An introduction to non-convex analysis of Robust PCA

Praneeth K Narayanamurthy

pkurpadn@iastate.edu

November 28, 2016

Overview

- Problem motivation
  - Classical Principal Components analysis (PCA)
  - Robust PCA
- Convex solution
- Non-convex solution
- Convex vs Non-Convex solutions
- Analysis of non-convex solution
- Simulation Results

Problem Motivation

- **Given:** Data $\mathbf{x}_i \in \mathbb{R}^n$, $i = 1,\ 2,\ \cdots,\ m$.
- **Assumption:** $\mathbf{x}_i$ lie in low-dimensional space, $\mathbb{R}^k$ where $k \ll n$.
- **Goal:** Estimate the k-dimensional subspace.
- Define

$$\mathbf{X} = \begin{bmatrix} -\mathbf{x}_1^T- \\ -\mathbf{x}_2^T- \\ \vdots \\ -\mathbf{x}_m^T- \end{bmatrix}$$

Problem Motivation

Classical Principal Components analysis (PCA)

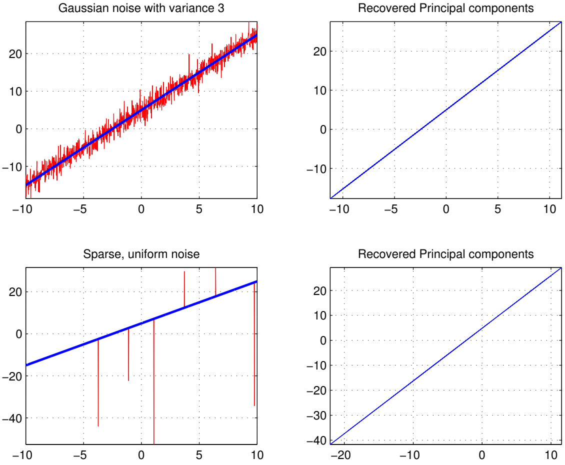
- **Input:** $\mathbf{X}$, $\text{rank}(\mathbf{X}) = r$, **Output:** $\hat{\mathbf{X}}$

$$\hat{\mathbf{X}} = \min \|\hat{\mathbf{X}} - \mathbf{X}\|_2$$
$$\text{subject to} \quad \text{rank}(\hat{\mathbf{X}}) \leq r$$

- Non-Convex Problem; but efficient algorithm to compute exact solution exists.
- **Algorithm:** Return the top r left singular vectors of $\mathbf{X}$ – using *Singular Value Decomposition (SVD)*
  - Advantages: Guaranteed Convergence*, Numerically stable
  - Drawback: Computationally intensive ($\mathcal{O}(n^2 r)$), Sensitive to outliers

Problem Motivation

Why is PCA sensitive to outliers?



Problem Motivation

Transition towards robust PCA

- Information Revolution – Very large-scale data, but intrinsically low dimensional.
- Examples: Image/Video/Multimedia processing, Web Search engines, Recommender Systems, Bio-Informatics etc.
- Physical limitations – Grossly corrupted, unreliable, missing data.
- Need for a more generic problem formulation.

Problem Motivation

Robust PCA

- Input: $\mathbf{M} = \mathbf{L}^* + \mathbf{S}^*$, Output: $\hat{\mathbf{L}}$, $\hat{\mathbf{S}}$
- $\hat{\mathbf{L}}$ is low-rank, and $\hat{\mathbf{S}}$ is sparse.
- Non-Convex problem.
- Ill-Posed – requires additional assumptions on the structure of individual components.

Convex Solution

- Under mild-assumptions, it is possible to recover $\mathbf{L}^*$ and $\mathbf{S}^*$ exactly.
- Solve the following convex relaxation
-

$$(\hat{\mathbf{L}},\ \hat{\mathbf{S}}) = \arg\min \quad \|\mathbf{L}\|_* + \lambda\|\mathbf{S}\|_1$$
$$\text{subject to} \quad \mathbf{M} = \mathbf{L} + \mathbf{S}$$

- Two approaches: Random sparsity model[1] and deterministic sparsity model[2]

[1] E. Candès et al., "Robust Principal Component Analysis?," *Journal of ACM*, 2011

[2] V. Chadrasekaran et al., "Rank Sparsity incoherence for matrix decomposition," *SIAM Journal of Optimization*, 2011

Non-Convex Solution[3]

- Alternating projections on to set of low-rank and sparse matrices
- Non-convex sets but the projection can be performed efficiently using Hard-thresholding and SVD
- Gives exact recovery under mild-assumptions

(L1) Rank of $\mathbf{L}^*$ is at most r

(L2) $\mathbf{L}^*$ is μ-incoherent, i.e., if $\mathbf{L}^* = U^* \Sigma^* (V^*)^T$ is the SVD then $\|(U^*)^i\| \leq \frac{\mu\sqrt{r}}{\sqrt{m}}$ and $\|(V^*)^j\| \leq \frac{\mu\sqrt{r}}{\sqrt{n}}$ $\forall i,\ j$

(S1) Each row and column of $\mathbf{S}^*$ has at most α fraction of non-zero entries, such that $\alpha \leq \frac{1}{512\mu^2 r}$

[3] P. Netrapalli et al., "Non-Convex Robust PCA," *NIPS*, 2014

Input: Matrix $M \in R^{m \times n}$, convergence criterion ϵ, target rank r, thresholding parameter β.

$P_k(A)$ denotes the best rank-k approximation of matrix A. $HT_\zeta(A)$ denotes hard-thresholding, i.e. $(HT_\zeta(A))_{ij} = A_{ij}$ if $|A_{ij}| \geq \zeta$ and 0 otherwise.

Set initial threshold $\zeta_0 \leftarrow \beta\sigma_1(M)$.

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

for Stage $k = 1$ to r **do**

    for Iteration $t = 0$ to $T = 10 \log \left(n\beta \| M - S^{(0)} \| / \epsilon \right)$ **do**

        Set threshold ζ as

$$\zeta \; = \; \beta \left(\sigma_{k+1}(M - S) + \left(\frac{1}{2} \right)^t \sigma_k(M - S) \right)$$

        $L^{(t+1)} = P_k(M - S^{(t)})$

        $S^{(t+1)} = HT_\zeta(M - L^{(t+1)})$

    end for

    if $\beta\sigma_{k+1}(L) < \frac{\epsilon}{2n}$ **then**

        Return: Ł$^{(T)}$, $S^{(T)}$ */* Return rank-k estimate if remaining part has small norm */*

    else

        $S^{(0)} = S^{(T)}$ */* Continue to the next stage */*

    end if

end for

Return: $L^{(T)}, S^{(T)}$

Convex vs. Non-Convex solutions

Algorithm	PCA*	Convex	Non-Convex
Run Time (per iteration)	$\mathcal{O}(rmn)$	$\mathcal{O}(m^2 n)$	$\mathcal{O}(r^2 mn)$
# iterations	$\mathcal{O}(\log(1/\epsilon))$	$\mathcal{O}(1/\epsilon)$	$\mathcal{O}(\log(1/\epsilon))$

- Above comparisons are for similar assumptions on sparsity and incoherence in convex and non-convex solutions.

  *Using the power method

Analysis of non-convex solution

- **Theorem:** Under conditions (L1), (L2), and (S1), and the choice of β as above, the outputs $\hat{L}$ and $\hat{S}$ of Algorithm satisfy:

$$\|\hat{L} - L^*\|_F \leq \epsilon, \; \|\hat{S} - S^*\|_\infty \leq \frac{\epsilon}{\sqrt{nm}}, \; \mathrm{supp}(\hat{S}) \subseteq \mathrm{supp}(S^*)$$

- Proof outline
  1. Reduce the problem to symmetric case, maintaining the assumptions
  2. Show decay in $\|L - L^*\|_\infty$ after projection onto set of rank-k matrices.
  3. Show decay in $\|S - S^*\|_\infty$ after projection onto set of sparse matrices
  4. Recurse the argument.

Analysis of non-convex solution

Some key ideas

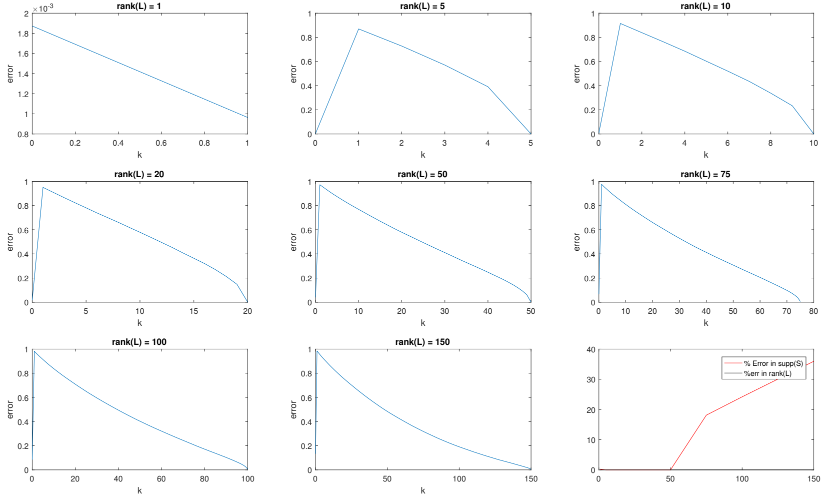
- Incoherence, sparsity assumption on the symmetrized versions (Remark)
- Fixed-point convergence characterization of eigenvalues of error term (Lemma 7)
- Counting p-hops on sparse graphs. (Lemma 5)

Simulation Results

Simulation conditions

- $m = 256$, $n = 256$.
- Generated $\text{supp}(S^*)$ uniformly at random with probability $p = 0.9 \implies \approx 6000$ non zero entries.
- T is the maximum of 50, value obtained by the formula given in Algorithm 1.

Simulation Results



Thank you.