Non-convex Robust PCA: Provable Bounds

Anima Anandkumar

U.C. Irvine

Joint work with Praneeth Netrapalli, U.N. Niranjan, Prateek Jain and Sujay Sanghavi.

Learning with Big Data

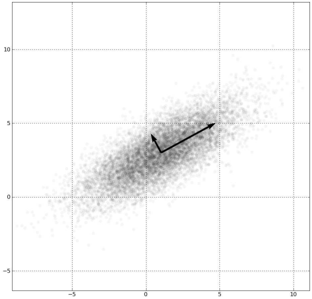
High Dimensional Regime

- Missing observations, gross corruptions, outliers, ill-posed problems.
- Needle in a haystack: finding low dimensional structures in high dimensional data.

Principled approaches for finding low dimensional structures?

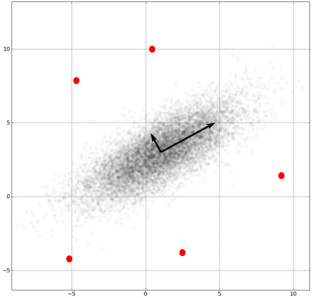
PCA: Classical Method

- Denoising: find hidden low rank structures in data.
- Efficient computation, perturbation analysis.



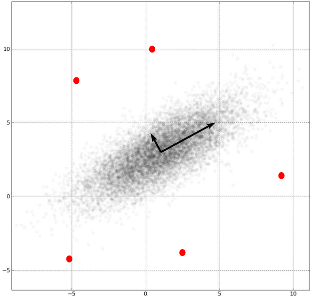
PCA: Classical Method

- Denoising: find hidden low rank structures in data.
- Efficient computation, perturbation analysis.



PCA: Classical Method

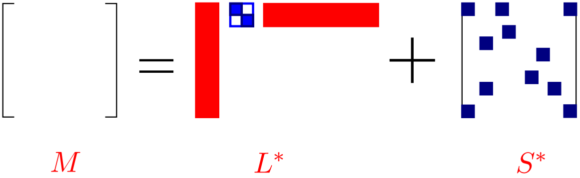
- Denoising: find hidden low rank structures in data.
- Efficient computation, perturbation analysis.



Not robust to even a few outliers

Robust PCA Problem

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



- $M \in \mathbb{R}^{n \times n}$, L^* is low rank and S^* is sparse.
- Applications in computer vision, topic and community modeling.

History

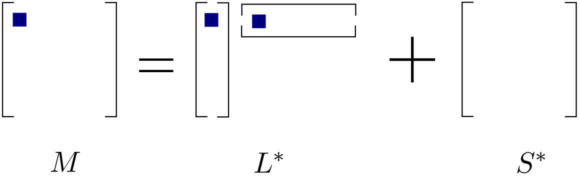
Heuristics without guarantees

- Multivariate trimming [Gnanadeskian+ Kettering 72]
- Random sampling [Fischler+ Bolles81].
- Alternating minimization [Ke+ Kanade03].
- Influence functions [de la Torre + Black 03]

Convex methods with Guarantees

- Chandrasekharan et. al, Candes et. al '11: seminal guarantees.
- Hsu et. al '11, Agarwal et. al '12: further guarantees.
- (Variants) Xu et. al '11: Outlier pursuit, Chen et. al '12: community detection.

Why is Robust PCA difficult?



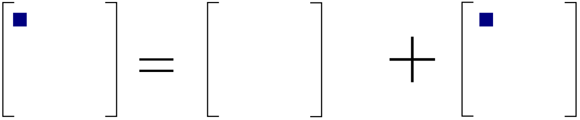
- No identifiability in general: Low rank matrices can also be sparse and vice versa.

Natural constraints for identifiability?

- Low rank matrix is NOT sparse and viceversa.
- Incoherent low rank matrix and sparse matrix with sparsity constraints.

Tractable methods for identifiable settings?

Why is Robust PCA difficult?



$M \qquad L^* \qquad S^*$

- No identifiability in general: Low rank matrices can also be sparse and vice versa.

Natural constraints for identifiability?

- Low rank matrix is NOT sparse and viceversa.
- Incoherent low rank matrix and sparse matrix with sparsity constraints.

Tractable methods for identifiable settings?

Convex Relaxation Techniques

(Hard) Optimization Problem, given $M \in \mathbb{R}^{n \times n}$

$$\min_{L,S} \mathrm{Rank}(L) + \gamma \|S\|_0, \quad M = L + S.$$

- $\mathrm{Rank}(L) = \{\#\sigma_i(L) : \sigma_i(L) \neq 0\}$, $\|S\|_0 = \{\#S(i,j) : S(i,j) \neq 0\}$ are not tractable.

Convex Relaxation Techniques

(Hard) Optimization Problem, given $M \in \mathbb{R}^{n\times n}$

$$\boxed{\min_{L,S} \text{Rank}(L) + \gamma \|S\|_0, \quad M = L + S.}$$

- $\text{Rank}(L) = \{\#\sigma_i(L) : \sigma_i(L) \neq 0\}$, $\|S\|_0 = \{\#S(i,j) : S(i,j) \neq 0\}$ are not tractable.

Convex Relaxation

$$\boxed{\min_{L,S} \|L\|_* + \gamma \|S\|_1, \quad M = L + S.}$$

- $\|L\|_* = \sum_i \sigma_i(L)$, $\|S\|_1 = \sum_{i,j} |S(i,j)|$ are convex sets.
- Chandrasekharan et. al, Candes et. al '11: seminal works.

Other Alternatives for Robust PCA?

$$\boxed{\min_{L,S} \|L\|_* + \gamma \|S\|_1, \quad M = L + S.}$$

Shortcomings of convex methods

Other Alternatives for Robust PCA?

$$\boxed{\min_{L,S} \|L\|_* + \gamma\|S\|_1, \quad M = L + S.}$$

Shortcomings of convex methods

- Computational cost: $O(n^3/\epsilon)$ to achieve error of ϵ
  - Requires SVD of $n \times n$ matrix.
- Analysis: requires dual witness style arguments.
- Conditions for success usually opaque.

Other Alternatives for Robust PCA?

$$\boxed{\min_{L,S} \|L\|_* + \gamma\|S\|_1, \quad M = L + S.}$$

Shortcomings of convex methods

- Computational cost: $O(n^3/\epsilon)$ to achieve error of ϵ
  - Requires SVD of $n \times n$ matrix.
- Analysis: requires dual witness style arguments.
- Conditions for success usually opaque.

Non-convex alternatives?

Proposal for Non-convex Robust PCA

$$\boxed{\min_{L,S} \|S\|_0, \quad s.t.\ M = L + S, \quad \text{Rank}(L) = r}$$

Proposal for Non-convex Robust PCA

$$\boxed{\min_{L,S} \|S\|_0, \quad s.t.\ M = L + S, \quad \text{Rank}(L) = r}$$

A non-convex heuristic (AltProj)

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

- Computationally efficient: each operation is just a rank-r SVD or thresholding.

Proposal for Non-convex Robust PCA

$$\boxed{\min_{L,S} \|S\|_0, \quad s.t.\ M = L + S, \quad \text{Rank}(L) = r}$$

A non-convex heuristic (AltProj)

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

- Computationally efficient: each operation is just a rank-r SVD or thresholding.

Any hope for proving guarantees?

Observations regarding non-convex analysis

Challenges

- Multiple stable points: bad local optima, solution depends on initialization.
- Method may have very slow convergence or may not converge at all!

Observations regarding non-convex analysis

Challenges

- Multiple stable points: bad local optima, solution depends on initialization.
- Method may have very slow convergence or may not converge at all!

Non-convex Projections vs. Convex Projections

- Projections on to non-convex sets: NP-hard in general.
  - Projections on to rank and sparse sets: tractable.
- Less information than convex projections: zero-order conditions.

$$\|P(M) - M\| \leq \|Y - M\|, \quad \forall\, Y \in C(\text{Non-convex}),$$

$$\|P(M) - M\|^2 \leq \langle Y - M, P(M) - M \rangle, \quad \forall\, Y \in C(\text{Convex}).$$

Non-convex success stories

Classical Result

- PCA: Convergence to global optima!

Non-convex success stories

Classical Result

- PCA: Convergence to global optima!

Recent results

- Tensor methods (Anandkumar et. al '12, '14): Local optima can be characterized in special cases.

Non-convex success stories

Classical Result

- PCA: Convergence to global optima!

Recent results

- Tensor methods (Anandkumar et. al ‘12, ‘14): Local optima can be characterized in special cases.
- Dictionary learning (Agarwal et. al ‘14, Arora et. al ‘14): Initialize using a “clustering style” method and do alternating minimization.

Non-convex success stories

Classical Result

- PCA: Convergence to global optima!

Recent results

- Tensor methods (Anandkumar et. al '12, '14): Local optima can be characterized in special cases.
- Dictionary learning (Agarwal et. al '14, Arora et. al '14): Initialize using a "clustering style" method and do alternating minimization.
- Matrix completion/phase retrieval: (Netrapalli et. al '13) Initialize with PCA and do alternating minimization.

Non-convex success stories

Classical Result

- PCA: Convergence to global optima!

Recent results

- Tensor methods (Anandkumar et. al '12, '14): Local optima can be characterized in special cases.
- Dictionary learning (Agarwal et. al '14, Arora et. al '14): Initialize using a "clustering style" method and do alternating minimization.
- Matrix completion/phase retrieval: (Netrapalli et. al '13) Initialize with PCA and do alternating minimization.

(Somewhat) common theme

- Characterize basin of attraction for global optimum.
- Obtain a good initialization to "land in the ball".

Non-convex Robust PCA

A non-convex heuristic (AltProj)

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

Observations regarding Robust PCA

- Projection on to rank and sparse subspaces: non-convex but tractable: SVD and hard thresholding.
- But alternating projections: challenging to analyze

Non-convex Robust PCA

A non-convex heuristic (AltProj)

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

Observations regarding Robust PCA

- Projection on to rank and sparse subspaces: non-convex but tractable: SVD and hard thresholding.
- But alternating projections: challenging to analyze

Our results for (a variant of) AltProj

- Guaranteed recovery of low rank L^* and sparse part S^*.
- Match the bounds for convex methods (deterministic sparsity).
- Reduced computation: only require low rank SVDs!

Non-convex Robust PCA

A non-convex heuristic (AltProj)

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

Observations regarding Robust PCA

- Projection on to rank and sparse subspaces: non-convex but tractable: SVD and hard thresholding.
- But alternating projections: challenging to analyze

Our results for (a variant of) AltProj

- Guaranteed recovery of low rank L^* and sparse part S^*.
- Match the bounds for convex methods (deterministic sparsity).
- Reduced computation: only require low rank SVDs!

Best of both worlds: reduced computation with guarantees!

Outline

Toy example: Rank-1 case

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

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 case

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

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 case

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

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 case

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

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 case

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

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.

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

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 $\|S^*\|$

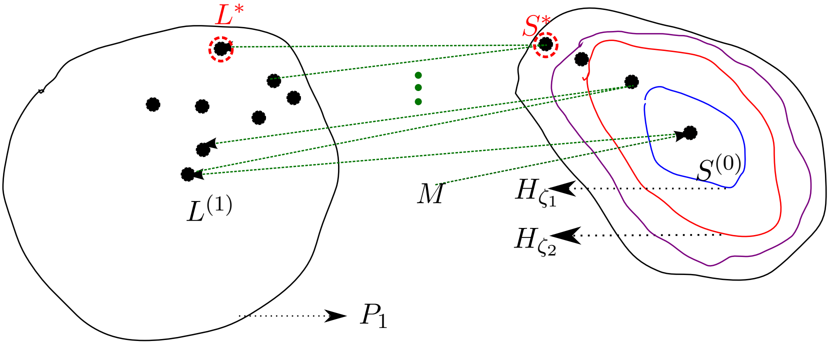
- First threshold M before rank-1 projection.
- Ensures large entries of S^* are identified.
- Choose threshold $\zeta_0 = \frac{4\mu^2}{n}$.

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

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

- $E^{(0)}$ 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.
- 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: $\boxed{\|(E^{(0)})^p u^*\|_\infty \leq \frac{\mu}{\sqrt{n}} (d \|E^{(0)}\|_\infty)^p}$.

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.

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

- $E^{(0)}$ is sparse (each row/column is d sparse) and u^* is μ-incoherent.
- We show: $\|(E^{(0)})^p u^*\|_\infty \leq \frac{\mu}{\sqrt{n}} (d \|E^{(0)}\|_\infty)^p$.
- 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 $\|E^{(0)}\|_\infty < \frac{4\mu^2}{n}$ 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.

Extension to general rank: challenges

Extension to general rank: challenges

A proposal for rank-r Non-convex method (AltProj)

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

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

Extension to general rank: challenges

A proposal for rank-r Non-convex method (AltProj)

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

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

Recall for rank-1 case

- Initial threshold controlled perturbation for rank-1 projection.

Extension to general rank: challenges

A proposal for rank-r Non-convex method (AltProj)

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

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

Recall for rank-1 case

- Initial threshold controlled perturbation for rank-1 projection.

Perturbation analysis in general rank case

- Small $\lambda^*_{\min}(L^*)$: no recovery of lower eigenvectors.

Extension to general rank: challenges

A proposal for rank-r Non-convex method (AltProj)

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

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

Recall for rank-1 case

- Initial threshold controlled perturbation for rank-1 projection.

Perturbation analysis in general rank case

- Small $\lambda^*_{\min}(L^*)$: no recovery of lower eigenvectors.
- Sparsity level depends on condition number $\lambda^*_{\max}/\lambda^*_{\min}$

Extension to general rank: challenges

A proposal for rank-r Non-convex method (AltProj)

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

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

Recall for rank-1 case

- Initial threshold controlled perturbation for rank-1 projection.

Perturbation analysis in general rank case

- Small $\lambda^*_{\min}(L^*)$: no recovery of lower eigenvectors.
- Sparsity level depends on condition number $\lambda^*_{\max}/\lambda^*_{\min}$

Guarantees without dependence on condition number?

Extension to general rank: challenges

A proposal for rank-r Non-convex method (AltProj)

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

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

Recall for rank-1 case

- Initial threshold controlled perturbation for rank-1 projection.

Perturbation analysis in general rank case

- Small $\lambda^*_{\min}(L^*)$: no recovery of lower eigenvectors.
- Sparsity level depends on condition number $\lambda^*_{\max}/\lambda^*_{\min}$

Guarantees without dependence on condition number?

- Lower eigenvectors subject to a large perturbation initially.

Extension to general rank: challenges

A proposal for rank-r Non-convex method (AltProj)

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

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

Recall for rank-1 case

- Initial threshold controlled perturbation for rank-1 projection.

Perturbation analysis in general rank case

- Small $\lambda^*_{\min}(L^*)$: no recovery of lower eigenvectors.
- Sparsity level depends on condition number $\lambda^*_{\max}/\lambda^*_{\min}$

Guarantees without dependence on condition number?

- Lower eigenvectors subject to a large perturbation initially.
- Reduce perturbation before recovering lower eigenvectors!

Improved Algorithm for General Rank Setting

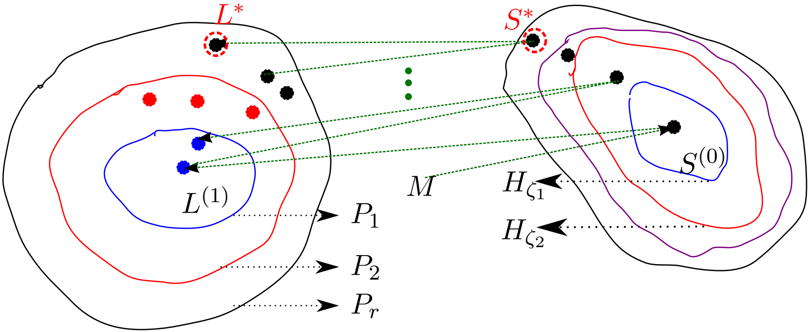
Stage-wise Projections

- Init $L^{(0)} = 0, S^{(0)} = H_{\zeta_0}(M)$.
- For stage $k = 1$ to r,
  - Iterate: $\boxed{L^{(t+1)} \leftarrow P_k(M - S^{(t)}), \quad S^{(t+1)} \leftarrow H_{\zeta}(M - L^{(t+1)})}$.

Improved Algorithm for General Rank Setting

Stage-wise Projections

- Init $L^{(0)} = 0, S^{(0)} = H_{\zeta_0}(M)$.
- For stage $k = 1$ to r,
  - Iterate: $\boxed{L^{(t+1)} \leftarrow P_k(M - S^{(t)}), \quad S^{(t+1)} \leftarrow H_\zeta(M - L^{(t+1)})}$.



Summary of Results

- Low rank part: $L^* = U^* \Lambda^* (V^*)^\top$ has rank r.
- Incoherence: $\|U^*(i,:)\|_2, \|V^*(i,:)\|_2 \leq \frac{\mu\sqrt{r}}{\sqrt{n}}$.
- Sparse part: S^* has at most d non-zeros per row/column.

Summary of Results

- Low rank part: $L^* = U^* \Lambda^* (V^*)^\top$ has rank r.
- Incoherence: $\|U^*(i,:)\|_2, \|V^*(i,:)\|_2 \leq \frac{\mu\sqrt{r}}{\sqrt{n}}$.
- Sparse part: S^* has at most d non-zeros per row/column.

Theorem: Guarantees for Stage-wise AltProj

- Exact recovery of L^*, S^* when $d = O\left(\frac{n}{\mu^2 r}\right)$

Summary of Results

- Low rank part: $L^* = U^* \Lambda^* (V^*)^\top$ has rank r.
- Incoherence: $\|U^*(i,:)\|_2, \|V^*(i,:)\|_2 \le \frac{\mu\sqrt{r}}{\sqrt{n}}$.
- Sparse part: S^* has at most d non-zeros per row/column.

Theorem: Guarantees for Stage-wise AltProj

- Exact recovery of L^*, S^* when $d = O\left(\frac{n}{\mu^2 r}\right)$
- Computational complexity: $O\left(r^2 n^2 \log(1/\epsilon)\right)$

Summary of Results

- Low rank part: $L^* = U^* \Lambda^* (V^*)^\top$ has rank r.
- Incoherence: $\|U^*(i,:)\|_2, \|V^*(i,:)\|_2 \leq \frac{\mu\sqrt{r}}{\sqrt{n}}$.
- Sparse part: S^* has at most d non-zeros per row/column.

Theorem: Guarantees for Stage-wise AltProj

- Exact recovery of L^*, S^* when $d = O\left(\frac{n}{\mu^2 r}\right)$
- Computational complexity: $O\left(r^2 n^2 \log(1/\epsilon)\right)$

Comparison to convex method

- Same (deterministic) condition on d. Running time: $O\left(n^3/\epsilon\right)$

Summary of Results

- Low rank part: $L^* = U^* \Lambda^* (V^*)^\top$ has rank r.
- Incoherence: $\|U^*(i,:)\|_2, \|V^*(i,:)\|_2 \leq \frac{\mu\sqrt{r}}{\sqrt{n}}$.
- Sparse part: S^* has at most d non-zeros per row/column.

Theorem: Guarantees for Stage-wise AltProj

- Exact recovery of L^*, S^* when $d = O\left(\frac{n}{\mu^2 r}\right)$
- Computational complexity: $O\left(r^2 n^2 \log(1/\epsilon)\right)$

Comparison to convex method

- Same (deterministic) condition on d. Running time: $O\left(n^3/\epsilon\right)$

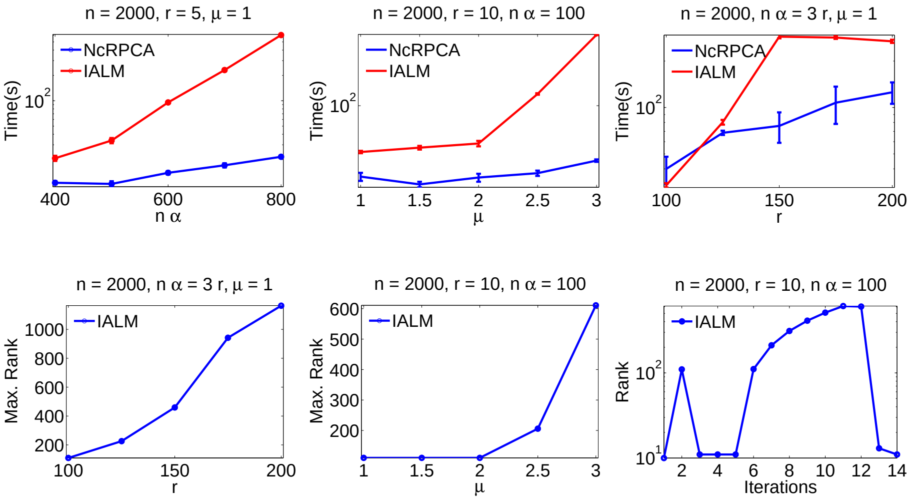
Best of both worlds: reduced computation with guarantees!

"Non-convex Robust PCA," P. Netrapalli, U.N. Niranjan, S. Sanghavi, A. , P. Jain, NIPS '14.

Outline

Synthetic Results

- NcRPCA: Non-convex Robust PCA.
- IALM: Inexact augmented Lagrange multipliers.



Real data: Foreground/background Separation

Original

Rank-10 PCA

AltProj

IALM

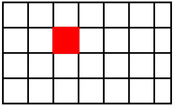
Real data: Foreground/background Separation

AltProj

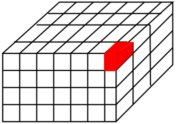
IALM

Outline

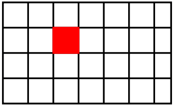
Robust Tensor PCA



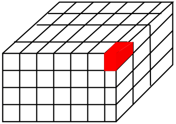
vs.



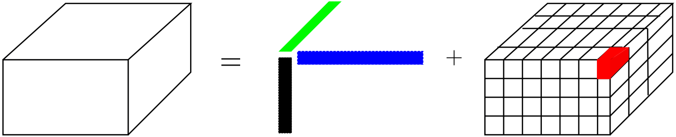
Robust Tensor PCA



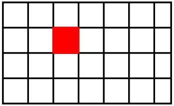
vs.



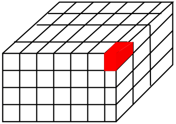
Robust Tensor Problem



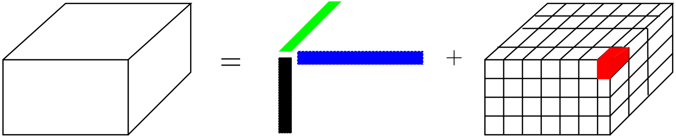
Robust Tensor PCA



vs.



Robust Tensor Problem



Applications: Robust Learning of Latent Variable Models.

A. , R. Ge, D. Hsu, S.M. Kakade and M. Telgarsky "Tensor Decompositions for Learning Latent Variable Models," Preprint, Oct. '12.

Challenges and Preliminary Observations

$$T = L^* + S^* \in \mathbb{R}^{n\times n\times n}, \quad L^* = \sum_{i\in[r]} a_i^{\otimes 3}.$$

Convex methods

- No natural convex surrogate for tensor (CP) rank.
- Matricization loses the tensor structure!

Challenges and Preliminary Observations

$$\boxed{T = L^* + S^* \in \mathbb{R}^{n\times n\times n}, \quad L^* = \textstyle\sum_{i\in[r]} a_i^{\otimes 3}.}$$

Convex methods

- No natural convex surrogate for tensor (CP) rank.
- Matricization loses the tensor structure!

Non-Convex Heuristic: Extension of Matrix AltProj

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

Challenges and Preliminary Observations

$$\boxed{T = L^* + S^* \in \mathbb{R}^{n\times n\times n}, \quad L^* = \sum_{i\in[r]} a_i^{\otimes 3}.}$$

Convex methods

- No natural convex surrogate for tensor (CP) rank.
- Matricization loses the tensor structure!

Non-Convex Heuristic: Extension of Matrix AltProj

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

Challenges in Non-Convex Analysis

- P_r for a general tensor is NP-hard!

Challenges and Preliminary Observations

$$\boxed{T = L^* + S^* \in \mathbb{R}^{n\times n\times n}, \quad L^* = \sum_{i\in[r]} a_i^{\otimes 3}.}$$

Convex methods

- No natural convex surrogate for tensor (CP) rank.
- Matricization loses the tensor structure!

Non-Convex Heuristic: Extension of Matrix AltProj

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

Challenges in Non-Convex Analysis

- P_r for a general tensor is NP-hard!
- Can be well approximated in special cases, e.g. full rank factors.

Challenges and Preliminary Observations

$$T = L^* + S^* \in \mathbb{R}^{n\times n\times n}, \quad L^* = \sum_{i\in[r]} a_i^{\otimes 3}.$$

Convex methods

- No natural convex surrogate for tensor (CP) rank.
- Matricization loses the tensor structure!

Non-Convex Heuristic: Extension of Matrix AltProj

$$L^{(t+1)} \leftarrow P_r(T - S^{(t)}), S^{(t+1)} \leftarrow H_\zeta(T - L^{(t+1)}).$$

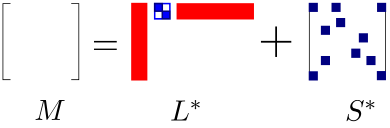
Challenges in Non-Convex Analysis

- P_r for a general tensor is NP-hard!
- Can be well approximated in special cases, e.g. full rank factors.

Guaranteed recovery possible!

Outline

Conclusion

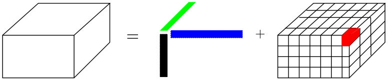


Guaranteed Non-Convex Robust PCA

- Simple non-convex method for robust PCA.
- Alternating rank projections and thresholding.
- Estimates for low rank and sparse parts "grown gradually".
- Guarantees match convex methods.
- Low computational complexity: scalable to large matrices.

Possible to have both: guarantees and low computation!

Outlook



- Reduce computational complexity? Skip stages in rank projections? Tight bounds for incoherent row-column subspaces?
- Extendable to the tensor setting with tight scaling guarantees.
- Other problems where non-convex methods have guarantees?
  - Csiszar's alternating minimization framework.
- (Laserre) hierarchy for convex methods: increasing complexity for "harder" problems.
- Analogous unified thinking for non-convex methods?

Holy grail: A general framework for non-convex methods?