boxed{M = L^* + S^*, \quad L^* = u^*(u^*)^\top}$$

Non-convex method (AltProj)

- Initialize $L, S = 0$ and iterate:
- $\boxed{L \leftarrow P_1(M - S)}$ and $\boxed{S \leftarrow H_\zeta(M - L)}$.
- $P_1(\cdot)$: rank-1 projection. $H_\zeta(\cdot)$: thresholding.

Immediate Observations

- First PCA: $L \leftarrow P_1(M)$.
- Matrix perturbation bound: $\|M - L\|_2 \leq O(\|S^*\|)$
- If $\|S^*\| \gg 1$, no progress!

Exploit incoherence of L^*?

Rank-1 Analysis Contd.

$$\boxed{M = L^* + S^*, \quad L^* = u^*(u^*)^\top}$$

Non-convex method (AltProj)

- Initialize $L, S = 0$ and iterate:
- $L \leftarrow P_1(M - S)$ and $S \leftarrow H_\zeta(M - L)$.

Rank-1 Analysis Contd.

$$\boxed{M = L^* + S^*, \quad L^* = u^*(u^*)^\top}$$

Non-convex method (AltProj)

- Initialize $L, S = 0$ and iterate:
- $L \leftarrow P_1(M - S)$ and $S \leftarrow H_\zeta(M - L)$.

Incoherence of L^*

- $L^* = u^*(u^*)^\top$ and $\|u^*\|_\infty \leq \frac{\mu}{\sqrt{n}}$ and $\|L^*\|_\infty \leq \frac{\mu^2}{n}$.

Rank-1 Analysis Contd.

$$\boxed{M = L^* + S^*, \quad L^* = u^*(u^*)^\top}$$

Non-convex method (AltProj)

- Initialize $L, S = 0$ and iterate:
- $L \leftarrow P_1(M - S)$ and $S \leftarrow H_\zeta(M - L)$.

Incoherence of L^*

- $L^* = u^*(u^*)^\top$ and $\|u^*\|_\infty \leq \frac{\mu}{\sqrt{n}}$ and $\|L^*\|_\infty \leq \frac{\mu^2}{n}$.

Solution for handling large $\|S^*\|$

- First threshold M before rank-1 projection.
- Ensures large entries of S^* are identified.

Rank-1 Analysis Contd.

$$\boxed{M = L^* + S^*, \quad L^* = u^*(u^*)^\top}$$

Non-convex method (AltProj)

- Initialize $L, S = 0$ and iterate:
- $L \leftarrow P_1(M - S)$ and $S \leftarrow H_\zeta(M - L)$.

Incoherence of L^*

- $L^* = u^*(u^*)^\top$ and $\|u^*\|_\infty \leq \frac{\mu}{\sqrt{n}}$ and $\|L^*\|_\infty \leq \frac{\mu^2}{n}$.

Solution for handling large $\|S^*\|$

- First threshold M before rank-1 projection.
- Ensures large entries of S^* are identified.
- Choose threshold $\zeta_0 = \frac{4\mu^2}{n}$.

Rank-1 Analysis Contd.

$$M = L^* + S^*, \quad L^* = u^*(u^*)^\top$$

Non-convex method (AltProj)

- Initialize $L = 0, S = H_{\zeta_0}(M)$ and iterate:
- $L \leftarrow P_1(M - S)$ and $S \leftarrow H_\zeta(M - L)$.

Incoherence of L^*

- $L^* = u^*(u^*)^\top$ and $\|u^*\|_\infty \leq \frac{\mu}{\sqrt{n}}$ and $\|L^*\|_\infty \leq \frac{\mu^2}{n}$.

Solution for handling large $\|S^*\|$

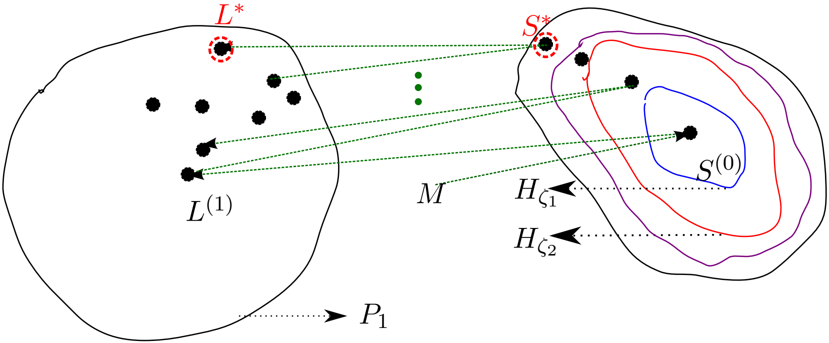
- First threshold M before rank-1 projection.
- Ensures large entries of S^* are identified.
- Choose threshold $\zeta_0 = \frac{4\mu^2}{n}$.

Rank-1 Analysis Contd.

Non-convex method (AltProj)

$L^{(0)} = 0, S^{(0)} = H_{\zeta_0}(M),$

$$\boxed{L^{(t+1)} \leftarrow P_1(M - S^{(t)}), S^{(t+1)} \leftarrow H_\zeta(M - L^{(t+1)})}.$$



- To analyze progress, track $E^{(t+1)} := S^* - S^{(t+1)}$

Rank-1 Analysis Contd.

One iteration of AltProj

$$L^{(0)} = 0, S^{(0)} = H_{\zeta_0}(M), \boxed{L^{(1)} \leftarrow P_1(M - S^{(0)}),\ S^{(1)} \leftarrow H_{\zeta}(M - L^{(1)})}.$$

Analyze $E^{(1)} := S^* - S^{(1)}$

- Thresholding is element-wise operation: require $\|L^{(1)} - L^*\|_\infty$.
- In general, no special bound for $\|L^{(1)} - L^*\|_\infty$.
- Exploit sparsity of S^* and incoherence of L^*?

Rank-1 Analysis Contd.

- $L^{(1)} = uu^\top = P_1(M - S^{(0)})$ and $E^{(0)} = S^* - S^{(0)}$.

Fixed point equation for eigenvectors $(M - S^{(0)})u = \lambda u$

Rank-1 Analysis Contd.

- $L^{(1)} = uu^\top = P_1(M - S^{(0)})$ and $E^{(0)} = S^* - S^{(0)}$.

Fixed point equation for eigenvectors $(M - S^{(0)})u = \lambda u$

- $\langle u^*, u\rangle u^* + (S^* - S^{(0)})u = \lambda u$ or $u = \lambda\langle u^*, u\rangle \left(I - \frac{E^{(0)}}{\lambda}\right)^{-1} u^*$

Taylor Series

$$\boxed{u = \lambda\langle u^*, u\rangle \left(I + \sum_{p\geq 1}\left(\frac{E^{(0)}}{\lambda}\right)^p\right) u^*}$$

Rank-1 Analysis Contd.

- $L^{(1)} = uu^\top = P_1(M - S^{(0)})$ and $E^{(0)} = S^* - S^{(0)}$.

Fixed point equation for eigenvectors $(M - S^{(0)})u = \lambda u$

- $\langle u^*, u\rangle u^* + (S^* - S^{(0)})u = \lambda u$ or $u = \lambda \langle u^*, u\rangle \left(I - \frac{E^{(0)}}{\lambda}\right)^{-1} u^*$

Taylor Series

$$\boxed{u = \lambda \langle u^*, u\rangle \left(I + \sum_{p\geq 1} \left(\frac{E^{(0)}}{\lambda}\right)^p\right) u^*}$$

- $E^{(0)}$ is sparse: $\mathrm{supp}(E^{(0)}) \subseteq \mathrm{supp}(S^*)$.

Rank-1 Analysis Contd.

- $L^{(1)} = uu^\top = P_1(M - S^{(0)})$ and $E^{(0)} = S^* - S^{(0)}$.

Fixed point equation for eigenvectors $(M - S^{(0)})u = \lambda u$

- $\langle u^*, u\rangle u^* + (S^* - S^{(0)})u = \lambda u$ or $u = \lambda \langle u^*, u\rangle \left(I - \frac{E^{(0)}}{\lambda}\right)^{-1} u^*$

Taylor Series

$$\boxed{u = \lambda \langle u^*, u\rangle \left(I + \sum_{p \geq 1} \left(\frac{E^{(0)}}{\lambda}\right)^p\right) u^*}$$

- $E^{(0)}$ is sparse: $\mathrm{supp}(E^{(0)}) \subseteq \mathrm{supp}(S^*)$.
- Exploiting sparsity: $(E^{(0)})^p$ is the p^{th}-hop adjacency matrix of $E^{(0)}$.

Rank-1 Analysis Contd.

- $L^{(1)} = uu^\top = P_1(M - S^{(0)})$ and $E^{(0)} = S^* - S^{(0)}$.

Fixed point equation for eigenvectors $(M - S^{(0)})u = \lambda u$

- $\langle u^*, u\rangle u^* + (S^* - S^{(0)})u = \lambda u$ or $u = \lambda \langle u^*, u\rangle \left(I - \frac{E^{(0)}}{\lambda}\right)^{-1} u^*$

Taylor Series

$$\boxed{u = \lambda \langle u^*, u\rangle \left(I + \sum_{p\geq 1} \left(\frac{E^{(0)}}{\lambda}\right)^p\right) u^*}$$

- $E^{(0)}$ is sparse: $\mathrm{supp}(E^{(0)}) \subseteq \mathrm{supp}(S^*)$.
- Exploiting sparsity: $(E^{(0)})^p$ is the p^{th}-hop adjacency matrix of $E^{(0)}$.
- Counting walks in sparse graphs.

Rank-1 Analysis Contd.

- $L^{(1)} = uu^\top = P_1(M - S^{(0)})$ and $E^{(0)} = S^* - S^{(0)}$.

Fixed point equation for eigenvectors $(M - S^{(0)})u = \lambda u$

- $\langle u^*, u\rangle u^* + (S^* - S^{(0)})u = \lambda u$ or $u = \lambda \langle u^*, u\rangle \left(I - \frac{E^{(0)}}{\lambda}\right)^{-1} u^*$

Taylor Series

$$u = \lambda \langle u^*, u\rangle \left(I + \sum_{p\geq 1} \left(\frac{E^{(0)}}{\lambda}\right)^p\right) u^*$$

- $E^{(0)}$ is sparse: $\mathrm{supp}(E^{(0)}) \subseteq \mathrm{supp}(S^*)$.
- Exploiting sparsity: $(E^{(0)})^p$ is the p^{th}-hop adjacency matrix of $E^{(0)}$.
- Counting walks in sparse graphs.
- In addition, u^* is incoherent: $\|u^*\|_\infty < \frac{\mu}{\sqrt{n}}$.

Rank-1 Analysis Contd.

$$\boxed{u = \lambda \langle u^*, u \rangle \left(I + \sum_{p \geq 1} \left(\frac{E^{(0)}}{\lambda} \right)^p \right) u^*}$$

- $E^{(0)}$ is sparse (each row/column is d sparse) and u^* is μ-incoherent.

Rank-1 Analysis Contd.

$$u = \lambda \langle u^*, u \rangle \left(I + \sum_{p \geq 1} \left(\frac{E^{(0)}}{\lambda} \right)^p \right) u^*$$

- $E^{(0)}$ is sparse (each row/column is d sparse) and u^* is μ-incoherent.
- We show: $\|(E^{(0)})^p u^*\|_\infty \leq \frac{\mu}{\sqrt{n}} (d \|E^{(0)}\|_\infty)^p$.

Rank-1 Analysis Contd.

$$\boxed{u = \lambda \langle u^*, u \rangle \left(I + \sum_{p \geq 1} \left(\frac{E^{(0)}}{\lambda} \right)^p \right) u^*}$$

- $E^{(0)}$ is sparse (each row/column is d sparse) and u^* is μ-incoherent.
- We show: $\boxed{\|(E^{(0)})^p u^*\|_\infty \leq \frac{\mu}{\sqrt{n}} (d \|E^{(0)}\|_\infty)^p}$.
- Convergence when terms are < 1, i.e. $d\|E^{(0)}\|_\infty < 1$.

Rank-1 Analysis Contd.

$$\boxed{u = \lambda \langle u^*, u \rangle \left(I + \sum_{p \geq 1} \left(\frac{E^{(0)}}{\lambda} \right)^p \right) u^*}$$

- $E^{(0)}$ is sparse (each row/column is d sparse) and u^* is μ-incoherent.
- We show: $\boxed{\|(E^{(0)})^p u^*\|_\infty \leq \frac{\mu}{\sqrt{n}} (d \|E^{(0)}\|_\infty)^p}$.
- Convergence when terms are < 1, i.e. $d\|E^{(0)}\|_\infty < 1$.
- Recall $\|E^{(0)}\|_\infty < \frac{4\mu^2}{n}$ due to thresholding.

Rank-1 Analysis Contd.

$$\boxed{u = \lambda \langle u^*, u \rangle \left(I + \sum_{p \geq 1} \left(\frac{E^{(0)}}{\lambda} \right)^p \right) u^*}$$

- $E^{(0)}$ is sparse (each row/column is d sparse) and u^* is μ-incoherent.
- We show: $\boxed{\|(E^{(0)})^p u^*\|_\infty \leq \frac{\mu}{\sqrt{n}} (d \|E^{(0)}\|_\infty)^p}$.
- Convergence when terms are < 1, i.e. $d\|E^{(0)}\|_\infty < 1$.
- Recall $\|E^{(0)}\|_\infty < \dfrac{4\mu^2}{n}$ due to thresholding.
- Require $\boxed{d < \frac{n}{4\mu^2}}$. Can tolerate $O(n)$ corruptions!

Rank-1 Analysis Contd.

$$\boxed{u = \lambda \langle u^*, u\rangle \left(I + \sum_{p \geq 1} \left(\frac{E^{(0)}}{\lambda} \right)^p \right) u^*}$$

- $E^{(0)}$ is sparse (each row/column is d sparse) and u^* is μ-incoherent.
- We show: $\boxed{\|(E^{(0)})^p u^*\|_\infty \leq \frac{\mu}{\sqrt{n}}(d\|E^{(0)}\|_\infty)^p}$.
- Convergence when terms are < 1, i.e. $d\|E^{(0)}\|_\infty < 1$.
- Recall $\|E^{(0)}\|_\infty < \frac{4\mu^2}{n}$ due to thresholding.
- Require $\boxed{d < \frac{n}{4\mu^2}}$. Can tolerate $O(n)$ corruptions!

Contraction of error $E^{(t)}$ when degree d is bounded.

Sketch of Further Steps

Extensions to higher rank

- Stage-wise algorithm: first only consider rank-1 estimates
- After sufficient denoising proceed to rank-2 estimates and so on.

Sketch of Further Steps

Extensions to higher rank

- Stage-wise algorithm: first only consider rank-1 estimates
- After sufficient denoising proceed to rank-2 estimates and so on.

Unique Challenges in Tensor Analysis

- Do not have guaranteed low rank decomposition of every tensor.
- Require perturbation to be small enough for guarantees.
- Tensor algebra quite different from matrix analysis.

Outline

Matrix RPCA: Foreground/background Separation

Original

Rank-10 PCA

AltProj

IALM

Tensor RPCA Results

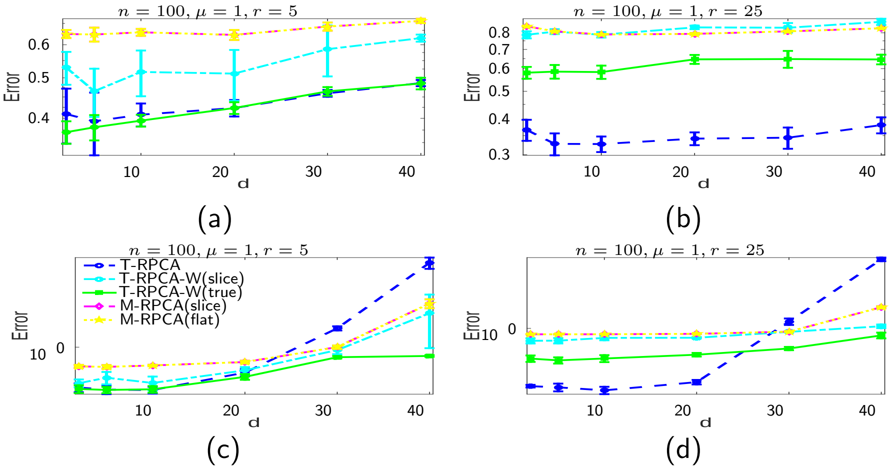
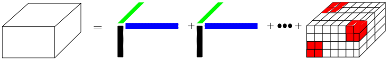


Figure : (a),(b) Error with non-block sparsity. (c),(d) Error with block sparsity.

Outline

Conclusion



Guaranteed Non-Convex Tensor Robust PCA

- Efficient non-convex method for tensor robust PCA.
- Alternating rank projections and thresholding.
- Estimates for low rank and sparse parts "grown gradually".
- Low computational complexity: scalable to large tensors.
- Advantage over matrix methods: tensor methods incorporate more constraints.

Possible to have both: guarantees and computational efficiency!