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, $m \times n$, $\mathcal{P} = \{1, \dots, n\}$

$$\text{(LP-P)} \qquad \begin{array}{ll} \max & b^\top y \\ \text{s.t.} & A^\top y \le c \end{array} \qquad\qquad \text{(LP-D)} \qquad \begin{array}{ll} \min & c^\top x \\ \text{s.t.} & Ax = b, \\ & x \ge 0. \end{array}$$

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

Exactly One is True:

(I) $\exists \hat{x}$ s.t. $A\hat{x} = b, \hat{x} > 0$ $\quad$ ($\hat{x} \in \operatorname{ri} F$, feas. set)
Slater point

(II) $0 \ne z = A^\top y \ge 0, \ b^\top y = 0$ $\quad$ $(\langle z, F \rangle = 0)$
exposing vector

Linear Programming Example, $x \in \mathbb{R}^5$

$$\min \quad \begin{pmatrix} 2 & 6 & -1 & -2 & 7 \end{pmatrix} x$$
$$\text{s.t.} \quad \begin{bmatrix} 1 & 1 & 1 & 1 & 0 \\ 1 & -1 & -1 & 0 & 1 \end{bmatrix} x = \begin{pmatrix} 1 \\ -1 \end{pmatrix}$$
$$x \geq 0$$

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/ fewer constr.

$\min 6x_2 - x_3$ s.t. $x_2 + x_3 = 1, x_2, x_3 \geq 0,$
$(x_1 = x_4 = x_5 = 0)$

Linear Programming, LP, $A^T y \leq c$

Slater's CQ for (LP-P) / Theorem of alternative

$\exists \hat{y}$ s.t. $c - A^\top \hat{y} > 0$, $\quad \left(\left(c - A^\top \hat{y}\right)_i > 0, \forall i \in \mathcal{P} =: \mathcal{P}^l \right)$

iff

$$Ad = 0,\ c^\top d = 0,\ d \geq 0 \implies d = 0 \qquad (*)$$

implicit equality constraints: $i \in \mathcal{P}^e$

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

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

(where $\mathcal{F}^y$ is primal feasible set)

$k = 1!$; we only need one step of FR for LP

d^* here exposes the minimal face (of slacks)

Make implicit-equalities explicit/ Regularizes LP

Facial Reduction: $A^\top y \leq_f c$; minimal face $f \trianglelefteq \mathbb{R}^n_+$

proper primal-dual pair; dual of dual is primal

$$\text{(LP}_{reg}\text{-P)} \qquad \begin{array}{ll} \max & b^\top y \\ \text{s.t.} & (A^I)^\top y \leq c^I \\ & (A^e)^\top y = c^e \end{array} \qquad \Bigg| \qquad \text{(LP}_{reg}\text{-D)} \qquad \begin{array}{ll} \min & (c^I)^\top x^I + (c^e)^\top x^e \\ \text{s.t.} & \begin{bmatrix} A^I & A^e \end{bmatrix} \begin{pmatrix} x^I \\ x^e \end{pmatrix} = b \\ & x^I \geq 0,\ x^e \text{ free} \end{array}$$

Generalized Slater CQ holds - And!

after deleting redundant equality constraints!

Mangasarian-Fromovitz CQ (MFCQ) holds

$$\left(\exists \hat{y} : \quad (A^I)^\top \hat{y} < c^I, \quad (A^e)^\top \hat{y} = c^e \right) \qquad (A^e)^\top \text{ is onto}$$

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)

$$\text{(OCP)} \qquad \sup_y \, b^\top y \text{ subject to } g(y) \leq 0$$

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

Slater's CQ; strict feasibility

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

Slater's CQ fails $\boxed{\iff}$ implicit equality constraints exist

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

Let $\mathcal{P}^l := \mathcal{P} \backslash \mathcal{P}^e$ and

$$g^l := (g_i)_{i \in \mathcal{P}^l}, \qquad g^e := (g_i)_{i \in \mathcal{P}^e}$$

implicit equalities to *equalities*/ Regularize OCP

Minimal face f

$$f = \{z \in \mathbb{R}^m_+ \ : \ z_i = 0, \forall i \in \mathcal{P}^e\} \trianglelefteq \mathbb{R}^m_+$$

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

$$(\mathrm{OCP_{reg}}) \qquad \begin{array}{ll} \sup & b^\top y \\ \text{s.t.} & g^l(y) \leq 0 \\ & y \in \mathcal{F}^e \end{array}$$

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

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

Slater's CQ holds for (OCP_{reg}) $\quad \boxed{\exists \hat{y} \in \mathcal{F}^e : g^l(\hat{y}) < 0}$

(Ben-Israel, Ben-Tal, Zlobec: BBZ Conditions '76-80)

* (FR full generality) Abstract convex program

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

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; $\Omega \subseteq \mathbb{R}^n$ convex set
  - $a \preceq_K b \iff b - a \in K$, $a \prec_K b \iff b - a \in \operatorname{int} K$
  - $g(\alpha x + (1-\alpha y)) \preceq_K \alpha g(x) + (1-\alpha) g(y)$, $\forall x, y \in \mathbb{R}^n, \forall \alpha \in [0,1]$

Slater's CQ: $\boxed{\exists \hat{x} \in \Omega}$ s.t. $g(\hat{x}) \in -\operatorname{int} K$ $(g(x) \prec_K 0)$

- guarantees strong duality
  (zero duality gap AND dual attainmment)
- (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, $\boxed{F \unlhd C}$

- $F \subseteq C$ 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, $F \unlhd K$, if (simplified)

$$\boxed{x, y \in K \text{ and } x + y \in F \implies x, y \in F}$$

Polar (Dual) Cone/Conjugate Face

- polar cone $K^* := \{\phi : \langle \phi, k \rangle \geq 0, \ \forall k \in K\}$
- If $F \unlhd K$, the conjugate face of F is

$$\boxed{F^c := F^\perp \cap K^* \unlhd K^*}$$

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 $\text{face}(C)$ denotes the minimal face (intersection of faces) containing C.

$F \trianglelefteq K$ 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 $\mathbb{R}^n_+$: $\mathcal{S}^n_+$ is a proper closed convex cone, self-dual and facially exposed.)

Regularize abstract convex program (full generality)

in memorium: Jonathan Borwein 20 May 1951 - 2 Aug 2016,

jonborwein.org

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

(Borwein-W.'78-79)

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

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!!

General Theorem of Alternative

Lemma (Facial Reduction (FR); find EXPOSING vector ϕ)

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

$$\left\{ \begin{array}{ll} (\Omega - \bar{x})^* \cap \partial\langle \phi, g(\bar{x})\rangle \neq \emptyset & \\ \phi \in K^*, & \langle \phi, g(\bar{x})\rangle = 0 \end{array} \right\} \quad \textit{implies} \quad K^f \subseteq \phi^\perp \cap K,$$

where: ∂ 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/Faces

SDP case/Replicating cone

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

Range/Nullspace representations

$$\textit{face}(X) = \{Y \in \mathcal{S}^n_+ : \text{Range}(Y) \subseteq \text{Range}(X)\}$$
$$\textit{face}(X) = \{Y \in \mathcal{S}^n_+ : \text{Null}(Y) \supseteq \text{Null}(X)\}$$
$$\text{ri}\, \textit{face}(X) = \{Y \in \mathcal{S}^n_+ : \text{Range}(Y) = \text{Range}(X)\}$$

Semidefinite Programming, SDP, $\mathcal{S}^n_+$

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

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

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

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 $\mathcal{A}^*$
- $\mathcal{A}x = (\operatorname{trace} A_i x) = (\langle A_i, x \rangle) \in \mathbb{R}^m, \quad A_i \in \mathcal{S}^n$
  $\mathcal{A}^* y = \sum_{i=1}^m A_i y_i \in \mathcal{S}^n$

Regularization Using Minimal Face

Borwein-W.'78-79 , $f_P = \text{face}\, \mathcal{F}_P^s$; min. face of feasible slacks

(SDP-P) is equivalent to the regularized

$$(\text{SDP}_{reg}\text{-P}) \qquad v_{RP} := \sup_y \{\langle b, y\rangle \ : \ \mathcal{A}^* y \preceq_{f_P} c\}$$

f_p is minimal face of primal feasible slacks

$\{s \succeq 0 : s = c - \mathcal{A}^* y\} \subseteq f_p \trianglelefteq \mathcal{S}_+^n$

Lagrangian dual of regularized problem satisfies strong duality:

$$(\text{SDP}_{reg}\text{-D}) \quad v_{DRP} \quad := \inf_x \{\langle c, x\rangle \ : \ \mathcal{A} x = b, \ x \succeq_{f_P^*} 0\}$$

$v_P = v_{RP} = v_{DRP}$ and v_{DRP} is attained.

regularized PROPER primal-dual pair | dual of dual is primal

If we take the dual of $(\text{SDP}_{reg}\text{-D})$ we recover the primal regularized problem $(\text{SDP}_{reg}\text{-P})$.

Slater's CQ/Theorem of Alternative for Dual

Assume feasibility: $\exists \tilde{x}$ s.t. $\mathcal{A}\tilde{x} = b, \tilde{x} \succeq 0$.

Exactly one of the following alternatives holds/is consistent:

$$(I) \quad \exists \hat{x} \text{ s.t. } \mathcal{A}\hat{x} = b, \hat{x} \succ 0 \qquad (\textit{Slater})$$

<u>or</u>

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

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

z exposes a proper face containing all the dual feasible points

$$\mathcal{A}x = b, x \succeq 0 \implies zx = 0. \quad (\text{equiv. trace } zx = 0)$$

Regularization of Dual Using Minimal Face

Borwein-W.'78-79 , $f_D = \text{face}\, \mathcal{F}_D^x$; min. face of dual feasible set

(SDP-D) is equivalent to the regularized

$$(\text{SDP}_{reg}\text{-D}) \qquad v_{RD} := \inf_x \{\langle c, x\rangle \;:\; \mathcal{A}x = b, x \succeq_{f_D} 0\}$$

f_D is mininimal face of dual feasible set

$\{x \succeq 0 : \mathcal{A}x = b, x \succeq 0\} \subseteq f_D \trianglelefteq \mathcal{S}_+^n$

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

$$(\text{SDP}_{reg}\text{-DD}) \qquad v_{DRD} \;:= \sup_y \{\langle b, y\rangle \;:\; \mathcal{A}^* y \preceq_{f_D^*} c\}$$

$v_D = v_{RD} = v_{DRD}$ and v_{DRD} is attained.

regularized primal-dual pair

If we take the dual of $(\text{SDP}_{reg}\text{-DD})$ we recover the dual regularized problem $(\text{SDP}_{reg}\text{-P})$.

View One for FR in SDP

$(SDP_D) \quad \min\{\text{trace } CX \text{ s.t. } \mathcal{A}X = b, X \in \mathcal{S}^n_+\}$

Step 1: Let $0 \neq Z \succeq 0$ be an exposing vector.

add constraint trace $ZX = 0$. (Equivalently $ZX = 0$)

from spectral decomposition of Z, with Range $P = \text{Null } Z$:

$$\text{substitute:} \quad X = P\mathbb{S}^{t_1}_+ P^T, \qquad t_1 = \text{nullity}(Z)$$

We get the equivalent smaller problem

$$(SDP_{D1}) \quad \begin{array}{ll} \min & \text{trace}(P^T C P)R \\ \text{s.t.} & \text{trace}(P^T A_i P)R = b_i, i = 1, \ldots, m \\ & R \in \mathbb{S}^{t_1}_+ \end{array}$$

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

View Two for FR in SDP

Lemma: Using exposing vectors

Let

$$Z_i \succeq 0, F_i = \mathcal{S}^n_+ \cap Z_i^\perp, i = 1, \ldots, m.$$

Then

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

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 $\mathcal{S}^n$, the pair $(\mathcal{V}, \mathcal{S}^n_+)$ is $\frac{1}{2^d}$-Holder regular, $\gamma = \frac{1}{2^d}$, with displacement, where d is the singularity degree of $(\mathcal{V}, \mathcal{S}^n_+)$ with displacement.
(e.g., for intersecting sets, for all compact sets U there exists a constant $c > 0$ such that
$\operatorname{dist}(x, \mathcal{V} \cap \mathcal{S}^n_+) \leq c\left(\operatorname{dist}^\gamma(x, \mathcal{V}) + \operatorname{dist}^\gamma(x, \mathcal{S}^n_+)\right), \quad \forall x \in U$)

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 $\mathcal{O}\left(k^{-\frac{1}{2^{d+1}-2}}\right)$
(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)

** FR - Motivation/Application; EDM, SNL

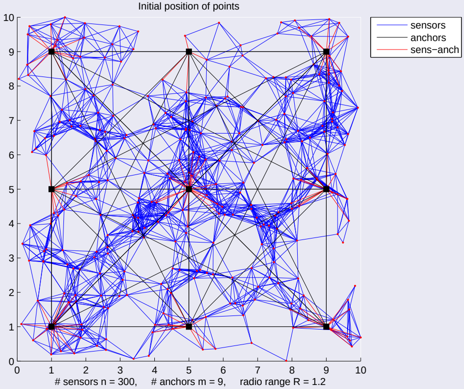
Highly (implicit) degenerate/low-rank problem

- 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 $R > 0$
- $$P^\top = \begin{bmatrix} p_1 & \ldots & p_n \end{bmatrix} = \begin{bmatrix} X^\top & A^\top \end{bmatrix} \in \mathbb{R}^{r \times n}$$

Sensors ○ and Anchors ■



Popular Techniques; SDP Relax.; Highly Degen.

Nearest, Weighted, SDP Approx. (relax/discard $\operatorname{rank} B$)

- $\min_{B \succeq 0} \| H \circ (\mathcal{K}(B) - D) \|$

  $\operatorname{rank} B = r; \qquad H_{ij} = \begin{cases} 1/\sqrt{D_{ij}} & \text{if } ij \in E, \\ H_{ij} = 0 & \text{otherwise} \end{cases}$
- with rank constraint: a non-convex, NP-hard program
- SDP relaxation is convex
  BUT: expensive/low accuracy/implicitly highly degenerate

cliques restrict ranks of feasible B

Random Noisless Problems, Krislock W. '**2010**

- 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

$$\text{RMSD} = \left(\frac{1}{n} \sum_{i=1}^{n} \|p_i - p_i^{\text{true}}\|^2 \right)^{1/2}$$

Results - Large n (SDP size $O(n^2)$)

n # of Sensors Located

n # sensors \ R	0.07	0.06	0.05	0.04
2000	2000	2000	1956	1374
6000	6000	6000	6000	6000
10000	10000	10000	10000	10000

CPU Seconds

# sensors \ R	0.07	0.06	0.05	0.04
2000	1	1	1	3
6000	5	5	4	4
10000	10	10	9	8

RMSD (over located sensors)

n # sensors \ R	0.07	0.06	0.05	0.04
2000	$4e{-}16$	$5e{-}16$	$6e{-}16$	$3e{-}16$
6000	$4e{-}16$	$4e{-}16$	$3e{-}16$	$3e{-}16$
10000	$3e{-}16$	$5e{-}16$	$4e{-}16$	$4e{-}16$

Results - N Huge SDPs Solved

Large-Scale Problems (results from 2010)

# sensors	# anchors	radio range	RMSD	Time
20000	9	.025	$5e{-16}$	25s
40000	9	.02	$8e{-16}$	1m 23s
60000	9	.015	$5e{-16}$	3m 13s
100000	9	.01	$6e{-16}$	9m 8s

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

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

Size of SDP Problems:

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

$N = 10^9 \, [0.2000 \quad 0.8000 \quad 1.8000 \quad 5.0000]$

View 2: Details with Exposing Vector/Numerics

Thm D.P.W. '15: $\mathcal{M} : \mathbb{E} \to \mathbb{Y}$, K proper convex cone

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

$$0 \neq \mathcal{M}^* v \in \mathcal{K}^*, \quad \langle v, b \rangle = 0.$$

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

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

Corollary

If Slater's condition fails, then $d = 1$ IFF the minimal $\text{face}(b, \mathcal{M}(K))$ is exposed.

Noisy Case

Using exposing vectors

Successful numerics recently Drusvyatskiy/Krislock/Vronin/W. 2015.

* FR for Low-Rank Matrix Completion, LRMC, (Huang-W.'16)

Intractable (nonconvex) minimum rank completion

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

$$(LRMC) \qquad \begin{array}{ll} \min & \operatorname{rank}(M) \\ \text{s.t.} & \|M_{\hat{E}} - Z_{\hat{E}}\| \leq \delta, \end{array}$$

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

convex nuclear norm relaxation

$$\begin{array}{ll} \min & \|M\|_* \\ \text{s.t.} & \|M_{\hat{E}} - Z_{\hat{E}}\| \leq \delta, \end{array}$$

where $\|M\|_* = \sum_i \sigma_i(M)$.

SDP Equivalent to Nuclear Norm Minimization

Trace minimization

$$\begin{array}{ll} \min & \|Y\|_* = \operatorname{trace}(Y) \\ \text{s.t.} & \|Y_{\bar{E}} - Q_{\bar{E}}\| \leq \delta \\ & Y \in \mathbb{S}_+^{m+n}, \end{array}$$

$Q = \begin{bmatrix} 0 & Z \\ Z^T & 0 \end{bmatrix} \in \mathbb{S}_+^{m+n}$ and $\bar{E}$ indices in Y 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 $V = \{1, \dots, m, m+1, \dots, m+n\}$ 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; $C = \{i_1, \dots, i_k\}$ with cardinalities

$$|C \cap \{1, \dots, m\}| = p \neq 0, \quad |C \cap \{m+1, \dots, m+n\}| = q \neq 0.$$

Exposing Vector for Low-Rank Completions

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 $\operatorname{rank} X = r_X$. Wlog

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

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

$$X = \bar{P}\bar{Q}^T = U_X \Sigma_X V_X^T, \, \Sigma_X \in \mathbb{S}_{++}^{r_X}, \quad \bar{P} = U_X \Sigma_X^{1/2}, \, \bar{Q} = V_X \Sigma_X^{1/2}.$$

$$C_X = \{i, \ldots, m, m+1, \ldots, m+k\}, \qquad r < \max\{p, q\},$$

target rank r.

(From HW) rewrite optimality conditions SDP as

$$0 \preceq Y = \begin{bmatrix} U \\ P \\ Q \\ V \end{bmatrix} D \begin{bmatrix} U \\ P \\ Q \\ V \end{bmatrix}^T = \left[\begin{array}{c|cc|c} UDU^T & UDP^T & UDQ^T & UDV^T \\ \hline PDU^T & PDP^T & PDQ^T & PDV^T \\ QDU^T & QDP^T & QDQ^T & QDV^T \\ \hline VDU^T & VDP^T & VDQ^T & VDV^T \end{array}\right].$$

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

$$\mathrm{FR}(\bar{P},\bar{Q}):\ \bar{P}\bar{P}^T + \boxed{\bar{U}\bar{U}^T} \succ 0,\ \bar{P}^T\bar{U} = 0,$$

$$\bar{Q}\bar{Q}^T + \boxed{\bar{V}\bar{V}^T} \succ 0,\ \bar{Q}^T\bar{V} = 0.$$

Numerics LRMC/average over 5 instances

Table: noiseless: $r = 2$; $m \times n$ size $\uparrow$.

Specifications			Time (s)	Rank	Residual (%Z)
m	n	mean(p)			
700	2000	0.30	9.00	2.0	4.4605e-14
1000	5000	0.30	28.76	2.0	3.0297e-13
1400	9000	0.30	77.59	2.0	7.8674e-14
1900	14000	0.30	192.14	2.0	6.7292e-14
2500	20000	0.30	727.99	2.0	4.2753e-10

Table: noiseless: $r = 4$; $m \times n$ size $\uparrow$.

Specifications			Time (s)	Rank	Residual (%Z)
m	n	mean(p)			
700	2000	0.36	12.80	4.0	1.5217e-12
1000	5000	0.36	49.66	4.0	1.0910e-12
1400	9000	0.36	131.53	4.0	6.0304e-13
1900	14000	0.36	291.22	4.0	3.4847e-11
2500	20000	0.36	798.70	4.0	7.2256e-08

Numerics LRMC/average over 5 instances

Table: noiseless: $r = 3$; $m \times n$ size $\uparrow$; noise $\uparrow$; density $\downarrow$.

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

** 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.

J.M. Borwein and H. Wolkowicz, *Characterization of optimality for the abstract convex program with finite-dimensional range*, J. Austral. Math. Soc. Ser. A **30** (1980/81), no. 4, 390–411. MR 83i:90156

Y-L. Cheung, S. Schurr, and H. Wolkowicz, *Preprocessing and regularization for degenerate semidefinite programs*, Computational and Analytical Mathematics, In Honor of Jonathan Borwein's 60th Birthday (D.H. Bailey, H.H. Bauschke, P. Borwein, F. Garvan, M. Thera, J. Vanderwerff, and H. Wolkowicz, eds.), Springer Proceedings in Mathematics & Statistics, vol. 50, Springer, 2013, pp. 225–276.

D. Drusvyatskiy, N. Krislock, Y-L. Cheung Voronin, and H. Wolkowicz, *Noisy sensor network localization: robust facial reduction and the Pareto frontier*, Tech. report, University of Waterloo, Waterloo, Ontario, 2014, arXiv:1410.6852, 20 pages.

D. Drusvyatskiy, G. Pataki, and H. Wolkowicz, *Coordinate shadows of semidefinite and Euclidean distance matrices*, SIAM J. Optim. **25** (2015), no. 2, 1160–1178. MR 3357643

D. Drusvyatskiy and H. Wolkowicz, *The many faces of degeneracy in conic optimization*, Tech. report, University of Waterloo, Waterloo, Ontario, 2016, in progress.

G.H. Golub and C.F. Van Loan, *Matrix computations*, 3^{nd} ed., Johns Hopkins University Press, Baltimore, Maryland, 1996.

B. Grone, C.R. Johnson, E. Marques de Sa, and H. Wolkowicz, *Positive definite completions of partial Hermitian matrices*, Linear Algebra Appl. **58** (1984), 109–124. MR 85d:05169

S. Huang, X. Ye, and H. Wolkowicz, *Low-rank matrix completion using nuclear norm with facial reduction*, Tech. report, University of Waterloo, Waterloo, Ontario, 2016, in progress.

N. Krislock and H. Wolkowicz, *Explicit sensor network localization using semidefinite representations and facial reductions*, SIAM Journal on Optimization **20** (2010), no. 5, 2679–2708.

G. Reid, F. Wang, H. Wolkowicz, and W. Wu, *Facial reduction and SDP methods for systems of polynomial equations*, Tech. report, University of Western Ontario, London, Ontario, 2014, submitted Dec. 2014, 38 pages.

R. Tyrrell Rockafellar, *Some convex programs whose duals are linearly constrained*, Nonlinear Programming (Proc. Sympos., Univ. of Wisconsin, Madison, Wis., 1970), Academic Press, New York, 1970, pp. 293–322.

Thanks for your attention!

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