Low-Rank Matrix Completion with Facial Reduction

Shimeng Huang\textsuperscript{1} \quad Henry Wolkowicz\textsuperscript{2}

\textsuperscript{1}Department of Statistics and Actuarial Science
University of Waterloo

\textsuperscript{2}Department of Combinatorics and Optimization
University of Waterloo

AT ECM2017, Thurs. June 1, 10:50AM
Finding low rank matrix completions is a numerically hard nonconvex problem.

A popular convex relaxation is the nuclear norm which is SDP-representable; and, both the SDP and its dual satisfy strict feasibility (Slater’s constraint qualification).

For inequality constrained optimization problems, perhaps the most important key is to identify the active constraints. We aim to do facial reduction for the optimal face of the SDP, i.e., identify the “active” face.

Thus we (try to) avoid a need for a SDP solver.
Low-Rank Matrix Completion

Example (Partial Matrix with Noise ——— BUT Low Rank)

\[
\begin{pmatrix}
1.01 & 2 & \?
\end{pmatrix}
\begin{pmatrix}
1 & ? & 2.99
\end{pmatrix}
\]
Low-Rank Matrix Completion

Example (Partial Matrix with Noise ——- BUT Low Rank)

\[
\begin{pmatrix}
1.01 & 2 & ? \\
1 & ? & 2.99
\end{pmatrix}
\quad \begin{pmatrix}
1 & 2 & 3 \\
1 & 2 & 3
\end{pmatrix}
\]
Low-Rank Matrix Completion

Example (Partial Matrix with Noise ——- BUT Low Rank)

<table>
<thead>
<tr>
<th>1.01</th>
<th>2</th>
<th>?</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>?</td>
<td>2.99</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>1</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
</tbody>
</table>

Problem Statement (non-convex & intractable)

Given a real partial matrix $z \in \mathbb{R}^{\hat{E}}$ with some level of noise,

$$(LRMC) \quad \min \quad \text{rank}(M) \quad \text{s.t.} \quad \|\mathcal{P}_{\hat{E}}(M) - z\| \leq \delta, \quad M \in \mathbb{R}^{m \times n}$$

- $\hat{E}$ indices for known entries (sampled data) in $Z \in \mathbb{R}^{m \times n}$;
- with coordinate projection/partial matrix $z = \mathcal{P}_{\hat{E}}(Z) \in \mathbb{R}^{\hat{E}}$
- $\delta > 0$ is a tuning parameter
Applications Include:

- data science
- model reduction
- collaborative filtering (Netflix problem)
- sensor network localization
- pattern recognition
- various machine learning scenarios
Minimizing rank is a hard nonconvex problem.

Rank is a lower semi-continuous function.

Nuclear Norm Minimization (convex relaxation)

The problem (LRMC) can be approximated by

\[
\begin{align*}
(\text{NN-LRMC}) & \quad \min & \|M\|_* \\
\text{s.t.} & & \|\mathcal{P}_{\hat{E}}(M) - z\| \leq \delta
\end{align*}
\]

- \(\|M\|_* = \sum \sigma_i(M)\), sum of singular values, nuclear norm (Schatten 1-norm, Ky-Fan \(r\)-norm, trace norm)
- \(\|UXV^T\|_* = \|X\|_*\) unitarily invariant
Theorem (Fazel, Hindi, Boyd ’01)

\[ \|X\|_\star \text{ is the convex envelope of rank } X \text{ on } \{X \in \mathbb{R}^{m \times n} : \|X\| \leq 1\} . \]

Properties of nuclear norm:

- “best” convex lower approximation of rank function
- The nuclear ball is the convex hull of the intersection of rank-1 matrices with the unit ball:
  \[ \text{conv}\{uv^T : u \in \mathbb{R}^n, v \in \mathbb{R}^m, \|u\| = 1, \|v\| = 1\} \]
- SDP-representable
- Related references by: Candes, Fazel, Parrilo, Recht
SDP Embedding Lemma

Let \( M \in \mathbb{R}^{m \times n} \) and \( t \in \mathbb{R} \). Then:

\[
\|M\|_* \leq t
\]

if, and only if,

there exist (symmetric) \( W_1 \) and \( W_2 \) such that

\[
\begin{bmatrix}
W_1 & M \\
M^T & W_2
\end{bmatrix} \succeq 0, \quad \text{trace}(W_1) + \text{trace}(W_2) \leq 2t.
\]

compact SVD: \( M = U\Sigma V^T \), \( \|M\|_* = \text{trace} \Sigma \leq t \)

\[
\begin{bmatrix}
U\Sigma^{1/2} \\
V\Sigma^{1/2}
\end{bmatrix}
\begin{bmatrix}
U\Sigma^{1/2} \\
V\Sigma^{1/2}
\end{bmatrix}^T =
\begin{bmatrix}
U\Sigma U^T & U\Sigma V^T \\
V\Sigma U^T & V\Sigma V^T
\end{bmatrix} \succeq 0
\]

For necessity, set \( W_1 = U\Sigma U^T \), \( W_2 = V\Sigma V^T \); for sufficiency, exploit

\[
\text{range } M \subseteq \text{range } W_1, \text{range } M^T \subseteq \text{range } W_2.
\]
Nuclear Norm Low Rank Problem, (NN-LRMC)

Semidefinite Embedding: Trace Minimization

Problem (NN-LRMC) can be formulated as:

\[(SDP-LRMC) \quad \min \quad \frac{1}{2} \text{trace}(Y) \]
\[s.t. \quad \|\mathcal{P}_{\bar{E}}(Y) - z\| \leq \delta \]
\[Y \succeq 0 \]

where \(Q = \begin{bmatrix} 0 & Z \\ Z^T & 0 \end{bmatrix}\), \(z = \mathcal{P}_{\bar{E}}(Z) = \mathcal{P}_{\bar{E}}(Q)\);
\(\bar{E}\) is set of indices in \(Q\) corresponding to known entries of \(Z\).

\[
Y = \begin{bmatrix} W1 & M \\ M' & W2 \end{bmatrix}
\]
First, an Example of Facial Reduction, FR

Example (Facial Reduction in Linear Programming)

\[
\begin{align*}
\text{min} & \quad (2 \quad 5 \quad -1 \quad 4 \quad 7) x \\
\text{s.t.} & \quad \begin{bmatrix} 1 & 1 & -1 & 3 & 1 \\ -1 & 1 & 2 & 2 & -1 \end{bmatrix} x = \begin{bmatrix} 1 \\ -1 \end{bmatrix} \\
x & \geq 0, \quad x \in \mathbb{R}^5
\end{align*}
\]
First, an Example of Facial Reduction, FR

Example (Facial Reduction in Linear Programming)

\[
\begin{align*}
\min & \quad (2 \quad 5 \quad -1 \quad 4 \quad 7) x \\
\text{s.t.} & \quad \begin{bmatrix} 1 & 1 & -1 & 3 & 1 \\ -1 & 1 & -2 & 2 & -1 \end{bmatrix} x = \begin{bmatrix} 1 \\ -1 \end{bmatrix} \\
x & \geq 0, \quad x \in \mathbb{R}^5
\end{align*}
\]

If we sum the two constraints we get a facial constraint

\[
2x_2 + x_3 + 5x_4 = 0 \quad \implies \quad x \in \mathcal{F} = \{x \in \mathbb{R}_+^5 : x_2 = x_3 = x_4 = 0\}
\]
First, an Example of Facial Reduction, FR

Example (Facial Reduction in Linear Programming)

\[
\begin{align*}
\text{min} \quad & (2 \quad 5 \quad -1 \quad 4 \quad 7)x \\
n\text{s.t.} \quad & \begin{bmatrix} 1 & 1 & -1 & 3 & 1 \\ -1 & 1 & 2 & 2 & -1 \end{bmatrix} x = \begin{pmatrix} 1 \\ -1 \end{pmatrix} \\
& x \geq 0, \ x \in \mathbb{R}^5
\end{align*}
\]

If we sum the two constraints we get a *facial* constraint

\[
2x_2 + x_3 + 5x_4 = 0 \implies x \in \mathcal{F} = \{ x \in \mathbb{R}_+^5 : x_2 = x_3 = x_4 = 0 \}
\]

strict feasibility fails; problem can be reduced

\[
\begin{align*}
\text{min} \quad & (2 \quad 7)v \\
n\text{s.t.} \quad & (1 \quad 1)v = 1 \\
v \geq 0
\end{align*}
\]
First Example of Facial Reduction, cont...

Example (Facial Reduction in Linear Programming)

\[
\begin{align*}
\min & \quad (2 & 5 & -1 & 4 & 7) x \\
\text{s.t.} & \quad \begin{bmatrix} 1 & 1 & -1 & 3 & 1 \\
-1 & 1 & 2 & 2 & -1 \end{bmatrix} x = \begin{pmatrix} 1 \\
-1 \end{pmatrix} \\
x & \geq 0
\end{align*}
\]
Example (Facial Reduction in Linear Programming)

\[
\begin{align*}
\text{min} & \quad (2 & 5 & -1 & 4 & 7) x \\
\text{s.t.} & \quad \begin{bmatrix} 1 & 1 & -1 & 3 & 1 \\ -1 & 1 & 2 & 2 & -1 \end{bmatrix} x = \begin{pmatrix} 1 \\ -1 \end{pmatrix} \\
x & \geq 0
\end{align*}
\]

Find \( y \) with \( y^T b = 0, 0 \neq w = A^T y \geq 0 \) to get:

\[
y = (1 \ 1)^T, \ 0 \neq w^T = (A^T y)^T = (0 \ 2 \ 1 \ 5 \ 0) \geq 0.
\]

Then \( w \) is an exposing vector of the feasible set:
**Example (Facial Reduction in Linear Programming)**

\[
\begin{align*}
\text{min} & \quad (2 & 5 & -1 & 4 & 7) x \\
\text{s.t.} & \quad \begin{bmatrix}
1 & 1 & -1 & 3 & 1 \\
-1 & 1 & 2 & 2 & -1
\end{bmatrix} x = \begin{pmatrix} 1 \\ -1 \end{pmatrix} \\
x & \geq 0
\end{align*}
\]

Find \( y \) with \( y^T b = 0 \), \( 0 \neq w = A^T y \geq 0 \) to get:

\[
y = (1 \ 1)^T, \quad 0 \neq w^T = (A^T y)^T = (0 \ 2 \ 1 \ 5 \ 0) \geq 0.
\]

Then \( w \) is an exposing vector of the feasible set:

\[
w^T x = 0, \quad \forall \text{ feasible } x \implies x = \begin{bmatrix} 1 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_5 \end{bmatrix}; \quad x_2 = x_3 = x_4 = 0;
\]

(simplified) FR problem is

\[
\min \{(2 \ 7) v : (1 \ 1) v = 1, \ v \geq 0\}
\]
Faces of a Closed Convex Cone, ccc

Face of a ccc $\mathcal{K}$, $\mathcal{K} + \mathcal{K} \subseteq \mathcal{K}$, $\mathbb{R}\mathcal{K} \subseteq \mathcal{K}$

Let $\mathcal{K}$ be a ccc. A cone $F \subseteq \mathcal{K}$ is a face of $\mathcal{K}$, $F \triangleleft \mathcal{K}$, if

$$x, y \in \mathcal{K}, \ x + y \in F \implies x, y \in F,$$

If $\emptyset \neq F \subsetneq \mathcal{K}$, then it is called a proper face.
Faces of a Closed Convex Cone, ccc

Face of a ccc $\mathcal{K}$, $\mathcal{K} + \mathcal{K} \subseteq \mathcal{K}, \mathbb{R}\mathcal{K} \subseteq \mathcal{K}$

Let $\mathcal{K}$ be a ccc. A cone $F \subseteq \mathcal{K}$ is a face of $\mathcal{K}$, $F \trianglelefteq \mathcal{K}$, if

$$x, y \in \mathcal{K}, \quad x + y \in F \quad \Rightarrow \quad x, y \in F,$$

If $\emptyset \neq F \varsubsetneq \mathcal{K}$, then it is called a proper face.

Characterization of Faces of PSD Cone $\mathbb{S}_+^n$

Let $X \in \text{relint}(F), \ F \trianglelefteq \mathbb{S}_+^n$;

let $X = \begin{bmatrix} U & V \end{bmatrix} \begin{bmatrix} D & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} U \\ V \end{bmatrix} , \ D \in \mathbb{S}_+^k$

be the spectral decomposition.

two views are: $F = U\mathbb{S}_+^k U^T = \mathbb{S}_+^n \cap (VV^T)^\perp$
Some Useful Facts about Faces

- a face of a face is a face;
- an intersection of two faces is a face
- \( F_i \subseteq K, F_i = K \cap \phi_i^\perp, i = 1, \ldots, k \), implies

\[
\bigcap_i F_i = K \cap \left( \sum_i \phi_i \right)^\perp
\]

i.e., intersection exposed faces - exposed by sum of exposing vectors
Some Useful Facts about Faces

- a face of a face is a face;
- an intersection of two faces is a face
- $F_i \subset K, F_i = K \cap \phi_i^\perp, \ i = 1, \ldots, k$, implies

$$\cap_i F_i = K \cap (\sum_i \phi_i)^\perp$$

i.e., intersection exposed faces - exposed by sum of exposing vectors

For PSD cone

- Self-replicating: a face of a PSD cone is still a PSD cone;
- Facially exposed: every face of PSD cone has exposing vector
- Self-dual: $\mathcal{K} = \mathcal{K}^* = \{x : \langle x, y \rangle \geq 0, \forall y \in \mathcal{K}\}$
Recall (SDP-LRMC) Problem: Given \( z \in \mathbb{R}^\hat{E} \) a partial matrix, find the matrix \( Z \) of minimum rank to complete \( z \), i.e., \( \mathcal{P}_\hat{E}(Z) = \mathcal{P}_\bar{E}(Q) = z \),

**Minimize nuclear norm using SDP**

\[
\begin{align*}
\text{(SDP-LRMC)} & \quad \min \| Y \|_* = \frac{1}{2} \text{trace}(Y) \\
\text{ s.t. } & \quad \mathcal{P}_\bar{E}(Y) = z \\
& \quad Y \succeq 0,
\end{align*}
\]

where \( \bar{E} \) is the set of indices in \( Y \) that correspond to \( \hat{E} \), the known entries of the upper right block of \( \begin{bmatrix} 0 & Z \\ Z^T & 0 \end{bmatrix} \in S_{m+n}^+ \).
Recall (SDP-LRMC) Problem: Given $z \in \mathbb{R}^{\hat{E}}$ a partial matrix, find the matrix $Z$ of minimum rank to complete $z$, i.e., $\mathcal{P}_{\hat{E}}(Z) = \mathcal{P}_{\bar{E}}(Q) = z$.

Minimize nuclear norm using SDP

$$\min \| Y \|_* = \frac{1}{2} \text{trace}(Y)$$
$$\text{s.t. } \mathcal{P}_{\bar{E}}(Y) = z$$
$$Y \succeq 0,$$

where $\bar{E}$ is the set of indices in $Y$ that correspond to $\hat{E}$, the known entries of the upper right block of

$$\begin{bmatrix} 0 & Z \\ Z^T & 0 \end{bmatrix} \in \mathbb{S}^{m+n}_{++}.$$

• Since the diagonal is free, note that the Slater condition (strict feasibility) does hold for (SDP-LRMC). (And it holds for its dual.)
Facial Reduction of (SDP-LRMC) for Optimal Face

Bipartite Graph, $G_Z = (U_m, V_n, \hat{E})$

With $Z$ and the sampled elements we get a bipartite graph $G_Z$.

Find Fully Known Submatrix $X$ – a biclique $\alpha$, $X \cong z[\alpha] \in \mathbb{R}^{p \times q}$

After permutation of rows and columns, WLOG

$$Z = \begin{bmatrix} Z_1 & Z_2 \\ X & Z_3 \end{bmatrix}, \quad z = Z[\hat{E}], \quad \alpha \subseteq \hat{E}, \quad X \cong z[\alpha].$$
Facial Reduction of (SDP-LRMC) for Optimal Face

**Bipartite Graph,** \( G_Z = (U_m, V_n, \hat{E}) \)

With \( Z \) and the sampled elements we get a bipartite graph \( G_Z \).

**Find Fully Known Submatrix \( X \) – a biclique \( \alpha, X \cong z[\alpha] \in \mathbb{R}^{p \times q} \)**

After permutation of rows and columns, WLOG

\[
Z = \begin{bmatrix}
Z_1 & Z_2 \\
X & Z_3
\end{bmatrix}, \quad z = Z[\hat{E}], \quad \alpha \subseteq \hat{E}, \quad X \cong z[\alpha].
\]

Our algorithm is based on finding **bicliques** in \( G_Z \); we do this by finding (nontrivial/nondiagonal-block) cliques within symmetric matrix \( Y \).

\[
Y = \begin{bmatrix}
W_1 & Z \\
Z^T & W_2
\end{bmatrix}
\]
Bipartite Graph and Biclique

Partial matrix

\[
\begin{bmatrix}
-5 & NA & 10 & -20 & NA & -6 \\
4 & 0 & 4 & 4 & 6 & 6 \\
-3 & NA & NA & 32 & 27 & NA \\
5 & NA & 0 & 10 & 12 & NA \\
NA & -30 & NA & NA & 27 & NA \\
3 & -5 & -2 & 8 & NA & 4 \\
5 & 5 & NA & 0 & 3 & NA
\end{bmatrix}
\]

biclique indices: \( \bar{U}_m = \{6, 1, 2\} \), \( \bar{V}_n = \{1, 4, 3, 6\} \), \( \alpha = \{61, 64, 63, 66, 11, \ldots, 26\} \)

\[
z[\alpha] \equiv X = \begin{bmatrix}
3 & 8 & -2 & 4 \\
-5 & -20 & 10 & -6 \\
4 & 4 & 4 & 6
\end{bmatrix}
\]

\[
Y[\alpha] = \begin{bmatrix}
FREE & 3 & 8 & -2 & 4 \\
-5 & -20 & 10 & -6 \\
4 & 4 & 4 & 6
\end{bmatrix}
\]

\( \hat{E} = \{11, 13, 14, 16, 21, \ldots, 74, 75\} \)
Our View of Facial Reduction and Exposed Faces

Theorem (Drusvyatskiy, Pataki, W. ’15)

Linear transformation $\mathcal{M}: \mathbb{S}^n \rightarrow \mathbb{R}^m$, adjoint $\mathcal{M}^*$; feasible set $\mathcal{F} := \{ X \in \mathbb{S}^n_+ : \mathcal{M}(X) = b \} \neq \emptyset$, $b \in \mathbb{R}^m$. Then a vector $v$ exposes a proper face of $\mathcal{M}(\mathbb{S}^n_+)$ containing $b$ $\iff$ $v$ satisfies the auxiliary system

$$0 \neq \mathcal{M}^*v \in \mathbb{S}^n_+ \quad \text{and} \quad \langle v, b \rangle = 0.$$ 

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

1. $\mathbb{S}^n_+ \cap \mathcal{M}^{-1}N = \text{face}(\mathcal{F})$, the smallest face containing $\mathcal{F}$.
2. For any vector $v \in \mathbb{R}^m$ the following equivalence holds:

$$v \text{ exposes } N \iff \mathcal{M}^*v \text{ exposes } \text{face}(\mathcal{F})$$

Noisy sensor network localization: robust facial reduction and the Pareto frontier

D. Drusvyatskiy, N. Krislock, Y-L. Cheung Voronin, and H. W. ’16
Facial Reduction for (SDP-LRMC), \( r \) is target rank for \( Z \)

**Biclique** \( \alpha \cong \) of \( G_Z \), \( z[\alpha] \equiv X \in \mathbb{R}^{p \times q} \)

**target rank** \( r \leq \min\{p, q\} < \max\{p, q\} \);

WLOG

\[
Z = \begin{bmatrix}
Z_1 & Z_2 \\
X & Z_3
\end{bmatrix},
\]

SVD:

\[
X = \begin{bmatrix} U_1 & U_X \end{bmatrix} \begin{bmatrix}
\Sigma \in S^{r}_{++} & 0 \\
0 & 0
\end{bmatrix} \begin{bmatrix} V_1 & V_X \end{bmatrix}^T
\]

We get full rank factorization

\[
X = \bar{P} \bar{Q}^T = U_1 \Sigma V_1^T, \quad \bar{P} = U_1 \Sigma^{1/2}, \quad \bar{Q} = V_1 \Sigma^{1/2}.
\]

Since \( \text{rank} \) is lower semi-continuous: \( \text{rank} \ X = \text{rank} \ Z \) generically.

In fact our tests form:

\[
Z = PQ^T
\]

with \( P, Q \) random, i.i.d. and full column rank \( r \).
FR using Optimal $Y$  

**Rewrite Optimal $Y$**

Assuming we have obtained the desired target rank $Y = r$

$$0 \preceq Y = \begin{bmatrix} U \\ P \\ Q \\ V \end{bmatrix} D \begin{bmatrix} U \\ P \\ Q \\ V \end{bmatrix}^T = \begin{bmatrix} UDU^T & UDP^T & UDQ^T & UDV^T \\ PDU^T & PDP^T & PDQ^T & PDV^T \\ QDU^T & QDP^T & QDQ^T & QDV^T \\ VDU^T & VDP^T & VDQ^T & VDV^T \end{bmatrix}$$

**And assume rank $X = r$**

$$X = PDQ^T = \bar{P} \bar{Q}^T.$$  

This implies the ranges satisfy

$$U_1^T U_X = P^T U_X = 0, \ V_1^T V_X = Q^T V_X = 0$$

$$\text{range}(X) = \text{range}(P) = \text{range}(\bar{P}) = \text{range}(U_1),$$

$$\text{range}(X^T) = \text{range}(Q) = \text{range}(\bar{Q}) = \text{range}(V_1).$$
Constructing Exposing Vectors

Key for facial reduction

We can use an exposing vector formed as $U_X U_X^T$ for block $PDP^T$ as well as $V_X V_X^T$ for block $QDQ^T$ and add appropriate blocks of zeros:

$$W_X = \begin{bmatrix}
0 & 0 & 0 & 0 \\
0 & U_X U_X^T & 0 & 0 \\
0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0
\end{bmatrix} + \begin{bmatrix}
0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 \\
0 & 0 & V_X V_X^T & 0 \\
0 & 0 & 0 & 0
\end{bmatrix}.$$

All three matrices provide exposing vectors.

Facial reduction from exposing vector

$$F^* \leq T \mathbb{S}^{((n+m)-(p+q-2r))} T^T, \quad \text{range } T = \text{null } W_X.$$
Measuring Noise of Biclique $\alpha \in \Theta$

**Biclique:** $\alpha \subseteq \hat{E}$, $z[\alpha] \cong X \in \mathbb{R}^{p \times q}$, target rank $r$

- Singular values of $X$: $\sigma_1 \geq \ldots \geq \sigma_{\min\{p,q\}}$

- Biclique noise:
  - $u_X^P := \frac{\sum_{i=r+1}^{\min\{p,q\}} \sigma_i^2}{0.5p(p-1)}$
  - $u_X^Q := \frac{\sum_{i=r+1}^{\min\{p,q\}} \sigma_i^2}{0.5q(q-1)}$

**Assign biclique weight**

Total noise of all bicliques: $S = \sum_{X \in \Theta} (u_X^P + u_X^Q)$

For each $\alpha \in \Theta$:

- $w_X^P = 1 - \frac{u_X^P}{S}$,
- $w_X^Q = 1 - \frac{u_X^Q}{S}$
Facial Reduction Process

Find set of bicliques $\Theta$, of appropriate sizes

Find corresponding exposing vectors $\{ Y_{\text{expo}}^\alpha \}$

$\alpha \in \Theta$

Calculate their weights $\{ \omega^\alpha \}$

$\alpha \in \Theta$

Calculate the weighted sum of all the exposing vectors

$Y_{\text{expo}}^\text{Final} = \sum_{\alpha \in \Theta} \omega^\alpha Y_{\text{expo}}^\alpha$

Find full column rank $V$ such that range $V = \text{null} Y_{\text{expo}}^\text{Final}$.

Solve equivalent smaller problem based on smaller dimensional matrix $R$, where

$Y = VRV^T$

(Follows the framework in Drusvyatskiy/Krislock/Cheung-Voronin/W.)
Facial Reduction Process

- Find set of bicliques $\Theta$, of appropriate sizes

Mathematical expressions:

$$\text{Find set of bicliques } \Theta, \text{ of appropriate sizes}$$
Facial Reduction Process

- Find set of bicliques $\Theta$, of appropriate sizes
- Find corresponding exposing vectors $\{Y_{\alpha}^{\text{expo}}\}_{\alpha \in \Theta}$
- Calculate their weights $\{\omega_{\alpha}\}_{\alpha \in \Theta}$

Calculate the weighted sum of all exposing vectors

$$\text{Final} = \sum_{\alpha \in \Theta} \omega_{\alpha} Y_{\alpha}^{\text{expo}}$$

Find full column rank $V$ such that range $V = \text{null}(Y_{\text{expo}}^{\text{Final}})$.

Solve equivalent smaller problem based on smaller dimensional matrix $R$, where

$$Y = VRV^T$$

(Follows the framework in Drusvyatskiy/Krislock/Cheung-Voronin/W.)
Facial Reduction Process

- Find set of bicliques $\Theta$, of appropriate sizes
- Find corresponding exposing vectors $\{Y^\text{expo}_\alpha\}_{\alpha \in \Theta}$
  calculate their weights $\{\omega_\alpha\}_{\alpha \in \Theta}$
- Calculate the weighted sum of all the exposing vectors

$$Y^\text{expo}_{\text{Final}} = \sum_{\alpha \in \Theta} \omega_\alpha Y^\text{expo}_\alpha$$
Facial Reduction Process

- Find set of bicliques $\Theta$, of appropriate sizes
- Find corresponding exposing vectors $\{Y_\alpha^{\text{expo}}\}_{\alpha \in \Theta}$
- Calculate their weights $\{\omega_\alpha\}_{\alpha \in \Theta}$
- Calculate the weighted sum of all the exposing vectors
  \[ Y^{\text{expo}}_{\text{Final}} = \sum_{\alpha \in \Theta} \omega_\alpha Y_\alpha^{\text{expo}} \]
- Find full column rank $V$ such that range $V = \text{null} Y^{\text{expo}}_{\text{Final}}$. 
Facial Reduction Process

- Find set of bicliques $\Theta$, of appropriate sizes
- Find corresponding exposing vectors $\{Y_\alpha^{expo}\}_{\alpha \in \Theta}$
- Calculate their weights $\{\omega_\alpha\}_{\alpha \in \Theta}$
- Calculate the weighted sum of all the exposing vectors

$$Y_{Final}^{expo} = \sum_{\alpha \in \Theta} \omega_\alpha Y_\alpha^{expo}$$

- Find full column rank $V$ such that range $V = \text{null } Y_{Final}^{expo}$.
- Solve equivalent smaller problem based on smaller dimensional matrix $R$, where

$$Y = VRV^T$$
Facial Reduction Process

- Find set of bicliques $\Theta$, of appropriate sizes
- Find corresponding exposing vectors $\{Y^{\text{expo}}_\alpha\}_{\alpha \in \Theta}$
  calculate their weights $\{\omega_\alpha\}_{\alpha \in \Theta}$
- Calculate the weighted sum of all the exposing vectors
  \[ Y_{\text{Final}}^{\text{expo}} = \sum_{\alpha \in \Theta} \omega_\alpha Y^{\text{expo}}_\alpha \]
- Find full column rank $V$ such that range $V = \text{null} Y_{\text{Final}}^{\text{expo}}$.
- Solve equivalent smaller problem based on smaller dimensional matrix $R$, where
  \[ Y = VRV^T \]
- (Follows the framework in Drusvyatskiy/Krislock/Cheung-Voronin/W.)
Exploit block structure

\( Y_{\text{Final}}^{\text{expo}} \) has block structure so \( V \) has a block structure too:

\[
Y_{\text{Final}}^{\text{expo}} = \begin{bmatrix}
\sum_{X \in C} w_X^P W_X^P & 0 \\
0 & \sum_{X \in C} w_X^Q W_X^Q
\end{bmatrix}, \quad V = \begin{bmatrix}
V_P & 0 \\
0 & V_Q
\end{bmatrix}
\]
$Y_{\text{Final}}^{\text{expo}}$ has block structure so $V$ has a block structure too:

$$
Y_{\text{Final}}^{\text{expo}} = \begin{bmatrix}
\sum_{X \in C} w_X^P W_X^P & 0 \\
0 & \sum_{X \in C} w_X^Q W_X^Q
\end{bmatrix}, \quad V = \begin{bmatrix} V_P & 0 \\
0 & V_Q \end{bmatrix}
$$

allows a computational speed up for eigenvalue subproblems.
FR dramatically reduces dimension of now overdetermined problem:

\[
\begin{aligned}
\min \quad & \text{trace}(R) \\
\text{s.t.} \quad & \mathcal{P}_E(V_P R_{pq} V_Q^T) = \tilde{z} \\
& R = \begin{bmatrix} R_p & R_{pq} \\ R_{pq}^T & R_q \end{bmatrix} \succeq 0.
\end{aligned}
\]

(\(\text{trace}(VRV^T)\))
FR dramatically reduces dimension of now overdetermined problem:

\[
\begin{align*}
\min & \quad \text{trace}(R) \\
\text{s.t.} & \quad \mathcal{P}_{\tilde{E}}(V_P R_{pq} V_Q^T) = z \\
& \quad R = \begin{bmatrix} R_p & R_{pq} \\ R_{pq}^T & R_q \end{bmatrix} \succeq 0.
\end{align*}
\]

remove the redundant constraints

Use a compact QR to find well-conditioned full rank matrix representation. A simple semidefinite constrained least squares solution may be enough!

\[
\min_{R \in \mathbb{S}^r_{++}} \| \mathcal{P}_{\tilde{E}}(V_P R_{pq} V_Q^T) - \tilde{z} \|.
\]

(Here \(\tilde{E}, \tilde{z}\) denote the corresponding entries after removing redundant constraints. Often \(R\) found explicitly.)
Cannot simply remove redundant constraints; use random **sketch matrix** $A$ to reduce the number of constraints; first solve:

$$\delta_0 = \min_{R \in S_r^+} \left\| A \left( P_{\hat{E}}(V_P R_{pq} V_Q^T) - z \right) \right\|.$$  

and hopefully obtain the target rank!
Noisy Case

Cannot simply remove redundant constraints; use random sketch matrix $A$ to reduce the number of constraints; first solve:

$$\delta_0 = \min_{R \in \mathbb{S}_+^{r_y}} \left\| A \left( \mathcal{P}_{\hat{E}}(V_P R_{pq} V_Q^T) - z \right) \right\|.$$  

and hopefully obtain the target rank! Otherwise, we use a refinement step.
Refinement Step in the Noisy Case

We would like to reduce the rank after the previous step using a parametric approach:

\[
\begin{align*}
\min & \quad \text{trace}(R) \\
\text{s.t.} & \quad \left\| A \left( \mathcal{P}_E(V_P R_{pq} V_Q^T) - b \right) \right\| \leq \delta_0 \\
& \quad R \succeq 0.
\end{align*}
\]
Refinement Step in the Noisy Case

We would like to reduce the rank after the previous step using a parametric approach:

$$\begin{align*}
\min & \quad \text{trace}(R) \\
\text{s.t.} & \quad \| A \left( \mathcal{P}_{\hat{E}}(V_P R_{pq} V_Q^T) - b \right) \| \leq \delta_0 \\
& \quad R \succeq 0.
\end{align*}$$

To ensure the rank can be reduced, we flip the problem:

$$\varphi(\tau) := \min \quad \| A \left( \mathcal{P}_{\hat{E}}(V_P R_{pq} V_Q^T) - b \right) \| + \gamma \| R \|_F$$
$$\text{s.t.} \quad \text{trace}(R) \leq \tau$$
$$\quad R \succeq 0.$$
Sample Results \( \approx 3 \times 10^6 \text{ variables} \)

**Table:** noiseless: \( r = 8; \; m \times n \) size; density \( p \); mean 20 instances.

<table>
<thead>
<tr>
<th>Specifications</th>
<th>( r_v )</th>
<th>Rcvrd (%Z)</th>
<th>Time (s)</th>
<th>Rank</th>
<th>Residual (%Z)</th>
</tr>
</thead>
<tbody>
<tr>
<td>( m \times n )</td>
<td>mean(( p ))</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1000 3000 0.53</td>
<td>16.10</td>
<td>96.39</td>
<td>37.29</td>
<td>8.0</td>
<td>1.1072e-10</td>
</tr>
<tr>
<td>1000 3000 0.50</td>
<td>17.65</td>
<td>88.99</td>
<td>36.50</td>
<td>8.0</td>
<td>4.6569e-10</td>
</tr>
<tr>
<td>1000 3000 0.48</td>
<td>32.15</td>
<td>71.66</td>
<td>72.14</td>
<td>8.5</td>
<td>2.0413e-07</td>
</tr>
</tbody>
</table>

**Table:** noisy: \( r = 2; \; m \times n \) size; density \( p \); mean 20 instances.

<table>
<thead>
<tr>
<th>Specifications</th>
<th>Rcvd (%Z)</th>
<th>Time (s)</th>
<th>Rank</th>
<th>Residual (%Z)</th>
</tr>
</thead>
<tbody>
<tr>
<td>( m \times n )</td>
<td>% noise</td>
<td>( p )</td>
<td>initial refine</td>
<td>initial refine</td>
</tr>
<tr>
<td>1100 3000 0.50</td>
<td>0.33</td>
<td>100.00</td>
<td>33.72</td>
<td>48.53</td>
</tr>
<tr>
<td>1100 3000 1.00</td>
<td>0.33</td>
<td>100.00</td>
<td>33.67</td>
<td>49.09</td>
</tr>
<tr>
<td>1100 3000 2.00</td>
<td>0.33</td>
<td>100.00</td>
<td>34.13</td>
<td>48.84</td>
</tr>
<tr>
<td>1100 3000 3.00</td>
<td>0.33</td>
<td>100.00</td>
<td>36.34</td>
<td>92.73</td>
</tr>
<tr>
<td>1100 3000 4.00</td>
<td>0.33</td>
<td>100.00</td>
<td>51.45</td>
<td>186.28</td>
</tr>
</tbody>
</table>
Preprocessing

- Though strict feasibility holds *generically*, failure appears in many applications. Preprocessing based on structure can both *regularize* and simplify the problem.
  (New Survey FR: Drusvyatskiy and W. ’17 )

Exploit structure at optimum

For low-rank matrix completion the structure at the optimum can be exploited to apply FR on the optimal face even though strict feasibility holds. In many cases one gets an optimal solution without the need of any SDP solver.

To do: reduce density/more refinement; real life applications
Low-Rank Matrix Completion with Facial Reduction

Shimeng Huang\textsuperscript{1} \hspace{1cm} Henry Wolkowicz\textsuperscript{2}

\textsuperscript{1}Department of Statistics and Actuarial Science
University of Waterloo

\textsuperscript{2}Department of Combinatorics and Optimization
University of Waterloo

AT ECM2017, Thurs. June 1, 10:50AM


