The many faces of degeneracy in conic optimization

Henry Wolkowicz

Dept. Combinatorics and Optimization, University of Waterloo, Canada

Wed. Sept. 7, 2016, 16:00-16:30

COCA16 Continuous Optimization: Challenges and Applications
Celebrating Ronny Ben-Tal’s 70th Birthday
Technion, Haifa, Israel
** Motivation: Loss of Slater CQ/Facial reduction

- **Slater condition** – existence of a **strictly feasible solution** – is at the heart of convex optimization.

- **Without Slater:** first-order optimality conditions may fail; dual problem may yield little information; small perturbations may result in infeasibility; many software packages can behave poorly.

- A pronounced phenomenon: though Slater holds **generically**, **surprisingly** many models arising from relaxations of hard nonconvex problems show loss of strict feasibility, e.g., Matrix completions/compressive sensing, **sensor network localization, SNL, EDM, POP, Molecular Conformation, QAP, GP, strengthened Max-Cut**

- We concentrate on appl. of Semidef. Progr., SDP. We look at various reasons and how to take advantage using two views of **FACIAL REDUCTION, FR**

*Main Ref: (in progress)*

*“The many faces of degeneracy in conic optimization”, Drusvyatskiy, Wolkowicz ’16*
** Facial Reduction/Preprocessing for LP

Primal-Dual Pair: $A$ onto, $m \times n$, $\mathcal{P} = \{1, \ldots, n\}$

(LP-P) \hspace{1cm} \begin{align*}
\max \quad & b^\top y \\
\text{s.t.} \quad & A^\top y \leq c
\end{align*}

(LP-D) \hspace{1cm} \begin{align*}
\min \quad & c^\top x \\
\text{s.t.} \quad & Ax = b, \\
& x \geq 0.
\end{align*}

Slater’s CQ for (LP-D) / Theorem of alternative

Exactly One is True:

(I) \hspace{1cm} \exists \hat{x} \text{ s.t. } A\hat{x} = b, \hat{x} > 0 \quad (\hat{x} \in \text{ri } F, \text{ feas. set})

Slater point

(II) \hspace{1cm} 0 \neq z = A^\top y \geq 0, \quad b^\top y = 0 \quad (\langle z, F \rangle = 0)

exposing vector
Linear Programming Example, \( x \in \mathbb{R}^5 \)

\[
\begin{align*}
\min & \quad (2 \ 6 \ -1 \ -2 \ 7) x \\
\text{s.t.} & \quad \begin{bmatrix} 1 & 1 & 1 & 1 & 0 \\ 1 & -1 & -1 & 0 & 1 \end{bmatrix} x = \begin{bmatrix} 1 \\ -1 \end{bmatrix} \\
& \quad x \geq 0
\end{align*}
\]

Sum the two constraints (multiply by: \( y^T = (1 \ 1) \)):
get: \( 2x_1 + x_4 + x_5 = 0 \implies x_1 = x_4 = x_5 = 0 \)
i.e., equiv. simplified problem/smaller face/\( \text{fewer} \) constr.

\[
\begin{align*}
\min & \quad 6x_2 - x_3 \\
\text{s.t.} & \quad x_2 + x_3 = 1, x_2, x_3 \geq 0, \\
& \quad (x_1 = x_4 = x_5 = 0)
\end{align*}
\]
Slater’s CQ for (LP-P) / Theorem of alternative

\[ \exists \hat{y} \text{ s.t. } c - A^\top \hat{y} \succ 0, \quad \left( (c - A^\top \hat{y})_i > 0, \forall i \in P =: P^I \right) \]

iff

\[ Ad = 0, \quad c^\top d = 0, \quad d \geq 0 \implies d = 0 \quad (\star) \]

implicit equality constraints: \( i \in P^e \)

Find \( 0 \neq d^* \) to (\( \star \)) with max number of non-zeros
(exposes minimal face containing feasible slacks)

\[ d_i^* > 0 \implies (c - A^\top y)_i = 0, \forall y \in F^y \quad i \in P^e \]

(where \( F^y \) is primal feasible set)

\( k = 1!; \quad \text{we only need one step of FR for LP} \)

\( d^* \) here exposes the minimal face (of slacks)
Facial Reduction: $A^\top y \leq_f c$; minimal face $f \subseteq \mathbb{R}_+^n$

proper primal-dual pair; dual of dual is primal

| (LP_{reg-P}) | max $b^\top y$ s.t. $(A_l)^\top y \leq c'$ $(A_e)^\top y = c^e$ | (LP_{reg-D}) | min $(c')^\top x' + (c^e)^\top x^e$ s.t. $\begin{bmatrix} A_l & A_e \end{bmatrix} \begin{bmatrix} x' \\ x^e \end{bmatrix} = b$ $x' \geq 0$, $x^e$ free |

Generalized Slater CQ holds - And!

after deleting redundant equality constraints!

Mangasarian-Fromovitz CQ (MFCQ) holds

$\left( \exists \hat{y} : (A_l)^\top \hat{y} < c', (A_e)^\top \hat{y} = c^e \right)$  $(A_e)^\top$ is onto

MFCQ holds iff dual optimal set is compact

Numerical difficulties if MFCQ fails; in particular for interior point methods!  Modelling issue!
** Convex Programming

Ordinary convex programming, (OCP)

\[
\text{(OCP)} \quad \sup_y b^\top y \text{ subject to } g(y) \leq 0
\]

\(b \in \mathbb{R}^m; \ g(y) = (g_i(y)) \in \mathbb{R}^n, \ g_i : \mathbb{R}^m \to \mathbb{R} \text{ convex, } \forall i \in P\)

Slater’s CQ; strict feasibility

\(\exists \hat{y} \text{ s.t. } g_i(\hat{y}) < 0, \forall i \) (implies MFCQ)

Slater’s CQ fails \(\iff\) implicit equality constraints exist

\(P^e := \{i \in P : g(y) \leq 0 \implies g_i(y) = 0\} \neq \emptyset\)

Let \(P^l := P \setminus P^e\) and

\(g^l := (g_i)_{i \in P^l}, \quad g^e := (g_i)_{i \in P^e}\)
**Minimal face** \( f \)

\[
f = \{ z \in \mathbb{R}_+^m : z_i = 0, \forall i \in \mathcal{P}^e \} \subseteq \mathbb{R}_+^m
\]

**Regularize OCP**

(OCP) is equivalent to \( g(y) \leq_f 0 \)

\[
\begin{align*}
\text{(OCP\textsubscript{reg})} & \quad \sup \quad b^\top y \\
\text{s.t.} & \quad g^l(y) \leq 0 \\
& \quad y \in \mathcal{F}^e
\end{align*}
\]

where \( \mathcal{F}^e := \{ y : g^e(y) = 0 \} \).

Then \( \mathcal{F}^e = \{ y : g^e(y) \leq 0 \} \), so is a convex set!!

**Slater’s CQ holds for (OCP\textsubscript{reg})**

\[
\exists \hat{y} \in \mathcal{F}^e : g^l(\hat{y}) < 0
\]

(Ben-Israel, Ben-Tal, Zlobec: BBZ Conditions ’76-80)
Abstract convex program

\[(ACP) \quad \inf_x f(x) \text{ s.t. } g(x) \preceq_K 0, x \in \Omega\]

where:

- \( f : \mathbb{R}^n \to \mathbb{R} \) convex; \( g : \mathbb{R}^n \to \mathbb{R}^m \) is \( K \)-convex
- \( K \subset \mathbb{R}^m \) closed convex cone; \( \Omega \subset \mathbb{R}^n \) convex set
- \( a \preceq_K b \iff b - a \in K, \quad a \prec_K b \iff b - a \in \text{int} K \)
- \( g(\alpha x + (1 - \alpha y)) \preceq_K \alpha g(x) + (1 - \alpha)g(y), \quad \forall x, y \in \mathbb{R}^n, \forall \alpha \in [0, 1] \)

Slater’s CQ: \( \exists \hat{x} \in \Omega \) s.t. \( g(\hat{x}) \in -\text{int} K \quad (g(x) \prec_K 0) \)

- guarantees strong duality (zero duality gap AND dual attainment)
- (near) loss of strict feasibility, nearness to infeasibility, correlates with number of iterations & loss of accuracy
- Recall that Slater (M-F) is equivalent to a nonempty bounded dual optimal set.
Faces of Convex Sets - Useful for Charact. of Opt.

Face of $C$, $F \subseteq C$

- $F \subseteq C$ is a face of $C$ if $F$ contains any line segment in $C$ whose relative interior intersects $F$.
- A convex cone $F \subseteq K$ is a face of a convex cone $K$, $F \subseteq K$, if (simplified)
  \[ x, y \in K \text{ and } x + y \in F \implies x, y \in F \]

Polar (Dual) Cone/Conjugate Face

- polar cone $K^* := \{ \phi : \langle \phi, k \rangle \geq 0, \ \forall k \in K \}$
- If $F \subseteq K$, the conjugate face of $F$ is
  \[ F^c := F^\perp \cap K^* \subseteq K^* \]
Properties of Faces

**General case**

- A face of a face is a face
- Intersection of a face with a face is a face.
- Let $C \subseteq K$, then face$(C)$ denotes the **minimal face** (intersection of faces) containing $C$.

**F** $\triangleleft K$ is an **exposed face** if there exists $\phi \in K^*$ with

$$F = K \cap \phi^\perp$$

$F^c$ is always exposed by $x \in \text{ri } F$.

The SDP cone is **facially exposed**, all its faces are exposed. (In fact like $\mathbb{R}_+^n : S^n_+$ is a proper closed convex cone, self-dual and facially exposed.)

\[ \text{(ACP) } \inf_x f(x) \text{ s.t. } g(x) \preceq_K 0, x \in \Omega \]

\[ \text{(Borwein-W.'78-79 )} \]

\[ \text{(ACP}_R\text{) } \inf_x f(x) \text{ s.t. } g(x) \preceq_{K^f} 0, x \in \Omega \]

where: \( K^f \) is the minimal face

Like LP, it is simple if we use the minimal face \( K^f \). We get a proper primal-dual pair!!
Lemma (Facial Reduction (FR); find EXPOSING vector $\phi$)

Suppose $\bar{x}$ is feasible. Then the LHS system

$$\begin{cases} (\Omega - \bar{x})^* \cap \partial \langle \phi, g(\bar{x}) \rangle \neq \emptyset \\ \phi \in K^*, \quad \langle \phi, g(\bar{x}) \rangle = 0 \end{cases}$$

implies $K^f \subseteq \phi^\perp \cap K$,

where: $\partial$ is subgradient; $\langle \cdot \rangle$ is inner-product.

Generally more than one step is needed to find $K^f$

Restrict to smaller face $\phi^\perp \cap K$; repeat till Slater is obtained
SDP case/Replicating cone

- Let \( X \in S^n_+ \) with spectral decomposition,
\[
X = [P \quad Q] \begin{bmatrix} D_+ & 0 \\ 0 & 0 \end{bmatrix} [P \quad Q]^T, \quad D_+ \in S^{r}_{++} \quad (\text{rank } X = r)
\]
- Then
\[
\text{Range}(X) = \text{Range}(P), \quad \text{Null}(X) = \text{Range}(Q)
\]
\[
\text{face}(X) = P S^r_+ P^T = (QQ^T)^\perp \cap S^n_+.
\]
\[
(\text{Z} = QQ^T \text{ exposing vector/matrix for face.})
\]
- \[
\text{face}(X)^c = QS^{n-r}_+ Q^T
\]

Range/Nullspace representations

- \[
\text{face}(X) = \left\{ Y \in S^n_+ : \text{Range}(Y) \subseteq \text{Range}(X) \right\}
\]
- \[
\text{face}(X) = \left\{ Y \in S^n_+ : \text{Null}(Y) \supseteq \text{Null}(X) \right\}
\]
- \[
\text{ri face}(X) = \left\{ Y \in S^n_+ : \text{Range}(Y) = \text{Range}(X) \right\}
\]
Semidefinite Programming, SDP, $\mathcal{S}_+^n$

$K = \mathcal{S}_+^n = K^*$: nonpolyhedral, self-polar, facially exposed

(SDP-P) $v_P = \sup_{y \in \mathbb{R}^m} b^\top y$ s.t. $g(y) := \mathcal{A}^* y - c \preceq_{\mathcal{S}_+^n} 0$

(SDP-D) $v_D = \inf_{x \in \mathcal{S}_+^n} \langle c, x \rangle$ s.t. $Ax = b$, $x \succeq_{\mathcal{S}_+^n} 0$

where:

- PSD cone $\mathcal{S}_+^n \subset \mathcal{S}^n$ symm. matrices
- $c \in \mathcal{S}^n$, $b \in \mathbb{R}^m$
- $\mathcal{A}: \mathcal{S}^n \to \mathbb{R}^m$ is an onto linear map, with adjoint $\mathcal{A}^*$
- $Ax = (\text{trace } A_i x) = (\langle A_i, x \rangle) \in \mathbb{R}^m$, $A_i \in \mathcal{S}^n$
- $\mathcal{A}^* y = \sum_{i=1}^m A_i y_i \in \mathcal{S}^n$
Regularization Using Minimal Face

Borwein-W.’78-79, \( f_P = \text{face } \mathcal{F}_P \); min. face of feasible slacks

(SDP-P) is equivalent to the regularized

(SDP_{reg}-P) \( \nu_{RP} := \sup_y \{ \langle b, y \rangle : A^* y \preceq_{f_P} c \} \)

\( f_P \) is minimal face of primal feasible slacks
\[ \{ s \succeq 0 : s = c - A^* y \} \subseteq f_P \subseteq S^n_+ \]

Lagrangian dual of regularized problem satisfies strong duality:

(SDP_{reg}-D) \( \nu_{DRP} := \inf_x \{ \langle c, x \rangle : A x = b, \ x \succeq_{f_P} 0 \} \)
\( \nu_P = \nu_{RP} = \nu_{DRP} \) and \( \nu_{DRP} \) is attained.

Regularized PROPER primal-dual pair dual of dual is primal

If we take the dual of \( (SDP_{reg}-D) \) we recover the primal regularized problem \( (SDP_{reg}-P) \).
Assume feasibility: ∃ \tilde{x} s.t. \mathcal{A}\tilde{x} = b, \tilde{x} \succeq 0.

Exactly one of the following alternatives holds/is consistent:

(I) ∃ \hat{x} s.t. \mathcal{A}\hat{x} = b, \hat{x} > 0 \quad (Slater)

or

(II) 0 \neq z = \mathcal{A}^*y \succeq 0, \langle b, y \rangle = 0, \quad (**)

(II) finds exposing vector: 0 \neq z \succeq 0

z exposes a proper face containing all the dual feasible points

\mathcal{A}x = b, x \succeq 0 \implies zx = 0. \quad (equiv. \ trace \ zx = 0)
Regularization of Dual Using Minimal Face

Borwein-W.’78-79, \( f_D = \text{face } \mathcal{F}_D \); min. face of dual feasible set

(SDP-D) is equivalent to the regularized

(SDP\textsubscript{reg}-D) \( \nu_{RD} := \inf_x \{ \langle c, x \rangle : A x = b, x \succeq f_D 0 \} \)

\( f_D \) is minimal face of dual feasible set

\( \{ x \succeq 0 : A x = b, x \succeq 0 \} \subseteq f_D \subseteq S^n_+ \)

Lagrang. dual of regulariz. dual problem satisfies strong duality:

(SDP\textsubscript{reg}-DD) \( \nu_{DRD} := \sup_y \{ \langle b, y \rangle : A^* y \preceq f_D^* c \} \)

\( \nu_D = \nu_{RD} = \nu_{DRD} \) and \( \nu_{DRD} \) is attained.

regularized primal-dual pair

If we take the dual of (SDP\textsubscript{reg}-DD) we recover the dual regularized problem (SDP\textsubscript{reg}-P).
\[(\text{SDP}_D) \quad \min \{ \text{trace } CX \text{ s.t. } AX = b, X \in S^n_+ \} \]

**Step 1:** Let \(0 \neq Z \succeq 0\) be an exposing vector.
- add constraint \(\text{trace } ZX = 0\). (Equivalently \(ZX = 0\))
- from spectral decomposition of \(Z\), with Range \(P = \text{Null } Z\):
  - substitute: \(X = P_{S_+^t}P^T, \quad t_1 = \text{nullity}(Z)\)

We get the equivalent smaller problem
\[
(\text{SDP}_{D1}) \quad \min \text{ trace}(P^T CP)R \\
\text{s.t. } \text{trace}(P^T A_i P)R = b_i, i = 1, \ldots, m \\
R \in S_+^{t_1}
\]

Remove/delete redundant linear constraints; repeat Step 1.
minimum number of steps is called the singularity degree
(ref. Sturm below)
Lemma: Using exposing vectors

Let

\[ Z_i \geq 0, F_i = S^n_+ \cap Z_i^\perp, i = 1, \ldots, m. \]

Then

\[ \cap_{i=1}^m F_i = S^n_+ \cap \left( \sum_{i=1}^m Z_i \right)^\perp \]

intersection of faces is exposed by sum of exposing vectors
Singularity Degree $d$ - Minimal Number of FR Steps

**Sturm’s error bounds Theorem for SDP, 2000**
Given an affine subspace $\mathcal{V}$ of $S^n$, the pair $(\mathcal{V}, S^n_+)$ is $\frac{1}{2^d}$-Holder regular, $\gamma = \frac{1}{2^d}$, with displacement, where $d$ is the singularity degree of $(\mathcal{V}, S^n_+)$ with displacement.
( e.g., for intersecting sets, for all compact sets $U$ there exists a constant $c > 0$ such that
$$\text{dist}(x, \mathcal{V} \cap S^n_+) \leq c \left( \text{dist}^\gamma(x, \mathcal{V}) + \text{dist}^\gamma(x, S^n_+) \right), \quad \forall x \in U$$)

**Cgnce rate alternating directions (MAP) for SDP**
Theorem (Drusvyatskiy, Li, W. 2015) If the sequence $X_k, Y_k$ converges, $d > 0$, then the rate is $O \left( k^{-\frac{1}{2d+1}-2} \right)$
(If Slater holds then cgnce is R-linear.)

(Paper includes Empirical Confirmation)
Applications?

- preprocessing is essential in commercial LP software.
- Can we do facial reduction in general?
- Is it efficient/worthwhile?
- important applications?
  - relation to feasibility questions, e.g., for matrix completion
  - iterative methods? convergence rates? (DR, MAP)
Highly (implicit) degenerate/low-rank problem
- high (implicit) degeneracy translates to low rank solutions
- take advantage of degeneracy; fast, high accuracy solutions

SNL - a Fundamental Problem of Distance Geometry;
easy to describe - dates back to Grasssmann 1886

- $r$: embedding dimension
- $n$ ad hoc wireless sensors $p_1, \ldots, p_n \in \mathbb{R}^r$ to locate in $\mathbb{R}^r$;
- $m$ of the sensors $p_{n-m+1}, \ldots, p_n$ are anchors (positions known, using e.g. GPS)
- pairwise distances $D_{ij} = ||p_i - p_j||^2$, $ij \in E$, are known within radio range $R > 0$

$$P^\top = [p_1 \ldots p_n] = [X^\top \ A^\top] \in \mathbb{R}^{r \times n}$$
Sensor Localization Problem/Partial EDM

Sensors « and Anchors »

Initial position of points

# sensors n = 300, # anchors m = 9, radio range R = 1.2
Nearest, Weighted, SDP Approx. (relax/discard rank $B$)

- $\min_{B \succeq 0} \| H \circ (\mathcal{K}(B) - D) \|$
- $\text{rank } B = r$; $H_{ij} = \begin{cases} 
1/\sqrt{D_{ij}} & \text{if } ij \in E, \\
0 & \text{otherwise}
\end{cases}$
- with rank constraint: a non-convex, NP-hard program
- SDP relaxation is convex

**BUT**: expensive/low accuracy/implicitly highly degenerate

cliqués restrict ranks of feasible $B$
2.16 GHz Intel Core 2 Duo, 2 GB of RAM
Dimension \( r = 2 \)
Square region: \([0, 1] \times [0, 1]\)
\( m = 9 \) anchors
Using only Rigid Clique Union and Rigid Node Absorption
Error measure: Root Mean Square Deviation

\[
\text{RMSD} = \left( \frac{1}{n} \sum_{i=1}^{n} \left\| p_i - p_{i}^{\text{true}} \right\|^2 \right)^{1/2}
\]
Results - Large $n$ (SDP size $O(n^2)$)

$n$ # of Sensors Located

<table>
<thead>
<tr>
<th>$n$ # sensors \ $R$</th>
<th>0.07</th>
<th>0.06</th>
<th>0.05</th>
<th>0.04</th>
</tr>
</thead>
<tbody>
<tr>
<td>6000</td>
<td>6000</td>
<td>6000</td>
<td>6000</td>
<td>6000</td>
</tr>
<tr>
<td>10000</td>
<td>10000</td>
<td>10000</td>
<td>10000</td>
<td>10000</td>
</tr>
</tbody>
</table>

CPU Seconds

<table>
<thead>
<tr>
<th># sensors \ $R$</th>
<th>0.07</th>
<th>0.06</th>
<th>0.05</th>
<th>0.04</th>
</tr>
</thead>
<tbody>
<tr>
<td>2000</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>6000</td>
<td>5</td>
<td>5</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>10000</td>
<td>10</td>
<td>10</td>
<td>9</td>
<td>8</td>
</tr>
</tbody>
</table>

RMSD (over located sensors)

<table>
<thead>
<tr>
<th>$n$ # sensors \ $R$</th>
<th>0.07</th>
<th>0.06</th>
<th>0.05</th>
<th>0.04</th>
</tr>
</thead>
<tbody>
<tr>
<td>2000</td>
<td>4e−16</td>
<td>5e−16</td>
<td>6e−16</td>
<td>3e−16</td>
</tr>
<tr>
<td>6000</td>
<td>4e−16</td>
<td>4e−16</td>
<td>3e−16</td>
<td>3e−16</td>
</tr>
<tr>
<td>10000</td>
<td>3e−16</td>
<td>5e−16</td>
<td>4e−16</td>
<td>4e−16</td>
</tr>
</tbody>
</table>
Results - $N$ Huge SDPs Solved

<table>
<thead>
<tr>
<th># sensors</th>
<th># anchors</th>
<th>radio range</th>
<th>RMSD</th>
<th>Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>20000</td>
<td>9</td>
<td>.025</td>
<td>$5 \times 10^{-16}$</td>
<td>25s</td>
</tr>
<tr>
<td>40000</td>
<td>9</td>
<td>.02</td>
<td>$8 \times 10^{-16}$</td>
<td>1m 23s</td>
</tr>
<tr>
<td>60000</td>
<td>9</td>
<td>.015</td>
<td>$5 \times 10^{-16}$</td>
<td>3m 13s</td>
</tr>
<tr>
<td>100000</td>
<td>9</td>
<td>.01</td>
<td>$6 \times 10^{-16}$</td>
<td>9m 8s</td>
</tr>
</tbody>
</table>

Size of SDPs Solved: $N = \binom{n}{2}$ (# vrbls)

$\mathcal{E}_n(\text{density of } G ) = \pi R^2$; $M = \mathcal{E}_n(|E|) = \pi R^2 N$ (# constraints)

Size of SDP Problems:

$M = [3,078,915, 12,315,351, 27,709,309, 76,969,790]$

$N = 10^9 \ [0.2000, 0.8000, 1.8000, 5.0000]$
Thm D.P.W. ’15: \( M : \mathbb{E} \to \mathbb{Y} \), \( K \) proper convex cone

\( \emptyset \neq F = \{ X \in K : M(X) = b \} \). Then a vector \( v \) exposes a proper face of \( M(K) \) containing \( b \) if, and only if, \( v \) satisfies the auxiliary system

\[
0 \neq M^* v \in K^*, \quad \langle v, b \rangle = 0.
\]

Let \( N = \text{face}(b, M(K)) \) (smallest face containing \( b \)). Then:

- \( K \cap M^{-1}(N) = \text{face}(F, K) \)
- \( v \) exposes \( N \) IFF \( M^*(v) \) exposes \( \text{face}(F, K) \).

**Corollary**

If Slater’s condition fails, then \( d = 1 \) IFF the minimal face \( \text{face}(b, M(K)) \) is exposed.
Using exposing vectors
Successful numerics recently Drusvyatskiy/Krislock/Vronin/W. 2015.
For Low-Rank Matrix Completion, LRMC, (Huang-W.’16)

Intractable (nonconvex) minimum rank completion

Given partial $m \times n$ real matrix $Z \in \mathbb{R}^{m \times n}$.

(LRMC) \[
\begin{align*}
\text{min} & \quad \text{rank}(M) \\
\text{s.t.} & \quad \|M_{\hat{E}} - Z_{\hat{E}}\| \leq \delta,
\end{align*}
\]

$\hat{E}$ sampled indices; $Z_{\hat{E}} \in \mathbb{R}^{\hat{E}}$; $\delta > 0$ tuning parameter

Convex nuclear norm relaxation

\[
\begin{align*}
\text{min} & \quad \|M\|_* \\
\text{s.t.} & \quad \|M_{\hat{E}} - Z_{\hat{E}}\| \leq \delta,
\end{align*}
\]

where $\|M\|_* = \sum_{i} \sigma_i(M)$.
**SDP Equivalent to Nuclear Norm Minimization**

**Trace minimization**

\[
\begin{align*}
\min & \quad \| Y \|_* = \text{trace}(Y) \\
\text{s.t.} & \quad \| Y_{\bar{E}} - Q_{\bar{E}} \| \leq \delta \\
& \quad Y \in S_{+}^{m+n},
\end{align*}
\]

\[
Q = \begin{bmatrix}
0 & Z \\
Z^T & 0
\end{bmatrix} \in S_{+}^{m+n} \text{ and } \bar{E} \text{ indices in } Y \text{ corresponding to } \hat{E}
\]

**Noiseless case: strict feasibility trivially holds**

\[
Y_{\bar{E}} = Q_{\bar{E}}
\]

choose diagonal of \( Y \) sufficiently large, positive.
(strict feas. holds for dual as well)

**Why consider this here?**

It has been shown recently by Huang-W. that one can exploit the structure at the optimum and efficiently apply FR.
Associated Undirected Weighted Graph $G = (V, E, W)$

node set $V = \{1, \ldots, m, m+1, \ldots, m+n\}$ Let:

$E_{1,m} := \{ij \in V \times V : i < j \leq m\}$

$E_{m+1,m+n} := \{ij \in V \times V : m+1 \leq i < j \leq m+n\}$

edge set $E := \bar{E} \cup E_{1,m} \cup E_{m+1,m+n}$.

weights for all $ij \in E$

$w_{ij} := \begin{cases} Z_{i(j-m)}, & \forall ij \in \bar{E} \\ 0, & \text{otherwise.} \end{cases}$

Corresponding adjacency matrix $A$; cliques $C$

nontrivial cliques of interest (after row/col perms) corresp. to full (specified) submatrix $X$ in $Z$; $C = \{i_1, \ldots, i_k\}$ with cardinalities

$|C \cap \{1, \ldots, m\}| = p \neq 0, \quad |C \cap \{m+1, \ldots, m+n\}| = q \neq 0$. 
Clique - $X$; generically rank $r$ by lsc of rank

$$X \equiv \{Z_{i(j-m)} : ij \in C\}, \quad \text{specified } p \times q \text{ submatrix.}$$

let $\text{rank } X = r_X$. Wlog

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

full rank factorization $X = \bar{P} \bar{Q}^T$ using SVD

$$X = \bar{P} \bar{Q}^T = U_X \Sigma_X V_X^T, \ \Sigma_X \in S_{++}^{r_X}, \ \bar{P} = U_X \Sigma_X^{1/2}, \ \bar{Q} = V_X \Sigma_X^{1/2}.$$
\[ C_X = \{i, \ldots, m, m+1, \ldots, m+k\}, \quad r < \max\{p, q\}, \]

target rank \( r \).

(From HW) rewrite optimality conditions SDP as

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

Lemma (Basic FR)

Let \( r < \min\{p, q\} \) and \( X = PDQ^T = \bar{P}\bar{Q}^T \) as above. We find a pair of exposing vectors using

\[
\text{FR}(\bar{P}, \bar{Q}) : \bar{P}\bar{P}^T + \bar{U}\bar{U}^T \succ 0, \quad \bar{P}^T\bar{U} = 0,
\]

\[
\bar{Q}\bar{Q}^T + \bar{V}\bar{V}^T \succ 0, \quad \bar{Q}^T\bar{V} = 0.
\]
Numerics LRMC/average over 5 instances

Table: noiseless: \( r = 2; \ m \times n \ \text{size} \uparrow. \)

<table>
<thead>
<tr>
<th>Specifications</th>
<th>Time (s)</th>
<th>Rank</th>
<th>Residual (%Z)</th>
</tr>
</thead>
<tbody>
<tr>
<td>( m \times n \ \text{mean}(p) )</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>700 2000 0.30</td>
<td>9.00</td>
<td>2.0</td>
<td>4.4605e-14</td>
</tr>
<tr>
<td>1000 5000 0.30</td>
<td>28.76</td>
<td>2.0</td>
<td>3.0297e-13</td>
</tr>
<tr>
<td>1400 9000 0.30</td>
<td>77.59</td>
<td>2.0</td>
<td>7.8674e-14</td>
</tr>
<tr>
<td>1900 14000 0.30</td>
<td>192.14</td>
<td>2.0</td>
<td>6.7292e-14</td>
</tr>
<tr>
<td>2500 20000 0.30</td>
<td>727.99</td>
<td>2.0</td>
<td>4.2753e-10</td>
</tr>
</tbody>
</table>

Table: noiseless: \( r = 4; \ m \times n \ \text{size} \uparrow. \)

<table>
<thead>
<tr>
<th>Specifications</th>
<th>Time (s)</th>
<th>Rank</th>
<th>Residual (%Z)</th>
</tr>
</thead>
<tbody>
<tr>
<td>( m \times n \ \text{mean}(p) )</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>700 2000 0.36</td>
<td>12.80</td>
<td>4.0</td>
<td>1.5217e-12</td>
</tr>
<tr>
<td>1000 5000 0.36</td>
<td>49.66</td>
<td>4.0</td>
<td>1.0910e-12</td>
</tr>
<tr>
<td>1400 9000 0.36</td>
<td>131.53</td>
<td>4.0</td>
<td>6.0304e-13</td>
</tr>
<tr>
<td>1900 14000 0.36</td>
<td>291.22</td>
<td>4.0</td>
<td>3.4847e-11</td>
</tr>
<tr>
<td>2500 20000 0.36</td>
<td>798.70</td>
<td>4.0</td>
<td>7.2256e-08</td>
</tr>
</tbody>
</table>
Table: noiseless: $r = 3$; $m \times n$ size $\uparrow$; noise $\uparrow$; density $\downarrow$.

<table>
<thead>
<tr>
<th>Specifications</th>
<th>Time (s)</th>
<th>Rank</th>
<th>Residual (%$Z$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$m$ $n$ % noise $p$</td>
<td>initial total</td>
<td>initial refine</td>
<td>initial refine</td>
</tr>
<tr>
<td>700 1000 0.00 0.40</td>
<td>2.22 1.82</td>
<td>2.40 2.40</td>
<td>3.961e-14 3.961e-14</td>
</tr>
<tr>
<td>700 1000 0.01 0.40</td>
<td>4.16 8.79</td>
<td>3.20 3.20</td>
<td>9.242e-01 9.360e-01</td>
</tr>
<tr>
<td>700 1000 0.15 0.40</td>
<td>3.64 6.32</td>
<td>2.40 2.40</td>
<td>9.416e-01 9.517e-01</td>
</tr>
<tr>
<td>700 1000 0.30 0.40</td>
<td>3.46 7.09</td>
<td>8.40 8.40</td>
<td>9.862e-01 9.862e-01</td>
</tr>
<tr>
<td>700 1000 0.45 0.40</td>
<td>3.45 4.26</td>
<td>3.80 3.80</td>
<td>9.539e-01 9.539e-01</td>
</tr>
<tr>
<td>1500 2000 10.00 0.40</td>
<td>14.07 19.13</td>
<td>2.40 2.40</td>
<td>9.281e-01 9.360e-01</td>
</tr>
<tr>
<td>1600 2100 10.00 0.35</td>
<td>13.85 18.03</td>
<td>2.40 2.40</td>
<td>9.535e-01 9.535e-01</td>
</tr>
<tr>
<td>1700 2200 10.00 0.30</td>
<td>10.48 30.81</td>
<td>11.00 11.00</td>
<td>8.000e-01 8.000e-01</td>
</tr>
<tr>
<td>1800 2300 10.00 0.25</td>
<td>4.22 15.22</td>
<td>4.60 4.60</td>
<td>4.000e-01 4.000e-01</td>
</tr>
<tr>
<td>1900 2500 10.00 0.40</td>
<td>21.39 29.03</td>
<td>2.20 2.20</td>
<td>9.506e-01 9.546e-01</td>
</tr>
<tr>
<td>2000 2600 10.00 0.35</td>
<td>18.58 50.70</td>
<td>10.20 10.20</td>
<td>9.894e-01 9.894e-01</td>
</tr>
<tr>
<td>2100 2700 10.00 0.30</td>
<td>22.75 40.97</td>
<td>6.40 6.40</td>
<td>9.759e-01 9.759e-01</td>
</tr>
<tr>
<td>2200 2800 10.00 0.25</td>
<td>6.61 26.14</td>
<td>5.20 5.20</td>
<td>4.000e-01 4.000e-01</td>
</tr>
</tbody>
</table>
**Conclusion**

**Preprocessing**
- Though strict feasibility holds *generically*, failure appears in many applications. Loss of strict feasibility is directly related to ill-posedness and difficulty in numerical methods.
- Preprocessing based on structure can both *regularize* and simplify the problem. In many cases one gets an optimal solution without the need of any SDP solver.

**Exploit structure at optimum**
For low-rank matrix completion the structure at the optimum can be exploited to apply FR even though strict feasibility holds.


The many faces of degeneracy in conic optimization

Henry Wolkowicz

Dept. Combinatorics and Optimization, University of Waterloo, Canada

Wed. Sept. 7, 2016, 16:00-16:30

COCA16 Continuous Optimization: Challenges and Applications
Celebrating Ronny Ben-Tal’s 70th Birthday
Technion, Haifa, Israel