Accelerated Alternating Projections for Robust Principal Component Analysis

HanQin Cai (UCLA) Jian-Feng Cai (HKUST) Ke Wei (Fudan U)

Highlights

- Novel Accelerated Low Rank Approximation via Manifold Tangent Space Projection
- State-Of-The-Art Computational Efficiency for Fully Observed RPCA
- Non-Convex Algorithm with Guaranteed Linear Convergence

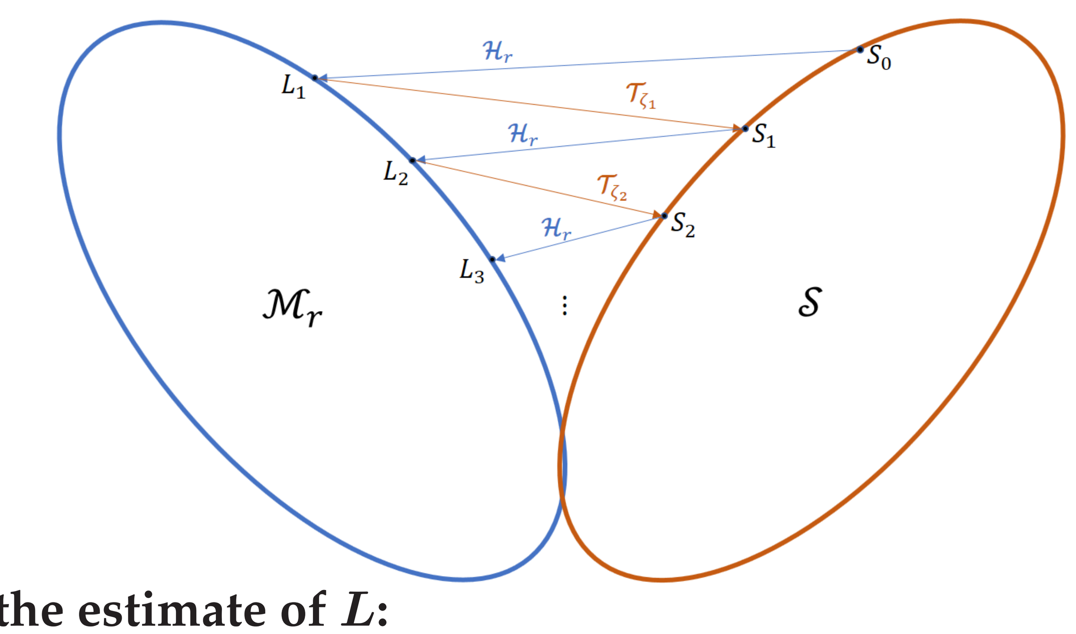
Alternating Projections For RPCA

Robust Principal Component Analysis (RPCA): Recover a low-rank matrix $\boldsymbol{L}$ and a sparse matrix $\boldsymbol{S}$ from their sum $\boldsymbol{D}=\boldsymbol{L}+\boldsymbol{S}\in\mathbb{R}^{m\times n}$.

Non-convex optimization problem:

$$\min_{\boldsymbol{L}',\boldsymbol{S}'} \|\boldsymbol{D}-\boldsymbol{L}'-\boldsymbol{S}'\|_F \quad \text{subject to } rank(\boldsymbol{L}')\le r \text{ and } \|\boldsymbol{S}'\|_0\le |supp(\boldsymbol{S})|.$$

Alternating Projections (AltProj) [1]: Projecting alternatively onto $\mathcal{M}_r$, the space of rank-r matrices, and $\mathcal{S}$, the space of sparse matrices.



Updating the estimate of L:

$$\boldsymbol{L}_k=\mathcal{H}_r(\boldsymbol{D}-\boldsymbol{S}_{k-1}), \tag{1}$$

where $\mathcal{H}_r$ is the best rank-r approximation via truncated SVD.

Updating the estimate of S:

$$\boldsymbol{S}_k=\mathcal{T}_{\zeta_k}(\boldsymbol{D}-\boldsymbol{L}_k), \tag{2}$$

where $\mathcal{T}_\zeta$ is the hard thresholding operator, and the thresholding value ζ_k is chose differently at each iteration.

Drawback: Truncated SVD is computational expansive when the matrix size is large and the rank is relatively small.

Assumptions

A1 *The underlying $\boldsymbol{L}\in\mathbb{R}^{m\times n}$ is a rank-r matrix with μ-incoherence, that is*

$$\|\boldsymbol{U}\|_{2,\infty}\le\sqrt{\mu r/m}, \quad \text{and} \quad \|\boldsymbol{V}\|_{2,\infty}\le\sqrt{\mu r/n}$$

hold for a positive numerical constant μ, where $\boldsymbol{L}=\boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^T$ is the SVD of $\boldsymbol{L}$.

A2 *The underlying $\boldsymbol{S}\in\mathbb{R}^{m\times n}$ is α-sparse. That is, $\boldsymbol{S}$ has at most αn non-zero entries in each row, and at most αm non-zero entries in each column.*

Accelerated Alternating Projections [2]

Tangent Space: $\mathcal{M}_r$ is indeed a Riemannian manifold. The tangent space of $\mathcal{M}_r$ at $\boldsymbol{L}$ is defined as

$$T=\{\boldsymbol{U}\boldsymbol{A}^T+\boldsymbol{B}\boldsymbol{V}^T \mid \boldsymbol{A}\in\mathbb{R}^{n\times r},\boldsymbol{B}\in\mathbb{R}^{m\times r}\},$$

where $\boldsymbol{L}_k=\boldsymbol{U}_k\boldsymbol{\Sigma}_k\boldsymbol{V}_k^T$ is the SVD of $\boldsymbol{L}$. The projection of a matrix $\boldsymbol{Z}$ onto T is given by

$$\mathcal{P}_T\boldsymbol{Z}=\boldsymbol{U}\boldsymbol{U}^T\boldsymbol{Z}+\boldsymbol{Z}\boldsymbol{V}\boldsymbol{V}^T-\boldsymbol{U}\boldsymbol{U}^T\boldsymbol{Z}\boldsymbol{V}\boldsymbol{V}^T.$$

Accelerating Low-Rank Approximation: When updating the estimate of $\boldsymbol{L}$, we first trim $\boldsymbol{L}_{k-1}$ to obtain an incoherent tangent space $\widetilde{T}_{k-1}$, then we replacing (1) by

$$\boldsymbol{L}_k=\mathcal{H}_r(\mathcal{P}_{\widetilde{T}_{k-1}}(\boldsymbol{D}-\boldsymbol{S}_{k-1})), \tag{3}$$

which can be computed by total $4n^2r+n^2+O(nr^2+r^3)$ flops. In contrast, (1) costs $O(n^2r)$ flops with a large hidden constant.

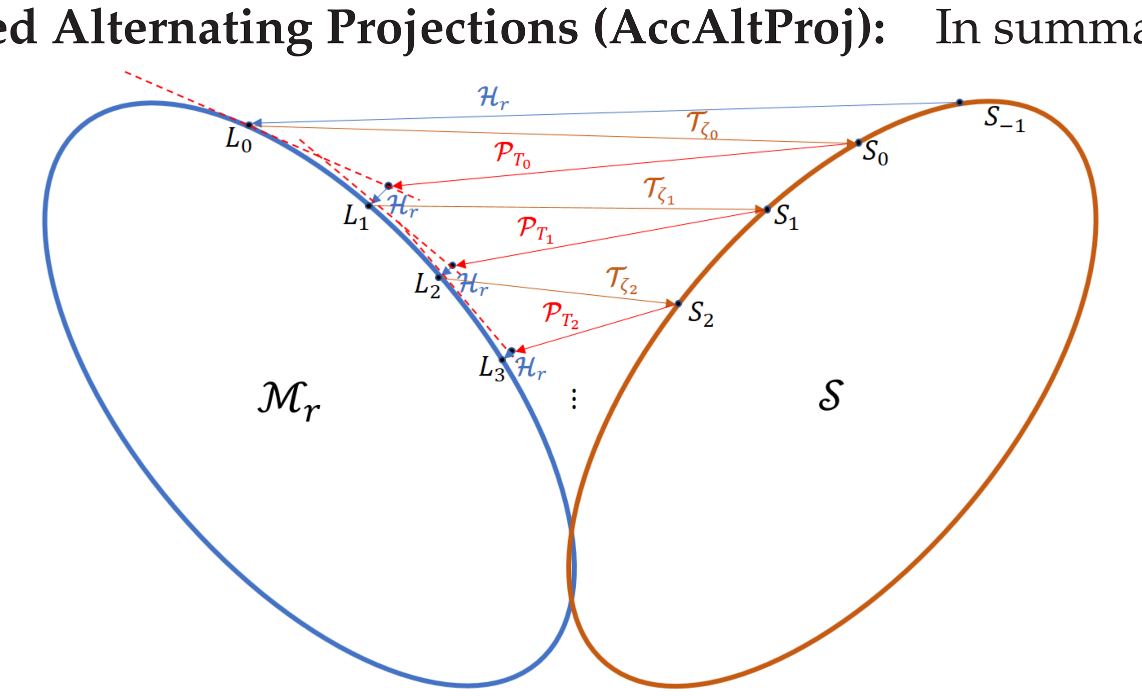
Proper Threshold Values: When updating the estimate of $\boldsymbol{S}$, we also employ the hard threshold operator as in (2). However, we choose ζ_k as

$$\zeta_k=\beta(\sigma_{r+1}(\mathcal{P}_{\widetilde{T}_{k-1}}(\boldsymbol{D}-\boldsymbol{S}_{k-1}))+\gamma^{k+1}\sigma_1(\mathcal{P}_{\widetilde{T}_{k-1}}(\boldsymbol{D}-\boldsymbol{S}_{k-1}))),$$

where β and γ are two positive parameters, and σ_i denotes the i^{th} singular value. The total cost of $\boldsymbol{S}_k$ updating is then $2n^2+O(1)$.

Initialization: With a different thresholding parameter β_{init}, we run 2 steps of AltProj as initialization, which will give us a sufficient close tangent space to start with.

Accelerated Alternating Projections (AccAltProj): In summary,



Theorem 1 (Recovery Guarantee) *Let $\boldsymbol{L}$ and $\boldsymbol{S}$ be two matrices satisfying Assumptions A1 and A2 with $\alpha\lesssim\min\{\frac{1}{\mu r^2\kappa^3},\frac{1}{\mu^{1.5}r^2\kappa},\frac{1}{\mu^2r^2}\}$. If the thresholding parameters obey $\frac{\mu r\sigma_1^L}{\sqrt{mn}\sigma_1^D}\le\beta_{init}\le\frac{3\mu r\sigma_1^L}{\sqrt{mn}\sigma_1^D}$ and $\beta=\frac{\mu r}{2\sqrt{mn}}$, alone with the convergence rate parameter $\gamma\in(\frac{1}{\sqrt{12}},1)$, then the outputs of AccAltProj satisfy*

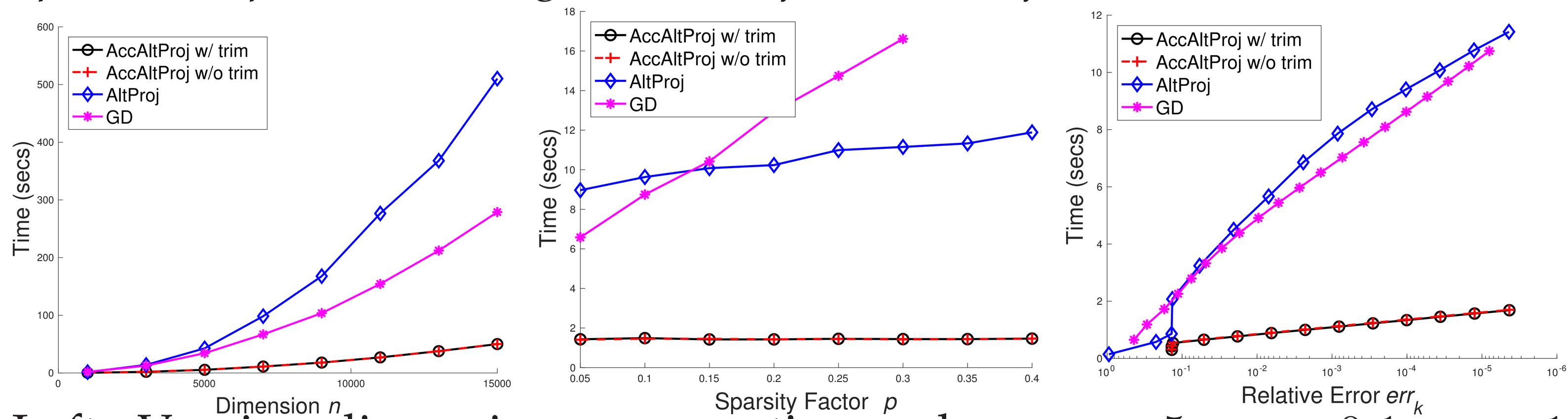
$$\|\boldsymbol{L}-\boldsymbol{L}_k\|_F\le\epsilon\sigma_1^L,\|\boldsymbol{S}-\boldsymbol{S}_k\|_\infty\le\frac{\epsilon}{\sqrt{mn}}\sigma_1^L, \text{ and } supp(\boldsymbol{S}_k)\subset supp(\boldsymbol{S})$$

in $O(\log_\gamma\epsilon)$ iterations.

Numerical Experiments

Synthetic Data: An $n\times n$ rank-r matrix $\boldsymbol{L}$ is formed via $\boldsymbol{L}=\boldsymbol{P}\boldsymbol{Q}^T$, where $\boldsymbol{P},\boldsymbol{Q}\in\mathbb{R}^{n\times r}$ are two random matrices having their entries drawn i.i.d from the standard normal distribution. The locations of the non-zero entries of the sparse matrix $\boldsymbol{S}$ are sampled uniformly and independently without replacement, while the values of the non-zero entries are drawn i.i.d from the uniform distribution over the interval $[-c\cdot\mathbb{E}(|[\boldsymbol{L}]_{ij}|),c\cdot\mathbb{E}(|[\boldsymbol{L}]_{ij}|)]$ for some constant $c>0$. The k^{th} iteration relative computing error is defined as $err_k=\|\boldsymbol{D}-\boldsymbol{L}_k-\boldsymbol{S}_k\|_F/\|\boldsymbol{D}\|_F$.

Speeds Comparisons among AccAltProj [2], AltProj [1] and GD [3]:



Left: Varying dimension n vs runtime, where $r=5$, $\alpha=0.1$, $c=1$, and n varies from 1000 to 15000. The algorithms are terminated after $err_k<1\times10^{-4}$ is satisfied. Center: Varying sparsity factor α vs runtime, where $r=5$, $c=1$ and $n=2500$. The algorithms are terminated when either $err_k<1\times10^{-4}$. Right: Relative error err_k vs runtime, where $r=5$, $\alpha=0.1$, $c=1$, and $n=2500$. The algorithms are terminated after $err_k<1\times10^{-5}$ is satisfied.

Video Background Subtraction: The two videos we have used for this test are *Shoppingmall* and *Restaurant* which can be found at [4]. Each video can be represented by a matrix, where each column of the matrix is a vectorized frame of the video. Then, we apply AccAltProj to decompose the matrix into a low-rank part which represents the static background of the video and a sparse part which represents the moving objects in the video.

References

[1] P. Netrapalli, U. Niranjan, S. Sanghavi, A. Anandkumar, and P. Jain, "Non-convex robust PCA," in *Advances in Neural Information Processing Systems*, 2014.

[2] H. Cai, J.-F. Cai, and K. Wei, "Accelerated alternating projections for robust principal component analysis," *Journal of Machine Learning Research*, 2019.

[3] X. Yi, D. Park, Y. Chen, and C. Caramanis, "Fast algorithms for robust PCA via gradient descent," in *Advances in neural information processing systems*, 2016.

[4] perception.i2r.a-star.edu.sg/bk_model/bk_index.html.