Robust Principal Component Analysis using Facial Reduction

Shiqian Ma, Fei Wang, Linchuan Wei, Henry Wolkowicz

Workshop on Modern Trends in Nonconvex Optimization, ICML 2018

Table of contents

Facial Reduction

Linear Programming

Well known LP formulation:

$$\theta_D = \sup\{b^T y | A^T y \leq c\}$$

And it's dual form:

$$\theta_P = \inf\{c^T x | Ax = b, x \geq 0\}$$

Cone Programming

Lets generalize linear programming by using cone based inequalities:

$$\theta_D = \sup\{b^T y | A^T y \leq_{\mathcal{K}} c\} = \sup\{b^T y | c - A^T y \in \mathcal{K}\}$$

And it's dual form:

$$\theta_P = \inf\{c^T x | Ax = b, x \in \mathcal{K}^*\}$$

This is equivalent to linear programming if $\mathcal{K} = \mathbb{R}^n_+$

Some Notation

Lets introduce some additional notations:

- $\mathcal{A} = \{c - A^T y | y \in \mathbb{R}^n\}$
- $\mathcal{F}_D = \mathcal{A} \cap \mathcal{K}$
- $\theta_D(\mathcal{F}) = \sup\{b^T y | c - A^T y \in \mathcal{F}\}$
- $H_c^- = \{x | c^T x \leq 0\}$

Possible problem

The set of all feasible F_D located inside the intersection of to sets: $\mathcal{A} = \{c - A^T y | y \in \mathbb{R}^n\}$ and $\mathcal{K}$.

In general difference between this two sets can be very big and leads to big duality gap and volatile problems.

The key behind Facial Reduction Algorithm is to reduce the size of $\mathcal{K}$

Problem Size Reduction and Faces

Let's formalize what does it mean to reduce the size of $\mathcal{K}$

- We are looking for a subset F such that $\mathcal{F}_D \subset \mathcal{F} \subset \mathcal{K}$
- $\mathcal{F}$ is a face of the cone $\mathcal{K}$ which means that for any x and y from $\mathcal{K}$ if $x + y \in \mathcal{F}$ it leads to $x, y \in \mathcal{F}$
- Ideally we want to find the smallest face $\mathcal{K}_{\min} = \text{face}(\mathcal{F}_D, \mathcal{K})$

Faces and Exposing Vectors

It'is possible to parametrize a face with a single vector. Let's start with defining dual cone:

$$\mathcal{K}^* = \{\varphi \in \mathbb{R}^n | (\varphi, k) \geq 0, \forall k \in \mathcal{K}\} \tag{1}$$

The face $\mathcal{F}$ is called an exposed face if:

$$\exists \varphi \in \mathcal{K}^* \text{ such that } \mathcal{F} = \varphi^\perp \cap \mathcal{K} \tag{2}$$

Some Properties

Facial reduction algorithms heavily use this two results:

Lemma 1

Let $\mathcal{F}$ be a face of K such that $\mathcal{F} \cap \mathcal{A} = \mathcal{F}_D$ If $ri(\mathcal{F}) \cap \mathcal{A} \neq \varnothing$, then $\mathcal{F} = \mathcal{K}_{\min}$

Lemma 2

Let $\mathcal{F}$ be a face of K such that $ri(\mathcal{F}) \cap \mathcal{A} = \varnothing$, then exist $w \in ker(A) \cap \mathcal{F}^*$ such that one of statement is true:

- $c^T w < 0$ and $\theta_D(\mathcal{F}) = -\infty$
- $c^T w = 0$ and $\mathcal{F} \cap \{w\}^\perp \cap \mathcal{A} = \mathcal{F} \cap \mathcal{A}$

Facial Reduction Algorithm

1. Set $i = 0$ and $\mathcal{F}_0 = \mathcal{K}$
2. If $ker(A) \cap H_c^- \cap \mathcal{F}_i^* \subseteq span(w_1, \ldots, w_i)$ then stop. $\mathcal{F}_i = \mathcal{K}_{\min}$
3. Find $w_{i+1} \in (ker(A) \cap H_c^- \cap \mathcal{F}_i^*) - span(w_1, \ldots, w_i)$
4. If $c^T w_{i+1} < 0$ then stop. The problem is infeasible
5. Set $\mathcal{F}_{i+1} = \mathcal{F}_i \cap \{w_{i+1}\}^\perp$ and $i = i + 1$. Go to step 2.

Robust PCA Formulation and Relaxation

Robust PCA

The main goal is to decompose matrix $Z \in \mathbb{R}^{m \times n}$ using lower rank approximation:

$$\begin{cases} \mathrm{rank}(L) + \mu \|S\|_0 \to \min\limits_{L,S} \\ L + S = Z. \end{cases} \tag{3}$$

- L – **dense** lower rank approximation,
- S – **sparse** noise component,
- $\mu > 0$ fixed number.

This is an NP-Hard problem. No easy solutions.

Partially Observed Data

If we observes only elements with indices $(i, j) \in \hat{E}$ problem formulation should be modified

$$\begin{cases} \text{rank}(L) + \mu \|S\|_0 \to \min\limits_{L,S} \\ \mathcal{P}_{\hat{E}}(L + S) = z. \end{cases} \tag{4}$$

Where $\mathcal{P}_{\hat{E}}$ keeps only elements with indices from $\hat{E}$

Straightforward approach is to approximate $\|\cdot\|_0$ norm with $\|\cdot\|_1$ norm. And $\operatorname{rank}(\cdot)$ with sum of eigenvalues also known as singular norm $\|\cdot\|_*$:

$$\begin{cases} \|L\|_* + \mu\|S\|_1 \to \min\limits_{L,S} \\ L + S = Z. \end{cases} \tag{5}$$

The followed problem is equivalent to the initial one:

$$\begin{cases} \operatorname{rank}(Y) + \mu\|S\|_0 \to \min\limits_{Y,S} \\ \mathcal{P}_{\hat{E}}(L+S) = z \\ Y = \begin{bmatrix} W_1 & L \\ L^T & W_2 \end{bmatrix} \succeq 0 \end{cases} \tag{6}$$

Optimal L^* is a submatrix of optimal Y^*

Optimization Algorithm

Faces in the Cone $\mathcal{S}_+^n$

Let V be a subspace in R^n, then the following subset is a face in the cone $\mathcal{S}_+^n$

$$\mathcal{F}_V = \{X \in \mathcal{S}_+^n | \mathrm{im}(Y) \subseteq V\} \tag{7}$$

$F_{\mathrm{im}(X)}$ is the smallest face containing matrix X.

Faces in the Cone $\mathcal{S}_+^n$

We want to find a subspace containing $\text{im}(X)$. Let use SVD for the matrix X:

$$X = \begin{bmatrix} P & Q \end{bmatrix} \begin{bmatrix} D & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} P \\ Q \end{bmatrix} \tag{8}$$

Matrix P can be used to build a projection on subspace formed by matrix X:

$$\text{face}(X) = P\mathcal{S}_+^r P^T = \mathcal{S}_+^n \cap (QQ^T)^\perp \tag{9}$$

Exposing Vectors in $\mathcal{S}_+^n$

Theorem

Consider a linear transformation $\mathcal{M} : \mathcal{S}_+^n \to \mathbb{R}^m$ and a nonempty feasible set

$$\mathcal{F} = \{X \in \mathcal{S}_+^n | \mathcal{M}(X) = b\}, \tag{10}$$

for some $b \in \mathbb{R}^m$. Then a vector v exposes a proper face of $\mathcal{M}(\mathcal{S}_+^m)$ if and only if:

$$0 \neq \mathcal{M}^* v \in \mathcal{S}_+^n \quad \text{and} \quad (v, b) = 0 \tag{11}$$

Let N denote the smallest face of $\mathcal{M}(\mathcal{S}_+^n)$ containing b. Then:

1. We always have
$$\mathcal{S}_+^n \cap \mathcal{M}^{-1} N = \text{face}(\mathcal{F}) \tag{12}$$
2. For any vector $v \in \mathbb{R}^m$:
$$v \text{ exposes } N \iff \mathcal{M}^* v \text{ exposes face}(\mathcal{F}) \tag{13}$$

Exposing Vectors in $\mathcal{S}_+^n$

We have to define several entities to find surface based on theorem 1.

- Linear mapping $\mathcal{M}$ is the coordinate projection onto the leading principal submatrix $\mathcal{S}_+^k$ of order k. Submatrix B is transformed into a vector $b = vec(B)$ of size $m = k(k+1)/2$
- $V \in \mathcal{S}_+^k$, trace$(VB) = 0$, $v = vec(V)$
- $Y = \mathcal{M}^* v$ – an exposing vector for the face $\mathcal{F}$

Graph Representation

Bipartie graph $G_Z((U_m, V_n), \hat{E})$ is associated with $Z \in \mathbb{R}^{m \times n}$

- Nodes $U_m = 1, \ldots, m \bigcup V_n = 1, \ldots, n$ represent different axes of Z
- $\hat{E}$ edges (i, j) corresponds to the elements presented in Z

Note that bicliques represent fully observed submatrices.

Let's find all bicliques in a graph G_Z

Submatrix Decomposition

In order to decompose fully observed submatrix $\bar{Z}$ let solve optimization problem:

$$\begin{cases} \frac{1}{2}\|\bar{L} + \bar{S} - \bar{Z}\|_F^2 \to \min\limits_{\bar{L},\bar{S}} \\ \text{rank}(\bar{L}) \leq \bar{r}, \qquad \|S\|_0 \leq \bar{s} \end{cases} \tag{14}$$

$\bar{r}$ and $\bar{s}$ are fixed parameters. This problem is much easier due to the size and optimization of Frobenius norm.

PALM Optimization

In order to find the solution we will basically do alternate gradient descent with projections:

1. $G_L^k = \bar{L} - \frac{1}{\gamma_1}(\bar{L}^k + \bar{S}^k - \bar{Z})$,
2. $\bar{L}^{k+1} = \arg\min_{\bar{L}}\{\|\bar{L} - G_L^k\|_F^2 : \mathrm{rank}(\bar{L}) \leq \bar{r}\}$,
3. $G_S^k = \bar{S}^k - \frac{1}{\gamma_2}(\bar{L}^{k+1} + \bar{S}^k - \bar{Z})$,
4. $\bar{S}^{k+1} = \arg\min_{\bar{S}}\{\|\bar{S} - G_S^k\|_F^2 : \|\bar{S}\|_0 \leq \bar{s}\}$.

Facial Reduction for Robust PCA

$\bar{L}$ processing

What we have:

- $\bar{Z}$ – fully observed submatrix of Z
- $\bar{Z} = \bar{L} + \bar{S}$, $\mathrm{rank}(\bar{L}) = r$ – submatrix decomposition
- Without loss of generality assume $L = \begin{bmatrix} L_1 & L_2 \\ \bar{L} & L_3 \end{bmatrix}$
- $\bar{L} = \begin{bmatrix} \bar{P} & \bar{U} \end{bmatrix} \begin{bmatrix} \Sigma_r & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} \bar{Q} \\ \bar{V} \end{bmatrix}$ – $\bar{L}$ SVD decomposition
- $\begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & \bar{U}\bar{U}^T & 0 & 0 \\ 0 & 0 & \bar{V}\bar{V}^T & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$ is an exposed vector for face containing $\bar{L}$

Reducing the Size of a Problem

Exposing vector allows us to decrease dim of a possible solution.

$$V = \mathrm{Null}(Y_{\text{expo}}) = \begin{bmatrix} V_P & 0 \\ 0 & V_Q \end{bmatrix}, \quad V_P^T V_P = I_{r_p}, \quad V_Q^T V_Q = I_{r_q} \tag{15}$$

Find the projection on this subspace:

$$Y^* = VRV^T = \begin{bmatrix} V_P R_p V_P^T & V_P R_{pq} V_Q^T \\ V_Q R_{pq}^T V_P^T & V_Q R_q V_Q^T \end{bmatrix} \tag{16}$$

Basically we are interesting in optimization matrix R_{pq}

Reduced Problem

We are interesting in optimization matrix R_{pq}

$$\begin{cases} \operatorname{rank}(R_{pq}) + \mu \|S\|_0 \to \min \\ \mathcal{P}_{\hat{E}}(V_P R_{pq} V_Q^T) + \mathcal{P}_{\hat{E}}(S) = z \end{cases} \tag{17}$$

Further Reducing the Size

On stage with PALM we exactly recovered the values of $\bar{S}$. This mean we can remove some entries from linear constrains. Let $\hat{E}_S$ be the set of exactly recovered s and $\hat{E}_{S^c}$ set of non recovered s

$$\begin{cases} \operatorname{rank}(R_{pq}) + \mu\|S\|_0 \to \min \\ \mathcal{P}_{\hat{E}_S}(V_P R_{pq} V_Q^T) = z_{\hat{E}_S} \\ \mathcal{P}_{\hat{E}_{S^c}}(V_P R_{pq} V_Q^T) + s = z_{\hat{E}_{S^c}} \end{cases} \tag{18}$$