The many faces of degeneracy in conic optimization

Henry Wolkowicz
Dept. Combinatorics and Optimization
University of Waterloo

Friday Mar. 18, 2016

at: Optimization and Learning: New Problems, New Challenges;
INFORMS Optimization Society 2016

(with: Dmitriy Drusvyatskiy, Univ. of Washington)

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 hard nonconvex problems show loss of strict feasibility, e.g., Matrix completions, SNL, EDM, POP, Molecular Conformation, QAP, GP, strengthened MC
- We look at various reasons and how to take advantage using FACIAL REDUCTION

  Refs: Borwein, W. '81; Cheung, Schurr, W.'11 Krislock, W.'10 , Drusvyatskiy, Pataki, W.'15 ; Cheung, Drusvyatskiy, Krislock, W.'14

Facial Reduction on LP: $F = \{x : Ax = b, x \geq 0\}$

Theorem of alternative, (with A full row rank)

Exactly one of the following is consistent:

(I) $\exists \hat{x}$ s.t. $A\hat{x} = b, \hat{x} > 0$

(II) $0 \neq z = A^\top y \geq 0, \; b^\top y = 0,$ $\quad (**) \quad$ (z exposes F)

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

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

Sum the two constraints (use $y^T = (1 \; 1)$ in (**)):

$2x_1 + x_4 + x_5 = 0 \implies x_1 = x_4 = x_5 = 0$

yields equivalent simplified problem:

$$\min \quad 6x_2 - x_3 \quad \text{s.t.} \quad x_2 + x_3 = 1, x_2, x_3 \geq 0$$

Facial Reduction on Primal, $A^T y \leq c$

Linear Programming Example, $y \in \mathbb{R}^2$

$$\max \quad \begin{pmatrix} 2 & 6 \end{pmatrix} y$$
$$\text{s.t.} \quad \begin{bmatrix} -1 & -1 \\ 1 & 1 \\ 1 & -1 \\ -2 & 2 \end{bmatrix} y \leq \begin{pmatrix} 1 \\ 2 \\ 1 \\ -2 \end{pmatrix},$$

active set $\{2, 3, 4\}$

$\begin{pmatrix} 3/2 \\ 1/2 \end{pmatrix}$ is optimal, $p^* = 6$

weighted last two rows $\begin{bmatrix} 1 & -1 & 1 \\ -2 & 2 & -2 \end{bmatrix}$ sum to zero:

set of implicit equalities: $\mathcal{P}^e := \{3, 4\}$

Facial reduction to 1 dim. after substit. for y

$$\begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \end{pmatrix} + t \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \quad \max \left\{ 2 + 8t : -1 \leq t \leq \tfrac{1}{2} \right\}, \quad t^* = \tfrac{1}{2}.$$

General Case?

- preprocessing is important in LP.
- Can we do facial reduction in general?
- Is it efficient/worthwhile?
- applications?

Background/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: $\exists \hat{x} \in \Omega$ s.t. $g(\hat{x}) \in -\operatorname{int} K$ $\quad (g(x) \prec_K 0)$

- guarantees strong duality
- (near) loss of strict feasibility, nearness to infeasibility, correlates with number of iterations & loss of accuracy

Faces of Cones - Useful for Charact. of Opt.

Face

A convex cone F is a face of convex cone K, denoted $F \trianglelefteq K$, if

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

Polar (Dual) Cone

$$K^* := \{\phi : \langle \phi, k \rangle \geq 0, \ \forall k \in K\}$$

Conjugate Face

If $F \trianglelefteq K$, the conjugate face of F is

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

If $x \in \mathrm{ri}(F)$, then $F^c = \{x\}^\perp \cap K^*$.

Recall: (ACP) $\inf_x f(x)$ s.t. $g(x) \preceq_K 0, x \in \Omega$

- polar cone: $K^* = \{\phi : \langle \phi, y \rangle \geq 0, \forall y \in K\}$.
- $K^f :=$ face(F) minimal face containing feasible set F.

Lemma (Facial Reduction; find EXPOSING vector ϕ)

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

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

Proof

line 1 of system implies $\bar{x}$ global min for convex function $\langle \phi, g(\cdot) \rangle$ on Ω; i.e., $0 = \langle \phi, g(\bar{x}) \rangle \leq \langle \phi, g(x) \rangle \leq 0, \forall x \in F$; implies $-g(F) \subseteq \phi^\perp \cap K$. □

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 $\mathcal{A}^*$

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

Slater's CQ/Theorem of Alternative

(Assume feasibility: $\exists\, \tilde{y}$ s.t. $c - \mathcal{A}^*\tilde{y} \succeq 0$.)
Exactly one of the following alternatives holds:

$$(I) \quad \exists\, \hat{y} \;\text{ s.t. } s = c - \mathcal{A}^*\hat{y} \succ 0 \qquad (\textit{Slater})$$

or

$$(II) \quad \mathcal{A}d = 0,\; \langle c, d\rangle = 0,\; 0 \neq d \succeq 0 \quad (*)$$

(d exposes a proper face containing all the feasible slacks $z = c - \mathcal{A}^*y \succeq 0$.)

Regularization Using Minimal Face

Borwein-W.'81 , $f_P = \text{face}\, \mathcal{F}_P^s$; 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 mininimal 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 primal-dual pair

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

SDP Regularization process

Alternative to Slater CQ

$$\mathcal{A}d = 0,\ \langle c, d\rangle = 0,\ 0 \neq d \succeq_{\mathcal{S}^n_+} 0 \qquad (*)$$

Determine a proper face $f_p \trianglelefteq f = Q\mathcal{S}^{\bar{n}}_+ Q^T \triangleleft \mathcal{S}^n_+$

- Let d solve $(*)$ with compact spectral decomosition $d = Pd_+P^\top$, $d_+ \succ 0$, and $[P\ Q] \in \mathbb{R}^{n\times n}$ orthogonal.
- Then d is an *exposing vector/matrix*

$$c - \mathcal{A}^*y \succeq_{\mathcal{S}^n_+} 0 \implies \langle c - \mathcal{A}^*y, d^*\rangle = 0$$

$$\implies \mathcal{F}^s_P \subseteq \mathcal{S}^n_+ \cap \{d^*\}^\perp = Q\mathcal{S}^{\bar{n}}_+ Q^\top \triangleleft \mathcal{S}^n_+$$

- (implicit rank reduction, $\bar{n} < n$)

Regularizing SDP

- at most $n - 1$ iterations to satisfy Slater's CQ.
- to check Theorem of Alternative

$$\mathcal{A}d = 0, \ \langle c, d \rangle = 0, \ 0 \neq d \succeq_{\mathcal{S}^n_+} 0, \qquad (*)$$

use stable auxiliary problem

$$(AP) \qquad \min_{\delta, d} \ \delta \ \text{ s.t. } \left\| \begin{bmatrix} \mathcal{A}d \\ \langle c, d \rangle \end{bmatrix} \right\|_2 \leq \delta, \quad \operatorname{trace}(d) = \sqrt{n}, \quad d \succeq 0.$$

- Both (AP) and its dual satisfy Slater's CQ.

Auxiliary Problem

$$(AP) \qquad \min_{\delta,d} \ \delta \ \text{ s.t. } \left\| \begin{bmatrix} \mathcal{A}d \\ \langle c, d\rangle \end{bmatrix} \right\|_2 \leq \delta, \quad \operatorname{trace}(d) = \sqrt{n}, d \succeq 0.$$

Both (AP) and its dual satisfy Slater's CQ ... but ...

Cheung-Schurr-W'11, a $k = 1$ step CQ

Strict complementarity holds for (AP)

iff

$k = 1$ steps are needed to regularize (SDP-P).

$k = 1$ always holds in LP case.

Singularity Degree - 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
$\text{dist}(x, \mathcal{V} \cap \mathcal{S}^n_+) \leq c\left(\text{dist}^\gamma(x, \mathcal{V}) + \text{dist}^\gamma(x, \mathcal{S}^n_+)\right), \quad \forall x \in U$)

Cgnce rate alternating directions 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.)

View of FR and Singularity Degree

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.

Applications of SDP where Slater's CQ fails

Instances SDP relaxations of NP-hard comb. opt.

- Quadratic Assignment (Zhao-Karish-Rendl-W.'96)
- Graph partitioning (W.-Zhao'99)
- Strengthened Max-Cut (Anjos-W'02)

Low rank problems

- Systems of polynomial equations (Reid-Wang-W.-Wu'15)
- Sensor network localization (SNL) problem (Krislock-W.'10 (Drusvyatskiy, Krislock, Veronin, W.'15)
- Molecular conformation (Burkowski-Cheung-W.'11)
- general SDP relaxation of low-rank matrix completion problems

Recent Application to QAP within ADMM Framework, D. Oliveira, Y. Xu, W'15

Quadratic Assignment Problem; "hardest" of NP-hard problems

$\min_{X\in\Pi} \operatorname{trace} AXBX^T + CX^T$; $\quad\Pi$ set of permutation matrices

SDP relaxation greatly simplifies after FR, facial reduction

FR: $Y = VRV^T$, $\quad Y \in \mathcal{S}_+^{n^2+1}$, $R \in \mathcal{S}_+^{(n-1)^2+1}$

$$\begin{array}{rl} \min_R & \langle L_Q, \hat{V}R\hat{V}^\top\rangle \\ \text{s.t.} & \mathcal{G}_J(\hat{V}R\hat{V}^\top) = E_{00} \\ & R \succeq 0, \end{array}$$

where L_Q linearizes the objective function; $\mathcal{G}_J$ is the gangster operator; E_{00} is the first unit matrix.

Implement ADMM (perfectly suited for FR)

$\min_{R,Y} \langle L_Q, Y\rangle$, s.t. $\mathcal{G}_J(Y) = E_{00}$, $Y = \hat{V}R\hat{V}^\top$, $R \succeq 0$.

augmented Lagrangian is

$$L_A := \langle L_Q, Y\rangle + \langle Z, Y - \hat{V}R\hat{V}^\top\rangle + \frac{\beta}{2}\|Y - \hat{V}R\hat{V}^\top\|_F^2.$$

alternating direction method of multipliers, ADMM

perform/repeat updates for (R_+, Y_+, Z_+)
('cheat' ... Eckert-Young for low rank psd)

$$R_+ = \text{argmin}_{R\succeq 0,\ \text{low rank}} L_A(R, Y, Z), \tag{1a}$$

$$Y_+ = \text{argmin}_{Y\in P} L_A(R_+, Y, Z), \tag{1b}$$

$$Z_+ = Z + \gamma \cdot \beta(Y_+ - \hat{V}R_+\hat{V}^\top), \tag{1c}$$

where P is the polyhedral constraints consisting of the gangster constraints and $0 \leq Y \leq 1$.

Sample Numerics: ADMM for SDP Relaxation of QAP

	1. opt value	2. Bundle [?] LowBnd	3. HKM-FR LowBnd	4. ADMM LowBnd	5. feas UpBnd	6. ADMM %gap	7. ADMM vs Bundle %Impr LowBnd	8 Tol5 cpusec HighRk	9 Tol5 cpusec LowRk	10 Tol12/5 cpuratio HighRk	11 HKM cpuratio Tol 9
Esc16a	68	59	50	64	72	11.76	7.35	2.30e+01	4.02	4.14	9.37
Esc16b	292	288	276	290	300	3.42	0.68	3.87e+00	4.55	2.15	8.08
Esc16c	160	142	132	154	188	21.25	7.50	1.09e+01	8.09	4.53	4.88
Esc16d	16	8	-12	13	18	31.25	31.25	2.14e+01	3.69	4.87	10.22
Esc16e	28	23	13	27	32	17.86	14.29	3.02e+01	4.29	4.80	8.79
Esc16g	26	20	11	25	28	11.54	19.23	4.24e+01	4.27	2.72	8.63
Esc16h	996	970	909	977	996	1.91	0.70	4.91e+00	3.53	2.33	10.60
Esc16i	14	9	-21	12	14	14.29	21.43	1.37e+02	4.30	2.39	8.76
Esc16j	8	7	-4	8	14	75.00	12.50	8.95e+01	4.80	3.83	7.93
Had12	1652	1643	1641	1652	1652	0.00	0.54	1.02e+01	1.08	1.06	5.91
Had14	2724	2715	2709	2724	2724	0.00	0.33	3.23e+01	1.69	1.19	10.46
Had16	3720	3699	3678	3720	3720	0.00	0.56	1.75e+02	3.15	1.04	12.51
Had18	5358	5317	5287	5358	5358	0.00	0.77	4.49e+02	6.00	2.22	13.28
Had20	6922	6885	6848	6922	6930	0.12	0.53	3.85e+02	12.15	4.20	14.53
Kra30a	149936	136059	-1111	143576	169708	17.43	5.01	5.88e+03	149.32	2.22	1111.11
Kra30b	91420	81156	-1111	87858	105740	19.56	7.33	4.36e+03	170.57	3.01	1111.11
Kra32	88700	79659	-1111	85775	103790	20.31	6.90	3.57e+03	200.26	4.28	1111.11
Nug12	578	557	530	568	632	11.07	1.90	2.60e+01	1.04	6.61	5.93
Nug14	1014	992	960	1011	1022	1.08	1.87	7.15e+01	1.87	5.06	8.43
Nug15	1150	1122	1071	1141	1306	14.35	1.65	9.10e+01	3.31	5.90	7.79
Nug16a	1610	1570	1528	1600	1610	0.62	1.86	1.81e+02	3.06	3.28	12.24
Nug16b	1240	1188	1139	1219	1356	11.05	2.50	9.35e+01	3.19	6.23	11.83
Nug17	1732	1669	1622	1708	1756	2.77	2.25	2.31e+02	4.34	3.63	13.13
Nug18	1930	1852	1802	1894	2160	13.78	2.18	4.16e+02	5.47	2.43	15.23
Nug20	2570	2451	2386	2507	2784	10.78	2.18	4.76e+02	11.56	3.75	14.35
Nug21	2438	2323	2386	2382	2706	13.29	2.42	1.41e+03	15.32	1.68	14.95
Nug22	3596	3440	3396	3529	3940	11.43	2.47	2.07e+03	21.82	1.39	13.90
Nug24	3488	3310	-1111	3402	3794	11.24	2.64	1.20e+03	29.64	3.29	1111.11
Nug25	3744	3535	-1111	3626	4060	11.59	2.43	3.12e+03	39.23	1.65	1111.11
Nug27	5234	4965	-1111	5130	5822	13.22	3.15	5.11e+03	78.18	1.58	1111.11
Nug28	5166	4901	-1111	5026	5730	13.63	2.42	4.11e+03	83.38	2.17	1111.11
Nug30	6124	5803	-1111	5950	6676	11.85	2.40	7.36e+03	133.38	1.76	1111.11
Rou12	235528	223680	221161	235528	235528	0.00	5.03	2.76e+01	0.93	0.98	6.90
Rou15	354210	333287	323235	350217	367782	4.96	4.78	3.12e+01	2.70	8.68	9.46
Rou20	725522	663833	642856	695181	765390	9.68	4.32	1.67e+02	10.31	10.90	16.08
Scr12	31410	29321	23973	31410	38806	23.55	6.65	4.40e+00	1.17	2.40	5.79
Scr15	51140	48836	42204	51140	58304	14.01	4.51	1.38e+01	2.41	1.84	10.75
Scr20	110030	94998	83302	106803	138474	28.78	10.73	1.53e+03	9.61	1.15	17.96
Tai12a	224416	222784	215637	224416	224416	0.00	0.73	1.79e+00	0.90	1.04	6.70
Tai15a	388214	364761	349586	377101	412760	9.19	3.18	2.74e+01	2.35	14.69	10.34
Tai17a	491812	451317	441294	476525	546366	14.20	5.13	6.50e+01	4.52	7.31	12.04
Tai20a	703482	637300	619092	671675	750450	11.20	4.89	1.28e+02	10.10	14.32	15.85
Tai25a	1167256	1041337	1096657	1096657	1271696	15.00	4.74	3.09e+02	38.48	5.58	1111.11
Tai30a	1818146	1652186	-1111	1706871	1942086	12.94	3.01	1.25e+03	142.55	10.51	1111.11
Tho30	88900	77647	-1111	86838	102760	17.91	10.34	2.83e+03	164.86	4.74	1111.11

Thanks for your attention!

The many faces of degeneracy in conic optimization

Henry Wolkowicz
Dept. Combinatorics and Optimization
University of Waterloo

Friday Mar. 18, 2016

at: Optimization and Learning: New Problems, New Challenges;
INFORMS Optimization Society 2016

(with: Dmitriy Drusvyatskiy, Univ. of Washington)