Three Views of Facial Reduction in Cone Optimization

Henry Wolkowicz

Dept. Combinatorics and Optimization, University of Waterloo

Thurs. May 5, 2016, 2PM

at: Applied Mathematics Spring Lecture Series
University of Western Ontario

(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, sensor network localization, SNL, EDM, POP, Molecular Conformation, QAP, GP, strengthened MC

- We look at various reasons and how to take advantage using three views of FACIAL REDUCTION, FR

  Refs: Borwein, W. '79-81'; Cheung, Schurr, W.'11 Krislock, W.'10 , Drusvyatskiy, Pataki, W.'15 ; Cheung, Drusvyatskiy, Krislock, W.'14
  ("And some things that should not have been forgotten were lost. History became legend. Legend became myth ...")

Facial Reduction (FR) on LP: feasible set $F = \{x : Ax = b, x \in \mathbb{R}^n_+ \ (x \geq 0)\}$

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

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:

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

Facial Reduction (FR) on , $A^T y \le c$

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

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

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

opt: $y^* = \begin{pmatrix} 3/2 \\ 1/2 \end{pmatrix}, p^* = 6$

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

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

Facial reduction, FR, to 1 dim. after substit. for y

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

General Case?

- preprocessing is important in LP.
- Can we do facial reduction in general?
- Is it efficient/worthwhile?
- important applications? .... relation to feasibility questions and iterative methods? (DR, MAP?)

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. $g(\hat{x}) \in -\operatorname{int} K$ $(g(x) \prec_K 0)$

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

Faces of Convex Sets - Useful for Charact. of Opt.

Face of C, $\boxed{F \trianglelefteq C}$

- $F \subseteq C$ 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 convex cone K, $F \trianglelefteq K$, if
  $\boxed{x, y \in K \text{and } x + y \in F \implies x, y \in F}$

Polar (Dual) Cone/Conjugate Face

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

FACES

- 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.
- Let $X \in K = \mathcal{S}^n_+$, PSD cone; $X = [P\ Q] \begin{bmatrix} D_+ & 0 \\ 0 & 0 \end{bmatrix} [P\ Q]^T$ be spectral decomposition of $X, \operatorname{rank} X = r, D_+ \in \mathcal{S}^r_{++}$. Then

  $$\text{face}(X) = P\mathcal{S}^r_+ P^T = (QQ^T)^\perp \cap \mathcal{S}^n_+.$$

  $Z = QQ^T$ is an *exposing vector* for the face.

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

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

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

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

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

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 $\mathcal{A}^*$

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

Slater's CQ/Theorem of Alternative

From Lemma for FR

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

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

<u>or</u>

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

(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

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

Lagrangian dual of regularized problem satisfies strong duality:

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

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

regularized primal-dual pair

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

Aside for motivation for SDP: Dantzig Story in ’47-48

- 1947: visits Von Neumann (Princeton) and developes Simplex Method for LP
- 1948: Econometric society presents simplex method to well-known statisticians, mathematicians, economists.
  - Hotelling (*giant of a man*): stood up and stated: “But we all know the world is nonlinear”
  - Von Neumann came to his defence: “The speaker titled his talk ’Linear Programming’ and he carefully stated his axioms. If you have an application that satisfies the axioms, use it. If it does not, then don’t”
- SIAM in 70’s: 70% of world computer time spent on LP
- Dantzig later admitted: “The world is nonlinear ...” Using nonlinear models can be better than restricting to linear models.

Motivation: Quadratic models are 'better'

Binary problems modelled by quadratics

E.g.

- binary: $x \in \{\pm 1\} \iff x^2 = 1$
- binary: $x \in \{0, 1\} \iff x^2 - x = 0$
- permutation matrices: $X \in \Pi \iff X^T X = I, X \circ X - X = 0$

QQP: quadratic objective; quadratic constraints

in GROUND SPACE

$q_i(x) := x^T Q_i x + g_i^T x + \alpha_i,\ i = 0, 1, \ldots, m$

$$(\text{QQP}) \qquad \begin{array}{rll} p^* := & \min & q_0(x) \\ & \text{s.t.} & q_i(x) = 0, \quad i = 1, \ldots, m \\ & & x \in \mathbb{R}^n \end{array}$$

Quadratic models lead to SDP

Lagrangian dual/relaxation of QQP/homogenized Lagrangian

$q_i(x) := x^T Q_i x + g_i^T x + \alpha_i,\ i = 0, 1, \ldots, m \qquad x \in \mathbb{R}^n.$

$$
\begin{array}{rcll}
p^* \geq d^* & := & \max_y \min_x & L(x, y) := q_0(x) + \sum_{i=1}^m q_i(x) \\
 & = & \max_{y,t} \min_{x,x_0} & L(x, x_0, y, t) := \\
 & & & \quad q_0(x, x_0) + \sum_{i=1}^m q_i(x, x_0) \\
 & & & \qquad + t(1 - x_0^2)
\end{array}
$$

$$
\begin{array}{rcll}
d^* & = & \max_{y,t} & \sum_{i=1}^m y_i \alpha_i + t \\
 & & \text{s.t.} & Z = Z(y, t) = \nabla^2 L(x, x_0, y, t) \succeq 0 \\
 & & & Z \in \mathcal{S}_+^{n+1}
\end{array}
$$

SDP Relaxation

Lagrangian relaxation equivalent to

$$
\begin{array}{rcll}
d^* & = & \max & b^T y \\
 & & \text{s.t.} & Z = C - \mathcal{A}^* y \succeq 0 \\
 & & & Z \in \mathcal{S}^n_+
\end{array}
$$

Dual of dual yields SDP relaxation

$$
\begin{array}{rcll}
d^{**} \geq d^* & = & \min & \text{trace}\, CY \\
 & & \text{s.t.} & \mathcal{A}\, Y = b \\
 & & & X \in \mathcal{S}^n_+
\end{array}
$$

SDP Regularization process

Recall 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

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

- Both (AP) with e.g. $d = I, \delta >> 0$, and its dual satisfy Slater's CQ.

Auxiliary Problem

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

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.)

Explains SLOW convergence seen for MAP applied to molecular conformation?

Highly (implicit) degenerate/low-rank problem

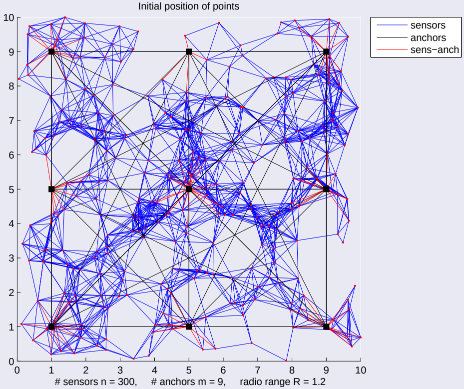
- high (implicit) degeneracy translates to low rank solutions
- take advantage of degeneracy; fast, high accuracy solutions

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

Sensor Localization Problem/Partial EDM

Sensors ○ and Anchors ■



Underlying Graph Realization/Partial EDM NP-Hard

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

- node set $\mathcal{V} = \{1, \ldots, n\}$
- 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)

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

$P^\top = \begin{bmatrix} p_1 & p_2 & \dots & p_n \end{bmatrix} \in \mathcal{M}^{r \times n}$;
$B := PP^\top \in \mathcal{S}^n_+$ (Gram matrix of inner products);
$\operatorname{rank} B = r$; let $D \in \mathcal{E}^n$ corresponding EDM; $e = \begin{pmatrix} 1 & \dots & 1 \end{pmatrix}^\top$

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

Moore-Penrose Generalized Inverse $\mathcal{K}^\dagger$

$$B \succeq 0 \quad \Longrightarrow \quad D = \mathcal{K}(B) = \mathrm{diag}\,(B)\, e^\top + e\, \mathrm{diag}\,(B)^\top - 2B \in \mathcal{E}$$

$$D \in \mathcal{E} \quad \Longrightarrow \quad B = \mathcal{K}^\dagger(D) = -\tfrac{1}{2} J\, \mathrm{offDiag}\,(D)\, J \succeq 0,\ Be = 0$$

$$(J = I - \tfrac{1}{n} ee^T)$$

Theorem (Schoenberg, 1935)

A (hollow) matrix D (with $\mathrm{diag}\,(D) = 0, D \in S_H$) is a
Euclidean distance matrix
if and only if

$$B = \mathcal{K}^\dagger(D) \succeq 0. \quad \text{(and centered } Be = 0)$$

And !!!!

$$\boxed{\mathrm{embdim}(D) = \mathrm{rank}\,\left(\mathcal{K}^\dagger(D)\right), \quad \forall D \in \mathcal{E}^n} \qquad (1)$$

Popular Techniques; SDP Relax.; Highly Degen.

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

- $\min_{B \succeq 0} \|H \circ (\mathcal{K}(B) - D)\|$; $\operatorname{rank} B = r$;
  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 Bs)

Instead: (Shall) Take Advantage of Degeneracy!

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

Basic Single Clique/Facial Reduction

Matrix with Fixed Principal Submatrix

For $Y \in \mathcal{S}^n$, $\alpha \subseteq \{1, \dots, n\}$: $Y[\alpha]$ denotes principal submatrix formed from rows & cols with indices α.

$\bar{D} \in \mathcal{E}^k$, $\alpha \subseteq 1:n$, $|\alpha| = k$

Define $\mathcal{E}^n(\alpha, \bar{D}) := \left\{ D \in \mathcal{E}^n : D[\alpha] = \bar{D} \right\}$. (completions)

Given $\bar{D}$; find a corresponding $B \succeq 0$; find the corresponding face; find the corresponding subspace.

if $\alpha = 1:k$; embedding dim $\operatorname{embdim}(\bar{D}) = t \le r$

$$D = \begin{bmatrix} \bar{D} & \cdot \\ \cdot & \cdot \end{bmatrix}.$$

BASIC THEOREM for Single Clique FR

Primal View

Let:

- $\bar{D} := D[1:k] \in \mathcal{E}^k$, $k < n$, $\text{embdim}(\bar{D}) = t \leq r$ be given;
- $B := \mathcal{K}^\dagger(\bar{D}) = \bar{U}_B S \bar{U}_B^\top$, $\bar{U}_B \in \mathcal{M}^{k \times t}$, $\bar{U}_B^\top \bar{U}_B = I_t$, $S \in \mathcal{S}_{++}^t$ be full rank orthogonal decomposition of Gram matrix;
- $U_B := \begin{bmatrix} \bar{U}_B & \frac{1}{\sqrt{k}} e \end{bmatrix} \in \mathcal{M}^{k \times (t+1)}$, $U := \begin{bmatrix} U_B & 0 \\ 0 & I_{n-k} \end{bmatrix}$, and $\begin{bmatrix} V & \frac{U^\top e}{\|U^\top e\|} \end{bmatrix} \in \mathcal{M}^{n-k+t+1}$ be orthogonal.

Then the minimal face:

- $$\begin{aligned} \text{face}\, \mathcal{K}^\dagger\left(\mathcal{E}^n(1:k, \bar{D})\right) &= \left(U \mathcal{S}_+^{n-k+t+1} U^\top\right) \cap \mathcal{S}_C \\ &= (UV) \mathcal{S}_+^{n-k+t} (UV)^\top \end{aligned}$$

The minimal face

Aside:

- $$\begin{aligned} \operatorname{face} \mathcal{K}^{\dagger}\left(\mathcal{E}^n(1:k,\bar{D})\right) &= \left(U\mathcal{S}_+^{n-k+t+1}U^\top\right) \cap \mathcal{S}_C \\ &= (UV)\mathcal{S}_+^{n-k+t}(UV)^\top \end{aligned}$$

Note that the minimal face is defined by the subspace $\mathcal{L} = \mathcal{R}(UV)$. We add $\frac{1}{\sqrt{k}}e$ to represent $\mathcal{N}(\mathcal{K})$; then we use V to eliminate e to recover a centered face.

Facial Reduction for Disjoint Cliques

Corollary from Basic Theorem

let $\alpha_1, \ldots, \alpha_\ell \subseteq 1{:}n$ pairwise disjoint sets, wlog:
$\alpha_i = (k_{i-1}+1){:}k_i$, $k_0 = 0$, $\alpha := \bigcup_{i=1}^{\ell} \alpha_i = 1{:}|\alpha|$ let
$\bar{U}_i \in \mathbb{R}^{|\alpha_i| \times (t_i+1)}$ with full column rank satisfy $e \in \mathcal{R}(\bar{U}_i)$ and

$$U_i := \begin{matrix} \\ k_{i-1} \\ |\alpha_i| \\ n-k_i \end{matrix} \begin{matrix} k_{i-1} & t_i+1 & n-k_i \\ \left[\begin{matrix} I \\ 0 \\ 0 \end{matrix}\right. & \begin{matrix} 0 \\ \bar{U}_i \\ 0 \end{matrix} & \left.\begin{matrix} 0 \\ 0 \\ I \end{matrix}\right] \end{matrix} \in \mathbb{R}^{n \times (n-|\alpha_i|+t_i+1)}$$

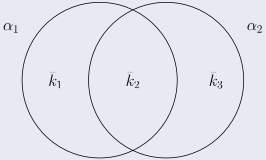
The minimal face is defined by $\mathcal{L} = \mathcal{R}(U)$:

$$U := \begin{matrix} \\ |\alpha_1| \\ \vdots \\ |\alpha_\ell| \\ n-|\alpha| \end{matrix} \begin{matrix} t_1+1 & \ldots & t_\ell+1 & n-|\alpha| \\ \left[\begin{matrix} \bar{U}_1 \\ \vdots \\ 0 \\ 0 \end{matrix}\right. & \begin{matrix} \ldots \\ \ddots \\ \ldots \\ \ldots \end{matrix} & \begin{matrix} 0 \\ \vdots \\ \bar{U}_\ell \\ 0 \end{matrix} & \left.\begin{matrix} 0 \\ \vdots \\ 0 \\ I \end{matrix}\right] \end{matrix} \in \mathbb{R}^{n \times (n-|\alpha|+t+1)},$$

where $t := \sum_{i=1}^{\ell} t_i + \ell - 1$. And $e \in \mathcal{R}(U)$.

Sets for Intersecting Cliques/Faces

$\alpha_1 := 1 : (\bar{k}_1 + \bar{k}_2); \quad \alpha_2 := (\bar{k}_1 + 1) : (\bar{k}_1 + \bar{k}_2 + \bar{k}_3)$



Let:

- $\alpha_1, \alpha_2 \subseteq 1:n; \quad k := |\alpha_1 \cup \alpha_2|$
- for $i = 1,2$: $\bar{D}_i := D[\alpha_i] \in \mathcal{E}^{k_i}$, embedding dimension t_i;
- $B_i := \mathcal{K}^\dagger(\bar{D}_i) = \bar{U}_i S_i \bar{U}_i^\top$, $\bar{U}_i \in \mathcal{M}^{k_i \times t_i}$, $\bar{U}_i^\top \bar{U}_i = I_{t_i}$, $S_i \in \mathcal{S}^{t_i}_{++}$;
- $U_i := \begin{bmatrix} \bar{U}_i & \frac{1}{\sqrt{k_i}} e \end{bmatrix} \in \mathcal{M}^{k_i \times (t_i+1)}$; and $\bar{U} \in \mathcal{M}^{k \times (t+1)}$ satisfies $\mathcal{R}(\bar{U}) = \mathcal{R}\left(\begin{bmatrix} U_1 & 0 \\ 0 & I_{\bar{k}_3} \end{bmatrix}\right) \cap \mathcal{R}\left(\begin{bmatrix} I_{\bar{k}_1} & 0 \\ 0 & U_2 \end{bmatrix}\right)$, with $\bar{U}^\top \bar{U} = I_{t+1}$
- $U := \begin{bmatrix} \bar{U} & 0 \\ 0 & I_{n-k} \end{bmatrix} \in \mathcal{M}^{n \times (n-k+t+1)}$ and $\begin{bmatrix} V & \frac{U^\top e}{\|U^\top e\|} \end{bmatrix} \in \mathcal{M}^{n-k+t+1}$ be orthogonal.

Then

$$\begin{array}{rcl} \underline{\underline{\bigcap_{i=1}^2 \text{face}\, \mathcal{K}^\dagger\left(\mathcal{E}^n(\alpha_i, \bar{D}_i)\right)}} & = & \left(U \mathcal{S}_+^{n-k+t+1} U^\top\right) \cap \mathcal{S}_C \\ & = & (UV)\mathcal{S}_+^{n-k+t}(UV)^\top \end{array}$$

Expense/Work of (Two) Clique/Facial Reductions

Subspace Intersection for Two Intersecting Cliques/Faces

Suppose:

$$U_1 = \begin{bmatrix} U_1' & 0 \\ U_1'' & 0 \\ 0 & I \end{bmatrix} \quad \text{and} \quad U_2 = \begin{bmatrix} I & 0 \\ 0 & U_2'' \\ 0 & U_2' \end{bmatrix}$$

Then:

$$U := \begin{bmatrix} U_1' \\ U_1'' \\ U_2'(U_2'')^\dagger U_1'' \end{bmatrix} \quad \text{or} \quad U := \begin{bmatrix} U_1'(U_1'')^\dagger U_2'' \\ U_2'' \\ U_2' \end{bmatrix}$$

($Q_1 =: (U_1'')^\dagger U_2'', Q_2 = (U_2'')^\dagger U_1''$ orthogonal/rotation)
(Efficiently) satisfies

$$\mathcal{R}(U) = \mathcal{R}(U_1) \cap \mathcal{R}(U_2)$$

Two (Intersecting) Clique Explicit Delayed Completion

Let:

- Hypotheses of intersecting Theorem (Thm 2) holds
- $\bar{D}_i := D[\alpha_i] \in \mathcal{E}^{k_i}$, for $i = 1, 2$, $\beta \subseteq \alpha_1 \cap \alpha_2, \gamma := \alpha_1 \cup \alpha_2$
- $\bar{D} := D[\beta]$ with embedding dimension r
- $B := \mathcal{K}^\dagger(\bar{D})$, $\quad \bar{U}_\beta := \bar{U}(\beta, :)$, where $\bar{U} \in \mathcal{M}^{k \times (t+1)}$ satisfies intersection equation of Thm 2
- $\begin{bmatrix} \bar{V} & \frac{\bar{U}^\top e}{\|\bar{U}^\top e\|} \end{bmatrix} \in \mathcal{M}^{t+1}$ be orthogonal.
- $\boxed{Z := (J\bar{U}_\beta \bar{V})^\dagger B((J\bar{U}_\beta \bar{V})^\dagger)^\top}$.

<u>THEN</u> $t = r$ in Thm 2, and $Z \in \mathcal{S}^r_+$ is the unique solution of the equation $(J\bar{U}_\beta \bar{V})Z(J\bar{U}_\beta \bar{V})^\top = B$, and the exact completion is

$$\boxed{D[\gamma] = \mathcal{K}\left(PP^\top\right)} \quad \text{where} \quad \boxed{P := UVZ^{\frac{1}{2}} \in \mathbb{R}^{|\gamma| \times r}}$$

Completing SNL (Delayed use of Anchor Locations)

Rotate to Align the Anchor Positions

- Given $P = \begin{bmatrix} P_1 \\ P_2 \end{bmatrix} \in \mathbb{R}^{n \times r}$ such that $D = \mathcal{K}(PP^\top)$
- Solve the orthogonal Procrustes problem:

$$\begin{array}{ll} \min & \|A - P_2 Q\| \\ \text{s.t.} & Q^\top Q = I \end{array}$$

$P_2^\top A = U \Sigma V^\top$ SVD decomposition; set $Q = UV^\top$;
(Golub/Van Loan'79-'12, Algorithm 12.4.1)

- Set $X := P_1 Q$

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

# 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

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

$\min_{X\in\Pi} \text{trace}\, AXBX^T + CX^T$; Π set of permutation matrices

SDP relaxation greatly simplifies after FR, facial reduction

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

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

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)

$$R_+ = \text{argmin}_{R \succeq 0,\ \text{low rank}} L_A(R, Y, Z), \quad (2a)$$

$$Y_+ = \text{argmin}_{Y \in P} L_A(R_+, Y, Z), \quad (2b)$$

$$Z_+ = Z + \gamma \cdot \beta (Y_+ - \hat{V} R_+ \hat{V}^\top), \quad (2c)$$

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

M.F. Anjos and H. Wolkowicz, *Strengthened semidefinite relaxations via a second lifting for the Max-Cut problem*, Discrete Appl. Math. **119** (2002), no. 1-2, 79–106, Foundations of heuristics in combinatorial optimization. MR MR1892412 (2003c:90084)

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, 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 semi-definite and Euclidean distance matrices*, Math. Programming **25** (2015), no. 2, 1160–1178, ArXiv:1405.2037.v1.

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!

Three Views of Facial Reduction in Cone Optimization

Henry Wolkowicz

Dept. Combinatorics and Optimization, University of Waterloo

Thurs. May 5, 2016, 2PM

at: Applied Mathematics Spring Lecture Series
University of Western Ontario

(with: Dmitriy Drusvyatskiy, Univ. of Washington)