Low rank SDP extreme points and Applications

Mohit Singh

Georgia Tech

SDP extreme points

- Pataki, Gábor. "On the rank of extreme matrices in semidefinite programs and the multiplicity of optimal eigenvalues." *Math of OR '98*.
- Barvinok, Alexander. "Problems of distance geometry and convex properties of quadratic maps." *Discrete & Computational Geometry '95*.

- Applications.
  - S-lemma [Yakubovich' 71]
  - Fair Dimensionality Reduction in Data Analysis [Tantipongpipat, Samadi, Morgernstern, Singh, Vempala '18,'19]

LP extreme points

- Consider the linear program where $A \in \Re^{m \times n}$, $b \in \Re^{m}$

$$\begin{aligned} \min c^T x \\ Ax \geq b \\ x \geq 0 \end{aligned}$$

Theorem: Every extreme point has at most m non-zero variables. Therefore, there exists an optimal solution that has at most m non-zero variables.

Numerous generalization, applications. [Neil's Talk tomorrow]

Semi-definite Programming

- A symmetric $n \times n$ matrix X is positive semi-definite (definite), i.e. $X \succcurlyeq 0$ $(X \succ 0)$ if
  - X has non-negative(positive) eigenvalues.
  - $X = UDU^T$ for some $U \in \Re^{n\times r}$ with orthogonal columns and $D \in \Re^{r\times r}$ symmetric, diagonal with positive entries. (r=n).
  - $v^T X v \geq 0$ for all $v \in \Re^n$. ($v^T X v > 0$ for all $v \in \Re^n \backslash \{0\}$.)

- Let $\langle A, B\rangle = \sum A_{ij} B_{ij} = Tr(A^T B)$

- $SDP\ (I)$:
- $$\min \langle C, X\rangle$$
  $$\langle A_i, X\rangle \geq b_i \quad \forall i = 1, \ldots, m$$
  $$X \succcurlyeq 0$$

Main Result

- $\min \langle C, X \rangle$

  $\langle A_i, X \rangle \geq b_i \quad \forall i = 1, \ldots, m$

  $X \succcurlyeq 0$
- Theorem[Barvinok'95, Pataki'98]: Every extreme point X of the above SDP has rank at most r where $t(r) \coloneqq \frac{r(r+1)}{2} \leq m$.

- Corollary 1: SDP has an optimal solution with rank at most $\sqrt{2m + \frac{1}{4}} - \frac{1}{2}$.
- Corollary 2: If m=2, then there is always a rank 1 optimal solution.

Proof

- Suppose not!
- Let X be rank r where $t(r) > m$. We find Y such that $X + Y$ and $X - Y$ are feasible. Contradiction.
- $X = UDU^T$ where $U \in \Re^{n \times r}$ with orthonormal columns, $D \in \Re^{r \times r}$ diagonal with positive entries.
- Search for $Z \in \Re^{r \times r}$, s.t. $U(D + Z)U^T$, $U(D - Z)U^T$ are feasible.

Want: $\langle A_i, U(D \mp Z)U^T \rangle \geq b_i \ \ \forall i = 1, \ldots, m$

$$U(D \mp Z)U^T \succcurlyeq 0$$

Proof (Contd)

- Claim: It is enough to ensure
  - $\langle A_i, UZU^T \rangle = 0 \quad \forall i = 1, \ldots, m$
  - Z symmetric.

Proof:

- $\langle A_i, U(D \mp Z)U^T \rangle = \langle A_i, UDU^T \rangle \mp \langle A_i, UZU^T \rangle \geq b \mp 0$.
- Eigenvalues of $D \mp Z$ are same as eigenvalues of $U(D \mp Z)U^T$.
- But $D \succ 0$ and therefore, if $\mathrm{Z} \leftarrow \epsilon Z$ will ensure $D \mp \epsilon Z \succ 0$.

But the above are $m + \frac{r(r-1)}{2}$ constraints overs r^2 variables.

If $t(r) > m$, then there is always a non-trivial solution.

Main Result

- $\min\langle C, X\rangle$

  $\langle A_i, X\rangle \geq b_i \quad \forall i = 1, \ldots, m$

  $X \succcurlyeq 0$

- Theorem[Barvinok'95, Pataki'98]: Every extreme point X of the above SDP has rank at most r where $t(r) \coloneqq \frac{r(r+1)}{2} \leq m$.

- Corollary 1: SDP (I) has an optimal solution with rank at most $\sqrt{2m + \frac{1}{4}} - \frac{1}{2}$.
- Corollary 2: If m=2, then there is always a rank 1 optimal solution.

Farkas Lemmas: Linear and Quadratic.

Farkas Lemma: Suppose $a_1, \ldots, a_m, b \in \mathfrak{R}^n$ such that $a_i^T x \geq 0 \ \forall i \Rightarrow b^T x \geq 0$.
Then, $b \geq \lambda_1 a_1 + \lambda_2 a_2 + \cdots + \lambda_m a_m$ some non-negative $\lambda_i \geq 0$.

S-Lemma (Yakubovich '79). Suppose $A, B \in \mathfrak{R}^{n \times n}$ symmetric matrices such that $x^T A\, x \geq 0 \Rightarrow x^T B\, x \geq 0$. Then $B \succcurlyeq \lambda\, A$ for some $\lambda \geq 0$.

Proof: Consider the SDP.

$$\min \langle X, B \rangle$$
$$\langle A, X \rangle \geq 0$$
$$Tr(X) = 1$$
$$X \succcurlyeq 0$$

Dual : Max z

$$y\, A + zI \preccurlyeq B$$
$$y \geq 0$$

Observe that primal optimal is rank 1. Thus $X = xx^T$. But then objective is at least 0. This implies dual is at least 0.

SDP extreme points

- Pataki, Gábor. "On the rank of extreme matrices in semidefinite programs and the multiplicity of optimal eigenvalues." *Math of OR '98*.
- Barvinok, Alexander. "Problems of distance geometry and convex properties of quadratic maps." *Discrete & Computational Geometry '95*.

- Applications.
  - S-lemma [Yakubovich' 71]
  - Fair Dimensionality Reduction in Data Analysis [Tantipongpipat, Samadi, Morgernstern, Singh, Vempala '18,'19]

Dimensionality Reduction

- Data is usually represented in high dimensions.
- There are few relevant directions.
- Dimensionality reduction leads to representation in relevant directions.
  - Also computationally useful for any data analysis or algorithms.

PCA (Principle Component Analysis)

- Data $D \in \mathbb{R}^{m \times n}$ of m data points. Want to reduce from n to d dimensions.
- Minimize reconstruction error:

$$\min_{U:\mathrm{rank}(U)=d} \|D - U\|_F^2$$

$\|D - U\|_F^2$ is the Frobenius norm, sum of square of error.

PCA (Principle Component Analysis)

- Data $D \in \mathbb{R}^{m\times n}$ of m data points. Want to reduce from n to d dimensions.
- Minimize reconstruction error:

$$\min_{U:\mathrm{rank}(U)=d} \|D-U\|_F^2$$

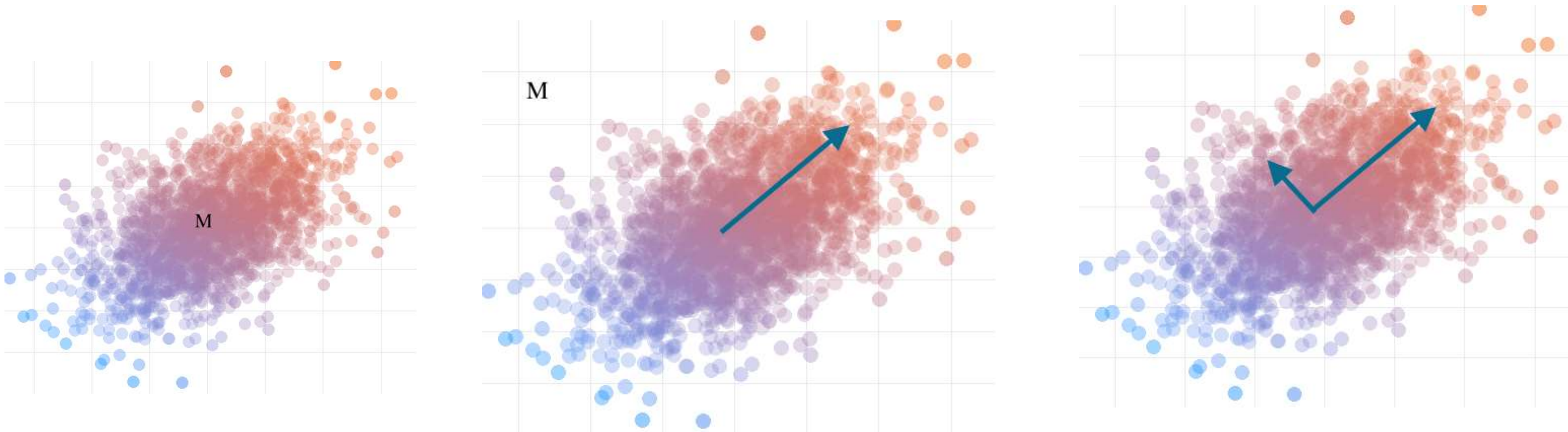
$\|D-U\|_F^2$ is the Frobenius norm, sum of square of error.

- Easily solved by SVD (Singular Value Decomposition). The optimal solution has a form

$$U = DP$$

where P is projection matrix on top d singular vectors of D.

PCA objective



$$\min_{U:\text{rank}(U)=d} \|D - U\|_F^2 = \min_{P \in \mathcal{P}_d} \|D - DP\|_F^2 = Tr(D^T D) - \max_{P \in \mathcal{P}_d} D^T D \cdot P$$

$$\mathcal{P}_d = \{P \in R^{n \times n}: P \text{ } symmetric, rank(P) = d, P^2 = P\}$$

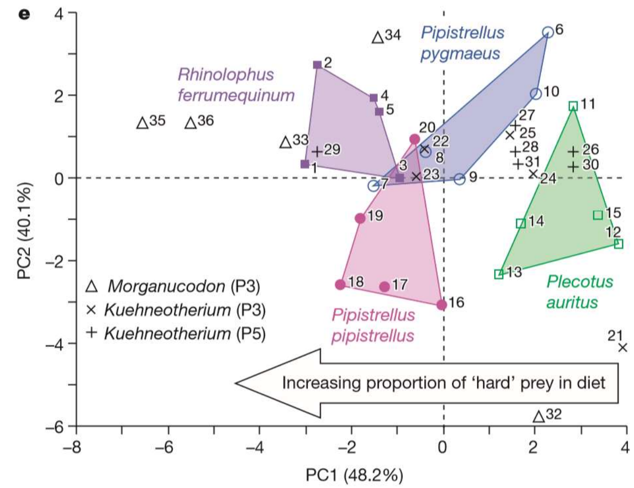
Applications and History of PCA

- Pearson’1901 and Hotelling’1931.
- Standard tool in data analysis.
  - Widely used in sciences, humanities, finance, image recognition.

  [Sirovich, Kirby’87, Turk, Pentland’91]

- Fossil Teeth Data: Kuehneotherium and Morganucodon Species.

  [Gill et al, Nature 2014]

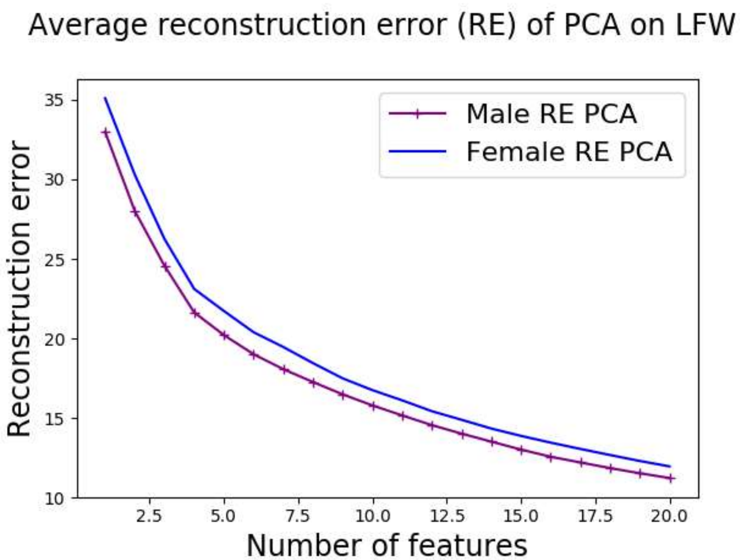


- Random Projection to lower dimensions.
  - Johnson-Lindenstrauss Lemma ‘1984: All distances are preserved up to a small error.

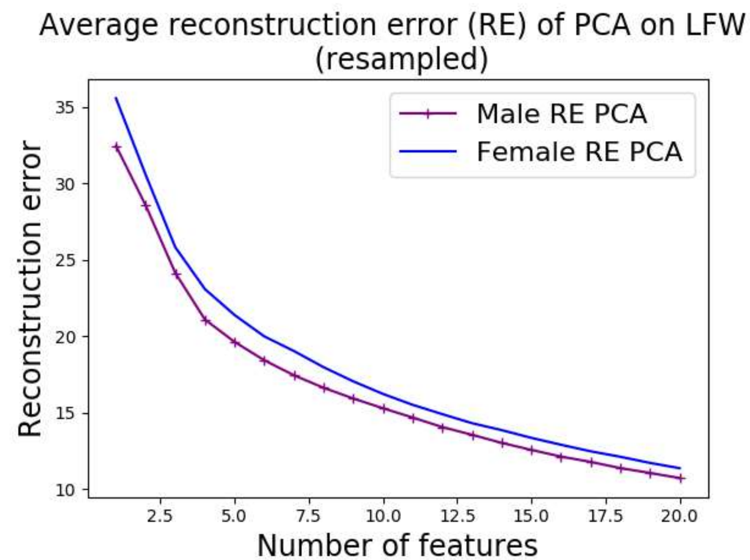
Unfairness of the PCA problem

Data belongs to users of different groups, say images of men and women.
PCA ensures average error in projection is small. Errors for two different groups are different.

Standard PCA on face data LFW of male and female.



Equalizing male and female weight before PCA



Fair PCA

- Given data matrices $D_i \in R^{m_i \times n}$ for $i = 1, \dots, k$ and a projection matrix $P \in R^{n \times n}$.
- $Err(D_i, P) = |D_i - D_i P|_F^2 = Tr\left(D_i^T D_i\right) - D_i^T D_i \cdot P$
- Given target dimension $d < n$.
- Fair PCA: Find a projection matrix P of rank at most d that minimizes the maximum error.

Fair PCA:= $\min_{P \in \mathcal{P}_d} \max_{i \in [k]} Err(D_i, P)$

$$\mathcal{P}_d = \{P \in R^{n \times n}: P\ symmetric, rank(P) = d, P^2 = P\}$$

Fair PCA as rank constrained SDP.

$$\min z$$

$$z \geq Tr\left(D_i^T D_i\right) - D_i^T D_i \cdot P \quad \forall i = 1, \dots, k$$
$$rank(P) = d$$
$$0 \preccurlyeq P \preccurlyeq I$$

Fair Dimensionality Reduction

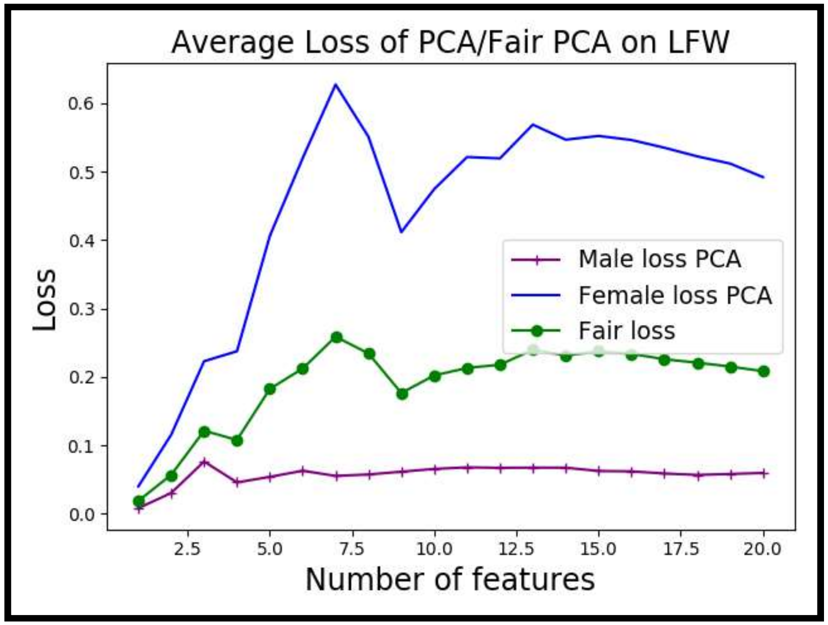
- More generally, we are given utility functions $u_i: \mathcal{P}_d \to \Re$ that measure the utility of each group.
- Moreover, we are given a function $g: \Re^k \to \Re$ that combines these utilities.

Fair DR:= $\max_{P \in \mathcal{P}_d} g\big(u_1(P), u_2(P), \dots, u_k(P)\big)$

$\mathcal{P}_d = \{P \in R^{n\times n}: P\ symmetric, rank(P) = d, P^2 = P\}$

Fair PCA: Special case with $u_i(P) = -Err(D_i, P)$ and $g(.) = \min$

$Loss_i(P) = \|D_i - D_i P\|_2^2 - \|D_i - D_i P_i^*\|_2^2$
where P_i^* is the best rank d projection for group i.
- Loss for being part of the other groups.



Related Work

- Rank Constrained SDPs are widely used.

- Signal processing[Davies and Eldar’12, Ahmed and Romberg’15]
- Distance Matrices: Localization sensors [So and Ye’07], nuclear magnetic resonance spectroscopy [Singer’08]
- Item Response Data, Recommendation Systems[Goldberg et al’93]
- Machine Learning: Multi-task Learning [Obozinski, Taskar, Jordan’10], Natural Language Processing[Blei’12]
- Survey by [Davenport, Romberg’2016]

- Work by Barvinok’95, Pataki’98 on characterizations of extreme points of SDPs.
  - Algorithmic work by [Burer, Monteiro’03]

Our Results

Fair PCA:= $\min_{P\in\mathcal{P}_d} \max_{i\in[k]} Err(D_i, P) \coloneqq |D_i - D_i P|_F^2$

$\mathcal{P}_d = \{P \in R^{n\times n}: P\ symmetric, rank(P) = d, P^2 = P\}$

- Theorem 1: The Fair PCA problem is polynomial time solvable for k=2.
  - "Integrality" of SDPs.
- Theorem 2: The Fair PCA problem is polynomial time solvable for constant k and d.
  - Algorithmic theory of quadratic maps. [Grigoriev and Pasechnik '05]
- Problem is NP-hard for general k, d=1.
- Results generalize to Fair Dimensionality reduction when u_i is linear and g is concave.

Fair DR:= $\max_{P\in\mathcal{P}_d} g\big(u_1(P), u_2(P), \dots, u_k(P)\big)$

$\mathcal{P}_d = \{P \in R^{n\times n}: P\ symmetric, rank(P) = d, P^2 = P\}$

Our Results: Approximation

Fair PCA:= $\min_{P \in \mathcal{P}_d} \max_{i \in [k]} Err(D_i, P) := |D_i - D_i P|_F^2$

$\mathcal{P}_d = \{P \in R^{n \times n}: P\ symmetric, rank(P) = d, P^2 = P\}$

- Theorem 3: There is a polynomial time algorithm for the Fair PCA problem that returns a rank at most $\mathrm{d} + \sqrt{2k + \frac{1}{4}} - \frac{3}{2}$ whose objective is better than the optimum.
  - Extreme Points of SDPs.

- Theorem 4: There is a polynomial time algorithm for the Fair PCA problem that returns a rank at most d whose objective is at most $OPT + \Delta$,

  where $\Delta := \max_{S \subseteq [m]} \sum_{i=1}^{\sqrt{2|S|}} \sigma_i \left(\frac{1}{|S|} \sum_{j \in S} D_j^T D_j \right)$ where $\sigma_i(B)$ is the i^{th} largest singular value of B.

  - Iterative Rounding Framework for SDPs.

SDP extreme points

Theorem 1: Every extreme point of the SDP-Relaxation has rank at most d. Thus the objective of the two programs are identical.

Rank-Constrained SDP

$$\begin{aligned} &\min C \cdot X \\ &A \cdot X \leq b \\ &rank(X) \leq d \\ &0 \preccurlyeq X \preccurlyeq I \end{aligned}$$

SDP-Relaxation

$$\begin{aligned} &\min C \cdot X \\ &A \cdot X \leq b \\ &Trace(X) \leq d \\ &0 \preccurlyeq X \preccurlyeq I \end{aligned}$$

Corollary: Fair PCA is polynomial time solvable for 2 groups.

Related: Barvinok’95, Pataki’98.

S-Lemma [Yakubovich’71].

SDP extreme points

Theorem 2: Every extreme point of the SDP-Relaxation has rank at most $d + \sqrt{2m + 9/4} - 3/2$.

Rank-Constrained SDP

$$\begin{gathered} \min C \cdot X \\ A_i \cdot X \leq b_i \; \forall i = 1, \dots, m \\ rank(X) \leq d \\ 0 \preccurlyeq X \preccurlyeq I \end{gathered}$$

SDP-Relaxation

$$\begin{gathered} \min C \cdot X \\ A_i \cdot X \leq b_i \; \forall i = 1, \dots, m \\ trace(X) \leq d \\ 0 \preccurlyeq X \preccurlyeq I \end{gathered}$$

- Corollary: There is a polynomial time algorithm for the Fair PCA problem that returns a rank at most $\mathrm{d} + \sqrt{2k + \frac{1}{4}} - \frac{3}{2}$ whose objective is better than the optimum.

SDP extreme points

Theorem: Every extreme point of the SDP-Relaxation has rank at most d. Thus the objective of the two programs are identical.

Rank-Constrained SDP	SDP-Relaxation
$\min C \cdot X$ $A \cdot X \leq b$ $rank(X) \leq d$ $0 \preccurlyeq X \preccurlyeq I$	$\min C \cdot X$ $A \cdot X \leq b$ $Trace(X) \leq d$ $0 \preccurlyeq X \preccurlyeq I$

Corollary: Fair PCA is polynomial time solvable for 2 groups.

Related: Barvinok'95, Pataki'98.

S-Lemma [Yakubovich'71].

$$\min C \cdot X$$
$$A_i \cdot X \le b_i \;\; \forall i = 1, \dots, m.$$
$$Tr(X) \le d$$
$$0 \preccurlyeq X \preccurlyeq I$$

Proof

- Let X be an extreme point with r fractional eigenvalues.
- $X = \left[U_1 \; U_f \; U_0\right] \left[diag(1) \; D \; 0\right] \left[U_1 \; U_f \; U_0\right]^{\mathrm{T}} = \mathrm{U_1 U_1^T + U_f D U_f^T}$
- D is $r \times r$ diagonal matrix with $0 < D_{ii} < 1$ and $\left[U_1 \; U_f \; U_0\right]$ is a orthogonal matrix of eigenvectors.

Claim: If $\frac{r(r+1)}{2} > m + 1$ then there exists a $r \times r$ symmetric matrix $F \neq 0$ such that

$\mathrm{Y = U_1 U_1^T + U_f (D + F) U_f^T}$ and $\mathrm{Z = U_1 U_1^T + U_f (D - F) U_f^T}$ are feasible.

Assuming the claim, we obtain a contradiction to extreme point.

Fact: Eigenvalues of Y are same as eigenvalues of $[diag(1) \; D + F \; 0]$ and eigenvalues of Z are same as eigenvalues of $[diag(1) \; D - F \; 0]$.

Proof

$$\min C \cdot X$$
$$A_i \cdot X \leq b_i \ \ \forall i = 1, \dots, m.$$
$$Tr(X) \leq d$$
$$0 \preccurlyeq X \preccurlyeq I$$

- Claim: If $\frac{r(r+1)}{2} > m + 1$ then there exists a $r \times r$ symmetric matrix $F \neq 0$ such that $\mathrm{U_1 U_1^T + U_f(D + F)U_f^T}$ and $\mathrm{U_1 U_1^T + U_f(D - F)U_f^T}$ are feasible.
- Proof: Consider the linear system.
  - $A_i \cdot U_f G U_f^T = 0 \ \forall i = 1, \dots, m.$
  - $Tr\left(U_f G U_f^T\right) = 0$
  - $G_{ij} = G_{ji} \quad \forall i \neq j$

  Number of equations $m + 1 + \frac{\mathrm{r(r-1)}}{2}$. Number of variables r^2.

  If $r^2 > m + 1 + \frac{\mathrm{r(r-1)}}{2}$, then there is a line of solutions, i.e. , G$\neq 0$ such that $\{\lambda\, G : \lambda \in R\}$ all satisfy the above constraints.

  Consider F=ϵG for small enough $\epsilon > 0$.
  - Observe that $\mathrm{U_1 U_1^T + U_f}(D \pm \mathrm{F})\mathrm{U_f^T}$ satisfies the linear constraints.
  - Eigenvalues of $[diag(1)\ D + F\ 0]$ = $[diag(1)\ D + \epsilon G\ 0]$ are bounded away from 0 and 1.

Technical Result SDP extreme points

Theorem: Every extreme point of the SDP

$$
\begin{aligned}
&\min C \cdot X \\
&\quad A_i \cdot X \leq b_i \quad \forall i = 1, \dots, m \\
&\quad Tr(X) \leq d \\
&\quad 0 \preccurlyeq X \preccurlyeq I
\end{aligned}
$$

has rank at most $d + \sqrt{2m + 9/4} - \frac{3}{2}$.

- m=1 we obtain rank is at most d.

Generalizes Barvinok'95, Pataki'98.

- Similar results for SDPs with affine constraints.

Iterative Rounding

Theorem: There is an iterative rounding algorithm that given

$$\begin{aligned} &\min C \cdot X \\ &A_i \cdot X \leq b_i \qquad \forall i = 1, \dots, m \\ &Tr(X) \leq d \\ &0 \preccurlyeq X \preccurlyeq I \end{aligned}$$

with optimal solution X^* returns a feasible solution Y s.t.

1. $\mathrm{rank}(Y) \leq d$.
2. $C \cdot Y \leq C \cdot X^*$.
3. $A_i \cdot Y \geq A_i \cdot X^* - \Delta$

Where $\Delta = \max_{S \subseteq [m]} \sum_{i=1}^{\sqrt{2|S|}} \sigma_i \left(\frac{1}{|S|} \sum_{j \in S} A_j \right)$ where $\sigma_i(B)$ is the i^{th} largest singular value of B.

Idea: Fix eigenvalues to 0 and 1.

Maintain two subspaces F_0 and F_1 for corresponding eigenfaces.

Update SDP to work only in the orthogonal space F.

Show a constraint can be removed or one of the eigenvalues is 0 or 1.

Iterative Algorithm

$$\max \langle F^T C F, X(r)\rangle$$
$$\langle F^T A_i F, X(r)\rangle \geq b_i - F_1^T A_i F_1 \quad i \in S$$
$$\mathrm{tr}(X) \leq d - \mathrm{rank}(F_1)$$
$$0 \preceq X(r) \preceq I_r$$

1. Initialize F_0, F_1 to be empty matrices and $F = I_n$, $S \leftarrow \{1, \ldots, m\}$.
2. If the $\mathbb{SDP}$ is infeasible, declare infeasibility. Else,
3. While F is not the empty matrix.
   (a) Solve $\mathbb{SDP}(r)$ to obtain extreme point $X^*(r) = \sum_{j=1}^{r} \lambda_j v_j v_j^T$ where λ_j are the eigenvalues and $v_j \in \mathbb{R}^r$ are the corresponding eigenvectors.
   (b) For any eigenvector v of $X^*(r)$ with eigenvalue 0, let $F_0 \leftarrow F_0 \cup \{Fv\}$.
   (c) For any eigenvector v of $X^*(r)$ with eigenvalue 1, let $F_1 \leftarrow F_1 \cup \{Fv\}$.
   (d) Let $X_f = \sum_{j:0<\lambda_j<1} \lambda_j v_j v_j^T$. If there exists a constraint $i \in S$ such that $\langle F^T A_i F, X_f\rangle < \Delta(\mathcal{A})$, then $S \leftarrow S \setminus \{i\}$.
   (e) For every eigenvector v of $X^*(r)$ with eigenvalue not equal to 0 or 1, consider the vectors Fv and form a matrix with these columns and use it as the new F.
4. Return $\tilde{X} = F_1 F_1^T$.

Conclusion

- Low rank SDP solutions under affine constraints. (Pataki, Barvinok).
- Low rank SDP solutions under PSD constraint $0 \preccurlyeq X \preccurlyeq I$.
  - Applications to fair PCA problem.
- In practice, these algorithms take 10-15 PCAs for 2 groups using MW update.
  - Code and data available at https://github.com/samirasamadi/Fair-PCA
- Other applications to low rank models in other areas?

Thanks!