Taking advantage of Degeneracy and Special Structure in Linear Cone Optimization

Yuen-Lam Cheung and Henry Wolkowicz

Dept. Combinatorics and Optimization, University of Waterloo

at: CanaDAM 2013, June 10-13,
Memorial University of Newfoundland

Motivation: Loss of Slater CQ/Facial reduction

- optimization algorithms rely on the KKT system;
  they require that some constraint qualification (CQ) holds
  (e.g. 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

- 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; for side chain positioning Burkowski, Cheung, W. '13)

Outline: Regularization/Facial Reduction

1. Motivation/Introduction
2. Preprocessing/Regularization
   - Abstract convex program
     - LP case
     - CP case
   - Cone optimization/SDP case
3. Applications: QAP, GP, SNL, Molecular conformation ...
   - Side Chain Positioning
   - Implementation
   - Numerics

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
  - $a \preceq_K b \iff b - a \in K$
  - $g(\alpha x + (1-\alpha y)) \preceq_K \alpha g(x) + (1-\alpha)g(y)$, $\forall x, y \in \mathbb{R}^n, \forall \alpha \in [0,1]$

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
- essential for efficiency/stability in primal-dual interior-point methods
  (((near) loss of strict feasibility correlates with number of iterations and loss of accuracy)

Case of Linear Programming, LP

Primal-Dual Pair: $A, m \times n$ / $\mathcal{P} = \{1, \ldots, n\}$ constr. matrix/set

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

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

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

iff

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

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

Finding solution $0 \neq d^*$ to $(*)$ with max number of non-zeros determines (where $\mathcal{F}^y$ is feasible set)

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

Rewrite implicit-equalities to equalities/ Regularize LP

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

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

Mangasarian-Fromovitz CQ (MFCQ) holds

(after deleting redundant equality constraints!)

$$\left(\exists \hat{y} : \quad \underbrace{(A^<)^\top \hat{y} < c^<}_{i \in \mathcal{P}^<} \quad \underbrace{(A^=)^\top \hat{y} = c^=}_{i \in \mathcal{P}^=} \right) \qquad (A^=)^\top \text{ is onto}$$

MFCQ holds iff dual optimal set is compact

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

Facial Reduction/Preprocessing

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

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

$\begin{pmatrix} 1 \\ 0 \end{pmatrix}$ feasible; weighted last two rows $\begin{bmatrix} 1 & -1 & 1 \\ -2 & 2 & -2 \end{bmatrix}$ sum to zero. $\mathcal{P}^< = \{1, 2\}, \mathcal{P}^= = \{3, 4\}$

Facial reduction; substit. for y; get 1 dim vrble; 2 dim slack

$$\begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \end{pmatrix} + t \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \begin{pmatrix} -1 \\ 1 \end{pmatrix} t \leq \begin{pmatrix} 1 \\ \frac{3}{2} \end{pmatrix}, t^* = -1, val^* = -6.$$

Case of ordinary convex programming, CP

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

where

- $b \in \mathbb{R}^m$; $g(y) = \big(g_i(y)\big) \in \mathbb{R}^n$, $g_i : \mathbb{R}^m \to \mathbb{R}$ 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 $\mathcal{P}^{<} := \mathcal{P} \backslash \mathcal{P}^{=}$ and

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

Rewrite implicit equalities to *equalities*/ Regularize CP

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

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

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

$\mathcal{F}^= = \{y : g^=(y) \le 0\}$, 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?

Faithfully convex case

Faithfully convex function f (Rockafellar'70)

f affine on a line segment only if affine on complete line containing the segment (e.g. analytic convex functions)

$\mathcal{F}^{=} = \{y : g^{=}(y) = 0\}$ is an affine set

Then:

$\mathcal{F}^{=} = \{y : Vy = V\hat{y}\}$ for some $\hat{y}$ and full-row-rank matrix V.

Then MFCQ holds for

$$(\text{CP}_{\text{reg}}) \qquad \begin{array}{lrcl} \sup & b^\top y & & \\ \text{s.t.} & g^{<}(y) & \leq & 0 \\ & Vy & = & V\hat{y} \end{array}$$

Semidefinite Programming, SDP

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

where $K^* := \{\phi : \langle \phi, x \rangle \geq 0, \forall x \in K\}$ dual/polar 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
- $c \in \mathcal{S}^n$, $b \in \mathbb{R}^m$
- $\mathcal{A} : \mathcal{S}^n \to \mathbb{R}^m$ is a linear map, with adjoint $\mathcal{A}^*$
  $\mathcal{A}x = (\text{trace}\, A_i x) \in \mathbb{R}^m$
  $\mathcal{A}^* y = \sum_{i=1}^m A_i y_i \in \mathcal{S}^n$

Slater's CQ/Theorem of Alternative

(Assume feasibility: $\exists \tilde{y}$ s.t. $c - \mathcal{A}^*\tilde{y} \succeq 0$.)

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

iff

$$\mathcal{A}d = 0,\ \langle c, d\rangle = 0,\ d \succeq 0 \implies d = 0 \qquad (*)$$

Faces of Cones - Useful for Charact. of Opt.

Face

A convex cone F is a face of K, denoted $F \trianglelefteq K$, if
$x, y \in K$ and $x + y \in F \implies x, y \in F$
($F \triangleleft K$ proper face)

Conjugate Face

If $F \trianglelefteq K$, the conjugate face (or complementary face) of F is
$F^c := F^\perp \cap K^* \trianglelefteq K^*$
If $x \in \operatorname{ri}(F)$, then $F^c = \{x\}^\perp \cap K^*$.

Minimal Faces

$f_P :=$ face $\mathcal{F}_P^s \trianglelefteq K$, $\quad \mathcal{F}_P^s$ is primal feasible set
$f_D :=$ face $\mathcal{F}_D^x \trianglelefteq K^*$, $\quad \mathcal{F}_D^x$ is dual feasible set
where: K^* denotes the dual (nonnegative polar) cone;
face S denotes the smallest face containing S.

Regularization Using Minimal Face

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

(SDP-P) is equivalent to the regularized

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

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

Lagrangian Dual DRP Satisfies Strong Duality:

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

and v_{DRP} is attained.

Conclusion Part I

- Minimal representations of the data regularize (P);
- Using the minimal face f_P regularizes SDPs.

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

Instances of SDP relaxations of NP-hard combinatorial optimization problems with fixed row and column sum and $0, 1$ constraints

- Quadratic Assignment (Zhao-Karish-Rendl-W.'96)
- Graph partitioning (W.-Zhao'99)

Low rank problems

- Sensor network localization (SNL) problem (Krislock-W.'10, Krislock-Rendl-W.'10) (SNL, highly (implicit) degenerate/low rank solutions)
- Molecular conformation (Burkowski-Cheung-W.'11)

Side Chain Positioning

- For our purposes, a protein macromolecule is a chain of amino acids, also called *residues*.
- For more tractable prediction, assume atoms in the backbone are fixed; then look for conformation of side chains for each residue.
- A further approximation inolves a discretization of possible side chain conformations that rely on rotamericity.

- Let $\mathcal{G} = (\mathcal{V}, \mathcal{E}, E)$ be a weighted, undirected graph with node set $\mathcal{V} = \bigcup_{i=1}^{p} \mathcal{V}_i$, where each subset $\mathcal{V}_i$ is a set consisting of *rotamers* for the i-th amino acid side chain/residue
- p is the number of residues; edge set $\mathcal{E}$ has weight (energy) E_{uv} associated with edge $uv \cong (u, v) \in \mathcal{E}$.

Integer Quadratic Program, (IQP)

$$
\text{(IQP)} \qquad
\begin{array}{rll}
val_{IQP} = & \min & \sum\limits_{(u,v)\in\mathcal{E}_n} E_{uv}x_u x_v \\
& \text{s.t.} & \sum\limits_{u\in\mathcal{V}_k} x_u = 1, \qquad \forall k = 1,\ldots,p \\
& & x_u \in \{0,1\}, \forall u \in \mathcal{V},
\end{array}
$$

where $x_u = \begin{cases} 1 & \text{if rotamer } u \text{ is chosen} \\ 0 & \text{otherwise} \end{cases}.$

Rewrite IQP as

$$
\text{(IQP)} \qquad
\begin{array}{rll}
val_{IQP} = & \min & x^T E x \\
& \text{s.t.} & Ax - \bar{e}_p = 0 \in \mathbb{R}^p \\
& & x = \begin{bmatrix} v_1^T & v_2^T & \cdots & v_p^T \end{bmatrix}^T \in \{0,1\}^{n_0} \\
& & v_k \in \{0,1\}^{m_k}, k = 1,\ldots,p.
\end{array}
$$

Quadratic, Quadratic Program, (QQP)

Redundant constraints within {}

$$(\text{QQP}) \quad \begin{array}{rll} val_{IQP} = val_{QQP} = & \min_x & x^T E x \\ & \text{s.t.} & \|\bar{e}_p - Ax\|^2 = 0 \\ & & x \circ x - x = 0 \\ & & \left\{ \begin{array}{l} \left(A^T A - I\right) \circ \left(xx^T\right) = 0 \\ (xx^T)_{ij} \geq 0, \ \forall (i,j) \in \mathcal{I}, \end{array} \right\} \end{array}$$

Recipe for SDP relaxation

- form the Lagrangian relaxation;
- apply homogenization;
- simplify to obtain the dual and an equivalent SDP;
- take the dual to obtain the SDP relaxation of the original IQP and remove any redundant (linearly dependent) constraints.

SCQ fails for SDP relaxation

Facially Reduced Primal-Dual Pair

$$\begin{array}{rl} \min\limits_{X \in \mathcal{S}^{n-p}} & \left\langle \hat{E}, X \right\rangle \\ \text{s.t.} & \operatorname{arrow}(X) = 0, \\ & {}^{d}\operatorname{bdiag}(X) = 0, \\ & X_{00} = 1, \\ & X \succeq 0, \end{array}$$

$$\begin{array}{rl} \max\limits_{t,w,\Lambda} & t \\ \text{s.t.} & {}^{1}\mathcal{O}(t) + \text{Arrow}(w) + {}^{\text{d}}\text{BDiag}(\Lambda) \preceq \hat{E}. \end{array}$$

Rounding to integral solution

Nearest feasible solution of IQP to $c \in \mathbb{R}^{n_0}$

$$\min_x \|x - c\| \text{ s.t. } Ax = \bar{e},\ x \in \{0,1\}^{n_0} \tag{1}$$

Obtaining IQP solution from SDP solution

- **Perron-Frobenis rounding**

  Let $u \in \mathbb{R}^n$ the principal eigvec. of Y^*, and $u' := \frac{p}{u_2+\ldots+u_n}\begin{pmatrix} u_2 \\ \vdots \\ u_n \end{pmatrix}$.

  $\Longrightarrow$ u' satisfies $Au' = \bar{e}$, and empirically $u' \in [0,1]^{n_0}$.

  $\Longrightarrow$ Take $c = u'$ and solve (1) for $\bar{u}'$.

- **Projection rounding**

  Let $\begin{pmatrix} 1 \\ u'' \end{pmatrix}$ be the diagonal of Y^*.

  $\Longrightarrow$ u'' satisfies $Au'' = \bar{e}$, $u'' \in [0,1]^{n_0}$.

  $\Longrightarrow$ Take $c = u''$ and solve (1) for $\bar{u}''$.

Adding nonnegativity constraints

- $Y_{ij} \geq 0$ is a valid constraint, $\forall (i,j)$, and tightens the SDP relaxation.
- But it is too expensive to enforce the constraint $Y \geq 0$ in the SDP relaxation.
- Use the cutting plane method:

repeat:
(1) solve SDP;
(2) add cutting planes (constraints $Y_{ij} \geq 0$).

How to choose cutting planes

- Cutting planes are not needed on diagonal blocks (which are diagonal).
- Some E_{ij} are very large $\implies$ Y_{ij} is likely to be negative.
- Rule: in each iter., choose (i,j) such that

(1) $Y_{ij} < 0$,
(2) $E_{ij}Y_{ij} << 0$ (i.e., $E_{ij} >> 0$).

Measuring the quality of rounded solutions

Metrics of IQP solution quality

Let x be a feasible solution of IQP. Then

$$x^T E x \geq val_{IQP} \geq d^*.$$

- The fraction $\frac{x^T Ex - val_{IQP}}{val_{IQP}}$ gives a measure of the quality of x.
- But val_{IQP} is not known.
- Use the *relative difference* instead:

$$\frac{x^T Ex - d^*}{\frac{1}{2}|x^T Ex + d^*|} \geq \frac{x^T Ex - val_{IQP}}{\frac{1}{2}|x^T Ex + val_{IQP}|}.$$

Computation results

Table: Results on medium-sized proteins

Protein	n_0	p	run time (min)		relative diff		relative gap	
			SCPCP	orig	SCPCP	orig	SCPCP	orig
1B9O	265	112	0.64	254.85	1.19E-11	2.14	1.45E-09	1.24
1C5E	200	71	2.59	70.63	4.93E-11	2.01	5.02E-09	1.00
1C9O	207	53	2.15	66.50	3.35E-12	2.00	2.77E-10	1.02
1CZP	237	83	1.90	143.95	8.30E-11	2.24	1.03E-08	1.00
1MFM	216	118	0.19	102.11	2.01E-11	2.00	1.24E-09	1.09
1QQ4	365	143	5.70	-	6.49E-11	-	2.27E-08	-
1QTN	302	134	5.04	-	2.24E-11	-	4.12E-09	-
1QU9	287	101	7.55	-	1.80E-11	-	5.52E-09	-

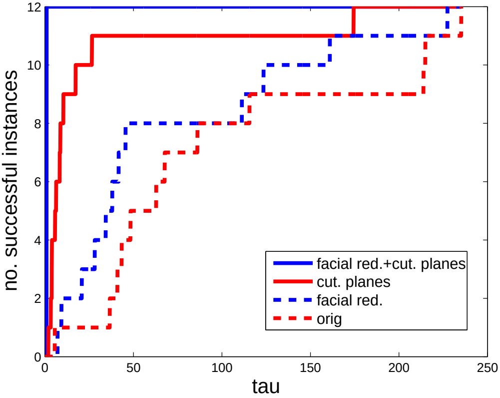
Computation results

Table: Results on large proteins (**SCPCP** only)

Protein	n_0	p	run time (hr)	rel. diff	rel. gap	numcut	# iter	Final # cuts
1CEX	435	146	0.08	1.26E-11	5.57E-09	40	9	485
1CZ9	615	111	3.96	2.98E-13	6.37E-10	60	25	1997
1QJ4	545	221	0.15	5.31E-12	1.14E-09	60	14	1027
1RCF	581	142	0.85	3.71E-12	1.15E-08	60	17	1305
2PTH	930	151	29.65	8.69E-09	7.63E-06	120	34	7247
5P21	464	144	0.31	1.39E-12	7.33E-10	40	16	822

Run times when using only facial red. or cutting planes

Figure: Performance profile for the use of facial reduction and cutting planes



Conclusion Part II

- SCQ fails for many SDP relaxations of hard combinatorial problems.
- facial reduction reduces size of problem and improves efficient/stability in particular when the structure is known.

Thanks for your attention!

Taking advantage of Degeneracy and Special Structure in Linear Cone Optimization

Yuen-Lam Cheung and Henry Wolkowicz

Dept. Combinatorics and Optimization, University of Waterloo

at: CanaDAM 2013, June 10-13,
Memorial University of Newfoundland