

Taking Advantage of Degeneracy in Cone Optimization: with Applications to Sensor Network Localization

Henry Wolkowicz

Dept. Combinatorics and Optimization, University of Waterloo



2015 JOINT MATHEMATICS MEETINGS

Largest Mathematics Meeting in the World

January 10 - 13 (Saturday - Tuesday), 2015 | [Henry B. Gonzalez Convention Center](#)

Motivation: Loss of Slater CQ/Facial reduction

- optimization algorithms rely on the KKT system;
and require that some constraint qualification (CQ) holds
(Slater's CQ/strict feasibility for convex conic optimization)
- However, surprisingly many conic opt, SDP relaxations,
instances arising from applications (QAP, GP, strengthened MC, SNL,
POP, Molecular Conformation)
do not satisfy Slater's CQ/are degenerate (ref. INFORMS meeting)
- lack of Slater's CQ results in: unbounded dual solutions;
theoretical and numerical difficulties,
in particular for *primal-dual interior-point methods*.
- solution:
 - theoretical *facial reduction* (Borwein, W:'81)
 - preprocess for regularized smaller problem (Cheung, Schurr, W:'11)
 - take advantage of degeneracy (for SNL)
(Krislock, W:'10; Krislock, Cheung, Drusvyatskiy, W:'14)

Outline: Regularization/Facial Reduction

- 1 Preprocessing/Regularization
 - Abstract convex program
 - LP case
 - CP case
 - Cone optimization/SDP case

- 2 Appl.: QAP, GP, Polyn Opt., SNL, Molecular conformation ...

Background/Abstract convex program

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

where:

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

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

- guarantees strong duality
- essential for efficiency/stability in p-d i-p methods
- ((near) loss of strict feasibility, **nearness to infeasibility** correlates with number of iterations & loss of accuracy)

Case of Linear Programming, LP

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

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

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

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

iff

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

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

expose minimal face containing feasible slacks; finding $0 \neq d^*$ to (*) with max number of non-zeros determines

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

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

Rewrite implicit-equalities to equalities/ Regularize LP

Facial Reduction: $A^T y \leq_f c$; minimal face $f \triangleq \mathbb{R}_+^n$

(LP_{reg-P})

$$\begin{array}{ll} \max & b^T y \\ \text{s.t.} & (A^<)^T y \leq c^< \\ & (A^=)^T y = c^= \end{array}$$

(LP_{reg-D})

$$\begin{array}{ll} \min & (c^<)^T x^< + (c^=)^T x^= \\ \text{s.t.} & [A^< \quad A^=] \begin{pmatrix} x^< \\ x^= \end{pmatrix} = b \\ & x^< \geq 0, x^= \text{ free} \end{array}$$

Mangasarian-Fromovitz CQ (MFCQ) holds

(after deleting redundant equality constraints!)

$$\left(\exists \hat{y} : \begin{array}{ll} \frac{i \in \mathcal{P}^<}{} & \frac{i \in \mathcal{P}^=}{} \\ (A^<)^T \hat{y} < c^< & (A^=)^T \hat{y} = c^= \end{array} \right) \quad (A^=)^T \text{ is onto}$$

MFCQ holds iff dual optimal set is compact

Numerical difficulties if MFCQ fails; in particular for interior point methods! Modelling issue?

Facial Reduction on Dual/Preprocessing

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

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

iff

$$z = A^T y \geq 0, b^T y = 0, \implies z = 0 \quad (**)$$

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

$$\min (2 \ 6 \ -1 \ -2 \ 7) x$$

$$\text{s.t. } \begin{bmatrix} 1 & 1 & 1 & 1 & 0 \\ 1 & -1 & -1 & 0 & 1 \end{bmatrix} x = \begin{pmatrix} 1 \\ -1 \end{pmatrix}, x \geq 0$$

Sum the two constraints ($y^T = (1 \ 1)$):

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

equiv. simplified problem/smaller face/less constr.

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

Case of ordinary convex programming, CP

$$(CP) \quad \sup_y b^\top y \text{ s.t. } g(y) \leq 0,$$

where

- $b \in \mathbb{R}^m$; $g(y) = (g_i(y)) \in \mathbb{R}^n$, $g_i : \mathbb{R}^m \rightarrow \mathbb{R}$ convex, $\forall i \in \mathbb{P}$
- Slater's CQ: $\exists \hat{y}$ s.t. $g_i(\hat{y}) < 0, \forall i$ (implies MFCQ)
- Slater's CQ fails implies implicit equality constraints exist,

i.e.:

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

Let $\mathcal{P}^< := \mathcal{P} \setminus \mathcal{P}^=$ and

$$g^< := (g_i)_{i \in \mathcal{P}^<}, \quad g^= := (g_i)_{i \in \mathcal{P}^=}$$

Rewrite implicit equalities to *equalities*/ Regularize CP

(CP) is equivalent to $g(y) \leq_f 0$, f is minimal face

$$\begin{array}{ll}
 (\text{CP}_{\text{reg}}) & \sup \quad b^\top y \\
 & \text{s.t.} \quad g^<(y) \leq 0 \\
 & \quad \quad y \in \mathcal{F}^= \quad \text{or } (g^=(y) = 0)
 \end{array}$$

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

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

Slater's CQ holds for (CP_{reg})

$$\exists \hat{y} \in \mathcal{F}^= : g^<(\hat{y}) < 0$$

modelling issue again?

Faces of Cones - Useful for Charact. of Opt.

Face

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

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

Conjugate Face

If $F \triangleleft K$, the **conjugate face** of F is

$$F^c := F^\perp \cap K^* \triangleleft K^*$$

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

where **polar cone**: $K^* = \{\phi : \langle \phi, y \rangle \geq 0, \forall y \in K\}$

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

- $K^f = \text{face}(F)$ minimal face containing feasible set F .

Lemma (Facial Reduction)

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

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

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

$K = \mathcal{S}_+^n = K^*$ nonpolyhedral cone!

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

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

where:

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

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 statements hold:

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

or

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

Φ exposes minimal face of slacks

If Φ is a non-zero max-rank solution to $(*\text{-SDP}(P))$, and $\Phi = PD_+P^T$ is compact spectral decomp., $[P \ Q]$ is orthogonal matrix, then

$$Z = C - \mathcal{A}^* y \succeq_{S_+^n} 0 \implies Z = QD_Z Q^T$$

so Φ determines/exposes a smaller face of S_+^n for feasible region/slacks.

Regularization Using Minimal Face

Borwein-W.'81 , $f_P = \text{face } \mathcal{F}_P^S$

(SDP-P) is equivalent to the **regularized**

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

f_P is minimal face of primal feasible slacks

slacks: $s = c - \mathcal{A}^* y \in f_P$

Lagrangian Dual DRP Satisfies Strong Duality:

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

$$= V_P = V_{RP}$$

and V_{DRP} is attained.

SDP Regularization process

Alternative to Slater CQ

$$\mathcal{A}d = 0, \langle c, d \rangle = 0, 0 \neq d \succeq_{S_+^n} 0 \quad (*)$$

Determine a proper face $f_p \trianglelefteq f = QS_+^{\bar{n}}Q^T \triangleleft S_+^n$

Let d solve (*) with compact spectral decomposition $d = Pd_+P^T$, $d_+ \succ 0$, and $[P \ Q] \in \mathbb{R}^{n \times n}$ orthogonal. Then

$$\begin{aligned} c - \mathcal{A}^*y \succeq_{S_+^n} 0 &\implies \langle c - \mathcal{A}^*y, d^* \rangle = 0 \\ &\implies \mathcal{F}_p^s \subseteq S_+^n \cap \{d^*\}^\perp = QS_+^{\bar{n}}Q^T \triangleleft S_+^n \end{aligned}$$

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

Conclusion

Part I

- **Minimal representations of the data regularize (P)**; use min. face f_P (and/or **implicit rank reduction**)
- goals:
 - exploit structure to find efficient exact facial reduction
 - a **backwards stable preprocessing algorithm** to handle (feasible) conic problems for which **Slater's CQ (almost) fails**

For Part II

- **Many instances of SDP** that arise from relaxations of hard combinatorial problems **fail the Slater CQ**.
- The **structure** of the problem can be **exploited** to allow for an explicit facial reduction of the problem. We obtain a **smaller (minimal) equivalent problem** for which the Slater CQ holds.

Part II: Applications of SDP where Slater's CQ fails

Instances of NP-hard combinatorial optimization problems

- Quadratic Assignment (Zhao-Karish-Rendl-W.'96)
- Graph partitioning (W.-Zhao'99)
- Min cut (Hao-Pong-Wang-W. '13)

Low rank problems

- Sensor network localization (SNL) problem (Krislock-W.'10, Drusvyatskiy-Cheung-Krislock-W'14)
- Molecular conformation (Burkowski-Cheung-W.'11)
- general SDP relaxation of low-rank matrix completion problem
- solving polynomial equations

SNL (K-W'10)

Highly (implicit) degenerate/low-rank problem

- SNL is a **graph realization** problem that can be modelled exactly as an SDP with rank restriction
- cliques in the graph correspond to principal submatrices that are singular; high (implicit) degeneracy translates to low rank solutions
- fast, high accuracy solutions for exact data

SNL - a Fundamental Problem of Distance Geometry

easy to describe - (dates back to Grassmann 1886)

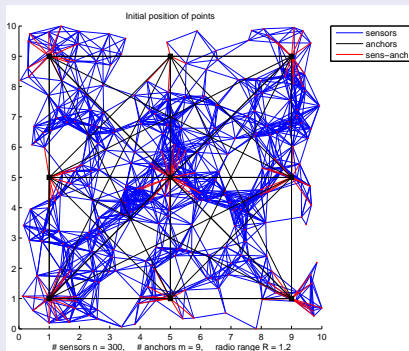
- r : embedding dimension
- n ad hoc wireless sensors $p_1, \dots, p_n \in \mathbb{R}^r$ to locate in \mathbb{R}^r ;
- m of the sensors p_{n-m+1}, \dots, p_n are anchors (positions known, using e.g. GPS)
- pairwise distances $D_{ij} = \|p_i - p_j\|^2, ij \in E$, are known within radio range $R > 0$



$$P^T = [p_1 \ \dots \ p_n] = [X^T \ A^T] \in \mathbb{R}^{r \times n}$$

Sensor Localization Problem/Partial EDM

Sensors \circ and Anchors \blacksquare



Underlying Graph Realization/Partial EDM

NP-Hard

Graph $\mathcal{G} = (\mathcal{V}, \mathcal{E}, \omega)$

- node set $\mathcal{V} = \{1, \dots, n\}$
- edge set $(i, j) \in \mathcal{E}$; $\omega_{ij} = \|p_i - p_j\|^2$ known approximately
- The anchors form a clique (complete subgraph)
- **Realization of \mathcal{G} in \mathbb{R}^r** : a mapping of nodes $v_i \mapsto p_i \in \mathbb{R}^r$ with squared distances given by ω .

Corresponding Partial Euclidean Distance Matrix, EDM

$$D_{ij} = \begin{cases} d_{ij}^2 & \text{if } (i, j) \in \mathcal{E} \\ 0 & \text{otherwise (unknown distance),} \end{cases}$$

$d_{ij}^2 = \omega_{ij}$ are known squared Euclidean distances between sensors p_i, p_j ; anchors correspond to a **clique**.

Connections to Semidefinite Programming (SDP)

$$D = \mathcal{K}(B) \in \mathcal{E}^n, B = \mathcal{K}^\dagger(D) \in \mathcal{S}^n \cap \mathcal{S}_C \text{ (centered } Be = 0)$$

$$P^\top = [p_1 \ p_2 \ \dots \ p_n] \in \mathcal{M}^{r \times n};$$

$$B := PP^\top \in \mathcal{S}_+^n \text{ (Gram matrix of inner products);}$$

$$\text{rank } B = r; \text{ let } D \in \mathcal{E}^n \text{ corresponding EDM; } e = (1 \ \dots \ 1)^\top$$

$$\begin{aligned}
 \text{(to } D \in \mathcal{E}^n) \quad D &= (\|p_i - p_j\|_2^2)_{i,j=1}^n \\
 &= (p_i^\top p_i + p_j^\top p_j - 2p_i^\top p_j)_{i,j=1}^n \\
 &= \boxed{\text{diag}(B) e^\top + e \text{diag}(B)^\top - 2B} \\
 &=: \mathcal{K}(B) \quad (\text{from } B \in \mathcal{S}_+^n).
 \end{aligned}$$

Popular Techniques; SDP Relax.; Highly Degen.

Nearest, Weighted, SDP Approx. (relax/discard rank B)

- $\min_{B \succeq 0} \|H \circ (\mathcal{K}(B) - D)\|$; rank $B = r$;
typical weights: $H_{ij} = 1/\sqrt{D_{ij}}$, if $ij \in E$, $H_{ij} = 0$ otherwise.
- 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 s)

Instead: (Shall) Take Advantage of Degeneracy!

clique α , $|\alpha| = k$ (corresp. $D[\alpha]$) with embed. dim. = $t \leq r < k$
 $\implies \text{rank } \mathcal{K}^\dagger(D[\alpha]) = t \leq r \implies \text{rank } B[\alpha] \leq \text{rank } \mathcal{K}^\dagger(D[\alpha]) + 1$
 $\implies \text{rank } B = \text{rank } \mathcal{K}^\dagger(D) \leq n - \boxed{(k - t - 1)} \implies$
 Slater's CQ (strict feasibility) **fails**

Facial Reduction for Cliques

- Using the basic theorem:
each clique corresponds to a
Gram matrix/corresponding subspace/corresponding face
of SDP cone (implicit rank reduction)
- In the case where
two cliques intersect, the union of the cliques correspond
to the (efficiently computable)
intersection of the corresponding faces/subspaces
- Finally, the
positions are determined using a
Procrustes problem

Results - Data for Random Noisless Problems

- 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

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

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

# 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:

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




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




Noisy SNL Case

200 Sensors; $[-0.5, 0.5]$ box; noise 0.05; radio range 0.1

- use **sum of exposing vectors** rather than **intersection of faces** obtained from cliques to do facial reduction
- use motivation: roundoff error cancels

show video

-  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
-  F. Burkowski, Y-L. Cheung, and H. Wolkowicz, *Efficient use of semidefinite programming for selection of rotamers in protein conformations*, INFORMS Journal on Computing **26** (2014), no. 4, 748–766.
-  Y-L. Cheung, D. Drusvyatskiy, N. Krislock, 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.

-  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.
-  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.
-  H. Wolkowicz and Q. Zhao, *Semidefinite programming relaxations for the graph partitioning problem*, Discrete Appl. Math. **96/97** (1999), 461–479, Selected for the special Editors' Choice, Edition 1999. MR 1 724 735



Q. Zhao, S.E. Karisch, F. Rendl, and H. Wolkowicz, *Semidefinite programming relaxations for the quadratic assignment problem*, J. Comb. Optim. **2** (1998), no. 1, 71–109, Semidefinite programming and interior-point approaches for combinatorial optimization problems (Fields Institute, Toronto, ON, 1996). MR 99f:90103

Thanks for your attention!

Taking Advantage of Degeneracy in Cone Optimization: with Applications to Sensor Network Localization

Henry Wolkowicz

Dept. Combinatorics and Optimization, University of Waterloo



2015 JOINT MATHEMATICS MEETINGS

Largest Mathematics Meeting in the World

January 10 - 13 (Saturday - Tuesday), 2015 | Henry B. Gonzalez Convention Center